

Mechanistic studies on the DNA methyltransferases DNMT3A and DNMT3B

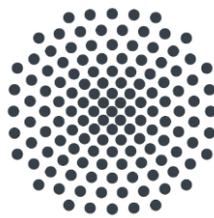
Von der Fakultät 3: Chemie der Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Michael Dukatz

Geboren am 18.04.1985 in Nürtingen

Hauptberichter: Prof. Dr. Albert Jeltsch
Mitberichter: Prof. Dr. Bernhard Hauer
Prüfungsvorsitzender: Prof. Dr. Andreas Köhn
Tag der mündlichen Prüfung: 25. Juni 2021



University of Stuttgart
Institute of Biochemistry and Technical
Biochemistry

2021

Erklärung über die Eigenständigkeit der Dissertation

Ich versichere, dass ich die vorliegende Arbeit mit dem Titel

‘Mechanistic studies on the DNA methyltransferases DNMT3A and DNMT3B’

selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe; aus fremden Quellen entnommene Passagen und Gedanken sind als solche kenntlich gemacht.

Declaration of Authorship

I hereby certify that the dissertation entitled

‘Mechanistic studies on the DNA methyltransferases DNMT3A and DNMT3B’

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

Name/Name: Michael Dukatz

Unterschrift/Signature:

Datum/Date: Tuesday, April 06, 2021

Table of Contents

Erklärung über die Eigenständigkeit der Dissertation.....	I
Declaration of Authorship	I
Acknowledgements.....	VII
List of publications	IX
Zusammenfassung	XI
Abstract	XIII
List of abbreviations.....	XV
1 Introduction.....	1
1.1 Basis of epigenetics.....	1
1.2 Chromatin structure	4
1.3 DNA methylation.....	6
1.3.1 Discovery and occurrence	6
1.3.2 The DNA methyltransferase family	8
1.3.3 The classical model of DNA methylation establishment and maintenance.....	9
1.3.4 Revision and adaptation of the classical model of DNA methylation inheritance	
11	
1.3.5 Catalytic function of the C-terminus of the DNMTs.....	14
1.3.6 Substrate recognition and catalytic preference of <i>de novo</i> DNMTs	15
1.3.7 Role of DNMT3 enzymes in non-CpG methylation.....	16
1.3.8 Recruitment of DNMTs and interaction with chromatin modifiers	16
1.3.9 Regulatory functions of the N-terminus of the <i>de novo</i> DNMTs.....	18
1.3.10 Effects of DNA methylation	21
1.4 Post-translational modifications of histone tails (PTMs)	23
2 Principal aims of the study.....	25
2.1 Investigation of the chromatin binding of the PWWP domain of DNMT3A.....	25
2.2 Investigation in the allosteric regulation of DNMT3A by MeCP2 and H3-tail peptides	
25	
2.3 Investigation of the catalytic mechanism of DNMT3A for cytosine-N3 methylation	
26	

2.4	Investigation of the flanking sequence preference of <i>de novo</i> DNA methyltransferases	26
3	Results	28
3.1	Targeting and chromatin interaction studies of the DNMT3A methyltransferase	28
3.1.1	H3K36me _{2/3} binding of the DNMT3A PWWP domain is disrupted by the D329A and K295E/I mutations	30
3.1.2	DNA binding of the DNMT3A PWWP domain is disrupted by the K295E/I and R362E mutations	30
3.1.3	Chromatin interaction of DNMT3A PWWP domain requires both, H3K36me _{2/3} as well as DNA binding	31
3.1.4	Influence of PWWP mutations on the sub-nuclear localization of DNMT3A	32
3.2	Investigation of the activity of DNMT3A with MeCP2 and H3	34
3.3	Mechanistic studies on cytosine N3 methylation by DNA methyltransferase DNMT3A	36
3.4	Investigation of differences in the sequence preference of DNMT3A and DNMT3B	39
3.5	Complex DNA sequence preference analysis of the DNMT3B DNA methyltransferase	42
4	Discussion	49
4.1	Targeting and chromatin interaction studies via the PWWP domain of the DNMT3A methyltransferase	50
4.2	Allosteric regulation of DNMT3A by competitive binding of MeCP2 and H3	53
4.3	Mechanistic studies on cytosine N3 methylation by DNA methyltransferase DNMT3A	55
4.4	Investigation of the flanking sequence preference of <i>de novo</i> DNA methyltransferases	58
4.5	Conclusion and perspectives	61
5	Materials and Methods	64
5.1	Cloning, mutagenesis, expression and purification of recombinant proteins	64
5.2	Biotin-avidin microplate assay	66
5.3	Flanking sequence preference analysis and library generation	67

5.4	Circular dichroism spectroscopy	68
5.5	Peptide arrays	68
5.6	Electrophoretic Mobility Shift Assay (EMSA).....	68
5.7	Tissue culture, harvesting and transfections.....	68
5.8	Isolation of nucleosomes.....	69
5.9	Nucleosomal pull-down assay.....	69
5.10	Western blot and antibody detection	69
6	References	71
7	Author contributions	86
8	Appendix (not included in the published thesis)	87

Acknowledgements

I am very grateful to everybody who supported me throughout my doctoral work! In particular I am thankful to Prof. Dr. Albert Jeltsch who gave me the opportunity to write my doctoral thesis at this institute. The time here was certainly not always a bed of roses, but with your analytical and rational scientific way of working and never lacking in ideas if something has not gone so well for a while you have become a good role model for my scientific work and certainly also for the future.

I am also thankful to Dr. Renata Jurkowska who initiated the launch of my work here.

I would also like to thank Prof. Dr. Bernhard Hauer for being the co-reporter of my thesis and Prof. Dr. Andreas Köhn for being the chair of my examination board.

A special thank you goes to the bachelor and master students I supervised, Max, Iris, Marianna, Evelin, Nora and Elias! The time with you was a lot of fun and I certainly learned more than it seemed for you.

I would also like to thank all my colleagues, be it for scientific or practical advice (special thanks to Pavel, Max and Cristiana!) or just for the nice working atmosphere that was present all the time.

Many thanks also to you Elisabeth, when there were battles to be fought with the administration you were always friendly and supportive with good advice. I also want to thank Dragica, who made the lab work much more comfortable!

Last but not least, I would like to thank my family and Franzi just for being there for me!

List of publications

Dukatz, M., Holzer, K., Choudalakis, M., Emperle, M., Lungu, C., Bashtrykov, P. & Jeltsch, A. (2019). "H3K36me2/3 Binding and DNA Binding of the DNA Methyltransferase DNMT3A PWWP Domain Both Contribute to Its Chromatin Interaction." *Journal of Molecular Biology* 431(24). doi: 10.1016/j.jmb.2019.09.006

Dukatz, M.*, Requena, C. E.*, Emperle, M., Hajkova, P., Sarkies, P. & Jeltsch, A. (2019). "Mechanistic Insights into Cytosine-N3 Methylation by DNA Methyltransferase DNMT3A." *Journal of Molecular Biology* 431(17). doi: 10.1016/j.jmb.2019.06.015

* These authors contributed equally to the work

Dukatz, M., Adam, S., Biswal, M., Song, J., Bashtrykov, P. & Jeltsch, A. (2020). "Complex DNA Sequence Readout Mechanisms of the DNMT3B DNA Methyltransferase." *Nucleic Acids Research* (14):1–15. doi: 10.1093/nar/gkaa938

Emperle, M., Adam, S., Kunert, S., Dukatz, M., Baude, A., Plass, C., Rathert, P., Bashtrykov, P. & Jeltsch, A. (2019). "Mutations of R882 Change Flanking Sequence Preferences of the DNA Methyltransferase DNMT3A and Cellular Methylation Patterns." *Nucleic Acids Research* 47(21). doi: 10.1093/nar/gkz911

(Not included in this thesis)

Emperle, M., Dukatz, M., Kunert, S., Holzer, K., Rajavelu, A., Jurkowska, R. Z. & Jeltsch, A. (2018). "The DNMT3A R882H Mutation Does Not Cause Dominant Negative Effects in Purified Mixed DNMT3A/R882H Complexes." *Scientific Reports* 8(1). doi: 10.1038/s41598-018-31635-8

(Not included in this thesis)

Gao, L.*, Emperle, M.*, Guo, Y.*, Grimm, S. A., Ren, W., Adam, S., Uryu, H., Zhang, Z. M., Chen, D., Yin D., Dukatz, M., Anteneh, H., Jurkowska, R. Z., Lu, J., Wang, Y., Bashtrykov, P., Wade, P. A., Wang, G. G., Jeltsch, A. & Song, J. (2020). "Comprehensive Structure-Function Characterization of DNMT3B and DNMT3A Reveals Distinctive de Novo DNA Methylation Mechanisms." *Nature Communications* 11(1). doi: 10.1038/s41467-020-17109-4

* These authors contributed equally to the work

Hassanzadeh, M., Kasymov, R., Mahernia, S., Adib, M., Emperle, M., Dukatz, M., Bashtrykov, P., Jeltsch, A. & Amanlou, M. (2017). "Discovery of Novel and Selective DNA Methyltransferase 1 Inhibitors by Pharmacophore and Docking-Based Virtual Screening." *ChemistrySelect* 2(27). doi: 10.1002/slct.201701734

(Not included in this thesis)

Rajavelu, A.*, Lungu, C.*, Emperle, M., **Dukat, M.**, Bröhm, A., Broche, J., Hanelt, I., Parsa, E., Schiffers, S., Karnik, R., Meissner, A., Carell, T., Rathert, P., Jurkowska, R. Z. & Jeltsch, A. (2018). **“Chromatin-Dependent Allosteric Regulation of DNMT3A Activity by MeCP2.”** *Nucleic Acids Research* 46(17). doi: 10.1093/nar/gky715

* These authors contributed equally to the work

Zusammenfassung

Untersuchungen von vererbbaaren Phänotyp Veränderungen, bei denen keine Veränderungen in der DNA-Sequenz stattfinden, fallen in das Fachgebiet der Epigenetik. Neben nicht-kodierenden RNAs und Histon Modifikationen ist die DNA-Methylierung eines der wichtigsten Elemente der epigenetischen Signalwege. In eukaryotischen Organismen wird diese Markierung hauptsächlich an der C5-Position von Cytosin Basen im CpG-Kontext gesetzt, kommt aber auch in geringerem Umfang im Nicht-CpG-Kontext vor. Die DNA Methylierung des menschlichen Genoms findet in zwei Wellen während der Embryonalentwicklung statt und ist für eine gesunde Entwicklung unerlässlich. Die Enzyme, welche für die Etablierung von diesen Methylierungsmustern auf der DNA verantwortlich sind, sind die *de novo* DNA-Methyltransferasen DNMT3A und DNMT3B. Die gezielte Etablierung von Methylierungsmustern erfordert eine präzise Rekrutierung und Kontrolle der Aktivität der DNA-Methyltransferasen. In dieser Arbeit wurden regulatorische und katalytische Mechanismen der *de novo* DNA-Methyltransferasen untersucht, welche die Interaktion mit anderen Proteinen und die spezifische Erkennung des Substrats einschlossen. Ein weiterer Teil dieser Arbeit zielt darauf ab aufzuklären, wie die enzymatische Generierung von 3-Methylcytosin durch DNMT3A erfolgen kann.

Die Erkennung von di- und trimethyliertem Lysin 36 an Histon H3-Schwänzen durch die PWWP-Domäne von DNMT3A ist essentiell für die heterochromatische Zielbestimmung des Enzyms. Im ersten Teil dieser Arbeit wurde gezeigt, dass die Rekrutierung von DNMT3A über diese Histon Modifikationen allein nicht ausreicht, um eine stabile heterochromatische Lokalisation zu gewährleisten, sondern die gleichzeitige Bindung der PWWP-Domäne an die DNA notwendig ist, um das Enzym in seiner Zielregion zu halten. Durch gezielte Mutationen auf der Oberfläche der PWWP-Domäne konnten zwei basische Reste (K295 und R362) identifiziert werden, welche maßgeblich an der DNA-Bindung beteiligt sind. In Nukleosomen-Pull-down-Experimenten wurde gezeigt, dass sowohl eine gestörte DNA-Bindung als auch eine gestörte H3K36me_{2/3}-Bindung zu einem massiven Verlust der Interaktion mit Chromatin führt. Lokalisierungsstudien von fluoreszenzmarkierten DNMT3A-Wildtyp und DNMT3A-PWWP-Mutanten in NIH 3T3-Zellen bestätigten diese Beobachtung in lebenden Zellen.

Das zweite Kapitel dieser Arbeit beschreibt, welche Beiträge ich zu einem laufenden Projekt von ehemaligen Institutsmitgliedern geleistet habe, in welchem eine starke Interaktion

zwischen der ADD-Domäne von DNMT3A und der TRD-Domäne des 5mC-Leseproteins MeCP2 aufgezeigt wurde. Mit einer Wiederholung von kritischen Pull-Down Experimenten der beiden Domänen in Gegenwart einer unspezifischen Endonuklease konnte ich letzte Zweifel an einer direkten Interaktion ausräumen. Weiterhin konnte ich zeigen, dass das durch MeCP2 inhibierte DNMT3A durch die Bindung an ein H3-Peptid mit unmodifiziertem Lysin 4 stärker stimuliert wurde als DNMT3A allein. So konnte ich zu der Entwicklung eines Modells beitragen, in welchem MeCP2 sowohl als Repressor oder als Aktivator von DNMT3A fungieren kann, abhängig davon welcher Methylierungsstatus an Lysin 4 von Histon H3 vorliegt.

Aufbauend auf der kürzlich erfolgten Entdeckung das DNMT3A neben 5-Methylcytosin auch geringe Mengen von 3-Methylcytosin generieren kann, wurde im dritten Projekt ein verstärktes Augenmerk auf die katalytische Tasche des Enzyms gelegt. Dazu wurden spezifische Kontaktstellen von DNMT3A für das aus der DNA herausgeflippte Cytosin entfernt und mit diesen Mutanten ein DNA-Substrat *in vitro* methyliert. In Zusammenarbeit mit der Gruppe von Peter Sarkies (MRC London) wurden diese Proben mittels LC-MS auf die Menge an generiertem 3-mC und 5-mC untersucht. Basierend auf den Ergebnissen wurde ein 'inverted base flipping' Mechanismus vorgestellt, der eine 180°-Drehung der Cytosin Base entlang der glykosidischen Bindung annimmt und so die Entstehung von 3-Methylcytosin erklären kann.

Der letzte Teil dieser Arbeit konzentrierte sich auf die Methylierungspräferenzen von DNMT3B in Bezug auf die flankierenden Basen um die zentrale CpG- oder CpX-Stelle. Durch Vergleich der erhaltenen Ergebnisse mit den Flankierungssequenzpräferenzen von DNMT3A und unter Verwendung verfügbarer Kristallstrukturen von DNMT3A und DNMT3B im Komplex mit DNA konnte unter anderem eine Erklärung für die geringere CpG-Spezifität von DNMT3B vorgestellt werden. Darüber hinaus wurde das Lysin 777 von DNMT3B als Sensor für die +1 flankierende Base erkannt, was dessen erhöhte Präferenz für Purine an dieser Position erklärt. Darüber konnte gezeigt werden, dass Threonin 775 eine wesentliche, dynamische Rolle im katalytischen Mechanismus von DNMT3B spielt. In Übereinstimmung mit der geringeren CpG-Spezifität von DNMT3B im Vergleich zu DNMT3A wurde auch eine hohe Korrelation der *in vitro* Nicht-CpG-Flankierungspräferenzen mit Nicht-CpG-Methylierungsmustern aus menschlichen Zellen nachgewiesen. Dies unterstreicht die biologische Relevanz dieser Nicht-CpG-Spezifität dieses Enzyms.

Abstract

Epigenetic regulation refers to the control of gene expression and chromatin states that is heritable but not encoded in the DNA sequence. In addition to non-coding RNAs and histone modifications, DNA methylation is one of the major elements of the epigenetic signaling pathways. In eukaryotic organisms, this mark is mainly placed at the C5 position of cytosine bases at CpG sites, but also occurs to a minor extent in non-CpG context. In the human genome, DNA methylation takes place in two waves during embryonic development and germ cell differentiation and it is indispensable for a healthy development. The enzymes responsible for establishing methylation patterns on DNA are the *de novo* DNA methyltransferases DNMT3A and DNMT3B. The targeted establishment of methylation patterns requires precise recruitment and control of the activity of the DNA-methyltransferases. In this work, both regulatory and catalytic mechanisms of *de novo* methyltransferases were investigated, which include interactions with other proteins and the specific recognition of the substrate sequence. Another part of this work strived to elucidate how enzymatic generation of 3-methylcytosine by DNMT3A can occur.

The recognition of di- and trimethylated K36 on histone H3 tails by the PWWP domain of DNMT3A is essential for heterochromatic targeting of the enzyme. In the first part of this work, it was shown that recruitment of DNMT3A by these histone modifications alone is not sufficient to ensure stable heterochromatic localization, but the simultaneous binding of the PWWP domain to DNA is necessary in both cases to keep the enzyme in its target region. With targeted mutations on the surface of the PWWP domain, two basic residues (K295 and R362) could be identified which are required for DNA binding. In nucleosome pull-down experiments, it became clear that disruption of DNA binding or disruption of H3K36me_{2/3} binding, resulted in a massive loss of interaction with chromatin. Localization studies of fluorescently labeled full length DNMT3A wildtype and DNMT3A PWWP mutants in NIH 3T3 cells confirmed this observation in living cells.

The second chapter of this thesis describes how I contributed to a running project initiated by former institute members where a strong interaction between the ADD domain of DNMT3A and the TRD domain of the 5mC reading protein MeCP2 was investigated. With a repetition of critical pull-down experiments of the two domains in the presence of a non-specific endonuclease, I was able to remove existing doubts regarding a direct interaction.

Furthermore, I was able to show that DNMT3A2 inhibited by MeCP2 could be stimulated to a greater extent by binding to an H3 peptide containing unmodified lysine 4 than DNMT3A2 alone. With this, I contributed to the development of a model in which MeCP2 can function as either a repressor or an activator of DNMT3A depending on the methylation status at histone H3 lysine 4.

Building on the recent discovery of the ability of DNMT3A to generate small amounts of 3-methyl cytosine, my third sub-project focused on the catalytic mechanism of the DNMT3A enzyme. In the catalytic pocket of DNMT3A, specific contact points of the enzyme to the cytosine flipped from the DNA were removed and a DNA substrate was methylated *in vitro* with these mutants. In cooperation with the group of Peter Sarkies (MRC London), these samples were analyzed for the amount of generated 3-mC and 5-mC by LC-MS. Based on the results, an 'inverted base flipping' mechanism was developed, which assumes a 180° rotation of the cytosine base along the glycosidic bond leading to its methylation at the N3 position.

The final part of this work focused on the methylation preferences of DNMT3B with respect to the flanking bases around the central CpG or CpX site. By comparing the obtained results with the flanking sequence preferences of DNMT3A and using available crystal structures of DNMT3A and DNMT3B in complex with DNA, it was possible to explain the lower CpG specificity of DNMT3B. In addition, lysine 777 of DNMT3B was recognized as a sensor for the +1 flanking base, explaining the increased preference for purines at this position. Furthermore, it could be shown that threonine 775 plays an essential, dynamic role in the catalytic mechanism of DNMT3B. Consistent with the lower CpG specificity of DNMT3B compared to DNMT3A, a high correlation of *in vitro* non-CpG flanking preferences with non-CpG methylation patterns in human cells was also observed. This underlines the biological relevance of the specificity of this enzyme.

List of abbreviations

A	adenine
ADD domain	ATRX-DNMT3-DNMT3L domain
AdoHcy / SAH	S-Adenosyl-L-homocysteine
AdoMet / SAM	S-Adenosyl methionine
BAH	bromo-adjacent homology domain
bps	base pairs
C	cytosine
cat.	catalytic
CD	catalytic domain
CGD	Chromo domain helicase DNA-binding
CGIs	CG islands
CoREST	REST corepressor 1
CXXC	CXXC domain
Cys	cysteine
DMAPD	DNA methyltransferase-associated protein 1-interacting domain
DMEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribonucleic acid
DNMT	de novo DNA methyltransferase
ECM	extracellular matrix
EMSA	Electrophoretic Mobility Shift Assay
ES cells or ESCs	embryonic stem cells
G	guanine
GK repeats	glycine lysine repeats
GLP	G9a-like protein
H	adenine or cytosine or guanine
HDAC	Histone deacetylases
HP1	heterochromatin protein 1
HRP	horseradish peroxidase
ICF syndrome	immunodeficiency, centromeric instability and facial anomalies syndrome
ICR	imprinting control region

iPSCs	induced pluripotent stem cells
ISWI	Imitation Switch
LC-MS	Liquid chromatography–mass spectrometry
LEDGF	lens epithelium-derived growth factor
m/v	mass per volume
mA	methyladenine
MBD	methyl-CpG-binding domain
mC	methylcytosine
MECP2	methyl-CpG-binding protein 2
MLA	Methyl-lysine analog
MNase	Micrococcal nuclease
MPP8	M phase phosphoprotein 8
MTase	methyltransferase
NCBI	National Center for Biotechnology Information
NGS	next generation sequencing
NLS	nuclear localization signal
PBD	PCNA-binding domain
PBS	phosphate buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PHD	plant homeodomain
PKMT	protein lysine methyltransferase
PWWP domain	Pro-Trp-Trp-Pro domain
RFTD	replication foci targeting sequence (RFTS) domain
ROS	reactive oxygen species
SatII	satellite II
SCNT-SCs	somatic cell nuclear transfer stem cells
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SEM	Standard deviation of the mean
T	thymine

TET	ten-eleven translocation
TPE	Tris-Phosphate-EDTA
TRD	transcriptional repression domain (in context of MECP2)
TRD-loop	target recognition domain loop (in context of DNMTs)
UCR	University of California Riverside
UHRF1	Ubiquitin-like containing PHD and RING finger domains, 1
X-ray	Roentgen radiation

1 Introduction

1.1 Basis of epigenetics

Biological research recognized the central role of DNA as the carrier of all genetic information a long time ago and reached its interim peak with the completion of the Human Genome Project in 2000 which decoded the entire human genome with its approximately 3.1 billion base pairs (Piovesan et al., 2019). However, the human body consists of about 200 different types of cells which share the same genetic code (review: Moris, Pina, and Arias 2016), and this (non-genetic) phenotypic and functional diversity is achieved by so-called epigenetic modifications. The term 'epigenetics' was first introduced by the British developmental biologist and geneticist Conrad Hal Waddington, who defined it as 'the study of the interactions between genes and their products which result in a particular phenotype' (Waddington, 1942). For the description of embryonic tissue development and differentiation, this theory was further conceptualized in an illustration known today as the 'Waddington's epigenetic landscape' (**Figure 1**). The first and most popular of the pictures painted by Waddington shows a cell in the form of a marble which has its start at the top of a hill and makes its way due to gravity over existing paths of the landscape. This leads to different outcomes which are illustrated there as valleys. When the cell has been guided in one direction at a fork, it is thereby restricted in its subsequent decisions, which means a reduced cellular potential and a restriction of the fate (**Figure 1A**). With the second picture, Waddington wanted to illustrate that the path taken or the shape of the epigenetic landscape is not random, but is shaped by a genetic regulatory network (**Figure 1B**). Even though these images were painted more than half a century ago, they can still be used today to illustrate alternative cell programming that follows epigenetic mechanisms.

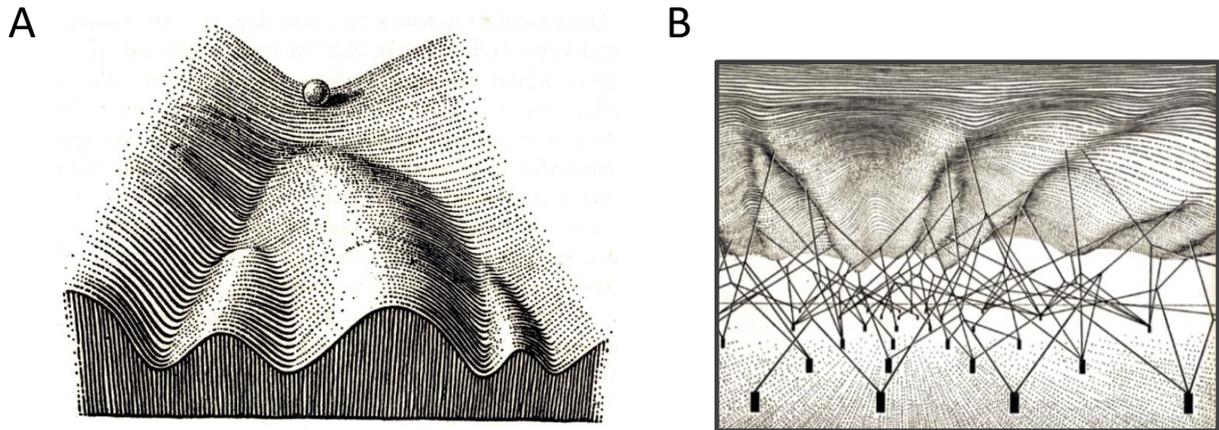


Figure 1 | Waddington's epigenetic landscape. A | Classical view in which the cell, represented as a ball rolling down a hill, has to pass some branching points due to the shape of the landscape, which are considered as development decision points. **B |** Waddington's idea that this landscape is underpinned by the activity of genes forming a regulatory network represented as pegs under the hills and valleys. Both images were adapted from Moris et al. (2016)

Since Waddington, the field of epigenetics has rapidly evolved and one of the more recent definitions of epigenetics is 'the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence' (review: Wu and Morris 2001). This definition is rather up to date, since it is meanwhile known that epigenetic signals contribute to the establishment and stabilization of cellular phenotypes by maintaining gene expression states. This is not realized by base sequence changes, but is achieved by structural changes of the chromatin which packages the DNA. Consequently, an identical set of genes is available in most cells in the human body, but only a small subset of these are active in a differentiated cell type at any given time (review: Allis and Jenuwein 2016).

Three categories of signals have been described for the establishment of a stable heritable epigenetic state and are depicted in **Figure 2** (Berger et al., 2009). The initiating signal, the 'epigenator' originates from the extracellular environment and can be of various types. This signal activates the intracellular pathway in the cytoplasm and acts there as a priming factor for the 'epigenetic initiator' which, as a result, targets a specific chromosomal locus. Arrived there, it is able to trigger a change in the epigenetic environment. To maintain this newly specified state, the 'epigenetic maintainers' then come into effect, which depend on their initiators for their locus-specific recruitment. In short, the maintainers are indispensable for the maintenance of the local epigenetic state, but are not independently able to initiate it.

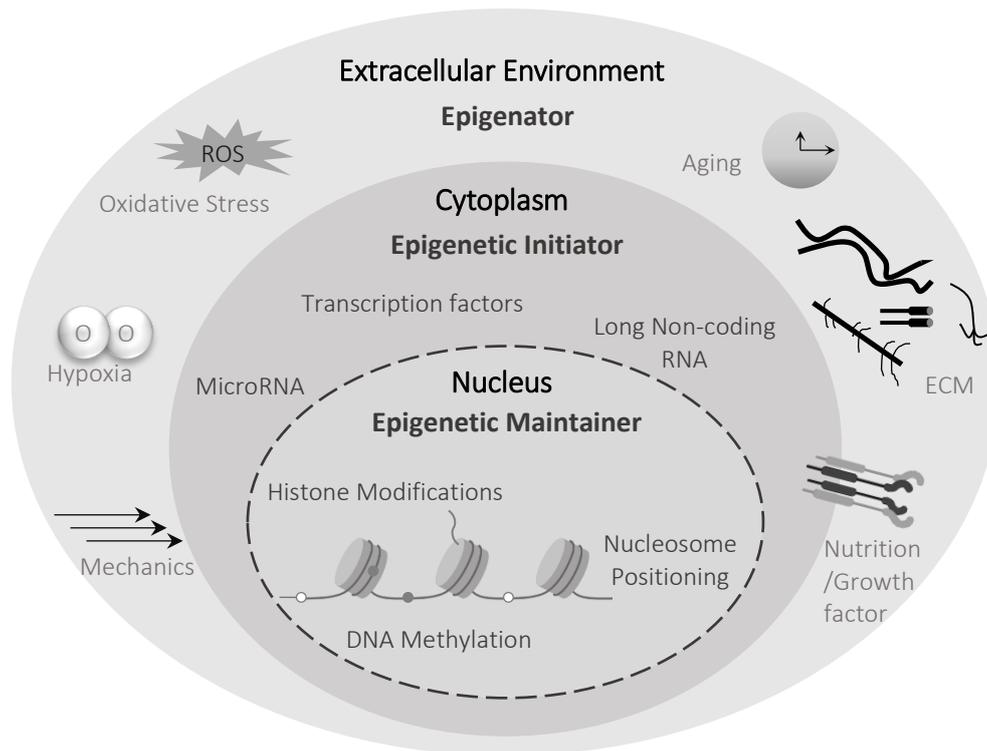


Figure 2 | Key players of the epigenetic signaling pathway. Three different functional elements are required for this pathway: First, the ‘Epigenator’ initiates the epigenetic pathway as an extracellular signal. Then the ‘Epigenetic Initiator’ targets a specific chromosomal locus with its sequence specificity to induce changes in the epigenetic landscape. Last, these changes are adopted there and inherited by the ‘Epigenetic Maintainer’. The image was adapted from Li, Ohliger, and Pei (2014).

Currently, the most studied epigenetic signals are DNA methylation, histone modifications, and non-coding RNAs. Each of these signals dynamically modulates chromatin structure in parallel and thus drives gene expression. Since each of these epigenetic signals represents a separate research area, only parts of the Epigenetic Initiators and Maintainers relevant to this work are introduced in the following sections.

1.2 Chromatin structure

The human genome consists of about 3.1 billion DNA base pairs, which means that the nucleus in every human cell contains DNA molecules which altogether have a length of about 2 meters. The nucleus, the organelle in which this genetic information is stored, has a diameter of only about 10 μm . This limited space in the nucleus requires a high degree of structural organization.

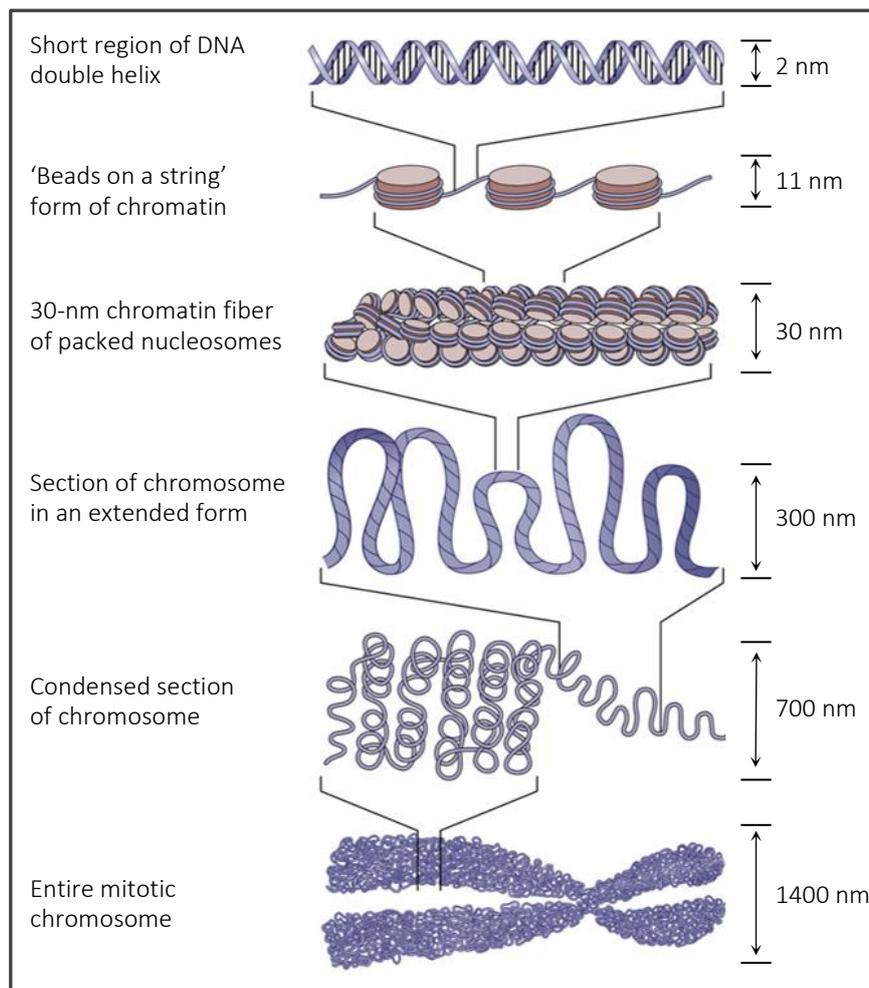


Figure 3 | Higher order packaging of the DNA within the nucleus of a eukaryotic cell. The figure is adapted from Jansen and Verstrepen (2011).

DNA is compacted with the help of specialized proteins and the resulting structure containing DNA and proteins is called chromatin (review: Zhu and Li 2016) (**Figure 3**). The smallest basic unit of the chromatin fiber is the nucleosome. This consists of 147 bp of DNA, which is wrapped around a protein octamer, composed of two copies each of the four core histone proteins H2A,

H2B, H3 and H4, in 1.7 superhelical turns, thus achieving a 7-fold linear compaction of the genomic DNA (Luger et al., 1997) (**Figure 4**).

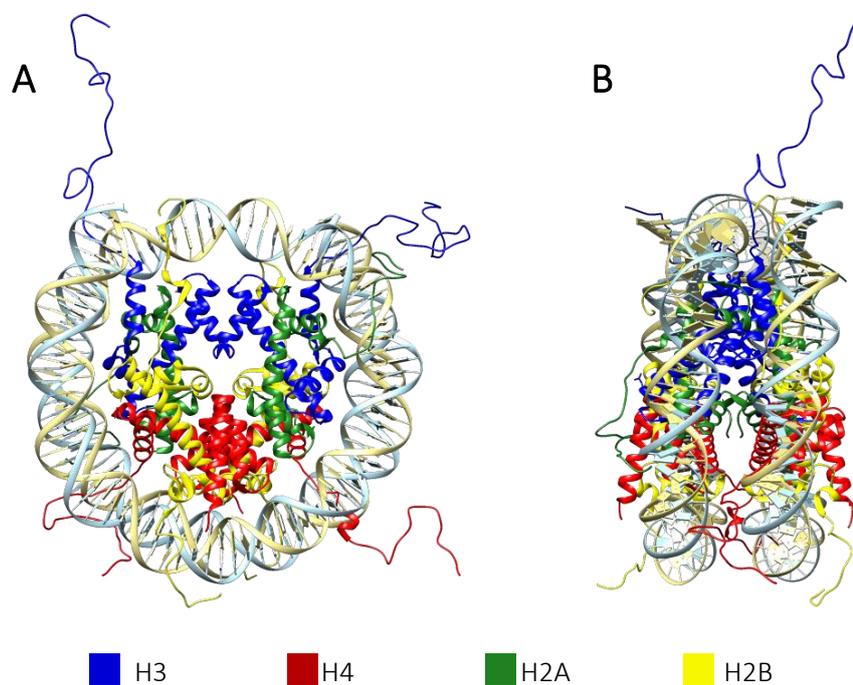


Figure 4| Nucleosome core particle (PDB ID 1KX5) in ribbon structure. The four core histone proteins H3, H4, H2A and H2B and their tails are highlighted in blue, red, green and yellow respectively. **A|** View down the DNA superhelix axis. **B|** View perpendicular to the DNA superhelix axis.

Histones are proteins with a high basicity, which explains their high affinity for DNA with its phosphate backbone. This leads to a very stable structure of the nucleosomes, which makes access of DNA-binding proteins difficult and acts as an inherent barrier for nuclear processes such as transcription, DNA replication and DNA repair (review: Grigoryev and Woodcock 2012). Using X-ray structural analysis, it was shown that per strand and helical turn, two major phosphates make direct contact between the DNA backbone and the core histone proteins. Paired-loop and paired-end-of-helix elements in the histone fold domains form contacts with the DNA helix. In this process, highly conserved arginine residues point toward the minor groove to facilitate DNA positioning, bending, and determine its superhelical shape (Davey et al., 2002; Luger et al., 1997). Through the connection of the individual nucleosomes by linker DNA, they are lined up to the so-called 'beads-on-a-string' structure. Further condensation of the structure is achieved by the connection of the nucleosomes to each other via linker histones (histone H1), resulting in more compact 30-nm fiber (Robinson & Rhodes, 2006). Condensed chromatin has a highly dynamic structure and the electrostatic interactions

between neighboring nucleosomes can either stabilize chromatin fiber formation or weaken it (review: Pepenella, Murphy, and Hayes 2014).

1.3 DNA methylation

1.3.1 Discovery and occurrence

The discovery of DNA methylation is attributed to Rollin Hotchkiss. To separate and quantify the components of DNA, he used paper chromatography and discovered that he could not only separate the four nucleobases thymine, adenine, cytosine and guanine but, to his surprise, also another constituent with a migration rate somewhat greater than that of cytosine, which he denoted as 'epicytosine' (Hotchkiss, 1948). As Hotchkiss had already assumed, epicytosine turned out to be the methylated form of cytosine, making the first description of an epigenetic mark only shortly after the identification of DNA as the carrier of genetic information (Oswald et al., 1944) and even years before its structure was solved (Crick & Watson, 1953). In most cases when DNA methylation is discussed, it is the methylation of the C5 position of the cytosine in the CpG context.

There are approximately 56 million CpG sites in the human genome, of which about 60-80% are methylated, representing 4-6% of all cytosines present (Deaton & Bird, 2011; Laurent et al., 2010; Lister & Ecker, 2009).

Of particular note is the fact that methylated CpG sites represent mutagenic hotspots, as hydrolytic deamination of the 5mC can result in TpG mismatches (review: Jurkowska, Jurkowski, and Jeltsch 2011) and 5mC deaminates up to four times faster than unmodified cytosine. In the case of unmodified cytosine, this deamination leads to the unnatural DNA base uracil, for which the uracil DNA glycosylase forms a special cell repair system that can recognize and correct this error, while in case of 5mC thymidine is formed, which cannot be recognized as DNA damage so easily (Shen et al., 1994). Presumably, it is precisely this particular chemical susceptibility of 5mC that has led to a selective reduction of CpG sites in the mammalian genome during molecular evolution, since, for example, in the human genome CpG dinucleotides are underrepresented by a factor of 5 compared to other dinucleotide combinations (review: Jurkowska et al. 2011). Interestingly, CpG dinucleotides are abundant at repetitive DNA elements as well as at gene regulatory elements, whereas, as the global distribution suggests, they are underrepresented in inter- and intragenic regions overall. For example, in gene promoters they cluster in so-called CpG islands (CGIs). CGIs are 500-2000 bp

long regions that have a GC content of more than 50% and a ratio of observed to expected number of CpG dinucleotides above 0.6 (Gardiner-Garden & Frommer, 1987; Takai & Jones, 2004). In about 70% of all human genes, CpG islands are found in their promoter regions, including most housekeeping and tissue-specific genes (Saxonov et al., 2006). The distribution of 5mC is bimodal. This means that 60-90% of the individual CpG sites (depending on the cell line) that do not occur in clusters are methylated, whereas CpG islands are mainly hypomethylated (review: Deaton and Bird 2011; review: Messerschmidt, Knowles, and Solter 2014). The fact that CGIs in the germline show no or only very low methylation implies a rather low mutagenic drift risk in this stadium, which could explain the survival of these CpG-rich clusters in the mammalian genome (L. Shen et al., 2007; Weber et al., 2007). Even though 5mC methylation is a very stable epigenetic mark, there are dynamic changes in methylation status at CpG sites during cellular differentiation. In the germline, CpG-dense promoters are usually unmethylated, but their methylation can increase during development. This results in a gene repressive state and the cells become more and more specified (Borgel et al., 2010; Meissner et al., 2008), or according to the illustration of Waddington the marble is about to reach the valley. However, DNA methylation not only plays a central role in the regulation of CGIs and repetitive elements, but is also associated with X-chromosome inactivation in females, alternative splicing, enhancer activity and parent-of-origin specific gene expression through imprinting (Anastasiadou et al., 2011; Aran et al., 2013; Hellman & Chess, 2007).

5mC DNA-methylation is not exclusive at CpG dinucleotides but also found at non-CpG sites, such as CpA, CpT and CpC although at a reduced extent. Initially this was long thought to be an artefact of the methodologies used to detect methylation (review: He and Ecker 2015). Reliable detection of non-CpG methylation is difficult due to several features. First, mCpH (H = A, T or C) is not found in most somatic cells but only in few cell types. Moreover, this modification mixes with the extensive genome-wide mCpG, making a high-resolution methylation assay such as bisulfite sequencing necessary to distinguish it from ubiquitous mCpG found in mammalian cells. Due to the fact that for each non-CpG site only a moderate proportion of alleles in the cell population are methylated, the mCpH was initially dismissed as an artefact of incomplete conversion of unmethylated cytosines during bisulfite conversion (Pinney, 2014). This misconception was disproved with the confirmation of the existence of this modification first in mice (Ramsahoye et al., 2000) and subsequently in the human genome (review: Lister et al. 2009). Finally, it was the development of Illumina sequencing that enabled accurate genome-

wide profiling of mCpH with single base resolution in a large number of mammalian cells and tissues (review: Metzker 2010). However, the biological role of mCpH is still unclear and controversial. Investigation of the DNA methylation profile of human ESCs revealed a strong correlation between mCpH and adjacent mCpG methylation, from which it could be concluded that the presence of mCpH may just be the result of nonspecific *de novo* methylation (Ziller et al., 2011).

But there are also observations speaking against this hypothesis. For example, large areas of mCpH hypomethylation were detected in SCNT-SCs and in iPCSs compared with ESCs, corresponding to regions that could neither be reprogrammed at the epigenetic nor the transcriptional level (Lister et al., 2011; Ma et al., 2014). In addition, CpH methylation in the gene body in brain cells is associated with transcriptional repression (Guo et al., 2014) and can enhance the binding of methyl-CpG binding protein 2 (MeCP2) (Guo et al. 2014), a protein critical to the cause of Rett syndrome (review: Chahrour and Zoghbi 2007; review: Guy et al. 2011). This indicates a connection between this type of DNA modification and this neurological disorder. In a mouse model of Rett syndrome, this connection was further tested and MeCP2 binding to mCpH was suggested to be important for the transcriptional regulation of genes related to neurological function (L. Chen et al., 2015; Gabel et al., 2015).

1.3.2 The DNA methyltransferase family

The family of DNA methyltransferases (DNMTs) share a similar catalytic mechanism which is characterized by the formation of a covalent reaction intermediate between the enzyme and the substrate base (review: Lyko 2018). In mammals, this family consists of the four active enzymes DNMT1, DNMT3A, DNMT3B, DNMT3C and the regulatory but inactive factor DNMT3L (**Figure 5**) (review: Jurkowska et al. 2011). The *de novo* methyltransferases DNMT3A and DNMT3B are mainly responsible for the establishment of DNA methylation in embryonic development (Okano et al., 1999). DNMT3C is found only in rodents and is a duplicate of DNMT3B which is responsible for DNA methylation in a subset of repetitive elements in the male germline (Barau et al., 2016; Jain et al., 2017). Dnmt3L is an essential regulatory protein that, although catalytically inactive, contributes as an accessory protein to *de novo* methylation in germ lines (review: Chen et al. 2019). To maintain the methylation pattern set by the *de novo* DNA methyltransferases, which would be lost after DNA replication, the maintenance DNA

methyltransferase DNMT1 comes into play. This methyltransferase has a preference for hemimethylated DNA and maintains the methylation pattern (Hermann et al., 2004).

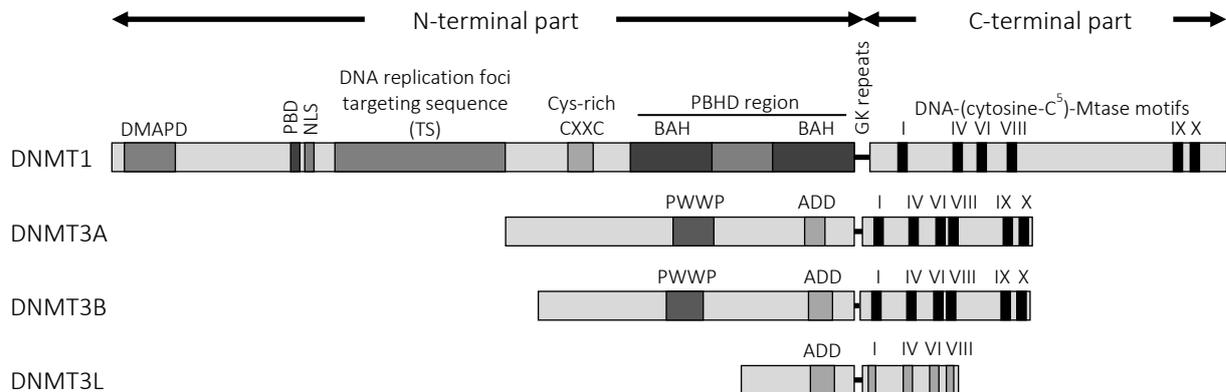


Figure 5] Domain structure of the mammalian DNMT enzymes. The human DNMT1, DNMT3A, DNMT3B and DNMT3L proteins consist of 1616, 912, 853 and 387 amino acid residues, respectively. DMAPD, DNA methyltransferase-associated protein 1-interacting domain; PBD, PCNA-binding domain; NLS, nuclear localization signal; RFTD, replication foci targeting sequence (RFTS) domain; CXXC, CXXC domain; BAH1 and BAH2, bromo-adjacent homology domains 1 and 2; GK, glycine lysine repeats; PWWP, PWWP domain; ADD, ATRX-DNMT3-DNMT3L domain (Figure adapted from Jurkowska et al. (2011)).

From a structural point of view, mammalian methyltransferases can be divided into two parts. A large N-terminal part with different domains that have regulatory functions, and a C-terminal catalytic part. The C-terminal catalytic domains of the two *de novo* methyltransferases DNMT3A and DNMT3B are not restricted in their function when the domains are separated from the rest of the enzymes (review: Gowher and Jeltsch 2001). Structural and catalytic properties of the individual domains are described in more detail in later paragraphs .

1.3.3 The classical model of DNA methylation establishment and maintenance

DNA methylation at CpG sites in mammals was considered the paradigm of epigenetic information transfer for a long time. The classical model of maintenance methylation emerged several decades ago and was proposed in two papers by Riggs and Holiday & Pugh (Holliday & Pugh, 1975; A.D. Riggs, 1975). It implies that DNA methylation is introduced on both DNA strands at palindromic CpG sites by *de novo* methyltransferases, creating a pattern of fully methylated and unmethylated CpG sites. Since DNA replication preserves the methylation pattern only on the parental strand, a mixture of hemimethylated and unmethylated CpG sites is generated. At this point, the maintenance methyltransferase recognizes the hemimethylated CpG sites to remethylate the DNA back to the initially set methylation pattern (**Figure 6**).

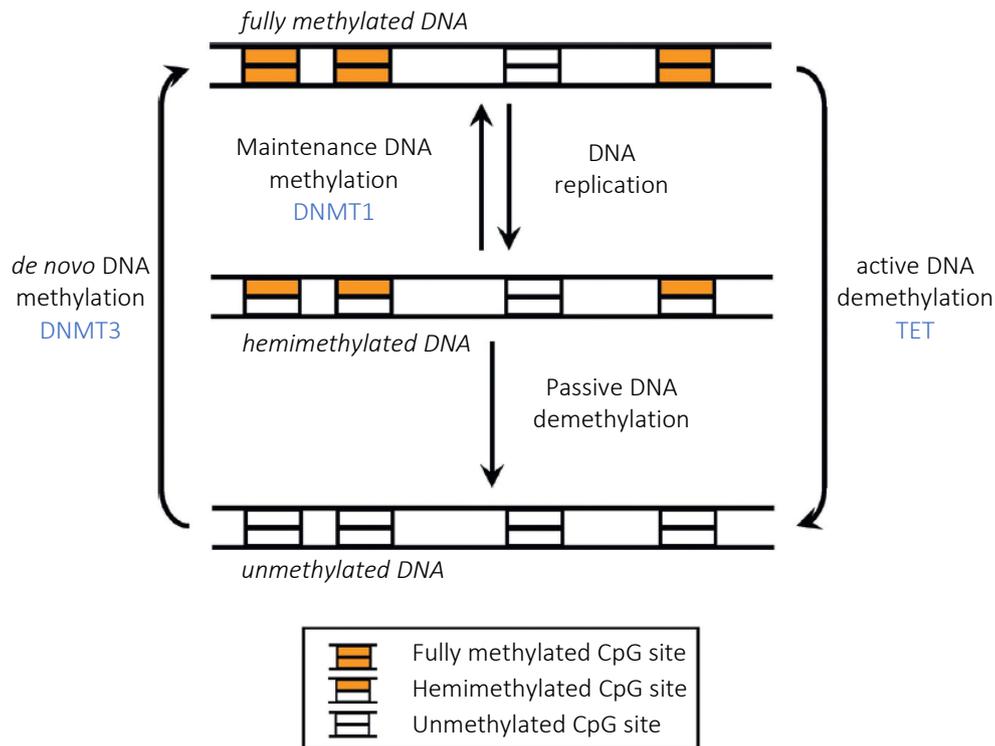


Figure 6| Classical site-specific DNA maintenance methylation model as a schematic drawing. Initial DNA methylation pattern consisting of fully unmethylated and fully methylated CpG sites is generated by *de novo* DNA methyltransferases. Maintenance of the methylation pattern after DNA replication is obtained by maintenance methylation or can get lost through passive demethylation. Active DNA demethylation can be catalyzed by TET enzymes. DNMT1 is considered to be the maintenance methyltransferase, DNMT3A and DNMT3B are regarded as the *de novo* methyltransferases. Abbreviations: DNMT, DNA methyltransferase; TET, ten eleven translocation enzyme. (Figure was adapted from Jeltsch and Jurkowska (2014))

As previously described in brief, each of the active DNA methyltransferases has essential tasks that cannot be completely taken over by another. The *de novo* methyltransferases DNMT3A and DNMT3B are the only known mammalian enzymes capable of creating a methylation pattern on unmethylated DNA thereby taking over the task of establishing the initial methylation pattern (Okano et al., 1999).

In mouse cells, DNMT3A is predominantly expressed in oocytes and early preimplantation embryos and is provided maternally. It plays an important role in determining the different methylation patterns in imprinting control regions (ICRs) in male and female germ cells (Kaneda et al., 2004; Kato et al., 2007). DNMT3B is transcribed in a later stage of embryonic development, during zygotic gene activation, and its expression is mainly observed in the blastocyst stage in the epiblast lineage (Watanabe et al., 2002). Thus, although at first glance the *de novo* enzymes are very similar, they have different functions. This has been shown in different knock-out studies. Whereas *Dnmt3a* knock-out produced partially viable mice, knock-

outs of *Dnmt3b* were lethal (Okano et al., 1999). DNMT3L also plays an essential role in *de novo* DNA methylation, being indispensable under certain conditions despite its lack of catalytic activity. Thus, loss of DNA methylation in the germ cell was observed in *Dnmt3l* knock-out mice, resulting in infertility male mice and embryonic lethality of maternally null-derived embryos (Bourc'his et al., 2001). With each cell division, a round of DNA replication takes place, whereby only one strand of DNA is methylated in the daughter cells. To maintain the DNA methylation pattern, DNMT1 recognizes this hemimethylated DNA as a preferred substrate and adapts the methylation pattern on the newly generated daughter strand to that of the mother strand. The expression of DNMT1 is tightly controlled in S phase by cell cycle dependent transcription factors and leads to high expression in mitotic cells (Inano et al., 2000). Proliferating cell nuclear antigen (PCNA), located at the replication fork, interacts with DNMT1 to recruit it to the vicinity of the hemimethylated DNA, to which DNMT1 is then further mediated via UHRF1 (Ubiquitin-like, containing PHD and RING finger domains, 1). DNMT1 is thus concentrated at sites of active DNA replication and kept away from sites without hemimethylated DNA (Avvakumov et al., 2008; Bostick et al., 2007; Sharif et al., 2007; Sharif & Koseki, 2011).

1.3.4 Revision and adaptation of the classical model of DNA methylation inheritance

Importantly, a wealth of publications today indicates that the 'classical model' of inheritance of the methylation pattern with a clear division into *de novo* methylation and maintenance methylation is somewhat too rigid, despite its logic and elegance. DNA methylation should rather be seen as a dynamic process with continuous methylation and demethylation events, in which the functions of DNMT1 and the DNMT3 enzymes partly overlap (review: Jeltsch and Jurkowska 2014; review: Jones and Liang 2009; Riggs and Xiong 2004). The revision of the 'classical model' was triggered by experimental observations that could not be reconciled with model reinforced, as outlined below.

At first, studies in mammalian cell lines and mice showed that the deletion of the *de novo* DNA methyltransferases DNMT3A and DNMT3B resulted in loss of DNA methylation at repetitive elements. These observations argue against, that a fully functional DNMT1 alone is capable of maintaining DNA methylation alone and also attribute some maintenance function to the *de novo* methyltransferases (T. Chen et al., 2003; Dodge et al., 2005; Egger et al., 2006).

Additionally, the classical model of inheritance would predict that identical DNA methylation patterns should be present in all cells of a tissue. However, this hypothesis could not be confirmed using bisulfite conversion and subsequent sequencing to elucidate the methylation status of individual CpG sites. Instead, analysis of the data revealed that the average methylation density profile of a given DNA region is conserved (Y. Zhang et al., 2009). These results are consistent with biochemical studies on DNMT1, in which a 10-40-fold preference for hemimethylated CpG sites was obtained. Although this is quite remarkable, it is not sufficient for an exact copying of the methylation status of all approximately 56 million CpG sites of the human genome (Bashtrykov, Jankevicius, et al., 2012; Bashtrykov, Ragozin, et al., 2012; Song et al., 2012).

Furthermore, work on human stem cells and differentiated neuronal progenitors showed methylation also at asymmetric sequences predominantly in a CpA context. For example, in induced pluripotent stem cells, in addition to 5mC methylation at about 67% CpG sites, more than 7% of CpA about 2% of CpT and roughly 1% of CpC sites had a 5mC mark (Barrès et al., 2009; Guo et al., 2014; W. Guo et al., 2014; Lister et al., 2013; Pinney, 2014). In somatic cells, the proportion of non-CpG methylation is only about 0.02% of the total 5mC methylation whereas in human male ES cells it can be up to 25% (Jang et al., 2017). This non-CpG methylation correlates with the expression levels of DNMT3A and DNMT3B, the two enzymes shown to introduce 5mC in a non-CpG context (Arand et al., 2012; Shirane et al., 2013). Since template strand information cannot be used for non-palindromic methylated sequences and thus these do not represent a substrate for DNMT1, the stable presence of CpA methylation in different cell types suggests a permanent presence of *de novo* DNMTs (review: Jeltsch and Jurkowska 2014).

Last but not least, the discovery of TET enzymes and their ability to actively demethylate CpG sites has highlighted the need for *de novo* methyltransferases as their counterparts to maintain or re-establish the methylation pattern (Arand et al., 2012; Métivier et al., 2008).

Standing alone, these individual arguments would only stand as critical questioning of the 'classical model' of DNA methylation inheritance, but taken together they form a strong body of arguments in favor of an expanded and revised model, in which the importance of regulatory factors in the targeting and control of the DNA methylation machinery is more strongly considered (review: Jeltsch and Jurkowska 2014).

In the revised model of stochastic DNA methylation inheritance, DNA methylation at each individual possible methylation site is determined by local rates of methylation and demethylation. The above rates are thereby dependent on the targeting and regulation of methyltransferases and demethylases, implying that DNA methylation is controlled by a complex network of epigenetic modifications. Other chromatin marks or their interaction with the aforementioned enzymes form a second level of feedback, leading to the stable establishment of larger repressed and activated chromatin domains. This unified approach allows to explain *de novo* and maintenance methylation, as well as non-CpG methylation in human cells and other organisms. It implies that not single methylation events are stably inherited, but the combined methylation state of larger DNA regions (review: Jeltsch and Jurkowska 2014).

1.3.5 Catalytic function of the C-terminus of the DNMTs

The C-terminal part of cytosine C5 DNA methyltransferases is conserved between prokaryotic and eukaryotic DNMTs and harbors the active site of the enzyme for the transfer of the methyl group of the cofactor S-adenosyl methionine (AdoMet) to the C5 position of cytosine (Xiaodong Cheng, 1995). This domain has an AdoMet-dependent MTase fold and contains 10 conserved amino acid motifs typical for all cytosine C5 DNMTs. It is constructed from a mixed seven-stranded sheet consisting of 6 parallel β -strands and an antiparallel β -strand which is inserted between strands 5 and 6. Motifs I and X are involved in cofactor binding, whereas motifs IV, VI, and VIII are responsible for catalysis. The non-conserved region between motifs VIII and IX plays a role in DNA recognition and is partly responsible for the specificity of each methyltransferase (review: Jurkowska, Jurkowski, et al. 2011). DNMT3L lacks motifs IX and X, making this protein enzymatically inactive (Bourc'his et al., 2001). The first crystal structures of the bacterial C5 methyltransferases HhaI and HaeIII in complex with DNA revealed for the first time one of the most interesting structural and mechanistic features of DNMTs, namely their mechanism of cytosine base modification (X. Cheng et al., 1993). Here, base flipping of the target cytosine out of the DNA helix occurs and the flipped base is introduced into a hydrophobic pocket in the active site, a process which is conserved among all DNA MTases (review: Jeltsch 2006).

In detail (**Figure 7**), a nucleophilic attack of the catalytic cysteine in the PCQ motif (motif IV) to the C6 position of the cytosine occurs, forming a covalent bond between the enzyme and the cytosine target base and deprotonating the glutamic acid in the ENV motif (motif VI). Subsequently, addition of the methyl group of the cofactor AdoMet to the C5 position can occur, returning glutamic acid to its protonated state. Finally, the covalent bond between the enzyme and the DNA is released via an elimination reaction with an unknown base deprotonating the cytosine at the C5 position and the cofactor is released from the reaction as S-adenosyl-L-homocysteine (AdoHcy) (review: Jurkowska, Jurkowski, et al. 2011).

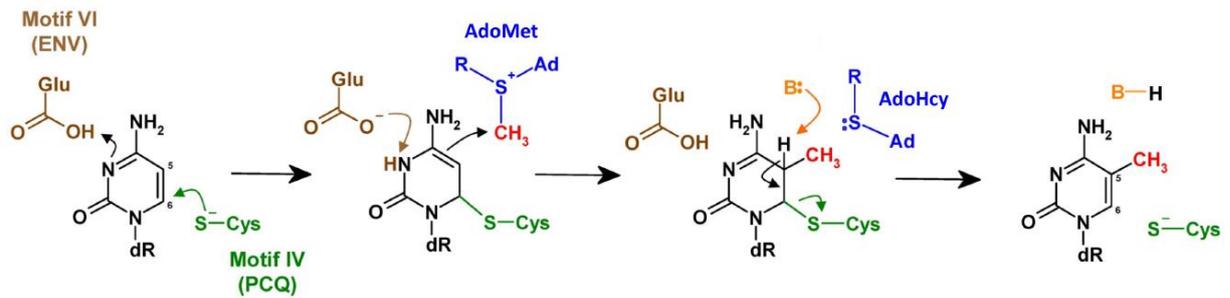


Figure 7 | Conserved catalytic mechanism of cytosine C5-DNA methyltransferases. The catalytic residue from the PCQ motif (motif IV) is colored in green, the brown colored ENV motif (motif VI) contains the catalytically important glutamate residue. The transferred methyl group is highlighted in red and initially bound to the cofactor AdoMet. The base for final proton abstraction (shown in orange) is not identified. (Figure taken from Gowher and Jeltsch (2018))

The changes due to the addition of the single methyl group at the C5 position initially appear very subtle, as the Watson-Crick base pairing is not altered in the modified base; however, the presence of the hydrophobic group within the major groove of the DNA has been shown to cause a slight bending and spoiling effect in crystal structures of methylated DNA oligonucleotides (Tippin & Sundaralingam, 1997). The 5mC signal also does not appear to be as chemically versatile as the PTMs of histone proteins, but its involvement in gene regulation is multifactorial. First, it depends on the number, status, and genomic position of the modified CpG sites (review: Reddington, Pennings, and Meehan 2013), and it is mediated by the recruitment of methylation-sensitive proteins and the binding of transcription factors to the methylated DNA (Machado et al., 2015; Patel, 2016).

1.3.6 Substrate recognition and catalytic preference of *de novo* DNMTs

As already mentioned, DNA methylation occurs mainly at the C-5 position in the CpG context but it was also observed that not all CpG sites are methylated to the same extent. The *de novo* DNMTs were shown to preferentially methylate the different flanked CpG sites in differing quantities. In particular, this was shown in the early 2000s for the first time, when the group of Chih-Lin Hsieh methylated plasmid DNA in vitro and analysed the methylation of each CpG site. This analysis revealed, that CpG sites flanked by pyrimidines are more preferred by murine DNMT3A than CpG sites flanked by purines (I. G. Lin et al., 2002). A few years later, first systematic experimental kinetic studies revealed more detailed flanking sequence preferences for both *de novo* DNMTs and a statistical analysis of previously published data from the human epigenome project was used to filter out the first flanking sequences that were either highly

methylated or had characteristically low methylation-levels in human cells (Handa & Jeltsch, 2005). These observations led to the question of how the *de novo* methyltransferases recognize their substrate and what determines their catalytic preferences.

1.3.7 Role of DNMT3 enzymes in non-CpG methylation

DNMT3A and DNMT3B were identified to be responsible for non-CpG methylation in embryonic stem cells and neurons. Moreover, the interaction of DNMT3B in ESCs with histone H3 lysine 36 in trimethylated state resulted in hyper-methylation at CpGs of actively transcribed genes, including those involved in embryo development (Lee et al., 2017). In summary, the presence of mCpG in different mammalian cell types and its possible functional implications is another challenge for the field of epigenetics to discover.

Furthermore, it must also be mentioned that other DNA methylations exist. In addition to 5-methylcytosine (5mC), N6-methyladenine (6mA) as well as N4-methylcytosine (4mC) have also been found in prokaryotes (Blow et al. 2016), and 6mA was also detected in some unicellular eukaryotes such as ciliates and green algae long ago, however, their biological significance of these modifications in these organisms remained unexplored for a long time (Luo et al., 2015).

Another long undiscovered modification in eukaryotes, is the toxic lesion 3-methylcytosine (3mC) which has been shown to be set by the DNA methyltransferases both in vitro and in vivo (Rošić et al., 2018). However, regulatory functions of this methylation are inconceivable because this mark disrupts DNA base pairing and the ALKB2 family may have co-evolved with the active DNMTs only for the repair of this toxic lesion in many species (Rošić et al., 2018). It can be seen that DNA methylation varies greatly in type and function in different organisms, certainly in mammals it plays a fundamental role in embryonic development as well as in adult tissue homeostasis (review: Messerschmidt et al. 2014).

1.3.8 Recruitment of DNMTs and interaction with chromatin modifiers

DNA methylation is known to influence the recruitment of transcriptional regulators (review: Laisné et al. 2018). On the contrary, it is also the case that transcription factors bound to DNA are involved in direct recruitment of DNA methyltransferases. The oncogenic transcription factor PML-RAR was the first to be reported to recruit DNMT1 and DNMT3A (Croce et al., 2003), but it did not take long till the list got extended, for example the recruitment of DNMT1 by p53 to silence the SURIVIN promoter (Chin et al., 2005) or the recruitment of DNMT3A by

MYC to silence p21/CDKN1A (Brenner et al., 2005). After that, many examples of cellular or viral transcription factors recruiting DNMTs to promoters via indirect interactions have been described and most of these concern the recruitment to promoter regions (review: Laisné et al. 2018). However, recruitment of DNMTs is not exclusive to promoter regions, as exemplified by the zinc finger protein ZBTB24 which recruits DNMT3B to specific regions in gene bodies (Thompson et al., 2018). Interestingly, both enzymes, DNMT3B and ZBTB24, are associated with autosomal recessive immunodeficiency, centromeric instability and facial anomalies syndrome (ICF). Roughly 50% of patients carry mutations in the DNA methyltransferase 3B (*DNMT3B*) gene (ICF1) while the remaining patients carried unknown genetic defects which partially could be addressed to mutations in the zinc-finger and BTB (bric-a-bric, tramtrack, broad complex)-domain-containing 24 (*ZBTB24*) gene (De Greef et al., 2011; Hansen et al., 1999).

In addition to the recruitment via transcription factors, there are also other interaction partners that contribute to the *de novo* methyltransferases reaching their destinations. Di- and Trimethylation of histone H3 lysine 9 (H3K9) is another major repressive epigenetic modification besides CpG DNA methylation, these two marks correlate positively with each other in chromatin although they have not been shown to interact directly with each other. With this in mind, it seems obvious that direct or indirect links of CpG methylation and this histone modification must exist. Indeed there are a couple of indirect connections of these two marks, for example the histone-lysine methyltransferase SETDB1, which methylates lysine 9 of Histone H3 until the trimethylated state, was shown to interact with DNMT3A via its plant homeodomain (PHD) zinc finger and therefore contributes to gene silencing (H. Li et al., 2006). Furthermore, another protein lysine methyltransferase (PKMT), namely G9a or G9a-like protein (GLP), which mono- and dimethylates lysine 9 from histone H3, was shown to methylate also murine DNMT3A on lysine 44 as a non-histone target. This modification of DNMT3A is recognized by the chromodomain of M-phase phosphoprotein 8 (MPP8) which forms dimers in solution and interacts with GLP. Thus, the MPP8 protein provides indirect linkage via GLP of H3K9 methylation and CpG DNA methylation and thereby recruitment of DNMT3A to chromatin (Chang et al., 2011). Moreover, H3K9me3 binds heterochromatin protein 1 (HP1), which, in addition to its primary function of transcriptional repression and heterochromatin formation and maintenance, also recruits DNA methyltransferase 3B to major satellite repeats on pericentric heterochromatin (Lehnertz et al., 2003).

1.3.9 Regulatory functions of the N-terminus of the *de novo* DNMTs

Both *de novo* DNMTs contain two conserved functional domains in their N-terminal part. One is the ATRX-DNMT3-DNMT3L (ADD) domain, and the other is the Pro-Trp-Trp-Pro (PWWP) domain. Towards the end of the N-terminus, both also contain a non-conserved region that is able to bind DNA and presumably contributes to the closer association of the enzymes with chromatin (Suetake et al., 2011). The role of this region is not yet known exactly. Structurally, DNMT3A and DNMT3B have only a low sequence identity in this part of the enzymes, which could lead to the conclusion that they might have different genomic targeting functions (Rondelet et al., 2016).

1.3.9.1 PWWP-domain

The structural organization of this domain consists of an antiparallel β -barrel-like fold formed by five β -strands (β 1- β 5). A short α 1-helix (η 2) is found between strands β 4 and β 5, and a short insertion motif is present between strands β 2 and β 3 (η 1). In both, DNMT3A and DNMT3B, the conserved Pro-Trp-Trp-Pro motif (where the domain got its name) has the first proline replaced by a serine. This position in the PWWP motif regulates the domain stability and oligomerization (Hung et al., 2015). For chromatin targeting of DNMTs, the PWWP domain plays a quite important role, as an S270P exchange in the aromatic pocket of DNMT3B abolishes H3K36me3 binding, resulting in decreased methylation levels at human pericentromeric repeats. In addition, using different PWWP pocket mutants within DNMT3A, partial loss of heterochromatin localization of the full length DNMT3A enzyme was observed in mouse fibroblasts (T. Chen et al., 2004; Dhayalan et al., 2010; Ge et al., 2004). In the crystal structure of the DNMT3B PWWP domain in complex with an H3K36me3 peptide (Rondelet et al., 2016), it can be seen that the trimethylammonium group of the peptide is introduced into a Phe-Trp-Trp-Asp cage, which is stabilized by intermolecular hydrogen bonds between the domain and the peptide backbone. In addition to binding of the K36 methylated H3-tail, both DNMT3A and DNMT3B were reported to have a relatively weak nonspecific interaction with DNA via a basic surface adjacent to the histone binding site (Purdy et al., 2010; Qiu et al., 2002). A positive correlation was found between H3K36me3 and DNA methylation by DNMT3B (Baubec et al., 2015). Consistent with this is also the biochemical comparison data of DNMT3A wildtype with the PWWP pocket mutants in which the wildtype is seen to be more active on native chromatin than the mutants (Dhayalan et al., 2010; Y. Zhang et al., 2010).

1.3.9.2 ADD-domain

The ADD-domain was first identified as a 'Cys-rich domain' in the ATRX protein (Gibbons et al., 1997). When DNMT3A and DNMT3B were discovered, a structurally very similar Cys-rich region was found in these proteins as well (Okano et al., 1998; Xie et al., 1999), and two years later with the discovery of DNMT3L it was also found in its N-terminal part (Aapola et al., 2000). Since no other human protein has a similar region, the name 'ADD' for ATRX-DNMT3-DNMT3L was chosen for this domain.

The structure of this domain consists of three subdomains. In the first subdomain four cysteine (Cys) residues bind a zinc ion to form a GATA-like zinc finger (GATA stand for the corresponding DNA bases). The second subdomain contains eight Cys residues binding two zinc ions which form a PHD finger with a cross-braced topology that is packed against the GATA-like finger. A long C-terminal α -helix that runs out from the PHD finger enters into intensive hydrophobic contact with the N-terminal GATA finger and forms the third subdomain (review: Jeltsch and Jurkowska 2016). Regulation of DNMT3 proteins is largely dependent on the ADD domain, as it specifically binds to the unmethylated lysine 4 of the H3 tail and also plays a role in protein-protein interactions (X. Guo et al., 2015; Ooi et al., 2007; Otani et al., 2009; Y. Zhang et al., 2010). However, when this H3-tail is di- or trimethylated, or acetylated at lysine 4, binding of DNMT3s no longer occurs (Otani et al., 2009; Y. Zhang et al., 2010). For DNMT3A, it was shown that binding to unmethylated H3K4 not only fulfills regulatory or chromatin targeting functions, but that this binding induces a structural change that directly affects the enzymatic activity of the enzyme (X. Guo et al., 2015). This same publication describes how, in DNMT3A, the structural arrangement of the ADD and catalytic domains creates an 'autoinhibitory' conformation. The two domains fold into two individual structural modules connected by a linker that is packed against a hydrophobic surface region of the CD domain (**Figure 8**). The amino acids 526-531 form a loop region which protrudes from the ADD domain and docks into a pocket of hydrophobic residues on the catalytic domain. This brings three acidic residues (D529, D530, and D531) and three hydrophobic residues (Y526, Y528, and Y533) from the ADD domain into the proximity of the DNA binding site, which blocks the binding of DNMT3A, inhibiting catalysis. However, upon binding of unmodified H3K4-peptide, this intramolecular linkage is reversed and a conformational rearrangement takes place, making the DNA binding site freely accessible again (X. Guo et al., 2015). Consequently, a stimulatory effect on

enzymatic activity of DNMT3A was obtained with an unmodified H3K4 peptide, whereas no stimulation was observed with the identical peptide sequence with trimethylated H3K4 (X. Guo et al., 2015). For several genome-wide studies in which an anti-correlation of H3K4me3 and DNA methylation was observed (Hodges et al., 2009; Meissner et al., 2008; Weber et al., 2007), this biochemical study could provide an explanation. H3K4me3 labeling is associated with active gene promoters and plays a key role in mammalian gene expression (Bernstein et al., 2012). It is found on many gene promoters of active genes, where the adjacent CpG islands are usually free of DNA methylation (Edwards et al., 2010; Meissner et al., 2008; Mikkelsen et al., 2007) indicating that H3K4me3 and 5mC are opposing epigenetic signals mediated in part by the ADD domain of the DNMT3 enzymes.

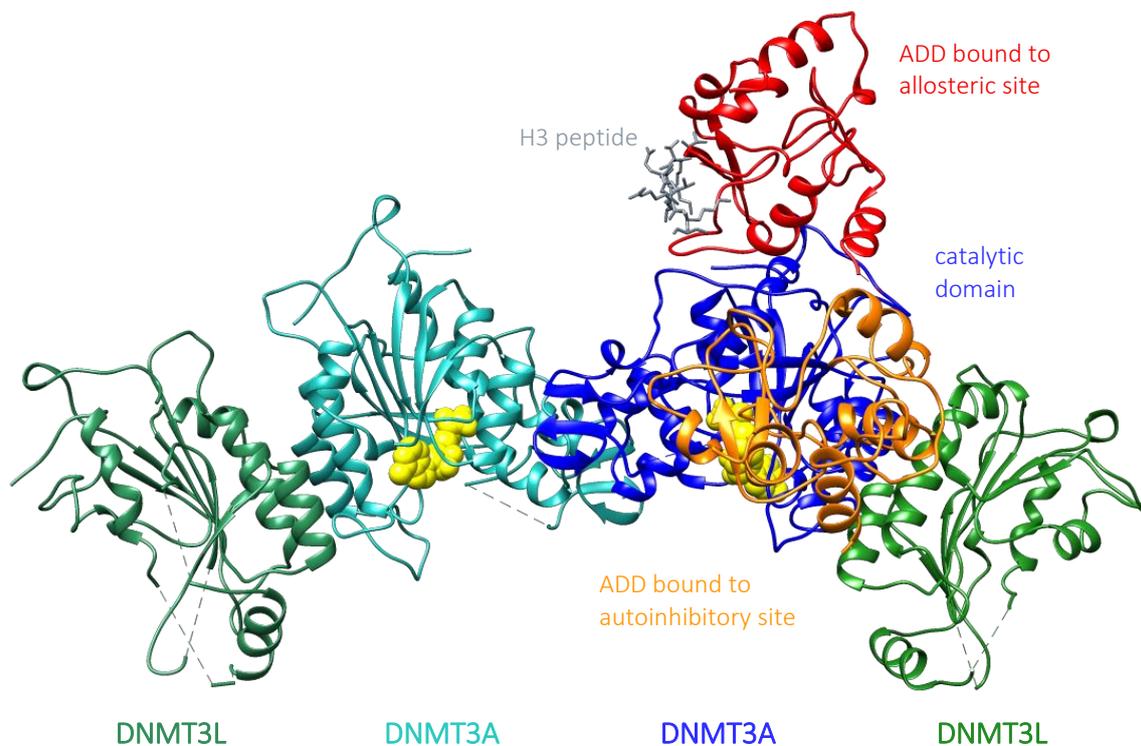


Figure 8 | Structure and allosteric regulation of DNMT3A. The picture shows the DNMT3A/3L heterotetramer. The autoinhibitory conformation is shown on the dark blue DNMT3A subunit with the ADD-domain in orange in close proximity to the catalytic pocket and the catalytically active conformation with the ADD domain in red bound to the H3 peptide (gray). For the binding of the H3 peptide the same residues are involved as for the binding of the ADD domain to the catalytic domain in the autoinhibitory conformation. The ADD domain of the cyan DNMT3A subunit has been omitted for clarity. The yellow sphere structure shows the cofactor SAH after methyl group transfer. (Figure adapted from Jeltsch and Jurkowska (2016), superposition of DNMT3A-DNMT3L complex (PDB:4U7P) on the DNMT3A-DNMT3L structure in complex with histone H3 (PDB: 4U7T))

This theory was experimentally supported by the exogenous introduction of DNMT3B into the model organism *Saccharomyces cerevisiae*. This yeast organism is well suited for this study because it lacks endogenous DNA methylation, but possesses conserved histone sequences

and many post-translational modifications that are also present in higher eukaryotes (Hu et al., 2009). With this setup, an anti-correlation between H3K4me3 levels and introduced 5mC was observed (Morselli et al., 2015).

When *set1*, the gene encoding the protein lysine methyltransferase responsible for the H3K4me3 mark in yeast, was deleted from the yeast genome in the same setup, this led to a spreading of DNA methylation into promoter regions (Morselli et al., 2015). For proper cellular differentiation and development, the results presented highlight the importance of a regulated DNMT3 activity. In **Figure 9** a brief summary illustration of the regulation of DNMT3 enzymes is presented.

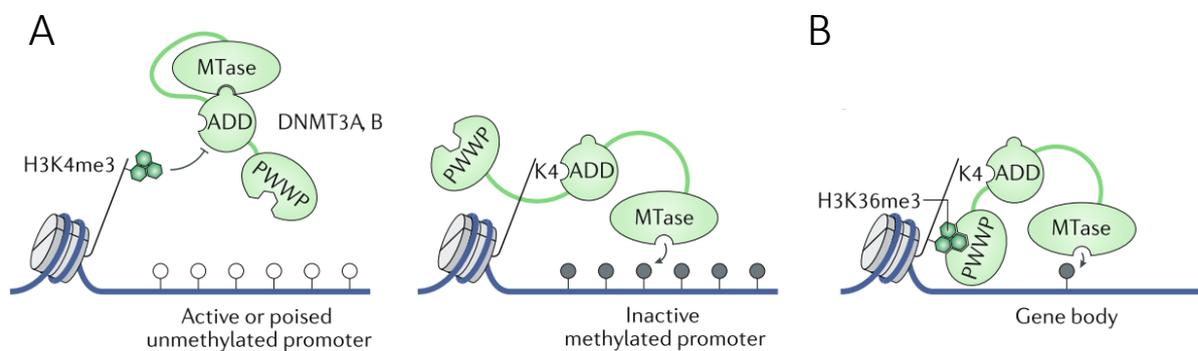


Figure 9 | DNA methylation by DNMT3A and DNMT3B. A | At promoters. Left: active or poised, marked by H3K4me3, with this modification preventing the binding and activating of the DNMTs. Right: inactive promoter, marked by unmethylated H3K4, where the DNMTs bind via the ADD-domain to H3K4 and are forced in the active conformation, thereby allowing the MTase domain to methylate the DNA. **B |** At gene bodies. H3K36me3 is a mark in active transcribed genes which recruits the DNMT3 enzymes via the PWWP domain, due to the close proximity the ADD domain is then also binding to the unmethylated H3K4 and the enzymes are forced in the active conformation and thereby methylate the DNA. (Figure taken from Greenberg and Bourc’his (2019))

1.3.10 Effects of DNA methylation

Methylation of DNA has only a minor effect on its structure, but two other regulatory mechanisms are based on this modification. As already mentioned earlier, DNA methylation can affect the binding properties of transcription factors and furthermore this mark can serve as a binding site for 5mC binding proteins. These proteins serve to recruit downstream effectors like for example chromatin remodelling complexes (Patel, 2016). Proteins capable of identifying fully methylated CpG sites are part of the methyl-CpG-binding domain (MBD) protein family (Du et al., 2015). Due to its relevance in this work, the MBD family member methyl-CpG-binding protein 2 (MeCP2) will be described briefly in the following section.

MeCP2 is a ubiquitously expressed chromatin architecture protein that nearly reaches the expression levels of histone octamers in postmitotic neurons (review: Ausió, de Paz, and

Esteller 2014). The *MECP2* gene is encoded on the long arm of the X chromosome and therefore also exposed to X chromosome inactivation (Amir et al., 1999). Mutations in this gene or aberrations in its expression are associated with the Rett syndrome, a severe neurodevelopmental disorder (review: Ausió et al. 2014). The protein structure of MeCP2 consists of six domains, the N-terminal domain, the methyl-binding domain, the intervening domain, the transcriptional repressor domain and the two C-terminal domains α and β (Martínez de Paz & Ausió, 2017). Among these domains, the TRD provides an interface for protein-protein contacts which arranges contact with numerous co-repressor complexes. One of the best characterized interactors of MeCP2 which dock on the TRD is the SIN3A/HDAC co-repressor complex.

This recruitment results in local changes on nucleosomes in the form of deacetylation, which in further stages leads to gene suppression.(P. L. Jones et al., 1998; Nan et al., 1996). Other well-known MeCP2 interacting proteins are among others the CoREST (co-repressor of elements-1-silencing transcription factor) complex, HP1 and ATRX. The common feature of these interaction partners is that they all have important roles in different gene silencing processes (Martínez de Paz & Ausió, 2017). With this in mind, it is not surprising that MeCP2 was classically considered a transcriptional repressor. However, when transcriptional profiles were generated from brain cells isolated from *Mecp2*-null mice, surprisingly, only a relatively low fraction of genes with changed expression was detected, despite the high level of MeCP2 expression and its interaction with some important epigenetic regulators (Tudor et al., 2002). However, when gene expression profiles of specific brain regions, such as the cerebellum and hypothalamus, were obtained, thousands of genes with altered expression were detectable and most of these genes were conspicuously down-regulated when MeCP2 was not present (Ben-Shachar et al., 2009), which may lead to the conclusion that the protein may also act as a transcriptional activator (Ben-Shachar et al., 2009; Sugino et al., 2014). Thus, MeCP2 seems to have an activating or inhibitory effect on gene expression depending on the chromatin locus. Although MeCP2 is a prominent reader of DNA methylation, there have been few studies investigating a possible direct interaction of the protein with the DNA methylation machinery. However, one of the few studies has shown that DNMT1 interacts directly with MeCP2 via its TRD domain and may support maintenance methylation (Kimura & Shiota, 2003). In addition, further studies showed that MeCP2 may not only be a reader for 5mCpG sites, but may also recognize the *de novo* methyltransferases established 5mCpH (H=A,T and C) modifications and

the hydroxy-methylated CpH modifications (Feng et al., 2011; Gabel et al., 2015; W. Guo et al., 2014; Hashimoto et al., 2012; Kinde et al., 2015). Although both, DNMT3A and MeCP2 are highly expressed in the human brain and both are members of the same epigenetic pathway, no studies have yet shown a direct connection of the two.

1.4 Post-translational modifications of histone tails (PTMs)

The regulation and recruitment of a variety of different nuclear proteins is driven by different modifications on the histone tails (review: Bannister and Kouzarides 2011). Part of these nuclear proteins which depend on this regulation and recruitment are *de novo* methyltransferases. To date, a variety of different histone tail modifications have been described, including methylation, phosphorylation, acetylation and ubiquitination (Y.-M. Xu et al., 2017). One of the best known histone tail modification to date is the methylation of lysine residues, which can directly affect gene expression (review: Greer and Shi 2012). The canonical positions for lysine methylation contain a total of eight residues on histones, of which six residues are on histones H3 (K4, K9, K26, K27, K36, and K79), K20 of histone H4, and K26 of the linker histone H1. With the exception of K79 on H3, all are localized on the N-terminal tails of the histone proteins (Musselman et al., 2012). Two of the lysines on histone H3 are directly related to *de novo* DNA methylation depending on their methylation state, H3K4 and H3K36. Depending on the methylation state of H3K4, different functions are attributed to this lysine which are also associated with different chromatin states. In the di- and trimethylated state these modifications are associated with actively transcribed chromatin regions, such as promoters and transcriptional start sites (Ernst et al., 2011; Liang et al., 2004; Mikkelsen et al., 2007; Schneider et al., 2004). In the monomethylated state, however, this histone modification is a mark of poised or active enhancers (Calo & Joanna, 2013; Ernst et al., 2011).

The unmethylation status of this lysine residue is the relevant form for the *de novo* methyltransferases. In 2010, histone H3 peptide (aa 1-21) with unmodified lysine 4 was shown stimulate DNMT3A and with increasing methylation status of this lysine residue the stimulation was lost (Y. Zhang et al., 2010). A structural explanation for the increased activity was provided by different crystal structures of DNMT3A, one in complex with and one without the H3 histone peptide (X. Guo et al., 2015). In this publication it was shown that DNMT3A exists in two different conformations. An autoinhibitory one which switches into an active conformation by binding to unmodified histone H3 with unmethylated lysine 4.

The other histone modification associated with the *de novo* DNA methyltransferases DNMT3A and DNMT3B is H3K36me3. Recruitment of the two enzymes by this modification then leads to DNA methylation in gene bodies and thereby silencing these genes (Baubec et al., 2015; Dhayalan et al., 2010). The solved crystal structure of the LEDGF PWWP domain in complex with H3K36-methylated nucleosome provided an initial visualization of how PWWP domain-containing proteins are bound to their chromatin targets (Wang et al., 2020).

2 Principal aims of the study

DNA methylation has emerged over the last decades as one of the most influential epigenetic mark. The introduction of this mark occurs through the *de novo* methyltransferases DNMT3A and DNMT3B. Therefore, better knowledge of the enzymology of mammalian methyltransferases is essential. This includes targeting and regulation as well as an understanding of how they set their preferred methylation patterns.

The aim of this thesis was to gain novel and detailed insights into the interactions of the *de novo* DNA methyltransferase DNMT3A with chromatin and furthermore to investigate mechanistic details of the catalytic domains of both *de novo* DNA methyltransferases DNMT3A and DNMT3B.

2.1 Investigation of the chromatin binding of the PWWP domain of DNMT3A

PWWP domains in general have been described as potential protein/protein interaction domains (Stec et al., 2000). However, they do not only interact with proteins, since for the PWWP domains of the DNMT3s nonspecific binding to DNA had been described as well (T. Chen et al., 2004; Purdy et al., 2010; Qiu et al., 2002). Furthermore, it had been shown that the PWWP domains of both *de novo* DNA methyltransferases are essential for chromatin targeting and heterochromatic localization (T. Chen et al., 2004; Ge et al., 2004). For the PWWP domain of DNMT3A was shown in vitro, that it binds the N-terminal tail of histone H3 when lysine 36 is trimethylated (Dhayalan et al., 2010). This shows that the PWWP domains are important for the targeting of methyltransferases, but further mechanistic insights are needed to understand their interaction with chromatin in more detail. This led to the aim of the study, to investigate the interaction of the DNMT3A PWWP domain with chromatin, its natural substrate, to gain deeper insights into the targeting and localization of DNMT3A.

2.2 Investigation in the allosteric regulation of DNMT3A by MeCP2 and H3-tail peptides

The ADD domain had been shown to be a major allosteric regulator of DNMT3A. Binding of the domain to unmodified histone H3 lysine 4 releases DNMT3A from its autoinhibitory conformation into its active conformation where the catalytic site of DNMT3A is accessible for DNA (X. Guo et al., 2015). Previous, at this time, unpublished work in the workgroup of Prof. Jeltsch identified the 5mC-reading protein MeCP2 as a strong interactor of DNMT3A and

mapped the interaction interface to the ADD domain of DNMT3A and the target recognition domain of MeCP2 (MeCP2-TRD). Interestingly, this interaction of MeCP2 with the ADD domain of DNMT3A led to an inhibition of the enzymatic activity.

The aim in this project was to validate the direct interaction of these two proteins and exclude indirect interaction mediated by DNA. Moreover, it was planned to investigate the catalytic activity of DNMT3A in presence of both interaction partners, the H3-tail peptide with unmodified lysine 4 and the TRD domain of MeCP2.

2.3 Investigation of the catalytic mechanism of DNMT3A for cytosine-N3 methylation

Across a wide range of eukaryotes, DNA methylation has been found to coevolve with the DNA alkylation repair enzyme ALKB2, whose preferred substrates are 1-methyladenine and 3-methylcytosine (Ringvoll et al., 2006; Rošić et al., 2018). Interestingly, it had been shown that DNA-(cytosine C5)-methyltransferases are also responsible for the introduction of the toxic lesion 3-methylcytosine into DNA both in vitro and in vivo (Rošić et al., 2018). Since the reason of the ‘mechanistic misbehavior’ of methyltransferases remained unanswered, a further goal of my work was to investigate the mechanistic details of the introduction of 3-methylcytosine by DNMT3A. The goal was to find an explanation of how the enzyme generates this unwanted by-product and, at best, to provide a mechanistic explanation for this.

2.4 Investigation of the flanking sequence preference of *de novo* DNA methyltransferases

In both, mammalian development and cellular differentiation, DNA methylation had been shown to be an essential epigenetic mechanism that not only regulates gene expression but also plays a key role in genomic stability (Z. Chen et al., 2019; Law & Jacobsen, 2010; Schübeler, 2015). The C-5 position of cytosine residues in CpG dinucleotides is the most dominant mark in mammalian DNA, affecting about 70-80% of all CpG sites of the genome (review: Lister et al. 2009). The *de novo* DNA methyltransferases DNMT3A and DNMT3B methylate DNA during gametogenesis and post-implantation development, but they also contribute to the conservation of DNA methylation at repetitive elements. DNMT3B had been shown to be essential for the methylation of Satellite II repeats in human cells (Hansen et al., 1999; Okano et al., 1999; G. L. Xu et al., 1999). This specification of the enzymes can be explained on the

one hand by different targeting, but possibly also by specific preferences for certain base sequences. Additionally, it had been found, that DNMT3 enzymes also methylate non-CpG sites (Gowher & Jeltsch, 2001; Ramsahoye et al., 2000; Suetake et al., 2003).

Until recently, direct investigations of different methyltransferase preferences have been very time-consuming and inconvenient and especially moderately methylated CpX sites were therefore not further examined. However, the recent development of the Illumina next-generation sequencing method opened up new possibilities.

In this part of the thesis, the aim was to investigate the detailed flanking sequence preference of the *de novo* DNA methyltransferase DNMT3B and to search for similarities and differences to the detailed flanking sequence preference of DNMT3A. Another goal was to extend the mechanistic understanding of this sequence preference of DNMT3B and to be able to explain obtained patterns via comparisons with available crystal structures.

3 Results

3.1 Targeting and chromatin interaction studies of the DNMT3A methyltransferase

The targeting of the DNMT3A and DNMT3B methyltransferases to chromatin is mediated by the N-terminal parts of the enzymes which are also involved in the regulation of their activities (Gowher & Jeltsch, 2018). As already mentioned, this N-terminal parts contain an unstructured region which is directly followed by a PWWP domain and an ADD domain. In general, PWWP domains are composed of roughly 100-130 amino acid residues and have been identified in 22 human proteins listed in the HPRD database (Goel et al., 2012). Initially, PWWP domains were considered as possible protein/protein interactions domains, but different publications also identified nonspecific DNA binding abilities of the DNMT3A and DNMT3B PWWP domains (T. Chen et al., 2004; Purdy et al., 2010; Qiu et al., 2002; Stec et al., 2000). Both PWWP domains were also shown to be required for chromatin targeting and heterochromatic localization of the enzymes (T. Chen et al., 2004; Ge et al., 2004). For DNMT3A, this interaction could be connected to binding of the PWWP domain to the H3 histone tail when lysine 36 is trimethylated, via an aromatic cage in the PWWP domain, which has been described in a similar structure for other trimethyl lysine binding proteins (Dhayalan et al., 2010). In 2016, the structure of the human DNMT3B PWWP domain bound to an H3 peptide containing H3K36me3 was solved and confirmed the previously identified binding pocket from the biochemical studies (Rondelet et al., 2016). An exchange of the D329 residue in the mouse DNMT3A PWWP domain by alanine, which is in close proximity to the aromatic cage, led to the loss of H3K36me3 binding (Bock et al., 2011; Heyn et al., 2019; Kungulovski et al., 2014).

On the surface of the DNMT3 PWWP domains there are different patches of basic residues, which represent potential DNA contact points. These residues are conserved among human and mouse DNMT3A and DNMT3B PWWP domains (Rondelet et al., 2016; H. Wu et al., 2011). However, the DNA and chromatin binding behaviour of PWWP domains after targeted disruption of these individual binding sites is unknown and thus also the influence of the individual interaction points on targeting and localisation. For this study two basic residues on the surface of the mouse DNMT3A PWWP domain, namely K295 and R362 (human K299 and R366 for full length localisation studies), were chosen and mutated both to glutamic acid and K295 additionally to isoleucine, because this exchange was found in *DNMT3A* in patients with

paraganglioma (Remacha et al., 2018). These two residues represent two potential DNA binding sites. For the disruption of the H3K36me3 binding the already described D329A mutation (human D333) was further investigated.

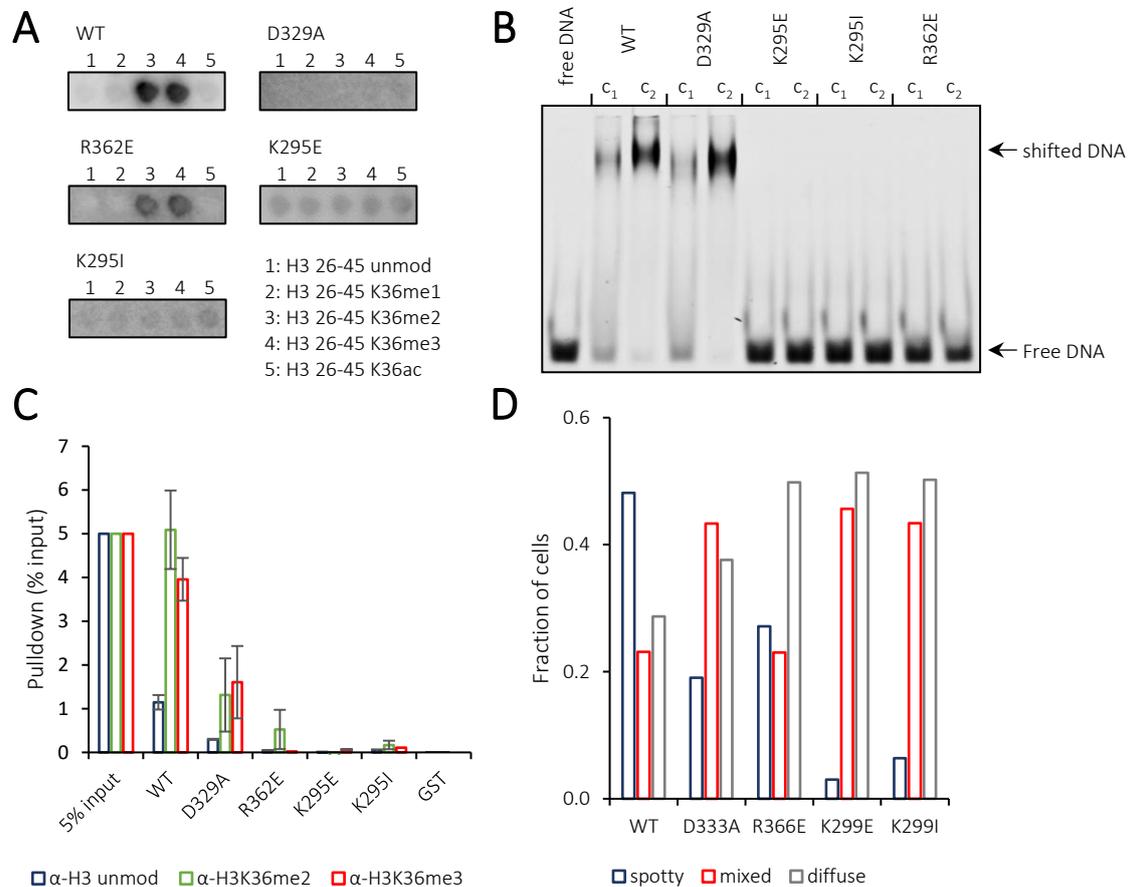


Figure 10| Summary of peptide and DNA binding experiments of GST-DNMT3A PWWP variants and localization studies of full length DNMT3A variants. A| Example images of wildtype and mutant PWWP domains binding to Modified™Histone Peptide Arrays. The picture shows the binding to the H3 26-45 peptide spots present on the array **B|** DNA binding of the DNMT3A PWWP domain proteins analyzed by a gel shift assay with two concentrations each (c₁ and c₂). **C|** Average pull-down efficiencies of oligonucleosomes pull-down experiments with GST-PWWP variants and standard deviations of 2 (unmodified) or 3 (K36me2 and K36me3) independent repetitions of each experiment **D|** Fractions of NiH-3T3 cells transfected with indicated mVenus labeled DNMT3A full length variant showing a spotty, mixed or diffuse nuclear distribution. Figure adapted from Appendix 1.

During this study different binding assays of the DNMT3A PWWP domain variants and localisation studies of the DNMT3A1 full length variants indicated a combined necessity of H3 tail binding and DNA binding for proper chromatin binding. Thus, a functional connection of both interactions has been proposed. The results have been published in the Journal of Molecular Biology and are briefly summarized here. For more details, see Dukatz et al. (2019) in **Appendix 1**.

3.1.1 H3K36me_{2/3} binding of the DNMT3A PWWP domain is disrupted by the D329A and K295E/I mutations

H3K36me₃ binding of DNMT3A PWWP was reported first (Dhayalan et al., 2010) before follow up studies revealed specific binding of both H3K36 di- and trimethylation (Bock et al., 2011; Heyn et al., 2019; Kungulovski et al., 2014; Mauser et al., 2018). Here, further investigations of the H3K36me_{2/3} binding of the PWWP mutants were conducted using Celluspot peptide arrays, which contain H3 26-45 peptide spots in different modifications (**Figure 10A**). For the wildtype PWWP domain, roughly equal binding to H3K36me₂ and H3K36me₃ was reproduced as previously reported. The R362E mutant showed similar binding behaviour as the wildtype with a slightly reduced intensity, indicating that this mutation has no drastic influence on the H3K36me_{2/3} interaction. As previously reported the D329A mutation (Bock et al., 2011; Heyn et al., 2019; Kungulovski et al., 2014; Mauser et al., 2018) as well as the K295 mutants first characterized here disrupted the peptide binding of H3K36me_{2/3} completely.

Further quantitative binding parameters of the wildtype and the R362E PWWP mutant were determined in equilibrium binding experiments to modified H3 27-43 peptides containing K36me₂ or K36me₃ using fluorescence anisotropy (**Appendix 1, Figure 1C and D**). The wildtype PWWP domain showed a slightly preferred binding to the H3K36me₂ compared to the H3K36me₃ peptide with binding constants of 63 μ M and 68 μ M respectively. The unmodified H3 27-43 control experiment showed no binding ($K_d > 400 \mu$ M). Compared to WT, the binding of the R362E mutant was 1.3-fold weaker but the relative preference for K36me₂ was similar. Taken together, the binding of the PWWP domain to H3K36me_{2/3} was lost for the K295 and D329 mutants and the WT and the R362E mutant could bind H3K36me_{2/3} with a slight preference for the di-methylated modification.

3.1.2 DNA binding of the DNMT3A PWWP domain is disrupted by the K295E/I and R362E mutations

For the DNA binding experiments of the PWWP WT and variants, a 30 bp long DNA oligonucleotide with a central CpG site was incubated with the PWWP domain variants followed by electrophoretic mobility shift assay (EMSA). The WT and the D329A mutant showed strong DNA binding, while the K295 and R362 had completely lost this ability (**Figure 10B**). The binding affinity to the 30mer DNA substrate stayed the same, no matter if it had an unmethylated or a methylated central CpG site (**Appendix 1, Figure 2B**) from which was

concluded, that the DNA binding of the PWWP domain is not influenced by DNA methylation at CpG sites.

Next it was investigated if the PWWP had any binding-preference for distinct DNA sequences. Therefore a PCR amplified 67 bp long DNA fragment with central CN site flanked by 10 randomized bases (**Appendix 1, Figure 2C**) was incubated with the GST-tagged wildtype PWWP domain and precipitated using GST-beads. The same was performed with the R362E mutant, which served as a negative control. The precipitated DNA was analysed in a semi-quantitative PCR, where a PCR product for the WT could be detected roughly 5 cycles earlier than for the mutant. This indicated an enrichment of about 32-fold of DNA for the WT compared to the control (**Appendix 1, Figure 2D**). The bound DNA and input DNA were both used for library preparation and subjected to Illumina sequencing. In the subsequent bioanalytical evaluation, no specificity for any dinucleotide starting with C could be determined, but a weak enrichment of A or T bases in the pull-down material when compared with the input was detected. In summary, a weak preference of the DNMT3A PWWP domain for AT-rich DNA sequences can be observed but no preferences for methylated or unmethylated CpG sites.

3.1.3 Chromatin interaction of DNMT3A PWWP domain requires both, H3K36me_{2/3} as well as DNA binding

In the next step, the influence of the selected amino acids of the PWWP domain on the binding behaviour to oligonucleosomes was investigated. For this purpose, oligonucleosomes were isolated from HepG2 cells which served as substrate in GST pull-down experiments with the PWWP variants. The pull-down experiments were analyzed using antibodies directed against unmodified H3, H3K36me₂ and H3K36me₃ (**Figure 10C and Appendix 1, Fig. 3B-D**). For the wildtype protein, a clear pull-down was observed in each experiment whereas no pull-down was obtained with the isolated GST protein. The strongest signal compared to the nucleosome input sample was obtained with the H3K36me₂ antibody and an almost equally strong signal with only slightly lower intensity compared to the input signal was obtained with the H3K36me₃ antibody. When the pull-down was detected with the nonspecific H3 antibody, an approximately 5-fold weaker signal was obtained compared to the H3K36me_{2/3} antibodies, which indicates a specific enrichment of nucleosomes containing H3K36me_{2/3}. For the PWWP mutants D329A and R362A, only weak interactions were detected, with reduced signals by a factor of 5 and 10, respectively. This suggests that the loss of H3K36me_{2/3} binding as well as

the loss of DNA binding results in a strong reduction of the interaction of the PWWP domain with oligonucleosomes. With both, K295E and K295I, mutants no oligonucleosomes were detected in the pull-down experiment, indicating that the combined loss of DNA and H3K36me2/3 binding causes an aggravated phenotype. The identical experimental set-up was also used with pure mononucleosomes. Interestingly, the mononucleosomes could not even be precipitated for the wild type, suggesting that the shorter linker DNA is not sufficient to support PWWP binding (**Appendix 1, Figure 5 Supplementary data**). In summary, it was clearly shown here that both H3K36me2/3 and DNA binding are important for the interaction of the DNMT3A PWWP domain with chromatin.

3.1.4 Influence of PWWP mutations on the sub-nuclear localization of DNMT3A

To link the previous *in vitro* data to a biological relevance and to investigate the effects of lost DNA binding of the DNMT3A PWWP domain at the cellular level, the sub-nuclear localizations of DNMT3A1 WT and the DNMT3A1 mutants were examined. Here, a well-established system was used, in which DNMT3A1 WT was shown to accumulate in heterochromatic regions which appear as well-defined spots in mouse NIH3T3 cells and the D333A exchange has led to a massive reduction in the heterochromatic localization (Dhayalan et al., 2010). For this, also the K299E, K299I as well as the R366E mutations were first inserted into mammalian expression vectors as Venus fusion constructs. It must be mentioned that in this section the numbering of the amino acids refers to the human DNMT3A which was used for the localization studies as Venus-tagged full-length protein. The mouse K295, D329 and R362 residues correspond to human K299, D333 and R366 respectively. Next, all five expression vectors (WT, D333A, K299E, K299I and R362E) were transfected into NIH3T3 cells and the sub-nuclear localization was observed by fluorescence microscopy (**Appendix 1, Figure 4**). The expected spotty sub-nuclear localization of wildtype DNMT3A1 which was observed in several previous studies was reproduced and co-staining with Hoechst confirmed that the spots are located at condensed heterochromatic regions (T. Chen et al., 2004; Dhayalan et al., 2010; Ge et al., 2004; Jurkowska, Rajavelu, et al., 2011; Rajavelu et al., 2012). To evaluate the microscopy data for all DNMT3A variants, the cells were categorized into three different phenotypes namely spotty, spotty/diffuse and diffuse. The strong reduction of the heterochromatic localization of DNMT3A1 D333A mutant was confirmed with an increased number of cells with diffuse or spotty/diffuse phenotype which has already been shown previously (Dhayalan et al., 2010).

The parallel experiments with the R366E mutant showed a similar nuclear distribution of phenotypes as the D333A mutant. For the heterochromatic localization of DNMT3A in mouse cells, these data indicate that both, DNA binding and H3K36me_{2/3} binding are of considerable importance. Reduced DNMT3A1 targeting to heterochromatic loci is observed for both, when the interaction with DNA is disrupted (R366E) and also when H3K36me_{2/3} binding is disrupted (D333A). For the K299E and K299I mutants, in which neither DNA nor H3K36me_{2/3} binding could be shown in the vitro experiments, an almost complete loss of heterochromatic localization was observed, suggesting a synergistic effect of both binding sites.

3.2 Investigation of the activity of DNMT3A with MeCP2 and H3

For the accurate setting of DNA methylation patterns, a precise chromatin targeting of the *de novo* methyltransferase DNMT3A is essential (review: Gowher and Jeltsch 2018). DNMT3A is an enzyme which has been shown to have two possible conformations, an autoinhibitory conformation which can be released into an active conformation when the enzyme binds to lysine 4 of the H3 tail (X. Guo et al., 2015). Previous work from former lab members of the Jeltsch research group uncovered a direct interaction of DNMT3A with the 5mC-reading protein MeCP2. This interaction could be mapped to the MeCP2-TRD binding to the ADD-domain of DNMT3A, and it was shown to lead to a strong inhibition of DNMT3A activity both, in vitro and in cell culture models. Since the regulation of the two different conformational states of DNMT3A and the binding to MeCP2 both are both mediated by the ADD domain, a combined regulation of DNMT3A by MeCP2 and H3-tails was hypothesized. Based on this, I investigated in a competitive methylation assay the impact of both interactions on DNMT3A activity. Additionally, to exclude that the observed interaction between MeCP2 and DNMT3A was indirectly due to DNA binding of both partners, I investigated the interaction in presence an unspecific endonuclease. Both results were included in the manuscript of Rajavelu et al. (2018) published in Nucleic Acid Research and can be found here in Appendix 2. A brief description of the results of these experiments is summarized in the following section.

First, to exclude the possibility of an indirect interaction via DNA, a GST-pull-down experiment with MeCP2-TRD as bait and MBP-DNMT3A-ADD as prey was performed in absence and presence of the highly active nonspecific endonuclease *Serratia marcescens* (DNase). The activity of the DNase was first validated in the pull-down experiment interaction buffer at the pull-down experiment reaction conditions and the DNA samples were separated on a 1% agarose TPE gel (**Figure 11A**). In the first lane 200 ng of a 509 bps DNA sample of a test digest without DNase and in the second lane 200 ng of a 509 bps DNA sample of a test digest with DNase are shown. In the first lane an intensive band of the DNA is visible, whereas no band is visible in the second lane, which confirmed the activity of the DNase in the pull-down experiment interaction buffer. The samples of the GST-pull-down experiments were separated on a 12% SDS-polyacrylamide gel and stained with Coomassie brilliant blue (**Figure 11B**) In the first lane, right next to the protein standard, a 10% input sample of MBP-DNMT3A-ADD domain with a size of roughly 58.4 kDa was loaded on the SDS-polyacrylamide gel as reference. The

second and third lanes are both GST-pull-down samples of GST-MeCP2-TRD with a mass of approximately 43.4 kDa showing >10% pull-down of MBP-DNMT3A-ADD, one without and the other with additional DNase in the incubation mixture. For both conditions, no difference in pull-down efficiency was observed. A parallel pull-down experiment, shown in the last lane, with the GST-tag alone had no detectable MBP-DNMT3A-ADD binding. These experiments confirm the direct interaction of DNMT3A-ADD and MeCP2-TRD, which is not mediated by DNA nor by the corresponding protein tags.

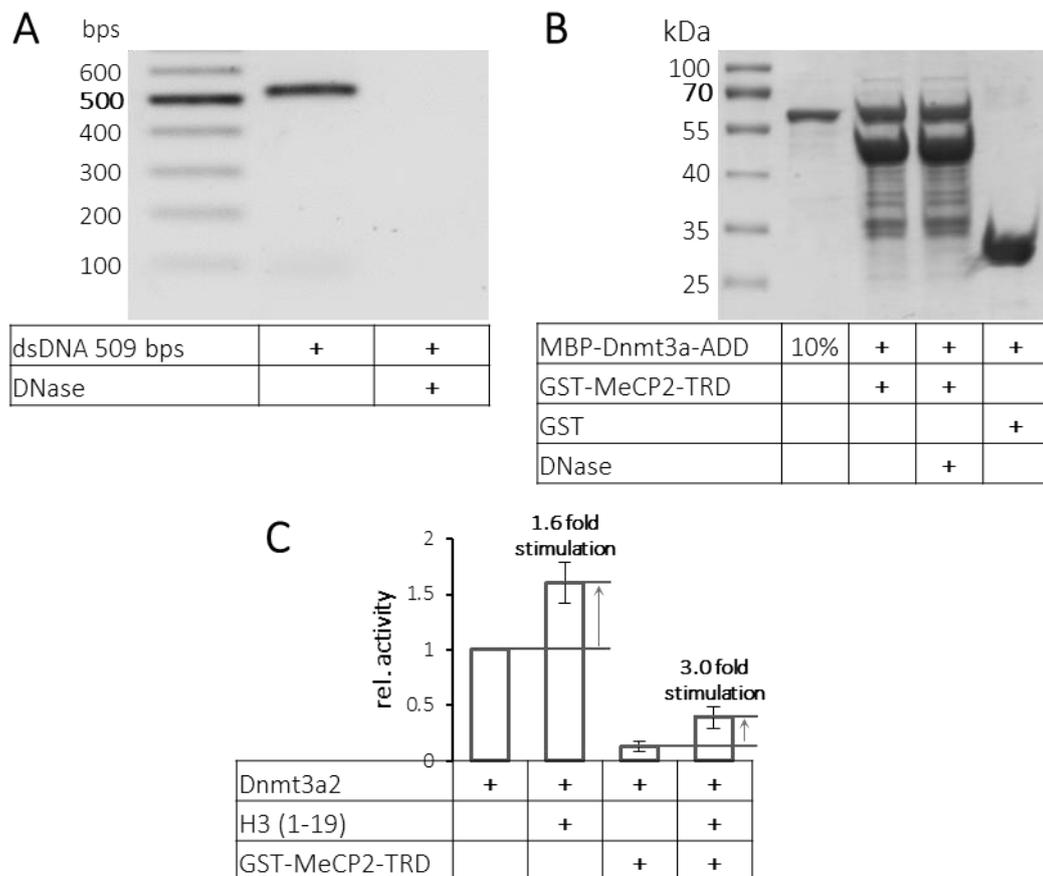


Figure 11 | Interaction of DNMT3A-ADD with MeCP2-TRD and relative activity of full length DNMT3A2 with MeCP2-TRD. A | Agarose gel with DNA samples of the activity validation of the nonspecific endonuclease *Serratia marcescens* (DNase) in the pull-down experiment interaction buffer **B |** Coomassie staining of the SDS-polyacrylamide gel with the samples from the validation pull-down experiment of GST-MeCP2-TRD with MBP-DNMT3A-ADD in absence and presence of the nonspecific *Serratia marcescens* endonuclease. **C |** DNA methylation activity of DNMT3A2 (1 μ M) in the absence or presence of the H3 peptide (amino acid sequence 1-19, 5 μ M) and/or MeCP2-TRD (1.2 μ M). Error bars represent SEM. Figure is adapted from Appendix 2.

Next, the activity of DNMT3A2 and its activation by binding to the unmodified lysine 4 of the H3 tail was investigated with the biotin-avidin microplate assay with and without preincubations of the enzyme with GST-MeCP2-TRD. It was observed that MeCP2-TRD-

mediated inhibition was attenuated in the presence of the unmodified H3-peptide (aa 1-19) (**Figure 11C**).

The data show that of the DNMT3A2-MeCP2-TRD complex by addition of the H3-peptide is stronger than activation of the free DNMT3A2. This is an indication that binding of H3 to the ADD domain may disrupt the DNMT3A-MeCP2-TRD interaction, which would reduce enzymatic inhibition.

3.3 Mechanistic studies on cytosine N3 methylation by DNA methyltransferase DNMT3A

A previous collaboration between the research group of Prof. Albert Jeltsch and the group of Dr. Peter Sarkies of the Medical Research Council (MRC) in London resulted in a publication describing the co-evolution of DNMTs and ALKB2/3 enzymes (Rošić et al., 2018). ALKB2/3 enzymes are glycosylases involved in DNA alkylation repair with a preference for 1-methyladenine and 3-methylcytosine. In the same publication it was shown that DNMT3A, apart from formation of the epigenetic modification 5mC, also produces small amounts of 3mC, thus providing an explanation for the co-evolution mentioned above. In the following part, it was investigated how it is mechanistically possible that the methylation of the cytosine base by DNMT3A does not exclusively occur at the C5 position, but it also generates small amounts of the toxic product 3mC. The results were published in the *Journal of Molecular Biology* and are briefly summarized in the following section. For further details see Dukatz, Requena, et al. (2019) in **Appendix 3**.

To understand how DNMT3A can generate 3mC besides 5mC, the different catalytic motifs that stabilize the transition state of the cytosine rotated out of the DNA helix were examined more closely. The following residues were known to be involved in the catalytic mechanism of cytosine-C5 methyltransferases: arginine 792 from the RXR motif, glutamic acid 756 from the ENV motif, and cysteine from the PCN motif (see **Figure 12A**).

First six active site catalytic domain mutants and the wildtype, which were already available as plasmids in the laboratory were overexpressed and purified, for the ENV motif the E756A mutant, for the RXR motif the R792A/H mutants and for the PCN motif the C710A/S mutants were obtained. Additionally, the C710S/E756A double mutant was investigated. For a general determination of the activity, the radioactive methylation assay was used for all variants. Similar activities to previously published data (Max Emperle et al., 2014; Gowher et al., 2006;

Lukashevich et al., 2016; Reither et al., 2003; Z. M. Zhang et al., 2018) for these mutants relative to WT were obtained with a maximum of ~5% residual activity of the E756A mutant (**Appendix 2, Figure 3A**). This confirmed the importance of each of these 3 residues for an efficient catalysis.

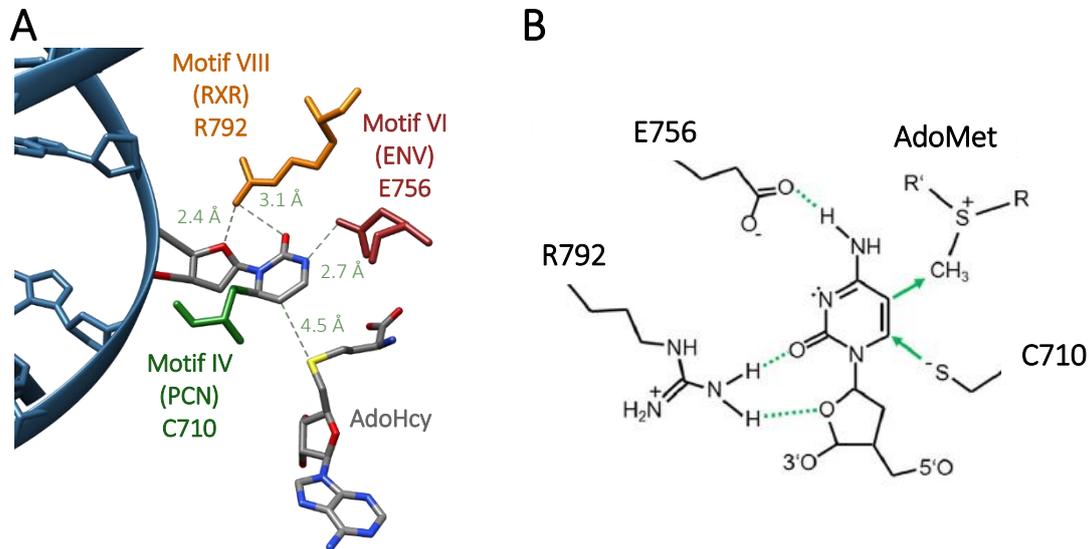


Figure 12 | Structural basis of 5mC methylation. A | Structure of the active site of DNMT3A with bound Zebularine (a cytidine analog lacking the amino group in position 4) rotated out of the DNA helix and adenosyl-L-homocysteine (AdoHcy), the cofactor product after methyl group transfer (PDB: 6F57) **B |** Schematic picture of the first catalytic steps of covalent complex formation and methyl group transfer. Figures are adapted from Appendix 3.

Next, another methylation assay was performed with all variants, where a plasmid DNA substrate from bacteria, lacking endogenous 5mC, was used. This assay needed to be performed at higher enzyme concentrations for longer incubation times, to support the 3mC detection in the later stage. The DNA was purified and further processed by our collaborators at the MRC, where it was degraded to nucleosides and the levels of 5mC and 3mC were determined by LC-MS/MS (**Appendix 2, Figure 2A and 2B**). Under the conditions of this assay, all mutants showed higher relative 5mC activities than in the radioactive kinetics. The methylation of the E756A mutant resulted in roughly 10% of 5mC methylation compared to the WT and the other mutants in about 1% of 5mC methylation compared to the WT (**Figure 13B**). The 3mC generation of the wildtype sample was comparable to previously published data (Rošić et al., 2018). For the E756A and R792A mutants the 3mC generation was reduced about 4-fold and for the C710A and R792H mutant about 10-fold. For both C710S mutants the 3mC generation was reduced below detectable limits (**Figure 13C**). To illustrate these differences and to explore the mechanism of 3mC generation, 3mC/5mC ratios were plotted relative to

wild type. It can be seen that mutations of the catalytic residues affect 3mC and 5mC differently. The E756A mutant showed a 2-fold increase in the 3mC/5mC ratio relative to wild type. The R792A/H and C710A had a 10-15-fold relative increase in the 3mC/5mC ratio and both C710S mutations affected 3mC formation more than 5mC formation, as they could show similar 5mC values to the C710A mutant, but its 3mC formation was below detectable limits. The conclusions of these findings will be discussed in section 4.3.

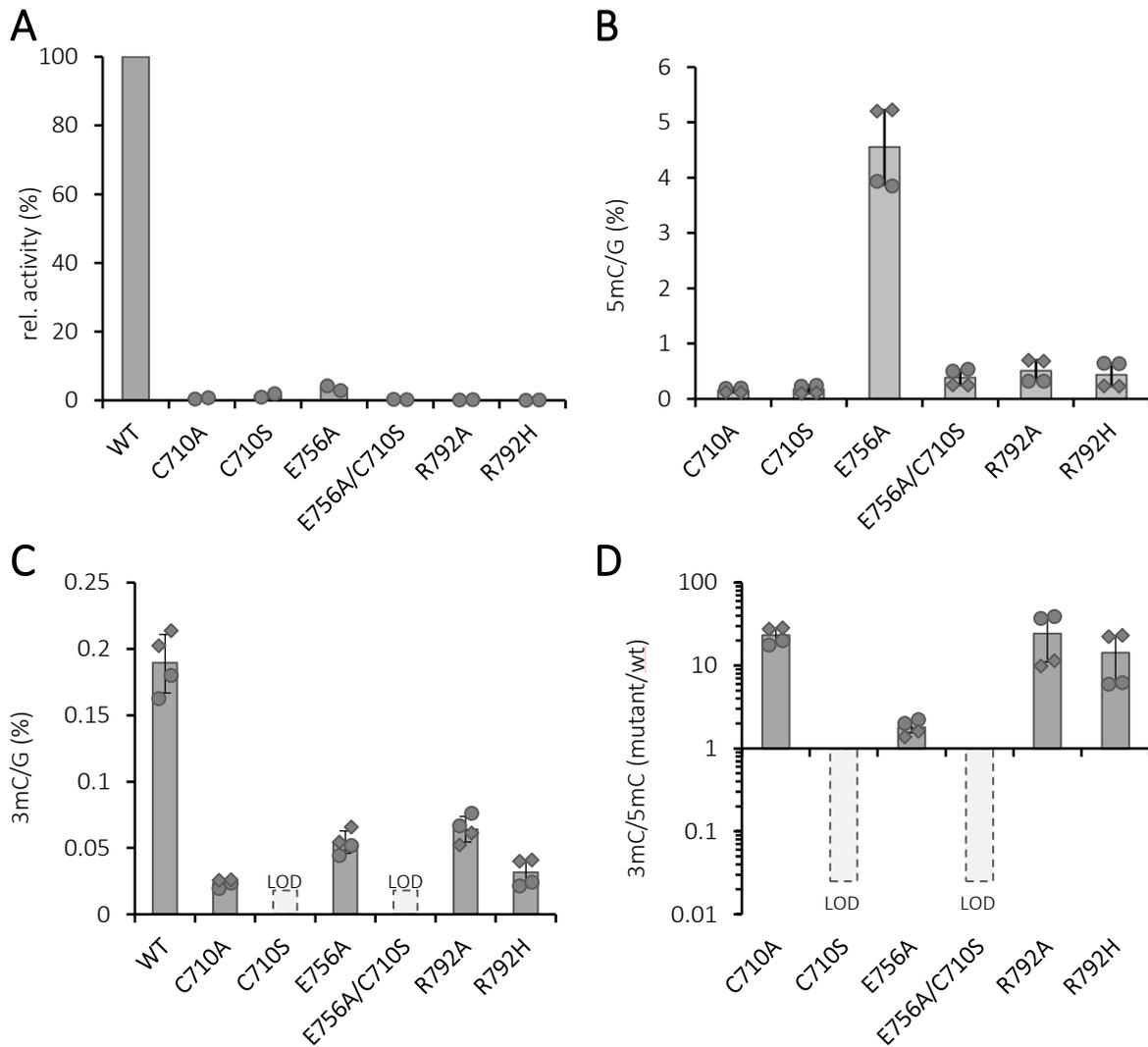


Figure 13 | Activity analysis of the DNMT3A mutants. **A** | Overall DNA methylation determined using the radioactive DNA methylation assay. **B** | 5mC generation determined by LC-MS/MS. 5mC generation displayed as ratio of 5mC and G. **C** | 3mC generation determined by LC-MS/MS. 3mC generation displayed as ratio of 3mC and G. **D** | Relative 3mC generation displayed as ratio of 3mC and 5mC indicated as relative value compared mutant and wild type. LOD, limit of detection. Figure is adapted from Appendix 3.

3.4 Investigation of differences in the sequence preference of DNMT3A and DNMT3B

The two *de novo* DNA methyltransferases DNMT3A and DNMT3B are closely related in their amino acid sequence (Okano et al., 1999). In earlier studies, there was evidence of partial redundancy of both enzymes in the establishment of methylation patterns throughout the genome (T. Chen et al., 2003; Okano et al., 1999). This would indicate that the enzymes are interchangeable, however single knock-outs of either DNMT3A or DNMT3B led to embryonic or postnatal lethality, indicating partially functional distinctions (Baubec et al., 2015; Challen et al., 2014; T. Chen et al., 2003). DNMT3A has been identified as the enzyme responsible for the establishment of methylation at major satellite repeats and for allele-specific imprinting during gametogenesis (T. Chen et al., 2003; Hata et al., 2002). DNMT3B, on the other hand, has a dominant role in early embryonic development and in the methylation of minor satellite repeats (T. Chen et al., 2003; Okano et al., 1999). Here, in cooperation with the research group of Jikui Song from UCR, differences of DNMT3A and DNMT3B regarding their crystal structure and their flanking sequence preference analyzed on genomic DNA from mouse embryonic stem cells, as well as in vitro methylation assays developed in the research group of Albert Jeltsch were investigated. The crystal structure of the DNMT3B/DNMT3L tetramer in complex with CGT DNA showed that lysine 777 in DNMT3B points towards the +1 flanking nucleotide, whereas the corresponding residue arginine 836 in the DNMT3A-CGT complex donates a hydrogen bond to Gua6-O6 (**Appendix 4, Figure 6a-c**). My task in this project, was the investigation of the sequence dependent methylation of the catalytic domain of DNMT3B and its K777A mutant with following sequence preference analysis. The results were included in the manuscript from Gao et al. (2020) published in nature communications and are attached in Appendix 4. A brief summary of these results is described in the following section. The analog results for DNMT3A was provided by one of the first authors of the paper, the results are depicted here for comparison.

First, a pool of DNA substrates with one target CpG site flanked by 10 randomized nucleotides on each side was generated and methylated by murine DNMT3B CD (aa 558 – 859). After methylation, the randomized substrate pool was subjected to hairpin ligation, bisulfite conversion, PCR amplification and next-generation sequencing (NGS) analysis. The analysis of base enrichments at all flank positions in the methylated sequences provided that methylation

of both *de novo* methyltransferases is significantly influenced by the CpG-flanking sequence from the -2 to the +3 site (**Figure 14A and Appendix 4, Figure 1a**). Because of the lower base enrichment at the outer flanking positions, the focus was placed on the ± 3 bp flanking positions on the activity of the enzymes. The effects of the flanking sequences on the methylation rates of both methyltransferases are very pronounced and we detected NNCGNNN flanks with very high and very low methylation levels in the same reaction mixtures (**Appendix 4, Supplementary Table 2**).

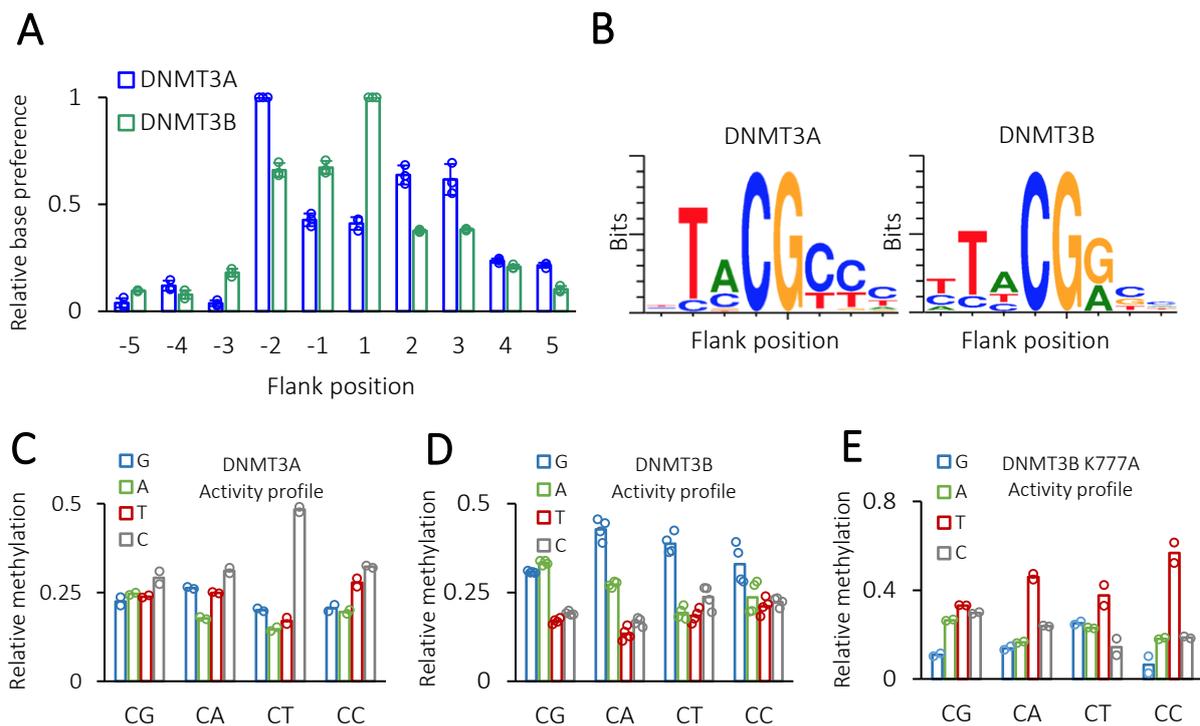


Figure 14 | Flanking sequence preference analysis of murine DNMT3A WT, murine DNMT3B WT and murine analog of human DNMT3B K777A. A | Relative base preferences at the -5 to +5 positions flanking CpG, indicating the strength of sequence readout at each side. The numbers refer to the standard deviations of the observed/expected base composition at each site among the methylated sequence reads, normalized to the highest value for each enzyme. N = 3 replicates. Data are mean \pm SD. **B |** Weblogos of the 50-200 most preferred NNCGNNN methylation sites **C-E |** relative CpG and non-CpG methylation averaged for the different +1 flanking base pairs of DNMT3A WT (**C**), and DNMT3B WT (**D**) and the mouse analog of human DNMT3B K777A (**E**) The cytosine-containing dinucleotides are displayed on the horizontal axis and +1 base is shown in the different bars, respectively. Data are mean of two replicates (**C** and **E**) or mean of 4 replicates (**D**). Figure is adapted from Appendix 4.

Using Weblogo analysis of the most highly methylated sites, a preference for a CG(C/T) motif could be determined for DNMT3A, whereas DNMT3B has a better catalytic effect on CG(G/A) motifs. Both methyltransferases prefer a T at the -2 flanking site (**Figure 14B and Appendix 4, Figure 2c**). Next, the flanking sequence preference of non-CpG methylation of DNMT3B was investigated, by extending the randomized DNA-substrate pool in the methylation for the NGS

approach, by adding equimolar amounts of same randomized substrates, but instead of a central CpG site carrying CpA, CpT and CpC sites (together CpH). In a targeted activity analysis, the preference of DNMT3B for a G at the +1 site showed to be strongly enhanced in the non-CpG methylation (**Figure 14D and Appendix 4, Figure 2b**). The same effect of DNMT3A was only significant in the CpT methylation for the favored C at the +1 site in the NGS methylation assay (**Figure 14C and Appendix 4, Figure 2a**).

The solved crystal structure of the DNMT3B-DNMT3L complex identified lysine 777 of DNMT3B as an amino acid that recognizes the contact with +1 flanking nucleotide (**Appendix 4, Figures 6e-l**). For this reason, the K777A mutant was cloned into the catalytic domain of the murine DNMT3B, overexpressed and purified, to determine the influence of the lysine on the flanking sequence preferences. The same deep enzymology workflow and analysis as performed for the DNMT3B WT enzyme was repeated with the K777A mutant. The distinct preference of DNMT3B for CpN methylation at sites flanked by G at the +1 site was completely lost with the K777A mutant and instead a preference for T at this site was observed (**Figure 13E and Appendix 4, Figure 7b**). Together with the crystal structure, these data demonstrate, that the lysine 777 of DNMT3B acts as a key determinant that recognizes the different sequence contexts flanking the methylation site, which is different from what was observed for DNMT3A. Since a previous study has demonstrated, that DNMT3L modulates cellular *de novo* methylation activities through focusing the DNA methylation machineries on well chromatinized DNA templates (Wienholz et al., 2010) the flanking sequence analyses were also performed with additional DNMT3L. Regarding the flanking sequence preferences the presence of DNMT3L made no significant differences for both DNMT3s as the data sets with and without additional DNMT3L highly correlated (**Appendix 4, Supplementary Figure 9**).

3.5 Complex DNA sequence preference analysis of the DNMT3B DNA methyltransferase

Besides CpG-DNA methylation during gametogenesis and post-implantation development (review: Chen et al. 2019; review: Zeng and Chen 2019) the two *de novo* DNA methyltransferases DNMT3A and DNMT3B also contribute to the preservation of CpG-DNA methylation at repetitive elements (review: Jeltsch and Jurkowska 2014). DNMT3B is particularly relevant in this process and has been shown to be essential for the methylation of Satellite II repeats in human cells (Hansen et al., 1999; Okano et al., 1999; G. L. Xu et al., 1999). With the solved crystal structure of the catalytic domain of DNMT3B in complex with the C-terminal domain of DNMT3L and 4 different DNA substrates (Gao et al., 2020; C. C. Lin et al., 2020), structural differences to DNMT3A could be identified and explained (Gao et al., 2020; Mallona et al., 2021). In this study, starting from the crystal structure, the focus was set on the investigation of the structural regions of DNMT3B that are directly involved in the interaction with DNA, these are the target recognition domain loop (TRD loop), the catalytic-loop and the RD homodimeric interface region (**Figure 15A and Appendix 5, Figure 1A**). The catalytic loop and TRD loop approach the DNA at the CpG target region from the minor and major groove side, respectively, generating a clamp holding the DNA, while the arginine 823 from the RD interface forms a hydrogen bond to the DNA-backbone. Interestingly, the CpG specificity of the two *de novo* methyltransferases is not exclusive since both enzymes also introduce low level of non-CpG methylation in vitro (Gowher & Jeltsch, 2001; Ramsahoye et al., 2000; Suetake et al., 2003). For DNMT3B it was shown, that non-CpG methylation is introduced in ESCs and neural cells, mainly in a CAG context (Laurent et al., 2010; Lee et al., 2017). Biological relevance of non-CpG methylation has so far been attributed to gene regulatory and chromatin modulatory functions, including X-chromosome inactivation, mainly in the nervous system (He & Ecker, 2015; Jeltsch et al., 2018). As already described in the previous section, the methylation of DNA by DNMT3B (as well as by DNMT3A) is also dependent of the flanking sequence context of the target CpG site. However, the mechanistic basis on which the methyltransferases determine their flanking preferences on CpX sites is still unknown. To expand our understanding of the mechanistic details leading to these preferences, detailed substrate preference analyses of the catalytic domain of human DNMT3B and selected DNMT3B mutants were performed in this study. For this purpose, the same experimental

procedure was used as described in the previous chapter. The results were published in *Nucleic Acid Research* and can be found in **Appendix 5**. A brief summary of the results is provided in the following sections.

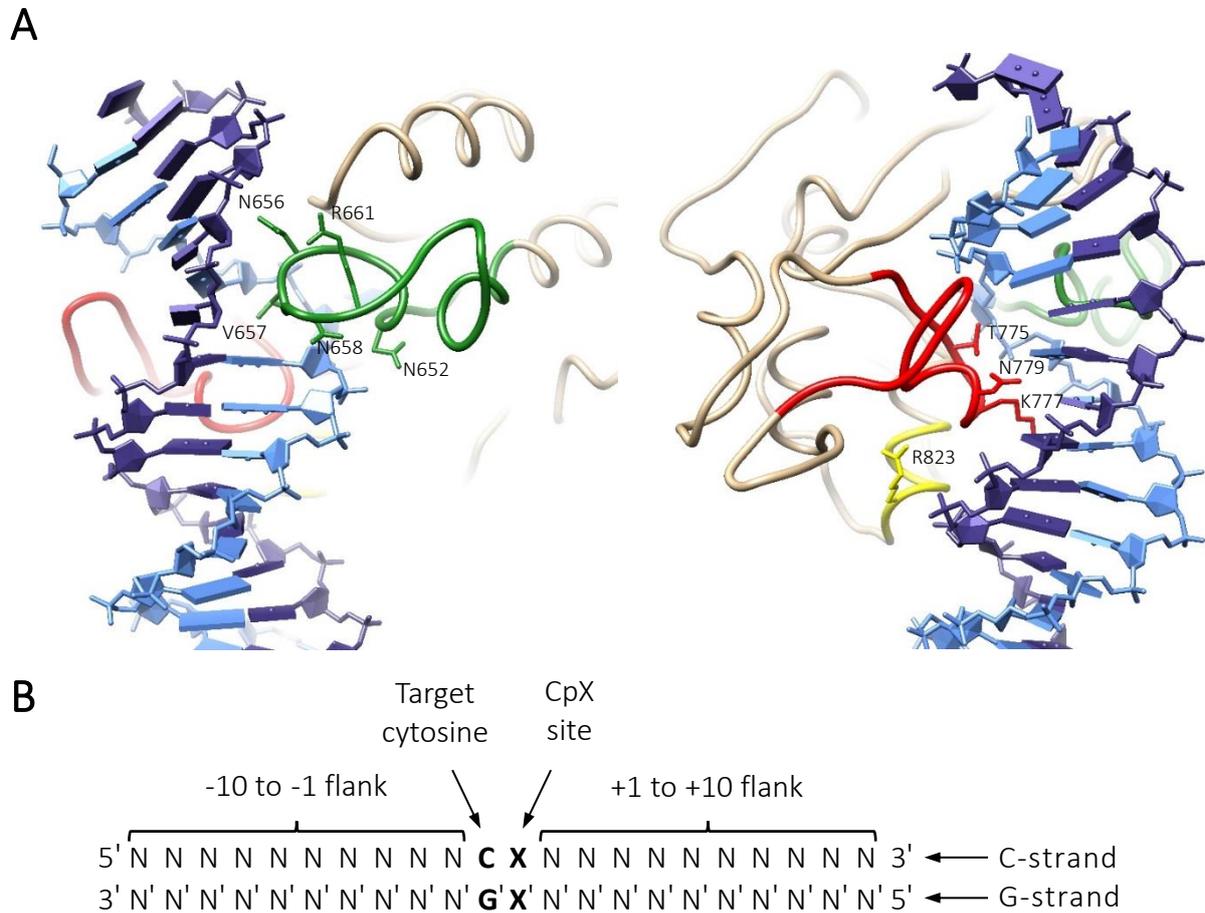


Figure 15| Localization of the designated DNA-contacting amino acids of DNMT3B and substrate design used in this study. A| Structure of the CGA complex of DNMT3B (6U8P) in two views focusing on the catalytic loop (green), TRD loop (red) and RD interface loop (yellow). The target strand for DNA methylation (C-strand) is shown in light blue, the non-target strand (G-strand) in purple. The amino acid residues mutated in this study are indicated. **B|** The terminology used in this study to describe the DNA substrate. Figure is adapted from Appendix 5.

Nine different DNA-contacting amino acids were selected for mutagenesis (**Figure 15A**) and the detailed effects of the mutations on the flanking sequence preferences were compared with the wild type. Five of the selected residues are located in the catalytic loop (N652, N656, V657, N658 and R661). The terminology used here is shown in **Figure 15B**. N652 contacts the phosphate backbone at the +1 flanking base pair (**Appendix 5, Supplementary Figure S1**). N656 contacts both, the -1 base and the DNA backbone at the -1 and -2 flanking positions (**Appendix 5, Supplementary Figure S1**). V657 partially occupies the space created by the rotated cytosine and contacts the G' together with the -1 flank and the CpX site (**Appendix 5, Supplementary Figure S1**). N658 contacts the CpX site and +1 flank (**Appendix 5, Supplementary Figure S1**).

Figure S1). R661 forms a side-chain hydrogen bond with N656, a structural feature distinct from that of DNMT3A (**Appendix 4, Figure 4b**). Three of the selected amino acid residues are located in the TRD loop. T775 approaches the sugar residue of the target cytosine and it forms a water mediated H-bond to the N7 atom of the CpX site G/A residue (**Appendix 4, Figure 3d and Appendix 5, Supplementary Figure S1**). K777 forms complex specific interactions at the +1 site (where it hydrogen bonds to the N7 of the G), and contacts to +2 and +3 flank sites (**Appendix 5, Supplementary Figure S1**). N779 forms complex specific interactions to the CpG site (including a hydrogen bond to the O6 atom of G) and +1 flank (**Appendix 5, Supplementary Figure S1**). The residue R823 from the RD interface forms a phosphate contact at the +3 flank site. The corresponding residue in DNMT3A is the R882 which is a hotspot for DNMT3A mutations in AML (Hamidi et al., 2015). Mutations of this residue have led to strong changes of flanking sequence preferences of DNMT3A (M. Emperle et al., 2019; Max Emperle et al., 2018) and structural adjustments of this region (Anteneh et al., 2020). In addition to the DNA interacting functions of these amino acids, they are also linked by two hydrogen bond networks. One of these networks is in the catalytic loop between residues N652, N656, N658 and R661 and the other between the TRD loop residues T775, K777 and N779. This also means that conformational changes of one residue can be transferred to others, which can lead to cooperative effects. Detailed information about DNA contacts and intramolecular contacts is provided in **Appendix 5** in **Supplementary Tables S2 and S3** and visualized in **Supplementary Figure S1**.

First all described residues were mutated to alanine, N656 and N779 additionally to aspartate and T775 additionally to asparagine and glutamine in context of the catalytic domain of human DNMT3B (aa 553-853). For a general activity determination of the used human DNMT3B CD variants, the radioactive DNA methylation assay was used employing a 30mer oligonucleotide substrate with a single CpG site in a TTCCGGGGA sequence context (**Figure 16A**). Next, a pool of DNA substrates with one target CpX site flanked by 10 randomized nucleotides on each side was generated and methylated with each DNMT3B variant. After the methylation, the randomized substrate pools were subjected to hairpin ligation, bisulfite conversion, PCR amplification and next-generation sequencing (NGS) analysis. Since the T775 mutants all showed very weak residual activities, only the T775A mutant was used for the deep enzymology experiments. For the initial analysis of the NGS data, the methylation levels of CpG sequences were compared with the results of the radioactive assay and both activity

determinations agree closely with each other (**Appendix 5, Figure 2A and B**). The activities of the N652A, N656A, N779A and N779D were similar to WT (<25% reduction compared to WT). V657A, R661A, K777A and R823A showed a residual activity of 25% - 75% compared to WT. N656D and N658A showed a strong reduction of activity (~5% and ~15% residual activity respectively) and the T775 mutants had almost no residual activity with only ~0.5% compared to WT. Next the global non-CpG methylation activities for the wildtype and mutants of DNMT3B were determined (**Appendix 5, Figure 2C**). For the WT enzyme the CpA methylation activity was 17% of the activity observed at CpG sites, 7% for CpT- and 6% for CpC methylation. In general, most mutants had similar overall non-CpG activities compared to the WT with some exceptions. K777A had increased relative non-CpG methylation in all three sequence contexts, the same was observed for T775A, but with high error margins between the repeats because of very low overall methylation levels. N658A had increased relative CpA methylation and V657A had increased relative CpT methylation. The opposite effect was observed for N656D and R823A with a global reduction of relative non-CpG methylation. Next, position specific enrichment of bases was calculated for the different types of CpX methylation for the -4 to +4 flanking region (**Appendix 5, Figure 3A**). The WT data for CpG methylation was very similar to the previous results described for murine DNMT3B (Gao et al., 2020). The profile showed the strongest preference for T at the -2 site and for G/A at the +1 site. Also, weaker effects in the previous data sets were confirmed, namely the preference for A and disfavor for G at -1 and favor for C and disfavor for A at +2. Different correlation analyses showed a high similarity of the flanking sequence preferences for the different CpX sites (**Appendix 5, Figures 3B-F**). When comparing the average methylation rates of the 15% most favored and 15% most disfavored sites, a clear trend for an enhanced flanking sequence preference for non-CpG methylation was revealed (**Appendix 5, Figure 3G**).

A more detailed comparison of the CpG methylation profiles in the -4 to +4 flanking base preferences with the methylation profiles of CpA, CpT, and CpC methylation revealed few but striking differences for the wildtype DNMT3B (**Appendix 5, Figure 3A**). Trends observed in CpG methylation were enhanced in CpA methylation. At the -2 position, T was more favored and G more disfavored, and at the -1 position, A was more favored. At the +1 position, G was more preferred. The preference of A observed at the +1 position was specific only to CpG methylation and was not observed for the other CpH profiles. CpT methylation gave a very similar profile compared to the profile of the CpA methylation. For CpC methylation, a favor

for T was observed at the -2 to +3 sites in comparison to the CpG methylation. It is also noticeable that the G disfavoring effect at the -2 position was enhanced in all CpH methylations.

For the comparison of genomic non-CpG methylation with the obtained *in vitro* preferences, publicly available whole genome bisulfite data of CpG and non-CpG methylation from human ES cells were compared with the same cell line after knock-out of DNMT3B (Tan et al., 2019). The difference of both data sets was used as indication of the activity of DNMT3B in native cells. Next, the flanking sequence dependent methylation levels of the DNMT3B associated genomic methylation were determined and compared with the *in vitro* flanking sequence preferences of DNMT3B (**Figure 16C and Appendix 5, Figure 4C and D**). No significant correlation was observed for CpG methylation, which agrees with the fact that DNMT3B does not strongly contribute to overall CpG methylation in this cell line, which is dominated by the activity of DNMT1. Strikingly, the flanking profiles of genomic CpA and CpT methylation were highly correlated with the activity profiles of the *in vitro* data (**Figure 16C and Appendix 5, Figure 4C and D**). The correlation for CpC methylation was weaker, but still highly significant. This clearly indicates that the cellular non-CpG methylation is highly influenced by the flanking sequence preference of DNMT3B, especially at CpA and CpT sites.

The detailed deep enzymology based flanking sequence preference analyses were next determined for the DNMT3B mutants except for T775A because of its low methylation rate. For most of the mutants, the key preferences of WT DNMT3B with the preferences for T at the -2, A at the -1 and G at the +1 site were reproduced (**Appendix 5, Figure 5**). As already seen and described for the murine K777A DNMT3B analog in **Figure 14**, the flanking sequence preference at the +1 site, also changed for the human DNMT3B K777A in the same way, G was highly disfavored in the CpG methylation and T strongly favored for the non-CpG methylation, which also illustrates the great similarity of both enzymes. Another interesting effect was observed at the +1 flanking site for some of the mutants, where the WT preferred G and A in the CpG methylation. The N656A, N658A, N779A and N779D had lost this preference for A at the +1 site, which was most obvious for the N779D mutant. Interestingly, this effect was already observed for the non-CpG methylation of the WT enzyme.

The SatII sequence (TTCGAT) is preferred by DNMT3B WT (ranked 19th for the WT among the 256 possible NNCGNN sites). When the SatII sequence preferences of the selected mutants were examined, a moderate reduction of preference for V657A, N658A, T775A, and N779A,

and a slightly stronger loss of preference for N656D, R661A, K777A, and N779D was revealed. In the case of N779D, the sequence can also no longer be counted among the favored substrates. Only for the RD interface mutant R823A, the preference stayed almost unchanged (**Figure 16B and Appendix 5, Figure 6C**). The reduced preference for the SatII sequence of N779A/D is correlated with the loss of the preference for A at the +1 site. The strongly reduced SatII preference for R661A and K777A can be connected to their loss of preference at the -1 and -2 sites, respectively, which are unique among all investigated mutants. Taken together, the results show that the SatII interaction depend on the positioning of both, the catalytic loop and the TRD loop and it is based on several protein-DNA contacts.

As briefly mentioned earlier, the catalytic domains of human and murine DNMT3B are very similar, but it must also be mentioned that 17 of 285 amino acids differ. One of these differences is a lysine to arginine exchange, which in human DNMT3B corresponds to K782 and is located in the TRD loop, where the lysine contacts the DNA backbone at the +2 and +3 flanking positions of the non-target strand. To determine differences in flanking sequence preference of human and mouse DNMT3B, the different average methylations of NNNCGNN sites were analyzed. Although the flanking sequences of both enzymes are strongly correlated on a global scale (**Appendix 5, Figure 7A**), they do differ significantly at some flanking sequences (**Figure 16D and Appendix 5, Figure 7B**), suggesting that some minor changes of flanking sequence preferences occurred in the evolution of mouse and human DNMT3B.

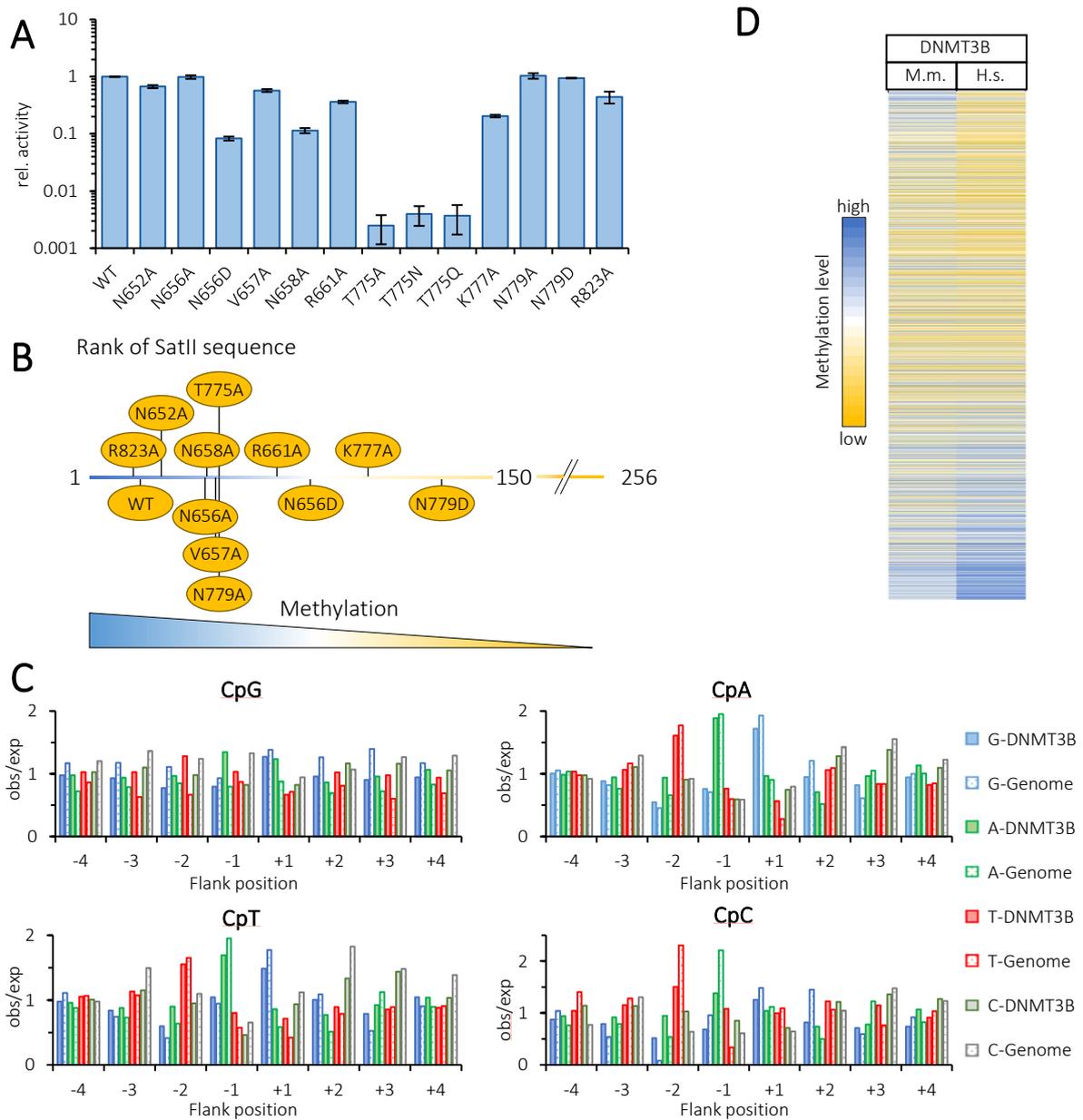


Figure 16 | Excerpt of the results of the complex DNA sequence readout mechanisms of DNMT3B. **A** | Relative CpG methylation activities determined by radioactive kinetics. Numbers present average of three experiments, error bars show the SD. **B** | Rank of the SatII sequence (TCGAT) methylation preference by WT and mutant DNMT3B in all 256 NNCGNN flanks. **C** | Enrichment and depletion of bases at -4 to +4 flank positions for CpX methylation in the genomic DNMT3B dependent methylation (Genome) and the DNMT3B in vitro preferences (DNMT3B). Data are shown as observed/expected values (obs/exp). **D** | Heatmap of the average methylation activities of mouse (M.m.) and human (H.s.) DNMT3B at NNCGNN sites sorted by the difference between human and mouse enzymes. Figure is adapted from Appendix 5.

4 Discussion

The basis of every living organism is the genetic information determining its phenotypic composition, which is stored in the genome in form of the DNA sequence. The stored information is passed on from one generation to the next and serves as the blueprint for how a multicellular organism containing a multitude of differentiated cells can develop from a single cell. While inheritance of DNA sequence information is the main subject of genetics, the field of epigenetics investigates those mechanisms that cause heritable changes of phenotypes, but do not change the underlying DNA sequence. This includes mechanisms that influence the regulation of transcription, whereby the organized decoding of DNA can be controlled. One of the epigenetic mechanisms is DNA methylation, which has been recognized over the last 50 years as one of the most important epigenetic modifications found in almost all species. In mammals, DNA methylation is essential for development and cellular differentiation. It is the only epigenetic modification that occurs directly at the DNA sequence level. There it arises from a transfer of a methyl group usually to the 5' position of the pyrimidine ring of cytosine generating 5-methylcytosine. During embryonic development, this modification is set by the *de novo* methyltransferases DNMT3A and DNMT3B, these enzymes also play a significant role in severe diseases such as acute myeloid leukemia or the ICF syndrome. Thus, it is fundamental to obtain a deeper understanding of the mechanism and regulation of these enzymes to acquire further knowledge for the treatment of those diseases. In this work, regulatory as well as catalytic mechanisms of the two *de novo* methyltransferases were investigated. The findings described in the results section are discussed and put in context of existing literature in this chapter.

First, the binding of the PWWP domain of DNMT3A to chromatin and the influence of H3K36me_{2/3} binding as well as direct DNA binding will be discussed. This is followed by an interpretation on the regulatory function of the DNMT3A ADD domain via H3K4 binding in the context of MeCP2 and its effect on the catalytic activity. In the third part of the discussion, the effects of different mutations in the catalytic domain of DNMT3A on the generation of 3-methyl cytosine will be compared and a mechanism for the introduction of this toxic base modification will be put forward. In the fourth part, differences of the two *de novo* methyltransferases DNMT3A and DNMT3B with respect to their different flanking preferences

are presented and discussed. Finally, sequence-specific methylation of DNMT3B is discussed in detail, and individual DNA contact-related regions and amino acids are analyzed.

4.1 Targeting and chromatin interaction studies via the PWWP domain of the DNMT3A methyltransferase

DNA methylation is a determinant of the cellular transcriptional program and is therefore closely linked to cell identity (Bogdanović & Lister, 2017). For these functions of DNA methylation, it requires the organized recruitment of methyltransferases via transcription factors as well as other DNA-binding proteins, chromatin modifiers, non-coding RNA and histone marks (review: Laisné et al. 2018). The latter, or more specifically the methylation of lysine 36 of histone H3 and the direct DNA binding property of DNMT3A, was the focus of this part of the work. In general, DNA interaction of PWWP domains could be shown for a variety of different proteins carrying this domain. These include for example the PWWP domains of LEDGF/p75 (Eidahl et al., 2013), HDGF (Lukasik, 2006), MSH6 (Laguri et al., 2008) but also those of DNMT3A (Purdy et al., 2010) and DNMT3B (Qiu et al., 2002). This property is based on the high content of basic residues, which causes the isoelectric point to be above 9, making the surface more positively charged and thus a favored contact site for DNA binding (Rona et al., 2016). In the present work, the focus was on the PWWP domain of DNMT3A, which in addition to its DNA-binding property has also been shown to bind H3K36me_{2/3} in vitro (Dhayalan et al., 2010). The close vicinity of the H3K36 side chain emerging from the histone octamer and the linker DNA from the nucleosome suggest that proteins binding to H3K36me_{2/3} might also interact with DNA. In addition to this indication, the PSIP1- PWWP domain was shown to bind to the H3K36 methylated H3 tail and nucleosomal DNA in a concerted binding reaction (Van Nuland et al., 2013).

The DNA binding ability of the DNMT3A PWWP domain was confirmed here and, in addition, a slight preference for AT-rich sequences was determined. A specificity for CpG sites or preferences for methylated or unmethylated substrates could not be observed. Apart from the general fact that basic residues must be available for DNA binding, further evidence of the importance of the K295 and R362 residues has only recently emerged. With the solved crystal structure of the LEDGF PWWP domain bound to a nucleosome containing a H3K36me₃ MLA (Wang et al., 2020), it was also possible to simulate a superposition of the DNMT3A PWWP domain on the LEDGF PWWP domain. It can be seen that these two basic residues point exactly

in the direction of the initial region of the linker DNA of the nucleosome with a very small distance to it, whereby not only the right charge but also the steric condition for an interaction is probable (**Figure 17**). Hence this structural study is in very good agreement with the results of my biochemical study that were described before the structure was available.

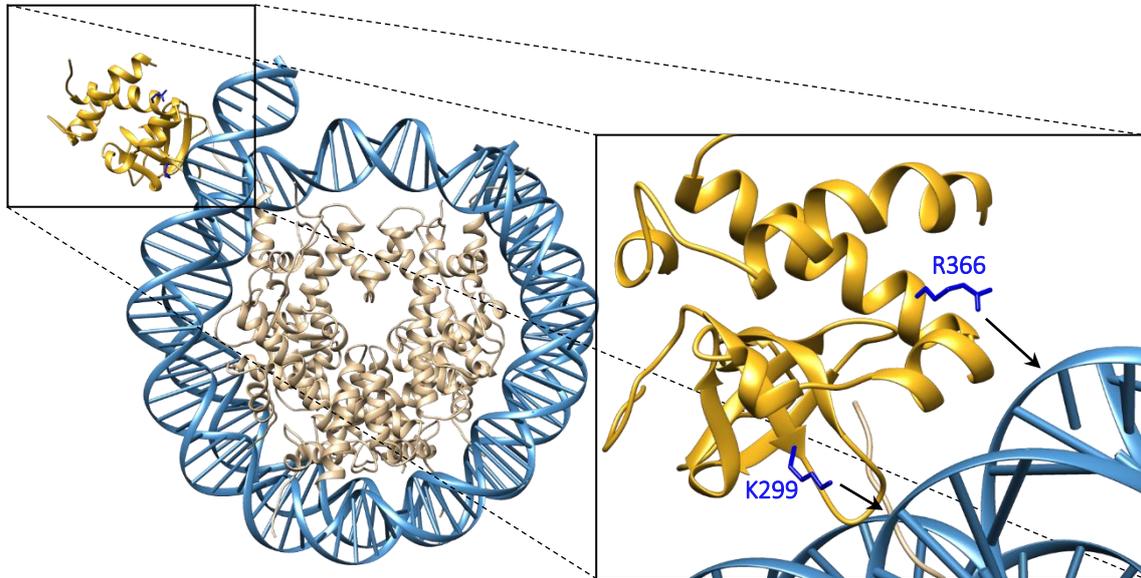


Figure 17 | Superposition of the DNMT3A PWWP domain (PBD: 3LLR) on the LEDGF PWWP structure bound to a nucleosome containing H3Kc36me3 (PDB: 6S01, Wang et al. NSBMB 27, 8–13, 2020). View down to the DNA superhelix axis. The two selected DNA binding residues are highlighted in blue and point into the direction of the nucleosomal DNA.

With the amount of data generated, it was not possible to determine further details of the potential sequence-dependent DNA interaction of the DNMT3A PWWP domain, for example, the size of the binding site or preferences for certain bases at specified positions would have been of great interest. With the charge reversal mutation of arginine 362 of the DNMT3A PWWP domain, it lost its DNA binding ability. The effect of the R362E mutation on the H3K36me_{2/3} interaction were very mild since almost equal binding compared to the wildtype to this modification was observed. This is in agreement with its position in the domain structure, since it has a large distance to the aromatic cage where K36me_{2/3} binding takes place (**Figure 18**).

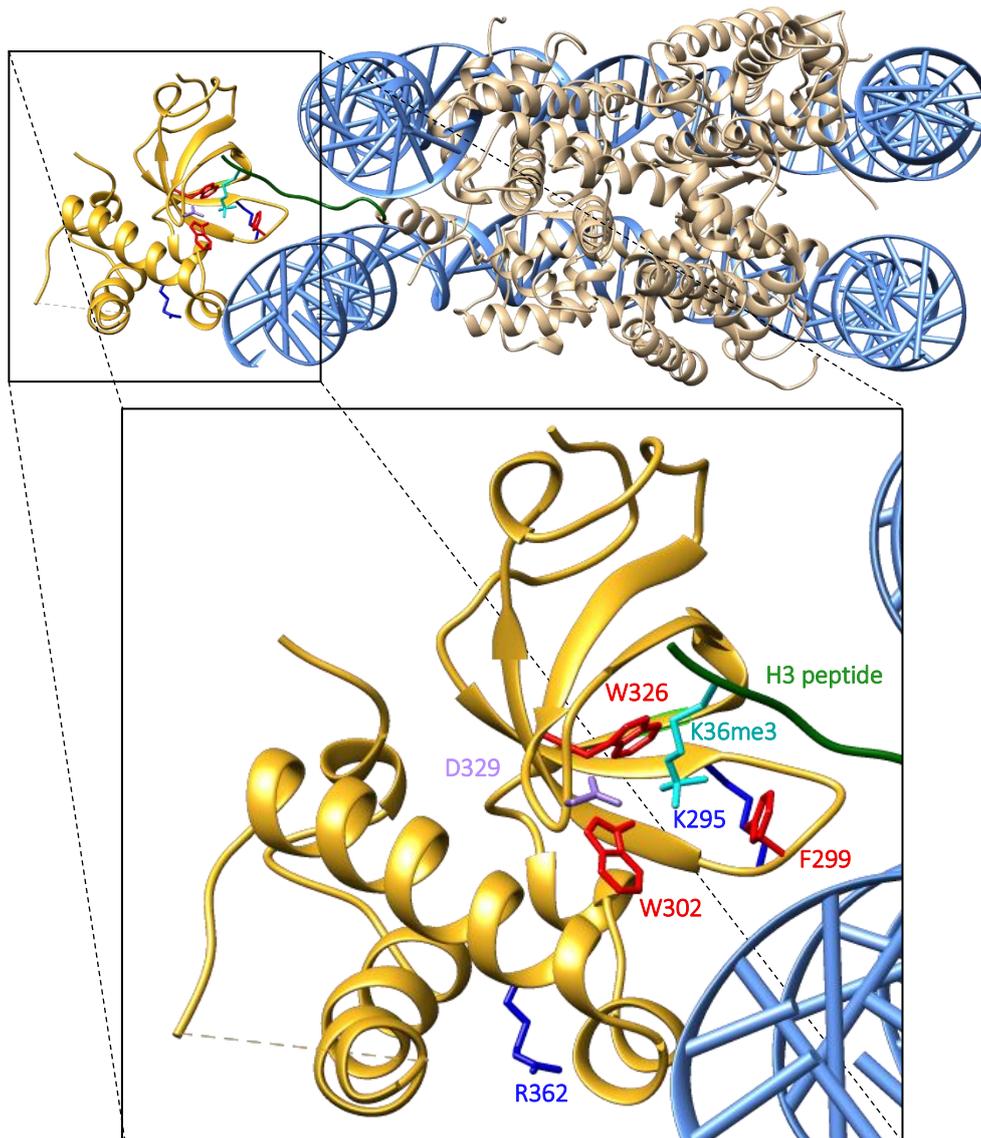


Figure 18 | Superposition of the DNMT3A PWWP domain (PBD: 3LLR) on the LEDGF PWWP structure bound to a nucleosome containing H3K36me3 (PDB: 6S01, Wang et al. NSBMB 27, 8–13, 2020). View perpendicular to the DNA superhelix axis. Red: residues of the aromatic cage for H3K36me2/3 binding, purple: D329 residue also required for H3K36me2/3 binding, green: H3 peptide, cyan: H3K36me3, blue: DNA binding residues K295 and R362.

However, chromatin interaction could no longer be observed with this mutation in the PWWP domain and the characteristic distribution of subnuclear localization with increased density in heterochromatic regions of transiently transfected DNMT3A1 full length was also reduced. This both indicates that DNA binding does play an important role in this process. The Lysine 295 mutation to glutamic acid showed neither detectable DNA binding nor H3K36me2/3 binding, which can be explained with its position in the same antiparallel β -sheet next to the aromatic cage residues F299 and W302 (**Figure 18**). Thus, it is possible that the mutation at K295 could cause, additionally to its loss of DNA binding, a change in the conformation of the aromatic

cage, even if the lysine 295 itself is not involved in direct contact with the trimethyl lysine of the H3 peptide or the H3 peptide itself. This double loss of both interactions resulted in an even more drastic delocalization of heterochromatic localization of the corresponding DNMT3A1 mutant.

Considering all these effects, combined binding of the DNMT3A PWWP domain via H3K36me_{2/3} and DNA to chromatin is very likely. Interestingly, the DNMT3A K299I mutation (corresponding to mouse K295I) was detected in paraganglioma (Remacha et al., 2018) and was therefore also investigated in more detail here with the available biochemical binding and localization assays. The observed effects of this mutant approximately mirrored those of the K295E mutant. Since both essential chromatin contact points are also disrupted in this disease-associated mutation, it strongly suggests that the pathogenic mechanism of this mutation is caused by the weakening of the DNMT3A-chromatin interaction. Similar effects were already shown for the W330R and D333N mutations in the K36me_{2/3} binding pocket of the human DNMT3A PWWP domain (Heyn et al., 2019).

4.2 Allosteric regulation of DNMT3A by competitive binding of MeCP2 and H3

The classical DNA methylation establishment and maintenance model, in which the *de novo* methyltransferases set the methylation pattern and the maintenance methyltransferase DNMT1 maintains it after each cell replication, has been revised based on several scientific findings and a dynamic, stochastic model has been proposed, in which the methylation status at each individual site depends on the local activity of methylating and demethylating enzymes as well as the DNA replication rate (review: Jeltsch and Jurkowska 2014). With this change in perspective, chromatin modifications and protein interactors have become more important in the recruitment and targeting of DNA methyltransferases. The dynamic DNA methylation landscape plays a particularly important role in non-dividing cells such as terminally differentiated neurons (Heyward & Sweatt, 2015; Shin et al., 2014; Weaver, 2014). Due to the lack of cell division, DNA methylation profiles can be controlled here exclusively by regulating DNA methyltransferases and demethylases. In this part of the thesis, I took a close look at the activity regulation of DNMT3A. In an already well-advanced project of Prof. Jeltsch's laboratory, the 5mC-reading protein MeCP2 was identified as a strong interaction partner of DNMT3A. In characterizing the interaction of these two epigenetic players, the TRD of MeCP2 and the ADD domain of DNMT3A were identified as mediators. This interaction resulted in a strong

inhibition of DNMT3A activity in vitro. Interestingly it was possible, in a competitive interaction assay, to disrupt the TRD-ADD binding by addition of unmodified histone H3 tail (aa 1-19) suggesting that H3 and TRD binding to the ADD domain are competitive.

I have further investigated this effect to test whether the strong inhibition of DNMT3A by the MeCP2 binding is released in presence of the H3 tail. My data have shown that the activity of free DNMT3A2 was increased by 1.6-fold by the unmodified H3 tail, whereas DNMT3A2 with bound MeCP2 showed a 3.0-fold stimulation by the unmodified H3 tail (**Appendix 2, Figure 6C**). This and the other data from this study provided novel insight into the complexities of regulation and control of *de novo* methyltransferases. DNMT3A alone is in an autoinhibitory conformation which upon binding to unmodified H3 tail changes into its active conformation, which allows it to set its marks (X. Guo et al., 2015). In certain gene regions, as for example in repetitive sequences, these methylation marks might recruit MeCP2 which could take over the recruitment of more DNMT3A enzymes. At present it cannot be decided if free MeCP2 first binds with its MBD-domain to the methylated DNA further recruiting DNMT3A or if a MeCP2-DNMT3A complex is recruited to the methylated DNA. If this process occurs near to nucleosomes with unmodified H3K4, the MeCP2-DNMT3A complex could release active DNMT3A from the complex via its interaction with the lysine 4. This could generate additional binding sites for MeCP2 through DNMT3A mediated DNA methylation, thereby creating a positive feedback loop due to new recruitment of MeCP2. This process could be further supported in neurons by DNMT3A set non-CpG methylation (mainly CpA methylation) which is also recognized by MeCP2. In contrast, the inhibitory effect of MeCP2 on DNMT3A in chromatin regions that have H3K4me2/3 modified marks (such as promoters of active genes) would prevent abnormal DNA methylation and thus untargeted gene silencing. Consequently, this work presents a model of allosteric regulation of DNMT3A (**Figure 19**) that, depending on the chromatin environment, uses MeCP2 as either a DNA methylation activator or repressor, depending on whether it recruits DNMT3A in a targeted manner or inhibits the enzyme.

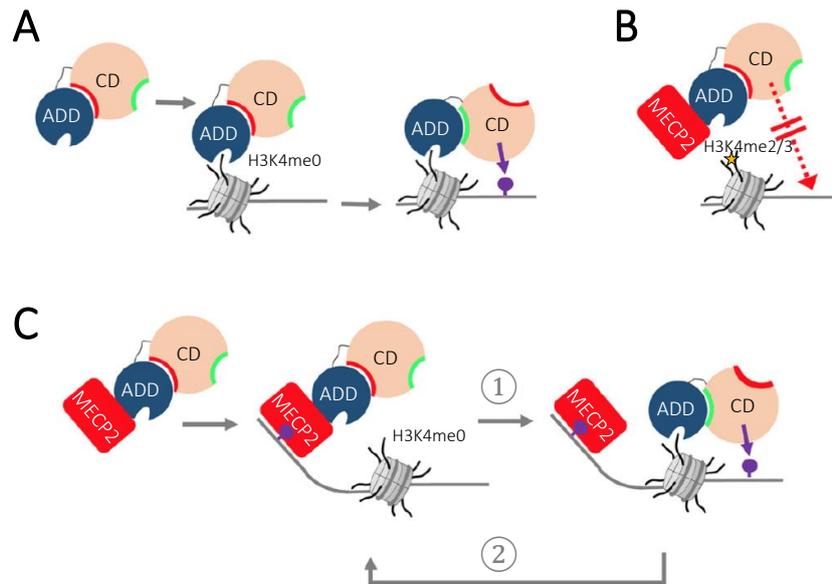


Figure 19| Model of the dual role of MeCP2 in the regulation and targeting of DNMT3A. A| Binding of unmodified H3K4 to the DNMT3A-ADD domain initiates a conformational switch where the ADD domain where the ADD domain moves from the autoinhibitory (red) to the allosteric (green) interaction site allowing DNA methylation to occur (purple). Model based on Guo et al. (2015). **B|** Inhibitory role of MeCP2 on DNMT3A. Binding of MeCP2 to DNMT3A-ADD stabilizes the autoinhibitory conformation and prevents untargeted activity **C|** Role of MeCP2 in targeted DNA methylation. At genomic sites with unmodified H3K4, H3 binding to the ADD domain disrupts the interaction between MeCP2 and DNMT3A leading to the activation and DNA methylation (step 1). MeCP2 further binds to the newly methylated CpG sites and recruits further DNMT3A, thereby initiating a positive feedback loop (step 2). Figure adapted from Appendix 2.

4.3 Mechanistic studies on cytosine N3 methylation by DNA methyltransferase DNMT3A

Until it was reported in 2018 that the *de novo* DNA methyltransferase DNMT3A and the bacterial methyltransferase M.SssI also generate small amounts of 3-N methylcytosine (Rošić et al., 2018), it was assumed that the methylation of cytosine at its N3 position does not arise enzymatically, but is mediated only chemically by endogenous or exogenous methyl donors (Pataillot-Meakin et al., 2016). This toxic base modification affects base pairing and has disruptive effects on local DNA functionality (Furrer & Van Loon, 2014). In non-malignant cells DNA repair enzymes of the Fe²⁺ dependent oxygenase family, in particular the ALKB2/3 homologs, are expected to remove this mark. In cancer however, DNA repair mechanisms and apoptotic pathways are impaired due to somatic mutations, allowing this lesion to persist and retain its influence on DNA (Pataillot-Meakin et al., 2016). The ability of DNMTs to introduce 3-methylcytosine has led to a coevolution of DNMTs and ALKB2/3 enzymes, which are needed to remove the DNA damage. However, the previous work left the question how this mark can

be introduced enzymatically and what is the exact enzymatic mechanism behind. To investigate this further, several amino acids from conserved motifs in the catalytic center of DNMT3A relevant for catalysis were mutated in this study. Specifically, these were the cysteine from motif IV (PCN motif), the glutamic acid from motif VI (ENV motif), and the second arginine from motif VIII (RXR motif) of the catalytic domain of DNMT3A. Each of these residues is involved in a temporary binding of the cytosine, when it is rotated out of the DNA and specific catalytic steps, thereby mediating the methyl transfer with the cofactor S-adenosyl-methionine (**Figure 12A**). As expected, the general activity of the 5mC generation was greatly reduced for all generated mutants, consistent with their conserved roles in the catalytic mechanism. With the reduction in catalytic 5mC generation, of course, no conclusions can yet be drawn about the mechanism of 3mC generation, but the relative rates of m5C and 3mC formation obtained from the LC-MS measurements were more informative in this regard. These ratios of the individual mutants served here as the basis for a proposed mechanism of 3mC formation, in which the cytosine is flipped out of the DNA helix, analogous to the 5mC formation, but the base is rotated 180° about the glycosidic bond and bound in this conformation into the catalytic pocket. The C5 position at which the cytosine is methylated in a regular arrangement is occupied in this position by the N-3 position of the base. With this arrangement, the methyl group of AdoMet is in close proximity to the N-3 of the cytosine and the transfer of the methyl group is possible (**Figure 20 right side**).

A closer look at the methylation results of the PCN mutants revealed that in the C710A mutant, C5 methylation was reduced more than N3 methylation, as illustrated by the increased 3mC/5mC ratio. It was concluded that N3 methylation does not include a C710-mediated covalent complex. The reduction in the absolute rate of 3mC formation from this mutant may be explained by the close distance of cysteine 710 to the cofactor S-adenosyl-methionine, since the exchange of cysteine to alanine may result in slight conformational changes, possibly affecting the binding of the cofactor. The exchange of the cysteine in the PCN motif to serine resulted in undetectable 3mC generation. This further supports the proposed model. The serine now is in perfect position to form a hydrogen bond with the cytosine N3 atom, if the base is in the rotated arrangement. Thus, the N3 of the cytosine would no longer be available for the methyl group transfer (**Appendix 3, Figure 1e**). When considering the regular mechanism of 5mC generation, it is noticeable that the R792 from the RXR motif, in addition to its contact to the deoxyribose, also forms a hydrogen bond to the O3 atom of the flipped

base, thus stabilizing the preferred orientation for 5mC formation (**Figure 12B and Figure 20**). Both exchanges of R792, either to alanine or to histidine, led to an extended 3mC/5mC ratio. This can be well explained by the loss of the hydrogen bond to the O3 atom and again points in the direction of the proposed mechanism. This time, the decrease in the absolute rate for 3mC generation of this mutant can be explained by the already mentioned, contact to the deoxyribose, which presumably supports base flipping for both, 5mC and 3mC generation, and is lost in the mutants. The ENV motif in the catalytic center of DNMT3A could form a hydrogen bond from the glutamic acid 756 to the exocyclic N4 amino group in both, the proposed mechanism of 3mC catalysis and the regular mechanism of 5mC catalysis. This would anchor the base in both mechanisms, highlighting the importance of this interaction for catalysis. At first glance, no change in the 3mC/5mC ratio would be expected by mutating E756 to alanine. Here, however, a slight increase of this ratio was observed, although not that strong as for the R792 mutants or the C706A mutant. At second glance, however, a bidental hydrogen bond from the glutamic acid to the flipped base would be possible in the case of 5mC generation. In addition to the contact with the N4 amino group of the base the arginine could also contact the N3 of the cytosine, since this is protonated during the reaction (**Figure 12B and Figure 20**). But as this contact is no longer present when E756 is mutated to alanine, the slight increase in the 3mC/5mC ratio during 3mC generation can be explained with this loss of interaction. In sum, the obtained results support the presented mechanism, in which the target cytosine is flipped by 180°, compared to the mechanism for the 5mC generation described in chapter 1.3.5, and anchored in the catalytic pocket in this position, whereby the methyl group is no longer transferred to the C5 but to the N3 of the base. Based on the highly conserved motifs in the different DNA methyltransferases, one might assume that a similar 'inverted base flipping' as mechanism described here for DNMT3A for the 3mC generation might also be valid for other 5mC methyltransferases. However, the active sites differ between these enzymes, which is why further experiments would be necessary for other 5mC methyltransferases to confirm or disprove further possible similarities between these enzymes.

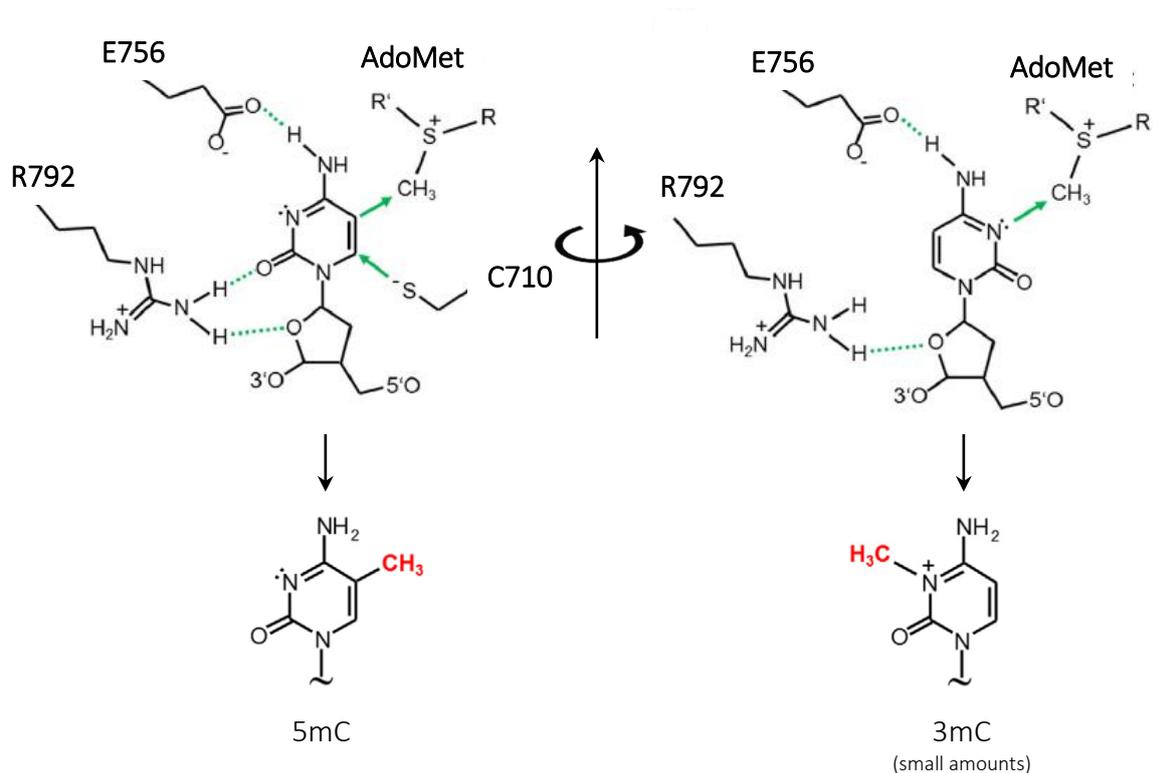


Figure 20| Proposed catalytic mechanisms for both, 5mC and 3mC generation. In the left part of the figure the cytosine base is in its preferred orientation with contacts to all three catalytical motifs which results in the main product 5-methyl cytosine. In the right part of the figure, the cytosine base is rotated by 180° around the z-axis, resulting in the non-preferred side-product of 3-methyl cytosine.

4.4 Investigation of the flanking sequence preference of *de novo* DNA methyltransferases

The deep enzymology approach developed in the Jeltsch laboratory, combined with the already in 2018 solved crystal structure of DNMT3A in complex with DNA (Z. M. Zhang et al., 2018) and three additional DNMT3B crystal structures in complex with DNA with different CpX target sites (Gao et al., 2020) helped to increase the understanding why these *de novo* C5-DNA-methyltransferases are responsible for the methylation of different target regions. With the use of the statistical analysis of the flanking sequence preferences of the individual flanking bases, clear preferences were obtained for both methyltransferases, especially in the plus minus 3 flanking range, which are partly overlapping (e.g. the high preference for T at the -2 site) but partly also distinct (in particular the preference for pyrimidines for DNMT3A at the +1 site versus the preference for purines for DNMT3B at the +1 site). This type of analysis also revealed, that absolute preferences are not just the sum of the preferences of the individual flanking positions. As for example one of the best-characterized targets for DNMT3B, the SatII

repeats (ATTCGATG), was among the top 4% of the most preferred targets of DNMT3B, although only the T at -2 is the only position out of six flanking nucleotides, where the base fits the to the most preferred target sequence. However, in this particular example, the flanking sequence preference provides an explanation for why the SatII repeats are so closely linked to DNMT3B. Hypomethylation of the SatII sequences, found in the ICF syndrome, is closely connected to mutations of DNMT3B and this cannot be compensated by DNMT3A due to their different target profiles. In addition, flanking sequence preferences now also provide an explanation for previously observed preferences in which DNMT3A was associated with major satellite repeats and DNMT3B with minor satellite repeats. In a previous publication based on the methylation levels of 48 CpG sites it was concluded that DNMT3L modulates cellular *de novo* methylation activities through focusing the DNA methylation machineries on chromatinized DNA templates and by influencing flanking preferences (Wienholz et al., 2010). In our work, a comparison of the complete and unbiased flanking sequence preferences of DNMT3A and DNMT3B in presence and absence of DNMT3L showed that the datasets with and without DNMT3L were highly correlated with each other (**Appendix 4, Supplementary Figure 9**). Therefore, it can be concluded that the flanking sequence preferences of the *de novo* DNMTs is shaped via direct substrate recognition by DNMT3A and DNMT3B and is unlikely to be modified by indirect interactions through DNMT3L.

In order to draw conclusions on the DNA recognition of individual amino acids of DNMT3B, several mutants were generated and their methylation profiles were compared with that of the wildtype. In general, the effects were more pronounced when residues of the TRD loop were mutated. This suggests a more dominant influence of this region on base recognition. The most dramatic effect was obtained by replacing T775. This exchange resulted in an almost complete loss of catalytic activity. A closer look at the crystal structures reveals that this threonine forms a water-mediated H-bridge to the N7 atom of the purine base (in both, the structure in CpG and in CpA context) and additionally it contacts the ribose of the target cytosine. The position of the threonine also occupies the intrahelical empty space created by the flipped cytosine base. This suggests that T775 has a supporting role in the stability of the transition state during target base flipping, because it can prevent the reverse rotation of the flipped cytosine back into the DNA where the methyl group cannot be transferred. The second largest effect of a mutation in DNMT3B was revealed at the level of flanking sequence

preference for the K777A mutation. The strong preference of DNMT3B for purines, especially for G, at the +1 position was completely removed by this mutation, which can be explained by the specific H-bond of this lysine side chain to the N7 of purine bases. In addition to the altered readout of the +1 flank of the K777A mutant, a pronounced loss of CpG specificity was also observed. This suggests a coupling of CpG and flank interactions. One explanation for this effect could be the approach of the C8 of the CpG guanine and the side chain methylene groups of K777. Like this the lysine 777 forms one wall of the binding pocket for the CpG guanine. The K777A mutation may have created additional space in the binding pocket and thereby increased the conformational freedom of the base at the CpX position, allowing DNMT3B to accept other bases than guanine. The N779 A/D mutations both had a strongly reduced preference for A at the +1 flanking site an effect, which was also present in the wildtype non-CpG methylation. An explanation for both reductions of preference for A at +1 may be the formation of an H-bond between N779 and the T'(+) O4 atom, which is only formed in the CpG complex. N779 seems to specifically important for the recognition of CGA trinucleotides, a sequence context which is also part of the SatII repeats where the CpG site is also followed by an A.

Comparison of the methylation patterns of the DNMT3B catalytic loop mutants with the methylation pattern of DNMT3B WT revealed increased CpT methylation for the V657A mutant and increased CpA methylation for the N658A mutant. N658 contacts ribose at the CpX site and the +1 flank in all complexes, and the data suggest that this contact is relevant for CpG recognition. Moreover, the observed changes in +1 flanking sequence preference of N658A were opposite to those of K777A, suggesting a role in buffering the effects of K777. Unfortunately, structural data of DNMT3B in complex with DNA in the CpT context do not yet exist, which leaves the increased CpT methylation of the V657A mutant difficult to be interpreted.

Interestingly, the N656A and N658A mutants from the catalytic loop and the N779A/D mutants from the TRD loop showed parallels in the reduction of A (+1) preference. This may suggest that a double-sided grip of DNA from the sides of the major and minor grooves is required to achieve a specific formation of the N779-AT' (+1) contact. For the methylation of GGCGGG sites, the N652, N658 and K777 residues seem to be obligatory, since this sequence context was methylated in neither of the mutants. Thus, it can be concluded that a tight grip of the DNA is necessary, especially for the binding of substrates with G-rich flanks. Moreover, the

residues of the catalytic loop appear to cooperate for the compensation of flanking sequence effects via an H-binding network which links them all together. For example, this is illustrated by a pronounced reduction in *Sat*II sequence preferences of the R661A mutant in which a disfavor is caused for T(-1). While R661 itself is not involved in DNA contacts, it appears to play an important role in stabilizing the conformation of the catalytic loop via H-bonds to N656 and N658.

4.5 Conclusion and perspectives

In summary, the results of this work provide new insights into the function of the two *de novo* DNA methyltransferases DNMT3A and DNMT3B. In addition to expanded insights into targeting and chromatin binding, mechanistic complexities were explored in more detail. These findings resulted from four different research approaches.

The first study revealed that efficient targeting of DNMT3A to heterochromatic chromatin regions with H3K36me_{2/3} marks is only possible, if next to the binding of this histone modification into the aromatic binding pocket of the PWWP-domain efficient DNA binding by two basic surface residues on the PWWP domain is also ensured. Next, it would be interesting to see if somatic cancer mutations of the PWWP domain have similar effects on the targeting of the enzyme. In the cosmic data base, it is noticeable that, for example, glycine 332, which is immediately adjacent to aspartic acid 333, and tryptophan 330, which is one of the residues for the aromatic binding pocket for H3K36me_{2/3}, are among the mutational hotspots in DNMT3A in cancers. On the basis of their positions, it can be speculated that similar effects can possibly be expected for these naturally occurring mutations as for the artificial mutants investigated here. Other mutation-prone amino acids of the DNMT3A PWWP domain are W313, L344, R320 and R326. Although these residues do not seem to be directly involved in the described binding properties of the enzyme, it still would be interesting to see what effects they have on the targeting through possible structural changes.

The second study investigated the interaction of DNMT3A and the 5mC-binding protein MeCP2, focusing on the regulation of the methyltransferase activity. The observed effects highlighted the high importance of the allosteric control of DNMT3A. It became clear that protein-protein interactions, here shown with MeCP2, can severely restrict activity, but at the same time it can also create new targeting possibilities. MeCP2 is also known to recognize CpA

methylation, a mark which is also introduced by DNMT3A, this raises the question to what extent non-CpG methylation might be involved in the recruitment of this and other chromatin reading proteins which recognize this mark. It would be interesting to map the interface on the surfaces of both proteins. With targeted mutations of the residues responsible for the binding on one of the two proteins loss of interaction could be shown by pull-down assays. With this information, it would be possible to study effects of MeCP2-DNMT3A interaction in neurons. For this one could introduce mutations in DNMT3A or MeCP2 that disrupt the interaction directly in neuronal cell lines by CRISPR/Cas mediated genome editing (Anzalone et al., 2019) and study the effects on genomic methylation patterns.

In the mechanistic study of DNMT3A in relation to the introduction of 3mC, a mechanism was proposed, based on the results of this work, which differs only by a 180° rotation of the cytosine longitudinally around the glycosidic bond to the deoxyribose. However, it is still unclear, if other DNA-C5-methyltransferases have a similar error rate as DNMT3A due to this 'inverted base flipping' mechanism for the generation of 3mC and further investigations would be necessary to solve this issue. Another interesting question in this regard would also be whether an increased 3mC generation is observed for commonly occurring DNMT3A mutations, such as the mutational hotspot R882. In addition to reduced activity and altered flanking sequence preference, as described for this DNMT3A mutant, increased 3mC generation could influence oncogenic signaling pathways.

The studies on the flanking sequence preferences of DNMT3A and DNMT3B provided the molecular basis of target recognition and could partly explain the preferences on the +1 site. The more in-depth study of DNMT3B revealed a complex balancing of flanking effects. In addition, a high correlation of non-CpG flanking sequence preferences of DNMT3B and non-CpG methylation in human cells was demonstrated which highlights this a major task of DNMT3B. The interesting differences in the flanking sequence preferences of human and mouse DNMT3B should be connected with particular amino acid residues as both enzymes only differ at 20 residues in their catalytic domains and the exchanges of human K782 to murine R788 and human G705 to murine N711 are most interesting, because they are close to the DNA and involve residues with DNA recognition potential. These mutants should be prepared and investigated regarding their flanking sequence preferences. A similarly detailed mutational study as conducted here for DNMT3B is still lacking for DNMT3A.

Last but not least, the results from this work highlight once more, that many different regulatory and mechanistic properties of *de novo* methyltransferases are important for a normal development and for correctly balanced methylation in differentiated cells. However, it remains a major task of our field to link data obtained from in vitro characterizations of the different epigenetic regulators with the cellular pathways in which they are involved. A better understanding of these pathways may find application in medical research, to ensure an even longer and healthier life than our society has already achieved.

5 Materials and Methods

All methods applied in this work are described in detail in the method parts in **Appendices 1 to 5**. A summary of the most important methods is provided in the following sections.

5.1 Cloning, mutagenesis, expression and purification of recombinant proteins

Every wild type DNA sequence mentioned in this study was already available in the lab either in the pET-28a(+) vector (Novagen) or in the pGEX-6P-2 vector (GE Healthcare). The generation of point mutations was either done using the megaprimer based site directed mutagenesis method (Jeltsch & Lanio, 2002) or performed by inverse PCR amplifying the entire vector encoding the WT sequence with Q5[®] High-Fidelity DNA Polymerase (NEB) using primers with up to three mismatches at the 5' end to introduce the desired mutation. The complete list of plasmids used in this thesis is shown in **Table 1**. Nomenclature for mouse genes and proteins was used according to the International Committee on Standardized Genetic Nomenclature for Mice (Jackson Lab) <http://www.informatics.jax.org/mgihome/nomen/gene.shtml> . Nomenclature for human genes and proteins was used according to Oxford University press http://www.oxfordjournals.org/our_journals/molehr/for_authors/gene_and_protein_nomenclature.html .

The competent *Escherichia coli* BL21-CodonPlus-RIL cells (Stratagene) were used for protein overexpression and usually grown in LB-broth with required antibiotics at 37°C until an OD₆₀₀ of 0.6 - 0.8 was reached. For the protein expression the temperature was lowered to 20°C and the induction was initiated by adding 500 mM Isopropyl-β-D-thiogalactopyranosid.

For the protein purification the cells were lysed by sonification and the cell debris was removed by centrifugation. The proteins of interest were isolated by affinity chromatography using the suitable agarose beads depending of the protein tag. The protein concentrations were measured by UV absorption at 280 nm and calculated via Lambert Beer's law using theoretical extinction coefficients which are based on the amino acid composition of the protein and determined with the ProtParam tool of <https://www.expasy.org/> . Storage took place at -80°C, in buffers containing 10% glycerol if not otherwise specified. The purity of the proteins was determined with Coomassie brilliant blue staining of SDS polyacrylamide gels for each batch of purification.

Table 1 : Plasmids used in this thesis listed with further information of boundaries, domain, accession number (NCBI), vector backbone and tag used as fluorophore or for purification.

insert	boundaries [aa]	domain	accession number	vector backbone	protein tag
<i>DNMT3A WT</i>	612-912	cat. domain	Q9Y6K1	pET-28a(+)	His
<i>DNMT3A C710A</i>	612-912	cat. domain	Q9Y6K1	pET-28a(+)	His
<i>DNMT3A C710S</i>	612-912	cat. domain	Q9Y6K1	pET-28a(+)	His
<i>DNMT3A C710S E756A</i>	612-912	cat. domain	Q9Y6K1	pET-28a(+)	His
<i>DNMT3A E756A</i>	612-912	cat. domain	Q9Y6K1	pET-28a(+)	His
<i>DNMT3A R792A</i>	612-912	cat. domain	Q9Y6K1	pET-28a(+)	His
<i>DNMT3A R792H</i>	612-912	cat. domain	Q9Y6K1	pET-28a(+)	His
<i>Dnmt3a WT</i>	279-420	PWWP	O88508	pGEX-6P-2	GST
<i>Dnmt3a K295E</i>	279-420	PWWP	O88508	pGEX-6P-2	GST
<i>Dnmt3a K295I</i>	279-420	PWWP	O88508	pGEX-6P-2	GST
<i>Dnmt3a D329A</i>	279-420	PWWP	O88508	pGEX-6P-2	GST
<i>Dnmt3a R362E</i>	279-420	PWWP	O88508	pGEX-6P-2	GST
<i>DNMT3A1 WT</i>	1-912	full length	Q9Y6K1	mVenus C1	mVenus
<i>DNMT3A1 K299E</i>	1-912	full length	Q9Y6K1	mVenus C1	mVenus
<i>DNMT3A1 K299I</i>	1-912	full length	Q9Y6K1	mVenus C1	mVenus
<i>DNMT3A1 D333A</i>	1-912	full length	Q9Y6K1	mVenus C1	mVenus
<i>DNMT3A1 R366E</i>	1-912	full length	Q9Y6K1	mVenus C1	mVenus
<i>DNMT3B WT</i>	553-853	cat. domain	Q9UBC3	pET-28a(+)	His
<i>DNMT3B N652A</i>	553-853	cat. domain	Q9UBC3	pET-28a(+)	His
<i>DNMT3B N656A</i>	553-853	cat. domain	Q9UBC3	pET-28a(+)	His
<i>DNMT3B N656D</i>	553-853	cat. domain	Q9UBC3	pET-28a(+)	His
<i>DNMT3B V657A</i>	553-853	cat. domain	Q9UBC3	pET-28a(+)	His
<i>DNMT3B N658A</i>	553-853	cat. domain	Q9UBC3	pET-28a(+)	His
<i>DNMT3B R661A</i>	553-853	cat. domain	Q9UBC3	pET-28a(+)	His
<i>DNMT3B T775A</i>	553-853	cat. domain	Q9UBC3	pET-28a(+)	His
<i>DNMT3B T775N</i>	553-853	cat. domain	Q9UBC3	pET-28a(+)	His
<i>DNMT3B T775Q</i>	553-853	cat. domain	Q9UBC3	pET-28a(+)	His
<i>DNMT3B K777A</i>	553-853	cat. domain	Q9UBC3	pET-28a(+)	His
<i>DNMT3B N779D</i>	553-853	cat. domain	Q9UBC3	pET-28a(+)	His
<i>DNMT3B R823A</i>	553-853	cat. domain	Q9UBC3	pET-28a(+)	His
<i>Dnmt3a WT</i>	221-908	full length isoform 2	O88508	pET-28a(+)	His
<i>MECP2 WT</i>	170-325	TRD	P51608	pGEX-6P-2	GST
<i>Dnmt3b WT</i>	558-859	cat. domain	O88509	pET-28a(+)	His
<i>Dnmt3b K783A</i>	558-859	cat. domain	O88509	pET-28a(+)	His

5.2 Biotin-avidin microplate assay

For the determinations of the catalytic activity of the *de novo* DNA methyltransferases, the biotin avidin microplate assay was used (Liebert & Jeltsch, 2008; Roth & Jeltsch, 2000). For this assay, the DNA substrate must be biotinylated. The basic idea behind this assay is to methylate the DNA in vitro with tritium-labelled AdoMet and then quench the reaction after certain incubations times in a solution with non-radioactively labelled AdoMet in wells of an avidin coated microplate. Now the high binding affinity of biotin to avidin causes the biotinylated substrates, methylated and unmethylated, to bind to the avidin-coated microplate. Excess of unreacted tritium labelled AdoMet, unbound substrate and enzymes are washed away. After these washing steps, the DNA is released by an unspecific endo-nuclease digestion. After transfer of the digested DNA into scintillation vials the activity can be detected by liquid scintillation counting. The illustrated workflow of this assay is shown in **Figure 21**.

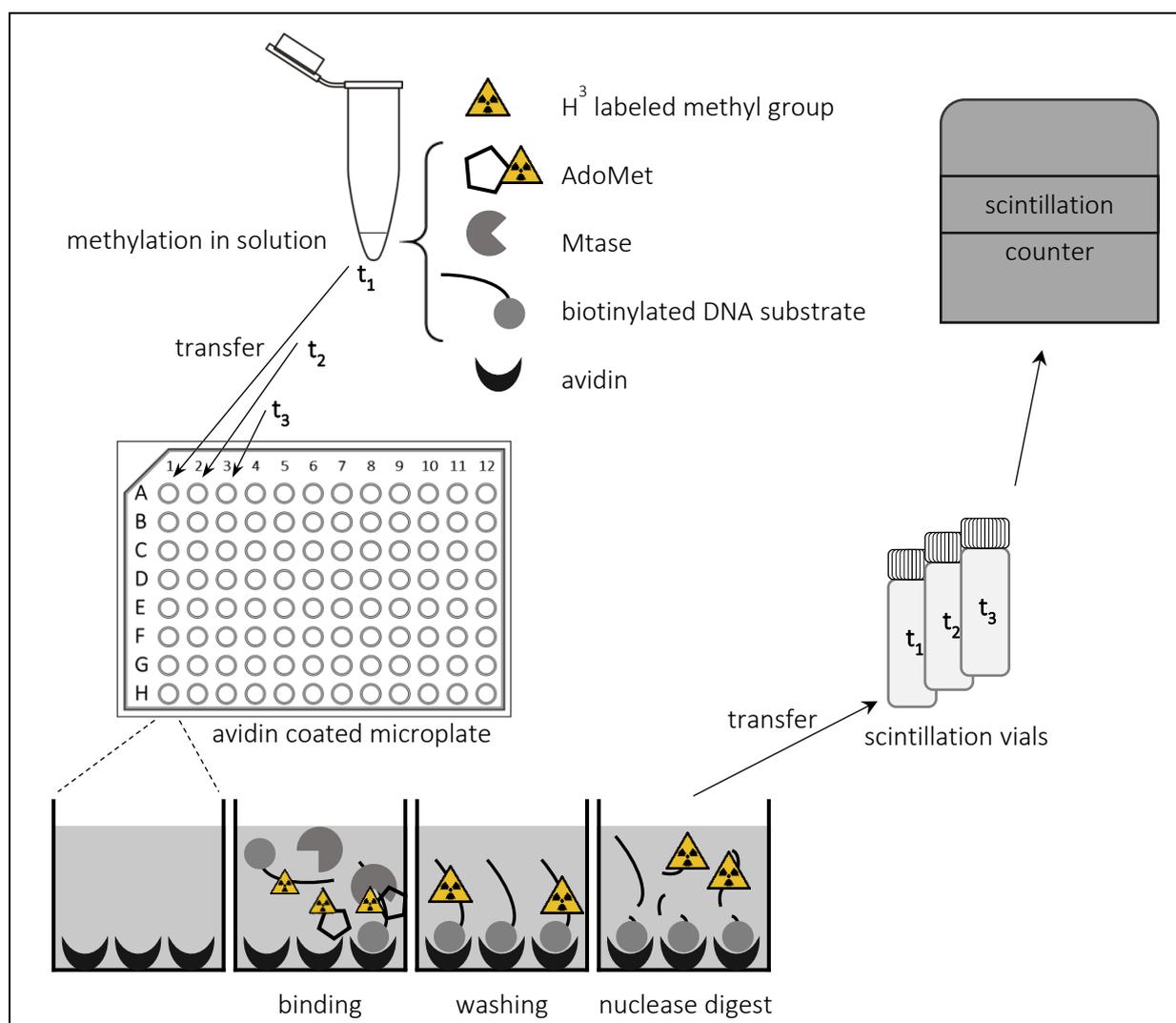


Figure 21| Schematic illustration of the biotin-avidin microplate assay. (Figure adapted from Monti (2012))

5.3 Flanking sequence preference analysis and library generation

The flanking sequence preferences of the DNMTs and their different variants were determined using DNA substrates with either centered hemi-methylated CpG or unmethylated CpN sites with ± 10 randomized flanking bases. The substrates were methylated by the DNMTs followed by bisulfite conversion to be able to distinguish later between the methylated and unmethylated cytosines in the CpG or CpN context. For the library generation each sample was further amplified with two short successive PCRs to add individual barcode and index combinations. After this the samples were pooled and sent for Illumina sequencing. A more detailed description of the method can be found in Gao et al. (2020) and a schematic illustration of the workflow is depicted in **Figure 22**.

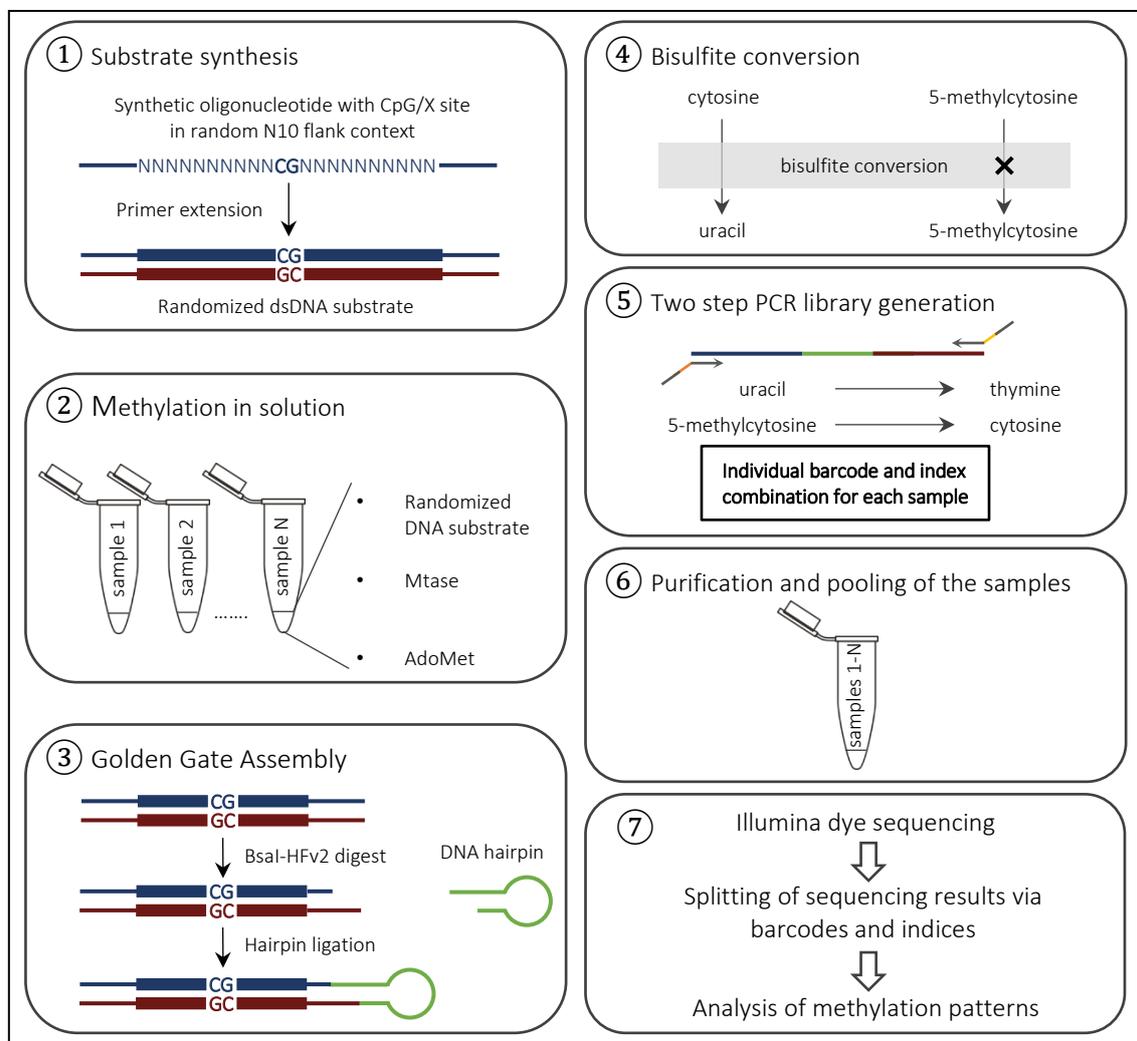


Figure 22 | Schematic process of the deep enzymology workflow.

5.4 Circular dichroism spectroscopy

Circular dichroism is the property of chiral molecules to absorb left- and right-handed polarized light differently. This property is used in CD spectroscopy in the UV range for the structural analysis of proteins. The folding of recombinant purified regulatory protein domains was analyzed using this analytical method with a J-815 CD spectrophotometer (JASCO Cooperation, Tokyo, Japan). The spectra were recorded at a scan speed of 100 nm/min over a wavelength range of 190 to 250 nm at 20°C. Further details of the measurements are described in **Appendix 1**.

5.5 Peptide arrays

To study binding of the DNMT3A PWWP domain to histone tails with different modifications, Celluspot peptide arrays (Active Motif, Carlsbad, USA) were used. These peptide arrays have 384 distinct peptide sequences in duplicates from 8 regions of the N-terminal histone tails (H3 1-19, 7-26, 16-35 and 26-45, H4 1-19 and 11-30, H2A 1-19 and H2B 1-19) with 59 different post-translational modifications. Protein binding experiments to the peptide arrays was performed using the protocol described in Bock et al. (2011).

Further details on buffers, antibodies and detection methods used can be found in the methods section of **Appendix 1**.

5.6 Electrophoretic Mobility Shift Assay (EMSA)

EMSA is a method for the detection of protein-DNA or protein-RNA interactions which takes advantage of the fact that nucleic acids bound to proteins migrate more slowly through polyacrylamide or agarose gels during gel electrophoresis than the same unbound nucleic acids. This assay was used for binding studies of the PWWP-domain of DNMT3A and a short Cy5-labeled 30-mer DNA with one central CpG site. A detailed description of the procedure is described in the method section of **Appendix 1**.

5.7 Tissue culture, harvesting and transfections

NiH-3T3 as well as HepG2 cells were grown at 37°C in a 5% CO₂ atmosphere using Dulbecco's modified Eagle's medium (DMEM) high glucose supplemented with 10% heat-inactivated fetal calf serum.

HepG2 cells were cultured for later isolation of nucleosomes and therefor grown in T175-flaks until they reached a confluency of >75%. For the harvesting of the cells they were washed once

with 10 mL PBS without calcium chloride and magnesium chloride and incubated for 10 minutes at 37°C with 20 $\mu\text{L}/1\text{cm}^2$ trypsin. To stop the trypsinization 10.5 mL DMEM were added and the detached cells were transferred into a 15 mL falcon tube and spun down with 300 x g for 5 min. After a washing step with 5 mL PBS the cells were again spun down with the same settings and the supernatant was discarded. The cells were then either directly used for further experiments or shock-frozen in liquid nitrogen and stored at -80°C.

NiH-3T3 cells were cultured for localization studies of Venus-tagged full length DNMT3A1 variants. The cells were grown until they reached a confluency of ~50% and then transfected with the corresponding DNMT3A1 expression plasmid using FuGENE®HD Transfection Reagent. A more detailed description of the conditions, imaging, DNA staining, and further analysis is listed in **Appendix 1**.

5.8 Isolation of nucleosomes

Native oligonucleosomes were obtained with a modified protocol described in Kasinathan et al. (2014). Since the description of the mentioned protocol refers to mononucleosomes, milder conditions for MNase digestion were chosen here to yield oligonucleosomes. A detailed description of the isolation of oligonucleosomes is given in the methods section of **Appendix 1**.

5.9 Nucleosomal pull-down assay

To study the binding of DNMT3A-PWWP to native oligonucleosomes, pull-down experiments were performed using GST-tagged DNMT3A-PWWP domain fusion proteins as bait and isolated nucleosomes as prey. Binding conditions and description of the protocol is shown in the methods section of **Appendix 1**.

5.10 Western blot and antibody detection

For the identification of specific protein sequences, the samples were separated by size on an SDS-polyacrylamide gel at constant amperage of ~ 30 milliampere/gel. After this, the proteins were transferred on a nitrocellulose membrane using a wet tank blotting system (Bio-Rad, USA). To reduce background signals due to unspecific binding of the antibodies, the membranes were blocked with 5% (m/v) milk in TBS-T for 1 hour at room temperature. The membranes were then incubated over night at 4°C together with protein-tag or sequence specific antibodies in 1% (m/v) milk in TBS-T. Host specific HRP coupled antibodies were used to detect the first antibodies. Signals were obtained with SuperSignal™ West Femto Maximum

Sensitivity Substrate (Thermo Fisher Scientific, USA). For imaging the chemiluminescent signals, either X-ray films or the digital imager FUSION advance solo 4 (Peqlab) were used. For more details see **Appendix 1**.

6 References

- Aapola, U., Shibuya, K., Scott, H. S., Ollila, J., Vihinen, M., Heino, M., Shintani, A., Kawasaki, K., Minoshima, S., Krohn, K., Antonarakis, S. E., Shimizu, N., Kudoh, J., & Peterson, P. (2000). Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics*, *65*(3), 293–298. <https://doi.org/10.1006/geno.2000.6168>
- Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*, *17*(8), 487–500. <https://doi.org/10.1038/nrg.2016.59>
- Amir, R. E., Van Den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, *23*(2), 185–188. <https://doi.org/10.1038/13810>
- Anastasiadou, C., Malousi, A., Maglaveras, N., & Kouidou, S. (2011). Human epigenome data reveal increased CpG methylation in alternatively spliced sites and putative exonic splicing enhancers. *DNA and Cell Biology*, *30*(5), 267–275. <https://doi.org/10.1089/dna.2010.1094>
- Anteneh, H., Fang, J., & Song, J. (2020). Structural basis for impairment of DNA methylation by the DNMT3A R882H mutation. *Nature Communications*, *11*(1), 1–12. <https://doi.org/10.1038/s41467-020-16213-9>
- Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), 149–157. <https://doi.org/10.1038/s41586-019-1711-4>
- Aran, D., Sabato, S., & Hellman, A. (2013). DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biology*, *14*(3). <https://doi.org/10.1186/gb-2013-14-3-r21>
- Arand, J., Spieler, D., Karius, T., Branco, M. R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V., & Walter, J. (2012). In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genetics*, *8*(6). <https://doi.org/10.1371/journal.pgen.1002750>
- Ausió, J., de Paz, A. M. artíne., & Esteller, M. (2014). MeCP2: the long trip from a chromatin protein to neurological disorders. *Trends in Molecular Medicine*, *20*(9), 487–498. <https://doi.org/10.1016/j.molmed.2014.03.004>
- Avvakumov, G. V., Walker, J. R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C. H., & Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*, *455*(7214), 822–825. <https://doi.org/10.1038/nature07273>
- Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395. <https://doi.org/10.1038/cr.2011.22>
- Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Héroult, Y., Guillou, F., & Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science*, *354*(6314), 909–912. <https://doi.org/10.1126/science.aah5143>
- Barrès, R., Osler, M. E., Yan, J., Rune, A., Fritz, T., Caidahl, K., Krook, A., & Zierath, J. R. (2009). Non-CpG Methylation of the PGC-1 α Promoter through DNMT3B Controls Mitochondrial Density. *Cell Metabolism*, *10*(3), 189–198. <https://doi.org/10.1016/j.cmet.2009.07.011>
- Bashtrykov, P., Jankevicius, G., Smarandache, A., Jurkowska, R. Z., Ragozin, S., & Jeltsch, A. (2012). Specificity of dnmt1 for methylation of hemimethylated CpG sites resides in its

- catalytic domain. *Chemistry and Biology*, 19(5), 572–578. <https://doi.org/10.1016/j.chembiol.2012.03.010>
- Bashtrykov, P., Ragozin, S., & Jeltsch, A. (2012). Mechanistic details of the DNA recognition by the Dnmt1 DNA methyltransferase. *FEBS Letters*, 586(13), 1821–1823. <https://doi.org/10.1016/j.febslet.2012.05.026>
- Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., Akalin, A., & Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*, 520(7546), 243–247. <https://doi.org/10.1038/nature14176>
- Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C. A., & Zoghbi, H. Y. (2009). Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. *Human Molecular Genetics*, 18(13), 2431–2442. <https://doi.org/10.1093/hmg/ddp181>
- Berger, S. L., Kouzarides, T., Shiekhhattar, R., & Shilatifard, A. (2009). An operational definition of epigenetics. *Genes and Development*, 23(7), 781–783. <https://doi.org/10.1101/gad.1787609>
- Bernstein, B. E., Birney, E., Dunham, I., E.D.Green, Gunter, C., & Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74. <https://doi.org/10.1038/nature11247>
- Bock, I., Dhayalan, A., Kudithipudi, S., Brandt, O., Rathert, P., & Jeltsch, A. (2011). Detailed specificity analysis of antibodies binding to modified histone tails with peptide arrays. *Epigenetics*, 6(2), 256–263. <https://doi.org/10.4161/epi.6.2.13837>
- Bogdanović, O., & Lister, R. (2017). DNA methylation and the preservation of cell identity. *Current Opinion in Genetics and Development*, 46, 9–14. <https://doi.org/10.1016/j.gde.2017.06.007>
- Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., & Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nature Genetics*, 42(12), 1093–1100. <https://doi.org/10.1038/ng.708>
- Bostick, M., Kim, J. K., Estève, P.-O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science*, 317(9), 1760–1764.
- Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science*, 294(5551), 2536–2539. <https://doi.org/10.1126/science.1065848>
- Brenner, C., Deplus, R., Didelot, C., Lorient, A., Viré, E., De Smet, C., Gutierrez, A., Danovi, D., Bernard, D., Boon, T., Pelicci, P. G., Amati, B., Kouzarides, T., De Launoit, Y., Di Croce, L., & Fuks, F. (2005). Myc represses transcription through recruitment of DNA methyltransferase corepressor. *EMBO Journal*, 24(2), 336–346. <https://doi.org/10.1038/sj.emboj.7600509>
- Calo, E., & Joanna, W. (2013). Modification of enhancer chromatin: what, how and why? *Molecular Cell*, 49(1), 825–837. <https://doi.org/10.1016/j.molcel.2013.01.038>
- Chahrour, M., & Zoghbi, H. Y. (2007). The Story of Rett Syndrome: From Clinic to Neurobiology. *Neuron*, 56(3), 422–437. <https://doi.org/10.1016/j.neuron.2007.10.001>
- Challen, G. A., Sun, D., Mayle, A., Jeong, M., Luo, M., Rodriguez, B., Mallaney, C., Celik, H., Yang, L., Xia, Z., Cullen, S., Berg, J., Zheng, Y., Darlington, G. J., Li, W., & Goodell, M. A. (2014). Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells. *Cell Stem Cell*, 15(3), 350–364. <https://doi.org/10.1016/j.stem.2014.06.018>
- Chang, Y., Sun, L., Kokura, K., Horton, J. R., Fukuda, M., Espejo, A., Izumi, V., Koomen, J. M.,

- Bedford, M. T., Zhang, X., Shinkai, Y., Fang, J., & Cheng, X. (2011). MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. *Nature Communications*, 2(1). <https://doi.org/10.1038/ncomms1549>
- Chen, L., Chen, K., Lavery, L. A., Baker, S. A., Shaw, C. A., Li, W., & Zoghbi, H. Y. (2015). MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, 112(22), E2982. <https://doi.org/10.1073/pnas.1507794112>
- Chen, T., Tsujimoto, N., & Li, E. (2004). The PWWP Domain of Dnmt3a and Dnmt3b Is Required for Directing DNA Methylation to the Major Satellite Repeats at Pericentric Heterochromatin. *Molecular and Cellular Biology*, 24(20), 9048–9058. <https://doi.org/10.1128/mcb.24.20.9048-9058.2004>
- Chen, T., Ueda, Y., Dodge, J. E., Wang, Z., & Li, E. (2003). Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Molecular and Cellular Biology*, 23(16), 5594–5605. <https://doi.org/10.1128/mcb.23.16.5594-5605.2003>
- Chen, Z., Chen, Z., Chen, Z., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, Y., & Zhang, Y. (2019). Role of Mammalian DNA Methyltransferases in Development. *Annual Review of Biochemistry*, 89, 135–158. <https://doi.org/10.1146/annurev-biochem-103019-102815>
- Cheng, X., Kumar, S., Klimasauskas, S., & Roberts, R. J. (1993). Crystal structure of the HhaI DNA methyltransferase. *Cold Spring Harbor Symposia on Quantitative Biology*, 58, 331–338. <https://doi.org/10.1101/SQB.1993.058.01.039>
- Cheng, Xiaodong. (1995). STRUCTURE AND FUNCTION OF DNA METHYLTRANSFERASES. *Annual Review of Biophysics and Biomolecular Structure*, 24, 293–318.
- Chin, H. G., Pradhan, S., & Este, P. (2005). Methyltransferase and P53 Modulate Expression of P53-Repressed Promoters. *Pnas*.
- Crick, F., & Watson, J. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171, 73.
- Croce, L. Di, Raker, V. A., Crosaro, M., Razi, F., Fanelli, M., Faretta, F., Fuks, F., Coco, F. Lo, Kouzarides, T., Nervi, C., Minucci, S., & Pelicci, P. G. (2003). Methyltransferase Recruitment and DNA Hypermethylation of Target Promoters by an Oncogenic Transcription Factor. *Chemtracts*, 16(13), 789–793.
- Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W., & Richmond, T. J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *Journal of Molecular Biology*, 319(5), 1097–1113. [https://doi.org/10.1016/S0022-2836\(02\)00386-8](https://doi.org/10.1016/S0022-2836(02)00386-8)
- De Greef, J. C., Wang, J., Balog, J., Den Dunnen, J. T., Frants, R. R., Straasheijm, K. R., Aytekin, C., Van Der Burg, M., Duprez, L., Ferster, A., Gennery, A. R., Gimelli, G., Reisli, I., Schuetz, C., Schulz, A., Smeets, D. F. C. M., Sznajer, Y., Wijmenga, C., Van Eggermond, M. C., ... Van Der Maarel, S. M. (2011). Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. *American Journal of Human Genetics*, 88(6), 796–804. <https://doi.org/10.1016/j.ajhg.2011.04.018>
- Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes and Development*, 25(10), 1010–1022. <https://doi.org/10.1101/gad.2037511>
- Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R. Z., Ragozin, S., & Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *Journal of Biological Chemistry*, 285(34), 26114–26120. <https://doi.org/10.1074/jbc.M109.089433>

- Dodge, J. E., Okano, M., Dick, F., Tsujimoto, N., Chen, T., Wang, S., Ueda, Y., Dyson, N., & Li, E. (2005). Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. *Journal of Biological Chemistry*, *280*(18), 17986–17991. <https://doi.org/10.1074/jbc.M413246200>
- Du, Q., Luu, P. L., Stirzaker, C., & Clark, S. J. (2015). Methyl-CpG-binding domain proteins: Readers of the epigenome. *Epigenomics*, *7*(6), 1051–1073. <https://doi.org/10.2217/epi.15.39>
- Dukat, M., Holzer, K., Choudalakis, M., Emperle, M., Lungu, C., Bashtrykov, P., & Jeltsch, A. (2019). H3K36me2/3 Binding and DNA Binding of the DNA Methyltransferase DNMT3A PWWP Domain Both Contribute to its Chromatin Interaction. *Journal of Molecular Biology*, *431*(24). <https://doi.org/10.1016/j.jmb.2019.09.006>
- Dukat, M., Requena, C. E., Emperle, M., Hajkova, P., Sarkies, P., & Jeltsch, A. (2019). Mechanistic Insights into Cytosine-N3 Methylation by DNA Methyltransferase DNMT3A. *Journal of Molecular Biology*, *431*(17). <https://doi.org/10.1016/j.jmb.2019.06.015>
- Edwards, J. R., O'Donnell, A. H., Rollins, R. A., Peckham, H. E., Lee, C., Milekic, M. H., Chanrion, B., Fu, Y., Su, T., Hibshoosh, H., Gingrich, J. A., Haghghi, F., Nutter, R., & Bestor, T. H. (2010). Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome Research*, *20*(7), 972–980. <https://doi.org/10.1101/gr.101535.109>
- Egger, G., Jeong, S., Escobar, S. G., Cortez, C. C., Li, T. W. H., Saito, Y., Yoo, C. B., Jones, P. A., & Liang, G. (2006). Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(38), 14080–14085. <https://doi.org/10.1073/pnas.0604602103>
- Eidahl, J. O., Crowe, B. L., North, J. A., McKee, C. J., Shkriabai, N., Feng, L., Plumb, M., Graham, R. L., Gorelick, R. J., Hess, S., Poirier, M. G., Foster, M. P., & Kvaratskhelia, M. (2013). Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic Acids Research*, *41*(6), 3924–3936. <https://doi.org/10.1093/nar/gkt074>
- Emperle, M., Adam, S., Kunert, S., Dukat, M., Baude, A., Plass, C., Rathert, P., Bashtrykov, P., & Jeltsch, A. (2019). Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns. *Nucleic Acids Research*, *47*(21). <https://doi.org/10.1093/nar/gkz911>
- Emperle, M., Rajavelu, A., Kunert, S., Arimondo, P. B., Reinhardt, R., Jurkowska, R. Z., & Jeltsch, A. (2018). The DNMT3A R882H mutant displays altered flanking sequence preferences. *Nucleic Acids Research*, *46*(6), 3130–3139. <https://doi.org/10.1093/nar/gky168>
- Emperle, M., Rajavelu, A., Reinhardt, R., Jurkowska, R. Z., & Jeltsch, A. (2014). Cooperative DNA binding and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase. *Journal of Biological Chemistry*, *289*(43), 29602–29613. <https://doi.org/10.1074/jbc.M114.572032>
- Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M., & Bernstein, B. E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, *473*(7345), 43–49. <https://doi.org/10.1038/nature09906>
- Feng, J., Zhou, Y., Campbell, S. L., Le, T., Li, E., Sweatt, J. D., Alcino, J., & Fan, G. (2011). Methylation and Synaptic Function in Adult Forebrain Neurons. *Nature Neuroscience*, *13*(4), 423–430. <https://doi.org/10.1038/nn.2514.Dnmt1>

- Furrer, A., & Van Loon, B. (2014). Handling the 3-methylcytosine lesion by six human DNA polymerases members of the B-, X- and Y-families. *Nucleic Acids Research*, *42*(1), 553–566. <https://doi.org/10.1093/nar/gkt889>
- Gabel, H. W., Kinde, B., Stroud, H., Gilbert, C. S., Harmin, D. A., Kastan, N. R., Hemberg, M., Ebert, D. H., & Greenberg, M. E. (2015). Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature*, *522*(7554), 89–93. <https://doi.org/10.1038/nature14319>
- Gao, L., Emperle, M., Guo, Y., Grimm, S. A., Ren, W., Adam, S., Uryu, H., Zhang, Z.-M., Chen, D., Yin, J., Dukatz, M., Anteneh, H., Jurkowska, R. Z., Lu, J., Wang, Y., Bashtrykov, P., Wade, P. A., Wang, G. G., Jeltsch, A., & Song, J. (2020). Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. *Nature Communications*, *11*(1). <https://doi.org/10.1038/s41467-020-17109-4>
- Gardiner-Garden, M., & Frommer, M. (1987). CpG Islands in vertebrate genomes. *Journal of Molecular Biology*, *196*(2), 261–282. [https://doi.org/10.1016/0022-2836\(87\)90689-9](https://doi.org/10.1016/0022-2836(87)90689-9)
- Ge, Y. Z., Pu, M. T., Gowher, H., Wu, H. P., Ding, J. P., Jeltsch, A., & Xu, G. L. (2004). Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *Journal of Biological Chemistry*, *279*(24), 25447–25454. <https://doi.org/10.1074/jbc.M312296200>
- Gibbons, R. J., Bachoo, S., Picketts, D. J., Aftimos, S., Asenbauer, B., Bergoffen, J., Berry, S. A., Dahl, N., Fryer, A., Keppler, K., Kurosawa, K., Levin, M. L., Masuno, M., Neri, G., Pierpont, M. E., Slaney, S. F., & Higgs, D. R. (1997). Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain. *Nature Genetics*, *17*(october), 146–148.
- Goel, R., Harsha, H. C., Pandey, A., & Prasad, T. S. K. (2012). Chapter 6 Human Protein Reference Database and Human Proteinpedia as Resources for Phosphoproteome Analysis. *Molecular BioSystems*, *8*(2), 453–463. <https://doi.org/10.1007/978-1-60761-232-2>
- 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-CpA sites. *Journal of Molecular Biology*, *309*(5), 1201–1208. <https://doi.org/10.1006/jmbi.2001.4710>
- Gowher, H., & Jeltsch, A. (2018). Mammalian DNA methyltransferases: New discoveries and open questions. *Biochemical Society Transactions*, *46*(5), 1191–1202. <https://doi.org/10.1042/BST20170574>
- Gowher, H., Loutchanwoot, P., Vorobjeva, O., Handa, V., Jurkowska, R. Z., Jurkowski, T. P., & Jeltsch, A. (2006). Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. *Journal of Molecular Biology*, *357*(3), 928–941. <https://doi.org/10.1016/j.jmb.2006.01.035>
- Greenberg, M. V. C., & Bourc’his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, *20*(10), 590–607. <https://doi.org/10.1038/s41580-019-0159-6>
- Greer, E. L., & Shi, Y. (2012). Histone methylation: A dynamic mark in health, disease and inheritance. *Nature Reviews Genetics*, *13*(5), 343–357. <https://doi.org/10.1038/nrg3173>
- Grigoryev, S. A., & Woodcock, C. L. (2012). Chromatin organization - The 30nm fiber. *Experimental Cell Research*, *318*(12), 1448–1455. <https://doi.org/10.1016/j.yexcr.2012.02.014>
- Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., Jin, X., Shi, X., Liu, P., Wang, X., Wang, W., Wei, Y., Li, X., Guo, F., Wu, X., ... Qiao, J. (2014). The DNA methylation

- landscape of human early embryos. *Nature*, 511(7511), 606–610. <https://doi.org/10.1038/nature13544>
- Guo, Su, Y., Shin, J. H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., Zhu, H., Chang, Q., Gao, Y., Ming, G. L., & Song, H. (2014). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature Neuroscience*, 17(2), 215–222. <https://doi.org/10.1038/nn.3607>
- Guo, W., Chung, W. Y., Qian, M., Pellegrini, M., & Zhang, M. Q. (2014). Characterizing the strand-specific distribution of non-CpG methylation in human pluripotent cells. *Nucleic Acids Research*, 42(5), 3009–3016. <https://doi.org/10.1093/nar/gkt1306>
- Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., Tian, C., Wang, J., Cong, Y., & Xu, Y. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature*, 517(7536), 640–644. <https://doi.org/10.1038/nature13899>
- Guy, J., Cheval, H., Selfridge, J., & Bird, A. (2011). The role of MeCP2 in the brain. *Annual Review of Cell and Developmental Biology*, 27, 631–652. <https://doi.org/10.1146/annurev-cellbio-092910-154121>
- Hamidi, T., Singh, A. K., & Chen, T. (2015). Genetic alterations of DNA methylation machinery in human diseases. *Epigenomics*, 7(2), 247–265. <https://doi.org/10.2217/epi.14.80>
- Handa, V., & Jeltsch, A. (2005). Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. *Journal of Molecular Biology*, 348(5), 1103–1112. <https://doi.org/10.1016/j.jmb.2005.02.044>
- Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K., Weemaes, C. M. R., & Gartler, S. M. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, 96(25), 14412–14417. <https://doi.org/10.1073/pnas.96.25.14412>
- Hashimoto, H., Liu, Y., Upadhyay, A. K., Chang, Y., Howerton, S. B., Vertino, P. M., Zhang, X., & Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Research*, 40(11), 4841–4849. <https://doi.org/10.1093/nar/gks155>
- Hata, K., Okano, M., Lei, H., & Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development*, 129(8), 1983–1993.
- He, Y., & Ecker, J. R. (2015). Non-CG Methylation in the Human Genome. *Annual Review of Genomics and Human Genetics*, 16, 55–77. <https://doi.org/10.1146/annurev-genom-090413-025437>
- Hellman, A., & Chess, A. (2007). Gene body-specific methylation on the active X chromosome. *Science*, 315(5815), 1141–1143. <https://doi.org/10.1126/science.1136352>
- Hermann, A., Goyal, R., & Jeltsch, A. (2004). DNA Processively with High Preference for Hemimethylated. *Journal of Biological Chemistry*, 279(46), 48350–48359. <https://doi.org/10.1074/jbc.M403427200>
- Heyn, P., Logan, C. V., Fluteau, A., Challis, R. C., Auchynnikava, T., Martin, C. A., Marsh, J. A., Taglini, F., Kilanowski, F., Parry, D. A., Cormier-Daire, V., Fong, C. T., Gibson, K., Hwa, V., Ibáñez, L., Robertson, S. P., Sebastiani, G., Rappsilber, J., Allshire, R. C., ... Jackson, A. P. (2019). Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. *Nature Genetics*, 51(1), 96–105. <https://doi.org/10.1038/s41588-018-0274-x>

- Heyward, F. D., & Sweatt, J. D. (2015). DNA Methylation in Memory Formation: Emerging Insights. *Neuroscientist*, *21*(5), 475–489. <https://doi.org/10.1177/1073858415579635>
- Hodges, E., Smith, A. D., Kendall, J., Xuan, Z., Ravi, K., Rooks, M., Zhang, M. Q., Ye, K., Bhattacharjee, A., Brizuela, L., McCombie, W. R., Wigler, M., Hannon, G. J., & Hicks, J. B. (2009). High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Research*, *19*(9), 1593–1605. <https://doi.org/10.1101/gr.095190.109>
- Holliday, R., & Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. *Science*, *187*(4173), 226–232. <https://doi.org/10.1126/science.1111098>
- Hotchkiss, R. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J Biol Chem.*, 315–332.
- Hu, J. L., Zhou, B. O., Zhang, R. R., Zhang, K. L., Zhou, J. Q., & Xu, G. L. (2009). The N-terminus of histone H3 is required for de novo DNA methylation in chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(52), 22187–22192. <https://doi.org/10.1073/pnas.0905767106>
- Hung, Y. L., Lee, H. J., Jiang, I., Lin, S. C., Lo, W. C., Lin, Y. J., & Sue, S. C. (2015). The First Residue of the PWWP Motif Modulates HATH Domain Binding, Stability, and Protein-Protein Interaction. *Biochemistry*, *54*(26), 4063–4074. <https://doi.org/10.1021/acs.biochem.5b00454>
- Inano, K., Suetake, I., Ueda, T., Miyake, Y., Nakamura, M., Okada, M., & Tajima, S. (2000). Maintenance-type DNA methyltransferase is highly expressed in post-mitotic neurons and localized in the cytoplasmic compartment. *Journal of Biochemistry*, *128*(2), 315–321. <https://doi.org/10.1093/oxfordjournals.jbchem.a022755>
- Jain, D., Meydan, C., Lange, J., Claeys Bouuaert, C., Mason, C. E., Anderson, K. V., & Keeney, S. (2017). rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. *PLoS Genetics*, *13*(8). <https://doi.org/10.1371/journal.pgen.1006964>
- Jang, H. S., Shin, W. J., Lee, J. E., & Do, J. T. (2017). CpG and non-CpG methylation in epigenetic gene regulation and brain function. *Genes*, *8*(6), 2–20. <https://doi.org/10.3390/genes8060148>
- Jansen, A., & Verstrepen, K. J. (2011). Nucleosome Positioning in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews*, *75*(2), 301–320. <https://doi.org/10.1128/mnbr.00046-10>
- Jeltsch, A. (2006). Molecular Enzymology of Mammalian DNA Methyltransferases. *DNA Methylation: Basic Mechanisms*, *301*, 203–225. https://doi.org/10.1007/978-3-319-43624-1_5
- Jeltsch, A., Broche, J., & Bashtrykov, P. (2018). Molecular processes connecting DNA methylation patterns with DNA methyltransferases and histone modifications in mammalian genomes. *Genes*, *9*(11), 1–20. <https://doi.org/10.3390/genes9110566>
- Jeltsch, A., & Jurkowska, R. Z. (2014). New concepts in DNA methylation. *Trends in Biochemical Sciences*, *39*(7), 310–318. <https://doi.org/10.1016/j.tibs.2014.05.002>
- Jeltsch, A., & Jurkowska, R. Z. (2016). Allosteric control of mammalian DNA methyltransferases - A new regulatory paradigm. *Nucleic Acids Research*, *44*(18), 8556–8575. <https://doi.org/10.1093/nar/gkw723>
- Jeltsch, A., & Lanio, T. (2002). Site-directed mutagenesis by polymerase chain reaction. *Methods in Molecular Biology (Clifton, N.J.)*. <https://doi.org/10.1385/1-59259-194-9:085>
- Jones, P. A., & Liang, G. (2009). Rethinking how DNA methylation patterns are maintained.

- Nature Reviews Genetics*, 10(11), 805–811. <https://doi.org/10.1038/nrg2651>
- Jones, P. L., Veenstra, G. J. C., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., & Wolffe, A. P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genetics*, 19(2), 187–191. <https://doi.org/10.1038/561>
- Jurkowska, Jurkowski, T. P., & Jeltsch, A. (2011). Structure and Function of Mammalian DNA Methyltransferases. *ChemBioChem*, 12(2), 206–222. <https://doi.org/10.1002/cbic.201000195>
- Jurkowska, Rajavelu, A., Anspach, N., Urbanke, C., Jankevicius, G., Ragozin, S., Nellen, W., & Jeltsch, A. (2011). Oligomerization and binding of the Dnmt3a DNA methyltransferase to parallel DNA molecules: Heterochromatic localization and role of Dnmt3L. *Journal of Biological Chemistry*, 286(27), 24200–24207. <https://doi.org/10.1074/jbc.M111.254987>
- Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, H., Li, E., & Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*, 429(6994), 900–903. <https://doi.org/10.1038/nature02633>
- Kasinathan, S., Orsi, G. A., Zentner, G. E., Ahmad, K., Henikoff, S., Division, S., Hutchinson, F., Program, S. T., & Recherche, C. De. (2014). High-resolution mapping of transcription factor binding sites on native chromatin. *Nature Methods*, 11(2), 203–209. <https://doi.org/10.1038/nmeth.2766>. High-resolution
- Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., & Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Human Molecular Genetics*, 16(19), 2272–2280. <https://doi.org/10.1093/hmg/ddm179>
- Kimura, H., & Shiota, K. (2003). Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *Journal of Biological Chemistry*, 278(7), 4806–4812. <https://doi.org/10.1074/jbc.M209923200>
- Kinde, B., Gabel, H. W., Gilbert, C. S., Griffith, E. C., & Greenberg, M. E. (2015). Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proceedings of the National Academy of Sciences of the United States of America*, 112(22), 6800–6806. <https://doi.org/10.1073/pnas.1411269112>
- Kungulovski, G., Kycia, I., Tamas, R., Jurkowska, R. Z., Kudithipudi, S., Henry, C., Reinhardt, R., Labhart, P., & Jeltsch, A. (2014). Application of histone modification-specific interaction domains as an alternative to antibodies. *Genome Research*, 24(11), 1842–1853. <https://doi.org/10.1101/gr.170985.113>
- Laguri, C., Duband-Goulet, I., Friedrich, N., Axt, M., Belin, P., Callebaut, I., Gilquin, B., Zinn-Justin, S., & Couprie, J. (2008). Human mismatch repair protein MSH6 contains a PWWP domain that targets double stranded DNA. *Biochemistry*, 47(23), 6199–6207. <https://doi.org/10.1021/bi7024639>
- Laisné, M., Gupta, N., Kirsh, O., Pradhan, S., & Defossez, P. A. (2018). Mechanisms of DNA methyltransferase recruitment in mammals. *Genes*, 9(12). <https://doi.org/10.3390/genes9120617>
- Laurent, L., Wong, E., Li, G., Huynh, T., Tsigos, A., Ong, C. T., Low, H. M., Sung, K. W. K., Rigoutsos, I., Loring, J., & Wei, C. L. (2010). Dynamic changes in the human methylome during differentiation. *Genome Research*, 20(3), 320–331. <https://doi.org/10.1101/gr.101907.109>
- Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, 11(3), 204–220.

<https://doi.org/10.1038/nrg2719>

- Lee, J. H., Park, S. J., & Nakai, K. (2017). Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s. *Scientific Reports*, *7*(1), 1–11. <https://doi.org/10.1038/s41598-017-11800-1>
- Lehnertz, B., Ueda, Y., Derijck, A. A. H. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., & Peters, A. H. F. M. (2003). Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Current Biology*, *13*(14), 1192–1200. <https://doi.org/10.1016/S>
- Li, H., Rauch, T., Chen, Z. X., Szabó, P. E., Riggs, A. D., & Pfeifer, G. P. (2006). The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. *Journal of Biological Chemistry*, *281*(28), 19489–19500. <https://doi.org/10.1074/jbc.M513249200>
- Li, J., Ohliger, J., & Pei, M. (2014). Significance of epigenetic landscape in cartilage regeneration from the cartilage development and pathology perspective. *Stem Cells and Development*, *23*(11), 1178–1194. <https://doi.org/10.1089/scd.2014.0002>
- Liang, G., Lin, J. C. Y., Wei, V., Yoo, C., Cheng, J. C., Nguyen, C. T., Weisenberger, D. J., Egger, G., Takai, D., Gonzales, F. A., & Jones, P. A. (2004). Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(19), 7357–7362. <https://doi.org/10.1073/pnas.0401866101>
- Liebert, K., & Jeltsch, A. (2008). Methyltransferases Using a Biotin / Avidin. *Methods in Molecular Biology*, *418*, 149–155.
- Lin, C. C., Chen, Y. P., Yang, W. Z., Shen, J. C. K., & Yuan, H. S. (2020). Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. *Nucleic Acids Research*, *48*(7), 3949–3961. <https://doi.org/10.1093/nar/gkaa111>
- Lin, I. G., Han, L., Taghva, A., O'Brien, L. E., & Hsieh, C.-L. (2002). Murine De Novo Methyltransferase Dnmt3a Demonstrates Strand Asymmetry and Site Preference in the Methylation of DNA In Vitro. *Molecular and Cellular Biology*, *22*(3), 704–723. <https://doi.org/10.1128/mcb.22.3.704-723.2002>
- Lister, R., & Ecker, J. R. (2009). Finding the fifth base: Genome-wide sequencing of cytosine methylation. *Genome Research*, *19*(6), 959–966. <https://doi.org/10.1101/gr.083451.108>
- Lister, R., Mukamel, E. A., Nery, J. R., Urich, M., Puddifoot, C. A., Johnson, N. D., Lucero, J., Huang, Y., Dwork, A. J., Schultz, M. D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu, J. C., Rao, A., Esteller, M., He, C., Haghghi, F. G., ... Ecker, J. R. (2013). Global epigenomic reconfiguration during mammalian brain development. *Science*, *341*(6146). <https://doi.org/10.1126/science.1237905>
- Lister, R., Pelizzola, M., Downen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–322. <https://doi.org/10.1038/nature08514>
- Lister, R., Pelizzola, M., Kida, Y. S., Hawkins, R. D., Nery, J. R., Hon, G., Antosiewicz-Bourget, J., O'Gmalley, R., Castanon, R., Klugman, S., Downes, M., Yu, R., Stewart, R., Ren, B., Thomson, J. A., Evans, R. M., & Ecker, J. R. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*, *471*(7336), 68–73. <https://doi.org/10.1038/nature09798>
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal

- structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), 251–260. <https://doi.org/10.1038/38444>
- Lukashevich, O. V., Cherepanova, N. A., Jurkowska, R. Z., Jeltsch, A., & Gromova, E. S. (2016). Conserved motif VIII of murine DNA methyltransferase Dnmt3a is essential for methylation activity. *BMC Biochemistry*, 17(1), 1–10. <https://doi.org/10.1186/s12858-016-0064-y>
- Lukasik, S. M. (2006). High resolution structure of the HDGF PWWP domain: A potential DNA binding domain. *Protein Science*, 15(2), 314–323. <https://doi.org/10.1110/ps.051751706>
- Luo, G. Z., Blanco, M. A., Greer, E. L., He, C., & Shi, Y. (2015). DNA N6-methyladenine: A new epigenetic mark in eukaryotes? *Nature Reviews Molecular Cell Biology*, 16(12), 705–710. <https://doi.org/10.1038/nrm4076>
- Lyko, F. (2018). The DNA methyltransferase family: A versatile toolkit for epigenetic regulation. *Nature Reviews Genetics*, 19(2), 81–92. <https://doi.org/10.1038/nrg.2017.80>
- Ma, H., Morey, R., O’Neil, R. C., He, Y., Daughtry, B., Schultz, M. D., Hariharan, M., Nery, J. R., Castanon, R., Sabatini, K., Thiagarajan, R. D., Tachibana, M., Kang, E., Tippner-Hedges, R., Ahmed, R., Gutierrez, N. M., Van Dyken, C., Polat, A., Sugawara, A., ... Mitalipov, S. (2014). Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature*, 511(7508), 177–183. <https://doi.org/10.1038/nature13551>
- Machado, A. C. D., Zhou, T., Rao, S., Goel, P., Rastogi, C., Lazarovici, A., Bussemaker, H. J., & Rohs, R. (2015). Evolving insights on how cytosine methylation affects protein-DNA binding. *Briefings in Functional Genomics*, 14(1), 61–73. <https://doi.org/10.1093/bfpg/elu040>
- Mallona, I., Ilie, I. M., Karemaker, I. D., Butz, S., Manzo, M., Caflisch, A., & Baubec, T. (2021). Flanking sequence preference modulates de novo DNA methylation in the mouse genome. *Nucleic Acids Research*, 49(1), 145–157. <https://doi.org/10.1093/nar/gkaa1168>
- Martínez de Paz, A., & Ausió, J. (2017). MeCP2, a modulator of neuronal chromatin organization involved in rett syndrome. *Advances in Experimental Medicine and Biology*, 978, 3–21. https://doi.org/10.1007/978-3-319-53889-1_1
- Mausser, R., Kungulovski, G., Meral, D., Maisch, D., & Jeltsch, A. (2018). Application of mixed peptide arrays to study combinatorial readout of chromatin modifications. *Biochimie*, 146, 14–19. <https://doi.org/10.1016/j.biochi.2017.11.008>
- Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., & Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205), 766–770. <https://doi.org/10.1038/nature07107>
- Messerschmidt, D. M., Knowles, B. B., & Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes and Development*, 28(8), 812–828. <https://doi.org/10.1101/gad.234294.113>
- Métivier, R., Gallais, R., Tiffoche, C., Le Péron, C., Jurkowska, R. Z., Carmouche, R. P., Ibberson, D., Barath, P., Demay, F., Reid, G., Benes, V., Jeltsch, A., Gannon, F., & Salbert, G. (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature*, 452(7183), 45–50. <https://doi.org/10.1038/nature06544>
- Metzker, M. L. (2010). Sequencing technologies the next generation. *Nature Reviews Genetics*, 11(1), 31–46. <https://doi.org/10.1038/nrg2626>
- Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., Lee, W., Mendenhall, E., O’Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., ... Bernstein, B. E. (2007).

- Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448(7153), 553–560. <https://doi.org/10.1038/nature06008>
- Monti, M. (2012). Epigenetics protocols. *European Journal of Histochemistry*. <https://doi.org/10.4081/ejh.2012.br8>
- Moris, N., Pina, C., & Arias, A. M. (2016). Transition states and cell fate decisions in epigenetic landscapes. *Nature Reviews Genetics*, 17(11), 693–703. <https://doi.org/10.1038/nrg.2016.98>
- Morselli, M., Pastor, W. A., Montanini, B., Nee, K., Ferrari, R., Fu, K., Bonora, G., Rubbi, L., Clark, A. T., Ottonello, S., Jacobsen, S. E., & Pellegrini, M. (2015). In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *ELife*, 2015(4), 1–21. <https://doi.org/10.7554/eLife.06205>
- Musselman, C. A., Lalonde, M.-E., Côte, J., & Kutateladze, T. G. (2012). Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol.*, 19(12)(1), 1218–1227. <https://doi.org/10.1038/nsmb.2436>. Perceiving
- Nan, X., Tate, P., Li, E., & Bird, A. (1996). DNA methylation specifies chromosomal localization of MeCP2. *Molecular and Cellular Biology*, 16(1), 414–421. <https://doi.org/10.1128/mcb.16.1.414>
- Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3), 247–257. [https://doi.org/10.1016/S0092-8674\(00\)81656-6](https://doi.org/10.1016/S0092-8674(00)81656-6)
- Okano, M., Xie, S., & Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases Non-invasive sexing of preimplantation stage mammalian embryos. *Nature America Inc.*, 19(july), 219–220.
- Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S. P., Allis, C. D., Cheng, X., & Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, 448(7154), 714–717. <https://doi.org/10.1038/nature05987>
- Oswald, T., Avery, M. D., Colin, M., MacLeod, M. D., & McCarty, M. M. D. (1944). Studies of the chemical nature of the substance inducing transformation of pneumococcal types. *Journal of Experimental Medicine*, 79(2)(6), 137–158.
- Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., & Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Reports*, 10(11), 1235–1241. <https://doi.org/10.1038/embor.2009.218>
- Pataillot-Meakin, T., Pillay, N., & Beck, S. (2016). 3-methylcytosine in cancer: An underappreciated methyl lesion? *Epigenomics*, 8(4), 451–454. <https://doi.org/10.2217/epi.15.121>
- Patel, D. J. (2016). A structural perspective on readout of epigenetic histone and DNA methylation marks. *Cold Spring Harbor Perspectives in Biology*, 8(3), 1–47. <https://doi.org/10.1101/cshperspect.a018754>
- Pepenella, S., Murphy, K., & Hayes, J. (2014). Intra- and Inter-nucleosome interactions of the core histone tail domains in higher-order chromatin structure. *Chromosoma*, 123(1), 3–13. <https://doi.org/10.1007/s00412-013-0435-8>. Intra-
- Pinney, S. E. (2014). Mammalian Non-CpG methylation: Stem cells and beyond. *Biology*, 3(4), 739–751. <https://doi.org/10.3390/biology3040739>
- Piovesan, A., Pelleri, M. C., Antonaros, F., Strippoli, P., Caracausi, M., & Vitale, L. (2019). On the length, weight and GC content of the human genome. *BMC Research Notes*, 12(1), 1–7.

<https://doi.org/10.1186/s13104-019-4137-z>

- Purdy, M. M., Holz-Schietinger, C., & Reich, N. O. (2010). Identification of a second DNA binding site in human DNA methyltransferase 3A by substrate inhibition and domain deletion. *Archives of Biochemistry and Biophysics*, 498(1), 13–22. <https://doi.org/10.1016/j.abb.2010.03.007>
- Qiu, C., Sawada, K., Zhang, X., & Cheng, X. (2002). The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nature Structural Biology*, 176(10), 139–148. <https://doi.org/10.1038/nsb759>
- Rajavelu, A., Lungu, C., Emperle, M., Dukatz, M., Bröhm, A., Broche, J., Hanelt, I., Parsa, E., Schiffers, S., Karnik, R., Meissner, A., Carell, T., Rathert, P., Jurkowska, R. Z., & Jeltsch, A. (2018). Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2. *Nucleic Acids Research*, 46(17). <https://doi.org/10.1093/nar/gky715>
- Rajavelu, Arumugam, Jurkowska, R. Z., Fritz, J., & Jeltsch, A. (2012). Function and disruption of DNA Methyltransferase 3a cooperative DNA binding and nucleoprotein filament formation. *Nucleic Acids Research*, 40(2), 569–580. <https://doi.org/10.1093/nar/gkr753>
- Ramsahoye, B. H., Biniszkiwicz, D., Lyko, F., Clark, V., Bird, A. P., & Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America*, 97(10), 5237–5242. <https://doi.org/10.1073/pnas.97.10.5237>
- Reddington, J. P., Pennings, S., & Meehan, R. R. (2013). Non-canonical functions of the DNA methylome in gene regulation. *Biochemical Journal*, 451(1), 13–23. <https://doi.org/10.1042/BJ20121585>
- Reither, S., Li, F., Gowher, H., & Jeltsch, A. (2003). Catalytic mechanism of DNA-(cytosine-C5)-methyltransferases revisited: Covalent intermediate formation is not essential for methyl group transfer by the murine Dnmt3a enzyme. *Journal of Molecular Biology*, 329(4), 675–684. [https://doi.org/10.1016/S0022-2836\(03\)00509-6](https://doi.org/10.1016/S0022-2836(03)00509-6)
- Remacha, L., Currás-Freixes, M., Torres-Ruiz, R., Schiavi, F., Torres-Pérez, R., Calsina, B., Letón, R., Comino-Méndez, I., Roldán-Romero, J. M., Montero-Conde, C., Santos, M., Pérez, L. I., Pita, G., Alonso, M. R., Honrado, E., Pedrinaci, S., Crespo-Facorro, B., Percesepe, A., Falcioni, M., ... Cascón, A. (2018). Gain-of-function mutations in DNMT3A in patients with paraganglioma. *Genetics in Medicine*, 20(12), 1644–1651. <https://doi.org/10.1038/s41436-018-0003-y>
- Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.*, 14(1), 9–25.
- Riggs, Arthur D., & Xiong, Z. (2004). Methylation and epigenetic fidelity. *Proceedings of the National Academy of Sciences of the United States of America*, 101(1), 4–5. <https://doi.org/10.1073/pnas.0307781100>
- Ringvoll, J., Nordstrand, L. M., Vågbø, C. B., Talstad, V., Reite, K., Aas, P. A., Lauritzen, K. H., Liabakk, N. B., Bjørk, A., Doughty, R. W., Falnes, P., Krokan, H. E., & Klungland, A. (2006). Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. *EMBO Journal*, 25(10), 2189–2198. <https://doi.org/10.1038/sj.emboj.7601109>
- Robinson, P. J., & Rhodes, D. (2006). Structure of the “30 nm” chromatin fibre: A key role for the linker histone. In *Current Opinion in Structural Biology* (Vol. 16, Issue 3, pp. 336–343). <https://doi.org/10.1016/j.sbi.2006.05.007>
- Rona, G. B., Eleutherio, E. C. A., & Pinheiro, A. S. (2016). PWWP domains and their modes of sensing DNA and histone methylated lysines. *Biophysical Reviews*, 8(1), 63–74.

<https://doi.org/10.1007/s12551-015-0190-6>

- Rondelet, G., Dal Maso, T., Willems, L., & Wouters, J. (2016). Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. *Journal of Structural Biology*, *194*(3), 357–367. <https://doi.org/10.1016/j.jsb.2016.03.013>
- Rošić, S., Amouroux, R., Requena, C. E., Gomes, A., Emperle, M., Beltran, T., Rane, J. K., Linnett, S., Selkirk, M. E., Schiffer, P. H., Bancroft, A. J., Grecis, R. K., Jeltsch, A., Hajkova, P., & Sarkies, P. (2018). Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. *Nature Genetics*, *50*(3), 452–459. <https://doi.org/10.1038/s41588-018-0061-8>
- Roth, M., & Jeltsch, A. (2000). Biotin-avidin microplate assay for the quantitative analysis of enzymatic methylation of DNA by DNA methyltransferases. *Biological Chemistry*, *381*(3), 269–272. <https://doi.org/10.1515/BC.2000.035>
- Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(5), 1412–1417. <https://doi.org/10.1073/pnas.0510310103>
- Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C., & Kouzarides, T. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nature Cell Biology*, *6*(1), 73–77. <https://doi.org/10.1038/ncb1076>
- Schübeler, D. (2015). Function and information content of DNA methylation. *Nature*, *517*(7534), 321–326. <https://doi.org/10.1038/nature14192>
- Sharif, J., & Koseki, H. (2011). Recruitment of Dnmt1: Roles of the SRA protein Np95 (Uhrf1) and other factors. In *Progress in Molecular Biology and Translational Science* (1st ed., Vol. 101). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-387685-0.00008-1>
- Sharif, J., Muto, M., Takebayashi, S. I., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M., & Koseki, H. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, *450*(7171), 908–912. <https://doi.org/10.1038/nature06397>
- Shen, J. cheng, Rideout, W. M., & Jones, P. A. (1994). The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Research*, *22*(6), 972–976. <https://doi.org/10.1093/nar/22.6.972>
- Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R. A., & Issa, J. P. J. (2007). Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genetics*, *3*(10), 2023–2026. <https://doi.org/10.1371/journal.pgen.0030181>
- Shin, A., Ming, G. L., & Song, H. (2014). DNA modifications in the mammalian brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1652). <https://doi.org/10.1098/rstb.2013.0512>
- Shirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., Kono, T., & Sasaki, H. (2013). Mouse Oocyte Methylomes at Base Resolution Reveal Genome-Wide Accumulation of Non-CpG Methylation and Role of DNA Methyltransferases. *PLoS Genetics*, *9*(4). <https://doi.org/10.1371/journal.pgen.1003439>
- Song, J., Teplova, M., Ishibe-Murakami, S., & Patel, D. J. (2012). Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science*, *335*(6069), 709–712. <https://doi.org/10.1126/science.1214453>
- Stec, I., Nagl, S. B., Van Ommen, G. J. B., & Den Dunnen, J. T. (2000). The PWWP domain: A potential protein-protein interaction domain in nuclear proteins influencing

- differentiation? *FEBS Letters*, 473(1), 1–5. [https://doi.org/10.1016/S0014-5793\(00\)01449-6](https://doi.org/10.1016/S0014-5793(00)01449-6)
- Suetake, I., Mishima, Y., Kimura, H., Lee, Y. H., Goto, Y., Takeshima, H., Ikegami, T., & Tajima, S. (2011). Characterization of DNA-binding activity in the N-terminal domain of the DNA methyltransferase Dnmt3a. *Biochemical Journal*, 437(1), 141–148. <https://doi.org/10.1042/BJ20110241>
- Suetake, I., Miyazaki, J., Murakami, C., Takeshima, H., & Tajima, S. (2003). Distinct enzymatic properties of recombinant mouse DNA methyltransferases Dnmt3a and Dnmt3b. *Journal of Biochemistry*, 133(6), 737–744. <https://doi.org/10.1093/jb/mvg095>
- Sugino, K., Hempel, C. M., Okaty, B. W., Arnsen, H. A., Kato, S., Dani, V. S., & Nelson, S. B. (2014). Cell-type-specific repression by methyl-CpG-binding protein 2 is biased toward long genes. *Journal of Neuroscience*, 34(38), 12877–12883. <https://doi.org/10.1523/JNEUROSCI.2674-14.2014>
- Takai, D., & Jones, P. A. (2004). Origins of Bidirectional Promoters: Computational Analyses of Intergenic Distance in the Human Genome. *Molecular Biology and Evolution*, 21(3), 463–467. <https://doi.org/10.1093/molbev/msh040>
- Tan, H. K., Wu, C. S., Li, J., Tan, Z. H., Hoffman, J. R., Fry, C. J., Yang, H., Di Ruscio, A., & Tenen, D. G. (2019). DNMT3B shapes the mCA landscape and regulates mCG for promoter bivalency in human embryonic stem cells. *Nucleic Acids Research*, 47(14), 7460–7475. <https://doi.org/10.1093/nar/gkz520>
- Thompson, J. J., Kaur, R., Sosa, C. P., Lee, J. H., Kashiwagi, K., Zhou, D., & Robertson, K. D. (2018). ZBTB24 is a transcriptional regulator that coordinates with DNMT3B to control DNA methylation. *Nucleic Acids Research*, 46(19), 10034–10051. <https://doi.org/10.1093/nar/gky682>
- Tippin, D. B., & Sundaralingam, M. (1997). Nine polymorphic crystal structures of d(CCGGGCCCGG), d(CCGGGCCm5CGG), d(Cm5CGGGCCm5CGG) and d(CCGGGCC(Br)5CGG) in three different conformations: Effects of spermine binding and methylation on the bending and condensation of A-DNA. *Journal of Molecular Biology*, 267(5), 1171–1185. <https://doi.org/10.1006/jmbi.1997.0945>
- Tudor, M., Akbarian, S., Chen, R. Z., & Jaenisch, R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24), 15536–15541. <https://doi.org/10.1073/pnas.242566899>
- Van Nuland, R., Van Schaik, F. M. A., Simonis, M., Van Heesch, S., Cuppen, E., Boelens, R., Timmers, H. T. M., & Van Ingen, H. (2013). Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics and Chromatin*, 6(1), 1–12. <https://doi.org/10.1186/1756-8935-6-12>
- Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature Publishing Group*.
- Wang, H., Farnung, L., Dienemann, C., & Cramer, P. (2020). Structure of H3K36-methylated nucleosome-PWWP complex reveals multivalent cross-gyre binding. *Nature Structural and Molecular Biology*, 27(1), 8–13. <https://doi.org/10.1038/s41594-019-0345-4>.Structure
- Watanabe, D., Suetake, I., Tada, T., & Tajima, S. (2002). Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mechanisms of Development*, 118(1–2), 187–190. [https://doi.org/10.1016/S0925-4773\(02\)00242-3](https://doi.org/10.1016/S0925-4773(02)00242-3)
- Weaver, I. C. G. (2014). Integrating early life experience, gene expression, brain development,

- and emergent phenotypes: Unraveling the thread of nature via nurture. In *Advances in Genetics* (Vol. 86). Elsevier. <https://doi.org/10.1016/B978-0-12-800222-3.00011-5>
- Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., & Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics*, *39*(4), 457–466. <https://doi.org/10.1038/ng1990>
- Wienholz, B. L., Karetka, M. S., Moarefi, A. H., Gordon, C. A., Ginno, P. A., & Chédin, F. (2010). DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo. *PLoS Genetics*, *6*(9). <https://doi.org/10.1371/journal.pgen.1001106>
- Wu, C. T., & Morris, J. R. (2001). Genes, genetics, and epigenetics: A correspondence. *Science*, *293*(5532), 1103–1105. <https://doi.org/10.1126/science.293.5532.1103>
- Wu, H., Zeng, H., Lam, R., Tempel, W., Amaya, M. F., Xu, C., Dombrovski, L., Qiu, W., Wang, Y., & Min, J. (2011). Structural and histone binding ability characterizations of human PWWP domains. *PLoS ONE*, *6*(6). <https://doi.org/10.1371/journal.pone.0018919>
- Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W. W., Okumura, K., & Li, E. (1999). Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene*, *236*(1), 87–95. [https://doi.org/10.1016/S0378-1119\(99\)00252-8](https://doi.org/10.1016/S0378-1119(99)00252-8)
- Xu, G. L., Bestor, T. H., Bourc'His, D., Hsieh, C. L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J. J., & Viegas-Péquignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature*, *402*(6758), 187–191. <https://doi.org/10.1038/46052>
- Xu, Y.-M., Yue, Y., Andy, T., & Lau, Y. (2017). Histone H3. *Encyclopedia of Signaling Molecules, Kouzarides*. <https://doi.org/10.1007/978-1-4614-6438-9>
- Zeng, Y., & Chen, T. (2019). DNA methylation reprogramming during mammalian development. *Genes*, *10*(4). <https://doi.org/10.3390/genes10040257>
- Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., & Jeltsch, A. (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Research*, *38*(13), 4246–4253. <https://doi.org/10.1093/nar/gkq147>
- Zhang, Y., Rohde, C., Tieriing, S., Jurkowski, T. P., Bock, C., Santacruz, D., Ragozin, S., Reinhardt, R., Groth, M., Waiter, J., & Jeltsch, A. (2009). DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. *PLoS Genetics*, *5*(3). <https://doi.org/10.1371/journal.pgen.1000438>
- Zhang, Z. M., Lu, R., Wang, P., Yu, Y., Chen, D., Gao, L., Liu, S., Ji, D., Rothbart, S. B., Wang, Y., Wang, G. G., & Song, J. (2018). Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature*, *554*(7692), 387–391. <https://doi.org/10.1038/nature25477>
- Zhu, P., & Li, G. (2016). Structural insights of nucleosome and the 30-nm chromatin fiber. *Current Opinion in Structural Biology*, *36*, 106–115. <https://doi.org/10.1016/j.sbi.2016.01.013>
- Ziller, M. J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C. B., Bernstein, B. E., Lengauer, T., Gnirke, A., & Meissner, A. (2011). Genomic distribution and Inter-Sample variation of Non-CpG methylation across human cell types. *PLoS Genetics*, *7*(12). <https://doi.org/10.1371/journal.pgen.1002389>

7 Author contributions

Dukat, M., Holzer, K., Choudalakis, M., Emperle, M., Lungu, C., Bashtrykov, P. & Jeltsch, A. (2019). **“H3K36me2/3 Binding and DNA Binding of the DNA Methyltransferase DNMT3A PWWP Domain Both Contribute to Its Chromatin Interaction.”** *Journal of Molecular Biology* 431(24)

I conducted the biochemical studies with contributions of KH except for the equilibrium peptide binding experiments which were performed by MC.

Together with AJ, I wrote the manuscript draft and prepared the figures. I contributed to data interpretation and discussion and writing of the final manuscript.

Rajavelu, A.* , Lungu, C.* , Emperle, M., **Dukat, M.,** Bröhm, A., Broche, J., Hanelt, I., Parsa, E., Schiffers, S., Karnik, R., Meissner, A., Carell, T., Rathert, P., Jurkowska, R. Z. & Jeltsch, A. (2018). **“Chromatin-Dependent Allosteric Regulation of DNMT3A Activity by MeCP2.”** *Nucleic Acids Research* 46(17).

* These authors contributed equally to the work.

I have confirmed the activity of *Serratia marcescens* nuclease in the applied interaction buffer and subsequently demonstrated the independence of the TRD-ADD binding in the pull-down experiment from DNA by addition of this DNase. Additionally, I performed the methylation assays with the unmodified H3-peptide for the stimulation of DNMT3A2 in presence and absence of the MeCP2-TRD domain. I contributed to data interpretation and discussion and writing of the final manuscript.

Dukat, M.* , Requena, C. E.* , Emperle, M., Hajkova, P., Sarkies, P. & Jeltsch, A. (2019). **“Mechanistic Insights into Cytosine-N3 Methylation by DNA Methyltransferase DNMT3A.”** *Journal of Molecular Biology* 431(17).

* These authors contributed equally to the work.

I conducted the protein purification, biochemical assays, and protein modelling with contributions of M.E. I contributed to data interpretation and discussion and writing of the final manuscript.

Gao, L.* , Emperle, M.* , Guo, Y.* , Grimm, S. A., Ren, W., Adam, S., Uryu, H., Zhang, Z. M., Chen, D., Yin D., **Dukat, M.,** Anteneh, H., Jurkowska, R. Z., Lu, J., Wang, Y., Bashtrykov, P., Wade, P. A., Wang, G. G., Jeltsch, A. & Song, J. (2020). **“Comprehensive Structure-Function Characterization of DNMT3B and DNMT3A Reveals Distinctive de Novo DNA Methylation Mechanisms.”** *Nature Communications* 11(1).

* These authors contributed equally to the work.

I contributed to enzyme purification and enzyme kinetics and deep enzymology methylation reactions with DNMT3B. Furthermore, I performed and analyzed the deep enzymology experiments with DNMT3B K777A.

Dukat, M., Adam, S., Biswal, M., Song, J., Bashtrykov, P. & Jeltsch, A. (2020). **“Complex DNA Sequence Readout Mechanisms of the DNMT3B DNA Methyltransferase.”** *Nucleic Acids Research* (14):1–15.

I performed the experimental work with support from S.A. I was involved in NGS data analysis and bioinformatic analyses. Together with A.J. I prepared the manuscript draft. I contributed to data interpretation and discussion and writing of the final manuscript.

8 Appendix (not included in the published thesis)

Appendix 1

Dukatz, M., Holzer, K., Choudalakis, M., Emperle, M., Lungu, C., Bashtrykov, P. & Jeltsch, A. (2019). **“H3K36me2/3 Binding and DNA Binding of the DNA Methyltransferase DNMT3A PWWP Domain Both Contribute to Its Chromatin Interaction.”** *Journal of Molecular Biology* 431(24)

Appendix 2

Rajavelu, A.* , Lungu, C.* , Emperle, M., Dukatz, M., Bröhm, A., Broche, J., Hanelt, I., Parsa, E., Schiffers, S., Karnik, R., Meissner, A., Carell, T., Rathert, P., Jurkowska, R. Z. & Jeltsch, A. (2018). **“Chromatin-Dependent Allosteric Regulation of DNMT3A Activity by MeCP2.”** *Nucleic Acids Research* 46(17).

* These authors contributed equally to the work.

Appendix 3

Dukatz, M.* , Requena, C. E.* , Emperle, M., Hajkova, P., Sarkies, P. & Jeltsch, A. (2019). **“Mechanistic Insights into Cytosine-N3 Methylation by DNA Methyltransferase DNMT3A.”** *Journal of Molecular Biology* 431(17).

* These authors contributed equally to the work.

Appendix 4

Gao, L.* , Emperle, M.* , Guo, Y.* , Grimm, S. A., Ren, W., Adam, S., Uryu, H., Zhang, Z. M., Chen, D., Yin D., Dukatz, M., Anteneh, H., Jurkowska, R. Z., Lu, J., Wang, Y., Bashtrykov, P., Wade, P. A., Wang, G. G., Jeltsch, A. & Song, J. (2020). **“Comprehensive Structure-Function Characterization of DNMT3B and DNMT3A Reveals Distinctive de Novo DNA Methylation Mechanisms.”** *Nature Communications* 11(1).

* These authors contributed equally to the work.

Appendix 5

Dukatz, M., Adam, S., Biswal, M., Song, J., Bashtrykov, P. & Jeltsch, A. (2020). **“Complex DNA Sequence Readout Mechanisms of the DNMT3B DNA Methyltransferase.”** *Nucleic Acids Research* (14):1–15.