Enzymatic characterization of protein lysine methyltransferases

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

Vorgelegt von
Sara Weirich
aus Nürtingen, Deutschland

Hauptberichter:	Prof. Dr. Albert Jeltsch
Mitberichter:	Prof. Dr. Stephan Nußberger

Tag der mündlichen Prüfung: 12.05.2017

Institut für Biochemie der Universität Stuttgart
2017


Universität Stuttgart

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbst verfasst und dabei keine anderen als die angegeben Quellen und Hilfsmittel verwendet habe.

Stuttgart, den 12.05.2017

Sara Weirich

Contents

Acknowledgements

First I would like to thank my supervisor Prof. Dr. Albert Jeltsch for giving me the opportunity to work in the exciting field of Epigenetics. I want to thank him for his continuous support, guidance and great suggestions during the entire research process and during writing of this thesis.

I would like to thank Prof. Dr. Stephan Nußberger and Prof. Dr. Holger Jeske for being co-referees of my PhD thesis.

I am also thankful to Prof. Dr. Christina Wege, Prof. Dr. Dieter Wolf, Prof. Dr. Holger Jeske and Prof. Dr. Markus Morrison for accepting to read and review my PhD thesis.

Many thanks to Dr. Srikanth Kudithipudi for his supervision during my whole PhD time. His suggestions and discussions were invaluable to my work.

I would like to extend my gratitude to PD Dr. Hans Rudolph for his confidence in me and for the support he provided me to be a radiation protection officer.

Special thanks to the "PKMTs": Denis, Srikanth, Maren and Alex for all the help, great teamwork, discussions and nice atmosphere in the lab.

I would also like to thank Miru and Rebekka for being good friends and for all the fun we had together.

I am also thankful to my colleagues in the North lab: Rebekka, Denis, Julian, Alex, Johannes, Maren, Srikanth and Pavel for the friendly atmosphere and great time in the lab.

I am grateful to a number of students with whom I worked with, namely Konrad, Alex, Nico, Marcel and Unni.

I am very thankful to Elisabeth Tosta and Regina Philipp for all the help regarding the technical support, organization and formalities, which made my work possible.

Very special thanks goes out to Dragica for her positive energy, for creating a familiar atmosphere and incredible help.

I would like to thank all my colleagues and former lab members for their help, support, interest and valuable advice.

Last but not least, I am very grateful to my parents and my brother for all their support, positive thoughts and for always believing in me.

List of Publications

Publications resulting from work during the doctoral studies:

Kudithipudi S, Kusevic D, **Weirich S** & Jeltsch A. (2014). **Specificity analysis of protein lysine methyltransferases using SPOT peptide arrays.** *Journal of Visualized Experiments 2014 Nov 29 (93), e52203, doi:10.3791/52203.*

Schuhmacher MK, Kudithipudi S, Kusevic D, **Weirich S**, & Jeltsch A. (2015). **Activity and specificity of the human SUV39H2 protein lysine methyltransferase**. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2015 Jan;1849(1):55-63. doi: 10.1016/j.bbagrm.2014.11.005. Epub 2014 Nov 22.*

Weirich S, Kudithipudi S, Jeltsch A. (2016*).* **Specificity of the Suv4-20H1 and Suv4-20H2 protein lysine methyltransferases and methylation of novel substrates,** *Journal of Molecular Biology, 2016 Jun 5;428(11):2344-58. doi: 10.1016/j.jmb.2016.04.015. Epub 2016 Apr 20.*

Weirich S, Kusevic D, Kudithipudi S & Jeltsch A. (2015). **Investigation of the methylation of Numb by the SET8 protein lysine methyltransferase**, *Scientific Reports*, *2015 Sep 22;5:13813. doi: 10.1038/srep13813.*

Weirich S, Kudithipudi S & Jeltsch A. (2017). **Somatic cancer mutations in MLL1 histone methyltransferase modulate its enzymatic activity and dependence on the WDR5/RBBP5/ASH2L complex**, *Molecular Oncology*. *2017 Apr;11(4):373-387. doi: 10.1002/1878-0261.12041. Epub 2017 Mar 10.*

Weirich S, Kudithipudi S, Kycia I & Jeltsch A. (2015). **Somatic cancer mutations in the MLL3-SET domain alter the catalytic properties of the enzyme**, *Clinical Epigenetics*, *2015 Mar 28;7:36. doi: 10.1186/s13148-015-0075-3.*

Publications included in the present doctoral thesis:

Weirich S, Kudithipudi S, Jeltsch A. (2016*).* **Specificity of the Suv4-20H1 and Suv4-20H2 protein lysine methyltransferases and methylation of novel substrates,** *Journal of Molecular Biology, 2016 Jun 5;428(11):2344-58. doi: 10.1016/j.jmb.2016.04.015. Epub 2016 Apr 20.*

Weirich S, Kusevic D, Kudithipudi S & Jeltsch A. (2015). **Investigation of the methylation of Numb by the SET8 protein lysine methyltransferase**, *Scientific Reports*, *2015 Sep 22;5:13813. doi: 10.1038/srep13813.*

Weirich S, Kudithipudi S & Jeltsch A. (2017). **Somatic cancer mutations in MLL1 histone methyltransferase modulate its enzymatic activity and dependence on the WDR5/RBBP5/ASH2L complex**, *Molecular Oncology*. *2017 Apr;11(4):373-387. doi: 10.1002/1878-0261.12041. Epub 2017 Mar 10.*

Weirich S, Kudithipudi S, Kycia I & Jeltsch A. (2015). **Somatic cancer mutations in the MLL3-SET domain alter the catalytic properties of the enzyme**, *Clinical Epigenetics*, *2015 Mar 28;7:36. doi: 10.1186/s13148-015-0075-3.*

Zusammenfassung

Die Lysin-Methylierung von Histonen ist ein epigenetischer Mechanismus, der an der Regulation vieler biologischer Prozesse beteiligt ist. In den letzten Jahren hat das Interesse an der Lysin-Methylierung von Proteinen deutlich zugenommen, da eine Vielzahl von Protein-Lysin-Methyltransferasen (PKMTs) und Lysin-Methylierungsstellen in den Genomen und Proteomen von verschiedenen Organismen identifiziert wurden. Zusätzlich wurden auch Nicht-Histon-Proteine als mögliche Substrate für PKMTs entdeckt. Diese rasante Entwicklung hat die Lysin-Methylierung und deren Verständnis in den Mittelpunkt der Forschung gerückt. Um unser Wissen über die Lysin-Methylierung weiter zu verbessern, ist es wichtig die spezifischen Lysin Methylierungsstellen mit den verantwortlichen PKMTs in Verbindung zu bringen und das gesamte Substrat-Spektrum von PKMTs zu untersuchen. In dieser Arbeit wurde eine Substratspezifitäts-Analyse durchgeführt, um dieser Herausforderung gerecht zu werden. Es wurde gezeigt, dass die Methylierung von Peptidarrays ein guter Ansatz ist, um die Substratpräferenz von PKMTs zu analysieren und um geringfügige Unterschiede zwischen verwandten Enzymen mit der gleichen Gesamtspezifität zu identifizieren. Des Weiteren wurde gezeigt, dass die Substratspezifitäts-Analyse für die Identifizierung neuer Substrate von Nutzen ist, was in der vorliegenden Arbeit für SUV4-20H1, SUV4-20H2, MLL1 und MLL3 erfolgreich nachgewiesen wurde.

In vitro Methylierungsexperimente haben gezeigt, dass SUV4-20H1 und SUV4-20H2 unter der Verwendung von monomethyliertem H4K20 als Substrat eine Dimethylierung von H4K20 katalysieren. SUV4-20H1 und SUV4-20H2 haben ein überlappendes Sequenzmotiv, wobei SUV4-20H2 eine geringere Spezifität aufweist. Dieses Ergebnis wurde durch die Identifizierung eines neuartigen Nicht-Histon-Substrats von SUV4-20H1 und von drei Nicht-Histon-Substraten für SUV4-20H2 bestätigt.

MLL1 und MLL3 sind H3K4-Methyltransferasen, aber sie gehören zu verschiedenen MLL-Unterfamilien. MLL1 katalysiert die H3K4-Trimethylierung an Promotoren von Entwicklungsgenen, während MLL3 die Monomethylierung von H3K4 an Enhancern katalysiert. MLL1 und MLL3 sind beide Teil eines Multiproteinkomplexes, der unter anderem aus WDR5, RBBP5 und ASH2L besteht. Die Substratspezifitäts-Analyse von MLL1 hat gezeigt, dass MLL1 neben den Aminosäuren der ursprünglichen Sequenzen von H3 auch mehrere andere Aminosäuren an zahlreichen Stellen der Zielsequenz toleriert. Auf Proteinebene wurden zwei neue Substrate (TICRR und ZNF862) von MLL1 methyliert. Der Vergleich der relativen Aktivität hat gezeigt, dass das H3 Protein das bevorzugte Substrat in

Abwesenheit der Komplexpartner war, wobei ZNF862 in Gegenwart von WRA bevorzugt wurde. Letztendlich haben die Ergebnisse gezeigt, dass sich die Substratspezifität von MLL3-WRA geringfügig von MLL1 unterscheidet, was darauf hindeutet, dass MLL3-WRA und MLL1 unterschiedliche Nicht-Histon-Substrate erkennen können.

In mehreren Publikationen wurde über die Zuordung von PKMTs und methylierte Histon- und Nicht-Histon-Substratproteine berichtet, deren Ergebnisse jedoch fragwürdig sind. Dies könnte zu einer falschen Interpretation biologischer Prozesse und zur Irreführung von Folgeuntersuchungen führen. In dieser Arbeit wurde anhand von zwei Beispielen gezeigt, dass die Substratspezifitäts-Analyse verwendet werden kann, um problematische Zuordnungen zwischen PKMT und Methylierungsereignis zu identifizieren, welche experimentell untersucht werden müssen, um die veröffentlichten Ergebnisse zu überprüfen.

Vougiouklakis und Mitarbeiter berichteten 2015, dass K302 und K361 von ERK1 durch SUV4-20H1 methyliert würden. Allerdings passen diese Methylierungsstellen nicht zum Spezifitätsprofil von SUV4-20H1. Tatsächlich konnte keine Methylierung von ERK1 durch SUV4-20H1 oder SUV4-20H2 auf Peptid- und Protein-Ebene nachgewiesen werden, obwohl Positivkontrollen die erwartete Methylierung aufweisen. Die Arbeitsgruppe um Dhami berichtete 2013, dass das Numb Protein durch SET8 an K158 und K163 methyliert wird, was wiederum nicht mit den Spezifitätsergebnissen von SET8 übereinstimmt. In dieser Arbeit wurde die Methylierung von Numb auf Peptid- und Proteinebene unter der Verwendung von rekombinantem SET8 (aufgereinigt aus *E.coli* oder HEK293-Zellen) untersucht. In beiden Fällen konnte keine Methylierung von Numb beobachtet werden. Diese Ergebnisse deuten darauf hin, dass diese Zuordnungen zwischen Methylierungssubstraten und PKMTs wahrscheinlich nicht richtig sind.

In Genom- und Transkriptom Sequenzierungsprojekten wurden häufig auftretende somatische Mutationen in epigenetischen Enzymen gefunden, die zur Krebsentwicklung führen. Somatische Krebsmutationen können „loss-of-function“ oder „gain-of-function“ Effekte auf die enzymatischen Eigenschaften von PKMTs haben. Insbesondere die „gain-of-function“ Effekte stellen für das Verständnis ihrer Rolle bei der Krebsentstehung eine Herausforderung dar. In dieser Arbeit wurden die Effekte von somatischen Krebsmutationen analysiert, die in der SET-Domäne von MLL1 und MLL3 gefunden wurden. Vier somatische Krebsmutationen von MLL1 und drei von MLL3 wurden aufgrund ihrer Position in der Nähe der Bindungsstellen von AdoMet, Peptid oder den Interaktionspartnern ausgewählt. Die Untersuchung der somatischen Krebsmutationen in MLL1 und MLL3 hat gezeigt, dass jede

spezifische Mutation einen speziellen Effekt auf die enzymatische Aktivität, die Produkt- oder Substratspezifität und den prinzipiellen regulatorischen Mechanismus hat. Das bedeutet, dass für jede Mutante spezifische und detaillierte experimentelle Untersuchungen erforderlich sind, um ihre karzinogene Wirkung zu verstehen. Darüber hinaus haben Inhibitorstudien gezeigt, dass jede Mutante experimentell untersucht werden muss, um die Entwicklung von mutationsspezifischen therapeutischen Strategien zu ermöglichen.

Abstract

Histone lysine methylation is an epigenetic mechanism that is involved in the regulation of many biological processes. Over the last decade, the global interest in protein lysine methylation events increased drastically, because several protein lysine methyltransferases (PKMTs) and lysine methylation sites were identified in the genomes and proteomes of many organisms also including non-histone proteins functioning as substrates for PKMTs. The fast development of this field has moved the understanding of the biological outcome of lysine methylation into the center of research. Most urgently, it is necessary to improve our knowledge about lysine methylation by connecting specific target sites with the responsible PKMT and identifying the full substrate spectrum of PKMTs. In this thesis substrate specificity analysis was performed to tackle this challenge. It was shown that methylation of substrate specificity arrays is a good approach to analyze the substrate preference of PKMTs and identify subtle differences between related enzymes with same overall specificity. Furthermore, substrate specificity analysis was shown to be useful for the identification of novel substrates, which was successfully demonstrated for SUV4-20H1, SUV4-20H2, MLL1 and MLL3 in the present study.

In vitro methylation experiments indicated that SUV4-20H1 and SUV4-20H2 introduce dimethylation on H4K20 using monomethylated H4K20 as substrate. SUV4-20H1 and SUV4-20H2 have an overlapping sequence motif, but SUV4-20H2 is less specific. This result was supported by the identification of one novel non-histone substrate for SUV4-20H1 and three non-histone targets for SUV4-20H2.

MLL1 and MLL3 are H3K4 methyltransferases, but they belong to different MLL subfamilies. MLL1 catalyzes H3K4 trimethylation at promotors of developmental genes, whereas MLL3 introduces H3K4 monomethylation at enhancers. MLL1 and MLL3 are parts of related multi protein complexes also containing WDR5, RBBP5 and ASH2L. Substrate specificity analysis of MLL1 showed that it accepts several other residues at many positions of the target sequence, in addition to the residues in the original sequences of H3. At the protein level two novel substrates (TICRR and ZNF862) were methylated by MLL1. Comparison of the relative activity showed that the H3 protein was the best target in the absence of complex partners, but ZNF862 was preferred in presence of WRA. Finally, my data indicate that the substrate specificity of MLL3-WRA differed slightly from MLL1, suggesting that they may have different non-histone substrates.

In several publications, assignments between PKMTs and methylated histone or non-histone target sites have been reported, but in some cases the data are questionable. This could lead to wrong interpretation of biological processes and misleading of follow-up studies. It has been shown for two examples in this study, that substrate specificity analysis can be used to identify problematic assignments between PKMT and methylation events, which need to be studied experimentally to confirm the published findings.

Vougiouklakis et al. (2015) reported that SUV4-20H1 methylates ERK1 at K302 and K361, but these target sites do not fit to the specificity profile of SUV4-20H1. Indeed, I could not detect methylation of ERK1 by SUV4-20H1 or SUV4-20H2 at peptide and protein level although positive controls showed the expected methylation. Dhami et al. (2013) reported that Numb protein is methylated by SET8 at K158 and K163, which was not in agreement with the specificity data of SET8. In this thesis, Numb peptide and protein methylation was studied using recombinant SET8 purified from *E.coli* or HEK293 cells. In both cases, no methylation of Numb could be observed. These data suggest that these assignments of methylation substrates and PKMTs are likely not correct.

Whole genome and whole transcriptome sequencing projects have frequently found somatic mutations in epigenetic enzymes in cancers. Somatic cancer mutations can have loss-of-function or gain-of-function effects on the enzymatic properties of PKMTs. Especially gain-of-function effects are a challenge in understanding their role in carcinogenesis. In this study, the effects of somatic cancer mutations found in the SET domain of MLL1 and MLL3 were analyzed. Four somatic cancer mutations of MLL1 and three of MLL3 were selected for analysis on the basis of their location close to binding sites of AdoMet, peptide or the interaction partners. The investigation of somatic cancer mutations in MLL1 and MLL3 indicated that each specific mutation has its unique effect on the enzymatic activity, product or substrate specificity and principle regulatory mechanism indicating that each mutant needs specific in depth experimental investigation in order to understand its carcinogenic effect. Moreover, inhibitor studies demonstrated that each mutant needs to be experimentally studied to allow for the development of mutation specific therapeutic strategies.

List of Abbreviations

Abbreviation	Discription
53BP1	Tumor suppressor p53 binding protein 1
ac	Acetylation
AdoHcy	S-adenosyl-L-homocysteine
AdoMet	S-adenosyl-L-methionine
AF4/9/10	ALL1-fused gene from chromosome 4/9/10
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
ASH2L	Absent small homeotic-2-like protein
Bromo	Bromodomain
CD	Circular Dichroism
Chromo	Chromatin organization modifier
COSMIC	Catalogue Of Somatic Mutations In Cancer
CpG	Cytosine phosphate guanine site
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dot1l/DOT1L	Disruptor of telomeric silencing 1-like protein
DPY30	Dumpy-30
dsRNA	Double stranded RNA
ENL	YEATS domain-containing protein 1
ERK1	Extracellular signal-regulated kinase 1
EZH2	Enhancer of Zeste 2
FAD	Flavin adenine dinucleotide
FYRN	Phenylalanine-Tyrosine rich domain N-terminal
FYRC	Phenylalanine-Tyrosine rich domain C-terminal
G9a	Euchromatic histone-lysine N-methyltransferase 2
GLP	G9a-like protein
GST	Glutathione S-transferase
H1	Histone 1
H2A	Histone 2A
H2B	Histone 2B
H3	Histone 3
H3K27	Histone 3 lysine 27

H3K36	Histone 3 lysine 36
H3K4	Histone 3 lysine 4
H3K79	Histone 3 lysine 79
H3K9	Histone 3 lysine 9
H4	Histone 4
H4K20	Histone 4 lysine 20
H4K44	Histone 4 lysine 44
HEK293	Human embryonic kidney 293 cells
Hox	Homeobox gene
HP1	Heterochromatin protein 1
iSET	Inserted region
JHDM	Jumonji domain-containing histone demethylase
KDM	Lysine demethylase
KMT	Lysine methyltransferase
L3MBTL1	Human lethal 3 malignant brain tumor repeat-like protein
LSD1/2	Lysine specific demethylase 1/2
MALDI	Matrix Assisted Laser Desorption Ionization
MBT	Malignant brain tumor
me	Methylation
me1	Monomethylation
me2	Dimethylation
me3	Trimethylation
MeK	Methyllysine
miRNA	Micro RNA
MLL	Mixed lineage leukemia
ncRNA	Noncoding RNA
NGS	Next generation sequencing
NHT	Non-histone target
NSD1/2	Nuclear receptor SET domain-containing protein 1/2
P	Phosphorylation
p53	Cellular tumor antigen p53
PRC2	Polycomb repressive complex 2
PHD	Plant homeodomain
piRNA	Piwi-interacting RNA

PKDM	Protein lysine demethylase
PKMT	Protein lysine methyltransferase
PRMT	Protein arginine methyltransferase
PTM	Posttranslational modification
PWWP	Proline-tryptophan-tryptophan-proline motif containing domain
RA	RBBP5-ASH2L heterodimer
RB	Retinoblastoma protein
RBBP5	Retinoblastoma-binding protein 5
RNA	Ribonucleic acid
RNAi	RNA interference
RRM	RNA recognition motif
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SET	Su(var)3-9; Enhancer of zeste (E(z)); Trithorax (Trx)
Set1	SET domain-containing protein 1
SET1A	SET domain-containing protein 1A
SET1B	SET domain-containing protein 1B
SET7/9	SET domain-containing protein 7/9
SET8	SET domain-containing protein 8
siRNA	Small interfering RNA
SMYD	SET and MYND domain-containing protein
Su(var) 3-9	Protein suppressor of variegation 3-9
SUV39H1/H2	Human protein suppressor of variegation 3-9 homolog 1/2
SUV4-20H1/H2	Human protein suppressor of variegation 4-20 homolog 1/2
Trr	Trithorax-related
Trx	Trithorax
UTX	Ubiquitously transcribed TPR protein transcribed on the X-chromosome
VDV motif	Valine-aspartate-valine motif
WDR5	Tryptophan-aspartate repeat protein 5
Wdr82	WD repeat-containing protein wdr-82
WIN	WDR5 interaction motif
WRA	WDR5/RBBP5/ASH2L
YFP	Yellow fluorescent protein

1. Introduction

1.1. Epigenetics

In 1942, Waddington introduced an influential definition of the research field of "Epigenetics". He defined epigenetics as "the branch of biology, which studies the causal interactions between genes and their products that bring the phenotype into being" (Waddington 2012). During the past years, research in epigenetics has gained more and more importance and also the definition of the field has changed over time. Riggs and colleagues described epigenetics as "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" (Riggs, 1996). Epigenetic mechanisms, including DNA methylation, histone modification and noncoding RNA molecules, alter the structure of chromatin and thus regulate the expression of specific genes. The epigenetic phenomenon describes the differentiation of cells into different cell types and the specialization into multicellular organism as well as the adaptation of cells and organism to environmental changes (Goldberg et al. 2007). The involvement of epigenetic mechanisms in the development of diseases, such as cancer, highlights the importance of epigenetics. In the past, cancers were mainly connected with mutations of oncogenes and tumor suppressor genes. Today it is known that mutations in the genome can influence enzymatic properties of epigenetic proteins that lead to changes in the readout of the genome. Stephen Baylin was one of the first researchers who identified epigenetic silencing of a tumor suppressor gene by DNA hypermethylation, which causes decreased protein expression and results in tumor growth (Brower 2011; Portela & Esteller 2010). Furthermore, environmental effects, like smoking, fatty food, etc. can influence the methylation level of genes and by this impact the development of cancers (Brower 2011). Hence, the participation of epigenetics in the development of cancers provides a new paradigm for cancer therapy and prevention, which is a topic of high interest these days.

1.2. Chromatin structure

In eukaryotic cells, chromatin is a highly organized polymer which represents the physiological template of all genetic information. The basic repeating unit of chromatin is the nucleosome, which consists of a histone octamer containing duplicates of each core histone protein, H3, H4, H2A and H2B, and a 147 bp long DNA fragment wrapped 1.7 times around the histone core (Luger et al. 1997; Kouzarides 2007). Adjacent nucleosomes are connected to each other by a 20-60 bp long linker DNA visualized as "beads-on-string" structure on isolated chromatin sample in electron microscopy images (Trojer & Reinberg 2007).

However, this first step of compaction is not enough to pack the 2 meter long human DNA filament into the cell nucleus. For further compaction, a linker histone of the H1 type binds to the linker DNA and nucleosomes, protecting the negative charge between adjacent nucleosomes and establishing the so-called “30 nm fiber”. The 30 nm fiber is a helix of 6 to 7 nucleosomes per 11 nm with a diameter of 30 nm (Jenuwein & Allis 2001; Luger et al. 1997; Song et al. 2014). This and further levels of compaction ultimately lead to the most condensed form of DNA called chromosomes (Figure 1) (Trojer & Reinberg 2007; Margueron et al. 2005).

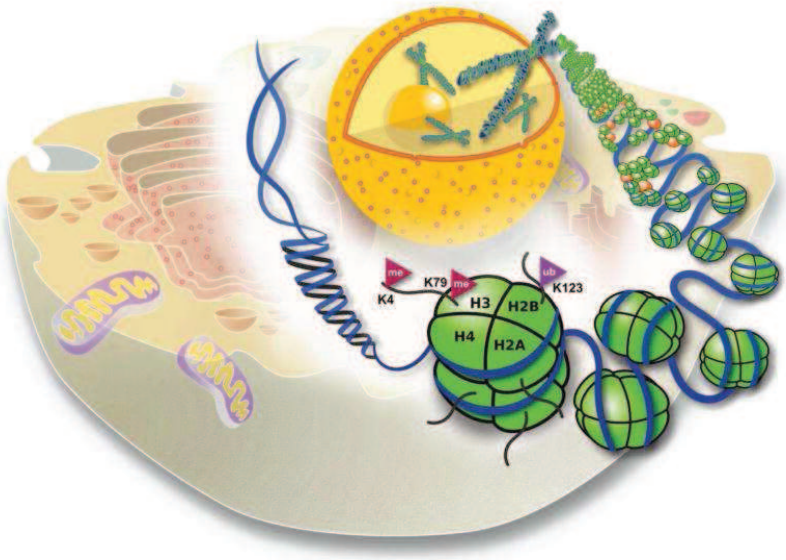


Figure 1: Schematic representation of eukaryotic DNA located in the cell nucleus. In the nucleus of a eukaryotic cell, the DNA is compacted due to the interaction of the histone proteins. Duplicates of the core histone proteins H3, H4, H2A and H2B form the nucleosome, which is surrounded by the DNA. The unstructured C-and N-terminal tails of the core histone proteins are protruding from the nucleosomes and are subject to numerous posttranslational modifications, like H3K4, H3K79 methylation or H2B monoubiquitination represented as examples in this figure. Further compactions lead to the formation of the highly condensed chromosome. The figure is taken from (Shilatifard 2012).

The core histone proteins are evolutionary conserved and form a globular structure with unstructured C- and N-terminal tails protruding from the nucleosomes (Jenuwein & Allis 2001). Both tails are subject to several posttranslational modifications (PTM) which regulate the chromatin structure and dynamics together with DNA methylation and the presence of non-coding RNA (Kouzarides 2007; Strahl & Allis 2000). Some parts of the chromatin, called euchromatin, are loosely packed and transcriptionally active, whereas heterochromatin, is more densely packed and contains inactive genes (Jenuwein & Allis 2001). Altogether the epigenetic marks affect a large number of very important cellular processes including transcription, DNA replication, DNA repair and genomic stability (Gelato & Fischle 2008).

1.3. DNA methylation

DNA methylation is an important epigenetic process that plays a critical role in gene regulation during embryogenesis and gametogenesis (Goll & Bestor 2005). DNA methyltransferases (DNMTs) catalyze the transfer of a methyl group from the cofactor S-adenosyl-L-methionine (AdoMet) to the C5 position of cytosine residues in CpG dinucleotides, generating 5-methylcytosine and S-adenosyl-L-homocysteine (AdoHcy) (Jeltsch & Jurkowska 2016; Jurkowska et al. 2011). CpG sites are often found in CpG islands, regions with a high density of CpG dinucleotides. An enrichment of methylated CpG islands is often discovered at noncoding regions like centromeric heterochromatin, imprinted genes or transposons to silence the expression of these elements and prevent genomic instability. Conversely, CpG islands are often unmethylated in regulatory elements like promotors and enhancers of specific genes (Jin & Robertson 2013). In mammals, four different DNMTs are present, DNMT1, DNMT3A, DNMT3B and DNMT3L (Jeltsch & Jurkowska 2016). DNMT3A and DNMT3B are so-called "de novo methyltransferases" (Mtases) and are responsible for setting the initial methylation pattern in early embryonic development. DNMT3L is catalytically inactive, but essential for the catalytic activity and genomic specificity of DNMT3A and DNMT3B. After DNA replication, fully methylated DNA is converted to hemimethylated DNA. The so-called maintenance DNA methyltransferase DNMT1 is responsible for remethylation of the DNA to retain inherit methylation pattern (Portela & Esteller 2010; Jurkowska et al. 2011). It has been reported that changes in the DNA methylation pattern contribute to the development of diseases, like breast, prostate and colorectal cancer (Miremadi et al. 2007; Portela & Esteller 2010).

1.4. Noncoding RNA

High-throughput genome wide studies have shown that only ~ 2% of the transcribed human genome encodes for proteins, whereas the rest is transcribed as noncoding RNA (Amaral et al. 2008). Noncoding RNA (ncRNA) molecules are important for the epigenetic regulation of gene expression and chromatin remodeling (Bernstein & Allis 2005). They are categorized into small and long ncRNAs. Small ncRNAs include micro RNAs (miRNAs), piwi-interacting RNAs (piRNAs), and small interfering RNAs (siRNAs) (Kaikkonen et al. 2011). In 1998, Fire discovered the RNA interference (RNAi) process in *Caenorhabditis elegans*, which describes the phenomenon that double stranded RNA (dsRNA) can trigger gene silencing (Fagard & Vaucheret 2000). Up to date this is one of the best-studied post-transcriptional gene silencing mechanism based on noncoding RNAs. Beside RNA silencing, which is used in many organism to fight against infections with RNA viruses or transposable

elements, noncoding RNAs play a role in heterochromatin formation, histone modification and DNA methylation by the recruitment of transcription factors, chromatin modifying complexes or by direct interaction with homologous RNA or DNA strands (Waterhouse et al. 2001; Gitlin & Andino 2003; Bonasio et al. 2010; Chen 2016).

1.5. Posttranslational modifications of histone proteins

Beside DNA methylation and noncoding RNAs, posttranslational modifications (PTM) of histone proteins play a key role in the regulation of chromatin and thereby influence many cellular processes, like transcription, DNA repair, replication and recombination (Bannister & Kouzarides 2011). As already mentioned, core histones have a predominantly globular structure with flexible C- and N-terminal tails, which are unstructured and protrude from the nucleosomes core body. The histones and particularly the N-terminal tails of H3 and H4 are subject to a large number and type of modifications; especially modifications of the N-terminal tails are well characterized (Figure 2) (Zhao & Garcia 2015). Altogether, at least 8 chemically different modifications are occurring at more than 60 different sites and the list is still growing (Rousseaux & Khochbin 2015; Xu et al. 2014). Extensively studied modifications are phosphorylation (S and T), acetylation (K) and methylation (K and R). Apart from that, also ubiquitination (K), symoylation (K), ADP ribosylation, glycosylation, biotinylation and carbonylation are established PTMs (Margueron et al. 2005). Histone modifications can directly influence the chromatin structure by changing the charge of the amino acid. For example, acetylation neutralizes the positive charge of a lysine residue and thereby weakens the electrostatic interaction between nucleosomes and the surrounding DNA. Hence, the chromatin structure is converted to a less compact state which facilitates the accessibility of chromatin and leads to transcriptional activation. Indirectly, these modifications can recruit effector protein complexes with specific binding properties, also called “reading domains” (Bannister & Kouzarides 2011; Kouzarides 2007; Fischle, Wang & Allis 2003). They are categorized in groups, depending on the detection of distinct histone modifications. Effector proteins containing a bromodomain recognize acetylation, while phosphorylation is known to be bound by 14-3-3 proteins. Methylation can be identified by chromo-like domains of the Royal family (chromo, tudor, PWWP and MBT), ankyrin repeats and PHD domains (Kouzarides 2007). The heterochromatin protein (HP1) is one example for an effector protein that recognizes the H3K9me3 mark with its chromodomain and majorly facilitates the heterochromatin formation and gene repression (Daniel et al. 2005).

Histone posttranslational modifications are very dynamic compared to DNA methylation and specific sites of the genome can undergo various modifications. Lysine residues, for example, are subject to methylation (mono-, di-, and tri-), acetylation, acylation, ubiquitination and sumoylation, but these different modifications cannot co-exist at the same time at the same residue. In agreement with this presumption, researchers identified complexes, which contain histone deacetylases together with histone methyltransferases. In this case an acetyl group is removed from a lysine, before a methyl group can be introduced (Vaute et al. 2002). Generally, lysine acetylation is always associated with activated genomic regions, whereas lysine methylation can be correlated with transcriptional activation and repression, depending on the genomic site and level of methylation (Bannister & Kouzarides 2011; Fischle, Wang, Jacobs, et al. 2003). Histone modifications can influence each other, when they are located on the same histone tail, a process called "cis effect". Interestingly, also communication exists between adjacent nucleosomes, characterized as "trans effect". For example, it was shown that ubiquitination of H2B is necessary for the methylation of H3K4 and H3K79 (Fischle, Wang & Allis 2003; Kouzarides 2007). Altogether, the sum of all possible PTM combinations at different sites is described as the "histone code", which may encode specific functional states (Jenuwein & Allis 2001).

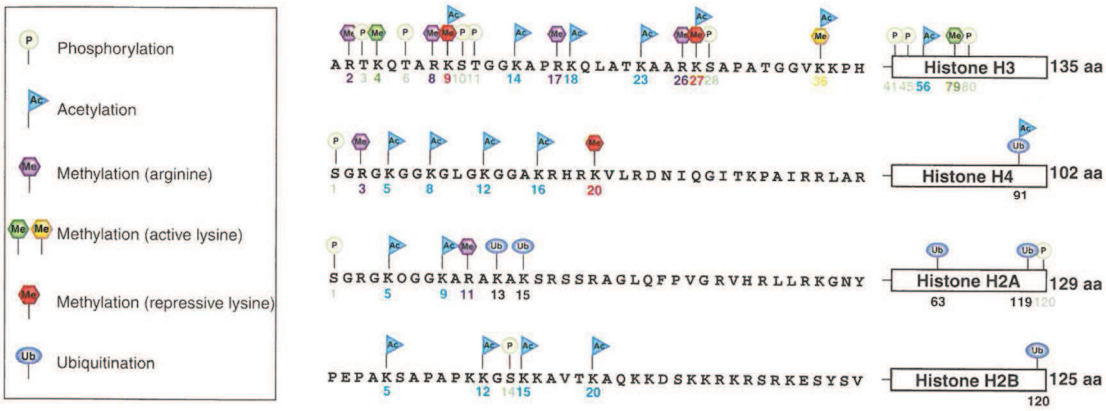


Figure 2: Schematic overview of posttranslational histone modifications. The main posttranslational modifications of the unstructured N-terminal tail and globular domain (boxed) of H3, H4, H2A and H2B are represented. Phosphorylation (P) is shown as light green circle, acetylation (Ac) as turquoise flag, arginine methylation (Me) as violette hexagon, active lysine methylation (Me) as green and yellow hexagon, repressive lysine methylation (Me) as red hexagon and ubiquitination (Ub) as blue ellipse. The figure is taken from the book "Epigenetics Second Edition". New York: Cold Spring Harbor Laboratory Press; Chapter 3, page 57; 2015.

1.6. Histone methylation

Histone methylation on lysine and arginine residues is one of the most studied posttranslational modification. It is found on all four histone proteins, but majorly on the H3

and H4 N-terminal tail. Lysine residues can be mono-, di- and trimethylated at the ε-amino group and arginine residues can be mono- or dimethylated at the ω-guanidino group, either symmetrically or asymmetrically (Bannister & Kouzarides 2011; Margueron et al. 2005). Lysine methylation is introduced by protein lysine methyltransferases (PKMT) and is removed by protein lysine demethylases (PKDM). Arginine methylation is set by protein arginine methyltransferases (PRMTs), but no specific enzyme has been identified to reverse arginine methylation (Allis et al. 2007). However, an indirect effect is known, where the protein arginine deiminase converts methylarginine to citrulline (Bannister & Kouzarides 2005; Kouzarides 2007). The subject of interest of this thesis mainly was lysine methylation, which was first described in 1964 by Murray (Murray 1964). Different from histone acetylation, methylation does not change the electrostatic interaction between DNA and histones. Instead, effector protein complexes are recruited to the specifically methylated lysine residues that subsequently alter the chromatin properties (Bannister & Kouzarides 2005; Kouzarides 2007; Gelato & Fischle 2008; Martin & Zhang 2005). As the biological outcome of histone methylation is dependent on the methylated site and level of methylation, its understanding is very complex. Five major lysine methylation sites on histones have been extensively studied, viz. H3K4, H3K9, H3K27, H3K36, H3K79 and H4K20 (Kouzarides 2007). Out of these, H3K4, H3K36 and H3K79 methylation are associated with transcriptional activity, while methylation of H3K9, H3K27 and H4K20 is found on repressed chromatin (Kouzarides 2007).

1.7. Protein lysine demethylases

For a long time, methylation of the ε-amino group of lysine was considered to be irreversible due to the formation of a strong C-N bond. Still, it was known that the methylation level changes during biological processes, for instance cellular differentiation and the existence of enzymes catalyzing the removal of methyl groups was expected (Bannister et al. 2002). Finally, in 2004 Shi et al. identified the first lysine specific demethylase 1 (LSD1, KDM1A), which catalyzes the process of H3K4 N^{ε}-demethylation (Shi et al. 2004). KDM1A catalyzes an amine oxidation reaction, using a flavin adenine dinucleotide as cosubstrate. After the transfer of a hydride from the N^{ε}-methyl group onto FAD, an imine is produced, which is unstable to hydrolysis and the methyl group is released as formaldehyde. LSD1 and its paralog LSD2 are reported to demethylate H3K4me1/me2 and H3K9me1/me2. However, demethylation of trimethylated lysine targets is precluded, because the amine oxidation reaction requires protonated nitrogen as substrate. Two years later, Tsukada et al. discovered

another histone demethylase, JHDM1 (JmjC domain-containing histone demethylase 1) that demethylates H3-methyl K36 (Tsukada et al. 2006). Over time many more Jumonji type histone demethylases were discovered and by now they are categorized in six main families called lysine demethylases (KDM2-7), dependent on their amino acid sequence (Thinnes et al. 2014). The Jumonji-containing lysine demethylases catalyze an Fe(II) dependent dioxygenase reaction using α-ketoglutarate and dioxygen as cosubstrates, generating formaldehyde and succinate. In contrast to the KDM1 subfamily, which is only able to catalyze demethylation of mono- and dimethylated lysine residues, KDM2-7 enzymes can catalyze the removal of all three methylation forms at multiple sites. Overall both subfamilies have overlapping functions and dysregulation seems to be linked to diseases, including cancer and genetic disorders (Shi et al. 2004; Thinnes et al. 2014; Tsukada et al. 2006; Upadhyay & Cheng 2011).

1.8. Protein lysine methyltransferases

The discovery of the first histone lysine methyltransferase Suv39h1 in 2000 by Jenuwein and colleagues was a milestone for understanding lysine methylation is the epigenetic control (Rea et al. 2000). During the last more than 15 years, the research on posttranslational modifications and enzymes, regulating these epigenetic mechanisms, made an impressive progress. By now approximately 80 PKMT enzymes have been identified (Zhang et al. 2012; Jenuwein 2006). Having more detailed information about each enzyme available, it became clear that the network of all known PKMTs is more complex than expected before. Different PKMTs can methylate the same target site, whereas also one single PKMT can catalyze the methylation of different target sites. Moreover, the methylation levels, catalyzed by specific protein lysine methyltransferases, can be influenced by the recruitment of interaction partners. Initially, only histones were considered as substrates for histone lysine methyltransferases, but subsequently also non-histone targets (NHT) were identified as target substrates for some PKMTs (G9a, SET7/9) (in more detail 1.9). The investigation of novel non-histone targets of different PKMTs is still in progress, which means this field will experience further progress in future (Zhang et al. 2012).

To keep an overview, enzymes catalyzing the lysine and arginine methylation of proteins are divided in three distinct protein families: the protein arginine methyltransferases (PRMT), the non-SET domain containing protein lysine methyltransferases like Dot1/DOT1L and the SET domain containing protein lysine methyltransferases (PKMT) (Feng et al. 2002; Dillon et al. 2005). Regarding their substrate specificity and sequence homology, PKMTs are divided in 8 families called lysine methyltransferases (KMT) 1-8 (Allis et al. 2007).

1.8.1. The overall SET domain structure

Most PKMTs contain a SET domain, which is a conserved 130 amino acids long structure found in enzymes from *Saccharomyces cerevisiae* to human (Dillon et al. 2005; Rea et al. 2000). The name **SET** arise from the first three identified proteins in *Drosophila melanogaster* containing this domain: Suppressor of variegation 3-9 (**S**u(var)3-9), a modifier of position-effect variegation; Enhancer of zeste (**E**(z)), a chromatin regulator of the Polycomb group and Trithorax (**T**rx), the chromatin regulator of the Trithorax group (Dillon et al. 2005; Rea et al. 2000). The structural fold of the SET domain is unique compared to other methyltransferases using AdoMet as methyl-group donor, like DNA methyltransferases or protein arginine methyltransferases. The core SET domain consists of several β-sheets, which are organized into three distinct sheets, surrounding a so-called "pseudo-knot" structure. This knot-like structure is formed by the C-terminal tail of the SET domain threading through a loop, which is formed by a sequence protruding from the structure. Moreover, the pseudoknot contains the two most conserved sequence motifs ELxF/YDY and NHS/CxxPN and thereby it is very important to bring the cofactor binding site and the peptide binding cleft in close proximity to the active pocket (Figure 3) (Qian & Zhou 2006).

At the N- and C-termini, the core SET domain is flanked by the pre- and post-SET domains, respectively, which are not conserved within the protein lysine methyltransferases. While the pre-SET domain works as a stabilizing element by interacting with domains of the core SET domain, the postSET domain completes the active site and forms the characteristic hydrophobic channel, which will be explained later in detail (Qian & Zhou 2006; Dillon et al. 2005; Xiao et al. 2003). Some PKMTs additionally contain an inserted region (iSET), which varies in length and is not conserved. In the three dimensional structures of SET domains bound to cofactor and substrate, it was shown that the iSET region directly binds the peptide substrate and thus contributes to the substrate specificity. For example, MLL1 and SET7/9 are both H3K4 methyltransferases, but contain enzyme specific iSET regions and thereby recognize different residues in the same target substrate (Xiao et al. 2003; Qian & Zhou 2006).

1.8.2. The catalytic active pocket

Compared to other AdoMet dependent methyltransferases, SET domain containing lysine methyltransferases have a unique rearrangement of the substrate peptide binding pocket and AdoMet binding site, which are located at opposite sides of the SET domain connected through a narrow hydrophobic channel. When the cofactor is bound, AdoMet is positioned in

a U-shaped conformation due to the hydrogen bond formation to residues in the conserved NHS/CxxPN motif of the pseudo-knot structure and the GxG motif in the preSET domain. In addition, the carboxylate of AdoMet forms a salt bridge with a conserved arginine or lysine in the SET domain. This conformation is necessary to place the AdoMet methyl group at the base of a hydrophobic channel. On the other side of this channel, the lysine residue of the substrate is inserted, such that the channel links the substrate peptide with the cofactor AdoMet. Within the hydrophobic channel, a conserved tyrosine residue deprotonates the ε-amino group of the lysine site chain and allows for a nucleophilic attack of lysine on the AdoMet methyl-group.

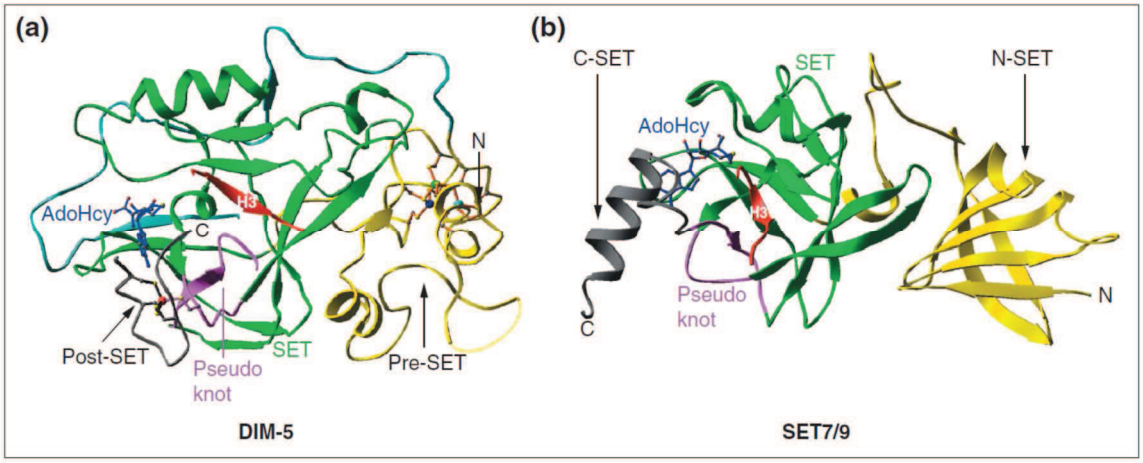


Figure 3: SET domain structures of (a) *Neurospora crassa* DIM-5 and (b) human SET7/9. The structures show the PostSET, SET and Pre-SET domain in grey, green and yellow, respectively. Also indicated is the characteristic pseudo-knot structure in pink. The substrate peptide H3 is visualized in red and the cofactor product AdoHcy as a stick model in blue. The figure is taken from (Dillon et al. 2005).

1.8.3. The Phe/Tyr switch controlling the product specificity of PKMTs

As already mentioned, the postSET domain contributes to the formation of the active site cavity of SET domain PKMTs. Mutational analyses have shown that the nature of one particular residue in the postSET region determines the methylation level of specific enzymes. Enzymes containing a tyrosine residue at this position, like Tyr245 of SET7/9, are monomethyltransferases, whereas enzymes containing a phenylalanine residue at this position, like Phe281 of Dim-5, catalyze trimethylation. This effect can be explained by the reaction mechanism of the enzymes. During the reaction of methyl transfer the tyrosine (Tyr245 of SET7/9) interacts via hydrogen bonds with the ε-amine of lysine and by this orients the lone pair of the lysine towards the methyl group of AdoMet. After deprotonation of the ε-amine of the lysine, an S_N2 nucleophilic attack of the lysine on the methyl group of AdoMet occurs and monomethyllysine together with AdoHcy are produced. As the hydrogen

bond is still present, the rotation of the lysine ε-carbon-nitrogen bond is hindered. Conversely, phenylalanine at this position (Phe281 in Dim-5) allows the lysine ε-carbon-nitrogen bond to rotate, so that the ε-amine can be further deprotonated without release of methyllysine in a processive methylation mechanism (Copeland et al. 2009; Smith & Denu 2009). Mutation of Phe281 to tyrosine has been shown to reduce the methylation level of Dim5 from trimethylation to mono- and dimethylation. In this aspect, the F/Y switch position is an important site to discriminate how many methyl groups can be transferred from the methyl donor cofactor AdoMet to the ε-amino lysine residue, dependent on the geometry and size of the lysine channel (Dillon et al. 2005; Qian & Zhou 2006; Upadhyay & Cheng 2011; Xiao et al. 2003; Collins et al. 2005).

1.8.4. Readout of lysine methylation marks

As described so far, posttranslational modifications of histones are important for epigenetic control mechanisms. However, the question remains how these different chemical modifications can be translated into specific biological outputs. Two models of "direct" and "effector mediated" readout are established for the translation of the epigenetic code. By charge removal induced chromatin decondensation, histone acetylation can directly stimulate expression. As an alternative modified residues can recruit reading domains for the translation of the modification into a biological output (Bannister & Kouzarides 2005; Kouzarides 2007; Gelato & Fischle 2008; Martin & Zhang 2005).

In case of histone lysine methylation, reading domains provide a specific surface to recognize the posttranslational modification and to discriminate different methylation states. Major families of domains involved in the readout of methyl lysine marks are the Royal superfamily (chromodomain, tudor, PWWP and MBT), ankyrin and PHD domains (Patel & Wang 2013; Taverna et al. 2007). Methyllysine (MeK) binders typically have a binding pocket consisting of 2-4 aromatic residues and additionally acidic residues, forming an aromatic cage (Taverna et al. 2007). Considering the amount of aromatic residues, MeK binding pockets can be classified as "half" aromatic cage or "full" aromatic cage. To discriminate between the methylation levels of target sites, binding pockets with different recognition modes exist. The cavity insertion mode binds mono- and dimethylated lysine residues through hydrogen bonds and electrostatic interactions between the methylammonium proton and the carboxylate group of acidic residues (aspartate), whereas higher methylation forms are structurally excluded (tandem tudor domain of 53BP1 and L3MBTL1). Trimethylated lysine residues are recognized by reading domains with a "surface groove" pocket, which is lined by three

conserved aromatic amino acids. It can stabilize the trimethylated lysine by cation-π interactions (Taverna et al. 2007; Yun et al. 2011). Hence domains binding lower methylation states tend to have a half pocket, while Kme3 specific domain contain a full pocket. However, this classification does not always correspond to the methylation level of specific genome sites. Unmethylated lysine residues are bound by intermolecular hydrogen bonds, but corresponding readers do not have a characteristic binding pocket (BHC80 PHD domain) (Yun et al. 2011).

All members of the Royal family fold in an incomplete β-barrel structure consisting of four antiparallel β-strands and the binding of the histone peptide, also folded as β-strand, complete the β-barrel structure. Well characterized is the β-barrel conformation of the heterochromatin protein 1 (HP1) bound to H3K9me3, which contributes to gene silencing. The solved crystal structure of this complex showed that reading domains are also dependent on flanking sequences of the Kme3 residue to distinguish sequence contexts, because HP1 also binds three amino acids preceding and one amino acid following the H3K9me3 mark (Figure 4) (Taverna et al. 2007; Yun et al. 2011; Patel & Wang 2013). This indicates that readout of posttranslational modifications is sequence specific and the recruitment of additional proteins allows the system to work in concert with other epigenetic systems.

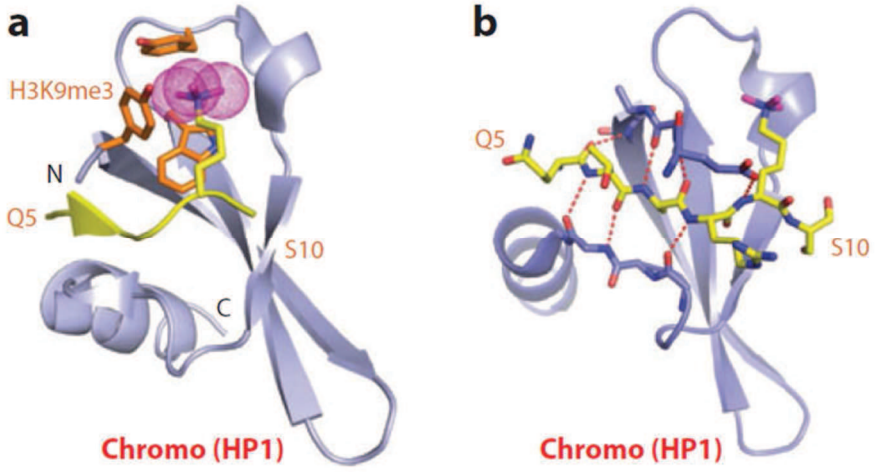
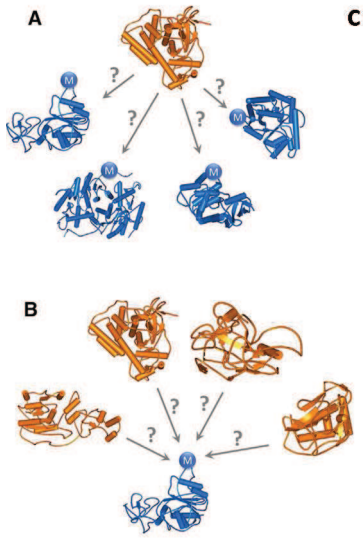


Figure 4: HP1 is a reading domain from the Royal superfamily. a) Crystal structure of HP1 bound to trimethylated H3(1-15)K9 peptide. It also recognizes flanking sequences of H3K9 from Q5 to S10. B) The peptide H3K9me3 folds as a β-strand and binds between two β-strands of HP1, resulting in the formation of a complete three stranded antiparallel β-barrel structure. The figure is taken from (Patel & Wang 2013).

1.9. Non-histone targets

High throughput proteomic studies led to the identification of many lysine methylation events at non-histone proteins, which are involved in the regulation of important biological

processes, like protein activity and stability, as well as interactions between different proteins (Zhang et al. 2015; Clarke 2013; Huang & Berger 2008; Lanouette et al. 2014). However, for many lysine methylation sites, the corresponding PKMT is not known or on the other hand, for many PKMTs, the complete variety of all possible cellular substrates is not yet identified (Figure 5A/B) (Kudithipudi & Jeltsch 2016). The first example of an histone lysine methyltransferase that can methylate a non-histone target was SET7/9, shown to methylate the tumor suppressor protein p53 (Chuikov et al. 2004). Beside SET7/9, also SMYD2, SET8 and G9a were recognized to methylate p53 at distinct lysine residues with different biological outcome. While SET7/9 methylation of p53 at lysine 372 results in activation of p53 regulated gene expression, methylation by SMYD2 (K370) and SET8 (K372) cause inactivation of p53 (Huang et al. 2006; Shi et al. 2007; Kurash et al. 2008). Afterwards additional PKMTs, like SMYD3, G9a, NSD1/2, SUV39H1/H2 etc. (Figure 5C) were shown to methylate non-histone substrates as well (Kudithipudi & Jeltsch 2016; Zhang et al. 2015). In addition, not only PKMTs can modify NHT, also the demethyltransferase LSD1 regulate the methylation level of p53 (Huang et al. 2007).



C

Table 1. Compilation of the PKMTs mentioned in this article

Name	Systematic Name	Alternate Name(s)	Example Substrates: Histone Substrates	Example Substrates: Non-histone Substrates
SET7/9	KMT7	SET7/SET9/ SETD7	H3K4	p53-K372 DNMT1-K142 MeCP2-K347
G9a	KMT1C	EHMT2	H3K9	CDYL1-K135 DNMT1-K70
SUV39H1	KMT1A		H3K9	
SUV39H2	KMT1B		H3K9	
SMYD2	KMT3C			p53-K370 RB-K860
SMYD3	KMT3E		H4K5	MAP3K2-K260 VEGFR1-K831
NSD1	KMT3B		H3K36 H1.5 K168	
NSD2	KMT3G	WHSC1/ MMSET	H3K36	
SET8	KMT5A	PR-SET7/ SETD8	H4K20	p53-K382
SUV4-20H1	KMT5B		H4K20	
SUV4-20H2	KMT5C		H4K20	
DOT1L	KMT4		H3K79	

Figure 5: Identification of novel targets for protein lysine methyltransferase. A) For most PKMTs, the list of possible targets is not identified yet. B) For most methylated target sites, the responsible PKMT is unknown. C) List of PKMTs, including their systematic names, alternate names, histone substrates and identified non-histone substrates. Figures are taken from (Kudithipudi & Jeltsch 2016).

With increasing interest in PKMTs, due to the discovery of non-histone proteins as possible targets, many methylation sites in the proteome of human cells were identified in a very short time, but also wrong assignment of PKMTs and methylation sites occurred. For example, NSD1 initially was characterized as H3K36 and H4K20 methyltransferase, however later sequence specificity analysis have shown that H4K20 does not fit to the specificity profile,

but rather H4K44 is methylated by NSD1 (Kudithipudi, Lungu, et al. 2014; Rayasam et al. 2003). Also SMYD2, initially reported as a H3K36 dimethyltransferase was shown to methylate p53 and retinoblastoma protein (RB) in addition to more NHT, whereas methylation of histone proteins could not be detected in later studies (Huang et al. 2006; Brown et al. 2006; Saddic et al. 2010). As dysregulation of protein lysine methyltransferases is often related to the development of diseases, wrong assignments of PKMT substrates can provoke big confusion, lead to wrong interpretation of biological processes and in the worst case to the development of misleading therapeutic strategies. To improve the reliability of the assignment of PKMTs with specific substrates and lysine methylation events, a guideline with seven rules was proposed recently (Kudithipudi & Jeltsch 2016). This guideline requires that methylation events of novel targets have to be further investigated. It is necessary to check if the methylation of the novel target is relevant by including positive controls. Furthermore, it has to be verified if the predicted target lysine is methylated and also if the protein lysine methyltransferase is really the methylating enzyme. The sequence of the putative non-histone target has to fit to the sequence preference of the enzyme. Beside quality analysis, also quantitative measurements have to be done to assess potential biological relevance of a methylation event.

Today, different approaches exist for the identification of novel PKMT substrates. Beside protein interaction analysis and biochemical purification, substrate specificity analysis of PKMTs is often used for a proteome wide screening of NHT of specific PKMTs. Independent which method was used, the critical part is always the validation of the methylation event *in vitro* and *in vivo*. Only with this confirmation relevant biological functions can be systematically connected to an identified methylation event and further improve the development of therapeutic drugs (Kudithipudi & Jeltsch 2016; Rathert, Dhayalan, Ma, et al. 2008). One main topic of this thesis was the investigation of novel non-histone targets of different histone methyltransferases, and it was also demonstrated that some assignments of PKMT substrates in the literature are incorrect.

1.10. Somatic cancer mutations in PKMTs

Over the past decades, sequence technologies have been massively improved and current next generating sequencing (NGS) allow analyses of whole exomes or even genomes at reasonable costs. In genome wide sequencing studies of cancer tissues, several somatic cancer mutations in PKMTs were discovered and proposed to have a central role in carcinogenesis (Watson et al. 2013). The “Catalogue Of Somatic Mutations In Cancer” (COSMIC) database has listed

all known somatic cancer mutations in various tumors and also provides information about the localization of mutations in the genome and in different cancers (Forbes et al. 2011). In general, somatic cancer mutations in epigenetic factors can be categorized in two functional groups, loss-of-function and gain-of-function mutations. Chromosomal translocations, which lead to alterations in the expression level of specific genes or recruit enzymes to aberrant genomic loci, represent a well-known mechanism for gain-of-function mutations. Non-sense or frameshift mutations in the coding region of genes induce loss of the protein function (Kudithipudi & Jeltsch 2014). However, some somatic cancer mutations cannot be strictly assigned in one of these two groups. Missense mutations are point mutations of single amino acids and can have various effects, dependent on the site of the mutation and the resulting amino acid substitution. In Figure 6A, possible effects of missense mutations in PKMTs are summarized. The disruption of AdoMet binding is one possible mechanism to inactivate an enzyme, as well as the disruption of the protein fold. Otherwise, point mutations can influence the interaction with other proteins, which may affect the regulation and targeting of PKMTs. Additionally, missense mutations can change the peptide recognition motif, due to that new substrates can be recognized, but probably the original substrate cannot be methylated anymore. Finally, the methylation level of the product can be altered (Kudithipudi & Jeltsch 2014). In B-cell lymphoma cell lines and follicular lymphoma samples, somatic cancer mutations of EZH2, the catalytically active component of the polycomb repressive complex 2 (PRC2), were found, which provide a well understood example of the latter mechanism. Wildtype EZH2 catalyzes mono- and dimethylation of H3K27 using unmethylated substrate as target. Mutation of Y641 changes the methylation level to H3K27me3 and also alters the substrate preference. Tyrosine 641 is located in the aromatic cage of the SET domain and formation of a hydrogen bond with the ε-amino group of lysine prevents trimethylation of H3K27. An amino acid exchange of Y641 leads to increased H3K27me3 levels and silencing of tumor suppressor genes (Morin et al. 2010; Sneeringer et al. 2010; Yap et al. 2011). Dependent on previous roles of a PKMT in regulating oncogenes or tumor suppressor genes, loss-of-function or gain-of-function mutations of PKMTs can reverse their function and play a fundamental role in tumor growth (Figure 6B) (Kudithipudi & Jeltsch 2014).

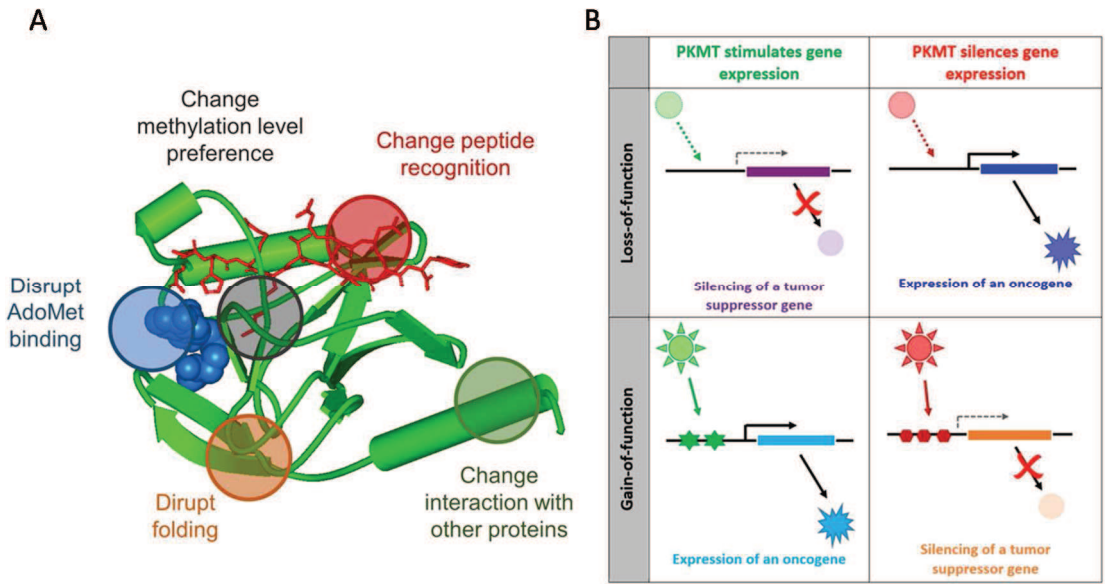


Figure 6: Somatic cancer mutations in PKMTs can have different biological effects. A) Summary of possible effects of missense mutations. B) Loss-of-function or gain-of-function mutations of PKMTs can reverse original functions of oncogenes or tumor suppressor genes. Figures are taken from (Kudithipudi & Jeltsch 2014).

For the understanding of the role of somatic cancer mutations in PKMTs, it is necessary to study the *in vitro* and *in vivo* effects of each mutation in detail. While loss-of-function mutations often can be easily explained on the background of the gene function, the understanding of the effect of gain-of-function mutations is more complex, because each mutation can have independent biochemical functions as described above. Hence, each mutant may also need specific therapeutic treatment. During carcinogenesis, aberrant histone methylation can be prevented, using PKMT inhibitors as a common cancer treatment strategy, but this approach cannot be used for the treatment of inactive PKMT mutants. Therefore, future biomedical research needs individual biochemical and cellular investigation of each somatic cancer mutation of PKMTs to know the consequence of somatic mutations in PKMTs and to focus on the establishment of individual cancer treatment.

1.11. Histone H4 lysine 20 methylation

Methylation of lysine 20 of histone H4 is correlated with transcriptional repression. H4K20 can be mono-, di- and trimethylated. Each methylation state has distinct biological roles; trimethylation of H4K20 leads to heterochromatin formation and gene silencing, monomethylation plays a role in cell cycle progression, genomic stability, DNA replication and DNA damage response (Jørgensen et al. 2013; Beck et al. 2012). In *Schizosaccharomyces pombe* one single methyltransferase is responsible for setting all three methylation states of H4K20. In human cells several enzymes have been reported to introduce the different methylation states of H4K20. The three most important PKMTs are SET8, catalyzing H4K20

monomethylation, SUV4-20H1 and SUV4-20H2 catalyzing H4K20 di- and trimethylation (Beck et al. 2012; Balakrishnan & Milavetz 2010). Lack of SET8 caused a complete loss of H4K20 methylation, including H4K20 di- and trimethylation indicating that SUV4-20H1/H2 use H4K20me1 as preferred substrate for methylation. Studies with double knock-out of SUV4-20H1 and H2 have shown a decrease in H4K20me2/me3 together with an increasing amount of H4K20me1, which suggests that SET8 can only introduce one single methyl group and it is not able to catalyze further methylation states (Schotta et al. 2008). *In vitro* methyltransferase activity analyses of SUV4-20H1 and SUV4-20H2 have shown that both enzymes transfer one methyl group to monomethylated H4K20 peptide substrates (Southall et al. 2014; Wu et al. 2013), but results differed in details. Wu et al. (2013) observed a ~3 fold preference for monomethylated H4K20 substrate, whereas Southall et al. (2014) reported that SUV4-20H enzymes prefer H4K20me1 by about 10 fold. In both studies dimethylation of monomethylated target substrate was detected as well, but no trimethylation was observed. In contrast, H4K20 trimethylation was shown with nucleosomes as substrate (Schotta et al. 2004). Knock-out studies of SUV4-20H1 and SUV4-20H2 in mice and zebrafish have shown massive loss of H4K20me2 and me3 indicating that the SUV4-20H enzymes are responsible for introducing both *in vivo* (Kuo et al. 2012; Schotta et al. 2008; Southall et al. 2014). Schotta et al. (2004) suggest a model where H4K20 trimethylation at pericentric heterochromatin is dependent on H3K9 trimethylation by Suv3-9 methyltransferase due to interaction with HP1 isoforms (Schotta et al. 2004; Southall et al. 2014).

Expression studies have shown that SUV4-20H1 is ubiquitously expressed during embryogenesis and is also found in all adult tissues, whereas SUV4-20H2 is only detectable in some adult tissues and expression during development is low. Knock-out experiments have shown that SUV4-20H1$^{-/-}$ mice have developmental defects and die due to perinatal lethality. In comparison to that, SUV4-20H2$^{-/-}$ mice have no apparent problems during development, although they have decreased H4K20me3 levels at pericentric chromatin (Schotta et al. 2008). These results show that the two paralogs of H4K20 methyltransferases, present in mammalian cells, have non-redundant functions (Schotta et al. 2008).

The expression level of SET8 is fluctuating during cell cycle progression. In the G1 phase the expression level of SET8 is decreased reaching a very low level at the beginning of the S phase. During S and G2 phase the SET8 level increases until a peak is reached in the M phase. In the same way, the H4K20 monomethylation mark oscillates during cell cycle progression, as cells entering the S phase start to incorporate new histone proteins without

H4K20 methylation. With increasing SET8 levels until G2, H4K20me1 is introduced, which remains stable during mitosis and is converted to H4K20me2/me3 in the G1 phase of the cell cycle (Beck et al. 2012). Loss of SET8 leads to embryonic lethality in mice, due to increased DNA damage and dysregulated cell cycle progression (Oda et al. 2009).

1.11.1. Catalytical activity and specificity of SET8

Similar to other histone lysine methyltransferases, SET8 also contains a catalytic SET domain and by resolving the crystal structure of SET8 in complex with AdoHcy and peptide substrate the monomethyltransferase activity could be explained. It was shown that two tyrosine residues (Tyr245 and Tyr334) located in the binding pocket of SET8 form direct or water-mediated hydrogen bond interactions with ε-amino group of lysine 20, respectively. Hydrogen bonds between the hydroxyl group of Tyr334, the carbonyl groups of Gly294 and Ile297 and the ε-amine group of K20me1 tightly interact with a water molecule. This strong binding of the water molecule hinders the K20me1 group from rotation in the cofactor binding site preventing further methylation reactions (Del Rizzo & Trievel 2014; Couture et al. 2005). Mutational analysis revealed that exchange of tyrosine at position 334 to phenylalanine enables SET8 to dimethylate H4K20 (Couture et al. 2005).

In agreement with the crystal structure of SET8, a substrate specificity analysis has indicated that SET8 has a long recognition motif ranging from R17 to R23 (R^{17}-H^{18}-(R^{19}KY)-K^{20}-(V^{21}ILFY)-(L^{22}FY)-R^{23}) and amino acid exchanges at R17 and H18 caused a complete loss of methyltransferase activity. Similar as other histone lysine methyltransferases, SET8 also methylate non-histone proteins. The methylation of p53 at K382 was discovered as the very first non-histone target of SET8. Methylated p53 is bound by L3MBTL1 (Lethal 3 malignant brain tumor 1) via its MBT domains, repress the expression of p53 target genes and finally leads to chromatin compaction (West et al. 2010). Furthermore, it was reported that SET8 dimethylates the Numb protein at K158 and K163. Numb was identified in *Drosophila* as a cell-fate determinant during development that interacts with p53 in the nucleus. Methylation of Numb was reported to disrupt its interaction with p53, which was described to result in increased p53 ubiquitination and reduced apoptosis (Dhami et al. 2013). Sequence comparisons of the residues flanking the reported target lysine residues of Numb indicated that these methylation sites differ from the SET8 requirements. In addition, SET8 was reported to act as a strict monomethyltransferase and dimethylation of lysine residues by SET8 was not identified before, which stands in contrast to the reported dimethylation activity of SET8 on Numb. Therefore, in this thesis I further investigated Numb methylation by SET8.

1.11.2. Catalytical activity and specificity of SUV4-20H1 and SUV4-20H2

The catalytic domains of SUV4-20H1 and H2 contain an N-terminal helical domain, the SET domain and Zn-binding postSET domain. Even if the overall fold of the catalytic SET domain is similar to other PKMTs, a sequence comparison of the SET domain indicates only 8-22% identity to other SET enzymes. However, the catalytic domain of SUV4-20H1/H2 harbor a sequence identity of 65%. The discovery of a unique mechanism of SUV4-20H enzymes, to catalyze dimethylation using monomethylated substrate as target, was quite interesting. As described above, methyltransferases known to catalyze monomethylation of lysines residues contain one tyrosine residue in the canonical SET domain that interacts with the ε-amino group of the target lysine. In case of SUV4-20H proteins, a serine residue (S251 for SUV4-20H1 and S161 for SUV4-20H2) forms a hydrogen bond between its hydroxyl side chain and the target lysine Nε (instead of Y245 in SET8). Another tyrosine (Y334 in SET8) is replaced by a cysteine in both enzymes (Cys305 in SUV4-20H1 and Cys215 in SUV4-20H2). The transfer of only one methyl group can be explained by the hydrogen bond formed by Ser251/161 and the ε-amino group of the target lysine which may hinder the transfer of a second methyl group, because it can prevent rotation of the amino group. Additionally, the methyl group of H4K20me1 is bound in a well-defined hydrophobic pocket formed by Ile204/Ile181 and Phe214/Phe191 in SUV4-20H1/H2, where unmethylated H4K20 cannot bind, which explains the preference for a monomethylated substrate (Figure 7).SET8 and SUV4-20H enzymes have differences in their enzymatic properties and substrate binding. Comparison of the substrate binding site and the active pocket of SET8 and SUV4-20H enzymes imply that even if they recognize the same substrate target H4K20, different interaction between the adjacent amino acids are present (Wu et al. 2013; Southall et al. 2014). This suggests that there might be also differences regarding the sequence specificity between SUV4-20H1 and SUV4-20H2, which has to be further investigated. It also indicates that both paralogs might recognize different non-histone targets which leads to new insights into unclear biological processes. Further analyses are necessary to discriminate the substrate specificity of SUV4-20H1 and H2, and to search for new non-histone targets, which has been part of this thesis.

Recently, the extracellular signal-regulated kinase 1 (ERK1) was reported to be trimethylated by SUV4-20H1 at two different lysine residues: K302 and K361. This methylation was connected to increased phosphorylation and transcription, followed by enhanced ERK1 signaling cascade finally resulting in tumorigenesis (Vougiouklakis et al. 2015). However, trimethylation activity of SUV4-20H1 using an unmethylated lysine as substrate is in

disagreement with the published data about the substrate specificity and product pattern of SUV4-20H1 described above (Southall et al. 2014). Also the sequences surrounding the predicted target lysines are not matching with the substrate specificity of both enzymes determined in this thesis. Therefore, a further analysis of ERK1 methylation by SUV4-20H enzymes was conducted as part of this thesis.

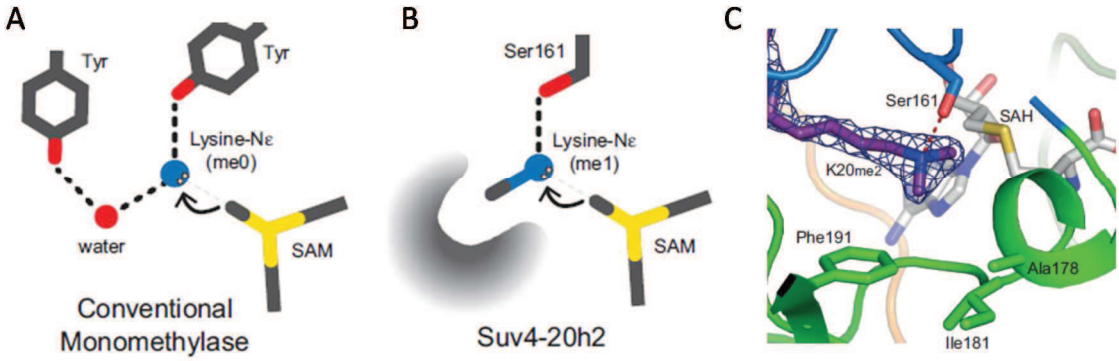


Figure 7: A) In conventional monomethylases (SET8) two tyrosine residues in the binding pocket form direct or water-mediated hydrogen bonds with the ε-amino group of lysine 20, respectively. B) In contrast to SET8, SUV4-20H enzymes have a serine residue that forms a hydrogen bond between its hydroxyl side chain and the target lysine Nε. C) A well-defined hydrophobic pocket formed by isoleucine and phenylalanine residues of SUV4-20H enzymes prefer monomethylated H4K20 so unmethylated H4K20. Figures are taken from (Southall et al. 2014).

1.12. The MLL histone lysine methyltransferase family

The mixed-lineage leukemia (MLL) family of histone lysine methyltransferases regulates the expression of homeotic and developmental genes (Smith et al. 2011; Ansari et al. 2009). In yeast, Set1 is the single histone lysine methyltransferase catalyzing all mono-, di- and trimethylation of lysine 4 on histone H3. In *Drosophila melanogaster*, three H3K4 methyltransferases: SET domain-containing 1 (Set1), Trithorax (Trx) and Trithorax-related (Trr) are present. In human cells each of the three H3K4 *Drosophila melanogaster* methyltransferases is represented by two homologs, so that the mammalian MLL protein family consists of six members in total: MLL1/MLL2 (related to Trx), MLL3/MLL4 (related to Trr) and SET1A/SET1B (related to Set1) (Piunti & Shilatifard 2016). Additionally MLL5 exists, but so far no enzymatic function could be discovered (Mas-y-Mas et al. 2016).

Each MLL protein has a specific domain organization. As shown in Figure 8A, all MLL proteins contain the catalytic SET domain, followed by a postSET domain at the C-terminus. MLL1-4 have several additional domains, like plant homeodomains (PHD) or bromodomains (Bromo), which facilitate modified histone interaction. Additionally, all four MLL proteins contain an FYRN-FYRC (Phe/Tyr residues N-terminal - Phe/Tyr residues C-terminal) domain. Crystal structures of MLL1-4 indicate that these parts fold as one single domain and

stay together in cells (Smith et al. 2011). During maturation, MLL1 and MLL2 are cleaved into a C- and N-terminal fragment by the protease Taspase 1. The Taspase cleavage site is located in a gap of 1595 amino acids in MLL1 and 631 amino acids in MLL2 between the FYRN and FYRC domains. In contrast, in MLL3 and MLL4, the FYRN and FYRC domains are located next to each other. At the N-terminal end, MLL1 and MLL2 carry a CxxC domain that has been shown to bind to unmethylated CpG DNA sites. Moreover AT-hooks are present in MLL1/2, which are involved in the localization to specific target gene loci (Macrini et al. 2003; Cierpicki et al. 2010). In contrast to the other enzymes of this family, the SET proteins (Set1, SET1A and SET1B) have only one or two RNA recognition motifs (RRM) located at the N-terminus.

Cellular studies indicated that the MLL proteins have non redundant functions, as mice containing MLL-deletion and -truncation constructs showing different phenotypes (Ansari et al. 2009), because they introduce different methylation states on H3K4 at distinct genomic regions. MLL1 and MLL2 catalyze trimethylation of H3K4 sites located at gene promotors, especially of Hox genes, whereas SET1A and SET1B proteins introduce genome-wide H3K4 trimethylation in mammals. MLL3 and MLL4 mainly introduce monomethylation marks on H3K4 sites at enhancers and promotors of target genes (Cheng et al. 2014).

Within all MLL family members, MLL1 is the most studied one, because it is involved in chromosomal translocations often found in patients with acute myeloid and lymphoid leukemia. In about 75% of MLL1 associated leukemia, the N-terminal part of MLL1 is genetically fused with 60 different partner genes, mostly AF4, AF9, AF10 and ENL. One part of the fusion protein is the DNA binding domain of MLL1, which delegates them to MLL1 specific target genes. The catalytic domain of MLL1 is exchanged by the fusion partners leading to an altered target gene regulation at the target loci. In addition, it was shown that some of the fusion partners recruit Dot1L, a histone methyltransferase that trimethylates H3K79 and unregulates the expression of Hox genes. The recruitment of Dot1L is linked to the chromosomal translocation of MLL1 with AF4, AF9, AF10 and ENL. The corresponding mistargeting of Dot1L deregulates the expression of HoxA9 and Meis1 genes ultimatively causing abnormal embryonic development (Anglin & Song 2013; Daser & Rabbitts 2005; Krivtsov & Armstrong 2007; Muntean & Hess 2012). Furthermore, MLL proteins are frequently mutated in cancer, which can lead to loss-of-function or gain-of-function. MLL3 is one of the most frequently mutated PKMT present in glioblastoma, melanoma, pancreatic and breast cancer (Kandoth et al. 2013; Kudithipudi & Jeltsch 2014). A detailed biochemical

investigation of effects of somatic cancer mutations located in the SET domain of MLL1 and MLL3 on the enzymatic properties was performed in this thesis.

1.12.1. MLL methyltransferases are associated with multi protein complexes

The MLL histone lysine methyltransferase family associates with a large protein complex, which is necessary for the optimal methyltransferase activity of each enzyme. In yeast, Set1 associates with a complex of proteins called COMPASS for "Complex of Proteins Associated with Set1", also found as COMPASS-like complexes in *D. melanogaster* and mammalian cells. All COMPASS complexes contain WDR5 (Tryptophan-aspartate repeat protein-5), RBBP5 (Retinoblastoma-binding protein-5), ASH2L (Absent small homeotic-2-like) and DPY30 (Dumpy-30), which are homologs of Cps30, Cps20, Cps60 and Cps25 in yeast, respectively. Additionally, each MLL COMPASS-like subfamily has specific additional complex partners. For example, the SET1A/SET1B COMPASS-like complex contains Wdr82, related to Cps35 in yeast, which was shown to cross-talk with H2B monoubiquitination and helps to install H3K4 trimethylation. Menin is specific for MLL1/2-COMPASS-like subfamily. It is a tumor suppressor that helps to localize to Hox genes, necessary for gene expression during development. The H3K27 demethylase UTX is a specific component for the MLL3/4 COMPASS-like structure. H3K27 methylation and H3K4 methylation are known to have antagonistic roles. Methylation of lysine 4 on histone H3 leads to gene activation, while methylation on lysine 27 on the same histone tail has silencing effects. Demethylation of H3K27 by UTX further helps to maintain the activation mark of H3K4 introduced by MLL3/MLL4 COMPASS-like structure (Figure 8B) (Smith et al. 2011; Piunti & Shilatifard 2016).

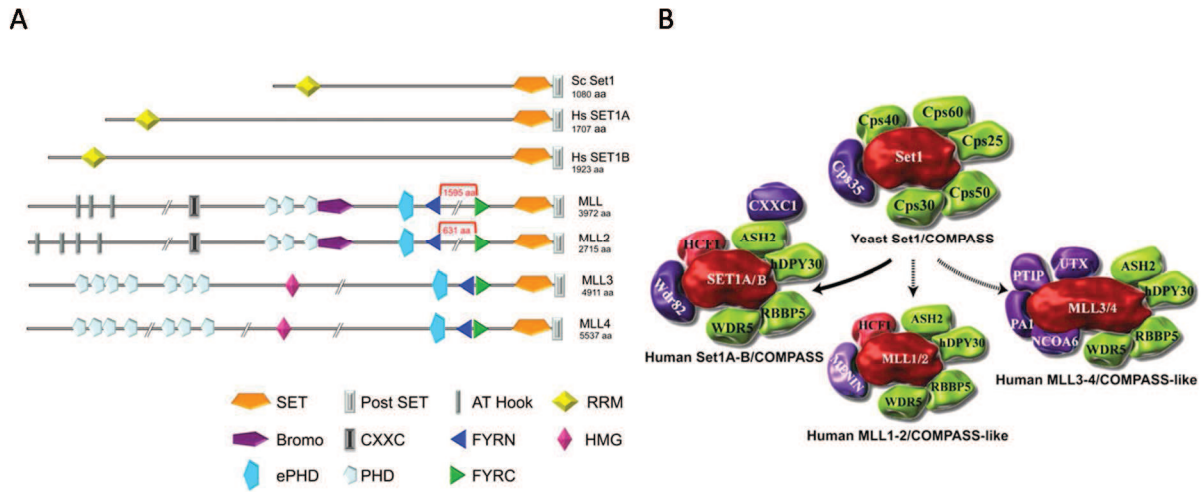


Figure 8: Domain organization of yeast Set1 and mammalian MLL family proteins. A) All presented proteins contain the catalytical SET domain, followed by the postSET domain at the C-terminus. Yeast Set1 (Sc Set1) and human SET1A/B (Hs SET1A/B) have a RNA recognition (RRM) motif at the N terminus. MLL1-4 has several protein-protein interaction domains like plant homeodomain (PHD) finger and bromodomain (Bromo). The FYRC and FYRN domains are also present in MLL1 to 4. In MLL1/2 these two domains are separated by a gap, whereas they are located next to each other in MLL3/4. Additionally, MLL1 and MLL2 contain a CxxC domain that has been shown to recruit to unmethylated CpG DNA sites and AT Hooks that helps to localize to specific target gene loci. B) Set1, the single H3K4 methyltransferase in yeast *Saccharomyces cerevisiae*, associates with a complex called COMPASS. Mammalian cells contain six homologous proteins viz. MLL1-4 and SET1A/B, which are all found associated with COMPASS-like homologous complex. All have in common the complex proteins WDR5 (Tryptophan-aspartate repeat protein-5), RBBP5 (Retinoblastoma-binding protein-5), ASH2L (Absent small homeotic-2-like) and DPY30 (Dumpy-30). Each MLL COMPASS-like subfamily has specific complex partners, like Wdr82 in SET1A/B, MENIN in MLL1/2 and UTX in MLL3/4. Figures are taken from (Smith et al. 2011).

1.12.2. Structural rearrangement of the COMPASS-like complexes

The crystal structure of MLL1 in complex with AdoHcy and a histone peptide substrate has shown that the active pocket of MLL1 has not the optimal conformation necessary for efficient methyl transfer from the cofactor AdoMet to the target lysine (Southall et al. 2009). Like other SET domain containing proteins, the active pocket of MLL1 consists of the preSET domain, SET domain (including SET-N, SET-I and SET-C) and the C-terminal postSET domain. Also, the essential residues of the active site are conserved and have an almost identical arrangement compared to other SET domain containing proteins. However, the orientation of the SET-I helix is different from highly active histone lysine methyltransferases like Dim5, which leads to a more open conformation. Southall et al. suggested a model in which the open conformation could close after binding of the complex partners WDR5, RBBP5 and ASH2L (WRA). The complex partners reorient the SET-I region and, thereby, a complete lysine binding channel is formed, where the target lysine chain is placed at a defined position for efficient methyl transfer (Southall et al. 2009). In 2016, Li and colleagues were able to resolve the crystal structure of MLL3 together with a minimized

structure of the RBBP5/ASH2L (RA) heterodimer and provided a two-step activation mechanism for MLL proteins (Li et al. 2016). In this study, measurements of the structural dynamics showed that the MLL proteins are highly dynamic without complex partners. By the addition of complex partners, a more stable catalytic competent conformation was achieved described as the first step for activation. In the second step, binding of the target substrate results in a rearrangement of the SET-I region that completes the hydrophobic tunnel (Avdic et al. 2011; Li et al. 2016; Southall et al. 2009).

1.12.3. Interaction between MLL and RBBP5/ASH2L heterodimer

The crystal structure of MLL together with complex proteins has led to a much better understanding of the regulation of the methyl transfer process in MLL proteins. It was shown that MLL proteins primarily interact with RA heterodimer. When MLL is bound to the RA heterodimer, together with the cofactor product AdoHcy and H3 peptide, RBBP5 adopts an extended conformation, with a rigid coil conformation in the ASH2L binding motif (ABM) and two β-strands in the activation segment (AS) (Figure 9A). The interface of ASH2L and RBBP5 contains two arginine residues (R343 and R367) of ASH2L interacting via hydrogen bonds with Asp353, Glu349 and Glu347 of RBBP5. A strong interaction between RBBP5 and MLL3 is achieved through intermolecular β-sheet interactions between two β-strands located in the L-shaped AS domain of RBBP5 and β4 and β7 of MLL3. In addition, a conserved arginine residue located in the SET-I region of MLL proteins (e.g. R3864 of MLL1) plays an important role in the complex assembly, as it fits perfectly into an acidic pocket formed by RBBP5 and ASH2L, which is stabilized by the formation of five salt-bridges and a network of hydrogen bonds. Sequence alignments have shown that the amino acid residues responsible for the binding with complex partners are conserved within the MLL protein family, indicating that all MLL proteins have the same electrostatic network to form the MLL-RA assembly. Based on a mutational analysis, Li and colleagues could show that RA heterodimer interacts with all MLL proteins identically and the activation mechanism is conserved within the MLL family.

1.12.4. Interaction between MLL and WDR5

WDR5 is a scaffolding protein that serves as bridging molecule and stabilizes the interaction between the catalytic MLL enzymes and the RBBP5/ASH2L heterodimer. Knockdown of WDR5 in mammalian cells resulted in a global loss of H3K4me3. It also led to a decreased expression of Hox genes and problems during development and hematopoiesis (Wysocka et al. 2005). The crystal structure of WDR5 indicates that it folds as a seven-bladed tryptophan-

aspartate (WD) repeat β-propeller and it binds MLL1 and RBBP5 at two well defined surface pockets, located at the opposite site of its β-propeller domain (Figure 9B) (Avdic et al. 2011). The MLL-WDR5-RBBP5 complex is stabilized by a network of hydrogen bonds and van der Waals interactions. RBBP5 interacts with WDR5 through three conserved amino acids valine-aspartate-valine, known as VDV motif. This motif adopts a kink conformation and binds into a V-shaped cleft formed by two β-strands of WDR5. On the opposite side, WDR5 has an arginine binding cavity that interacts with a highly conserved WDR5-interacting-motif (WIN), located in the N-SET region of MLL proteins. The interaction between MLL1 and WDR5 is important for the core complex assembly and optimal methyltransferase activity. As already mentioned, MLL1 is often found as MLL1-AF9 fusion protein in ALL and AML patients. For treatment of leukemogenesis, scientists developed inhibitors, mimicking the WIN motif of MLL1. These inhibitors disrupt the interaction between MLL1 and WDR5 and by this decrease its catalytic activity. The inhibitor MM-102, was shown to inhibit the expression of HoxA9 and Meis-1. Further, it inhibits cell growth in leukemia cells carrying MLL1 fusion proteins (Karatas et al. 2013). MM-102 was also used in this thesis to study the inhibitory effect on somatic cancer mutations located in the SET domain of MLL1.

1.12.5. Differences between MLL1 and the other MLL family proteins

Previous studies have shown that MLL1 behave differently than other MLL proteins, regarding its methyltransferase activity (Li et al. 2016). Unlike all other MLL proteins, MLL1 is dependent on WDR5 to exhibit full methyltransferase activity. The methyltransferase activity of MLL1 is increased by the addition of RA heterodimer, but the addition of WDR5 further stimulates it, finally leading to the optimal methyltransferase activity. Compared to that, all other members of the MLL family are fully activated by the addition of RA alone and WDR5 is not necessary for further stimulation. Two amino acids (Asn3861 and Gln3867) located in the binding region of MLL1 to RA are necessary for these differences. MLL1 contains a glutamine residue at position 3867, instead of a valine in the case of MLL3 that cannot fit into the well-defined pocket formed by MLL and RBBP5. At position 3861, MLL1 contains asparagine, which cannot interact with the hydrophobic residues Leu339, Val343 and Tyr345 of RBBP5 surrounding a threonine at position 4803 in MLL3. This sequence differences weaken the interaction of MLL1 with the RA heterodimer such that WDR5 is needed as bridging factor to further stabilize the interaction.

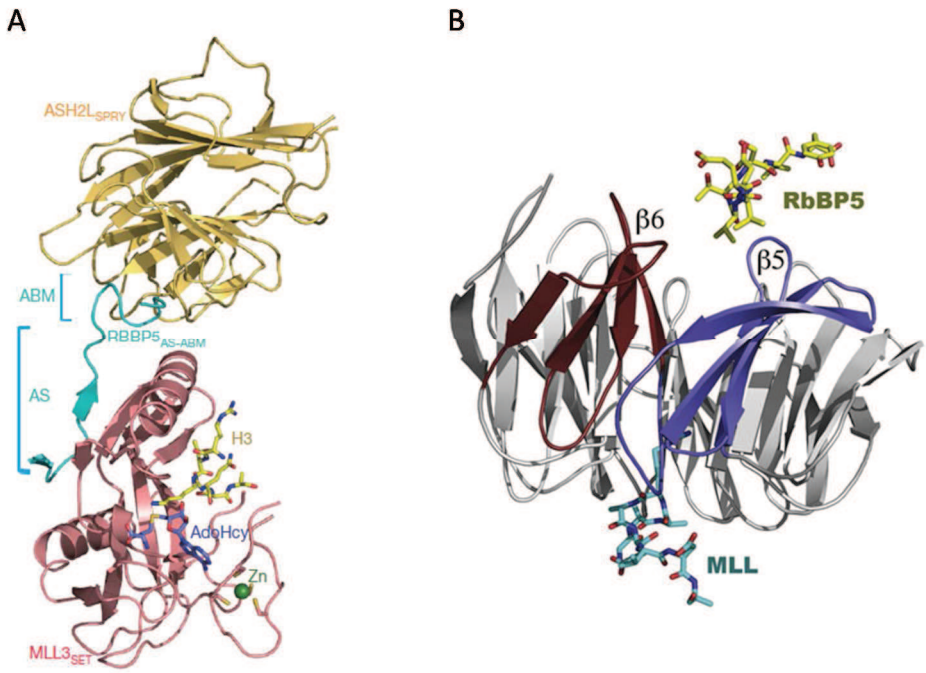


Figure 9: Complex conformation of MLL-RA. A) Crystal structure of MLL3 SET domain (red) in complex with RBBP5 (blue) and ASH2L (yellow). RBBP5 represents two β-strands located in the activation segment (AS) and a rigid coil located in the ASH2L binding motif (ABM). The crystal structure reveals the SPRY (SPIa and Ryanodine Receptor) domain located in the N terminus of ASH2L. B) WDR5 binds MLL and WDR5 at two distinct binding pockets, located at opposite sides of its β-propeller domain. Two β-blades (β5 and β6), which are necessary for the interaction between WDR5 and RBBP5 are highlighted. Carbon atoms of $MLL1_{WIN}$ peptide colored in cyan and RBBP5 $_{371-380}$ colored in yellow. Figures are taken from (Avdic et al. 2011; Li et al. 2016).

2. Aims of the study

It was one general aim of this PhD thesis to investigate the substrate specificity of the SUV4-20H1, SUV4-20H2, MLL1 and MLL3 PKMTs in detail. Depending on the results, it was planned to use the specificity data for the identification of novel non-histone substrates of all four enzymes. During the course of the work two examples of unexpected assignments of PKMT and methylated protein were published in which the methylation site did not fit to the specificity profile of the PKMT. It was one aim of this thesis to reinvestigate methylation of these substrates. Over the past decade, somatic mutations in PKMTs were discovered by genome-wide sequencing studies of cancer tissues. To check the role of these mutations in carcinogenesis, it was planned to investigate the effects of them on the enzymatic properties of the MLL proteins in detail.

2.1. Specific goals and achievements of the project

2.1.1. Substrate specificity analysis of the SUV4-20H1 and SUV4-20H2 enzymes

Previous studies raised questions regarding the product and substrate specificity of SUV4-20H1 and SUV4-20H2. A ~10 fold preference for monomethylated H4K20 substrate was reported by Southall et al. (2014), whereas a ~3 fold preference for H4K20me1 was observed by Wu et al. (2013). Both researchers identified that SUV4-20H enzymes only transfer one methyl group on H4K20me1 (Southall et al. 2014; Wu et al. 2013). Contrary, H4K20 trimethylation was observed, when nucleosomes were used as substrates (Schotta et al. 2004). In this study, it was planned to perform peptide array methylation experiments and methylation of purified peptides in solution to solve this discrepancy. To get more insights into the different functions of these two paralogs, substrate specificity profile arrays of both enzymes should be analyzed and further used for the identification of potential non-histone targets in a proteome wide search.

2.1.2. Identification of wrong assignments between PKMT and methylated target site

In the course of this work, two examples of potential mis-assignments of protein methylation sites and PKMTs were published. ERK1 was reported to be trimethylated at K302 and K361 by SUV4-20H1 (Vougiouklakis et al. 2015). This result was surprising, because it was reported that SUV4-20H1 uses an unmethylated target substrate to introduce trimethylation, which is not comparable with our results on histone H4 peptide substrates. In addition, the surrounding sequences of both target ERK1 lysines are not matching with the specificity profile of SUV4-20H1. The aim of this project was to verify methylation of ERK1 by SUV4-20H enzymes at peptide and protein level using purified SUV4-20H enzymes.

Dhami et al. (2013) reported that SET8 dimethylates Numb at K158 and K163 (Dhami et al. 2013). Also this finding was surprising, because it is known that SET8 is a strict monomethyltransferase. Comparison of the reported methylation sites of Numb with the substrate specificity of SET8 showed that the surrounding sequences of the predicted target lysine residues are not fitting. It was one aim of this project to confirm methylation of Numb by SET8 purified from *E.coli* and HEK293 cells at peptide and protein level.

2.1.3. Substrate specificity analysis of MLL1 and MLL3

The MLL protein family consists of 6 members: MLL1-4, SET1A and SET1B. They can be further classified in 3 subfamilies dependent on different enzymatic functions. MLL1/2 catalyze trimethylation of H3K4 at promotors of specific genes, e.g. Hox genes. MLL3/4 are monomethyltransferases of H3K4 located at enhancers elements and SET1A/B catalyze genome-wide trimethylation of H3K4. All MLL proteins are associated with a multi protein complex to obtain full methyltransferase activity. One aim of this study was to determine the detailed substrate specificity of MLL1 and MLL3 in absence or presence of complex proteins to find out if the complex partners influence the specificity of MLL enzymes. Based on the resulting recognition motif, it was intended to identify possible non-histone substrates in the human proteome.

2.1.4. Influence of somatic cancer mutations on the enzymatic properties of MLL1 and MLL3

Genome-wide sequencing studies of cancer tissues uncovered several somatic mutations in PKMTs. Somatic cancer mutations can have various effects on the enzymatic properties of PKMTs including loss-of-function or gain-of-function phenotypes. In this project, it was planned to investigate the functional consequences of selected somatic cancer mutations in MLL1 and MLL3, using detailed biochemical analyses and by this improve our understanding of the role of these mutations and the corresponding PKMTs in carcinogenesis in general.

3. Materials and Methods

This section describes only methods which have not been covered in manuscripts 1-4, provided in the attachment of this thesis.

3.1. Cloning, Expression and Purification of Proteins

The DNA sequences encoding the non-histone proteins were amplified from cDNA derived from HEK293 cells and cloned into the corresponding restriction sites of the pGEX-6p2 vector (GE Healthcare) as GST fusion proteins. Detailed information about the non-histone substrates are listed in Table 1 and 2 in the results part.

The C3882S mutation of MLL1 protein variants and the K to R mutations of TICRR and ZNF862 were introduced using a megaprimer PCR mutagenesis method. Mutagenesis was confirmed by restriction site digestion and followed by DNA sequencing.

For protein expression *E.coli* BL21 Codon + cells (Novagen) were transformed with the corresponding plasmid and grown in Luria-Bertani media at 37°C until they reached an optical density at 600 nm of 0.6 to 0.8. Then, the cells were transferred to 20°C for 10 min and induced overnight by addition of 1 mM iso-propyl-beta-D-thiogalactopyranoside to the growth medium. The cells were harvested by centrifugation (5000 g).

The GST-fusion proteins were purified as described (Dhayalan et al. 2011).

3.2. Synthesis of peptide SPOT arrays

Peptide arrays were synthesized on cellulose membrane using the SPOT synthesis method (Frank 2002) with a Autospot Multipep system (Intavis AG) by Dr. Srikanth Kudithipudi. Each spot contained approximately 9 nmol peptide (Autospot Reference Handbook, Intavis AG) and the successful synthesis of the peptides on the cellulose membrane was qualitatively confirmed by bromophenol blue staining. Data analysis and derivation of sequence motif and discrimination factor calculation was performed as described previously (Kudithipudi, Kusevic, et al. 2014; Rathert, Zhang, et al. 2008).

3.3. Peptide array methylation

The peptide arrays were pre-incubated in methylation buffer (50 mM Tris/HCl pH 8, 200 mM NaCl, 5 mM $MgCl_2$ and 3 mM DTT) for 5 min at 25°C. The arrays were then incubated for 60 min at 25°C in methylation buffer containing 840 nM MLL1 SET protein or 50 nM MLL3-SET protein and 0.76 µM radioactive labeled AdoMet (Perkin Elmer). Afterwards, the

membranes were washed 5 times for 5 minutes with wash buffer (100 mM NH_4HCO_3 and 1% SDS) and finally incubated with Amplify NAMP100V solution (GE Healthcare) for 5 minutes. The peptide arrays were exposed to Hyperfilm^{TM} high performance autoradiography films (GE Healthcare) in the dark at -80°C for 1-8 days. The autoradiography films were developed in Optimus TR developing machine.

3.4. Protein methylation reaction

Protein methylation was performed by incubating recombinant H3.1 (New England Biolabs) or non-histone protein substrates with MLL protein variants in the presence or absence of equimolar amounts of complex partners WDR5, RBBP5 and ASH2L in methylation buffer containing 50 mM Tris/HCl pH 8, 200 mM NaCl, 5 mM $MgCl_2$ and 3 mM DTT, supplemented with 0.76 µM labeled [methyl-H3]-AdoMet (Perkin Elmer) for 2 h at room temperature. Afterwards, the methylation reactions were stopped by heating to 95°C for 5 min and the samples were separated on a 16% SDS PAGE gel. Then, the gels were dried, after soaked with Amplify NAMP100V solution (GE Healthcare) for 5 minutes. Subsequently, the gels were exposed to Hyperfilm^{TM} high performance autoradiography films (GE Healthcare) in the dark at -80°C for 1-8 days. The autoradiography films were developed in Optimus TR developing machine.

4. Results

Lysine methylation is an important posttranslational modification, which is introduced by protein lysine methyltransferases. It affects many important processes of proteins including protein-protein interaction, protein localization and protein degradation, which are involved in the regulation of many biological processes. Over the last decades, many PKMTs and also protein lysine methylation sites were identified in the genomes and proteomes of many organisms in a very short time. While initially histones were considered the main targets of protein lysine methylation, more recently non-histone proteins were detected as possible substrates of methyltransferases. The understanding of the biological outcome of lysine methylation events is very complex. On each lysine residue up to three methyl groups can be transferred and, dependent on the methylation state and the methylated target site, lysine methylation leads to different effects, like transcriptional activation or repression if histone methylation is considered.

To advance our understanding of the biological role of protein lysine methylation, it is a great challenge to connect a specific target site with the responsible methyltransferase. A good approach to tackle this challenge is to analyze the substrate specificity of PKMTs. This can be achieved by methylation of substrate specificity arrays, which allow the methylation of many different peptide substrates in competition within one experiment. This method provides detailed information about the specificity of an enzyme which is helpful for the identification of novel PKMT substrates in the human proteome. However, with increasing global interest in protein lysine methylation also wrong assignments between enzymes and target substrates were published, which could be sorted out with better knowledge of the substrate sequence preference of PKMTs.

PKMTs also play a fundamental role in carcinogenesis. Expression changes of PKMTs were detected in cancers and in several cases somatic cancer mutations were identified in PKMTs. Since each of these mutations can have specific effects on the enzymatic properties, a detailed experimental examination of the effects of mutations in PKMTs is required for the better understanding of the role of these enzymes in cancer and the development of an effective therapeutic strategy.

In this work, the substrate specificity of different protein lysine methyltransferases was investigated. Two examples of unexpected assignments of methylation events to PKMTs were studied and could be shown to represent wrong assignments. Furthermore, somatic cancer

mutations of MLL1 and MLL3 were identified and mutant specific effects on the enzymatic properties of these enzymes were described in detail.

4.1. Specificity of the SUV4-20H1 and SUV4-20H2 protein lysine methyltransferases and methylation of novel substrates

(Manuscript 1 in the attachment of this thesis)

Weirich S, Kudithipudi S, Jeltsch A. (2016*). Journal of Molecular Biology, 2016 Jun 5;428(11):2344-58. doi: 10.1016/j.jmb.2016.04.015. Epub 2016 Apr 20.*

In human cells, two SUV4-20 paralogs, SUV4-20H1 and SUV4-20H2, catalyze the methylation of H4 at K20 and play an important role in heterochromatin formation. Both enzymes have overlapping and distinct biological effects. In this study, critical enzymatic properties of SUV4-20H1 and SUV4-20H2 (sometimes collectively abbreviated as SUV4-20H) were investigated to get more insights into their different cellular functions. Previous published data showed slight disagreements with respect to the substrate and product specificity of the SUV4-20H enzymes. As described in 1.11., *in vitro* methyltransferase activity analyses of SUV4-20H1 and SUV4-20H2 have shown that both enzymes transfer one methyl group to monomethylated H4K20 peptide substrates (Southall et al. 2014; Wu et al. 2013), but results differed in details. While Wu et al. (2013) observed a ~3 fold preference for monomethylated H4K20 substrate, Southall et al. (2014) reported that SUV4-20H enzymes prefer H4K20me1 by about 10 fold. In order to resolve these discrepancies, mouse SUV4-20H1 and SUV4-20H2 were cloned as GST-fusion proteins and purified by affinity chromatography (manuscript 1, Figure 1a). Methylation of peptide arrays containing different modified forms of H4K20 peptide substrates indicated that both enzymes strongly prefer H4K20me1 as substrate (manuscript 1, Figure 1b). Methylation assays with soluble peptides showed a 10-fold higher preference of SUV4-20H1 for the monomethylated H4K20 peptide when compared with the unmethylated H4K20 peptide substrate. Similarly, SUV4-20H2 had a 28-fold higher activity on the H4K20me1 peptide than on the unmethylated H4K20 peptide (manuscript 1, Figure 1c). Further analysis of peptide methylation by SUV4-20H1/H2 using MALDI mass spectrometry failed to detect methyl transfer to the unmethylated peptide substrate. Methylation of the H4K20me1 substrate by SUV4-20H1 and H2 resulted in the appearance of a second peak with 14 kDa difference, which revealed the transfer of a single methyl group to the H4K20me1 peptide. Higher methylation states were not introduced by SUV4-20H enzymes even after long incubation times (manuscript 1, Figure 2). These peptide

methylation experiments imply that SUV4-20H1 and SUV4-20H2 introduce one single methyl group on the monomethylated H4K20 peptide substrate *in vitro*.

Even if SUV4-20H1 and SUV4-20H2 have the same product and substrate specificity with respect to the methylation state of the target lysine, different biological functions can be explained by sequence specific interactions of each enzyme with the substrate peptide. Substrate specificity profile analysis of both paralogs were performed to analyze their specific interaction with the substrate peptide using the monomethylated H4 (13-27) sequence as template (Figure 10). The substrate specificity peptide arrays were methylated by SUV4-20H1 or SUV4-20H2 using radioactively labeled AdoMet. The methyl transfer to the immobilized peptide was detected by autoradiography. As all the peptides are methylated in competition, the observed methylation signal directly reflects the preference of SUV4-20H enzymes towards a specific peptide sequence. The data revealed that SUV4-20H1 prefers a (RY)-Kme1-(IVLM)-(LFI)-X-D sequence motif, while SUV4-20H2 recognizes a slightly different X-Kme1-(IVLMK)-(LVFI)-X-(DEV) sequence motif (manuscript 1, Table 1). Overall the specificity profile analysis demonstrates that the SUV4-20H enzymes interact with the same residues of H4 tail, but SUV4-20H1 is more specific and accepts only a smaller number of alternative residues at the corresponding positions than SUV4-20H2. Compared to SUV4-20H2, SUV4-20H1 is more specific at the -1 and +4 positions of the substrate (considering the target lysine as position 0). At position -1, SUV4-20H1 is only active with the canonical arginine, whereas in the case of SUV4-20H2 arginine can be exchanged by several amino acids at this position. At position +4, SUV4-20H1 only recognizes aspartic acid, whereas SUV4-20H2 also accepts E and V (manuscript 1, Figure 3 and 4). These deviations suggest that there might be differences between both enzymes regarding the recognition of non-histone targets.

Based on the more relaxed substrate recognition motif of SUV4-20H2, which also includes all potential substrates of SUV4-20H1, the Phosphosite PLUS database was used to search for potential novel methylation substrates of these enzymes, because it also contains several protein lysine methylation events identified in proteomics studies. 7 novel substrates were identified, which were known to be methylated within the SUV4-20H core sequence recognition motif by a not yet identified PKMT (manuscript 1, Table 2). Peptide array methylation experiments showed that one protein (CASZ1) can be methylated by SUV4-20H1 and three proteins (CASZ1, OIP5 and CENPU) by SUV4-20H2 (manuscript 1, Figure 5 and Suppl. Figure S1).

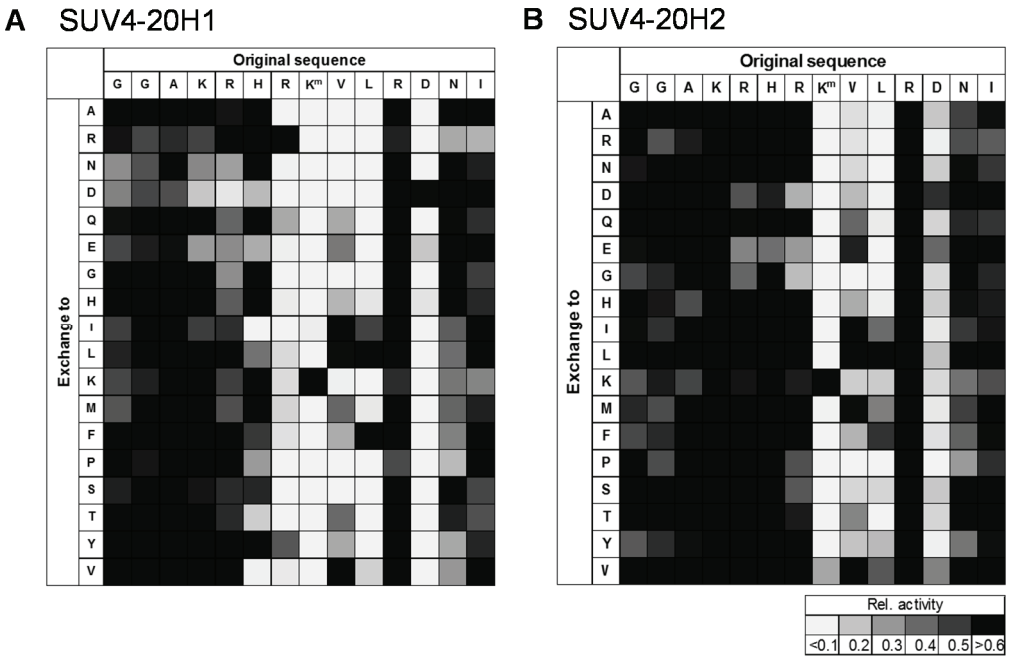


Figure 10: Substrate specificity analysis of SUV4-20H1 and SUV4-20H2. A+B) Three independent peptide array methylation experiments were conducted with SUV4-20H1 or SUV4-20H2 and the data were averaged after normalizing the full activity to 1. The activity is displayed in the grey scale as indicated. The horizontal axis represents the template sequence (H4 13-27) and the vertical axis indicates the residues that were sequentially introduced at the position corresponding to the row. The target lysine H4K20 was monomethylated (K^m). (Taken from manuscript 1, Figure 3b and Figure 4b)

Vougiouklakis et al. (2015) reported recently, that ERK1 is methylated by SUV4-20H1 at K302 and K361 (Vougiouklakis et al. 2015). As described in 1.11.2., the surrounding sequences of both lysine residues are not fitting to the identified substrate recognition motif of SUV4-20H enzymes. To proof methylation of ERK1 by SUV4-20H enzymes, methylation reactions at peptide and protein level were performed. Methylation experiments on peptide arrays containing different methylation states of the predicted target lysine as well as K to A mutants as negative controls showed no methylation signal (manuscript 1, Figure 6). Also at the protein level, methylation of GST fused ERK1 or isolated ERK1 by the SUV4-20H enzymes could not be detected (Figure 11). These findings suggest that the reported methylation of ERK1 by SUV4-20H1 is a wrong assessment of PKMT and target. This finding was important, because methylated ERK1 was reported to play a central role during carcinogenesis and its connection to SUV4-20H enzymes could lead to pointless follow up studies and wrong therapeutic approaches.

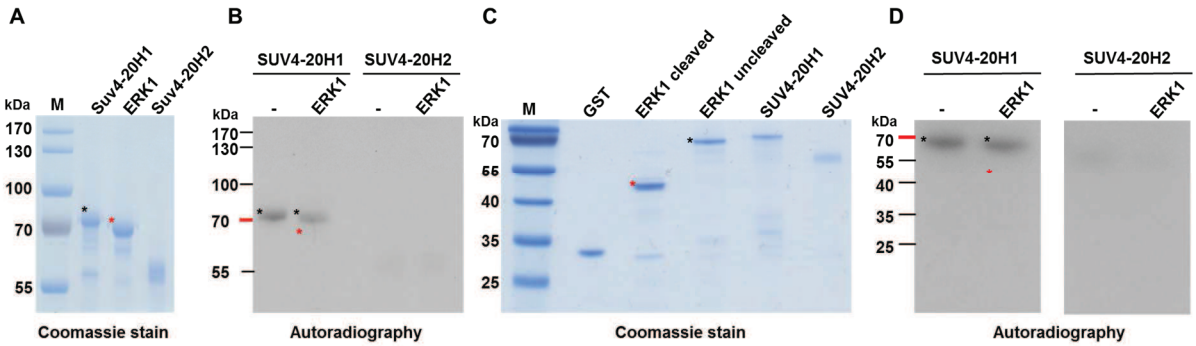


Figure 11: *In vitro* protein methylation of ERK1 by SUV4-20H enzymes. A) Coomassie stained SDS-PAGE gel showing the proteins used in the methylation assay shown in panel B. SUV4-20H1 and GST fused ERK1 protein are labeled with black and red asterisks. B) SUV4-20H proteins were incubated with radioactively labeled AdoMet in the methylation buffer in the absence and presence of GST-ERK1 and methylation was analyzed by autoradiography. Sizes of proteins are indicated as described in A. C) Coomassie stained SDS-PAGE of GST cleaved ERK1 protein (indicated with the red asterisk), GST and uncleaved ERK1 (indicated with the black asterisk) were loaded as controls. SUV4-20H proteins were also loaded to indicate the protein sizes. D) SUV4-20H proteins were incubated with radioactively labeled AdoMet in the methylation buffer in the absence and presence of cleaved ERK1 and methylation was analyzed by autoradiography. The sizes of SUV4-20H1 and the cleaved ERK1 protein are indicated with black and red asterisks. A longer film exposure of this gel is shown in manuscript 1 Suppl. Fig. 2. (Taken from manuscript 1, Figure 7)

4.2. Investigation of the methylation of Numb by the SET8 protein lysine methyltransferase

(Manuscript 2 in the attachment of this thesis)

Weirich S, Kusevic D, Kudithipudi S & Jeltsch A. (2015). *Scientific Reports*, *2015 Sep 22;5:13813. doi: 10.1038/srep13813.*

Compared to SUV4-20H1 and SUV4-20H2, SET8 is a highly specific H4K20 monomethyltransferase with a long recognition motif extending from R17 to R23 on H4 [R^{17}-H^{18}-(R^{19}KY)-K^{20}-(V^{21}ILFY)-(L^{22}FY)-R^{23}] in which it tolerates only few alternative residues at the positions -1, +1 and +2 (considering the target lysine as position 0) (Figure 12A boxed) (Kudithipudi et al. 2012). Like other protein lysine methyltransferases, SET8 was also found to methylate non-histone targets and the tumor suppressor protein p53 was the first identified non-histone target of SET8. Recently it has been reported, that SET8 dimethylates Numb protein at lysine 158 and 163 (Dhami et al. 2013). Due to discrepancies at least at 4 out of 6 positions between the sequence recognition motif of SET8 and the surrounding sequences of the predicted target lysines (K158 and K163) in Numb (Figure 12A), a further investigation of Numb methylation by SET8 was performed in this thesis.

Peptide methylation experiments using radioactively labeled AdoMet showed that SET8 cannot methylate the predicted target lysine residues K158 and K163, because only very weak

radioactive signals were observed and lysine to alanine mutations of these reported target sites did not lead to a loss of the radioactive signal (manuscript 2, Figure 2A). To further analyze, if the weak observed signals were due to binding of the SET8-AdoMet complex to these peptides or if another amino acid residue was methylated, mutational scanning peptide arrays of the Numb (151-165) and (156-170) peptides were prepared, in which each individual peptide residue was exchanged to alanine or serine if the original residue was alanine. Incubation of these arrays with SET8 indicated that SET8 probably catalyzes a weak cysteine methylation (manuscript 2, Figure 2B), a side reaction that was already reported for other PKMTs like MLL1 (Patel et al. 2014). To further investigate Numb methylation by SET8 at protein level, the Numb protein domain containing the reported methylation targets (residues 12-272) was overexpressed as GST fusion protein in *E.coli* and purified by affinity chromatography. *In vitro* methylation assays were performed and clear methylation signals were detected with the positive controls H4 and p53, but no methylation of Numb was observed (Figure 12 B/C). We then noticed that for their *in vitro* methylation reactions, Dhami et al. used recombinant SET8 protein purified from human cells by immunoprecipitation. As SET8 purified from human cells might contain posttranslational modifications or interactors influencing its activity and specificity might be copurified, methylation reactions were repeated using recombinant YFP-SET8 purified from HEK293 cells after transient overexpression. However, even using YFP-SET8 preparations, no methylation of Numb was detected (manuscript 2, Figure 4). In summary, in this study I could not observe methylation of Numb peptide or Numb protein by SET8 overexpressed and purified from *E.coli* or HEK293 cells. Therefore, more evidence is needed to support the claim by Dhami et al. and the reported downstream results (like the reported influence of Numb methylation on its interaction with p53 and subsequent effects on apoptosis) should be taken with care.

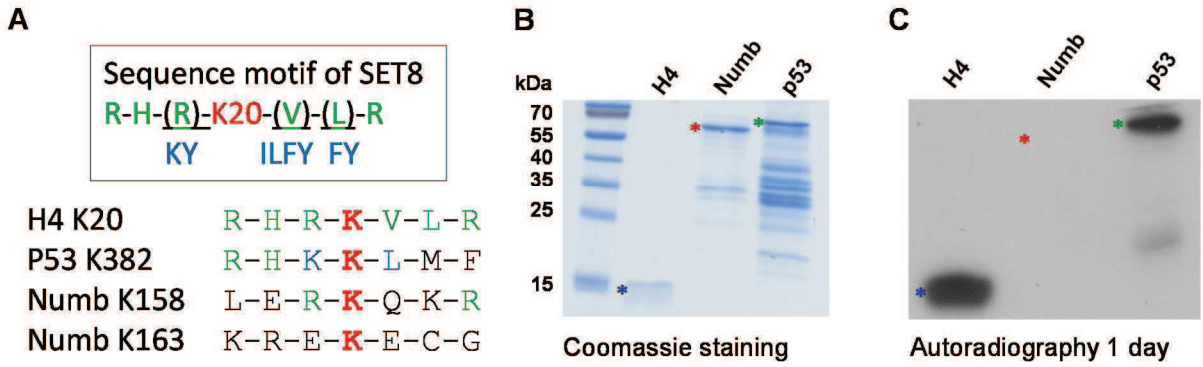


Figure 12: Sequence comparison and protein methylation of histone H4, p53 and Numb. A) The substrate specificity profile of SET8 was adopted from Kudithipudi et al. (2012) (boxed). The upper row represents the amino acid sequence of the H4 tail surrounding K20 (printed in green), the lower row specifies the other amino acids that are also accepted at the corresponding position (printed in blue). Below, sequence alignment of the H4, p53 and Numb target lysine residues (Taken from manuscript 2, Figure 1) B) Coomassie BB stained SDS polyacrylamide gel of the purified GST-Numb, GST-p53 and recombinant H4 used as methylation substrates. The corresponding bands of H4, GST-Numb and GST-p53 are labeled by blue, red and green asterisks. (Taken from manuscript 2, Figure 3A) C) Methylation of GST-Numb, GST-p53 and H4 with recombinant SET8. The positions of the corresponding protein bands are indicated as described in panel B. Methylation signals from H4 and GST-p53 were detected, whereas no methylation signal was visible for GST-Numb. (Taken from manuscript 2, Figure 3B)

4.3. Studies with MLL protein lysine methyltransferase family members

The mammalian mixed-lineage leukemia (MLL) protein family consists of 6 members: MLL1-4, SET1A and SET1B (Piunti & Shilatifard 2016). Three pairs of MLL proteins form three subfamilies representing the homologs of the three H3K4 *Drosophila melanogaster* methyltransferases SET1, TRX and TRR. All MLL proteins catalyze the methylation of H3K4 and thereby play an important role in early development and hematopoiesis (Ansari et al. 2009; Smith et al. 2011). Previous studies have shown that the different MLL PKMTs have non-overlapping functions. MLL1/2 introduce H3K4 trimethylation at promotors of active genes, MLL3/4 catalyze H3K4 monomethylation at active enhancers and SET1A/B introduce genome-wide H3K4 trimethylation (Cheng et al. 2014). Each MLL protein has its specific domain architecture, whereas the MLL subfamilies are similar constructed (see Figure 8A in the introduction of this thesis). In cells, the MLL proteins are part of a large protein complex also containing the WDR5 (Tryptophan-aspartate repeat protein-5), RBBP5 (Retinoblastoma-binding protein-5), ASH2L (Absent small homeotic-2-like) and DPY30 (Dumpy-30) core members, which are needed to achieve optimal methyltransferase activity (Avdic et al. 2011; Li et al. 2016). Li et al. reported that the MLL proteins primarily interact with the RA heterodimer and WDR5 serves as a bridging molecule between MLL and RA. DPY30 is another component, which is common for all MLL complexes. However, previous studies have shown that it only interacts with ASH2L and increases the overall reaction rate only by

2-fold compared to MLL-WRA (Patel et al. 2009). In this study two MLL proteins, MLL1 and MLL3, which are members of different subgroups of the MLL proteins, were selected and further analyzed. My main interest was to investigate the individual sequence specificities of both paralogs, as it was already shown that they catalyze H3K4 methylation at distinct genomic regions. Moreover, the specificity profiles of MLL1 and MLL3 were used to search for additional possible substrates of both enzymes.

In addition, somatic cancer mutations of MLL proteins have been found in several cancer tissues. In this thesis, identified somatic cancer mutations located in the SET domains of MLL1 and MLL3 were experimentally investigated regarding their effects on the enzymatic activity in presence or absence of the WRA complex partners. Differences in the substrate specificity of MLL1 cancer mutants were studied on MLL1 specific non-histone targets. Finally, potential effects of somatic cancer mutations on the product methylation pattern were further investigated for MLL1 and MLL3 in this thesis.

4.3.1. Investigation of the MLL1 protein lysine methyltransferase

4.3.1.1. Somatic cancer mutations in MLL1 induce conformational changes and modulate the enzymatic activity

(Manuscript 3 in the attachment of this thesis)
Weirich S, Kudithipudi S & Jeltsch A. (2017). *Molecular Oncology. 2017 Apr;11(4):373-387. doi: 10.1002/1878-0261.12041. Epub 2017 Mar 10.*

The protein lysine methyltransferase MLL1 (KMT2A) is a large, 3969 amino acids long protein which catalyzes trimethylation of H3K4 at promotors of active genes (Cheng et al. 2014; Smith et al. 2011). Within the MLL protein family, MLL1 is one of the most studied members, because it has been found to be frequently involved in chromosomal translocations in acute lymphoid and myelogenous leukemia, which lead to dysregulation of developmental genes like HoxA9 and Meis1 (Anglin & Song 2013; Daser & Rabbitts 2005; Krivtsov & Armstrong 2007; Muntean & Hess 2012). Beside chromosomal translocation several somatic cancer mutations have been identified in MLL1, which have not been studied yet (Kudithipudi & Jeltsch 2014). Using the COSMIC database (August 2016), 665 mutations were reported in MLL1, out of them 78% missense and 22% non-sense mutations and frame-shifts. To better understand the effect of somatic cancer mutations, four mutations in the SET domain were selected and studied here in detail, which are located close to the peptide, AdoMet or complex partner interaction site and, therefore, could have direct effects on

enzymatic properties of MLL1: R3841W (found in prostate cancer) (Barbieri et al. 2012), R3864C (found in lung cancer) (Muzny et al. 2012), S3865F (found in skin cancer) (Durinck et al. 2011) and R3903H (found in large intestine cancer) (Muzny et al. 2012) (manuscript 3, Figure 1). The catalytic SET domains of MLL1 wildtype and all four MLL1 cancer variants were cloned as GST-fusion proteins and purified via affinity chromatography (manuscript 3, Figure 2A). Circular dichroism spectroscopy (CD) measurements of the four purified MLL1 cancer variants showed that all mutant proteins have similar fold like wildtype, except R3841W (manuscript 3, Figure 2B). Changes in the CD spectra of R3841W could be due to changes in conformation, folding or aggregation state. To assess the catalytic activity of the purified proteins, methylation experiments were performed. Two mutants (R3864C and R3841W) had an increased activity in comparison to wildtype MLL1, whereas S3865F showed a reduced methyltransferase activity and R3903H was inactive (manuscript 3, Figure 2C). To test the stimulatory effect of the WRA complex members on the methyltransferase activity of MLL1, peptide methylation assays were performed. As expected, a strong stimulation of the methyltransferase activity of MLL1 was detected in presence of the WRA complex. Further methylation experiments were implemented to analyze the catalytic properties of all 4 cancer mutants in presence of equimolar amounts of complex partners (manuscript 3, Figure 2D). The results demonstrated that R3903H was inactive also in the presence of the WRA complex proteins. MLL1 wildtype and S3865F showed higher methyltransferase activity in presence of complex partners, but the overall catalytic activity of S3865F was lower. Interestingly, two of the mutants, R3864C and R3841W, displayed a reduced activity in the presence of complex partners, which represents a pronounced change in properties when compared to wildtype MLL1 (Figure 13). This result indicates that both enzymes are not dependent on complex partners to achieve their full catalytic activity, as they were already fully active in isolated form.

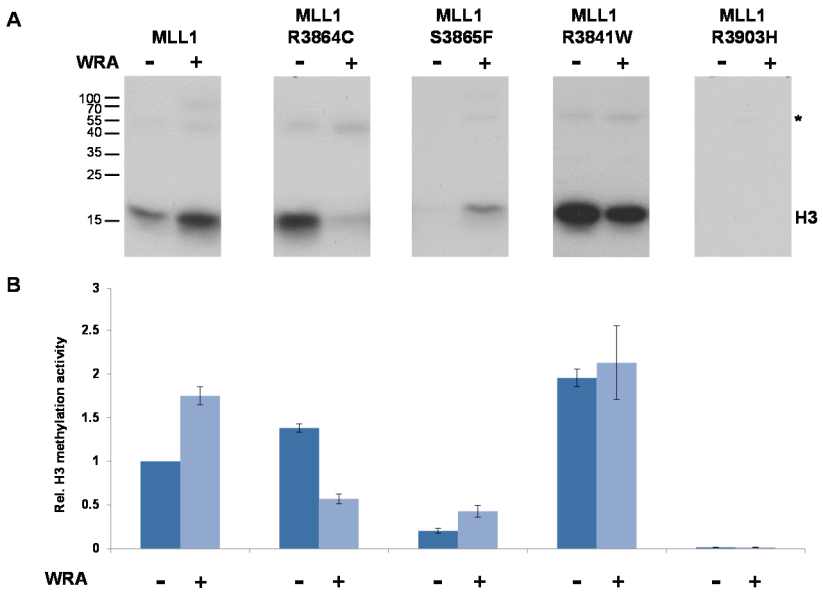


Figure 13: Methylation activity of MLL1 cancer variants in complex with the WRA proteins. Recombinant H3 was methylated by MLL1 and mutant proteins in absence or presence of the WRA complex. A) Autoradiographic image of an SDS polyacrylamide gel. Samples with (+) or without (-) complex partners were loaded next to each other. The methylation signal of H3 is indicated. * represents automethylation of the MLL1 mutant proteins. B) Quantitative analysis of the H3 methylation signal using duplicate experiments. The activity of isolated MLL1 was set to 1 and the other signals were normalized accordingly. Error bars indicate the standard error of the mean. (Taken from manuscript 3, Figure 3)

In this respect, the question arose if the different MLL1 variants were interacting properly with the complex partners. While MLL2-4 are mainly dependent on the interaction with RA heterodimer to be fully active (Li et al. 2016; Cao et al. 2014; Southall et al. 2009), MLL1 requires WDR5 as a bridging molecule to interact with RA and achieve full methyltransferase activity (Li et al. 2016). Interaction between MLL1 and WDR5 is based on the WIN motif, located in the N-SET region of MLL proteins (Patel et al. 2009). The WIN motif contains an arginine residue that is conserved within the MLL protein family. Previous studies showed that this arginine is crucial for the complex assembly as exchange of R3765 against alanine abolishes the interaction between MLL1 and WDR5 (Patel et al. 2008). Small molecular inhibitors mimicking this WIN sequence (MM-102 inhibitor) were investigated to disrupt the MLL1-WDR5 interaction. These inhibitors can be used as a new therapeutic approach to reduce MLL1 hyperactivity in leukemia (Cao et al. 2014; Karatas et al. 2013; Li et al. 2016). The WDR5 interaction was confirmed for all 4 somatic cancer mutations with GST pull-down experiments, using the MM-102 inhibitor to disrupt persisting interactions as additional control (manuscript 3, Figure 4A). This result was expected as the mutation did not include residues in the WIN motif. In the next step, AlphaScreen assay measurements were used to investigate the interaction between MLL1 and the RA or WRA complexes. For all MLL1 variants similar AlphaScreen signals were obtained, indicating that all MLL1 proteins interact

similarly with the RA heterodimer. The AlphaScreen signals were increased with the WRA complex, in agreement with the fact that WDR5 further stabilizes the interaction between RA and all MLL1 protein variants. These findings demonstrate similar interactions of all MLL somatic cancer mutations with the RA heterodimer (manuscript 3, Figure 4B). Altogether, these interaction studies confirmed that the mutations have no effect on the complex formation and altered complex formation cannot be the reason for changes in the catalytic activity of the MLL1 mutants.

To further analyze the effects of each complex member on the MLL1 activity, methylation experiments were carried out using all individual, binary and ternary combination of complex partners together with the MLL1 proteins. Main differences were detected when the RA heterodimer was added to the MLL1 protein variants. The MLL1 wildtype methyltransferase activity was stimulated by the addition of RA heterodimer and further stimulation was achieved by the addition of WDR5. In contrast to that, addition of the RA heterodimer to the R3864C mutant led to a strong inhibition of the methyltransferase activity. When WDR5 was added to the R3864C-RA complex an increase in activity was detected, but the final activity of the R3864C-WRA complex was still lower than that of isolated R3864C. In the case of R3841W, the addition of RBBP5, ASH2L and RA heterodimer caused reductions in the methyltransferase activity, but addition of WDR5 to the R3841W-RA complex increased the activity to the same level like isolated R3841W (manuscript 3, Figure 5).

Finally, the inhibitory effect of MM-102 on the methyltransferase activity of the different MLL1 mutants was investigated. Inhibition was detected in the case of MLL1 wildtype and S3865F, whereas no effect was observed with R3864C and R3841W (Figure 14). This result is in agreement with the previously described biochemical data, which demonstrate that the methyltransferase activity of MLL1 and S3865F is dependent on the interaction with WRA. In contrast, R3864C and R3841W exhibit full methyltransferase activity in absence of complex partners and because of this they are not affected by the inhibitory effect of MM-102 on the MLL1-WDR5 interaction. In summary, these results demonstrate that the selected somatic cancer mutations either increase or reduce the activity of MLL1. Also the dependency on complex partners to achieve full methyltransferase activity was changed. Moreover, these data illustrate MM-102 cannot be used as universal MLL1 inhibitor, because no effect was detected with R3864C and R3841W. More generally these data show that individual biochemical investigation for each mutant is necessary for the development of therapeutic drugs.

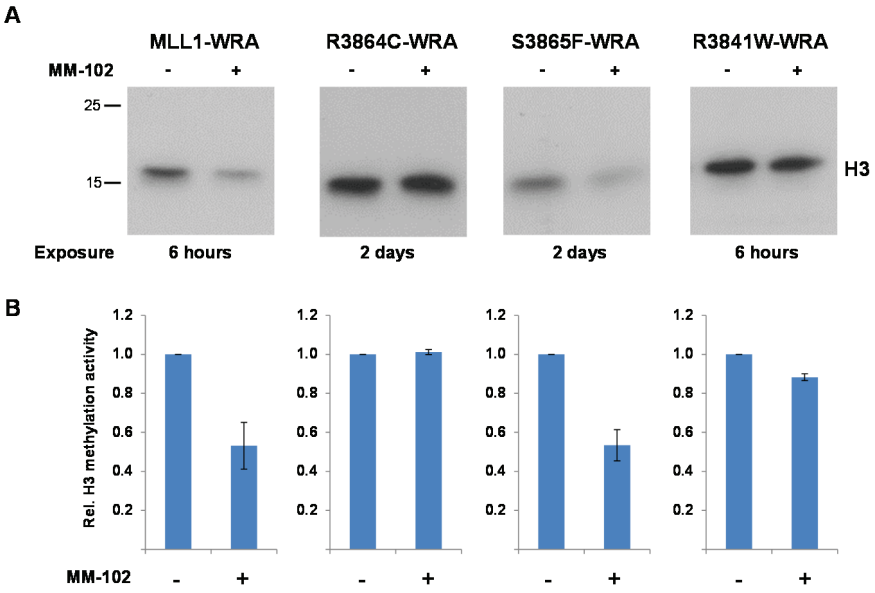


Figure 14: Inhibition of the MLL1 proteins by MM-102. A) Recombinant histone H3 was methylated by MLL1 cancer variants together with WRA complex in the presence and absence of inhibitor. The methylation signal of H3 is shown. Different exposure times are indicated. B) Quantitative analysis of H3 methylation signals using duplicates of experiments. For better visualization of the inhibitory effect by MM-102, the activities of the MLL1 mutants were normalized to the corresponding activity without inhibitor treatment. The error bars indicate the standard error of the mean. (Taken from manuscript 3, Figure 6)

4.3.1.2. Specificity analysis of MLL1

Lysine methylation is an important epigenetic mark with an essential role in many biological processes. With the identification of non-histone proteins as possible substrates for many lysine methyltransferases like SET8, SET7/9, SMYD3, G9a and NSD1/2, lysine methylation gained even more importance (Fu et al. 2016; Rathert, Dhayalan, Murakami, et al. 2008; Kudithipudi, Lungu, et al. 2014; Kudithipudi et al. 2012; Dhayalan et al. 2011). Up to date, many methylation sites have been identified in proteome-wide studies, but the responsible methyltransferases are often unknown. Conversely, for many protein lysine methyltransferases the complete list of all relevant substrates has not been identified yet. In this chapter, substrate specificity analysis was used to investigate the specific interaction of MLL1 with the substrate peptide and search for novel substrates for MLL1.

4.3.1.2.1. Substrate sequence specificity of MLL1

To investigate the substrate specificity of MLL1 in detail, a peptide library was synthesized on a cellulose membrane using the SPOT synthesis method and methylated by MLL1. The libraries consisted of 320 individual peptides, in which each peptide contains an exchange of a single amino acid of the H3 (1-15) sequence against any of the 20 natural amino acids. Due to that the specific influence of each amino acid at every position of the substrate peptide on enzymatic activity can be analyzed in detail.

Substrate specificity peptide arrays were methylated by MLL1 using radioactively labeled AdoMet as described in chapter 4.1. Two independent methylation experiments were normalized and average methylation levels were calculated together with the standard deviation (SD) for the methylation activity on each peptide (Figure 15A). The distribution of errors shows that the experiments were highly reproducible. 93% of the peptides showed an SD smaller than ± 20%, 11 peptides had an SD between 20-30% and 7 peptides between 30-40% (Figure 15B). The substrate specificity profile indicates that MLL1 discriminates residues of the H3 sequence from A1 to R8. High specificity was observed at position -2 and +4. At position -2, MLL1 only tolerates the canonical arginine, whereas at position +4 it weakly accepts cysteine or tyrosine, beside the canonical arginine. At -1 position, MLL1 mainly interacts with hydrophobic residues (I, L, F, W, Y). MLL1 has a high preference for isoleucine or valine and it weakly accepts alanine and cysteine at position +2, in addition to the canonical threonine. In summary, the following substrate recognition motif was identified: A(PRFWY)-R-T(ILFWY)-K-Q(RHMFWY)-T(IVAC)-X-R(CY)

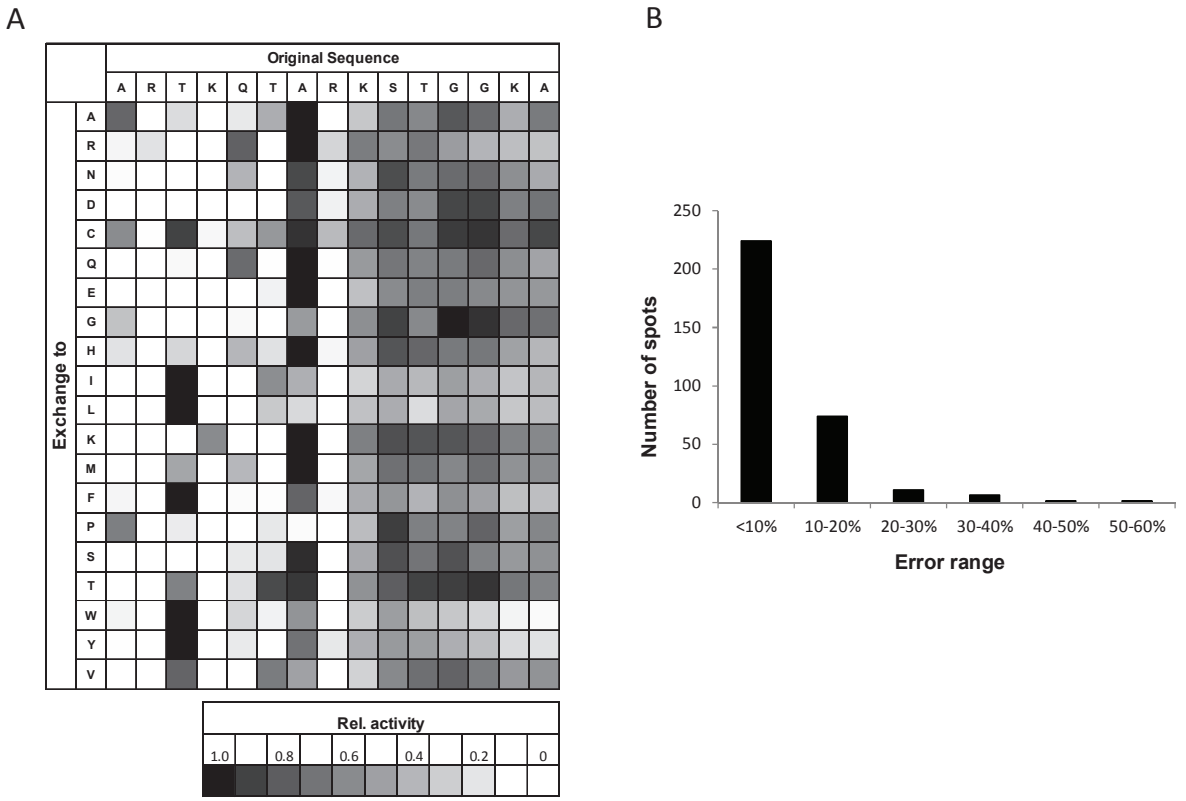


Figure 15: Substrate specificity analysis of MLL1. A) Two independent peptide array methylation experiments were performed with MLL1 and the data were averaged after normalizing the full activity to 1. The activity is displayed in a grey scale as indicated. The horizontal axis represents the template sequence (H3 1-15). Each residue was exchanged against all 20 natural amino acid residues as indicated by the vertical axis. B) Distribution of standard errors of the mean of MLL1 activity on all peptides tested in the two independent assays.

4.3.1.2.2. Methylation of potential non-histone substrates

The substrate specificity analysis showed that MLL1 accepts several other residues at many positions of the target sequence, in addition to the residues in the original sequence of H3. Because of that, the identified substrate recognition motif was used to search for potential novel non-histone targets in the human proteome using Scansite (Obenauer et al. 2003). 98 potential new peptide substrates containing the predicted sequence motif were found (Table 4, see appendix). To investigate if these targets can be methylated by MLL1, 15 amino acid long peptides with the target lysine at the center were synthesized on a cellulose membrane, including H3 (1-15) as positive control and H3K4A as negative control. The peptide array was incubated with MLL1 in a reaction mixture containing radioactively labeled AdoMet and the methyl transfer was detected by autoradiography. The data showed that 23 out of 98 putative non-histone targets were methylated on the peptide array (Figure 16 and Table 4).

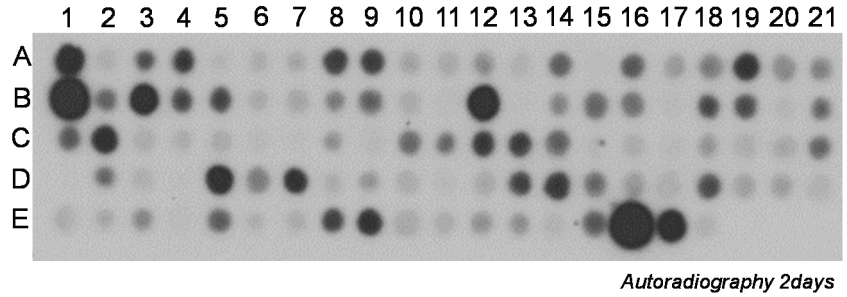


Figure 16: Screening of MLL1 non-histone peptide substrates. The peptide array was methylated with MLL1 in presence of radioactively labeled AdoMet. The transfer of the methyl group to the immobilized peptides was detected by autoradiography. Target information is listed in Table 4.

To further analyze the methylation of the validated peptide substrates at protein level, 15 of them were selected and cloned as GST fusion proteins (Table 1). 10 out of 15 targets were successfully cloned. Afterwards, these targets were overexpressed and purified by affinity chromatography. Out of 10 target protein domains, 8 could be successfully overexpressed and purified (Figure 17A).

Table 1: Selected non-histone targets from the peptide array (Figure 16), which were further analyzed.

	Swiss Prot. Nr.	Original name	Name	Synonyme	Sequence	Protein-length	MW (kDa)	Target K position
A3	Q9BR61	ACBD6_HUMAN	ACBD6 = Acyl-CoA-binding domain-containing protein 6	ACBD6	LYLYARYKQVKVGNC	282	31	73
A14	Q9H159	CAD19_HUMAN	CDH19 = Cadherin-19	CDH19	NELGPRFKRLACMFG	772	87	759
A16	A6NFT4	CC42B_HUMAN	CCDC42B = Coiled-coil domain-containing protein 42B	CCDC42B	ARLQRRLKRLEPCAR	308	36	138
A17	P22674	CCNO_HUMAN	CCNO= Cyclin-O	CCNO	EVHPPRVKQLLALCC	350	38	198
B1	Q7Z2Z1	CO042_HUMAN	TICRR= Treslin	TICRR	LRRSPRIKQLSFSRT	1910	211	1019
B2	O15263	DEFB2_HUMAN	DEFB4A= Beta-defensin 4A	DEFB4A	VFCPRRYKQIGTCGL	64	7	48
B3	Q6QHC5	DEGS2_HUMAN	DEGS2 = Sphingolipid delta(4)-desaturase/C4-hydroxylase DES2	DEGS2	LGPYARVKRVYRLAK	323	37	313
B12	O60258	FGF17_HUMAN	FGF17= Fibroblast growth factor 17	FGF17	SAPTRRTKRTRRPQP	216	25	207
B19	O00219	HAS3_HUMAN	HAS3= Hyaluronan synthase 3	HAS3	LSLGYRTKYTARSKC	553	63	332
C2	Q01113	IL9R_HUMAN	IL9R= Interleukin-9 receptor	IL9R	FKLSPRVKRIFYQNV	521	57	298
D5	P11086	PNMT_HUMAN	PNMT= Phenylethanolamine N-methyltransferase	PNMT	RQLRARVKRVLPIDV	282	31	152
D7	Q09MP3	R51A2_HUMAN	RAD51AP2= RAD51-associated protein 2	RAD51	LSRKARIKQLHPYLK	1159	134	1141
D18	Q9NYW4	TA2R5_HUMAN	TAS2R5= Taste receptor type 2 member 5	TAS2R5	IMGIPRVKQTCQKIL	299	35	280
E5	Q5VYS8	TUT7_HUMAN	ZCCHC6= Terminal uridylyltransferase 7	ZCCHC6	SAIDPRVKYLCYTMK	1495	171	1145
E16	O60290	ZN862_HUMAN	ZNF862= Zinc finger protein 862	ZNF862	GDGPRRIKRTYRPRS	1169	132	457

For methylation of the purified target protein domains, roughly equal amounts of each protein was incubated with the MLL1 SET domain in the presence of radioactively labeled AdoMet. As positive control recombinant H3 was included. To check for MLL1 automethylation, MLL1 was incubated in reaction buffer without including any additional substrate. After the methylation reaction, the samples were loaded on a 16% SDS gel and methyl transfer was detected by autoradiography. The result showed a strong automethylation signal of MLL1, because a methylation signal was observed at the size of the MLL1 SET domain even when no additional substrate was included in the methylation reaction. As many of the non-histone protein domains had nearly the same size as the MLL1 protein, the automethylation signal of MLL1 potentially overlapped with the substrate methylation signal making them more difficult to detect. For two substrate proteins, TICRR and ZNF862, clear methylation signals were observed. However for further analysis, the next step was trying to prevent or at least reduce the automethylation of MLL1 to increase the sensitivity of the detection of substrate protein methylation.

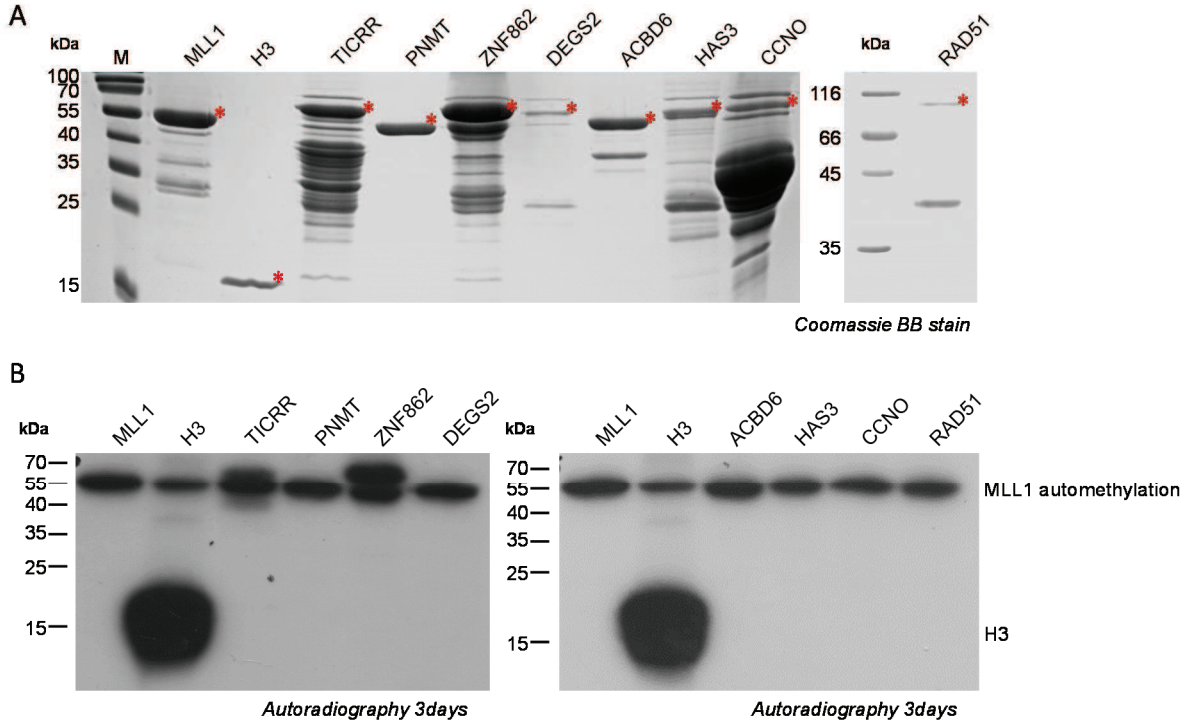


Figure 17: Methylation of MLL1 non-histone candidate substrate proteins. A) 16% SDS gel of MLL1 with all selected non-histone proteins (TICRR, PNMT, ZNF862, DEGS2, ACBD6, HAS3, CCNO and RAD51) serving as loading control of the methylation experiments. B) Methylation of all proteins by MLL1. The corresponding bands of the expected size are labeled with asterisk. The methylation of recombinant H3, automethylation of MLL1 and exposure times are indicated.

In 2014, Patel et al. identified that MLL1 undergoes automethylation at cysteine 3882 and they reported that an exchange of cysteine at position 3882 against serine prevents this reaction (Patel et al. 2014). Therefore, site directed mutagenesis was performed to create the MLL1 C3882S mutation in the GST-SET domain protein. In parallel, the C3882S mutation was also introduced into the MLL1 R3841W, S3865F and R3864C mutants. Successful mutagenesis was confirmed in all cases by sequencing. Afterwards, the MLL1 protein variants were overexpressed and purified by affinity chromatography (Figure 18A). To test automethylation, equal amounts of the original MLL1 wildtype and cancer mutants and the corresponding C3882S mutants were incubated with recombinant H3 in presence of radioactively labeled AdoMet. Methylated samples were separated via SDS-PAGE and transfer of the methyl group to the target lysine was analyzed by autoradiography.

In Figure 18A, the upper band at 55 kDa represents the automethylation of the MLL1 protein variants, whereas the methylation signal at 15 kDa corresponds to methylation of the target substrate H3. The cysteine 3882 to serine mutation resulted in a reduction of automethylation of MLL1, R3841W and S3865F. In the case of the R3864C/C3882S double mutant, no loss of automethylation was detected (Figure 18A). One possible reason for that could have been that the introduced cysteine at position 3864 could be a new point of automethylation in this

MLL1 mutant. For all MLL1 proteins methylation of recombinant histone H3 was detected. Methylation of H3 by MLL1 C3882S was weaker compared to the methylation of H3 by wildtype MLL1. A similar decrease in activity on H3 was observed for all MLL1 C3882S somatic cancer mutants. No methylation of H3 was detected with MLL1 S3865F/C3882S, but after long exposure of the autoradiogram, H3 methylation by this protein variant was confirmed (Figure 19B). As shown previously, the somatic cancer mutations MLL1 R3864C and MLL1 R3841W exhibited higher methyltransferase activity on H3, whereas MLL1 S3865F was less active compared to MLL1 wildtype (Figure 18A).

To answer the question if cysteine at position 3864 is a new point of automethylation, another double mutant was cloned containing a glutamine at position 3864 and a serine at position 3882. After the confirmation of successful cloning by sequencing, the MLL1 R3864Q/C3882S protein was overexpressed and successfully purified (Figure 18B). Similar amounts of MLL1 R3864C/C3882S and MLL1 R3864Q/C3882S were used for methylation of recombinant H3. As control MLL1 wildtype and MLL1 C3882S was included. After the methylation reaction, the samples were separated via SDS-PAGE (Figure 18B). For both proteins, MLL1 C3882S and MLL1 R3864Q/C3882S, loss of automethylation was observed and also H3 methylation was identified, which means that both proteins were still active. With this result it was confirmed that the cysteine residue at position 3864 can be automethylated in the R3864C mutant.

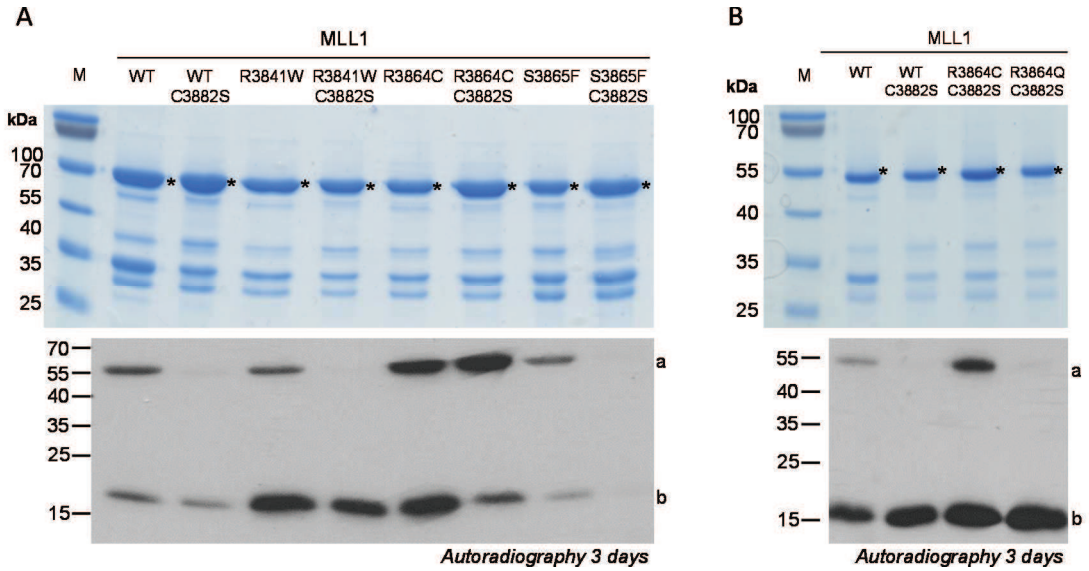


Figure 18: MLL1 automethylation assays. A) 16% SDS gel of MLL1 wildtype and MLL1 somatic cancer mutants with the corresponding C3882S mutant showing that roughly equal protein amounts of all MLL1 protein variants were used for methylation. B) 16% SDS gel of MLL1 wildtype (WT), MLL1 C3882S, MLL1 R3864C/C3882S and MLL1 R3864Q/C3882S showing that roughly equal protein amounts of all MLL1 protein variants were used for methylation. Methyl transfer was detected by autoradiography. The corresponding bands of the expected size are labeled with asterisk. Exposure times are indicated. a= automethylation of MLL1; b= methylation of recombinant H3.

As described above, it was shown that MLL1 only exhibits full methyltransferase activity, when it is associated with its complex partners WDR5, RBBP5 and ASH2L. In this context, the question arose, whether the exchange of cysteine against serine at position 3882 might influence the interaction with the complex partners. To answer this question, methylation of recombinant H3 by MLL1 wildtype and MLL1 C3882S was conducted in absence or in presence of the WRA complex partners. The same experiment was performed with MLL1 R3841W, MLL1 S3865F and the corresponding C3882S mutants. As described above, MLL1 was more active in presence of WRA. The same result was obtained with the MLL1 C3882S mutant. Also for MLL1 S3865F and MLL1 S3865F/C3882S higher methyltransferase activity was observed together with complex partners. In contrast to that, the somatic cancer mutant MLL1 R3841W was more active in absence of WRA. No change in the response to the presence of complex partners was observed with the MLL1 R3841W/C3882S mutant. These findings are in agreement with the previously described results in chapter 4.3.1.1 Figure 13.

For MLL1, MLL1 R3841W and MLL1 S3865F an increase in automethylation was observed in presence of complex partners. Interestingly, automethylation of MLL1 R3841W was increased with WRA, but H3 methylation was weaker. This result suggests that WRA somehow inhibits the H3 interaction of MLL1 R3841W, which has to be further investigated. In summary, the mutation C3882S did not change the enzymatic properties of MLL1 wildtype and its somatic cancer mutations in presence or absence of complex partners (Figure 19).

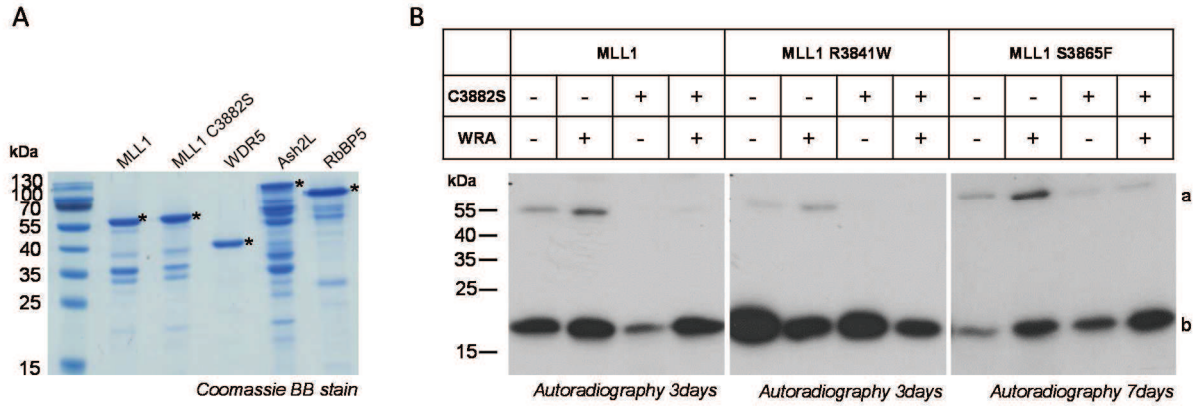


Figure 19: Automethylation of MLL protein variants and H3 methylation in presence or absence of complex proteins WDR5, RBBP5 and ASH2L (WRA). A) 16% SDS gel showing that roughly equal amounts of MLL1 wildtype, MLL1 C3882S and the complex proteins WDR5, RBBP5 and ASH2L were used for methylation. B) Methylation of H3 by MLL1 wildtype, MLL R3841W, MLL S3865F and the corresponding C3882S mutants in presence or absence of WRA. Methyl transfer was detected by autoradiography. The corresponding bands of the expected size are labeled with asterisk. Exposure times are indicated. a= automethylation of MLL1; b= methylation of recombinant H3.

4.3.1.2.3. Methylation of the non-histone targets by the MLL1 C3882S wildtype and cancer variants

Using the established C3882S mutants with reduced automethylation, the next aim of this study was to investigate if the purified non-histone protein domains are substrates for MLL1 protein variants and if the substrate preferences differ between wildtype MLL1 and the cancer mutants. To ensure almost equal protein amounts in the methylation assay, all selected non-histone targets were analyzed on SDS gels (Figure 20A). H3 was included as positive control. The targets DEGS2 and RADAP were only obtained in low concentrations, which resulted in a lower sensitivity in the methylation assay. For methylation, nearly equal amounts of the target protein domains were incubated with MLL1 C3882S in the presence of radioactively labeled AdoMet. After methylation, samples were loaded on a 16% SDS gel and methylation was analyzed by autoradiography. Despite the reduction of automethylation, only methylation of the two targets, TICRR and ZNF862, was confirmed, which were already observed before (compare Figure 17). Moreover, the data in this experiment showed that automethylation was not fully removed as shown after longer exposure times of films (Figure 20B/D).

As already mentioned, somatic cancer mutations can have different effects on the enzymatic properties of PKMTs. Changes in the catalytic properties of MLL1 S3865F and R3841W were already shown (4.3.1.1). Now, it was aimed to analyze if any of these MLL1 cancer mutations led to changes in the recognition of individual non-histone substrates. In this respect, the same methylation reactions of non-histone substrates as just described for MLL1 wildtype C3882S was performed with MLL1 R3841W/C3882S and MLL1 S3865F/C3882S. As shown in Figure 20, MLL1 R3841W and MLL1 S3865F methylate the same non-histone targets as MLL1 C3882S. To see if the substrate specificity of the MLL1 protein variants changes by the interaction with the complex partners WDR5, RBBP5 and ASH2L, the potential non-histone targets were screened with the same MLL1 constructs after complex formation with WRA. However, also in presence of WRA, the two non-histone targets, TICRR and ZNF862, were methylated by all the MLL1 cancer variants and no additional targets were detected (Figure 21). Altogether, these results indicate that the specificity of MLL1 protein variants was not changed by the cancer mutations and by the addition of complex partners.

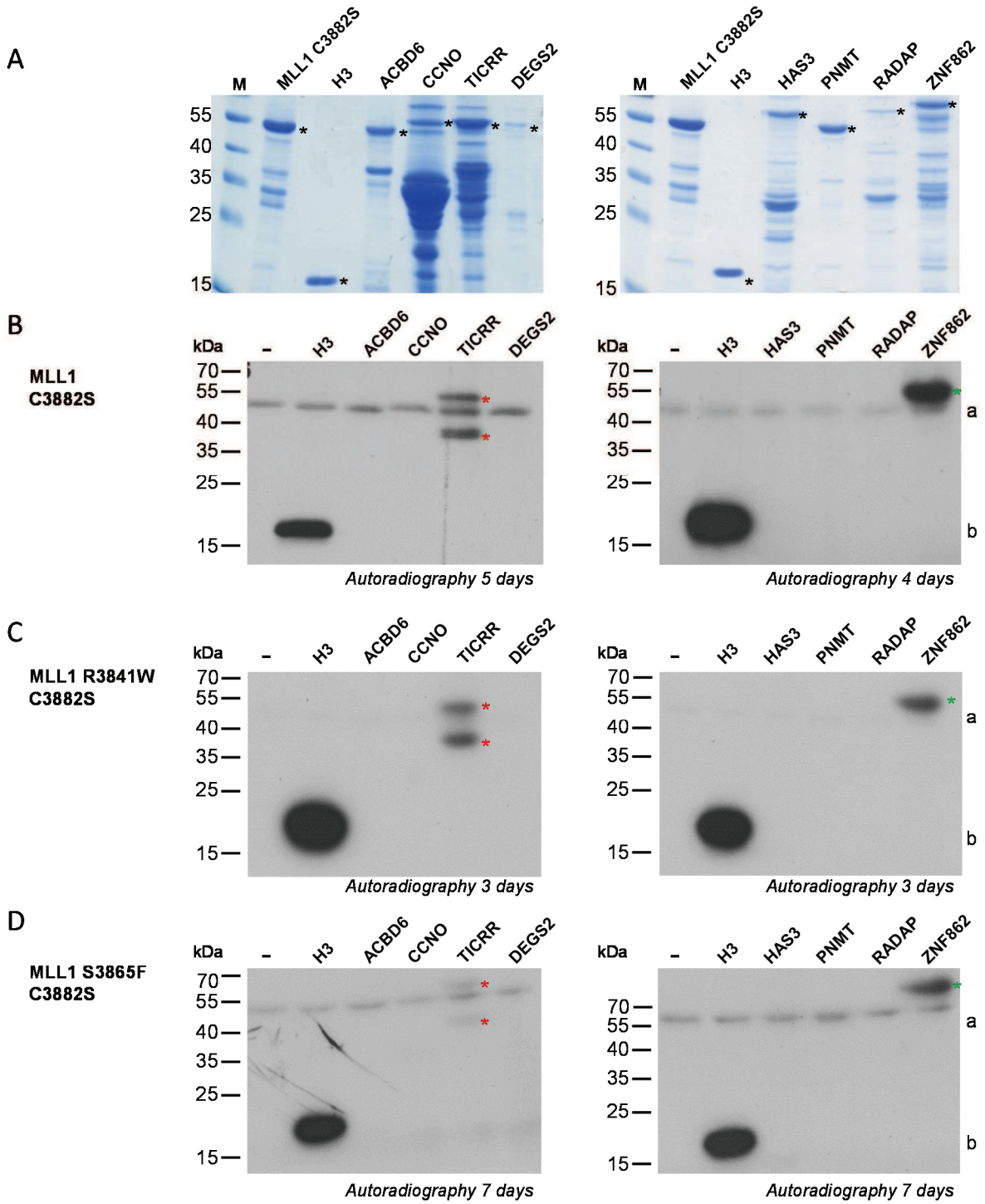


Figure 20: Screening of the non-histone targets with MLL1 C3882S and its somatic cancer mutants. A) 16% SDS gel of MLL1 C3882S with all selected non-histone targets serving as input control for the protein amounts of ACBD6, CCNO, TICRR, DEGS2, HAS3, PNMT, RADAP and ZNF862 that were used for methylation. B) Methylation of all proteins by the MLL1 C3882S. C) Methylation of all proteins by MLL1 R3841W/C3882S. D) Methylation of all proteins by MLL1 S3865F/C3882S. Methyl transfer was detected by autoradiography. The corresponding bands of the expected size are labeled with asterisk. Exposure times are indicated. a= automethylation of MLL1; b= methylation of recombinant H3.

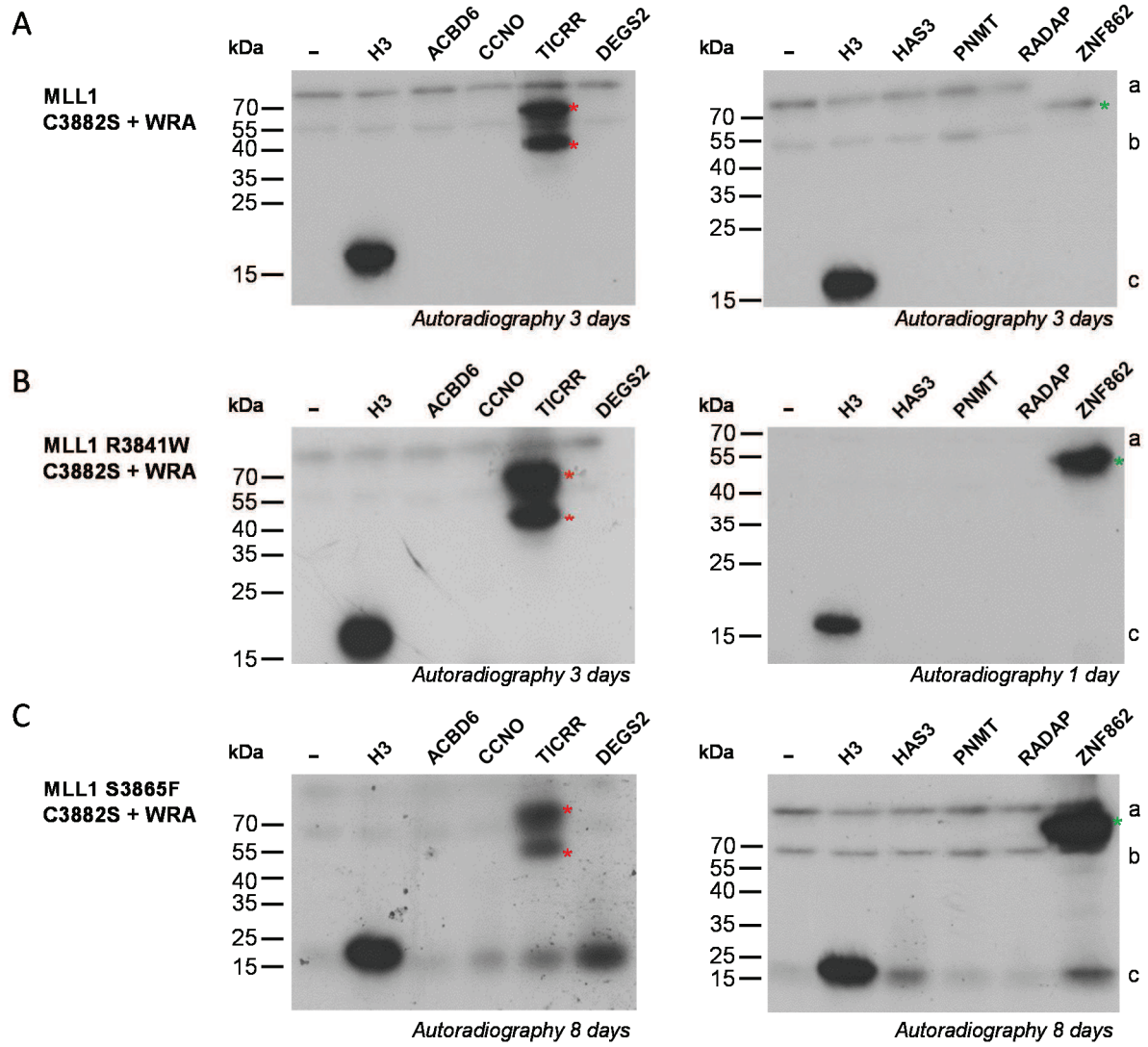


Figure 21: Screening of the non-histone targets with MLL1 C3882S and its somatic cancer mutants in presence of WRA. For methylation the same protein amounts of ACBD6, CCNO, TICRR, DEGS2, HAS3, PNMT, RADAP and ZNF862 were used as shown in Figure 20A. A) Methylation of all proteins by the MLL1 C3882S in presence of WRA. B) Methylation of all proteins by MLL1 R3841W/C3882S in presence of WRA. C) Methylation of all proteins by MLL1 S3865F/C3882S in presence of WRA. Methyl transfer was detected by autoradiography. The corresponding bands of the expected size are labeled with asterisk. Exposure times are indicated. a= methylation of ASH2L; b= automethylation of MLL1; c= methylation of recombinant H3.

4.3.1.2.4. Verification of the predicted target lysine residues of TICRR and ZNF862

An important step in the identification of novel non-histone PKMT substrates is the verification that the predicted target lysine of the novel non-histone substrates is indeed the place of methylation. For this, the putative target lysine residue was exchanged against an arginine by site directed mutagenesis. After successful cloning, the K to R mutants were overexpressed and purified by affinity chromatography. Equal amounts of wildtype TICRR or ZNF862 and their corresponding target lysine mutants were incubated with MLL1 C3882S with or without complex partners. The autoradiography image showed clear loss of the

methylation signal for TICRR K1019R and ZNF862 K457R with and without WRA complex (Figure 22). This result confirms that both proteins were methylated at their predicted target lysine, K1019 in TICRR and K457 in ZNF862.

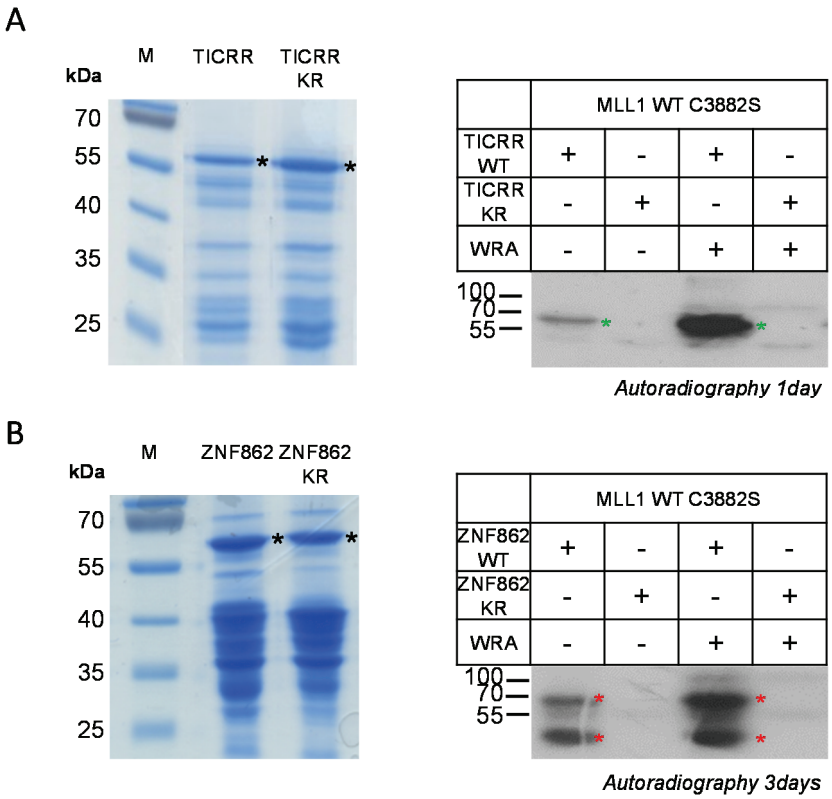


Figure 22: Verification of the target lysine methylation of TICRR and ZNF862. A) 16% SDS gel of TICRR and TICRR K to R mutant showing that roughly equal protein amounts of both proteins were used for methylation. Methylation of TICRR or TICRR KR by the MLL1 WT C3882S in presence or absence of WRA. B) 16% SDS gel of ZNF862 and ZNF862 K to R mutant documents that roughly equal protein amounts of both proteins were used for methylation. Methylation of ZNF862 or ZNF862 KR by MLL1 WT C3882S in presence or absence of WRA. Methyl transfer was detected by autoradiography. The corresponding bands of the expected size are labeled with asterisk. Exposure times are indicated.

4.3.1.2.5. Comparison of methylation activity of MLL1 C3882S cancer variants on H3 and the two novel non-histone targets.

In the next step, more quantitative methylation experiments with the identified non-histone targets were conducted to find out if they are better or worse substrates than H3. As shown before, the methylation of H3 with MLL1 C3882S was increased by the addition of complex partners. Also, the methylation of TICRR or ZNF862 with MLL1 C3882S was stronger when WRA was added. This result indicates that methylation of all substrates by MLL1 C3882S was stimulated by WRA, but the stimulatory effect was most pronounced with ZNF862. The comparison of the relative activity demonstrates that recombinant H3 was a better substrate than TICRR with or without complex proteins for MLL1 C3882S. In contrast, ZNF862 was a better substrate than H3 in presence of complex partners (Figure 23/24A).

As shown before, methylation of H3 by MLL1 R3841W/C3882S showed higher methyltransferase activity alone and addition of complex partners causes a decrease in activity. Interestingly, methylation of TICRR and ZNF862 by this mutant was increased after addition of WRA. This finding indicates that the methylation of the novel targets was stimulated by WRA. Also automethylation of MLL1 R3841W/C3882S was stimulated in the presence of complex partners. In contrast, the methylation of H3 by this mutant was reduced with WRA. It seems that the interaction of WRA inhibits the binding of H3 to MLL1 R3841W. The comparison of the relative activities shows that the recombinant H3 was a better target than TICRR with or without complex partners, whereas ZNF862 was a better substrate than H3 in presence of WRA (Figure 23/24B).

MLL1 S3865F/C3882S behave similar as the MLL1 C3882S enzyme. The methylation of recombinant H3, TICRR and ZNF862 was stimulated by the addition of complex proteins. The methylation signal of TICRR in presence of complex partners was comparable to H3 methylation in absence of WRA. Also for MLL1 S3865F/C3882S, H3 was a better substrate than TICRR, with or without WRA. No difference was observed between ZNF862 and H3 in presence of complex partners, whereas H3 was a better target than ZNF862 in absence of WRA. Altogether, MLL1 C3882S and the MLL1 C3882S cancer variants prefer H3 as target. Intriguingly, MLL1 R3841W/C3882S was more active on TICRR and ZNF862 with complex partners, whereas on H3 it was more active alone (Figure 23/24C).

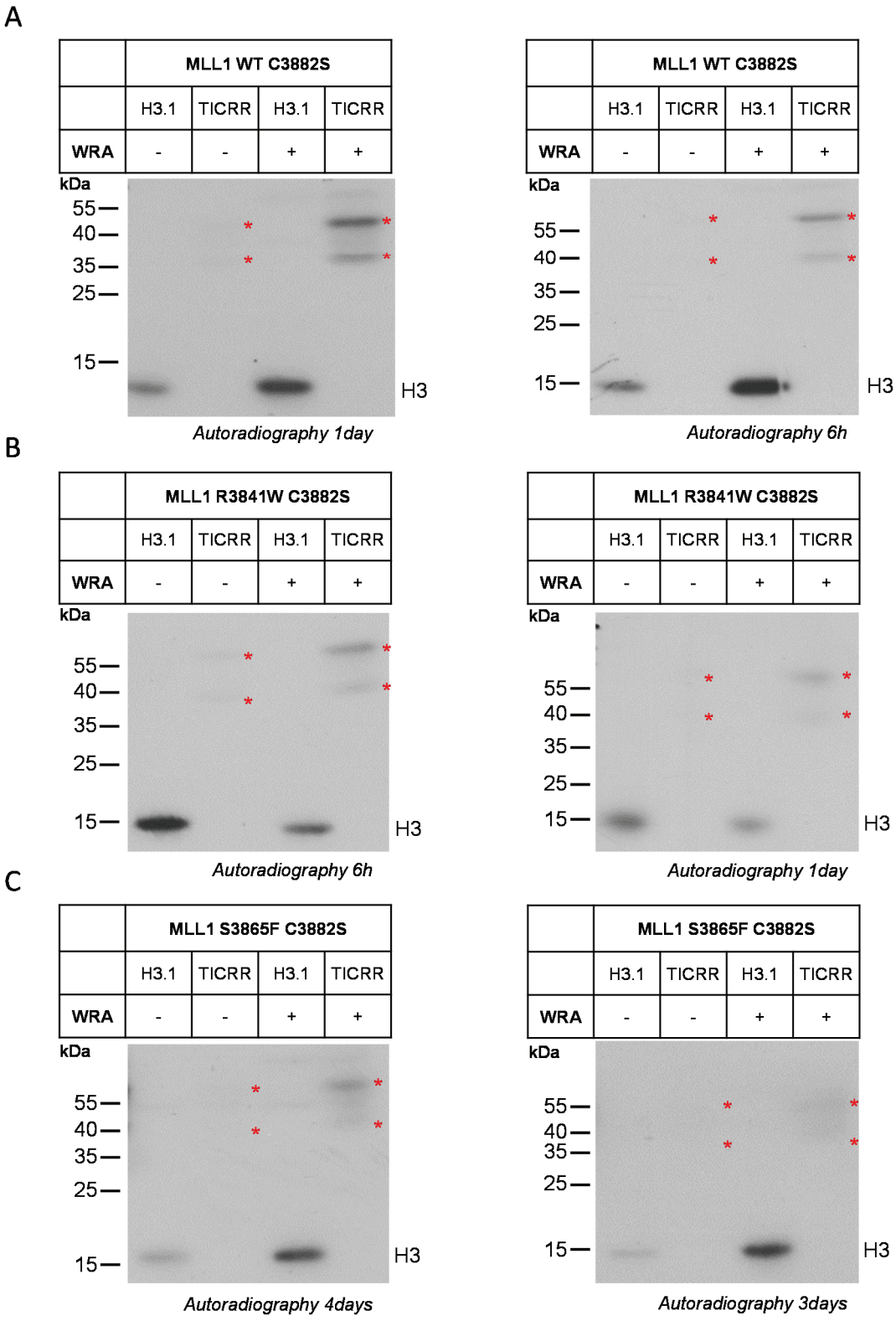


Figure 23: Comparison of the methylation activity of MLL1 and its somatic cancer mutants on H3 and TICRR. Duplicates of each experiment with different film exposition time are presented. A) Methylation of H3 or TICRR by MLL1 C3882S in absence (-) or presence (+) of WRA. B) Methylation of H3 or TICRR by MLL1 R3841W/C3882S in absence (-) or presence (+) of WRA. D) Methylation of H3 or TICRR by MLL1 S3865F/C3882S in absence (-) or presence (+) of WRA. Methyl transfer was detected by autoradiography. The corresponding bands of the expected size are labeled with asterisk. Exposure times and methylation of H3 are indicated.

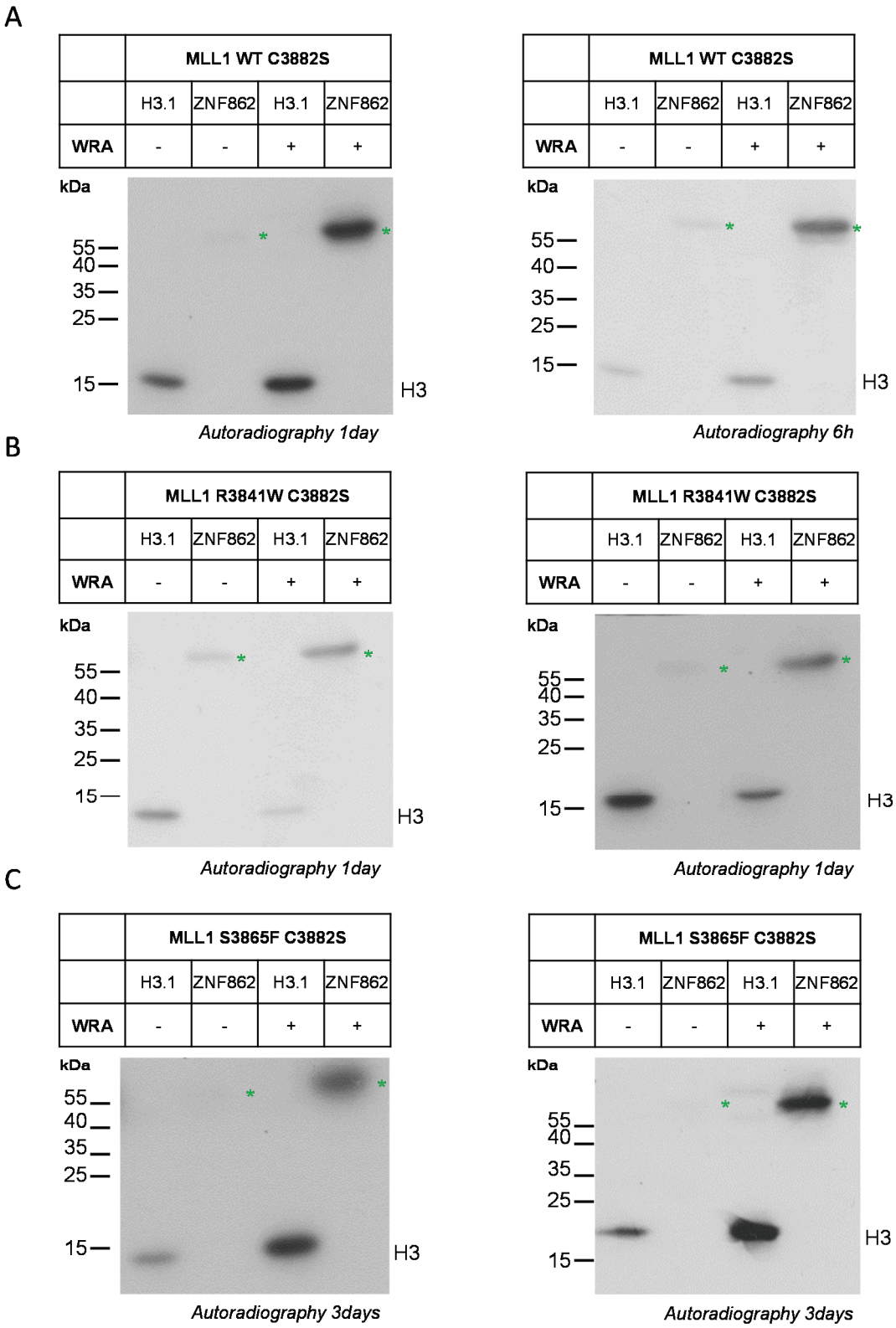


Figure 24: Comparison of methylation activity of MLL1 and its somatic cancer mutants on H3 and ZNF862. Duplicates of each experiment with different film exposition time are presented. A) Methylation of H3 or ZNF862 by MLL1 C3882S in absence (-) or presence (+) of WRA. B) Methylation of H3 or ZNF862 by MLL1 R3841W/C3882S in absence (-) or presence (+) of WRA. D) Methylation of H3 or ZNF862 by MLL1 S3865F/C3882S in absence (-) or presence (+) of WRA. Methyl transfer was detected by autoradiography. The corresponding bands of the expected size are labeled with asterisk. Exposure times and methylation of H3 are indicated.

4.3.1.2.6. Conclusion and outlook of the experiments with MLL1

Using substrate specificity arrays the recognition of each amino acid in the H3(1-15) peptide sequence by MLL1 was analyzed in detail. With the identified substrate recognition motif, 98 potential non-histone substrates were identified. Out of these, 23 targets were methylated by MLL1 at the peptide level. To test the methylation at the protein level, 8 non-histone targets could be successfully overexpressed and purified. Out of 8 potential new protein substrates tested, two were methylated at the predicted target lysine residue by MLL1 variants in presence or absence of WRA. Unfortunately, MLL1 showed strong automethylation, which reduced the sensitivity of the methylation assay with non-histone targets in the first attempt. By mutation of cysteine to serine at position 3882, automethylation of MLL1 and its mutants was strongly reduced. Only in the case of R3864C, automethylation could not be removed, because the newly introduced cysteine at position 3864 also was methylated. It was confirmed that the C3882S mutation does not change the enzymatic properties of MLL1, MLL1 R3841W or MLL1 S3865F. Out of 8 proteins tested, two non-histone targets were methylated at the predicted target lysine residue by MLL1 variants in presence or absence of WRA. However, the removal of automethylation did not reveal any additional methylation targets. In absence of complex proteins, all MLL1 variants prefer H3 as substrate. With complex partners, all MLL1 variants showed stronger methylation activity on ZNF862 than on H3.

In future experiments, it would be necessary to analyze, if the strong methylation activity on ZNF862 is due to stronger binding to MLL1. This could be performed by pull-down experiments. Furthermore, it would be interesting to determine the methylation degree of the identified non-histone targets. Finally, cellular studies with TICRR and ZNF862 would give further insights into the potential biological functions of these methylation events.

4.3.2. Investigation of the MLL3 protein lysine methyltransferase

4.3.2.1. Somatic cancer mutations in the MLL3-SET domain alter the catalytic properties of the enzyme

(Manuscript 4 in the attachment of this thesis)

Weirich S, Kudithipudi S, Kycia I & Jeltsch A. (2015). *Clinical Epigenetics, 2015 Mar 28;7:36. doi: 10.1186/s13148-015-0075-3.*

The MLL3 (KMT2C) protein is another member of the MLL protein family. It is a 4911 amino acids long enzyme (Smith et al. 2011) that functions mainly as a monomethyltransferase introducing H3K4me1 at active enhancers (Cheng et al. 2014). MLL3

has been found mutated in several cancers including glioblastoma, melanoma, pancreatic and breast cancer (Chen et al. 2014; Kandoth et al. 2013; Kudithipudi & Jeltsch 2014) and recent studies reported that MLL3 is one of the most frequently mutated protein lysine methyltransferase in cancer (Chen et al. 2014; Kandoth et al. 2013; Kudithipudi & Jeltsch 2014). Using the COSMIC database, three missense mutations located in the catalytic SET domain of MLL3 (4771 to 4911) were selected for biochemical analysis based on their location either close to the peptide or AdoMet binding sites of the enzyme: S4757C (found in lung cancer), N4848S (found in endometrium, central nervous system) and Y4884C (found in large intestine) (manuscript 4, Figure 1). MLL3 wildtype and the selected cancer mutants were cloned as GST-fusion proteins and purified via affinity chromatography (manuscript 4, Suppl. Figure 1A). Methylation experiments indicated that MLL3 wildtype is strongly stimulated by the presence of the WRA complex proteins. The two mutants, S4757C and Y4884C, behave similar to MLL3 wildtype by showing an increased catalytic activity in the presence of complex partners. However, the activity of Y4884C was 4-fold reduced by WRA. The amino acid exchange from asparagine to serine at position 4848 converts MLL3 to an inactive methyltransferase (Figure 25). Correct protein folding of all proteins was confirmed by analyzing the secondary structure composition with CD spectroscopy (manuscript 4, Suppl. Figure 1B).

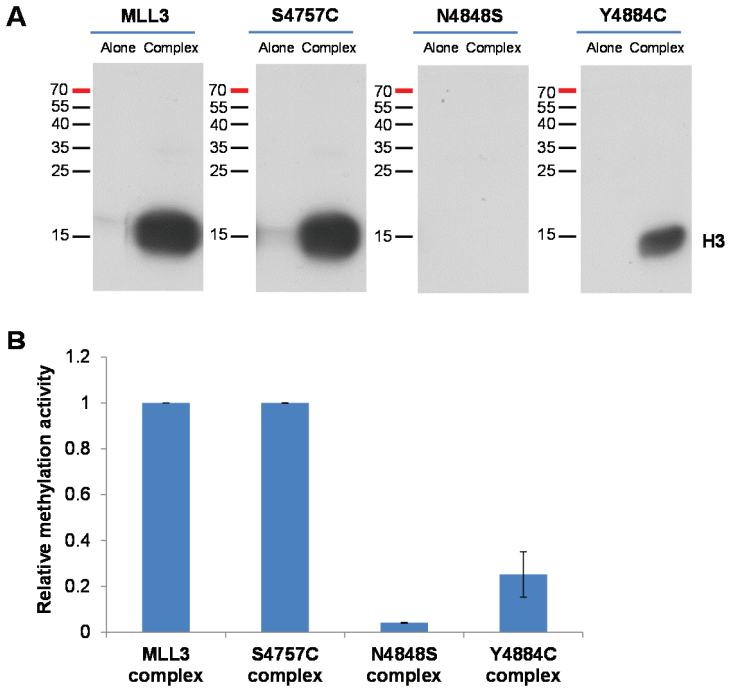


Figure 25: Activity of MLL3 protein variants. Recombinant histone H3 protein was methylated with radioactively labeled AdoMet by MLL3-SET wildtype and MLL3-SET mutant proteins either alone or in the presence of complex member proteins. A) Example of an autoradiographic image of the SDS polyacrylamide gel. The methylation signal of H3 is indicated. B) Quantitative analysis of the averages of duplicate experiments. The error bars indicate the standard error of the mean. (Taken from manuscript 4, Figure 2)

To further analyze the effects of these mutations on the substrate specificity and product pattern, peptide methylation experiment were carried out. SPOT peptide arrays with different methylated forms of H3K4 and a K4A mutant peptide as negative control were incubated with the MLL3 variants in presence of complex partners and radioactively labeled AdoMet. In agreement with the previous result, N4848S mutant was inactive. MLL3 wildtype and S4757C preferred the unmethylated H3K4 peptide as substrate, and no methylation was observed with methylated H3K4 peptides. However, a strong change in the substrate specificity for Y4884C was observed. The Y4884C variant preferred the H3K4me1 peptide as substrate and additionally weak methylation was observed with H3K4me2 substrate. These results show that MLL3 and S4757C transfer one single methyl group on unmethylated H3K4, whereas Y4884C can transfer up to three methyl groups and it prefers H3K4me1 as substrate (Figure 26A). Mass spectroscopy measurements were used to further analyze the change in the substrate and product specificity of Y4884C variant. MLL3 wildtype and Y4884C variant were incubated in presence of complex partners with soluble unmethylated or monomethylated H3 peptides and the methylation reaction was analyzed at different time points. MLL3 wildtype was shown to be a strict monomethyltransferase only using unmethylated H3K4 peptide as substrate (Hu et al. 2013; Herz et al. 2012). The same results were obtained using the S4757C mutant. Interestingly, Y4884C was only weakly active on H3K4me0 peptide. Still, no accumulation of H3K4me1 appeared which implies that this mutated enzyme immediately converts monomethylated H3K4 peptide into the di- and trimethylated forms. Using monomethylated H3K4 peptide substrate, strong methylation signals were observed and the dimethylated and trimethylated H3K4 peptide products were generated (manuscript 4, Figure 4).

Change in the product specificity of MLL3 Y4884C was also observed at the protein level using recombinant histone H3 as substrate. After the methylation reactions, the samples were separated via SDS-PAGE and H3K4 trimethylation was detected using an H3K4me3 antibody. In presence of complex partners, a strong H3K4 trimethylation signal was detected after H3 methylation with the Y4884C mutant, whereas with MLL3 wildtype only a weak signal was observed. Without complex partners both enzymes did not show activity (Figure 26B). Furthermore, the global H3K4me3 level was analyzed in cells after overexpression of the different cancer mutants. Therefore, histones were isolated from HEK293 cells, which were transfected with expression constructs of the individual MLL3 cancer variants. The histones were separated via SDS-PAGE and methylation analyzed with H3K4me3 antibody. Similarly as in the *in vitro* reaction, transiently expressed Y4884C increased the H3K4me3

signal (Figure 26C), while the other mutants did not show clear differences to the wildtype result. In summary, this research shows that somatic cancer mutations can have individual and specific effects on the enzymatic properties of PKMTs. In the case of MLL3, the mutation N4848S converts MLL3 to an inactive enzyme, whereas S4757C shows no significant changes compared to MLL3 wildtype. Still, there can be differences, which were not detected with these experimental approaches. Interestingly, exchange of tyrosine to cysteine at position 4884 results in the conversion of MLL3 from a monomethyltransferase to a trimethyltransferase, which was confirmed *in vitro* and in cells. This observation is in analogy to EZH2 Y641 as described in chapter 1.10.

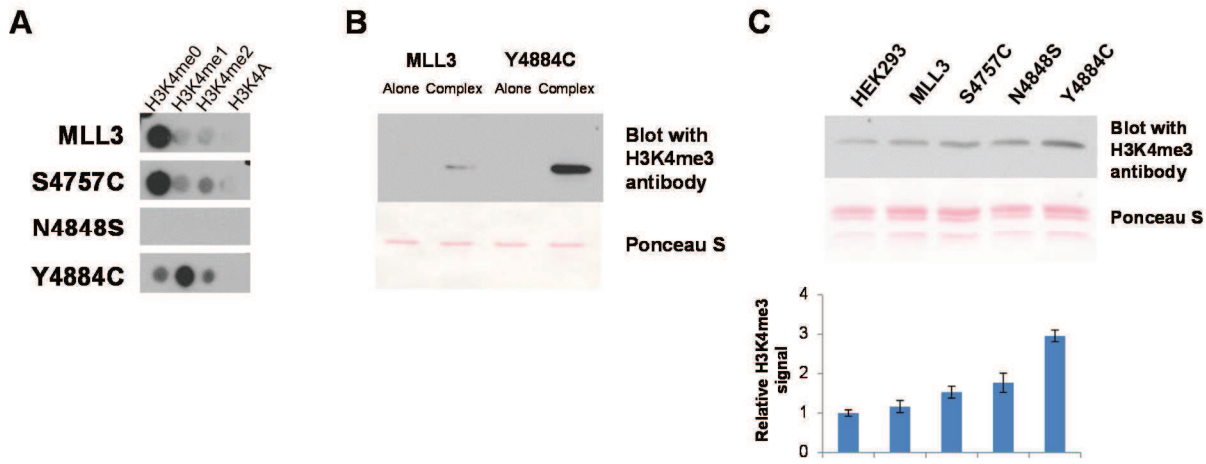


Figure 26: Substrate and product specificity of MLL3 protein variants. A) H3(1-15) peptide arrays containing H3K4 at different methylation states and a K4A variant peptide were methylated with MLL3-SET wildtype and mutant proteins in the presence of complex members using radioactively labeled AdoMet. The figures show an example of an autoradiographic image of the methylated peptide SPOT arrays, peptides with the corresponding lysine variants are indicated (Taken from manuscript 4 Figure 3A). B) Methylation of recombinant histone H3 protein by MLL3-SET wildtype and Y4884C alone and in the presence of complex members using unlabeled AdoMet as cofactor. After methylation the proteins were separated on a SDS polyacrylamide gel, blotted and the methylation was detected using an H3K4 trimethylated antibody (upper panel). The lower panel shows the Ponceau S stained image of the blot (Taken from manuscript 4 Figure 6A). C) Global histone H3K4me3 methylation analysis from HEK293 cells. The cells were transfected with different MLL3 variant plasmids, histones were isolated and H3K4me3 methylation probed by western blot. The upper image shows the H3K4me3 signal and a Ponceau S stain as loading control. The bar diagram shows the quantification from 3 experiments. The error bars display the standard error of the mean. The signal obtained from the MLL3 was set to 1 and the other signals were normalized accordingly (Taken from manuscript 4 Figure 6B).

4.3.2.2. Specificity analysis of MLL3 in complex with WDR5, RBBP5 and ASH2L

As it was already described for MLL1 in 4.3.1.2.1, a peptide array was used to investigate the influence of each amino acid in the H3(1-15) template for the substrate recognition by MLL3. In the case of MLL3, methylation of the substrate specificity array was performed in presence of complex partners WDR5, RBBP5 and ASH2L, as MLL3 was only active in presence of complex partners (4.3.2.1). After methylation of the peptide array with the MLL3-WRA

protein complex in the presence of radioactively labeled AdoMet, the methyl transfer was analyzed by autoradiography. The experiment was carried out twice and the average activities were calculated and plotted in Figure 27A. The calculation of the standard deviation for the methylation activity of each peptide from both experiments showed that the results are highly reproducible with a SD smaller than 20% for 92.5% of the peptides. 99% of the peptides have a SD < 30% and only 2 peptides showed a SD between 30 and 40% (Figure 27B).

The results show that MLL3-WRA recognizes the template H3 (1-15) residues from A1 to T6. At position -2, MLL3-WRA mainly tolerates the canonical arginine and only weakly accepts leucine. Alanine at position -3 can be replaced by all other hydrophobic residues, additionally serine and threonine were accepted. Beside glutamine at position +1, MLL3-WRA has also a good activity with asparagine, histidine and methionine. In summary, the substrate specificity recognition motif can be described as follows: A(ILFP)-R(L)-T(AILMFYV)-K-Q(NHM)-T(AQEILSV).

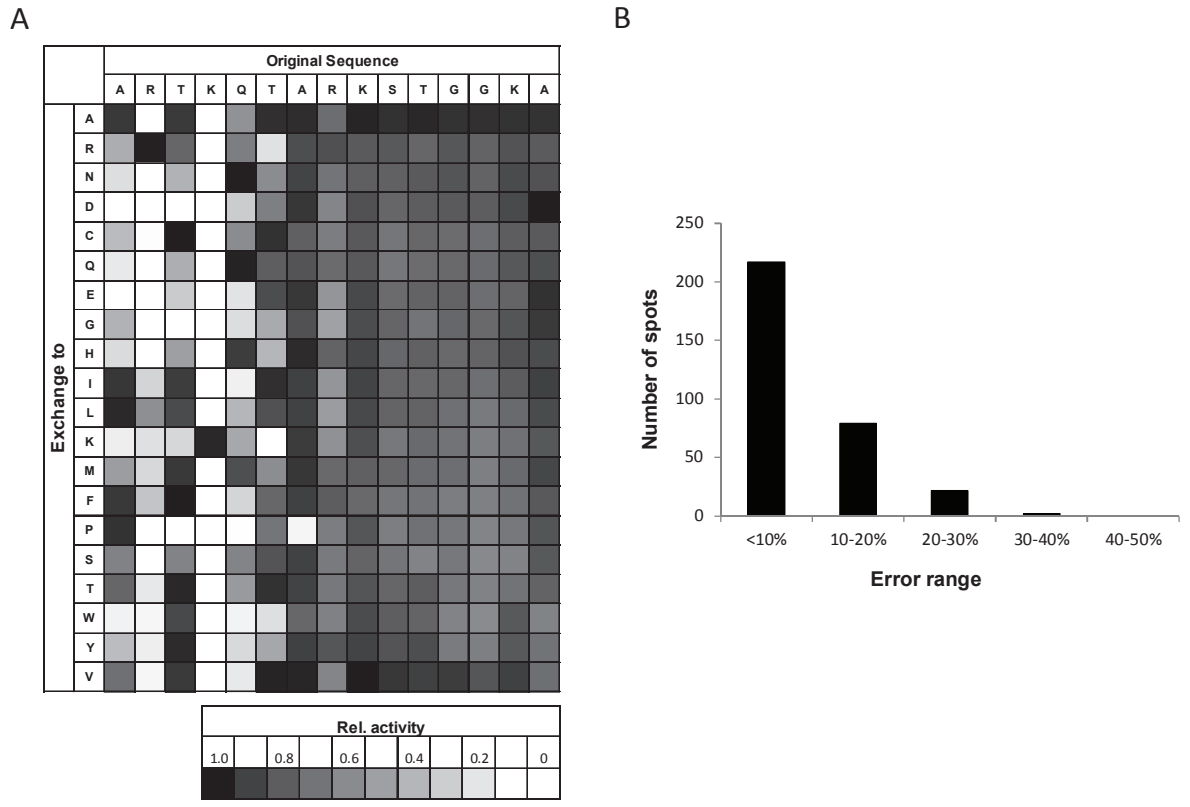


Figure 27: Substrate specificity analysis of MLL3 in complex with WRA. A) Two independent peptide array methylation experiments were performed with MLL3+WRA and the data were averaged after normalizing the full activity to 1. The activity is displayed in the grey scale as indicated. The horizontal axis represents the template sequence (H3 1-15). Each residue was exchanged against all 20 natural amino acid residues as indicated by the vertical axis. B) Distribution of standard errors of the mean of MLL3+WRA on all peptides tested in the two independent assays.

To screen the human proteome for possible non-histone targets of MLL3, a Scansite search with the substrate recognition motif was conducted and 42 potential non-histone substrates were identified (Obenauer et al. 2003). To check if the identified non-histone targets can be methylated at peptide level, 15 amino acid long peptides were synthesized on a cellulose membrane. The peptide array was incubated with MLL3 complex containing WDR5, RBBP5 and ASH2L in presence of radioactively labeled AdoMet and the methyl group transfer was detected by autoradiography. Out of the 42 putative non-histone targets (Table 5, see appendix), 24 were methylated (Figure 28A). Out of these, 10 target protein domains were selected (Table 2), overexpressed and purified by affinity chromatography (Figure 28B).

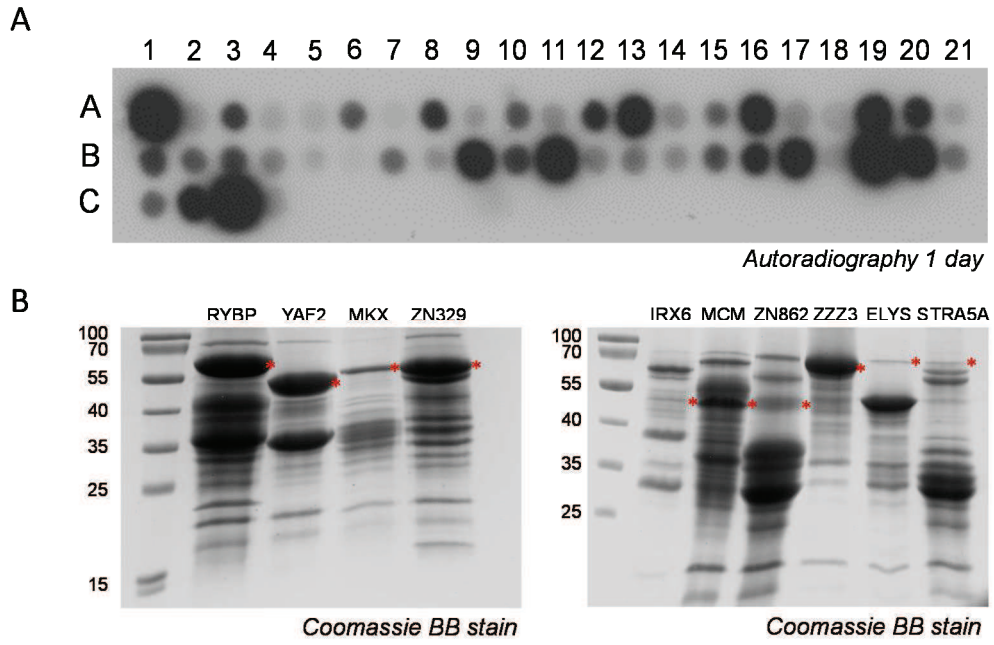


Figure 28: Screening of MLL3 non-histone targets. A) The peptide array was methylated with MLL3 together with WRA in presence of radioactively labeled AdoMet. The transfer of the methyl group to the immobilized target lysine residues was detected by autoradiography. Target information is listed in Table 5. B) 16% SDS gel documents the purification of 10 non-histone targets. The corresponding bands of the expected size are labeled with asterisk.

Table 2: Selected non-histone targets from the peptide array (Figure 28A), which were further analyzed.

	Swiss Prot. Nr.	Original name	Name	Sequence	Protein-length	MW (kDa)	Target K position
A6	Q8WYP5	ELYS_HUMAN	Protein ELYS	DKQLRIKHVRRVRGR	2266	253	1933
A13	P78412	IRX6_HUMAN	Iroquois-class homeodomain protein IRX-6	SGAGRRKNATRETTS	446	48	151
A16	Q7L590	MCM10_HUMAN	Protein MCM10 homolog	PALPRTKRVARTPKA	875	98	187
A19	Q8IYA7	MKX_HUMAN	Homeobox protein Mohawk	NARRRLKNTVRQPDL	352	39	129
B9	Q8N488	RYBP_HUMAN	RING1 and YY1-binding protein	TSRPRLKNVDRSTAQ	228	25	149
B11	P42229	STA5A_HUMAN	Signal transducer and activator of transcription 5A / 5B	MSLKRIKRADRRGAE	794	91	425
B17	Q8IY57	YAF2_HUMAN	YY1-associated factor 2	KTRPRLKNVDRSSAQ	180	20	106
B19	Q86UD4	ZN329_HUMAN	Zinc finger protein 329	MRLKMTTRNFPEREV	541	62	4
B20	O60290	ZN862_HUMAN	Zinc finger protein 862	DGPRRIKRTYRPRSI	1169	132	457
C2	Q8IYH5	ZZZ3_HUMAN	ZZ-type zinc finger-containing protein 3	PVLKRIKRCLRSEAP	903	102	117/120

Equal amounts of RYBP, YAF2, MKX and ZN329 (Figure 29A) were incubated with the MLL3-WRA complex (Figure 29B) in a reaction mixture containing radioactively labeled AdoMet, to investigate methylation of the candidate substrate proteins. As positive control, H3 was included. To rule out any methylation signal from the MLL3 or its complex partners, MLL3-WRA was also incubated under the same conditions without any substrate. Strong methylation of the recombinant H3 was detected. In addition, methylation of ASH2L was observed. Further investigation is necessary to analyze if MLL1 catalyze the methylation of ASH2L or if ASH2L automethylation is observed. However, even after long exposure, none of the four selected non-histone targets was methylated by MLL3 in presence of complex partners (Figure 29C).

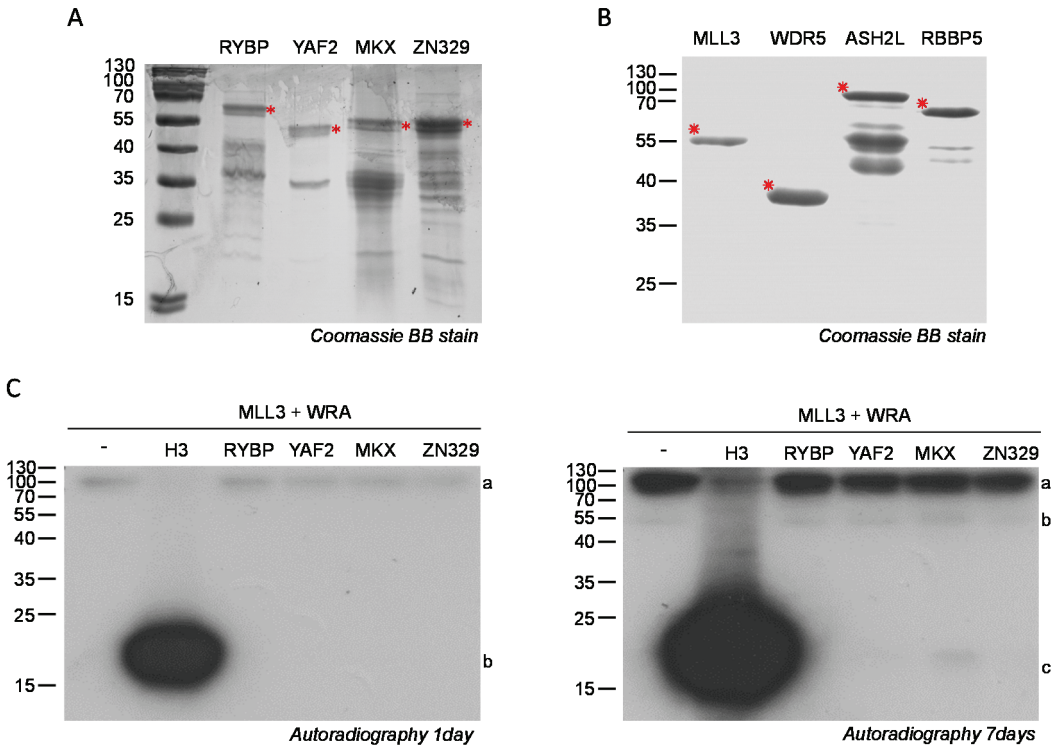


Figure 29: Methylation of non-histone proteins by MLL3. A) 16% SDS gel of all selected non-histone targets serving as input control of the protein amounts of RYBP, YAF2, MKX and ZN329 that were used for methylation. B) 16% SDS gel of MLL3 with the complex proteins WDR5, RBBP5 and ASH2L showing that roughly equal protein amounts of all proteins were used for methylation C) Methylation of RYBP, YAF2, MKX and ZN329 by MLL3 in presence of WRA. Methyl transfer was detected by autoradiography. The corresponding bands of the expected size are labeled with asterisk. Exposure times are indicated. a= methylation of ASH2L; b= automethylation of MLL3; c= methylation of recombinant H3.

4.3.2.2.1. Conclusion and outlook of the MLL3 experiments

MLL3 substrate specificity peptide arrays were used to analyze the interaction of MLL3 complex with the amino acids of the H3 peptide (1-15). Using a Scansite search, 42 potential non-histone targets with the identified substrate recognition motif were found. Out of these, 24 peptides were methylated and 10 corresponding proteins were successfully overexpressed and purified. Methylation was performed for RYBP, YAF2, MKX and ZN329 with MLL3 in complex, but no methylation of any of the non-histone substrates could be identified. More research is necessary to test the potential methylation of the other non-histone protein domains (IRX, MCM, ZN862, ZZZ3, ELYS and STRA5A). In addition, more targets can be selected on the peptide array to test methylation at protein level. Furthermore, methylation of the targets should be performed with MLL3 in absence of complex partners, because it is possible that WRA influences the specificity of MLL3.

5. Discussion

Histone lysine methylation is an abundant posttranslational modification with very important roles in the regulation of epigenetic mechanisms. Different protein lysine methyltransferases are known to catalyze mono-, di- and trimethylation of lysine residues. Dependent on the methylation degree and the methylated residue, lysine methylation is associated with transcriptional activity, repression or other functions. Since the discovery of the first protein lysine methyltransferase Suv39h1 by Jenuwein (Rea et al. 2000), the insight into the biological relevance of protein lysine methylation has increased drastically. Many PKMTs were identified and it was shown that different enzymes methylate the same target site, but sometimes also different target sites are methylated by the same PKMT. In addition, non-histone proteins were identified as possible substrates for methyltransferases. By now, several hundreds of methylated target lysine residues in the human proteome have been detected, but the responsible PKMTs are not yet identified. On the other hand, for many PKMTs the variety of relevant targets is unknown. However, due to the fast development of the field wrong assignments occurred between PKMTs and methylated target sites, which led to wrong interpretation of biological processes.

In this study, the substrate specificity of SUV4-20H1, SUV4-20H2, MLL1 and MLL3 was determined and it was demonstrated that specificity analysis is a powerful strategy to identify novel non-histone targets for PKMTs and to prevent mis-assignment. In addition, detailed biochemical investigations of the effects of somatic cancer mutations in PKMTs showed mutation related changes of the enzymatic properties. This discovery suggests the development of cancer specific therapeutic strategies.

The following chapters represent a summarized discussion of all projects. A more detailed discussion of individual results is presented in the different publications attached in the appendix.

5.1. Substrate and product specificity of SUV4-20H1 and SUV4-20H2

In mammals, many chromatin modifying enzymes are present in two or more paralogs, like MLL1/2, MLL3/4, and G9A/GLP. All of them have overlapping, but also non-redundant biological functions. In this study, the two paralogs SUV4-20H1 and SUV4-20H2 were further investigated. In human cells, three enzymes SET8, SUV4-20H1 and SUV4-20H2 are the most important H4K20 methyltransferases. SET8 is known as a strict H4K20 monomethyltransferase (Zhang & Bruice 2008). In contrast, previous studies raised questions regarding the product and substrate specificity of the two paralogs SUV4-20H1 and SUV4-

20H2. Southall et al. (2014) reported a ~ 10 fold preference, whereas Wu et al. (2013) observed a ~ 3 fold preference of SUV4-20H enzymes for monomethylated H4K20 substrate over unmethylated ones. Even though both groups showed different preferences, they reported that SUV4-20H enzymes transfer one single methyl group on H4K20me1. However, H4K20 trimethylation was detected with an antibody even in an *in vitro* reaction, when recombinant nucleosomes were used as substrates (Schotta et al. 2004).

5.1.1. The methylation pattern of SUV4-20H1 and SUV4-20H2

In this study, peptide array methylation experiments and methylation of purified peptides in solution have shown very similar substrate and product methylation state preferences for both SUV4-20H enzymes. Both enzymes prefer monomethylated H4K20 as substrate to catalyze higher methylation forms. These findings were underscored by knock-out studies of SET8 that led to global losses of H4K20me1, but also H4K20me2 and H4K20me3 (Oda et al. 2009; Wu et al. 2013; Southall et al. 2014). Moreover, different lines of evidence indicated in this work that SUV4-20H enzymes transfer one methyl group on monomethylated target substrate *in vitro*. This result is also in agreement with other studies and can be explained by the unique active site geometry of SUV4-20H1 and SUV4-20H2. In canonical monomethyltransferases (e.g. SET8), a tyrosine residue is located in the SET domain and interacts with the ε-amino group of the target lysine. Compared to that, SUV4-20H enzymes contain a serine residue which enlarges the pocket and allow binding of monomethylated substrates. Similarly as the tyrosine in SET8, the serine forms a hydrogen bond with the Nε of the target lysine and this interaction prevents rotation of the lysine residue and limits the enzyme to the transfer of only one methyl group on the monomethylated H4K20 substrate (Southall et al. 2014; Wu et al. 2013). This result is not compatible with knock-out studies of SUV4-20H enzymes, which led to a genome wide loss of H4K20me2 and –me3 in mice and zebrafish (Schotta et al. 2004; Kuo et al. 2012). As mentioned above, H4K20 trimethylation was also observed when nucleosomes were used as substrates (Schotta et al. 2004), which may explain this discrepancy. Alternatively (but less likely) another enzyme may use dimethylated H4K20 as substrate to introduce H4K20me3 *in vivo* or the product pattern of SUV4-20H enzymes is different in cells. Altogether, it was shown that both paralogs have similar properties regarding the methylation state preference and the number of transferred methyl groups.

5.1.2. The substrate specificity of SUV4-20H1 and SUV4-20H2

Substrate specificity analysis was performed to analyze the influence of each single amino acid in the histone template sequence on the enzymatic activity and identify possible

differences between the substrate peptide interaction of SUV4-20H1 and SUV4-20H2. Interestingly, the data indicated that SUV4-20H1 and SUV4-20H2 recognize (RY)-Kme1-(IVLM)-(LFI)-X-D and X-Kme1-(IVLMK)-(LVFI)-X-(DEV) sequences. Hence, both enzymes have overlapping sequence recognition motifs, but at two positions SUV4-20H1 is more specific towards H4K20 than SUV4-20H2. Especially at position -1, SUV4-20H1 only tolerates an arginine, whereas SUV4-20H2 accepts several other residues. Similarly, at position +4 SUV4-20H1 is only active when aspartic acid is present, but in the case of SUV4-20H2 aspartic acid at position +4 can be exchanged by glutamic acid or asparagine. This difference in the recognition motifs indicates that SUV4-20H2 is less specific than SUV4-20H1 and suggest that SUV4-20H2 can methylate more non-histone targets than SUV4-20H1.

5.1.3. Identification of NHT for SUV4-20H1 and SUV4-20H2

The identified sequence recognition motif of SUV4-20H2 was used to search for potential non-histone substrates, because it also covers the recognition motif of SUV4-20H1. The Phosphosite PLUS database has listed all human proteins, which have been found to be methylated in proteomic experiments, but often the responsible PKMT is not yet identified (Hornbeck et al. 2015). A search in Phosphosite PLUS revealed seven non-histone proteins that are lysine methylated within the sequence recognition motif of SUV4-20H enzymes. SUV4-20H2 methylates three (CASZ1, OIP5 and CENPU) out of the 7 potential NHT. In comparison to that, SUV4-20H1 methylates only one target (CASZ1), which is in agreement with the observation that it is more specific. Interestingly, among all 7 proteins only CASZ1 carries an R at position -1, a finding that underscores the importance of R at position -1 for SUV4-20H1. The two non-histone substrates OIP5 and CENPU, which are methylated by SUV4-20H2, are both centromeric proteins, suggesting that they may have a role in heterochromatin organization and may come in contact with SUV4-20H2 in cells. CASZ1 is a zinc finger transcription factor and may support SUV4-20H1 and H2 in gene regulation. Overall, these data indicate that substrate recognition analysis is a promising strategy to identify even subtle differences in properties of PKMT paralogs. This information can help to get further insights into their individual biological roles.

5.2. Investigation of problematic assignments between PKMTs and methylated target sites

5.2.1. Investigation of the methylation of ERK1 by SUV4-20H1

Recently, it was reported that ERK1 is methylated by SUV4-20H1 at K302 and K361 (Vougiouklakis et al. 2015). However, a comparison with the identified specificity profile of SUV4-20H1 has shown that the surrounding sequences of both reported target lysine residues are not fitting to the requirements of SUV4-20H1. In addition, it was reported that SUV4-20H1 introduces a trimethylation on the unmethylated ERK1 target substrate, which is in disagreement with results presented above regarding the substrate methylation state preferences of SUV4-20H1. To solve this discrepancy, further methylation experiments were performed in this study. However methylation experiments with ERK1 peptides and protein did not provide any evidence for methylation of ERK1 by SUV4-20H1 or SUV4-20H2.

While working on ERK1 methylation, additional weaknesses of the Vougiouklakis et al. (2015) study were recognized. When searching for novel non-histone targets, it is important to include known targets as positive controls to compare the relative activity. Vougiouklakis et al. (2015) performed methylation of ERK1 with recombinant SUV4-20H1 using radioactively labeled AdoMet, but no positive control was included and also no quantification was done. Thus, no statement can be made about the strength of the methylation signal on ERK1. By MS/MS measurements, they identified K302 and K361 as target lysine sites for the methylation by SUV4-20H1. However, surprisingly individual mutations of single target residues to alanine completely abolished ERK1 methylation, although the second lysine residue was still present. Moreover, when searching for new PKMT substrates, it is helpful to use inactive enzyme variants to check if detected methylation events are really catalyzed by the enzyme of interest or if methylation signals are coming from contaminating PKMT activity copurified with the enzyme. This speculation appeared, because Vougiouklakis et al. (2015) did not show the purified protein and also the source and purification method were not described. In addition, the reported methylation of ERK1 by SUV4-20H1 was very weak, as judged by the images provided by the authors. Weak methylation signals can occur from binding of AdoMet to the protein or non-specific methylation detected after long exposure. This possibility can be excluded if parallel reactions are conducted with an inactive enzyme variant as well as positive controls are included. One potential problem in my study was that mouse SUV4-20H enzymes were used to methylate human ERK1 protein, which could be the reason for the missing ERK1 methylation. However, sequence comparison of human and

mouse SUV4-20H1/H2 as well as human and mouse ERK1 ruled out across species discrepancies. Human and mouse SUV4-20H1 differ only at two residues, which are located far away from the peptide binding site. No difference was observed within the peptide recognition motif of SUV4-20H enzymes between human and mouse ERK1. In summary, if true, methylation of ERK1 by SUV4-20H1 would be a very relevant finding, because ERK1 activity plays a central role in carcinogenesis and a potential role of SUV4-20H1 in this process would be very important. One may speculate that substrate and product methylation state preferences may change with different substrates, but there is no precedence case for this speculation. Therefore it is not possible to completely rule out the methylation of ERK1 by SUV4-20H1, based on the data presented here, but one need to take the claim of ERK1 methylation by SUV4-20H1 with a lot of care, and additional supportive evidence would be needed before follow up studies on biological roles of this methylation event are recommended.

5.2.2. Investigation of the methylation of Numb by SET8

The methylation of the Numb protein by the SET8 protein lysine methyltransferase, is another example for a problematic assignment between methylation site and PKMT (Dhami et al. 2013). Numb was shown to have a tumor-suppressive function acting in a p53 dependent manner. Dhami et al. (2013) reported that the interaction between p53 and Numb is regulated by the methyltransferase SET8. They reported that SET8 dimethylates Numb at K158 and K163. Methylated Numb was found to disrupt the interaction with p53, which then is ubiquitinated and finally degraded. However, also in this case, the surrounding sequences of the predicted target lysine residues differ from the specificity profile of SET8 at 4 out of 6 positions. Additionally, SET8 is a strict monomethyltransferase and dimethylation was not reported before. In this study, a further investigation of Numb methylation by SET8 revealed that this enzyme is not able to methylate Numb at K158 or K163 *in vitro*. Similarly as described in the previous example, Dhami et al. (2013) also missed to include essential controls in their study. For methylation experiments, SET8 protein was used that was purified from cells by immunoprecipitation. The purity of the sample was not shown and no catalytically inactive SET8 enzyme was included to confirm the purity of the immunoprecipitated enzyme. Therefore, it cannot be excluded that the reported methylation signal comes from another PKMT co-immunoprecipitated with SET8. Furthermore, positive controls were missing as well, which would have allowed to relate the methylation activity at novel sites to that at already known and well characterized sites. In my study, H4 and p53 proteins were included as positive controls, because previous studies indicate methylation of

these two proteins by SET8 (Shi et al. 2007; Zhang & Bruice 2008). Both substrates were methylated, showing that SET8 was active and the methylation assay was sensitive. Although Numb methylation could not be detected *in vitro*, *in vivo* methylation cannot be completely ruled out, because interaction partners or posttranslational modifications can influence the substrate preference of PKMTs. However, up to date changes in the substrate specificity due to interaction partners or PTMs were only shown for PKMTs, which are not very specific regarding the peptide sequence (like SET7/9 (Dhayalan et al. 2011) or PRMTs (Herrmann et al. 2009)). Since SET8 is highly specific towards H4K20, changes in its substrate specificity are not expected. After including all mentioned controls, it appears more likely that Numb is methylated by another not yet identified PKMT.

More examples exist in which the substrate preference was used to identify wrong assignment of methylation substrates to PKMTs. NSD1 was reported to methylate H3K36, H4K20 and the p65 non-histone target (Berdasco et al. 2009; Lu et al. 2010; Rayasam et al. 2003). Kudithipudi et al. (2014) identified the substrate recognition motif (YFG)(VLI)K(QRKNM)(IVP) that differs from the surrounding residues of H4K20 and p65 (Kudithipudi, Lungu, et al. 2014) . Indeed, absence of H4K20 and p65 methylation was confirmed, whereas methylation of H4K44 was detected. The sequence context of H4K44 matches with the sequence preference of NSD1. Sone et al. (2014) described that SUV39H2 catalyze methylation of H2AX at lysine 134 (Sone et al. 2014). It was shown that H2AX methylation imply enhanced phosphorylation of serine at position 139, which led to increased activation of DNA damage repair pathway. Again, a mismatch between the H2AX sequence and SUV39H2 sequence preference revealed a wrong assignment between PKMT and methylated target site (Schuhmacher et al. 2016). Altogether, these findings demonstrate that substrate specificity analyses are important to characterize PKMTs, because they are very predictive in the validation of previously reported PKMT substrates and can guide the discovery of novel PKMT substrates.

5.3. Substrate specificity analysis of MLL family PKMTs

The MLL PKMT family consists of six enzymes, MLL1/2, MLL3/4 and SET1A/B, all catalyzing H3K4 methylation that is associated with actively transcribed genes. Each pair of human paralogs corresponds to one of the three H3K4 methyltransferases *Drosophila melanogaster*, namely SET domain-containing 1 (Set1), Trithorax (Trx) and Trithorax-related (Trr) (Piunti & Shilatifard 2016). The human paralogs are characterized by a high sequence similarity and similar domain organization, which led to their classification (Smith et al.

2011). MLL proteins are found in multi protein complexes in eukaryotic cells. The WRA complex is common for all MLL proteins and is responsible for the stimulation of the methyltransferase activity of the enzyme. DPY30 is also a common member of all MLL complexes, but is not an essential complex partner for the stimulatory effect of the methyltransferase activity (Patel et al. 2009). In addition, each MLL subfamily has specific additional complex partners related to their individual biological roles (Smith et al. 2011). The three subfamilies introduce different methylation states on H3K4 at specific target gene loci as described in 1.12.

As described above, in the past many PKMTs were found to methylate non-histone proteins as well. In this study, MLL1 and MLL3 were selected to be further characterized in this respect. The substrate specificity profile was used for the identification of potential PKMT substrates in proteome wide searches. In addition, somatic cancer mutations were found in PKMTs, which have a central role in carcinogenesis. In previous studies it was shown that somatic cancer mutations have several effects that influence the properties of an enzyme, like activity, product and substrate specificity. Here, the effect of somatic cancer mutations on the properties of MLL1 and MLL3 were investigated.

5.3.1. Substrate specificity analysis of MLL1

In the present study a substrate specificity analysis was performed for MLL1. One aim of this analysis was the identification of novel non-histone targets of this enzyme. Moreover, the analysis allowed to investigate if WRA complex formation influences the substrate recognition and to study the potential effect of somatic cancer mutations in this enzyme on its specificity. With peptide array methylation experiments it was shown that MLL1 specifically discriminates residues of the H3 sequence from A1 to R8. Especially at position -2 and +4 high specificity was identified. Exchange of the canonical arginine by any other amino acid at position -2 completely abolished the methyltransferase activity of MLL1. At position +4, MLL1 also prefers arginine, but weakly accepts cysteine or tyrosine. The crystal structure of MLL1 reveals that R8 forms a water mediated hydrogen bond with serine 3915, which may explain the sequence recognition of this residue (Southall et al. 2009). Using the identified recognition motif, 98 putative non-histone targets were found with the Scansite database and methylation by MLL1 was investigated at the peptide level. Out of the 98 non-histone peptides, 23 targets were methylated. Especially two peptide targets, TICRR and ZNF862, indicated strong methylation signal. 15 of these 23 peptide substrates were selected for protein methylation analysis, 8 of them were successfully cloned, overexpressed and purified.

Unfortunately, while screening the purified non-histone protein methylation, a strong automethylation of MLL1 was observed that overlapped with the potential substrate methylation signal. Still, methylation signals were observed for two proteins, TICRR and ZNF862, corresponding to the strongly methylated peptides mentioned above. Both targets contain an arginine residue at position -2. At position +4, only ZNF862 has the preferred arginine, whereas TICRR has a phenylalanine. Hence, stronger methylation of ZNF862 compared to TICRR illustrates the accuracy of the substrate specificity profile analysis.

In future cellular methylation of these proteins should be investigated and the methylation degree of TICRR and ZNF862 has to be analyzed by MALDI measurements of samples methylated *in vitro* and *in vivo*. The methylation of the additional 98 non-histone peptide targets should also be analyzed in presence of complex partners, to see if different targets are methylated.

5.4. Screening of NHT methylation by the MLL1 wildtype

To further analyze methylation of non-histone proteins by MLL1, the automethylation signal of MLL1 had to be removed or at least reduced. Patel et al. (2014) reported that the automethylation occurs at cysteine 3882, which is located in the SET-I subdomain of the SET domain (Patel et al. 2014). The SET-I region in MLL proteins is differently oriented than in other SET domain containing proteins, because it exists in open conformation where the target lysine cannot be properly oriented for optimal methyl transfer. Cysteine 3882 belongs to the channel tetrapeptide from Cys3882 to Phe3885 that forms one side of the substrate lysine alkyl binding channel, which is shifted away from the SET-C region in the open conformation (Southall et al. 2009). Interestingly, the distance between this cysteine and the AdoMet methyl group is about 10 Å in the open conformation, indicating that the SET-I region must close down during turnover. It was shown that binding of unmodified H3 reduces the automethylation signal, because the substrate bound in the active site pocket prevents an approach of the Cys to the AdoMet. This is in agreement with the results reported in the present thesis. Patel and colleagues reported that an exchange of cysteine 3882 against serine removed most of the automethylation, but it was not diminished completely.

In this study, the automethylation was successfully reduced for MLL1 wildtype by introduction of the C3882S mutation. However, the H3 methylation of the MLL1 C3882S protein was weaker compared to the methylation with the MLL1 wildtype protein. As already mentioned, cysteine 3882 is located in the SET-I lobe within the catalytically active pocket. It is possible that the exchange of cysteine against serine shifts the conformation into the

inactive state, for example, serine is more hydrophilic than cysteine, which may induce an opening of the structure. In presence of complex partners, the methyltransferase activity of MLL1 is getting stimulated. Therefore, in the next step it was investigated if the C3882S mutation influences the enzymatic properties of the MLL1 protein in presence of complex partners. It was shown that the exchange of cysteine against serine did not change the interaction with the complex partners. Interestingly, the methylation signals of H3 were comparable between MLL1 C3882S variants and the MLL1 wildtype protein in presence of WRA, whereas reduced methylation activity was observed for the MLL1 C3882S variant in absence of complex partners as mentioned above. This result additionally supports the suggestion that the C3882S mutation pushes the conformation of the SET-I helix more into the inactive state in absence of complex partners. Apparently, the binding of WRA can overcome this effect. Furthermore, in presence of WRA an increased automethylation signal was detected for the MLL1 protein. This finding is in agreement with the model that WRA stabilize the SET-I helix in the closed conformation and thereby diminishes the distance between C3882 and AdoMet, which is the reason for stronger automethylation.

After successful reduction of the automethylation signal for MLL1, the C3882S mutant was used to investigate the methylation of the selected non-histone proteins in presence or in absence of complex partners. The MLL1 protein with or without WRA methylated the two targets, TICRR and ZNF862, which were already observed before.

Methylation of the target lysine residues of both proteins was verified by mutation of TICRR K1017 and ZNF862 K457 to Arg, both resulting in a complete loss of target methylation.

Next, a direct comparison of the efficiency of MLL1 methylation was performed between the known H3 target and the two novel identified substrates. In the absence of complex partners, H3 was the best target, but in presence of complex partners, ZNF862 was stronger methylated than H3. This means methylation of ZNF862 must be tested in cells and it could be physiologically very relevant.

5.5. Screening of NHT methylation by the somatic cancer mutants of MLL1

In addition, the C3882S mutation was included into the MLL1 somatic cancer mutants R3841W, R3864C and S3865F, to examine if these mutations influence the substrate specificity of MLL1. Reduction of the automethylation signal was detected for MLL1 R3841W and MLL1 S3865F, whereas automethylation could not be removed for MLL1 R3864C. When the newly introduced cysteine 3864 (resulting from the R3864C mutation) was exchanged against glutamine, a loss of automethylation was observed. This result indicated that Cys3864 was a new point of automethylation in the MLL1 R3864C mutant. Similar to MLL1 wildtype, the H3 methylation was weaker for all MLL1 C3882S proteins compared to the methylation with the MLL1 protein variants without C to S mutation. Interestingly, the automethylation signal was comparable between MLL1 R3864C and MLL1 R3864C/C3882S, whereas the H3 methylation was weaker with MLL1 R3864C/C3882S. This result supports the hypothesis that the serine induces an opening of the structure and H3 cannot be properly aligned into the active pocket of MLL1. The influence of the CS mutation on the interaction with the complex partners was further investigated for the cancer mutants as well, but same as with wildtype MLL1 no change was observed.

In the next step, MLL1 R3841W/C3882S and MLL1 S3865F/C3882S were used to screen the MLL1 specific non-histone targets in presence or in absence of WRA. Both proteins with or without complex partners methylated the same targets as MLL1 wildtype. While the target preference of MLL1 S3865F was similar to MLL1 wildtype, MLL1 R3841W/C3882S behaved differently with H3 or NHT in presence of WRA. A reduction on H3 methylation was observed, but methylation of TICRR and ZNF862 was increased in the presence of WRA. Also in presence of complex partners stronger automethylation of MLL1 R3841W was observed, although H3 methylation was reduced. It seems that the interaction of WRA and MLL1 R3841W inhibits the binding or methylation of H3. The mechanism of this inhibition has to be further investigated.

For one of the new MLL1 substrates a biological role has been already reported (Kumagai et al. 2010). TICRR, also known as Treslin, interacts with TopBP1 and regulates the initiation of DNA replication in vertebrates. Perhaps, TICRR also has a role in the transcriptional regulation of genes involved in hematopoiesis and development. Moreover, it is striking that DNA replication initiation is also regulated by chromatin modifications and it is known that open chromatin replicates early and heterochromatin replicates late. About the zinc finger

protein ZNF862 not much is known. It is maybe involved in transcriptional regulation and may support MLL1 in its function. To get further insights into the biological roles of these two novel MLL1 substrates, cellular studies have to be performed. For example, it needs to be investigated if the methylated residue 1017 is involved in the biological functions of TICRR, and also if MLL1 inhibitors would affect TICRR.

5.6. The substrate specificity analysis of MLL3

A substrate specificity profile analysis was performed with MLL3 methyltransferase as well to discover its substrate recognition motif. Since the MLL3 protein is dependent on WRA to achieve detectable methyltransferase activity, peptide array analysis was performed with MLL3 in presence of complex partners. MLL3-WRA recognized the template H3 (1-15) residues from A1 to T6 with a sequence recognition motif of A(ILFP)-R(L)-T(AILMFYV)-K-Q(NHM)-T(AQEILSV). MLL3-WRA is highly specific at position -2, which is in agreement with the crystal structure of MLL3-RA in complex with cofactor product AdoHcy and the H3 peptide (Li et al. 2016). Beside the canonical arginine, it weakly accepts leucine at this place. At position -3, MLL3-WRA prefers hydrophobic residues, whereas at position +1 glutamine can be exchanged by asparagine, histidine and methionine without loss of activity. Although MLL1 and MLL3 belong to the same PKMT family, differences were observed in their specificity profiles (Table 3).

Table 3: The substrate recognition preference of MLL1 and MLL3-WRA

Position	**-3 A1**	**-2 R2**	**-1 T3**	**0 K4**	**+1 Q5**	**+2 T6**	**+3 A7**	**+4 R8**
MLL1 Preference	PRFWY	R	ILFWY	K	RHMFWY	IVAC	X	CY
MLL3-WRA Preference	ILFP	L	AILMFYV	K	NHM	AQEILSV	X	X

First of all, MLL1 recognizes a longer sequence than MLL3. A big difference was detected at position +4, where MLL1 is highly specific towards arginine, cysteine or tyrosine, while MLL3-WRA accepts all 20 natural amino acids. Nevertheless, the complex partners may change the specificity of MLL1, so that differences between the specificity profile arrays of MLL1 and MLL3-WRA has to be regarded as preliminary observation and further analysis is necessary, for example using the MLL1-WRA complex. Using Scansite, 42 potential novel non-histone MLL3 substrates were identified and 24 of them were methylated at the peptide level. One of the methylated non-histone targets was ZNF862, which was already identified

for MLL1. This indicates that overlapping targets of both enzymes exist despite slight differences in the recognition motifs were identified. 10 methylated peptides were selected, and the corresponding proteins cloned, overexpressed and purified. Out of these ten proteins, four non-histone targets were selected for the first round of methylation experiments by MLL3-WRA, but none of them was methylated. The remaining 6 targets still have to be tested for methylation. This finding illustrates that methylation at peptide level cannot always predict methylation at protein level. One reason for this is that peptides synthesized on a cellulose membrane are unfolded and easy to access for enzymes. In contrast, proteins are folded and the target site can be located inside of the protein and thereby prevent methylation. Another possibility is that MLL3 may need the target lysine residue next to N-terminal end, which is not the case in a protein. Further experiments are necessary to identify possible targets for MLL3-WRA. Moreover, cellular experiments on methylation of ZNF862 would be interesting as it is common for MLL1 and MLL3-WRA. In future, a detailed comparison of the substrate specificity arrays between MLL1 and MLL3 is necessary to find more common and differential targets.

5.7. Investigation of the somatic cancer mutations of MLL1 and MLL3

With the improvement of genome-wide sequencing strategies, several somatic mutations in PKMTs were detected in cancer tissues. For many PKMTs (EZH2, GLP, NSD2) it was shown that somatic cancer mutations influence their enzymatic properties and change the chromatin modification patterns introduced by these enzymes (Kudithipudi & Jeltsch 2014). In principle, somatic cancer mutations can lead to gain-of-function or loss-of-function effects. Non-sense or frameshifts in the coding region of genes generally lead to loss-of-function effect which can be understood on the background of the function of the normal gene product. Missense mutations can also lead to loss of enzymatic activity by the disruption of critical enzymatic functions. This was illustrated by the MLL3 N4848S mutant in this thesis. N4848 is located in the NHXC motif in the AdoMet binding pocket of MLL3. The hydrogen bond between asparagine and the cofactor is affected after exchange of asparagine against serine, which explains loss of activity.

In contrast, gain-of-function effects can influence different overall enzymatic properties of PKMTs, like their activity, product pattern or substrate specificity. The understanding of gain-of-function mutations, therefore, requires detailed investigations of affected proteins and in particular of each specific mutation. Better understanding of the effects of somatic cancer mutations in PKMTs will help to understand the role of PKMTs in carcinogenesis and provide

foundations for the development of tumor specific therapies. Somatic cancer mutations can influence the methylation level preference of PKMTs, which was already discovered for EZH2 (Morin et al. 2010; Sneeringer et al. 2010; Yap et al. 2011). EZH2 is the catalytic subunit of the PRC2 complex and it mainly introduces H3K27me1 and –me2 in unmethylated substrates. It is most frequently mutated at Y641, which is associated with a change in the substrate preference and increased H3K27 trimethylation levels (see 1.10). It was observed here, that in a similar manner the MLL3 Y4884C mutation converts the MLL3 enzyme from a monomethyltransferase to a trimethyltransferase. Y4884 is located in the aromatic cage of MLL3 and represents the well-known Phe/Tyr switch position. The tyrosine residue forms a hydrogen bond with the ε-amino group of lysine and prevents trimethylation of H3K4. After the exchange of tyrosine against Cys, the rotation of the lysine ε-carbon-nitrogen bond is possible and further methyl groups can be transferred on the lysine. In addition, the Tyr to Cys exchange increases the size of the active site pocket, allowing to reach higher methylation states. For several PKMTs it was reported that similar mutations of aromatic cage residues affect the substrate preference and product pattern (Collins et al. 2005). The change of the methylation pattern of MLL3 may influence the expression of tumor suppressor genes or oncogenes and thereby could play a role in carcinogenesis.

Additionally, somatic cancer mutations may alter the regulation of PKMTs. This was illustrated by two somatic cancer mutants in MLL1 which are no longer regulated by WRA. MLL1 R3841W and MLL1 R3864C are both hyperactive in isolated form but reduced activity was observed when complex partners were added. R3864 is located in the SET-I region of MLL1, whereas R3841 is located in the SET-N region and forms a main-chain H-bond to AdoMet. Both mutants, R3864C and R3841W, were fully active without complex partners, which suggest that these missense mutations induce conformational changes from the open conformation into a more closed and more active conformation. Interestingly, the addition of complex partners inhibits the methyltransferase activity, likely because an inactive conformation is adopted. Changes in the conformation of MLL1 R3841W compared to MLL1 wildtype were already observed with circular dichroism measurements. Altogether, these results indicate that somatic cancer mutations can influence conserved regulatory mechanisms of PKMTs. This result was further supported by studying the inhibitory effect of the MM-102 MLL1 inhibitor. The MM-102 drug is a specific MLL1 inhibitor that mimics the GSARAE residues in the WIN motif of MLL1 and disrupts the interaction between MLL1 and WDR5 (Karatas et al. 2013). Enzymatic inhibition was detected for MLL1 wildtype and MLL1 S3865F. No effect was observed for MLL1 R3841W and MLL1 R3864C. This is in

agreement with the discovery that the enzymatic activities of these two enzymes are not regulated by the complex partners anymore. For MLL1 wildtype it was shown that MM-102 leads to reduced expression of MLL1 target genes, like HoxA9 and Meis-1, in leukemia cell lines. The data obtained in this study clearly demonstrate that MM-102 has no inhibitory effect when R3864C and R3841W mutations are present in MLL1 proteins. This indicates that MM-102 cannot be used as therapeutic drug in cancers with these MLL1 mutations. Consequently, somatic cancer mutants have to be analyzed in detail to develop personalized therapeutic strategies.

6. Conclusion

This chapter presents a summarized conclusion of all projects. A more detailed conclusion of individual results is presented in the different publications attached in the appendix or in the sections 4.3.1.2.6 and 4.3.2.2.1.

Up to date 80 PKMTs and hundreds of lysine methylation sites in proteins have been identified. However, with the discovery of each single PKMT the following questions appear: How is the enzyme regulated and how many methyl groups can be transferred on its substrate? How does the enzyme recognize its substrate? Can it methylate further substrates? Similarly, the discovery of a lysine methlyation event leads to the following questions: Which PKMT can introduce this modification? And what is the biological effect of this methylation event? Since aberrant methylation is involved in many diseases, it is important to improve our knowledge about each specific PKMT including the investigation of its substrate specificity as an important step in the characterization of PKMTs. In this PhD work, substrate peptide arrays were used to analyze the detailed substrate preferences of four PKMTs, SUV4-20H1, SUV4-20H2, MLL1 and MLL3, revealing the specific influence of each amino acid at every position of the substrate peptide on the enzymatic activity. The investigation of the substrate preference of SUV4-20H1 and SUV4-20H2 showed that even subtle differences in the specificity of these paralogs can be identified with this method. This is important for the understanding of the biological roles of closely related gene paralogs in human cells.

Furthermore, the specificity profile allows the identification of novel substrates for PKMTs and supports the connection of methylation sites with the responsible PKMT in a systematic fashion. By the identification of novel substrates, additional roles of the enzyme can be revealed in various cellular processes. In this thesis, one new protein substrate (CASZ1) was identified for SUV4-20H1 and three substrates (CASZ1, OIP5 and CENPU) for SUV4-20H2. All three proteins were already known to be methylated at the target lysine residue, but the responsible methyltransferase had not been identified. In future, it would be interesting to discover these unknown methyltransferases that catalyze the monomethylation of the identified NHT, which are further used as substrates for the SUV4-20H enzymes. In this respect, comparison between the surrounding sequences of the target lysine residues and the substrate preference of different methyltransferases will offer initial insight into suitable enzymes. Perhaps, SUV4-20H enzymes are forming complexes with the unknown monomethyltransferases, which work together to regulate different regulatory pathways. As already concluded in 4.3.1.2.6, two novel protein substrates have been identified. One of

them, TICRR interacts with TopBP1 and regulates the initiation of DNA replication in vertebrates. In general, methylation of target substrates can influence their function, interaction, properties and localization. It would be interesting to analyze, if the methylation of TICRR influences its interaction with TopBP1 and thereby affect its function in cells. Furthermore, it would be worth to study if novel interaction partners can be identified, which may lead to unknown regulatory mechanisms in cells.

The knowledge of the substrate preference of PKMTs can help to identify problematic assignments between enzyme and reported target sites in the literature and by this guide further work to confirm or disprove described PKMT substrates. The uncovering of wrong methylation events is very important, because lysine methylation signaling is involved in the regulation of many biological processes. Misinterpretation of lysine methylation processes can have various effects on follow-up studies, including misleading therapeutic strategies, since methylation reactions are often involved in diseases.

To advance the understanding of PKMTs it is also important to analyze their product and substrate specificity. In PKMTs, the substrate binding site and cofactor binding site are situated on opposite surfaces of the catalytic SET domain, connected through a hydrophobic tunnel. The product and substrate specificity of a PKMT is dependent on the polarity, shape and geometry of this channel. SET8, for example, is a monomethyltransferase using unmethylated H4K20 as substrate. It contains a tyrosine residue in the active pocket that limits the enzyme to the transfer of one single methyl group. In comparison, in this thesis it was shown that SUV4-20H enzymes transfer one methyl group on monomethylated H4K20 *in vitro*. SUV4-20H enzymes contain a serine residue instead of the tyrosine in their active pocket. Serine is smaller than tyrosine and therefore a monomethylated substrate better fits into the well-defined hydrophobic pocket. Similar to tyrosine, serine also forms a hydrogen bond with the target lysine Nε that limits the rotation of the amino group and hinders the transfer of a second methyl group.

The MLL protein family is another example illustrating the importance of the structural arrangement of the active pocket for the enzymatic activity. In comparison to other highly active SET domain containing proteins, the orientation of the SET-I helix is different in MLL proteins. The binding of the complex partners, WDR5, RBBP5 and ASH2L reorients the SET-I region and completes the lysine binding channel, which strongly stimulates the methyltransferase activity.

In recent years several somatic cancer mutations located in the SET domain of PKMTs were discovered, which were reported to play an important role in carcinogenesis. The exchange of single amino acids can cause a reduction or complete loss of the enzymatic activity, but it can also change the product and substrate specificity or important regulatory mechanisms of the PKMT, which are necessary for the optimal methyltransferase activity. For example, I discovered one MLL1 mutant which was active without the complex members and one MLL3 mutant which introduces trimethylation at H3K4, where the wildtype MLL3 only generates monomethylation. But, why these somatic mutations transform cells into malignancies? Changes in the enzymatic properties can affect multiple pathways in cells. For the development of successful cancer treatment, the first step is to understand the effects of somatic cancer mutations on the enzymatic properties, which was done in the present thesis. In the second step, the cellular effects of somatic cancer mutations have to be analyzed. Only when the altered regulatory mechanisms are known and understood, one can fight against it and develop precise cancer treatment. Inhibitor studies in this thesis with somatic cancer mutants indicated, that some compounds which are highly active with wildtype PKMTs are no longer able to inhibit mutants. Therefore, it is important to investigate the effect of each specific somatic cancer mutant in detail for the development of mutation specific therapeutic strategies.

References

Allis, C.D. et al., 2007. New Nomenclature for Chromatin-Modifying Enzymes. *Cell*, 131(4), pp.633–636.

Allis, C.D.; Caparros, M-L.; Jenuwein,T.; Reinberg,D. editors. Epigenetics Second Edition. New York: *Cold Spring Harbor Laboratory Press*; Chapter 3, page 57; 2015.

Amaral, P.P. et al., 2008. The eukaryotic genome as an RNA machine. *Science (New York, N.Y.)*, 319(5871), pp.1787–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18369136.

Anglin, J.L. & Song, Y., 2013. A medicinal chemistry perspective for targeting histone H3 lysine-79 methyltransferase DOT1L. *Journal of Medicinal Chemistry*, 56(22), pp.8972–8983.

Ansari, K.I., Mishra, B.P. & Mandal, S.S., 2009. MLL histone methylases in gene expression, hormone signaling and cell cycle. *Frontiers in bioscience (Landmark edition)*, 14, pp.3483–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19273288 [Accessed December 22, 2016].

Avdic, V. et al., 2011. Structural and biochemical insights into MLL1 core complex assembly. *Structure (London, England : 1993)*, 19(1), pp.101–8. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0969212610003965 [Accessed December 22, 2016].

Balakrishnan, L. & Milavetz, B., 2010. Decoding the histone H4 lysine 20 methylation mark. *Critical reviews in biochemistry and molecular biology*, 45(5), pp.440–52. Available at: http://informahealthcare.com/doi/abs/10.3109/10409238.2010.504700.

Bannister, A.J. & Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell research*, 21(3), pp.381–395. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3193420&tool=pmcentrez&rendertype=abstract.

Bannister, A.J. & Kouzarides, T., 2005. Reversing histone methylation. *Nature*, 436(7054), pp.1103–1106.

Bannister, A.J., Schneider, R. & Kouzarides, T., 2002. Histone methylation: dynamic or static? *Cell*, 109(7), pp.801–806. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12110177.

Barbieri, C.E. et al., 2012. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nature Genetics*, 44(6), pp.685–689. Available at: http://www.nature.com/doifinder/10.1038/ng.2279 [Accessed November 4, 2016].

Beck, D.B. et al., 2012. PR-set7 and H4K20me1: At the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes and Development*, 26(4), pp.325–337.

Berdasco, M. et al., 2009. Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51), pp.21830–5. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0906831106%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2793312&tool=pmcentrez&rendertype=abstract.

Bernstein, E. & Allis, C.D., 2005. RNA meets chromatin. *Genes & Development*, 19(14), pp.1635–1655. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16024654 [Accessed December 22, 2016].

Bonasio, R., Tu, S. & Reinberg, D., 2010. Molecular Signals of Epigenetic States. *Science*, 330(6004), pp.612–616. Available at: http://www.sciencemag.org/content/330/6004/612%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/21030644%5Cnhttp://www.sciencemag.org/content/330/6004/612.full.pdf%5Cnhttp://www.sciencemag.org/content/330/6004/612.long.

Brower, V., 2011. Unravelling the cancer code. *PLoS Genetics*, pp.6–7.

Brown, M. a et al., 2006. Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. *Molecular cancer*, 5, p.26.

Cao, F. et al., 2014. Targeting MLL1 H3K4 Methyltransferase Activity in Mixed-Lineage Leukemia. *Molecular Cell*, 53(2), pp.247–261.

Chen, C. et al., 2014. MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. *Cancer Cell*, 25(5), pp.652–665. Available at: http://dx.doi.org/10.1016/j.ccr.2014.03.016.

Chen, L.-L., 2016. Ling-Ling Chen: Linking Long Noncoding RNA Processing and Function to RNA Biology. *Trends in Biochemical Sciences*, 0(0), pp.1–12. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0968000416300780.

Cheng, J. et al., 2014. A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. *Molecular Cell*, 53(6), pp.979–992. Available at: http://dx.doi.org/10.1016/j.molcel.2014.02.032.

Chuikov, S. et al., 2004. Regulation of p53 activity through lysine methylation. *Nature*, 432(November), pp.353–360.

Cierpicki, T. et al., 2010. Structure of the MLL CXXC domain–DNA complex and its functional role in MLL-AF9 leukemia. *Nature Structural & Molecular Biology*, 17(1), pp.62–68. Available at: http://www.nature.com/doifinder/10.1038/nsmb.1714 [Accessed October 20, 2016].

Clarke, S.G., 2013. Protein methylation at the surface and buried deep: Thinking outside the histone box. *Trends in Biochemical Sciences*, 38(5), pp.243–252.

Collins, R.E. et al., 2005. In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *Journal of Biological Chemistry*, 280(7), pp.5563–5570.

Copeland, R. a, Solomon, M.E. & Richon, V.M., 2009. Protein methyltransferases as a target class for drug discovery. *Nature reviews. Drug discovery*, 8(9), pp.724–732. Available at: http://dx.doi.org/10.1038/nrd2974.

Couture, J.F. et al., 2005. Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes and Development*, 19(12), pp.1455–1465.

Daniel, J.A., Pray-Grant, M.G. & Grant, P.A., 2005. Effector proteins for methylated histones: An expanding family. *Cell Cycle*, 4(7), pp.919–926.

Daser, A. & Rabbitts, T.H., 2005. The versatile mixed lineage leukaemia gene MLL and its many associations in leukaemogenesis. *Seminars in Cancer Biology*, 15(3), pp.175–188.

Dhami, G. et al., 2013. Dynamic Methylation of Numb by Set8 Regulates Its Binding to p53 and Apoptosis. *Molecular Cell*, 50(4), pp.565–576. Available at: http://dx.doi.org/10.1016/j.molcel.2013.04.028.

Dhayalan, A. et al., 2011. Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. *Chemistry and Biology*, 18(1), pp.111–120. Available at: http://dx.doi.org/10.1016/j.chembiol.2010.11.014.

Dillon, S.C. et al., 2005. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome biology*, 6(8), p.227. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16086857.

Durinck, S. et al., 2011. Temporal dissection of tumorigenesis in primary cancers. *Cancer Discovery*, 1(2), pp.137–143.

Fagard, M. & Vaucheret, H., 2000. (Trans)gene silencing in plants: How many mechanisms? *Annual Review of Plant Biology*, 51, pp.167–194. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-0034485171&partnerID=40&md5=27eadf7e4c0d5363009630bf20fff184.

Feng, Q. et al., 2002. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Current Biology*, 12(12), pp.1052–1058.

Fischle, W., Wang, Y., Jacobs, S.A., et al., 2003. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by polycomb and HP1 chromodomains. *Genes and Development*, 17(15), pp.1870–1881.

Fischle, W., Wang, Y. & Allis, C.D., 2003. Histone and chromatin cross-talk. *Current Opinion in Cell Biology*, 15(2), pp.172–183.

Forbes, S.A. et al., 2011. COSMIC: Mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Research*, 39(SUPPL. 1), pp.945–950.

Frank, R., 2002. The SPOT-synthesis Technqie. Synthethic Peptide Arrays on Membrane Support - Principles and Applications. *J. Immunol. Methods*, 267, pp.13–26.

Fu, W. et al., 2016. Structural basis for substrate preference of SMYD3, a SET domain-containing protein lysine methyltransferase. *Journal of Biological Chemistry*, 291(17), pp.9173–9180.

Gelato, K.A. & Fischle, W., 2008. Role of histone modifications in defining chromatin structure and function. *Biological Chemistry*, 389(4), pp.353–363.

Gitlin, L. & Andino, R., 2003. Nucleic Acid-Based Immune System : the Antiviral Potential of Mammalian RNA Silencing MINIREVIEW Nucleic Acid-Based Immune System : the Antiviral Potential of Mammalian RNA Silencing. , 77(13), pp.7159–7165.

Goldberg, A.D., Allis, C.D. & Bernstein, E., 2007. Epigenetics: A Landscape Takes Shape. *Cell*, 128(4), pp.635–638.

Goll, M.G. & Bestor, T.H., 2005. Eukaryotic Cytosine Methyltransferases. *Annual Review of Biochemistry*, 74(1), pp.481–514. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev.biochem.74.010904.153721.

Herrmann, F. et al., 2009. Human protein arginine methyltransferases in vivo--distinct properties of eight canonical members of the PRMT family. *Journal of cell science*, 122(Pt 5), pp.667–677.

Herz, H.M. et al., 2012. Enhancer-associated H3K4 monomethylation by trithorax-related, the drosophila homolog of mammalian MLL3/MLL4. *Genes and Development*, 26(23), pp.2604–2620.

Hornbeck, P. V. et al., 2015. PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Research*, 43(D1), pp.D512–D520.

Hu, D. et al., 2013. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Molecular and cellular biology*, 33(23), pp.4745–54. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3838007&tool=pmcentrez&rendertype=abstract.

Huang, J. et al., 2007. p53 is regulated by the lysine demethylase LSD1. *Nature*, 449(7158), pp.105–108.

Huang, J. et al., 2006. Repression of p53 activity by Smyd2-mediated methylation. *Nature*, 444(7119), pp.629–632.

Huang, J. & Berger, S.L., 2008. The emerging field of dynamic lysine methylation of non-histone proteins. *Current Opinion in Genetics and Development*, 18(2), pp.152–158.

Jeltsch, A. & Jurkowska, R.Z., 2016. Allosteric control of mammalian DNA methyltransferases – a new regulatory paradigm. *Nucleic Acids Research*, (23), p.gkw723. Available at: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkw723.

Jenuwein, T., 2006. The epigenetic magic of histone lysine methylation: Delivered on 6 July 2005 at the 30th FEBS Congress in Budapest, Hungary. *FEBS Journal*, 273(14), pp.3121–3135.

Jenuwein, T. & Allis, C.D., 2001. Thomas Jenuwein 1 and C. David Allis 2. *Science*, 293(August), pp.1074–1081.

Jin, B. & Robertson, K.D., 2013. Epigenetic Alterations in Oncogenesis. *Advances in experimental medicine and biology*, 754, pp.3–29. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3707278&tool=pmcentrez&rendertype=abstract.

Jørgensen, S., Schotta, G. & Sørensen, C.S., 2013. Histone H4 Lysine 20 methylation: Key player in epigenetic regulation of genomic integrity. *Nucleic Acids Research*, 41(5), pp.2797–2806.

Jurkowska, R.Z., Jurkowski, T.P. & Jeltsch, A., 2011. Structure and Function of Mammalian DNA Methyltransferases. *ChemBioChem*, 12(2), pp.206–222.

Kaikkonen, M.U., Lam, M.T.Y. & Glass, C.K., 2011. Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular Research*, 90(3), pp.430–440.

Kandoth, C. et al., 2013. Mutational landscape and significance across 12 major cancer types. *Nature*, 502(7471), pp.333–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3927368&tool=pmcentrez&rendertype=abstract.

Karatas, H. et al., 2013. High-A ffi nity, Small-Molecule Peptidomimetic Inhibitors of MLL1/WDR5 Protein − Protein Interaction. *Journal of the American Chemical Society*.

Kouzarides, T., 2007. Chromatin Modifications and Their Function. *Cell*, 128(4), pp.693–705.

Krivtsov, A. V. & Armstrong, S.A., 2007. MLL translocations, histone modifications and leukaemia stem-cell development. *Nature Reviews Cancer*, 7(11), pp.823–833. Available at: http://www.nature.com/nrc/journal/v7/n11/full/nrc2253.html%5Cnhttp://www.nature.com/nrc/journal/v7/n11/pdf/nrc2253.pdf.

Kudithipudi, S., Kusevic, D., et al., 2014. Specificity analysis of protein lysine methyltransferases using SPOT peptide arrays. *Journal of visualized experiments : JoVE*, di(93), pp.e52203–e52203. Available at: http://www.jove.com/video/52203/specificity-analysis-protein-lysine-methyltransferases-using-spot%5Cnpapers3://publication/doi/10.3791/52203.

Kudithipudi, S., Lungu, C., et al., 2014. Substrate specificity analysis and novel substrates of the protein lysine methyltransferase NSD1. *Chemistry and Biology*, 21(2), pp.226–237. Available at: http://dx.doi.org/10.1016/j.chembiol.2013.10.016.

Kudithipudi, S. et al., 2012. The SET8 H4K20 protein lysine methyltransferase has a long recognition sequence covering seven amino acid residues. *Biochimie*, 94(11), pp.2212–2218. Available at: http://dx.doi.org/10.1016/j.biochi.2012.04.024.

Kudithipudi, S. & Jeltsch, A., 2016. Approaches and Guidelines for the Identification of Novel Substrates of Protein Lysine Methyltransferases. *Cell Chemical Biology*, 9, pp.1–7. Available at: http://linkinghub.elsevier.com/retrieve/pii/S2451945616302434.

Kudithipudi, S. & Jeltsch, A., 2014. Role of somatic cancer mutations in human protein lysine methyltransferases. *Biochimica et Biophysica Acta - Reviews on Cancer*, 1846(2), pp.366–379. Available at: http://dx.doi.org/10.1016/j.bbcan.2014.08.002.

Kumagai, A. et al., 2010. Treslin Collaborates with TopBP1 in Triggering the Initiation of DNA Replication. *Cell*, 140(3), pp.349–359. Available at: http://dx.doi.org/10.1016/j.cell.2009.12.049.

Kuo, A.J. et al., 2012. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature*, 484(7392), pp.115–9. Available at: http://www.nature.com.libproxy.uoregon.edu/nature/journal/v484/n7392/full/nature10956.html.

Kurash, J.K. et al., 2008. Methylation of p53 by Set7/9 Mediates p53 Acetylation and Activity In Vivo. *Molecular Cell*, 29(3), pp.392–400.

Lanouette, S. et al., 2014. The functional diversity of protein lysine methylation. *Molecular Systems Biology*, 10(4), pp.1–26.

Li, Y. et al., 2016. Structural basis for activity regulation of MLL family methyltransferases. *Nature*, 530(7591), pp.447–452. Available at: http://dx.doi.org/10.1038/nature16952.

Lu, T. et al., 2010. Regulation of NF-kappaB by NSD1/FBXL11-dependent reversible lysine methylation of p65. *Proceedings of the National Academy of Sciences of the United States of America*, 107(1), pp.46–51. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0912493107%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2806709&tool=pmcentrez&rendertype=abstract.

Luger, K. et al., 1997. Crystal structure of the nucleosome core particle at 2.8|[thinsp]||[Aring]| resolution. *Nature*, 389(6648), p.251. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9305837%5Cnpapers3://publication/doi/doi:10.1038/38444.

Macrini, C.M.T. et al., 2003. MLL AT-hook sequence is strongly conserved in infant acute leukemia with or without MLL gene rearrangement. *Leukemia*, 17(7), pp.1432–3. Available at: http://dx.doi.org/10.1038/sj.leu.2402966.

Margueron, R., Trojer, P. & Reinberg, D., 2005. The key to development: Interpreting the histone code? *Current Opinion in Genetics and Development*, 15(2), pp.163–176.

Martin, C. & Zhang, Y., 2005. The diverse functions of histone lysine methylation. *Nature Rev.*, 6(11), pp.838–49.

Mas-y-Mas, S. et al., 2016. The Human Mixed Lineage Leukemia 5 (MLL5), a Sequentially and Structurally Divergent SET Domain-Containing Protein with No Intrinsic Catalytic Activity. *Plos One*, 11(11), p.e0165139. Available at: http://dx.plos.org/10.1371/journal.pone.0165139.

Miremadi, A. et al., 2007. Cancer genetics of epigenetic genes. *Human Molecular Genetics*, 16(R1).

Morin, R.D. et al., 2010. Somatic mutation of EZH2 (Y641) in Follicular and Diffuse Large B-cell Lymphomas of Germinal Center Origin. *Nature genetics*, 42(2), pp.181–185.

Muntean, A.G. & Hess, J.L., 2012. The pathogenesis of mixed-lineage leukemia. *Annual review of pathology*, 7, pp.283–301. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22017583%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3381338.

Murray, K., 1964. The Occurrence of iε-N-Methyl Lysine in Histones. *Biochemistry*, 3(1), pp.10–15. Available at: http://dx.doi.org/10.1021/bi00889a003.

Muzny, D.M. et al., 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407), pp.330–337. Available at: http://dx.doi.org/10.1038/nature11252.

Obenauer, J.C., Cantley, L.C. & Yaffe, M.B., 2003. Scansite 2.0: Proteome-wide prediction of cell signalling interactions using short sequence motifs. *Nucleic Acids Research*, 31(13), pp.3635–3641.

Oda, H. et al., 2009. Monomethylation of histone H4-lysine 20 is involved in chromosome structure and stability and is essential for mouse development. *Molecular and cellular biology*, 29(8), pp.2278–2295.

Patel, A. et al., 2008. A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. *Journal of Biological Chemistry*, 283(47), pp.32162–32175.

Patel, A. et al., 2014. Automethylation activities within the mixed lineage leukemia-1 (MLL1) core complex reveal evidence supporting a "two-active site" model for multiple histone H3 lysine 4 methylation. *Journal of Biological Chemistry*, 289(2), pp.868–884.

Patel, A. et al., 2009. On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. *Journal of Biological Chemistry*, 284(36), pp.24242–24256.

Patel, D.J. & Wang, Z., 2013. Readout of epigenetic modifications. *Annual review of biochemistry*, 82, pp.81–118. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4696766&tool=pmcentrez&rendertype=abstract.

Piunti, A. & Shilatifard, A., 2016. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science*.

Portela, A. & Esteller, M., 2010. Epigenetic modifications and human disease. *Nature biotechnology*, 28(10), pp.1057–1068. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20944598%5Cnhttp://dx.doi.org/10.1038/nbt.1685.

Qian, C. & Zhou, M.M., 2006. SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cellular and Molecular Life Sciences*, 63(23), pp.2755–2763.

Rathert, P., Zhang, X., et al., 2008. Analysis of the Substrate Specificity of the Dim-5 Histone Lysine Methyltransferase Using Peptide Arrays. *Chemistry and Biology*, 15(1), pp.5–11.

Rathert, P., Dhayalan, A., Murakami, M., et al., 2008. Protein lysine methyltransferase G9a acts on non-histone targets. *Nature chemical biology*, 4(6), pp.344–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18438403%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2696268.

Rathert, P., Dhayalan, A., Ma, H., et al., 2008. Specificity of protein lysine methyltransferases and methods for detection of lysine methylation of non-histone proteins. *Molecular BioSystems*, 4(12), p.1186. Available at: http://xlink.rsc.org/?DOI=b811673c.

Rayasam, G. V. et al., 2003. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *The EMBO Journal*, 22(12), pp.3153–3163. Available at: http://emboj.embopress.org/cgi/doi/10.1093/emboj/cdg288.

Rea, S. et al., 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406(6796), pp.593–599.

Del Rizzo, P.A. & Trievel, R.C., 2014. Molecular basis for substrate recognition by lysine methyltransferases and demethylases. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1839(12), pp.1404–1415.

Rousseaux, S. & Khochbin, S., 2015. Histone Acylation beyond Acetylation: Terra Incognita in Chromatin Biology. *Cell Journal*, 17(1), pp.1–6.

Saddic, L.A. et al., 2010. Methylation of the retinoblastoma tumor suppressor by SMYD2. *Journal of Biological Chemistry*, 285(48), pp.37733–37740.

Schotta, G. et al., 2008. A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes and Development*, 22(15), pp.2048–2061.

Schotta, G. et al., 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *GENES & DEVELOPMENT*, 18(1), pp.1251–1262.

Schuhmacher, M.K., Kudithipudi, S. & Jeltsch, A., 2016. Investigation of H2AX methylation by the SUV39H2 protein lysine methyltransferase. *FEBS Letters*, 590(12), pp.1713–1719. Available at: http://doi.wiley.com/10.1002/1873-3468.12216.

Shi, X. et al., 2007. Modulation of p53 Function by SET8-Mediated Methylation at Lysine 382. *Molecular Cell*, 27(4), pp.636–646.

Shi, Y. et al., 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*, 119(7), pp.941–953.

Shilatifard, A., 2012. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Biochemistry*, 81(1), pp.65–95. Available at: http://pubget.com/paper/22663077?institution=utsouthwestern.edu%5Cnpapers3://publication/doi/10.1146/annurev-biochem-051710-134100.

Smith, B.C. & Denu, J.M., 2009. Chemical mechanisms of histone lysine and arginine modifications. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1789(1), pp.45–57. Available at: http://dx.doi.org/10.1016/j.bbagrm.2008.06.005.

Smith, E., Lin, C. & Shilatifard, A., 2011. The super elongation complex (SEC) and MLL in development and disease. *Genes & development*, 25(7), pp.661–672.

Smith, E., Lin, C. & Shilatifard, A., 2011. The super elongation complex (SEC) and MLL in development and disease. *Genes & Development*, 25(7), pp.661–672. Available at: http://genesdev.cshlp.org/cgi/doi/10.1101/gad.2015411 [Accessed December 23, 2016].

Sneeringer, C.J. et al., 2010. Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proceedings of the National Academy of Sciences of the United States of America*, 107(49), pp.20980–5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3000297&tool=pmcentrez&rendertype=abstract.

Sone, K. et al., 2014. Critical role of lysine 134 methylation on histone H2AX for γ-H2AX production and DNA repair. *Nature communications*, 5, p.5691. Available at: http://www.nature.com/ncomms/2014/141209/ncomms6691/full/ncomms6691.html.

Song, F. et al., 2014. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science*, 344(6182), pp.376–380. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24763583.

Southall, S.M. et al., 2009. Structural Basis for the Requirement of Additional Factors for MLL1 SET Domain Activity and Recognition of Epigenetic Marks. *Molecular Cell*, 33(2), pp.181–191. Available at: http://dx.doi.org/10.1016/j.molcel.2008.12.029.

Southall, S.M., Cronin, N.B. & Wilson, J.R., 2014. A novel route to product specificity in the Suv4-20 family of histone H4K20 methyltransferases. *Nucleic acids research*, 42(1), pp.661–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24049080 [Accessed October 19, 2016].

Strahl, B.D. & Allis, C.D., 2000. The language of covalent histone modifications. *Nature*, 403(6765), pp.41–45.

Taverna, S.D. et al., 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Structural & Molecular Biology*, 14(11), pp.1025–1040. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=17984965&retmode=ref&cmd=prlinks.

Thinnes, C.C. et al., 2014. Targeting histone lysine demethylases - Progress, challenges, and the future. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1839(12), pp.1416–1432.

Trojer, P. & Reinberg, D., 2007. Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Molecular Cell*, 28(1), pp.1–13.

Tsukada, Y. et al., 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature*, 439(February), pp.811–816.

Upadhyay, A.K. & Cheng, X., 2011. Dynamics of histone lysine methylation: structures of methyl writers and erasers. *Progress in drug research. Fortschritte der Arzneimittelforschung. Progres des recherches pharmaceutiques*, 67, pp.107–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21141727 [Accessed December 22, 2016].

Vaute, O. et al., 2002. Functional and physical interaction between the histone methyl transferase Suv39H1 and histone deacetylases. *Nucleic acids research*, 30(2), pp.475–481.

Vougiouklakis, T. et al., 2015. SUV420H1 enhances the phosphorylation and transcription of ERK1 in cancer cells. *Oncotarget*, 6(41), pp.43162–71. Available at: http://www.oncotarget.com/abstract/6351 [Accessed December 22, 2016].

Waddington, C.H., 2012. The epigenotype. 1942. *International journal of epidemiology*, 41(1), pp.10–13.

Waterhouse, P.M., Wang, M.B. & Finnegan, E.J., 2001. Role of short RNAs in gene silencing. *Trends in Plant Science*, 6(7), pp.297–301.

Watson, I. et al., 2013. Emerging patterns of somatic mutations in cancer. *Nature Reviews Genetics*, 14(10), pp.703–718. Available at: http://www.nature.com/nrg/journal/vaop/ncurrent/full/nrg3539.html?WT.mc_id=TWT_NatureRevGenet.

West, L.E. et al., 2010. The MBT repeats of L3MBTL1 link SET8-mediated p53 methylation at lysine 382 to target gene repression. *Journal of Biological Chemistry*, 285(48), pp.37725–37732.

Wu, H. et al., 2013. Crystal structures of the human histone H4K20 methyltransferases SUV420H1 and SUV420H2. *FEBS Letters*, 587(23), pp.3859–3868. Available at: http://dx.doi.org/10.1016/j.febslet.2013.10.020.

Wysocka, J. et al., 2005. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*, 121(6), pp.859–872.

Xiao, B., Wilson, J.R. & Gamblin, S.J., 2003. SET domains and histone methylation. *Current Opinion in Structural Biology*, 13(6), pp.699–705.

Xu, Y.M., Du, J.Y. & Lau, A.T.Y., 2014. Posttranslational modifications of human histone H3: An update. *Proteomics*, 14(17–18), pp.2047–2060.

Yap, D.B. et al., 2011. Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood*, 117(8), pp.2451–2459.

Yun, M. et al., 2011. Readers of histone modifications. *Cell research*, 21(4), pp.564–78. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3131977&tool=pmcentrez&rendertype=abstract.

Zhang, X. & Bruice, T.C., 2008. Product specificity and mechanism of protein lysine methyltransferases: Insights from the histone lysine methyltransferase SET8. *Biochemistry*, 47(25), pp.6671–6677.

Zhang, X., Huang, Y. & Shi, X., 2015. Emerging roles of lysine methylation on non-histone proteins. *Cellular and Molecular Life Sciences*, 72(22), pp.4257–4272.

Zhang, X., Wen, H. & Shi, X., 2012. Lysine methylation : beyond histones Modifying Enzymes for Lysine Methylation. *Acta Biochim Biophys Sin*, 44, pp.14–27.

Zhao, Y. & Garcia, B.A., 2015. Comprehensive catalog of currently documented histone modifications. *Cold Spring Harbor Perspectives in Biology*, 7(9).

Appendix I

Table 4:

List of potential novel non-histone targets of MLL1

Table 5:

List of potential novel non-histone targets of MLL3+WRA

Appendix II (not included in the published thesis)

Manuscript 1:

Weirich S, Kudithipudi S, Jeltsch A. (2016*).* **Specificity of the Suv4-20H1 and Suv4-20H2 protein lysine methyltransferases and methylation of novel substrates,** *Journal of Molecular Biology, 2016 Jun 5;428(11):2344-58. doi: 10.1016/j.jmb.2016.04.015. Epub 2016 Apr 20.*

Manuscript 2:

Weirich S, Kusevic D, Kudithipudi S & Jeltsch A. (2015). **Investigation of the methylation of Numb by the SET8 protein lysine methyltransferase**, *Scientific Reports*, *2015 Sep 22;5:13813. doi: 10.1038/srep13813.*

Manuscript 3:

Weirich S, Kudithipudi S & Jeltsch A. (2017). **Somatic cancer mutations in MLL1 histone methyltransferase modulate its enzymatic activity and dependence on the WDR5/RBBP5/ASH2L complex**, *Molecular Oncology. 2017 Apr;11(4):373-387. doi: 10.1002/1878-0261.12041. Epub 2017 Mar 10.*

Manuscript 4:

Weirich S, Kudithipudi S, Kycia I & Jeltsch A. (2015). **Somatic cancer mutations in the MLL3-SET domain alter the catalytic properties of the enzyme**, *Clinical Epigenetics*, *2015 Mar 28;7:36. doi: 10.1186/s13148-015-0075-3.*

Table 4: List of potential novel non-histone targets identified in Scansite search using the recognition motif of MLL1. The predicted target lysine is labeled in red

	Swiss Prot. Nr.	Original name	Name	Sequence	Protein-length	MW (kDa)	Target K position
A3	Q9BR61	ACBD6_HUMAN	ACBD6 = Acyl-CoA-binding domain-containing protein 6	LYLYARYKQVKVGNC	282	31	73
A4	O00767	ACOD_HUMAN	SCD = Acyl-CoA desaturase	AAILARIKRTGDGNY	359	42	349
A5	Q96LA8	ANM6_HUMAN	PRMT6 = Protein arginine N-methyltransferase 6	SVLHARTKWLKEGGL	375	42	178
A6	O15085	ARHGB_HUMAN	ARHGEF11 = Rho guanine nucleotide exchange factor 11	QIHDYRTKRTLGLGS	1522	168	422
A7	P35613	BASI_HUMAN	BSG = Basigin	LTRAPRVKWVRAQAV	385	42	126
A8	Q8TDC3	BRSK1_HUMAN	BRSK1 = Serine/threonine-protein kinase BRSK1	SGPSRRFKRVVETIQ	778	85	703
A9	Q8IWQ3	BRSK2_HUMAN	BRSK2 = Serine/threonine-protein kinase BRSK2	SGPSRRFKRVVETIQ	736	82	623
A10	Q7L1Q6	BZW1_HUMAN	BZW1 = Basic leucine zipper and W2 domain-containing protein 1	NKLIRRYKYLEKGFE	419	48	231
A11	Q9Y6E2	BZW2_HUMAN	BZW2 = Basic leucine zipper lomain-containing protein 2	NKLIRRYKYLEKAFE	419	48	123
A12	P11586	C1TC_HUMAN	MTHFD1= C-1-tetrahydrofolate synthase, cytoplasmic	DIQIRRLKRLGIEKT	935	102	498
A13	Q5T3J3	CA103_HUMAN	LRIF1 = Ligand-dependent nuclear receptor-interacting factor 1	DEKIRRLKQVLREKE	769	85	749
A14	Q9H159	CAD19_HUMAN	CDH19 = Cadherin-19	NELGPRFKRLACMFG	772	87	759
A15	P55286	CADH8_HUMAN	CDH8= Cadherin-8	SDWGPRFKRLGELYS	799	88	784
A16	A6NFT4	CC42B_HUMAN	CCDC42B = Coiled-coil domain-containing protein 42B	ARLQRRLKRLEPCAR	308	36	138
A17	P22674	CCNO_HUMAN	CCNO= Cyclin-O	EVHPPRVKQLLALCC	350	38	198
A18	P20273	CD22_HUMAN	CD22= B-cell receptor CD22	LKLQRRWKRTQSQQG	847	95	713
A19	Q03188	CENPC_HUMAN	CENPC1= Centromere protein C 1	TPNVRRTKRTRLKPL	943	107	742
A20	Q52M58	CN177_HUMAN	C14orf177 =Putative uncharacterized protein C14orf177	FCATPRFKQLFKGTV	125	14	112
A21	Q2M3C6	CO027_HUMAN	C15orf27= Transmembrane protein C15orf27	MLRIWRVKRVIDAYV	531	58	213
B1	Q7Z2Z1	CO042_HUMAN	TICRR= Treslin	LRRSPRIKQLSFSRT	1910	211	1019
B2	O15263	DEFB2_HUMAN	DEFB4A= Beta-defensin 4A	VFCPRRYKQIGTCGL	64	7	48
B3	Q6QHC5	DEGS2_HUMAN	DEGS2 = Sphingolipid delta(4)-desaturase/C4-hydroxylase DES2	LGPYARVKRVYRLAK	323	37	313
B4	Q16760	DGKD_HUMAN	DGKD= Diacylglycerol kinase delta	VKFLRRFKQLLNPAQ	1214	135	343
B5	Q86XP1	DGKH_HUMAN	DGKH= Diacylglycerol kinase eta	VKFLRRFKQLLNPAQ	1220	135	354
B6	Q9P265	DIP2B_HUMAN	DIP2B= Disco-interacting protein 2 homolog B	FKGWPRLKWVVTDSK	1576	171	486
B7	O95672	ECEL1_HUMAN	ECEL1= Endothelin-converting enzyme-like 1	EHPLPRLKYTHDQLF	775	88	704
B8	Q14156	EFR3A_HUMAN	EFR3A =Protein EFR3 homolog A	SALRPRYKRLVDNIF	821	92.924	17
B9	Q9Y2G0	EFR3B_HUMAN	EFR3B= Protein EFR3 homolog B	GALRPRYKRLVDNIF	817	92	16
B10	Q14674	ESPL1_HUMAN	ESPL1= Separin	ARQAPRLKYLIGAAP	2120	233	2103
B11	Q9NYZ1	FA18B_HUMAN	FAM18B1= Protein FAM18B1	ALFSFRVKWLAVVIM	205	24	152
B12	O60258	FGF17_HUMAN	FGF17= Fibroblast growth factor 17	SAPTRRTKRTRRPQP	216	25	207
B13	Q2V2M9	FHOD3_HUMAN	FHOD3= FH1/FH2 domain-containing protein 3	QAREERYKYLEQLAA	1422	159	484
B14	P42685	FRK_HUMAN	FRK= Tyrosine-protein kinase FRK	VVKHYRIKRLDEGGF	505	58	168

B15	Q9BTV5	FSD1_HUMAN	FSD1= Fibronectin type III and SPRY domain-containing protein 1	SFYNARTKQVLHTFK	496	56	441
B16	A6NG13	GL54D_HUMAN	Glycosyltransferase 54 domain-containing protein	KLASWRIKQVLDFCI	374	44	233
B17	Q5T7V8	GORAB_HUMAN	GORAB = RAB6-interacting golgin	EEELRRLKQTKDPFE	394	45	42
B18	Q8TAX9	GSDMB_HUMAN	GSDMB= Gasdermin-B	RVLSYRVKQLVFPNK	411	47	210
B19	O00219	HAS3_HUMAN	HAS3= Hyaluronan synthase 3	LSLGYRTKYTARSKC	553	63	332
B20	P51610	HCFC1_HUMAN	HCFC1= Host cell factor 1	VLLQPRWKRVVGWSG	2035	209	20
B21	Q03164	HRX_HUMAN	Histone-lysine N-methyltransferase MLL	ARSNARLKQLSFAGV	3969	432	3707
C1	Q7Z6Z7	HUWE1_HUMAN	HUWE1= E3 ubiquitin-protein ligase HUWE1	PAMAARIKQIKPLLS	4374	482	1047
C2	Q01113	IL9R_HUMAN	IL9R= Interleukin-9 receptor	FKLSPRVKRIFYQNV	521	57	298
C3	P51617	IRAK1_HUMAN	IRAK1= Interleukin-1 receptor-associated kinase 1	KTHGARTKYLKDLVE	712	77	432
C4	P38570	ITAE_HUMAN	ITGAE= Integrin alpha-E	LVTSPRTKRTPGPLH	1179	130	61
C5	O14713	ITBP1_HUMAN	ITGB1BP1= Integrin beta-1-binding protein 1	TCAEFRIKYVGAIEK	200	22	66
C6	Q9UPP5	K1107_HUMAN	KIAA1107= Uncharacterized protein KIAA1107	IISLPRVKWTEAALT	1409	156	14
C7	Q2LD37	K1109_HUMAN	KIAA1109= Uncharacterized protein KIAA1109	ETHDARTKRLWFLWP	5005	555	1107
C8	O15550	KDM6A_HUMAN	KDM6A= Lysine-specific demethylase 6A	SALAARIKYLQAQLC	1401	154	395
C9	Q96QE4	LR37B_HUMAN	LRRC37B= Leucine-rich repeat-containing protein 37B	LVQLPRLKWVQTTDL	947	106	153
C10	P22888	LSHR_HUMAN	LHCGR= Lutropin-choriogonadotropic hormone receptor	FINLPRLKYLSICNT	699	79	126
C11	O00453	LST1_HUMAN	LST1= Leukocyte-specific transcript 1 protein	CWLHRRVKRLERSWA	97	11	39
C12	Q7L590	MCM10_HUMAN	MCM10= Protein MCM10 homolog	VPALPRTKRVARTPK	875	98	197
C13	O60318	MCM3A_HUMAN	MCM3AP= 80 kDa MCM3-associated protein	IMRQARVKRTDLDKA	1980	218	623
C14	P30307	MPIP3_HUMAN	CDC25C= M-phase inducer phosphatase 3	NLNRPRLKQVEKFKD	473	53	227
C15	Q5VT25	MRCKA_HUMAN	CDC42BPA= Serine/threonine-protein kinase MRCK alpha	EAYERRIKRLEQEKL	1732	197	449
C16	P35749	MYH11_HUMAN	MYH11= Myosin-11	EKGNARVKQLKRQLE	1972	227	1880
C17	Q9NZM1	MYOF_HUMAN	MYOF= Myoferlin	FIVWRRFKWVIIGLL	2061	235	2025
C18	O14594	NCAN_HUMAN	NCAN= Neurocan core protein	ARDAPRIKWTKVRTA	1321	143	77
C19	P17677	NEUM_HUMAN	GAP43= Neuromodulin	LCCMRRTKQVEKNDD	238	25	9
C20	Q99784	NOE1_HUMAN	OLFM1=Noelin	CSRDARTKQLRQLLE	485	55	92
C21	Q6IF63	O52W1_HUMAN	OR52W1= Olfactory receptor 52W1	LIYGARTKQIRDRLL	320	34	303
D1	Q8IXS6	PALM2_HUMAN	PALM2= Paralemmin-2	EEDEFRVKQLEDNIQ	379	42	80
D2	Q9NPG4	PCD12_HUMAN	PCDH12=Protocadherin-12	ELGHFRLKRTNGNTY	1184	129	403
D3	Q96GW7	PGCB_HUMAN	BCAN= Brevican core protein	VLGSPRVKWTFLSRG	911	99	78
D4	Q13371	PHLP_HUMAN	PDCL= Phosducin-like protein	INDWRRFKQLETEQR	301	34	70
D5	P11086	PNMT_HUMAN	PNMT= Phenylethanolamine N-methyltransferase	RQLRARVKRVLPIDV	282	31	152
D6	O43900	PRIC3_HUMAN	PRICKLE3= Prickle-like protein 3	PGEKYRIKQLLHQLP	615	69	130
D7	Q09MP3	R51A2_HUMAN	RAD51AP2= RAD51-associated protein 2	LSRKARIKQLHPYLK	1159	134	1141
D8	Q96S59	RANB9_HUMAN	RANBP9= Ran-binding protein 9	KELQRRLKRLYPAVD	729	78	152
D9	Q86WA9	S2611_HUMAN	SLC26A11= Sodium-independent sulfate anion transporter	FRTLWRVKRLDLLPL	606	65	431

D10	P78381	S35A2_HUMAN	SLC35A2= UDP-galactose translocator	SAAHRRLKYISLAVL	396	41	36
D11	Q96H72	S39AD_HUMAN	SLC39A13= Zinc transporter ZIP13	EAGAWRLKQLLSFAL	371	39	107
D12	Q9UHD8	SEPT9_HUMAN	SEPT9= Septin-9	HFEAYRVKRLNEGSS	586	65	563
D13	A8K8P3	SFI1_HUMAN	SFI1= Protein SFI1 homolog	YTQKRRYKQLLQARA	1242	148	470
D14	Q9Y5B9	SP16H_HUMAN	SUPT16H= FACT complex subunit SPT16	DAYYRRVKRLYSNWR	1047	120	15
D15	Q9Y5Y6	ST14_HUMAN	ST14= Suppressor of tumorigenicity 14 protein	GVQERRLKRIISHPF	855	95	695
D16	P47897	SYQ_HUMAN	QARS= Glutamine--tRNA ligase	DPVAYRVKYTPHHRT	775	88	421
D17	A6NC51	T150B_HUMAN	TMEM150B= Transmembrane protein 150B	QLLLWRLKRLPQPGA	233	26	141
D18	Q9NYW4	TA2R5_HUMAN	TAS2R5= Taste receptor type 2 member 5	IMGIPRVKQTCQKIL	299	35	280
D19	O95759	TBCD8_HUMAN	TBC1D8= TBC1 domain family member 8	EALLARLKQVHANHP	1140	131	392
D20	Q3YBR2	TBRG1_HUMAN	TBRG1= Transforming growth factor beta regulator 1	QNEKYRLKYLRLRKA	411	45	37
D21	Q96BS2	TESC_HUMAN	TESC= Calcineurin B homologous protein 3	EQLHRRFKQLSGDQP	214	25	33
E1	Q6NT04	TIGD7_HUMAN	TIGD7= Tigger transposable element-derived protein 7	CKRLYRWKQLEESL	549	63	345
E2	O94876	TMCC1_HUMAN	TMCC1= Transmembrane and coiled-coil domains protein 1	KQQAARIKQVFEKKN	653	72	278
E3	Q96QT4	TRPM7_HUMAN	TRPM7= Transient receptor potential cation channel subfamily M member 7	ELHPRIKQLLGKGL	1865	213	153
E4	Q92574	TSC1_HUMAN	TSC1= Hamartin	ELDPRRWKRLETHDV	1164	130	248
E5	Q5VYS8	TUT7_HUMAN	ZCCHC6= Terminal uridylyltransferase 7	SAIDPRVKYLCYTMK	1495	171	1145
E6	O75631	UPK3A_HUMAN	UPK3A= Uroplakin-3a	AATEYRFKYVLVNMS	287	31	165
E7	Q502W6	VWA3B_HUMAN	VWA3B= von Willebrand factor A domain-containing protein 3B	ARIRRRIKWLQDGSQ	1294	146	491
E8	Q64LD2	WDR25_HUMAN	WDR25= WD repeat-containing protein 25	PLA AARFKQVKLSRN	544	60	135
E9	A8MVM7	YD021_HUMAN	Putative uncharacterized protein ENSP00000382790	NLAIARLKRTEFKRL	634	73	293
E10	P49750	YLPM1_HUMAN	YLPM1= YLP motif-containing protein 1	PEEDARLKQLQAAAA	1951	220	381
E11	O96006	ZBED1_HUMAN	ZBED1= Zinc finger BED domain-containing protein 1	TFLDPRYKRLPFLSA	694	78	480
E12	P15822	ZEP1_HUMAN	HIVEP1= Zinc finger protein 40	MPRTKQIHPR NL	2718	297	5
E13	Q63HK3	ZKSC2_HUMAN	ZKSCAN2= Zinc finger protein with KRAB and SCAN domains 2	MIPVPRLKRIAISAK	967	111	448
E14	Q9UBW7	ZMYM2_HUMAN	ZMYM2= Zinc finger MYM-type protein 2	EDYLWRIKQLGSHSP	1377	155	1215
E15	Q8TD17	ZN398_HUMAN	ZNF398= Zinc finger protein 398	CGRSFRYKQTLKDHL	642	71	579
E16	O60290	ZN862_HUMAN	ZNF862= Zinc finger protein 862	GDGPRRIKRTYRPRS	1169	132	457

Table 5: List of potential novel non-histone targets identified in Scansite search using the recognition motif of MLL3-WRA. The predicted target lysine is labeled in red

	Swiss Prot. Nr.	Original name	Name	Sequence	Protein-length	Mw (kDa)	Target K position
A3	Q12873	CHD3_HUMAN	Chromodomain-helicase-DNA-binding protein 3	DGPVRTKKLKRGRPG	2000	227	345
A4	P15336	ATF2_HUMAN	Cyclic AMP-dependent transcription factor ATF-2	PDEKRRKFLERNRAA	505	55	357
A5	Q02930	CREB5_HUMAN	Cyclic AMP-responsive element-binding protein 5	PDERRRKFLERNRAA	508	57	380
A6	Q8WYP5	ELYS_HUMAN	Protein ELYS	DKQLRIKHVRRVRGR	2266	253	1933
A7	Q56NI9	ESCO_HUMAN	N-acetyltransferase ESCO2	FRLKRRKRIARRLVD	601	68	547
A8	Q8IZT9	FAM9C_HUMAN	Protein FAM9C	EIKNRIKNVLRTTQL	166	19	100
A9	Q6QHK4	FIGLA_HUMAN	Factor in the germline alpha	KERERIKNLNRGFAR	219	24	79
A10	P41235	HNF4A_HUMAN	Hepatocyte nuclear factor 4-alpha	CRYCRLKKCFRAGMK	474	53	118
A11	Q92551	IP6K1_HUMAN	Inositol hexakisphosphate kinase 1	REQPRRKHSRRSLHR	441	50	116
A12	P78413	IRX4_HUMAN	Iroquois-class homeodomain protein IRX-4	DSGTRRKNATRETTS	519	54	147
A13	P78412	IRX6_HUMAN	Iroquois-class homeodomain protein IRX-6	SGAGRRKNATRETTS	446	48	151
A14	P41743	KPCI_HUMAN	Protein kinase C iota type	VLLVRLKKTDRIYAM	596	68	274/275
A15	Q07954	LRP1_HUMAN	Prolow-density lipoprotein receptor-related protein 1	CNSRRCKKTFRQCSN	4544	505	2568
A16	Q7L590	MCM10_HUMAN	Protein MCM10 homolog	PALPRTKRVARTPKA	875	98	187
A17	O60318	MCM3A_HUMAN	80 kDa MCM3-associated protein	CLGERLKHLERLIRS	1980	218	1951
A18	O95402	MED26_HUMAN	Mediator of RNA polymerase II transcription subunit 26	ELAKRAKKLLRSWQK	600	66	74/75
A19	Q8IYA7	MKX_HUMAN	Homeobox protein Mohawk	NARRRLKNTVRQPDL	352	39	129
A20	Q9UMN6	MLL4_HUMAN	Histone-lysine N-methyltransferase MLL4	VQGPRIKHVCRHAAV	2715	294	879
A21	Q9NYL2	MLTK_HUMAN	Mitogen-activated protein kinase kinase kinase MLT	ATLERLKKLERDLSF	800	91	295/296
B1	Q8IUG5	MY18B_HUMAN	Myosin-XVIIIb	VEGLRRKRAQRGQGS	2567	285	2260
B2	Q6X4W1	NELF_HUMAN	Nasal embryonic luteinizing hormone-releasing hormone factor	ETSGRRKKLERMYSV	530	60	197/198
B3	Q9H2A3	NGN2_HUMAN	Neurogenin-2	ETVQRIKKTRRLKAN	272	30	110/111
B4	Q9Y3T9	NOC2L_HUMAN	Nucleolar complex protein 2 homolog	PEIKRRKMADRKDED	749	85	652
B5	Q99743	NPAS2_HUMAN	Neuronal PAS domain-containing protein 2	DEKDRAKRASRNKSE	824	92	10
B6	Q13133	NR1H3_HUMAN	Oxysterols receptor LXR-alpha	EEQIRLKKLKRQEEE	447	50	177/178
B7	P10588	NR2F6_HUMAN	Nuclear receptor subfamily 2 group F member 6	CQYCRLKKCFRVGMR	404	43	114/115
B8	Q8WW12	PCNP_HUMAN	PEST proteolytic signal-containing nuclear protein	EAKMRMKNIGRDTPT	178	19	133
B9	Q8N488	RYBP_HUMAN	RING1 and YY1-binding protein	TSRPRLKNVDRSTAQ	228	25	149
B10	Q9NQZ2	SAS10_HUMAN	Something about silencing protein 10	GLTPRRKKIDRNPRV	479	55	426/427
B11	P42229	STA5A_HUMAN	Signal transducer and activator of transcription 5A / 5B	MSLKRIKRADRRGAE	794	91	425
B12	Q9UEW8	STK39_HUMAN	STE20/SPS1-related proline-alanine-rich protein kinase	DIAQRAKKVRRVPGS	545	60	362/363
B13	O75683	SURF6_HUMAN	Surfeit locus protein 6	RQNLRRKKAARAERR	361	41	332/333
B14	O75478	TAD2A_HUMAN	Transcriptional adapter 2-alpha	KERQRRKKIIRDHGL	443	52	222/223
B15	Q86TJ2	TAD2B_HUMAN	Transcriptional adapter 2-beta	KERQRRKNIARDYNL	420	49	220
B16	Q17RP2	TIGD6_HUMAN	Tigger transposable element-derived protein 6	QVDARMKRAERRILL	521	59	273

B17	Q8IY57	YAF2_HUMAN	YY1-associated factor 2	KTRPRLKNVDRSSAQ	180	20	106
B18	P51786	ZN157_HUMAN	Zinc finger protein 157	TGFVRRKRTPRGDKN	506	58	153
B19	Q86UD4	ZN329_HUMAN	Zinc finger protein 329	MRLKMTTRNFPEREV	541	62	4
B20	O60290	ZN862_HUMAN	Zinc finger protein 862	DGPRRIKRTYRPRSI	1169	132	457
B21	P17040	ZSC20_HUMAN	Zinc finger and SCAN domain-containing protein 20	QCRYRVKNLLRNYRK	1043	118	381
C1			Zinc finger and SCAN domain-containing protein 20	QCRYRFKNLLRSYRK			
C2	Q8IYH5	ZZZ3_HUMAN	ZZ-type zinc finger-containing protein 3	PVLKRIKRCLRSEAP	903	102	117/120