# Mechanistic study on the DNA methyltransferase DNMT3A

Von der Fakultät 4: Energie-, Verfahrens- und Biotechnik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung


Vorgelegt von

**Max Ulrich Emperle**

Geboren am 30.04.1982 in Stuttgart-Bad Cannstatt

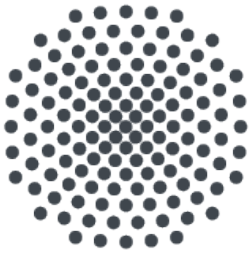


Hauptberichter: Prof. Dr. Albert Jeltsch

Mitberichterin: Prof. Dr. Monilola Olayioye

Prüfungsvorsitzender: Prof. Dr. Stephan Nussberger

Tag der mündlichen Prüfung: 25.07.2019


Institut für Biochemie und Technische Biochemie

Abteilung Biochemie

Universität Stuttgart


2019

## Eidesstattliche Erklärung

Ich versichere, dass ich die vorliegende Arbeit mit dem Titel
***"Mechanistic study on the DNA methyltransferase DNMT3A"***
selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe; aus fremden Quellen entnommene Passagen und Gedanken sind als solche kenntlich gemacht.

## Declaration of Authorship

I hereby certify that the dissertation entitled
***„Mechanistic study on the DNA methyltransferase DNMT3A"***
is entirely my own work except where otherwise indicated. Passages and ideas from other sources have been clearly indicated.

Name/Name: Max Ulrich Emperle

Unterschrift/Signed: 

Datum/Date: 14.02.2020

# Table of Contents

# Acknowledgements

I am grateful to everybody who supported me during my PhD. In particular I am very thankful to my mentor Prof. Dr. Albert Jeltsch for his excellent guidance, intellectual support and invaluable influence on my personal and scientific development.
I am also thankful to Dr. Renata Jurkowska for the great supervision during the first half of my PhD.

I also would like to thank Prof. Dr. Moniola Olayioye for being the 2. proofreader of my thesis and Prof. Dr. Stephan Nussberger for being the head of my committee.

A special thanks also goes to my MSc students who are now colleagues, Michael and Stefan. I personally learnt a lot during the supervision time and we grew up as a team bringing science forward.

I would like to thank the IBC gang from the middle lab, Stefan, Julian, Michael und Rebekka for having a great atmosphere always in the lab and not to forget the nice people from the right Lab like Alex, Sabine, Katrin and Philipp. Of course I am also thankful to the former lab members like Denis, Johannes, Raluca and Miru. A special thanks also goes to the Wolf group, Dieter, Alex, Ingo, Nicole and Ruth. During my PhD I have found really good friends and the IBC became my second home.

My special thanks also goes to the “old” lab members, Hans for becoming my climbing and mountaineering partner, and Elisabeth for being always there and helping in all situations.

Last but not least, my deepest gratitude goes to my family for supporting me in every situation of my life during the last 36 years. I am proud to be your son.
As being part of my family and the center of my life I am very grateful to the love of my life Cristiana for the immense and enthusiastic support during my whole scientific career and keeping me always motivated during the preparation time of this thesis. I am very proud to be your partner.

# List of publications

**Emperle, M.**, Rajavelu, A., Reinhardt, R., Jurkowska, R. Z., & Jeltsch, A. (2014). **Cooperative DNA binding and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase.** *The Journal of biological chemistry*, *289*(43), 29602-13.

**Emperle, M.**, Rajavelu, A., Kunert, S., Arimondo, P. B., Reinhardt, R., Jurkowska, R. Z., & Jeltsch, A. (2018). **The DNMT3A R882H mutant displays altered flanking sequence preferences.** *Nucleic acids research*, *46*(6), 3130-3139.

**Emperle, M.**, Dukatz, M., Kunert, S., Holzer, K., Rajavelu, A., Jurkowska, R. Z., & Jeltsch, A. (2018). **The DNMT3A R882H mutation does not cause dominant negative effects in purified mixed DNMT3A/R882H complexes.** *Scientific reports*, *8*(1), 13242.

Rajavelu, A.*, Lungu, C.*, **Emperle, M.**, Dukatz, M., Bröhm, A., Broche, J., Hanelt, I., Parsa, E., Schiffers, S., Karnik, R., Meissner, A., Carell, T., Rathert, P., Jurkowska, R. Z., & Jeltsch, A. (2018). **Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2.** *Nucleic acids research*, *46*(17), 9044-9056. *co-shared first authors

Rošić, S., Amouroux, R., Requena, C. E., Gomes, A., **Emperle, M.**, Beltran, T., Rane, J. K., Linnett, S., Selkirk, M. E., Schiffer, P. H., Bancroft, A. J., Grencis, R. K., Jeltsch, A., Hajkova, P., & Sarkies, P. (2018). **Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity.** *Nature genetics*, *50*(3), 452-459.

Hassanzadeh, M., Kasymov, R., Mahernia, S., Adib, M., **Emperle, M.**, Dukatz, M., Bashtrykov, P., Jeltsch, A., & Amanlou, M. (2017). **Discovery of Novel and Selective DNA Methyltransferase 1 Inhibitors by Pharmacophore and Docking-Based Virtual Screening.** *ChemistrySelect*, *2*(27), 8383-8392. (Not included in this thesis)

# Zusammenfassung

Die phänotypische und funktionelle Vielfalt der Säugetierzelltypen ist zu einem großen Teil auf epigenetische Signale zurückzuführen, die Genexpressionsprofile bestimmen und stabilisieren. Eines der wichtigsten epigenetischen Signale ist die DNA Cytosin-C5 Methylierung. Diese Modifikation wird früh in der Entwicklung durch die *de novo* DNA-Methyltransferasen DNMT3A und DNMT3B überwiegend in einem CpG-Dinukleotid-Kontext gesetzt. Eine exakte Festlegung der DNA-Methylierungsmuster ist für eine normale Entwicklung entscheidend und wird durch die präzise Regulation der DNMT-Aktivität am Chromatin bestimmt. Das Ziel dieser Arbeit war es, neue mechanistische Erkenntnisse über den Mechanismus von DNMT3A, seine Regulation durch Chromatin-Signale und Interaktionspartner sowie dessen Fehlregulation bei Krebs zu gewinnen. Des Weiteren wurde das Potenzial von DNMT3A zur Erzeugung von 3-Methylcytosin als Nebenreaktion untersucht.

Es konnte gezeigt werden, das DNMT3A an DNA ausgedehnte multimere Protein-/DNA-Fasern bildet. In der Literatur wurde jedoch postuliert, dass dieses Enzym DNA auf prozessive Weise methyliert, eine Eigenschaft, die mit der Faserbildung nicht vereinbar ist. Definierte biochemischer Experimente ermöglichten es zu zeigen, dass die DNA-Methylierungsrate von DNMT3A mit zunehmender Enzymkonzentration auf einem langen DNA-Substrat mehr als linear ansteigt, nicht aber auf einem kurzen 30-mer-Oligonukleotid, auf welchem größere DNMT3A-Polymere keinen Platz finden. Methylierungs-Experimente mit verschiedenen Enzymkonzentrationen und Substraten, die ein oder zwei CpG-Stellen enthalten, lieferten keine Hinweise auf einen prozessiven Mechanismus. Die Zugabe einer katalytisch inaktiven DNMT3A-Mutante konnte die DNA-Methylierungsrate von DNMT3A auf dem langen Substrat erhöhen, nicht aber auf dem Kurzen. Zusammengenommen zeigen diese Daten, dass DNMT3A in einer kooperativen Reaktion an DNA bindet und die Bildung von stabilen Protein/DNA-Fasern die DNA-Methylierungsrate erhöht.

Das zweite Projekt beschäftigte sich mit der Charakterisierung der Auswirkungen der R882H Mutante von DNMT3A. Diese Mutation befindet sich in der DNA-Bindungsstelle von DNMT3A und wird häufig bei akuter myeloischer Leukämie (AML) beobachtet. Die Etablierung eines speziellen Aufreinigungsverfahrens ermöglichte es mir in

R882H/Wildtyp DNMT3A-Komplexen zu zeigen, das diese Mutation nur eine geringe Reduktion der Methylierungsaktivität zur Folge hat. Jedoch konnte eine ausgeprägte Veränderung in der Flankierungssequenz-Präferenz von DNMT3A R882H gefunden werden. Basierend darauf wurde ein Substrat erstellt, dessen zentrale CpG Stelle sich im für R882H idealen Flankierungskontext befand. Auf diesem Substrat zeigte die R882H Mutante eine höhere Methylierungsaktivität als Wildtyp DNMT3A. Zusammengenommen, liefern diese Daten keinen Hinweis auf einen dominant-negativen Effekt der R882H Mutation, vielmehr deuten sie auf einen ortspezifischen Hyperaktivitäts-Effekt hin. Eine kürzlich ermittelte Struktur von DNMT3A im Komplex mit DNA stimmt mit diesen Ergebnissen überein und könnte die hohe Prävalenz dieser spezifischen Punktmutation in AML erklären.

Das dritte Projekt basiert auf früheren Experimenten des Labors, die eine starke und direkte Interaktion zwischen der ADD-Domäne von DNMT3A und der TRD-Domäne des 5mC-Leseproteins MECP2 aufdeckten. Kinetische Studien zeigten, dass die MECP2 Bindung die autoinhibitorische Konformation von DNMT3A allosterisch stabilisiert und die enzymatische Aktivität *in vitro* hemmt. Diese Effekte konnten durch die Bindung von unmodifiziertem Histon H3 aufgehoben werden. In meiner Arbeit habe ich die Interaktion dieser Domänen durch Größenausschlusschromatographie weiter validiert. Außerdem habe ich Zelllinien mit stabiler Überexpression von MECP2 generiert, und konnte so zeigen, dass MECP2 die DNMT3A-Aktivität auch in Zellen hemmt. Zusammenfassend, geben diese Daten detaillierte Einblicke in die Regulation von DNMT3A durch die Kombination von Chromatin-Modifikationen und Interaktionspartnern. MECP2 kann je nach Modifikationsstatus des H3-Schwanzes an seinem Zielort als Repressor oder Aktivator der DNA-Methylierung fungieren.

Das letzte Projekt beschäftigte sich mit der Koevolution von DNA-Methylierung und ALKB DNA-Reparatursystemen, das in enger Zusammenarbeit mit dem Labor von Dr. Peter Sarkies (MRC London) durchgeführt wurde. In *in vitro* Methylierungsexperimente mit der katalytischen Domäne von DNMT3A konnte ich zeigen, dass DNMT3A neben 5mC auch 3mC anfügen kann, eine Modifikation, die einen Alkylierungsschaden der DNA darstellt. Diese Studie bietet eine neue evolutionäre Perspektive auf den Verlust der DNA-Methylierung, die bei vielen Arten beobachtet wird.

# Abstract

The phenotypical and functional diversity of mammalian cell types can be attributed to a large extent to epigenetic signals that determine and stabilize gene expression profiles. One of the most important types of epigenetic signals is DNA methylation. This modification is set early in development by the *de novo* DNA methyltransferases DNMT3A and DNMT3B, and is found predominantly at the C5 position of cytosine bases in a CpG dinucleotide context. The accurate setting of DNA methylation patterns is critical for normal development and is determined by the precise recruitment and control of DNMT activity on chromatin. In this work, four main directions of research were undertaken, with the ultimate goal of shedding novel mechanistic insights into the mechanism of DNMT3A, its regulation by chromatin signals and interaction partners, as well as the dysregulation of this enzyme in cancer. Furthermore, the potential of DNMT3A to generate 3-methylcytosine as a side reaction was explored.

The DNA methyltransferase DNMT3A has been shown to multimerize on DNA and to form large multimeric protein/DNA fibers. However, it has also been postulated that this enzyme can methylate DNA in a processive manner, a property incompatible with fiber formation. By using a dedicated set of biochemical experiments, I was able to show that the DNA methylation rate of DNMT3A increases more than linearly with increasing enzyme concentration on a long DNA substrate, but not on a short 30-mer oligonucleotide, which cannot accommodate DNMT3A polymers. Methylation experiments over a range of enzyme concentrations and with substrates containing one or two CpG sites did not provide evidence for a processive mechanism. The addition of a catalytically inactive DNMT3A mutant was found to increase the DNA methylation rate by DNMT3A on the long substrate but not on the short one. Together, these data clearly indicate that DNMT3A binds to DNA in a cooperative reaction and the formation of protein/DNA fibers increases the DNA methylation rate. These results contribute mechanistic insights into the mode by which DNA methylation patterns are established during development.

The second project dealt with characterizing the effects of the R882H exchange on DNMT3A. The R882H mutation is found in the DNA binding interface of DNMT3A and

is frequently observed in acute myeloid leukemia (AML). By establishing a double-tag affinity purification system, I was able to show that the mutation only leads to a minor reduction in overall DNA methylation activity in mixed R882H/wildtype DNMT3A complexes. However, a pronounced change in flanking sequence preference of the DNMT3A-R882H mutant was found. Accordingly, a substrate designed to contain the target CpG site flanked by sequences preferred by R882H was better methylated by the variant than by the wildtype enzyme. Together, these data strongly argue against a dominant-negative effect of the R882H mutation and rather propose a site-specific gain-of-activity effect. These findings are in agreement with a recently determined structure of DNMT3A in complex with DNA and they might explain the high prevalence of this specific point mutation in AML.

The third project was built on previous data from the lab, documenting a strong and direct interaction between the ADD domain of DNMT3A and the TRD domain of the 5mC reading protein MECP2. These experiments revealed that through its binding, MECP2 allosterically stabilizes the autoinhibitory conformation of DNMT3A, resulting in a strong inhibition of enzymatic activity *in vitro*. The interaction between these two proteins and its associated inhibition could be disrupted by unmodified histone H3. In my work, I further validated the interaction between the ADD and the TRD domains by size exclusion chromatography. Also, by generating cell lines with stable over-expression of MECP2, I could show that MECP2 inhibits DNMT3A activity in cells. Together, the data from this study offer unprecedented insights into the regulation of DNMT3A by the combined action of chromatin modifications and interaction partners. Accordingly, depending on the modification status of the H3 tail at the target site, MECP2 can act as either a repressor or activator of DNA methylation.

The last project dealt with the coevolution between DNA methylation and DNA repair systems, a very exciting topic that was addressed in close collaboration with the laboratory of Dr. Peter Sarkies (MRC London). By performing *in vitro* methylation experiments with the catalytic domain of DNMT3A, I could show that, in addition to 5mC, DNMT3A can also introduce 3mC, a modification which represents an alkylation damage of DNA. This study provides a new evolutionary perspective on the loss of DNA methylation that is observed in many species.

# List of Abbreviations

| | |
|---|---|
| Ab | Antibody |
| ADD | ATRX-DNMT3-DNMT3L |
| AdoHyc | S-Adenosyl-L-homocysteine |
| AdoMet | S-Adenosyl-L-methionine |
| ALKB | Alkane 1-monooxygenase |
| AML | Acute myeloid leukemia |
| ATRX | Alpha thalassemia/mental retardation syndrome X-linked |
| BAH1 | Bromo-adjacent homology domains 1 |
| BAH2 | Bromo-adjacent homology domains 2 |
| BISMA | Bisulfite Sequencing DNA Methylation Analysis |
| BRCA | Breast cancer gene |
| CD | Catalytic domain |
| *CDH1* | Cadherin 1 |
| *CDKN1A* | Cyclin-dependent kinase inhibitor 1A |
| CGI | CG island |
| *CIMP* | CGI methylator phenotype |
| CN-AML | Cytogenically normal AML |
| CoREST | Co-repressor for elements-1-silencing transcription factor |
| COSMIC | Catalogue Of Somatic Mutations In Cancer |
| CREB1 | CAMP responsive element binding protein 1 |
| CTCF | CCCTC-binding factor |
| CTD | C-terminal domain |
| CXXC domain | cysteine rich domain; Cys-X-X-Cys |
| dCas9 | Nuclease deficient Cas9 |
| Ddm1 | Decrease in DNA methylation 1 |
| *Ddx4* | DEAD-Box helicase 4 |
| DMAPD | DNA methyltransferase-associated protein 1-interacting domain |
| DMR | Differentially methylated region |

| | |
|---|---|
| DMVs | DNA methylation valleys |
| DNMT | DNA (cytosine-5)-methyltransferase |
| *Dppa2* | Developmental pluripotency-associated protein 2 |
| DTT | Dithiothreitol |
| EcoR | Receptor for ecotropic host-range murine retroviruses |
| EDTA | Ethylenediaminetetraacetic acid |
| *ERBBS* | Enhanced reduced representation bisulfite sequencing |
| ES cells | Embryonic stem cells |
| *EYFP* | Enhanced yellow fluorescent protein |
| FACS | Fluorescence-activated cell sorting |
| FCS | Fetal calf serum |
| FRET | Förster resonance energy transfer |
| G4s | G-quadruplexes |
| G9a/EHMT2 | Euchromatic histone lysine methyltransferase 2 |
| GKn | Glycine lysine repeats |
| GLP | G9a-like protein |
| GST | Glutathiontransferasen |
| H3K36me3 | histone H3 tri-methylated on lysine 36 |
| H3K4me3 | histone H3 tri-methylated on lysine 4 |
| H3K9me3 | histone H3 tri-methylated on lysine 9 |
| HCT116 | Human colorectal carcinoma cell line |
| HDAC | Histone deacetylase |
| HELLS | Helicase, lymphoid specific |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| HP1 | Heterochromatin protein 1 |
| ICRs | Imprinting control regions |
| ID | Intervening domain |
| IDH | Isocitrate dehydrogenase |
| iPS cells | Induced pluripotent stem cell |
| IPTG | isopropyl-β-D-thiogalactopyranosid |
| KCl | Potassium chloride |
| KDM | Lysine demethylase |
| Klf4 | Krüppel-like factor 4 |

| | |
|---|---|
| KMT | Lysine methyltransferase |
| LC-MS/MS | Liquid chromatography electrospray ionization tandem mass spectrometry |
| *LINE* | Long interspersed nuclear elements |
| LSH | Lymphoid-specific helicase |
| LT-HSC | Long-term hematopoietic stem cell |
| *LTR* | Long terminal repeat |
| *MAGE* | Melanoma antigen |
| MBD | Methyl-CpG-binding domain |
| MBP | Maltose binding protein |
| MECP2 | Methyl-CpG-binding protein 2 |
| MLH1 | MutL homolog 1 |
| MPP8 | M-phase phosphoprotein 8 |
| MTases | Methyltransferase |
| MYC | Myelocytomatosis oncogene cellular homolog |
| N-CoR-SMRT | Nuclear receptor corepressor and silencing mediator of retinoic acid and thyroid hormone receptor |
| NGC | Next generation chromatographie |
| NGS | Next generation sequencing |
| Ni-NTA | Ni2+ nitrilotriacetic acid agarose |
| NLS | Nuclear localization signal |
| NTD | N-terminal domain |
| *OCT4* | Octamer-binding transcription factor 4 |
| PBD | PCNA-binding domain |
| PBST | Phosphate buffered saline with Tween-20 |
| PCNA | proliferating cell nuclear antigen |
| PWWP | Pro-Trp-Trp-Pro |
| qPCR | quantitative polymerase chain reaction |
| *R-RAS* | Ras-related protein |
| *RAC1* | Ras-related C3 botulinum toxin substrate 1 |
| RAD51 | Recombinase |
| RFTD | Replication foci targeting sequence (RFTS) domain |
| *RHOC* | Ras homolog family member C |

| | |
|---|---|
| RT | Room temperature |
| rtTA3 | Reverse tetracycline-controlled transactivator 3 |
| SAM | S-adenosyl-L-methionine |
| *SDHA* | Succinate dehydrogenase complex, subunit A |
| SDS | Sodium dodecyl sulfate |
| SET domain | Su(var)3-9, enhancer-of-zeste and trithorax domain |
| SETD2 | SET domain containing 2 |
| SETDB1 | SET domain bifurcated 1 |
| SFM | Scanning force microscopy |
| sgRNA | Single guide RNA |
| SIN3A | Swi-independent 3A |
| *SINE* | Short interspersed nuclear elements |
| *Slc25a31* | Solute carrier family 25 member 31 |
| SRA | SET-and-Ring finger-associated domain |
| SUHW1 | Suppressor of hairy wing homolog 1 |
| *Syce1* | Synaptonemal complex central element protein 1 |
| TET | Ten-eleven translocation |
| *Tex11* | Testis-expressed sequence 11 protein |
| TFs | Transcription factors |
| *TRD* | Target recognition domain |
| TRD | Transcriptional repressor domain |
| UDG | Uracil-DNA glycosylase |
| UHRF1 | Ubiquitin like with PHD and RING finger domains 1 |
| *UTR* | Untranscribed region |
| ZBTB24 | Zinc finger And BTB domain-containing protein 24 |
| ZF protein | Zinc finger protein |

# 1. Introduction

## 1.1 Foundations of Epigenetics

The development of multicellular organisms requires a precisely synchronized cascade of events that involves a step-wise functional specification of cellular populations, from the totipotent cells of the zygote to the over 200 differentiated cell types that make up the various organs (Moris et al., 2016). Noteworthy, while almost all the cells in the human body share a nearly identical set of genes, they are phenotypically and functionally clearly distinct. A main regulator of this fascinating cellular diversity is epigenetic information. The term “Epigenetics” was coined in 1942 by Conrad Waddington and was defined as “the study of the interactions between genes and their products which result in a particular phenotype” (Waddington, 1942). The aim of this description was to explain the changes in phenotype, for which little mechanistic understanding was available (Allis and Jenuwein, 2016; Waddington, 1942).

Because of the rapid development of the field in the last decade the definition of the term “Epigenetics” was further modified and refined to “the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence” (Wu and Morris, 2001).

Today we know that epigenetic signals establish and stabilize cellular phenotypes by maintaining gene expression states. Importantly, this is achieved through changes in the local interpretation of the genetic code and not by changes in the sequence of the DNA (*i.e.* mutations). Accordingly, while most cells of the human body have access to an identical set of genes, only a subset of these will be active in a particular cell type at a particular time (Allis and Jenuwein, 2016).

As depicted in **Figure 1**, the accurate establishment of an inheritable epigenetic state requires three important factors (Berger et al., 2009). The “Epigenator”, which is an extracellular signal that is able to activate an intracellular signalling pathway, acts as a priming factor for the “epigenetic Initiator”. This is able to bind to a particular genomic region and subsequently induce a local change in the epigenetic landscape. Finally, this newly established state is perpetuated by the “epigenetic Maintainer”. To keep in

mind is that the maintainer lacks absolute sequence specificity, thereby relying on the Initiator for its locus-specific recruitment.

There are three main types of epigenetic signals, namely DNA methylation, covalent modifications of histone proteins and non-coding RNAs. All of these operate simultaneously and non-randomly to modulate chromatin structure and impinge on gene expression. Due to its high relevance for this thesis, the next sections will focus in particular on DNA methylation.

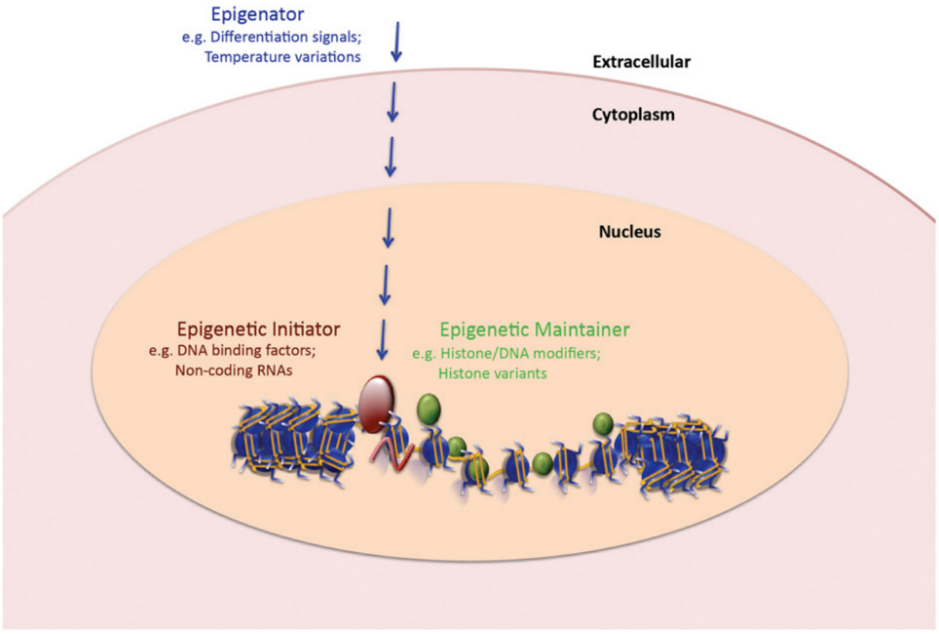



**Figure 1 Principle actors in the epigenetic pathway.** The pathway is based on 3 different functional elements: first, the "Epigenator" (blue), an extracellular stimulus, triggers the epigenetic pathway. Second, the "epigenetic Initiator" (red), with its DNA sequence specificity, targets a defined locus on a chromosome and triggers changes in the epigenetic landscape. These changes are then taken over and inherited by the "epigenetic Maintainer" (green)(Image was taken from Berger et al., 2009).

## 1.2 Evolutionary impact of DNA methylation

DNA methylation was for the first time discovered in 1948 by Hotchkiss and since then this modification was intensively studied (Hotchkiss, 1948). It is found primarily at the C5 position of cytosine bases and is present with few exceptions in most eukaryotes ranging from fungi to vertebrates, where it has various functions. For mammals, DNA methylation is a fundamental and vital modification, which plays essential roles in embryonic development and adult tissue homeostasis (Messerschmidt et al., 2014).

DNA methylation occurs predominantly at palindromic CpG sites on both DNA strands. There are around 56 million CpG sites in the human genome, 60-80% of which are methylated. This corresponds to 4-7% of all cytosines (Laurent et al. 2010; Lister et al. 2009). To be noted is the fact that modified CpG sites are mutagenic hotspots because their hydrolytic deamination can lead to TpG mismatches (Jurkowska et al., 2011a). By comparison, spontaneous deamination of unmethylated cytosine bases takes place up to 4 times more slowly *in vitro,* and gives rise to the unnatural DNA base, uracil, which can be much easier detected and corrected by a dedicated repair systems in cells like Uracil-DNA glycosylase (UDG) (Krokan et al., 2000; Shen et al., 1994). The described mutagenic effect of C5 methylated cytosine has led to a selective evolutionary depletion of CpG sites from mammalian genomes. Indeed, the CpG dinucleotides are underrepresented in the human genome by a factor of 5 compared to other dinucleotide combinations (Jurkowska et al., 2011a).

CpG dinucleotides show a bimodal distribution, being globally underrepresented in inter- and intra-genic regions while being abundant at repetitive and gene regulatory elements. At gene promoters, CpG dinucleotides cluster together and form so called CpG islands (CGI), regions with a length of 500-2000 bps, a GC content of more than 50%, and a ratio of observed over expected number of CpG dinucleotides above 0.6 (Gardiner-Garden and Frommer, 1987; Takai and Jones, 2003). Looking at the distribution of methylated CpG dinucleotides, it can be seen that 60-80% of “single” CpG sites are methylated, depending on the cell type, while CGIs remain predominantly unmethylated (Deaton and Bird, 2011; Messerschmidt et al., 2014).

About 70% of all human genes, including housekeeping and tissue-specific genes have CGIs at their promoters (Saxonov et al., 2006). The fact that these CpG-rich clusters survived in the mammalian genome might be explained by the fact that in the germline, CGIs are lowly methylated or completely unmethylated, thereby posing little mutagenic drift risk at this stage (Shen et al., 2007; Weber et al., 2007). As cells differentiate, the methylation status of CpG sites is subjected to dynamic changes. Accordingly, while CpG dense promoters are unmethylated in the germline, some gain methylation progressively during development. This ensures a stable gene repression (*e.g.* by silencing of pluripotency genes) as cells become increasingly differentiated (Borgel et al., 2010; Meissner et al., 2008).

Repetitive elements like pericentromeric repeats are also heavily decorated with methylated CpG dinucleotides.

Based on the number of the modified CpG sites and their genomic position, the 5mC signal can regulate gene expression in various ways. For instance, at promoters, DNA methylation is known to silence gene expression, which plays important roles in processes like X-chromosome inactivation and cellular differentiation (Reddington et al., 2013). Also, through their methylation, the activation of endogenous retroviruses and transposons is prevented. On the other hand, when found at bodies of actively transcribed genes, DNA methylation is known to prevent spurious intragenic transcription initiation. Methylation of gene bodies was also found to be involved in the regulation of alternative splicing. The most direct downstream effect of DNA methylation is its impact on the recruitment of methylation sensitive proteins and on the binding of transcription factors (Anastasiadou et al., 2011; Aran and Hellman, 2013; Bartolomei, 2009; Dantas Machado et al., 2015; Ferguson-Smith, 2011; Hellman and Chess, 2007; Jones, 2012; Jurkowska et al., 2011a; Klose and Bird, 2006; Patel, 2016)

## 1.3 The mammalian DNA methyltransferases: mechanisms of catalysis

DNA methylation patterns are introduced and maintained by the family of DNA methyltransferases (DNMTs / DNA MTases), which contains 3 active enzymes: DNMT1, DNMT3A, and DNMT3B. In addition the family encompasses the catalytically inactive variant DNMT3L, which acts as a regulatory factor (**Figure 2**) (Gowher and Jeltsch, 2018). In addition, the recently discovered DNMT3C is present exclusively in rodents where it is responsible for the methylation of evolutionary young retrotransposons in the male germ line (Barau et al., 2016). The catalytic domains of both DNMT3A and DNMT3B are active in isolated form and were used in biochemical studies to gain insights into the catalytic mechanisms of DNA methylation (Gowher and Jeltsch, 2002). To keep in mind is that the catalytic domain of DNMT3A is identical in aa sequence between the mouse and human enzyme.

From a structural point of view, DNMTs consist of two main parts: a large N-terminal part containing multiple domains with regulatory functions and a C-terminal part, where the transfer of the methyl group from the cofactor S-adenosyl-L-methionine (SAM) to

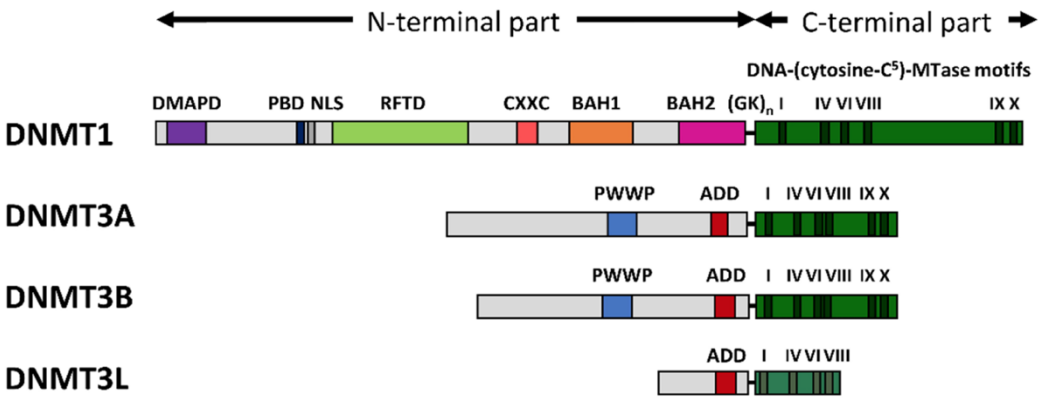



**Figure 2: Domain structure of the mammalian DNMT enzymes.** The human DNMT1, DNMT3A, DNMT3B and DNMT3L proteins consist of 1616, 912, 853 and 387 amino acid residues, respectively. DMAPD, DNA methyltransferase-associated protein 1-interacting domain; PBD, PCNA-binding domain; NLS, nuclear localization signal; RFTD, replication foci targeting sequence (RFTS) domain; CXXC, CXXC domain; BAH1 and BAH2, bromo-adjacent homology domains 1 and 2; GKn, glycine lysine repeats; PWWP, PWWP domain; ADD, ATRX-DNMT3-DNMT3L domain (Figure taken from Gowher and Jeltsch, 2018).

the C5 position of cytosine takes place. This part has an MTase fold that is conserved among prokaryotic and eukaryotic C5 DNMTs and contains the active center of the enzyme and 10 conserved amino acid motifs that are diagnostic for all C5 DNMTs (Jurkowska et al., 2011a; Keck, 1995). The MTase fold consists of 7 β strands, 6 of which being oriented in parallel, while the 7$^{th}$ strand is inserted in an anti-parallel orientation between strands 5 and 6. Inside, the motifs X and I are involved in SAM binding while the motifs IV (PCN**)**, VI (ENV) and IX form the unit responsible for catalysis. The non-conserved region between the motifs VIII and IX are involved in DNA recognition and specificity (Jeltsch, 2002, 2006; Jurkowska et al., 2011a; Keck, 1995). DNMT3L, which is catalytically inactive, contains a crippled catalytic domain, which lost motifs IX and X (Bourc'his et al., 2001).

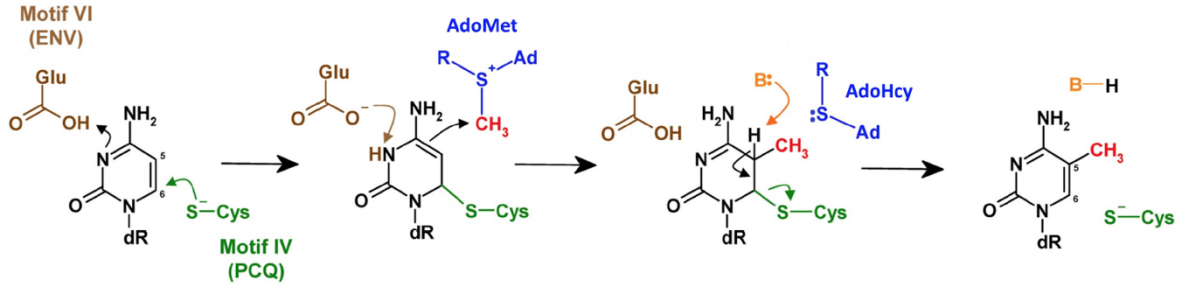



**Figure 3: Catalytic mechanism of cytosine C5 DNA methyltransferases**. The brown colored ENV motif (Motif VI) contains the catalytic glutamate residue, the catalytic cysteine residue from PCQ motif (motif IV) is colored green. The cofactor SAM (AdoMet) is colored in blue with the methyl group highlighted in red. The base for final proton abstraction (shown in orange) is not identified; it may be a water molecule (Figure taken from Gowher and Jeltsch, 2018).

In more detail, the transfer of the methyl group involves base flipping of the target cytosine out of the DNA helix into a hydrophobic pocket in the active center of the enzyme (Jeltsch, 2002, 2006). Afterwards, a nucleophilic attack of the catalytically active cysteine of the PCQ motif (IV) occurs, which leads to the formation of a covalent bond between the enzyme and the target cytosine (**Figure 3**). Subsequently, the methyl group is transferred to the C5 position, deprotonation takes place and finally the covalent bound formed between the enzyme and the DNA base is cleaved (Jeltsch, 2006; Jurkowska et al., 2011a). The addition of a methyl group to C5 does not disturb the Watson-Crick paring of the modified base, but the presence of the hydrophobic group within the major groove of the DNA leads to a subtle bending and twisting effect in the crystal structure of methylated DNA oligonucleotides (Tippin and Sundaralingam, 1997).

## 1.4 The classical and revised models of DNA methylation inheritance

Classically, all DNMTs have important and non-overlapping functions. The establishment of methylation patterns is done by DNMT3A and DNMT3B, which are also called *de novo* DNMTs; they use unmethylated DNA as substrate (**Figure 4**) (Chédin, 2011; Okano et al., 1999). *Dnmt3a* is maternally provided and is primarily expressed in the oocytes and early preimplantation embryos. The resulting enzyme is involved in establishing the differential methylation patterns at imprinting control regions (ICRs) in male and female gametes (Kaneda et al., 2004; Kato et al., 2007; Messerschmidt et al., 2014). In comparison, *Dnmt3b* is transcribed upon zygotic gene activation is mainly expressed during the blastocyst stage, mostly in the epiblast lineage (Watanabe et al., 2002).

Knockout studies have shown that despite their structural similarity, DNMT3A and DNMT3B have partial non-overlapping functions. Accordingly, while deleting *Dnmt3b* is embryonically lethal, *Dnmt3a* knockouts are partially viable (Okano et al., 1999). Despite its lack of catalytic activity, DNMT3L is indispensable for the activity of *de novo* DNA methyltransferases under specific conditions. Indeed, *Dnmt3l* knock out mice show loss of *de novo* methylation in the germline, which causes sterility in males and embryonic lethality of maternal null-derived embryos (Bourc'his et al., 2001;

Messerschmidt et al., 2014). This demonstrates the activity and the importance of the DNMTs and their crosstalk.

After the initial methylation pattern is set by the *de novo* DNMTs, the maintenance methyltransferase DNMT1 takes over (**Figure 4**). This fact has a high importance because after each round of DNA replication, hemimethylated DNA is created (**Figure 4**), where just the parental strand harbours the 5mC mark. DNMT1, which has hemimethylated DNA as substrate, is responsible for restoring the methylation pattern on the newly synthesized daughter strand. DNMT1 has a high preference for hemimethylated DNA and this property is essential for restricting its methylation activity to the maintenance functions. DNMT1 expression is tightly controlled, it is activated in the S phase by cell-cycle-dependent transcription factors, which lead to a high expression in mitotic cells (Kishikawa et al., 2003).

The recruitment of DNMT1 to replication foci takes place through its interaction with PCNA. The binding to hemimethylated DNA is further promoted by UHRF1. PCNA and UHRF1 concentrate DNMT1 at sites of active DNA replication and away from areas that do not contain hemimethylated DNA. (Avvakumov et al., 2008; Bostick et al., 2007; Sharif and Koseki, 2011; Sharif et al., 2007).

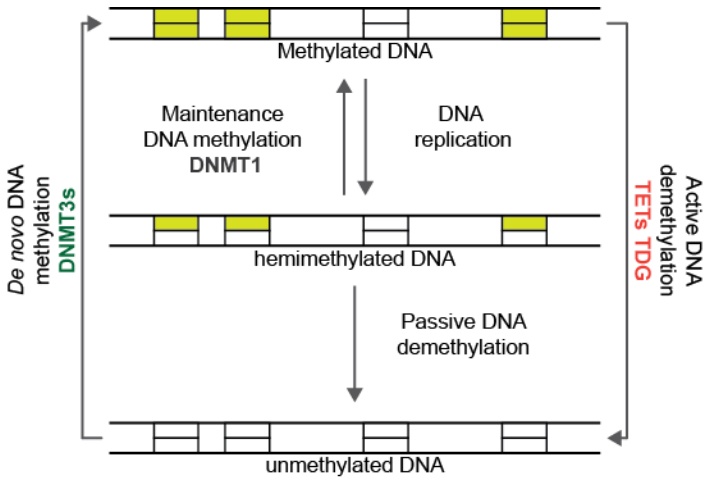



**Figure 4: The classical model of DNA methylation.** Initially, the DNA methylation pattern consists out of unmethylated and fully methylated CpG sites, the latter of which being generated by the DNMT3s *de novo* methyltransferases. With each round of cell division and DNA replication, the newly generated hemimethylated DNA is converted back to a fully methylated state by the maintenance MTase DNMT1. In addition to passive demethylation, the methylation pattern also can be actively removed through the action of TET enzymes. In this model, DNMT1 is seen as the maintenance enzyme, while DNMT3A and DNMT3B are seen as *de novo* methyltransferases. (Figure adapted from Jeltsch and Jurkowska, 2014).

Additionally, it was reported that UHRF1 stimulates the activity of DNMT1 by a factor of 5 *in vitro*, increasing the specificity of DNMT1 for hemimethylated CpG sites (Bashtrykov et al., 2014). In mice the deletion of *Dnmt1* is lethal at the gastrula state, because of a drastic loss of DNA methylation at a global level (Brown and Robertson, 2007; Kurihara et al., 2008). Interestingly, an almost similar effect can be observed when UHRF1 is knocked out (Sharif and Koseki, 2011; Sharif et al., 2007).

If the function of DNMT1 is getting suppressed, passive DNA demethylation takes place, which results in around 50% loss of 5mC in each round of DNA replication. In parallel to this, there is also active demethylation taking place by the activity of TET enzymes (**Figure 4**) (Ito et al., 2011; Véron and Peters, 2011). One of the proposed functions of active demethylation is for instance the prevention of erroneous methylation of CpG islands (Williams et al., 2012).

Despite its elegance, recent studies demonstrated that the “classical model” of DNA methylation, with its clear division of labour between the *de novo* DNMT3s and the maintenance DNMT1, is too crude (Jeltsch and Jurkowska, 2014). Instead, DNA methylation has to be seen as a dynamic process of on-going methylation and demethylation events in which the functions of DNMT1 and DNMT3 enzymes partially overlap (Jeltsch and Jurkowska, 2014; Jones and Liang, 2009; Riggs and Xiong, 2004). Several lines of experimental observations initiated the revision of the classical model of DNA methylation.

First, the prediction of the classical model that all cells coming from one tissue would show identical 5mC patterns could not be confirmed by bisulfite conversion followed by sequencing. Instead, it was found that, rather than conserving the methylation status of individual CpG sites, the average methylation density profile of a particular DNA region is preserved (Jones and Liang, 2009; Zhang et al., 2009a). These cellular studies found support in biochemical work showing that DNMT1 has a 10-40 fold higher preference for hemimethylated DNA. Although this accuracy is high, it is not enough to flawlessly copy the methylation status of each of the 56 million CpG sites in the human genome (Bashtrykov et al., 2012; Jeltsch and Jurkowska, 2014; Song et al., 2012).

The second argument that supports the revised DNA methylation model stems from studies in mammalian cell lines and mice, where it was demonstrated that deletion of the *de novo* DNA methyltransferases DNMT3A and DNM3B results in a loss of DNA methylation at repetitive elements. These observations argue for a maintenance function of DNMT3A and DNMT3B, as the presence of a fully functional DNMT1 was not sufficient to maintain DNA methylation at these sites (Chen et al., 2003; Dodge et al., 2005; Egger et al., 2006).

The third argument in favour of the revised DNA methylation model comes from recent work in human stem cells and differentiated neuronal progenitors, where DNA methylation was also found at asymmetric sequences, particularly in a CA context. For instance, in human iPS cells, 5mCs are found in approximately 68.31%, 7.81%, 1.99%, and 1.05% of CpG, CA, CT, and CC sites, respectively (Barrès et al., 2009; Guo et al., 2014a, 2014b, Lister et al., 2009, 2013; Pinney et al., 2014). While non-CpG methylation only accounts for a mere 0.02% of the overall 5mCs in somatic cells, it can make up to 25% at non-CpG sites in human male ES cells (Jang et al., 2017). This can be correlated with the expression level of DNMT3A and DNMT3B, the enzymes that can introduce this type of modification (Arand et al., 2012; Shirane et al., 2013). The fact, that the template strand information cannot be used for non-palindromic methylated sequences, hence they cannot be a substrate for DNMT1, demonstrates that there is a permanent need for *de novo* activity of DNMT3A for the stable presence of CA methylation in several cell types (Jeltsch and Jurkowska, 2014).

The last argument for the revised model was the discovery of TET enzymes, because since then it is clear that there is a continuous need of active *de novo* MTase to counteract active DNA demethylation performed by TET enzymes (Arand et al., 2012; Métivier et al., 2008; Williams et al., 2012).

Taken together these four key arguments described above support the revision of the “classical model” of DNA methylation inheritance and highlight the importance of regulatory factors in the targeting and control of the DNA methylation machinery (Jeltsch and Jurkowska, 2014).Based on this refined concept, the corrections of methylation errors that happen during the setting and maintenance of DNA methylation patterns can be explained. Accordingly, DNA methylation is rather a dynamic process,

in which the methylation of each individual CpG site is determined by the local activity of the DNA methyltransferases. Chromatin-associated signals, DNA demethylases and the DNA replication rate are factors that are in place to control the DNA methylation patterns and act as feedback loops. An important implication of the revised model is that the average DNA methylation levels of a particular DNA region are maintained and not the methylation status of each individual CpG site. Mistakes in the DNA methylation patterns at these sites can be corrected *via* the crosstalk with the different chromatin marks (Jeltsch and Jurkowska, 2014).

## 1.5 Regulation of the activity of the DNMT3 enzymes

### 1.5.1 The N-terminus of *de novo* DNMTs and its regulatory functions

The N-terminal part of the active *de novo* DNMTs 3A and 3B contains two conserved functional domains, the ADD (ATRX-DNMT3-DNMT3L) and the PWWP (Pro-Trp-Trp-Pro) domain, the latter being not present in DNMT3L (**Figure 2**, page 5). After the PWWP domain, towards the end of the N-terminus, DNMT3A and DNMT3B contain a non-conserved region, which was shown to be able to bind DNA and likely contributes to the tight association of the enzymes to chromatin (Suetake et al., 2011). Until today, the exact function of this region is still unknown, based on the low sequence identity in this region between both proteins (10-11%) it is possible that this part is responsible for the differential recruitment of DNMT3A and DNMT3B activities to chromatin (Jeltsch and Jurkowska, 2016; Rondelet et al., 2016).

Both, the ADD and the PWWP play important roles in guiding the DNMT activity on chromatin and ensure a continuous cross talk between DNA methylation and histone tail modifications. The PWWP domain is part of the ‘Royal’ superfamily of domains, which can recognize methylated lysines on the histone tail through a conserved aromatic cage (Qin and Min, 2014; Rondelet et al., 2016). In case of the DNMT3 enzymes this domain binds specifically the H3K36me2/3 modification, which is found in the bodies of expressed genes (Baubec et al., 2015; Dhayalan et al., 2010; Hahn et al., 2011; Lee and Shilatifard, 2007). Here it is introduced by the lysine methyltransferase SETD2, which is recruited by the elongating RNAPII (Edmunds et al., 2008). The PWWP plays very important roles in the chromatin targeting of

DNMT3s, since cells expressing a DNMT3B mutant that is not capable to bind to H3K36me3 were found to have reduced DNA methylation levels. In addition, mutating the H3K36me3 binding pocket in DNMT3A lead to a partial delocalization of the enzyme from pericentromeric heterochromatin (Chen et al., 2004; Dhayalan et al., 2010). Also, Dhayalan et al., (2010) showed that the PWWP-H3K36me3 interaction increases the activity of DNMT3A for the *in vitro* methylation of nucleosomal DNA.

Not only the PWWP domain, but also the ADD domain was found to be an important regulator of DNMT3 activity. This domain was shown to specifically read the modification status of the H3K4 tail, keeping in check the activity of DNMTs particularly at promoter elements. Accordingly, histone modifications associated with active genes such as di-and trimethylation of H3K4 (H3K4me3/me3) disrupt the binding of the ADD domain, while H3K4me0 and me1 allow binding (Otani et al., 2009; Zhang et al., 2010). The relevance of this was elegantly demonstrated in a recent structural study by Guo et al., (2015). The authors found that DNM3A is allosterically inhibited, adopting a closed conformation where the ADD domain docks against a hydrophobic surface of the catalytic domain and as a result, blocks catalysis. By contrast, binding of the unmodified H3K4 tail to the ADD domain, allosterically releases this inhibition by disruption the intramolecular interaction between the catalytic and the ADD domain. Importantly, the findings of this structural work are supported by several cellular studies (Morselli et al., 2015; Noh et al., 2015). Accordingly, expression of a DNMT3A mutant variant designed to be insensitive to the status of the H3K4 methylation resulted in the introduction of ectopic DNA methylation at both H3K4me3-marked CGIs as well as intergenic regions (Noh et al., 2015). Of note, the CGIs were more resistant to DNA methylation by comparison to intergenic regions, indicating that mechanisms other than H3K4me3 are in place to prevent abnormal silencing of active genes (Noh et al., 2015). Indeed, the ADD domain is known to function as a protein-protein interaction interface (Laisné et al., 2018). Not only DNMT3A, but also DNMT3B was found to be sensitive to the modification status of H3K4 (Zhang et al., 2010). Exogenous expression of this enzyme in the budding yeast *Saccharomyces cerevisiae,* which lacks endogenous DNA methylation, led to the deposition of DNA methylation in a manner that was strongly anti-correlated with H3K4me3. Conversely, deletion of the H3K4 MTase *set1,* culminated into a spreading of DNA methylation into promoter regions (Morselli et al., 2015).

### 1.5.2 Regulation of the activity of DNMT3s through multimerization

The first hints at a role of multimerization in regulating the activity of the *de novo* DNA MTases was provided by the catalytic inactive member of the family, namely DNMT3L. As described above, although DNMT3L lacks amino acid motifs that are critical for catalysis, it is an important regulator of the DNA methylation landscape. This is due to the fact that DNMT3L can directly interact with the catalytic domains of both DNMT3A and DNMT3B and as a result, stimulate their enzymatic activity both in *in vitro* and *in vivo* (Chedin et al., 2002; Chen et al., 2005; Gowher et al., 2005; Kareta et al., 2006; Suetake et al., 2004). The interaction between DNMT3L and in particular DNMT3A is required for normal germ cell development, by enforcing the correct methylation pattern at imprinted regions (Hata et al., 2002).

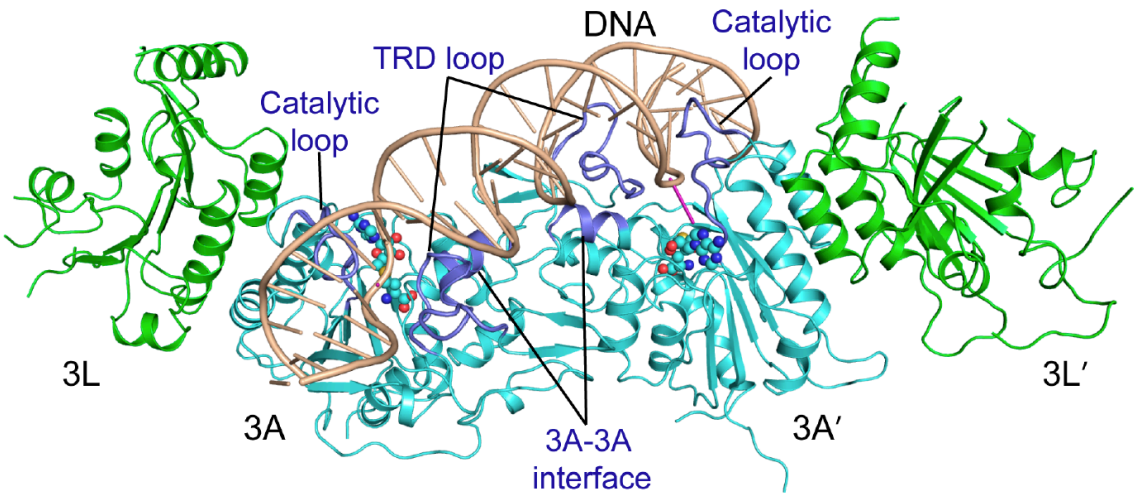



**Figure 5: Crystal structure of the DNMT3A-3L tetramer in complex with DNA.** The image shows a DNMT3A(Cyan) -3L(Green) tetramer covalently bound to a 25mer DNA duplex, which contains two CpG/ZpG sites (Z = Zebularine, colored in purple). The DNA binding interface is formed by residues located in the catalytic and the TRD loop as well as at the 3A-3A interface, as indicated in the figure. Abbreviation: TRD - target recognition domain (Adapted from Ren et al., 2018).

Important mechanistic insights into the multimerization of DNMT3 enzymes have been provided in 2007 by Jia and colleagues, who solved the crystal structure of the catalytic domain of DNMT3A in complex with DNMT3L (Jia et al., 2007). The main findings of this study were recently confirmed by Zhang et al. (2018) who were able to crystalize the DNMT3A-DNMT3L tetramer in complex with DNA (**Figure 5**). The study by Jia et al. (2007) did not only provide a structural explanation for the interaction between DNMT3A and DNMT3L, but also identified a previously uncharacterized dimerization

property of DNMT3A. Accordingly, it revealed that DNMT3A forms a linear heterotetramer with DNMT3L, where the inactive partner flanks the two centrally located DNMT3A subunits. This arrangement is the result of two important interfaces. First, four stacked phenylalanine residues (two from each subunit) form one hydrophobic “FF” DNMT3A/3L interface. Here, residues from the DNMT3L side make contacts with key catalytic or SAM binding residues of DNMT3A, explaining the stimulatory effect of DNMT3L on DNMT3A activity. The structural symmetry of the interface suggests that DNMT3A subunits can also form a stable FF interface, which has been validated experimentally (see below).

The second interface is polar and is referred to as the “RD” interface. This is formed by a network of salt bridges and hydrogen bonds between arginine and aspartate residues that are located at the interface between the two DNMT3A molecules. Importantly, the RD interface generates the DNA binding cleft. Due to the lack of motifs IX, X in DNMT3L, this interface can only be formed by the DNMT3A-DNMT3A homodimer. In contrast, the FF interface is present in both the homo and heterodimer.

Notably, although CpG sites are palindromic, the arrangement of the active sites between the two DNMT3A monomers does not allow one complex to simultaneously methylate one CpG site in both strands. Since the monomers are separated by ~40 Å, a distance equivalent to one helical turn of DNA, the crystal structure suggested instead that the DNMT3A dimer is capable of methylating two CpG sites located across the opposite strands of one DNA duplex, separated by ~10 bp DNA, in one binding event. This 10 bp periodicity model was subsequently supported by several biochemical and cellular studies, and is in line with the observation that this spacing is found in the differentially methylated regions (DMRs) of 12 maternally imprinted mouse genes, which are biological substrates of the DNMT3A/3L complex (Jia et al., 2007; Jurkowska et al., 2008).

Recently, the crystal structure of DNMT3A-DNMT3L in complex with DNA was solved (Zhang et al., 2018). To this end, a DNA duplex containing two separate CpG sites was used, where the target cytosine was replaced by zebularine, leading to the formation of covalent bonds between the flipped zebularine bases and the active site cysteine residues in the two active centers. This structure revealed a productive state

of the DNMT3A-DNA complex and closely resembles the DNA-free DNMT3A-DNMT3L complex solved by Jia et al. a decade earlier (**Figure 5**). The most notable structural difference arises from a disordered loop that is part of the target recognition domain (TRD) and which becomes ordered upon DNA binding. The structure revealed that the interaction between the enzyme and the DNA is mediated by the catalytic loop, the TRD loop and the RD interface generated by the DNMT3A-DNMT3A molecules. Interestingly, this work found that the two CpG sites are separated by 14 bps and not 10 bps, as in the previous model and experiments. These conflicting observations await further investigation.

The fact that DNMT3A possess both a FF and a RD interface has important structural consequences. Accordingly, in the absence of DNMT3L, the enzyme can further oligomerize, several potential DNA binding sites are generated in this process (**Figure 6**) (Jurkowska et al., 2011b). Several biochemical studies, addressing the mode of binding of the DNMT3A oligomer to DNA, have revealed an unusual mode of binding. Accordingly, Förster resonance energy transfer (FRET) assays using a labelled, flexible DNA substrate, as well as scanning force microscopy (SFM) work, revealed that DNMT3A oligomers are able to bind simultaneously to two DNA molecules oriented in parallel (Jurkowska et al., 2008, 2011b; Rajavelu et al., 2012). This unusual property of DNMT3A oligomers to bind DNA was found to be important for targeting of the enzyme to chromatin *in vivo*. Wildtype DNMT3A is enriched at pericentromeric heterochromatic repeats, which in mouse fibroblasts cluster into distinct foci (Bachman et al., 2001). By contrast, DNMT3A variants with mutations that affect the integrity of the FF interface, lost this heterochromatic localization. This was in spite of the fact that the ADD and PWWP domains were intact, indicating that DNMT3A polymerization is critical for the heterochromatic localization of this enzyme (Jurkowska et al., 2011b). While the biological relevance of this finding is not clear, one hypothesis is that this might contribute to the long-term maintenance of DNA methylation at heterochromatic loci, which are highly enriched in repeats and other highly methylated sequences (Jones and Liang, 2009; Jurkowska et al., 2011a). Noteworthy, co-transfection of DNMT3A with DNMT3L, which lacks an intact RD interface, distributed the enzyme away from heterochromatin foci into euchromatin. This property of DNMT3L to regulate the oligomerization status of DNMT3A is supported by *in vitro* work, the precipitation

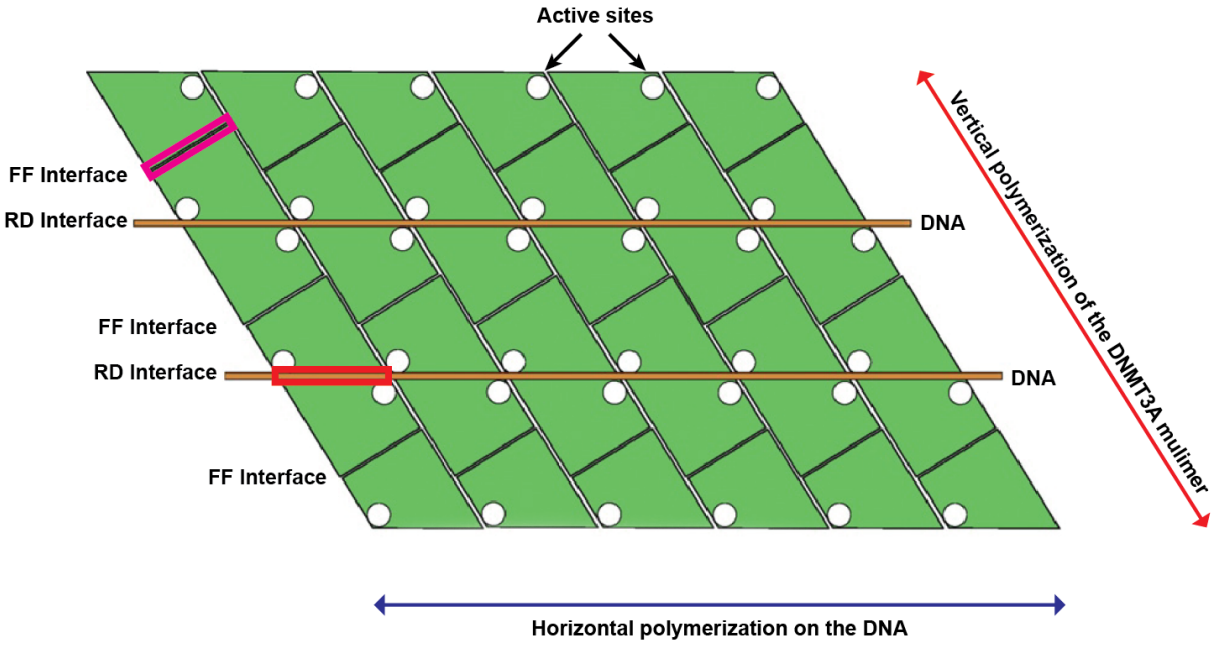



**Figure 6: Schematic illustration of two dimensional multimerization of DNMT3A on DNA.** Here, both process are demonstrated, the vertical polymerization of DNMT3A leading to parallel DNA binding and the horizontal multimerization along the DNA (Figure adapted from Jeltsch and Jurkowska, 2013b).

of DNMT3A with decreasing concentrations of salt being reduced by the presence of DNMT3L (Jurkowska et al., 2011b).

Importantly, since the RD and the FF interfaces are conserved between DNMT3A and DNMT3B, these two enzymes can also form heterooligomers. Indeed, measurements on cells transfected with enzymes fused with FRET-compatible fluorescent proteins have demonstrated that such heterooligomers can form *in vivo*. Common regulated targets are for instance, the promoters of the *OCT4* and *NANOG* genes (Li et al., 2007).

Worth noting is that DNMT3A and DNMT3B were found to have distinct flanking sequence preferences, *i.e.* depending on the sequence flanking the central CpG site, the two enzymes display different methylation efficiencies (Handa and Jeltsch, 2005; Jurkowska et al., 2011c; Lin et al., 2002). Since the RD interface forms the DNA binding cleft of DNMT3A, the flanking sequence preferences are most probably related to processes at the protein/DNA interface in the RD region. Indeed, site directed mutagenesis work on recombinantly purified DNMT3A revealed an altered sequence preferences when the residues in the RD interface were exchanged (Gowher et al., 2006). Accordingly, the R720A variant had completely lost the flanking preferences observed with the wildtype enzyme, while the R882A variant displayed a reduced

sensitivity towards changes of the inner flank and it did not show any preference for the outer flanks. Based on these data, it is tempting to speculate that distinct RD interfaces when one compares DNMT3A/3A with DNMT3A/3B complexes might influence the flanking sequence preferences of the corresponding complexes, which would translate into distinct DNA methylation patterns in cells (Jeltsch and Jurkowska, 2013).

DNA is a polymer that can provide several binding sites for DNMT3A oligomers to bind non-specifically in an organized manner. Several studies have shown that binding of several DNMT3A and DNMT3A/3L complexes to DNA does not occur independently but in a cooperative manner, this results in the formation of large nucleoprotein filaments on the DNA (**Figure 6**) (Jia et al., 2007; Jurkowska et al., 2008, 2011b). Cooperative DNA binding implies that an unbound enzyme will preferentially associate to a DNA molecule next to an already bound protein (Jurkowska et al., 2011b). That DNMT3A is able to bind DNA cooperatively was indicated by gel retardation assays, where no binding intermediates could be observed with increasing protein concentrations. In addition, Hill-analysis of the binding behaviour of DNMT3A to fluorescently labelled oligonucleotides in solution revealed sigmoidal binding, with at least two DNMT3A complexes binding a 30mer and at least three complexes binding to a longer 60mer substrate. Finally, the formation of polymers on DNA could be directly visualized in SFM studies. DNMT3L further supports the presence of linear active heterotetrameric complexes with two DNMT3A subunits in the center being connected by the RD interface (Jia et al., 2007; Liebert and Jeltsch, 2008).

Importantly, while the RD interface is largely conserved between DNMT3A and DNMT3B, the enzymes do differ in their catalytic mechanism. Accordingly, while DNMT3A binds DNA in a cooperatively and forms stable protein/DNA fibers, DNMT3B methylates multiple CpG sites in a processive manner (Jia et al., 2007; Jurkowska et al., 2008; Rajavelu et al., 2012). This difference to DNMT3A was addressed in more detail in a recent study, where the authors have used DNA substrates of various lengths to distinguish between a processive and a cooperative mechanism of DNMT3B action (Norvil et al., 2018). With this experimental setup, it was found that binding of the enzyme to specific sites on DNA is favoured by SAM binding, and that during the processive catalysis, DNMT3B is capable of cofactor exchange without dissociating

from DNA. By contrast, early studies have found that DNMT3A follows an ordered Bi-Bi mechanism in which DNA binding occurs first and SAM binding is second (Yokochi and Robertson, 2002).

Noteworthy, despite intensive investigations the exact kinetic mechanism of DNMT3A is still under debate. While there is solid evidence of a cooperative way of action, another group reported that DNMT3A methylates its substrate in a processive manner and that DNMT3L acts as a positive processivity factor (Holz-Schietinger and Reich, 2010). Mechanistically, processive catalysis and cooperative DNA binding exclude each other, because processive turnovers go along with a single enzyme complexes moving along a DNA substrate and is not compatible with protein complexes, which are multimerizing on DNA.

In summary, despite significant progress in dissecting protein multimerization of the DNMT3 enzymes, many questions still await experimental clarification. For instance, based on the sequence similarity and the interface conservation, DNMT3B is expected to self-oligomerize as well. However, to date, no evidence for an oligomeric state of DNMT3B is known. In addition, owing to the conflicting reports, the exact mode of action of DNMT3A awaits elucidation. This last point was of central importance for this work, where the functional role of DNMT3A multimerization on DNA was investigated.

### 1.5.3 Recruitment and regulation of DNMT3s by interaction partners

The accurate setting of DNA methylation patterns is critical for normal development and is determined by the precise recruitment and control of DNMT activity on chromatin. While many studies have contributed to a clear understanding of the mechanism of catalysis of DNMT enzymes, our insight into the targeting and regulation of these enzymes in the context of chromatin is still in its infancy (Gowher and Jeltsch, 2018; Jeltsch and Jurkowska, 2016). The following paragraphs will summarize key recent findings that have addressed this point, with particular focus on the DNMT3 enzymes.

As described above, DNA methylation can influence the recruitment of transcription factors to chromatin. Conversely, transcription factors can also directly recruit the DNMT machinery to target genomic sites (Laisné et al., 2018). For instance, MYC was

shown to actively recruit DNMT3A to silence the *p21/CDKN1A* locus (Brenner et al., 2005). The transcription factor E2F6 was reported to mediate the silencing of the germline genes *Slc25a31*, *Syce1*, *Tex11* and *Ddx4* by direct recruitment of DNMT3B (Velasco et al., 2010). Recently, the zinc finger protein ZBTB24 was found to be important for the gene body methylation of a subset of genes involved in the maintenance of cellular homeostasis. The authors show that this is due to the ZBTB24-mediated recruitment of DNMT3B, where absence of ZBTB24 led to a loss of DNA methylation at gene bodies (Thompson et al., 2018). Importantly, this effect was restricted to DNMT3B, and no effects on DNMT3A or DNMT1 were observed.

As mentioned before, chromatin modifications were found to play important roles in controlling the recruitment the activity of DNMT3 enzymes. Accordingly, recognition of H3K36me3 modification is important for the methylation of major satellite repeats and gene bodies of actively transcribed genes. Also, the H3K4me3 modification, present at active promoters, allosterically inhibits DNMT3A activity (Jeltsch et al., 2018). Both DNMT3A and DNMT3B were found to associate not only with the modified chromatin but also with some of the enzymes that introduce these modifications (Laisné et al., 2018). For instance, DNMT3A and DNMT3B, but not DNMT1, were reported to interact with the H3K9 methyltransferase SETDB1. The interaction with DNMT3A enforced the silencing of genes that are frequently down-regulated in human cancers (Li et al., 2006). The H3K9me3-binding protein MPP8 provides an interesting type of DNMT3A interactor as it may act as a bridge between the G9A/GLP H3K9 protein methyltransferase complex and DNMT3A (Chang et al., 2011). Accordingly, in addition to its binding to the repressive histone modification, the chromodomain of MMP8 can also recognize dimethylated DNMT3A K44me2 and self-methylated GLP. This mark on DNMT3A is introduced by G9A/GLP. *Via* homodimerization, MMP8 could thereby bring these two epigenetic silencers in close spatial proximity and further promote gene repression (Chang et al., 2011).

Non-coding RNAs can also recruit DNMT proteins. Here, most groups have focused on the recruitment of DNMT1, data on DNMT3A and DNMT3B being still sparse. In this respect, one of the best-characterized examples is given by RNA-DNA triplex structures, which are found at the promoters of rDNA, and were observed to promote the DNMT3B-mediated transcriptional silencing of these loci (Schmitz et al., 2010).

Recently, a group identified the *lncRNA Dum* to interact with DNMTs and regulate *Dppa2* expression during myogenic differentiation and muscle regeneration (Wang et al., 2015).

Chromatin remodellers also have a strong historical connection with the DNMT machinery. Accordingly, a screen performed in *Arabidopsis thaliana* for mutants that lose DNA methylation have led to the identification of the Ddm1 chromatin remodeller (Jeddeloh et al., 1999). Follow up investigations performed in mammalian systems have revealed that the mammalian Ddm1 homolog HELLS/LSH is essential for DNA methylation (Yu et al., 2014; Zhu et al., 2006). HELLS was found to directly interact with DNMT3A and probably stimulates the DNA methylation by facilitating the access of the methylation machinery to sites of highly compacted chromatin (Briones and Muegge, 2012; Lungu et al., 2015). Indeed, several studies have shown that DNMTs cannot efficiently bind and methylate nucleosomal substrates (Felle et al., 2011; Lyons and Zilberman, 2017; Takeshima et al., 2006; Zemach et al., 2013).

In summary, recent work has underscored the importance of interaction partners in the accurate deposition and maintenance of DNA methylation patterns. Nevertheless, many questions are still unanswered. For instance, with the exception of the regulatory factor DNMT3L, no others regulators of DNMT3 polymerization have been found to date (Jeltsch and Jurkowska, 2013). In addition, while most studies have focused on mechanisms of recruitment and activation of DNMT3 enzymes, very few have focused on factors that have an inhibitory function on DNMT3 activity. Inhibition of DNMT3 activity is essential to prevent abnormal DNA methylation and subsequent gene silencing.

## 1.6 Genomic distribution and variability of DNA methylation

DNA methylation shows a bimodal distribution, in which CpG-containing areas are found in either a ‘largely methylated’ or ‘largely unmethylated’ state (Jeltsch et al., 2018). DNA methylation  it is uniquely distributed among the different genomic elements and its deposition is influenced by the presence of other epigenetic marks, such as histone tail modifications (**Figure 7**) (Jeltsch et al., 2018).

| Genomic region | Methylation level (%) |
|---|---|
| All | 83.52 |
| Canyons | 4.33 |
| CGIs | 7.71 |
| CGI shore | 48.02 |
| Gene | 82.68 |
| Intron | 84.45 |
| Exon | 72.4 |
| Promoter | 23.78 |
| 5'-UTR | 68.61 |
| 3'-UTR | 83.07 |
| LINE | 88.53 |
| LTR | 89.79 |
| SINE | 90.03 |
| H3K4me3 regions | 9.35 |
| H3K27me3 regions | 43.04 |
| H3K36me3 regions | 94.03 |
| CTCF binding regions | 41.04 |
| Gata2 binding regions | 47.26 |
| PU.1 binding regions | 17.48 |

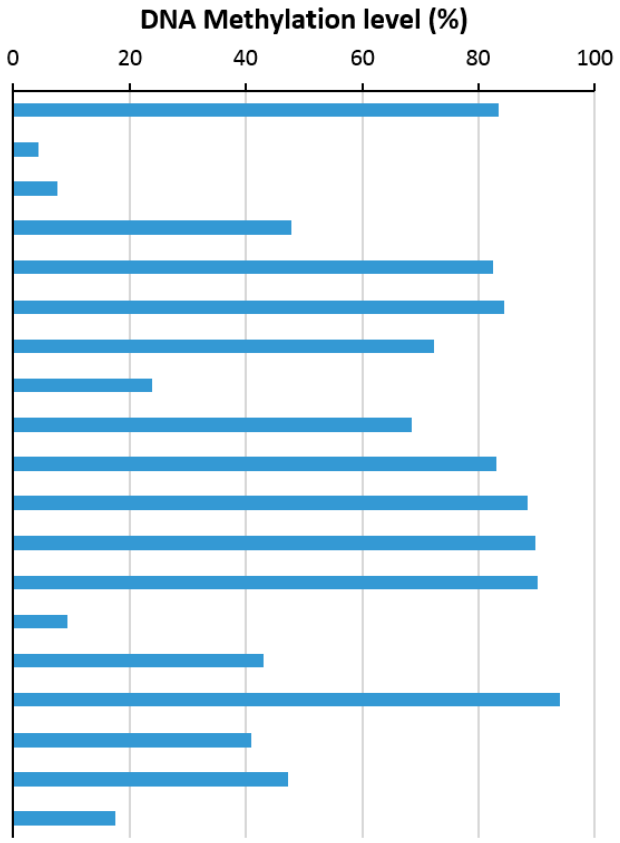



**Figure 7: Exemplary DNA methylation levels in mouse hematopoietic stem cells in various genome regions.** CGI: CpG islands, UTR: untranscribed region, SINE: Short Interspersed Nuclear Elements, LINE: Long Interspersed Nuclear Elements, LTR: Long Terminal Repeat (Figure taken from Jeltsch et al., 2018).

70% of all gene promoters contain CGIs. Their methylation is tightly controlled during development and is usually restricted to stable silencing as in X-chromosome inactivation, silencing of imprinted genes and transposable elements, but aberrant CGI methylation is observed in many diseases. For instance, a CGI methylator phenotype (CIMP), where hundreds of CGIs become hypermethylated, is found in a subset of colorectal cancers and results in the inactivation of tumour suppressor genes (Yi et al., 2016). Depending on their CpG content, promoters can be further subdivided into high, intermediate and low CpG promoters, genome-wide methylome analysis demonstrating a clear anti-correlation between CpG density and DNA methylation levels (Meissner et al., 2008; Weber et al., 2007; Zhang et al., 2009a). Recently, it has become clear that CpG sites that are located distally to traditional CGIs can also control gene expression. For example, tissue- and cancer-specific differentially methylated regions occur more frequently within CGI shores. These are regions that flank CGIs (up to 2kb distant) and have lower CpG density than the CGIs themselves (Doi et al., 2009; Irizarry et al., 2009).

CGIs are generally protected from DNA methylation, different mechanisms being in place for this (Jeltsch et al., 2018). In an elegant study on trans-chromosomic animal models, Klose and co-workers (2016) have shown that promoter-associated CGIs are refractory to DNA methylation regardless of the host species, demonstrating that the DNA sequence plays a central role in keeping the CGIs in an unmethylated state by mechanisms conserved between mouse and zebrafish (Long et al., 2016). Furthermore, in a recent study, Mao *et al* (2018) showed that particular DNA secondary structures, known as G-quadruplexes (G4s), which are present at circa 23% of all CGIs, inhibit DNMT1 activity and can protect the CGIs from local methylation (Mao et al., 2018).

In contrast to CGIs, elements distal to gene promoters exhibit more variable methylation between species, uncovering a widespread dependence on nucleotide frequency and occupancy of DNA-binding transcription factors (TFs) in shaping the DNA methylation landscape away from gene promoters (Long et al., 2016). The tissue-specific expression of TFs can also explain the strong overrepresentation of regulatory elements in regions that show differential methylation during development (Ziller et al., 2013). Mechanistically, transcription factors can physically block the access of DNMTs to the underlying DNA sequence, or recruit epigenetic modifiers that antagonize DNA methylation (Jeltsch et al., 2018; Stadler et al., 2011). Proteins that contain CXXC domains can recognize unmethylated CGIs and protect their unmethylated state. These structural folds are present in many chromatin regulators such as the KDM2A and KDM2B H3K36-specific lysine demethylases, the KMT2A and KMT2B H3K4-specific protein lysine methyltransferases, as well as the methylcytosine dioxygenases TET1 and TET3 (Hashimoto et al., 2010; Long et al., 2013; Tanaka et al., 2014; Xu et al., 2018). H3K36me3 accumulates in the bodies of expressed genes and can be recognized by the PWWP domain of DNMT3 enzymes. As described above, this modification plays important roles in guiding the activity of the methyltransferases on chromatin (Chen et al., 2004; Dhayalan et al., 2010). Also, H3K4me2/3 is known to antagonize *de novo* DNA methylation by an allosteric inhibition mechanism (Jeltsch and Jurkowska, 2016).

A particular class of gene regulatory elements are enhancers. They functionally resemble promoters by containing binding sites for TFs, but they show depletion of

CpG sites roughly corresponding to the average genome (Jeltsch et al., 2018). In a recent paper, Cedar and co-workers (2018) documented the existence of programmed postnatal changes in enhancer methylation in many different tissues of the body and, by focusing on the liver, found that hormone signalling induces many of these alterations (Reizel et al., 2018). Unlike the methylation of CGIs, the methylation of TF binding sites located in CpG-poor promoter and at enhancers can have variable effects on gene expression. The outcome of the modification depends on the specific role of the TF (activating or repressing) and on the influence of DNA methylation on the binding of the TF (Yin et al., 2017). CTCF is a zinc finger protein, which is known to promote the interaction between enhancer and cognate promoters by forming DNA loops, while buffering interactions between sequences located inside and outside the loops (Ong and Corces, 2014). The binding of this TF to DNA is sensitive to DNA methylation, as investigated in more detail by a recent structural study (Hashimoto et al., 2017). Interestingly, depending on the position of the methylated cytosine within the 15bp-long motif, the authors found that this has distinct effects on CTCF binding. Accordingly, CTCF binding could be disrupted by increased $C_2$ methylation or enhanced by increased $C_{12}$ methylation in the 15bp-long motif (Hashimoto et al., 2017).

DNA methylation can promote the binding of several transcription factors. For instance, the Krüppel-like factor 4 (Klf4) is a zinc-finger transcription factor that preferentially recognized methylated CpG sites. In depth analysis revealed that Klf4 is recruited to methylated enhancer elements, where it triggers chromatin remodelling and transcriptional activation (Wan et al., 2017). Noteworthy, among the affected targets are *RHOC* and *RAC1*, which encode proteins that promote cell migration and motility. This study proposed a new paradigm for DNA-methylation mediated chromatin remodelling and gene activation (Wan et al., 2017). Recently, DNMT3A and DNMT3B were found to associate in an H3K36me3-dependent manner with active enhancers in epidermal stem cells (Rinaldi et al., 2016). Interestingly, while both proteins were required for the production of enhancer RNA, they had different effects on enhancer DNA modification. While DNMT3B was found to be involved in enhancer body methylation, DNMT3A was shown to cooperate with TET2 and promote enhancer DNA hydroxymethylation (Rinaldi et al., 2016). These data illustrate a dual regulatory potential of DNMT3A and DNMT3B for gene repression and activation.

Arguably, among of the best-characterized targets of DNMTs are repetitive elements. These are abundantly decorated with DNA methylation to prevent their activation and subsequent genome instability. More than two-thirds of the mammalian genome consists of repetitive elements (de Koning et al., 2011). These are Short Interspersed Nuclear Elements (SINEs), Long Interspersed Nuclear Elements (LINEs), Long Terminal Repeats (LTRs), major satellites and simple repeats (Papin et al., 2017). All of these show a depletion of CpG sites and their methylation varies throughout development. For instance, CpG-poor SINEs show high DNA methylation, while the methylation of CpG-rich LTRs is highly dynamic during differentiation (Papin et al., 2017).

Genome-wide methylome analyses have also revealed the existence of a new class of large hypomethylated regions, referred to as “canyons” or DNA methylation “valleys” (DMVs) (Jeong et al., 2014; Xie et al., 2013). These canyons span up to 1MB, they contain both CGI and non-CGI regions, are hypomethylated throughout development, and are highly conserved across vertebrates. These sites form hotspots of regulatory regions, as they are highly enriched for TF binding sites (Li et al., 2018). Importantly, the borders of canyons are demarked by 5-hydroxymethylcytosine and become eroded in the absence of DNMT3A, suggesting that these are regions of dynamic competition between DNMTs and TETs to maintain a *status quo*. About half of the genes in these methylation canyons are coated with repressive histone marks, whereas the remainder are covered by activating histone marks and are highly expressed in hematopoietic stem cells. Interestingly, genes deregulated in human leukemias such as AML were found enriched at canyons (Jeong et al., 2014).

## 1.7 Interpretation of DNA methylation patterns

The deposition of the 5mC modification does not dramatically alter the structure of the DNA helix. Instead, this mark exerts its function through two main mechanisms. As described above, DNA methylation can influence the binding affinities of transcription factors. Alternatively, the mark can act as a docking platform for 5mC binding proteins that recruit downstream effectors, such as chromatin remodelling complexes (Patel, 2016). There are several groups of proteins that can interpret and locally modulate the DNA methylation pattern. These include the families of methyl-CpG-binding domain (MBD) proteins, which recognize fully methylated CpG dinucleotides, the Kaiso family of ZF proteins, which preferentially bind to CpG sites that are embedded in long, specific sequences, and the SET-and-Ring finger-associated (SRA) domain family, which recognize the hemimethylated DNA generated behind the replication fork (Du et al., 2015; Filion et al., 2006; Hendrich and Bird, 1998; Unoki et al., 2004). Owing to its relevance for this thesis, the following subsection will focus on the founding member of the MBD family, the methyl-CpG-binding protein 2 (MECP2).

MECP2 is an important chromatin architectural protein that is ubiquitously expressed, reaching near histone octamer levels in postmitotic neurons (Ausió et al., 2014; Guy et al., 2011). The gene encoding for the protein is located on the X-chromosome and is thereby subjected to X-chromosome inactivation (Amir et al., 1999). Aberrations in the MECP2 protein sequence or dosage are associated with the Rett syndrome, a severe neurodevelopmental disorder (Ausió et al., 2014). There are two main MECP2 isoforms that are generated through alternative splicing; isoform 1 is translated from exon one and is the predominant isoform expressed in the brain, isoform 2 is translated from exon 2 (Ausió et al., 2014). Structurally, the protein consists of six well-defined domains: the N-terminal domain (NTD), the methyl-binding domain (MBD), intervening domain (ID), transcriptional repressor domain (TRD), and the C-terminal domains α and β (CTDα and CTDβ) (Martínez de Paz and Ausió, 2017). Traditionally viewed as a reader of 5mCpG sites, recent work has uncovered that MECP2 might have a broader specificity and could be involved in the readout of mCH (H= A, T, C) and hmCH modification (Feng et al., 2010; Gabel et al., 2015; Guo et al., 2014a; Hashimoto et al., 2012; Kinde et al., 2015).

Among the domains of MECP2, the TRD is known to act as a protein-protein interaction interface, mediating the interaction with several co-repressor complexes. Several studies have highlighted the unusual amino acid composition of MECP2. Accordingly, 60-65% of the protein is intrinsically disordered and displays random coil-like hydrodynamic properties (Adams et al., 2007; Hite et al., 2009). This suggests that MECP2 might heavily rely on interacting partners to adopt a folded structure. This hypothesis is supported by an increasingly diverse list of factors reported to interact with MECP2 and which connect this protein to a plethora of functions and biological effects (Hite et al., 2009; Della Ragione et al., 2016).

MECP2 has been traditionally viewed as a transcriptional repressor, although this has been challenged by a series of recent reports (Della Ragione et al., 2016). Arguably, the best-characterized interactor of MECP2 is the SIN3A/HDAC co-repressor complex, which docks on the TRD domain of the protein. This recruitment of the SIN3A/HDAC complex to chromatin leads to the local deacetylation of nucleosomes and subsequent gene repression (Jones et al., 1998; Nan et al., 1996). Other interacting proteins are, for example, the N-CoR-SMRT (nuclear receptor corepressor and silencing mediator of retinoic acid and thyroid hormone receptor) complex, the CoREST (co-repressor for elements-1-silencing transcription factor) complex, ATRX and HP1; all of these are important members of several gene silencing machineries (Martínez de Paz and Ausió, 2017). Surprisingly, despite its high expression levels as well as interaction with a cohort of high-profile epigenetic regulators, transcriptional profiles of brain isolated from *Mecp2*-null mice has found a disappointingly small subset of genes with altered expression (Tudor et al., 2002). On the other hand, gene expression profiling of specific areas of the brain, like the hypothalamus and cerebellum, was more informative, with thousands of genes showing altered expression. Strikingly, most of these appeared to be downregulated in the absence of MECP2, suggesting that the protein might act as a transcriptional activator for them (Ben-Shachar et al., 2009) (Chahrour et al., 2008; Sugino et al., 2014). Indeed, MECP2 was found to associate with the transcriptional activator CREB1 at the promoter of activated target genes but not at repressed promoters (Chahrour et al., 2008). Together, these data indicate that depending on the chromatin locus and interaction partner, MECP2 can either activate or inhibit gene expression.

Surprisingly, despite the fact that MECP2 is a well-established reader of DNA methylation, only few studies have addressed a possible interplay between this protein and the DNA methyltransferase machinery. Kimura *et al* (2003) found that MECP2 can directly *via* its TRD domain with DNMT1; this interaction may promote maintenance methylation *in vivo* (Kimura and Shiota, 2003). Despite the fact that both DNMT3A and MECP2 are highly expressed in the adult brain and are part of the same epigenetic pathway, a cross talk between these two proteins has not been addressed to date. In the present work, the interaction between these two important epigenetic factors was addressed by combining a set of biochemical and cellular methods.

## 1.8 DNA methylation and cancer

The accurate establishment and maintenance of DNA methylation patterns are essential for normal mammalian development. Their importance is highlighted by the growing number of diseases known to develop when epigenetic processes go awry (Robertson, 2005) (**Figure 8**). For instance, a connection between DNA methylation and cancer was already established in 1983, where Feinberg and Vogelstein found that by comparison to their normal counterparts, cancer cells show globally reduced levels of DNA methylation (Feinberg and Vogelstein, 1983). This drastic effect is due to the loss of the mark at repetitive elements, sequences that comprise approximately half of the human genome (**Figure 8**) (Weisenberger et al., 2005). These alterations in the epigenetic landscape have drastic cellular consequences, as they lead to the activation of transposable elements and subsequent genomic instability, a hallmark of tumour cells. Perturbations in the epigenetic landscape were observed at focal regions as well. Accordingly, silencing of tumour suppressor genes as well as activation of oncogenes were identified (**Figure 8**). For instance, abnormal growth advantages have been linked to hypomethylation and thereby activation of the oncogenes *R-RAS and MAP-SIN* in gastric cancer, *S-100* in colon cancer and *MAGE* in melanoma (Wilson et al., 2007). This is complemented by a hypermethylation of the CGIs found at the promoters of tumour suppressors, genes that are essential for cell-cycle maintenance, DNA repair, cell adhesion, apoptosis and angiogenesis. CGI hypermethylation has been described

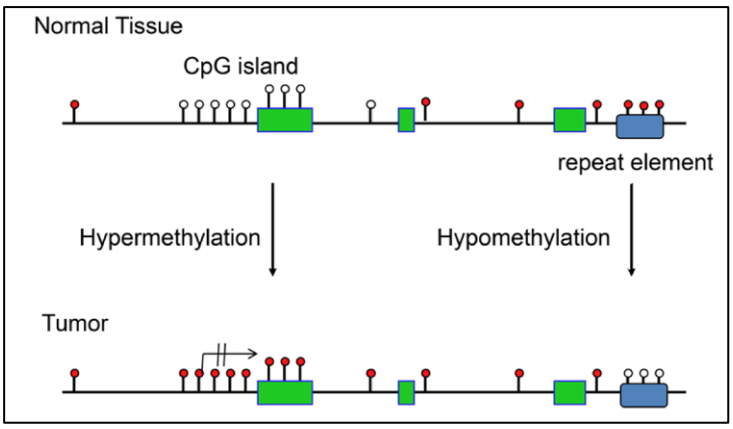



**Figure 8: Schematic illustration of the most important changes in DNA methylation observed in cancers.** Loss of DNA methylation occurs on a genome wide level and is often observed at repetitive regions. These become activated and give rise to genome instability, a hallmark of cancer. In addition to the global hypomethylation, focal DNA hypermethylation events are found at a subset of CGIs, this leads to gene silencing. Tumor suppressors can be silenced through this mechanism. White circles symbolizes unmethylated CpG sites, red circles show methylated CpG sites. The transcription start site and the constant lack of transcription after DNA methylation is marked by the crossed out arrow. Repetitive elements are represented by the blue rectangle and the green boxes highlight exons (Figure taken from Pfeifer et al., 2018).

for almost every tumour type and encompasses epigenetic lesions that affect, for instance, DNA repair (*MLH1*), cell cycle (*INK4a-ARF* locus) and cell adhesion (*CDH1*) genes (Esteller, 2002). DNA hypermethylation of one allele of a tumour suppressor is a key component of the Knudson's two-hit hypothesis, where a somatic or germline mutation is responsible for the inactivation of one allele, while DNA methylation can inactivate the second allele (Di Ruscio et al., 2016). Importantly, with the recent advent of high-throughput technologies, it has become clear that not only the methylation of CGIs but also that of CpG sites found at regulatory elements outside of CGIs play critical roles in cancer development. For instance, a recent study on breast cancer patients has shown that DNA methylation of enhancer elements can identify distinct breast cancer lineages (Fleischer et al., 2017).

Changes in the DNA methylation landscape are a frequent and early event in cancer, they correlate with the disease progression state and with the metastatic potential of many tumour types (Robertson, 2005; Widschwendter et al., 2004). Although recent OMICS studies have revealed that epigenetic alterations in cancer can surpass the number of genetic variations, it is currently unknown what are the mechanisms responsible for the accumulation of such global epigenetic alterations, and which of these are causes or rather consequences of tumour development (Lechner et al., 2010). An important point to keep in mind is that although the importance of the DNA

methylation landscape in cancer progression is well documented, it is technically very difficult to distinguish between 'driver' alterations, which provide a selective growth advantage and consequently promote cancer development, and 'passenger' changes, which passively accumulate during cancer progression (Pon and Marra, 2015; Roy et al., 2014). As discussed in more detail in the next section, it is also possible that the early stages of cancer development are triggered by 'driver' mutations, which occur in key epigenetic enzymes and as a consequence predispose the cells to the accumulation of additional epigenetic changes. Conversely, the initial event could also be an epigenetic change that would subsequently make a certain genomic area more prone to the accumulation of genetic mutations (Pon and Marra, 2015).

### 1.8.1 Genetic alterations of the DNA methylation machinery in AML

AML is a type of haematological cancer that affects the myeloid lineage and causes a clonal, malignant proliferation of white blood cells (Ferrara and Schiffer, 2013). This uncontrolled proliferation, which is coupled with impaired cellular differentiation, leads to the accumulation of immature blood cells in the bone marrow, which culminates in the development of cytopenia, neutropenia, and thrombocytopenia. Clinical manifestations of symptoms include fatigue, dyspnoea, susceptibility to infections, and haemorrhage (Chaudry and Chevassut, 2017). More than half of the reported AML cases affect patients older than 65, some infrequent cases have also been reported for children (Ferrara and Schiffer, 2013). In adults, AML is the most common type of acute leukemia and its prognosis is dependent on the age of the patient (Li et al., 2017). Accordingly, the 5-year survival rate of patients older than 65 is around 15–20%, while younger patients were reported to have a better survival rate (Schlenk and Döhner, 2013).

As revealed by early karyotypic studies and further characterized by NGS work, AML is a disease of genetic heterogeneity, with no single mutation being found to drive the disease; instead several mutations contribute to leukaemogenesis (Li et al., 2016; Papaemmanuil et al., 2016; Testa et al., 1979). For instance, cytogenetic or chromosomal abnormalities, such as translocations, are reported in 55% of patients with AML. Using a cohort a of 1540 patients, Papaemmanuil et al. (2016) were able to identify distinct molecular subgroups that reflect discrete paths in the evolution of AML, an analysis that may help improve the classification of the disease as well as its prognostic stratification. Noteworthy, nearly 50% of AML samples have a normal

karyotype and are cytogenetically normal. Instead, these patients carry various molecular mutations that contribute to AML pathogenesis and form an intermediate-risk group (Chaudry and Chevassut, 2017; Kumar, 2011).

Findings of several recent next-generation sequencing studies have provided important insights into the pattern of mutation acquisition, the existence of preleukemic stem cells and the critical involvement of epigenetic enzymes in AML development. Several lines of evidence propose a model where mutations in “landscaping” genes, genes which are involved in global chromatin changes, happen early in the evolution of AML, while mutations in “proliferative” genes occur late (Corces-Zimmerman et al., 2014; Krönke et al., 2013; Shlush et al., 2014). Accordingly, mutations in genes that are involved in epigenetic regulation (*e.g. DNMT3A* and *TET2*) could be identified in preleukemic hematopoietic stem cells, a subpopulation of cells that are capable of multilineage differentiation, can survive chemotherapy and contribute to remission (Corces-Zimmerman et al., 2014; Schlenk and Döhner, 2013). In line with this, aberrant DNA methylation was found to be a hallmark of AML. For instance, a methylation profiling analysis performed by the TCGA Research Network on 192 samples of AML, using Illumina 450K arrays, identified significant changes in DNA methylation at 160,519 CpG loci, which accounted for 42% of the sites tested, with 67% resulting in hypermethylation and 33% resulting in hypomethylation (Ley et al., 2013). A pairwise AML cohort study examining the DNA methylation by enhanced reduced representation bisulfite sequencing (ERBBS), proposed that global DNA methylation allele shifting was a universal feature of AML relative to normal bone marrow controls (Li et al., 2016). Another DNA methylation profiling study on a cohort of 344 AML patients could stratify these into 16 distinct subgroups, each of these displaying a unique epigenetic signature when compared with the normal bone marrow CD34+ cells (Figueroa et al., 2010). Interestingly, a genome-wide differential methylation study on cytogenically normal AML (CN-AML), found that the most pronounced changes in DNA methylation occurred in non-CpG island regions, whereas hypermethylation enrichment was only represented in CpG islands (Qu et al., 2014). Importantly, this study also proposed that mutations in *DNMT3A* have a dominant effect in determining the methylome in CN-AML, which results in a genome-wide hypomethylated phenotype.

A connection between DNMT3A and hematopoiesis has already been observed in early knockout studies in mice, where *Dnmt3a* deletion led to a preference for self-renewal over differentiation and gradual expansion of the long-term hematopoietic stem cell (LT-HSC) compartment on serial transplantation (Challen et al., 2012). By comparison, deletion of *Dnmt3b* did not drastically affect the LT-HSC compartment, which can be explained by the fact that the predominant DNMT3B isoform, which is expressed in the hematopoietic system, is catalytically inactive (Challen et al., 2014). Interestingly, as shown by Goodell and co-workers (2015), *Dnmt3a* knockout predisposed mice to the development of myeloid and lymphoid malignancies, with distinct DNA methylation abnormalities, depending on the lineage. Accordingly, global hypomethylation was observed in all of the malignancies, but lymphoid malignancies also exhibited hypermethylation, particularly at promoter regions (Mayle et al., 2015) .

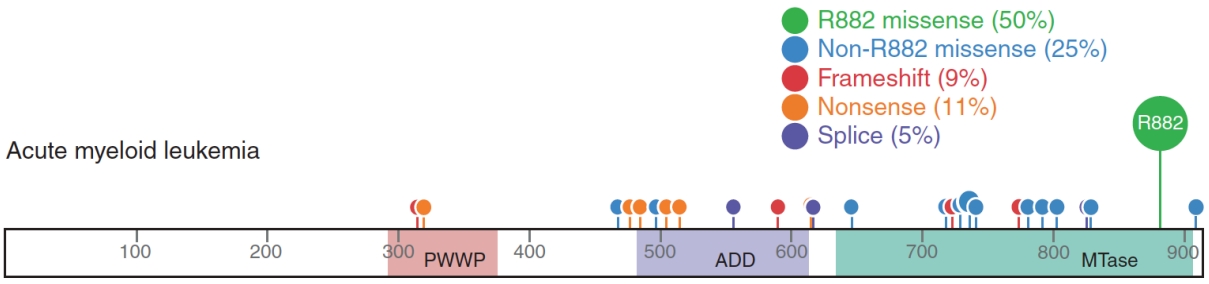



**Figure 9: *DNMT3A* mutations in AML.** The shown mutations of DNMT3A in AML are based on the TCGA database and are mapped onto DNMT3A1. The database includes 57 different mutations appearing in DNMT3A derived from 51 individuals. Nonsynonymous mutations are depicted as lollipops, where the color stands for the type of mutation and the size correlates with the mutation count (Adapted from Brunetti et al., 2017).

Since the first description of *DNMT3A* mutations in AML, multiple exome and targeted resequencing studies have identified *DNMT3A* mutations in this disease (Ivey et al., 2016; Klco et al., 2015; Ley et al., 2010; Shlush et al., 2014). *DNMT3A* mutations could be detected in around 20% of the patients with *de novo* AML, as well as in secondary AML, which evolves from previous myelodysplastic syndromes or myeloproliferative neoplasms (Fried et al., 2012; Gaidzik et al., 2013; Ley et al., 2010; Thol et al., 2011). Interestingly, these mutations are not randomly distributed over the coding sequence of *DNMT3A*, but rather cluster in the catalytic domain of the enzyme, with very high occurrence at codon R882 (**Figure 9**) (Brunetti et al., 2017). Although non-R882 missense and truncating mutations are found in each of the major domains (Ley et al., 2013; Roller et al., 2013), circa 65% of mutations are heterozygous missense

mutations at R882. The most frequent exchange seems to be R882H. Since this mutation correlates with a poor prognosis for the patients, several groups have mechanistically addressed the consequences of this exchange for the function of DNMT3A (Brunetti et al., 2017; Marcucci et al., 2012), but these studies have led to conflicting reports.

Reich and co-workers (2012) proposed that the R882H mutation disrupts the dimerization interface of the enzyme, causes a 10-fold increase in the dissociation rate of DNMT3A from DNA and as a result leads to a decrease in enzymatic activity *in vitro* (Holz-Schietinger et al., 2012). The authors proposal that this mutation interferes with the processivity of DNMT3A, is not in agreement with the cooperative methylation model proposed by other groups (see above) (Jeltsch and Jurkowska, 2013; Rajavelu et al., 2012). In line with the findings of Holz-Schietinger et al. (2012), work by Kim et al. (2013) showed that co-expression of DNMT3A R878H (the mouse equivalent of R882H) impaired the enzymatic activity of wildtype DNMT3A and DNMT3B leading the authors to propose a dominant-negative role for this mutation (Kim et al., 2013). Interestingly, Russler-Germain and colleagues (2014) showed that while the *in vitro* mixing of wildtype and R882H DNMT3A did not affect the activity of the wildtype enzyme, the co-expression of the two proteins in cells disrupted the ability of the DNMT3A to heterotetramerize, indicating that R882H acts in a dominant negative manner (Russler-Germain et al., 2014).

Of note, while also other groups have observed a global hypomethylation of islands and shores associated with the R882H mutation, hypermethylation of CpG promoter regions has also been associated with the same mutation in AML (Qu et al., 2014; Yan et al., 2011). This, together with the fact that some AML patients successfully respond to the demethylating agent decitabine, contradict the dominant-negative model and rather argues for a model where the mutation causes alterations and not reductions in the DNA methylation patterns (Chowdhury et al., 2015; Metzeler et al., 2012). More mechanistic studies are needed to elucidate the origin and impact of the R882H mutation on the DNA methylation landscape. In this thesis, the effect of the R882H mutation on the DNMT3A function was mechanistically addressed.

# 2. Principal aims of the study

Since 1948, when Hotchkiss discovered DNA methylation with the help of paper chromatography, this epigenetic signal has become the subject of intense investigations. Although the importance of DNA methylation was already highlighted by early genetic studies and subsequently demonstrated in decades of active research, detailed knowledge on the mechanisms that are in place to control the setting and maintenance of DNA methylation patterns is still sparse (Jurkowska and Jeltsch, 2016). Indeed, while mammalian DNMTs were identified around 2 decades ago (Bestor et al., 1988; Okano et al., 1998) our understanding of their exact mode of action, as well as their regulation *via* chromatin modifications and interaction partners is still insufficient (Jurkowska and Jeltsch, 2016; Li et al., 1992; Okano et al., 1998, 1999). Furthermore, little mechanistic understanding is available on how alterations of the DNA methylation machinery, as the mutations observed in some types of cancers, influence the function of these enzymes (Yang et al., 2015).

The aims of this work were to explore four main directions of research, with the ultimate goal of provide novel mechanistic insights into the mode of action of the *de novo* DNA methyltransferase 3A, the regulation of this enzyme by chromatin signals and interaction partners, as well as obtaining a deeper understanding of the dysregulation of this enzyme in cancer.

The DNA methyltransferase DNMT3A has been shown to multimerize on DNA and to form large multimeric protein/ DNA fibers (Jeltsch and Jurkowska, 2013). However, it has also been reported that this enzyme can methylate DNA in a processive manner (Holz-Schietinger and Reich, 2010). These properties appear mutually exclusive because the concept of a processive turnover is based on individual enzyme molecules moving along a DNA substrate. It was one aim of this project was to investigate the functional role of DNMT3A multimerization on DNA and discriminate between these two distinct modes of action by dedicated biochemical experiments.

Cancers acquire somatic DNA mutations as well as epigenetic changes, which ultimately confer a growth advantage to the affected cells (Baylin and Jones, 2011). These two mechanisms of carcinogenesis are intimately connected in cases when

mutations occur in epigenetic enzymes like DNA and histone methyltransferases and demethylases (Flavahan et al., 2017; Hattori and Ushijima, 2014). The heterozygous R882H mutation in the DNA methyltransferase DNMT3A occurs frequently in patients diagnosed with the hematological disorder AML and it has a role in early tumorigenesis. Models propose that the R882H mutation has a dominant negative effect in complex with wildtype DNMT3A and leads to a substantial reduction of activity (Yang et al., 2015). The aim of this project was to biochemically address the effect of the R882H mutation on the function of DNMT3A and better understand the mechanistic details of the proposed dominant negative effect. To this end protein stability measurements were to be performed. To this end, it was planned to establish a double-tag affinity purification workflow, which would enable the direct purification of wildtype-R882H heterotypic complexes. Since the mutation is located at the RD interface, where the DNA binding cleft is formed, it was planned to conduct a flanking site specificity analysis.

While many studies have shown that DNA methylation is a vital modification, which plays essential roles in embryonic development and tissue homeostasis, the mechanisms that are in place to control the chromatin targeting and activity of DNMT enzymes are still not completely understood (Laisné et al., 2018). Previous work from our lab have identified the 5mC-reading protein MECP2, as a direct and strong interactor of DNMT3A and mapped the interaction interface to the TRD domain of MECP2 and the ADD domain of DNMT3A. In *in vitro* DNA methylation assays it was found that through its docking, MECP2 allosterically inhibited the activity of DNMT3A. Interestingly, addition of unmodified histone H3 was found to disrupt the complex and consequently relieve the enzymatic inhibition. One aim of this work was to further validate the interaction between the TRD and ADD and to investigate the effect of MECP2 on DNMT3A in a cellular context. For this, it was planned to generate cell lines with stable and doxycycline inducible expression of MECP2 and study their DNA methylation.

During the course of this work, our lab was contacted by Dr. Peter Sarkies (MRC London) who had uncovered a fascinating co-evolutionary relationship between DNA methylation and the DNA alkylation repair enzyme ALKB2 across eukaryotes. Interestingly, the preferred substrates for ALKB2/3 in DNA are 1mA and 3mC (Nay et

al., 2012; Ringvoll et al., 2006). Based on this, one aim of my work was to perform *in vitro* DNA methylation assays with the catalytic domain of recombinantly purified DNMT3A to investigate if DNMT3A can introduce 3mC in addition to 5mC.

# 3. Materials and Methods

## 3.1 Cloning, site-directed mutagenesis, protein expression and purification

The sequences encoding for the DNMT3 proteins and their variants used in this study were either already available in the lab or were newly cloned in the pMAL p2X backbone (New England BioLabs) as an N-terminal MBP-tag fusions using the Gibson assembly approach (Gibson et al., 2009). All constructs used in the study can be found in **Table 1**. Point mutations of interest were generated using the megaprimer based site directed mutagenesis method (Jeltsch and Lanio, 2002). The integrity of all constructs was confirmed by DNA sequencing.

BL21 (DE3) Codon+ RIL *Escherichia coli* cells (Stratagene) were used for protein overexpression. Shortly described, cells were grown until an OD 0.6 in terrific broth (TB) medium and protein expression was induced at 20°C for 12h by addition of 0.5 mM isopropyl-1-thio-ß-D-galactopyranoside.

Protein purification was performed by affinity chromatography using the appropriate matrix molecules and the proteins were stored in 20mM HEPES at pH 7.5. 200mM KCl, 0.2mM DTT, 1mM EDTA and 10% glycerol at -80°C. The Concentration and purity of each preparation were determined by UV absorption at 280nm and Coomassie BB stained SDS-polyacrylamide gels.

More detailed information, like details of the purification of mixed DNMT3AC/R882H complexes is available in the attached papers (**Appendix 3**).

**Table 1: Constructs used in this study:** Listed with additional information about backbone, domain, accession number (NCBI), boundaries, and tag used for purification or identification.

| Insert | Vector Backbone | Domain | Accession Number | Boundaries [aa] | Tag |
|---|---|---|---|---|---|
| Dnmt3aC | pET28+ (Novagen) | Cat. domain | Q9Y6K1 | 612-912 | His |
| Dnmt3aC | pMAL p2X (NEB) | Cat. domain | Q9Y6K1 | 612-912 | MBP |
| Dnmt3aC R882H | pET28+ (Novagen) | Cat. domain | Q9Y6K1 | 612-912 | His |
| Dnmt3aC R882H | pMAL p2X (NEB) | Cat. domain | Q9Y6K1 | 612-912 | MBP |
| Dnmt3aC R885A | pET28+ (Novagen) | Cat. domain | Q9Y6K1 | 612-912 | His |
| Dnmt3aC F732D | pET28+ (Novagen) | Cat. domain | Q9Y6K1 | 612-912 | His |
| Dnmt3a2 | pET28+ (Novagen) | Isoform 2 | NP_001258682 | 220-908 | His |
| Dnmt3a2-R878H | pET28+ (Novagen) | Isoform 2 | NP_001258682 | 220-908 | His |
| Dnmt3aC | pET28+ (Novagen) | Cat. domain | O88508 | 608-908 | His |
| Dnmt3aC E752A | pET28+ (Novagen) | Cat. domain | O88508 | 608-908 | His |
| Dnmt3bC | pET28+ (Novagen) | Cat. domain | O88509 | 551-859 | His |
| Dnmt3bC K887R | pET28+ (Novagen) | Cat. domain | O88509 | 551-859 | His |
| Dnmt3bC G883L | pET28+ (Novagen) | Cat. domain | O88509 | 551-859 | His |
| Dnmt3bC G881S | pET28+ (Novagen) | Cat. domain | O88509 | 551-859 | His |
| Dnmt3a ADD | pMAL p2X (NEB) | ADD domain | NP_001258682 | 472-610 | MBP |
| MeCP2-TRD | pGEX-6p2 | TRD domain | NP_004983 | 170-325 | GST |
| MeCP2-EYFP | TRE3G-PGK-NEO | Isoform 2 | NP_004983 | 1-486 | EYFP |
| M.SssI | pBAD24 | Fulllength | P15840 | 1-386 | His |

## 3.2 Methyltransferase activity assay

The biotin-avidin microplate assay was used to measure the *in vitro* activity of the DNMTs (Liebert and Jeltsch, 2008; Roth and Jeltsch, 2000).

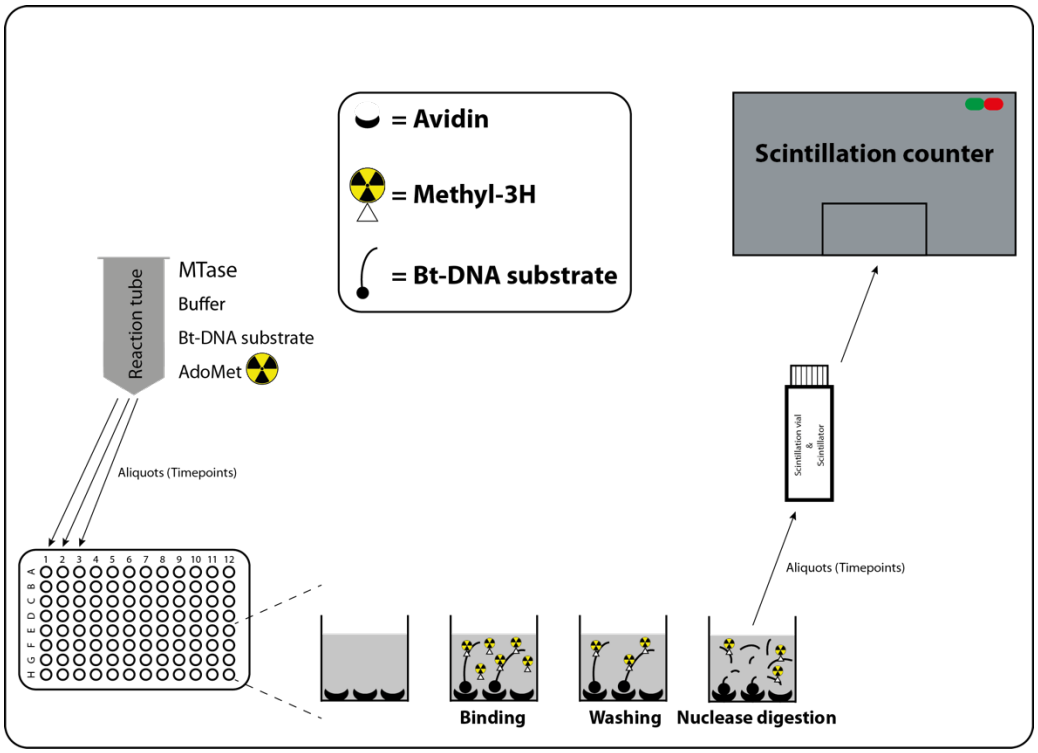



**Figure 10: Schematic illustration of the biotin-avidin microplate assay** (Figure adapted from Monti 2012).

This assay takes advantage of the high binding affinity between biotin and avidin. A schematic illustration of the workflow can be seen in **Figure 10.** In this study, two main substrates were used a 509 bp long PCR fragment originally amplified from a CpG island upstream of human *SUHW1* (Suppressor of hairy wing homolog 1) gene containing 58CpG sites and a 30 bp long double stranded oligo containing just one CpG site as described in **Appendix 1, 2 & 3**. A microplate coated with avidin is used for the selective binding of biotinylated DNA. The methylation of DNA is detected by the incorporation of titrated methyl groups from radioactively labeled Cofactor SAM (S-Adenosyl-L-methionine) (Perkin Elmer) into the biotinylated substrate. For project specific details on the composition and sequence of the substrates please see **Table 2** and **Appendix 1, 2 & 3.**

**Table 2: Setup of in vitro DNMT3 activity based experiments**: Detailed list of specific experimental approaches used in this study based on enzymatic kinetics.

| Approach/concept | Protein(s) | Substrate(s) | Add. Information | Reference |
|---|---|---|---|---|
| Standard Kinetics | 1 or 2 µM Enzyme | 30-mer or 509-mer | TP: 2, 4, 8 und 16min | Appendix 1<br>Appendix 2<br>Appendix 3 |
| Catalytic activity of Dnmt3a-C on the 509-mer-longDNA substrate and on the 30-mer oligonucleotide. | 3aC<br>0.25, 0.5, 0.75, 1, 1.5, and 2 µM | 250nM 30mer 100nM 509mer | | Appendix 1 |
| Stimulation of DNA methylation of Dnmt3a-C WT by the addition of the catalytically inactive Dnmt3a-C E752A variant | 1µM 3aC (1W), 1µM 3aC & 1µM E752A (1W1M) and 2µM 3aC (2W) | 1µM 30mer | | Appendix 1 |
| Activity of Dnmt3a-C on the 509-mer and 30-mer substrates after preincubation and chase with competitor substrate | 1.5 µM 3aC | 1 & 1 µM 30mer (+/- Bt)<br>50 & 50 nM 509mer (+/- Bt) | | Appendix 1 |
| Methylation of the 1-site and 2-site substrates at diff. Concentrations of Dnmt3aC | 3aC<br>0.0625, 0.25, 0.5, 1 and 2 µM | 30-mer containing either one or two CpG sites in the same flanking context | | Appendix 1 |
| Protein stability of DNMT3AC and R882H analyzed by the loss of enz. Activity a pre-incubation the protein for 30min at diff. Temperatures. | 2 µM DNMT3AC<br>2µM DNMT3AC-R882H | 100nM 509-mer | Master-Mix: One tube started direcrtly the second tube was kept at 22°C or 37°C for 30min before starting | Appendix 2 |
| Design of a preferred R882H substrate. | 1µM DNMT3AC<br>1µM DNMT3AC-R882H | 30-mer containingt 1 CpG site either preferred or disfavored by R882H and in a hemimethylated or fully methylated version | | Appendix 2 |
| Characterization of activity of DNMT3AC R882H homomultimers | 1 µM 3AC, 1µM R882H<br>&<br>0.5 µM 3AC & 0.5 µM R882H mixed | 30-mer | | Appendix 3 |
| Purification and activity analysis of mixed R882H/wildtype DNMT3A complexes | 1 µM 3AC-His/3AC-MBP Complex<br>1 µM 3AC-His/R882H-MBP Complex<br>1 µM R882H-His/3AC-MBP Complex | 30-mer | | Appendix 3 |
| Effect of the RH mutation in the context of DNMT3A2 | 1 µM 3A2, 1µM RH<br>&<br>0.5 µM 3A2 & 0.5 µM RH mixed | 30-mer | | Appendix 3 |

## 3.3 Bisulfite sequencing and DNA methylation analysis

For this experimental approach, the 509 bp PCR fragment containing 58 CpG sites described in **section 3.2** was used. The reaction was performed in methylation buffer using 100nM substrate and different concentrations of enzyme as described (Zhang et al., 2009b) . Each reaction was stopped by flash-freezing in liquid nitrogen and the DNA was isolated and bisulfite converted using the EZ DNA Methylation Kit (ZYMO RESEARCH, Freiburg). The converted DNA was then amplified using bisulfite specific

primers (GTG GGA TTT GGT TTT GTT TTG TAT TTT and GAA TAT TAC TAC TAC CCT CCT TCT CAA TTT AAC). PCR products were then subcloned using the StrataClone PCR cloning kit (AGILENT, Santa Clara, USA) individual clones were sequenced and the results were analyzed using the web based BISMA software (**Appendix 1 & 2**; Rohde et al., 2010).

## 3.4 Circular dichroism (CD) spectroscopy analysis

CD was used to analyze the folding of the purified proteins. The CD spectroscopy measurements were performed using a circular dichroism spectrophotometer (J-815, JASCO Corporation, Tokyo, Japan). The measurements were performed in a buffer containing 200 mM KCl, 20 mM HEPES pH 7.5, 10% glycerol and 10 µM protein. The spectra were collected at 21°C in a wavelength range of 195-240 mm using a 0.1 mm cuvette as described in **Appendix 2**.
For the CD melting experiment, the protein was kept in storage buffer. The signal was measured at 222nm in a 0.1mm cuvette in a temperature range from 20 to 70°C with an increase of 2°C/min. The determination of the melting temperature was performed as described (Greenfield, 2007).

## 3.5 Western blot

Protein samples were run on 12 or 15% SDS gels with a constant conduction current of 35mA. In next step, the proteins were blotted on a nitrocellulose membrane using a wet tank blotting system (Bio-Rad, USA). Next, the membranes were blocked in 5% milk in PBST for 1h at RT and incubated with the primary antibody (Ab) of interest over night at 4°C. For more details on project specific experimental approaches please see **Appendix 1 & 4**.

## 3.6 Quantitative reverse transcription PCR

This method was used for gene expression analysis. The assays were performed using SsoFast EvaGreen supermix (Bio-Rad, Hercules, CA USA) run on a CFX96 Connect Real-Time detection system (Bio-Rad, Hercules, CA, USA). For each sample, RNA

from $10^6$ cells was isolated using RNeasy Mini Kit (Qiagen, Limburg, Netherlands). The cDNA preparation was performed using MultiScribe™ Reverse Transcriptase (Applied Biosystems, Thermo Fisher, USA) with oligo d(T) 18 primers (New England Biolabs, Ipswich, MA, USA) and 500 ng of RNA. The specific primer sets used for the targets of interest are shown in **Table 3**. For normalization, the housekeeping gene SDHA (TGG GAA CAA GAG GGC ATC TG and CCA CCA CTG CAT CAA ATT CAT) was used. For more information please see the corresponding publication (**Appendix 4)**

**Table 3: Quantitative RT-PCR primer sets:** Detailed list containing infomation like the mRNA accession number, amplicon size and the sequence of each primer.

| Primer Sets | mRNA accession number | Target | Amplicon size (bps) | Sequence |
|---|---|---|---|---|
| qRT-Dnmt3a_Forward | NM_175629 | DNMT3A | 91 | CGATTTCTCGAGTCCAACCCTG |
| qRT-Dnmt3a_Reverse | | | | TACCGGGAAGGTTACCCCA |
| qRT-Dnmt3b_Forward | NM_006892 | DNMT3B | 125 | CAGTGACACGGGGCTTGAATATG |
| qRT-Dnmt3b_Reverse | | | | CTTTGAGGACTAGGTAGCCTGTCGCG |
| qRT-Dnmt1_Forward | NM_001130823 | DNMT1 | 159 | GAGACACGATGTCCGACCTG |
| qRT-Dnmt1_Reverse | | | | CCAATGCACTCATGTCCTTACAG |
| SDHA_Forward | NM_004168 | succinate dehydrogenase complex flavoprotein subunit A | 86 | TGGGAACAAGAGGGCATCTG |
| SDHA_Reverse | | | | CCACCACTGCATCAAATTCAT |

## 3.7 Generation of HCT116 DNMT1 hypomorphic cells with stable expression of MECP2 and mVenus

The HCT116 DNMT1 hypomorphic cells (Rhee et al., 2000) were cultured at 37°C under a $CO_2$ atmosphere of 5% in McCoy 5A medium (Sigma) supplemented with 10% FCS, 2mM L-glutamine, 100 U/ml penicillin and 100 ug/ml streptomycin. In the first experimental step the cells got modified to express the ecotropic receptor EcoR and rtTA3 were introduced using retroviral transduction followed by a drug selection as described (Rathert et al., 2015). In the second step this newly generated EcoR positive cell line was transduced with ecotropically packaged retroviruses produced in PlatiniumE cells (Cell Biolabs) as described (Fellmann et al., 2013). All constructs used in this study can be found in **Table 1**. To achieve single copy insertions than 20% of the initial population were infected. 48h after infection the cells were selected for 7 days using 500 µg/ml G418 (Gibco Life technologies, Thermo Fisher, USA). For all information belonging to this experimental approach please see **Appendix 4.**

## 3.8 Amplified Luminescent Proximity Homogenous assay

This method relies on the use of donor and acceptor beads to which the protein of interest can be bound (e.g. 3AC-His & 3AC-R882H-MBP). When the 2 proteins interact, the donor and acceptor beads are brought in close proximity.Upon laser excitation of the donor beads, $O_2$ is converted to a singlet state which can diffuse over a max. distance of 200 nm and induce chemiluminescence in the acceptor beads. This further activates a fluorophore that emits light in the 520-620 nm range (Eglen et al., 2008). A schematic illustration of the assay system is provided in **Figure 11**.

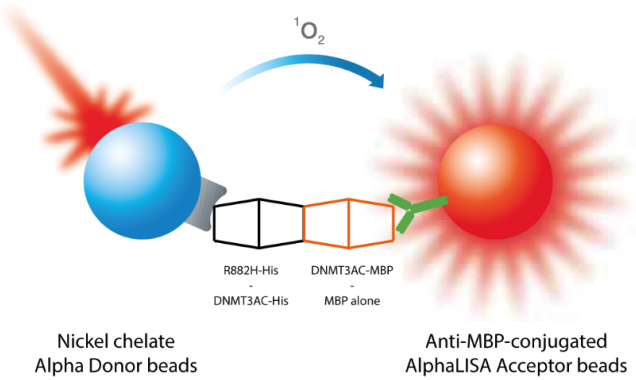



**Figure 11: Alpha assay principle:** Demonstration chemical concept behind the alpha assay modified from PerkinElmer (Figure adapted from Eglen et al., 2008).

In this study, the method was used to observe the exchange of subunits of DNMT3AC and DNMT3AC/3L complexes and show the formation of heterotypic complexes. For this nickel chelate donor beads (PerkinElmer) for His-tagged proteins and anti-MBP AlphaLISA acceptor beads (PerkinElmer) for MBP-tagged proteins were used.
For more details information about the experimental settings please see **Appendix 3**.

## 3.9 Size exclusion chromatography

This approach can be used to demonstrate if two recombinant proteins are able to form a stable complex. To address a potential complex formation, a Superdex$^{TM}$ 200 10/300 GL column (GE Healthcare) was used. The column was connected to an NGC Quest$^{TM}$ 10 system from Bio-Rad. For calibration, the gel filtration calibration kit HMW (GE Healthcare) was used. To visualize the results as chromatograms Chromlab 5.0 was

used. The total run was collected in fractions of 500µl and which afterwards were separated on 12% SDS gels and analyzed using western blot. For more detailed information please see **Appendix 4.**

# 4. Results

## 4.1 Mechanistic investigation of the mode of DNMT3A action *in vitro*

Seminal structural work has shown that in complex with its regulatory factor, Dnmt3L, the catalytic domain of DNMT3A can form a linear heterotetramer in which two central DNMT3A molecules are flanked by two Dnmt3L units (Jia et al., 2007). The formation of the heterotetramer is mediated by two distinct and critical interfaces: a 3A–3A central 'RD' interface, which generates the DNA binding surface, and two peripheral 3A–3L 'FF' interfaces (Jia et al., 2007). If DNMT3L is present, DNMT3A can only polymerize along the DNA to give rise to large multimeric protein/DNA fibers, but if DNMT3L is absent, DNMT3A can additionally polymerize vertically leading to large multimeric protein/DNA complexes. Several lines of evidence have demonstrated that the formation of the protein/DNA fibers is determined by the cooperative binding of DNMT3A to DNA (Jurkowska et al., 2011b). Conversely, other reports have proposed that DNMT3A methylates DNA in a processive manner, a property which is antagonistic with protein/DNA fiber formation. In the course of this project, these conflicting models were addressed by investigating the functional role of DNMT3A multimerization. The results of this work were described in a paper published in the Journal of Biological Chemistry and attached here as **Appendix 1**. The main findings of this study are succinctly described below.

In all *in vitro* experiments the catalytic domain of DNMT3A was used, entitled as DNMT3AC. *In vitro* methylation experiments using DNMT3AC and two different DNA substrates, a 30bp-short double stranded oligodeoxynucleotide have indicated that DNMT3AC shows an increased methylation rate on the long substrate (**Figure 12A**). Based on this observation it was hypothesized that the long DNA substrate but not the short one supports the multimerization of DNMT3AC. To address this in more detail, *in vitro* methylation experiments were performed with increasing concentrations of DNMT3AC to determine the increase of the methylation rate on the short and on the long substrate (**Figure 12A**). The more than linear increase in the methylation rate at increasing concentrations of DNMT3AC on the long DNA substrate seen in **Figure 12A**

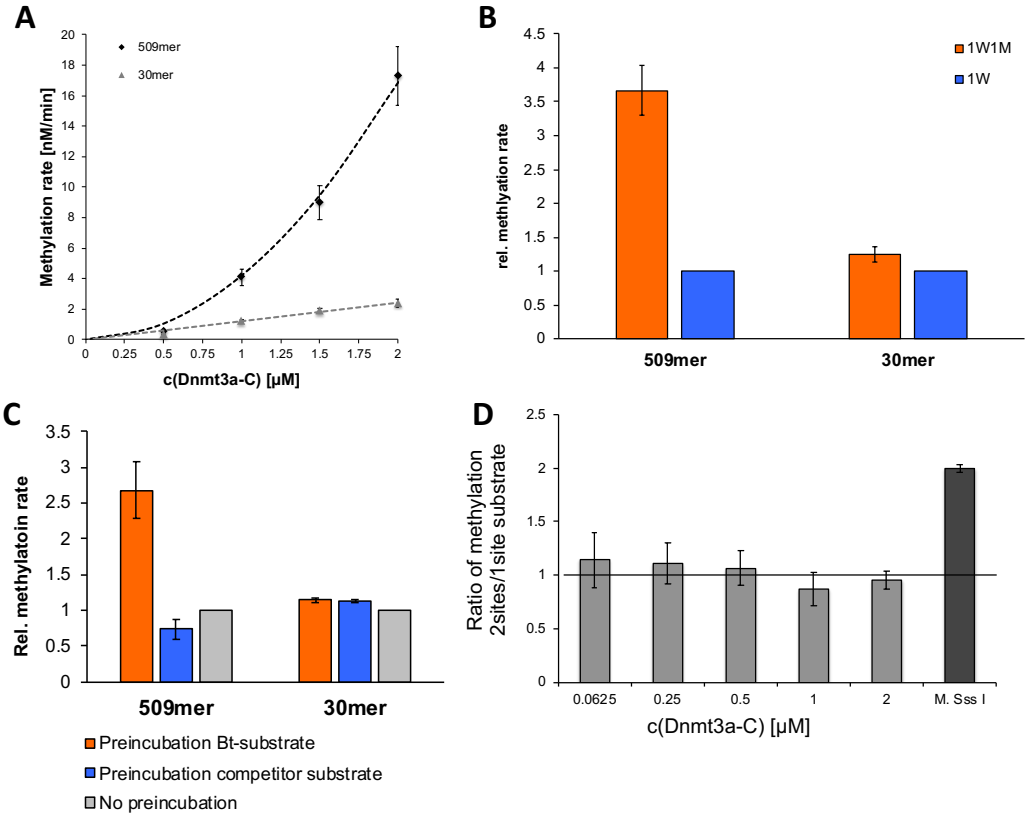



**Figure 12: Key activity assays of DNMT3AC. A** Catalytic activity of DNMT3AC with increasing enzyme concentration with the 509mer-long DNA (0.1 µM) substrate (dark gray diamonds) and with the oligonucleotide (0.25 µM) (light gray triangles). All reaction mixtures were preincubated for 25 min at room temperature before the starting the reaction by addition of SAM (S-Adenosyl-L-methionine). The light gray line shows a linear regression of the methylation rates. The dark gray line, included for illustration, shows a second-order polynomial fit of the 509mer methylation rates. **B** Stimulation of DNA methylation of DNMT3AC wildtype (1 µM) by the addition of the catalytically inactive DNMT3AC E752A variant (1 µM) with the long 509mer substrate (0.1 µM). **C** Activity of DNMT3AC on the 509mer (0.1 µM)) and 30mer (2 µM) substrates after preincubation & chase with competitor substrate. All reaction mixtures contained half of the substrate in a biotinylated form and the other half unbiotinylated. Experimental setup: (i) Preincubation with biotinylated substrate, afterwards addition of unbiotinylated competitor substrate and starting of the reaction(Orange). (ii) Preincubation with unbiotinylated competitor substrate afterwards addition of the biotinylated substrate and starting of the reaction (Blue). (iii) 25min preincubation without any substrate, then addition of both substrates together and starting of the reaction (Light grey). **D** Methylation of the 1-site and 2-sites substrates at different concentrations of DNMT3AC. Sample incubation time 60 min, and the final incorporation of radioactivity was determined using 0.25 µM oligonucleotide as substrate. M.SssI served as a control experiment confirming a 2-fold higher incorporation of radioactivity into the 2-sites substrate when compared with the 1-site substrate (dark gray bar). For more details please refer to **Appendix 1**.

can be explained by the cooperative binding of DNMT3AC to DNA resulting in a sigmoidal DNA binding curve. Based on this finding the conclusion was that the multimerization of DNMT3AC on the long DNA substrate is stimulating the rate of DNA methylation. As mentioned above, the catalytic domain of DNMT3A forms a heterotetramer. It was thereby interesting to test if the addition of a catalytically inactive mutant of DNMT3A would also lead to an increased rate of DNA methylation. To this

end, *in vitro* methylation kinetics using the long DNA substrate and following reaction setups were performed: 1µM wildtype (1W) alone, 1µM wildtype and 1µM of an inactive variant DNMT3AC-E752A (1W1M) and 1µM DNMT3AC-E752A alone. As shown in **Figure 12B** addition of the catalytically inactive DNMT3AC caused a strong, 4-fold stimulation of wildtype DNMT3AC activity. This result was very striking since it was expected that due to competition for the DNA substrate and the cofactor SAM, the addition of DNMT3AC-E752A would reduce the activity of the wildtype enzyme. This extraordinary finding can be nevertheless easily explained by the cooperative DNA binding and multimerization of DNMT3AC on DNA.

A consequence of the formation of stable protein/DNA fibers would be the reduction of the dissociation rate of DNMT3AC from DNA. To address this hypothesis preincubation/chase experiments were conducted using either the oligonucleotide or the 509mer as a substrate. Both substrates were used in biotinylated form, for retrieval via the avidin-coated plated, or unbiotinylated state as competitor DNA. Three different experimental setups were conducted in parallel (**Figure 12C**). In the first setup, DNMT3AC was preincubated with the biotinylated substrate, the unbiotinylated competitor substrate was added before starting the reaction with radioactively labeled SAM. The second setup was prepared in the opposite mixing order. In the third setting DNMT3AC was preincubated without any substrate and just before starting the reaction both substrates were added. The reaction rate was determined over 16min. In case of the oligonucleotide substrate all three settings showed the same results, which is an indication for a rapid binding equilibrium between DNMT3AC and both labeled and unlabeled substrates (**Figure 12C**). In contrast, for the 509mer substrate, it could be clearly seen that for the setup where DNMT3AC was preincubated with the biotinylated substrate, the methylation rate was ~ 3 fold higher than for the other two reaction mixtures. This indicates that the acceleration of methylation is based on the binding of DNMT3AC molecules on DNA. Looking at the results of setting two, the methylation rate of the biotinylated substrate was reduced, indicating that some DNMT3AC molecules were trapped on the competitor substrate. Taken together, these results demonstrate the stability of the formed protein/DNA fiber on long DNA. Importantly, these fibers cannot form on the oligonucleotide substrate since this does not provide enough space for the formation of a stable DNMT3AC multimers.

All results described above together with other experiments shown in **Appendix 1** strongly support the hypothesis that DNMT3A cooperatively binds to DNA and this

results in an increased methylation activity at higher enzyme concentrations. To finally exclude the postulated processivity of DNMT3A and to also address this point from another side, an experiment was conducted using two different substrates one with one CpG site and one with two CpG sites, both in an identical flanking sequence context. Their methylation was measured at different DNMT3AC concentrations and a fixed concentration of 0.25 µM of the substrate, which corresponds to either 0.5 µM target cytosines in case of one CpG site and 1 µM in case of the two site substrate. In case of processivity the second target site in the two site substrate should be methylated faster than the first one, which should translate into an increased rate of methyl group transfer to the two site substrate. As it can be seen in **Figure 12D** the ratio between the 2 sites and the 1 site substrate is around 1 at any concentration of DNMT3AC indicating that processivity is undetectable. As a positive control, M.SssI which is known to be a processive enzyme was used (**Figure 12D**, dark gray bar) and M.SssI clearly showed an increased methylation rate on the two site substrate.
To summarize, the data presented here (**Appendix 1**) are perfectly consistent with a cooperative DNA binding and fiber formation model of DNMT3A and they are not compatible with a processive DNA methylation model.

## 4.2 Flanking sequence preference analysis of the AML relevant DNMT3A mutant R882H

In the COSMIC database (Catalogue Of Somatic Mutations In Cancer, (https://cancer.sanger.ac.uk/cosmic) the mutation of arginine (R) 882 to histidine (H) in the catalytic domain of DNMT3A is by far the most frequent mutation reported for the enzyme. Interestingly, although this mutation was repeatedly associated with the development of AML (acute myeloid leukemia), the molecular and biochemical mechanism of leukemia initiation and progression through this mutation has not been elucidated. In the course of this project, the mechanistic consequences of the DNMT3A R882H mutation on the functions of the enzyme were analyzed. The results of this work were described in a paper published in Nucleic Acids Research and attached here as **Appendix 2**. The main findings of this study are summarized below.

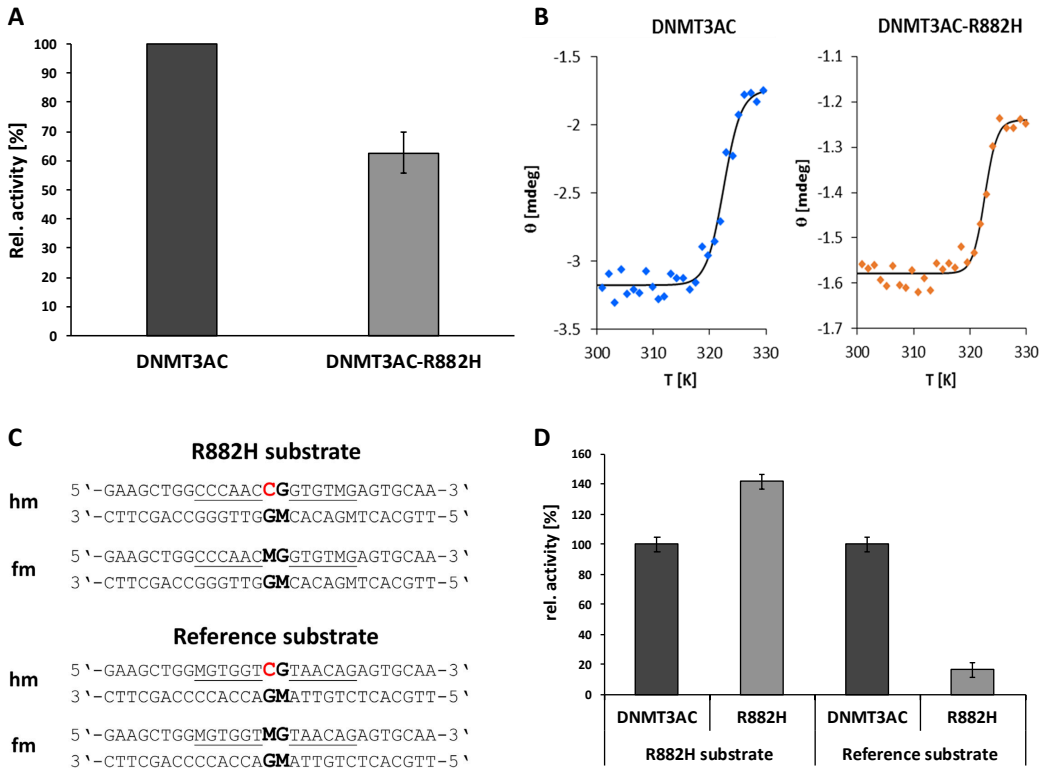



**Figure 13: Characterization and flanking sequence preference of the R882H mutant. A** DNA methylation activity of DNMT3AC and the R882H mutant using a 509mer DNA substrate. The bars represent the average of 5 experiments, error bars indicate the standard error of the mean. **B** Thermal stability of DNMT3AC and R882H analyzed by CD melting at protein concentrations of 22 µM for DNMT3AC and 8 µM for R882H. Quantitative analysis revealed identical melting temperatures of 49.3 °± 0.4◦C and 49.5 °± 0.6◦C for DNMT3AC and R882H. **C** Based on the sequences of the 10 most preferred methylation sites by R882H, a substrate with R882H optimized flanks was designed in paralel with a reference substrate not containing R882H preferred residues. For each substrate, a hemimethylated version (hm) and a fully methylated (fm) version was used to allow measurement of the specific methylation activity of the central CpG site in the upper DNA strand. 'M' denotes 5-methyldeoxycytosine, the target C is highlighted in red. **D** Average target site methylation rates of the R882H substrate and reference substrate by wildtype DNMT3AC and DNMT3AC-R882H. To determine the specific methylation of the central CpG site in the upper DNA strand, the methylation of the fm substrate is subtracted from the methylation of the hm substrate. Error bars indicate the SEM based on four independent experiments. For more details please refer to **Appendix 2**.

The R882H mutation is located at the RD interface of DNMT3A, where the DNA binding interface is formed. To assess whether this mutation has an influence on the activity of the enzyme, *in vitro* methylation experiments were performed using the catalytic domain of DNMT3A (DNMT3AC) and a 509bp-long substrate derived from the CpG island upstream of the human *SUHW1* (Suppressor of hairy wing homolog 1) gene. As documented in **Figure 13A**, DNMT3AC-R882H displayed ca. 40% less activity than the wildtype enzyme. To investigate if the mutational change from R to H at position 882 potentially influences protein stability, CD melting experiments were conducted. The results showed identical melting temperatures for DNMT3AC-R882H ($T_m$ = 49.3 ±

0.4°C) and the wildtype DNMT3AC ($T_m$ = 49.5 ± 0.6°C) as it can be seen in **Figure 13B**.

The fact that the mutation R882H is located in the DNA binding interface of DNMT3A and that is already known that DNMT3A shows different methylation activity of CpG sites depending on their flanking sequence context (Handa and Jeltsch, 2005; Jurkowska et al., 2011c; Lin et al., 2002; Wienholz et al., 2010) raised the question if this mutation might influence the flanking sequence preference of DNMT3A. To address this point methylation kinetics were conducted using the 509mer substrate, which contains 56 CpG sites in different flanking contexts. After methylation with DNMT3AC or DNMT3AC-R882H, the substrate was bisulfite converted and sequenced. Comparison of the relative methylation levels of different CpG sites between wildtype DNMT3AC and R882H uncovered a strong change in the flanking sequence preference for the mutant. Certain sites showed a much stronger reduction of methylation activity than the general 38% reported before. Much more striking was the fact that particular sites were methylated up to 5 times more efficiently by R882H mutant than by the wildtype DNMT3AC. These particular sites were observed to be weakly methylated by the wildtype enzyme but still methylated with good yields by the mutant. Based on this result the -6 to +6 flanks of the 10 sites most preferred by DNMT3AC-R882H were used to design a substrate ideal for the mutant. For this, the base of each flanking position in the preferred sites was compared to the overall distribution in all sites. This statistical analysis indicated clearly that the R882H mutant differs substantially in its flanking sequence preference from the wildtype DNMT3AC.
Based on this analysis model substrates were designed containing one CpG site in the center with flanks optimized for the methylation of the upper DNA strand by R882H ("R882H substrate") and flanks without features preferred by R882H ("reference substrate") (**Figure 13C**). To investigate specifically the methylation of the central CpG site in the upper strand, the designed substrates were used in hemimethylated form. Because it is known that DNMT3A also methylates other cytosine residues in a non CpG context (Arand et al., 2012; Gowher and Jeltsch, 2001; Jurkowska et al., 2011c; Ramsahoye et al., 2000), an additional control setup for both designed substrates were used, where the central CpG sites is methylated in both strands (fully methylated). To extrapolate the methylation activity at the central CpG site in the upper strand the activity measured with the fully methylated substrates was subtracted from the activity

measured with the corresponding hemimethylated substrates. Using this system, the R882H mutant showed 45% more activity than the wildtype DNMT3AC on its “R882H substrate”. By contrast, less than 15% activity, when compared to wildtype, was observed when the “reference substrate” was used (**Figure 13D**). Together with the data included in **Appendix 2**, these results demonstrate that it is possible to design an R882H preferred and disfavored substrate and uncovered that the consequence of the R to H mutation at position 882 is a strong change in the flanking sequence preference of DNMT3A.

## 4.3 Mechanistic investigation of the postulated dominant negative effect of DNMT3A-R882H

The DNMT3A R882H mutation is observed in 25% of all AML patients. This mutation occurs usually heterozygous, with one wildtype and one affected allele. The fact that R882 is located in the center of the RD interface of the enzyme, which forms the DNA binding pocket, gives rise to the hypothesis that the R882H exchange might affect DNTM3A tetramerization and/or DNA binding. During the course of this project, mixed wildtype/R882H complexes were used to investigate in *in vitro* assays the influence of this mutation on DNMT3A tetramerization. The main findings of this work were described in a manuscript published in Scientific Reports and attached here as **Appendix 3**.

A previous publication reported that mixed DNMT3A wildtype/R882H complexes had a strongly reduced enzymatic activity in *in vitro* DNA methylation assays (Russler-Germain et al., 2014). Based on these data, a dominant negative effect of DNMT3A-R882H on its wildtype counterpart was postulated by several groups. To investigate mechanistic details of this postulated dominant negative effect of the R882H mutation in complex with the wildtype enzyme, I conducted further experiments with either complexes which were formed by mixing of individually purified proteins or directly co-purified after co-expression of both subunit types. For this, DNMT3AC-R882H and wildtype were separately purified and subsequently pre-incubated for 30min as a mixture to allow the exchange of subunits. In parallel, each protein preparation was pre-incubated separately under the same exact conditions. Next, the catalytic activities of all samples were measured and the result of the mixed sample was compared to

the activities of the separately incubated samples (**Figure 14A**). As it can be seen in **Figure 14A** there was not a detectable difference between the activity of the mixed complexes compared to the sum of the activities of the individual enzymes. This result demonstrates that under these experimental settings, no dominant negative effect of DNMT3AC R882H can be observed. Due to a potential impact coming from the N-

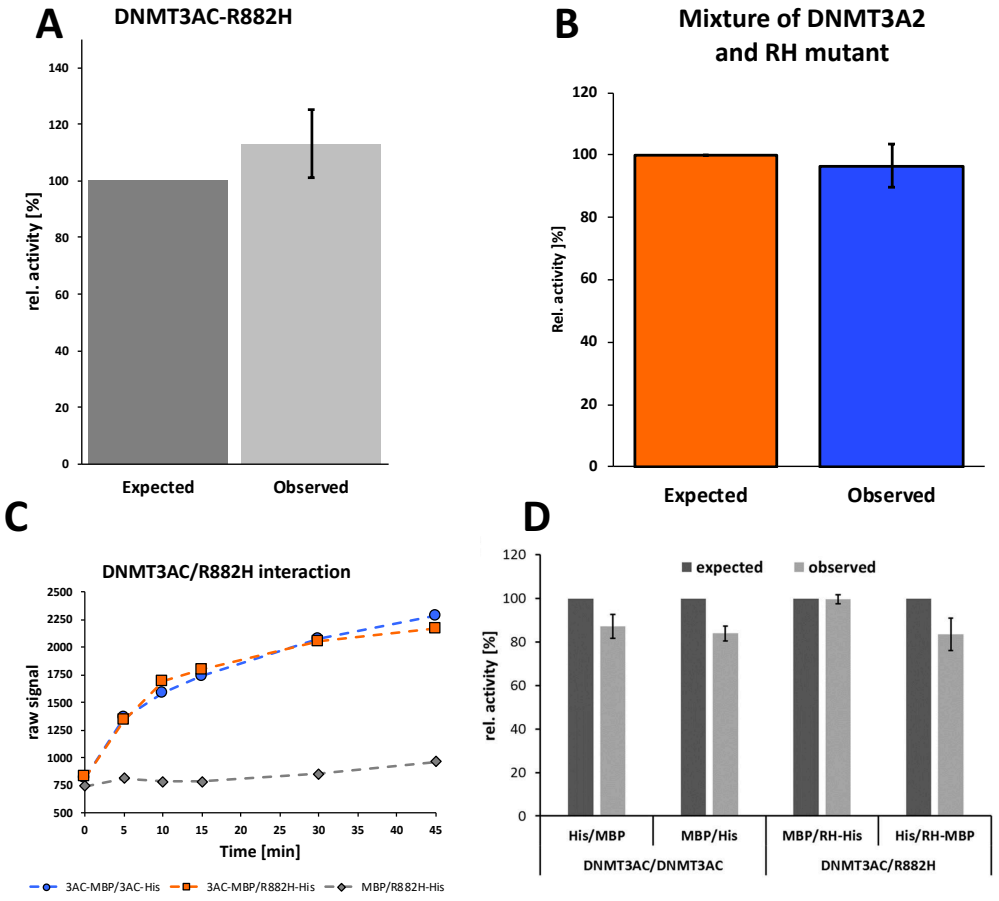



**Figure 14: Activity and interaction of R882H in complex with wildtype. A** Activity of mixed DNMT3AC/R882H complexes generated by pre-incubation. The grey bar indicates the expected activity corresponding to the sum of the activities of the two subunit types, the light grey bar shows the observed activity. **B** Activity of mixed wildtype and mutant DNMT3A2 preparations after pre-incubation. The orange bar indicates the expected activity corresponding to the sum of the activities of the two subunit types, the blue bar shows the observed activity. **C** Documentation of heterocomplex formation after pre-incubation of MBP- and His-tagged DNMT3AC subunits. Alpha acceptor and donor beads were bound to the tags and the samples were mixed. Formation of mixed complexes leads to an approximation of the two beads causing the emission of a light signal. The image shows Alpha-screen raw data observed after mixing of His and MBP-tagged DNMT3AC (blue curve) or MBP-DNMT3AC with His-R882H (orange curve). The signal observed after mixing of MBP with His-DNMT3AC (grey curve) illustrates the background. The averaged data revealed half lives for the subunit exchange of 9.7 ± 2.2 min (average ± SEM) for the wildtype MBP/wildtype-His combination and 7.4 ± 0.4 min for the wildtype-MBP/R882H-His combination. **D** Catalytic activities of purified His/MBP tagged heterotypic DNMT3AC complexes. On the left side, data for heterotypic His/MBP complexes containing two wildtype subunits are shown, on the right side, data for heterotypic wildtype/R882H complexes are displayed. The dark grey bars indicate expected activities corresponding to the sum of the activities of the two subunits, the light grey bars show the observed activities. Error bars indicate the SEM based on 3 independent experiments. For more details please refer to **Appendix 3**.

terminus of the protein, the same experimental setup was used to investigate the impact of the R882H (mouse R878H) mutation of the behavior of the full length murine DNMT3A2, a naturally occurring variant which is lacking a part of the N-terminus, while still containing intact ADD and PWWP domains. Also here, no dominant negative effect could be observed (**Figure 14B**).

To confirm that during the pre-incubation step exchange of the enzyme subunits occurs, an Alpha-screen assay was performed. For this, His-tagged wildtype DNMT3AC and R882H mutant were bound to Nickel chelate Alpha donor beads and MBP tagged wildtype DNMT3AC was bound to Anti-MBP acceptor beads. After mixing of both sample preparations and incubation at ambient temperature the Alpha signal was recorded over 45min. As displayed in **Figure 14C**, the alpha signal for the combination DNMT3AC-His/DNMT3AC-MBP increases over the time frame of 45min. The exactly same results were achieved when DNMT3AC-R882H-His/DNMT3AC-MBP were combined. Both combinations showed after exponential fitting of the curves identical half-lives of 7-10min. No subunit exchange was observed for the negative control DNMT3AC-R882H-His/MBP.
Next, heterocomplexes consisting out of one wildtype and one R882H subunit were isolated by co-expression in bacteria followed by purification using a double tag approach. As control, a corresponding double tag homocomplexes consisting of DNMT3AC-His/ DNMT3AC-MBP were co-expressed and purified according to the same protocol. The composition of each complex preparation was determined by SDS-gel electrophoresis and the exact amounts of both subunits were determined based on the Coomassie- stained gels (**Figure 14D**). By taking this together with the experimentally measured activities of the His- and MBP-tagged wildtype and R882H, it was possible to calculate the expected theoretical activities of the mixed complexes, shown in **Figure 14D**. All effects coming from the R882H mutation and of both tags on the overall activity are incorporated in these calculated activities, assuming that the R882H and the wildtype subunits do not influence each other. If a dominant negative effect existed, the observed activity of the mixed wildtype/R882H complexes should have dropped below the expected activity level. However, this was not the case, as can be seen in **Figure 14D**. Taken together, the data from this study convincingly demonstrate that under these experimental settings a dominant negative effect of the R882H mutant on the wildtype enzyme cannot be observed.

## 4.4 Investigation of the interplay between MECP2 and DNMT3A

The precise chromatin targeting and enzymatic activity of the *de novo* methyltransferase DNMT3A is essential for the accurate setting of DNA methylation patterns (Gowher and Jeltsch, 2018). Previous at that time unpublished work from our group had uncovered that the activity of DNMT3A is tightly controlled by the interaction of the enzyme with the 5mC-reading protein, MECP2. This direct interaction was shown to lead to a strong inhibition of DNMT3A activity both *in vitro* and in cell culture models. Based on these lines of evidence, I used size exclusion chromatography as an independent validation method to show the interaction between DNMT3A and MECP2. In addition, cell lines with inducible expression of MECP2 were generated to further characterize the effect that the expression of this protein has on the global levels of DNA methylation. Together, these data were included in a manuscript published in Nucleic Acid Research and attached here as **Appendix 4**.

In the following paragraph, a short description of the experiments that I have contributed to this study is provided. Previous *in vitro* pulldown experiments in our group have shown that recombinant DNMT3A and MECP2 directly interact *via* the ADD domain of the enzyme and the TRD of MECP2. To analyze this interaction in further detail, I have used size exclusion chromatography to analyze the ADD-TRD complex. For this, the MBP-tagged ADD domain of DNMT3A and the GST-tagged TRD of MECP2 were pre-incubated for 1h at 22°C to establish an interaction. As it can be seen in **Figure 15A** the isolated ADD-MBP domain alone (yellow curve) is eluting at 56kDa what would correspond to a monomer with a theoretical molecular weight of 58 kDa. In case of the TRD-GST domain (orange curve) the discrepancy in weight between the elution at 79 kDa and the theoretical weight of the monomer of 43 kDa could be explained by the tendency of GST to dimerize. The red dashed line indicates the theoretical signal expected for the mixed TRD-GST and ADD-MBP samples, calculated under the assumption that the two proteins would not interact. However, in the elution profile of the pre-incubated proteins ADD-MBP/TRD-GST (black line) an additional peak at 156 kDa was observed that would roughly correspond to a mixture of 1:2 (56 kDa + 79 kDa = 135 kDa) and 2:2 complexes (2x 56 kDa + 79 kDa = 191 kDa).

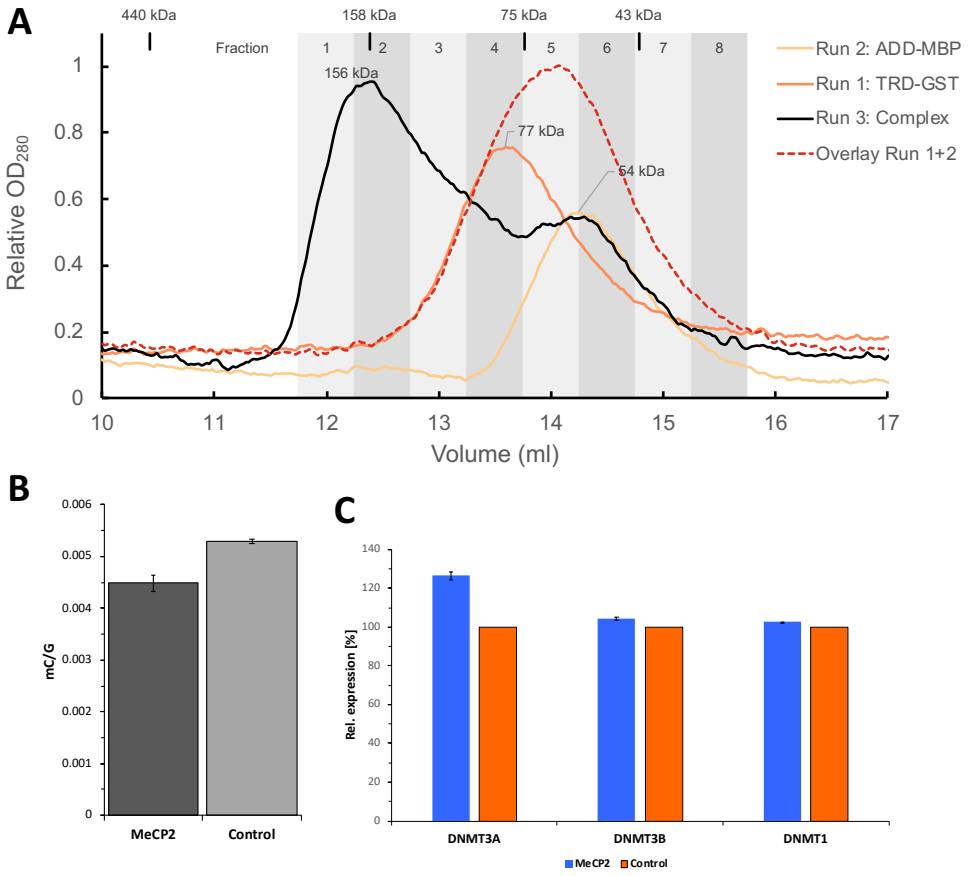



**Figure 15: Size exclusion chromatography of individual samples of GST-TRD, MBP-ADD and their mixture.** **A** The isolated MBP-ADD domain (yellow curve) elutes at 56 kDa fitting to a monomer (theoretical weight 58 kDa). The isolated GST-TRD domain (orange curve) elutes at 79 kDa approximately fitting to a dimer (theoretical weight of the monomer 43 kDa), because it is known that GST has the tendency to dimerize. The sum of both signals is shown by the red, dashed line. The elution profile of both pre-incubated proteins is shown in black, showing an additional peak at 156 kDa, roughly corresponding to a mixture of 1:2 (56 kDa + 79 kDa = 135 kDa) and 2:2 complexes (2x56 kDa + 79 kDa = 191 kDa). **B** Global DNA methylation levels after overexpression of EYFP-MeCP2 in human HCT116$^{D1hypo}$ cells. EYFP-transfected cells were used as control. **C** Expression levels of all DNMTs in the HCT116$^{D1hypo}$ cells expressing EYFP-MeCP2 or EYFP control. SDHA was used as reference gene. The error bars represent the SD based on two repeats. For more details please refer to **Appendix 4.**

This result clearly demonstrates the strong and direct interaction between the TRD domain of MECP2 and the ADD domain of DNMT3A.

Next, the goal was to see if this inhibitory effect of MECP2 onto DNMT3A can be also detected in a cellular model system. For this, the HCT116 DNMT1 hypomorphic (HCT116$^{D1hypo}$) colon cancer cell line was used which contains a truncated DNMT1 with reduced activity but with intact DNMT3A and DNMT3B (Egger et al., 2006; Rhee et al., 2002). Because of the reduced maintenance DNA methylation activity, these cells have around 20% less DNA methylation and this depends more on the methylation activity of DNMT3A and DNMT3B. For this reason, the HCT116$^{D1hypo}$ cell line is an appropriate model system to investigate the inhibitory effect of MECP2 onto

DNMT3A activity. Using viral transduction, it was possible to genomically integrate EYFP-fused MECP2 or the fluorophore as control and select for stably expressing clones. The expression of the transgenes was induced for 14 days by addition of doxycycline. Afterwards, EYFP-MECP2 or fluorophore expressing clones were enriched by FACS (fluorescence-activated cell sorting), genomic DNA was isolated and the global levels of 5mC were quantified by liquid chromatopraphy-mass spectrometry (LC-MS/MS). The mass spectrometry analysis was performed in collaboration with the group of Dr. Thomas Carell, at the Ludwig-Maximilians University in Munich. This analysis revealed a reduction of 15% of the global DNA methylation levels when MECP2 was expressed compared to the control, as it can be seen in **Figure 15B.** To exclude any influence coming from the expression level of the DNMTs, their expression levels were determined by qPCR. As reported in **Figure 15C**, a slight increase in DNMT3A expression and no changes in DNMT1 and DNMT3B could be seen.

In summary, together with the other data included shown in **Appendix 4** it could be demonstrated that DNMT3A and MECP2 directly interact and this interaction leads to a strong inhibition of DNMT3A activity not only *in vitro* but also in cells.

## 4.5 Evolutionary analysis of the potential link between DNA alkylation damage and the cytosine DNA methyltransferase activity

This project was performed in collaboration with Dr. Peter Sarkies from the Medical Research Council (MRC) in London as the main project coordinator, the experiments being summarized in a manuscript published in Nature Genetics and attached here as **Appendix 5**. Experiments from the Sarkies group had uncovered a robust and widespread co-evolution between DNMTs and ALKB2/3, enzymes involved in DNA alkylation repair. Importantly, ALKB2/3 enzymes have as a preferred substrate 1-methyladenine and 3mC. In my work, I have addressed the question of whether DNMT3A can introduce 3mC, apart from the well-established 5mC modification.

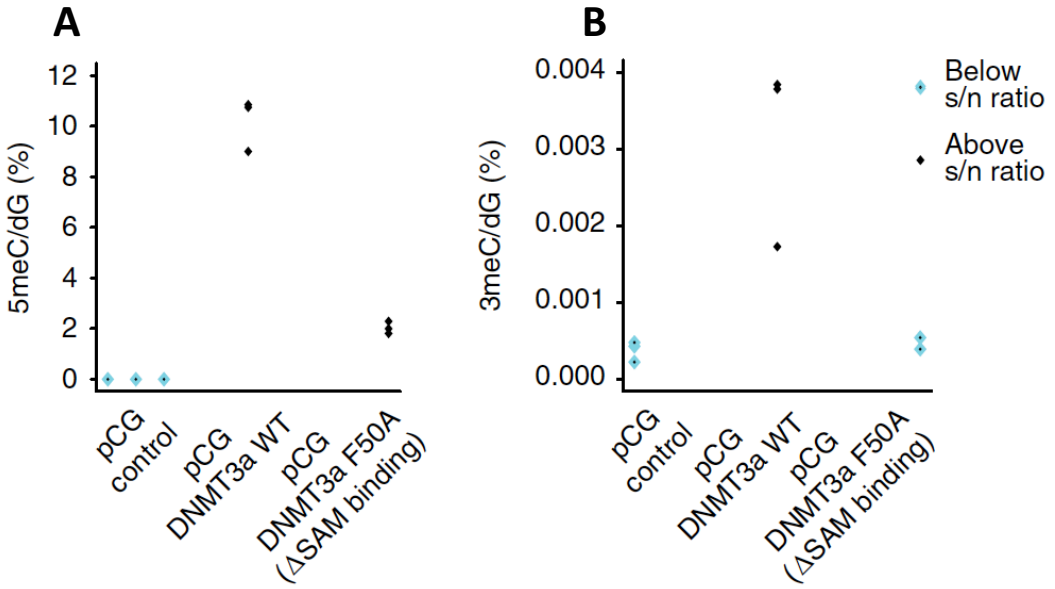



**Figure 16: Validation of the method to detect 3meC specifically in the presence of 5meC using LC/MS. A** shows the measured 5meC level introduced by the catalytic domain of 3A. The F50A mutant, which does not bind the cofactor SAM and was used as a negative control. **B** LC/MS measurement of 3meC introduced by the catalytic domain of DNMT3A *in vitro* compared with 3meC generation by the F50A mutant. Each of the three individual points for each sample shows the mean of two technical replicates independent in vitro reactions. Measurements below the signal-to-noise (s/n) ratio are shown in cyan. For more details please refer to **Appendix 5.**

In the context of this collaborative project, I have performed *in vitro* methylation assays using recombinant DNMT3AC as well as a mutated variant of DNMT3AC, F50A (F646A), which lost its ability to bind the cofactor SAM and, as a consequence, cannot methylate DNA. The *in vitro* methylated substrate was sent for ultrasensitive mass spectroscopy (LC/MS) analysis to the group Dr. Peter Sarkies. Strikingly, as shown in **Figure 16**, this analysis revealed that DNMT3A is not only able to introduce the 5mC (**Figure 16A**) modification, but also 3mC **(Figure 16B**). However, the 3mC product was much less abundant than 5mC, with a ratio of 1:2850 between 3mC and 5meC. By contrast, the DNMT3AC F50A mutant did not show any 3mC activity. This ability of DNMT3A to introduce 3mC can serve as the mechanistic link between DNA methylation and the presence of alkylation DNA damage, because 3meC is next to 1-methyladenine (1mA) the preferred substrate of ALKB2/3.

# 5. Discussion

The accurate writing of DNA methylation patterns is essential for mammalian development and proper cellular differentiation. The activity of *de novo* DNA MTases is central to this process, the abnormal setting and interpretation of DNA methylation patterns contributing to the development of severe diseases such as cancer. The ultimate aim of this work was to gain deeper insights into the mechanisms that are in place to control the activity of DNMT3A. As described in the Results section and discussed in more detail below, four main directions of research were undertaken during this thesis: 1) the regulation of DNMT3A activity by multimerization and 2) interaction partners was assessed. 3) The effect of the R882H exchange on DNMT3A activity was mechanistically characterized. This mutation is frequently found in AML patients, and was previously reported to act in a dominant negative manner. 4) Finally, together with the laboratory of Dr. Peter Sarkies, the generation of 3-methylcytosine as a side reaction of DNMT3A was explored. It is interesting to note that although DNMT3A is an essential enzyme for the accurate setting of DNA methylation patterns, the work in my thesis touched controversial findings in several sub-projects and in project 4 a novel, unprecedented activity of DNMT3A with pronounced evolutionary impact was described for the first time.

## 5.1 Mechanistic investigation of the mode of DNMT3A action *in vitro*

Strikingly, a detailed functional understanding of the mechanism of DNMT3A activity is currently lacking, because two modes of action have been put forward in literature for the mechanism by which DNMT3A methylates DNA. In our group previous experimental approaches like gel shift DNA binding experiments (Jia et al., 2007; Jurkowska et al., 2008), fluorescence polarization equilibrium DNA binding experiments (Jurkowska et al., 2011b), atomic force microscopy experiments (Jurkowska et al., 2011b), and activity studies using a natural multimeric DNA substrate (Jurkowska et al., 2008, 2011b) showed that DNMT3A multimerizes on DNA and is able to form stable protein/DNA fibers. In stark contrast, work by Reich and co-workers postulated that DNMT3A methylates DNA in a processive manner (Holz-

Schietinger and Reich, 2010; Holz-Schietinger et al., 2011, 2012), which was in contrast to earlier findings of our group showing a non-processive mechanism for DNMT3A (Gowher and Jeltsch, 2001, 2002). Mechanistically, these two models are mutually exclusive because the core of a processive model is that individual enzyme complexes move along the DNA and undergo multiple turnovers. This is not compatible with a multimerization, fiber-formation model. The goal of this study was to dissect these two very different views of the mode of action of DNMT3A.

To this end, dedicated *in vitro* biochemical experiments with the catalytic domain of DNMT3A (DNMT3AC) were designed. Accordingly, different concentrations of enzyme were used to determine the activity of DNMT3AC on long and short DNA substrates, and in the presence of a catalytically inactive variant of the enzyme. These experiments revealed that on a long DNA substrate the DNA methylation rate of DNMT3AC increased more than linearly with increasing enzyme concentration supporting an enzyme multimerization mechanism on DNA. The same could not be observed on a short 30-mer oligonucleotide substrate, where a stable protein/DNA fiber cannot be formed for spatial reasons. Noteworthy, a similar result was also obtained by Reich and co-workers (2011), where a sigmoidal concentration dependence of DNMT3A activity at lower enzyme concentration could be detected. However, these authors interpreted this finding as a reflection of the dimer/tetramer equilibrium of DNMT3A (Holz-Schietinger et al., 2011). This interpretation has two serious problems: 1) If it were correct, one would expect that the methylation levels on both the long and the short substrate should be influenced to the same extent. The fact that the long substrate is preferentially affected, supports a mode of cooperative binding and protein/DNA fiber formation. 2) Structural and biochemical data showed that the smallest active species of DNMT3A must be a tetramer (Jia et al., 2007; Jurkowska et al., 2008), because dimers are formed through the FF interface and therefore are lacking the DNA binding interface.

The multimerization model was further supported by the results of *in vitro* methylation kinetics where the catalytically almost inactive DNMT3AC-E752A variant was added to the wildtype enzyme. Strikingly, the presence of this mutant stimulated the activity of the wildtype enzyme up to 3.5-fold on the long substrate. This stimulation could be only observed on the long substrate but not when using the short 30-mer. These

counterintuitive results can be easily explained in the context of cooperative DNA binding and multimerization of DNMT3AC on DNA, because this reaction is supported by the higher combined concentration of the wildtype and the E752A mutant. This experiment cannot be explained by a processive model, because when the number of binding sites is large in excess, as was the case in the experimental setup, additional inactive enzymes moving along the DNA are not expected to influence the active enzyme molecules.

The claim of Reich and co-workers that DNMT3A is a processive enzyme is based on only one type of experimental approaches, where the authors perform methylation reactions for 2 hours using substrates of different lengths (Holz-Schietinger and Reich, 2010). After 20 min, an excess of a plasmid that did not contain any CpG sites ("chase") was added; in parallel, a reaction without chasing was performed. The absence of a significant drop in DNA methylation after chasing was taken by Holz-Schietinger and Reich as supporting evidence for the processive methylation model. As a control, the same reaction was performed, with the modification that the chase substrate was added at the beginning. This resulted in a strong reduction of the methylation rate. These observations can be easily rationalized in the context of the multimerization model if the dissociation rate of DNMT3A from DNA is slow, which is in agreement with the results of the premixing/chase experiments performed in the context of this thesis. Accordingly, after preincubation of the enzyme with the unlabelled long DNA substrate, a drop in methylation levels of the biotinylated substrate used for chasing was observed. Importantly, this reduction was found only when the experiments were performed with the 509mer and not with the short 30mer. In line with previous findings, this result suggests that the DNMT3AC oligomers stably bind the long substrate, while the short oligo cannot stably accommodate a DNMT3AC fiber.

Of note, in the experiments performed in my study just a single turnover was observed. Based on their calibration of the radioactive assay and enzyme concentration measurement, Reich and co-workers (Holz-Schietinger et al., 2011) postulated instead that during chasing with the non-CpG plasmid, several turnover events took place, which would be indicative of a processive methylation model. To address this, I measured the *in vitro* methylation levels of two different substrates: one with one CpG site and the other with two CpG sites (all in an identical flank context) at different

DNMT3AC concentrations. Under conditions of processive methylation, it is expected that the second site is methylated faster. However, both substrates were found to be methylated with the same relative efficiencies, even at low enzyme concentrations, indicating that processivity is undetectable. Importantly, in the same publication Reich and coworkers (Holz-Schietinger et al., 2011) described that only after preincubation of the enzyme with DNA they could detected processivity. This observation actually can be much easier interpreted by the formation of stable protein/DNA fibers than by a conceptual framework of a processive turnover.

Taken together, the data obtained in this study (**Appendix 1**) fully support a cooperative DNA binding and fiber formation model of DNMT3A and they do not provide any evidence for a processive DNA methylation model. Methylation of DNA is instead started by a cooperative DNA binding reaction and depending on the enzyme concentration stable fiber or small oligomers, which are not stably bound, will be formed (**Figure 17**). Inside of the protein/DNA fiber, the methylation of CpG sites is set stochastically and depends on the flanking sequence context of each individual CpG site and their access to the individual active centers.

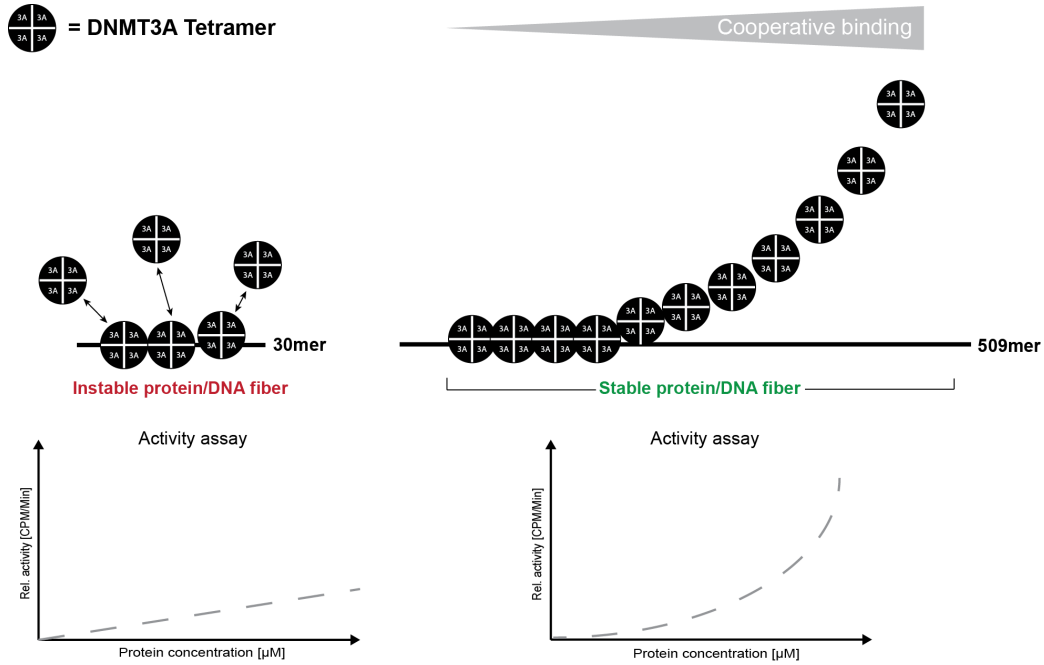



**Figure 17: Schematic illustration of DNMT3A mode of action.** On a short substrate (e. g. 30mer) the formation of a stable protein/DNA fiber is not possible for spatial reasons, which can be detected in an activity assay as a linear increase in methylation. By contrast, a long DNA substrate (e.g. 509mer) supports the formation of a stable protein/DNA fiber via cooperative binding, which can be detected in an activity assay as a more than linear increase in methylation.

As a direct continuation of my work it would be interesting to investigate, if fiber formation enhances catalysis mainly by increasing the residence time of the enzyme at its target site, or whether it induces a conformational change of the enzyme that increases the catalytic activity. This question could be addressed by determining DNA methylation rates at high enzyme concentrations after saturating the DNA with enzyme and comparing rates for short oligonucleotides (in the size range of 10-15 nucleotides) that do not allow multimerization and longer one, where more than on DNMT3A complex can bind.

On a long-term perspective, it would also be of great interest to demonstrate the existence of DNMT3A fibers in cells and to uncover other mechanisms that are in place to regulate the polymerization status of the enzyme. Of note, indirect evidence for the formation of DNMT3A multimers in cells was recently provided by a study where a DNMT3A–DNMT3L construct was fused to the nuclease-inactivated dCas9 and used for targeted gene silencing purposes (Stepper et al., 2017). Accordingly, wildtype DNMT3A was found to cause widespread methylation that extended outside of the sgRNA target sites. By contrast, the DNMT3A R832E mutant, which is not capable of horizontal polymerization, introduced DNA methylation that was spatially confined to the nearest vicinity of the target site.

One direct approach to provide evidence for DNMT3A fibers on DNA could make use of the increased half-life of the protein on the DNA after fiber formation. This could be measured *in vitro* by Fluorescence Correlation Spectroscopy, one of the advanced single molecule fluorescence techniques, using fluorescently labelled DNA molecules of different length. In addition the R832E mutant can be used, which was shown not to form fibers on DNA. Afterwards, the residence time of the protein on DNA could be determined in cells as well. To this end, single molecule tracking measurements could be employed to compare the residence time of the wildtype enzyme to that of the DNMT3A R832E mutant, which is not capable of horizontal polymerization (Clauß et al., 2017; Rajavelu et al., 2012).

Although technically still very challenging, recent advances in single-molecule super resolution microscopy might enable a direct visualization of DNMT3A multimers in cells. For instance, the RAD51 recombinase is known to assemble into ordered

filaments at sites of DNA damage (Liu et al., 2011). Recently, Haas et al. (2018) (Haas et al., 2018) developed a method that involves de-noising and filtering of single molecule microscopy data, as well as the quantification of structure size at nanoscale resolution. With this approach, the authors were able to visualize the initiation, elongation and disassembly stages of RAD51 filament dynamics after radiation-induced damage. The resolution of the method was limited to 5-10 molecules in the smallest cluster (Haas et al., 2018). For clusters larger than this, the median length of the filaments as well as the number of constituent molecules could be determined. It is conceivable that a similar approach would enable a direct visualization of DNMT3A multimers in cells. The challenge here would rather lie in the generation of a DNMT3A mutant protein that is not able to form fibers and would serve as a negative control. This variant should recapitulate all features of the wildtype protein except from its property to form fibers. The R832E mutant described above would not be a suitable candidate for this approach because of its weaker DNA binding (Rajavelu et al., 2012).

An alternative approach to document the formation of DNMT3A fibers *in vivo,* would be to immunoprecipitate the protein out of crosslinked cells and look at the length of the recovered DNA fragments. This would be done in parallel for cells expressing either the wildtype or a non-fiber forming mutant of DNMT3A. The expectation is that the wildtype protein would be associated with the recovery of longer DNA fragments.

## 5.2 Biochemical characterization of the DNMT3A-R882H mutation and its function in AML

Diseases such as cancers acquire somatic mutations that ultimately confer a growth advantage to the carrier cells and endow these with tumour initiating properties. When occurring in epigenetic regulators, such as DNMT3A, these mutations can lead to a global alteration of the chromatin landscape, which favours cancer progression, without leaving any fingerprints in the genome. For instance, around 20% of all patients that are diagnosed with the haematological condition acute myeloid leukemia (AML) carry mutations in DNMT3A (Ley et al., 2010). These mutations occur heterozygously in only one allele of the gene that is accompanied by an intact wildtype allele and they show a strong enrichment of missense mutations (**Figure 9, page 30**). The exchange of R to H at position 882 in the DNA binding interface of DNMT3A is a hotspot mutation,

encompassing 65% of all missense mutations in DNMT3A. DNMT3A mutations are a negative prognostic marker correlated with a reduced time to treatment failure, shorter duration of complete remission and decreased disease-free survival in cytogenetically normal AML (Roy et al., 2014; Yang et al., 2015). Importantly, DNMT3A mutations are often found at early stages of the disease and were proposed to work as oncogenic drivers (Shlush et al., 2014; Yang et al., 2015).

The fact that the R882H mutation is located at the RD interface of DNMT3A suggests that this exchange might influence the multimerization and/or DNA binding of DNMT3A. Despite intensive efforts, the exact pathogenic function of this mutant is still not understood at a molecular level (Roy et al., 2014). Different studies have determined the catalytic activity of the purified R882H mutant and often observed 50–70% residual activity of the mutant enzyme (Holz-Schietinger et al., 2012; Yamashita et al., 2010; Yan et al., 2011). Recently, Russler-Germain et al. (2014) postulated a functional model where the subunits of the R882H mutant affect the wildtype subunits in heterotypic wildtype/R882H complexes in a dominant negative manner. Specifically they observed and reported a 12% residual activity of mixed wildtype/R882H complexes purified from a human cell line (Russler-Germain et al., 2014). However, this model was only based on indirect biochemical studies performed with mixed enzyme preparations obtained after co-expressing the wildtype and the R882H variant in mammalian cells. Also, a general loss-of-function mechanism is not in agreement with the strong prevalence of this particular mutation in cells. If a general mechanism for loss-of-function would be in place, deletions, insertions or a broader mutational spectrum of *DNMT3A* would be expected. In addition, homozygous losses would accompany the heterozygous mutation. In the context of this thesis, a defined set of biochemical and cellular assays was used to mechanistically pinpoint the consequences of the R882H mutation on DNMT3A activity. This work resulted in two published articles (**Appendix 2** und **Appendix 3**), the main findings of which will be discussed in a combined manner below.

To characterize the biochemical properties of the DNMT3A-R882H mutant, *in vitro* methylation experiments, UV circular dichroism spectra measurements and DNA binding assays were performed with the recombinantly purified catalytic domain of the enzyme (DNMT3AC-R882H). Overall, the lack of changes in DNA binding and protein

stability indicated that the R882H mutation does not dramatically alter the DNMT3A/3A RD interface. This conclusion agrees with the moderate reduction in catalytic activity of the R882H mutant, because disruption of the RD interface by a R881A mutation in the catalytic domain of murine DNMT3A (corresponding to human R885) completely inactivated the enzyme (Jurkowska et al., 2008).

The RD interface is not only important for generating the DNA binding cleft, but it also plays a role in determining the flanking sequence specificity of DNMT3A (Gowher et al., 2006; Jurkowska et al., 2011c). Strikingly, *in vitro* methylation kinetics on a substrate, which allows to study the methylation of 56 CpG sites in different flanking contexts, and subsequent bisulfite sequencing revealed a strong change in flanking sequence preference for the mutant. Detailed analysis of the methylation data showed that some sites were methylated up to five times faster by R882H than by the wildtype enzyme. This was mainly observed at sites which were weakly methylated by wildtype DNMT3AC, but still methylated with good yields by R882H. Among the flanks of CpG sites preferred by R882H, a strong overrepresentation of G at the position +3 was observed. These findings clearly showed that the R882H exchange directly affects the DNA interaction of DNMT3A, which would explain the very high prevalence of this specific mutation in AML patients. This result is supported by previous data from the lab, where it was found that DNMT3A shows a pronounced flanking sequence preferences in the methylation of different CpG sites (Gowher et al., 2006). To explore this in more detail, a model of the DNMT3A/DNA complex was created by using the DNA from the M.HhaI-DNA complex (Klimasauskas et al., 1994) after superposition of the conserved catalytic methyltransferase residues of DNMT3A and M.HhaI. In line with the methylation data, this model revealed that the R882 side chain is directly positioned next to the +3 base in an ideal position to contact the DNA backbone at this particular site. Recently, the structure of the DNMT3A-DNMT3L heterotetramer in complex with a DNA was solved (Zhang et al., 2018) for the first time. This work revealed that the DNMT3A-DNA interaction involves a continuous interface formed by a loop part of the target recognition domain (residues R831–F848), the catalytic loop (residues G707–K721), and the DNMT3A homodimeric interface. Of note, the authors found that the R882 residue is important for the stabilization of the TRD loop through hydrogen-bonding interactions. Also, as shown in **Figure 18** and in line with my work, the two R882 residues, each provided by one DNMT3A subunit, were found to make

contacts with the +3 position of both CpG sites contained in the DNA sequence (**Figure 18**). Based on these findings, it could be speculated that the R882 residue could probe sequence dependent conformational preferences of the DNA substrate in a process that reminds of the classic indirect readout of DNA sequence by some transcription factors (Garvie and Wolberger, 2001; Rohs et al., 2010).

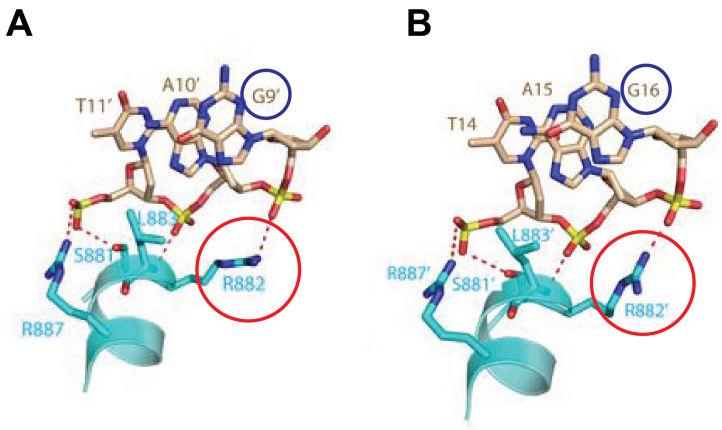



**Figure 18: Structural basis for the interaction of DNMT3A with DNA**. The images show the intermolecular interactions between the homodimeric interface of DNMT3A and the DNA substrate, which contains two CpG sites. The hydrogen bonding interaction between the R882 residue (Red circle) and the +3 position (Blue circle) relative to the two CpG sites is depicted in panel **A** and **B**, respectively (Figure was adapted from Zhang et al., 2018).

To further complement the methylation data, the flanking sequence preferences extracted from the analysis mentioned above were used to generate two model substrates: an R882H substrate and a reference substrate. The results of methylation experiments of these substrates confirmed the conclusions of the flanking sequence analysis of methylated sites by demonstrating that it is possible to design an R882H preferred and a disfavoured substrate on the basis of the flanking sequence preferences. The fact that a substrate could be designed on which the R882H mutant was more active than the wildtype enzyme provides further evidence against a dominant negative influence of this mutation.

While these experiments were performed separately with the wildtype and the R882H mutant, *in vivo* these proteins co-occur in the same cell and they have the potential to form heterotetramers with properties that could be distinct from the homotetramers formed by wildtype DNMT3A. To address this point, mixed DNMT3AC wildtype/R882H complexes were directly purified *via* a double tag affinity purification and subjected to activity analysis. This approach allowed for the first time the preparation of defined DNMT3AC/DNMT3AC-R882H heterocomplexes. In parallel, heterocomplexes were

generated by incubating the separately purified proteins for 30 minutes. This was performed with the catalytic domains of the enzymes but also with DNMT3A2, a naturally occurring DNMT3A variant that lacks a part of the N-terminal part, while having intact ADD and PWWP domains. Importantly, no dominant negative effects of the R882H mutant on the activity of the wildtype enzyme could be observed in any of these experimental setups. This result is not in agreement with the reported results of Russler-Germain et al. (2014), but it is relevant to discuss some problems in the data interpretation of their work. Russler-Germain et al. (2014) have determined the residual activity of R882H to be around 29% of wildtype DNMT3A, lower than results published by other labs (**Table 4**). Subsequently, the authors have co-purified WT-R882H mixed complexes and found that these have a residual activity of only 12%, a finding that led them to propose a dominant-negative effect of the R882H on wildtype DNMT3A. Such a strong reduction in enzymatic activity does not fit with what is expected based on a stochastic mixture of wildtype/R882H complexes. Accordingly, based on the premise that the R882H exchange does not influence the formation of the RD interface, 3 types of complexes are expected to be formed as follows: 25% wildtype/wildtype, 50% wildtype/R882H and 25% R882H/R882H. Assuming that both complex species wich contain the R882H mutant are completely inactive, a residual activity of at least 32.25% would have been expected for the mixed population. Based on this reasoning it is highly possible that the low activity, of only 12%, as observed by Russler-Germain et al. (2014) is rather due to technical reasons than to the presence of the R882H mutant in the heterocomplex.

**Table 4: Collection of DNMT3A-R882H catalytic activities reported in literature.** Taken from **Appendix 3**.

| Rel. activity of R882H | Reference |
|---|---|
| 44% | (Yamashita et al., 2010), Fig. 3 |
| 68% | (Yan et al., 2011), Fig. 2A (In presence of DNMT3L) |
| 54% | (Holz-Schietinger et al., 2012), Table 3 |
| 29% | (Russler-Germain et al., 2014), Fig. 6C |
| 62% | Appendix 2, Fig. 1A, 509mer DNA substrate |
| 65% | Appendix 3, Fig. 2C |

To summarize, the data of the studies in **Appendix 2** and **Appendix 3** does not agree with a dominant negative effect of the R882H exchange on the function of DNMT3A. Instead they show that this mutation leads to a drastic change in the flanking sequence preference of the mutated protein. That the R882H mutation is changing the flanking sequence preferences of DNMT3A indicates that the effects of this mutation on the methylation activity of DNMT3A must be CpG site specific and as a result R882H might cause combined hypomethylation and hypermethylation effects depending on the target site. Accordingly, at sites where the central CpG site is flanked by nucleotides that are preferred by the R882H mutant (like a G at position +3), a local gain of methylation may be observed. By contrast, at other sequences, a reduction of DNA methylation is expected.

A gain- of-function effect coming through the R882H mutation could lead to a very punctuated hyperactivity at particular sites and this together with its positive correlation with other mutations blocking DNA demethylation would boost these effects. Indeed, the R882H mutation in DNMT3A was found to co-occur with TET2 inactivating mutations in peripheral T-cell lymphomas (Scourzic et al., 2016; Yang et al., 2015) and both were shown to cooperate in the induction of lymphoid malignancies in mice (Scourzic et al., 2016). In line with these findings, IDH mutations (which lead to an indirect reduction of TET activity) are positively correlated with DNMT3A-R882H in AML and myeloproliferative neoplasm (Alexandrov et al., 2013; Shen et al., 2011; Yang et al., 2015).

To further refine the ideal flanking sequence preferred by the R882H mutant it would be next interesting to expand the experimental procedure used in **Appendix 2**. So far, a PCR product containing 56 CpG sites and a length of 509 bps was used to identify the change in flanking sequence preferences coming from the R882H mutation. To further increase the predictive power of this method it would be beneficial to design an experimental approach where each base at each position plus and minus 3 of the CpG site of the template DNA can be randomized, while still being able to retrieve strand specific information. Like this, the caveats stemming from the 509mer containing just 56 CpG sites in a fixed flanking sequence context would be circumvented. Using this new derived “ideal” flanking sequence it could be worth searching for similarities and matches between these sequences and DNA methylation data obtained from cellular

studies. For this, a 450K chip analysis could be performed on cell lines that overexpress either the wildtype DNMT3A or DNMT3A-R882H.
In line with this, Bera et al. (2018) have recently generated U937 leukaemia cells, where either the wildtype or the R882H variant were exogenously introduced (Bera et al., 2018). The R882H-expressing cells displayed enhanced colony formation abilities, self-renewal activity, and cell proliferation. Also, they showed an impaired stimulant-induced differentiation with inactivation of apoptosis. Gene expression microarray analysis identified genes that are both up and down-regulated by the DNMT3A mutant. Interestingly, many dysregulated genes are involved in gene translocation or chromosome aberrations that occur frequently in AML and mixed-lineage leukaemia. In addition, genes with roles in DNA damage repair, apoptosis and the cell cycle were found to be deregulated (Bera et al., 2018). By performing DNA methylation microarrays, Bera et al. (2018) also identified areas that showed both hyper- and hypomethylation in the presence of the R882H mutant. It would be next interesting to corroborate the expression data with the methylome analysis to investigate if the promoters/bodies of these genes with change in expression are enriched for motifs preferred be the R882H mutant.

From a more clinical perspective, the R882H preferred motif could also be used to mine the cancer patient databases. For instance, Glass et al. (2017) performed cytosine methylation sequencing on a set of 119 adult patients with AML. The authors found a dominant hypermethylation at promotors whereas a dominant hypomethylation was discovered at distal and intronic regions (Glass et al., 2017). It would be next interesting to further investigate the flanking sequence context of the new DMRs, identified by Glass and co-workers. Accordingly, based on the results of my thesis it is expected that patients that contain a DNMT3A-R882H mutation show a local gain of methylation at genes that contain the R882H preferred motif. Depending on the position of the methylated sequence (promoter *vs.* gene body *vs.* regulatory elements), both gene silencing or gene activation effects could be observed. Next, functional/gene ontology analysis of the genes with a change of expression could be performed. With this approach one could address if the DNMT3A-R882H variant directly contributes to the activation of genes or even pathways that are known contributors to carcinogenesis. This could be a result of the combined action of promoter hypomethylation and gene body hypermethylation. Also, the abnormal DNA

methylation pathway could translate into loss or gain of binding sites for transcription factors, which would as a consequence lead to a local remodelling of the chromatin landscape. In line with this, a compound screen for sensitivity to inhibitors of epigenetic regulators found that leukaemia stem cells carrying a *DNMT3A* mutation are more sensitive to DOT1L inhibitors (Lu et al., 2016). This histone H3 lysine 79 (H3K79) methyltransferase, belongs to a transcription elongation regulatory complex and was found to introduce this mark at R882H-associated hypo-differentially methylated regions (Li et al., 2014; Lu et al., 2016). This type of analysis could lay the basis for targeted therapy approaches for patients carrying the R882H mutation.

To keep in mind is that as shown in **Appendix 2**, the R882H mutant still responds to small molecule inhibitors that target DNMT3A. The fact that some sites become hypermethylated in the presence of the mutant could also explain why some patients were found to respond better to treatment with hypomethlyating agents (Metzeler et al., 2012; Traina et al., 2014). Finally, while the work included in this thesis laid a solid basis for the mechanistic understanding of the effects of the R882H mutation on the properties of DNMT3A, the potential clinical implications of this new view on the R882H mutation are currently still not clear.

## 5.3 Investigation of the interplay between MECP2 and DNMT3A

An increasing number of studies has shown that DNA methylation patterns are not static but are subjected to dynamic changes that arise though the continuous interplay between the DNA methylation and DNA demethylation machineries (Jeltsch and Jurkowska, 2014). The accurate setting and remodelling of this landscape are particularly important in non-dividing cells, such as terminally differentiated neurons (Heyward and Sweatt, 2015; Shin et al., 2014; Weaver, 2014). Here, in the absence of cell division, and consequently lack of DNA replication, the alteration of the DNA methylation profiles largely depends on the action of the *de novo* DNA methyltransferases. Importantly, despite its utmost importance for the setting of DNA methylation patterns, our understanding of how the activity of these enzyme is controlled *via* interaction partners and chromatin-associated signals is very limited. In this part of the thesis, I took a closer look at the regulation of the DNMT3A

methyltransferase by the 5mC-reading protein MECP2. This project was based on previous experiments from our laboratory, identifying MECP2 as a strong and direct interactor of DNMT3A. This work revealed that the interaction between these two epigenetic players is mediated by the TRD domain of MECP2 and the ADD domain of DNMT3A and translates into a strong allosteric inhibition of DNMT3A activity *in vitro*. Interestingly, binding of unmodified histone H3 to the complex was found to disrupt the association between DNMT3A and MECP2 and subsequently relieve the enzymatic inhibition *in vitro*. In my work, I have further characterized the interaction between DNMT3A and MECP2 *via* gel filtration chromatography. In addition, a cellular model was generated and used to pinpoint the effect of MECP2 overexpression on the DNA methylation landscape *in vivo*.

To validate the interaction between the ADD domain of DNMT3A and the TRD of MECP2, a gel filtration experiment was performed. Despite the fact that TRD is notoriously known for being largely unstructured, I was successful at detecting a robust interaction between the TRD and the ADD domains with this methodology.

Through its docking, MECP2 was previously found to give rise to a strong inhibition of DNMT3A activity *in vitro*. To test whether this strong effect can also be recapitulated in a cellular context, I have retrovirally generated cell lines with stable and doxycycline inducible expression of MECP2. For this, I resorted to the HCT116 DNMT1 hypomorphic colon cancer cell line, which contains a truncated DNMT1 with reduced activity, but active copies of DNMT3A and DNMT3B. Because of the impaired maintenance DNA methylation activity, these cells have a ~20% reduced level of global DNA methylation, which is more dependent on the activity of DNMT3A and DNMT3B (Egger et al., 2006; Rhee et al., 2002). For this reason, this cellular model is a suitable model system to pinpoint the influence of MECP2 expression on DNMT3A activity. Following a 14-day induction time, genomic DNA was extracted and global 5mC levels were analysed by liquid chromatography-mass spectrometry (LC-MS/MS) in cooperation with the group of Prof. Dr. Thomas Carell (LMU München). This analysis revealed a strong 15% decrease in global DNA methylation upon MECP2 expression. This experiment provided important cellular evidence for the inhibition of DNMT3A activity by MECP2 and complements the *in vitro* work previously performed in the lab.

Together, the data of this study provide unprecedented insights into a complex regulatory network, which is in place to control the activity of the *de novo* DNMT machinery. Accordingly, on a global level, the overexpression of MECP2 in a cell culture system led to a strong inhibition of DNMT3A activity. This global response could be a potential safeguard mechanism to protect the genome from aberrant DNA methylation. At specific target sites such as repetitive sequences, where H3 lacks activating marks, MECP2 can function as a recruiter of DNMT3s. Here, unmodified H3 can disrupt the interaction between MECP2 and DNMT3A, consequently releasing the allosteric inhibition. This could initiate a positive feedback loop, where active DNMT3A generates novel 5mCpG sites, which could act as further recruitment signals for more MECP2 molecules. In neurons, this process may be further supported by the non-CpG methylation (mainly CpA) set by DNMT3A and which is bound by MECP2 as well. By contrast, at sites that carry H3K4me2/3-modified chromatin (*e.g.* promoters of active genes), the interaction between MECP2 and DNMT3A would prevent abnormal DNA methylation and untargeted gene silencing.

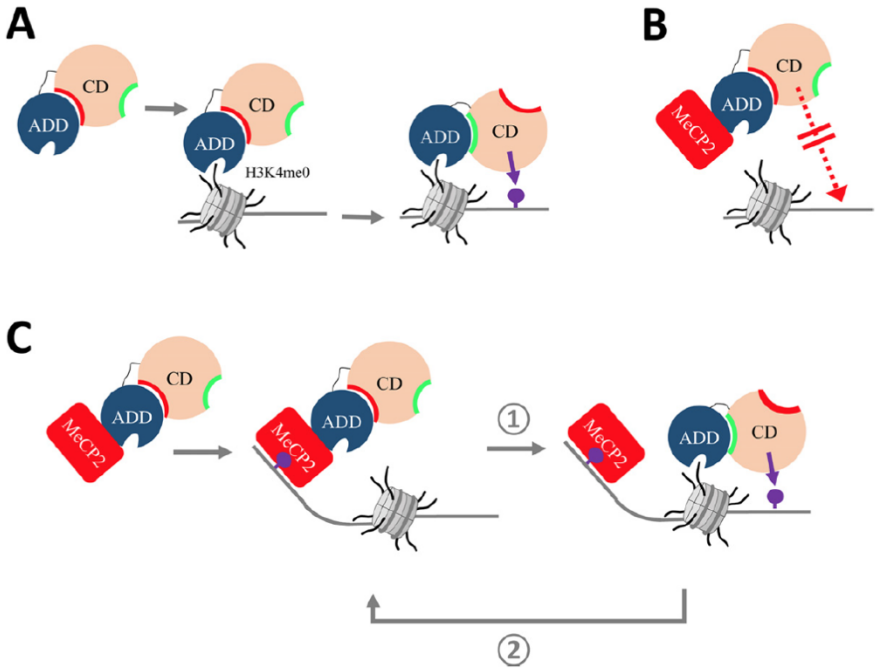



**Figure 19: Model of allosteric regulation of DNMT3A by the combined action of MECP2 and histone H3. A** shows the transition of the ADD domain of DNMT3A from the autoinhibitory (red) into the allosteric (green) interaction site, which is triggered by the binding of the unmodified H3 tail to the ADD domain. DNA methylation (purple) is subsequently introduced. **B** The setting of DNA methylation is inhibited when MECP2 binds to DNMT3A and stabilizes the autoinhibitory conformation of the enzyme. **C** The newly generated 5mC mark can act as a docking site for another MECP2 molecule that recruits additional DNMT3A enzymes. The presence of unmodified H3 tail disrupts the inhibitory interaction between MECP2 and DNMT3A, activates the enzyme (step 1) and initiates a positive feedback loop (step 2).

This work is the first **(Figure 19)** to present a comprehensive molecular characterization of a negative regulator of DNMT3A activity and provides a conceptual framework that can accommodate the conflicting reports found in literature on the function of MECP2. Accordingly, depending on the local chromatin environment, MECP2 can act as either a transcriptional activator or repressor via either targeting or inhibiting DNA methyltransferases (Ausió et al., 2014; Chahrour et al., 2008; Lyst and Bird, 2015; Della Ragione et al., 2016).

Next, it would be interesting to characterize the interaction between DNMT3A and MECP2 in more detail. For instance, peptide scanning could be performed to pinpoint the exact amino acids that are involved in the formation of the DNMT3A-MECP2 complex. Once these residues are identified, protein variants where these amino acids are mutated could be generated. Alternatively, charge reversal mutations could be placed on the surface of the ADD domain or in defined distances in the TRD (where no structure is available) and the ADD-TRD interaction tested. These new protein variants could be next used to test in a series of cellular experiments the mutual influence of MECP2 and DNMT3A upon each other. For instance, by taking advantage of the typical localization of DNMT3A at mouse heterochromatic foci (Bachman et al., 2001; Chen et al., 2004), NIH3T3 cells could be used to test if MECP2 variants carrying mutations in the TRD domain would still be recruited to these sites by the ADD domain. Equally interesting would be to test if MECP2 variants that do not longer interact with DNMT3A, but have an intact 5mC binding-pocket, would still localize to pericentromeric repeats. Also, as already successfully used in my work, the HCT116 DNMT1 hypomorphic colon cancer cell line, could be used to test whether this mutant still manages to inhibit the activity of DNMT3A in cells.

One of the paramount findings of the study in **Appendix 4** was that MECP2 stabilizes the allosterically locked conformation of DNMT3A. In the context of this work, DNMT3A variants that preferentially adopt either the allosterically locked or the allosterically open conformation were generated. These tools could be next used to identify proteins that preferentially interact with one of these two conformations. To this end, recombinant, tagged DNMT3A mutants could be immobilized on an affinity resin support and incubated with total cell lysates. The eluted fraction would then be subjected to mass spectrometry measurements to identify novel allosteric regulators

of DNMT3A activity. If successful, this approach could next be performed in parallel with lysate of non-transformed and transformed/ drug resistant cells to identify potential interaction partners that preferentially exploit the allosteric regulation of DNMT3A in a disease context. Methylation assays *in vitro* and in cells could then be used to study the effect of these novel interactors on the activity of DNMT3A.

## 5.4 Evolutionary analysis of the potential link between DNA alkylation damage and the cytosine DNA methyltransferase activity

The last project of this thesis dealt with the coevolution between DNA methylation and DNA repair systems, a very exciting topic that was addressed in close collaboration with the laboratory of Peter Sarkies (MRC London). Our laboratory has a very strong expertise in DNA methylation analysis as well as in the production and purification of recombinant, active DNA methyltransferases. By teaming up with the laboratory of Peter Sarkies, who has a keen interest in investigating the connections between epigenetic gene regulation and evolution, it could be shown that DNMT activity is associated with the generation of 3mC both *in vitro* and *in vivo* and that the demethylase ALKB2 is needed to repair this type of alkylation damage (**Appendix 5**) (**Figure 20**).

This correlation between the catalytic side-activity of the DNMTs and the resulting DNA damage may explain why organisms that contain 5mC generally need ALKB2. But even when ALKB2/3 are present the introduction of 3mC by the DNMTs can lead to

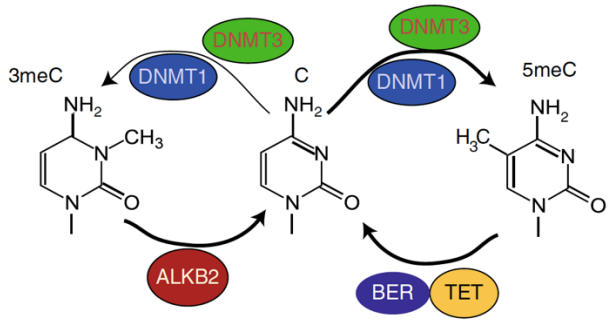



**Figure 20: Model for how DNMTs influence methylation on different positions of the target cytosine.** DNMTs predominantly methylate the C5 position of cytosine, generating 5mC (Right hand side). This modification can be removed by the TET/BER enzymes. As an off-target activity, DNMTs can also methylate the N3 position of the cytosine base, generating 3mC (Left hand side). This harmful modification can be removed by ALKB2. Abbreviations: DNMT – DNA methyltransferase; BER – Base excision repair; TET – ten eleven translocase; ALKB2 – Alkane 1-monooxygenase 2 (Figure taken from Rošić et al., 2018).

genome instability by causing DNA polymerases to stall which leads to appearance of double strand DNA breaks. In line with this hypothesis, members of the BRCA complex and RAD18, both having key functions in the DNA double-strand break repair system (Shrivastav et al., 2008) have also coevolved with the DNMTs.

In the future it would be interesting to dissect the mechanism by which DNMTs can generate 3mC and pinpoint which sidechains are involved in this off-target mechanism. This is important for a better understanding of how the DNMTs evolved and it would contribute another aspect on the involvement of the DNMTs into the development of diseases like cancer.

## 5. 5 Final conclusions and outlook

Taken together, the results included in this doctoral thesis offer unprecedented insights into the mechanisms that are in place to regulate the activity and chromatin targeting of *de novo* DNA methyltransferases, with particular focus on DNMT3A. Four main directions of research were undertaken in this respect.

In the first project, solid evidence was provided to support the cooperative DNA binding and fiber formation model of DNMT3A. It will be next interesting to investigate, if fiber formation enhances catalysis mainly by increasing the residence time of the enzyme at its target site, or whether it induces a conformational change of the enzyme that increases the catalytic activity. Also, the existence of DNMT3A fibers in cells awaits to be shown.

The second project dealt with the mechanistic characterization of the effect of the R882H exchange in DNMT3A, a driver mutation that it found in many patients diagnosed with AML. Two important findings stemmed out of this work: It could be shown that this mutation does not behave in a dominant - negative manner; instead it leads to a strong change in target site flanking sequence preference of the enzyme. These data expand the model of the potential carcinogenic effect of the R882H mutation. In the future it would be interesting to identify relevant sites which are hyper- and hypomethylated by R882H in cellular studies, and most importantly in methylome

datasets obtained from AML patients. These findings might improve the diagnosis and therapy options of these patients.

The third project focused on characterizing the molecular cross-talk between DNMT3A and the 5mC reader MECP2. The findings of this study demonstrate the importance of allosteric control of DNMT3A activity *via* the combined effect of chromatin modifications and interaction partners, highlighting the importance of characterizing novel DNMT3 interactors and their effects on the enzymatic activity.

The fourth project was performed in close collaboration with the group of Dr. Peter Sarkies and dealt with the coevolution between DNA methylation and DNA repair systems. In this work it could be shown that the activity of DNMT enzymes is associated with the generation of 3mC both *in vitro* and *in vivo* and that the demethylase ALKB2 is needed to repair this type of alkylation damage. In the future it would be interesting to pinpoint the molecular mechanisms that enable the generation of 3mC and explore the potential connections between this toxic side-product of DNA methylation and diseases such as cancer.

Finally, as highlighted by the 6 publications stemming from this doctoral work, mechanistic studies based on a balanced use of biochemical and cellular methodologies are essential to advance our understanding of the mechanisms that control the chromatin targeting and activity of *de novo* DNA methyltransferases. Combining the mechanistic insights derived from this Thesis with the recent developments in the Systems Biology field as well as the continuously increasing amount of NGS datasets, it is expected to provide unprecedented insights into the workings of the DNA methylation machinery both in normal development but also in diseases such as cancer.

# 6. References

Adams, V.H., McBryant, S.J., Wade, P.A., Woodcock, C.L., and Hansen, J.C. (2007). Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. J. Biol. Chem.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A. V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat. Rev. Genet. *17*, 487–500.

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

Anastasiadou, C., Malousi, A., Maglaveras, N., and Kouidou, S. (2011). Human Epigenome Data Reveal Increased CpG Methylation in Alternatively Spliced Sites and Putative Exonic Splicing Enhancers. DNA Cell Biol. *30*, 267–275.

Aran, D., and Hellman, A. (2013). DNA methylation of transcriptional enhancers and cancer predisposition. Cell *154*, 11–13.

Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V., et al. (2012). In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases. PLoS Genet. *8*, e1002750.

Ausió, J., Paz, A.M. de, and Esteller, M. (2014). MeCP2: the long trip from a chromatin protein to neurological disorders. Trends Mol. Med. *20*, 487–498.

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.

Bachman, K.E., Rountree, M.R., and Baylin, S.B. (2001). Dnmt3a and Dnmt3b Are Transcriptional Repressors That Exhibit Unique Localization Properties to Heterochromatin. J. Biol. Chem. *276*, 32282–32287.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science *354*, 909–912.

Barrès, R., Osler, M.E., Yan, J., Rune, A., Fritz, T., Caidahl, K., Krook, A., and Zierath, J.R. (2009). Non-CpG Methylation of the PGC-1α Promoter through DNMT3B Controls Mitochondrial Density. Cell Metab. Bartolomei, M.S. (2009). Genomic imprinting: employing and avoiding epigenetic processes. Genes Dev. *23*, 2124–2133.

Bashtrykov, P., Jankevicius, G., biology, A.S.-C.&, and 2012, undefined Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. Elsevier.

Bashtrykov, P., Ragozin, S., and Jeltsch, A. (2012). Mechanistic details of the DNA recognition by the

Dnmt1 DNA methyltransferase. FEBS Lett. *586*, 1821–1823.

Bashtrykov, P., Rajavelu, A., Hackner, B., Ragozin, S., Carell, T., and Jeltsch, A. (2014). Targeted Mutagenesis Results in an Activation of DNA Methyltransferase 1 and Confirms an Autoinhibitory Role of its RFTS Domain. ChemBioChem *15*, 743–748.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature *520*, 243–247.

Baylin, S.B., and Jones, P.A. (2011). A decade of exploring the cancer epigenome — biological and translational implications. Nat. Rev. Cancer *11*, 726–734.

Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C.A., and Zoghbi, H.Y. (2009). Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. Hum. Mol. Genet.

Bera, R., Chiu, M.-C., Huang, Y.-J., Liang, D.-C., Lee, Y.-S., and Shih, L.-Y. (2018). Genetic and Epigenetic Perturbations by DNMT3A-R882 Mutants Impaired Apoptosis through Augmentation of PRDX2 in Myeloid Leukemia Cells. Neoplasia *20*, 1106–1120.

Berger, S.L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. Genes Dev. *23*, 781–783.

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.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nat. Genet. *42*, 1093–1100.

Bostick, M., Kim, J.K., Estève, P.-O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. Science (80-. ). *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 (80-. ). *294*, 2536–2539.

Brenner, C., Deplus, R., Didelot, C., Loriot, A., Viré, E., De Smet, C., Gutierrez, A., Danovi, D., Bernard, D., Boon, T., et al. (2005). Myc represses transcription through recruitment of DNA methyltransferase corepressor. EMBO J.

Briones, V., and Muegge, K. (2012). The ghosts in the machine: DNA methylation and the mystery of differentiation. Biochim. Biophys. Acta - Gene Regul. Mech.

Brown, K.D., and Robertson, K.D. (2007). DNMT1 knockout delivers a strong blow to genome stability and cell viability. Nat. Genet. *39*, 289–290.

Brunetti, L., Gundry, M.C., and Goodell, M.A. (2017). DNMT3A in Leukemia. Cold Spring Harb. Perspect. Med. *7*.

Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T.C., Qin, J., and Zoghbi, H.Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. Science *320*, 1224–

1229.

Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., et al. (2012). Dnmt3a is essential for hematopoietic stem cell differentiation. Nat. Genet.

Challen, G.A., Sun, D., Mayle, A., Jeong, M., Luo, M., Rodriguez, B., Mallaney, C., Celik, H., Yang, L., Xia, Z., et al. (2014). Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells. Cell Stem Cell.

Chang, Y., Sun, L., Kokura, K., Horton, J.R., Fukuda, M., Espejo, A., Izumi, V., Koomen, J.M., Bedford, M.T., Zhang, X., et al. (2011). MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. Nat. Commun.

Chaudry, S.F., and Chevassut, T.J.T. (2017). Epigenetic Guardian: A Review of the DNA Methyltransferase DNMT3A in Acute Myeloid Leukaemia and Clonal Haematopoiesis. Biomed Res. Int. *2017*.

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

Chédin, F. (2011). The DNMT3 Family of Mammalian De Novo DNA Methyltransferases. Prog. Mol. Biol. Transl. Sci. *101*, 255–285.

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.

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, Z.X., Mann, J.R., Hsieh, C.L., Riggs, A.D., and Chédin, F. (2005). Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. J. Cell. Biochem.

Chowdhury, B., McGovern, A., Cui, Y., Choudhury, S.R., Cho, I.H., Cooper, B., Chevassut, T., Lossie, A.C., and Irudayaraj, J. (2015). The hypomethylating agent Decitabine causes a paradoxical increase in 5-hydroxymethylcytosine in human leukemia cells. Sci. Rep.

Clauß, K., Popp, A.P., Schulze, L., Hettich, J., Reisser, M., Escoter Torres, L., Uhlenhaut, N.H., and Gebhardt, J.C.M. (2017). DNA residence time is a regulatory factor of transcription repression. Nucleic Acids Res. *45*, 11121–11130.

Corces-Zimmerman, M.R., Hong, W.-J., Weissman, I.L., Medeiros, B.C., and Majeti, R. (2014). Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. Proc. Natl. Acad. Sci.

Dantas Machado, A.C., Zhou, T., Rao, S., Goel, P., Rastogi, C., Lazarovici, A., Bussemaker, H.J., and Rohs, R. (2015). Evolving insights on how cytosine methylation affects protein-DNA binding. Brief. Funct. Genomics *14*, 61–73.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010).

The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J. Biol. Chem. *285*, 26114–26120.
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.

Doi, A., Park, I.H., Wen, B., Murakami, P., Aryee, M.J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., et al. (2009). Differential methylation of tissue-and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat. Genet.

Du, Q., Luu, P.L., Stirzaker, C., and Clark, S.J. (2015). Methyl-CpG-binding domain proteins: Readers of the epigenome. Epigenomics.

Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. EMBO J. *27*, 406–420.

Egger, G., Jeong, S., Escobar, S.G., Cortez, C.C., Li, T.W.H., Saito, Y., Yoo, C.B., Jones, P.A., and Liang, G. (2006). Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. Proc. Natl. Acad. Sci. *103*, 14080–14085.

Eglen, R.M., Reisine, T., Roby, P., Rouleau, N., Illy, C., Bossé, R., and Bielefeld, M. (2008). The use of AlphaScreen technology in HTS: current status. Curr. Chem. Genomics *1*, 2–10.

Emperle, M., Rajavelu, A., Reinhardt, R., Jurkowska, R.Z., and Jeltsch, A. (2014). Cooperative DNA binding and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase. J. Biol. Chem. *289*, 29602–29613.

Emperle, M., Rajavelu, A., Kunert, S., Arimondo, P.B., Reinhardt, R., Jurkowska, R.Z., and Jeltsch, A. (2018). The DNMT3A R882H mutant displays altered flanking sequence preferences. Nucleic Acids Res. *46*, 3130–3139.

Esteller, M. (2002). CpG island hypermethylation and tumor suppressor genes: A booming present, a brighter future. Oncogene.

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

Felle, M., Hoffmeister, H., Rothammer, J., Fuchs, A., Exler, J.H., and Längst, G. (2011). Nucleosomes protect DNA from DNA methylation in vivo and in vitro. Nucleic Acids Res.

Fellmann, C., Hoffmann, T., Sridhar, V., Hopfgartner, B., Muhar, M., Roth, M., Lai, D.Y., Barbosa, I.A.M., Kwon, J.S., Guan, Y., et al. (2013). An Optimized microRNA Backbone for Effective Single-Copy RNAi. Cell Rep. *5*, 1704–1713.

Feng, J., Zhou, Y., Campbell, S.L., Le, T., Li, E., Sweatt, J.D., Silva, A.J., and Fan, G. (2010). Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. Nat. Neurosci.

Ferguson-Smith, A.C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. Nat. Rev. Genet. *12*, 565–575.

Ferrara, F., and Schiffer, C.A. (2013). Acute myeloid leukaemia in adults. In The Lancet, p.
Figueroa, M.E., Lugthart, S., Li, Y., Erpelinck-Verschueren, C., Deng, X., Christos, P.J., Schifano, E., Booth, J., van Putten, W., Skrabanek, L., et al. (2010). DNA Methylation Signatures Identify Biologically

Distinct Subtypes in Acute Myeloid Leukemia. Cancer Cell.

Filion, G.J.P., 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.

Flavahan, W.A., Gaskell, E., and Bernstein, B.E. (2017). Epigenetic plasticity and the hallmarks of cancer. Science (80-. ). *357*, eaal2380.

Fleischer, T., Tekpli, X., Mathelier, A., Wang, S., Nebdal, D., Dhakal, H.P., Sahlberg, K.K., Schlichting, E., Sauer, T., Geisler, J., et al. (2017). DNA methylation at enhancers identifies distinct breast cancer lineages. Nat. Commun.

Fried, I., Bodner, C., Pichler, M.M., Lind, K., Beham-Schmid, C., Quehenberger, F., Sperr, W.R., Linkesch, W., Sill, H., and Wölfler, A. (2012). Frequency, onset and clinical impact of somatic DNMT3A mutations in therapy-related and secondary acute myeloid leukemia. Haematologica.

Gabel, H.W., Kinde, B., Stroud, H., Gilbert, C.S., Harmin, D.A., Kastan, N.R., Hemberg, M., Ebert, D.H., and Greenberg, M.E. (2015). Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. Nature.

Gaidzik, V.I., Schlenk, R.F., Paschka, P., Stölzle, A., Späth, D., Kuendgen, A., Von Lilienfeld-Toal, M., Brugger, W., Derigs, H.G., Kremers, S., et al. (2013). Clinical impact of DNMT3A mutations in younger adult patients with acute myeloid leukemia: Results of the AML Study Group (AMLSG). Blood.

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

Garvie, C.W., and Wolberger, C. (2001). Recognition of Specific DNA Sequences. Mol. Cell *8*, 937–946.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Glass, J.L., Hassane, D., Wouters, B.J., Kunimoto, H., Avellino, R., Garrett-Bakelman, F.E., Guryanova, O.A., Bowman, R., Redlich, S., Intlekofer, A.M., et al. (2017). Epigenetic Identity in AML Depends on Disruption of Nonpromoter Regulatory Elements and Is Affected by Antagonistic Effects of Mutations in Epigenetic Modifiers. Cancer Discov. *7*, 868–883.

Gowher, H., and Jeltsch, A. (2001). Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpA sites. J. Mol. Biol. *309*, 1201–1208.

Gowher, H., and Jeltsch, A. (2002). Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases Downloaded from (JBC Papers in Press).

Gowher, H., and Jeltsch, A. (2018). Mammalian DNA methyltransferases: new discoveries and open questions. Biochem. Soc. Trans. BST20170574.

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

Gowher, H., Loutchanwoot, P., Vorobjeva, O., Handa, V., Jurkowska, R.Z., Jurkowski, T.P., and Jeltsch,

A. (2006). Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. J. Mol. Biol. *357*, 928–941.
Greenfield, N.J. (2007). Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat. Protoc. *1*, 2527–2535.

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., et al. (2014a). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat. Neurosci. *17*, 215–222.

Guo, W., Chung, W.-Y., Qian, M., Pellegrini, M., and Zhang, M.Q. (2014b). Characterizing the strand-specific distribution of non-CpG methylation in human pluripotent cells. Nucleic Acids Res. *42*, 3009–3016.

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature *517*, 640–644.

Guy, J., Cheval, H., Selfridge, J., and Bird, A. (2011). The Role of MeCP2 in the Brain. Annu. Rev. Cell Dev. Biol.

Haas, K.T., Lee, M., Esposito, A., and Venkitaraman, A.R. (2018). Single-molecule localization microscopy reveals molecular transactions during RAD51 filament assembly at cellular DNA damage sites. Nucleic Acids Res. *46*, 2398–2416.

Hahn, M.A., Wu, X., Li, A.X., Hahn, T., and Pfeifer, G.P. (2011). Relationship between Gene Body DNA Methylation and Intragenic H3K9me3 and H3K36me3 Chromatin Marks. PLoS One *6*, e18844.

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.
Hashimoto, H., Vertino, P.M., and Cheng, X. (2010). Molecular coupling of DNA methylation and histone methylation. Epigenomics.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res.

Hashimoto, H., Wang, D., Horton, J.R., Zhang, X., Corces, V.G., and Cheng, X. (2017). Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. Mol. Cell.

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

Hattori, N., and Ushijima, T. (2014). Compendium of aberrant DNA methylation and histone modifications in cancer. Biochem. Biophys. Res. Commun. *455*, 3–9.

Hellman, A., and Chess, A. (2007). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science (80-. ). *211*, 393–396.

Hendrich, B., and Bird, A. (1998). Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. Mol. Cell. Biol.

Heyward, F.D., and Sweatt, J.D. (2015). DNA Methylation in Memory Formation. Neurosci. *21*, 475–489.
Hite, K.C., Adams, V.H., and Hansen, J.C. (2009). Recent advances in MeCP2 structure and function.

Biochem Cell Biol.

Holz-Schietinger, C., and Reich, N.O. (2010). The inherent processivity of the human de novo methyltransferase 3A (DNMT3A) is enhanced by DNMT3L. J. Biol. Chem. *285*, 29091–29100.

Holz-Schietinger, C., Matje, D.M., Harrison, M.F., and Reich, N.O. (2011). Oligomerization of DNMT3A controls the mechanism of de novo DNA methylation. J. Biol. Chem. *286*, 41479–41488.

Holz-Schietinger, C., Matje, D.M., and Reich, N.O. (2012). Mutations in DNA methyltransferase (DNMT3A) observed in acute myeloid leukemia patients disrupt processive methylation. J. Biol. Chem. *287*, 30941–30951.

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

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303.

Ivey, A., Hills, R.K., Simpson, M.A., Jovanovic, J. V., Gilkes, A., Grech, A., Patel, Y., Bhudia, N., Farah, H., Mason, J., et al. (2016). Assessment of Minimal Residual Disease in Standard-Risk AML. N. Engl. J. Med.

Jang, H.S., Shin, W.J., Lee, J.E., and Do, J.T. (2017). CpG and non-CpG methylation in epigenetic gene regulation and brain function. Genes (Basel).

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

Jeltsch, A. (2002). Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases. ChemBioChem *3*, 274–293.

Jeltsch, A. (2006). Molecular Enzymology of Mammalian DNA Methyltransferases. In DNA Methylation: Basic Mechanisms, (Berlin/Heidelberg: Springer-Verlag), pp. 203–225.

Jeltsch, A., and Jurkowska, R.Z. (2013). Multimerization of the Dnmt3a DNA methyltransferase and its functional implications. Prog. Mol. Biol. Transl. Sci. *117*, 445–464.

Jeltsch, A., and Jurkowska, R.Z. (2014). New concepts in DNA methylation. Trends Biochem. Sci. *39*, 310–318.

Jeltsch, A., and Jurkowska, R.Z. (2016). Allosteric control of mammalian DNA methyltransferases – a new regulatory paradigm. Nucleic Acids Res. *44*, 8556–8575.

Jeltsch, A., and Lanio, T. (2002). Site-Directed Mutagenesis by Polymerase Chain Reaction. In In Vitro Mutagenesis Protocols, (Totowa, NJ: Humana Press), pp. 85–94.

Jeltsch, A., Broche, J., and Bashtrykov, P. (2018). Molecular Processes Connecting DNA Methylation Patterns with DNA Methyltransferases and Histone Modifications in Mammalian Genomes. Genes

(Basel). *9*, 566.

Jeong, M., Sun, D., Luo, M., Huang, Y., Challen, G.A., Rodriguez, B., Zhang, X., Chavez, L., Wang, H., Hannah, R., et al. (2014). Large conserved domains of low DNA methylation maintained by Dnmt3a. Nat. Genet.

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.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

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.C., 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.

Jurkowska, R.Z., and Jeltsch, A. (2016). Mechanisms and Biological Roles of DNA Methyltransferases and DNA Methylation: From Past Achievements to Future Challenges. (Springer, Cham), pp. 1–17.
Jurkowska, R.Z., Anspach, N., Urbanke, C., Jia, D., Reinhardt, R., Nellen, W., Cheng, X., and Jeltsch, A. (2008). Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. Nucleic Acids Res. *36*, 6656–6663.

Jurkowska, R.Z., Jurkowski, T.P., and Jeltsch, A. (2011a). Structure and Function of Mammalian DNA Methyltransferases. ChemBioChem *12*, 206–222.

Jurkowska, R.Z., Rajavelu, A., Anspach, N., Urbanke, C., Jankevicius, G., Ragozin, S., Nellen, W., and Jeltsch, A. (2011b). Oligomerization and binding of the Dnmt3a DNA methyltransferase to parallel DNA molecules: heterochromatic localization and role of Dnmt3L. J. Biol. Chem. *286*, 24200–24207.

Jurkowska, R.Z., Siddique, A.N., Jurkowski, T.P., and Jeltsch, A. (2011c). Approaches to Enzyme and Substrate Design of the Murine Dnmt3a DNA Methyltransferase. ChemBioChem *12*, 1589–1594.

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.

Kareta, M.S., Botello, Z.M., Ennis, J.J., Chou, C., and Chédin, F. (2006). Reconstitution and mechanism of the stimulation of de novo methylation by human DNMT3L. J. Biol. Chem.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Hum. Mol. Genet. *16*, 2272–2280.

Keck, W.M. (1995). STRUCTURE AND FUNCTION OF DNA METHYLTRANSFERASES Xiaodong Cheng.

Kim, S.J., Zhao, H., Hardikar, S., Singh, A.K., Goodell, M.A., and Chen, T. (2013). A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. Blood.

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

Kinde, B., Gabel, H.W., Gilbert, C.S., Griffith, E.C., and Greenberg, M.E. (2015). Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. Proc. Natl. Acad. Sci.

Kishikawa, S., Murata, T., Ugai, H., Yamazaki, T., and Yokoyama, K.K. (2003). Control elements of Dnmt1 gene are regulated in cell-cycle dependent manner. Nucleic Acids Symp. Ser. *3*, 307–308.

Klco, J.M., Miller, C.A., Griffith, M., Petti, A., Spencer, D.H., Ketkar-Kulkarni, S., Wartman, L.D., Christopher, M., Lamprecht, T.L., Helton, N.M., et al. (2015). Association between mutation clearance after induction therapy and outcomes in acute myeloid leukemia. JAMA - J. Am. Med. Assoc.

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

Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends Biochem. Sci. *31*, 89–97.

de Koning, A.P.J., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive elements may comprise over Two-Thirds of the human genome. PLoS Genet.

Krokan, H.E., Nilsen, H., Skorpen, F., Otterlei, M., and Slupphaug, G. (2000). Base excision repair of DNA in mammalian cells. FEBS Lett. *476*, 73–77.

Krönke, J., Bullinger, L., Teleanu, V., Tschürtz, F., Gaidzik, V.I., Kühn, M.W.M., Rücker, F.G., Holzmann, K., Paschka, P., Kapp-Schwörer, S., et al. (2013). Clonal evolution in relapsed NPM1-mutated acute myeloid leukemia. Blood.

Kumar, C.C. (2011). Genetic abnormalities and challenges in the treatment of acute myeloid Leukemia. Genes and Cancer.

Kurihara, Y., Kawamura, Y., Uchijima, Y., Amamo, T., Kobayashi, H., Asano, T., and Kurihara, H. (2008). Maintenance of genomic methylation patterns during preimplantation development requires the somatic form of DNA methyltransferase 1. Dev. Biol. *313*, 335–346.

Laisné, M., Gupta, N., Kirsh, O., Pradhan, S., Defossez, P.-A., Laisné, M., Gupta, N., Kirsh, O., Pradhan, S., and Defossez, P.-A. (2018). Mechanisms of DNA Methyltransferase Recruitment in Mammals. Genes (Basel). *9*, 617.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J., et al. (2010). Dynamic changes in the human methylome during differentiation. Genome Res. *20*, 320–331.

Lechner, M., Boshoff, C., and Beck, S. (2010). Cancer Epigenome.

Lee, J.-S., and Shilatifard, A. (2007). A site to remember: H3K36 methylation a mark for histone deacetylation. Mutat. Res. Mol. Mech. Mutagen. *618*, 130–134.

Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J., et al. (2010). *DNMT3A* Mutations in Acute Myeloid Leukemia. N. Engl. J. Med. *363*, 2424–2433.

Ley, T.J., Ding, L., Raphael, B.J., Mungall, A.J., and et al. (2013). Genomic and Epigenomic Landscapes

of Adult De Novo Acute Myeloid Leukemia The Cancer Genome Atlas Research Network. N. Engl. J. Med.
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., Szabó, 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.

Li, J.-Y., Pu, M.-T., Hirasawa, R., Li, B.-Z., Huang, Y.-N., Zeng, R., Jing, N.-H., Chen, T., Li, E., Sasaki, H., et al. (2007). Synergistic Function of DNA Methyltransferases Dnmt3a and Dnmt3b in the Methylation of Oct4 and Nanog. Mol. Cell. Biol.

Li, S., Garrett-Bakelman, F.E., Chung, S.S., Sanders, M.A., Hricik, T., Rapaport, F., Patel, J., Dillon, R., Vijay, P., Brown, A.L., et al. (2016). Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. Nat. Med.

Li, Y., Wen, H., Xi, Y., Tanaka, K., Wang, H., Peng, D., Ren, Y., Jin, Q., Dent, S.Y.R., Li, W., et al. (2014). AF9 YEATS Domain Links Histone Acetylation to DOT1L-Mediated H3K79 Methylation. Cell *159*, 558–571.

Li, Y., Xu, Q., Lv, N., Wang, L., Zhao, H., Wang, X., Guo, J., Chen, C., Li, Y., and Yu, L. (2017). Clinical implications of genome-wide DNA methylation studies in acute myeloid leukemia. J. Hematol. Oncol.
Li, Y., Zheng, H., Wang, Q., Zhou, C., Wei, L., Liu, X., Zhang, W., Zhang, Y., Du, Z., Wang, X., et al. (2018). Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. Genome Biol.

Liebert, K., and Jeltsch, A. (2008). Detection and Quantitation of the Activity of DNA Methyltransferases Using a Biotin/Avidin Microplate Assay. In Avidin-Biotin Interactions, (Totowa, NJ: Humana Press), pp. 149–155.

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.

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.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. Science *341*, 1237905.

Liu, J., Ehmsen, K.T., Heyer, W.-D., and Morrical, S.W. (2011). Presynaptic filament dynamics in homologous recombination and DNA repair. Crit. Rev. Biochem. Mol. Biol. *46*, 240–270.
Long, H.K., Blackledge, N.P., and Klose, R.J. (2013). ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochem. Soc. Trans.

Long, H.K., King, H.W., Patient, R.K., Odom, D.T., and Klose, R.J. (2016). Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. Nucleic Acids Res.

Lu, R., Wang, P., Parton, T., Zhou, Y., Chrysovergis, K., Rockowitz, S., Chen, W.-Y., Abdel-Wahab, O., Wade, P.A., Zheng, D., et al. (2016). Epigenetic Perturbations by Arg882-Mutated DNMT3A Potentiate Aberrant Stem Cell Gene-Expression Program and Acute Leukemia Development. Cancer Cell *30*, 92–

107.

Lungu, C., Muegge, K., Jeltsch, A., and Jurkowska, R.Z. (2015). An ATPase-Deficient Variant of the SNF2 Family Member HELLS Shows Altered Dynamics at Pericentromeric Heterochromatin. J. Mol. Biol. *427*, 1903–1915.

Lyons, D.B., and Zilberman, D. (2017). DDM1 and lsh remodelers allow methylation of DNA wrapped in nucleosomes. Elife.

Lyst, M.J., and Bird, A. (2015). Rett syndrome: a complex disorder with simple roots. Nat. Rev. Genet. *16*, 261–275.

Mao, S.Q., Ghanbarian, A.T., Spiegel, J., Martínez Cuesta, S., Beraldi, D., Di Antonio, M., Marsico, G., Hänsel-Hertsch, R., Tannahill, D., and Balasubramanian, S. (2018). DNA G-quadruplex structures mold the DNA methylome. Nat. Struct. Mol. Biol.

Marcucci, G., Metzeler, K.H., Schwind, S., Becker, H., Maharry, K., Mrożek, K., Radmacher, M.D., Kohlschmidt, J., Nicolet, D., Whitman, S.P., et al. (2012). Age-related prognostic impact of different types of DNMT3A mutations in adults with primary cytogenetically normal acute myeloid leukemia. J. Clin. Oncol.

Martínez de Paz, A., and Ausió, J. (2017). MeCP2, a modulator of neuronal chromatin organization involved in rett syndrome. In Advances in Experimental Medicine and Biology, p.

Mayle, A., Yang, L., Rodriguez, B., Zhou, T., Chang, E., Curry, C. V, Challen, G.A., Li, W., Wheeler, D., Rebel, V.I., et al. (2015). Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation. Blood *125*, 629–638.

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.

Messerschmidt, D.M., Knowles, B.B., and Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev. *28*, 812–828.
Métivier, R., Gallais, R., Tiffoche, C., Le Péron, 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.

Metzeler, K.H., Walker, A., Geyer, S., Garzon, R., Klisovic, R.B., Bloomfield, C.D., Blum, W., and Marcucci, G. (2012). DNMT3A mutations and response to the hypomethylating agent decitabine in acute myeloid leukemia. Leukemia *26*, 1106–1107.

Monti, M. (2012). Epigenetics protocols. Eur. J. Histochem. *56*, 8.

Moris, N., Pina, C., and Arias, A.M. (2016). Transition states and cell fate decisions in epigenetic landscapes. Nat. Rev. Genet. *17*, 693–703.

Morselli, M., Pastor, W.A., Montanini, B., Nee, K., Ferrari, R., Fu, K., Bonora, G., Rubbi, L., Clark, A.T., Ottonello, S., et al. (2015). In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. Elife.

Nan, X., Tate, P., Li, E., Bird, A., Nan, X., Tate, P., and Li, E.N. (1996). DNA methylation specifies chromosomal localization of MeCP2 . These include : DNA Methylation Specifies Chromosomal Localization of MeCP2. Mol. Cell. Biol.

Nay, S.L., Lee, D.-H., Bates, S.E., and O'Connor, T.R. (2012). Alkbh2 protects against lethality and mutation in primary mouse embryonic fibroblasts. DNA Repair (Amst). *11*, 502–510.

Noh, K.M., Wang, H., Kim, H.R., Wenderski, W., Fang, F., Li, C.H., Dewell, S., Hughes, S.H., Melnick, A.M., Patel, D.J., et al. (2015). Engineering of a Histone-Recognition Domain in Dnmt3a Alters the Epigenetic Landscape and Phenotypic Features of Mouse ESCs. Mol. Cell.

Norvil, A.B., Petell, C.J., Alabdi, L., Wu, L., Rossie, S., and Gowher, H. (2018). Dnmt3b Methylates DNA by a Noncooperative Mechanism, and Its Activity Is Unaffected by Manipulations at the Predicted Dimer Interface. Biochemistry.

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

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.

Ong, C.T., and Corces, V.G. (2014). CTCF: An architectural protein bridging genome topology and function. Nat. Rev. Genet.

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.

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V.I., Paschka, P., Roberts, N.D., Potter, N.E., Heuser, M., Thol, F., Bolli, N., et al. (2016). Genomic Classification and Prognosis in Acute Myeloid Leukemia. N. Engl. J. Med.

Papin, C., Ibrahim, A., Le Gras, S., Velt, A., Stoll, I., Jost, B., Menoni, H., Bronner, C., Dimitrov, S., and Hamiche, A. (2017). Combinatorial DNA methylation codes at repetitive elements. Genome Res.
Patel, D.J. (2016). A Structural Perspective on Readout of Epigenetic Histone and DNA Methylation Marks. Cold Spring Harb. Perspect. Biol. *8*, a018754.

Pfeifer, G., Pfeifer, and P., G. (2018). Defining Driver DNA Methylation Changes in Human Cancer. Int. J. Mol. Sci. *19*, 1166.

Pinney, S., Pinney, and E., S. (2014). Mammalian Non-CpG Methylation: Stem Cells and Beyond. Biology (Basel). *3*, 739–751.

Pon, J.R., and Marra, M.A. (2015). Driver and Passenger Mutations in Cancer. Annu. Rev. Pathol. Mech. Dis.

Qin, S., and Min, J. (2014). Structure and function of the nucleosome-binding PWWP domain. Trends Biochem. Sci. *39*, 536–547.

Qu, Y., Lennartsson, A., Gaidzik, V.I., Deneberg, S., Karimi, M., Bengtzén, S., Höglund, M., Bullinger, L., Döhner, K., and Lehmann, S. (2014). Differential methylation in CN-AML preferentially targets non-CGI regions and is dictated by DNMT3A mutational status and associated with predominant hypomethylation of HOX genes. Epigenetics.

Della Ragione, F., Vacca, M., Fioriniello, S., Pepe, G., and D'Esposito, M. (2016). MECP2, a multi-talented modulator of chromatin architecture. Brief. Funct. Genomics *15*, elw023.

Rajavelu, A., Jurkowska, R.Z., Fritz, J., and Jeltsch, A. (2012). Function and disruption of DNA Methyltransferase 3a cooperative DNA binding and nucleoprotein filament formation. Nucleic Acids Res. *40*, 569–580.

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

Rathert, P., Roth, M., Neumann, T., Muerdter, F., Roe, J.-S., Muhar, M., Deswal, S., Cerny-Reiterer, S., Peter, B., Jude, J., et al. (2015). Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. Nature *525*, 543–547.

Reddington, J.P., Pennings, S., and Meehan, R.R. (2013). Non-canonical functions of the DNA methylome in gene regulation. Biochem. J. *451*, 13–23.

Reizel, Y., Sabag, O., Skversky, Y., Spiro, A., Steinberg, B., Bernstein, D., Wang, A., Kieckhaefer, J., Li, C., Pikarsky, E., et al. (2018). Postnatal DNA demethylation and its role in tissue maturation. Nat. Commun.

Ren, W., Gao, L., Song, J., Ren, W., Gao, L., and Song, J. (2018). Structural Basis of DNMT1 and DNMT3A-Mediated DNA Methylation. Genes (Basel). *9*, 620.

Rhee, I., Jair, K.-W., Yen, R.-W.C., Lengauer, C., Herman, J.G., Kinzler, K.W., Vogelstein, B., Baylin, S.B., and Schuebel, K.E. (2000). CpG methylation is maintained in human cancer cells lacking DNMT1. Nature *404*, 1003–1007.

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

Riggs, A.D., and Xiong, Z. (2004). Methylation and epigenetic fidelity. Proc. Natl. Acad. Sci. U. S. A. *101*, 4–5.

Rinaldi, L., Datta, D., Serrat, J., Morey, L., Solanas, G., Avgustinova, A., Blanco, E., Pons, J.I., Matallanas, D., Von Kriegsheim, A., et al. (2016). Dnmt3a and Dnmt3b Associate with Enhancers to Regulate Human Epidermal Stem Cell Homeostasis. Cell Stem Cell.

Ringvoll, J., Nordstrand, L.M., Vågbø, C.B., Talstad, V., Reite, K., Aas, P.A., Lauritzen, K.H., Liabakk, N.B., Bjørk, A., Doughty, R.W., et al. (2006). Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. EMBO J. *25*, 2189–2198.
Robertson, K.D. (2005). DNA methylation and human disease. Nat. Rev. Genet.

Rohde, C., Zhang, Y., Reinhardt, R., and Jeltsch, A. (2010). BISMA - Fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. BMC Bioinformatics *11*, 230.

Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., and Mann, R.S. (2010). Origins of Specificity in Protein-DNA Recognition. Annu. Rev. Biochem. *79*, 233–269.

Roller, A., Grossmann, V., Bacher, U., Poetzinger, F., Weissmann, S., Nadarajah, N., Boeck, L., Kern, W., Haferlach, C., Schnittger, S., et al. (2013). Landmark analysis of DNMT3A mutations in hematological malignancies. Leukemia *27*, 1573–1578.

Rondelet, G., Dal Maso, T., Willems, L., and Wouters, J. (2016). Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. J. Struct. Biol. *194*, 357–367.

Rošić, S., Amouroux, R., Requena, C.E., Gomes, A., Emperle, M., Beltran, T., Rane, J.K., Linnett, S., Selkirk, M.E., Schiffer, P.H., et al. (2018). Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. Nat. Genet. *50*, 452–459.

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.

Roy, D.M., Walsh, L.A., and Chan, T.A. (2014). Driver mutations of cancer epigenomes. Protein Cell *5*, 265–296.

Di Ruscio, A., Welner, R.S., Tenen, D.G., and Amabile, G. (2016). The second hit of DNA methylation. Mol. Cell. Oncol. *3*, e1093690.

Russler-Germain, D.A., Spencer, D.H., Young, M.A., Lamprecht, T.L., Miller, C.A., Fulton, R., Meyer, M.R., Erdmann-Gilmore, P., Townsend, R.R., Wilson, R.K., et al. (2014). The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-Type DNMT3A by Blocking Its Ability to Form Active Tetramers. Cancer Cell *25*, 442–454.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc. Natl. Acad. Sci. U. S. A. *103*, 1412–1417.

Schlenk, R.F., and Döhner, H. (2013). Genomic applications in the clinic: use in treatment paradigm of acute myeloid leukemia. Hematology Am. Soc. Hematol. Educ. Program.

Schmitz, K.M., Mayer, C., Postepska, A., and Grummt, I. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev.

Scourzic, L., Couronné, L., Pedersen, M.T., Della Valle, V., Diop, M., Mylonas, E., Calvo, J., Mouly, E., Lopez, C.K., Martin, N., et al. (2016). DNMT3A R882H mutant and Tet2 inactivation cooperate in the deregulation of DNA methylation control to induce lymphoid malignancies in mice. Leukemia *30*, 1388–1398.

Sharif, J., and Koseki, H. (2011). Recruitment of Dnmt1: Roles of the SRA Protein Np95 (Uhrf1) and Other Factors. Prog. Mol. Biol. Transl. Sci. *101*, 289–310.

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., and Jones, P.A. (1994). The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. Nucleic Acids Res. *22*, 972–976.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A., and Issa, J.-P.J. (2007). Genome-Wide Profiling of DNA Methylation Reveals a Class of Normally Methylated CpG Island Promoters. PLoS Genet. *3*, e181.

Shen, Y., Zhu, Y.-M., Fan, X., Shi, J.-Y., Wang, Q.-R., Yan, X.-J., Gu, Z.-H., Wang, Y.-Y., Chen, B., Jiang, C.-L., et al. (2011). Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia. Blood *118*, 5593–5603.

Shin, J., Ming, G. -l., and Song, H. (2014). DNA modifications in the mammalian brain. Philos. Trans. R. Soc. B Biol. Sci. *369*, 20130512–20130512.

Shirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., Kono, T., and Sasaki, H. (2013). Mouse Oocyte Methylomes at Base Resolution Reveal Genome-Wide Accumulation of Non-CpG Methylation and Role of DNA Methyltransferases. PLoS Genet. *9*, e1003439.

Shlush, L.I., Zandi, S., Mitchell, A., Chen, W.C., Brandwein, J.M., Gupta, V., Kennedy, J.A., Schimmer, A.D., Schuh, A.C., Yee, K.W., et al. (2014). Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature *506*, 328–333.

Shrivastav, M., De Haro, L.P., and Nickoloff, J.A. (2008). Regulation of DNA double-strand break repair pathway choice. Cell Res. *18*, 134–147.

Song, J., Teplova, M., Ishibe-Murakami, S., and Patel, D.J. (2012). Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science *335*, 709–712.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., Tiwari, V.K., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature.

Stepper, P., Kungulovski, G., Jurkowska, R.Z., Chandra, T., Krueger, F., Reinhardt, R., Reik, W., Jeltsch, A., and Jurkowski, T.P. (2017). Efficient targeted DNA methylation with chimeric dCas9–Dnmt3a–Dnmt3L methyltransferase. Nucleic Acids Res. *45*, 1703–1713.

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.

Suetake, I., Mishima, Y., Kimura, H., Lee, Y.-H., Goto, Y., Takeshima, H., Ikegami, T., and Tajima, S. (2011). Characterization of DNA-binding activity in the N-terminal domain of the DNA methyltransferase Dnmt3a. Biochem. J. *437*, 141–148.

Sugino, K., Hempel, C.M., Okaty, B.W., Arnson, H.A., Kato, S., Dani, V.S., and Nelson, S.B. (2014). Cell-Type-Specific Repression by Methyl-CpG-Binding Protein 2 Is Biased toward Long Genes. J. Neurosci.

Takai, D., and Jones, P.A. (2003). Origins of Bidirectional Promoters: Computational Analyses of Intergenic Distance in the Human Genome. Mol. Biol. Evol. *21*, 463–467.

Takeshima, H., Suetake, I., Shimahara, H., Ura, K., Tate, S.I., and Tajima, S. (2006). Distinct DNA methylation activity of Dnmt3a and Dnmt3b towards naked and nucleosomal DNA. J. Biochem.

Tanaka, Y., Umata, T., Okamoto, K., Obuse, C., and Tsuneoka, M. (2014). CxxC-ZF Domain Is Needed for KDM2A to Demethylate Histone in rDNA Promoter in Response to Starvation. Cell Struct. Funct.

Testa, J.R., Uri Mintz Rowley, J.D., Vardiman, J.W., and Golomb, H.M. (1979). Evolution of Karyotypes in Acute Nonlymphocytic Leukemia'. Cancer Res.

Thol, F., Damm, F., Lüdeking, A., Winschel, C., Wagner, K., Morgan, M., Yun, H., Göhring, G., Schlegelberger, B., Hoelzer, D., et al. (2011). Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia. J. Clin. Oncol.

Thompson, J.J., Kaur, R., Sosa, C.P., Lee, J.H., Kashiwagi, K., Zhou, D., and Robertson, K.D. (2018). ZBTB24 is a transcriptional regulator that coordinates with DNMT3B to control DNA methylation. Nucleic Acids Res.

Tippin, D.., and Sundaralingam, M. (1997). Nine polymorphic crystal structures of d(CCGGGCCCGG), d(CCGGGCCm5CGG), d(Cm5CGGGCCm5CGG) and d(CCGGGCC(Br)5CGG) in three different conformations: effects of spermine binding and methylation on the bending and condensation of A-DNA. J. Mol. Biol. *267*, 1171–1185.

Traina, F., Visconte, V., Elson, P., Tabarroki, A., Jankowska, A.M., Hasrouni, E., Sugimoto, Y., Szpurka, H., Makishima, H., O'Keefe, C.L., et al. (2014). Impact of molecular mutations on treatment response to DNMT inhibitors in myelodysplasia and related neoplasms. Leukemia *28*, 78–87.

Tudor, M., Akbarian, S., Chen, R.Z., and Jaenisch, R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. Proc. Natl. Acad. Sci.
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.

Velasco, G., Hube, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Pequignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. Proc. Natl. Acad. Sci.

Véron, N., and Peters, A.H.F.M. (2011). Tet proteins in the limelight. Nature *473*, 293–294.

Waddington, C.H. (1942). Canalization of Development and the Inheritance of Acquired Characters. Nature *150*, 563–565.

Wan, J., Su, Y., Song, Q., Tung, B., Oyinlade, O., Liu, S., Ying, M., Ming, G.L., Song, H., Qian, J., et al. (2017). Methylated cis-regulatory elements mediate KLF4-dependent gene transactivation and cell migration. Elife.

Wang, L., Zhao, Y., Bao, X., Zhu, X., Kwok, Y.K.Y., Sun, K., Chen, X., Huang, Y., Jauch, R., Esteban, M.A., et al. (2015). LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. Cell Res.

Watanabe, D., Suetake, I., Tada, T., and Tajima, S. (2002). Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. Mech. Dev. *118*, 187–190.

Weaver, I.C.G. (2014). Integrating Early Life Experience, Gene Expression, Brain Development, and Emergent Phenotypes: Unraveling the Thread of Nature via Nurture. Adv. Genet. *86*, 277–307.

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

Weisenberger, D.J., Campan, M., Long, T.I., Kim, M., Woods, C., Fiala, E., Ehrlich, M., and Laird, P.W. (2005). Analysis of repetitive element DNA methylation by MethyLight. Nucleic Acids Res.

Widschwendter, M., Jiang, G., Woods, C., Müller, H.M., Fiegl, H., Goebel, G., Marth, C., Müller-Holzner, E., Zeimet, A.G., Laird, P.W., et al. (2004). DNA hypomethylation and ovarian cancer biology. Cancer Res.

Wienholz, B.L., Kareta, M.S., Moarefi, A.H., Gordon, C.A., Ginno, P.A., and Chédin, F. (2010). DNMT3L Modulates Significant and Distinct Flanking Sequence Preference for DNA Methylation by DNMT3A and DNMT3B In Vivo. PLoS Genet. *6*, e1001106.

Williams, K., Christensen, J., and Helin, K. (2012). DNA methylation: TET proteins—guardians of CpG islands? EMBO Rep. *13*, 28–35.

Wilson, A.S., Power, B.E., and Molloy, P.L. (2007). DNA hypomethylation and human diseases. Biochim. Biophys. Acta - Rev. Cancer.

Wu, C.T., and Morris, J.R. (2001). Genes, genetics, and epigenetics: A correspondence. Science (80-. ).

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell.

Xu, C., Liu, K., Lei, M., Yang, A., Li, Y., Hughes, T.R., and Min, J. (2018). DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. Structure.

Yamashita, Y., Yuan, J., Suetake, I., Suzuki, H., Ishikawa, Y., Choi, Y.L., Ueno, T., Soda, M., Hamada, T., Haruta, H., et al. (2010). Array-based genomic resequencing of human leukemia. Oncogene *29*, 3723–3731.

Yan, X.J., Xu, J., Gu, Z.H., Pan, C.M., Lu, G., Shen, Y., Shi, J.Y., Zhu, Y.M., Tang, L., Zhang, X.W., et al. (2011). Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat. Genet.

Yang, L., Rau, R., and Goodell, M.A. (2015). DNMT3A in haematological malignancies. Nat. Rev. Cancer *15*, 152–165.

Yi, J.H., Liu, J., and Wang, K.H. (2016). CpG island methylator phenotype in colorectal cancer. World Chinese J. Dig.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science (80-. ).

Yokochi, T., and Robertson, K.D. (2002). Preferential methylation of unmethylated DNA by mammalian de novo DNA methyltransferase Dnmt3a. J. Biol. Chem.

Yu, W., McIntosh, C., Lister, R., Zhu, I., Han, Y., Ren, J., Landsman, D., Lee, E., Briones, V., Terashima, M., et al. (2014). Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. Genome Res.

Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell.

Zhang, Y., Rohde, C., Tierling, S., Jurkowski, T.P., Bock, C., Santacruz, D., Ragozin, S., Reinhardt, R., Groth, M., Walter, J., et al. (2009a). DNA Methylation Analysis of Chromosome 21 Gene Promoters at Single Base Pair and Single Allele Resolution. PLoS Genet. *5*, e1000438.

Zhang, Y., Rohde, C., Tierling, S., Stamerjohanns, H., Reinhardt, R., Walter, J., and Jeltsch, A. (2009b). DNA Methylation Analysis by Bisulfite Conversion, Cloning, and Sequencing of Individual Clones. (Humana Press), pp. 177–187.

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. *38*, 4246–4253.
Zhang, Z.M., Lu, R., Wang, P., Yu, Y., Chen, D., Gao, L., Liu, S., Ji, D., Rothbart, S.B., Wang, Y., et al. (2018). Structural basis for DNMT3A-mediated de novo DNA methylation. Nature.

Zhu, H., Geiman, T.M., Xi, S., Jiang, Q., Schmidtmann, A., Chen, T., Li, E., and Muegge, K. (2006). Lsh is involved in de novo methylation of DNA. EMBO J.

Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature.

# 7. Author´s contribution

• **Emperle, M.**, Rajavelu, A., Reinhardt, R., Jurkowska, R. Z., & Jeltsch, A. (2014). Cooperative DNA binding and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase. *The Journal of biological chemistry*, *289*(43), 29602-13.

A.J., M.E. and A.R.. devised the project. M.E. conducted the biochemical assays with contributions from A.R. and R.Z.J., R.R. conducted DNA sequencing for the bisulfite methylation analysis. R.Z.J. and A.J. supervised research. A.J. wrote the manuscript draft. All authors contributed to data interpretation and discussion, read and approved the final manuscript.

• **Emperle, M.**, Rajavelu, A., Kunert, S., Arimondo, P. B., Reinhardt, R., Jurkowska, R. Z., & Jeltsch, A. (2018). The DNMT3A R882H mutant displays altered flanking sequence preferences. *Nucleic acids research*, *46*(6), 3130-3139.

A.J., M.E. and A.R. devised the project. M.E. and A.R. conducted the biochemical assays with contributions from S.K. R.R. conducted DNA sequencing for the bisulfite methylation analysis. P.B.A. provided the DNMT3A inhibitors. R.Z.J. and A.J. supervised research. A.J. wrote the manuscript draft. All authors contributed to data interpretation and discussion, read and approved the final manuscript.

• **Emperle, M.**, Dukatz, M., Kunert, S., Holzer, K., Rajavelu, A., Jurkowska, R. Z., & Jeltsch, A. (2018). The DNMT3A R882H mutation does not cause dominant negative effects in purified mixed DNMT3A/R882H complexes. *Scientific reports*, *8*(1), 13242.

A.J. and M.E. devised the project. M.E. and M.D. conducted the biochemical assays with contributions from A.R., S.K. and K.H. A.J. and R.Z.J. supervised research. A.J. wrote the manuscript draft. All authors contributed to data interpretation and discussion, read and approved the final manuscript.

• Rajavelu, A.*, Lungu, C.*, **Emperle, M.**, Dukatz, M., Bröhm, A., Broche, J., Hanelt, I., Parsa, E., Schiffers, S., Karnik, R., Meissner, A., Carell, T., Rathert, P., Jurkowska, R. Z., & Jeltsch, A. (2018). Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2. *Nucleic acids research*, *46*(17), 9044-9056. *co-shared first authors

A.J., A.R., C.L. and R.Z.J. devised the project and analyzed the data. C.L. and A.R. conducted the biochemical assays with contributions from I.H., M.D., A.B. and M.E. C.L. and A.R. performed the fluorescence microscopy experiments. C.L. and M.E. performed the cell culture and biochemical work for the data shown in Figure 5 with contribution from P.R. and J.B. S.S., E.P. and T.C. performed the LC-ESI-MS/MS. All authors contributed to data interpretation and discussion, read and approved the final manuscript.

• Rošić, S., Amouroux, R., Requena, C. E., Gomes, A., **Emperle, M.**, Beltran, T., Rane, J. K., Linnett, S., Selkirk, M. E., Schiffer, P. H., Bancroft, A. J., Grencis, R. K., Jeltsch, A., Hajkova, P., & Sarkies, P. (2018). Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. *Nature genetics*, *50*(3), 452-459.

P.S. and P.H. conceived the study. P.S., P.H. and A.J. designed the experiments. DNA extraction and bisulfite sequencing were carried out by S.R. and P.S. P.S. performed bioinformatic and computational analyses. 3meC analysis by LC/MS was carried out by R.A., C.E.R., S.L. and P.S. ESC CRISPR deletion and analysis was performed by A.G., J.K.R. and P.S. M.E. and A.J. carried out the in vitro DNMT3a analysis. T.B. and P.H.S. performed genome assembly. S.R., M.E.S., R.K.G. and A.J.B. were responsible for nematode culture. P.S., P.H. and A.J. analyzed the data and prepared the manuscript.

# 8 Appendix (not included in the published thesis)

### Appendix 1

**Emperle, M.**, Rajavelu, A., Reinhardt, R., Jurkowska, R. Z., & Jeltsch, A. (2014). Cooperative DNA binding and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase. *The Journal of biological chemistry*, *289*(43), 29602-13.

### Appendix 2

**Emperle, M.**, Rajavelu, A., Kunert, S., Arimondo, P. B., Reinhardt, R., Jurkowska, R. Z., & Jeltsch, A. (2018). The DNMT3A R882H mutant displays altered flanking sequence preferences. *Nucleic acids research*, *46*(6), 3130-3139.

### Appendix 3

**Emperle, M.**, Dukatz, M., Kunert, S., Holzer, K., Rajavelu, A., Jurkowska, R. Z., & Jeltsch, A. (2018). The DNMT3A R882H mutation does not cause dominant negative effects in purified mixed DNMT3A/R882H complexes. *Scientific reports*, *8*(1), 13242. doi:10.1038/s41598-018-31635-8.

### Appendix 4

Rajavelu, A.*, Lungu, C.*, **Emperle, M.**, Dukatz, M., Bröhm, A., Broche, J., Hanelt, I., Parsa, E., Schiffers, S., Karnik, R., Meissner, A., Carell, T., Rathert, P., Jurkowska, R. Z., & Jeltsch, A. (2018). Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2. *Nucleic acids research*, *46*(17), 9044-9056. *co-shared first authors

### Appendix 5

Rošić, S., Amouroux, R., Requena, C. E., Gomes, A., **Emperle, M.**, Beltran, T., Rane, J. K., Linnett, S., Selkirk, M. E., Schiffer, P. H., Bancroft, A. J., Grencis, R. K., Jeltsch, A., Hajkova, P., & Sarkies, P. (2018). Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. *Nature genetics*, *50*(3), 452-459.