Deep Enzymology studies on the mammalian DNA methyltransferases and methylcytosine dioxygenases

Von der Fakultät 3: Chemie der Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von
Sabrina Adam

Geboren am 08.07.1994 in Backnang

Hauptberichter:	Prof. Dr. Albert Jeltsch
Mitberichter:	Prof. Dr. Jens Brockmeyer
Prüfungsvorsitzender:	Prof. Dr. Andreas Köhn

Tag der mündlichen Prüfung: 15.09.2022

Universität Stuttgart

Institut für Biochemie und Technische Biochemie

Abteilung Biochemie

2022

Erklärung über die Eigenständigkeit der Dissertation

Ich versichere, dass ich die vorliegende Arbeit mit dem Titel

„Deep Enzymology studies on the mammalian DNA methyltransferases and methylcytosine dioxygenases“

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

“Deep Enzymology studies on the mammalian DNA methyltransferases and methylcytosine dioxygenases”

is entirely my own work except where otherwise indicated. Passages and ideas from other sources have been clearly indicated.

Name/Name: Sabrina Adam

Unterschrift/Signed: _____

Datum/Date: 20.06.2022

Acknowledgements

First, I want to thank Prof. Dr. Albert Jeltsch for the opportunity to be a part of his research group. It was exactly what I needed to further grow my love of biochemistry and to be able to pursue this love in my future career. I also thank you for your guidance and repeated encouragement, your essential assistance in all writing processes and that you always have an open ear for any problems.

Furthermore, I would like to thank Prof. Dr. Brockmeyer for being the co-referee of my PhD thesis and Prof. Dr. Köhn for taking up the role as the head of the committee.

Next, a special thanks goes to Dr. Dr. Pavel Bashtrykov, who never stopped to inspire me during all the years I've known him. It was always a great pleasure to not only discuss scientific topics but also hiking routes and vacation spots with you and I will definitely remember you as a mentor that helped me grow into the scientific person I am today (starting from my bachelor days).

In addition, I would like to thank all of my colleagues in the lab, who always helped me if I had any issues and made my stay so enjoyable. I will never forget how pleasant the work atmosphere was and how many fun hours we spent together. I also thank my great students Lukas and Greta, who showed me how nice it is if you can start a scientific spark in someone and who were a great help during my research.

Finally, I would like to thank my family and friends, who never understood why I liked biochemistry so much, but never stopped trying to understand what it's all about (except for Mira, who was the only one who got it, luckily). Last but not least, I want to thank my boyfriend Jonas, who always supported me throughout these years and encouraged me to pursue what I love.

List of publications

Adam, Sabrina*; Anteneh, Hiwot*; Hornisch, Maximilian; Wagner, Vincent; Lu, Jiuwei; Radde, Nicole E.; Bashtrykov, Pavel; Song, Jikui; Jeltsch, Albert (2020): DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. In Nature communications 11 (1), p. 3723. DOI: 10.1038/s41467-020-17531-8.

* These authors contributed equally to the work.

Adam, Sabrina; Bräcker, Julia; Klingel, Viviane; Osteresch, Bernd; Radde, Nicole E.; Brockmeyer, Jens; Bashtrykov, Pavel; Jeltsch, Albert (2022): Flanking sequences influence the activity of TET1 and TET2 methylcytosine dioxygenases and affect genomic 5hmC patterns. In Communications biology 5 (1), p. 92. DOI: 10.1038/s42003-022-03033-4.

Bröhm, Alexander; Schoch, Tabea; Dukatz, Michael; Graf, Nora; Dorscht, Franziska; Mantai, Evelin; **Adam, Sabrina**; Bashtrykov, Pavel; Jeltsch, Albert (2022): Methylation of recombinant mononucleosomes by DNMT3A demonstrates efficient linker DNA methylation and a role of H3K36me3. In Communications biology 5 (1), p. 192. DOI: 10.1038/s42003-022-03119-z.

Data not included in this thesis.

Dukatz, Michael; **Adam, Sabrina**; Biswal, Mahamaya; Song, Jikui; Bashtrykov, Pavel; Jeltsch, Albert (2020): Complex DNA sequence readout mechanisms of the DNMT3B DNA methyltransferase. In Nucleic acids research 48 (20), pp. 11495–11509. DOI: 10.1093/nar/gkaa938.

Data not included in this thesis.

Emperle, Max*; **Adam, Sabrina***; Kunert, Stefan; Dukatz, Michael; Baude, Annika; Plass, Christoph; Rathert, Phillip; Bashtrykov, Pavel; Jeltsch, Albert (2019): Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns. In Nucleic acids research 47 (21), pp. 11355–11367. DOI: 10.1093/nar/gkz911.

* These authors contributed equally to the work.

Emperle, Max*; Bangalore, Disha M.*; **Adam, Sabrina**; Kunert, Stefan; Heil, Hannah S.; Heinze, Katrin G.; Bashtrykov, Pavel; Tessmer, Ingrid; Jeltsch, Albert (2021): Structural and biochemical insight into the mechanism of dual CpG site binding and methylation by the DNMT3A DNA methyltransferase. In Nucleic acids research 49 (14), pp. 8294–8308. DOI: 10.1093/nar/gkab600.

* These authors contributed equally to the work.

Gao, Linfeng*; Emperle, Max*; Guo, Yiran; Grimm, Sara A.; Ren, Wendan; **Adam, Sabrina**; Uryu, Hidetaka; Zhang, Zhi-Min; Chen, Dongliang; Yin, Jiekai; Dukatz, Michael; Anteneh, Hiwot; Jurkowska, Renata Z.; Lu, Jiuwei; Wang, Yinsheng; Bashtrykov, Pavel; Wade, Paul A.; Wang, Gang Greg; Jeltsch, Albert; Song, Jikui (2020b): Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. In Nature communications 11 (1), p. 3355. DOI: 10.1038/s41467-020-17109-4.

* These authors contributed equally to the work.

Hofacker, Daniel*; Broche, Julian*; Laistner, Laura; **Adam, Sabrina**; Bashtrykov, Pavel; Jeltsch, Albert (2020): Engineering of Effector Domains for Targeted DNA Methylation with Reduced Off-Target Effects. In International journal of molecular sciences 21 (2). DOI: 10.3390/ijms21020502.

* These authors contributed equally to the work. Data not included in this thesis.

Jeltsch, Albert; **Adam, Sabrina**; Dukatz, Michael; Emperle, Max; Bashtrykov, Pavel (2021): Deep Enzymology Studies on DNA Methyltransferases Reveal Novel Connections between Flanking Sequences and Enzyme Activity. In Journal of molecular biology 433 (19), p. 167186. DOI: 10.1016/j.jmb.2021.167186.

Review of all Deep Enzymology publications.

Mack, Alexandra; Emperle, Max; Schnee, Philipp; **Adam, Sabrina**; Pleiss, Jürgen; Bashtrykov, Pavel; Jeltsch, Albert (2022): Preferential Self-interaction of DNA Methyltransferase DNMT3A Subunits Containing the R882H Cancer Mutation Leads to Dominant Changes of Flanking Sequence Preferences. In Journal of molecular biology 434 (7), p. 167482. DOI: 10.1016/j.jmb.2022.167482

Table of contents

List of figures

List of tables

List of abbreviations

2OG	2-oxoglutarate
30mer	DNA substrate with 30 nucleotides
ca5C	5-carboxylcytosine
f5C	5-formylcytosine
hm5C	5-hydroxymethylcytosine
m5C	5-methylcytosine
A	Adenine
ADD	ATRX-DNMT3A-DNMT3L domain
AdoHcy (SAH)	S-adenosyl-L-homocysteine
AdoMet (SAM)	S-adenosyl-L-methionine
AML	Acute myeloid leukaemia
BAH (1/2)	Bromo-adjacent homology domain (1 or 2)
BC	Barcode
BER	Base excision repair
bps	Base pairs
C	Cytosine
CpA	Shorthand for 5'-cytosine-phosphate-adenine-3'
CpC	Shorthand for 5'-cytosine-phosphate-cytosine-3'
CpG	Shorthand for 5'-cytosine-phosphate-guanine-3'
CpH (non-CpG)	Shorthand for 5'-cytosine-phosphate-adenine/cytosine/guanine -3'
CpN	Shorthand for 5'-cytosine-phosphate-any base-3'
CpT	Shorthand for 5'-cytosine-phosphate-thymine-3'
Cryo-EM	Cryogenic electron microscopy
CXXC	Cysteine-X-X-cysteine
DBSH	Double-stranded beta-helix
DMAB-seq	DNMT1 methylation activity-assisted bisulfite sequencing
DNA	Deoxyribonucleic acid
DNMT(s) (1/3/3A/3B/3C/3L)	DNA methyltransferase(s) (1 or 3 or 3A or 3B or 3C or 3L)
DOT1L	DOT1 like histone lysine methyltransferase
dsDNA	Double-stranded DNA
E. coli	Escherichia coli
ESCs	Embryonic stem cells
Fe(II)/Fe(III)/Fe(IV)	Iron in oxidation state +2 or +3 or +4
FF interface	Four stacked phenylalanine residues forming the interface
G	Guanine
GFP	Green fluorescent protein
GK	Glycine-lysine
H3K9	Histon 3 lysine 9
H3R2	Histone 3 arginine 2
HCT116 cells	Human colon cancer cell line
HEK293 cells	Human embryonic kidney 293 cells
His	Histidine (as tag referring to a string of six histidines)
hmCpA	Hydroxymethyladenine
HPLC-MS/MS	High-performance liquid chromatography coupled with mass spectrometry

HSPC	Human pluripotent stem cells
ICF	Immunodeficiency, centromere region instability, facial anomalies
iPSCs	Induced pluripotent stem cells
JBP (1/2)	J binding protein (1 or 2)
KO (1KO/DKO/TKO)	Knockout (single or double or triple knockout)
LB	Luria-Bertani
MBD	Methyl-CpG-binding domain
MD simulation	Molecular dynamics simulation
MeCP2	Methyl-CpG-binding protein 2
MTase	Methyltransferase
NGS	Next-Generation Sequencing
NgTET1	TET1 from *Naegleria gruberi*
Nickel-NTA	Ni(II) ions coupled to nitrilotriacetic acid
NLS	Nuclear localization signal
Oct4	Octamer-binding transcription factor 4
p53	Tumour suppressor protein 53
PCNA	Proliferating cell nuclear antigen
PCR (1/2)	Polymerase chain reaction (step 1 or 2 for library preparation)
PDBI	Protein database index
PGCs	Primordial germ cells
PIP box	PCNA-interacting protein element
Pre-LSC	Cells in a pre-leukemic state
PWWP	Proline-tryptophan-tryptophan-proline domain
RD interface	Arginine and aspartate residues forming the interface
RFTS	Replication foci targeting sequence domain
RING	Really interesting new gene domain
RMSD	Root mean squared deviation
RNA	Ribonucleic acid
SatII	Satellite 2
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	The standard deviation of the mean
Sf21 cells	Ovarian cells isolated from *Spodoptera frugiperda*
SFM	Scanning force microscopy
SRA	SET and RING-associated domain
ssDNA	Single-stranded DNA
T	Thymine
T2/T4	Enterobacteria phage 2 or 4
TDG	Thymine DNA glycosylase
TET (TET1/2/3)	Ten-eleven translocation (enzyme 1 or 2 or 3)
TpG	Shorthand for 5'-thymine-phosphate-guanine-3'
TRD	Target recognition domain
tRNA	Transfer ribonucleic acid
TSS	Transcription start site
U2O2 cells	Human bone osteosarcoma epithelial cells
UHRF (1/2)	Ubiquitin-like, containing PHD and RING finger domains (1 or 2)
UV	Ultraviolet
WT	Wild-type

Zusammenfassung

Das sich rapide entwickelnde Feld der Epigenetik untersucht vererbliche Veränderungen von zellulären Phänotypen, die nicht mit einer Veränderung der DNA Sequenz einhergehen. In Säugetieren werden diese Veränderungen von verschiedenen epigenetischen Signalen kontrolliert, wie beispielsweise Modifikationen von Histonen, Inkorporation von Histonvarianten, nicht-kodierender RNA und Modifikationen der DNA, wobei letzteres den Fokus dieser Arbeit darstellt. Zusammen regulieren diese Modifikationen die Gentranskription über Veränderungen des Chromatinzustandes, sowie der Zugänglichkeit der DNA für Transkriptionsfaktoren. In Säugetieren tritt die Methylierung von Cytosin-Basen meistens an der C5 Position (m5C) auf. Die daraus folgenden zellspezifischen DNA Methylierungsmuster umfassen hauptsächlich Methylierungen im CpG Kontext, welcher insgesamt zu 70 % methyliert vorliegt, jedoch sind auch geringere Mengen an CpH Methylierungen vorhanden. Dieser Prozess der DNA Methylierung, welcher mithilfe der *de novo* DNMT3 DNA Methyltransferasen erfolgt, ist ein essentieller Schritt in der Entwicklung von Säugetieren. Die entstandenen DNA Methylierungsmuster werden anschließend mithilfe der Erhaltungsmethyltransferase DNMT1 in einem replikationsabhängigen Prozess aufrechterhalten. Über die Methylierung von DNA hinaus wird die aktive Form der DNA Demethylierung von den TET Methylcytosin-Dioxygenasen initiiert, welche die sequentielle Oxidation von 5-Methylcytosin (m5C) über 5-Hydroxymethylcytosin (hm5C) und 5-Formylcytosin (f5C) zu 5-Carboxylcytosin (ca5C) katalysieren. Durch die Entdeckung dieses Vorgangs wurden neue Fragen aufgeworfen hinsichtlich des mechanistischen Zusammenspiels aller beteiligten Enzyme an den gemeinsamen Regulationen der dynamischen DNA Methylierungsmuster in Zellen.

Im Vergleich zu früheren Studien, welche sich hauptsächlich auf die Rekrutierung und die biologische Funktion der Enzyme fokussierten, war es das Hauptziel dieser Arbeit die fundamentale molekulare Basis der DNA Erkennung der DNMTs und TETs aus Säugetieren systematisch zu untersuchen. Zu diesem Zweck wurde deren Aktivität an verschiedenen Methylierungsstellen in Kombination mit allen möglichen Flankierungssequenzen bestimmt. Um diese detaillierten mechanistischen Einblicke zu erhalten, wurde ein neuer experimenteller Ansatz mit dem Namen Deep Enzymology entwickelt und angewendet. Dieser Ansatz basiert auf der Einzelmolekül-

Analyse von Enzymaktivitäten, welche sich durch die Kombination von Reaktionen auf randomisierten Substraten und der anschließenden Analyse der Modifikationen einzelner DNA Moleküle durch Bisulfit-Konversion und Next-Generation Sequencing auszeichnet. Mithilfe dieser Methode war es möglich, ausgeprägte und bisher unbekannte Flankierungssequenz-Präferenzen von DNMT1, DNMT3A und DNMT3B zusammen mit einigen ihrer Mutanten, sowie von TET1 und TET2 zu bestimmen. Darüber hinaus konnten in vielen Fällen die strukturelle Erklärungen und biologische oder pathologische Konsequenzen dieser Effekte identifiziert werden.

Im ersten Teil dieser Arbeit habe ich festgestellt, dass DNMT1 bisher unbekannte aber über 100-fache Unterschiede in der Methylierung von hemimethylierten CpG Stellen mit verschiedenen Flankierungssequenzen aufweist. Mithilfe von publizierten und neuen DNMT1 Kristallstrukturen im Komplex mit verschiedenen DNA Substraten konnten diese Erkenntnisse zusätzlich verschiedenen Biegungszuständen sowie Konformationsänderungen der DNA und des Enzyms während der Komplexbildung zugeordnet werden. Es wurde auch gezeigt, dass die Präferenzen von DNMT1 für bestimmte Flankierungssequenzen stark mit zellulären m5C Profilen korrelieren, was deutlich darauf hinweißt, dass das zelluläre DNA Methylom von diesen Präferenzen mitbestimmt wird.

Im zweiten Projekt habe ich bei der Validierung und Präzisierung der Präferenzen der *de novo* Enzyme DNMT3A und DNMT3B im CpG und CpN Kontext mitgewirkt, welche zuvor bereits in kleineren bis mittelgroßen Studien beobachtet wurden. Mittels des Deep Enzymology Ansatzes wurden für beide Enzyme sehr stark ausgeprägte Präferenzen für bestimmte Flankierungssequenzen beobachtet und es konnten mechanistisch wichtige Schlüsselregionen oder Aminosäuren (z.B. K777 und T775) in einer vorhandenen Kristallstruktur von DNMT3A bzw. neuen Strukturen von DNMT3B identifiziert werden. Auf dieser Basis konnten anschließend die beobachteten Unterschiede der DNMT3 Präferenzen im CpG Kontext sowie die stärkere Methylierung durch DNMT3B im CpH Kontext erklärt werden. Die ausgeprägte Korrelation der experimentellen Flankierungspräferenzen mit zellulären DNA Methylierungsdaten betonte zusätzlich die biologische Relevanz der enzymatischen Präferenzen. Des Weiteren war ich an einem Unterprojekt beteiligt, in welche

verschiedene DNA Interaktionsmodi von DNMT3A/3L Heterotetrameren während der Methylierung von zwei CpG Stellen in verschiedenen Abständen aufgedeckt wurden.

Im Zusammenhang mit DNMT3A zielte das dritte Projekt dieser Arbeit darauf ab, den pathogenen Mechanismus der heterozygot exprimierten Krebsmutation R882H aufzudecken, welche besonders häufig in Patienten mit akuter myeloischer Leukämie gefunden wurde. Hierbei wurden die biochemischen Untersuchungen auf früheren Beobachtungen im Labor aufgebaut, welche in kleineren Studien eine Änderung der Präferenz-Profile von DNMT3A nach der Mutation von R882 zu Histidin gezeigt hatten. Systematische Untersuchungen dieses Effekts zeigten eine mehr als 70-fache Veränderung der Flankierungssequenz-Präferenzen von DNMT3A ausgelöst durch die Mutation dieser einen Aminosäure. Auf der Basis dieses Effekts konnte zudem gezeigt werden, dass R882H Untereinheiten präferentiell mit sich selbst interagieren, was eine mechanistische Interpretation für den beobachteten dominanten Effekt der Mutante gegenüber dem Wildtyp DNMT3A bei deren gleichzeitiger Expression in Zellen lieferte. Zusammengenommen boten diese zwei Teilprojekte wichtige Einblicke in den karzinogenen Effekt von R882H.

Abschließend wurde im vierten Projekt dieser Arbeit der Einfluss von Flankierungssequenzen auf die Aktivität von TET1 und TET2 untersucht, welche einen ähnlichen Reaktionsmechanismus zum Ausklappen des Cytosins verwenden wie die Methyltransferasen, bei denen jedoch bisher aufgrund vorhandener Kristallstrukturen ein Fehlen solcher Präferenzen postuliert wurde. Tatsächlich konnte ich ausgeprägte Präferenzen für beide Enzyme bestimmen, sowohl auf Substraten mit einer CpG als auch CpH Stelle, welche m5C oder hm5C Modifikation enthielten. Zudem reflektierten die bestimmten Effekte lokale sowie genomweite hm5C Muster und sie konnten im Fall von TET2 mit indirekten DNA Interaktionen eines spezifischen Arginin-Restes mit der +1 Flankierungsposition oder Stabilisierungseffekte des Cytosins über das Basenpaar an der -1 Flankierungsposition erklärt werden.

Abstract

The emerging field of epigenetics investigates heritable changes in cellular phenotypes that do not include changes in the DNA sequence. In mammals, these changes are controlled by various epigenetic signals such as histone modifications or incorporation of histone variants, non-coding RNAs and DNA modifications, with the latter being the focus of this work. Together, these modifications control the transcription states of cells through the modulation of chromatin states and the accessibility of DNA to transcription factors. In mammals, methylation of cytosine bases at the C5 position (m5C) predominantly occurs in the context of CpG sites, which are methylated to approximately 70% in a cell type-specific pattern, but lower levels of cytosine methylation are also present in CpH context. The introduction of m5C is mediated by the *de novo* DNMT3 DNA methyltransferases in a process that was proven to be essential for mammalian development. Later, the maintenance methyltransferase DNMT1 preserves the DNA methylation patterns in a replication-coupled manner. Beyond the methylation of DNA, active DNA demethylation is initiated by the TET methylcytosine dioxygenases through the stepwise oxidation of 5-methylcytosine (m5C) over 5-hydroxymethylcytosine (hm5C) and 5-formylcytosine (f5C) to 5-carboxylcytosine (ca5C) giving rise to new questions about the mechanistic interplay of all involved enzymes that altogether regulate the dynamic DNA methylation landscape in cells.

In contrast to previous studies, which focused more on the targeting of the enzymes and their biological functions, it was the ultimate purpose of this work to systematically determine the fundamental molecular basis for DNA recognition of mammalian DNMTs and TETs. To this end, their activity was studied on different target sites embedded into all possible flanking sequence contexts. To gain such detailed mechanistic insights, a new experimental approach termed Deep Enzymology was developed and applied. It is based on the single molecule analysis of enzyme activity, which is achieved by coupling enzymatic reactions on randomized substrates with hairpin-bisulfite conversion followed by Next-Generation Sequencing for the readout of the modification state of individual DNA molecules. Using this method, it was possible to discover distinct and previously unknown flanking sequence preferences for DNMT1, DNMT3A and DNMT3B as well as some of their mutants, together with TET1 and

TET2. Moreover, in many cases, the structural explanation and biological or pathogenic consequences of these effects could be uncovered.

In the first part of my work, I discovered that DNMT1 shows profound differences of up to 100-fold in the methylation of hemimethylated CpG sites with different flanking sequences, which were previously unknown. The findings could further be connected to different degrees of DNA bending and conformational rearrangements of the DNA and enzyme during complex formation using published but also new DNMT1 crystal structures in complex with DNA molecules of different sequences. The flanking sequence preferences of DNMT1 were shown to be highly correlated with cellular m5C profiles, clearly indicating that DNMT1 preferences shape the cellular methylome.

In the second project, I contributed to the validation and refinement of the CpG and CpN flanking sequence preferences of the *de novo* enzymes DNMT3A and DNMT3B previously observed in small- to mid-scale studies. Using the Deep Enzymology approach, very strong and distinct flanking sequence preferences were observed for both enzymes and mechanistic key regions or residues (e.g. K777 and T775) could be identified using available crystal structures of DNMT3A and newly provided structures of DNMT3B, which can explain the difference in CpG methylation preferences as well as the stronger CpH methylation of DNMT3B. Correlated methylation effects were also observed in cellular data, which emphasizes the biological relevance of the determined flanking preferences. Furthermore, I was included in a sub-project that unravelled the different DNA interaction modes applied by DNMT3A/3L heterotetramers to perform co-methylation of CpG sites in different distances.

In the context of DNMT3A, the third project of this thesis strived to elucidate the pathogenic mechanism of the heterozygous cancer mutation R882H, which was shown to be enriched in patients with acute myeloid leukaemia. Here, the biochemical investigations were built on previous observations of the lab that demonstrated a change in the flanking sequence preference profiles upon the mutation of R882 to histidine in small-scale studies. This effect was systematically investigated showing more than a 70-fold change in the flanking sequence preferences of DNMT3A due to the single amino acid mutation. Exploiting these effects, we could also demonstrate that R882H subunits preferentially interact with each other, providing a mechanistic interpretation for the observed dominant effect of this mutant over wild-type DNMT3A

when both are expressed in one cell. Together, these two sub-projects offered important insights into the carcinogenic effect of R882H.

Lastly, the fourth project examined potential flanking sequence effects on the activity of TET1 and TET2, which use a similar base-flipping mechanism as the DNMTs but were so far assumed to lack flanking effects based on the published crystal structures. Indeed, I was able to determine distinct preferences for both enzymes using substrates with CpG as well as non-CpG target sites containing m5C or hm5C modifications. Moreover, these findings recapitulated local and genome-wide hm5C patterns and they could also be connected to indirect interactions of the DNA with a specific arginine residue in TET2 on the +1 flanking position and stabilization effects of the target cytosine by the -1 base pair.

1 Introduction

1.1 Key players of epigenetics

Every human being is made up of trillions of cells, making us a very complex organism to be studied. During development, precursor cells originating from one single zygote differentiate into the roughly 200 different cell types the human body consists of (Moris et al., 2016). In this process, a gradual transition of embryonic stem cells (ESCs) into groups of more differentiated cells with more specialized morphology and functionality takes place. Despite the tremendous variety of cellular phenotypes, almost all of these differentiated cells contain the same genetic information encoded in their DNA sequence. Hence, cellular specialization must be by the regulation of gene expression realised by an additional layer of epigenetic information.

The term “epigenetics” was first introduced by Conrad Waddington in 1942 and described as the investigation of these additional modifications as “the study of the interactions between genes and their products which result in a particular phenotype” (Waddington, 1942). Over the last decades, this particular field of research has developed rapidly so that the definition had to be redefined to “the study of changes in gene function that are mitotically and/or meiotically heritable and do not entail a change in DNA sequence” (Wu and Morris, 2001).

Until today, it is known that epigenetic signals include non-coding RNAs, histone post-translational modifications such as acetylation, methylation or phosphorylation as well as the incorporation of histone variants and the introduction of chemical modifications to the DNA, e.g. DNA methylation. Together, these modifications lead to spatial and temporal modulations in the chromatin structure, which subsequently regulate the gene expression state of differentiated cells (Allis and Jenuwein, 2016). One gene regulatory principle is that condensation of chromatin into heterochromatin results in transcriptional repression, whereas remodelling into a more open chromatin structure (euchromatin) enables gene expression and DNA repair (Bonasio et al., 2010).

As depicted in Figure 1, stable inheritance of gene expression states and therefore the maintenance of cellular identities requires three main factors (Berger et al., 2009; Wallace et al., 2018). Firstly, an extracellular signal termed “Epigenator” (e.g. oxidative stress, lack of oxygen or other cellular conditions) activates the epigenetic signalling

cascade in the cell. Then, “Initiators” such as DNA-binding proteins or non-coding RNAs define the regions on the chromosome to be modulated based on their locus and sequence specificity. Lastly, different “Maintainers” such as DNA methyltransferases or histone-modifying enzymes recruited by the “Initiator” are responsible to generate and maintain the local changes in chromatin organization. DNA methylation as an epigenetic “Maintainer” as well as the different pathways of its removal is the focus of this thesis and will be further discussed in the following sections.

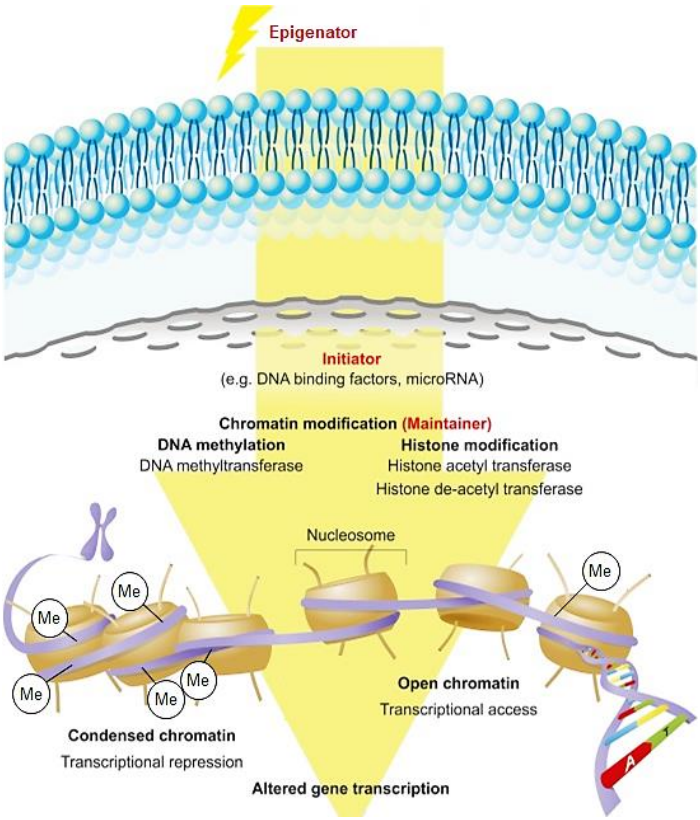


Figure 1: Schematic summary of key epigenetic players. The interplay of three different factors is needed to stably maintain cellular identity. An extracellular “Epigenator” initiates the signalling cascade in the cell that starts with an “Initiator” that locus- and sequence-specifically recruits different “Maintainers” to establish and maintain epigenetic modifications such as DNA methylation (Me). In turn, these modifications lead to changes in the chromatin organization and influence the gene expression state at this locus (adapted from Wallace et al., 2018).

1.2 DNA methylation

1.2.1 DNA methylation as an epigenetic modification

Methylation as a chemical modification of DNA bases was first discovered by Hotchkiss in 1948 and its product was defined as “epicytosine” (Mattei et al., 2022). As the name suggests, this specific epigenetic modification occurs mostly at the C5 position of cytosines in mammals in the context of palindromic CpG dinucleotides (Ambrosi et al., 2017; Schübeler, 2015), but during the last years asymmetric DNA methylation at non-CpG sites has also been observed (He and Ecker, 2015).

In the human genome, which was completely sequenced during the Human Genome Project in 2000 (Piovesan et al., 2019), roughly 60-80% of all 56 million CpG sites are methylated corresponding to 4-6% of all cytosines (Laurent et al., 2010; Lister and Ecker, 2009). Based on this high number it was hardly surprising that DNA methylation was found to play a role in various fundamental developmental processes in mammals such as inactivation of the female X-chromosome during embryonic development, silencing of repetitive elements and retrotransposons to maintain cellular integrity as well as genetic imprinting to ensure paternal-origin-specific expression of genes (Breiling and Lyko, 2015). In addition, aberrant DNA methylation has been connected to the emergence and progression of different cancer types (Baylin and Jones, 2011; Bergman and Cedar, 2013) and it plays a role in several other diseases such as psychiatric disorders or immune dysfunctions (Jin and Liu, 2018).

Despite the importance of DNA methylation, its preferred sequence context of CpG dinucleotides was found to be evolutionarily depleted by a factor of 5-10 from the human genome except for specific regions called CpG islands (Gardiner-Garden and Frommer, 1987). The reason for this phenomenon is the mutagenic property of methylated CpG sites. Hydrolytic deamination of methylated cytosines results in TpG mismatches accounting for approximately 35% of all point mutations (Cooper and Youssoufian, 1988). Such mispairs are more difficult to correct than UpG mismatches, resulting from the 4-fold slower deamination of unmethylated cytosine nucleobases, due to the standard occurrence of thymine in DNA in comparison to the unnatural base uracil (Shen et al., 1994).

Throughout the human genome, approximately 29,000 so-called CpG islands were found which were defined as regions of 500-2,000 bps with a GC content of at least 50% and an observed versus expected ratio of CpG dinucleotides above 0.6 (Gardiner-Garden and Frommer, 1987; Takai and Jones, 2002). These regions of clustered CpG sites occur mostly at gene promotors (about 70% of all genes) including housekeeping and tissue-specific genes (Saxonov et al., 2006).

Overall, the methylation state of the CpG sites present at a specific locus determines the accessibility of the regional chromatin landscape and therefore the transcription of the locus-specific genes (Deaton and Bird, 2011). CpG-rich regions such as CpG islands are mostly hypomethylated and enable the recruitment of transcription factors and other chromatin remodelling factors. In contrast, regions lacking these clusters are generally hypermethylated (with 60-80% of singly occurring CpG sites being methylated, depending on the cell type) to silence gene expression of repetitive and transposable elements (Pehrsson et al., 2019) or to prevent initiation of intragenic transcription (Neri et al., 2017). This is not only accomplished by the blockage of transcription activating factors but also through the recruitment of proteins that specifically bind to methylated cytosine residues, leading to the formation of repressive complexes and the formation of heterochromatin (Allis and Jenuwein, 2016).

1.2.2 Classical and extended models of cytosine DNA methylation

DNA methylation as an epigenetic modification of cytosine nucleobases has been the subject of many studies throughout the last decades (Ambrosi et al., 2017; Schübeler, 2015), but the principle process of how the cell type-specific DNA methylation patterns are set and maintained was already proposed independently by two research groups in 1975 (Holliday and Pugh, 1975; Riggs, 1975). This classical model, which is shown in an updated form in Figure 2 (Jeltsch et al., 2018), comprised two groups of DNA methyltransferases (DNMTs), one setting the methylation marks and one restoring the marks after DNA replication. Later on, it was discovered that the family of DNA methyltransferase 3 (DNMT3) enzymes introduces DNA methylation on CpG dinucleotides. Like this, the *de novo* enzymes create methylation patterns of fully methylated and fully unmethylated CpG sites (Okano et al., 1999). During DNA replication, the newly synthesized daughter strands do not contain this epigenetic

information any longer, which would lead to dilution of the signal over time. Therefore, it is essential to restore the cytosine methylation, a process which is performed by the DNA methyltransferase 1 (DNMT1). This maintenance methyltransferase specifically recognizes the hemimethylated CpG sites (one strand methylated, one strand not methylated) occurring after DNA replication and methylates the daughter strand (Jeltsch, 2006). For this process to occur accurately, the enzyme harbours intrinsic properties that are discussed in more detail in section 1.2.3.2.

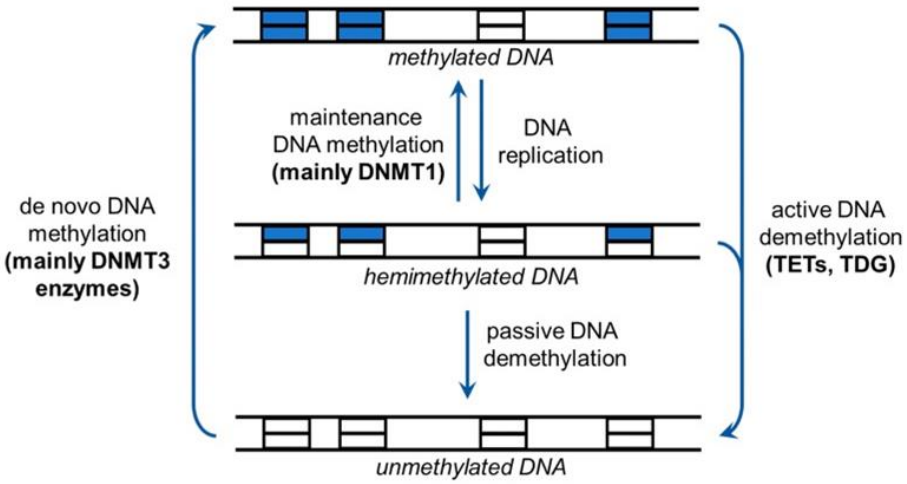


Figure 2: Classical model of DNA methylation setting and maintenance. Cell type-specific methylation patterns are established through de novo methylation of DNA by enzymes of the DNA methyltransferase 3 (DNMT3) family. This creates patterns of completely methylated (blue) or unmethylated (white) sites. During DNA replication the information of the parental strand is transferred to the new strand by DNA methyltransferase 1 (DNMT1) which specifically recognizes the hemimethylated (one strand blue, one strand white) sites occurring in DNA replication and maintains the methylation pattern. Without this process, DNA methylation could be passively lost. Furthermore, active DNA demethylation pathways involving the Ten-eleven translocation enzymes (TETs) and thymine DNA glycosylase (TDG) are known (taken from Jeltsch et al., 2018).

In addition to a passive loss of DNA methylation prevented by DNMT1 (see section 1.3.3), there is also an active pathway of demethylation known since 2009 that involves the family of Ten-eleven translocation enzymes (TETs) (Tahiliani et al., 2009). Following a pathway of step-wise oxidation, which is the subject of section 1.3.1, TET enzymes initiate the removal of the methylation mark with the help of thymine DNA glycosylase (TDG) and the base excision repair machinery (He et al., 2011; Ito et al., 2011).

The classical model of DNA methylation based on the early studies from 1975 has been challenged over the past years by experimental data suggesting a more cooperative function of DNMT3 and DNMT1 (Jeltsch and Jurkowska, 2014). Evidence for this new model came from studies in mammalian cell lines and mice where deletion of DNMT3A and DNMT3B resulted in a reduction of DNA methylation present at repetitive elements, even though DNMT1 was fully functional (Chen et al., 2003; Dodge et al., 2005). Based on this observation, both DNMT3 enzymes share the maintenance function with DNMT1 at least for some specific targets. Moreover, DNMT1 was shown to be essential for the process of *de novo* methylation, since DNMT3 enzymes were observed to preferentially generate hemimethylated CpG sites, which are then further methylated by DNMT1 (Fatemi et al., 2002). The main cause of this is the strong flanking sequence dependency of the DNMT3 enzymes (Handa and Jeltsch, 2005; Lin et al., 2002), which often leads to preferred methylation in one strand but disfavoured methylation in the other strand. In addition, it was proven that DNMT1 displays direct *de novo* methyltransferase activity on specific transposable elements (Haggerty et al., 2021) which would enhance the long-term repression of these regions.

Another argument against the old DNA methylation model arose from studies in human stem cells and differentiated neuronal progenitor cells, in which DNA methylation was also observed in asymmetric non-CpG (CpA, CpC, CpT) context, although to a reduced extent (Jang et al., 2017). The existence of non-CpG methylation was long suspected to be a methodical artefact (He and Ecker, 2015). However, the presence of this epigenetic mark was proven later on first in mice (Ramsahoye et al., 2000) and then in the human genome (Lister et al., 2009). Until today, the detection of methylation in CpH context is still difficult due to the high amount of genomic CpG methylation and requires high-resolution sequencing such as Illumina Next-Generation Sequencing (NGS) (Metzker, 2010). In comparison to somatic cells that contain only about 0.02% of their total cytosine methylation in CpH-context, this number increases up to 25% in human ESCs (Jang et al., 2017). In human induced pluripotent stem cells (iPSCs) that carry roughly 68% of methylation at CpG sites, methylation of more than 8% was observed in CpA context followed by 2% in CpT and 1% in CpC context (Ziller et al., 2011). Furthermore, non-dividing cells such as neurons showed gene bodies and transposons enriched with non-CpG methylation at >2% of all cytosines (Lister et al., 2013). Overall, non-CpG methylation was shown to correlate with repressed

transcription (Guo et al., 2014b) and states of pluripotency (Butcher et al., 2016). These asymmetric sites are no targets for DNMT1 methylation due to the missing methylation information on the template strand, leaving only DNMT3A and DNMT3B as possible methyltransferases for their maintenance. Indeed, biochemical experiments showed that DNMT3A and DNMT3B can introduce this epigenetic mark due to their less stringent specificity and the expression levels of both enzymes were shown to correlate with DNA methylation levels in non-CpG the context in ESCs and mammalian oocytes (Arand et al., 2012; Shirane et al., 2013).

Furthermore, active demethylation through the TET-TDG pathway can take place at both upper and lower strand, leading to fully unmethylated CpG sites that are no longer targets for DNMT1. At such sites, the *de novo* methyltransferases would need to reintroduce the DNA methylation marks again to maintain the methylation pattern (Jeltsch and Jurkowska, 2014). The specific interplay of DNMT-mediated methylation and TET oxidation has been recently shown to be critical for mammalian stem cells during their exit from pluripotency (Parry et al., 2021). In contrast to the global DNA methylation and demethylation waves that occur during other stages of embryonic development (Messerschmidt et al., 2014), which are generally attributed to either upregulation of TET or DNMT3A/DNMT3B expression, both enzyme families are co-expressed in this special transition stage towards differentiation (Parry et al., 2021). Strikingly, ESCs that were lacking DNMTs or TET enzymes were still pluripotent, but could not differentiate anymore (Dawlaty et al., 2014; Tsumura et al., 2006). The basis for this effect was shown to be the combination of the two highly active machineries together with passive DNA demethylation, which leads to fast and continuous turnover of the individual DNA methylation states, especially at distal regulatory elements of poor to mediate CpG content (Parry et al., 2021).

Together, these experimental findings forced the epigenetic field to change the old static concept into a more dynamic one. DNA methylation is now seen as a continuous process determined at each CpG site and genomic region by the local concentrations and catalytic activities of DNA methyltransferases with partly-overlapping functions as well as TET enzymes and DNA replication rates (Jeltsch and Jurkowska, 2014). This revised model is now able to explain that cells originating from the same tissue can have different methylation patterns as shown by bisulfite sequencing (Zhang et al.,

2009) due to stochastic changes in the enzymatic activities that lead only to the preservation of average methylation density profiles. Following this model, methylation mistakes can be corrected through the help of various feedback loops based on the crosstalk between DNA methylation and a complex network of chromatin marks. This either includes further recruitment of DNMTs at regions with repressive marks or reduction of methylation at regions of activating chromatin marks (Jeltsch and Jurkowska, 2014).

1.2.3 The family of DNA methyltransferases

In mammals, the family of DNA methyltransferases comprises four active enzymes and one catalytically inactive but regulatory protein (Figure 3). The *de novo* methyltransferases DNMT3A and DNMT3B set the cell-specific methylation patterns during early embryogenesis as well as gametogenesis which are then maintained by the maintenance methyltransferase DNMT1 (Jurkowska et al., 2011d). Furthermore, a methyltransferase termed DNMT3C that evolved from DNMT3B was recently discovered in rodents and shown to exclusively methylate retrotransposons in male germ cells (Barau et al., 2016). In addition to the active enzymes, the catalytically inactive DNMT3L was shown to act as a regulatory factor during *de novo* methylation (Jurkowska et al., 2011d).

As shown in Figure 3, all DNMTs share the same structural composition of two main parts which are connected by a linker sequence like a glycine-lysine repeat region in the case of DNMT1. The smaller C-terminal parts of all enzymes contain the highly conserved active centre as well as the cofactor binding site. The larger N-terminal part varies between the enzymes and contains different regulatory domains responsible for the nuclear localization of the enzymes as well as their interaction with DNA, specific histone post-translational modifications or other proteins (Jeltsch and Jurkowska, 2014). Unlike DNMT1 which was shown to lose its function after removal of the N-terminal domains (Fatemi et al., 2001; Zimmermann et al., 1997), DNMT3A and DNMT3B were observed to also be active in their isolated C-terminal form (Gowher and Jeltsch, 2002).

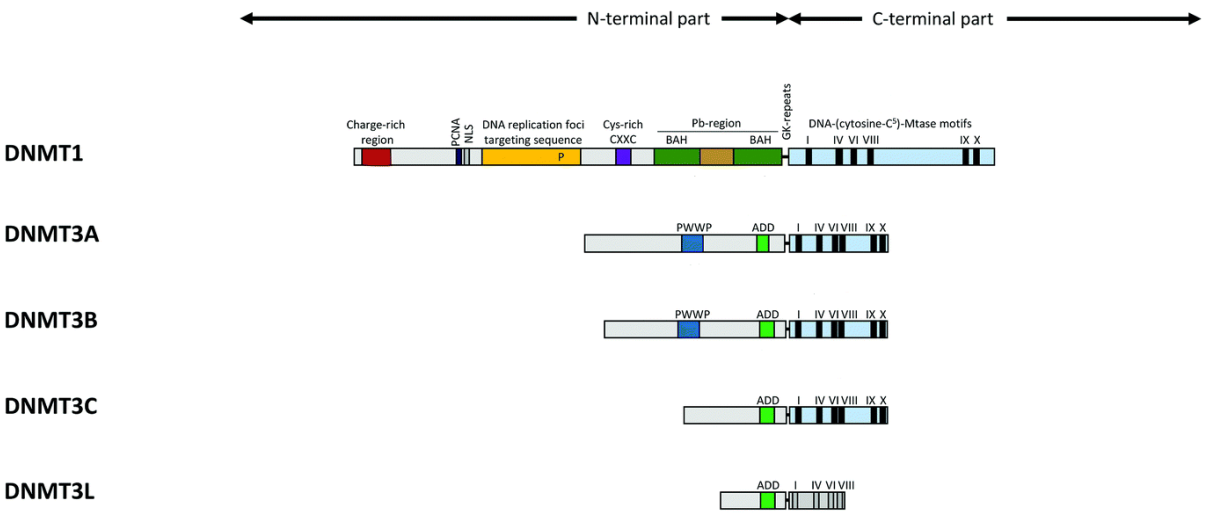


Figure 3: Schematic drawing of the domain arrangement in the DNMT family. All enzymes consist of a C-terminal part containing the ten conserved amino acid motifs responsible for the methyltransferase activity (except for the inactive DNMT3L) and a regulatory N-terminal part with several domains for interaction with DNA, chromatin or other proteins. PCNA, binding domain for proliferating cell nuclear antigen; NLS, nuclear localization signal; CXXC, cysteine-X-X-cysteine domain; BAH, Bromo-adjacent homology domains; GK, glycine-lysine repeats; PWWP, proline-tryptophan-tryptophan-proline domain; ADD, ATRX-DNMT3-DNMT3L domain (adapted from Ravichandran et al., 2018).

All prokaryotic and eukaryotic cytosine C5 DNA methyltransferases contain ten evolutionary conserved amino acid motifs that are essential for the methyltransferase activity (Gowher and Jeltsch, 2018). From these motifs, motif IV (also known as PCQ motif; PCN in case of DNMT3s), VI (also known as ENV motif) as well as IX are essential for catalysis. The binding of the cofactor is mediated by motifs I and X and DNA binding as well as the specificity of the DNMTs is dependent on the non-conserved region between the motifs VIII and IX. Overall, the domain has a typical MTase fold made up of a mixed seven-stranded sheet of six parallel β-strands and one antiparallel strand inserted between strands 5 and 6.

Structurally, all the DNMTs use a common principle base-flipping mechanism as shown for the bacterial methyltransferases HhaI (Klimasauskas et al., 1994) and HaeIII (Reinisch et al., 1995) in which the target cytosine is flipped out of the DNA double helix and inserted into the hydrophobic pocket of the active site. The transfer of the methyl group from the cofactor S-adenosyl-L-methionine (AdoMet) to position 5 of cytosine nucleobases then follows the catalytic mechanism depicted in Figure 4 (Lyko, 2018). At first, a nucleophilic attack of the catalytic cysteine residue from motif IV at

the C6 position takes place that covalently links the enzyme to the flipped-out cytosine. This leads to transient protonation of the N3 through the catalytic glutamate residue from motif VI. The next step includes the deprotonation of N3 by the glutamate residue that activates the C5 position and enables the nucleophilic attack on the methyl group of the cofactor AdoMet. In the end, β-elimination of the H5 proton by a base that has not been identified yet but could be a water molecule leads to the release of the methylated DNA and S-adenosyl-L-homocysteine (AdoHcy) from the enzyme. The fact that the initial nucleophilic attack of the cysteine and the attack on the methyl group must occur at opposite sides of the ring system is one of the reasons for the usage of the already described base-flipping mechanism.

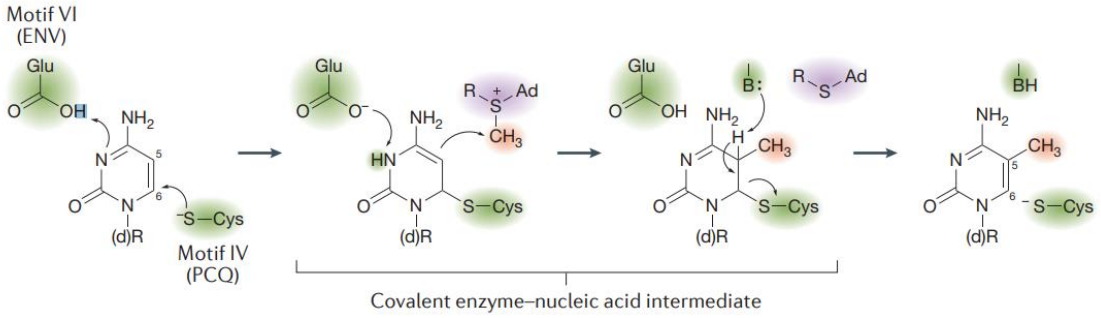


Figure 4: Catalytic mechanism of cytosine C5 DNA methyltransferases. Nucleophilic attack of the catalytic cysteine from motif IV (PCQ motif) coloured in green together with the catalytic glutamate from motif VI (ENV motif) leads to transfer of the red coloured methyl group from the cofactor AdoMet coloured in violet to the C5 position of cytosine followed by deprotonation with an unknown base (taken from Lyko, 2018).

Methylation-sensitive DNA-interacting proteins (Du et al., 2015) are then able to detect the incorporated methyl group in the major groove. Although the introduction of this modification does not alter the Watson-Crick pairing, it changes the contact profile of the major groove. Moreover, the hydrophobic character of the methyl group leads to a bending of the DNA together with changes in flexibility, thermostability and solvation state (Rausch et al., 2021).

1.2.3.1 *De novo* methyltransferases DNMT3A and DNMT3B

The establishment of CpG methylation patterns, as well as the preservation of non-CpG methylation, is mainly the role of the DNMT3 enzymes DNMT3A and DNMT3B assisted by DNMT3L. Although both active methyltransferases show high overlap in the sequence of their MTase domain (Okano et al., 1998), the enzymes were shown to have distinct temporal expression patterns and functions. While expression of DNMT3A mainly takes place in oocytes and early stages of embryonic preimplantation where the enzyme is responsible for allele-specific imprinting control during gametogenesis, DNMT3B is predominantly expressed in the blastocyst stage during post-implantation development (Kaneda et al., 2004; Kato et al., 2007; Watanabe et al., 2002). Furthermore, both enzymes are involved in the preservation of methylation states of various repetitive elements, examples being the human Satellite II (SatII) repeats in case of DNMT3B or major satellite repeats in case of DNMT3A (Chen et al., 2003). Indeed, several knockout studies emphasized the critical roles of all three DNMT3 enzymes, since mice were only partly viable after DNMT3A knockout, the absence of DNMT3B was lethal (Okano et al., 1999) and loss of DNMT3L also led to embryonic lethality or infertility (Bourc´his et al., 2001). The latter is an indirect effect on *de novo* methylation, since DNMT3L lacks catalytic activity itself but enhances the activity of DNMT3A and DNMT3B and thereby regulates their structural organization and nuclear localization (Jurkowska et al., 2011b).

Mechanistic insights into the interaction of DNMT3A and DNMT3L were obtained later on through several crystal structures that were published without (Jia et al., 2007) or with co-crystallized DNA (Zhang et al., 2018). These structures revealed that DNMT3A can dimerize and form distinct linear heterotetramers with DNMT3L with the specific arrangement of 3L-3A-3A-3L, a fact that was also biochemically shown through ultracentrifugation and size exclusion chromatography (Jurkowska et al., 2008; Jurkowska et al., 2011b; Nguyen et al., 2019). Accordingly, the two active subunits located in the centre of the tetramer mediate the 3A-3A interaction at the so-called RD interface, while the two inactive DNMT3L subunits located at the outer positions are involved in the formation of two peripheral 3A-3L or FF interfaces (Figure 5A) (Jeltsch and Jurkowska, 2016). The naming is hereby based on the amino acids that predominantly make up these interfaces, being arginine and aspartate residues building a contact network in case of the polar RD interface, or four stacked

phenylalanine residues forming the hydrophobic FF interface. Like this, the central subunits build up the DNA binding site of the complex, while the outer subunits are not in contact with the DNA but rather interact with residues that are involved in catalysis or binding of the cofactor. The contacts through the FF interface result in the increased methyltransferase activity of DNMT3A in the presence of DNMT3L, which is not able to form RD interfaces by itself due to the lack of essential motifs in its MTase domain.

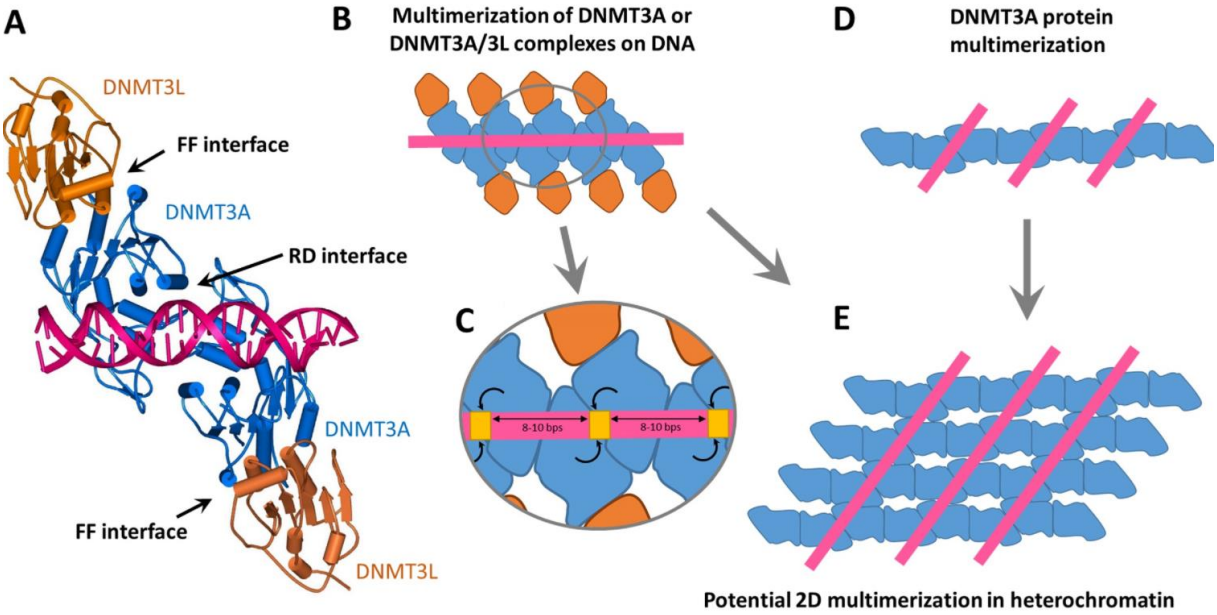


Figure 5: Schematic illustrations of DNMT3A and DNMT3A/3L complex structure and multimerization. Heterotetramers of two DNMT3A and two DNMT3L subunits (shown in A in blue and orange, respectively), as well as homotetramers, can form and bind to one DNA strand or various DNA strands (pink) in parallel orientation with different potential for multimerization (shown in B and D, respectively). Like this, co-methylation can occur on both strands of CpG sites (shown in C) and a combination of both multimerization processes can lead to the formation of 2D methylation networks (shown in E) (taken from Jeltsch and Jurkowska, 2016).

In contrast to DNMT3L, DNMT3A can also form homotetramers followed by potential multimerization in not only a horizontal (Figure 5B) but also a vertical direction (Figure 5D) with several DNA binding sites located on one or more DNA molecules in parallel arrangement (Figure 5C). This can lead to the formation of DNMT3A filaments or even DNMT3A networks (Figure 5E), which was also experimentally verified by scanning force microscopy and Förster resonance energy transfer assays (Jurkowska et al., 2008; Jurkowska et al., 2011b; Rajavelu et al., 2012).

Overall, DNA binding of multiple homo- or heterotetramer complexes occurs non-specifically (Rajavelu et al., 2012) but in a cooperative manner, meaning that binding of new complexes to a DNA molecule preferentially takes place next to already bound ones. This property of DNMT3A has been studied intensively (Emperle et al., 2014; Jia et al., 2007; Jurkowska et al., 2008; Rajavelu et al., 2012), nevertheless, contradicting results can also be found in the literature (Holz-Schietinger and Reich, 2010).

In addition to the subunit organization and formation of the essential interfaces, the crystal structure published in 2007 already gave some hints about how CpG sites are methylated even though there was no co-crystallized DNA present in this complex (Jia et al., 2007). Interestingly, structural constraints on the two central active subunits hinder the simultaneous methylation of both strands of a CpG target site despite its palindromic nature. Instead, parallel co-methylation of two CpG sites in opposite strands was suggested, whereby a CpG distance of roughly 10 bps would fit the ~40 Å separation of the inner subunits. Consistent with this hypothesis, biochemical as well as cellular data of different mammalian studies showed the same 10 bps periodicity for CpG (Jia et al., 2007; Jurkowska et al., 2008) and even non-CpG methylation (Lee et al., 2017), creating characteristic DNA methylation profiles. Nevertheless, adjacent DNMT3 complexes in an oligomeric structure would still be able to methylate the upper or lower strand of one CpG target site, respectively, giving rise to some flexibility during the DNA methylation process.

Recently, new insights could be obtained into the complex mechanism of DNMT3A/DNMT3L when a new crystal structure was solved in 2018 with co-crystallized DNA containing two CpG target sites or two short DNA substrates containing one site each (Zhang et al., 2018). To obtain stable structures, target cytosines were replaced with the nucleotide analogue zebularine. Usage of this cytosine analogue has the advantage that a stable covalent complex is formed between the cysteine in the active centres of the enzyme and the target cytosines due to the prevention of the β-elimination step in the catalytic mechanism (Osterman et al., 1988). This new crystal structure was in good agreement with the DNA-free structure published almost 10 years earlier (Jia et al., 2007) except for a part of the target recognition domain (TRD) which appeared to be disordered in the absence of DNA but ordered upon its binding. In addition to the TRD loop, the catalytic loop, as well as the

RD interface, were determined to be the essential DNA interacting regions. Furthermore, a distance of 12 bps instead of 10 bps between the two target sites on one DNA strand was found to be preferred for co-methylation, a discrepancy which still had to be resolved. Surprisingly, the two different structures determined in this new study also showed variance regarding the bending of the DNA upon binding to the active centre. One structure showed the DNMT3A/3L heterotetramer in contact with two short individual DNA molecules, which were both unbent and bound to one inner subunit of the complex, respectively. In contrast, the longer DNA with two target CpG sites showed compression of the major groove due to approximately 40° bending in the middle of the substrate. Lastly, several DNMT3A residues were observed to be involved in specific contacts with the bound DNA, leading to distinct flanking sequence preferences, as further discussed in section 1.4.

Despite the tremendous knowledge obtained for DNMT3A or DNM3A/3L methylation, little was known about the mechanism or organization of DNMT3B complexes due to the absence of a crystal structure until the start of this thesis. Several biochemical studies have shown that DNMT3B methylates DNA in a processive manner (Norvil et al., 2018) instead of the cooperative mechanism of DNMT3A (Jia et al., 2007; Jurkowska et al., 2008; Rajavelu et al., 2012) and that DNMT3B also harbours intrinsic preferences for specific flanking sequences, although they differ from DNMT3B (see section 1.4). Moreover, the sequence similarity of both DNMT3 enzymes especially in the conserved interfaces (Okano et al., 1998) would suggest a potential self-oligomerization of DNMT3B, but this could not be proven so far. Therefore, further insights into the mechanism of the *de novo* methylation by DNMT3B needed to be obtained through future experiments and for DNMT3A, the described discrepancies had to be clarified.

1.2.3.2 The maintenance methyltransferase DNMT1

DNMT1 was the first mammalian cytosine C5 DNA methyltransferase to be cloned and sequenced (Bestor et al., 1988). The protein is a large polypeptide of around 180 kDa comprising 1616 or 1620 amino acids in the human and mouse enzyme, respectively. Its amino acid sequence was later shown to be highly conserved between different species (Jurkowski and Jeltsch, 2011). Since DNMT1 is the major enzyme responsible for the maintenance of DNA methylation, it was not surprising that knockout of the enzyme or loss of its catalytic function were observed to be fatal in various studies. Disruption of the *DNMT1* gene in mice led to delays in embryonic development and death of the embryos shortly after gastrulation (Li et al., 1992). A similar phenotype was observed in mice containing a mutation of *DNMT1* in both alleles, rendering the enzyme catalytically inactive (Takebayashi et al., 2007). In addition, deletion of DNMT1 was shown to be lethal in all proliferating somatic cells (Fan et al., 2001; Sen et al., 2010; Trowbridge et al., 2009) and a strong global reduction in the DNA methylation level was observed after disruption of the gene in ESCs (Liao et al., 2015). Overall, these results support the major role of DNMT1 during all embryonic developmental stages.

Expression levels of DNMT1 depend strongly on the cell type and the cell cycle stage. Non-dividing cells were found to express only low levels of the enzyme in contrast to the high expression levels in all mitotic cells (Kishikawa et al., 2003; Lee et al., 1996; Robertson et al., 2000). Furthermore, the levels of DNMT1 expression, as well as its localization in the nucleus, are strongly regulated during cell proliferation. The protein shows a dynamic subnuclear localization pattern, including various spotty patterns in the early, middle and late S-phase corresponding to sites of active DNA replication in specific chromatin states. The structures develop from small spots of DNMT1 associated with foci with ongoing replication in the euchromatin to larger toroidal structures of DNMT1 at less abundant replication sites on centromeric heterochromatin. This specific localization pattern is lost during other stages of the cell cycle, in which the enzyme shows a diffuse distribution throughout the nucleus (Schneider et al., 2013).

In all S-phase stages, the punctuate pattern of DNMT1 in the cells is a result of its association with replication forks and the hemimethylated CpG sites that are occurring after DNA replication and for which the methylation state has to be restored. For this specific maintenance function, DNMT1 displays a very high preference for hemimethylated over unmethylated CpG sites, ranging from 15-fold in *in vitro* experiments using a long 634 bps substrate with multiple hemimethylated target sites to 30-40-fold on short 30mer oligonucleotides containing one hemimethylated CpG site (Fatemi et al., 2001; Goyal et al., 2006; Pradhan et al., 1999) depending on the experimental conditions. In contrast to that, DNMT3A was shown to not discriminate between unmethylated and hemimethylated target sites (Gowher and Jeltsch, 2001).

As a maintenance methyltransferase, DNMT1 was also proven to be highly processive referring to its ability to consecutively introduce methylation at up to 30 CpG sites without dissociation from the DNA (Goyal et al., 2006; Hermann et al., 2004; Vilkaitis et al., 2005) with a very low skipping rate of under 0.3% (Goyal et al., 2006). Since it was shown in these studies that processive methylation only occurs on one DNA strand without changing the target strand, DNMT1 was proposed to slide along the daughter strand while DNA replication takes place.

Despite all the enzymatic properties unravelled so far, stable maintenance methylation of all 56 million CpG sites in the human genome by DNMT1 seems to be impossible during S-phase. Therefore, additional mechanisms involving other proteins and cofactors present in the cell are required to further enhance the activity and specificity of the methyltransferase. During the last 20 years, this complex network around DNMT1 has been studied intensively and several key players have been identified (Petryk et al., 2021).

Ubiquitin-like, containing PHD and RING finger domains 1 (UHRF1), an E3 ubiquitin ligase, was shown to preferentially recognize hemimethylated CpG sites through its SET-and RING-associated (SRA) domain, which leads to allosteric activation of the protein that allows ubiquitination of histone H3 and/or auto-ubiquitination (Fang et al., 2016). Both modifications are recognized by the replication-focus-targeting-sequence (RFTS) domain of DNMT1, which guides the methyltransferase to the hemimethylated sites (Li et al., 2018) and leads to allosteric activation of the enzyme through the release of the auto-inhibitory RFTS domain from the catalytic centre (Jeltsch and

Jurkowska, 2016). UHRF1 also recognizes repressive chromatin marks such as H3K9 trimethylation (Nady et al., 2011) together with unmodified H3R2 (Hu et al., 2011; Wang et al., 2011) using a combinatorial readout involving its tandem tudor and plant homeodomain, which was demonstrated to be essential for H3 ubiquitination and subsequent DNA methylation (Qin et al., 2015).

In addition, proliferating cell nuclear antigen (PCNA), a ring-shaped DNA clamp interacting with DNA polymerase δ, was shown to colocalize with DNMT1 at replication foci (Chuang et al., 1997; Easwaran et al., 2004) and the presence of PCNA enhances the binding of the methyltransferase to DNA (Iida et al., 2002). DNMT1 was observed to interact with PCNA through its PCNA-binding domain in the N-terminal part of the protein, but the detailed mechanism of this interaction was unknown until the crystal structure of PCNA with the interacting domain of DNMT1 (PIP box motif) was solved (Jimenji et al., 2019). Indeed, mutation of this binding domain was shown to reduce the maintenance methylation by a factor of 2 in comparison to the wild-type DNMT1 (Egger et al., 2006; Spada et al., 2007). Overall, disruption of the complex network around DNMT1, UHRF1 and PCNA led to global DNA hypomethylation in brain, lung, breast and mesothelial cells which is an oncogenic event in human tumorigenesis (Pacaud et al., 2014).

Regarding the specificity of DNMT1, several crystal structures published in 2011 and especially in 2012 provided mechanistic insights into substrate recognition of the methyltransferase (Song et al., 2011; Song et al., 2012). The 2012 work provided the structure of a truncated mouse DNMT1 variant (amino acids 731-1602, PDBI: 4DA4) co-crystallized with a 12 bps oligonucleotide with a single 5-fluorocytosine containing hemimethylated CpG site. As depicted in Figure 6, the DNA was bound in the catalytic cleft with the methylated target 5-fluorocytosine flipped out of the DNA double helix and placed in the active site. In addition, the recognition of the hemimethylated CpG site was shown to be mediated via a hydrophobic surface in the TRD subdomain around the methyl group in the major groove. DNA contacts were formed by two loops of this subdomain in the major and one catalytic loop penetrating the minor groove of the DNA. Surprisingly, in this structure, the space of the flipped-out cytosine was occupied by two amino acids from DNMT1 and several structural reorganizations of the DNA around the target site took place.

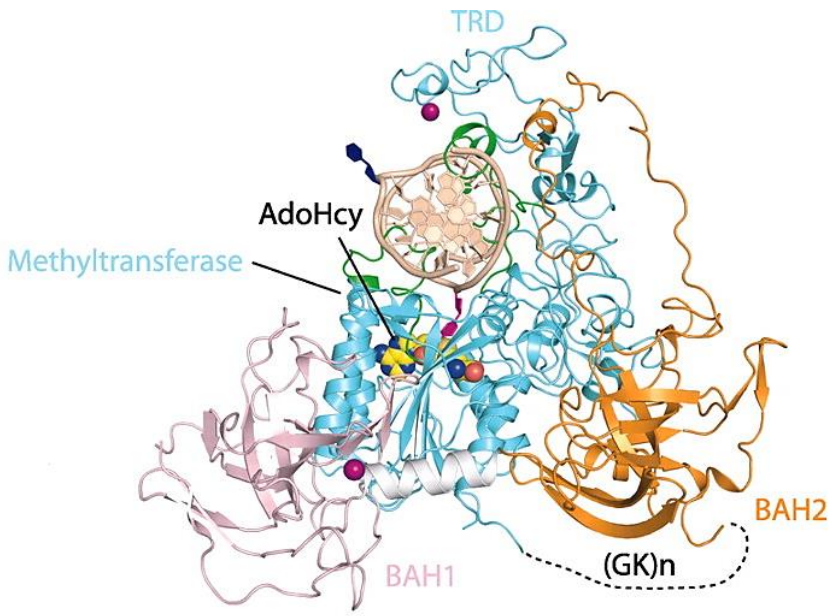


Figure 6: Structure of mouse DNMT1 co-crystallized with hemimethylated DNA. Bromo-adjacent homology domains 1 and 2 (BAH1 and BAH2) are coloured light pink and orange, the catalytic and target recognition domain (TRD) loops 1 and 2 are coloured green and the methyltransferase domain is shown in cyan; black dashed lines are used to show the disordered [(GK)n] linker; the DNA is coloured barley with the flipped-out 5-fluorocytosine and the parental cytosine shown in purple and blue, respectively; Coordinated zinc ions are shown in purple and the bound cofactor S-adenosyl-L-homocysteine (AdoHcy) is shown in a space-filling representation (adapted from Song et al., 2012; PDBI: 4DA4).

1.2.4 Acute myeloid leukaemia and the role of the DNMT3A R882H mutation

The relevance of faithful establishment and maintenance of epigenetic marks became increasingly clear during the last decades, in which the development and progression of many diseases and disorders could be linked to aberrant epigenetic processes (Robertson, 2005). In the case of DNA methylation, reduction of global methylation levels in human cancer cells compared to normal cells had already been observed in 1983 (Feinberg and Vogelstein, 1983), a hallmark which was later connected to the hypomethylation of high numbers of repetitive elements (Weisenberger et al., 2005) leading to genomic instability due to inefficient silencing of transposable elements.

Generally, two types of carcinogenic alterations have to be distinguished: “driver” alterations that lead to enhanced proliferation of cancer cells and therefore promote the development of the disease, and “passenger” alterations that do not have this influence but accumulate by chance in the clonal selection of cell lines as cancer progresses (Pon and Marra, 2015; Roy et al., 2014). Although the exact classification of the different epigenetic changes is difficult, aberrant global DNA methylation levels were shown to occur frequently and in an early stage of cancer development (Sonnet et al., 2014). The cause for these changes often lies in different somatic mutations of DNMT3A, a condition that was shown to be prevalent in different types of malignancies such as lung (Gao et al., 2011) and haematological cancers (Yang et al., 2015) or developmental disorders (Tatton-Brown et al., 2014). Strikingly, mutations of DNMT3A were observed to be especially enriched in 20-40% of patients with acute myeloid leukaemia (AML) (Ley et al., 2010; Yamashita et al., 2010; Yan et al., 2011), a type of haematological cancer that originates from granulocytes or monocytes in the bone marrow and causes unregulated proliferation and subsequent accumulation of malignant, immature white blood cells (Ferrara and Schiffer, 2013). The pathogenesis of AML is a multistage process as depicted in Figure 7, where mutations in epigenetic regulators such as DNMT3A tend to occur in the early stages, leading to a pre-leukemic state (Li et al., 2016a; Papaemmanuil et al., 2016; Sato et al., 2016) which develops into heterogenic cancer cells if even more mutations accumulate. Like this, DNMT3A mutations drive the carcinogenic progression, which negatively influences the prognosis of treatment efficiency, remission time and survival rate (Brunetti et al., 2017).

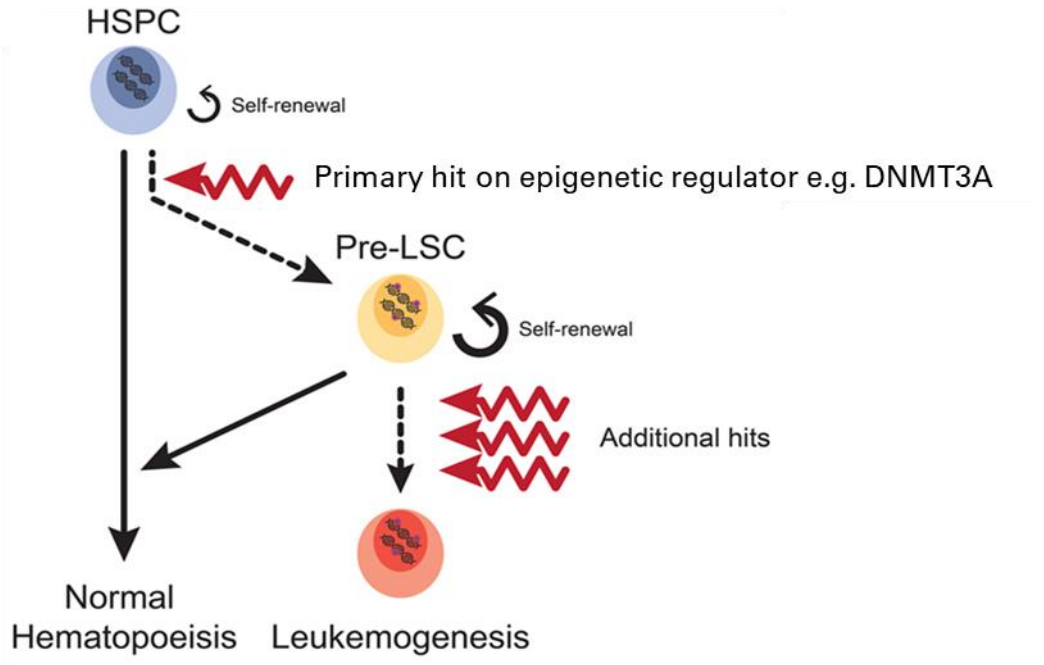


Figure 7: Multistage process of AML development. A pre-leukemic state (Pre-LSC) is induced first through primary mutations of epigenetic regulators such as DNMT3A and TET2 occurring in hematopoietic stem cells or committed progenitor cells (HSPC). In this state, cells still contribute to self-renew and normal haematopoiesis. However, further mutations drive the pathogenesis, leading to different subpopulations of AML cells (adapted from Sato et al., 2016).

A more detailed look at the DNMT3A mutations frequently observed in AML patients revealed that most of them are clustered at the MTase as well as the proline-tryptophan-tryptophan-proline (PWWP) and ATRX-DNMT3A-DNMT3L (ADD) domains (Figure 8). Generally, a strong enrichment of 73% missense mutations could be determined, with most of them occurring in a heterozygous manner combined with an intact wild-type allele that still expresses the unmutated DNMT3A (Brunetti et al., 2017). Hereby, the most common mutated residue is R882, which was shown to account for roughly 60% of the missense mutations and is predominantly converted to histidine (two-thirds) or cysteine (one-third) and only rarely to serine and proline.

The location of R882 in the RD interface of DNMT3A complexes and the involvement in DNA binding through contacts to the DNA backbone (Zhang et al., 2018) hint towards a specific molecular effect of this mutation. For this reason, and due to the high prevalence of the mutation, several groups have investigated the mechanistic and pathogenic mechanisms of the R882H mutation, but with partly contradicting results (Brunetti et al., 2017; Marcucci et al., 2012). First, the catalytic activity of the mutant

was shown to be reduced by 30-50% in *in vitro* experiments compared to the wild-type enzyme (Emperle et al., 2018a, Emperle et al., 2018b; Holz-Schietinger et al., 2012; Yan et al., 2011). Observed effects were even more striking *in vivo*, although conflicting models of interaction with the co-expressed wild-type enzyme were proposed, including a dominant-negative effect (Kim et al., 2013; Russler-Germain et al., 2014). Similarly, the influence of the mutation on the potential oligomerization of DNMT3A is still under debate, with one group presenting data that supports an impaired dimerization of R882H subunits leading to a subsequent decrease in the processivity of the enzyme (Holz-Schietinger et al., 2012). In contrast, other groups showed increased multimerization of the mutated DNMT3A (Nguyen et al., 2019), changes in the enzymatic activity under different pH conditions (Holz-Schietinger and Reich, 2015) or the cooperative binding to DNA (Norvil et al., 2018). Furthermore, R882H was not only linked to global hypomethylation of CpG island and shore regions in AML (Qu et al., 2014) but it was also associated with hypermethylated promotors (Yan et al., 2011). Finally, the location of R882H in the RD interface was shown to influence recognition of the CpG target sites, with the observation of strong flanking sequence preferences distinct from wild-type DNMT3A (Emperle et al., 2018b) as further discussed in section 1.4. In summary, several consequences of the DNMT3A R882H mutation are still not fully unravelled, so further insights into this highly relevant AML mutation need to be obtained through future experiments.

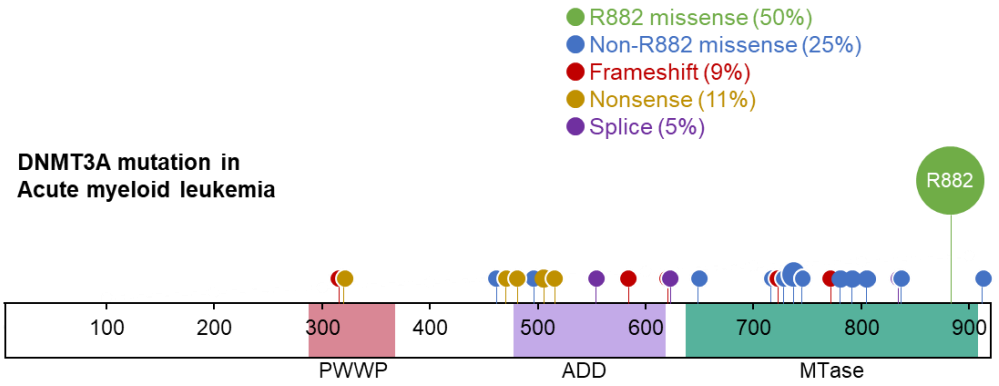


Figure 8: Mutations of DNMT3A occurring in AML. Nonsynonymous mutations are depicted on the schematic protein domain arrangement and presented as lollipops with colours referring to the type of mutation and size correlating with the mutation count (based on Brunetti et al., 2017).

1.3 DNA Demethylation

1.3.1 Principles of DNA demethylation and roles of its products

DNA methylation as an epigenetic modification was long thought to be relatively stable due to the chemically inert character of the C-C-bond, therefore removal of this mark was suspected to take place through replication-dependent loss of the methylation mark in the absence of proper DNMT1 function (see section 1.3.3). However, this mechanism appears to be insufficient to account for the two main demethylation events happening during embryonic development, demethylation in the paternal genome after fertilization but before the first DNA replication occurs and genome-wide demethylation during germ cell specification (Messerschmidt et al., 2014).

The existence of an active demethylation pathway was proven in 2000 when two independent groups observed a replication-independent global loss of DNA methylation in mouse zygotes (Mayer et al., 2000; Oswald et al., 2000), but the mechanism remained unclear until the Ten-eleven translocation (TET) enzyme TET1 was discovered in 2009. This enzyme was shown to oxidize 5-methylcytosine (m5C) to 5-hydroxymethylcytosine (hm5C) *in vitro* and *in vivo* (Tahiliani et al., 2009). The enzyme was identified in a computational screen as the mammalian analogue of the trypanosome dioxygenases JBP1 and JBP2, which oxidize thymine to 5-hydroxymethyluracil using the cofactors Fe(II) and 2-oxoglutarate (2OG) (Cliffe et al., 2009; Yu et al., 2007). Later on, the same oxidation activity was confirmed for the other two members of the TET family, TET2 and TET3 (Ito et al., 2010).

5-hydroxymethylcytosine as a modified base had already been discovered in 1952 in the genomes of T-even bacteriophages T2 and T4 as part of their restriction-modification system (Wyatt and Cohen, 1952). Due to contradicting experiments focused on the abundance of hm5C in mammals (Kothari and Shankar, 1976; Penn et al., 1972), the formation of this modified base was suspected to result from oxidative damage until the breakthrough in 2009. Using mouse ESCs, the authors could show that the oxidation product hm5C accounts for 0.03% of all nucleotide bases genome-wide with m5C being 14-fold more abundant (Tahiliani et al., 2009). The hm5C content was even higher in mouse granule cells and purkinje neurons, in which it constitutes 0.2% and 0.6% of all nucleotides, respectively (Kriaucionis and Heintz, 2009). Lastly, the abundance of hm5C was found to be highest in brain cells and ESCs (0.7% and

0.4% of dG, respectively), but also other mouse tissues such as lung, kidney, heart and muscle were shown to contain this modification (Globisch et al., 2010). Overall, hm5C levels varied strongly between the different tissue types, whereas m5C levels were relatively constant. Throughout the genome, hm5C was shown to occur non-overlapping with m5C and the modification was found especially enriched in euchromatic regions. In brain tissue and ESCs, hm5C was mostly observed at transcription start sites (TSS), promotors with moderate to low CpG content or in gene bodies (Shi et al., 2017), where a positive correlation with gene expression was observed (Ficz et al., 2011). Given the fact that special reader proteins such as UHRF2 (Spruijt, et al., 2013; Zhou et al., 2014) have been identified for hm5C, the oxidized cytosine species is nowadays also considered to play a role as a stable epigenetic mark whose specific biological role is still unclear. Unsurprisingly, the loss of this epigenetic modification is discussed as another hallmark of cancer (Ficz and Gribben, 2014).

The full pathway of active demethylation (schematically shown in Figure 9) was unravelled two years later when two groups (He et al., 2011; Ito et al., 2011) independently showed that the TETs can further oxidize hm5C to 5-formylcytosine (f5C) and 5-carboxylcytosine (ca5C), similar to the stepwise oxidation of thymine by thymine-7-hydroxylase (Liu et al., 1973; Neidigh et al., 2009). Thymine DNA glycosylase (TDG) then recognizes these higher oxidized bases and hydrolyses the bond between the base and the deoxyribose, thereby creating an abasic site in the DNA which is replaced with an unmodified cytosine by the base excision repair machinery (BER) (He et al., 2011). The role of TDG was confirmed since knockdown of the enzyme led to a 10-fold increase in the levels of f5C and ca5C in mouse ESCs (Cortellino et al., 2011; Raiber et al., 2012). However, since its expression levels are low in the zygote and loss of TDG did not change the zygotic demethylation (Guo et al., 2014a), the enzyme cannot be the only DNA glycosylase involved. Levels of f5C and ca5C were found to be 100-1,000-fold lower than hm5C (Carell et al., 2018), but evidence for their biological relevance has been found (Lu et al., 2015; Song et al., 2013) so these less abundant bases could still have roles as epigenetic marks.

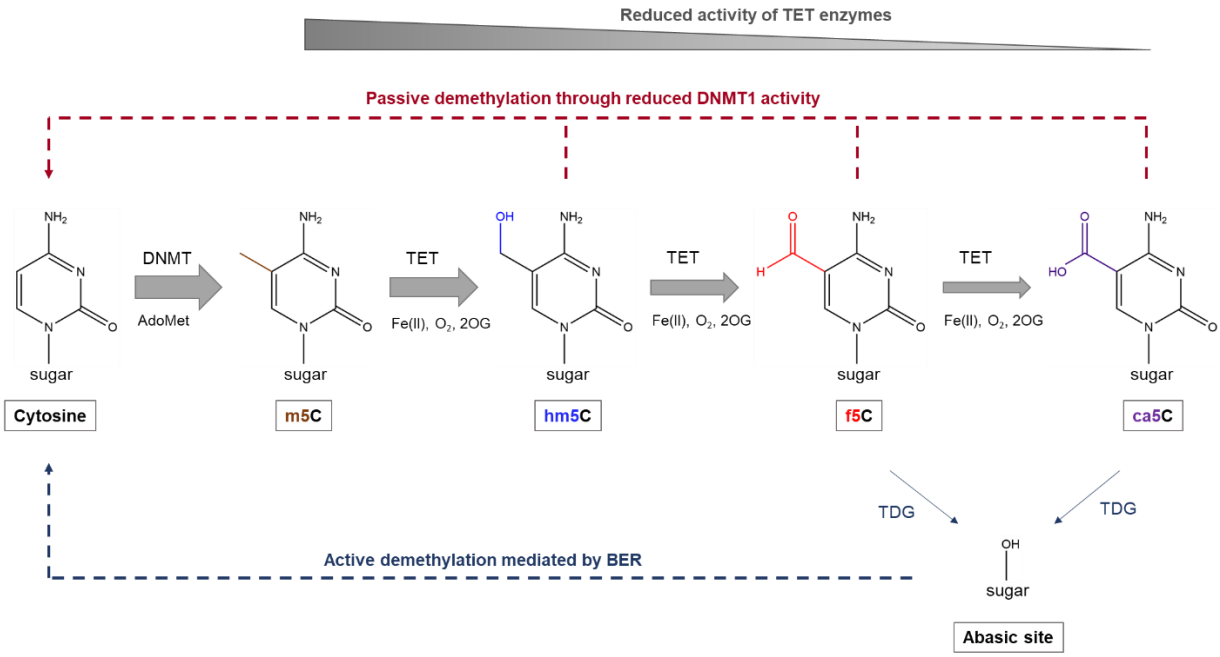


Figure 9: Pathways of passive and active DNA demethylation. DNA methylation is set by DNA methyltransferases (DNMT) at the C5 position of cytosine creating 5-methylcytosine (m5C). Using Fe(II), 2-oxoglutarate (2OG) and molecular oxygen as cofactors, the Ten-eleven translocation (TET) enzymes sequentially oxidize the methyl group yielding 5-hydroxymethylcytosine (hm5C) followed by 5-formylcytosine (f5C) and 5-carboxylmethylcytosine (ca5C). The last two modified bases are targets for the thymine DNA glycosylase (TDG), which excises the modified base which is then replaced with unmodified cytosine by the base excision repair (BER) machinery. This active demethylation pathway occurs in parallel to passive demethylation through inhibition of DNMT1 by hemi-modified CpG sites containing oxidized m5C species (based on An et al., 2017).

1.3.2 Ten-eleven translocation enzymes

In mammals, there are three different members of the TET family, namely TET1, TET2 and TET3, which have both overlapping and distinct functions depending on the cell type. While TET1 and TET2 were shown to be highly expressed in ESCs and the inner cell mass, the expression of TET3 was the highest in the oocyte and zygote, making this enzyme the main candidate responsible for the major demethylation during embryonic development. After differentiation, the levels of all TET enzymes generally decrease (Rasmussen and Helin, 2016).

All members of the TET family consist of a highly conserved C-terminal part and a less-conserved N-terminal part as shown in Figure 10 for human TETs (Rasmussen and Helin, 2016). The core catalytic domain contains a double-stranded β-helix domain

(DSBH) with binding sites for Fe(II) and 2OG, which adopts a characteristic fold for dioxygenases dependent on these cofactors (Iyer et al., 2009). The domain also harbours a low-complexity region increasing in size from TET1 to TET3, whose function remains unknown but was shown to not affect the enzymatic activity upon deletion (Hu et al., 2013). As shown for the DNMT3s, the isolated C-terminal domain on the TET enzymes alone can localize to the nucleus where it converts m5C to hm5C (Ito et al., 2010; Tahiliani et al., 2009). Overall, a large number of isoforms were found over the last years generated by alternative splicing or differential usage of the promotors (Melamed et al., 2018).

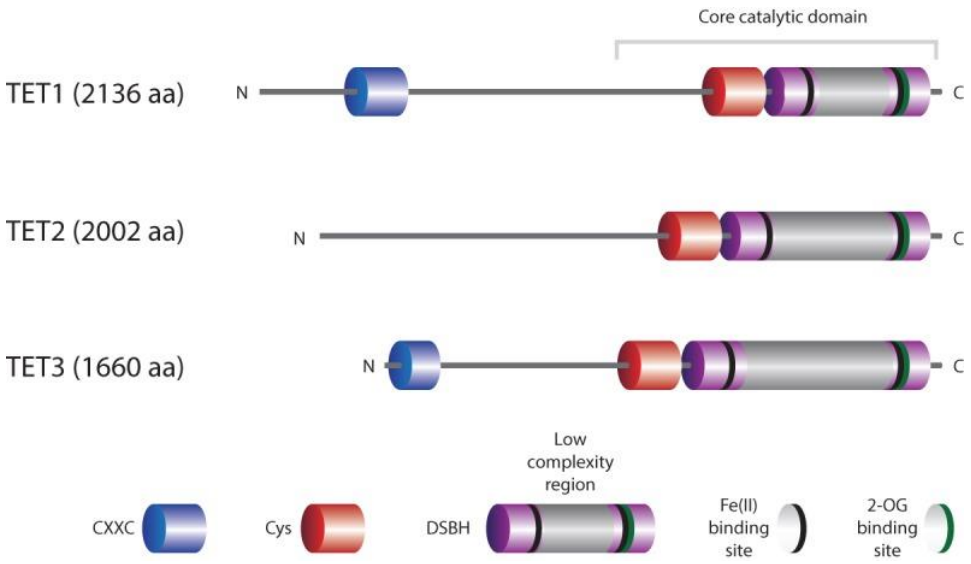


Figure 10: Schematic drawing of the domain arrangement in the human TET family. The C-terminal core catalytic domain consists of a cysteine-rich (Cys, shown in red) region and the double-stranded β-helix domain (DSBH, shown in violet), which contains the two binding sites for the cofactors Fe(II) and 2-oxoglutarate (2OG) (shown in black and green, respectively) and a low-complexity region (shown in grey). The N-terminal part of TET1 and TET3 contains a zinc finger cysteine-X-X-cysteine (CXXC, shown in blue) domain that was evolutionarily lost for TET2 (taken from Rasmussen and Helin, 2016).

Similar to the DNMTs, the TET enzymes use a base-flipping mechanism, in which the target cytosine is flipped out of the DNA double helix and inserted into the active site, resulting in a DNA bending of about 40° (Hu et al., 2013). Generally, the catalytic mechanism of the TET enzymes consists of two main parts, the activation of dioxygen followed by the oxidation of the respective substrate as depicted in Figure 11 (Parker et al., 2019). Firstly, Fe(II) and water are coordinated in the active centre through a HXD triad with X being any amino acid residue. The catalytic cycle then starts with the

binding of 2OG resulting in an active form of the enzyme, which in turn leads to the binding of the DNA substrate. After the replacement of one coordinated water molecule, one oxygen atom of molecular dioxygen binds to the Fe(II). Activation of the dioxygen occurs through the insertion of the unbound oxygen atom after decarboxylation of 2OG to succinate, followed by the cleavage of the bond between the two oxygen atoms. Like this, a highly reactive Fe(IV)-oxo intermediate is formed (Krebs et al., 2007; Valegard et al., 2004), which abstracts a hydrogen from the (modified) methyl group of the DNA substrate. The substrate radical then attacks the Fe(III)-hydroxide complex, resulting in the release of the oxidized substrate and the reduction to Fe(II) (Hoffart et al., 2006; Price et al., 2003). After the dissociation of succinate, the oxidation cycle can start again after the binding of 2OG and oxygen following the same mechanism. Since the +2 oxidation state of Fe(II) is essential for the activity of the TET enzymes, reducing agents such as ascorbic acid can enhance the oxidation reaction (Blaschke et al., 2013; Minor et al., 2013; Yin et al., 2013).

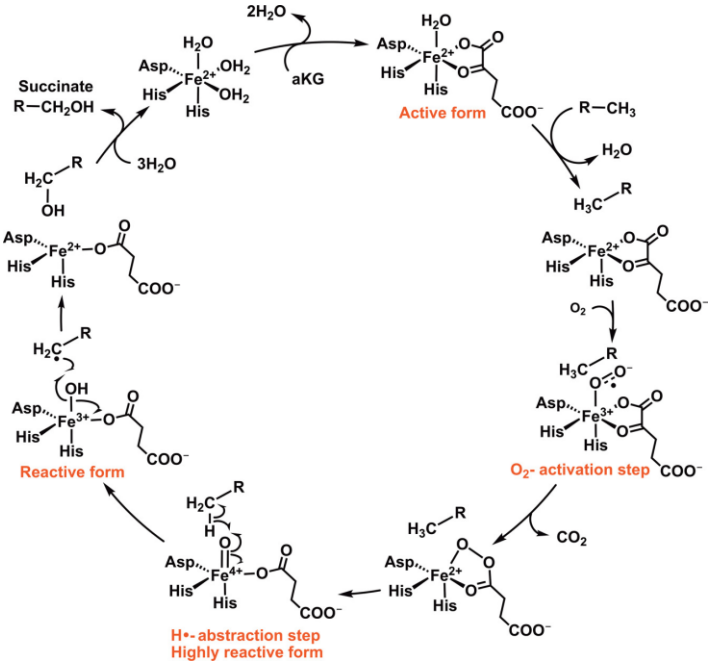


Figure 11: Catalytic cycle of TET enzymes. The oxidation follows a radical mechanism that is initiated through the binding of the cofactors Fe(II) anoxoglutaraterate to the active site of the TET enzyme (taken from Parker et al., 2019).

1.3.2.1 TET1

The first member of the TET enzyme family, TET1, is the largest enzyme in the group with a length of 2136 amino acids. It contains the largest N-terminal part but the shortest C-terminal part due to the smallest size of the unstructured region in the DSBH domain (see Figure 10 for domain arrangements of human TET1). Therefore, the catalytic domain of TET1 was the easiest to purify and the first TET enzyme for which oxidation of m5C to hm5C was demonstrated (Tahiliani et al., 2009). Analyses of the hm5C distribution across various tissue and cell types revealed that TET1 shows high expression levels in mouse ESCs and primordial germ cells (PGCs) correlating with hm5C enrichment (Yamaguchi et al., 2012), but its levels are decreasing during differentiation. In ESCs, it was shown to be involved in the regulation of stem cell maintenance by suppressing the expression of factors related to differentiation (Dawlaty et al., 2011; Williams et al., 2011; Wu et al., 2011). In PGCs, TET1 is highly expressed and it contributes to the massive demethylation that occurs genome-wide to regulate the methylation of imprinting genes (Hackett et al., 2013; Yamaguchi et al., 2013). Compared to that, either no or only low expression levels were observed in the oocyte and zygote (Wossidlo et al., 2011). TET1 was observed to be present in high levels in iPSCs from mouse embryonic fibroblasts, where deletion of all TET enzymes was shown to prevent iPSC formation (Gao et al., 2013). Finally, overexpression of the enzyme was shown to increase hm5C levels in the central nervous system (Guo et al., 2011).

Immunostaining of HEK293 cells showed that TET1 localizes in the nucleus (Tahiliani et al., 2009). Looking at the specific placement of TET1 binding sites in the genome, several groups demonstrated the colocalization with hm5C patterns in euchromatin with special enrichment at hypomethylated promotors of high CG-content (Williams et al., 2011; Wu et al., 2011). Although hm5C levels were demonstrated to be reduced after knockout of TET1 in ESCs and m5C levels were increased especially at CpG-rich promotors after knockdown or knockout of the enzyme, TET1 deficient mice were shown to be viable and fertile, with normal brain development except that they are smaller in size (Dawlaty et al., 2011) and have poor learning and memory function due to impaired neurogenesis of the hippocampus (Gao et al., 2013).

Despite its early characterisation compared to the other TET enzymes, no crystal structure of mouse or human TET1 is available until today. Nevertheless, several remarks about the reactivity of the enzyme can be made based on available TET2 and *Naegleria gruberi* NgTET1 structures due to the partly conserved catalytic domain (Hashimoto et al., 2014). The NgTET1 enzyme was shown to oxidize m5C following the same mechanism, with similar overall structure and DNA recognition as observed for TET2 (Hashimoto et al., 2014). DNA is bound on the basic surface of NgTET1, contacted by hairpin loop L1 (equivalent to loop L2 in human TET2) and the methylated cytosine is flipped out of the DNA helix and inserted into the active site cavity, inducing DNA bending of 65°. The size of this hydrophobic pocket was demonstrated to be responsible to control the activity of the enzyme towards higher oxidation since the oxidation capacity was shown to be reduced if a pocket size reducing point mutation was introduced (Hashimoto et al., 2015). Overall, 5-10-fold lower activities on hm5C and f5C containing substrates were observed in comparison to m5C substrates (Hashimoto et al., 2015). In terms of substrate specificity, NgTET1 was shown to prefer CpG sites over non-CpG sites (Hashimoto et al., 2014). Nevertheless, the enzyme is also suspected to oxidize target sites in CpH context to a lower extent, which would fit the abundance of hm5C in non-CpG (especially CpA) context observed in neurons (Mellén et al., 2017) and the role of TET1 in demethylation of neurons (Guo et al., 2011).

1.3.2.2 TET2

Compared to TET1, the N-terminal part of TET2 is smaller and lacks the cysteine-X-X-cysteine (CXXC) domain, which was lost due to gene duplication and inversion during evolution. It is now known that TET1 binds the inhibition of the dvl and axin complex (IDAX) protein, which contains a CXXC domain that presumably replaces the endogenous one (Iyer et al., 2009; Pastor et al., 2013). The C-terminal part of TET2 contains a larger unstructured region in its DSBH domain (see Figure 10 for domain arrangements of human TET2), which was shown not to influence activity but to hinder expression and purification of the protein, hence it was deleted in some structural and biochemical studies (Hu et al., 2013; Hu et al., 2015).

In terms of protein expression in different cell types, TET2 was shown to have similar expression levels as TET1 (Ficz et al., 2011; Ito et al., 2010; Wassidlo et al., 2011) with additional abundance in hematopoietic cells (Ko et al., 2010). The enzyme was also found to localize to the nucleus, which was shown, among others, for U2OS and HEK293 cells (Ito et al., 2010). Generally, TET2 is the family member most frequently mutated and misregulated in AML and other types of myeloid cancers, with an occurrence of 10-30% (Jiang, 2020; Langemeijer et al., 2009; Tefferi et al., 2009) and observed loss-of-function alterations resulting in decreased levels of hm5C. In accordance, mice harbouring TET2 mutations developed myeloid malignancies and TET2 knockout mice showed myeloid and also lymphoid disorders (Ito et al., 2019). Furthermore, DKO of TET1 and TET2 led to a 40% reduced birth rate (Dawlaty et al., 2011; Dawlaty et al., 2013) and TET1-3 TKO ESCs showed altered capacity in differentiation and embryonic development (Dawlaty et al., 2014).

In the case of TET2, several crystal structures of the human enzyme are available with different modified versions of DNA (Hu et al., 2013; Hu et al., 2015), allowing for specific characterisation of protein-DNA interactions. As depicted in Figure 12, a crystal structure including methylated dsDNA shows that the substrate is bound above the core of the DSBH domain (enriched in basic and hydrophobic amino acids) with stabilization by the loops L1 and L2 from the cysteine-rich region. The target cytosine is flipped out of the DNA helix and the emptied space is occupied by a hydrophobic loop with a highly conserved tyrosine residue for stabilization. This is followed by the orientation of the methyl group towards the Fe(II) and the 2OG cofactors to enable catalytic turnover. As for NgTET1 (Hashimoto et al., 2015), the catalytic cavity of TET2 was shown to be large enough to accommodate both hm5C and f5C, with recognition of these modified target sites being the same as for m5C and almost identical conformation of the flipped-out base in the active centre (Hu et al., 2013; Hu et al., 2015). The reason for the 5-10-fold lower conversion rates, which were experimentally determined for hm5C and f5C containing DNA substrates (Hu et al., 2013; Hu et al., 2015; Ito et al., 2011), lies in the structure of the catalytic pocket. Additional hydrogen bonds are formed in presence of these oxidized species, hindering free rotation of the C-C-bond between the C5 cytosine and the modified methyl group. This leads to restrained conformations of the oxidized groups that prevent a fast abstraction of hydrogen during the catalytic mechanism, therefore slowing down the whole turnover

process. Whether the enzyme uses a distributive or processive mechanism for the stepwise oxidation of one site or between different sites on the same DNA substrate is still under debate (Crawford et al., 2016; Tamanaha et al., 2016).

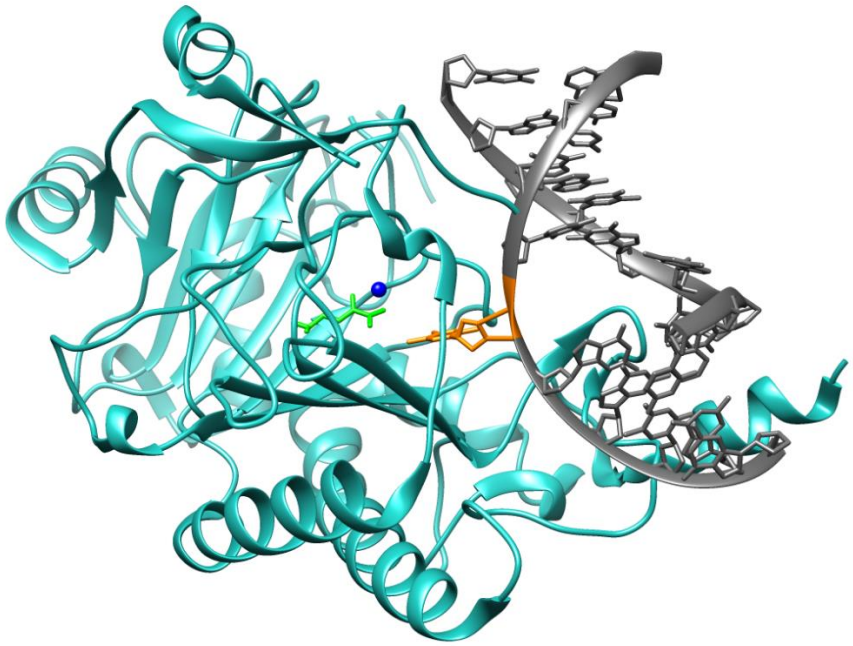


Figure 12: Structure of human TET2 co-crystallized with hemimethylated DNA. TET2 is shown in cyan, with co-crystallized DNA shown in grey and the flipped-out 5-methylcytosine coloured in orange; coordinated Fe(II) and 2-oxoglutarate are coloured in dark blue and light green, respectively (based on Hu et al., 2013; PDBI: 4NM6).

Overall, it was shown that TET2 prefers CpG sites over target sites in a non-CpG context, as demonstrated for NgTET1 (Hashimoto et al., 2014; Hu et al., 2013). Since most cytosine methylation occurs in the CpG context, this preference of the TET enzymes fits nicely to the preference of the DNMTs. Nevertheless, there was also methylation found in non-CpG context with the highest levels in embryonic stem and brain cells (Ku et al., 2011), accompanied by hm5C in CpH context (Ficz et al., 2011; Lister et al., 2013). Following this observation, TET2 was proven to have moderate activity on CpH substrates, with reported preferences for either CpA (DeNizio et al., 2021) or CpC (Hu et al., 2013).

1.3.3 Passive DNA demethylation pathways

In addition to the active demethylation pathway including sequential oxidation of m5C by the TET family enzymes and base excision, there is another pathway of demethylation known as replication-dependent or passive DNA demethylation. The basis for this process is the loss or inhibition of the maintenance methyltransferase DNMT1, which normally restores the methylation patterns after each DNA replication cycle (Jurkowska et al., 2011d). If the activity of this enzyme is disturbed (drastic cases would be the expression of a catalytically inactive mutant or gene knockout), the newly synthesized daughter strand will lose the epigenetic information and 50% of DNA methylation in the genome will be lost with every round. However, changes in the recognition of the target sites due to their oxidation by TET enzymes were also shown to promote passive demethylation. Up to 60-fold reduced efficiency of maintenance methylation was observed *in vitro* after replacement of m5C in a CpG site by hm5C (Hashimoto et al., 2012). In another study, the authors systematically analysed the effects of m5C substitutions by hm5C, f5C or ca5C at a hemimethylated CpG site in a CGXG context on DNMT1 activity (Ji et al., 2014), where X marks the modified nucleobase. Not only did they observe an influence of the substitution on the methylation of the modified CpG site, but methylation of the adjacent 5' CpG site was also affected. In this experiment, a significant reduction of up to 70% of the DNMT1 activity was shown in GCGC context after the exchange of m5C with hm5C, which increased even further with the incorporation of higher oxidation products. Interestingly, the introduction of f5C and ca5C even blocked methylation of the neighbouring unmodified CpG site, which could suggest that these modifications spread the loss of methylation signal over the genome. The effect was even more elevated in an earlier study, in which they observed a reduction of over 90% in DNMT1 methylation in the context of CCGG when the hydrophobic methyl group was replaced by the hydrophilic hydroxymethyl group (Valinluck and Sowers, 2007). Until today, many groups have shown that the general trend in DNMT1 preference follows m5C>hm5C>C (Hashimoto et al., 2012; Otani et al., 2013; Valinluck and Sowers, 2007), a fact which was even used to develop a new sequencing method called DMAB-seq to discriminate between the three nucleobases (Takahashi et al., 2015).

Considering that faithful inheritance of DNA methylation also includes different cofactors of DNMT1, several studies investigated the binding affinity of the cofactor

UHRF1 to hemi-hm5C instead of hemi-m5C containing CpG sites. Hashimoto et al. observed that under their experimental conditions UHRF1 showed a more than 10-fold reduced affinity towards hemi-hm5C (Hashimoto et al., 2012). Consistent with this finding, another group also observed a reduced binding of the SRA-domain of UHRF1 if m5C was substituted with hm5C (Otani et al., 2013)

In contrast to the reduced activity of DNMT1 on oxidized target sites, DNMT3A and DNMT3B were shown to have similar activities on both hemi-m5C and hemi-hm5C containing CpG sites (Gowher and Jeltsch, 2001; Hashimoto et al., 2012; Otani et al., 2013). Nevertheless, this is still under debate since an extended study from Ji et al. showed a reduction of methylation with DNMT3A after substitution of m5C with hm5C but elevated methylation with f5C and no changes in the case of ca5C (Ji et al., 2014). Overall, a combination of active and passive demethylation pathways may be the explanation for how rapid and efficient removal of cytosine methylation can take place during specific developmental stages (Messerschmidt et al., 2014) or in specific cell and tissue types.

1.4 Flanking sequence preferences of DNMTs and TET enzymes

Mammalian DNA methyltransferases and TET enzymes all share their preference for methylation of cytosines in the context of CpG dinucleotides, which is much shorter than the recognition sites of most restriction-modification enzymes or transcription factors that span up to 6 nucleotides (Kumar et al., 1994). Efficient recognition of the target sites is therefore crucial for DNA modification, which is mediated by different intrinsic properties of the enzymes such as the preference for CpG sites, in the case of DNMT1 even for a hemimethylated CpG site (see section 1.2.3.2). In addition to these features, it was first discovered in 2002 that DNA methyltransferases (in this case DNMT3A) also discriminate *in vitro* between the sequences flanking the target site (Lin et al., 2002). Based on the methylation of plasmid DNA with DNMT3A, which was investigated with bisulfite sequencing, it was shown that flanks containing pyrimidine bases were strongly preferred by DNMT3A (loose consensus sequence YNCGY) over flanks containing purines.

Following this early study, the flanking sequence preferences of DNMT3s have been the subject of many studies throughout the last 20 years. In 2005, an analysis of the data obtained during the Human Epigenome project further defined the preferences of DNMT3A and DNMT3B since high methylation was observed in a consensus sequence of 5'-CTTGCGCAAG-3', whereas low methylation was shown in the context of 5'-TGTTCGGTGG-3' (Handa and Jeltsch, 2005). These consensus sequences were also verified in methylation kinetics using oligonucleotides designed to be preferred or disfavoured, which revealed a more than 13-fold difference in their methylation rates. Furthermore, an up to 500-fold difference was observed when these contexts were combined with specific ±4 outer flank sequences. Overall, the authors noticed a depletion of C and G in the preferred ±1 flanking sequence contexts for both DNMT3A and DNMT3B.

A study by Wienholz et al. in 2010 expanded this view by elaborating the differences in the flanking sequence preferences of the two *de novo* enzymes. The authors showed even higher selectivity for DNMT3B which is more influenced by the ±1 position in the flanks, whereas DNMT3A was more influenced by the ±2 position despite the highly conserved catalytic domain of both enzymes (Wienholz et al., 2010).

Afterwards, a complete sequence profile of DNMT3A ranging from the -6 to the +6 position was determined through a more systematic analysis of over 1300 methylated CpG sites by bisulfite sequencing (Jurkowska et al., 2011c). This profile revealed strong enrichment of T at position -2 and A at position -1 as well as pyrimidines at position +1 followed by A and T at position +3 in the preferred flanking contexts. In contrast to wild-type DNMT3A, mutations of residues in the RD interface were shown to alter the flanking sequence preferences of the enzyme (Gowher et al., 2006). This was also observed in cellular DNA methylation levels, for instance somatic DNMT3A mutations such as R882H frequently observed in AML were shown to promote cancer through changes in the flanking sequence preference profile (Emperle et al., 2018b).

In contrast to the extensively studied flanking sequence preferences of DNMT3A and DNMT3B, little was known about the influence of the CpG adjacent DNA context on the activity of DNMT1. Although some CpG sites seemed to be resistant to DNMT1 methylation (Feltus et al., 2003) and high GC content in the flanks of CpG sites was observed to enhance the activity of the maintenance methyltransferase (Flynn et al.,

1998), a small-scale analysis of the ±1 flanks did not provide any evidence for a strong flanking sequence preference of DNMT1 (Bashtrykov et al., 2012b).

Similar to DNMT1, no flanking sequence preferences were known for the TET enzymes until the start of this work and based on the crystal structure of TET2 the enzymes were even suspected to lack this intrinsic property (Hu et al., 2013; Hu et al., 2015). In line with this, comparable activities were observed on DNA containing AT- or GC-rich sequences flanking a methylated CpG site (Hu et al., 2015). Despite the m5CpG dinucleotide, where the cytosine is also flipped out of the DNA helix as in the DNMT mechanism, only phosphate groups were shown to be involved in the TET-DNA contacts through a network of extensive hydrogen bonds and hydrophobic interactions (Hu et al., 2013, Hu et al., 2015). Only two other studies suggested an influence of the flanking sequences on the TET activity, but those were either small-scaled (Pais et al., 2015) or based on potentially unreliable antibody pulldowns (Kizaki et al., 2016). Due to the lack of information about the DNA sequence readout mechanisms of DNMT1 and the TET enzymes, which are needed to fully unravel the dynamics of DNA methylation at different genomic regions, detailed investigations of the flanking sequence preferences of the involved enzymes were urgently required at the start of this thesis.

2 Aims of this study

DNA methylation and its role as an epigenetic mark have been the focus of numerous studies throughout the last decades (Kim and Costello, 2017). Recently, the emerging field of active DNA demethylation and the investigation of the resulting oxidation products (Wu and Zhang, 2017) gave rise to even more questions about this epigenetic change. Nevertheless, many properties of the enzymes setting and removing the methylation marks are still under debate or have not been investigated so far, although dysregulation and mutations of the involved enzymes have been clearly linked to cancer development and propagation (Rasmussen and Helin, 2016; Zhang et al., 2020).

To specifically unravel the details of target site recognition, DNA interaction and the influence of different flanking sequence contexts on the activity of DNA methyltransferases and methylcytosine dioxygenases, it was planned to apply a new Deep Enzymology approach, which I started to develop in my M.Sc. thesis. Here, DNA modification reactions are performed using pools of double-stranded DNA substrates which contain one or more target sites in the context of randomized bases, followed by bisulfite conversion coupled to Next-Generation Sequencing readout of millions of reaction products allowing to determine the modification state and specific sequence of each product molecule. Using this approach, it was the main goal of this study to systematically analyse the DNA sequence readout mechanisms and flanking sequence preferences of enzymes involved in DNA methylation and demethylation. Furthermore, it was planned to develop variations of the workflow to be used for investigations of the specific roles of DNMT3 mutations in this readout process and specific mechanistic questions.

The DNA methyltransferase DNMT1 has been observed to play a crucial role in maintenance DNA methylation (Petryk et al., 2021) where it efficiently restores the cell-specific DNA methylation patterns after every replication cycle. For this specific purpose, the enzyme is highly processive with preferences for methylation of hemimethylated CpG sites (Jeltsch, 2006) whereby the range varied due to the applied experimental conditions. Although structural evidence for this specificity was provided later (Song et al., 2011; Song et al., 2012), major distortions of the DNA structure around the target CpG site were observed in the crystal structures including the

formation of a non-canonical base-pair unknown for enzymes using the described base-flipping mechanism. Since the flexibility of the DNA helix is dependent on its sequence, it was suspected that the DNA sequence readout of DNMT1 might be dependent on the flanking sequences of the target site as already shown for DNMT3A and DNMT3B (Handa and Jeltsch, 2005; Lin et al., 2002) followed by specific structural consequences. Evidence for this had already been shown in two studies that observed CpG sites prone to or resistant to DNMT1 methylation (Feltus et al., 2003; Flynn et al., 1998), but a small-scale study of the influence of the ±1 position on the methylation activity did not support this concept (Bashtrykov et al., 2012b). Therefore, it was the main goal of the first project in this thesis to investigate the flanking sequence preferences of DNMT1 systematically and more broadly.

Compared to the maintenance methyltransferase DNMT1, flanking sequence effects on the activity of DNMT3A and DNMT3B have been the focus of investigations since 2002 (Lin et al., 2002). Starting with a loose consensus sequence of YNCGY for DNMT3A, further studies revealed a more than 13-fold difference in the methylation rates of different sequences for both enzymes, (Handa and Jeltsch, 2005) including investigations of the most relevant flanking positions (Wienholz et al., 2010), which finally lead to a complete sequence profile for DNMT3A (Jurkowska et al., 2011c). Based on these findings, it was the general goal of the second project in this thesis to verify and refine the flanking sequence preference analysis for DNMT3A and DNMT3B and to bring these results into a structural and biological context.

In addition to the DNA sequence readout mechanisms of wild-type DNMT3A, it was the third goal of this thesis to investigate the role of the somatic cancer mutation R882H frequently observed in AML (Ley et al., 2010). The position of this residue in the active site (RD interface) of the DNMT3A tetramer was previously shown to be important for target site recognition and the catalytic activity of the enzyme since its mutation drastically altered the flanking sequence preferences (Emperle et al., 2018b; Gowher et al., 2006) and decreased the methylation levels by 30-50% (Emperle et al., 2018a; Emperle et al., 2018b; Holz-Schietinger et al., 2012; Yan et al., 2011). Although the R882H mutation has been previously studied in our laboratory, only a small number of CpG flanking sequences were investigated until the start of this project, resulting in the

necessity to conduct a detailed large-scaled study on the flanking sequence preferences of DNMT3A R882H as part of this thesis.

Since their discovery in 2009 (Tahiliani et al., 2009), properties of the Ten-eleven translocation enzymes such as their activity, processivity as well as substrate recognition and the role of the enzymes in active DNA demethylation have been under debate (Rasmussen and Helin, 2016). Intriguingly, hm5C and the other oxidation products are present in all tissue and cell types, whereby levels varied strongly and hm5C was found 100-1,000-fold more often than f5C and ca5C (Carell et al., 2018). Due to its abundance, hm5C is considered as a stable epigenetic mark today with its exact biological roles still undisclosed. Based on the crystal structures of TET2 (Hu et al., 2013; Hu et al., 2015) and the TET1 homolog from *Naegleria gruberi* (Hashimoto et al., 2014), target site recognition of the TET enzymes was concluded to be mediated only through the CpG site and hydrogen bond networks involving the phosphate backbone of the DNA. However, the TET enzymes use the same base-flipping mechanism as the DNMTs, leading to the question if the demethylation also follows specific DNA sequence readout mechanisms. Hence, it was the fourth goal of this study to systematically investigate the flanking sequence preferences of TET1 and TET2 in CpG as well as non-CpG contexts.

3 Materials and methods

In the following paragraphs, the most important materials and methods are summarized. Detailed descriptions of all methods applied throughout this thesis are provided in the respective method parts in the publications provided in the appendices 1 to 6.

3.1 Cloning, overexpression and protein purification

3.1.1 DNMT1

The plasmid encoding His- and GFP-tagged full-length mouse DNMT1 was available in the laboratory (pFastBacHT vector, UniProtKB P13864) and was directly used for overexpression following the Bac-to-Bac baculovirus expression system (Invitrogen) as described (Bashtrykov et al., 2012a; Goyal et al., 2006). Shortly, the transformation of the vector into *E. coli* DH10Bac cells was performed to generate a recombinant Bacmid followed by amplification of the baculovirus particles and subsequent expression of the protein in Sf21 cells. After harvesting, cells were lysed by sonication, cell debris was removed by centrifugation and DNMT1 was purified using affinity chromatography with nickel-NTA agarose beads (Genaxxon bioscience GmbH). The concentration of each purified batch of DNMT1 was determined by UV absorption at 280 nm and the purity was verified by SDS-PAGE stained with Coomassie-Brilliant Blue. For more detailed information, refer to the methods section of Appendix 1.

3.1.2 TET1 and TET2 isoforms

Plasmids encoding the His-tagged catalytic domain of mouse TET1 (amino acids 1367-2057 of XP_006513930.1) and full-length mouse TET2 were purchased from Addgene (plasmids #81053 and #89735). In the case of TET1, the plasmid was directly used for overexpression of the protein. In the case of TET2, the catalytic domain (amino acids 915-1920) was cloned into the TET1 expression vector using the Gibson Assembly approach (Gibson, 2009) followed by the replacement of amino acids 1401-1764 with a 15 amino acid linker (yielding amino acids 915-1400-linker-1765-1920 of XP_006501349.1). Furthermore, another isoform of TET2 was prepared using the megaprimer based site-directed mutagenesis method (Jeltsch and Lanio, 2002) to

insert a serine (yielding amino acids 915-1401-linker-1765-1920 of NP_001333665.1). A schematic overview is shown in Appendix 6 Figure S10.

Overexpression of all three TET constructs was performed by growing BL21 (DE3) Codon+ RIL *E. coli* cells (Stratagene) in LB medium supplemented with trace metals (Studier, 2005) at 37°C until an $A^{600\ nm}$ of 0.6-0.8 was reached, then 0.5 mM isopropyl 1-thio-D-galactopyranoside (Roth) was added to induce protein expression for 12-14 h at 20°C. Purification was carried out similarly as described for DNMT1 with the special addition of the enzymatic cofactor 2OG (Sigma Aldrich) to all buffers. Again, all concentrations were determined by UV absorption at 280 nm and the purity of the three proteins was verified by SDS-PAGE stained with Coomassie-Brilliant Blue. A more detailed version of the purification protocol is given in the methods section in Appendix 6.

3.2 Biotin-avidin microplate assay

The catalytic activities of the purified DNMT1 and the other DNA methyltransferases investigated in this thesis were determined using a biotin-avidin microplate assay (Liebert and Jeltsch, 2008; Roth and Jeltsch, 2000) as illustrated in Figure 13. This assay is based on the strong interaction of avidin, which is used to coat the microplate, and biotin-labelled DNA substrates with specifically designed sequences. For the DNMT1 project, *in vitro* methylation reactions with different enzyme concentrations and DNA sequences (obtained from IDT) were performed using tritium-labelled AdoMet (Perkin Elmer) as a cofactor. The reactions were then stopped after different time intervals by quenching with unlabelled cofactor (Sigma Aldrich) present in the microplate and the biotinylated DNA substrates were bound to the avidin (Sigma Aldrich) coated plate. After several washing steps to remove the remaining cofactor and other reaction components, nuclease digestion was used to release the DNA molecules from the plate. The solution was then transferred into scintillation vials and the number of incorporated tritium-labelled methyl groups was determined for each time interval by liquid scintillation counting. Finally, the obtained counts per minute were plotted against the incubation time with Microsoft Excel (2019) followed by linear regression analysis. This assay was also used to verify the enzymatic preferences of DNMT1 determined with the newly applied Deep Enzymology method as shown in

Appendix 1. For the other investigated methyltransferases, radioactive kinetics were performed by other colleagues as described in Appendices 2, 4 and 5.

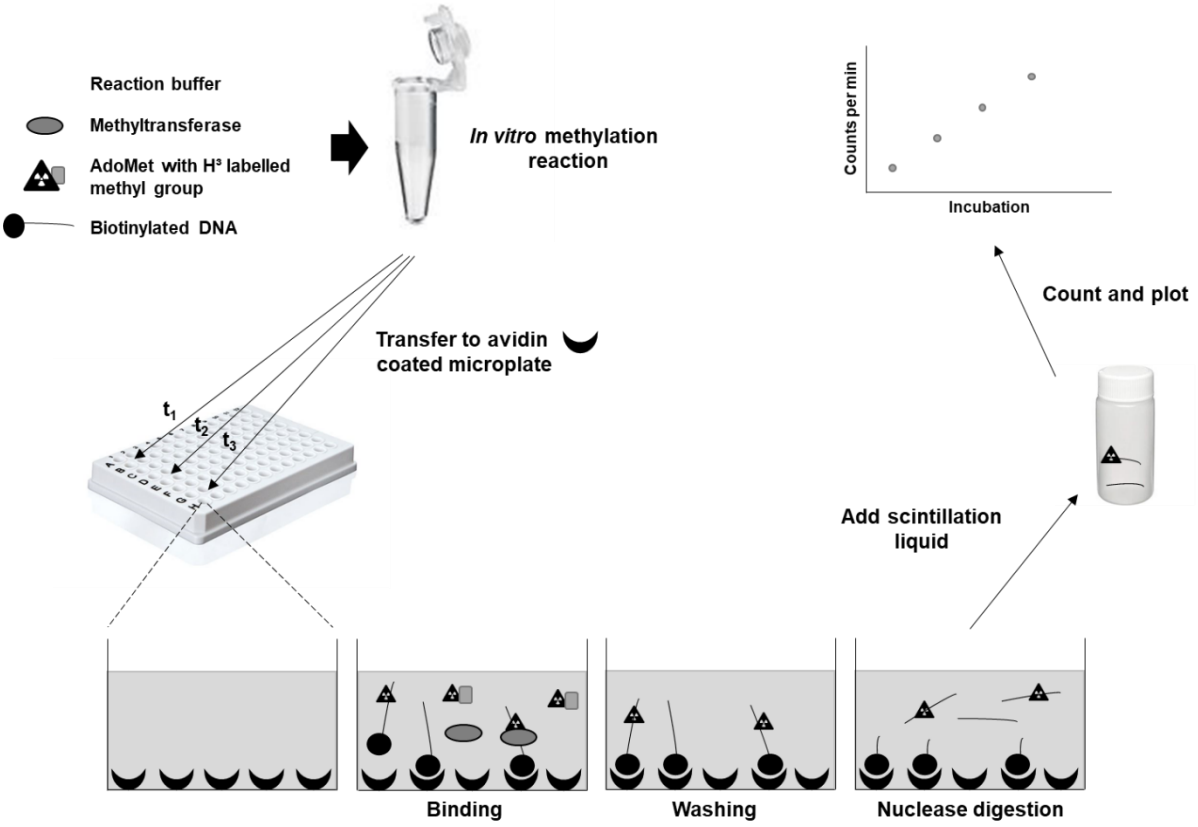


Figure 13: Schematic illustration of the biotin-avidin microplate assay. Methylation reactions were performed using tritium-labelled cofactor AdoMet to determine the catalytic activity of DNMT1 (based on Jurkowska et al., 2011a).

3.3 HPLC-MS/MS analysis

The oxidation activities of the purified TET1 and TET2 proteins on m5C and hm5C containing DNA substrates were determined using liquid chromatography coupled to mass spectrometry. For this, *in vitro* oxidation reactions were performed in the presence of 2OG (Sigma Aldrich) and ammonium iron(II) sulfate (Roth) using different enzyme concentrations, DNA sequences (obtained from IDT) as well as different substrate modifications as described in detail in Appendix 6. The reactions were stopped after different time intervals by freezing in liquid nitrogen. Enzymatic digestion to nucleosides and filtration were performed to prepare the samples for HPLC-MS/MS analysis (Traube et al., 2019). The digested samples, as well as different dilutions of

external standards (unmodified dATP, dGTP, dCTP, and dTTP from Genaxxon bioscience GmbH; 5fdCTP, 5cadCTP, and 5mdCTP from tebu-bio GmbH and 5hmdCTP from Zymo Research Europe GmbH), were then diluted in a 1:1 ratio with a mixture composing of 1% acetonitrile and 0.1% formic acid and spectra were measured in the group of Prof. Jens Brockmeyer (Department of Food Chemistry, Institute of Biochemistry and Technical Biochemistry, University of Stuttgart) (for further details about the measurements see the respective section in Appendix 6). Data analysis was carried out using the Skyline (version 20.2) software by calculation of the peak area of the transitions given in Appendix 6 Figure S6. Depending on the injection volume, the peak integrals were normalized for all reactions and plotted against the reaction time using Microsoft Excel (2019). For input normalization, the integrals were next divided by the area obtained for cytosine nucleosides in each reaction, which do not participate in the oxidation pathway. Afterwards, calibration factors obtained from the external calibration curves were applied and the sum of all modified cytosine nucleosides was set to 1 for each reaction.

3.4 Deep Enzymology reactions with randomized substrates

The flanking sequence preferences of different DNA methyltransferases, as well as cytosine dioxygenases, were investigated during this study with the usage of randomized DNA substrates (obtained from IDT) containing one central CpG or CpN site or two target sites in a specific distance (for further details see Table 1). These target sites harboured different cytosine modifications (unmethylated, methylated or hydroxymethylated) and were embedded into a context of randomized nucleotides. To obtain the double-stranded DNA substrates, primer extension of purchased oligonucleotides was performed as depicted in step 1 in Figure 14 followed by a purification step using either DNA-binding columns (Macherey-Nagel) (unmethylated and hemimethylated substrates) or streptavidin Dynabeads M-280 (ThermoFisher Scientific) (hemihydroxymethylated substrate labelled with a Desthiobiotin-TEG). The latter had to be introduced due to the weak binding of substrates containing this cytosine modification to silica membranes.

Table 1: Summary of the main randomized substrates used during the projects of this thesis with annotated number of target sites and their respective context and modification.

Enzyme	**Context**	**Specification**	**Modification**	**Appendix**
DNMT1	CG	One target site	Hemimethylated	1
DNMT3A/3B	CG/CN	One target site	Un-/hemimethylated	2
DNMT3A	CG	Two sites in a variable distance	Unmethylated	3
DNMT3A R882H	CG	One target site	Unmethylated	4
DNMT3A R882H	CG	Two sites in 12 bps distance	Unmethylated	5
TET1/2	CG/CN	One target site	Hemi-/hemihydroxymethylated	6

For the *in vitro* reactions, the randomized double-stranded substrates were incubated with individual DNMTs or TET enzymes and their respective cofactors using various experimental conditions (steps 2 and 3 in Figure 14, respectively) and reactions were stopped after different incubation times by freezing in liquid nitrogen. Proteinase K (NEB) treatment was performed afterwards to release the enzyme-bound DNA. To be able to discriminate between different modification states of the central cytosines as well as to reconstitute the randomized sequence parts during the analysis, hairpin-bisulfite sequencing was used (Laird et al., 2004). Therefore, both DNA strands were linked with a short hairpin oligonucleotide using a Golden-Gate digestion approach (Engler et al., 2008) (step 4 in Figure 14). Afterwards, bisulfite conversion (EZ DNA Methylation-Lightning kit from ZYMO) of the linked strands was performed which leads to the deamination of unmodified, formyl- and carboxylcytosines to uracil but leaves methyl- and hydroxymethylcytosines intact (step 5 in Figure 14). For great-depth analysis of the reaction products, Illumina Next-Generation Sequencing was used for which DNA libraries were built with a successive PCR approach (Hess et al., 2020) that introduces unique combinations of barcode and index sequences (step 6 in Figure 14). Like this, samples from different reactions could be pooled for sequencing but still distinguished in the following bioinformatic analyses of the obtained data sets. In the end, the original DNA sequences were reconstituted using the information from

both strands and the modification states of the upper or lower strand of the target site were determined for all possible N2 or N3 flanking sequence combinations. In addition, fitting of the NNCGNN oxidation activities to monoexponential reaction progress curves was performed in the group of Prof. Nicole Radde (Institute for Systems Theory and Automatic Control, University of Stuttgart) to determine the flanking sequence-dependent rate constants. The method used to determine the flanking sequence preferences of different enzymes was first published in Appendix 4, but also used for all other projects with potential adjustments and termed Deep Enzymology due to its great statistical power.

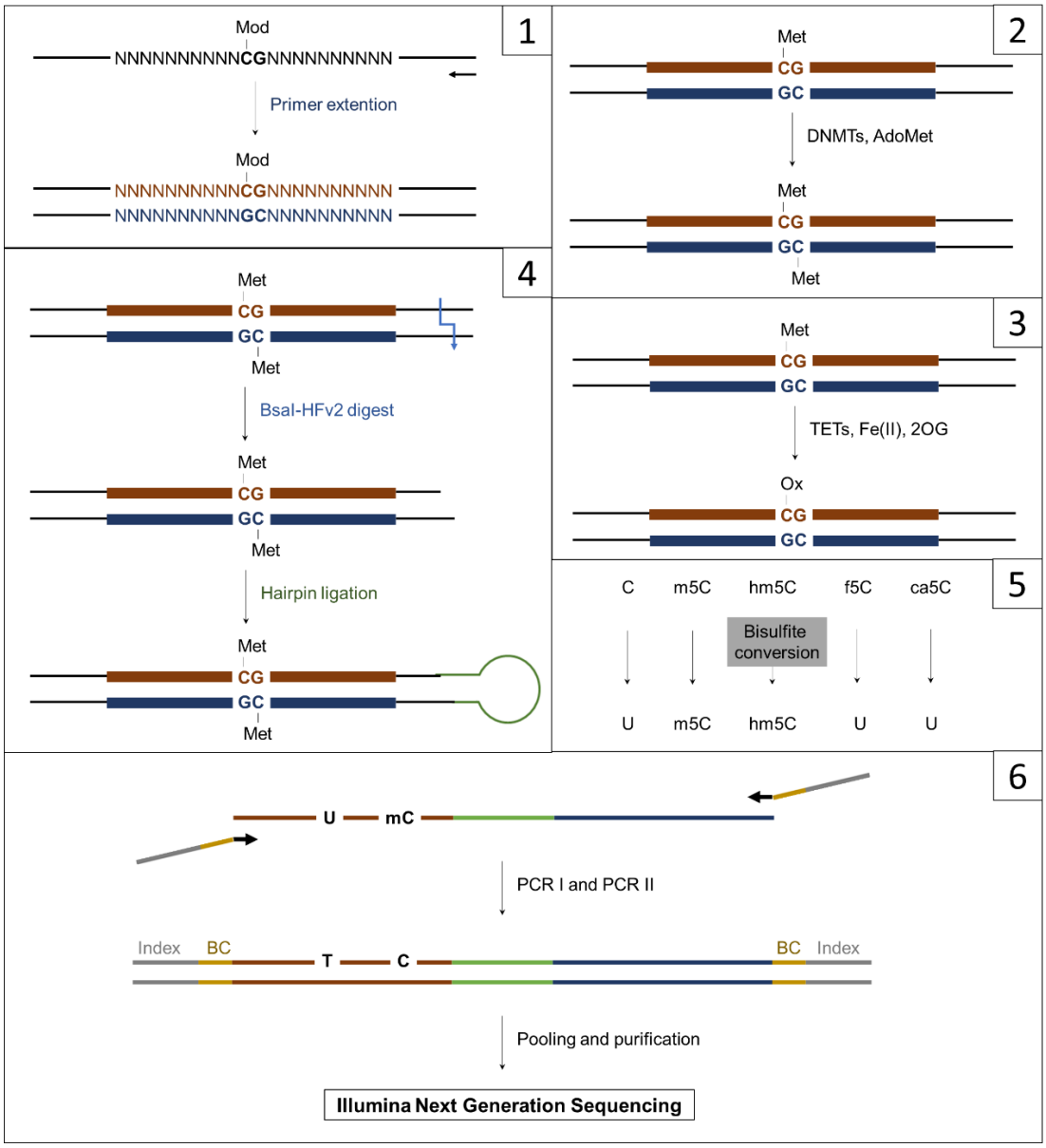


Figure 14: Schematic illustration of the Deep Enzymology approach used to study the flanking sequence preferences of different DNA methyltransferases and methylcytosine dioxygenases. Step 1) Substrate synthesis through primer extension to obtain the double-stranded hemi-modified (Mod) randomized substrates. Steps 2 and 3) Exemplary reaction of hemimethylated (Met) randomized substrates in the presence of DNA methyltransferases (DNMTs) and their cofactor S-adenosyl-L-methionine (AdoMet) or cytosine dioxygenases (TETs) and their cofactors Fe(II) and 2-oxoglutarate (2OG). The oxidized cytosine forms originating from the TET reactions are indicated by “Ox”. Step 4) Golden-Gate Assembly with the type II restriction enzyme BsaI-HFv2 used for the ligation of a phosphorylated DNA hairpin. Step 5) Schematic illustration of the outcome of differently modified cytosines after treatment with a bisulfite reagent. Step 6) Library generation for Illumina Next-Generation Sequencing through two subsequent PCRs that attach unique barcode (BC) and index primer combinations to each sample. Incorporation of thymine and cytosine in place of uracil and methylcytosine, respectively, is indicated.

4 Results

4.1 Investigation of the flanking sequence preferences of DNMT1

Accurate maintenance of cell-specific methylation patterns by the DNA methyltransferase 1 (DNMT1) is essential for the stable inheritance of cellular identities (Petryk et al., 2021). For this purpose, the enzyme specifically recognizes hemimethylated CpG sites, which occur during DNA replication, and restores the DNA methylation in a processive manner (Jeltsch, 2006). Although the CpG target sites are short in size compared to DNA recognition motifs of other enzymes, for instance several bacterial methyltransferases (Kumar et al., 1994), their efficient recognition is a crucial step during the DNA methylation process. Furthermore, extended preferences for specific nucleotides flanking the CpG target sites have been revealed in 2002 for the other two DNA methyltransferases 3A and 3B (Lin et al., 2002) and further investigated since (Jeltsch, 2021). In contrast to these two enzymes, potential flanking sequence preferences of DNMT1 have not been analysed until today except for small-scale experiments that showed no significant influence (Bashtrykov et al., 2012b). Nevertheless, indications for flanking effects on the maintenance DNA methylation were already found in early biochemical studies (Feltus et al., 2003; Flynn et al., 1998) and during previous work conducted in the Jeltsch research group. In addition, the same base-flipping mechanism as used by the DNMT3 enzymes was shown for DNMT1 (Song et al., 2012). Surprisingly, these DNMT1 structures showed unusual structural reorganizations of the DNA after flipping of the cytosine, strengthening the hypothesis that the nucleotide composition around the CpG target site is mechanistically relevant. Therefore, it was the goal of the first project in this thesis to elucidate the effect of different flanking sequences on the activity of DNMT1. To approach this question, I performed a systematic investigation using the Deep Enzymology approach established during my previous work in the Jeltsch research group. This approach uses pools of double-stranded DNA substrates that contain specifically modified CpG target sites in the context of 10 random nucleotides on both sides (for further information see section 3.4). The results of this biochemical investigation were published in Nature Communications together with data from other research groups and can be found in Appendix 1, while the main results are provided in the following paragraphs.

On the basis that DNMT1 is a maintenance methyltransferase, the flanking sequence preference analysis for this enzyme was performed using substrates with one hemimethylated target site as shown in Figure 15A. After *in vitro* methylation kinetics were performed with the enzyme under different reaction conditions, hairpin ligation was carried out to join the sequence information of both DNA strands followed by bisulfite conversion to convert unmethylated cytosines into uracil. After the preparation of libraries for Next-Generation Sequencing via a two-step PCR approach, thousands of product molecules could be subjected to a detailed investigation of the methylation state of the central target site and its correlation with the flanking context of that particular DNA molecule. Like this, two independent experimental repeats were conducted with varying enzyme concentrations (Appendix 1 Table S2).

To determine which flanking positions exhibit the greatest influence on the maintenance activity, observed versus expected frequencies of each nucleotide at each flanking position were calculated. These values refer to how often a nucleotide appeared at this specific position in the methylated (observed) and total (expected) pool of sequenced DNA molecules. As shown in Figure 15A, the numbering -1 to -10 refers to the randomized positions 5' of the target site, whereas +1 to +10 refers to the randomized positions 3' of the target site. DNMT1 activity was strongly influenced by the -2 to +2 flanking positions (Figure 15B and Appendix 1 Figure 3A), therefore further analyses were focused on this range. Combined effects of specific nucleotide compositions were investigated for the ±2 flanks (NNCGNN) by determining the average methylation levels of all 256 NNCGNN flanks, which again showed a high overall correlation between the two experiments (Appendix 1 Figures 3B, C) but only weak correlation with the flanks determined for DNMT3A and DNMT3B in other parts of this thesis (Appendix 1 Figure 3B and section 4.2). For more detailed analyses, data from both DNMT1 repeats were normalized based on their average methylation and merged, revealing an almost 100-fold difference in the relative methylation rates of specific NNCGNN flanks (Appendix 1 Figure 3D). Weblogos were prepared using the 20 most preferred or disfavoured sites, which showed very distinct preferences of DNMT1, especially a preference for T at the -2 and A or T at the +2 position as well as strong disfavour for C at the -2 and G at the -1 position (Figure 15C and Appendix 1 Figure 3E). These observations could be validated by radioactive kinetics with substrates containing flanks specifically designed to be highly preferred or mildly to

highly disfavoured by DNMT1, which showed an up to 12-fold difference in the methylation rates (Appendix 1 Table 1 and Figure S4A).

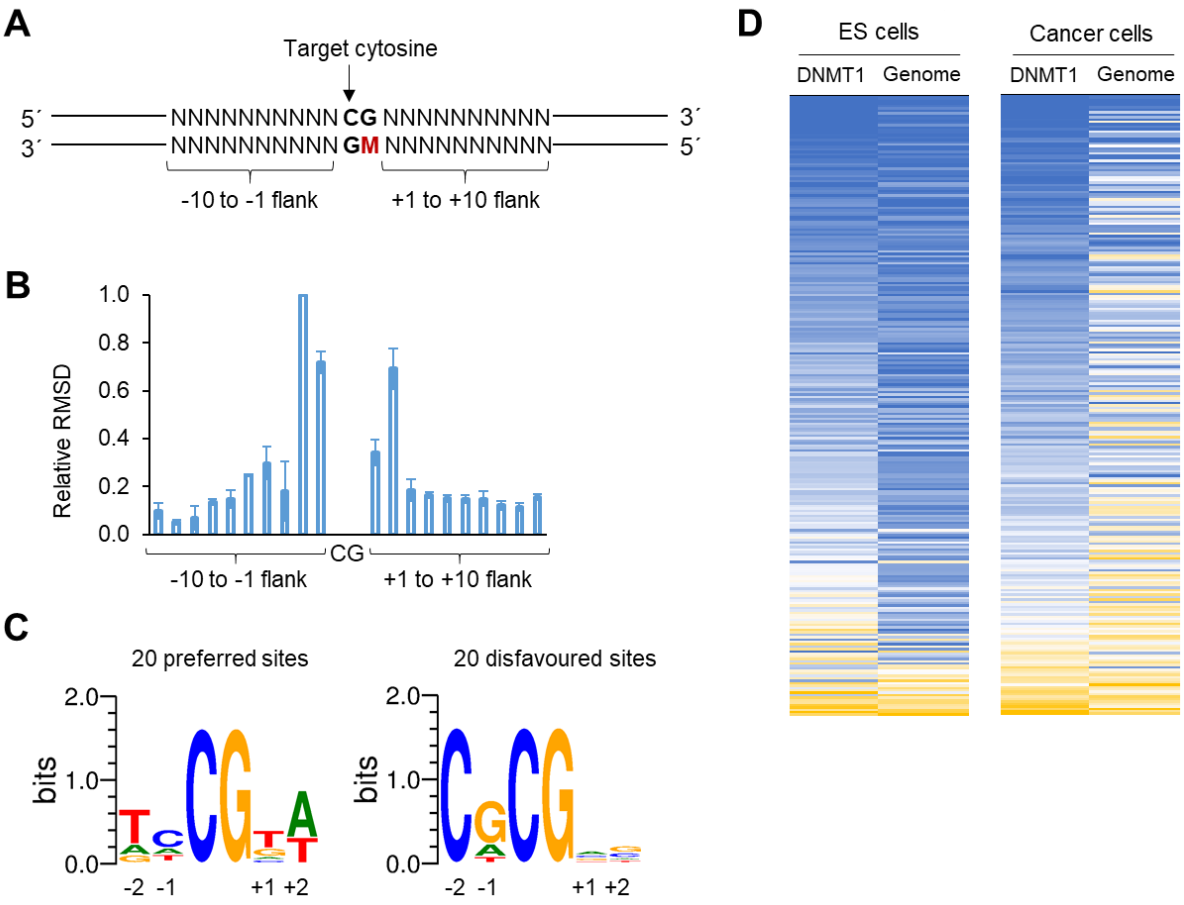


Figure 15: Flanking sequence preferences of DNMT1. A) Design of the substrate pool used in the Deep Enzymology approach for DNMT1 kinetics with one hemimethylated CpG target site (methylation in the lower strand marked as red "M") embedded into a context of 10 randomized nucleotides on either side. B) Relative base preferences were determined for all randomized flanking positions and given as root mean squared deviations (RMSD) of the observed/expected base composition within the methylated reads, normalized to the highest effect at position -2. C) Weblogos of the 20 NNCGNN sites most preferred or disfavoured by DNMT1. D) Correlation of the averaged NNCGNN methylation levels determined by NGS (DNMT1) with CpG methylation patterns extracted from whole genome bisulfite data of human ESCs (Genome, left) (Charlton et al., 2018) or reduced representation genome bisulfite data of lung cancer cells (Genome, right) (Hascher et al., 2014) given as heatmaps. Panels A and B were prepared based on Appendix 1 and panels C and D were adapted from Appendix 1.

Strikingly, the flanking sequence preferences of DNMT1 derived from the Deep Enzymology analysis could also be observed during the methylation analysis of a longer substrate derived from a CpG island with 44 hemimethylated CpG sites, which

was methylated with the enzyme for different incubation times (Appendix 1 Figures 1A, B). Here, an up to 30-fold divergence in the methylation rates of target sites embedded in different flanking contexts was observed and the extracted preferences (Appendix 1 Figures 1B, C) showed high congruence with the respective NNCGNN flanks from the Deep Enzymology approach except for small deviations as a result of low coverage of certain flank sites in the 44 site substrate (Appendix 1 Figures 3F, S4B). In addition to the flanking sequence analysis, the long substrate methylation kinetics were analysed with two kinetic models in the research group of Prof. Nicole Radde (Institute for Systems Theory and Automatic Control, University of Stuttgart), whereby one model postulated DNMT1 to methylate only in a distributive manner and one model proposed mixed distributive and processive methylation (Appendix 1 Figures 1D, 2). Only Model 2 was able to capture the experimental methylation data quantitively, which once again strengthened the conclusion made in previous studies (Jeltsch et al., 2006) of DNMT1 being a processive enzyme, but also gave new mechanistic insights based on calculated parameters for the binding and dissociation as well as processive and distributive methylation rates.

Due to the strong flanking sequence preferences of DNMT1 discovered with the Deep Enzymology approach, mechanistic differences in the methylation of different DNA contexts were postulated during this project. This hypothesis was based on the unusual crystal structure of DNMT1 published in 2012, in which the enzyme was co-crystallized with DNA containing a relatively disfavoured NNCGNN flank, that showed striking reorganizations of the DNA and amino acid residues after the target cytosine was flipped into the active site (Song et al., 2012). To investigate if different DNA contexts have other structural consequences for the binding and methylation step of DNMT1 reactions, two new crystal structures were prepared by members of the research group of Prof. Jikui Song (Department of Biochemistry, University of California) containing co-crystallized DNA fragments with relatively favoured NNCGNN flanks. Comparison of the new complexes with the published crystal structure revealed overall structural similarity, except for the catalytic helix that was observed to adopt a rather kinked conformation in the preferred sequence context compared to the rather straight conformation in the disfavoured sequence context (Appendix 1 Figures 4A, 5A-F). In addition, the DNA molecules themselves showed sequence-dependent conformational reorganization after base-flipping of the target cytosine (Appendix 1

Figures 4B, 4F-H) especially influenced by the 5' flank of the target CpG, which was accompanied by a change in solvent accessibility of the cofactor binding pocket (Appendix 1 Figures 5G-I). These effects will be described in more detail and discussed in the discussion section of this thesis (5.1).

Lastly, flanking sequence preference profiles determined as part of this study were compared with published genomic DNA methylation data to investigate how DNMT1 preferences shape the global DNA methylation landscape. In general, very strong correlations were observed both for NNCGNN flanks extracted from whole genome bisulfite sequencing data of human ESCs (Charlton et al., 2018) as well as reduced representation bisulfite sequencing of metastatic lung cancer cells (Hascher et al., 2014) with the Deep Enzymology data (Figure 15D and Appendix 1 Figures 6A, B). For the latter, treatment with DNA methyltransferase inhibitor 5-azacytidine resulted in a switch to a highly significant inverse correlation with the DNMT1 NNCGNN preferences determined with the randomized substrates (Appendix 1 Figure 6C), which indicated that CpG sites disfavoured by the enzyme were more efficiently demethylated under treatment. Finally, a comparison with genome-wide methylation data from mouse ESCs (Li et al., 2015) verified that DNMT1 shapes the global DNA methylation landscape in wild-type as well as DNMT3A and DNMT3B double knockout cells, whereas DNMT1 knockout cells showed a strong correlation with the NNCGNN flanking profiles determined for the DNMT3 enzymes during another project of this thesis (Appendix 1 Figure 7 and section 4.2).

4.2 Investigation of the DNA sequence readout mechanisms of DNMT3A and DNMT3B

The two *de novo* DNA methyltransferases 3A and 3B are the enzymes mainly responsible for the accurate setting of mammalian DNA methylation patterns during early developmental stages (Okano et al., 1999). In addition to CpG sites, the enzymes have been shown to introduce non-CpG methylation in ESCs and neural cells to varying degrees (Lee et al., 2017). The activity of DNMT3A and DNMT3B is enhanced by the catalytically inactive DNMT3L which interacts with both methyltransferases in specific regulatory manners (Jeltsch and Jurkowska, 2016). In the case of DNMT3A, a

crystal structure published in 2018 revealed the mechanistic details of DNMT3A/3L interaction in a linear heterotetramer complex bound to DNA (Zhang et al., 2018). These heterotetramers assemble in a specific 3L-3A-3A-3L arrangement, where the central positions responsible for the formation of the DNA binding site are taken up by the active DNMT3A subunits. Contrary to this so-called RD interface, the two peripheral 3A-3L interaction sites, also referred to as FF interfaces (Figure 5A) (Jeltsch and Jurkowska, 2016), mediate the general stimulation of the DNMT3A methyltransferase activity. Furthermore, interactions between the DNA and both DNMT3A subunits in the DNMT3A/3L heterotetramer are also mediated by parts of the catalytic core and the target recognition domain (TRD). Although the RD interface was shown to be largely conserved between both DNMT3A and DNMT3B (Okano et al., 1998), distinct biological functions have been discovered for both DNMT3 enzymes. However, structural details about the DNMT3B/3L methylation of DNA as well as biochemical consequences remained unknown until the start of this thesis because the crystal structure of this complex was not available. Therefore, one aim of this study was to gain more insights into the specific DNA interaction of DNMT3B and to further deepen our understanding of the mechanism of DNMT3A.

4.2.1 General insights into the *de novo* methylation mechanisms

In the past, close relation in the amino acid sequence of DNMT3A and DNMT3B (roughly 80% for the MTase domain) has led to the hypothesis that both enzymes could be interchangeable during global *de novo* DNA methylation (Okano et al., 1999). However, single knockout studies resulting in embryonic or postnatal lethality disproved this assumption later on (Challen et al., 2014; Chen et al., 2003; Okano et al., 1999) and distinct roles of both active DNMT3s were uncovered. Compared to DNMT3A, which prefers methylation of major satellite repeats and is the main enzyme involved in the setup of imprinting during gametogenesis (Chen et al., 2003; Kato et al., 2007; Watanabe et al., 2002), DNMT3B preferably methylates minor satellite repeats and is essential for embryonic development (Chen et al., 2003; Okano et al., 1999). As a consequence, mutations in these two enzymes lead to the development of different diseases such as AML in the case of DNMT3A (Brunetti et al., 2017) or immunodeficiency, centromeric instability, facial anomalies (ICF) syndrome in the case

of DNMT3B (Ehrlich et al., 2008). Based on the fact that the distinct functional properties of the enzymes have to have an underlying mechanistic reason, structural and functional investigations of DNMT3A and DNMT3B were carried out as part of this thesis in cooperation with the research group of Prof. Jikui Song (Department of Biochemistry, University of California) that contributed the structural data. All results were then published in Nature Communications and are attached as Appendix 2. The focus of my work during this project laid on the analysis of the flanking sequence preferences of the two enzymes, for which I was involved in the design and validation of the specific workflow applied in this study as well as the bioinformatic analyses of the obtained data.

Although it has already been known from various studies that DNMT3A and DNMT3B differ in their preferred DNA sequence context (Handa and Jeltsch, 2005; Jurkowska et al., 2011c; Lin et al., 2002; Wienholz et al., 2010), all conclusions so far were based on a small- to midscale number of flanking sequences used for biochemical investigations or extracted from genomic data. To ultimately define the preference profiles of both enzymes, which can then be connected to structural features and biological targets, a systematic analysis was performed with the Deep Enzymology approach. Here, pools of DNA substrates containing a CpG or CpN site flanked by 10 random nucleotides on either side were methylated with the catalytic domains of DNMTA or DNMT3B by other colleagues from the Jeltsch research group. Afterwards, I performed hairpin ligation, bisulfite conversion and prepared the libraries for Next-Generation Sequencing (Appendix 2 Figure S1). Collectively performed bioinformatic analyses of the different experiments showed high correlations of the flanking sequence preferences of experimental repeats for both enzymes, respectively (Appendix 2 Figures 1B, S2E). Moreover, interesting differences between the two *de novo* DNA methyltransferases were worked out which will be described in the following section.

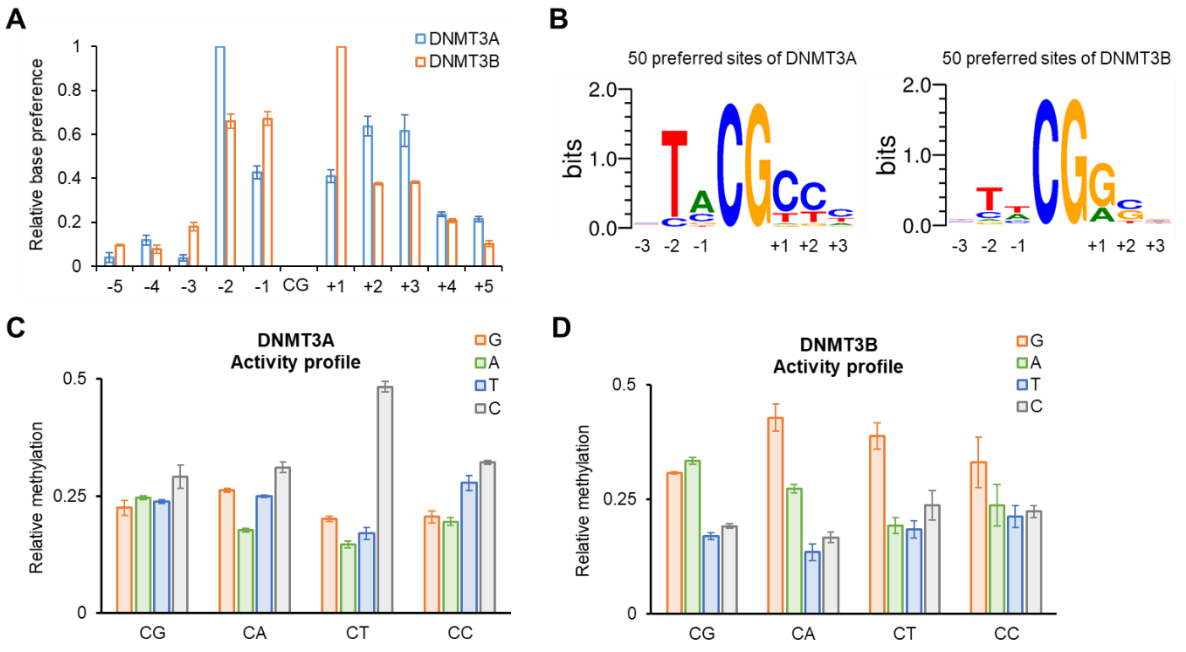


Figure 16: Flanking sequence preferences of DNMT3A and DNMT3B. A) Relative base preferences were determined for the -5 to +5 flanking positions and given as mean standard deviations of the observed/expected base composition within the methylated reads, normalized to the highest effect at position -2 (DNMT3A) or +1 (DNMT3B). B) Weblogos of the 50 NNNCGNNN sites most preferred by DNMT3A or DNMT3B. C) Averaged CpN methylation by DNMT3A determined by NGS for all four dinucleotides concerning the +1 flanking base. D) Averaged CpN methylation by DNMT3B determined by NGS for all four dinucleotides concerning the +1 flanking base. Panels A, C and D were prepared based on Appendix 2 and panel B was adapted from Appendix 2.

First, base enrichments were calculated for all flank positions in the -5 to +5 range to investigate which positions show the highest impact on DNMT3 activity. As depicted in Figure 16A, the significant influence of the -2 to +3 positions was observed for both DNMT3A and DNMT3B, although to a varying degree with -2 being the most important position in the case of DNMT3A, whereas DNMT3B activity was most influenced by the +1 position. Due to these results, further investigations were focused on the averaged NNNCGNNN sites and Weblogos were prepared based on the 50 preferred sites each (Figure 16B and Appendix 2 Figure 1C). In line with previous studies (Jeltsch et al., 2021), pronounced flanking sequence preferences were observed for both enzymes, showing a common preference for T at the -2 position but different flanking preferences at the 3' flank. Strikingly, more than 100-fold differences in the methylation rates of specific sequences could be determined when directly comparing both DNMT3 flanking preferences (Appendix 2 Figure 1D). As anticipated, DNMT3L did not change

these determined preferences in the case of heterotetramer formation since it only occupies the outer positions of the heterotetramer not in contact with the DNA (Appendix 2 Figure S9). To validate the NGS findings, radioactive *in vitro* methylation kinetics were performed with substrates containing flanks specifically designed to be preferred or disfavoured by DNMTA or DNMT3B by other colleagues in the Jeltsch laboratory, which showed similar results as the extracted Deep Enzymology data (Appendix 2 Figure S3A).

In addition to the preference profiles in CpG context, DNMT3A and DNMT3B activities on substrates containing CpN sites in different flanking contexts were determined using an extended version of the Deep Enzymology approach that included all target sites in equimolar ratio. Although methylation of CpG sites was highly preferred by both enzymes (Appendix 2 Figure 4D and Table S1), activity profiles could be put together for non-CpG methylation, which was up to 2-fold higher for DNMT3B than for DNMT3A. Since the relative preferences of the enzymes differed the most at the +1 position, detailed analyses were conducted for all target sites regarding the base enrichments at this position (Figure 16C, D and Appendix 2 Figures 2A, B). Intriguingly, strongly enhanced preferences for G at +1 were observed in the case of non-CpG methylation by DNMT3B, whereas DNMT3A rather showed elevated activities for C at this position, especially in CpT context.

Together with other data provided by different collaborators, the biochemically determined flanking sequence preferences could also be brought into a structural and biological context. The new crystal structure of DNMT3B/3L showed conformational differences in the catalytic loop and the RD interface compared to the DNMT3A/3L structure (Zhang et al., 2018) and identified lysine 777 in DNMT3B as a key residue for recognition of the base at the +1 flanking position (Appendix 2 Figures 4-6). Furthermore, similar CpN flanking context preferences for DNMT3A and DNMT3B as in the NGS data could also be found, respectively, in enhanced reduced representation bisulfite sequencing data of DNMT knockout mouse ESCs which were stably transduced with either DNMT3A or DNMT3B (Appendix Figures 2C, D and S5). Lastly, connections to specific biological targets and genetic diseases could be drawn based on the determined DNMT3 flanking sequence profiles (Appendix 2 Figures 1D-G),

including the SatII repeats as a well-characterised DNMT3B target that is known to be hypomethylated in patients with ICF syndrome (Ehrlich et al., 2008).

4.2.2 Dual CpG site methylation mechanism of DNMT3A

Another interesting property of the DNMT3 enzymes is the potential co-methylation of two CpG target sites during one methylation event. This specific mode of action can take place through the two central subunits of a tetramer each methylating one cytosine in their active site, respectively. As shown in the crystal structure of DNMT3A/3L with co-crystallized DNA (Zhang et al., 2018), structural constraints of the inner subunits hinder the methylation of both cytosines in one palindromic target site, but co-methylation of two CpG sites that occur in a specific distance and on opposite strands can be realized. This co-methylation was already investigated earlier in biochemical and cellular studies that suggested 10 bps as the preferred distance (Jia et al., 2007; Jurkowska et al., 2008). Overall, different degrees of DNA helix bending dependent on the distance of the co-methylated CpG sites were observed in the crystallographic study from 2018. While two short co-crystallized DNA fragments containing one target site each showed almost no conformational deviation from B-DNA in the bound state, a DNA fragment with two CpG sites in a distance of 12 bps showed approximately 40° bending with the centre of bending between the two sites (Zhang et al., 2018). Since the flexibility of the DNA helix needed for co-methylation of two target sites on one DNA substrate is dependent on the DNA sequence and strong flanking sequence preferences of DNMT3A were determined in the scope of this study, it was interesting to see if those preferences could be linked to specific bending and binding modes of the enzyme. In addition to the interaction with DNMT3L over a hydrophobic interface called the FF interface, DNMT3A has been shown to also form homotetramers or even higher-order structures in the absence of DNMT3L and these oligomers exhibit unusual DNA interaction properties (Jeltsch and Jurkowska, 2016). To unravel the different binding modes responsible for *de novo* methylation of the whole genome, co-methylation of two CpG sites by DNMT3A and DNMT3A/3L tetramers was investigated as part of this thesis, for which I contributed to the design and validation of the adapted workflow as well as the bioinformatic analyses of the obtained data. The obtained results, which will be summarized in the following sections, were published in Nucleic

Acids Research (Appendix 3) together with scanning force microscopy measurements performed in the research group of Prof. Ingrid Tessmer (Rudolf Virchow Center for Integrative and Translational Bioimaging, University of Würzburg).

To address the question of DNA sequence-dependent co-methylation by DNMT3A, adjustments were made to the previously described Deep Enzymology approach used for the flanking sequence preference analysis of the DNMTs (Appendix 1 and 2). In this specific project, pools of substrates containing two unmethylated CpG sites embedded in a randomized sequence context and with a variable distance of 2-15 bps between them were used for methylation with either DNMT3A or DNMT3A/3L (Figure 17A and Appendix 3 Figure 1A). For the data analysis of the conducted experiments after NGS, structurally and mechanistically different types of co-methylation of the two target sites had to be distinguished as depicted: co-methylation of the first CpG site in the upper strand and the second site in the lower strand (MW), the reverse situation (WM) and co-methylation in the same strand (MM or WW) whereby methylation of both upper or both lower sites were taken together. The conducted experiments showed increasing overall methylation levels with increasing enzyme concentrations and further enhanced methylation was observed in the presence of DNMT3L (Appendix 3 Figure 1B) as known from the literature (Chédin et al., 2002; Suetake et al., 2004). Using the averaged and normalized data sets, a detailed analysis was conducted regarding the distance dependency of co-methylation. In addition to the peak at distances of 2 bps, which would suggest short-range movements of the tetramer complexes, additional peaks of the relative co-methylation levels were observed for MW and WM co-methylation at different distances of 5-6, 9 or 12 bps dependent on the experimental conditions (Figure 17B and Appendix 3 Figures 2, 3 and S4).

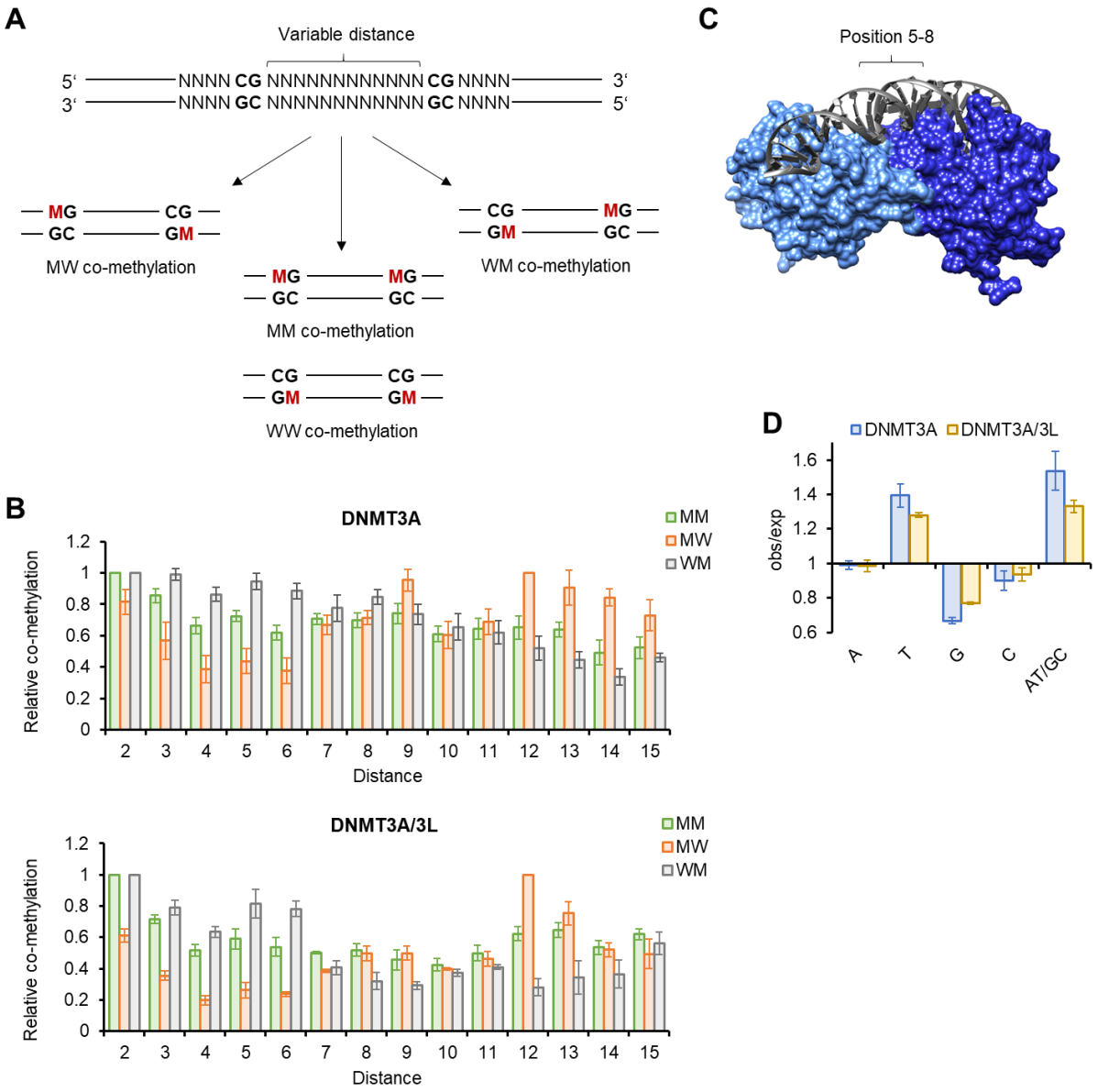


Figure 17: Co-methylation in different distances catalysed by DNMT3A or DNMT3A/3L. A) Design of the substrate pool used to investigate co-methylation of two CpG target sites in variable distances, which are embedded into a context of randomized nucleotides. Three different types of co-methylation of the two target sites can be distinguished (methylation marked as red “M”, respectively) whereby MM and WW methylation are mechanistically equivalent. B) Relative co-methylation by DNMT3A or DNMT3A/3L in variable distances with given standard deviations and normalization to the highest effect for every type of co-methylation. C) Schematic view of the analysed positions 5-8 at the centre of bending at the RD interface formed by two DNMT3A subunits coloured in light and dark blue, respectively (Zhang et al., 2018; PDBI: 6BRR). D) Averaged base enrichment and depletion determined for substrates that showed co-methylation by DNMT3A or DNMT3A/3L with error bars given as SEM. Panels B and C were prepared based on data from Appendix 3 and panel D was adapted from Appendix 3.

Since MW co-methylation at a distance of 12 bps is in good agreement with the published crystal structure (Zhang et al., 2018) that also demonstrated the need for DNA bending during methylation, the influence of the DNA sequence composition between the two CpG sites was further investigated while focusing on the positions 5-8 as the centre of bending (Figure 17C and Appendix 3 Figure 4A). Here, strong sequence preferences were observed with highly significant enrichment of T and depletion of G in the upper strand of the MW co-methylated DNA molecules consistent with facilitated bending in the presence of A/T nucleotides (Figure 17D and Appendix 3 Figure 4B). Based on these findings and scanning force microscopy data provided by our collaborators (Appendix 1 Figures 5-7), different DNA sequence- and distance-dependent binding modes of DNMT3A/3L heterotetramers were postulated for target sites with 6, 9 or 12 bps spacing, which would result in different types of co-methylation accompanied by varying degree of DNA bending (see discussion section 5.2.2).

4.3 Investigation of the DNMT3A R882H mutation consequences

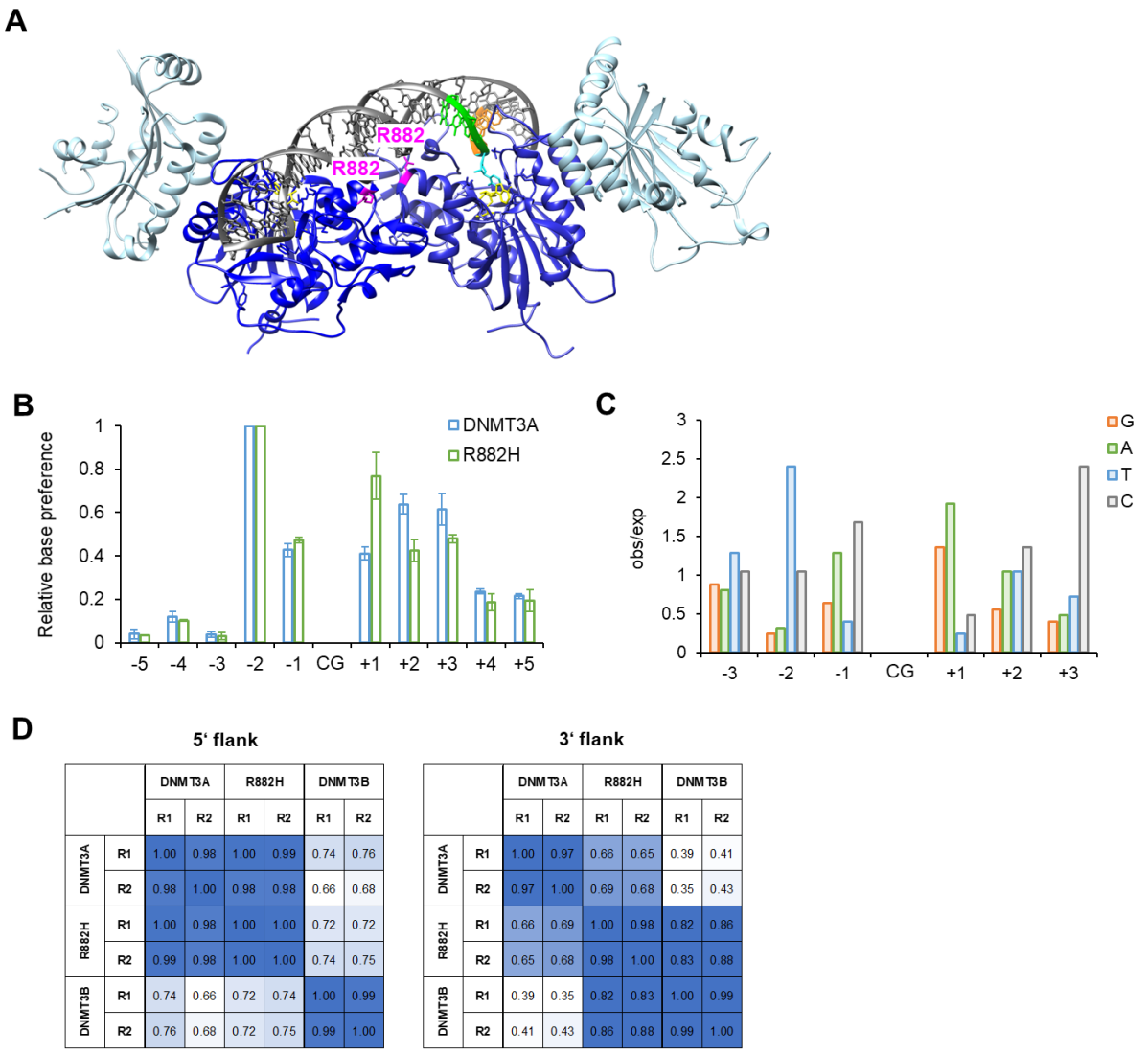
For mammalian development, it is essential that the DNA methylation landscape is accurately established and maintained in all cells as discussed in the previous sections. Hence, changes in the methylation patterns represent potential reasons for the development and further progression of different diseases such as cancer (Robertson, 2005). One of these cancer types correlated with aberrant DNA methylation is the haematological cancer AML which causes uncontrolled proliferation of abnormal blood cells (Ferrara and Schiffer, 2013). Several mutations were found to contribute to the development of AML, most of them in enzymes playing an epigenetic role such as TET2 or DNMT3A (DiNardo and Cortes, 2016). In the case of the *de novo* methyltransferase 3A, up to 25% of all AML patients carried mutations that were found to be clustered in the catalytic domain of the enzyme (Brunetti et al., 2017). Overall, 70% of the mutations are heterozygous missense mutations, resulting in the expression of mutated DNMT3A from one allele and wild-type enzyme from the other. The most abundant mutation occurs at R882, where the arginine was shown to be predominantly replaced by histidine followed by cytosine, serine or proline in descending order. As shown in Figure 18A, this residue interacts with the DNA backbone since it is located in the RD interface which builds up the catalytic centre of

the DNMT3A tetramer (Zhang et al., 2018). Due to the strong negative impact of the R882H mutation on the life expectancy of AML patients, several groups have focused on this specific DNMT3A mutation and its underlying mechanisms. Although reduced catalytic activity was observed for the mutated protein in different studies (Emperle et al., 2018a; Emperle et al., 2018b; Holz-Schietinger et al., 2012; Yan et al., 2011), its interaction with DNA and influence on the wild-type enzyme are still highly under debate. Therefore, detailed mechanistic studies of DNMT3A R882H were conducted as part of this thesis to reveal its impact on the DNA *de novo* methylation process.

4.3.1 Flanking sequence preferences of DNMT3A R882H

In terms of wild-type DNMT3A, flanking sequence preferences have been under discussion since their first documentation in 2002 (Lin et al., 2002) and further investigations were conducted as part of this thesis as well (see section 4.2). Changes in these preferences were further researched during a mutational analysis of the catalytic domain of the enzyme using a small subset of designed substrates (Gowher et al., 2006). In this study, altered preferences were shown for a R882A mutation, leading to the hypothesis that the frequently observed R882H mutation in AML might have similar effects. Previous work in the Jeltsch group confirmed this assumption using a substrate that contained 56 CpG sites in a different context for methylation kinetics with the wild-type and mutant enzyme (Emperle et al., 2018b). In this analysis, the authors showed that some CpG sites were methylated even better by the mutated enzyme, although the overall activity of R882H was reduced compared to the wild type. However, the number of sequence contexts investigated in this study was not high enough to make definite conclusions about the cellular impact of the mutation. Therefore, a detailed analysis of the flanking sequence preference changes caused by mutations at R882 of DNMT3A was conducted in this thesis using the same Deep Enzymology approach as for the wild-type enzymes in section 4.2.1. In this project, I was involved in the design and validation of the applied workflow as well as the bioinformatic analyses of the obtained data. The results of this study were published in Nucleic Acids Research together with data from experiments conducted by other colleagues (for further details see Appendix 4).

To systematically determine the exact changes in the flanking sequence preferences of the R882H mutant protein, all observed effects were compared either to the wild-type enzyme or to both DNMT3A and DNMT3B. For a first analysis, the overall impact of each flanking position on the activity of R882H was determined and compared with the effects obtained for the DNMT3 WT. Interestingly, both enzymes showed high influences of flanking residues in the -2 to +3 region. There was almost no change in the relative base preferences at the 5' flank of the target site, whereas strong differences were observed for the 3' flank (Figure 18B and Appendix 4 Figure 1A). A more detailed analysis of the relative depletion or enrichment of individual bases at each flanking position for R882H showed strong favour for T and disfavour of G at the -2 position, preference for C at the -1 as well as G or A at the +1 position (Figure 18C). Furthermore, combined analyses were conducted for all NNCGNN and NNNCGNNN flanks, which showed a high correlation of the experimental repeats performed with one enzyme, but only a weak correlation between R882H, DNMT3A and DNMT3B (Appendix 4 Figures 1B, C). After averaging and normalizing the flanking sequence preferences for both DNMT3A enzymes, up to 70-fold differences in the preferences of R882H for specific NNNCGNNN flanks could be detected (Appendix 4 Figure 2A). Using these data, Weblogos were prepared with the most preferred and most disfavoured sites of R882H when compared with wild-type DNMT3A, illustrating once again how drastic the changes in the flanking profiles are (Appendix 4 Figure 2B). For validation of the Deep Enzymology results, radioactive methylation kinetics were performed with both R882H and DNMT3A WT using various exemplary substrates with a target site in different flanking contexts by other colleagues from the Jeltsch group, and the calculated methylation rates were in good agreement with the flanking sequence data (Appendix 4 Figure 2C).



D

5‘ flank

		DNMT3A		R882H		DNMT3B	
		R1	R2	R1	R2	R1	R2
DNMT3A	R1	1.00	0.98	1.00	0.99	0.74	0.76
	R2	0.98	1.00	0.98	0.98	0.66	0.68
R882H	R1	1.00	0.98	1.00	1.00	0.72	0.72
	R2	0.99	0.98	1.00	1.00	0.74	0.75
DNMT3B	R1	0.74	0.66	0.72	0.74	1.00	0.99
	R2	0.76	0.68	0.72	0.75	0.99	1.00

3‘ flank

		DNMT3A		R882H		DNMT3B	
		R1	R2	R1	R2	R1	R2
DNMT3A	R1	1.00	0.97	0.66	0.65	0.39	0.41
	R2	0.97	1.00	0.69	0.68	0.35	0.43
R882H	R1	0.66	0.69	1.00	0.98	0.82	0.86
	R2	0.65	0.68	0.98	1.00	0.83	0.88
DNMT3B	R1	0.39	0.35	0.82	0.83	1.00	0.99
	R2	0.41	0.43	0.86	0.88	0.99	1.00

Figure 18: Flanking sequence preference analysis for the DNMT3A R882H mutation. A) Schematic view of the position of R882 at the RD interface of a DNMT3A/3L tetramer (Zhang et al., 2018; PDBI: 5YX2). DNMT3A subunits are shown in shades of darker blue, respectively, with R882H marked in magenta and DNMT3L subunits are shown in light blue. The cofactor S-adenosyl-L-homocysteine is shown in yellow and the bound DNA is coloured in dark grey with 5’ and 3’ flank of one CpG site marked in green and orange, respectively, and the flipped-out cytosine shown in cyan. B) Relative base preferences of DNMT3A and R882H were determined for the -5 to +5 flanking positions and given as standard deviations of the observed/expected base composition within the methylated reads, normalized to the highest effect at position -2. C) Averaged methylation levels of the mCpG substrates determined after kinetic reactions with DNMT3A R882H. Values are given as observed/expected (obs/exp) for every base at the -3 to +3 flank positions. D) Correlations of the 5’ and 3’ flanking preferences obtained for R882H with results obtained in section 4.2.1 for DNMT3A and DNMT3B. Panels A and C were prepared based on data from Appendix 4 and panels B and D were adapted from Appendix 4.

To gain further insights into flanking preference changes at the 3’ or 5’ flank, profiles obtained for all three DNMT3 enzymes were compared, respectively (Figure 18D and Appendix 4 Figure 3B). Surprisingly, high correlations were observed for DNMT3A WT and R882H on the 5’ flank, whereas the mutant protein resembled DNMT3B on the 3’ flank. As seen for wild-type DNMT3A and DNMT3B, differences in the DNA sequence readout mostly have a structural basis and the loop that contains the R882 residue greatly differs between the two enzymes in amino acid sequence as well as conformation (see section 4.2.1). Although only one amino acid is mutated in the R882H complex compared to the wild-type enzyme, this arginine was shown to engage in a backbone contact to the 3’ flank of the target site while leaving the 5’ flank untouched (see discussion section 5.3.1).

To bring our Deep Enzymology data into a biological context, cellular experiments were conducted by the other first author using HCT116 DNMT1 hypomorph cells with inducible expression of DNMT3A WT or R882H. Strikingly, methylation patterns extracted from the genomic DNA of these treated cells were in very good agreement with the flanking sequence profiles of the two enzymes, respectively (Appendix 4 Figure 5). In addition, published enhanced reduced representation bisulfite data from AML patients (Glass et al., 2017) showed the same alterations in the DNA methylation patterns, although only some patients included in the averaged data set harboured the R882H mutation (Appendix 4 Figure 6). Overall, investigations of other mutations occurring at this position in AML (R882C, R882S, R882P) with the Deep Enzymology approach showed similar changes in the flanking preferences as observed for R882H, which suggests that the specific trigger for these effects is the loss of the backbone contact resulting from the removal of the arginine 882 side chain (Appendix 4 Figure 4).

4.3.2 Assembly of DNMT3A WT and R882H heterotetramers

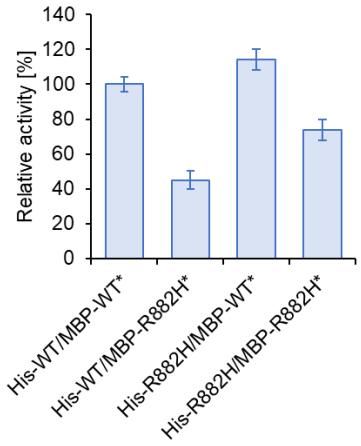
In the presence of catalytically inactive DNMT3L, heterotetramer formation with active DNMT3A subunits takes place in which the two DNMT3As are connected through a central RD interface that generates the DNA binding site of the 3L-3A-3A-3L complex (Jia et al., 2007; Zhang et al., 2018). As an alternative to DNMT3L, other DNMT3A subunits can occupy the outer positions, building up DNMT3A homotetramers as the smallest active form with the potential for multimerization (Jurkowska et al., 2011b). Since R882 is located in the DNA-binding RD interface of DNMT3A/3L tetramers with its mutation to histidine occurring in a heterozygous manner, the influence of mutated subunits on the activity of wild-type subunits after the formation of mixed tetramers has been investigated by various groups during the last years but with partly contradicting results (Emperle et al., 2018a; Russler-Germain et al., 2014). During the conduct of this thesis, strong flanking sequence preference changes were observed for R882H especially at the 3' flank of the target site as discussed in section 5.3.1. These specific changes were used as a basis to characterise the structural organization of DNMT3A WT/R882H subunits in homo- or heterotetramers and how the subunits influence each other's activity. For this purpose, the Deep Enzymology approach was adjusted and pools of DNA substrates containing two CpG sites embedded in a random sequence context with a distance of 12 bps, which was shown to be favourable for co-methylation (see section 4.2.2), were used for methylation with specially prepared tetramers. The main results of this project, that were published in the Journal of Molecular Biology (attached as Appendix 5), will be described in the following paragraphs.

To investigate the properties of DNMT3A homo- and heterotetramers, a previously developed double-tag purification procedure (Emperle et al., 2018a) was used with co-expression of His- and MBP-tagged wild-type and/or R882H subunits followed by two steps of affinity chromatography under high salt conditions. These conditions inhibit the formation of the RD interface, leading to the formation and purification of FF interface dimers, which are then dialyzed against low salt buffer to obtain different homo- or heterotetramers. Using this method, homotypic tetramers were purified first, for which the catalytic activity was shown to be reduced (Appendix 5 Figures S1A, B) as seen in previous studies (Emperle et al., 2018a; Emperle et al., 2018b; Holz-Schietinger et al., 2012; Yan et al., 2011). Next, homotetrameric and mixed complexes were purified including one active His-tagged subunit and one inactive MBP-tagged

subunit (marked with an asterisk) with a C710S mutation in the catalytic pocket removing the cysteine residue essential for catalysis (Appendix 5 Figure 1B). Using this approach, direct investigation of the impact of the R882H mutation on the activity of mixed complexes was possible, since all catalytic activity could be assigned to the active His-tagged subunit. Radioactive methylation kinetics were performed to determine the overall activity and the initial reaction rates were used for comparison of the four complex combinations (Figure 19A and Appendix 5 Figure 1C). As previously seen, homotetrameric R882H showed reduced activity compared to the homotetrameric wild-type complex. In contrast, the combination of His-WT/MBP-R882H* dramatically reduced the activity of the wild-type subunit and even more striking, the combination of His-R882H/MBP-WT* did not just rescue the loss of activity with the mutant subunit but showed enhanced activity.

For interpretation of these unexpected results, we proposed a model in which both subunits that make up the RD interface are independent and can methylate one DNA strand each. Since the random formation of the interface would not lead to enzymatic activities that are in agreement with our experimental data, preferential formation of R882H/R882H RD interfaces was postulated (Appendix 5 Figure 1D). Like this, the activity of His-WT/MBP-R882H* would be especially low due to the inactive subunits occupying the positions relevant to the catalytic activity of the complex. In contrast, the inactive wild-type subunits would take up the irrelevant outer positions in His-R882H/MBP-WT*, which would make this complex more active than His-WT/MBP-WT* with an equal distribution of both active and inactive subunits.

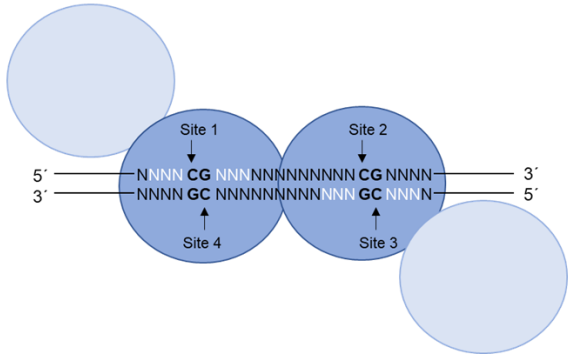
A



C

		Site 1, flank position						Site 3, flank position					
		-3	-2	-1	+1	+2	+3	-3	-2	-1	+1	+2	+3
His-WT/ MBP-WT	A	1.03	0.88	1.12	0.99	1.04	1.09	1.05	0.86	1.14	1.05	1.06	1.10
	T	1.02	1.25	0.90	0.98	1.00	0.99	1.04	1.41	0.89	1.02	1.01	0.99
	G	0.98	0.69	0.93	0.86	0.82	0.83	0.96	0.78	0.90	0.83	0.86	0.80
	C	0.97	1.06	1.06	1.13	1.10	1.06	0.95	1.10	1.05	1.11	1.09	1.13
His-R882H/ MBP-R882H	A	1.02	0.61	1.25	1.18	0.84	1.00	1.05	0.68	1.22	1.16	0.88	1.00
	T	1.03	1.61	0.76	0.65	0.93	0.89	1.01	1.80	0.75	0.57	0.98	0.97
	G	0.99	0.46	0.85	1.25	0.78	0.76	1.00	0.63	0.85	1.17	0.87	0.72
	C	0.96	1.03	1.17	1.10	1.36	1.30	0.91	1.17	1.15	1.07	1.43	1.43
His-R882H/ MBP-WT	A	1.04	0.68	1.13	1.08	0.83	1.02	1.06	0.77	1.15	1.12	0.94	0.99
	T	1.01	1.47	0.83	0.78	0.95	0.97	1.06	1.62	0.76	0.71	1.04	1.05
	G	0.97	0.51	0.94	1.05	0.79	0.79	0.96	0.69	0.90	1.04	0.83	0.71
	C	0.98	1.09	1.13	1.16	1.36	1.18	0.90	1.13	1.18	1.12	1.29	1.37
His-WT/ MBP-R882H	A	1.04	0.74	1.16	1.04	0.93	1.06	1.02	0.77	1.17	1.08	0.97	1.05
	T	1.02	1.41	0.84	0.82	0.99	0.94	1.02	1.57	0.80	0.80	1.03	0.98
	G	0.99	0.57	0.88	1.03	0.76	0.77	0.99	0.71	0.89	1.00	0.82	0.76
	C	0.95	1.06	1.14	1.14	1.24	1.17	0.95	1.13	1.12	1.13	1.29	1.28

B



D

Plus flank correlation

		His-WT/ MBP-WT		His-R882H/ MBP-R882H		His-R882H/ MBP-WT		His-WT/ MBP-R882H	
		site1	site3	site 1	site 3	site1	site3	site 1	site 3
His-WT/ MBP-WT	site1	1.00	0.96	0.43	0.44	0.63	0.63	0.76	0.73
	site3	0.96	1.00	0.38	0.42	0.55	0.62	0.69	0.71
His-R882H/ MBP-R882H	site 1	0.43	0.38	1.00	0.97	0.93	0.90	0.90	0.90
	site 3	0.44	0.42	0.97	1.00	0.91	0.96	0.88	0.93
His-R882H/ MBP-WT	site1	0.63	0.55	0.93	0.91	1.00	0.91	0.96	0.94
	site3	0.63	0.62	0.90	0.96	0.91	1.00	0.92	0.98
His-WT/ MBP-R882H	site 1	0.76	0.69	0.90	0.88	0.96	0.92	1.00	0.97
	site 3	0.73	0.71	0.90	0.93	0.94	0.98	0.97	1.00

Figure 19: Compilation of the results from the flanking sequence preference analysis of homo- and heterotetrameric DNMT3A WT/R882H complexes. A) Mean catalytic activity of tetrameric DNMT3A complexes determined using radioactive kinetics with normalization to the His-WT/MBP-WT* activity (inactive subunit marked with an asterisk) and error bars given as SEM. B) Schematic view of the substrate pool used for the preference analysis of DNMT3A complexes (inner subunits in darker blue, outer subunits in light blue), with two CpG sites in a 12 bps distance and embedded into a randomized nucleotide context. The investigated co-methylation takes place at both target sites but in opposite strands (Site 1 and Site 3) and the -3 to +3 flank of these sites (coloured in white) is used for analysis. C) Base enrichment and depletion profiles of substrates co-methylated at Site 1 and Site 3 after methylation with different tetramer complexes. Values are given as observed/expected for the -3 to +3 flanking positions of both sites with a colour code associated with values >1 (blue) or <1 (red). D) Correlations of the -3 to +3 flanking profiles of Site 1 and Site 3 were obtained in panel C for all four complex combinations. Panels A, C and D were adapted from Appendix 5 and panel B was prepared based on Appendix 5.

To further investigate the hypothesis of preferential RD interface formation of R882H subunits, an adjusted Deep Enzymology approach was used to determine the flanking sequence preference profiles of homo- and heterotetramers. Due to the strong changes in the preferences of DNMT3A upon R882H mutation observed in a previous

part of this thesis (Appendix 5 Figure 2A and section 4.3.1), which took place especially on the 3' flank, this change in the preference profiles was used to determine the nature of the RD interface subunits. For this purpose, homo- and heterotypic complexes were prepared with all subunits being active (Appendix 5 Figure S4) and Deep Enzymology reactions were conducted after treatment with factor Xa to remove the activity-lowering MBP-tag (Appendix 5 Figure S3). As depicted in Figure 19B, the randomized substrates contained two CpG sites in a distance of 12 bps, where co-methylation should be observed in opposite strands (Site 1 and Site 3). Using this setup, the flanking sequence preferences of the inner subunits could be determined by hairpin-bisulfite conversion and NGS. Data from different experimental repeats were filtered for product molecules that were methylated at both CpG sites, averaged due to their high overall correlation (Appendix 5 Table S1), and the obs/exp ratios were calculated for the -3 to +3 flanks of the co-methylated substrates. As already seen in Appendix 4, a comparison of the different tetramer reactions showed very similar patterns for the 5' flanks with strong disfavour of A and G and preference for T at the -2 position and preference for A and C at -1 for all complexes (Figure 19C and Appendix 5 Figure 2C). However, strong differences were observed for the 3' flanks, especially at the +1 position, where C was preferred and G disfavoured by the wild-type homotetramer, while A or G were preferred and T disfavoured by the R882H homotetramer.

Intriguingly, both mixed complexes showed high correlations with the 3' flank patterns of R882H alone (Figure 19D and Appendix 5 Figures 2C, D), confirming the hypothesis of the preferential formation of R882H/R882H RD interfaces. This analysis was then further extended and refined with a closer look at the +1 to +4 preferences of both enzymes. Extraction of the methylation levels of all CGNNNN sites showed once more strong differences between DNMT3A WT and R882H mutant for all four positions, with up to 5-fold preferences of the wild type and 2-fold preferences of the mutant enzyme for some specific sequences (Appendix 5 Figure 3). Finally, our findings were brought into a structural context using recently published crystal structures (Anteneh et al., 2020) and MD simulations, confirming the preferred formation of mutant RD interfaces since the R882H/R882H complex showed a tighter packing due to higher numbers of inter-subunit contacts and buried more solvent accessible surface area than WT/WT RD interfaces (Appendix 5 Figure 4).

4.4 Investigation of the flanking sequence preferences of TET1 and TET2

DNA methylation is a dynamic epigenetic signal and its removal via sequential oxidation of the methyl group is initiated in mammals by three active Ten-eleven translocation enzymes (Ito et al., 2011; He et al., 2011; Tahiliani et al., 2009). Despite the highly conserved catalytic domains, the enzymes were shown to be expressed differently during mammalian development, with TET3 mostly present in the early stages while TET1 and TET2 play roles during PGC development and processes in the neural system (Rasmussen and Helin, 2016), suggesting distinct roles for the different enzymes. Furthermore, TET2 is the member most frequently found mutated in cancer (Jiang, 2020). Similar to the methylation mark, the first oxidation product hm5C is a stable DNA modification, which functions among other roles as a marker of active genes (Ficz et al., 2011) and has specific reader proteins such as UHRF2 (Spruijt, et al., 2013; Zhou et al., 2014).

To unravel the mechanism behind the sequential oxidation, several crystal structures of TET2 (Hu et al., 2013; Hu et al., 2015) and NgTET1 from *Naegleria gruberi* (Hashimoto et al., 2014; Hashimoto et al., 2015) have been published including different modifications of the co-crystallized DNA. All structures show the target cytosine flipped out into the active site of the enzyme accompanied by DNA bending, similar to the mechanism used by the DNA methyltransferases. Detailed characterization of the TET-DNA interactions indicated that only the phosphate backbone of the bound DNA and the CpG site are involved in target site recognition. This conclusion was in line with a biochemical study that showed no further influence of the nucleotides flanking the CpG site on the enzymatic activity (Hu et al., 2013, Hu et al., 2015), but the mechanistic similarity to the DNMT3 enzymes and two other small-scale studies (Kizaki et al., 2016; Pais et al., 2015) might suggest otherwise. Beyond the preferred reaction in CpG context (Hashimoto et al., 2014; Hu et al., 2013), both TET1 and TET2 can oxidize non-CpG target sites, although the specific order is still under debate (DeNizio et al., 2021; Hu et al., 2013). To elucidate the potential influence of flanking sequence preferences on those two TET family members, further biochemical studies were conducted as part of this thesis and the results were published in Communications Biology (see attached Appendix 6). For the biochemical characterization, the same Deep Enzymology approach as for DNMT1 was used with

pools of randomized DNA substrates containing either CpG or CpN target sites but with m5C (M) or hm5C (O) modification (Figure 20A) on the upper strands. Oxidation reactions were then performed with the catalytic domains of TET1 and TET2 (for more information see Appendix 6 Figure S10) followed by hairpin ligation and bisulfite conversion coupled to deep Next-Generation Sequencing. This approach allowed the discrimination between the hm5C and f5C reaction state but not between the other modifications (Appendix 6 Figure S1C).

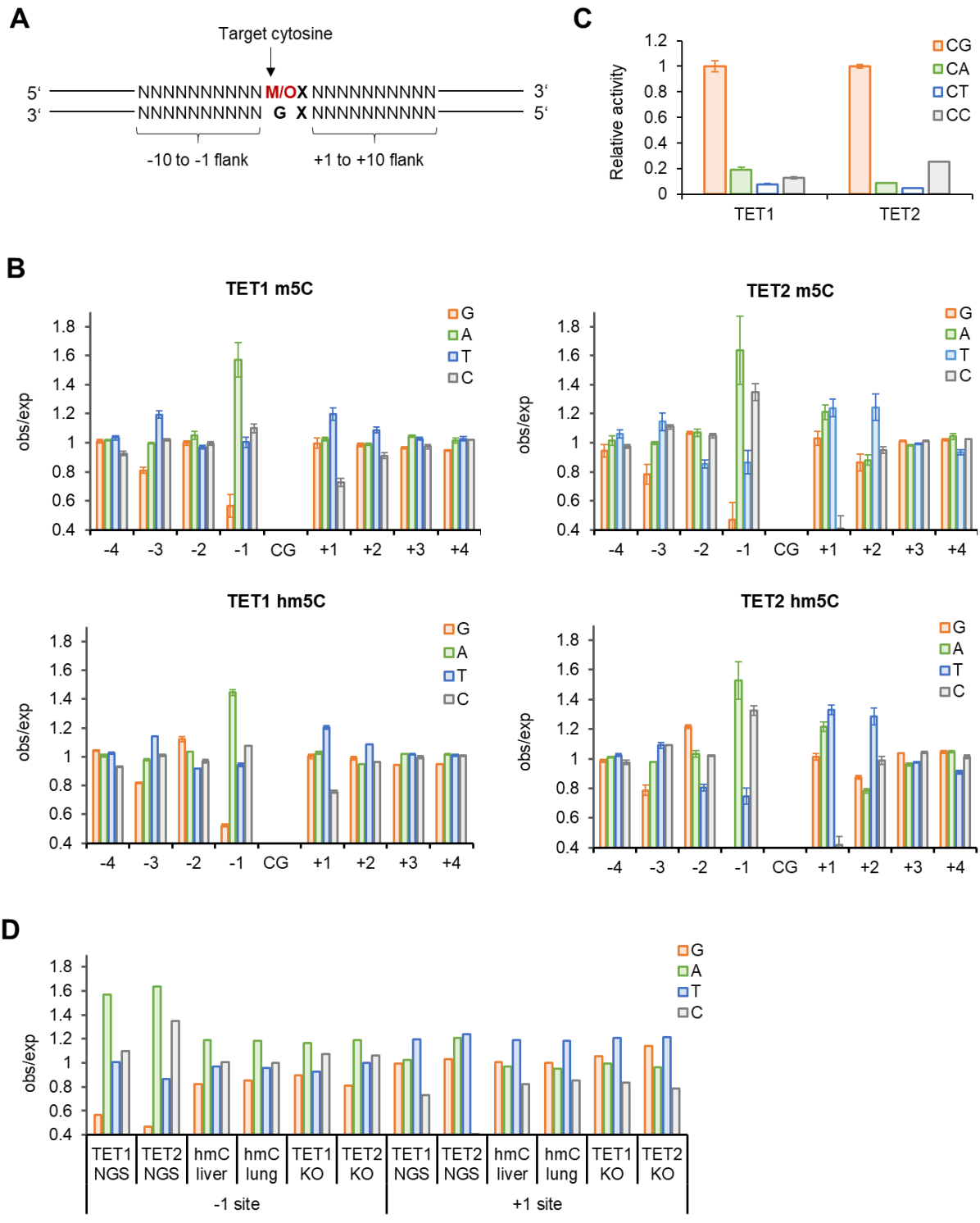


Figure 20: Flanking sequence preference analysis of TET1 and TET2. A) Design of the substrate pool used in the Deep Enzymology approach with one hemimethylated or hemihydroxymethylated CpG target site (methylation or hydroxymethylation in the upper strand marked as red “M” or “O”, respectively) embedded into a context of 10 randomized nucleotides on either side. B) Averaged oxidation levels of the mCpG (m5C) and hmCpG (hm5C) substrates determined after kinetic reactions with TET1 and TET2. Values are given as observed/expected (obs/exp) for every base

at the -4 to +4 flank positions. C) Relative oxidation activity of TET1 and TET2 on CpN substrates with normalization to CpG activity for both enzymes and error bars given as standard deviations. D) Comparison of the TET1 and TET2 preference profiles determined for the -1 and +1 flanking position (NGS) with the enrichment or depletion of specific bases in genome-wide hm5C patterns (hmC) extracted from liver and lung cells (Li et al., 2016b) or at sites that showed an increase in DNA methylation after knockout (KO) of TET1 or TET2 in mouse ESCs (Reimer et al., 2019). Panel A was prepared based on Appendix 6 and panels B, C and D were adapted from Appendix 6.

First, oxidation states were extracted for all sequenced DNA molecules with a CpG target site and the respective flanking sequence contexts were determined. Due to the high correlation of the independent experimental repeats performed with both enzymes (Appendix 6 Figures S2A, S3A), data sets were averaged and used for an initial analysis of the relative depletion or enrichment of individual bases at each flanking position (Appendix 6 Figures 1D, F). As shown in Figure 20B, the -1 position was observed to have the highest influence on TET activity with strong disfavour for G and favour for A, but effects were also visible to a lesser extent at the -2 and -3 positions as well as for +1 and +2 (for TET2). Furthermore, combined effects were determined for both substrates for each NNCGNN flank with monoexponential fitting of the averaged oxidation activities conducted by the research group of Prof. Nicole Radde (Institute for Systems Theory and Automatic Control, University of Stuttgart). With this analysis, average rate constants for the flanking sequence-dependent oxidation reactions were determined, revealing up to 22-fold and 69-fold differences for the m5C containing substrates for TET1 and TET2, respectively, and up to 8-fold and 25-fold differences in the oxidation rates for the hm5C modified substrates (Appendix 6 Figure 2A). Overall, flanking sequence effects were stronger for TET2, although the preference profiles were similar for both enzymes (Appendix 6 Figures 2A, B). Nevertheless, small differences regarding the most preferred and most disfavoured NNCGNN contexts could be worked out by visualization as Weblogos, which further refined the effects for both enzymes and cytosine modifications (Appendix 6 Figure 2C).

To verify the overall flanking sequence preferences determined with the Deep Enzymology approach, oxidation reactions were performed with TET1 and TET2 on m5C or hm5C hemi-modified substrates containing flanks specifically designed to be highly preferred or disfavoured. The reaction progress was hereby detected using HPLC-MS/MS conducted in the group of Prof. Jens Brockmeyer (Department of Food

Chemistry, Institute of Biochemistry and Technical Biochemistry, University of Stuttgart) (example kinetics in Appendix 6 Figure 3A). In agreement with known literature, decreasing oxidation rates were observed from m5C to f5C, but interestingly, an appropriate fit for the experimental data from the preferred substrate could only be achieved through the inclusion of a processive oxidation step from m5C to f5C (Appendix 6 Figures 3B, C and S7). In general, methylation rates were increased for the preferred flanking context and similar to the Deep Enzymology data stronger effects were observed for TET2.

Since previous studies also indicated TET activity in non-CpG context (DeNizio et al., 2021; Hu et al., 2013), the Deep Enzymology investigation was then extended by using a pool of substrates that contained one m5C modified CpN site for competitive methylation. Experimental repeats again showed a high overall correlation (Appendix 6 Figure S8) and averaged data revealed striking differences in the order of CpH methylation. Both TET1 and TET2 highly preferred methylation in CpG context, however, the next best target site was CpA in the case of TET1 followed by CpC and very low CpT methylation, whereas TET2 preferred CpC over CpA (Figure 20C and Appendix 6 Figure 4A). Although the preference profiles were very similar for all three higher methylated target sites, the variability was greater for the oxidation rates of CpA and CpC than for CpG sites suggesting greater importance of flanking sequence preferences in non-CpG context (Appendix 6 Figure 4B). Furthermore, frequency plots compiled for comparison of all NNCNNN flanks except for the ones in CpT context revealed a large overlap in the activity between the different substrates (Appendix 6 Figure 4C). This demonstrates that CpG sites in disfavoured flanking sequence context are less suitable for oxidation by TET1 and TET2 than non-CpG sites in preferred flanking context.

To bring all our findings into a biological context, the effects of TET flanking sequence preferences on the DNA methylation and hydroxymethylation landscape in cells were investigated focusing on the -1 and +1 flank that showed the greatest influence on the enzymatic activity. hm5C levels were extracted for all CpG sites from published whole genome bisulfite data of human lung and liver cells (Li et al., 2016b) as well as from reduced representation bisulfite data of wild-type and TET1 or TET2 knockout mouse ESCs (Reimer et al., 2019). For both data sets, enrichment and depletion of bases at

hm5C sites at the -1 and +1 position were in very good agreement with the Deep Enzymology data (Figure 20D and Appendix 6 Figure 5A). Similar correlations were also found with the average increase in methylation after TET knockout. Furthermore, m5C and hm5C levels determined for all genomic NNCGNN flanks in the first published data sets also showed high correlations with the combined TET flanks determined as part of this study (Appendix 6 Figure 5C). Generally, flanking sequence effects were not only visible on a global scale but also for local hm5C patterns as exemplarily shown for some chromosomal regions (Appendix 6 Figures 5E-G).

In addition to the flanking sequence dependence of the writer proteins TET1 and TET2, EMSA experiments conducted in this project also discovered preferences for the oxidation product readout by UHRF2 (Appendix 6 Figures 3D, E), which has previously been shown to specifically bind to hm5C (Spruijt, et al., 2013; Zhou et al., 2014). Overall, the findings clearly demonstrate that the flanking sequence preferences of the TET enzymes shape DNA modification patterns on a local and genome-wide scale.

5 Discussion

Over the past decades, increasing insights into the dynamic regulation of DNA methylation patterns in mammalian cells were obtained, which is an essential process in differentiation determining the fate of every cell in a multicellular organism. Accurate establishment as well as faithful inheritance of the specific epigenetic information, which goes beyond the information of the DNA sequence itself, was shown to be crucial since aberrant DNA methylation has been linked to the pathogenesis of various diseases, most notably cancer (Robertson, 2005). In general, the transfer of the methyl group from the cofactor AdoMet to the C5 position of cytosine nucleobases to generate 5-methylcytosine is mediated by two classes of DNA methyltransferases (DNMTs). On the one hand, DNMT3A and DNMT3B generate the *de novo* methylation patterns during development, whereas, on the other hand, DNMT1 preserves the genomic methylation state during cell proliferation to prevent replication-dependent loss of the epigenetic information (Jurkowska et al., 2011d). Nevertheless, a necessity for active removal of DNA methylation marks was shown to emerge at different developmental stages, which is accomplished by the family of Ten-eleven translocation (TET) enzymes (Rasmussen and Helin, 2016).

Hence, the activities of these different DNA methyltransferases, as well as the methylcytosine dioxygenases at different target sites, are fundamental for the dynamic regulation of gene expression. Mutations of the involved enzymes were shown to play a role in different severe diseases such as the ICF syndrome (Ehrlich et al., 2008), Tatton-Brown-Rahman syndrome (Tatton-Brown et al., 2014) or different cancers, most notably AML (Brunetti et al., 2017), although their exact pathogenic mechanisms were not completely unravelled yet. Therefore, this work aimed to obtain further insights into the exact DNA sequence readout mechanisms of the DNMTs and TET enzymes on a biochemical, structural and cellular level, including not only how the enzymes recognize their target sites but also how the interactions with the different flanking sequence contexts of these target sites influence their activity. Deepening our understanding of these mechanisms is essential since sufficient enzymatic activity is only possible with productive DNA binding mediated by direct enzyme-DNA contacts that typically span over 10 bps. In the following paragraphs, the findings of the four projects presented in the Results section are discussed and further compared to the existing literature in the respective fields. Notably, usage of the new Deep Enzymology

approach implemented in this work led to the discovery of previously unknown flanking preferences of DNMT1, TET1 and TET2, striking connections between mechanistic DNMT3A and DNMT3B differences and their individual genomic targets as well as important insights into the fundamental basis for altered DNA methylation patterns in cancer cells containing the specific DNMT3A R882H mutation.

5.1 Profound flanking sequence preferences of DNMT1

Faithful inheritance of the cellular DNA methylation patterns has been proven to be essential in mammalian cells in several knockout and deletion studies (Li et al., 1992; Fan et al., 2001; Takebayashi et al., 2007), where lack of the required DNMT1 activity was shown to be lethal. Intrinsic properties of the maintenance methyltransferase such as its preference for hemimethylated over unmethylated CpG target sites as well as its processivity (Jeltsch, 2006) have been set as the basis for the specific role of DNMT1, but how the enzyme directly copes with the different flanking contexts of the target site remained elusive. Even though all three DNMTs use the same base-flipping mechanism for methylation of the target cytosine (Song et al., 2012; Zhang et al., 2018), no real systematic attempts have been made yet to investigate the DNA sequence readout of DNMT1 (Jeltsch et al., 2021). Furthermore, although some early studies hinted at flanking sequence effects (Feltus et al., 2003; Flynn et al., 1998), prior investigations of a small subset of substrate sequences with different 5’ flanks conducted in our laboratory could not unravel any relevant flanking preferences of DNMT1 (Bashtrykov et al., 2012b). Things changed with the results of the methylation experiment on a long hemimethylated DNA substrate, which was also performed in this study to investigate the processivity of DNMT1 in more detail. In this experiment, very striking differences in the methylation rates of different CpG sites were observed, which most easily could be connected to the corresponding flanking sequences. To follow up on this observation, comprehensive biochemical studies were performed in our laboratory with the conduction of DNMT1 catalysed methylation reactions on a randomized hemimethylated DNA pool. The results were then brought into structural and molecular context with help from the research groups of Prof. Jikui Song (Department of Biochemistry, University of California) and Prof. Nicole Radde (Institute

for Systems Theory and Automatic Control, University of Stuttgart) and were published as attached in Appendix 1.

As shown in Figure 15, unrecognized flanking sequence preferences of DNMT1 were identified in the scope of this project, showing up to 100-fold differences in the methylation rates of the most and least preferred NNCGNN sequence. Overall, DNMT1 activity on the hemimethylated CpG target sites was mostly influenced by the -2 to +2 flanking positions, where the enzyme was shown to highly prefer T at -2 and A or T at +2, whereas methylation was hampered at sequences with C at -2 or G at -1. These preferences extend and refine the findings of another study (Yoder et al., 1997), which only identified disfavour for G at the -1 position (based on a rather small data set) but did not observe any of the other effects. Intriguingly, the preferences determined with the Deep Enzymology approach, which could also be verified with targeted kinetics, differed from the flanking profiles of DNMT3A and DNMT3B (see section 4.2.1) indicating that they are determined by details of the structure of the DNMT-DNA complex.

Similar preferences as determined with the randomized substrate pool were also obtained in the single molecule kinetics on the long DNA substrate mentioned above. Some minor deviations in the sequence rankings between both data sets could be explained by the small number of investigated sequences on the long DNA substrate. Considering how DNMT1 enzymatic activity is influenced by the different DNA sequences surrounding the target sites, kinetic models were determined by the research group of Prof. Nicole Radde (Institute for Systems Theory and Automatic Control, University of Stuttgart) (Appendix 1 Figures 1D and 2) that once again verified that DNMT1 is indeed processive since a solely distributive kinetic model could not capture the biochemical DNA methylation data. In addition to the CpG site-specific kinetic parameters, the new model also identified conformational changes of the enzyme to a closed state as the basis for the high processivity of DNMT1. The final mathematical model based on differential equations provides much greater insight into the molecular mechanism of DNMT1 than previous models that neglected this intrinsic enzymatic property.

Structural interpretation of the kinetic data was started using an existing DNMT1 crystal structure in complex with a hemimethylated DNA substrate in a strongly disfavoured

GG**CG**GC (GCG complex) flanking context (Song et al., 2012). The new information that this flanking context is unfavourable could explain the large rearrangements of the orphan guanine, which was engaged in a non-canonical GG base pair with the G(-1), and other surrounding bases observed after flipping the target cytosine into the active site, which altogether disrupted the B-DNA helix over 3 base pairs. Due to these findings, we were motivated to investigate crystal structures of DNMT1 in complex with substrates containing CpG sites in a neutral or preferred flanking context, which were prepared by the research group of Prof. Jikui Song (Department of Biochemistry, University of California). Strikingly, almost no (CCG complex) or only moderate (ACG complex) conformational changes around the target base were observed in these newly determined crystal structures, as shown in Figure 21A. Another divergence of the complexes was observed regarding the conformational state of the catalytic helix, which transitioned from a mainly kinked conformation in the case of the favoured substrates into a rather straight conformation in the case of the disfavoured substrate as shown in Figure 21B (Song et al., 2012). Based on recently published enzymatic assays and MD simulations, this structural effect was concluded to directly impact the catalytic activity of DNMT1 (Ye et al., 2018). These findings led to the conclusion that larger conformational changes of the DNA or active site of DNMT1 result in lower enzymatic activity due to decreased efficiency of active complex formation. So far, the newly observed conformational changes of DNMT1 are unique among the mammalian DNA methyltransferases. In addition to the general effects, specific details about the mechanism of DNA sequence readout at the -1 and -2 positions could be pinpointed to certain inter-molecular DNA base interactions (e.g. stacking interactions) that also affect the minor groove width (Appendix 1 Figure S8). Together, the biochemical and mechanistic data obtained in the scope of this project provided significant insights into the maintenance methylation process.

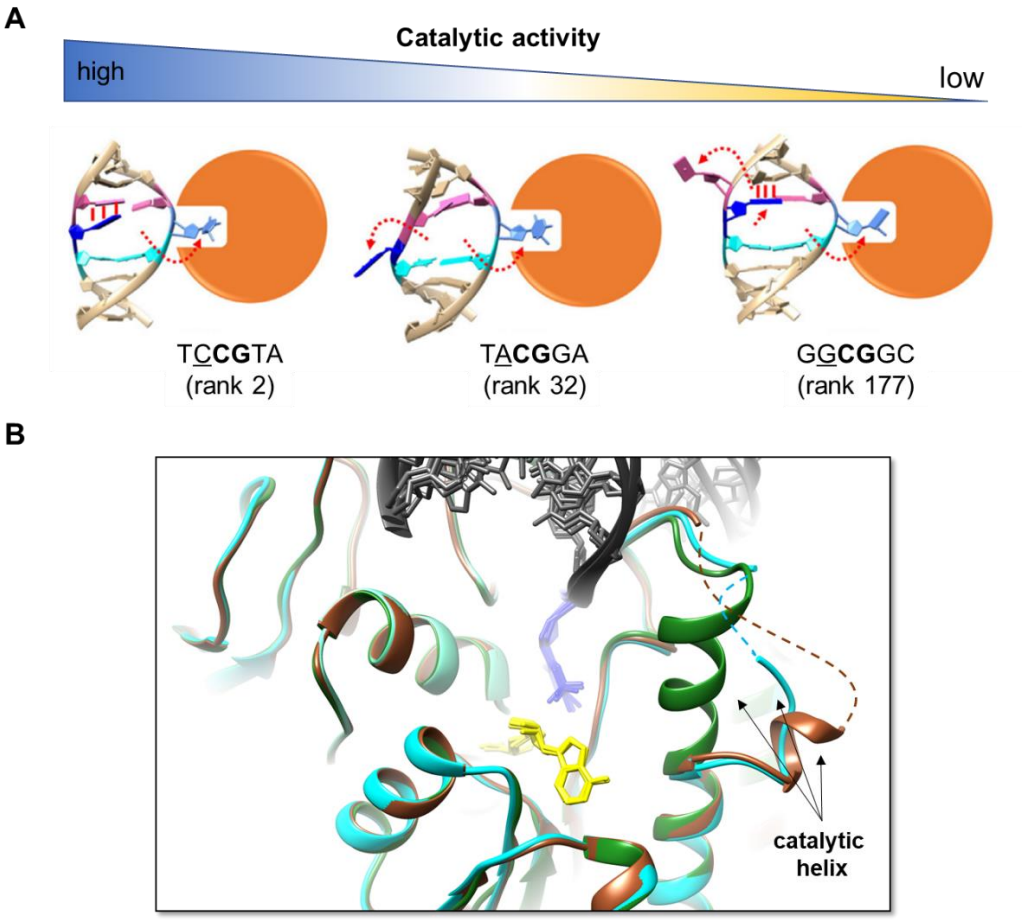


Figure 21: Schematic overview of the structural changes of DNMT1 resulting from the binding to different flanking sequence contexts. A) Flipping-out of the target cytosine (light blue) into the active site of DNMT1 (orange circle) is accompanied by small rearrangements (shown with red dashed arrows) in the case of the preferred CCG complex (PMID: 6W8W), whereas the middle-ranked ACG (PMID: 6W8V) and disfavoured GCG (PMID: 4DA4) complexes showed strong rearrangements of the orphan guanine (dark blue) and the 5' positioned basepair (pink). The 3' positioned basepair is shown in cyan. B) Different conformations of the catalytic helix in the preferred CCG complex (PMID: 6W8W, shown in brown), middle-ranked ACG (PMID: 6W8V, shown in cyan) and disfavoured GCG complex (PMID: 4DA4, shown in green). For all three complexes, overlays of the DNA (grey), the flipped-out target cytosine (dark blue) and S-adenosyl-L-homocysteine (yellow) are shown. Dashed lines are used to represent the disordered N-terminal regions of the catalytic helices. Panel A was adapted from Jeltsch et al. (Jeltsch et al., 2021) and Panel B was prepared based on Appendix 1.

To investigate the relevance of these novel findings for cellular DNA methylation, the NNCGNN preference profiles obtained for DNMT1 with the Deep Enzymology approach were compared to different publicly available whole genome bisulfite sequencing data sets. Not only were they strongly correlated with genomic DNA methylation patterns extracted from human ESCs (Charlton et al., 2018) and lung

cancer cells (Hascher et al., 2014), treatment of the latter with the potent DNA methyltransferase inhibitor 5-azacytidine had a significantly higher influence on the methylation of CpG sites in disfavoured flanking sequence context (Appendix 1 Figure 6C). Furthermore, a comparison of the obtained preferences with data from DNMT knockout studies in mouse ESCs (Li et al., 2015) verified that a knockout of DNMT1 in DNMT1 single knockout cells (1KO) or DNMT1/3A/3B triple knockout cells (TKO) leads to loss of the significant correlation of the average genomic methylation levels of NNCGNN sites with the flanking sequence preferences of DNMT1. As expected, DNA methylation patterns in the 1KO cells still strongly correlated with DNMT3A and DNMT3B preferences obtained in another project of this thesis (Appendix 2), which was lost in DNMT3 double knockout (DKO) or the TKO cells (Appendix 1 Figure 7). With this, it was proven that DNMT1 flanking sequence preferences shape the dynamic DNA methylation landscape in cells, which perfectly fits its role as the maintenance methyltransferase (Jeltsch and Jurkowska, 2014).

Despite the fascinating insights obtained throughout this project, many interesting factors influencing DNMT1 activity and specificity still await further investigation, one being the impact of different chromatin modifications or interaction partners such as UHRF1 and PCNA on DNMT1 activity and flanking sequence preferences. Based on our findings that the full-length and catalytically active domain of DNMT1 showed highly overlapping preference profiles (Appendix 1 Figure S4D), interactions mediated by the N-terminal domains of the protein are suspected to only enhance the enzymatic activity but not alter the flanking sequence preferences. Nevertheless, this hypothesis has to be validated, including mutational studies to determine the influence of other domains such as the CXXC domain on the recognition of hemimethylated CpG sites. The role of this specific region for the specificity of DNMT1 was debated (Bashtrykov et al., 2012a; Song et al., 2011), therefore it would be interesting to use the Deep Enzymology approach to clear up this discrepancy. For therapeutic usage, different DNMT1 inhibitors could be investigated with the Deep Enzymology workflow combined with structural and cellular studies to determine their potential influence on DNMT1 specificity, flanking sequence preferences or other enzymatic properties. Furthermore, the exact influence of oxidized cytosine bases emerging from TET oxidation on the activity of DNMT1 could be investigated using modified substrates as described in section 4.4. This question is important to unravel the effect of TET enzymes in passive

DNA demethylation executed by blocking DNMT1. Finally, transfer of the knowledge obtained for DNMT1 to bacterial DNA methyltransferases such as M.HhaI (Klimasauskas et al., 1994) and M.HaeIII (Reinisch et al., 1995) would be possible, since these enzymes are known to utilize a similar base-flipping mechanism on the target cytosine.

5.2 Different DNA sequence readout mechanisms of DNMT3A and DNMT3B

The precisely regulated activity of DNMT3A and DNMT3B is the fundamental basis for the correct establishment of mammalian DNA methylation patterns during early development, where these enzymes introduce methylation in CpG as well as non-CpG context (He and Ecker, 2015; Okano et al., 1998). Organized chromatin recruitment of the enzymes was shown to be mediated by their interaction with other epigenetic signals such as posttranslational histone tail modifications, transcription factors, non-coding RNAs or other DNA and chromatin binding proteins (Laisné et al., 2018). Although both enzymes share their largely conserved MTase domain as well as increased enzymatic activity in the presence of the regulatory family member DNMT3L (Jeltsch and Jurkowska, 2016), distinct functionalities have been identified that lead to the development of different diseases (Brunetti et al., 2017; Gagliardi et al., 2018). Despite the tremendous efforts made during the past decades, several aspects of the underlying mechanisms that determine the enzymatic activity of the *de novo* DNA methyltransferases are still unknown or under debate. Some of these open questions were resolved in the scope of this work with the usage of the newly developed Deep Enzymology approach together with structural, cellular and biochemical data leading to the publications attached as Appendices 2 and 3.

5.2.1 Distinct *de novo* methylation mechanisms and their biological connections

Detailed insights into the mechanisms of DNMT3A and DNMT3B mediated DNA methylation were obtained through the determination of the refined flanking sequence preference profiles of the two enzymes, which were then connected to specific DNA sequence readout mechanisms using the already published crystal structure of DNMT3A/3L in complex with DNA (Zhang et al., 2018) and new crystal structures of DNMT3B/3L co-crystallized with different DNA sequences (Appendix 2) which were provided by the research group of Prof. Jikui Song (Department of Biochemistry, University of California). In all complexes, the formation of linear heterotetramers was observed, with two DNMT3A or DNMT3B subunits at the central positions, and DNMT3L subunits positioned on the outer sites of the complexes (Appendix 2 Figure 10A). Since DNA binding at the central RD interface is only mediated by DNMT3A or DNMT3B subunits, flanking sequence preferences were determined for the catalytic domains of these enzymes without DNMT3L in most experiments through methylation of a pool of randomized substrates that contained one CpG target site. With this approach, profound and distinct preferences were determined for both DNMT3s as shown in Figure 16, showing the strongest influence ranging from the -2 to +3 positions. The preferences of both enzymes differed mostly on the 3' flank (where DNMT3A prefers pyrimidines at the +1 position, whereas DNMT3B prefers purines), while the 5' flank preferences were highly similar. These findings, which were also verified in targeted kinetics with designed preferred and disfavoured substrates (Appendix 2 Figures S3A, B) are in perfect agreement with data obtained in several small- to midscale studies conducted in the past (Handa and Jeltsch, 2005; Jurkowska et al., 2011c; Lin et al., 2002; Wienholz et al., 2010). However, the applied Deep Enzymology approach provides quantitative methylation information for all possible flanking sequences in an NNNCGNNN or NNCGNN context (depending on the sequencing depth). As expected from the crystal structures, the preferences were not altered in the presence of DNMT3L (Appendix 2 Figure S9), which is in line with other publications (Mao et al., 2020; Norvil et al., 2020) but contradicts previously obtained biochemical data derived from a small study including only 48 CpG sites (Wienholz et al., 2010). Strikingly, the over 100-fold differences in the DNMT3 preferences could be traced back to subtle conformational differences in the RD interface as well as the catalytic and TRD loop of the central subunits resulting in a decreased number of

DNMT3B DNA backbone contacts compared to DNMT3A as shown in Figure 22. The structural differences between DNMT3A and DNMT3B were most pronounced for the TRD loop, involving DNA interacting key residues such as T775, K777 and N779 in DNMT3B (Appendix 2 Figure 3). Notably, DNMT3B was seen to specifically recognize CGA sequences, a motif which is already known from a prominent target of DNMT3B, the SatII repeats. Hypomethylation of these repeats due to loss of DNMT3B activity, which cannot be rescued by DNMT3A, was shown to cause the ICF syndrome (Ehrlich et al., 2008). Indeed, the Deep Enzymology data obtained in this thesis provide a mechanistic explanation for this pathologic effect, since the motif of the SatII repeats was proven to be one of the sequences most preferentially methylated by DNMT3B (Appendix 2 Figure 1D). Similar biological connections between the sequences of their most relevant repeat region substrates and enzyme preferences could be obtained for both DNMT3A and DNMT3B as well as the human and mouse enzymes of the latter (Appendix 2 Figure 1G), which clearly showed a co-evolution of the flanking sequence preferences with their respective physiological targets.

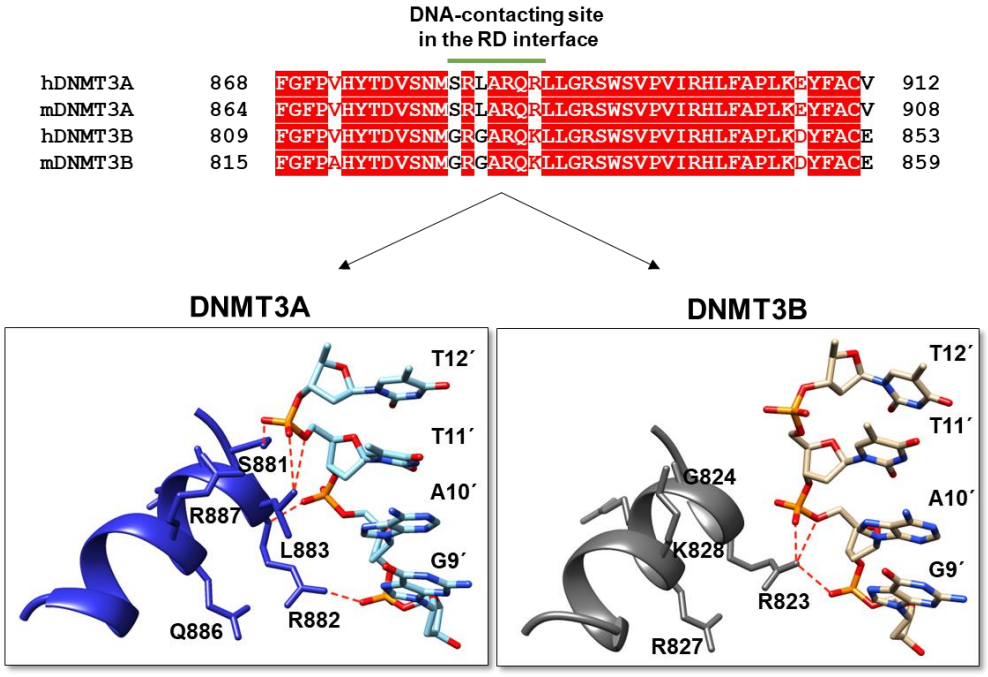


Figure 22: Structural basis for the different flanking sequence preferences of the DNMT3s. The divergence in the preference profiles of the two enzymes mainly results from the different amino acid sequences in the RD interface (upper part) of human and mouse DNMT3A and DNMT3B, leading to more DNA phosphate backbone contacts in the respective loop structure of DNMT3A (PDBI: 5YX2) compared to DNMT3B (PDBI: 6U8V) (lower part). The Figure was prepared based on Appendix 2.

To further expand our knowledge about the mechanism of non-CpG methylation, flanking sequence preferences were also determined for both DNMT3 enzymes using substrate pools with CpN target sites in a randomized context. With this approach, the known CpG specificity of both enzymes (Okano et al., 1998) was confirmed, this time in the context of all possible flanking sequences. Both enzymes showed the greatest difference in flanking sequence preferences of non-CpG methylation at the +1 position. Interestingly, this effect could be linked to the amino acid K777 in DNMT3B (Appendix 2 Figure 6), a key residue which was also investigated in much greater detail in another publication I contributed to (Dukatz et al., 2020). Overall, the higher non-CpG methylation activity of DNMT3B than of DNMT3A observed by others (Aoki et al., 2001; Suetake et al., 2003) was verified although to a different extent, hinting at a combined readout of the target site and flanking sequence context. Meanwhile, similar conclusions regarding the structural basis of the flanking sequence preferences in CpG- and CpH-context as well as connections to the same key amino acids K777 and N779 were also drawn by other groups (Lin et al., 2020; Mallona et al., 2021; Mao et al., 2020), which once more verifies the results obtained in the scope of this project. Despite the available crystal structures discussed here, which already include different DNA sequence contexts, many more would be needed for sufficient explanation of the underlying mechanisms of the observed biochemical data, e.g. for other flanking positions than the +1 position. To achieve this goal, tremendous improvements in the field of cryo-EM and other crystallization technologies are necessary.

Remarkably, the determined DNMT3 CpN flanking sequence preferences were also recapitulated in genomic data from different cell types. For example, Lee et al. discovered that distinct methylation in neurons and brain cells in CAC context can be attributed to DNMT3A, whereas methylation in ESCs in CAG context is mainly set by DNMT3B (Lee et al., 2017). Nevertheless, interactions of the DNMT3 enzymes with additional cellular factors such as histone tail modifications and transcription factors, which are mediated through the different N-terminal domains of the proteins, remain to be investigated. For this, full-length proteins instead of the catalytically active MTase domains would be needed. In line with this, studies of DNMT3A methylation in nucleosomal context (Bröhm et al., 2022), which were also conducted in our laboratory and hinted towards similar flanking sequence preferences shaping the linker DNA

methylation levels, could be extended by introducing randomized sequence parts into the nucleosomal DNA.

5.2.2 Mechanisms of co-methylation of two CpG sites by DNMT3A

Due to the tetrameric structure of DNMT3A and DNMT3A/3L complexes, the enzyme is capable of introducing two methylation marks into one DNA molecule through one binding event as seen in the crystal structure of DNMT3A/3L bound to a DNA molecule with two CpG sites (Zhang et al., 2018). Regarding these co-methylation events, three different types can be distinguished, being co-methylation of the first CpG site in the upper strand and the second site in the lower strand (MW), the reverse situation (WM) and co-methylation in the same strand (MM or WW) with methylation of both upper or lower sites taken together. MW co-methylation by DNMT3A can be realized through the individual binding of each of the central DNMT3A subunits to one of the two target sites, but this interaction is restricted towards methylation on opposite strands in MW mode due to structural constraints. In addition, the distance of the co-methylated cytosines influences the methylation efficiency, which was shown to be the highest at a distance of 12 bps (Gao et al., 2020a). In contrast, previous *in vitro* and *in vivo* experiments hinted at a preferred distance of 8 to 10 bps for co-methylation by DNMT3A (Jia et al., 2007; Jurkowska et al., 2008). To resolve this discrepancy and to understand how the enzyme can methylate differently placed target sites on a genome-wide scale, the Deep Enzymology approach was adapted to study the co-methylation of two target sites in a randomized context and in variable distance of 2 to 15 bps. Like this, different modes of co-methylation (MW, WM or MM as shown in Figure 17) could be studied in all possible DNA contexts, which was intended since bending of the DNA in the crystal structures (Zhang et al., 2018) was shown to be not only distance but also sequence-dependent. Using this method, new insights into various co-methylation mechanisms of DNMT3A and DNMT3A/3L tetrameric complexes were obtained (attached as Appendix 3) through connected biochemical studies from our laboratory and scanning force microscopy (SFM) studies performed with the research group of Prof. Ingrid Tessmer (Rudolf Virchow Center for Integrative and Translational Bioimaging, University of Würzburg).

In comparison to the already mentioned biochemical studies (Jia et al., 2007; Jurkowska et al., 2008), one of the big advantages of the used method was the possibility to discriminate between simultaneous co-methylation events and independent methylation of the two target sites. The latter would result in a statistically evenly distributed methylation independent of the target site distance, which was not in agreement with our data (Appendix 3 Figures S9, S10) and could therefore be ruled out.

For interpretation of the observed short-range co-methylation, two well-established models of one-dimensional diffusion (Halford and Szczelkun, 2002; Halford and Marko, 2004) were taken into consideration. Firstly, short distance MM co-methylation can be explained with a model, where DNMT3A slides along the DNA, introducing more than just one methylation without releasing the substrate. According to this model, the described effect would be fading out with increasing distances of the target sites due to decreasing diffusion efficiency of the enzyme, which was also observed in our experiments. Secondly, short distance MW and WM co-methylation, where DNA methylation occurs on opposite DNA strands, can be explained by a “hopping” of the enzyme complex. In this model, the central subunits would dissociate and re-associate multiple times and hereby stochastically switch strands and target sites. Nevertheless, both of these models could not explain the observed preferences of the enzyme to co-methylate CpGs at greater distances of 6, 9 and 12 bps.

In agreement with the published crystal structure (Zhang et al., 2018), our Deep Enzymology data showed a clear preference for MW co-methylation of target sites in the distance of 12 bps for both DNMT3A alone and DNMT3A in the presence of DNMT3L (model and crystal structure shown in Figure 23A). Intriguingly, the DNMT3A/3L complex even showed an enhanced favour for this distance, which could result from the structural limitations of the DNMT3A/3L complex due to the outer DNMT3L subunits. Hypothetically, a DNMT3A tetramer or even higher multimer structures (Jurkowska et al., 2011b) would rise the opportunity for more relaxed co-methylation at different distances due to the increased number of subunits that can participate in the DNA interaction. Using SFM, preferential binding of two target sites in 12 bps distance was verified, with either one single tetramer bound to the DNA as determined by volume analyses resulting in DNA bending of roughly 40° or two single

tetramers which then causes a doubling of the volume and bending angle (Appendix 3 Figures 5-7). Such additive effects would indicate the binding of two tetramers side-by-side, where only one tetramer methylates both sites while the other binds non-specifically next to it. Notably, this perfectly fits previous experiments demonstrating the cooperative binding of DNMT3A/3L to DNA (Jurkowska et al., 2008; Rajavelu et al., 2012). Similar co-methylation investigations were also performed in a simultaneously published paper, where the authors kept one target site fixed introducing a zebularine and then placed CpG sites at different distances on the same substrate (Gao et al., 2020a). In agreement with our data, they also observed preferred binding of the DNA and co-methylation of CpG sites in a distance of 12 bps and also another but weaker peak for shorter distances.

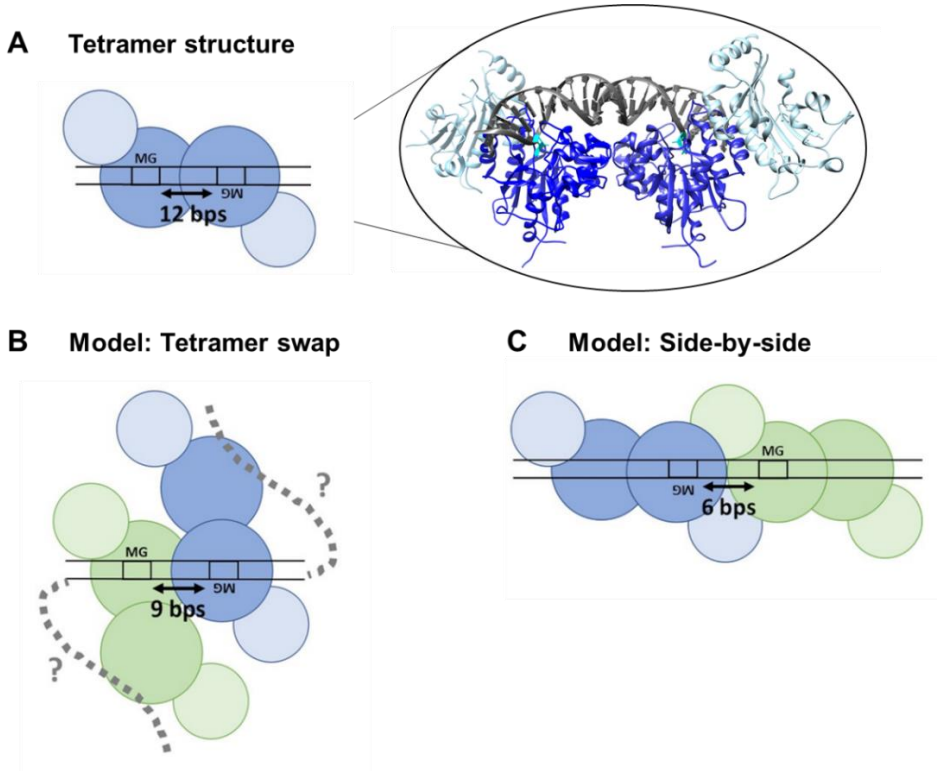


Figure 23: Schematic illustrations of the DNMT3A tetramer structure and potential models to explain different types of co-methylation by the enzyme complex. A) The DNMT3A/3L crystal structure (PDBI: 6BRR; 3A subunits in dark blue and 3L subunits in light blue, DNA in grey with flipped-out cytosines in cyan) is in accordance with MW co-methylation in 12 bps distance (M representing the methylated cytosine). B) Model of MW co-methylation in shorter distances by two adjacent but swapped DNMT3A/3L tetramers. C) Model of WM co-methylation of two target sites in short distances by a linear side-by-side organization of two adjacent tetramer complexes. Panels A-C were adapted from Appendix 3 and the crystal structure in Panel A was prepared based on Zhang et al. (Zhang et al., 2018).

In addition to the peak at 12 bps, preferences for co-methylation at distances of 5-6 (WM) and 8-9 (MW) bps were observed during the analysis of our NGS data. For the shorter distances, the effect was also reflected by the SFM data although to a weaker extent, with long elongated shapes of the complexes on the DNA accompanied by roughly 100° DNA bending. An explanation for these observations would be a similar side-by-side model as described for 12 bps, but in this case, each of the adjacent tetramers would interact with one target site, respectively (Figure 23C).

In contrast, MW co-methylation by DNMT3A/3L tetramers in a distance of 8-9 bps can not be explained by a side-to-side model, therefore we came up with a structural model that includes a swap of the DNA between the two bound tetramers (Figure 23B). In this arrangement, one tetramer would also interact with just one of the two target sites and very compact complexes of greater height would be formed, which was verified by SFM analyses (Appendix 3 Figures 5-7). Strong DNA bending of roughly 100° was determined in the SFM studies in contrast to the hypothesized lack of bending, which could result from additional binding of the outer DNMT3A subunits to the DNA substrates, leading to the formation of secondary loop structures. Due to these unexpected results, further investigation of the proposed tetramer swap model would be needed, potentially involving binding mutants of the enzyme or mutations that change other properties of DNMT3A such as R882H (see section 4.3.1).

Finally, the DNA sequence dependency of the observed co-methylation modes was investigated for the 12 bps distance substrate through a detailed look at the randomized regions between the two target sites in the methylated fraction of molecules. Since DNA bending of up to 100° was proven to be necessary for co-methylation in the SFM studies and the crystal structure also showed compression of the central minor groove (Zhang et al., 2018; crystal structure in Figure 22A), we hypothesized that co-methylation should be dependent on the sequence context at the centre of bending. Indeed, a clear enrichment of AT base pairs was observed, which fits the intrinsic bending known to be introduced by A-tracts and the fact that an overall high AT/GC ratio leads to a minor groove compression through the absence of the N2-amino group of guanosines (Allemann and Egli, 1997). Notably, these influences of the DNA sequence context on co-methylation were not investigated so far, since other studies were always performed with fixed and mostly highly repetitive sequences

including several CpG sites (Gao et al., 2020a; Jurkowska et al., 2008; Rajavelu et al., 20012). Overall, the newly obtained data and DNA interaction models can connect all previously published preferences, explaining why DNMT3A/3L is able to introduce co-methylation in various distances despite its intrinsic preference for a distance of 12 bps resulting from the structural organization of the tetramer subunits (Zhang et al., 2018). This finding is also more in agreement with the biological role of DNMT3A/3L, which has to generate genome-wide DNA methylation patterns flexibly (Jeltsch and Jurkowska, 2016; Okano et al., 1999). In the future, this adaptation of the Deep Enzymology approach could also be used in the context of different mutant subunits to see, if co-methylation preferences are altered by critical DNMT3A mutations.

5.3 Pathogenic mechanism of the DNMT3A R882H mutation

Changes in the genetic and epigenetic state of cells are associated with different diseases, above all the development and progression of cancer (Baylin and Jones, 2011; Bergman and Cedar, 2013). Here, somatic mutations have been acquired that result in altered structural and functional features of the tumour cells in line with their upregulated growth rate. In these tumour cells, abnormal transcriptional repression or activation of genes, so-called epimutations, are the consequences of specific mutations occurring in epigenetic regulators such as the family of DNA methyltransferases or methylcytosine dioxygenases followed by global changes in the chromatin landscape (DiNardo and Cortes, 2016). One of the most prominent factors illustrating this mechanism is DNMT3A, which was found to be mutated in around 25% of all patients with AML and normal karyotype. Most of these mutations were shown to be missense mutations, with R882H accounting for over 60% of them (Brunetti et al., 2017). Since DNMT3A mutations, in general, have been connected to the bad prognosis of patients suffering from AML e.g. due to decreased treatment efficiency, extensive efforts have been made to unravel the underlying pathogenic mechanism of the very frequent R882H mutation. Structural data suggested that the location of this residue in the DNA binding cleft of the DNMT3A/3L complex (Zhang et al., 2018) provides an explanation for potential changes of some mutant properties. Indeed, previously published data from our laboratory confirmed that the activity of DNMT3A in different flanking sequence contexts was strongly altered upon R882H mutation

(Emperle et al., 2018b). Substantial extensions of these conclusions, which so far were only made based on the methylation levels of 56 different CpG sites, were obtained in the third project of this thesis using Deep Enzymology data combined with structural and cellular investigations as shown in Appendices 4 and 5.

5.3.1 Altered flanking sequence preferences of DNMT3A R882H

The homodimeric RD interface, which builds up the central interface of the DNMT3A/3L heterotetramer (Zhang et al., 2018), has been shown through mutational analyses to be involved in DNA binding as well as the variance of catalytic activities of the enzyme complex on target sites embedded into different DNA sequence contexts (Gowher et al., 2006). Additionally, stabilization of the catalytic loop, as well as a loop from the target recognition domain that harbours the R882 residue, were explicitly shown to impact DNA methylation by DNMT3A (Zhang et al., 2018). In the crystal structure, R882 was observed to engage in specific phosphate backbone contacts with the +3 position of the CpG site in the bound DNA substrate, indicating that an indirect readout of the DNA sequence-dependent conformations could be mediated by this residue among others. In agreement with these findings, flanking sequence effects ranging from the -2 to the +3 position were observed in this project as shown in Figure 18, similarly to what was shown on a smaller scale before (Emperle et al., 2018b). Furthermore, drastic up to 70-fold differences in the NNNCGNNN flanking sequence preferences of R882H compared to DNMT3A WT were determined using the Deep Enzymology approach, which were also verified using targeted kinetics on designed DNA substrates (Appendix 4 Figures 2 and 4A). Notably, our findings recapitulated all effects previously observed with 56 CpG sites, including the preference of R882H for G and A at the +1 position where the wild-type enzyme prefers C, as well as the loss of G disfavour at the +2 and +3 positions by the mutant enzyme (Emperle et al., 2018b). Interestingly, other mutations of R882 found in AML (R882C, S, P) showed very similar flanking sequence preference profiles as determined for R882H (Appendix 4 Figure 4), suggesting a common pathomechanism resulting from the loss of the backbone contact due to removal of the R882 side chain.

A more detailed analysis of the data obtained in the scope of this thesis revealed that the R882H mutation does not affect the preference profile at the 5’ flank, but strong

alterations are caused at the 3’ flank. These findings fit the orientation of the 882 residue in the wild-type and mutant crystal structures, as shown in Figure 24.

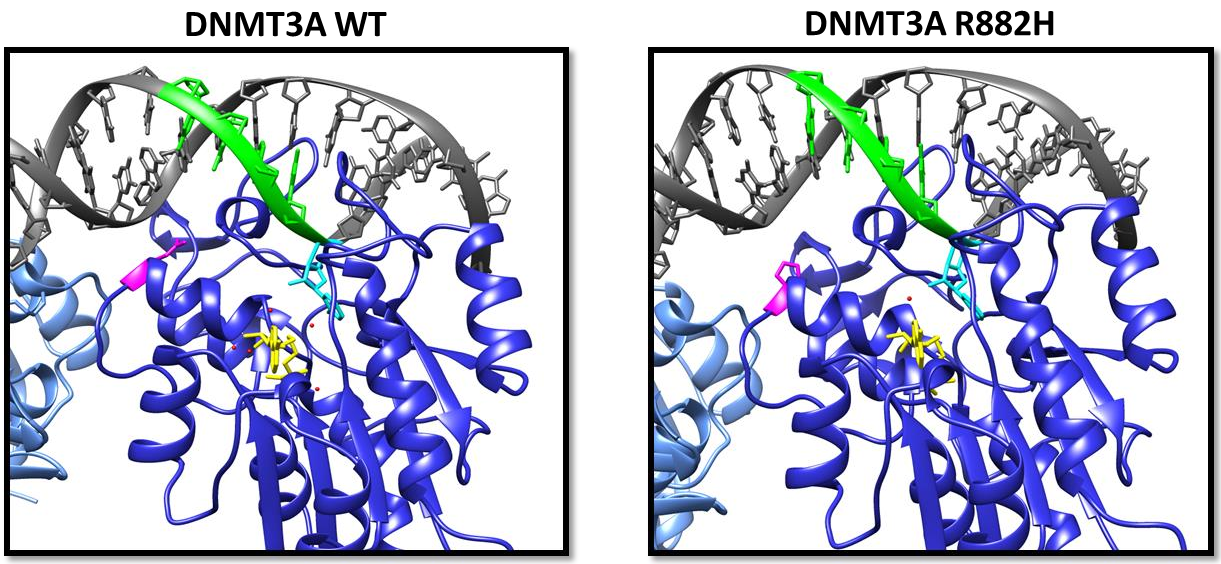


Figure 24: Orientation of the 882 amino acid in DNMT3A WT and R882H crystal structures regarding the 3’ flank of the target cytosine. Schematic view of the position of R882 (left) and H882 (right) at the RD interface of a DNMT3A/3L tetramer (Anteneh et al., 2020; PDBI: 6W8B, 6W89). DNMT3A subunits are shown in light and dark blue, respectively, with R882H marked in magenta. The cofactor S-adenosyl-L-homocysteine is shown in yellow, water molecules are shown in red and the bound DNA is coloured in dark grey, with the flipped-out cytosine in cyan and its 3’ flank marked in green.

Strikingly, R882H flanking sequence preferences at these positions reflected the preferences of DNMT3B rather than the preferences of DNMT3A WT. This chimeric behaviour of R882H, which was also observed by others (Norvil et al., 2020), could be rationalized in our second project by the fact, that the R882 containing TRD loop was shown to be a key region determining the different preferences of the two DNMT3 enzymes on this side of the CpG site (Appendix 2). In parallel, Anteneh et al. published crystal structures of DNMT3A WT or R882H mutant in complex with DNA substrates of different CpG and CpA flanking contexts (Anteneh et al., 2020), where the authors observed different sequence-dependent flexibility modes of the TRD loop. Conformational changes were hereby associated with altered inter-residue or DNA-protein contacts, resulting in the verification of different target sites and +1 base pair readouts. Overall, increased TRD dynamics were connected with decreased CpG

specificity of the mutant enzyme together with a shift towards methylation of CG(G/A) sites overlapping with our Deep Enzymology results.

Finally, our kinetic data was brought into a biological context using either cellular experiments or published DNA methylation data from AML patients (Appendix 4 Figures 5, 6). For the *in vivo* investigations, DNMT3A WT or R882H were overexpressed in DNMT1 hypomorph HCT116 cells, for which we could show that global DNA methylation levels highly correlated with the determined symmetrical R882H/DNMT3A* preference profiles (averaged preferences of pairs of corresponding complementary flanks). Overall, our data rationalized the specific loss of methylation at major satellite repeats shown in mouse DNMT3 DKO cells overexpressing R882H (Norvil et al., 2020), since these specific target site contexts were proven to be highly disfavoured by R882H with the Deep Enzymology approach. The authors of this paper also showed that R882H expression in DNMT3A and DNMT3B DKO cells maintained DNA methylation at the minor satellite repeats, which are a prominent target of DNMT3B, once again emphasizing the DNMT3B-like switch of this specific DNMT3A mutant. Furthermore, sites harbouring strong hypermethylation in R882H-containing tumours from AML patients (Glass et al., 2017) were also demonstrated to be in very good agreement with the mutant flanking sequence preferences. In summary, detailed gain-of-function effects resulting from the introduction of a R882H mutation in DNMT3A were determined in this project, which explained the hypermethylation of one-third of the differentially methylated target sites through changes in the flanking sequence preferences of the mutant enzyme.

Taken together, our new data provide a significant contribution to the understanding of the pathogenic effect of the R882H mutation in patients with AML. However, further biochemical, structural and cellular investigations would be needed to ultimately unravel all mechanistic bases behind this specific mutation and its interplay with other DNMT3A mutations (Brunetti et al., 2017), p53 WT or mutants (Muller and Vousden, 2013), histone methyltransferases such as DOT1L (Rau et al., 2016) or other regulatory enzymes such as TET2 and its mutants (Jiang et al., 2020). In addition, Anteneh et al. provided some insights into the structural effect of the R882H mutation (Anteneh et al., 2020), nevertheless, more crystal structures including DNA substrates with e.g. different DNA bases at the +3 position would be needed to fully connect our

biochemical data with structural effects. In the end, potential therapeutic usages could be based on the knowledge obtained in this project, examining the DNA sequence-dependent effects of R882H when using potent inhibitors similarly to what has been tried in our laboratory before (Emperle et al., 2018b).

5.3.2 Preferred self-assembly of DNMT3A R882H homodimeric interfaces

Despite the essential insights obtained about the changes in flanking sequence preferences of the R882H mutant protein with the usage of the Deep Enzymology approach, these alterations are still not strong enough to fully explain the pathogenic mechanism of the mutant protein in AML patients. Especially the fact, that this mutation occurs in a heterozygous manner (Brunetti et al., 2017; Yang et al., 2015), meaning in combination with fully functional DNMT3A WT expressed from the non-mutated allele, would argue against the repeatedly observed strong effect of R882H, which has been demonstrated to be at least 50-70% as active as the wild-type enzyme (Emperle et al., 2018a; Emperle et al., 2018b; Holz-Schietinger et al., 2012; Yan et al., 2011). Consequently, different functional models of the mutant and wild-type subunit interactions have been postulated in the past, one being a dominant-negative effect of R882H subunits on the WT subunits in heterotetrameric complexes through blocking of interface formation (Holz-Schietinger et al., 2012; Russler-Germain et al., 2014), but this proposal lacked a conclusive mechanistic explanation and it was challenged in another work from our laboratory (Emperle et al., 2018a). In contrast, stabilization of DNMT3A oligomers including R882H mutant subunits was shown by another group (Nguyen et al., 2019) resulting in a decrease in the mutant's enzymatic activity. Surprisingly, our kinetic data obtained using mixed DNMT3A complexes of active and inactive wild-type and R882H subunits showed that active R882H combined with an inactive WT* subunit (asterisk marks catalytically inactive C710S subunits) in double-tag purified complexes had an even higher methylation activity than WT/WT*, while WT/R882H* and R882H/R882H* complexes showed reduced activity (Figure 19A). The only model in agreement with these kinetic data postulated a preferential formation of RD interfaces built from two independent R882H mutant subunits (Figure 25), resulting in an enrichment of these specific complexes. This model would explain why the WT/R882H* shows even less activity than the homodimeric R882H/R882H

complex since the preferred R882H* subunits in the centre would further lower the DNA methylation. In contrast, the boost in activity seen for R882H/WT* compared to the homodimeric WT/WT* complex could be the result of the inactive wild-type subunits taking the outer positions, whereas the active R882H subunits would preferably take the inner positions in the heterotetramer. In contrast, an equal distribution of subunits is expected in WT/WT*, resulting in 50% of WT/WT activity.

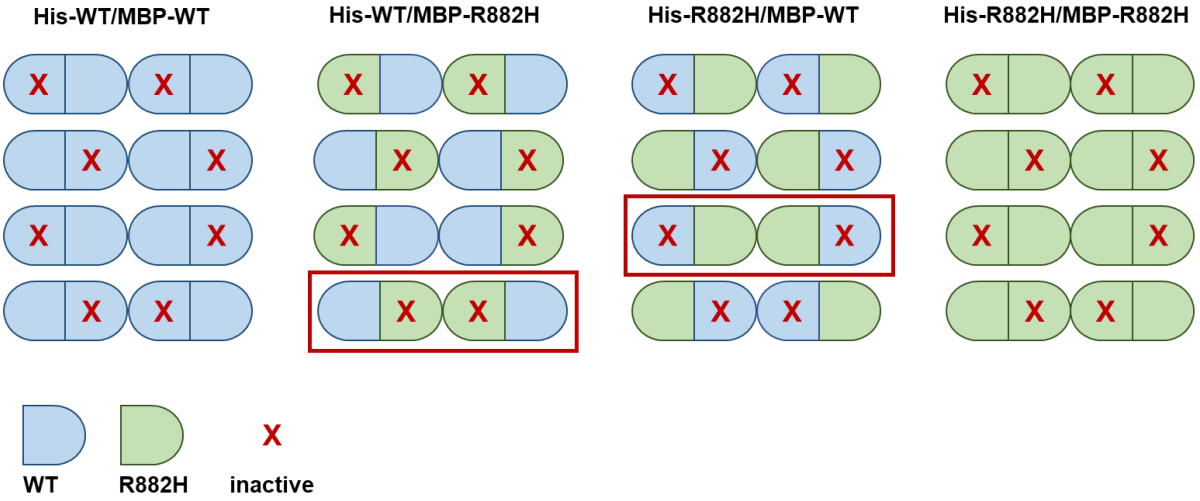


Figure 25: Schematic illustration of all possibly formed homo- and heterotetrameric DNMT3A complexes, including active and inactive WT and R882H subunits. The formation of all potentially build-up RD interfaces is shown with WT subunits (WT) in blue and R882H mutant subunits (R882H) in green. Inactive subunits with the MBP-tag are marked with a red cross. Enrichment of the R882H homodimeric RD interface complexes is indicated by the red boxes. The scheme was prepared based on Appendix 5.

To verify this new model, which for the first time could explain the mechanistic basis of dominant effects observed for R882H, the adjusted Deep Enzymology approach from section 4.2.2 was used to investigate the structural organization in homo- and heterotetrameric DNMT3A complexes. In this experiment, co-methylation of two CpG sites by mixed complexes containing different combinations of active DNMT3A WT and R882H subunits (WT/WT, R882H/R882H, WT/R882H and R882H/WT) was studied and the 5' and 3' flanking sequence preference profiles were determined for both target sites. As expected, all complexes showed similar 5' profiles, but differences at the 3' flanks. Intriguingly, the 3' preferences of all mixed complex profiles strongly correlated with the profile obtained for the R882H homotetramers. Consequently, the results of this adjusted Deep Enzymology approach verified that R882H subunits preferentially form homodimeric RD interfaces in mixed DNMT3A WT/R882H tetramers.

Structurally, the formation of mixed RD interfaces could be disfavoured since inter-subunit contacts are increased through a concerted and symmetrical conformational rearrangement of the RD loops when H882 is present in both central subunits (Anteneh et al., 2020). The disfavour for WT/WT RD interfaces was verified in our project using MD simulations that demonstrated a decreased solvent accessibility of R882H/R882H RD interfaces due to the tighter packing of the area (Appendix 5 Figure 4). As a next step, further systematic *in vivo* experiments would be needed to validate our findings in a cellular context.

In conclusion, strong alterations in the flanking sequence preferences, as well as the preferred self-assembly of homodimeric R882H interfaces can cause hyper- as well as hypomethylation at specific CpG sites compared to DNMT3A WT. Accordingly, not only dominant-negative but also dominant positive effects of these specific mutations can be mechanistically explained by this model for the first time. Therefore, this model is in very good agreement with high correlations of R882H hypermethylated regions found in AML patients (Glass et al., 2017) and our mutant preference profiles (Appendix 4) as well as the lack of major satellite repeats methylation caused by R882H mutation in mouse ESCs (Kim et al., 2013). However, despite the elegance of this model, it can not fully explain the loss of DNA methylation generally observed in tumour cells from AML patients (Russler-Germain et al., 2014), indicating that further investigations are needed regarding the interplay of R882H and other regulatory mutations that leads to the DNA hypomethylation of cancer cells (Baylin and Jones, 2011; Bergman and Cedar, 2013). Finally, potential targeted clinical approaches could be undertaken based on this preferential self-assembly of the R882H RD interfaces.

5.4 Distinct flanking sequence preferences of TET1 and TET2

Dynamic regulation of the DNA methylation patterns in cells on a local and global level is mediated through the interplay of DNA replication rates, DNMT activity, as well as the TET enzymes responsible for active DNA methylation removal (Jeltsch and Jurkowska, 2014). All steps of the sequential oxidation of m5C over hm5C to f5C and ca5C (Ito et al., 2011; He et al., 2011; Tahiliani et al., 2009) are catalysed by one of the three members of this dioxygenase family, TET1-3, which show differential expression patterns throughout the different stages of mammalian development

(Rasmussen and Helin, 2016). Especially mutations of TET2 were demonstrated to play a role in different cancers through the global loss of hm5C levels (Jiang, 2020). This specific cytosine modification has been associated with active gene transcription over the last several years and the view that it can act as a stable DNA modification with specific functions was further supported by the identification of specific hm5C readers such as the SRA domain of UHRF2 (Spruijt et al., 2013; Zhou et al., 2014). Structural investigations of the TET enzymes bound to DNA substrates in different DNA contexts and modification states (Hashimoto et al., 2014; Hu et al., 2013; Hu et al., 2015) shed light on the oxidation mechanism of these enzymes, which interestingly use the same base-flipping of the target cytosine into the catalytic pocket as the DNMTs. However, these mechanistic studies suggested that TET enzymes solely form direct contacts with the phosphate backbone of the bound DNA as well as the target CpG site, which was supported by one of the groups with small-scale biochemical data revealing an apparent lack of flanking sequence effects (Hu et al., 2015). Contrary, another group stated that the -1 position influences the enzymatic activity of the TET enzymes (Pais et al., 2015). So far, only one study tried to obtain information on potential TET preferences on a larger scale, but this approach included antibody enrichment and stringent filtering of the results, which could introduce a strong bias (Kizaki et al., 2016). In addition to the flanking effects, the target site specificity of TET1 and TET2 in CpN context beyond their high preference for CpG sites has been under debate, with CpC (Hu et al., 2013; Kizaki et al., 2016) or CpA (DeNizio et al., 2021) described as the next preferred cytosine context for TET2. Hence, further studies were needed to elucidate the specificity as well as possible flanking sequence effects of these TET enzymes, for which the Deep Enzymology approach provided highly significant novel insights as shown in Appendix 6.

In contrast to the mentioned literature findings, more than 50-fold differences in the NNCGNN flanking sequence preferences of TET1 and TET2 were determined on m5C containing randomized substrates with a monoexponential fitting of the obtained data performed by the group of Prof. Nicole Radde (Institute for Systems Theory and Automatic Control, University of Stuttgart), and to a lesser degree also on hm5C containing substrates. Overall, flanking sequence preferences were highly similar for both modifications (Figure 20), with effects ranging from -3 to +2 including the most pronounced effect on the -1 position. However, effects were stronger for TET2 and

smaller differences could be defined by analysis of subsets of highly preferred or disfavoured sequences (Appendix 6 Figure 2C). Verification of the determined preferences was achieved using HPLC-MS/MS measurements of the oxidation products obtained through TET1 and TET2 reactions with designed substrates in both m5C and hm5C modification states (Appendix 6 Figures 3A-C). The measurements, which were performed in the department of Prof. Jens Brockmeyer (Department of Food Chemistry, Institute of Biochemistry and Technical Biochemistry, University of Stuttgart), also provided further details about the reaction progress under the applied experimental conditions. In agreement with previous enzymatic assays (Hashimoto et al., 2014; Hu et al., 2015; Ito et al., 2011), decreasing conversion rates were observed with increasing oxidation state of the substrate for both enzymes, leaving the first m5C to hm5C conversion as the fastest reaction. A suitable explanation for this effect has already been provided by the crystal structure of TET2, which showed decreased efficiency of the hydrogen abstraction step due to more constrained binding of the more oxidized bases (Hu et al., 2015). Nevertheless, the size of the catalytic pocket was shown to be large enough to fit all modified states of the methyl group (Hu et al., 2013), which provides a basis for processive oxidation by the enzyme. Notably, the experimentally obtained HPLC-MS/MS data could only be fitted appropriately if a processive reaction step from m5C to f5C was included. Therefore, these data would agree with another publication on processive TET oxidation (Crawford et al., 2016), but also a distributive mechanism has been reported before (Tamanaha et al., 2016).

For structural interpretation of our Deep Enzymology results, published crystal structures of NgTET1 (Hashimoto et al., 2014), as well as two structures of human TET2 (Hu et al., 2013; Hu et al., 2015), were inspected regarding the DNA-protein interactions surrounding the target CpG site as depicted in Figure 26. Generally, conformational changes around the flipped-out cytosine ranged from the -3 to the +2 position. Focusing on the -1/+1 positions, one could directly connect the obtained preference profiles with structural elements. For the +1 position, the conserved amino acid R1302 was identified in the two TET2 structures to form either direct (CG(+1) in 5DEU) or water-mediated hydrogen bonds (TA(+1) in 5D9Y) to the minor groove, leading to a DNA context-dependent shift of the base pair in this position. Hypothetically, this shift towards the axis of the DNA helix could result in a decreased efficiency of target cytosine base-flipping, e.g. in the case of disfavoured C(+1). Our

hypothesis was indeed verified by a very recently published paper (Ravichandran et al., 2021), which also showed the “in”/“out” conformations of R1302 and the resulting hydrogen bond interactions in TET2 complexes with preferably favoured (A or T at +1) or disfavoured (C or G at +1) substrates. Interestingly, this essential +1 position effect has been neglected so far, since a previous small-scale publication stated that only the -1 position influences the enzymatic activity of the TET enzymes (Pais et al., 2015).

For the specific effect of the -1 flanking position, no amino acid mediated contact could be found, but stabilization of the orphan guanine G´ after base-flipping was proven to be influenced by the base pair at this position when all three crystal structures were compared. Strikingly, less preferred substrates bound by the 5DEU complex that harboured a G(-1´) and by the 4LT5 complex with a T(-1´) showed strong conformational changes of the orphan guanine after flipping of the target cytosine, whereas the most preferred substrate in the NgTET1 complex showed intrahelical stabilization of the orphan guanine. These structural effects are in perfect agreement with our Deep Enzymology data and following this line, even more pronounced effects should be observed for a DNA with C(-1´). Again, the conclusions about the influence of the non-target strand bases at the -1 position were verified in the recently published paper for the most and least preferred -1 base compositions (Ravichandran et al., 2021). Overall, this specific publication also demonstrated strong CpG flanking sequence preferences of all three TET enzymes, which mainly fits our obtained data, including only minor deviations in the preference ranking that probably resulted from their smaller scale of investigated sequence contexts.

In addition to the CpG flanking preferences, recognition of other target sites, as well as extended flanking effects, were also determined in the scope of this project, revealing differences in the order of mCpN oxidation for TET1 and TET2. Although both enzymes highly prefer oxidation of mCpG sites and mCpT oxidation was very low in both cases, TET1 was observed on average to prefer mCpA over mCpC sites, whereas TET2 showed inverse preferences. Overall, the strong preference for CpG target sites can be rationalized by specific structural effects. On the one hand, TET2 was shown to form direct interactions with the m5C that stabilize the flipping of the cytosine out of the DNA helix. On the other hand, further stabilization of the extrahelical conformation is achieved by hydrogen bonds and base-stacking interactions of the G:C base pair with

a tyrosine residue that is inserted into the emptied cytosine space (Hashimoto et al., 2014; Hu et al., 2013).

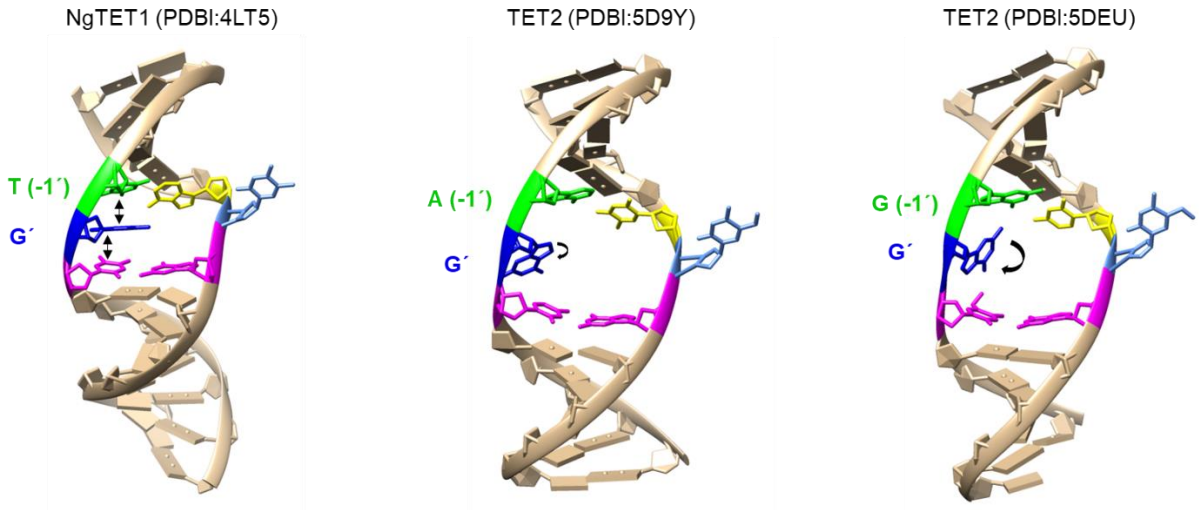


Figure 26: Comparison of three TET crystal structures determined with different flanking contexts of the target site. All three complexes show flipping of the target cytosine (shown in light blue) out of the DNA helix. The conformation of the orphan guanine (G´, shown in dark blue) was observed to be dependent on the -1 flanking base pair (shown in green and yellow), especially the base at the indicated -1´ position. In some cases, interactions with the +1´ position (shown in pink) also contribute to the stability of the active complex, e.g. in the NgTET1 complex. The figure was prepared based on Appendix 6.

In the non-CpG context, differences between the oxidation rates of various DNA sequence contexts were even more pronounced than for the CpG context, indicating that for these less preferred target sites, effects of the flanking sequences play an even bigger role in the reaction. Moreover, there was a significant overlap between the different preference profiles for CpG, CpA and CpC, which demonstrates that some of the most favoured non-CpG sites are more preferred by the enzymes than the most disfavoured CpG sites (Appendix 6 Figure 4C). Especially this finding was previously unknown since all other publications on this topic described a fundamental mCpG preference over all other DNA sequences (DeNizio et al., 2021; Dey et al., 2021; Hu et al., 2013; Kizaki et al., 2016; Ravichandran et al., 2021). Intriguingly, the preference of the different target sites was demonstrated to be modulated through the +1 flanking position, since e.g. T at this position enhanced mCpA oxidation by TET2, whereas A at this position resulted in elevated mCpC oxidation. Together, these flanking effects can be used to explain why previous publications stated contradicting results regarding

the non-CpG activity of the TET enzymes. Hu et al. observed 85% oxidation in CpG context by TET2 followed by 10% oxidation in CpC and 2% oxidation in CpA context with G(+1) (Hu et al., 2013), a CpG>CpC>CpA>CpT target specificity order that was also found by two other groups (Dey et al., 2021; Ravichandran et al., 2021) using a greater variance of DNA sequences. In contrast, other recent experiments on TET2 specificity reported mCpA as the next best target sequence of this enzyme, which fits our data since the used DNA substrates here contained the already annotated T(+1) (DeNizio et al., 2021).

Nevertheless, all these publications solely based their findings on arbitrarily chosen fixed sequences or small pools of different DNA sequences, which provide only a fraction of information compared to the data obtained with the Deep Enzymology approach. Only one other study tried a similar approach based on randomized flanks as used in our project but with an additional hm5C-antibody enrichment step of the reaction products (Kizaki et al., 2016). With this method, the authors determined strong preferences of TET1 for oxidation in CpG context, but they also observed activity in CpC context. However, their concluded preference profiles do not reflect our obtained data, probably because especially good or bad flanking sequences were enriched through the antibody pulldown, which would lead to a strong bias in the sequencing results. In the context of all these previous publications, we concluded that the competitive Deep Enzymology approach provides the most detailed CpN flanking sequence preferences so far based on the combined effects of all NNCNNN flanks.

To determine, if our obtained TET preference profiles also reflect local and global DNA methylation and hydroxymethylation patterns, we correlated our obtained NCGN or NNCGNN preferences with patterns extracted from published data sets. Based on whole genome bisulfite (detection of m5C and hm5C) as well as oxidative bisulfite (detection of m5C alone) sequencing data of human lung and liver cells (Li et al., 2016b), DNA methylation and hydroxymethylation levels of these cell types were determined for all possible NNCGNN flanking contexts, which showed strong overlap with our Deep Enzymology preference profiles. Furthermore, sites that showed a gain of methylation upon TET knockout determined through enhanced reduced representation bisulfite sequencing of mouse ESCs (Reimer et al., 2019) were in good agreement with our biochemical data, which could also be illustrated locally for

selected chromosomal regions Appendix 6 Figure 5G). Finally, we could show that cellular hm5C levels were most accurately predicted by a combination of the TET preferences combined with m5C levels (Appendix 6 Figure 5C), which is self-evident since DNA methylation first has to be set by DNMTs which also harbour specific flanking sequence preferences. Finally, the preferred oxidation of TET1 in CpA context would fit the reported hm5C levels determined in this sequence context in neurons (Mellén et al., 2017), where TET1 was proven to be essential (Guo et al., 2011). In conclusion, the detailed flanking sequence information of TET1 and TET2 obtained with the Deep Enzymology approach was proven to shape the cellular hydroxymethylome of different cell types.

Further investigations regarding the specificity and the preferences of the TET enzymes will be needed to fully grasp the cellular impact of these enzymes. Above all, TET3 preferences have to be systematically investigated, which were stated so far to be similar to the preferences of TET1 (Ravichandran et al., 2021). In addition, the bisulfite conversion step included in our Deep Enzymology approach can only distinguish m5C and hm5C from f5C and ca5C, leaving us without information about the first reaction step. Although very similar preferences for both m5C and hm5C containing substrate pools were observed, other methods to quantify oxidized bases (Liu et al., 2016) need to be introduced into the approach to verify these findings. Regarding the TET enzymes, changes in the determined preferences upon mutation or effects of different splicing isoforms (Melamed et al., 2018) could be studied to understand the potential mechanistic effects of these protein alterations. In addition to DNA, both TET1 (Basanta-Sanchez et al., 2017) and TET2 (Fu et al., 2014; He et al., 2021) were also shown to oxidize cytosine in an RNA context, especially in tRNAs, therefore the Deep Enzymology approach could be extended towards the usage of RNA as a substrate. For a more cellular context, systematic investigations of oxidation reactions in the nucleosome context could be performed as described before (Kizaki et al., 2016), potentially even including different histone modifications (Chrysanthou et al., 2022; Yamagata and Kobayashi, 2017). It is known from the literature that targeting and regulation of the TET enzymes are mediated by their interaction with posttranslational modifications, transcription factors or non-coding RNAs (Rasmussen and Helin, 2016), however, we hypothesized so far that these processes only recruit and activate the dioxygenases but the specificity and flanking sequence preferences

are dependent on the catalytic domains. In addition, specific biological targets and their respective roles could be investigated that fit the determined TET preferences, similar to the published connections drawn to transcription factor and Oct4 binding sites (Ravichandran et al., 2021). Finally, more detailed information about potential readers of the TET oxidation products could be obtained beyond UHRF2 binding to hm5C, with one potential candidate being MeCP2 binding to sites that contain hmCpA in neurons (Kinde et al., 2015; Sperlazza et al., 2017).

5.5 Conclusion and perspectives for the Deep Enzymology approach

The results of this doctoral thesis systematically determined the fundamental basis for variances in the enzymatic activity of mammalian DNA methyltransferases and methylcytosine dioxygenases on different target sites embedded into all possible flanking sequence contexts. In four different projects, new unprecedented insights into the biochemical properties of the involved enzymes were obtained with the developed Deep Enzymology approach, which couples single molecule kinetics with deep NGS readout. Strikingly, most effects could be connected to direct or indirect DNA sequence readout mechanisms of the enzymes (Garvie and Wolberger, 2001; Seeman et al., 1976) such as different degrees of conformational rearrangements of DNMT1 and TET1/2 dependent on the DNA context. In addition, the molecular basis for different DNA recognition mechanisms of the DNMT3 enzymes and the DNA interaction modes of DNMT3A applied for co-methylation of two target sites were determined. Furthermore, biological and above all pathogenic consequences were identified or validated including specific biological targets of DNMT3A or DNMT3B and the altered flanking sequence preferences and preferred self-assembly of DNMT3A subunits harbouring a R882H mutation that is often found in AML patients (Ley et al., 2010).

Taken together, the number of publications using the Deep Enzymology approach or partially adjusted versions of it (6 shown in this thesis) highlight the significant contribution that could be made to the epigenetic field with the development of this technology. One of the big advantages of this method is that it is hypothesis-free and all possible flanking sequence contexts are reacted on in competition, which nullifies the impact of specifically designed DNA substrates. Still, it is possible to design randomized substrate pools with specific target site density, arrangement or

modifications (and various combinations of those) to target defined research questions optimally. Due to the high amount of great-depth NGS data, that can be produced in a relatively short time, high levels of statistical relevance can be reached that allow significant comparisons with cellular data and direct determination of biological targets. In the end, using the knowledge obtained in this thesis on the properties of R882H and the effects of azacytidine treatment on DNMT1, new disease perspectives could also be provided, which might lead to an improvement in the diagnosis and therapy of patients in the future.

With NGS getting less expensive every day, this method will be the approach of choice to study the biochemical properties of DNMTs from other organisms (bacteria, plants, zebrafish etc.) that often share features with the already investigated mammalian DNMTs (Klimasauskas et al., 1994; Reinisch et al., 1995). It would also be interesting to include other mutational studies of DNMT3A or other DNMTs as already conducted for DNMT3B (Dukatz et al., 2020) to understand which enzymatic strategies are used to minimize flanking sequence effects on the enzymes. Furthermore, details about the mechanisms of DNA repair proteins such as TDG (DeNizio et al., 2021), RNA editing enzymes (Meier et al., 2016) or the DNA binding sequence specificity of different transcription factors and readers of methylated (or oxidized forms of) cytosines such as the members of the MBD family (Fatemi and Wade, 2006) await to be investigated. For this purpose, EMSA using randomized substrates or pulldown approaches could be introduced into the Deep Enzymology approach.

However, despite the seminal results that were obtained using the Deep Enzymology approach, there are still some minor drawbacks that have to be considered. Firstly, relatively high sequencing depths are needed to support statistical analyses of the flanking sequence effects, which can be overcome at least partly by the design of the DNA substrate pool. In addition, the approach so far includes only bisulfite conversion, rendering it impossible to distinguish between oxidized cytosine modifications stemming from TET oxidation. This could be achieved by the inclusion of adapted sequencing methods such as oxidative bisulfite sequencing or others to distinguish between the higher oxidized forms (Liu et al., 2016). Combined readout of modifications such as m5C and hm5C together with the respective DNA sequence is already possible with the help of Nanopore Sequencing (Laszlo et al., 2013; Schreiber

et al., 2013), which would make bisulfite conversion obsolete and simplify the bioinformatic analysis. In the future, it might also be possible to detect f5C and ca5C with this method, possibly even all cytosine modifications simultaneously. Furthermore, it is unfeasible so far to study the combined effect of modifications on both strands of the double-stranded DNA substrates on the enzymatic activities due to technical reasons, which have to be overcome in the future to reflect biological DNA states.

Moreover, there is also still a great discrepancy between the number of available crystal structures and the possible target site DNA contexts that can be investigated with the Deep Enzymology approach, so many biochemical effects could not be connected to their structural basis so far. To solve this issue, hundreds of additional structures would be needed, which could only be obtained through automated and parallelized structure determination methods. One attempt could be the usage of advanced crystallization techniques such as cryo-EM and especially molecular dynamics modelling, two methods that could also be combined in the future to achieve sufficient overlap of structural and biochemical data (Nierzwicki and Palermo, 2021). Nevertheless, it remains challenging to model the large conformational changes observed upon base-flipping of the target cytosine.

In the long-term perspective, targeting mechanisms of the involved enzymes by chromatin modifications or non-coding RNAs (Gowher and Jeltsch, 2018; Rasmussen and Helin, 2016) still need to be included in the Deep Enzymology approach to mimic cellular conditions more closely. Using chromatin-bound randomized DNA, it could be possible to get one step closer to ultimately unravelling the crosstalk between DNMTs, TETs and other involved proteins and how they dynamically regulate DNA methylation patterns in normal tissues as well as in the context of fundamental processes such as ageing (Bell et al., 2019) or cancer (Baylin and Jones, 2011; Bergman and Cedar, 2013).

6 References

Adam, Sabrina; Anteneh, Hiwot; Hornisch, Maximilian; Wagner, Vincent; Lu, Jiuwei; Radde, Nicole E. et al. (2020): DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. In Nat Commun 11 (1), p. 3723. DOI: 10.1038/s41467-020-17531-8.

Allemann, Rudolf K.; Egli, Martin (1997): DNA recognition and bending. In Chemistry & biology 4 (9), pp. 643–650. DOI: 10.1016/S1074-5521(97)90218-0.

Allis, C. David; Jenuwein, Thomas (2016): The molecular hallmarks of epigenetic control. In Nature reviews. Genetics 17 (8), pp. 487–500. DOI: 10.1038/nrg.2016.59.

Ambrosi, Christina; Manzo, Massimiliano; Baubec, Tuncay (2017): Dynamics and Context-Dependent Roles of DNA Methylation. In Journal of molecular biology 429 (10), pp. 1459–1475. DOI: 10.1016/j.jmb.2017.02.008.

An, Jungeun; Rao, Anjana; Ko, Myunggon (2017): TET family dioxygenases and DNA demethylation in stem cells and cancers. In Experimental & molecular medicine 49 (4), e323. DOI: 10.1038/emm.2017.5.

Anteneh, Hiwot; Fang, Jian; Song, Jikui (2020): Structural basis for impairment of DNA methylation by the DNMT3A R882H mutation. In Nature communications 11 (1), p. 2294. DOI: 10.1038/s41467-020-16213-9.

Aoki, A.; Suetake, I.; Miyagawa, J.; Fujio, T.; Chijiwa, T.; Sasaki, H.; Tajima, S. (2001): Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. In Nucleic acids research 29 (17), pp. 3506–3512. DOI: 10.1093/nar/29.17.3506.

Arand, Julia; Spieler, David; Karius, Tommy; Branco, Miguel R.; Meilinger, Daniela; Meissner, Alexander et al. (2012): In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. In PLoS genetics 8 (6), e1002750. DOI: 10.1371/journal.pgen.1002750.

Barau, Joan; Teissandier, Aurélie; Zamudio, Natasha; Roy, Stéphanie; Nalesso, Valérie; Hérault, Yann et al. (2016): The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. In Science (New York, N.Y.) 354 (6314), pp. 909–912. DOI: 10.1126/science.aah5143.

Basanta-Sanchez, Maria; Wang, Rui; Liu, Zhenzhen; Ye, Xiaohan; Li, Minyong; Shi, Xiaodong et al. (2017): TET1-Mediated Oxidation of 5-Formylcytosine (5fC) to 5-Carboxycytosine (5caC) in RNA. In Chembiochem : a European journal of chemical biology 18 (1), pp. 72–76. DOI: 10.1002/cbic.201600328.

Bashtrykov, Pavel; Jankevicius, Gytis; Smarandache, Anita; Jurkowska, Renata Z.; Ragozin, Sergey; Jeltsch, Albert (2012a): Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. In Chemistry & biology 19 (5), pp. 572–578. DOI: 10.1016/j.chembiol.2012.03.010.

Bashtrykov, Pavel; Ragozin, Sergey; Jeltsch, Albert (2012b): Mechanistic details of the DNA recognition by the Dnmt1 DNA methyltransferase. In FEBS letters 586 (13), pp. 1821–1823. DOI: 10.1016/j.febslet.2012.05.026.

Baylin, Stephen B.; Jones, Peter A. (2011): A decade of exploring the cancer epigenome - biological and translational implications. In Nature reviews. Cancer 11 (10), pp. 726–734. DOI: 10.1038/nrc3130.

Bell, Christopher G.; Lowe, Robert; Adams, Peter D.; Baccarelli, Andrea A.; Beck, Stephan; Bell, Jordana T. et al. (2019): DNA methylation aging clocks: challenges and recommendations. In Genome biology 20 (1), p. 249. DOI: 10.1186/s13059-019-1824-y.

Berger, Shelley L.; Kouzarides, Tony; Shiekhattar, Ramin; Shilatifard, Ali (2009): An operational definition of epigenetics. In Genes & Development 23 (7), pp. 781–783. DOI: 10.1101/gad.1787609.

Bergman, Yehudit; Cedar, Howard (2013): DNA methylation dynamics in health and disease. In Nature structural & molecular biology 20 (3), pp. 274–281. DOI: 10.1038/nsmb.2518.

Bestor, Timothy; Laudano, Andrew; Mattaliano, Robert; Ingram, Vernon (1988): Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. In Journal of molecular biology 203 (4), pp. 971–983. DOI: 10.1016/0022-2836(88)90122-2.

Blaschke, Kathryn; Ebata, Kevin T.; Karimi, Mohammad M.; Zepeda-Martínez, Jorge A.; Goyal, Preeti; Mahapatra, Sahasransu et al. (2013): Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. In Nature 500 (7461), pp. 222–226. DOI: 10.1038/nature12362.

Bonasio, Roberto; Tu, Shengjiang; Reinberg, Danny (2010): Molecular signals of epigenetic states. In Science (New York, N.Y.) 330 (6004), pp. 612–616. DOI: 10.1126/science.1191078.

Bourc'his, D.; Xu, G. L.; Lin, C. S.; Bollman, B.; Bestor, T. H. (2001): Dnmt3L and the establishment of maternal genomic imprints. In Science (New York, N.Y.) 294 (5551), pp. 2536–2539. DOI: 10.1126/science.1065848.

Breiling, Achim; Lyko, Frank (2015): Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. In Epigenetics & chromatin 8, p. 24. DOI: 10.1186/s13072-015-0016-6.

Bröhm, Alexander; Schoch, Tabea; Dukatz, Michael; Graf, Nora; Dorscht, Franziska; Mantai, Evelin et al. (2022): Methylation of recombinant mononucleosomes by DNMT3A demonstrates efficient linker DNA methylation and a role of H3K36me3. In Communications biology 5 (1), p. 192. DOI: 10.1038/s42003-022-03119-z.

Brunetti, Lorenzo; Gundry, Michael C.; Goodell, Margaret A. (2017): DNMT3A in Leukemia. In Cold Spring Harbor perspectives in medicine 7 (2). DOI: 10.1101/cshperspect.a030320.

Butcher, Lee M.; Ito, Mitsuteru; Brimpari, Minodora; Morris, Tiffany J.; Soares, Filipa A. C.; Ährlund-Richter, Lars et al. (2016): Non-CG DNA methylation is a biomarker for assessing endodermal differentiation capacity in pluripotent stem cells. In Nature communications 7, p. 10458. DOI: 10.1038/ncomms10458.

Carell, Thomas; Kurz, Matthias Q.; Müller, Markus; Rossa, Martin; Spada, Fabio (2018): Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA. In Angewandte Chemie (International ed. in English) 57 (16), pp. 4296–4312. DOI: 10.1002/anie.201708228.

Challen, Grant A.; Sun, Deqiang; Mayle, Allison; Jeong, Mira; Luo, Min; Rodriguez, Benjamin et al. (2014): Dnmt3a and Dnmt3b have overlapping and distinct functions in

hematopoietic stem cells. In Cell stem cell 15 (3), pp. 350–364. DOI: 10.1016/j.stem.2014.06.018.

Charlton, Jocelyn; Downing, Timothy L.; Smith, Zachary D.; Gu, Hongcang; Clement, Kendell; Pop, Ramona et al. (2018): Global delay in nascent strand DNA methylation. In Nature structural & molecular biology 25 (4), pp. 327–332. DOI: 10.1038/s41594-018-0046-4.

Chédin, Frederic; Lieber, Michael R.; Hsieh, Chih-Lin (2002): The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. In Proceedings of the National Academy of Sciences of the United States of America 99 (26), pp. 16916–16921. DOI: 10.1073/pnas.262443999.

Chen, Taiping; Ueda, Yoshihide; Dodge, Jonathan E.; Wang, Zhenjuan; Li, En (2003): Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. In Molecular and cellular biology 23 (16), pp. 5594–5605. DOI: 10.1128/MCB.23.16.5594-5605.2003.

Chrysanthou, Stephanie; Tang, Qin; Lee, Joun; Taylor, Samuel J.; Zhao, Yilin; Steidl, Ulrich et al. (2022): The DNA dioxygenase Tet1 regulates H3K27 modification and embryonic stem cell biology independent of its catalytic activity. In Nucleic acids research 50 (6), pp. 3169–3189. DOI: 10.1093/nar/gkac089.

Chuang, L. S.; Ian, H. I.; Koh, T. W.; Ng, H. H.; Xu, G.; Li, B. F. (1997): Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. In Science (New York, N.Y.) 277 (5334), pp. 1996–2000. DOI: 10.1126/science.277.5334.1996.

Cliffe, Laura J.; Kieft, Rudo; Southern, Timothy; Birkeland, Shanda R.; Marshall, Marion; Sweeney, Kate; Sabatini, Robert (2009): JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. In Nucleic acids research 37 (5), pp. 1452–1462. DOI: 10.1093/nar/gkn1067.

Cooper, D. N.; Youssoufian, H. (1988): The CpG dinucleotide and human genetic disease. In Human genetics 78 (2), pp. 151–155. DOI: 10.1007/BF00278187.

Cortellino, Salvatore; Xu, Jinfei; Sannai, Mara; Moore, Robert; Caretti, Elena; Cigliano, Antonio et al. (2011): Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. In Cell 146 (1), pp. 67–79. DOI: 10.1016/j.cell.2011.06.020.

Crawford, Daniel J.; Liu, Monica Yun; Nabel, Christopher S.; Cao, Xing-Jun; Garcia, Benjamin A.; Kohli, Rahul M. (2016): Tet2 Catalyzes Stepwise 5-Methylcytosine Oxidation by an Iterative and de novo Mechanism. In Journal of the American Chemical Society 138 (3), pp. 730–733. DOI: 10.1021/jacs.5b10554.

Dawlaty, Meelad M.; Breiling, Achim; Le, Thuc; Barrasa, M. Inmaculada; Raddatz, Günter; Gao, Qing et al. (2014): Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. In Developmental cell 29 (1), pp. 102–111. DOI: 10.1016/j.devcel.2014.03.003.

Dawlaty, Meelad M.; Breiling, Achim; Le, Thuc; Raddatz, Günter; Barrasa, M. Inmaculada; Cheng, Albert W. et al. (2013): Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. In Developmental cell 24 (3), pp. 310–323. DOI: 10.1016/j.devcel.2012.12.015.

Dawlaty, Meelad M.; Ganz, Kibibi; Powell, Benjamin E.; Hu, Yueh-Chiang; Markoulaki, Styliani; Cheng, Albert W. et al. (2011): Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. In Cell stem cell 9 (2), pp. 166–175. DOI: 10.1016/j.stem.2011.07.010.

Deaton, Aimée M.; Bird, Adrian (2011): CpG islands and the regulation of transcription. In Genes & Development 25 (10), pp. 1010–1022. DOI: 10.1101/gad.2037511.

DeNizio, Jamie E.; Dow, Blaine J.; Serrano, Juan C.; Ghanty, Uday; Drohat, Alexander C.; Kohli, Rahul M. (2021): TET-TDG Active DNA Demethylation at CpG and Non-CpG Sites. In Journal of molecular biology 433 (8), p. 166877. DOI: 10.1016/j.jmb.2021.166877.

Dey, Aninda Sundar; Bhattacharya, Chayan; Guan, Yihong; Jha, Babal Kant; Mukherji, Mridul (2021): Demethylation of Non-CpG Sites in DNA Is Initiated by TET2 5-Methylcytosine Dioxygenase. In DNA 1 (1), pp. 26–36. DOI: 10.3390/dna1010004.

DiNardo, Courtney D.; Cortes, Jorge E. (2016): Mutations in AML: prognostic and therapeutic implications. In Hematology. American Society of Hematology. Education Program 2016 (1), pp. 348–355. DOI: 10.1182/asheducation-2016.1.348.

Dodge, Jonathan E.; Okano, Masaki; Dick, Fred; Tsujimoto, Naomi; Chen, Taiping; Wang, Shumei et al. (2005): Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. In The Journal of biological chemistry 280 (18), pp. 17986–17991. DOI: 10.1074/jbc.M413246200.

Du, Qian; Luu, Phuc-Loi; Stirzaker, Clare; Clark, Susan J. (2015): Methyl-CpG-binding domain proteins: readers of the epigenome. In Epigenomics 7 (6), pp. 1051–1073. DOI: 10.2217/epi.15.39.

Dukatz, Michael; Adam, Sabrina; Biswal, Mahamaya; Song, Jikui; Bashtrykov, Pavel; Jeltsch, Albert (2020): Complex DNA sequence readout mechanisms of the DNMT3B DNA methyltransferase. In Nucleic acids research 48 (20), pp. 11495–11509. DOI: 10.1093/nar/gkaa938.

Easwaran, Hariharan P.; Schermelleh, Lothar; Leonhardt, Heinrich; Cardoso, M. Cristina (2004): Replication-independent chromatin loading of Dnmt1 during G2 and M phases. In EMBO Rep 5 (12), pp. 1181–1186. DOI: 10.1038/sj.embor.7400295.

Egger, Gerda; Jeong, Shinwu; Escobar, Sonia G.; Cortez, Connie C.; Li, Tony W. H.; Saito, Yoshimasa et al. (2006): Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. In Proceedings of the National Academy of Sciences of the United States of America 103 (38), pp. 14080–14085. DOI: 10.1073/pnas.0604602103.

Ehrlich, Melanie; Sanchez, Cecilia; Shao, Chunbo; Nishiyama, Rie; Kehrl, John; Kuick, Rork et al. (2008): ICF, an immunodeficiency syndrome: DNA methyltransferase 3B involvement, chromosome anomalies, and gene dysregulation. In Autoimmunity 41 (4), pp. 253–271. DOI: 10.1080/08916930802024202.

Emperle, Max; Adam, Sabrina; Kunert, Stefan; Dukatz, Michael; Baude, Annika; Plass, Christoph et al. (2019): Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns. In Nucleic acids research 47 (21), pp. 11355–11367. DOI: 10.1093/nar/gkz911.

Emperle, Max; Bangalore, Disha M.; Adam, Sabrina; Kunert, Stefan; Heil, Hannah S.; Heinze, Katrin G. et al. (2021): Structural and biochemical insight into the mechanism of dual CpG site binding and methylation by the DNMT3A DNA methyltransferase. In Nucleic acids research 49 (14), pp. 8294–8308. DOI: 10.1093/nar/gkab600.

Emperle, Max; Dukatz, Michael; Kunert, Stefan; Holzer, Katharina; Rajavelu, Arumugam; Jurkowska, Renata Z.; Jeltsch, Albert (2018a): The DNMT3A R882H mutation does not cause dominant negative effects in purified mixed DNMT3A/R882H complexes. In Scientific reports 8 (1), p. 13242. DOI: 10.1038/s41598-018-31635-8.

Emperle, Max; Rajavelu, Arumugam; Kunert, Stefan; Arimondo, Paola B.; Reinhardt, Richard; Jurkowska, Renata Z.; Jeltsch, Albert (2018b): The DNMT3A R882H mutant displays altered flanking sequence preferences. In Nucleic acids research 46 (6), pp. 3130–3139. DOI: 10.1093/nar/gky168.

Emperle, Max; Rajavelu, Arumugam; Reinhardt, Richard; Jurkowska, Renata Z.; Jeltsch, Albert (2014): Cooperative DNA binding and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase. In The Journal of biological chemistry 289 (43), pp. 29602–29613. DOI: 10.1074/jbc.M114.572032.

Engler, Carola; Kandzia, Romy; Marillonnet, Sylvestre (2008): A one pot, one step, precision cloning method with high throughput capability. In PloS one 3 (11), e3647. DOI: 10.1371/journal.pone.0003647.

Fan, Guoping; Beard, Caroline; Chen, Richard Z.; Csankovszki, Györgyi; Sun, Yi; Siniaia, Marina et al. (2001): DNA Hypomethylation Perturbs the Function and Survival of CNS Neurons in Postnatal Animals. In J. Neurosci. 21 (3), pp. 788–797. DOI: 10.1523/JNEUROSCI.21-03-00788.2001.

Fang, Jian; Cheng, Jingdong; Wang, Jiaolong; Zhang, Qiao; Liu, Mengjie; Gong, Rui et al. (2016): Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition. In Nature communications 7 (1), p. 11197. DOI: 10.1038/ncomms11197.

Fatemi, M.; Hermann, A.; Pradhan, S.; Jeltsch, A. (2001): The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. In Journal of molecular biology 309 (5), pp. 1189–1199. DOI: 10.1006/jmbi.2001.4709.

Fatemi, Mehrnaz; Hermann, Andrea; Gowher, Humaira; Jeltsch, Albert (2002): Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA. In European journal of biochemistry 269 (20), pp. 4981–4984. DOI: 10.1046/j.1432-1033.2002.03198.x.

Fatemi, Mehrnaz; Wade, Paul A. (2006): MBD family proteins: reading the epigenetic code. In Journal of cell science 119 (Pt 15), pp. 3033–3037. DOI: 10.1242/jcs.03099.

Feinberg, A. P.; Vogelstein, B. (1983): Hypomethylation distinguishes genes of some human cancers from their normal counterparts. In Nature 301 (5895), pp. 89–92. DOI: 10.1038/301089a0.

Feltus, F. A.; Lee, E. K.; Costello, J. F.; Plass, C.; Vertino, P. M. (2003): Predicting aberrant CpG island methylation. In Proceedings of the National Academy of Sciences of the United States of America 100 (21), pp. 12253–12258. DOI: 10.1073/pnas.2037852100.

Ferrara, Felicetto; Schiffer, Charles A. (2013): Acute myeloid leukaemia in adults. In The Lancet 381 (9865), pp. 484–495. DOI: 10.1016/S0140-6736(12)61727-9.

Ficz, Gabriella; Branco, Miguel R.; Seisenberger, Stefanie; Santos, Fátima; Krueger, Felix; Hore, Timothy A. et al. (2011): Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. In Nature 473 (7347), pp. 398–402. DOI: 10.1038/nature10008.

Ficz, Gabriella; Gribben, John G. (2014): Loss of 5-hydroxymethylcytosine in cancer: cause or consequence? In Genomics 104 (5), pp. 352–357. DOI: 10.1016/j.ygeno.2014.08.017.

Flynn, J.; Azzam, R.; Reich, N. (1998): DNA binding discrimination of the murine DNA cytosine-C5 methyltransferase. In Journal of molecular biology 279 (1), pp. 101–116. DOI: 10.1006/jmbi.1998.1761.

Fu, Lijuan; Guerrero, Candace R.; Zhong, Na; Amato, Nicholas J.; Liu, Yunhua; Liu, Shuo et al. (2014): Tet-mediated formation of 5-hydroxymethylcytosine in RNA. In Journal of the American Chemical Society 136 (33), pp. 11582–11585. DOI: 10.1021/ja505305z.

Gagliardi, Miriam; Strazzullo, Maria; Matarazzo, Maria R. (2018): DNMT3B Functions: Novel Insights From Human Disease. In Frontiers in cell and developmental biology 6, p. 140. DOI: 10.3389/fcell.2018.00140.

Gao, Linfeng; Anteneh, Hiwot; Song, Jikui (2020a): Dissect the DNMT3A- and DNMT3B-mediated DNA Co-methylation through a Covalent Complex Approach. In Journal of molecular biology 432 (2), pp. 569–575. DOI: 10.1016/j.jmb.2019.11.004.

Gao, Linfeng; Emperle, Max; Guo, Yiran; Grimm, Sara A.; Ren, Wendan; Adam, Sabrina et al. (2020b): Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. In Nat Commun 11 (1), p. 3355. DOI: 10.1038/s41467-020-17109-4.

Gao, Qing; Steine, Eveline J.; Barrasa, M. Inmaculada; Hockemeyer, Dirk; Pawlak, Mathias; Fu, Dongdong et al. (2011): Deletion of the de novo DNA methyltransferase Dnmt3a promotes lung tumor progression. In Proceedings of the National Academy of Sciences of the United States of America 108 (44), pp. 18061–18066. DOI: 10.1073/pnas.1114946108.

Gao, Yawei; Chen, Jiayu; Li, Ke; Wu, Tong; Huang, Bo; Liu, Wenqiang et al. (2013): Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. In Cell stem cell 12 (4), pp. 453–469. DOI: 10.1016/j.stem.2013.02.005.

Gardiner-Garden, M.; Frommer, M. (1987): CpG Islands in vertebrate genomes. In Journal of molecular biology 196 (2), pp. 261–282. DOI: 10.1016/0022-2836(87)90689-9.

Garvie, Colin W.; Wolberger, Cynthia (2001): Recognition of Specific DNA Sequences. In Molecular cell 8 (5), pp. 937–946. DOI: 10.1016/S1097-2765(01)00392-6.

Gibson, Daniel G.; Young, Lei; Chuang, Ray-Yuan; Venter, J. Craig; Hutchison, Clyde A.; Smith, Hamilton O. (2009): Enzymatic assembly of DNA molecules up to several hundred kilobases. In Nature methods 6 (5), pp. 343–345. DOI: 10.1038/nmeth.1318.

Glass, Jacob L.; Hassane, Duane; Wouters, Bas J.; Kunimoto, Hiroyoshi; Avellino, Roberto; Garrett-Bakelman, Francine E. 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. In Cancer discovery 7 (8), pp. 868–883. DOI: 10.1158/2159-8290.CD-16-1032.

Globisch, Daniel; Münzel, Martin; Müller, Markus; Michalakis, Stylianos; Wagner, Mirko; Koch, Susanne et al. (2010): Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. In PloS one 5 (12), e15367. DOI: 10.1371/journal.pone.0015367.

Gowher, H.; 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-CpG correction of non-CpA sites. In Journal of molecular biology 309 (5), pp. 1201–1208. DOI: 10.1006/jmbi.2001.4710.

Gowher, Humaira; Jeltsch, Albert (2002): Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. In The Journal of biological chemistry 277 (23), pp. 20409–20414. DOI: 10.1074/jbc.M202148200.

Gowher, Humaira; Jeltsch, Albert (2018): Mammalian DNA methyltransferases: new discoveries and open questions. In Biochemical Society transactions 46 (5), pp. 1191–1202. DOI: 10.1042/BST20170574.

Gowher, Humaira; Loutchanwoot, Panida; Vorobjeva, Olga; Handa, Vikas; Jurkowska, Renata Z.; Jurkowski, Tomasz P.; Jeltsch, Albert (2006): Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. In Journal of molecular biology 357 (3), pp. 928–941. DOI: 10.1016/j.jmb.2006.01.035.

Goyal, Rachna; Reinhardt, Richard; Jeltsch, Albert (2006): Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. In Nucleic acids research 34 (4), pp. 1182–1188. DOI: 10.1093/nar/gkl002.

Guo, Fan; Li, Xianlong; Liang, Dan; Li, Tong; Zhu, Ping; Guo, Hongshan et al. (2014a): Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. In Cell stem cell 15 (4), pp. 447–459. DOI: 10.1016/j.stem.2014.08.003.

Guo, Hongshan; Zhu, Ping; Yan, Liying; Li, Rong; Hu, Boqiang; Lian, Ying et al. (2014b): The DNA methylation landscape of human early embryos. In Nature 511 (7511), pp. 606–610. DOI: 10.1038/nature13544.

Guo, Junjie U.; Su, Yijing; Zhong, Chun; Ming, Guo-li; Song, Hongjun (2011): Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. In Cell 145 (3), pp. 423–434. DOI: 10.1016/j.cell.2011.03.022.

Hackett, Jamie A.; Sengupta, Roopsha; Zylicz, Jan J.; Murakami, Kazuhiro; Lee, Caroline; Down, Thomas A.; Surani, M. Azim (2013): Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. In Science (New York, N.Y.) 339 (6118), pp. 448–452. DOI: 10.1126/science.1229277.

Haggerty, Chuck; Kretzmer, Helene; Riemenschneider, Christina; Kumar, Abhishek Sampath; Mattei, Alexandra L.; Bailly, Nina et al. (2021): Dnmt1 has de novo activity targeted to transposable elements. In Nature structural & molecular biology 28 (7), pp. 594–603. DOI: 10.1038/s41594-021-00603-8.

Halford, Stephen E.; Marko, John F. (2004): How do site-specific DNA-binding proteins find their targets? In Nucleic acids research 32 (10), pp. 3040–3052. DOI: 10.1093/nar/gkh624.

Halford, Stephen E.; Szczelkun, Mark D. (2002): How to get from A to B: strategies for analysing protein motion on DNA. In European biophysics journal : EBJ 31 (4), pp. 257–267. DOI: 10.1007/s00249-002-0224-4.

Handa, Vikas; Jeltsch, Albert (2005): Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. In Journal of molecular biology 348 (5), pp. 1103–1112. DOI: 10.1016/j.jmb.2005.02.044.

Hascher, Antje; Haase, Ann-Kristin; Hebestreit, Katja; Rohde, Christian; Klein, Hans-Ulrich; Rius, Maria et al. (2014): DNA methyltransferase inhibition reverses epigenetically embedded phenotypes in lung cancer preferentially affecting polycomb target genes. In Clinical cancer research : an official journal of the American Association for Cancer Research 20 (4), pp. 814–826. DOI: 10.1158/1078-0432.CCR-13-1483.

Hashimoto, Hideharu; Liu, Yiwei; Upadhyay, Anup K.; Chang, Yanqi; Howerton, Shelley B.; Vertino, Paula M. et al. (2012): Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. In Nucleic acids research 40 (11), pp. 4841–4849. DOI: 10.1093/nar/gks155.

Hashimoto, Hideharu; Pais, June E.; Dai, Nan; Corrêa, Ivan R.; Zhang, Xing; Zheng, Yu; Cheng, Xiaodong (2015): Structure of Naegleria Tet-like dioxygenase (NgTet1) in complexes with a reaction intermediate 5-hydroxymethylcytosine DNA. In Nucleic acids research 43 (22), pp. 10713–10721. DOI: 10.1093/nar/gkv870.

Hashimoto, Hideharu; Pais, June E.; Zhang, Xing; Saleh, Lana; Fu, Zheng-Qing; Dai, Nan et al. (2014): Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. In Nature 506 (7488), pp. 391–395. DOI: 10.1038/nature12905.

He, Chongsheng; Bozler, Julianna; Janssen, Kevin A.; Wilusz, Jeremy E.; Garcia, Benjamin A.; Schorn, Andrea J.; Bonasio, Roberto (2021): TET2 chemically modifies tRNAs and regulates tRNA fragment levels. In Nature structural & molecular biology 28 (1), pp. 62–70. DOI: 10.1038/s41594-020-00526-w.

He, Yu-Fei; Li, Bin-Zhong; Li, Zheng; Liu, Peng; Wang, Yang; Tang, Qingyu et al. (2011): Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. In Science (New York, N.Y.) 333 (6047), pp. 1303–1307. DOI: 10.1126/science.1210944.

He, Yupeng; Ecker, Joseph R. (2015): Non-CG Methylation in the Human Genome. In Annual review of genomics and human genetics 16, pp. 55–77. DOI: 10.1146/annurev-genom-090413-025437.

Hermann, Andrea; Goyal, Rachna; Jeltsch, Albert (2004): The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. In The Journal of biological chemistry 279 (46), pp. 48350–48359. DOI: 10.1074/jbc.M403427200.

Hess, J. F.; Kohl, T. A.; Kotrová, M.; Rönsch, K.; Paprotka, T.; Mohr, V. et al. (2020): Library preparation for Next-Generation sequencing: A review of automation

strategies. In Biotechnology advances 41, p. 107537. DOI: 10.1016/j.biotechadv.2020.107537.

Hoffart, Lee M.; Barr, Eric W.; Guyer, Robert B.; Bollinger, J. Martin; Krebs, Carsten (2006): Direct spectroscopic detection of a C-H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. In Proceedings of the National Academy of Sciences of the United States of America 103 (40), pp. 14738–14743. DOI: 10.1073/pnas.0604005103.

Holliday, R.; Pugh, J. E. (1975): DNA modification mechanisms and gene activity during development. In Science (New York, N.Y.) 187 (4173), pp. 226–232.

Holz-Schietinger, Celeste; Matje, Doug M.; Reich, Norbert O. (2012): Mutations in DNA methyltransferase (DNMT3A) observed in acute myeloid leukemia patients disrupt processive methylation. In The Journal of biological chemistry 287 (37), pp. 30941–30951. DOI: 10.1074/jbc.M112.366625.

Holz-Schietinger, Celeste; Reich, Norbert O. (2010): The inherent processivity of the human de novo methyltransferase 3A (DNMT3A) is enhanced by DNMT3L. In The Journal of biological chemistry 285 (38), pp. 29091–29100. DOI: 10.1074/jbc.M110.142513.

Holz-Schietinger, Celeste; Reich, Norbert O. (2015): De novo DNA methyltransferase DNMT3A: Regulation of oligomeric state and mechanism of action in response to pH changes. In Biochimica et biophysica acta 1850 (6), pp. 1131–1139. DOI: 10.1016/j.bbagen.2015.02.003.

Hu, Lulu; Li, Ze; Cheng, Jingdong; Rao, Qinhui; Gong, Wei; Liu, Mengjie et al. (2013): Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. In Cell 155 (7), pp. 1545–1555. DOI: 10.1016/j.cell.2013.11.020.

Hu, Lulu; Lu, Junyan; Cheng, Jingdong; Rao, Qinhui; Li, Ze; Hou, Haifeng et al. (2015): Structural insight into substrate preference for TET-mediated oxidation. In Nature 527 (7576), pp. 118–122. DOI: 10.1038/nature15713.

Iida, Tetsuo; Suetake, Isao; Tajima, Shoji; Morioka, Hiroshi; Ohta, Satoshi; Obuse, Chikashi; Tsurimoto, Toshiki (2002): PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA. In Genes to cells : devoted to molecular & cellular mechanisms 7 (10), pp. 997–1007. DOI: 10.1046/j.1365-2443.2002.00584.x.

Ito, Kyoko; Lee, Joun; Chrysanthou, Stephanie; Zhao, Yilin; Josephs, Katherine; Sato, Hiroyo et al. (2019): Non-catalytic Roles of Tet2 Are Essential to Regulate Hematopoietic Stem and Progenitor Cell Homeostasis. In Cell reports 28 (10), 2480-2490.e4. DOI: 10.1016/j.celrep.2019.07.094.

Ito, Shinsuke; D'Alessio, Ana C.; Taranova, Olena V.; Hong, Kwonho; Sowers, Lawrence C.; Zhang, Yi (2010): Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. In Nature 466 (7310), pp. 1129–1133. DOI: 10.1038/nature09303.

Ito, Shinsuke; Shen, Li; Dai, Qing; Wu, Susan C.; Collins, Leonard B.; Swenberg, James A. et al. (2011): Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. In Science (New York, N.Y.) 333 (6047), pp. 1300–1303. DOI: 10.1126/science.1210597.

Iyer, Lakshminarayan M.; Tahiliani, Mamta; Rao, Anjana; Aravind, L. (2009): Prediction of novel families of enzymes involved in oxidative and other complex modifications of

bases in nucleic acids. In Cell cycle (Georgetown, Tex.) 8 (11), pp. 1698–1710. DOI: 10.4161/cc.8.11.8580.

Jang, Hyun Sik; Shin, Woo Jung; Lee, Jeong Eon; Do, Jeong Tae (2017): CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. In Genes 8 (6). DOI: 10.3390/genes8060148.

Jeltsch, Albert (2006): On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. In Epigenetics 1 (2), pp. 63–66. DOI: 10.4161/epi.1.2.2767.

Jeltsch, Albert; Adam, Sabrina; Dukatz, Michael; Emperle, Max; Bashtrykov, Pavel (2021): Deep Enzymology Studies on DNA Methyltransferases Reveal Novel Connections between Flanking Sequences and Enzyme Activity. In Journal of molecular biology 433 (19), p. 167186. DOI: 10.1016/j.jmb.2021.167186.

Jeltsch, Albert; Broche, Julian; Bashtrykov, Pavel (2018): Molecular Processes Connecting DNA Methylation Patterns with DNA Methyltransferases and Histone Modifications in Mammalian Genomes. In Genes 9 (11). DOI: 10.3390/genes9110566.

Jeltsch, Albert; Jurkowska, Renata Z. (2014): New concepts in DNA methylation. In Trends in biochemical sciences 39 (7), pp. 310–318. DOI: 10.1016/j.tibs.2014.05.002.

Jeltsch, Albert; Jurkowska, Renata Z. (2016): Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. In Nucleic acids research 44 (18), pp. 8556–8575. DOI: 10.1093/nar/gkw723.

Jeltsch, Albert; Lanio, Thomas (2002): Site-directed mutagenesis by polymerase chain reaction. In Methods in molecular biology (Clifton, N.J.) 182, pp. 85–94. DOI: 10.1385/1-59259-194-9:085.

Ji, Debin; Lin, Krystal; Song, Jikui; Wang, Yinsheng (2014): Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation. In Molecular bioSystems 10 (7), pp. 1749–1752. DOI: 10.1039/c4mb00150h.

Jia, Da; Jurkowska, Renata Z.; Zhang, Xing; Jeltsch, Albert; Cheng, Xiaodong (2007): Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. In Nature 449 (7159), pp. 248–251. DOI: 10.1038/nature06146.

Jiang, Shuai (2020): Tet2 at the interface between cancer and immunity. In Communications biology 3 (1), p. 667. DOI: 10.1038/s42003-020-01391-5.

Jimenji, Tomohiro; Matsumura, Rumie; Kori, Satomi; Arita, Kyohei (2019): Structure of PCNA in complex with DNMT1 PIP box reveals the basis for the molecular mechanism of the interaction. In Biochemical and biophysical research communications 516 (2), pp. 578–583. DOI: 10.1016/j.bbrc.2019.06.060.

Jin, Zelin; Liu, Yun (2018): DNA methylation in human diseases. In Genes & diseases 5 (1), pp. 1–8. DOI: 10.1016/j.gendis.2018.01.002.

Jurkowska, Renata Z.; Anspach, Nils; Urbanke, Claus; Da Jia; Reinhardt, Richard; Nellen, Wolfgang et al. (2008): Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. In Nucleic acids research 36 (21), pp. 6656–6663. DOI: 10.1093/nar/gkn747.

Jurkowska, Renata Z.; Ceccaldi, Alexandre; Zhang, Yingying; Arimondo, Paola B.; Jeltsch, Albert (2011a): DNA methyltransferase assays. In Methods in molecular biology (Clifton, N.J.) 791, pp. 157–177. DOI: 10.1007/978-1-61779-316-5_13.

Jurkowska, Renata Z.; Rajavelu, Arumugam; Anspach, Nils; Urbanke, Claus; Jankevicius, Gytis; Ragozin, Sergey et al. (2011b): Oligomerization and binding of the Dnmt3a DNA methyltransferase to parallel DNA molecules: heterochromatic localization and role of Dnmt3L. In The Journal of biological chemistry 286 (27), pp. 24200–24207. DOI: 10.1074/jbc.M111.254987.

Jurkowska, Renata Z.; Siddique, Abu Nasar; Jurkowski, Tomasz P.; Jeltsch, Albert (2011c): Approaches to enzyme and substrate design of the murine Dnmt3a DNA methyltransferase. In Chembiochem : a European journal of chemical biology 12 (10), pp. 1589–1594. DOI: 10.1002/cbic.201000673.

Jurkowska, Renata Zofia; Jurkowski, Tomasz Piotr; Jeltsch, Albert (2011d): Structure and function of mammalian DNA methyltransferases. In Chembiochem : a European journal of chemical biology 12 (2), pp. 206–222. DOI: 10.1002/cbic.201000195.

Jurkowski, Tomasz P.; Jeltsch, Albert (2011): On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. In PloS one 6 (11), e28104. DOI: 10.1371/journal.pone.0028104.

Kaneda, Masahiro; Okano, Masaki; Hata, Kenichiro; Sado, Takashi; Tsujimoto, Naomi; Li, En; Sasaki, Hiroyuki (2004): Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. In Nature 429 (6994), pp. 900–903. DOI: 10.1038/nature02633.

Kato, Yuzuru; Kaneda, Masahiro; Hata, Kenichiro; Kumaki, Kenji; Hisano, Mizue; Kohara, Yuji et al. (2007): Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. In Human molecular genetics 16 (19), pp. 2272–2280. DOI: 10.1093/hmg/ddm179.

Kim, Mirang; Costello, Joseph (2017): DNA methylation: an epigenetic mark of cellular memory. In Experimental & molecular medicine 49 (4), e322. DOI: 10.1038/emm.2017.10.

Kim, Soo Jin; Zhao, Hongbo; Hardikar, Swanand; Singh, Anup Kumar; Goodell, Margaret A.; Chen, Taiping (2013): A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. In Blood 122 (25), pp. 4086–4089. DOI: 10.1182/blood-2013-02-483487.

Kinde, Benyam; Gabel, Harrison W.; Gilbert, Caitlin S.; Griffith, Eric C.; Greenberg, Michael E. (2015): Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. In Proceedings of the National Academy of Sciences of the United States of America 112 (22), pp. 6800–6806. DOI: 10.1073/pnas.1411269112.

Kishikawa, Shotaro; Murata, Takehide; Ugai, Hideyo; Yamazaki, Takahito; Yokoyama, Kazunari K. (2003): Control elements of Dnmt1 gene are regulated in cell-cycle dependent manner. In Nucleic acids research. Supplement (2001) (3), pp. 307–308. DOI: 10.1093/nass/3.1.307.

Kizaki, Seiichiro; Chandran, Anandhakumar; Sugiyama, Hiroshi (2016): Identification of Sequence Specificity of 5-Methylcytosine Oxidation by Tet1 Protein with High-

Throughput Sequencing. In Chembiochem : a European journal of chemical biology 17 (5), pp. 403–406. DOI: 10.1002/cbic.201500646.

Klimasauskas, Saulius; Kumar, Sanjay; Roberts, Richard J.; Cheng, Xiaodong (1994): Hhal methyltransferase flips its target base out of the DNA helix. In Cell 76 (2), pp. 357–369. DOI: 10.1016/0092-8674(94)90342-5.

Ko, Myunggon; Huang, Yun; Jankowska, Anna M.; Pape, Utz J.; Tahiliani, Mamta; Bandukwala, Hozefa S. et al. (2010): Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. In Nature 468 (7325), pp. 839–843. DOI: 10.1038/nature09586.

Kothari, R. M.; Shankar, V. (1976): 5-Methylcytosine content in the vertebrate deoxyribonucleic acids: species specificity. In Journal of molecular evolution 7 (4), pp. 325–329. DOI: 10.1007/BF01743628.

Krebs, Carsten; Galonić Fujimori, Danica; Walsh, Christopher T.; Bollinger, J. Martin (2007): Non-heme Fe(IV)-oxo intermediates. In Accounts of chemical research 40 (7), pp. 484–492. DOI: 10.1021/ar700066p.

Kriaucionis, Skirmantas; Heintz, Nathaniel (2009): The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. In Science (New York, N.Y.) 324 (5929), pp. 929–930. DOI: 10.1126/science.1169786.

Ku, Chee Seng; Naidoo, Nasheen; Wu, Mengchu; Soong, Richie (2011): Studying the epigenome using Next-Generation sequencing. In Journal of medical genetics 48 (11), pp. 721–730. DOI: 10.1136/jmedgenet-2011-100242.

Kumar, S.; Cheng, X.; Klimasauskas, S.; Mi, S.; Posfai, J.; Roberts, R. J.; Wilson, G. G. (1994): The DNA (cytosine-5) methyltransferases. In Nucleic acids research 22 (1), pp. 1–10. DOI: 10.1093/nar/22.1.1.

Laird, Charles D.; Pleasant, Nicole D.; Clark, Aaron D.; Sneeden, Jessica L.; Hassan, K. M. Anwarul; Manley, Nathan C. et al. (2004): Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. In Proceedings of the National Academy of Sciences of the United States of America 101 (1), pp. 204–209. DOI: 10.1073/pnas.2536758100.

Laisné, Marthe; Gupta, Nikhil; Kirsh, Olivier; Pradhan, Sriharsa; Defossez, Pierre-Antoine (2018): Mechanisms of DNA Methyltransferase Recruitment in Mammals. In Genes 9 (12), p. 617. DOI: 10.3390/genes9120617.

Langemeijer, Saskia M. C.; Aslanyan, Mariam G.; Jansen, Joop H. (2009): TET proteins in malignant hematopoiesis. In Cell cycle (Georgetown, Tex.) 8 (24), pp. 4044–4048. DOI: 10.4161/cc.8.24.10239.

Laszlo, Andrew H.; Derrington, Ian M.; Brinkerhoff, Henry; Langford, Kyle W.; Nova, Ian C.; Samson, Jenny Mae et al. (2013): Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. In Proceedings of the National Academy of Sciences of the United States of America 110 (47), pp. 18904–18909. DOI: 10.1073/pnas.1310240110.

Laurent, Louise; Wong, Eleanor; Li, Guoliang; Huynh, Tien; Tsirigos, Aristotelis; Ong, Chin Thing et al. (2010): Dynamic changes in the human methylome during differentiation. In Genome research 20 (3), pp. 320–331. DOI: 10.1101/gr.101907.109.

Lee, Jong-Hun; Park, Sung-Joon; Nakai, Kenta (2017): Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s. In Scientific reports 7 (1), p. 11295. DOI: 10.1038/s41598-017-11800-1.

Lee, P. J.; Washer, L. L.; Law, D. J.; Boland, C. R.; Horon, I. L.; Feinberg, A. P. (1996): Limited up-regulation of DNA methyltransferase in human colon cancer reflecting increased cell proliferation. In Proceedings of the National Academy of Sciences of the United States of America 93 (19), pp. 10366–10370. DOI: 10.1073/pnas.93.19.10366.

Ley, Timothy J.; Ding, Li; Walter, Matthew J.; McLellan, Michael D.; Lamprecht, Tamara; Larson, David E. et al. (2010): DNMT3A mutations in acute myeloid leukemia. In The New England journal of medicine 363 (25), pp. 2424–2433. DOI: 10.1056/NEJMoa1005143.

Li, En; Bestor, Timothy H.; Jaenisch, Rudolf (1992): Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. In Cell 69 (6), pp. 915–926. DOI: 10.1016/0092-8674(92)90611-F.

Li, Sheng; Garrett-Bakelman, Francine E.; Chung, Stephen S.; Sanders, Mathijs A.; Hricik, Todd; Rapaport, Franck et al. (2016a): Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. In Nature medicine 22 (7), pp. 792–799. DOI: 10.1038/nm.4125.

Li, Tao; Wang, Linsheng; Du, Yongming; Xie, Si; Yang, Xi; Lian, Fuming et al. (2018): Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation. In Nucleic acids research 46 (6), pp. 3218–3231. DOI: 10.1093/nar/gky104.

Li, Xin; Liu, Yun; Salz, Tal; Hansen, Kasper D.; Feinberg, Andrew (2016b): Whole-genome analysis of the methylome and hydroxymethylome in normal and malignant lung and liver. In Genome research 26 (12), pp. 1730–1741. DOI: 10.1101/gr.211854.116.

Li, Zhiguang; Dai, Hongzheng; Martos, Suzanne N.; Xu, Beisi; Gao, Yang; Li, Teng et al. (2015): Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells. In Genome biology 16, p. 115. DOI: 10.1186/s13059-015-0685-2.

Liao, Jing; Karnik, Rahul; Gu, Hongcang; Ziller, Michael J.; Clement, Kendell; Tsankov, Alexander M. et al. (2015): Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. In Nature genetics 47 (5), pp. 469–478. DOI: 10.1038/ng.3258.

Liebert, Kirsten; Jeltsch, Albert (2008): Detection and quantitation of the activity of DNA methyltransferases using a biotin/avidin microplate assay. In Methods in molecular biology (Clifton, N.J.) 418, pp. 149–156. DOI: 10.1007/978-1-59745-579-4_13.

Lin, Chien-Chu; Chen, Yi-Ping; Yang, Wei-Zen; Shen, James C. K.; Yuan, Hanna S. (2020): Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. In Nucleic acids research 48 (7), pp. 3949–3961. DOI: 10.1093/nar/gkaa111.

Lin, Iping G.; Han, Li; Taghva, Alexander; O'Brien, Laura E.; Hsieh, Chih-Lin (2002): Murine de novo methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA in vitro. In Molecular and cellular biology 22 (3), pp. 704–723. DOI: 10.1128/MCB.22.3.704-723.2002.

Lister, Ryan; Ecker, Joseph R. (2009): Finding the fifth base: genome-wide sequencing of cytosine methylation. In Genome research 19 (6), pp. 959–966. DOI: 10.1101/gr.083451.108.

Lister, Ryan; Mukamel, Eran A.; Nery, Joseph R.; Urich, Mark; Puddifoot, Clare A.; Johnson, Nicholas D. et al. (2013): Global epigenomic reconfiguration during mammalian brain development. In Science (New York, N.Y.) 341 (6146), p. 1237905. DOI: 10.1126/science.1237905.

Lister, Ryan; Pelizzola, Mattia; Dowen, Robert H.; Hawkins, R. David; Hon, Gary; Tonti-Filippini, Julian et al. (2009): Human DNA methylomes at base resolution show widespread epigenomic differences. In Nature 462 (7271), pp. 315–322. DOI: 10.1038/nature08514.

Liu, Chen-Kao; Hsu, Chin-An; Abbott, Mitchel T. (1973): Catalysis of three sequential dioxygenase reactions by thymine 7-hydroxylase. In Archives of Biochemistry and Biophysics 159 (1), pp. 180–187. DOI: 10.1016/0003-9861(73)90443-8.

Liu, M. Y.; DeNizio, J. E.; Kohli, R. M. (2016): Quantification of Oxidized 5-Methylcytosine Bases and TET Enzyme Activity. In Methods in enzymology 573, pp. 365–385. DOI: 10.1016/bs.mie.2015.12.006.

Lu, Xingyu; Han, Dali; Zhao, Boxuan Simen; Song, Chun-Xiao; Zhang, Li-Sheng; Doré, Louis C.; He, Chuan (2015): Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. In Cell research 25 (3), pp. 386–389. DOI: 10.1038/cr.2015.5.

Lyko, Frank (2018): The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. In Nature reviews. Genetics 19 (2), pp. 81–92. DOI: 10.1038/nrg.2017.80.

Mack, Alexandra; Emperle, Max; Schnee, Philipp; Adam, Sabrina; Pleiss, Jürgen; Bashtrykov, Pavel; Jeltsch, Albert (2022): Preferential Self-interaction of DNA Methyltransferase DNMT3A Subunits Containing the R882H Cancer Mutation Leads to Dominant Changes of Flanking Sequence Preferences. In Journal of molecular biology 434 (7), p. 167482. DOI: 10.1016/j.jmb.2022.167482.

Mallona, Izaskun; Ilie, Ioana Mariuca; Karemaker, Ino Dominiek; Butz, Stefan; Manzo, Massimiliano; Caflisch, Amedeo; Baubec, Tuncay (2021): Flanking sequence preference modulates de novo DNA methylation in the mouse genome. In Nucleic acids research 49 (1), pp. 145–157. DOI: 10.1093/nar/gkaa1168.

Mao, Shi-Qing; Cuesta, Sergio Martínez; Tannahill, David; Balasubramanian, Shankar (2020): Genome-wide DNA Methylation Signatures Are Determined by DNMT3A/B Sequence Preferences. In Biochemistry 59 (27), pp. 2541–2550. DOI: 10.1021/acs.biochem.0c00339.

Marcucci, Guido; Metzeler, Klaus H.; Schwind, Sebastian; Becker, Heiko; Maharry, Kati; Mrózek, Krzysztof et al. (2012): Age-related prognostic impact of different types of DNMT3A mutations in adults with primary cytogenetically normal acute myeloid leukemia. In Journal of clinical oncology : official journal of the American Society of Clinical Oncology 30 (7), pp. 742–750. DOI: 10.1200/JCO.2011.39.2092.

Mattei, Alexandra L.; Bailly, Nina; Meissner, Alexander (2022): DNA methylation: a historical perspective. In Trends in genetics : TIG. DOI: 10.1016/j.tig.2022.03.010.

Mayer, W.; Niveleau, A.; Walter, J.; Fundele, R.; Haaf, T. (2000): Demethylation of the zygotic paternal genome. In Nature 403 (6769), pp. 501–502. DOI: 10.1038/35000656.

Meier, Jochen C.; Kankowski, Svenja; Krestel, Heinz; Hetsch, Florian (2016): RNA Editing-Systemic Relevance and Clue to Disease Mechanisms? In Frontiers in Molecular Neuroscience 9, p. 124. DOI: 10.3389/fnmol.2016.00124.

Melamed, Philippa; Yosefzon, Yahav; David, Cfir; Tsukerman, Anna; Pnueli, Lilach (2018): Tet Enzymes, Variants, and Differential Effects on Function. In Frontiers in cell and developmental biology 6, p. 22. DOI: 10.3389/fcell.2018.00022.

Mellén, Marian; Ayata, Pinar; Heintz, Nathaniel (2017): 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes. In Proceedings of the National Academy of Sciences of the United States of America 114 (37), E7812-E7821. DOI: 10.1073/pnas.1708044114.

Messerschmidt, Daniel M.; Knowles, Barbara B.; Solter, Davor (2014): DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. In Genes & Development 28 (8), pp. 812–828. DOI: 10.1101/gad.234294.113.

Metzker, Michael L. (2010): Sequencing technologies - the Next-Generation. In Nature reviews. Genetics 11 (1), pp. 31–46. DOI: 10.1038/nrg2626.

Minor, Emily A.; Court, Brenda L.; Young, Juan I.; Wang, Gaofeng (2013): Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. In The Journal of biological chemistry 288 (19), pp. 13669–13674. DOI: 10.1074/jbc.C113.464800.

Moris, Naomi; Pina, Cristina; Arias, Alfonso Martinez (2016): Transition states and cell fate decisions in epigenetic landscapes. In Nature reviews. Genetics 17 (11), pp. 693–703. DOI: 10.1038/nrg.2016.98.

Muller, Patricia A. J.; Vousden, Karen H. (2013): p53 mutations in cancer. In Nature cell biology 15 (1), pp. 2–8. DOI: 10.1038/ncb2641.

Nady, Nataliya; Lemak, Alexander; Walker, John R.; Avvakumov, George V.; Kareta, Michael S.; Achour, Mayada et al. (2011): Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. In The Journal of biological chemistry 286 (27), pp. 24300–24311. DOI: 10.1074/jbc.M111.234104.

Neidigh, Jonathan W.; Darwanto, Agus; Williams, Adides A.; Wall, Nathan R.; Sowers, Lawrence C. (2009): Cloning and characterization of Rhodotorula glutinis thymine hydroxylase. In Chemical research in toxicology 22 (5), pp. 885–893. DOI: 10.1021/tx8004482.

Neri, Francesco; Rapelli, Stefania; Krepelova, Anna; Incarnato, Danny; Parlato, Caterina; Basile, Giulia et al. (2017): Intragenic DNA methylation prevents spurious transcription initiation. In Nature 543 (7643), pp. 72–77. DOI: 10.1038/nature21373.

Nguyen, Tuong-Vi; Yao, Shihua; Wang, Yahong; Rolfe, Alan; Selvaraj, Anand; Darman, Rachel et al. (2019): The R882H DNMT3A hot spot mutation stabilizes the formation of large DNMT3A oligomers with low DNA methyltransferase activity. In The Journal of biological chemistry 294 (45), pp. 16966–16977. DOI: 10.1074/jbc.RA119.010126.

Nierzwicki, Łukasz; Palermo, Giulia (2021): Molecular Dynamics to Predict Cryo-EM: Capturing Transitions and Short-Lived Conformational States of Biomolecules. In Frontiers in molecular biosciences 8, p. 641208. DOI: 10.3389/fmolb.2021.641208.

Norvil, Allison B.; Alabdi, Lama; Liu, Bigang; Tu, Yu Han; Forstoffer, Nicole E.; Michie, Amie R. et al. (2020): The acute myeloid leukemia variant DNMT3A Arg882His is a DNMT3B-like enzyme. In Nucleic acids research 48 (7), pp. 3761–3775. DOI: 10.1093/nar/gkaa139.

Norvil, Allison B.; Petell, Christopher J.; Alabdi, Lama; Wu, Lanchen; Rossie, Sandra; Gowher, Humaira (2018): Dnmt3b Methylates DNA by a Noncooperative Mechanism, and Its Activity Is Unaffected by Manipulations at the Predicted Dimer Interface. In Biochemistry 57 (29), pp. 4312–4324. DOI: 10.1021/acs.biochem.6b00964.

Okano, M.; Xie, S.; Li, E. (1998): Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. In Nature genetics 19 (3), pp. 219–220. DOI: 10.1038/890.

Okano, Masaki; Bell, Daphne W.; Haber, Daniel A.; Li, En (1999): DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. In Cell 99 (3), pp. 247–257. DOI: 10.1016/S0092-8674(00)81656-6.

Osterman, D. G.; DePillis, G. D.; Wu, J. C.; Matsuda, A.; Santi, D. V. (1988): 5-Fluorocytosine in DNA is a mechanism-based inhibitor of HhaI methylase. In Biochemistry 27 (14), pp. 5204–5210. DOI: 10.1021/bi00414a039.

Oswald, J.; Engemann, S.; Lane, N.; Mayer, W.; Olek, A.; Fundele, R. et al. (2000): Active demethylation of the paternal genome in the mouse zygote. In Current Biology 10 (8), pp. 475–478. DOI: 10.1016/S0960-9822(00)00448-6.

Otani, Junji; Kimura, Hironobu; Sharif, Jafar; Endo, Takaho A.; Mishima, Yuichi; Kawakami, Toru et al. (2013): Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. In PloS one 8 (12), e82961. DOI: 10.1371/journal.pone.0082961.

Pacaud, Romain; Brocard, Emeline; Lalier, Lisenn; Hervouet, Eric; Vallette, François M.; Cartron, Pierre-François (2014): The DNMT1/PCNA/UHRF1 disruption induces tumorigenesis characterized by similar genetic and epigenetic signatures. In Scientific reports 4, p. 4230. DOI: 10.1038/srep04230.

Pais, June E.; Dai, Nan; Tamanaha, Esta; Vaisvila, Romualdas; Fomenkov, Alexey I.; Bitinaite, Jurate et al. (2015): Biochemical characterization of a Naegleria TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. In Proceedings of the National Academy of Sciences of the United States of America 112 (14), pp. 4316–4321. DOI: 10.1073/pnas.1417939112.

Papaemmanuil, Elli; Gerstung, Moritz; Bullinger, Lars; Gaidzik, Verena I.; Paschka, Peter; Roberts, Nicola D. et al. (2016): Genomic Classification and Prognosis in Acute Myeloid Leukemia. In The New England journal of medicine 374 (23), pp. 2209–2221. DOI: 10.1056/NEJMoa1516192.

Parker, Mackenzie J.; Weigele, Peter R.; Saleh, Lana (2019): Insights into the Biochemistry, Evolution, and Biotechnological Applications of the Ten-Eleven Translocation (TET) Enzymes. In Biochemistry 58 (6), pp. 450–467. DOI: 10.1021/acs.biochem.8b01185.

Parry, Aled; Rulands, Steffen; Reik, Wolf (2021): Active turnover of DNA methylation during cell fate decisions. In Nature reviews. Genetics 22 (1), pp. 59–66. DOI: 10.1038/s41576-020-00287-8.

Pastor, William A.; Aravind, L.; Rao, Anjana (2013): TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. In Nature reviews. Molecular cell biology 14 (6), pp. 341–356. DOI: 10.1038/nrm3589.

Pehrsson, Erica C.; Choudhary, Mayank N. K.; Sundaram, Vasavi; Wang, Ting (2019): The epigenomic landscape of transposable elements across normal human development and anatomy. In Nature communications 10 (1), p. 5640. DOI: 10.1038/s41467-019-13555-x.

Penn, N. W.; Suwalski, R.; O'Riley, C.; Bojanowski, K.; Yura, R. (1972): The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. In The Biochemical journal 126 (4), pp. 781–790. DOI: 10.1042/bj1260781.

Petryk, Nataliya; Bultmann, Sebastian; Bartke, Till; Defossez, Pierre-Antoine (2021): Staying true to yourself: mechanisms of DNA methylation maintenance in mammals. In Nucleic acids research 49 (6), pp. 3020–3032. DOI: 10.1093/nar/gkaa1154.

Piovesan, Allison; Antonaros, Francesca; Vitale, Lorenza; Strippoli, Pierluigi; Pelleri, Maria Chiara; Caracausi, Maria (2019): Human protein-coding genes and gene feature statistics in 2019. In BMC research notes 12 (1), p. 315. DOI: 10.1186/s13104-019-4343-8.

Pon, Julia R.; Marra, Marco A. (2015): Driver and passenger mutations in cancer. In Annual review of pathology 10, pp. 25–50. DOI: 10.1146/annurev-pathol-012414-040312.

Pradhan, S.; Bacolla, A.; Wells, R. D.; Roberts, R. J. (1999): Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. In The Journal of biological chemistry 274 (46), pp. 33002–33010. DOI: 10.1074/jbc.274.46.33002.

Price, John C.; Barr, Eric W.; Glass, Timothy E.; Krebs, Carsten; Bollinger, J. Martin (2003): Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine:alpha-ketoglutarate dioxygenase (TauD). In Journal of the American Chemical Society 125 (43), pp. 13008–13009. DOI: 10.1021/ja037400h.

Qin, Weihua; Wolf, Patricia; Liu, Nan; Link, Stephanie; Smets, Martha; La Mastra, Federica et al. (2015): DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. In Cell Res 25 (8), pp. 911–929. DOI: 10.1038/cr.2015.72.

Qu, Ying; Lennartsson, Andreas; Gaidzik, Verena I.; Deneberg, Stefan; Karimi, Mohsen; Bengtzén, Sofia et al. (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. In Epigenetics 9 (8), pp. 1108–1119. DOI: 10.4161/epi.29315.

Raiber, Eun-Ang; Beraldi, Dario; Ficz, Gabriella; Burgess, Heather E.; Branco, Miguel R.; Murat, Pierre et al. (2012): Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. In Genome biology 13 (8), R69. DOI: 10.1186/gb-2012-13-8-r69.

Rajavelu, Arumugam; Jurkowska, Renata Z.; Fritz, Jürgen; Jeltsch, Albert (2012): Function and disruption of DNA methyltransferase 3a cooperative DNA binding and

nucleoprotein filament formation. In Nucleic acids research 40 (2), pp. 569–580. DOI: 10.1093/nar/gkr753.

Rasmussen, Kasper Dindler; Helin, Kristian (2016): Role of TET enzymes in DNA methylation, development, and cancer. In Genes & Development 30 (7), pp. 733–750. DOI: 10.1101/gad.276568.115.

Rau, Rachel E.; Rodriguez, Benjamin A.; Luo, Min; Jeong, Mira; Rosen, Allison; Rogers, Jason H. et al. (2016): DOT1L as a therapeutic target for the treatment of DNMT3A-mutant acute myeloid leukemia. In Blood 128 (7), pp. 971–981. DOI: 10.1182/blood-2015-11-684225.

Rausch, Cathia; Zhang, Peng; Casas-Delucchi, Corella S.; Daiß, Julia L.; Engel, Christoph; Coster, Gideon et al. (2021): Cytosine base modifications regulate DNA duplex stability and metabolism. In Nucleic acids research 49 (22), pp. 12870–12894. DOI: 10.1093/nar/gkab509.

Ravichandran, M.; Jurkowska, R. Z.; Jurkowski, T. P. (2018): Target specificity of mammalian DNA methylation and demethylation machinery. In Organic & biomolecular chemistry 16 (9), pp. 1419–1435. DOI: 10.1039/c7ob02574b.

Ravichandran, Mirunalini; Rafalski, Dominik; Ortega-Recalde, Oscar; Davies, Claudia I.; Glanfield, Cassandra R.; Kotter, Annika et al. (2021): Pronounced sequence specificity of the TET enzyme catalytic domain guides its cellular function.

Reimer, Michael; Pulakanti, Kirthi; Shi, Linzheng; Abel, Alex; Liang, Mingyu; Malarkannan, Subramaniam; Rao, Sridhar (2019): Deletion of Tet proteins results in quantitative disparities during ESC differentiation partially attributable to alterations in gene expression. In BMC Developmental Biology 19 (1), p. 16. DOI: 10.1186/s12861-019-0196-6.

Reinisch, Karin M.; Chen, Lin; Verdine, Gregory L.; Lipscomb, William N. (1995): The crystal structure of HaeIII methyltransferase covalently complexed to DNA: An extrahelical cytosine and rearranged base pairing. In Cell 82 (1), pp. 143–153. DOI: 10.1016/0092-8674(95)90060-8.

Riggs, A. D. (1975): X inactivation, differentiation, and DNA methylation. In Cytogenetics and cell genetics 14 (1), pp. 9–25. DOI: 10.1159/000130315.

Robertson, K. D.; Ait-Si-Ali, S.; Yokochi, T.; Wade, P. A.; Jones, P. L.; Wolffe, A. P. (2000): DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. In Nature genetics 25 (3), pp. 338–342. DOI: 10.1038/77124.

Robertson, Keith D. (2005): DNA methylation and human disease. In Nature reviews. Genetics 6 (8), pp. 597–610. DOI: 10.1038/nrg1655.

Roth, M.; Jeltsch, A. (2000): Biotin-avidin microplate assay for the quantitative analysis of enzymatic methylation of DNA by DNA methyltransferases. In Biological chemistry 381 (3), pp. 269–272. DOI: 10.1515/BC.2000.035.

Roy, David M.; Walsh, Logan A.; Chan, Timothy A. (2014): Driver mutations of cancer epigenomes. In Protein & cell 5 (4), pp. 265–296. DOI: 10.1007/s13238-014-0031-6.

Russler-Germain, David A.; Spencer, David H.; Young, Margaret A.; Lamprecht, Tamara L.; Miller, Christopher A.; Fulton, Robert et al. (2014): The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its

ability to form active tetramers. In Cancer cell 25 (4), pp. 442–454. DOI: 10.1016/j.ccr.2014.02.010.

Sato, Hanae; Wheat, Justin C.; Steidl, Ulrich; Ito, Keisuke (2016): DNMT3A and TET2 in the Pre-Leukemic Phase of Hematopoietic Disorders. In Frontiers in oncology 6, p. 187. DOI: 10.3389/fonc.2016.00187.

Saxonov, Serge; Berg, Paul; Brutlag, Douglas L. (2006): A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. In Proceedings of the National Academy of Sciences of the United States of America 103 (5), pp. 1412–1417. DOI: 10.1073/pnas.0510310103.

Schneider, Katrin; Fuchs, Christiane; Dobay, Akos; Rottach, Andrea; Qin, Weihua; Wolf, Patricia et al. (2013): Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. In Nucleic acids research 41 (9), pp. 4860–4876. DOI: 10.1093/nar/gkt191.

Schreiber, Jacob; Wescoe, Zachary L.; Abu-Shumays, Robin; Vivian, John T.; Baatar, Baldandorj; Karplus, Kevin; Akeson, Mark (2013): Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. In Proceedings of the National Academy of Sciences of the United States of America 110 (47), pp. 18910–18915. DOI: 10.1073/pnas.1310615110.

Schübeler, Dirk (2015): Function and information content of DNA methylation. In Nature 517 (7534), pp. 321–326. DOI: 10.1038/nature14192.

Seeman, N. C.; Rosenberg, J. M.; Rich, A. (1976): Sequence-specific recognition of double helical nucleic acids by proteins. In Proceedings of the National Academy of Sciences of the United States of America 73 (3), pp. 804–808. DOI: 10.1073/pnas.73.3.804.

Sen, George L.; Reuter, Jason A.; Webster, Daniel E.; Zhu, Lilly; Khavari, Paul A. (2010): DNMT1 maintains progenitor function in self-renewing somatic tissue. In Nature 463 (7280), pp. 563–567. DOI: 10.1038/nature08683.

Shen, J. C.; Rideout, W. M.; Jones, P. A. (1994): The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. In Nucleic acids research 22 (6), pp. 972–976. DOI: 10.1093/nar/22.6.972.

Shi, Dong-Qiao; Ali, Iftikhar; Tang, Jun; Yang, Wei-Cai (2017): New Insights into 5hmC DNA Modification: Generation, Distribution and Function. In Frontiers in genetics 8, p. 100. DOI: 10.3389/fgene.2017.00100.

Shirane, Kenjiro; Toh, Hidehiro; Kobayashi, Hisato; Miura, Fumihito; Chiba, Hatsune; Ito, Takashi et al. (2013): Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. In PLoS genetics 9 (4), e1003439. DOI: 10.1371/journal.pgen.1003439.

Song, Chun-Xiao; Szulwach, Keith E.; Dai, Qing; Fu, Ye; Mao, Shi-Qing; Lin, Li et al. (2013): Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. In Cell 153 (3), pp. 678–691. DOI: 10.1016/j.cell.2013.04.001.

Song, Jikui; Rechkoblit, Olga; Bestor, Timothy H.; Patel, Dinshaw J. (2011): Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. In Science (New York, N.Y.) 331 (6020), pp. 1036–1040. DOI: 10.1126/science.1195380.

Song, Jikui; Teplova, Marianna; Ishibe-Murakami, Satoko; Patel, Dinshaw J. (2012): Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. In Science (New York, N.Y.) 335 (6069), pp. 709–712. DOI: 10.1126/science.1214453.

Sonnet, Miriam; Claus, Rainer; Becker, Natalia; Zucknick, Manuela; Petersen, Jana; Lipka, Daniel B. et al. (2014): Early aberrant DNA methylation events in a mouse model of acute myeloid leukemia. In Genome medicine 6 (4), p. 34. DOI: 10.1186/gm551.

Spada, Fabio; Haemmer, Andrea; Kuch, David; Rothbauer, Ulrich; Schermelleh, Lothar; Kremmer, Elisabeth et al. (2007): DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. In The Journal of cell biology 176 (5), pp. 565–571. DOI: 10.1083/jcb.200610062.

Sperlazza, M. Jeannette; Bilinovich, Stephanie M.; Sinanan, Leander M.; Javier, Fatima R.; Williams, David C. (2017): Structural Basis of MeCP2 Distribution on Non-CpG Methylated and Hydroxymethylated DNA. In Journal of molecular biology 429 (10), pp. 1581–1594. DOI: 10.1016/j.jmb.2017.04.009.

Spruijt, Cornelia G.; Gnerlich, Felix; Smits, Arne H.; Pfaffeneder, Toni; Jansen, Pascal W. T. C.; Bauer, Christina et al. (2013): Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. In Cell 152 (5), pp. 1146–1159. DOI: 10.1016/j.cell.2013.02.004.

Studier, F. William (2005): Protein production by auto-induction in high density shaking cultures. In Protein expression and purification 41 (1), pp. 207–234. DOI: 10.1016/j.pep.2005.01.016.

Suetake, Isao; Miyazaki, Junko; Murakami, Chikako; Takeshima, Hideyuki; Tajima, Shoji (2003): Distinct enzymatic properties of recombinant mouse DNA methyltransferases Dnmt3a and Dnmt3b. In Journal of biochemistry 133 (6), pp. 737–744. DOI: 10.1093/jb/mvg095.

Suetake, Isao; Shinozaki, Fuminori; Miyagawa, Junichi; Takeshima, Hideyuki; Tajima, Shoji (2004): DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. In The Journal of biological chemistry 279 (26), pp. 27816–27823. DOI: 10.1074/jbc.M400181200.

Tahiliani, Mamta; Koh, Kian Peng; Shen, Yinghua; Pastor, William A.; Bandukwala, Hozefa; Brudno, Yevgeny et al. (2009): Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. In Science (New York, N.Y.) 324 (5929), pp. 930–935. DOI: 10.1126/science.1170116.

Takahashi, Saori; Suetake, Isao; Engelhardt, Jan; Tajima, Shoji (2015): A novel method to analyze 5-hydroxymethylcytosine in CpG sequences using maintenance DNA methyltransferase, DNMT1. In FEBS open bio 5, pp. 741–747. DOI: 10.1016/j.fob.2015.09.003.

Takai, Daiya; Jones, Peter A. (2002): Comprehensive analysis of CpG islands in human chromosomes 21 and 22. In Proceedings of the National Academy of Sciences of the United States of America 99 (6), pp. 3740–3745. DOI: 10.1073/pnas.052410099.

Takebayashi, Shin-ichiro; Tamura, Takashi; Matsuoka, Chisa; Okano, Masaki (2007): Major and essential role for the DNA methylation mark in mouse embryogenesis and stable association of DNMT1 with newly replicated regions. In Molecular and cellular biology 27 (23), pp. 8243–8258. DOI: 10.1128/MCB.00899-07.

Tamanaha, Esta; Guan, Shengxi; Marks, Katherine; Saleh, Lana (2016): Distributive Processing by the Iron(II)/α-Ketoglutarate-Dependent Catalytic Domains of the TET Enzymes Is Consistent with Epigenetic Roles for Oxidized 5-Methylcytosine Bases. In Journal of the American Chemical Society 138 (30), pp. 9345–9348. DOI: 10.1021/jacs.6b03243.

Tatton-Brown, Katrina; Seal, Sheila; Ruark, Elise; Harmer, Jenny; Ramsay, Emma; Del Vecchio Duarte, Silvana et al. (2014): Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. In Nature genetics 46 (4), pp. 385–388. DOI: 10.1038/ng.2917.

Tefferi, A.; Pardanani, A.; Lim, K-H; Abdel-Wahab, O.; Lasho, T. L.; Patel, J. et al. (2009): TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. In Leukemia 23 (5), pp. 905–911. DOI: 10.1038/leu.2009.47.

Traube, Franziska R.; Schiffers, Sarah; Iwan, Katharina; Kellner, Stefanie; Spada, Fabio; Müller, Markus; Carell, Thomas (2019): Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides. In Nature protocols 14 (1), pp. 283–312. DOI: 10.1038/s41596-018-0094-6.

Trowbridge, Jennifer J.; Snow, Jonathan W.; Kim, Jonghwan; Orkin, Stuart H. (2009): DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. In Cell stem cell 5 (4), pp. 442–449. DOI: 10.1016/j.stem.2009.08.016.

Tsumura, Akiko; Hayakawa, Tomohiro; Kumaki, Yuichi; Takebayashi, Shin-ichiro; Sakaue, Morito; Matsuoka, Chisa et al. (2006): Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. In Genes to cells : devoted to molecular & cellular mechanisms 11 (7), pp. 805–814. DOI: 10.1111/j.1365-2443.2006.00984.x.

Valegård, Karin; van Terwisscha Scheltinga, Anke C.; Dubus, Alain; Ranghino, Graziella; Oster, Linda M.; Hajdu, Janos; Andersson, Inger (2004): The structural basis of cephalosporin formation in a mononuclear ferrous enzyme. In Nature structural & molecular biology 11 (1), pp. 95–101. DOI: 10.1038/nsmb712.

Valinluck, Victoria; Sowers, Lawrence C. (2007): Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. In Cancer research 67 (3), pp. 946–950. DOI: 10.1158/0008-5472.CAN-06-3123.

Vilkaitis, Giedrius; Suetake, Isao; Klimasauskas, Saulius; Tajima, Shoji (2005): Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. In The Journal of biological chemistry 280 (1), pp. 64–72. DOI: 10.1074/jbc.M411126200.

Waddington, C. H. (2012): The epigenotype. 1942. In International journal of epidemiology 41 (1), pp. 10–13. DOI: 10.1093/ije/dyr184.

Wallace, Robert G.; Twomey, Laura C.; Custaud, Marc-Antoine; Turner, Jonathan D.; Moyna, Niall; Cummins, Philip M.; Murphy, Ronan P. (2018): The role of epigenetics in cardiovascular health and ageing: A focus on physical activity and nutrition. In Mechanisms of ageing and development 174, pp. 76–85. DOI: 10.1016/j.mad.2017.11.013.

Wang, Chengkun; Shen, Jie; Yang, Zhongzheng; Chen, Ping; Zhao, Bin; Hu, Wei et al. (2011): Structural basis for site-specific reading of unmodified R2 of histone H3 tail by UHRF1 PHD finger. In Cell research 21 (9), pp. 1379–1382. DOI: 10.1038/cr.2011.123.

Watanabe, Daisuke; Suetake, Isao; Tada, Takashi; Tajima, Shoji (2002): Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. In Mechanisms of Development 118 (1-2), pp. 187–190. DOI: 10.1016/S0925-4773(02)00242-3.

Weisenberger, Daniel J.; Campan, Mihaela; Long, Tiffany I.; Kim, Myungjin; Woods, Christian; Fiala, Emerich et al. (2005): Analysis of repetitive element DNA methylation by MethyLight. In Nucleic acids research 33 (21), pp. 6823–6836. DOI: 10.1093/nar/gki987.

Wienholz, Bethany L.; Kareta, Michael S.; Moarefi, Amir H.; Gordon, Catherine A.; Ginno, Paul A.; Chédin, Frédéric (2010): DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo. In PLoS genetics 6 (9), e1001106. DOI: 10.1371/journal.pgen.1001106.

Williams, Kristine; Christensen, Jesper; Pedersen, Marianne Terndrup; Johansen, Jens V.; Cloos, Paul A. C.; Rappsilber, Juri; Helin, Kristian (2011): TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. In Nature 473 (7347), pp. 343–348. DOI: 10.1038/nature10066.

Wossidlo, Mark; Nakamura, Toshinobu; Lepikhov, Konstantin; Marques, C. Joana; Zakhartchenko, Valeri; Boiani, Michele et al. (2011): 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. In Nature communications 2, p. 241. DOI: 10.1038/ncomms1240.

Wu, Ct; Morris, J. R. (2001): Genes, genetics, and epigenetics: a correspondence. In Science (New York, N.Y.) 293 (5532), pp. 1103–1105. DOI: 10.1126/science.293.5532.1103.

Wu, Hao; D'Alessio, Ana C.; Ito, Shinsuke; Xia, Kai; Wang, Zhibin; Cui, Kairong et al. (2011): Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. In Nature 473 (7347), pp. 389–393. DOI: 10.1038/nature09934.

Wu, Xiaoji; Zhang, Yi (2017): TET-mediated active DNA demethylation: mechanism, function and beyond. In Nature reviews. Genetics 18 (9), pp. 517–534. DOI: 10.1038/nrg.2017.33.

Wyatt, G. R.; Cohen, S. S. (1952): A new pyrimidine base from bacteriophage nucleic acids. In Nature 170 (4338), pp. 1072–1073. DOI: 10.1038/1701072a0.

Yamagata, Kazuyuki; Kobayashi, Akira (2017): The cysteine-rich domain of TET2 binds preferentially to mono- and dimethylated histone H3K36. In Journal of biochemistry 161 (4), pp. 327–330. DOI: 10.1093/jb/mvx004.

Yamaguchi, Shinpei; Hong, Kwonho; Liu, Rui; Shen, Li; Inoue, Azusa; Diep, Dinh et al. (2012): Tet1 controls meiosis by regulating meiotic gene expression. In Nature 492 (7429), pp. 443–447. DOI: 10.1038/nature11709.

Yamaguchi, Shinpei; Shen, Li; Liu, Yuting; Sendler, Damian; Zhang, Yi (2013): Role of Tet1 in erasure of genomic imprinting. In Nature 504 (7480), pp. 460–464. DOI: 10.1038/nature12805.

Yamashita, Y.; Yuan, J.; Suetake, I.; Suzuki, H.; Ishikawa, Y.; Choi, Y. L. et al. (2010): Array-based genomic resequencing of human leukemia. In Oncogene 29 (25), pp. 3723–3731. DOI: 10.1038/onc.2010.117.

Yan, Xiao-Jing; Xu, Jie; Gu, Zhao-Hui; Pan, Chun-Ming; Lu, Gang; Shen, Yang et al. (2011): Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. In Nature genetics 43 (4), pp. 309–315. DOI: 10.1038/ng.788.

Yang, Liubin; Rau, Rachel; Goodell, Margaret A. (2015): DNMT3A in haematological malignancies. In Nature reviews. Cancer 15 (3), pp. 152–165. DOI: 10.1038/nrc3895.

Ye, Fei; Kong, Xiangqian; Zhang, Hao; Liu, Yan; Shao, Zhiyuan; Jin, Jia et al. (2018): Biochemical Studies and Molecular Dynamic Simulations Reveal the Molecular Basis of Conformational Changes in DNA Methyltransferase-1. In ACS chemical biology 13 (3), pp. 772–781. DOI: 10.1021/acschembio.7b00890.

Yin, Ruichuan; Mao, Shi-Qing; Zhao, Bailin; Chong, Zechen; Yang, Ying; Zhao, Chao et al. (2013): Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. In Journal of the American Chemical Society 135 (28), pp. 10396–10403. DOI: 10.1021/ja4028346.

Yoder, J. A.; Soman, N. S.; Verdine, G. L.; Bestor, T. H. (1997): DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. In Journal of molecular biology 270 (3), pp. 385–395. DOI: 10.1006/jmbi.1997.1125.

Yu, Zhong; Genest, Paul-André; Riet, Bas ter; Sweeney, Kate; DiPaolo, Courtney; Kieft, Rudo et al. (2007): The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. In Nucleic acids research 35 (7), pp. 2107–2115. DOI: 10.1093/nar/gkm049.

Zhang, Jiayu; Yang, Cheng; Wu, Chunfu; Cui, Wei; Wang, Lihui (2020): DNA Methyltransferases in Cancer: Biology, Paradox, Aberrations, and Targeted Therapy. In Cancers 12 (8). DOI: 10.3390/cancers12082123.

Zhang, Yingying; Rohde, Christian; Tierling, Sascha; Jurkowski, Tomasz P.; Bock, Christoph; Santacruz, Diana et al. (2009): DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. In PLoS genetics 5 (3), e1000438. DOI: 10.1371/journal.pgen.1000438.

Zhang, Zhi-Min; Lu, Rui; Wang, Pengcheng; Yu, Yang; Chen, Dongliang; Gao, Linfeng et al. (2018): Structural basis for DNMT3A-mediated de novo DNA methylation. In Nature 554 (7692), pp. 387–391. DOI: 10.1038/nature25477.

Zhou, Ting; Xiong, Jun; Wang, Mingzhu; Yang, Na; Wong, Jiemin; Zhu, Bing; Xu, Rui-Ming (2014): Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. In Molecular cell 54 (5), pp. 879–886. DOI: 10.1016/j.molcel.2014.04.003.

Ziller, Michael J.; Müller, Fabian; Liao, Jing; Zhang, Yingying; Gu, Hongcang; Bock, Christoph et al. (2011): Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. In PLoS genetics 7 (12), e1002389. DOI: 10.1371/journal.pgen.1002389.

Zimmermann, C.; Guhl, E.; Graessmann, A. (1997): Mouse DNA methyltransferase (MTase) deletion mutants that retain the catalytic domain display neither de novo nor

maintenance methylation activity in vivo. In Biological chemistry 378 (5), pp. 393–405. DOI: 10.1515/bchm.1997.378.5.393.

7 Author contributions

Adam, Sabrina*; Anteneh, Hiwot*; Hornisch, Maximilian; Wagner, Vincent; Lu, Jiuwei; Radde, Nicole E. et al. (2020): DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. In Nature communications 11 (1), p. 3723. DOI: 10.1038/s41467-020-17531-8.

* These authors contributed equally to the work.

I have performed the biochemical work with the help of MH. I conducted the bioinformatic work with the help of PB and AJ and contributed to the analysis of the data and their interpretation.

Gao, Linfeng*; Emperle, Max*; Guo, Yiran; Grimm, Sara A.; Ren, Wendan; **Adam, Sabrina** et al. (2020b): Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. In Nature communications 11 (1), p. 3355. DOI: 10.1038/s41467-020-17109-4.

* These authors contributed equally to the work.

I implemented the Deep Enzymology approach used in this work and conducted the bioinformatic work with the help of PB and AJ. Additionally, I contributed to the analysis of the data and their interpretation.

Emperle, Max*; Bangalore, Disha M.*; **Adam, Sabrina**; Kunert, Stefan; Heil, Hannah S.; Heinze, Katrin G. et al. (2021): Structural and biochemical insight into the mechanism of dual CpG site binding and methylation by the DNMT3A DNA methyltransferase. In Nucleic acids research 49 (14), pp. 8294–8308. DOI: 10.1093/nar/gkab600.

* These authors contributed equally to the work.

I contributed to the design of the adapted Deep Enzymology approach used in this work and conducted the NGS library preparation. Additionally, I conducted bioinformatic work with the help of PB and contributed to the analysis of the data and their interpretation.

Emperle, Max*; **Adam, Sabrina***; Kunert, Stefan; Dukatz, Michael; Baude, Annika; Plass, Christoph et al. (2019): Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns. In Nucleic acids research 47 (21), pp. 11355–11367. DOI: 10.1093/nar/gkz911.

* These authors contributed equally to the work.

I implemented the Deep Enzymology approach used in this work and conducted the bioinformatic work with the help of PB and AJ. Additionally, I contributed to the analysis of the data and their interpretation.

Mack, Alexandra; Emperle, Max; Schnee, Philipp; **Adam, Sabrina**; Pleiss, Jürgen; Bashtrykov, Pavel; Jeltsch, Albert (2022): Preferential Self-interaction of DNA Methyltransferase DNMT3A Subunits Containing the R882H Cancer Mutation Leads to Dominant Changes of Flanking Sequence Preferences. In Journal of molecular biology 434 (7), p. 167482. DOI: 10.1016/j.jmb.2022.167482.

I contributed to the design of the adapted Deep Enzymology approach used in this work and conducted the bioinformatic work with the help of ME, PB and AJ. Additionally, I contributed to the analysis of the data and their interpretation.

Adam, Sabrina; Bräcker, Julia; Klingel, Viviane; Osteresch, Bernd; Radde, Nicole E.; Brockmeyer, Jens et al. (2022): Flanking sequences influence the activity of TET1 and TET2 methylcytosine dioxygenases and affect genomic hm5C patterns. In Communications biology 5 (1), p. 92. DOI: 10.1038/s42003-022-03033-4.

I have performed protein purification and enzyme kinetics (for HPLC-MS/MS or Deep Enzymology) and conducted bioinformatic work with the help of PB and AJ. Additionally, I contributed to the analysis of the data and their interpretation and the preparation of the manuscript together with AJ.

8 Appendix (not included in the published thesis)

Appendix 1:

Adam, Sabrina*; Anteneh, Hiwot*; Hornisch, Maximilian; Wagner, Vincent; Lu, Jiuwei; Radde, Nicole E. et al. (2020): DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. In Nature communications 11 (1), p. 3723. DOI: 10.1038/s41467-020-17531-8.

* These authors contributed equally to the work.

Appendix 2:

Gao, Linfeng*; Emperle, Max*; Guo, Yiran; Grimm, Sara A.; Ren, Wendan; **Adam, Sabrina** et al. (2020b): Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. In Nature communications 11 (1), p. 3355. DOI: 10.1038/s41467-020-17109-4.

* These authors contributed equally to the work.

Appendix 3:

Emperle, Max*; Bangalore, Disha M.*; **Adam, Sabrina**; Kunert, Stefan; Heil, Hannah S.; Heinze, Katrin G. et al. (2021): Structural and biochemical insight into the mechanism of dual CpG site binding and methylation by the DNMT3A DNA methyltransferase. In Nucleic acids research 49 (14), pp. 8294–8308. DOI: 10.1093/nar/gkab600.

* These authors contributed equally to the work.

Appendix 4:

Emperle, Max*; **Adam, Sabrina***; Kunert, Stefan; Dukatz, Michael; Baude, Annika; Plass, Christoph et al. (2019): Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns. In Nucleic acids research 47 (21), pp. 11355–11367. DOI: 10.1093/nar/gkz911.

* These authors contributed equally to the work.

Appendix 5:

Mack, Alexandra; Emperle, Max; Schnee, Philipp; **Adam, Sabrina**; Pleiss, Jürgen; Bashtrykov, Pavel; Jeltsch, Albert (2022): Preferential Self-interaction of DNA Methyltransferase DNMT3A Subunits Containing the R882H Cancer Mutation Leads to Dominant Changes of Flanking Sequence Preferences. In Journal of molecular biology 434 (7), p. 167482. DOI: 10.1016/j.jmb.2022.167482.

Appendix 6:

Adam, Sabrina; Bräcker, Julia; Klingel, Viviane; Osteresch, Bernd; Radde, Nicole E.; Brockmeyer, Jens et al. (2022): Flanking sequences influence the activity of TET1 and TET2 methylcytosine dioxygenases and affect genomic hm5C patterns. In Communications biology 5 (1), p. 92. DOI: 10.1038/s42003-022-03033-4.