Biochemical characterization and identification of novel substrates of protein lysine methyltransferases

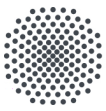
Von der Fakultät 3: Chemie der Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

von

Maren Kirstin Schuhmacher
aus Ilshofen, Deutschland

Hauptberichter:	Prof. Dr. Albert Jeltsch
Mitberichter:	Prof. Dr. Bernhard Hauer
Tag der mündlichen Prüfung:	13.12.2018

Institut für Biochemie und Technische Biochemie der Universität Stuttgart
Abteilung für Biochemie
2019



Universität Stuttgart

Erklärung über die Eigenständigkeit der Dissertation

Ich versichere, dass ich die vorliegende Arbeit mit dem Titel

Biochemical characterization and identification of novel substrates of protein lysine methyltransferases

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

Declaration of Authorship

I hereby certify that the dissertation entitled

Biochemical characterization and identification of novel substrates of protein lysine methyltransferases

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

Name/Name: Maren Kirstin Schuhmacher

Unterschrift/Signed:

Datum/Date: 16.10.2019

Contents

Acknowledgments

First I want to thank Prof. Dr. Albert Jeltsch for the opportunity to make my PhD thesis in his lab. I also want to thank him for his support, motivation, patience and help which he provided me all the time. And as well for his time which he spends for scientific discussions.

I wanted to thank Prof. Dr. Bernhard Hauer and Prof. Dr. Bernd Plietker for being members of my PhD thesis committee.

My sincere thanks also goes to Dr. Srikanth Kudithipudi for his supervision and help. He showed, explained and helped me to get familiar with the biochemical working procedure.

I would like to thank especially Dr. Sara Weirich, Nicole Berner and Dr. Rebekka Mauser for their time and help. Thank you for being there in a time when I needed you.

I thank all my colleagues, especially Dr. Ruth Menssen-Franz, Alexander Bröhm, Julian Broche and Mina Khella for providing me a nice working atmosphere and for their help.

A thank also goes to my students: Nadia Evgenieva Petkova, Felix Richard Fischer and Jan Ludwig for their engaged work.

I also want to thank Dragica Kapucija for creating a nice familiar atmosphere and for the nice and funny early talks in the morning and as well for being there when I needed her. A thank also goes to Regina Phillipp and Elisabeth Tosta for the help in ordering chemicals and the organization of many different things.

Last but not least I would like to thank my parents for their love, support, encouragement and as well to tolerate my mood over my whole studies. I want to dedicate this thesis to my dad who died too early. Besides my parents, I also want to thank my aunt Marlene Frenz-Krummel and my uncles Dr. med. Eberhard Krummel and Friedrich Frenz for their help and time. A thank also goes to Sebastian, Anja and Leni Naundorf for the nice time at the weekends. I would like to thank my friends as well for their motivation and the funny weekends.

List of publications

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.*

Schuhmacher MK, Kudithipudi S, Jeltsch A. (2016). **Investigation of H2AX methylation by the SUV39H2 protein lysine methyltransferase.** *FEBS Letters, 2016 Jun;590(12):1713-9. doi: 10.1002/1873-3468.12216. Epub 2016 May 30.*

Kudithipudi S, **Schuhmacher MK**, Kebede AF, Jeltsch A. (2017). **The SUV39H1 Protein Lysine Methyltransferase Methylates Chromatin Proteins Involved in Heterochromatin Formation and VDJ Recombination.** *ACS Chemical Biology, 2017 Apr 21;12(4):958-968. doi: 10.1021/acschembio.6b01076. Epub 2017 Feb 16.*

Schuhmacher MK, Kusevic D, Kudithipudi S, Jeltsch A. (2017). **Kinetic Analysis of the Inhibition of the NSD1, NSD2 and SETD2 Protein Lysine Methyltransferases by a K36M Oncohistone Peptide**. *Chemistry Select, 2017 Oct 18;2(29):9532-9536 doi: 10.1002/slct.201701940. Epub 2017 Oct 18.*

Schuhmacher MK, Rolando M, Bröhm A, Weirich S, Kudithipudi S, Buchrieser C, Jeltsch A. (2018). **The *Legionella pneumophila* methyltransferase RomA methylates also non-histone proteins during infection.** *Journal of Molecular Biology, 2018 June 22;430(13):1912-1925 doi: 10.1016/j.jmb.2018.04.032. Epub 2018 May 04.*

Authors contribution

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.*

MKS conducted all experiments with the help of SK, DK and SW. The data were analyzed by MKS and SK. AJ drafted the manuscript and all authors contributed to the writing and editing of the final manuscript.

Schuhmacher MK, Kudithipudi S, Jeltsch A. (2016). **Investigation of H2AX methylation by the SUV39H2 protein lysine methyltransferase.** *FEBS Letters, 2016 Jun;590(12):1713-9. doi: 10.1002/1873-3468.12216. Epub 2016 May 30.*

MKS performed all methylation experiments. MKS contributed to data analysis and interpretation, the writing and editing of the manuscript.

Kudithipudi S, **Schuhmacher MK**, Kebede AF, Jeltsch A. (2017). **The SUV39H1 Protein Lysine Methyltransferase Methylates Chromatin Proteins Involved in Heterochromatin Formation and VDJ Recombination.** *ACS Chemical Biology, 2017 Apr 21;12(4):958-968. doi: 10.1021/acschembio.6b01076. Epub 2017 Feb 16.*

MKS performed the experiments shown in Figure 2 A, the DOT1L and MLL1 experiments in Figure 2 B and as well the DOT1L experiments in Figure 3 C. MKS conducted the SET8 studies shown in Figure 5 and Suppl. Figure 7. In addition, the Coomassie stain in Suppl. Figure 1 A, the non-histone target investigation of DOT1L, MLL1, RSF1 and ZNF583 in Suppl. Figure 3 A and the SUV39H2 experiments in Suppl. Figure 3 B were conducted by MKS. MKS contributed to data analysis and interpretation and editing of the manuscript.

Schuhmacher MK, Kusevic D, Kudithipudi S, Jeltsch A. (2017). **Kinetic Analysis of the Inhibition of the NSD1, NSD2 and SETD2 Protein Lysine Methyltransferases by a K36M Oncohistone Peptide**. *Chemistry Select, 2017 Oct 18;2(29):9532-9536 doi: 10.1002/slct.201701940. Epub 2017 Oct 17.*

MKS performed the experiments with SETD2, the steady-state methylation experiments with NSD1 and the SDS-gels. MKS contributed to data analysis and interpretation, the writing and editing of the manuscript.

Schuhmacher MK, Rolando M, Bröhm A, Weirich S, Kudithipudi S, Buchrieser C, Jeltsch A. (2018). **The *Legionella pneumophila* methyltransferase RomA methylates also non-histone proteins during infection.** *Journal of Molecular Biology, 2018 June 22;430(13):1912-1925 doi: 10.1016/j.jmb.2018.04.032. Epub 2018 May 04.*

The *in vitro* experiments were conducted by MKS except that AB purified the RomA protein and performed the target lysine scan and SW conducted the initial activity test of RomA. The cellular studies were performed by MR and CB. MKS, MR, SK, CB and AJ were involved in the design and analysis of the experiments. MKS, MR, CB and AJ drafted the manuscript. All authors contributed to the editing of the manuscript.

List of Abbreviations

ac	acetylation
Acetyl-CoA	Acetyl-coenzyme A
AdoHcy	*S*-Adenosyl-*L*-homocystein
AdoMet	*S*-Adenosyl-*L*-methionine
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
ANKAR	Ankyrin and armadillo repeat-containing protein
AROS	Active regulator of SIRT1
AWS	Associated with SET-domain
bp	Base pairs
BBX	HMG box transcription factor BBX
CAC1G	Voltage-dependent T-type calcium channel subunit alpha-1G
CBFA2T2	Protein CBFA2T2
CC	Coiled-coiled domain
CD	Circular Dichroism or catalytic domain
cDNA	Complementary deoxyribonucleic acid
C. elegans	*Caenorhabditis elegans*
CENP-A	Centromere protein A or Histone H3-like centromeric protein A
CFP	Cyan fluorescent protein
chromo	Chromatin organization modifier
COMA	Collagen alpha-1(XXII) chain
Coomassie BB	Coomassie brilliant blue
cRCC	Cell renal cell carcinoma
CTD	C-terminal domain
Clr4	Histone-lysine N-methyltransferase, H3 lysine-9 specific
Cyc	Cycline C
Dim-5	Histone-lysine N-methyltransferase, H3 lysine-9 specific dim-5
DNMT	DNA methyltransferase
DOT1L	Histone-lysine N-methyltransferase, H3 lysine-79 specific
	Disruptor of telomeric silencing 1-like protein
DSB	Double strand break

DYSF	Dysferlin
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FBN1	Fibrillin-1
FOXO4	Forkhead box protein O4
G9a	Euchromatic histone-lysine N-methyltransferase
GFP	Green fluorescent protein
GLP	G9a-like protein
GST	Glutathion *S*-transferase
GCN5	Lysine acetyltransferase 2A
H-bond	Hydrogen bond
H1	Histone 1 or linker histone
H2A	Histone 2A
H2A.Bbd	Histone 2A.Barr body deficient
H2AX	Histone 2AX
H2AZ	Histone 2AZ
H2B	Histone 2B
H3	Histone 3
H3.1	Histone 3.1
H3.2	Histone 3.2
H3.3	Histone 3.3
H3K9	Histone 3 lysine 9
H3K14	Histone 3 lysine 14
H3K27	Histone 3 lysine 27
H3K36	Histone 3 lysine 36
H3K79	Histone 3 lysine 79
H3S10	Histone 3 serine 10
H3T11	Histone 3 threonine 11
H4	Histone 4
H4K20	Histone 4 lysine 20
HAT	Histone acetyl transferase
HEK293	Human embryonic kidney 293 cells

HMCN1	Hemicentin-1
HKMT	Histone lysine methyltransferase
HIP1	Huntingtin interacting protein 1
HP1	Heterochromatin protein 1
INT6	Integrator complex subunit 6
IPTG	Isopropyl-β-D-thiogalactopyranoside
i-SET	Inserted region
JARID	Jumonji/ARID domain-containing protein demethylase
JHDM	Jumonji domain-containing histone demethylase
JmjC	Jumonji C
KDM	Lysine demethylase
KIFC1	Kinesin-like protein KIFC1
L3MBTL1	Human lethal 3 malignant brain tumor repeat-like protein
L. pneumophila	*Legionella pneumophila*
LSD1	Lysine specific demethylase 1
MALDI	Matrix Assisted Laser Desorption/Ionization
MBT	Malignant brain tumor
me1	Monomethylation
me2	Dimethylation
me3	Trimethylation
MED13	Mediator of RNA polymerase II transcription subunit 13
MLL1	Histone-lysine N-methyltransferase 2A
MN	Mononucleosomes
MPP8	M-phase phosphoprotein 8
MS/MS	Tandem mass spectrometry
mut	Mutation
NAD	Nicotinamide adenine dinucleotide
N. crassa	*Neurospora crassa*
NEB	NEW ENGLAND BioLabs
NFκB	Nuclear factor of kappa light polypeptide gene enhancer in B-cells
NHEJ	Non-homologues end-joining
Ni-NTA	Nickel-Nitrilotriacetic acid

NSD1/2	Nuclear receptor-binding SET-domain-containing protein 1/2
OCT4	Octamer-binding protein 4
OD_{600}	Optical density at 600 nm
p53	Cellular tumor antigen p53 or tumor suppressor p53
PCR	Polymerase chain reaction
PHD	Plant homeodomain
PKMT	Protein lysine methyltransferase
PRC2	Polycomb repressive complex 2
Prdm14	PR domain zinc finger protein 14
Prdm16	PR domain zinc finger protein 16
PRMT	Protein arginine methyltransferase
PTM	Post-translational modification
PWWP	Proline-tryptophan-tryptophan-proline motif containing domain
RAG2	V(D)J recombination-activating protein 2
RAI1	Retinoic acid induced protein 1
RNAPII	RNA polymerase II
RSF1	Remodeling and spacing factor 1
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
S_N2	Bimolecular nucleophilic substitution
SAH	*S*-Adenosyl-*L*-homocystein
SDS	Sodium dodecylsulfate
SDS-PAGE	Sodium-Dodecylsulfate polyacrylamide gel electrophoresis
SET	Su(var)3-9, Enhancer of Zeste and Trithorax
SETDB1	Histone H3K9 methyltransferase 4
SETDB2	Lysine N-methyltransferase 1F
SET7/9	SET-domain containing protein 7
SET8	SET-domain containing protein 8
SETD2	Su(var), Enhancer of zeste, Trithorax-domain containing protein 2
SETMAR	SET domain and mariner transposase fusion gene-containing protein
SIA8D	CMP-N-acetyl-neuraminate-poly-alpha-2,8-sialyltransferase
SIRT1	NAD-dependent protein deacetylase sirtuin-1

SRI-domain	Set2-Rpb1 interacting-domain
Su(var) 3-9	Suppressor of variegation 3-9
SUV39H1	Human suppressor of variegation 3-9 homolog 1
SUV39H2	Human suppressor of variegation 3-9 homolog 2
SUV4-20H1	Human suppressor of variegation 4-20 homolog 1
SUV4-20H2	Human suppressor of variegation 4-20 homolog 2
T4SS	Type-4 secretion system
TAF3	Transcription initiation factor TFIID subunit 3
WIZ	Widely interspaced zinc finger motifs protein
WT	Wild type
WW	Tryptophan-tryptophan containing domain
YFP	Yellow fluorescent protein
7BS	seven-β-strand methyltransferase fold
ZNF583	Zinc finger protein 583

Zusammenfassung

Die Methylierung von Lysinseitenketten ist eine wichtige post-translationale Modifikation (PTM) von Proteinen, die durch Protein-Lysin-Methyltransferasen (PKMTs) eingefügt wird. Die Lysin-Methylierung von Histonproteinen kann unterschiedliche Auswirkungen auf die Chromatinstruktur haben und die Methylierung von anderen nicht-Histon-Proteinen kann Protein/Protein-Wechselwirkungen und die Proteinstabilität beeinflussen. Für die meisten PKMTs sind derzeit noch nicht alle Methylierungssubstrate bekannt. Dies führt dazu, dass der Einfluss dieser Enzyme in verschiedenen Prozessen in der Zelle nicht komplett verstanden wird. Daher ist es ein wichtiges Forschungsziel, mehr Informationen über das Substratspektrum von PKMTs zu erhalten. Die Untersuchung der Substratspezifität verschiedener PKMTs ist ein wichtiger Schritt auf dem Weg zur Identifizierung neuer Methylierungsstellen. Der Fokus dieser Arbeit lag auf der Analyse der Substratspezifität verschiedener PKMTs unter Verwendung von SPOT-Peptidarrays und im Folgenden auf der Identifizierung neuer Substrate, und deren Validierung.

Die Analyse der Substratspezifität von humaner SUV39H2 ergab signifikante Unterschiede zwischen SUV39H2 und seinem Homolog SUV39H1, obwohl beide Enzyme das gleiche Histon-Substrat (H3K9) methylieren. SUV39H2 ist spezifischer als SUV39H1, was durch die fehlende Methylierung der SUV39H1 nicht-Histon-Substrate durch SUV39H2 belegt werden konnte und dadurch, dass in der vorliegenden Arbeit keine neuen Substrate für SUV39H2 identifiziert werden konnten. Kinetische Studien zeigten, dass SUV39H2 unmethyliertes H3K9 als Substrat bevorzugt. Darüber hinaus wurde gezeigt, dass die Mutation N324K in SUV39H2, die als genetische Krankheit bei Labrador-Retrievern nachgewiesen wurde, eine Veränderung der Faltung und in Folge dessen eine Inaktivierung des Enzyms hervorruft. Studien anderer Labore zeigten, dass die Histonvariante H2AX durch SUV39H2 methyliert wird. Die Sequenz um K134 von H2AX stimmt jedoch nicht mit der hier ermittelten Substratspezifität von SUV39H2 überein. Die in der vorliegenden Arbeit durchgeführten *in vitro* Studien auf Peptid- und Proteinebene zeigten, dass H2AX K134 nicht durch SUV39H2 methyliert wird. Dies deutet darauf hin, dass die H2AX-Methylierung durch SUV39H2 höchstwahrscheinlich eine falsche Zuordnung eines Substrats zu einer PKMT ist. Basierend auf bereits verfügbaren Spezifitätsdaten wurde SET8 als neues Substrat für SUV39H1 auf zellulärer Ebene validiert. SET8 gehört zu den PKMTs, welche H4K20 methylieren können. In dieser Arbeit konnte gezeigt

werden, dass die Methylierung von SET8 an K210 durch SUV39H1 die Aktivität von SET8 stimuliert.

Im menschlichen Organismus existieren verschiedene PKMTs, die H3K36 methylieren. Als Beispiel sind hier NSD1, NSD2 und SETD2 zu nennen, welche hier untersucht wurden. In der Literatur wurde gezeigt, dass die K36M Onkohistonmutation NSD2 und SETD2 inaktiviert. *Steady-state* Methylierungskinetiken unter Verwendung eines Peptidsubstrats und eines K36M-Peptidinhibitors zeigten, dass NSD1 ebenfalls durch diese Mutation inhibiert wird. Die *steady-state* Inhibitionsparameter für alle Enzyme zeigten eine bessere Bindung des Inhibitorpeptids an die PKMTs als das Substratpeptid, was auf mechanistische Ähnlichkeiten in der Peptidinteraktion hindeutet. SETD2 ist eine Methyltransferase, die H3K36 dreifach methylieren kann. In dieser Arbeit wurden zwei Substratspezifitätsmotive von SETD2 mittels Peptidarray-Methylierungsexperimenten bestimmt. Basierend auf den Substratspezifitätsuntersuchungen wurde zusätzlich ein Super-Substrat auf Peptid- und Proteinebene hergestellt. Des Weiteren wurde ein neues Substrat (FBN1) für SETD2 ermittelt und validiert.

Studien von unseren Kooperationspartnern zeigten, dass die PKMT RomA aus *Legionella pneumophila* in der Lage ist, H3 an K14 zu methylieren. Basierend auf dem Spezifitätsprofil von RomA, welches in dieser Arbeit ermittelt wurde, konnte gezeigt werden, dass dieses Enzym sieben weitere menschliche nicht-Histon-Proteine methyliert. Unsere Kooperationspartner untersuchten die Methylierung eines dieser nicht-Histon-Substrate (AROS) während der Infektion von menschlichen Zellen mit *L. Pneumophila* und zeigten die Methylierung von AROS. Die Rolle dieser Methylierungsereignisse im Infektionsprozess muss in zukünftigen Experimenten weiter untersucht werden.

Abstract

The methylation of lysine side chains is a prevalent post-translational modification (PTM) of proteins, which is introduced by protein lysine methyltransferases (PKMTs). Histone methylation can have different effects on chromatin structure, lysine methylation of non-histone proteins can regulate protein/protein interactions and protein stability. For most PKMTs currently not all methylation sites are known which limits our understanding of the regulatory role of these enzymes in cells. Therefore, it is an important research aim to gain more information about the substrate spectrum of PKMTs. The identification of the substrate specificity of a PKMT is a very important step on the way to identify new PKMT methylation sites. The focus of this study was the analysis of the substrate specificity of different PKMTs by SPOT peptide arrays and based on this on the identification and validation of possible new methylation substrates.

The analysis of the substrate specificity of human SUV39H2 revealed significant differences to its human homolog SUV39H1, although both enzymes methylate the same histone substrate (H3K9). SUV39H2 is more stringent than the SUV39H1, which could be demonstrated by the lack of methylation of SUV39H1 non-histone targets by SUV39H2 and by the fact that it was not possible in this study to identify non-histone substrates for SUV39H2. Kinetic studies showed that SUV39H2 prefers the unmethylated H3K9 as substrate. Moreover, it was shown that the N324K mutation of SUV39H2 which leads to a genetic disease in Labrador retrievers causes a change in folding finally leading to the inactivation of the enzyme. It had been reported by another group that the histone variant H2AX is methylated by SUV39H2. However, the sequence of H2AX K134 does not fit to the substrate specificity profile of SUV39H2 determined in the present work. Follow-up *in vitro* peptide and protein methylation studies indeed showed that H2AX K134 is not methylated by SUV39H2. This indicates that H2AX methylation by SUV39H2 is most probably a wrong assignment of a substrate to a PKMT. Based on already available specificity data for the SUV39H1 PKMT, the SET8 protein was validated as novel substrate in cellular studies. SET8 is a PKMT itself and it could be shown in this thesis that methylation of SET8 at residue K210 by SUV39H1 stimulated the SET8 activity.

In humans, there exist different PKMTs, which methylate H3K36. For example, NSD1, NSD2 and SETD2 which were investigated in this thesis. In literature, it was shown that the

oncohistone mutation K36M inactivates NSD2 and SETD2. *Steady-state* methylation kinetics using a peptide substrate and a K36M peptide as inhibitor revealed that NSD1 is inhibited by this histone oncomutation as well. The *steady-state* inhibition parameters for all enzymes showed a better binding of the PKMTs to the inhibitor peptide than to the substrate, suggesting some mechanistic similarities in target peptide interaction. The SETD2 is a methyltransferase, which is able to introduce trimethylation of H3K36. During this thesis two substrate specificity motifs of SETD2 were determined using peptide array methylation experiments. Additionally, based on the substrate specificity investigations a super-substrate at peptide and protein level was determined. Furthermore, one novel substrate (FBN1) for SETD2 was discovered and validated.

The *Legionella pneumophila* RomA PKMT was shown previously by our collaborators to methylate H3 at K14. Based on the specificity profile of RomA determined in this study it could be shown that this enzyme methylates seven additional human non-histone proteins. Collaborators tested the methylation of one of the non-histone targets (AROS) and could demonstrate its methylation during the infection of human cells with *L. pneumophila*. The role of these methylation events in the infection process must be studied in future experiments.

1. Introduction

1.1 Epigenetics

The first definition of epigenetics was introduced by C. H. Waddington in 1942 as “the branch of biology that studies the causal interactions between genes and their products which bring the phenotype into being” (Waddington 1968; Dupont et al. 2009). After the definition by Waddington, more insights into this research field led to some refinements in the definition. Today epigenetics is considered as “the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence” (Wu and Morris 2001).

For a better visualization, Waddington used the epigenetic landscape (Figure 1). In this, a cell is represented as ball on a landscape, in which it undergoes different stable decisions during development that lead to different stages of the final differentiated cell. DNA methylation is the best studied epigenetic mark so far. Other examples of epigenetic marks are histone modifications including histone methylation, one of the subjects of this PhD thesis. Non-coding RNAs play a role in epigenetic signaling as well.

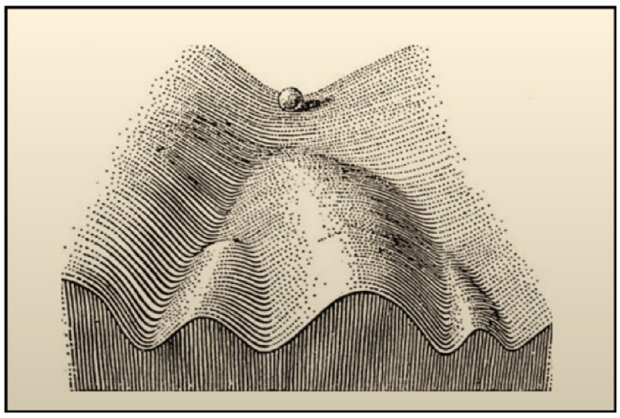


Figure 1: The epigenetic landscape proposed by C. H. Waddington. A cell is displayed as ball, which can go into different directions during development. The picture is taken from (Goldberg et al. 2007).

1.2 Chromatin

Chromatin was first described around 1880 by W. Flemming (Flemming 1882; Olins and Olins 2003). The chromatin contains the cellular DNA and with this all the genetic information of the cell. The basic building block of chromatin is a nucleosome consisting of a histone octamer (also called histone core particle) and a 147 base pair (bp) long DNA fragment, which is wrapped around the histone core. The histone octamer comprises a tetramer of histone 3 (H3) and histone 4 (H4), and two dimers of histone 2A (H2A) and histone 2B (H2B) (Luger et al. 1997; Kouzarides 2007). Each nucleosome is connected with another, neighboring nucleosome, by a 10-50 bp long linker DNA fragment (Segal et al. 2006). This connection of the nucleosomes and the generated structure is called “beads on a string“, also known as 10-nm fiber (Francastel et al. 2000). The 10-nm fiber is the first level of the compaction of chromatin, followed by the 30-nm fiber, which is described as “packed nucleosomes” (Figure 2) (Francastel et al. 2000). In the construction of the 30-nm fiber another histone is involved, which is known as histone 1 (H1) or linker histone. H1 has an influence on the nucleosome assembly, by neutralizing the negative charge of the DNA (Luger 2001).

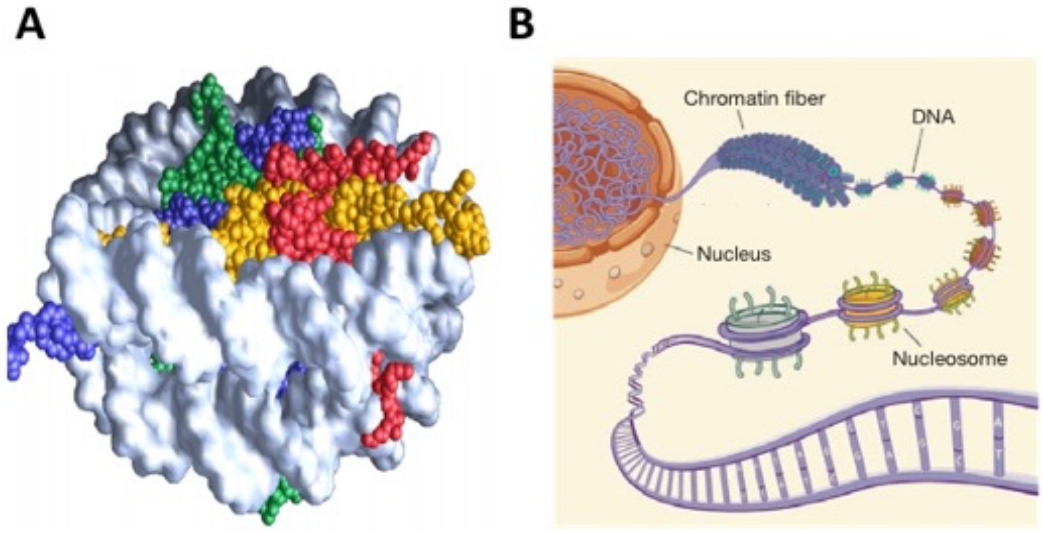


Figure 2: Representation of the histone core particle and overview of the DNA compaction in the cell nucleus. A) Sphere model of the histone core particle. The blue color represents the atoms of H3, the red color represents the atoms of H2B, the green color shows the H4 atoms, and yellow represents H2A atoms. DNA is shown as molecular surface. The figure is adopted from (Luger 2001). B) Schematic illustration of the compaction of the DNA in the cell nucleus. First the DNA is wrapped around a histone core particle generating a 10-nm fiber, which is then itself further compacted into a 30-nm fiber, resulting in chromatin. The figure is taken from (Baker 2011).

In literature, two different forms of chromatin are described called euchromatin and heterochromatin. The difference between both forms is in the compaction of the chromatin. Heterochromatin represents a condensed, closed form of chromatin, whereas the euchromatin is a more decondensed, open form. Heterochromatin is marked by DNA methylation, trimethylation of H3K9 and H3K27 as well as H4K20 and a low level of acetylated histones (Kouzarides 2007). Heterochromatin normally is transcriptionally silent. In contrast, euchromatin is associated with a high level of acetylated histones and a trimethylation mark present on H3K4, H3K36 and H3K79 (Kouzarides 2007) and actively transcribed genes. Another difference between both chromatin states is that euchromatin regions are also characterized by a lack of the linker histone H1.

1.2.1 Histone variants

Beside the canonical histone proteins, so-called "histone variants" exist for H2A and H3, but not for H4 and H2B. For H3, four variants are known so far namely, H3.1, H3.2, H3.3 and the centromere protein A (CENP-A) (Chen et al. 2013). Most of the variants have a more or less similar amino acid sequence compared to the canonical histone proteins. For example, the first three H3 variants (H3.1, H3.2 and H3.3) only have a few amino acid variations when compared with the canonical H3. Only CENP-A has a large amino acid sequence dissimilarity. For H2A, the following variations are known: H2AZ, macroH2A, H2A.Barr body deficient (H2A.Bbd) and H2AX. H2AZ has a 60 % identity with H2A and is mostly found in actively transcribed chromatin (Luger 2001; Chen et al. 2013). The H2AX variant was firstly described in the 1980s and is conserved over many species, from *Saccharomyces cerevisiae* to human (Redon et al. 2002). It differs from the canonical H2A only at few positions (Figure 3), mainly at amino acids, which extend the C-terminus of H2A and terminate it with an S - Q - E - Y motif. The serine in this sequence (S139) can be phosphorylated in response to double strand breaks (DSB) in the DNA and to a lesser extend T136 located in an TQ motif is phosphorylated as well (Redon et al. 2002). The example of phosphorylation of S139 in H2AX shows that histone protein variant specific modifications can be very important for cellular regulation.

```
P0C0S8 H2A1_HUMAN      1   SGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYLAAVLEYLTA   60
P16104 H2AX_HUMAN      1   SGRGKTGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGHYAERVGAGAPVYLAAVLEYLTA   60
                           ***** *********:**********************:**********************

P0C0S8 H2A1_HUMAN     61   EILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKT   120
P16104 H2AX_HUMAN     61   EILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGGVTIAQGGVLPNIQAVLLPKKT   120
                           ************************************ ********************

P0C0S8 H2A1_HUMAN    121   ESHHKAKGK-------------                                      129
P16104 H2AX_HUMAN    121   SATVGPKAPSGGKKATQASQEY                                      142
                           .:    *.
```

Figure 3: Sequence alignment of human histone H2A (P0C0S8) and its histone variant H2AX (P16104). The stars represent amino acid positions, which are identical between both sequences. Colons indicate amino acids with similar properties. H2AX has a C-terminal extension of thirteen amino acids, in which the $T^{136}Q$ and $S^{139}Q$ motifs are localized. The serine of the SQ motif can be phosphorylated in response to double strand breaks (Redon et al. 2002). The sequence alignment was prepared with the UniProt sequence alignment tool.

1.3 Protein post-translational modifications

Post-translational modifications (PTM) are covalent alterations of proteins, which occur after protein synthesis. The N-terminal tails of histone proteins are hotspots of post-translational modifications. This was firstly described in 1964 by Allfrey and Mirsky (Allfrey and Mirsky 1964). For the histones H2A and H2B also PTMs at the C-terminal tails are known (Margueron et al. 2005). These PTMs include acetylation of lysine residues, methylation of lysine or arginine residues, phosphorylation of serine or threonine side chains, ubiquitination and sumoylation of lysine residues, ADP ribosylation, glycosylation, biotinylation and carbonylation (Margueron et al. 2005). In general, PTMs can have an influence on protein/protein interaction, protein stability and protein localization and the activity of the protein. Another interesting point is that PTMs can affect each other. For example, it was previously reported that H3K14 can be methylated and acetylated, but methylation of H3K14 by the bacterial methyltransferase RomA prevents its acetylation (Rolando and Buchrieser 2014). Histone PTMs can also have an effect on the binding of other proteins as it was described for the heterochromatin protein 1 (HP1) which binds to H3K9me3 but binding is released by H3S10 phosphorylation (Fischle et al. 2003). Figure 4 shows an overview of the methylation, acetylation, phosphorylation and ubiquitination sites on histone tails.

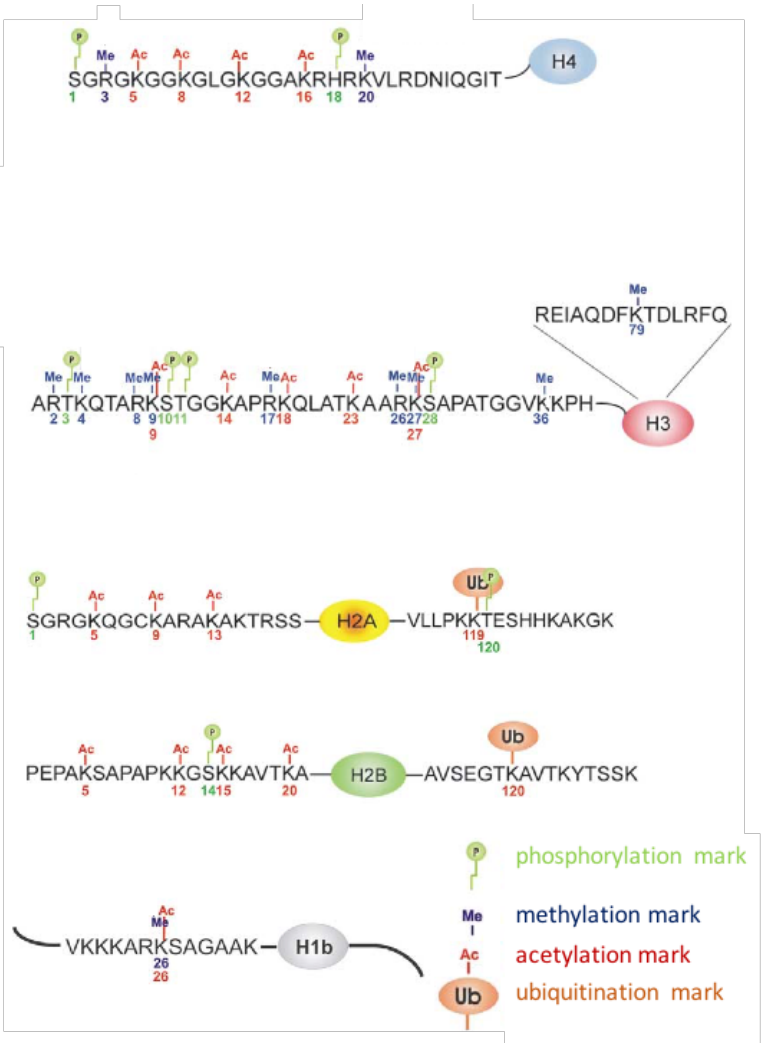


Figure 4: Overview of different post-translational modifications on the N- and C-terminal tails of histone proteins. The core particle of H4 is represented in light blue, H3 in red, H2A in yellow, H2B in green and H1b in grey. PTMs are represented in green (phosphorylation), dark blue (methylation), red (acetylation) and brown (ubiquitination). The numbers indicate the amino acid position in the histone tails. The figure was adopted and changed from (Margueron et al. 2005).

1.3.1 Protein acetylation

The first described enzymes which modify the N-terminal tail of histones were histone acetyltransferases (HATs) which transfer an acetyl-group from the cofactor Acetyl-coenzyme A (Acetyl-CoA) to the ε-amino group of a lysine residue. Acetylation of the ε-amino group of lysine causes the removal of the positive charge of the lysine and by this it changes its biochemical properties leading to weaker interaction with the DNA. Lysine acetylation is in most cases associated with chromatin decondensation and gene activation (Marmorstein 2001). The lysine acetylation of the ε-amino group is a reversible mark which can be removed

by histone deacetylases. The acetylation of lysine residues not only occurs on histone tails, but it was also described for other proteins at the α-amino-group of the N-terminus.

1.3.2 Histone lysine methylation

The main focus in this thesis is the lysine methylation (Murray 1964). The methylation of the ε-amino group of a lysine residue is catalysed by protein lysine methyltransferases (PKMTs), which can introduce with high specificity up to three methyl groups using *S*-Adenosyl-*L*-methionine (AdoMet) as methyl group donor (Figure 5). The best characterized lysine methylation events occur on the N-terminal tails of H3 and H4 and were extensively studied in the last couple of years.

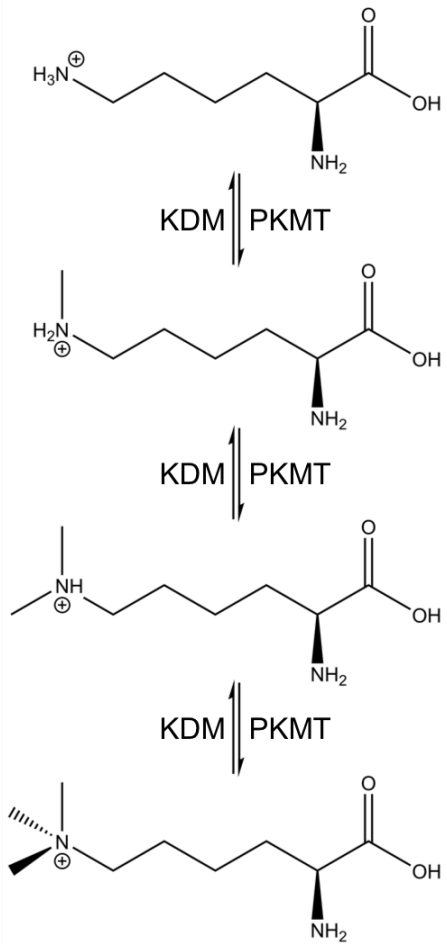


Figure 5: Introduction and removal of a methyl group on the ε-amino group of a lysine. The methylation of a lysine can go up to the trimethylation state. It is introduced by a PKMT using AdoMet as cofactor. The histone methylation is a reversible mark which can be removed by so-called lysine demethylases (KDM).

In general, histone methylation by PKMTs has a big influence on important processes in human cells, for example gene transcription or DNA repair. As an example, trimethylation of H3K4, H3K36 and H3K79 is associated with active transcribed genes in euchromatin, whereas H3K9 and H4K20 trimethylation is correlated with pericentric heterochromatin and gene silencing (Peters et al. 2003; Schotta et al. 2004; Kouzarides 2007).

The H3K9 methylation is introduced by different enzymes of the suppressor of variegation 3-9 enzyme family (SUV39 family), including for example the human suppressor of variegation 3-9 homolog 1 and 2 (SUV39H1/H2), the euchromatic histone-lysine N-methyltransferase (G9a) or the G9a-like protein (GLP) and the histone H3K9 methyltransferase 4 (SETDB1). It is an important mark for heterochromatin formation and transcriptional silencing (Tachibana et al. 2007). Mono- and dimethylation of H3K9 is found in silent domains in the euchromatin (Rice et al. 2003). The H3K9me3 is enriched at pericentric heterochromatin which is an important marker for heterochromatin (Rice et al. 2003). In *Neurospora crassa* Tamaru et al. (2003) showed that H3K9me3 methylation is an important mark for cytosine methylation (Tamaru et al. 2003). Later studies in mice showed, that H3K9me3 is important in the inactivation of the X chromosome (Fukuda et al. 2014). The HP1 protein family members bind with their chromodomain to methylated H3K9, which then recruits the human suppressor of variegation 4-20 homolog proteins (SUV4-20H) (Lachner et al. 2003). After targeting of SUV4-20H enzymes by HP1 they methylate H4K20me1 up to the trimethylation state thereby generating H4K20me3 another prominent heterochromatic mark (Balakrishnan and Milavetz 2010). The H4K20me1 mark is set by another SET-domain containing methyltransferase, SET8 (SET-domain containing protein 8), which uses H4K20 as substrate (Jorgensen et al. 2013). Another observation was that HP1 binding to H3K9me3 recruits SUV39H1. This recruitment of SUV39H1 then leads to further methylation of histones and to a maintenance of heterochromatin (Bannister et al. 2001). These processes illustrate the intimate interaction between the different chromatin marks and the enzymes, which are setting and removing them.

Many PKMTs are able to methylate H3K36. Two examples are the Su(var), Enhancer of zeste, Trithorax (SET) domain containing protein 2 (SETD2) and the nuclear receptor-binding SET-domain-containing protein 1 (NSD1) and 2 (NSD2), which have been investigated in the

present thesis. Many studies associate H3K36 methylation with actively transcribed genes. Bannister et al. (2005) observed a shift of the mono- to trimethylation localization from the promotor to the 3´ ends of expressed genes (Bannister and Kouzarides 2005). Previously, it was shown by Kizer et al. (2005) that the Set2-Rpb1 interacting-domain (SRI-domain) of yeast Set2 binds to the C-terminal repeat domain of the RNA polymerase II (RNAPII). This binding recruits SETD2 leading to the enrichment of the H3K36me3 mark in gene bodies of actively transcribed genes. As downstream signal, H3K36me3 is associated with DNA methylation because the Proline-tryptophan-tryptophan-proline (PWWP) domain of the DNA methyltransferases DNMT3A and DNMT3B bind this mark. Through this binding they recruit DNA methylation next to H3K36me3 (Dhayalan et al. 2010). This observation links the DNA and histone methylation together. Moreover, H3K36me2 is able to enhance the DNA double strand break repair by the non-homologues end-joining (NHEJ) pathway, because H3K36me2 is important for the recruitment of the components of the DNA repair complex (Fnu et al. 2011). This methylation mark is introduced by another H3K36 methyltransferase which is called SET domain and mariner transposase fusion gene-containing protein (SETMAR).

1.3.3 Non-histone protein methylation

In the last couple of years, not only histone methylation was described, but it was discovered that non-histone proteins are also methylated by PKMTs. The identification of non-histone substrates of PKMTs or *vice versa* the identification of a PKMT, which sets a known methylation mark is an important field of research and many studies were performed to identify new substrates of the methyltransferases. By using proteomic studies, many methylated proteins were identified. Like for example the tumor suppressor protein p53 (p53), which can be methylated at K370, K372, K373 and K382 by different PKMTs (Chuikov et al. 2004; Huang et al. 2006; Shi et al. 2007; Huang et al. 2010). p53 is a tumor suppressor and transcription factor, which is a highly mutated gene in human cancers (West and Gozani 2011). The K372 methylation of p53 by the SET-domain containing protein 7 (SET7/9) was the first assignment of a non-histone target to a PKMT (Chuikov et al. 2004). The p53 K732 methylation leads to transcription enhancement of p53 target genes (Egorova et al. 2010). The non-histone lysine methylation can have an influence on the stability or activity of the methylated protein. It could also have an impact on the regulation of target genes of the non-histone proteins like it was mentioned above for p53.

Until now not all methylation sites identified in proteomic studies in human proteins could be assigned to a specific PKMT. For the different PKMTs not all of the substrates are known. In the present thesis this issue will be addressed by the determination of the substrate specificity profiles of different PKMTs and the identification and validation of novel non-histone protein methylation substrates.

1.3.4 Readout of lysine methylation

The lysine methylation is read by binding or reading domains. These reading domains distinguish between the different lysine methylation states (un-, mono-, di- and trimethylated). Examples for such domains are the Plant Homeodomain (PHD)-, the Chromo-, the Tudor-, the PWWP- or Malignant Brain Tumor (MBT)-domains. These binding domains contain a methyllysine binding pocket formed by aromatic amino acids, which form an aromatic cage (Yun et al. 2011). The binding pocket of mono- or dimethyl readers have a smaller binding pocket compared to di- and trimethyl binding domains and in general also form hydrogen-bonds (H-bonds) with the lysine amino group (Figure 6).

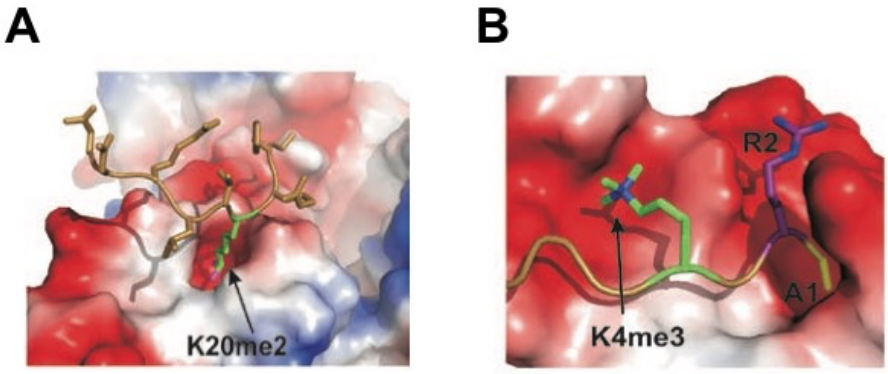


Figure 6: Lysine binding pocket of binding domains or readers. A) MBT-domain of L3MBTL1 which targets the H4K20me2 mark. B) H3K4me3 binding by the PHD-domain of TAF3. Adopted from (Yun et al. 2011).

Another important aspect of the binding of reading domains to the methylated lysine is the sequence context of the lysine and the location of the lysine. Lysine side chains located at the tails of histones or in disordered loops on the surface of proteins can be better accessed by reading domains than lysine residues engaged in stable salt bridges for example. The reading domains are important to target proteins to their place of action, like it was demonstrated for the Tudor-domain of KDM4a. The Tudor-domain of KDM4a binds to H3K4me3 or H4K20me3

and by this the KDM4a is able to demethylate H3K9me2/me3 or H3K36me2/me3 in nearby or the same histone tails (Yun et al. 2011).

1.3.5 Lysine demethylation

In the last couple of years, it was shown that lysine methylation can be removed by so-called lysine demethylases (KDMs), which makes lysine methylation a reversible modification. The first identified KDM was the lysine specific demethylase 1 (LSD1), which is highly specific for the demethylation of H3K4me1 or H3K4me2 but not H3K4me3 (Shi et al. 2004; Bannister and Kouzarides 2005). H3K4 demethylation by LSD1 is performed by an amine oxidase reaction leading to unmethylated lysine, hydrogen peroxide and formaldehyde as products (Shi et al. 2004). In the following years, another class of KDMs was discovered which contains a Jumonji C (JmjC) domain as active part, the Jumonji domain-containing histone demethylase (JHDM) family or the Jumonji/ARID domain-containing protein demethylases (JARID). In contrast to LSD1, the JmjC containing demethylases use an oxidative hydroxylation mechanism for the demethylation of lysine residues (Trewick et al. 2005). H3K9 methylation is removed by three classes of demethylases (JHDM2, JHDM3 and PHF8) (Hyun et al. 2017). For the removal of the H3K36 methylation, two enzyme classes are known (JHDM1 and JHDM3) (Hyun et al. 2017).

1.3.6 Protein arginine methylation

Methylation has been observed at H3R8 and H4R3 (Pal et al. 2004). The protein arginine methyltransferases (PRMTs) can introduce methyl groups on arginine residues. PRMTs can introduce up to two methyl groups onto the ω–guanidino group of arginine (Yang and Bedford 2013), that results in three different forms of arginine methylation. The monomethylation, which serves as an intermediate state, and dimethylation in a symmetrical or asymmetrical way (Yang and Bedford 2013). To date, nine PRMTs are known, which all contain a catalytically active seven-β-strand methyltransferase fold (7BS) (Yang and Bedford 2013).

1.4 Proteome wide studies and PKMT assays

Proteome wide studies are employed for the detection of lysine methylation of proteins. The identification of cellular methylated lysine residues can be performed by different methods. One possible method is the addition of radioactively labeled methyl donors to cell culture

media and afterwards the detection of the methylated lysine proteins by SDS-PAGE and autoradiography (Lanouette et al. 2014). However, this method is time-consuming and the identification of the methylated proteins is not easy. Another method which is used for lysine methylation is immunoprecipitation of methylated proteins followed by mass spec analysis (Iwabata et al. 2005). This approach relies on the usage of pan-methyllysine antibodies. The quality of this antibodies may vary from company to company and lot to lot which makes it difficult to reproduce obtained results (Lanouette et al. 2014). Today the method which is used for the identification of proteome wide lysine methylation events is the mass spectrometry. It is a powerful method because a high number of samples can be measured in a short period of time and this method can be used to distinguish between the different methylation states. A recent study developed an antibody mixture to enrich methylated peptides, because the amount of methylated lysine in the human cell is much lower than unmethylated lysine residues (Guo et al. 2014). Another approach which is used is the prediction of lysine methylation events (Lanouette et al. 2014). This approach relies on the knowledge of the sequence which is recognized by the investigated PKMT. Based on this sequence, a selection of putative targets can by investigated for methylation on peptides and proteins.

The determination of the activity of PKMTs is important for any functional study. For this different activity assays can be used, like for example radioactive based *in vitro* methylation assays. In this assay, radioactively labeled AdoMet is used as cofactor and incubated with the known or putative substrate protein or peptide of the enzyme and the enzyme itself. The radioactive methyl groups transferred to the substrate can then be detected by different methods like for example autoradiography. Another possible method for the detection of the activity and the specificity of a PKMT is mass spectrometry conducted after incubation of enzyme and substrate in the presence of unlabeled AdoMet.

1.5 Protein lysine methyltransferases

The PKMTs were first identified as histone lysine methyltransferases (HKMTs). Later it was shown that these enzymes can not only methylate histone tails of mainly H3 or H4, but also so-called non-histone targets. Concerning this important observation, they were renamed

from HKMTs to PKMTs. There exist two groups of PKMTs. One group of PKMTs contains a conserved SET-domain and the other group a 7BS fold. The disruptor of telomeric silencing 1-like protein (DOT1L), a H3K79 methyltransferase, is a member of the 7BS family. The 7BS fold consists of seven twisted β-sheets as core domain, which are arranged in a conserved orientation, in which the first six sheets have a parallel orientation, whereas the last is directed in the opposite way (Falnes et al. 2016). In contrast to the SET-domain containing PKMTs, the 7BS enzymes bind the AdoMet in a deep pocket. This proposes a mechanism in which a conformational change of the enzyme is needed for the release of the *S*-Adenosyl-*L*-homocysteine (AdoHcy) and binding of a new AdoMet (Cheng et al. 2005). The SET-domain and examples of the SET-domain containing PKMTs will be described in the following paragraphs.

1.5.1 SET-domain PKMTs

The SET-domain was first identified in *Drosophila melanogaster* proteins (Suppressor of variegation 3-9 (Su(var)3-9), Enhancer of zeste (E(z)) and Trithorax (Trx)), from which it received the name (Dillon et al. 2005). It has a size of around 130 amino acids (Jenuwein et al. 1998) and is a well conserved structure observed in many different species e.g. *Homo sapiens*, *Saccharomyces cerevisiae* and others (Cheng et al. 2005). Over many years of research, the structures of many SET-domain PKMTs were investigated and it was shown that the SET-domain folds into several β-strands, which are divided into three main sheets that itself form a knot-like structure (Figure 7) (Qian and Zhou 2006). The C-terminal part of the domain plays an important role in the formation of the pseudo knot-like structure, because it passes through a loop formed by a stretch of the sequence (Qian and Zhou 2006). The solved structures of SET-domain PKMTs showed that the SET-domain (core-SET) is often surrounded by two additional domains called the pre- and post-SET-domains. The pre-SET-domain stabilize the core-SET-domain by the interaction with parts of the core-SET-domain (Qian and Zhou 2006). The post-SET-domain is critical for the formation of the active center, into which it contributes with an aromatic residue (Zhang et al. 2002). In addition, the SET-domains of some lysine methyltransferases contain an i-SET-domain, which has a different size in distinct lysine methyltransferases. It was shown that it interacts with the substrate lysine and contributes by this to the substrate specificity of the SET-domain. The SET-domain structure of SET7/9 is shown in Figure 7.

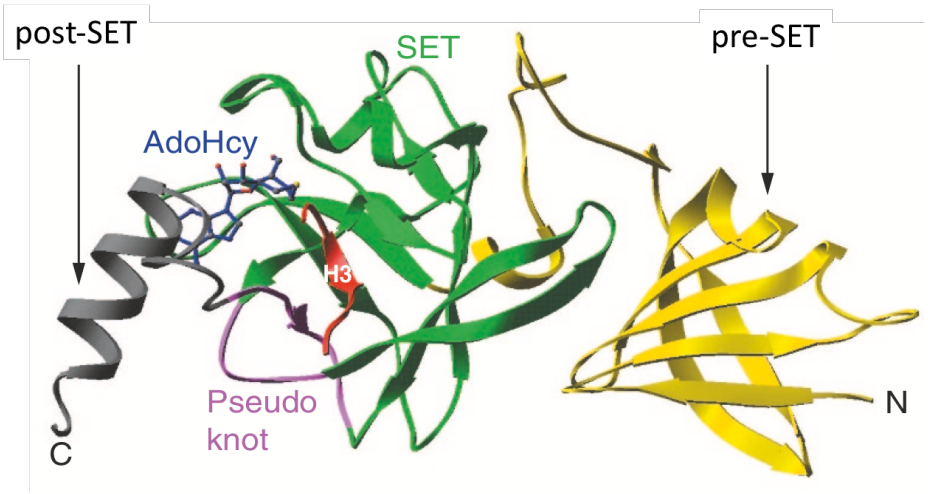


Figure 7: SET-domain structure of the protein lysine methyltransferase SET7/9. The pseudo-knot (knot-like)-structure is shown in purple, the AdoHcy in blue and the H3 peptide in red. The SET-domain is shown in green, the post-SET in grey and the pre-SET in yellow. Adopted and changed from (Dillon et al. 2005).

Cheng et al. (2005) showed by sequence alignment, that the SET-domain contains four important sequence motifs (Cheng et al. 2005). These motifs are named SET motif I (GxG), SET motif II (YxG), SET motif III (RFINHxCxPN) and SET motif IV (ELxFDY). SET motif I, the RFINH sequence of motif III and the tyrosine of motif IV are involved in AdoMet binding. For the catalytic activity of the SET-domain, the tyrosine of motif II is important. To build the hydrophobic binding-channel for a target lysine residue, the motif IV and the CxPN part of motif III are necessary.

The structure of the core SET-domain and its associated domains create a narrow tunnel through which the target lysine and the cofactor AdoMet come in contact, because both substrate and cofactor are bound on opposite sides of the domain. Mutation experiments in the methyltransferase SET7/9 and others showed that alteration of amino acids in the active center leads to an exchange of the methylation degree of the substrate or to an inactivation of the enzyme (Xiao et al. 2003). The SET7/9 can only monomethylate H3K4, whereas the Dim-5 enzyme is able to create H3K9me3 (Tamaru et al. 2003). By comparison of the active site of these enzymes it was shown that they have different amino acids at a specific position located near the ε-amino group of the target lysine. SET7/9 has a Y and Dim-5 contains a F. Mutation experiments in which these two residues where exchanged against each other demonstrated that this F/Y switch determines the methylation degree of the substrate.

Introduction of the Y305F mutation into SET7/9 converted it to a dimethyltransferase. The same was observed for Dim-5 in which the mutation of F281Y converted it to a monomethyltransferase (Cheng et al. 2005). These findings can be explained, because the hydroxyl group of Y prevents the rotation of the monomethylated lysine in the active center and by this block further methylation (Cheng et al. 2005). Interestingly, this F/Y switch has no influence on the substrate specificity.

The catalytic mechanism of PKMTs follows an S_N2 nucleophilic attack of the lone electron pair of the nitrogen atom of lysine onto the electrophilic methylsulphonium cation of AdoMet (Figure 8) (Copeland et al. 2009). The lone electron pair of the lysine residue is produced in the hydrophobic binding pocket where the positive charge of the AdoMet lowers the pK_a of the lysine residue (Zhang and Bruice 2008).

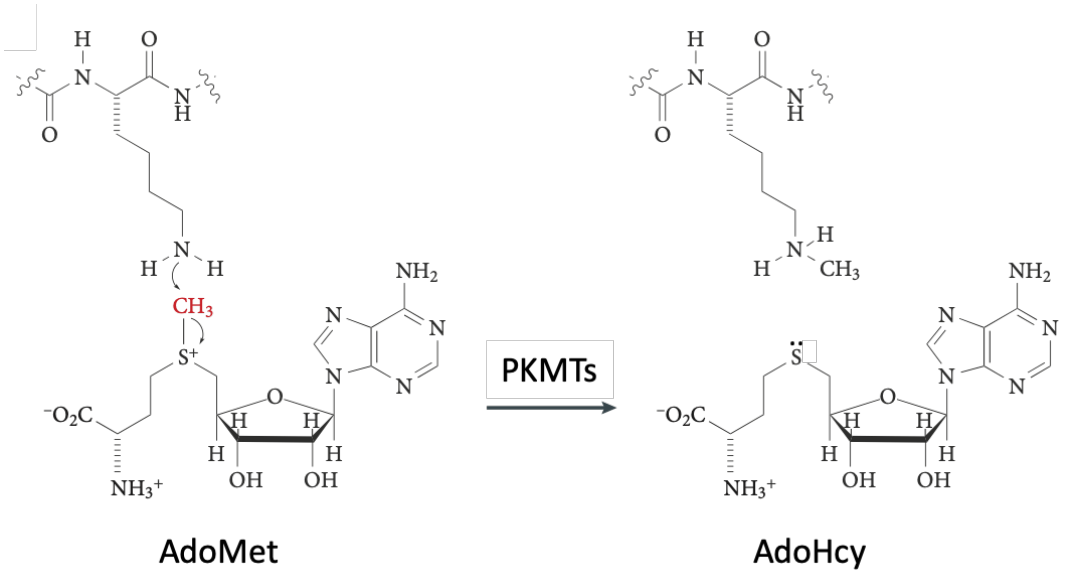


Figure 8: General bimolecular nucleophilic substitution reaction of PKMTs. The lone pair of the nitrogen atom of the side chain amine group of lysine attacks the cofactor AdoMet. As result of the reaction AdoHcy and a methylated lysine is produced. The picture was taken from (Copeland et al. 2009) and changed.

1.6 SUV39H family

The Su(var) 3-9 enzyme was initially identified in *Drosophila melanogaster*. The enzyme family is conserved in many different organisms, for example, the Clr4 methyltransferase of *Schizosaccharomyces pombe* (Nakayama et al. 2001) or Dim-5 of *Neurospora* (Tamaru and Selker 2001) are members of this family. The human genome contains two suppressor of

variegation 3-9 homologs, the SUV39H1 and the SUV39H2 enzymes (Figure 9). The SUV39H enzyme family is a sub-class of the SUV39 PKMTs which also contains the following methyltransferases: G9a (EHMT2/KMT1C), GLP (EHMT1/KMT1D), SETDB1 (KMT1E) and SETDB2 (KMT1F) (Rao et al. 2017). Both SUV39H enzymes catalyze the methylation of H3K9 up to the trimethylation state (Rea et al. 2000).

To investigate the role of SUV39H enzymes in mice, Peters et al. (2001) conducted knock out studies with Suv39h1 or Suv39h2. They found that the mice are viable after the single knock outs. In contrast, when both Suv39h enzymes were knocked out in the mice germ line, most of the embryos were not viable. The mice which were born showed a retarded growth (Rea et al. 2000; Peters et al. 2001). This observation clearly demonstrates that both SUV39H enzymes have a partially redundant and important role in the development of mice.

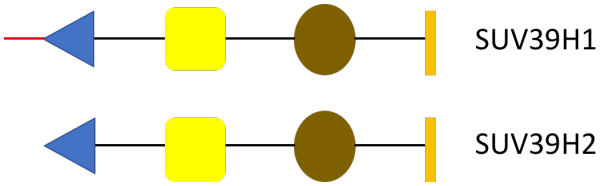


Figure 9: Schematic overview of the domain structure of human SUV39H enzymes. Both enzymes consist of a chromodomain (blue), a pre-SET-domain (yellow), a SET-domain (brown) and a post-SET-domain (orange).

1.6.1 SUV39H1

The SUV39H1 methyltransferase is also known as KMT1A and was the first identified human methyltransferase. SUV39H1 has a size of 412 amino acids and it specifically methylates lysine 9 of histone 3 up to the trimethylation state (Rea et al. 2000). SUV39H1 consists of two domains, a chromodomain in the N-terminal part and a C-terminal catalytic SET-domain (Figure 9).

By deletion of different parts of the SUV39H1 enzyme, Melcher et al. (2000) showed that the N-terminal part till amino acid 118, which includes the chromodomain, is important for the localization of this enzyme to heterochromatin (Melcher et al. 2000). This finding supports the importance of the chromodomain in SUV39H1. The crystal structure of the SUV39H1 chromodomain showed a similar folding as reported for other chromodomains (Figure 10). One main difference compared to other chromodomains is that the chromodomain of

SUV39H1 has a longer α-helix at the C-terminal part (Wang et al. 2012). Previous studies showed the binding of the SUV39H1 chromodomain to H3K9me2/3 (Wang et al. 2012). Through this binding of the SUV39H1 chromodomain to H3K9me2/3, a feedback loop for the generation of H3K9me3 is generated. The binding of the N-terminal part of SUV39H1, including the chromodomain to H3K9me3 induces an allosteric activation of the enzyme, because in the ground-state the enzyme is autoinhibited by its N-terminal part (Muller et al. 2016). An analogous feedback loop was also documented for other enzymes like the polycomb repressive complex 2 (PRC2) (Margueron et al. 2009).

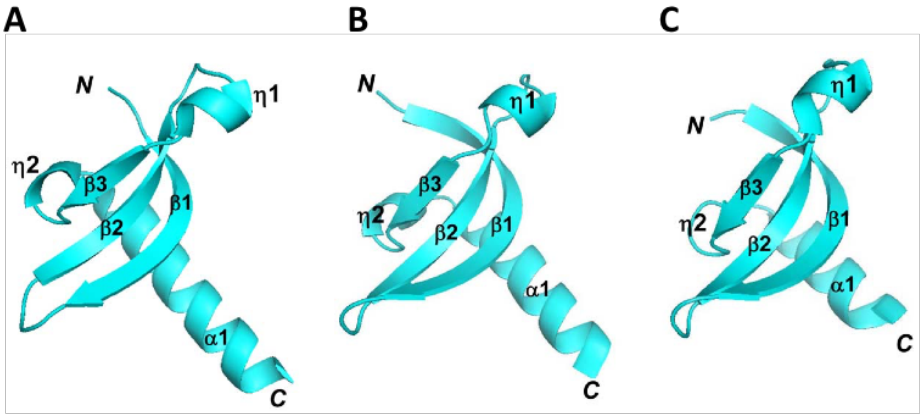


Figure 10: Comparison of different chromodomain structures. A) Panel A) represents the chromodomain of human SUV39H1 (PDB ID: 3MTS). B) Solved structure of the chromodomain of human MPP8 (PDB ID: 3R93). C) Identified structure of the Drosophila HP1 chromodomain (PDB ID: 1KNE). The N and C indicates the N- or C-terminal part of the domain, whereas α, β and η indicate the different motifs. Adopted from (Wang et al. 2012).

The chromodomain of SUV39H1 has another important function. It is able to bind to nucleic acid in which it prefers the binding to single stranded RNA over the binding to double stranded DNA (Shirai et al. 2017). The chromodomain contributes to the localization of SUV39H1 to heterochromatic regions by RNA binding (Johnson et al. 2017; Shirai et al. 2017). The binding to RNA is independent of H3K9me3 binding but the recognition of both signals is important for the SUV39H1 activity in heterochromatin formation (Shirai et al. 2017).

The SET-domain of SUV39H1 contains the catalytic center of the enzyme, but it needs the N-terminal pre-SET-domain and the C-terminal post-SET-domain for its activity as described above (Kouzarides 2002). Rea et al. (2000) introduced mutations into the SET-domain of

SUV39H1, in which they showed that a mutation of histidine 320 to arginine resulted in an increased methylation activity of the enzyme. In contrast, the mutations of H324L, H324K and C326A led to an inactivation of SUV39H1, indicating the importance of these amino acids in the SET-domain structure (Rea et al. 2000).

Interestingly, SUV39H1 is regulated by PTMs. The K266 of SUV39H1 can be acetylated. Vaquero et al. (2007) showed that the NAD-dependent protein deacetylase sirtuin-1 (SIRT1) deacetylates lysine 266 of SUV39H1 and it recruits the methyltransferase to chromatin. Deacetylation of SUV39H1 on K266 leads to an activation of the enzyme. Therefore, loss of SIRT1 has an influence on the H3K9me3 level, which then has a direct impact in HP1 localization to heterochromatin (Vaquero et al. 2007). Moreover, SUV39H1 is methylated by SET7/9 on K105 and K123. This methylation leads to a reduced activity of SUV39H1 and a loss of H3K9me3. Loss of the H3K9me3 mark induces genome instability in cancer cells (Wang et al. 2013; Rao et al. 2017). These findings show that SUV39H1 is regulated by other enzymes as well and that this regulation has an important influence on other processes in the cell.

1.6.2 SUV39H2

Using a mouse model, O´Carroll and colleagues showed that there exists a second H3K9 methyltransferase in mammals, namely Suv39h2 which in mice is encoded on the chromosome 2. The sequence identity of mouse Suv39h1 and Suv39h2 is 59 %, however mouse Suv39h2 has a basic N-terminal extension which is not present in Suv39h1. In mice, Suv39h2 is expressed in adult testis, which is not the case for murine Suv39h1. In immunolocalization studies an enrichment of Suv39h2 was detected at the first meiotic prophase and early stages of spermiogenesis (O'Carroll et al. 2000) overall suggesting that mouse Suv39h2 has a role in the organization of meiotic heterochromatin, which may have an influence in the male germ line. However, as mentioned above, during embryogenesis both Suv39h enzymes have an overlapping expression and function (Peters et al. 2001).

The human SUV39H2 methyltransferase comprises 410 amino acids and, like SUV39H1, it consists of a chromodomain and a catalytically active SET-domain. In contrast to the chromodomain of SUV39H1 (Wang et al. 2012), the function of the chromodomain in SUV39H2 has not been experimentally validated so far. SUV39H2 methylates H3K9 up to the

trimethylation state. One study reported that it prefers the unmethylated peptide over the mono- and dimethylated peptide as substrate (Allali-Hassani et al. 2012). Another study showed that the SUV39H enzymes prefer the H3K9me1 as substrate *in vivo* (Peters et al. 2003). This discrepancy regarding substrate preference was addressed in this thesis.

The catalytic center of SUV39H2 consist of a SET-domain, including N-SET and i-SET-subdomains, which is surrounded by pre- and post-SET regions (Figure 11). The structure of the SUV39H2 SET-domain has been solved (Wu et al. 2010) showing that the post-SET and i-SET domains build a groove in which the substrate peptide is bound. This binding groove has a negative surface charge which fits to the H3 peptide because of its high amount of lysine and arginine residues. The post-SET-domain is important as well for peptide binding, because it closes the peptide binding channel upon binding of the substrate (Wu et al. 2010). In previous peptide methylation studies using mouse GST-tagged Suv39h2 it was shown that preexisting histone tail modifications have an influence on the enzyme activity. For example, S10 phosphorylation and K9 acetylation inhibited Suv39h2, and the K4L mutation and K14ac led to a decrease in activity (O'Carroll et al. 2000). This indicates that other PTMs on histone tails can influence the activity of SUV39H2.

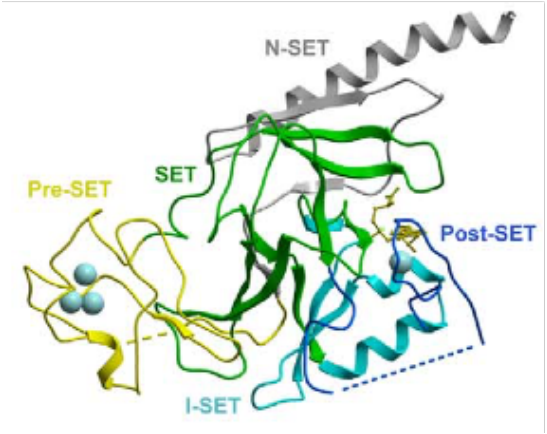


Figure 11: Structure of human SUV39H2 in complex with AdoMet (yellow sticks). The different domains of SUV39H2 are colored in yellow (pre-SET), cyan (i-SET), blue (post-SET), grey (N-SET) and green (SET). The picture was taken from (Wu et al. 2010).

Previously, it was shown that SUV39H2 methylates the non-histone target LSD1 at K322 and the histone variant H2AX at K134 (Sone et al. 2014; Piao et al. 2015). This finding indicated

that SUV39H2 could have other functions in human cells beyond the methylation of histone 3. SUV39H2 is also implicated in cancer. For example, in acute lymphoblastic leukemia (ALL) it has been reported to be overexpressed. Mutonga et al. (2015) showed in ALL cell lines that knockdown of SUV39H2 led to a decrease in cell viability. Interestingly, overexpression of SUV39H2 was shown to lead to a better protection against chemotherapy reagents (Mutonga et al. 2015). All these findings suggest that SUV39H2 could be a drug target for cancer treatment.

1.7 SETD2

The SETD2 PKMT was first identified as huntingtin interacting protein 1 (HYPB, HIP1) (Rega et al. 2001; Duns et al. 2010). It has a size of 230 kDa corresponding to 2564 amino acids (Xie et al. 2008). The reason for the initial name of SETD2 was its involvement in the pathogenesis of Huntington disease by its interaction with the Huntingtin protein (Xie et al. 2008). SETD2 is another example of a PKMT, which contains a catalytically active SET-domain. This enzyme introduces up to three methyl groups on its histone target H3K36 (Eram et al. 2015). This ability for trimethylation of H3K36 makes SETD2 unique, because no other trimethyltransferase for this lysine is known so far (Edmunds et al. 2008; Eram et al. 2015). As mentioned above, the methylation of H3K36 is related to actively transcribed genes. The H3K36 mono- and dimethylation are mainly introduced by the SET-domain containing proteins NSD1 and NSD2, which are members of the NSD enzyme family. In 2016, Park et al. (2016) showed that SETD2 is able to methylate a non-histone target as well. They demonstrated the methylation of α-tubulin at lysine 40 by SETD2 indicating that SETD2 has substrates in the chromatin and in the cytoskeleton. Indeed, SETD2 is crucial for mitosis and cytokinesis, because deletion of SETD2 leads to mitosis spindle and cytokinesis defects (Park et al. 2016). The SETD2 enzyme interacts also with p53 and regulates its activity. This interaction has also impact on other genes like for example *puma* and *fas*. However, it is not known if SETD2 also methylates p53 (Xie et al. 2008). SETD2 is very important in many processes, because of its methyltransferase activity on H3K36, which is a signal or binding point for other proteins. For example, SETD2 plays an important role in the regulation of DNA repair in the G1 and early S phase and DNA double-strand break repair (Li et al. 2013a; Carvalho et al. 2014).

Other important domains of SETD2 are the Associated with SET-domain (AWS), the post-SET-domain, the Tryptophan-tryptophan (WW)-domain (Sun et al. 2005), the Coiled-coiled-domain (CC) (Hacker et al. 2016) and the SRI-domain (Kizer et al. 2005) (Figure 12). SETD2 is an example of a methyltransferase, which has an AWS-domain instead of a pre-SET-domain. The WW-domain is positioned at the C-terminal part of SETD2 and is known as a protein-protein interacting motif, which prefers a proline rich region for interaction (Sudol and Hunter 2000; Sun et al. 2005).

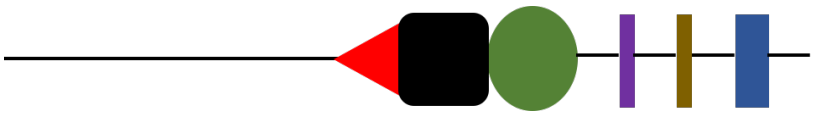


Figure 12: Schematic representation of mammalian SETD2 and its domain architecture. Red indicates the AWS-domain, the black box shows the SET-domain, the green circle indicates the post-SET-domain, the CC-domain is highlighted in purple, the WW-domain is marked in brown, and the blue box represents the Set2-Rpb1 interacting-domain of SETD2.

Yang et al. (2016) solved the structure of the SETD2 catalytic domain (CD) in complex with an H3K36M (29-42) peptide and AdoHcy. SETD2-CD forms several contacts to the amino acids 29 to 35 and 37 to 42 of the H3 tail substrate peptide. The first part of the peptide 29-35 contains mostly uncharged amino acids which are bound by SETD2-CD as an "pseudo-β-sheet" between the loop connecting α6 and β5 of the SET-domain and the L_{IN} that combines the SET-domain and the post-SET-domain. The second fragment 37-42 consist of amino acids with basic and large side chains which are bound into an acidic groove of SETD2-CD. In addition, the amino acids 39-41 can stabilize peptide binding by the formation of hydrogen bonds with E1636 and T1637 of SETD2-CD. The amino acids 33-36 including the target lysine and aliphatic amino acids are located in a hydrophobic environment constituted by different amino acids of SETD2-CD (Yang et al. 2016). The importance of the amino acids which contribute to the hydrophobic environment was demonstrated by mutational analyses showing that mutation of these amino acids leads to inactivation of SETD2 (Yang et al. 2016).

Previous investigations revealed many mutations of the SETD2 gene in cancer tissues, for example in pediatric high grade gliomas and adult high grade gliomas (Fontebasso et al. 2013). Another cancer type in which SETD2 plays an important role is cell renal cell carcinoma (cRCC).

In this cancer type frameshift, non-sense and missense mutations in SETD2 were observed, suggesting a loss-of-function mechanism. This model was supported by the observation that the level of H3K36 trimethylation is reduced in cRCC but the H3K36me2 level is not affected (Duns et al. 2010).

Another mechanism of SETD2 inhibition is by expression of the so-called H3K36M oncohistone, which contains a K36M mutation frequently found in chondroblastomas (Behjati et al. 2013). The H3K36M mutation inhibits the methyltransferase activity of SETD2, because of its ability to block the active center (Figure 13) (Yang et al. 2016). It was shown that NSD2, another H3K36 methyltransferase, is inhibited by H3K36M as well (Lu et al. 2016).

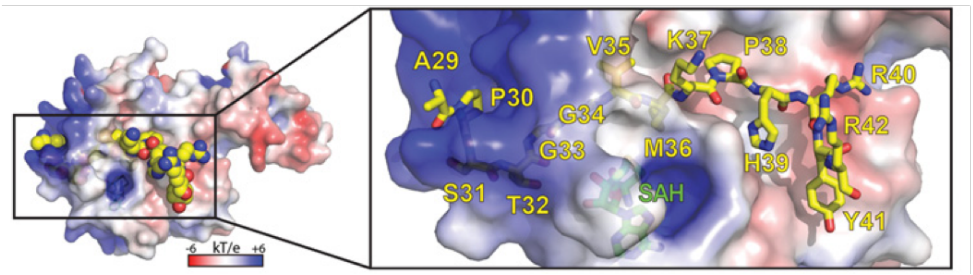


Figure 13: Recognition of H3K36M peptide by SETD2 catalytic domain (amino acids 1434-1711). The yellow sticks represent the H3K36M peptide and the green sticks show the *S*-Adenosyl-*L*-homocysteine (SAH, AdoHcy). Adopted from (Yang et al. 2016).

1.8 NSD1

The NSD1 PKMT (KMT3B) has a size of 1696 amino acids. It belongs to the NSD subfamily of SET-domain containing PKMTs. Other members of the NSD family are NSD2 (WHSC1/MMSET) and NSD3 (WHSC1L1). Previously, it was reported that NSD1 methylates histone 3 at lysine 36, histone 4 at lysine 20 and a non-histone target p65, which is a subunit of the nuclear factor-κB (NFκB) (Rayasam et al. 2003; Lu et al. 2010). However, the methylation of H4K20 and p65 could not be confirmed by other groups (Yang et al. 2008; Kudithipudi et al. 2014b). It is well documented that NSD1 introduces mono- and dimethylation on H3K36 (Lucio-Eterovic et al. 2010; Qiao et al. 2011), but it is not able to introduce a third methyl group. Kudithipudi et al. (2014) showed that lysine 44 of H4, lysine 168 of H1.5 and H1.2, lysine 169 of H1.3 and other non-histone peptides are methylated by NSD1 (Kudithipudi et al. 2014b).

The catalytically active part of NSD1 consists of an AWS-domain, a SET-domain, a post-SET-domain and an N-terminal fragment of about 40 amino acid length (Figure 14) (Qiao et al. 2011). This catalytic domain is placed in the protein between several PHD-domains and two PWWP-domains (Kudithipudi et al. 2014b). NSD1 contains an autoinhibitory post-SET-loop (amino acids 2060-2066), which can block the substrate binding site. NSD1 undergoes a spontaneous rearrangement in which it opens the binding site for an unmethylated lysine. Upon binding of lysine, the post-SET-loop moves further away giving the possibility to introduce two methyl groups onto the target lysine (Qiao et al. 2011).

The NSD1 PKMT is linked to many diseases, e.g. the Sotos syndrome. This hereditary disease is characterized by a developmental and cognitive disorder and a typical facial abnormality (Leventopoulos et al. 2009; Pasillas et al. 2011). Sotos syndrome patients contain many mutations in the catalytic domain of NSD1, which have an influence on cofactor binding and the folding of this PKMT.

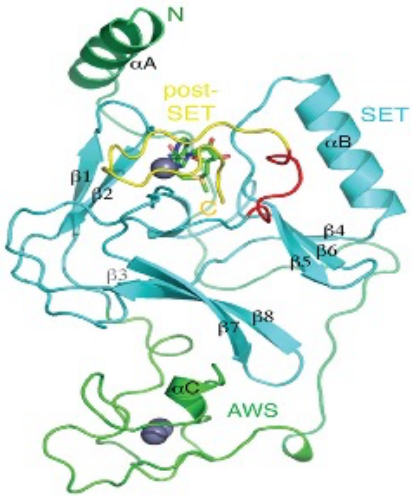


Figure 14: Structure of the binary complex of the NSD1 catalytic domain with AdoMet represented as stick model. The post-SET-domain (yellow), the SET-domain (cyan), the AWS-domain (light green), the post-SET loop (red) and the N-terminal extension (dark green) are represented as well. The bound zinc ions are represented as grey dots. The picture was taken from (Qiao et al. 2011).

1.9 RomA

RomA is a bacterial PKMT that was identified in the Gram-negative bacteria *Legionella pneumophila* (*L. pneumophila*) of the strain Paris, which cause the so-called Legionnaire´s disease. In humans, Legionnaire´s disease infection can occur through the uptake of water contaminated with bacteria (Fraser et al. 1977). After the bacteria have entered the human body, a Type-4 secretion system (T4SS) is important for *L. pneumophila* to infect the human host cells, which then secretes several proteins including RomA into the host cell (Marra et al. 1992; Berger and Isberg 1993; Segal et al. 1998).

The RomA methyltransferase is encoded by the gene *lpp1683* and it has a size of 532 amino acids. It contains a eukaryotic-type SET-domain, which is responsible for the methyltransferase activity. Like in eukaryotes, the SET-domain of RomA contains a GxG motif that is important for AdoMet binding. The sequence alignment of the *L. pneumophila* RomA SET-domain with eukaryotic SET-domains, like for example the human SUV39H1 is shown in Figure 15, where the partially conserved amino acids of the different enzymes are highlighted by blue boxes.

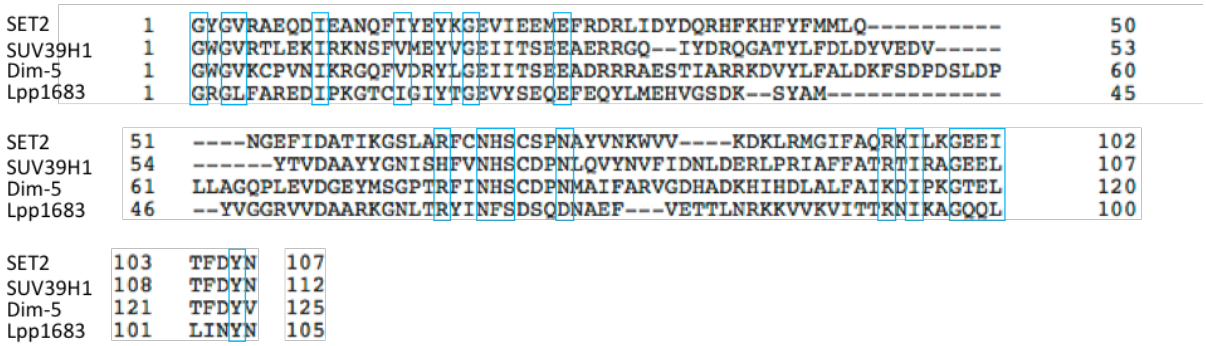


Figure 15: Superimposition of the SET-domain of RomA with eukaryotic SET-domains. Sequence alignment of SET-domains of *Saccharomyces cerevisiae* (SET2, P46995), *Homo sapiens* (SUV39H1, NP_003164.1), *Neurospora crassa* (Dim-5, AAL35215.1) and *Legionella pneumophila* RomA (Lpp1683, CAH12835.1). The partially conserved amino acids of the SET-domains are highlighted by blue boxes. The sequence alignment was prepared with the UniProt sequence alignment tool.

It had been shown that RomA trimethylates human H3K14 (Rolando et al. 2013). This histone methylation mark was not previously identified in mammals. However, it was known that H3K14 can be acetylated, for example by the lysine acetyltransferase 2A (GCN5) (Grant et al.

1999). This acetylation is blocked by the methylation of H3K14 via RomA (Rolando et al. 2013). This change of post-translational modifications on a specific amino acid in the histone tail leads to changes in gene expression, because H3K14 acetylation is associated with an active promotor state (Karmodiya et al. 2012).

Another bacterial methyltransferase, LegAS4 of the *L. pneumophila* strain Philadelphia, was reported to methylate human H3 at lysine 4 and lysine 9 (Li et al. 2013b). In the following year, Rolando and Buchrieser showed that LegAS4 methylate H3K14 during infection as well (Rolando and Buchrieser 2014). The structure of LegAS4 in complex with AdoMet was solved. Overlay of the LegAS4 structure with human PKMTs e.g. GLP or SET8 showed a highly similar architecture (Figure 16) (Son et al. 2015). The structure shows that the amino acids E203 and E206 of LegAS4 are important for the substrate-binding-site, which presumably form contacts to positively charged amino acids of the substrate peptide. As the SET-domains of RomA and LegAS4 have a very high sequence identity they must have the same or similar structure, recognition of the target lysine and methylation mechanism (Son et al. 2015).

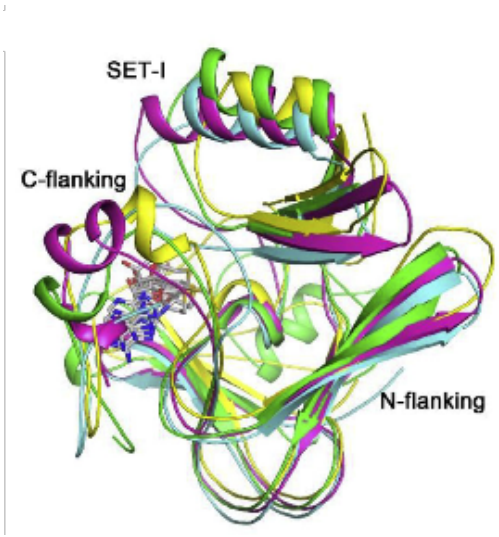


Figure 16: SET-domain structure of LegAS4. Overlay of the SET-domain structures of LegAS4 (green) and the human PKMTs GLP (yellow, PDB ID: 3MO0), SET8 (magenta, PDB ID: 3F9W) and ASH1L (cyan, PDB ID: 3OPE). The picture is taken from (Son et al. 2015).

2. Aims of the study

The investigation of the enzymatic properties and specificity of PKTMs and their influence on other proteins are important research topics mandatory to gain more insights into the regulatory mechanisms in human cells. Until now not all protein lysine methylation sites are known for most PKMTs and for many methylated sites the enzymes that introduce the methylation have not yet been identified (Figure 17). One approach to increase our understanding of PKMTs is to analyze their substrate specificity in great detail. This provides important information about the molecular interaction of PKMTs with their substrates. Moreover, the knowledge of the substrate specificity of PKMTs is important, because based on this information a search for possible novel substrates can be performed.

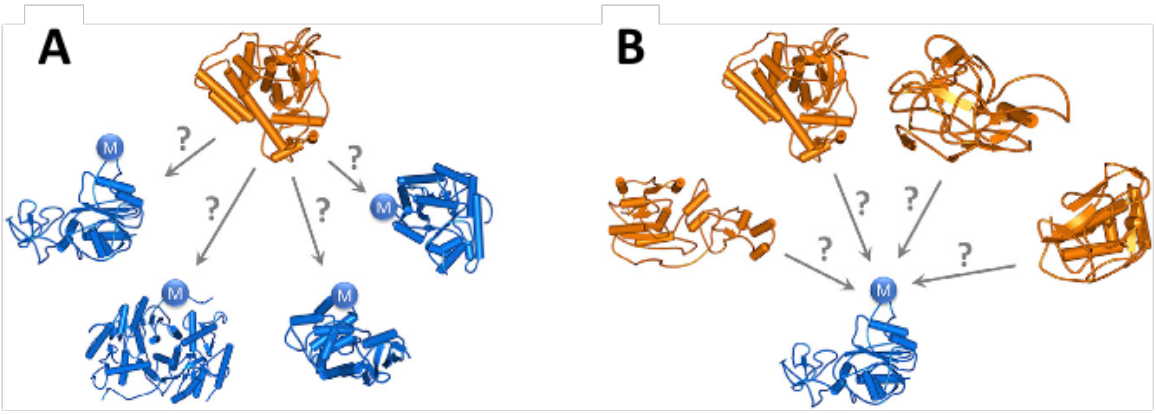


Figure 17: Overview of the open questions in substrate and PKMT assignment. A) For most PKMTs (orange) it is likely that additional not yet identified methylation substrates (blue) exist. B) For many known methylation events (blue), the PKMT (orange) which sets the mark is unknown. The picture has been adopted from (Kudithipudi and Jeltsch 2016).

To avoid wrong PKMT-substrate assignments Kudithipudi and Jeltsch proposed guidelines for the identification and validation of novel PKMT substrates (Kudithipudi and Jeltsch 2016). These guidelines consist of seven rules demanding to include positive controls to compare the relevance of the methylation of a novel substrate. Also, negative controls should be included like the target lysine mutation to confirm the methylation of the predicted target site and experiments with PKMT variant containing mutations which inactivate the PKMT. It was also recommended to confirm that the specificity of the PKMT fits to the sequence of possible methylation sites. If antibodies are used for the detection of methylation events, it was

highlighted that it is important to validate the antibodies before their application. Another important step is the connection of *in vitro* results with cellular studies to show the biological relevance in cells. Lastly not only qualitative data are important but rather also quantitative data. Whenever appropriate these guidelines were followed in the present thesis.

The identification of the complete list of its protein substrates of every PKMT is an essential step in the understanding of the involvement of PKMTs in different regulatory mechanisms. The main aim of the present thesis was the investigation of four different PKMTs:

- SUV39H1, a human PKMT known to methylate histone 3 at K9
- SUV39H2, a human PKMT known to methylate histone 3 at K9
- SETD2, a human PKMT known to methylate histone 3 at K36
- RomA, a bacterial PKMT from *Legionella pneumophila* known to methylate histone 3 at K14

Depending on available previous data in the work group, the enzymes should be expressed, purified and their substrate specificity investigated. Additionally, for some of the enzymes a basic biochemical and enzymatic characterization was planned. The next aim of this thesis was the identification and validation of non-histone substrates of these PKMTs based on the beforehand identified substrate specificity motifs. The identified non-histone substrate candidates should be investigated at peptide and protein level and the methylation of the predicted target lysine residue should be verified and if possible information about the methylation of the novel non-histone substrates in cells should be obtained. Another aim of the present thesis was the validation of already reported substrates in the context of the new data obtained here. Such validation is important because in the last years there were examples of wrongly attributed substrates to PKMTs, e.g. methylation of Numb by SET8 (Weirich et al. 2015).

3. Results and Discussion

In the present PhD thesis, different PKMTs have been investigated and the obtained results will be summarized in the following chapters. The subproject specific discussion of the results is also included in this chapter. As mentioned above, PKMTs are enzymes which catalyze the transfer of a methyl group from the cofactor AdoMet to specific lysine residues in peptide and protein substrates. The SET-domain is responsible for the transfer of the methyl group. In addition, PKMTs contain other important domains like for example a chromodomain in SUV39H1 and SUV39H2. In the past, the unstructured tails of histones were described as the main PKMT substrates, but in the last couple of years it has been shown that many PKMTs have additional targets, also known as non-histone targets. G9a is one example of a PKMT, which methylates a non-histone target called Widely interspaced zinc finger motifs protein (WIZ) (Rathert et al. 2008a). In this thesis, other PKMTs were investigated and a search for potential new substrates of these PKMTs at peptide and protein level was performed.

For the identification of novel non-histone targets of SUV39H2, RomA and SETD2, the substrate specificity profile of these enzymes was investigated using radioactive peptide array methylation experiments. Based on the determined profiles, the Scansite database (Obenauer et al. 2003) was used to search for potential non-histone targets present in the human proteome. Methylation of the identified potential non-histone targets was first investigated at peptide level and later confirmed at protein level *in vitro*. The target site methylation of the different PKMTs was validated by site-directed mutagenesis in which the target lysine was mutated to arginine. Based on the obtained results, further experiments were conducted, as for example cellular studies for SUV39H1. For RomA, methylation of one identified non-histone target was investigated during infection with *L. pneumophila* by our collaborators.

Another part of this study concerned the validation of a known PKMT non-histone substrate (H2AX) which was previously reported to be methylated by SUV39H2 (Sone et al. 2014). This type of validation experiments is important, because in the past a lot of wrong assignments between non-histone targets and PKMTs have been published (Weirich et al. 2015; Weirich et al. 2016), which hinder future research.

A final part of this thesis studied the inhibition of PKMTs. In the human genome, many PKMTs are responsible for H3K36 methylation. As described above, the NSD1 and NSD2 are responsible for H3K36 mono- and dimethylation, whereas SETD2 is able to introduce the trimethylation state of H3K36. Previously, it was documented that mutation of lysine 36 to methionine blocks the methyltransferase activity of SETD2 and NSD2 (Lu et al. 2016; Yang et al. 2016). Two of the three H3K36 methyltransferases (NSD1 and SETD2) were investigated in inhibition kinetic studies using H3K36M peptide inhibitors, especially NSD1, for which this inhibition had not been reported so far.

3.1 Investigation of the SUV39H PKMTs

3.1.1 Activity and specificity of SUV39H2

(manuscript 1 in the attachment of this thesis)

Schuhmacher MK, Kudithipudi S, Kusevic D, Weirich S, Jeltsch A. (2015) *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms 1849(1):55-63. doi: 10.1016/j.bbagrm.2014.11.005.*

The SUV39H2 enzyme consists of 410 amino acids. It is the human homolog of the *Drosophila* Su(var) 3-9 enzymes together with its paralog SUV39H1. SUV39H2 is also called KMT1B and is mainly found in heterochromatin regions. It catalyzes the methylation of histone 3 at lysine 9 up to the trimethylation state using unmethylated H3K9 as a preferred substrate (Allali-Hassani et al. 2012). At the beginning of this thesis the substrate specificity profiles had already been investigated for many PKMTs (Rathert et al. 2008b; Dhayalan et al. 2011; Kudithipudi et al. 2014b). However, for SUV39H2, the substrate specificity profile and potential other targets beside of histone 3 remained an open research question, which was approached in this thesis. Three different constructs of SUV39H2 were used (Figure 18): a His_6-tagged full-length (amino acids 1-410) or SET-domain SUV39H2 (amino acids 112-410) and a GST-tagged SET-domain (amino acids 112-410) construct, which all were cloned with N-terminal tags. The SET-domain constructs were cloned by Dr. Srikanth Kudithipudi. All enzymes were purified by affinity chromatography (manuscript 1, Figure 1 B).

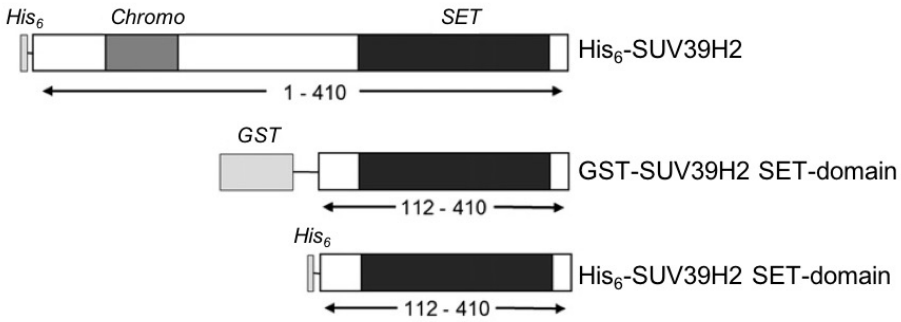


Figure 18: Schematic representation of the SUV39H2 constructs used in this thesis. The numbers indicate the cloned boundaries in amino acids. The figure was taken and changed from (manuscript 1, Figure 1 A).

Initial radioactive methylation experiments using the H3 protein as substrate showed that the GST-tagged SUV39H2 SET-domain had the highest activity, followed by the His_6-tagged SET-domain, and the lowest activity was observed for the His_6-tagged full-length SUV39H2 protein (manuscript 1, Figure 1 C). Radioactive based peptide array methylation experiments demonstrated that there is no difference in substrate specificity between the His_6-tagged full-length and the GST-tagged SET-domain SUV39H2 proteins (manuscript 1, Figure 1 D). Both enzymes showed a strong readout of the amino acids R8, S10 and G12 surrounding the K9 target, which was indicated by a loss of methylation after mutation of these amino acids to alanine. This demonstrates that the GST-tagged SET-domain of SUV39H2 is suitable for the activity and specificity analysis conducted here.

As a next step, a substrate specificity analysis of the GST-tagged SUV39H2 SET-domain was performed using H3 (2-16) as template sequence (Figure 19). For this, fifteen amino acid long peptides were synthesized on a cellulose membrane using the SPOT synthesis method. For each amino acid of the template sequence, single amino acid mutations were created in which each individual residue of the template sequence H3 (2-16) was exchanged against all natural amino acids. The 20 natural amino acids are represented in the vertical axis in Figure 19, whereas the horizontal axis represent the template sequence H3 (2-16). The membranes containing the substrate specificity scan were incubated with SUV39H2 and radioactively labeled AdoMet, and the transfer of the methyl groups from the cofactor AdoMet to the peptides was detected by autoradiography. The peptide array methylation experiments were performed in triplicates. Afterwards the activity was normalized and the data of three independent replicates was averaged (Figure 19 A). For a better visualization of the obtained

substrate specificity profile, the discrimination factors were calculated (Kudithipudi et al. 2014a) showing the preference of SUV39H2 for different amino acids at specific positions (Figure 19 B).

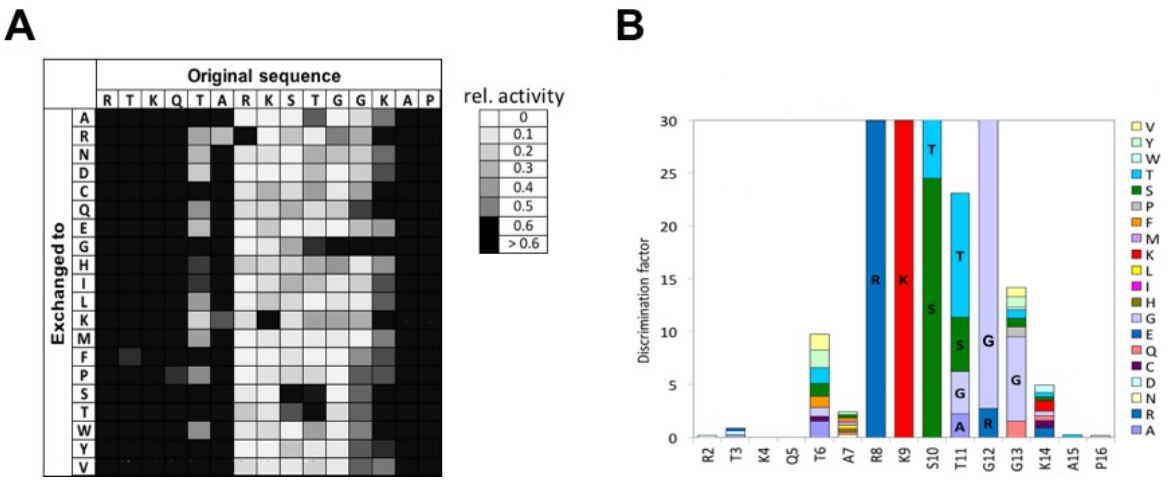


Figure 19: Substrate specificity analysis of SUV39H2 SET-domain. A) Results of peptide array methylation experiments with SUV39H2 based on the quantification of three different experiments. The vertical axis represents the 20 amino acids which were used to create single amino acid mutations of the H3 (2-16) template sequence shown on the horizontal axis. The signal intensity is indicated in a greyscale on the right side. B) Discrimination factors of SUV39H2 substrate recognition showing the amino acid preference of SUV39H2 in different positions. Both pictures were taken from (manuscript 1, Figure 2).

These data revealed that the SUV39H2 SET-domain prefers a core sequence motif ranging from the -1 to +4 position of the H3 tail when K9 is annotated as position zero. It shows that R8 of H3 is very important for the specificity of SUV39H2. Other important amino acids for SUV39H2 recognition are S10, T11 and G12 (positions +1 to +3 of the H3 tail), because at these positions only few amino acids apart from the natural amino acids are accepted by SUV39H2. The S in position +1 can only be exchanged against T, whereas the T in position +2 can be exchanged against S, G and A. For the G in position +3 only an R is accepted by SUV39H2. Additionally, SUV39H2 showed a more relaxed but still detectable recognition of T6, A7, G13 and K14. The specificity profile was as follows (the natural amino acids are underlined, the target lysine is marked red and X represents any amino acid):

R - K - (S>T) - (T>S, G, A) - (G>R) - (G>X)

Based on the highly preferred amino acids of the substrate specificity profile investigation ((A, C, G, F, S, T, Y, V) - X - R - K - S - (T, S) - G - G), a Scansite search for possible non-histone targets in the human proteome was performed, but no candidate proteins were retrieved. Because of this, a more relaxed specificity motif was used, in which the -3 and +4 positions were not included and in the +1 position T and in +2 A and G were allowed as well. With this profile, a second Scansite search, was performed, retrieving seven possible non-histone substrate candidates. To test the methylation of the corresponding peptides, fifteen amino acid long peptides were synthesized on a cellulose membrane by SPOT synthesis method and methylated by SUV39H2 using radioactively labeled AdoMet as cofactor. The signal of the transferred radioactively labeled methyl groups was detected by autoradiography. However, all investigated possible targets were either not methylated or methylation was much weaker than that of the known H3K9 substrate (manuscript 1, Figure 3).

It is known that the H3 tail carries many different post-translational modifications like methylation, phosphorylation or acetylation of different residues. The influence of these modifications on the SUV39H2 SET-domain activity was investigated using CelluSpot peptide arrays containing H3 tail peptides with these post-translational modifications (manuscript 1, Figure 4). The CelluSpot peptide array methylation was conducted as described above for SPOT peptide arrays. The investigation revealed that the activity of SUV39H2 was abolished by trimethylation of K9, which is trivial, as this prevents further methylation. In addition, acetylation of K9 also inhibited the SUV39H2 activity. Moreover, it was shown that phosphorylation of H3S10 and H3T11 blocks the activity of SUV39H2. A reduction in SUV39H2 activity was also observed by H3R8 methylation.

To investigate the preferences of SUV39H2 regarding the initial methylation state of the substrate at the target lysine, the efficiency of the methylation of H3K9 unmethylated, H3K9me1 and H3K9me2 substrates on the CelluSpot arrays were compared. H3K9 unmethylated and H3K9me1 were equally strong methylated indicating that both are good substrates for SUV39H2. The methylation intensity of H3K9me2 was weaker, which shows that SUV39H2 introduces H3K9me3, but H3K9me2 is not a preferred substrate of SUV39H2. Since CelluSpot peptide array gave a significant difference in the substrate methylation between H3K9me1 and H3K9me2, the product pattern of GST-tagged SUV39H2 SET-domain was

further investigated using mass spectrometry (manuscript 1, Figure 5 A, B, Suppl. Figure S1). For this experiment, the SUV39H2 SET-domain was incubated with the respective H3 (1-19 or 1-17) peptides in solution and unlabeled AdoMet was used as cofactor. At specific time points samples were taken. The obtained data confirmed, that SUV39H2 introduces all three methylation states on H3K9 and the preferred substrate of SUV39H2 is the unmethylated H3K9 peptide (manuscript 1, Figure 5 B, Suppl. Figure S1). In addition, this showed that the introduction of the first methyl group was fast, and the conversion of H3K9me1 into H3K9me2 even faster, but the addition of the third methyl group was very slow.

Thereafter, the N324K mutation of SUV39H2 SET-domain was investigated. This mutation was previously described in Labrador Retrievers, where it leads to a hereditary disease (Jagannathan et al. 2013). To investigate the effect of the N324K mutation, the mutant protein was cloned by site-directed mutagenesis and purified by affinity chromatography (manuscript 1, Figure 6 A). Afterwards equal amounts of wildtype and mutant proteins were used for methylation reactions showing that the N324K mutation leads to inactivation of the enzyme (manuscript 1, Figure 6 B). As a next step, circular dichroism (CD) was used to determine the folding of both proteins to show that the inactivation of the SUV39H2 N324K mutant is not due to unfolding (manuscript 1, Figure 6 C). Although the CD spectra showed significant deviations in the folding of both proteins, no unfolding for the mutant protein was detected. A super-position of SUV39H2 and two other PKMTs (Dim-5 and G9a) showing the localization of the N324K mutation and the recognition of the H3 peptide by amino acids in SUV39H2 is provided in the manuscript as well (manuscript 1, Figure 7).

3.1.1.1 Discussion

In the present study a comparison of the activity of the SET-domain construct of SUV39H2 and its full length construct showed that the SET-domain is more active than the full-length SUV39H2. This observation is maybe explained by a previous observation of Muller et al. (2016). They reported that the N-terminal part of SUV39H1 has an autoinhibitory function which is only relieved by nucleosome interaction. This observation suggests that the N-terminal part of SUV39H2 could have a similar autoinhibitory function as reported for SUV39H1. In a follow up experiment, the specificity investigation of the full-length and

SET-domain constructs of SUV39H2 showed no difference, suggesting that the chromodomain of SUV39H2 has no influence on its specificity.

As it was shown above that the SET-domain of SUV39H2 has the same specificity as the full-length construct, the substrate specificity motif of SET-domain SUV39H2 was determined and used for the identification of novel non-histone substrates. The search revealed seven possible candidates but experiments showed that no possible non-histone substrate was equally strong methylated compared to the positive control H3K9. This observation indicates that H3K9 is the main target of SUV39H2 and that this enzyme has a very specific recognition pattern of the target site. A closer look into the structure of SUV39H2 reveals that the amino acids D307 and D309 of the SET-domain are very important for the recognition of the peptide. D307 of SUV39H2 can form a hydrogen bond with the R8 of H3. In addition, D309 can form a hydrogen bond with the S10 of H3. These interactions strengthen the binding of the substrate to the enzyme. However, the sequences of the retrieved peptides are largely consistent with the motif, which implies that further interactions play a role in the recognition of the methylation target by SUV39H2. The influence of other interactions in substrate recognition has to be investigated in further experiments. Previous studies have also shown that some PKMTs prefer histones as targets. Examples of PKMTs, which mainly recognize and methylate histone targets are SET8 and NSD1 (Kudithipudi et al. 2012; Kudithipudi et al. 2014b) and the data obtained here suggest that SUV39H2 also belongs to this group.

Since SUV39H2 substrate specificity showed a strong readout of R9, S10 and T11 it was interesting to investigate the influences of different PTMs on these residues on its enzymatic activity. It was shown that phosphorylation of S10 and T11 completely abolished the methyltransferase activity of SUV39H2, indicating that these amino acids in unmodified forms are important contact points for the enzyme. In addition, modifications of R8 leads to a reduced activity of SUV39H2 which is in agreement with the substrate specificity investigations and also highlights the importance of the recognition site for SUV39H2. These observations are in agreement with other results, because G9a which also has a strong readout of R8 is blocked by an asymmetric methylation of R8 at peptide level (Rathert et al. 2008a). Similar to SUV39H2 the PKMTs SUV39H1 and Dim-5 were inactivated by modifications of S10 and T11 (O'Carroll et al. 2000; Rea et al. 2000; Rathert et al. 2008b). This shows that

SUV39H2 resembles the other members of the SUV39 methyltransferase family in the point of the recognition of R8, S10 and T11. Another observation was that K9 acetylation inhibited the SUV39H2 activity, which is also in agreement with previous studies (Rea et al. 2000) showing that PKMTs generally do not accept acetyllysine as substrate for methylation. This observation may be explained by the fact that the keto group reduces the electron density at the nitrogen atom.

Previously, three groups published different results for the substrate specificity of SUV39H2. Peters et al. (2003) described that both SUV39H enzymes prefer the monomethylated H3K9 as substrate (Peters et al. 2003). This observation was supported by Pinheiro et al. (2012). They showed that depletion of Prdm3 and Prdm16, which introduce H3K9 monomethylation, leads to loss of H3K9 trimethylation (Pinheiro et al. 2012). By this, it can be concluded that H3K9me1 is used as substrate by the SUV39H enzymes to produce H3K9me3. However, the study of Allali-Hassani et al. (2012) showed that SUV39H2 prefers the unmethylated H3K9 as substrate. This discrepancy in substrate preference of SUV39H2 was addressed in the present thesis by CelluSpot and MALDI experiments. The obtained results, showed that SUV39H2 clearly prefers the unmethylated H3K9 as substrate. Furthermore, it was shown that SUV39H2 is able to introduce trimethylation on H3K9 *in vitro*. Another observation was that the conversion from di- to trimethylation of H3K9 by SUV39H2 was very slow. The obtained results and the results from Allali-Hassani et al. (2012) were in general in a good agreement. However, the Pinheiro et al. (2012) results could not be excluded because *in vivo* other factors may have an influence on SUV39H2 specificity.

Another research topic of SUV39H2 was the investigation of the N324K mutation which was not described in the human protein so far. It could be shown that this mutation inactivates the SUV39H2 enzyme activity which was not related to an unfolded protein. A closer look into the structure of SUV39H2 reveals that N324 can form a hydrogen bond to E168 that is important for the connection of two β-strands in the SET-domain fold. The exchange of N324 to lysine which is a bigger amino acid may lead to structural changes. This observation indicates that this connection is important for the methyltransferase activity of SUV39H2.

3.1.2 Investigation of H2AX methylation by SUV39H2

(manuscript 2 in the annex to this thesis)

Schuhmacher MK, Kudithipudi S, Jeltsch A. (2016) *FEBS Letters 590(12):1713-9. doi: 10.1002/1873-3468.12216.*

In 2014, a new non-histone target of SUV39H2 was reported by Sone et al. (2014). They published that lysine 134 of H2AX is methylated by SUV39H2 up to the dimethylation state. However, the comparison of the sequence of H2AX (manuscript 2, Figure 1 A) and the above-mentioned substrate specificity profile of SUV39H2 (Figure 19) revealed a big discrepancy, because apart from the target lysine only the T at position +2 in H2AX was matching to the SUV39H2 substrate specificity profile. Therefore, this sequence discrepancy and methylation of H2AX by SUV39H2 was analyzed in more detail.

Initially methylation of the H2AX peptide by SUV39H2 SET-domain was studied on peptide arrays as described above. The H3 (1-15) peptide was included as positive control, whereas the H3K9A (1-15) and H2AX K134A (127-141) peptides as negative controls, respectively. No methylation signal of the peptides containing the H2AX K134 (127-141) and H2AX K134A (127-141) was observed, indicating that SUV39H2 is not able to methylate K134 of H2AX at peptide level (manuscript 2, Figure 1 B). On the other hand, a high methylation signal of H3K9 (1-15) was observed, whereas the negative control H3K9A mutation showed no signal as well. This observation confirmed the activity of the SUV39H2 SET-domain and the technical quality of the entire experiment. The same result was obtained with the His_6-tagged full-length SUV39H2, albeit the overall enzyme activity was much weaker (manuscript 2, Figure 1 D).

As a next step, a target lysine scan of H2AX with peptide arrays was performed. This experiment was conducted to investigate, if other lysine residues of H2AX were eventually methylated by SUV39H2. In this peptide array, the fifteen amino acid sequence was stepwise shifted by four amino acids to investigate the whole H2AX sequence and the above mentioned H3 (1-15) positive and negative controls were included as well. The peptide array was processed as described in 3.1.1. Unfortunately, no methylation signal for the expected target lysine of H2AX (spots B12-14) or any other spot of H2AX was observed (Figure 20). Only the

positive control H3K9 was methylated by the SUV39H2 SET-domain. These results indicate that K134 of H2AX and all other lysine residues of H2AX are no possible targets for SUV39H2 SET-domain at peptide level.

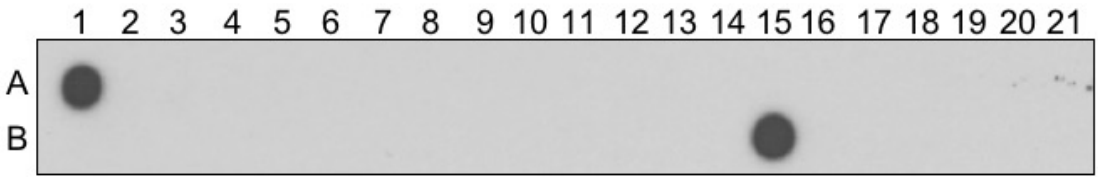


Figure 20: Target lysine scan of H2AX by SUV39H2 enzyme at peptide level. The immobilized peptides were methylated with SUV39H2 and radioactively labeled AdoMet as cofactor and the signal was detected by autoradiography. Spots A1 and B15 contain the positive controls, A2 and B16 the negative controls. In A3-B14, the peptide sequence was shifted by four amino acids through the entire H2AX sequence as detailed in Table 1. The picture was taken from (manuscript 2, Figure 1 E).

Table 1: Fifteen amino acid long peptide sequences used for the lysine scan of H2AX by SUV39H2 shown in Figure 20. The target lysine (K9) in the H3K9 (Uniprot no: P68431) positive control is labeled in red and the K134 of H2AX (Uniprot no: P16104) is shown in blue. It is given the Spot. No, the peptide fragment in amino acids and the peptide sequence. The table was adopted from (manuscript 2, Table 1).

Spot No.	Peptide fragment	Sequence
A 1	H3K9 (1-15)	ARTKQTARKSTGGKA
A 2	H3K9A (1-15)	ARTKQTARASTGGKA
A 3	H2AX (1-15)	MSGRGKTGGKARAKA
A 4	H2AX (5-19)	GKTGGKARAKAKSRS
A 5	H2AX (9-23)	GKARAKAKSRSSRAG
A 6	H2AX (13-27)	AKAKSRSSRAGLQFP
A 7	H2AX (17-31)	SRSSRAGLQFPVGRV
A 8	H2AX (21-35)	RAGLQFPVGRVHRLL
A 9	H2AX (25-39)	QFPVGRVHRLLRKGH
A10	H2AX (29-43)	GRVHRLLRKGHYAER
A11	H2AX (33-47)	RLLRKGHYAERVGAG
A12	H2AX (37-51)	KGHYAERVGAGAPVY
A13	H2AX (41-55)	AERVGAGAPVYLAAV
A14	H2AX (45-59)	GAGAPVYLAAVLEYL
A15	H2AX (49-63)	PVYLAAVLEYLTAEI
A16	H2AX (53-67)	AAVLEYLTAEILELA
A17	H2AX (57-71)	EYLTAEILELAGNAA

A18	H2AX (61-75)	AEILELAGNAARDNK
A19	H2AX (65-79)	ELAGNAARDNKKTRI
A20	H2AX (69-83)	NAARDNKKTRIIPRH
A21	H2AX (73-87)	DNKKTRIIPRHLQLA
B 1	H2AX (77-91)	TRIIPRHLQLAIRND
B 2	H2AX (81-95)	PRHLQLAIRNDEELN
B 3	H2AX (85-99)	QLAIRNDEELNKLLG
B 4	H2AX (89-103)	RNDEELNKLLGGVTI
B 5	H2AX (93-107)	ELNKLLGGVTIAQGG
B 6	H2AX (97-111)	LLGGVTIAQGGVLPN
B 7	H2AX (101-115)	VTIAQGGVLPNIQAV
B 8	H2AX (105-119)	QGGVLPNIQAVLLPK
B 9	H2AX (109-123)	LPNIQAVLLPKKTSA
B10	H2AX (113-127)	QAVLLPKKTSATVGP
B11	H2AX (117-131)	LPKKTSATVGPKAPS
B12	H2AX (121-135)	TSATVGPKAPSGGKK
B13	H2AX (125-139)	VGPKAPSGGKKATQA
B14	H2AX (129-143)	APSGGKKATQASQEY
B15	H3K9 (1-15)	ARTKQTARKSTGGKA
B16	H3K9A (1-15)	ARTKQTARASTGGKA

Another interesting question was, if SUV39H1 a homolog of SUV39H2, is able to methylate K134 of H2AX. To shed light on this, the same peptide array methylation experiment was performed using GST-tagged SUV39H1 SET-domain (amino acids 82-412) (manuscript 2, Figure 1 C). It was observed that SUV39H1 is not able to methylate lysine 134 of H2AX at peptide level as well. As the study from Sone et al. (2014) was performed at the protein level, the C-terminal domain (CTD) (amino acids 118-143) of H2AX was cloned C-terminally to the GST-tag (Figure 21) and methylation of this GST-H2AX-CTD protein was investigated in radioactive methylation experiments.

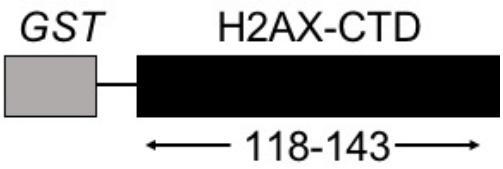


Figure 21: Schematic representation of the GST-H2AX-CTD protein. The H2AX-CTD was cloned C-terminal to the GST-tag. The numbers indicate the cloned boundaries in amino acids.

The methylation reaction was performed with both SUV39H enzymes and radioactively labeled AdoMet as stated above (Figure 22). For both SUV39H enzymes, no methylation signal of the H2AX-CTD protein was observed, indicating that H2AX-CTD is not a substrate for any of the SUV39H enzymes at protein level. The positive control H3 was methylated by both enzymes *in vitro* indicating that the overall reaction was technically correct.

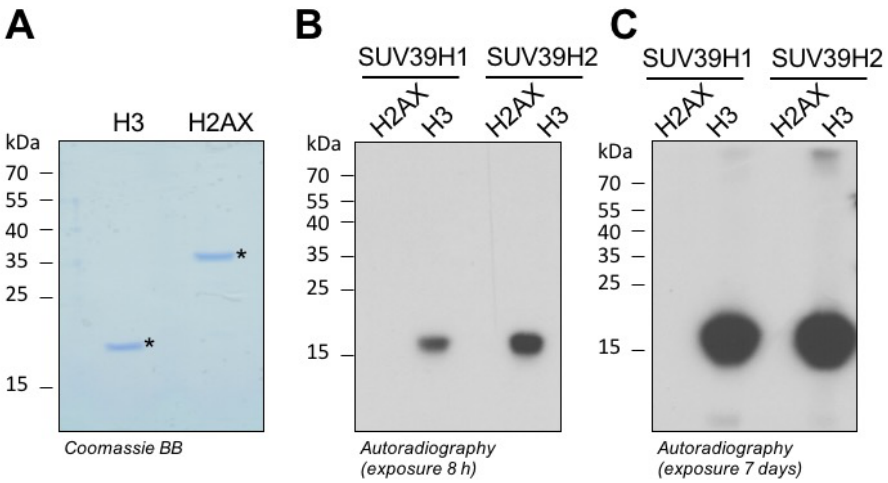


Figure 22: *In vitro* methylation analysis of GST-H2AX-CTD by SET-domains of SUV39H enzymes at protein level. A) Panel A shows the equal loading of GST-H2AX-CTD and H3 proteins. B) Obtained autoradiography image after radioactive methylation of GST-H2AX-CTD and H3 protein with both SUV39H enzymes. C) Longer exposed autoradiography image of panel B). The figure was adopted from (manuscript 2, Figure 2).

3.1.2.1 Discussion

In the above mentioned part it was shown that H2AX is not methylated by SUV39H2 under the used conditions at protein and peptide level *in vitro*. This could be explained by the fact that the sequence of H2AX contains an A instead of an R at position -1 which is not tolerated by SUV39H2. As pointed out above, this position is important for the interaction between SUV39H2 and the substrate, because a hydrogen bond will be formed between D307 and R8

of H3. The S at position +1 is missing as well which is an important recognition point of the SUV39H2 enzyme. Although SUV39H1 has a less stringent substrate specificity motif it was also not able to methylate H2AX, because important contacts could not be formed. One has to mention, that further *in vivo* analysis of H2AX methylation by SUV39H2 are necessary to formally exclude the hypothesis that H2AX is a SUV39H2 substrate, because in cells other enzymes or cofactors may have an impact on the specificity of SUV39H2 and on the interaction of H2AX and SUV39H2.

Of note, the published data of Sone et al. (2014) lack important controls. For example, they showed the methylation of H2AX by SUV39H2 using MS/MS, but they did not include the well-known target H3K9 as positive control. In addition, they did not provide the details for the cloned domain boundaries and the species in which they expressed the recombinant protein. The study by Sone et al. (2014) has been cited 47 times till now (Google Scholar retrieved on 01.10.2018). Three of these citations are from our group, which are not supporting the published data, but 44 citations are coming from other researchers who interpret their data in the light of the results of Sone et al. (2014). Because of the wrong assignment it is likely that other researchers interpret their data in a wrong way.

3.1.3 Activity and specificity of the SUV39H1 PKMT and investigation of novel methylation substrates

(manuscript 3 in the attachment of this thesis)

Kudithipudi S, **Schuhmacher MK**, Kebede AF, Jeltsch A. (2017) *ACS Chemical Biology 12(4):958-968. doi: 10.1021/acschembio.6b01076.*

SUV39H1 was the first identified human methyltransferase (Rea et al. 2000). As described above, it is the human homolog of the *Drosophila* Su(var) 3-9 enzymes, which catalyze the H3K9 methylation. Since interesting non-histone targets were identified for other PKMTs, like G9a (Rathert et al. 2008a), it was interesting to investigate this topic for SUV39H1 as well.

Beforehand different coworkers lead by Dr. Srikanth Kudithipudi purified and investigated the SUV39H1 enzyme. They determined the substrate specificity motif by peptide array

methylation experiments and searched for non-histone targets. These non-histone targets were validated at peptide and protein level. The methylation of RAG2 and SET8 was further analyzed in cellular studies.

Based on the existing preliminary data, the non-histone targets MLL1, DOT1L, RSF1 and ZNF583 were cloned as N-terminal GST-tagged proteins. After successful cloning and purification, the target proteins were investigated in protein methylation experiments using radioactively labeled AdoMet and SUV39H1 enzyme as described above (manuscript 3, Figure 2 A, Suppl. Figure 3 A). It was shown that MLL1 and DOT1L were methylated by SUV39H1. However, only DOT1L showed a strong methylation signal and was further investigated. A DOT1L N-terminal YFP tagged protein was cloned for cellular studies. Expression of DOT1L with or without co-expression of SUV39H1 in HEK293 cells revealed a methylation of this substrate which was detected by an anti-pan di/trimethyllysine antibody (manuscript 3, Figure 3 C). In summary, this experiment showed that SUV39H1 is able to methylate DOT1L up to the di- or trimethylation state in cells.

Next the non-histone target SET8 was investigated. SET8 is a PKMT itself, which catalyzes the monomethylation of H4K20. For SET8 it was interesting to test if the K210 methylation by SUV39H1 has an influence on its own endogenous methyltransferase activity, because K210 is located next to the SET-domain of SET8. To gain more insight into this question, HEK293 cells were co-transfected with YFP tagged full-length SET8 and CFP-tagged SUV39H1 or empty vector. After co-transfection, the cells were collected and the SET8 proteins were purified by GFP-Trap. Thereafter, equal amounts of both purified proteins were used in a radioactive methylation assay to detect SET8 activity. To do this, GFP-Trap purified SET8 was incubated with recombinant H4 as substrate and radioactively labeled AdoMet as cofactor. Interestingly, it was observed that SET8 which is methylated at K210 by SUV39H1 showed an increased activity on recombinant H4 (Figure 23 A). Co-transfection of the SET8 K210R mutant and SUV39H1 into HEK293 cells revealed no stimulation of SET8 methyltransferase activity, indicating that the methylation of K210 is important (manuscript 3, Suppl. Figure 7).

The stimulation of SET8 activity was investigated *in vitro* using purified proteins as well. For this, YFP-tagged SET8 was transfected into HEK293 cells and purified by GFP-Trap. Afterwards,

the SET8 was incubated with unlabeled AdoMet in presence or absence of SUV39H1 SET-domain and processed further as described above (Figure 23 B). This experiment confirmed the stimulation of the SET8 methyltransferase activity by K210 methylation by around 2-fold.

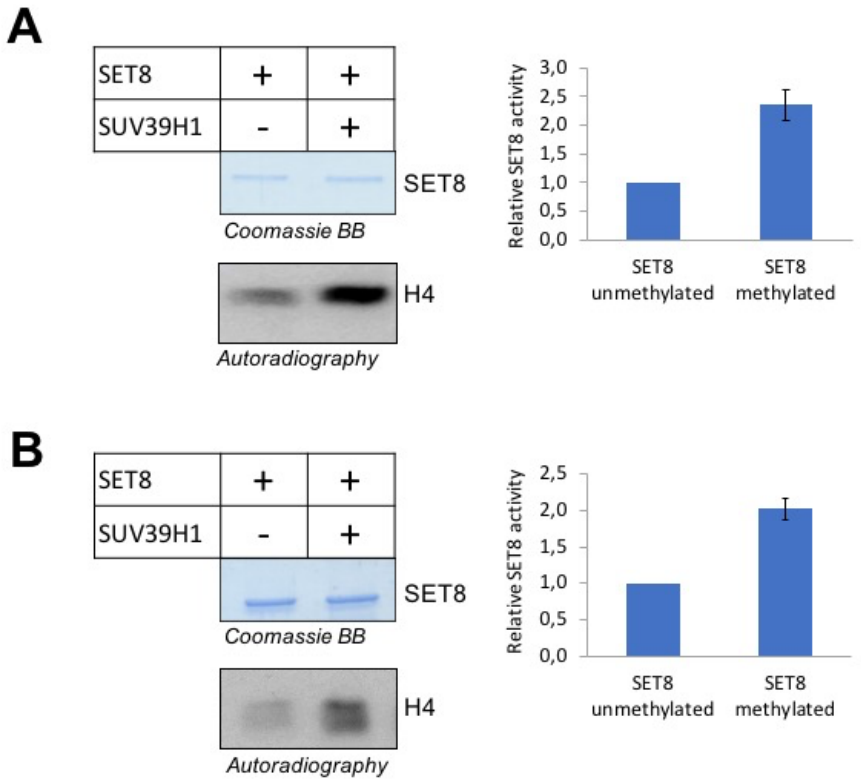


Figure 23: SET8 methylation by SUV39H1 and investigation of SET8 activity in cells and *in vitro*. A) Panel A shows results obtained after cellular methylation of SET8 by SUV39H1. The SET8 protein was incubated with recombinant H4 and radioactively labeled AdoMet. The bar diagram, represents the obtained averaged activity of SET8 over three experiments. The error bar shows the standard error of the mean. B) Methylation of YFP tagged SET8 isolated from HEK293 cells by SUV39H1 *in vitro*. The *in vitro* methylated SET8 was processed as mentioned in panel A. The bar diagram shows the averaged activity of SET8, which was obtained from two individual experiments. Whereas the error bar represents the standard error of the mean. The figure was taken from (manuscript 3, Figure 5).

As described above, both SUV39H enzymes can catalyze the methylation of H3K9. The substrate specificity motif of SUV39H2 enzyme was determined in this thesis. The SUV39H2 and SUV39H1 (analyzed by Dr. Srikanth Kudithipudi) substrate specificity motifs are summarized in Table 2.

Table 2: Summary of the substrate specificity motifs of the SUV39H enzymes. The natural amino acids in the H3K9 sequence are underlined. Adopted and changed from (manuscript 3, Table 1).

Position	SUV39H1	SUV39H2
- 5	K>R>>N	
- 4	X	
- 3	A, S, T, Y	
- 2	A, N, L, P, W>Q, H	
- 1	R	R
0	K	K
+ 1	R, K, S>T	S>T
+ 2	A, G, S, T	T>S, G, A
+ 3	Q, G, K>X	G>R
+ 4	X	G>X

The comparison of both substrate specificity profile arrays revealed, that both enzymes need the contact to the target lysine 9 of H3. This observation highlights the importance of the methylation site. For both enzymes, the position -1 (K9 is position zero) is equally important as the lysine 9, indicating a very strong readout of RK. However, the comparison of both specificity profiles showed a stringent recognition of the N-terminal part by SUV39H1, whereas SUV39H2 has a relaxed readout of this sequence part. This observation changes in the C-terminal part, where SUV39H2 has a more stringent readout than SUV39H1 especially in the positions S10, T11 and G12. The position +1 is very important in SUV39H2, because in this position only T is tolerated apart of the native S. In contrast, SUV39H1 tolerates R, K and to a lesser extent T as well. All these information point in the direction that SUV39H1 and H2 can methylate the same target lysine 9 of histone 3, but maybe contact it in a different way.

After the identification of differences in the substrate specificity motif of both enzymes and new non-histone targets of SUV39H1 at the protein level, it was interesting to see, if its paralog SUV39H2 is also able to methylate these targets. For this, the above described radioactive based methylation experiments at protein level were repeated with SUV39H2 revealing that it was not able to methylate the non-histone target proteins of SUV39H1 (manuscript 3, Suppl. Figure 3 B).

3.1.3.1 Discussion

As mentioned above, two non-histone substrates (DOT1L and SET8) were investigated by SUV39H1 in cellular studies. The cellular studies confirmed the target methylation of the non-histone substrates by SUV39H1. It is important to confirm the target site methylation in cellular studies because in cells there exist other factors which may influence the methylation which cannot be mimicked by *in vitro* studies. Also the determined methylation state of the non-histone substrates are in agreement with studies that showed the trimethylation of H3K9 by SUV39H1 (Rea et al. 2000) and it demonstrates that the non-histone protein substrates are methylated up to the trimethylation state in cells by SUV39H1.

The methylation of DOT1L by SUV39H1 demonstrates a connection between both methyltransferases. This methylation event may has an impact on the DOT1L interaction with DNA or nucleosomes, because lysine 410 is localized in a region which is important to perform this interaction (Min et al. 2003). However, the potential influence of the DOT1L methylation by SUV39H1 on the activity and function of DOT1L is not known so far and needs further investigation.

Previous studies demonstrated that the SET8 monomethylated H4K20 is used as substrate by the methyltransferases SUV4-20H1 and SUV4-20H2 to finally generate H4K20me3 (Jorgensen et al. 2013). On the other side, it was shown that SUV39H1 has an indirect influence on the H4K20 methylation by the creation of H3K9me3. The H3K9me3 mark is recognized by HP1 which then can recruit the SUV4-20H enzymes for H4K20 methylation (Lachner et al. 2003). In this study, it was shown that SET8 is methylated at K210 by SUV39H1 which increases its activity. This observation is very important because by this increase of activity, more H4K20me1 is present which can be converted into higher methylation states indicating that SUV39H1 has another more direct way for the regulation of the H4K20 methylation.

The non-histone targets of SUV39H1 were investigated with SUV39H2 revealing that SUV39H2 is not able to methylate these substrates. A possible explanation is that the sequence of the investigated non-histone targets of SUV39H1 do not correlate with the identified substrate specificity motif of SUV39H2. Because of that the important interactions, as described above, between the amino acids of the target and the enzyme could not be formed. This observation

that SUV39H1 has different additional substrates than SUV39H2 could be one explanation, why the biological role of both enzymes is not fully overlapping.

3.2 Investigation of the substrate specificity of SETD2

The SETD2 methyltransferase was first identified as huntingtin interacting protein (Rega et al. 2001). Later it was shown to methylate H3K36 up to the trimethylation state (Eram et al. 2015). Whereas, the mono- and dimethylated state of H3K36 is introduced by other enzymes like NSD1 or NSD2 as well, SETD2 is the only known methyltransferase in mammalian cells, which introduces H3K36me3 (Edmunds et al. 2008; Eram et al. 2015). This observation makes SETD2 a very interesting research target. By now, little is known about the target specificity of this enzyme and its possible other targets in addition to H3K36 and the already described non-histone target α-tubulin (Park et al. 2016). In the next chapter the unpublished results regarding the specificity analysis of SETD2 and its potential non-histone substrates are described in detail. A discussion of the obtained results is also provided in this chapter.

3.2.1 Purification and activity determination of SETD2

The His_6-tagged SETD2 (amino acids 1347-1711) plasmid was transformed into *E.coli* DE3 Codon Plus cells followed by the overexpression of the protein. After affinity chromatography purification using Nickel-Nitrilotriacetic acid (Ni-NTA) beads, the purified protein was analyzed by SDS-PAGE stained with Coomassie BB, showing the purity of the protein preparation (Figure 24 A). Next, the activity of the purified SETD2 protein was determined using radioactively labeled AdoMet as cofactor and mononucleosomes as substrate, which were isolated from HEK293 cells using a standard protocol by Dr. Sara Weirich. After methylation, the protein sample was processed as described above (Figure 24 B). As a second method to determine the activity of SETD2, peptide arrays were used. As template sequence, H3 (29-43) with centered K36 was used and as a negative control the target lysine to alanine mutation (H3K36A) was synthesized as well. The synthesized peptide arrays were treated as outlined above (Figure 24 C).

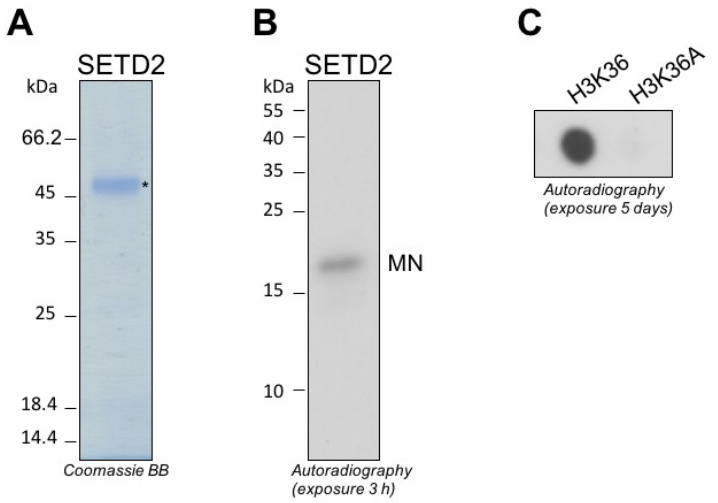


Figure 24: Purification and activity of SETD2 at protein- and peptide level. A) Purified His_6-tagged SETD2 was visualized by Coomassie BB staining. The asterisk indicates the size of SETD2 (50 kDa). B) Activity test of SETD2 using radioactively labeled AdoMet as cofactor and isolated HEK293 mononucleosomes (MN) as substrate. The signal was detected by autoradiography. C) SETD2 activity on the H3K36 (29-43) peptide substrate. The radioactive methylation of the peptides by SETD2 was detected by autoradiography.

After confirmation of the SETD2 activity at protein- and peptide level, its substrate specificity profile was determined. As template sequence H3 (29-43) was utilized which is shown in the horizontal axis of Figure 25 A. For the SETD2 substrate specificity investigation the amino acids C and W were excluded because of their unfavorable coupling properties. The peptide array methylation experiments were performed in two replicates and analyzed as described in 3.1.1 and 5.3 (Figure 25 B, C, D). For a better visualization of the obtained substrate specificity profile, the discrimination factors were calculated (Kudithipudi et al. 2014a) showing the preference of SETD2 for different amino acids at specific positions (Figure 25 D). In addition, a Weblogo of the substrate specificity profile was prepared as well (Figure 25 E) (Crooks et al. 2004).

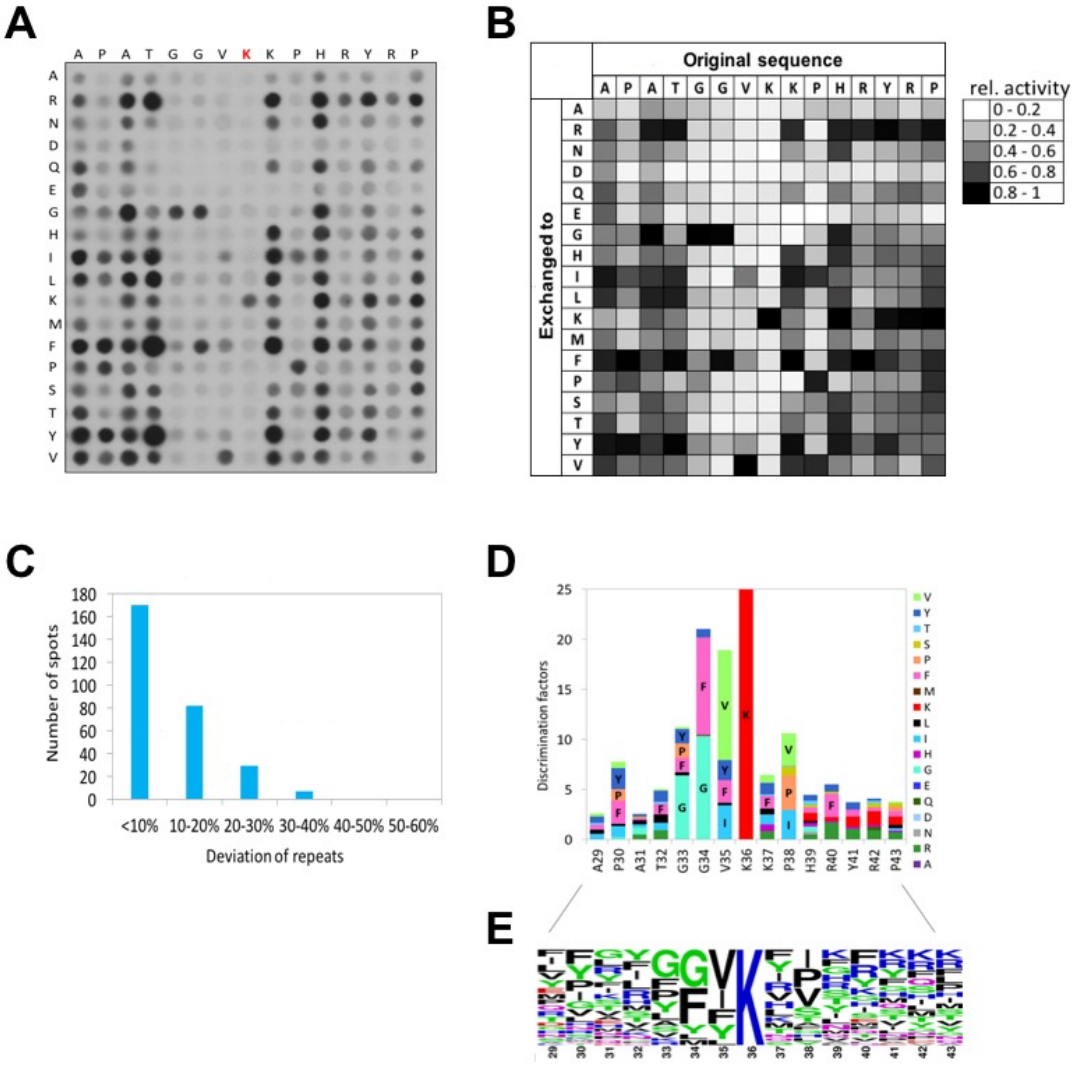


Figure 25: Substrate specificity analysis of SETD2. A) Peptide array methylation with SETD2 and radioactively labeled AdoMet. The vertical axis represents the 18 amino acids which were used to create single amino acid mutations of the H3 (29-43) template sequence shown on the horizontal axis. The red labeled K indicates K36. B) Average of data obtained after quantification of two independent substrate specificity peptide arrays of SETD2. The signal intensity is indicated in a greyscale on the right side. C) Standard deviations of relative methylation activities of individual peptides in both peptide arrays. D) Calculated discrimination factors of SETD2 substrate specificity investigation. E) Weblogo of the substrate specificity profile.

Unexpectedly, the substrate specificity investigation revealed that SETD2 prefers amino acids different from the known target sequence surrounding H3K36 at many sites of the profile. For example, at position -4 (K36 is annotated as 0) SETD2 prefers an F, Y, L, I and R much more than the naturally occurring amino acid T. At position +1, SETD2 has a higher preference for R, H, I, L, F, Y and V compared to the natural K. The same observation was made at other positions, like the -5, in which R, G and L are more preferred than the natural A. At position

+3, R, N, G, K, F and Y are more favored than the naturally occurring H. At position +5 R is more preferred than Y.

To further follow up on these observations, fifteen amino acid long peptides were synthesized on a cellulose membrane using H3K36 (29-43) as positive control and H3K36A mutant as negative control. On this membrane, peptides with single, double, triple, quadrupole and quintuple mutations of H3 (29-43) were synthesized to search for a “super-substrate” (Figure 26). For creation of the mutations, some of the above-mentioned preferred amino acids were used at the corresponding sites. This membrane was processed as described above.

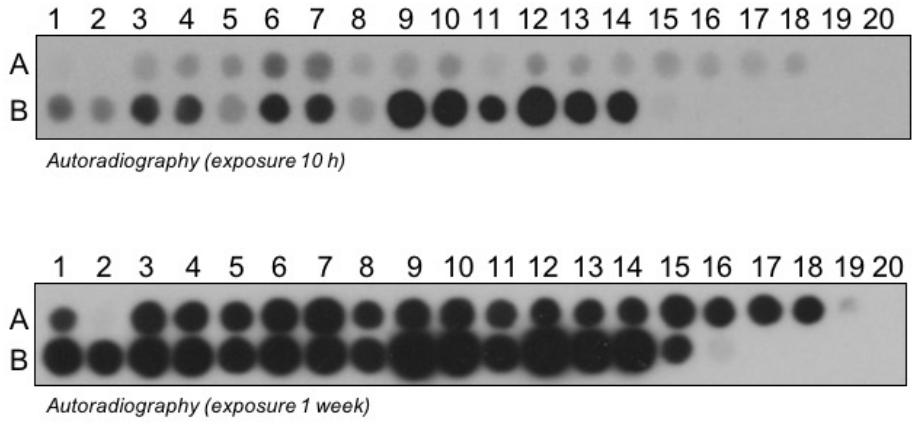


Figure 26: Identification of a peptide “super-substrate” of SETD2. Fifteen amino acid long peptides of H3K36 (29-43) and single, double, triple, quadrupole and quintuple mutations of it were synthesized on a cellulose membrane and incubated with SETD2 and radioactively labeled AdoMet. The signal of the transferred radioactive methyl groups was detected by autoradiography. Two films with different exposure times are shown to visualize weaker methylated spots as well. The synthesized amino acid sequences are listed in Table 3.

Table 3: Fifteen amino acid long H3 (29-43) peptides (Uniprot no. P68431) used for the "super-substrate" peptide scan of SETD2 shown in Figure 26. The target lysine (K36) is labeled in red and the introduced amino acid mutations are shown in blue.

Spot No.	Mutation(s)	Sequence
A 1	-	APATGGVKKPHRYRP
A 2	K36A	APATGGVAKPHRYRP
A 3	A31R	APRTGGVKKPHRYRP
A 4	A31G	APGTGGVKKPHRYRP
A 5	T32R	APARGGVKKPHRYRP
A 6	T32F	APAFGGVKKPHRYRP
A 7	T32Y	APAYGGVKKPHRYRP
A 8	K37R	APATGGVKRPHRYRP
A 9	K37H	APATGGVKHPHRYRP
A 10	K37I	APATGGVKIPHRYRP
A 11	K37L	APATGGVKLPHRYRP
A 12	K37F	APATGGVKFPHRYRP
A 13	K37Y	APATGGVKYPHRYRP
A 14	K37V	APATGGVKVPHRYRP
A 15	H39R	APATGGVKKPRRYRP
A 16	H39N	APATGGVKKPNRYRP
A 17	H39G	APATGGVKKPGRYRP
A 18	H39K	APATGGVKKPKRYRP
A 19	H39Y	APATGGVKKPYRYRP
A 20	A31R, K37R	APRTGGVKRPHRYRP
B 1	A31G, K37R	APGTGGVKRPHRYRP
B 2	A31R, T32R	APRRGGVKKPHRYRP
B 3	A31R, T32F	APRFGGVKKPHRYRP
B 4	A31R, T32Y	APRYGGVKKPHRYRP
B 5	T32R	APARGGVKKPHRYRP
B 6	A31R, T32F, K37R	APRFGGVKRPHRYRP
B 7	A31R, T32Y, K37R	APRYGGVKRPHRYRP
B 8	T32R	APARGGVKKPHRYRP
B 9	A31R, T32F, K37Y, H39K	APRFGGVKYPKRYRP
B 10	A31R, T32F, K37R, H39N	APRFGGVKRPNRYRP
B 11	A31R, T32F, K37F	APRFGGVKFPHRYRP
B 12	A31R, T32F, K37Y, H39K, Y41R	APRFGGVKYPKRRRP
B 13	A31R, T32F, K37R, H39N, Y41R	APRFGGVKRPNRRRP
B 14	A31R, T32F, K37F, Y41R	APRFGGVKFPHRRRP
B 15	-	APATGGVKKPHRYRP
B 16	K36A	APATGGVAKPHRYRP

As shown in Figure 26 the spots A3-A18 and B1-14 were much stronger methylated by SETD2 than the H3K36 (29-43) peptide (spots A1 and B15), which indicates that these peptide substrates are highly favored by SETD2. Another observation is that most of the created mutations were better substrates for SETD2 than the H3K36 (29-43) peptide. Among them, spot B10 was one of the highest methylated spots. Its sequence was, therefore, selected as new target for further investigation and was named as H3K36 (29-43) “super-substrate”. The sequence of this peptide is as follows (the natural amino acids in the H3 (29-43) sequence are underlined and the red labeled amino acid represents the target lysine 36):

A - P - R - F - G - G - V - K - R - P - N - R - Y - R - P

Afterwards, a second substrate specificity profile based on the H3K36 (29-43) “super-substrate” sequence was determined to investigate if the observed preferences are influenced by the sequence template. The preparation, the working procedure and the analysis of the obtained data was performed as described in 3.1.1 and 5.3. Figure 27 represents the results, which were obtained from two independent replicates.

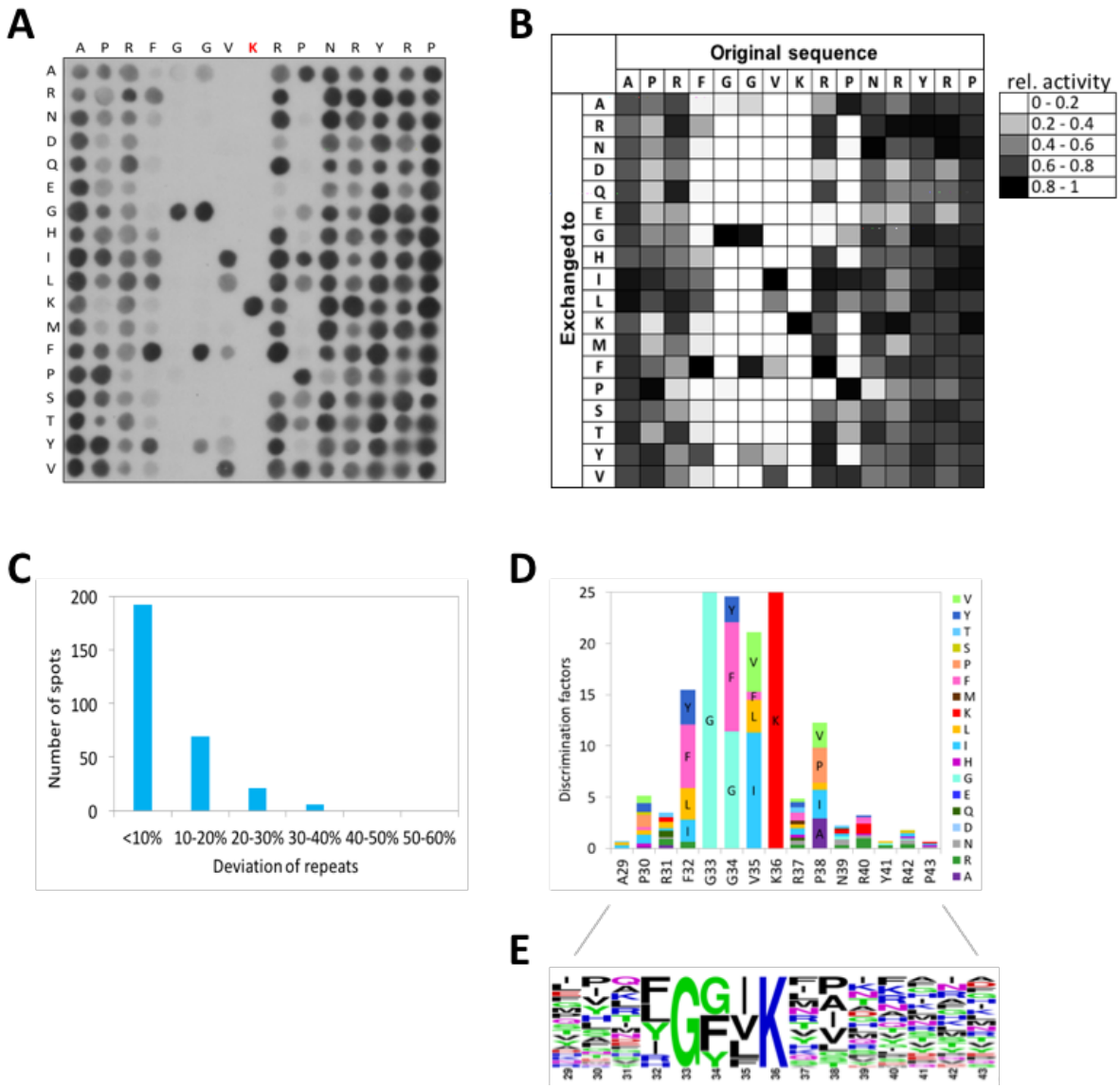


Figure 27: Substrate specificity analysis of SETD2 based on the H3K36 (29-43) "super-substrate". A) Peptide array methylation with SETD2 and radioactively labeled AdoMet. The vertical axis represents the 18 amino acids which were used to create single amino acid mutations of the H3K36 (29-43) "super-substrate" template sequence shown on the horizontal axis. The red labeled K indicates K36. B) Average of data obtained after quantification of two independent substrate specificity peptide arrays of SETD2. The signal intensity is indicated in a greyscale on the right side. C) Standard deviations of relative methylation activities of individual peptides in both peptide arrays. D) Calculated discrimination factors of SETD2 substrate specificity investigation. E) Weblogo of the substrate specificity profile.

The substrate specificity of SETD2 based on the H3K36 (29-43) "super-substrate" sequence shows a highly specific and stringent profile (Figure 27). The -4 position (when K36 is position zero) is specifically recognized by SETD2 where it only accepts F, R, I, L and Y. Interestingly, the -3 position is highly specific and only G is tolerated. At position -2 a strong specificity was

observed as well, where F and G are similarly accepted, and Y and A weaker. The I at position -1 is better recognized by SETD2 as the V. In addition, L, F and Y are very weakly tolerated at this position. At position +1, SETD2 shows a relaxed specificity although the amino acids E, G and P are not tolerated. The position +2 is also specifically recognized by SETD2. Only A, G, I, L, S, T and V are tolerated apart from P. However, I, A and P are preferred compared to G, L, S, T and V at this position. The following specificity motif based on the H3K36 (29-43) "super-substrate" sequence of SETD2 was determined (the amino acids of the super-substrate are bolded and red represents the target lysine 36):

(R, I, L, Y, **F**) - **G** - (F, **G**>A, Y) - (I, **V**, L, F, Y) - K - **R** (not E, G, P) - (A, I, **P**>G, L, S, T, V)

Interestingly, both substrate specificity profile analysis resulted in very similar amino acid preferences (Figure 28). A small deviation between both analyses was observed at position -3 and +2. In the H3K36 (29-43) peptide substrate specificity SETD2 tolerated Y, P and F apart from the G, but G was the preferred residue at position -3. In position +2 no A is accepted compared to the super-substrate specificity profile. Also a small discrepancy was observed at position -1.

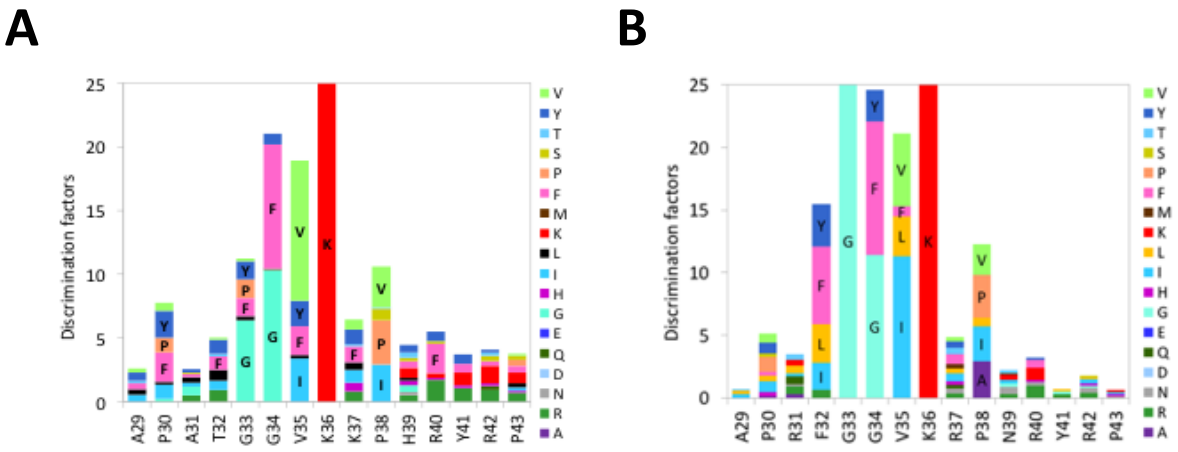


Figure 28: Comparison between both substrate specificity profiles of SETD2. A) Discrimination factors of the substrate specificity investigation of SETD2 using the H3K36 (29-43) sequence. The substrate specificity motif of SETD2 showed a recognition from position -4 to +2. B) Discrimination factors based on the H3K36 (29-43) "super-substrate" specificity profile investigation of SETD2. The substrate specificity motif of SETD2 showed a recognition from position -4 to +2.

3.2.2 Identification of SETD2 non-histone targets at peptide level

The substrate specificity profile based on the H3K36 (29-43) “super-substrate” sequence was used to perform a Scansite search to identify possible human non-histone targets of SETD2. In this search, 166 possible targets were retrieved, which contain the substrate specificity motif of SETD2. Methylation of these 166 possible targets was first investigated at the peptide level using radioactive AdoMet as described above (Figure 29). As positive controls the H3K36 (29-43) peptide, the H3K36 (29-43) “super-substrate” peptide (spot A1, I9 and A3, I11) and as negative controls the respective lysine to alanine mutations (spot A2, I10 and A4, I12) were included.

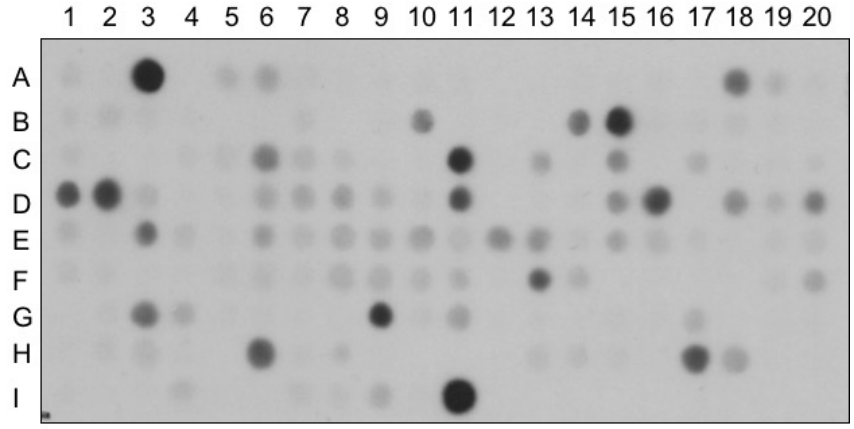


Figure 29: Non-histone target scan, based on the H3K36 (29-43) “super-substrate” specificity profile. The peptide array was incubated in methylation buffer, supplemented with SETD2 enzyme and radioactively labeled AdoMet and the transfer of radioactive methyl groups was detected by autoradiography. The information about the synthesized amino acid sequence is provided in Table S1 in the attachment of this thesis.

Among these 166 putative targets, 20 peptides were selected although they showed a weaker methylation than the “super-substrate” peptide. For the 20 selected putative non-histone targets, another membrane was synthesized on which the target lysine to alanine mutation was included. This second peptide array methylation assay was performed to confirm that the methylation occurs at the predicted target lysine. The methylation reaction of the membrane was conducted as described in 5.3 (Figure 30).

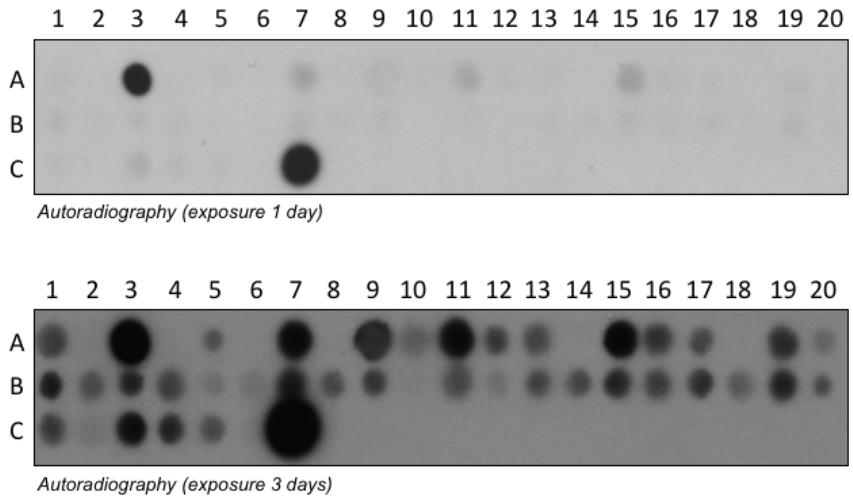


Figure 30: Non-histone targets scan with lysine to alanine mutations. In this scan only the selected peptides from Figure 29 were used and target lysine to alanine mutations were introduced. Each second spot represents the alanine mutation. Spots A1 and C5 represents the natural target H3K36 (29-43) peptide and the two spots A2 and C6 its alanine mutation. The spots A3 and C7 represents the H3K36 (29-43) "super-substrate" peptide and spots A4 and C8 its lysine to alanine mutation. Two different exposure times are shown to visualize the weaker methylated spots as well. The amino acid sequence is provided in Table S2 in the attachment of this thesis.

In this scan, the target site methylation for eight non-histone target peptides and the positive controls were validated (Table 4), but all of them were weaker methylated than the "super-substrate" peptide (spots A3 and C7). This indicates that the H3K36 (29-43) "super-substrate" peptide is the best known methylation substrate of SETD2. The intensities of the methylation signal of the H3K36 (29-43) peptide and H3K36 (29-43) "super-substrate" peptide spots and their respective alanine mutants were measured in two independent experiments. After subtraction of the background, the averaged ratio of H3K36 (29-43) and H3K36 (29-43) "super-substrate" was calculated revealing a 50 times stronger methylation of the H3K36 (29-43) "super-substrate". However, this sequence is not present in the human proteome, which makes it to be a hypothetical substrate.

Table 4: Summary of the selected non-histone targets based on Figure 30. It is given the Uniprot No., the target lysine position, the synthesized peptide sequence and the spot No. in Figure 30. The red labeled K represents the target lysine.

Protein name	Uniprot No.	Target lysine	Sequence	Spot No.
H3K36 (29-43) WT	P68431	36	APATGGVKKPHRYRP	A1, C5
H3K36 (29-43) super-substrate	P68431	36	APRFGGVKRPNRYRP	A3, C7
Ankyrin and armadillo repeat-containing protein	Q7Z5J8	317	RRGIGYLKLICFLIP	A5
Voltage-dependent T-type calcium channel subunit alpha-1G	O43497	804	YGPFGYIKNPYNIFD	A7
Collagen alpha-1(XXII) chain	Q8NFW1	472	SEQIGFLKTINCSCP	A13
Dysferlin	O75923	338	AGARGYLKTSLCVLG	A17
Fibrillin-1	P35555	666	STCYGGYKRGQCIKP	A19
Hemicentin-1	Q96RW7	127	EMSIGAIKIALEISL	B9
Integrator complex subunit 6	Q9UL03	369	GHPFGYLKASTALNC	B11
CMP-N-acetyl-neuraminate-poly-alpha-2,8-sialyltransferase	Q92187	350	LHNRGALKLTTGKCV	C1

3.2.3 Analysis of the selected non-histone substrates and H3K36 (29-43)-GST variants at protein level

The eight putative non-histone targets, H3K36 (29-43) WT, the super-substrate and H3K36M super-substrate mutant as negative control were cloned into a pGEX-6p2 bacterial expression vector for protein expression by Jan Ludwig in his B.Sc. work under my supervision. Previously, it was shown that the H3K36M mutation inhibits the enzyme activity of SETD2 (Yang et al. 2016), see also chapter 3.3 of this PhD thesis). To study this at protein level the H3K36 (29-43) sequence variants were cloned with C-terminal GST tag (Figure 31). The possible non-histone targets were cloned C-terminal to the GST tag. An overview of the selected proteins with their cloned domain boundaries is provided in Table 5.

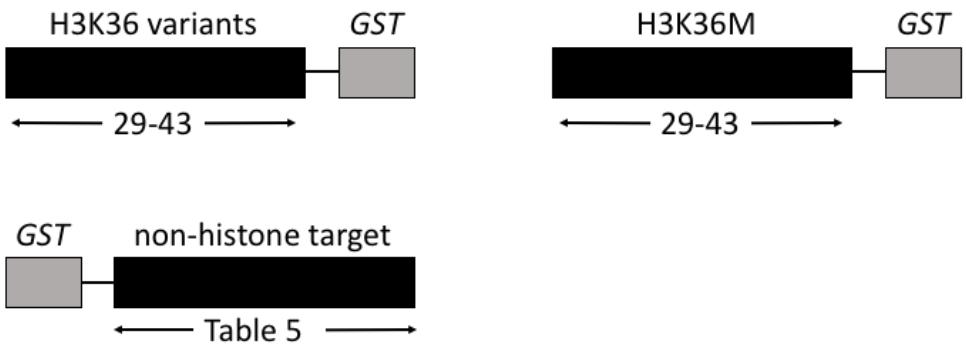


Figure 31: Schematic representation of the H3K36 (29-43)-GST variants and non-histone targets. The H3K36 (29-43)-GST variant proteins and GST-H3K36M (29-43) mutant are shown with the cloned amino acids. For the non-histone targets the cloned boundaries are provided in Table 5.

Table 5: Putative non-histone targets and H3K36 (29-43)-GST variants shown in Figure 31. "Abb" is the short version for abbreviation and "Domain" represents the domain boundaries in amino acids, which were used to clone the different non-histone targets at protein level. The Uniprot No., the target lysine position, the molecular weight (MW), the synthesized sequence and the spot no. in Figure 30 is provided as well. The target K is printed in red.

Protein name	Abb	Uniprot No.	Target lysine	Domain	Sequence	MW [kDa]	Spot No.
H3K36 (29-43) WT	GST-H3K36	P68431	36	29-43	APATGGVKK PHRYRP	28	A1, C5
H3K36 (29-43) super-substrate	GST-super-substrate	P68431	36	29-43	APRFGGVKR PNRYRP	28	A3, C7
H3K36M (29-43) super-substrate mutant	GST-H3K36M super	P68431	36	29-43	APRFGGVM RPNRYRP	28	
Ankyrin and armadillo repeat-containing protein	ANKAR	Q7Z5J8	317	173-459	RRGIGYLKLIC FLIP	58	A5
Voltage-dependent T-type calcium channel subunit alpha-1G	CAC1G	O43497	804	695-924	YGPFGYIKNP YNIFD	51	A7
Collagen alpha-1(XXII) chain	COMA	Q8NFW1	472	429-603	SEQIGFLKTIN CSCP	46	A13
Dysferlin	DYSF	O75923	338	218-516	AGARGYLKT SLCVLG	58	A17
Fibrillin-1	FBN1	P35555	666	529-763	STCYGGYKR GQCIKP	52	A19
Hemicentin-1	HMCN1	Q96RW7	127	5-217	EMSIGAIKIAL EISL	50	B9
Integrator complex subunit 6	INT6	Q9UL03	369	296-463	GHPFGYLKA STALNC	45	B11
CMP-N-acetyl-neuraminate-poly-alpha-2,8-sialyltransferase	SIA8D	Q92187	350	98-358	LHNRGALKLT TGKCV	55	C1

The following cloning and methylation experiments were conducted by Jan Ludwig under my supervision. The super-substrate mutations and K36M were introduced into the GST-H3K36 sequence by site-directed mutagenesis method. Six non-histone targets and all H3K36 (29-43)-GST variant proteins were cloned, expressed and purified. The possible ANKAR non-histone target could not be cloned and the obtained protein for CAC1G showed an incorrect size after SDS-PAGE analysis. After affinity chromatography purification, the GST-tagged H3K36 (29-43)-GST variant substrate proteins were used in equal amounts, as shown by Coomassie BB staining (Figure 32). Afterwards a methylation reaction was performed as described in 5.4 (Figure 32). Mononucleosomes (MN) isolated from HEK293 cells were used as positive control. The MN were isolated by Dr. Sara Weirich.

Similarly, equal amounts of the putative non-histone target proteins and the GST-super-substrate, as positive control, were loaded on a SDS-gel and stained with Coomassie BB (Figure 33). Afterwards protein methylation reactions were conducted as described above (Figure 33).

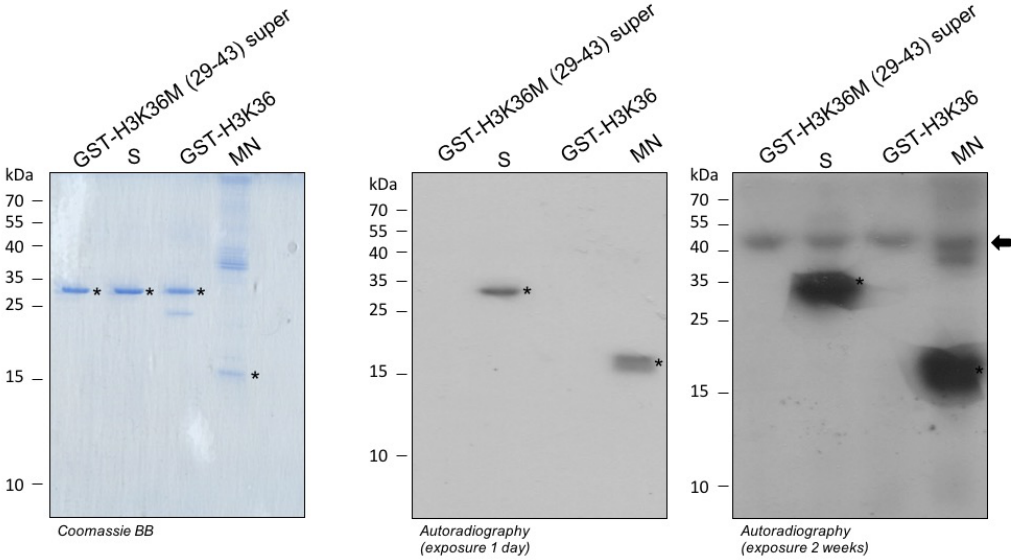


Figure 32: Investigation of H3K36 (29-43)-GST variant protein methylation by SETD2. The left panel shows the equal loading of the GST tagged proteins under investigation. The right panel represents the obtained autoradiography films after methylation of the H3K36 (29-43)-GST variant proteins (28 kDa) and native mononucleosomes (MN) (15 kDa) with SETD2 and radioactively labeled AdoMet. Two different exposure times are shown to visualize the weaker methylated proteins as well. The arrow represents the size of the SETD2 automethylation (50 kDa), the asterisks the correct size of the investigated proteins and S is the abbreviation of the GST-super-substrate.

As shown in Figure 32 only the GST-super-substrate was methylated by SETD2 at protein level *in vitro*. No signal was observed for the GST-H3K36 protein and GST-H3K36M (29-43) super-substrate mutant protein.

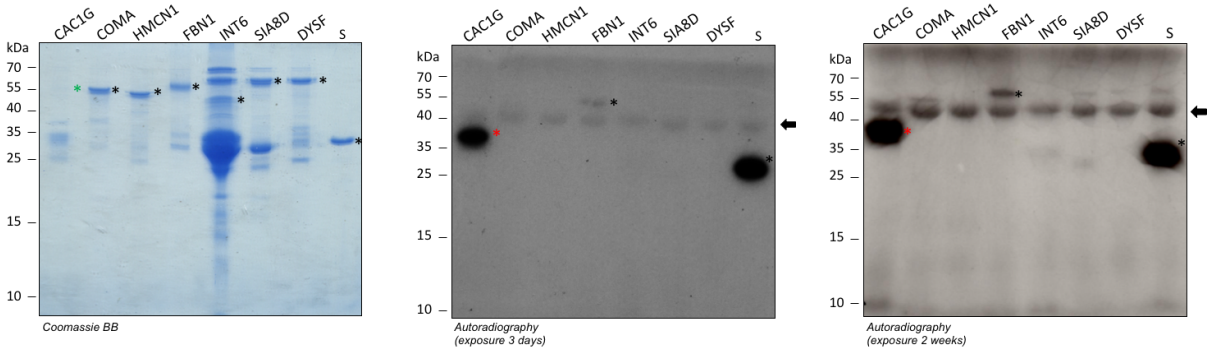


Figure 33: Investigation of putative non-histone targets of SETD2 at protein level *in vitro*. The left panel represents the equal loading of the non-histone proteins and the positive control GST-super-substrate (S) (28 kDa). The right pictures show the autoradiography image after methylation of the proteins with radioactively labeled AdoMet and SETD2. The arrow represents the size of SETD2 automethylation (50 kDa) and the black asterisk indicates the expected size of the non-histone substrate proteins. Two different exposure times are shown to visualize the weaker methylated proteins as well. The green asterisk indicates the expected size of CAC1G (51 kDa) and the red asterisk visualize the incorrect methylation size of CAC1G. The S is the abbreviation of the GST-super-substrate.

Among the seven putative non-histone substrate proteins, one target (FBN1) was methylated very weakly at protein level *in vitro*. The CAC1G was methylated in a comparable level as the positive control GST-super-substrate, despite showing an incorrect size, probably because it represents a degraded protein. This finding likely indicates that SETD2 methylated a short unstructured peptide fragment attached to the GST. This possible non-histone target was not investigated further.

3.2.4 Confirmation of target lysine methylation

Based on preliminary data, the target lysine methylation of the FBN1 protein was confirmed. A target lysine to arginine mutation was introduced by site-directed mutagenesis and confirmed by sequencing. The overexpressed wildtype and mutant proteins were purified by affinity chromatography with similar purity (Figure 34). Equal amounts of WT and mutant proteins were methylated with SETD2 and radioactively labeled AdoMet as described in 5.4.

After mutation of the target lysine 666 of FBN1, a complete loss of the methylation signal was observed. This confirms that SETD2 methylates FBN1 at K666 and that FBN1 is a substrate of SETD2.

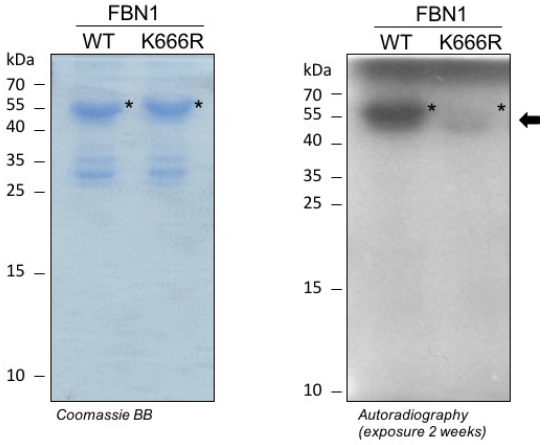


Figure 34: Confirmation of the target lysine methylation of FBN1 by SETD2 at protein level. The left panel shows the equal loading of the wild type and mutant proteins. The image on the right side shows the autoradiography film after SETD2 methylation of the proteins with radioactively labeled AdoMet. The black arrow indicates the automethylation of SETD2 (50 kDa) and the asterisks the correct size of the investigated FBN1 WT and mutant proteins (53 kDa).

3.2.5 Discussion

The substrate specificity profile of SETD2 was determined by peptide array methylation using the H3K36 (29-43) peptide sequence as template. The investigation of the substrate recognition motif revealed that other amino acids different from the natural target sequence are much more preferred by SETD2. This finding suggests that H3K36 might not be the best substrate of SETD2 and that there could be other SETD2 substrates beside of H3K36 and α-tubulin in the human proteome. Based on the substrate specificity profile analysis a so-called super-substrate of SETD2 was created. This super-substrate was stronger methylated by SETD2 than H3K36 which supports the suggestion that H3K36 is may be not the best substrate of SETD2. The better methylation shows as well that the design of the super-substrate was successful. However, this sequence is not present in the human proteome, which makes it to be a hypothetical substrate.

Based on this super-substrate peptide, a second substrate specificity profile scan was performed. In this profile, only one amino acid was more preferred. This suggests that the

super substrate specificity motif resembles the preferred recognition sequence of SETD2. Comparison of both substrate specificity motifs showed that both are very similar. This indicates that these analyses are not strongly dependent on the sequence context, at least not for SETD2.

The optimized substrate specificity motif is interesting because the T at position 32 of the histone 3 is not preferred, but an F is the best amino acid at this site. This exchange may lead to a different interaction of the substrate with the SETD2 enzyme, because a polar and neutral amino acid is exchanged against a nonpolar and hydrophobic one. Another interesting exchange which shows a big influence happened at position 31 in which A was exchanged by R. A deeper look into the crystal structure of SETD2 bound to H3K36M peptide (Yang et al. 2016) showed that E1674 of SETD2 is capable to contact the R of the mutated A31 to R (Figure 35 A, B). Additionally, exchange of T to F at position 32 may lead to an interaction with Y1604 of SETD2 which could stabilize the binding between enzyme and substrate (Figure 35 C). Another observation is that Q1667 and Q1669 of SETD2 are able to form hydrogen bonds to the mutated K37 to R (Figure 35 D). The C-terminal part of SETD2 forms a loop which could be important for the binding of the substrate to the enzyme and it might be involved in pushing the substrate into the binding pocket (Figure 35 E).

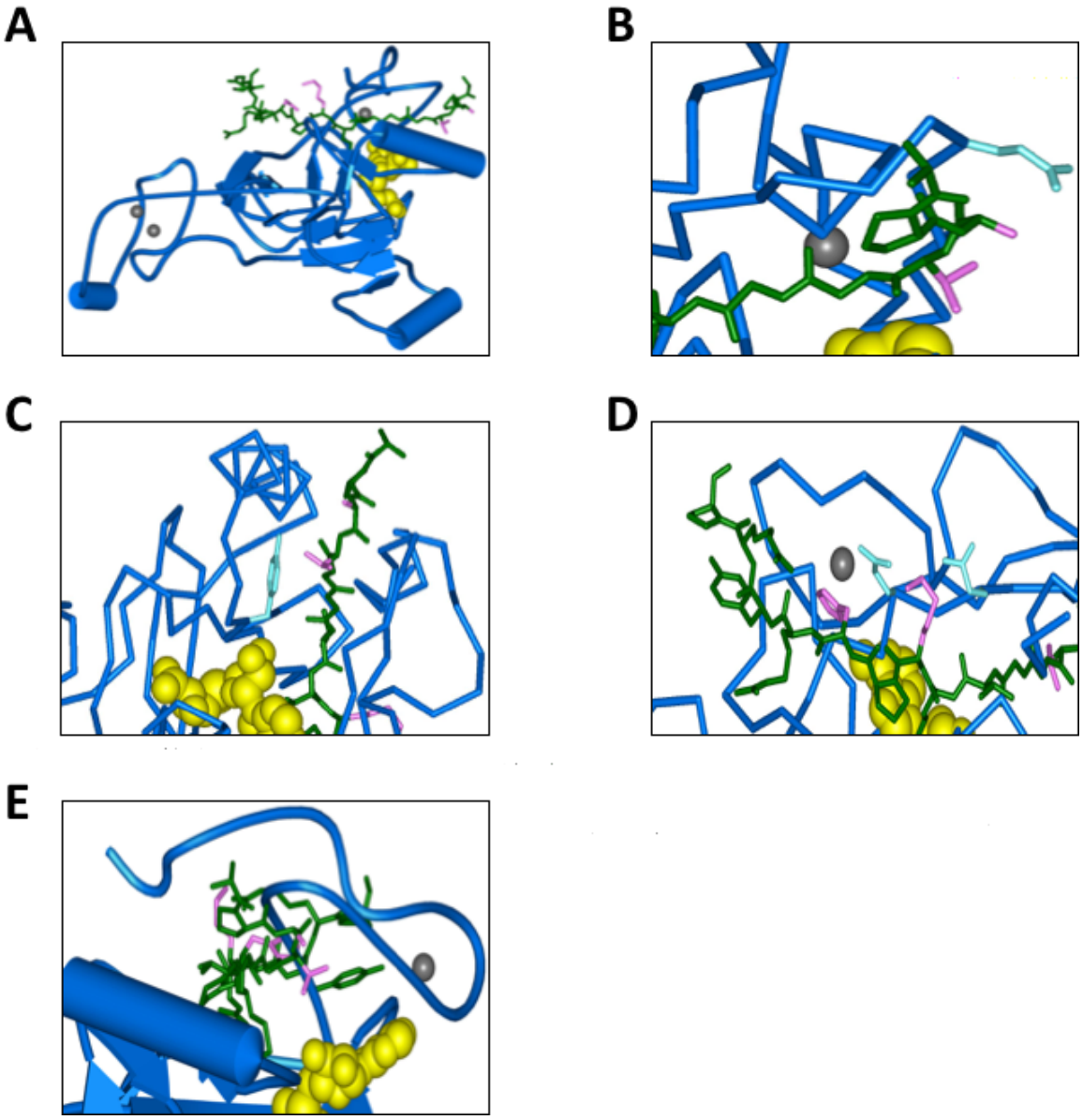


Figure 35: Structure of SETD2 (blue) in complex with the H3K36M (29-43) peptide by atom type (green) and the cofactor AdoMet as sphere model (yellow) (PDB ID: 5V21). The zinc ions are represented as grey dots. A) Structure of the ternary complex of H3K36M (29-43) peptide, AdoMet and SETD2 enzyme. B) Zoom to the A31 (pink) peptide position which is replaced by an R in the super-substrate and the possible interaction with E1674 (cyan) in SETD2. C) Possible interaction between T32 (pink) mutated to F in the super-substrate and the Y1604 (cyan) of SETD2. D) Potential hydrogen bond between the K37 (pink) of the H3 peptide that is mutated to R in the super-substrate and the Q1667 and Q1669 (cyan) of SETD2. E) C-terminal loop (blue) of SETD2 which may fix the peptide into the active pocket of the enzyme.

In a next step additional non-histone targets of SETD2 were identified and the methylation of the H3K36 (29-43) variants (GST-super-substrate, GST-H3K36 and GST-H3K36M) were investigated. The non-histone target search based on the optimized specificity profile revealed

166 possible targets. Eight were methylated at peptide level but unfortunately only one non-histone substrate protein (FBN1) was methylated. The methylation of FBN1 at K666 was confirmed by K to R mutation. For the H3K36 (29-43) variants only the GST-super-substrate was methylated by SETD2. One explanation for the methylation discrepancy between peptide and protein level is, that the fifteen amino acid long peptide substrates are immobilized on a cellulose membrane and because of that the peptides are very good accessible for the SETD2 enzyme. In contrast to this, a protein substrate is folded and the target K is maybe not that good accessible for the enzyme which results in an unmethylated substrate protein. This shows that the data of the non-histone peptide array methylation experiments cannot always be extrapolated to protein level. By now, SETD2 is so far, the only known methyltransferase which methylates H3K36 up to the trimethylation state. Because of this outstanding important methylation activity, it is possible that not many non-histone targets of SETD2 are present in the human genome beside of α-tubulin and FBN1. The observed unmethylated H3K36M (29-43) mutant protein was expected as the target lysine was mutated to methionine.

The GST-H3K36 (29-43) was cloned and purified to check if the short peptide sequence is methylated at protein level, because preliminary studies showed that SETD2 is not active on recombinant H3 protein. Interestingly, the GST-H3K36 was not methylated by SETD2, which indicates that the GST-super-substrate is a better substrate of SETD2 than the GST-H3K36. This result is in agreement with the peptide array methylation experiments in which the super-substrate peptide shows a higher methylation signal than the H3K36 (29-43) peptide.

In summary, this investigation shows that SETD2 is able to methylate a second non-histone target (FBN1). The FBN1 is a major component of the 10-12 nm diameter microfibrils of the extracellular matrix (ECM) (Sakai et al. 1986). These microfibrils are highly conserved macromolecular assemblies (Jensen and Handford 2016). Previously it was also shown that mutations of FBN1 causes different diseases like for example the Marfan syndrome (Dietz et al. 1991). The methylation of FBN1 by SETD2 may implies the involvement of SETD2 in other processes than chromatin and cytoskeleton remodeling. The better methylation of the GST-super-substrate suggests a different binding of the substrate to the enzyme which needs further validation.

3.3 Kinetic Analysis of the Inhibition of the NSD1, NSD2 and SETD2 Protein Lysine Methyltransferases by a K36M Oncohistone Peptide

(manuscript 4 in the annex to this thesis)

Schuhmacher MK, Kusevic D, Kudithipudi S, Jeltsch A. (2017). *Chemistry Select 11;2(29):9532-9536. doi: 10.1002/slct.201701940.*

The NSD1, NSD2 and SETD2 enzymes are responsible for the introduction of methyl groups on H3K36. However, the NSD1 and NSD2 are only able to methylate H3K36 up to the dimethylation state, whereas SETD2 is able to introduce a third methyl group (Eram et al. 2015). Previously, it was shown that oncohistone mutations in which the target lysine is replaced by methionine inhibit the enzymatic activity of different PKMTs like for example the PRC2 complex (Justin et al. 2016), which is inhibited by the K27M mutation of H3. Other examples for the inactivation of the enzyme by oncohistone mutations are the H3K36 methyltransferases NSD2 and SETD2 (Lu et al. 2016; Yang et al. 2016). Both enzymes are inhibited by the K36M mutation. As the inhibition of NSD1 by the oncohistone H3K36M was not investigated, it was tested here to see if this K36 methyltransferase is blocked by this mutation as well.

The experiments for NSD2, the peptide array methylation of NSD1 and the data analysis of NSD2 were performed by Dr. Denis Kusevic. For all three enzymes (NSD1, NSD2 and SETD2) steady-state kinetic inhibition analyses were performed, in the presence of radioactively labeled AdoMet and H3K36M inhibitor peptide (27-43). After successful purification of the enzymes, the activity was tested by peptide array methylation (manuscript 4, Figure 1 A, B) using arrays with positive (H3K36 and H4K44) and negative controls (H3K36A and H4K44A) in which the target lysine is mutated to alanine. The peptide arrays were incubated with the respective enzyme and radioactively labeled AdoMet and processed as described above. All positive controls were methylated by the enzymes, whereas the negative controls remained unmethylated indicating the activity of the enzymes.

As this initial test confirmed the activity of the three investigated methyltransferases, the kinetic inhibition parameters were determined. The first step was to perform radioactive methylation experiments with increasing H3K36 (27-43) peptide concentrations for all three enzymes. For this, the different enzymes were incubated with the H3K36 (27-43) peptide and radioactively labeled AdoMet as cofactor. The methylation samples were then separated by Tricine-SDS-PAGE and the signal was captured by autoradiography and afterwards quantified (manuscript 4, Figure 3). This experiment was repeated in the presence of different H3K36M (27-43) peptide inhibitor concentrations at a constant H3K36 (27-43) peptide concentration (manuscript 4, Figure 3). Interestingly, the determined K_I values for the K36M peptide inhibitor binding were about 1.5 fold lower than the K_M for all three enzymes (Table 6).

Table 6: K_M, K_I and K_M/K_I values of the methyltransferases NSD1, NSD2 and SETD2. The table was adopted from (manuscript 4, Table 1).

PKMT	K_M [µM]	K_I [µM]	K_M/K_I
NSD1	205 ± 16	142 ± 16	1.45 ± 0.07
NSD2	70.6 ± 2.8	46.5 ± 1.7	1.52 ± 0.05
SETD2	65.3 ± 3.0	44.6 ± 3.3	1.47 ± 0.03

3.3.1 Discussion

Overall, this investigation showed for the first time the inhibition of NSD1 by the oncohistone H3K36M. Also, previous results of the inhibition of NSD2 and SETD2 by H3K36M were confirmed. The present study shows, that NSD2 and SETD2 have very similar K_M and K_I values, whereas the two values of NSD1 differ from the other ones (Table 6). This finding could be explained by the fact that H3K36 is not the best substrate of NSD1 (Kudithipudi et al. 2014b). The achieved data indicates that the H3K36M (27-43) inhibitor peptide is stronger bound by the enzymes than the H3K36 (27-43) WT peptide. This observation could be explained by the fact that the lysine has to be deprotonated before its insertion into the hydrophobic substrate binding pocket. In contrast to this the methionine does not need the deprotonation step. Therefore, the hydrophobic side chain of methionine may fit better into the substrate binding pocket. Another explanation of the strong binding of H3K36M to the enzymes could be the sulfur-aromatic interactions between the Y1666 of SETD2 and the side chain of K36M (Yang et al. 2016). This interpretation could also be valid for the NSD1 and NSD2 methyltransferases, because the Y1666 of SETD2 is a well conserved amino acid in the H3K36 methyltransferases.

3.4 Investigation of the RomA PKMT

(manuscript 5 in the attachment of this thesis)

Schuhmacher MK, Rolando M, Bröhm A, Weirich S, Kudithipudi S, Buchrieser C, Jeltsch A. (2018) *Journal of Molecular Biology 430(13):1912-1925. doi: 10.1016/j.jmb.2018.04.032.*

The RomA is a methyltransferase encoded by the Gram-negative bacterium *Legionella pneumophila (L. pneumophila)*. It consists, like human methyltransferases, of a catalytically active SET-domain, which is responsible for the transfer of the methyl group from the cofactor AdoMet to the protein and peptide substrates. Previously, it was shown that RomA methylates lysine 14 of H3 up to the trimethylation state and by this methylation the K14 acetylation site is blocked (Rolando et al. 2013). This change in PTMs results in different gene regulation, because H3K14ac is an activating chromatin modification. As this was not yet investigated for RomA, it was interesting to determine the substrate recognition motif of this PKMT and search for possible human non-histone targets. Another interesting question was, if RomA methylates human non-histone targets during infection with *L. pneumophila*.

As an initial step in the RomA investigation, Alexander Bröhm and Dr. Sara Weirich purified the enzyme and validated the activity of the purified protein. Afterwards they performed a target lysine scan at peptide level to confirm the H3K14 methylation. After verification of the target site methylation and enzyme activity, peptide arrays were used to determine the substrate specificity profile of RomA based on the H3 (7-21) sequence (Figure 36). These experiments were conducted as part of this thesis as described in 5.3.

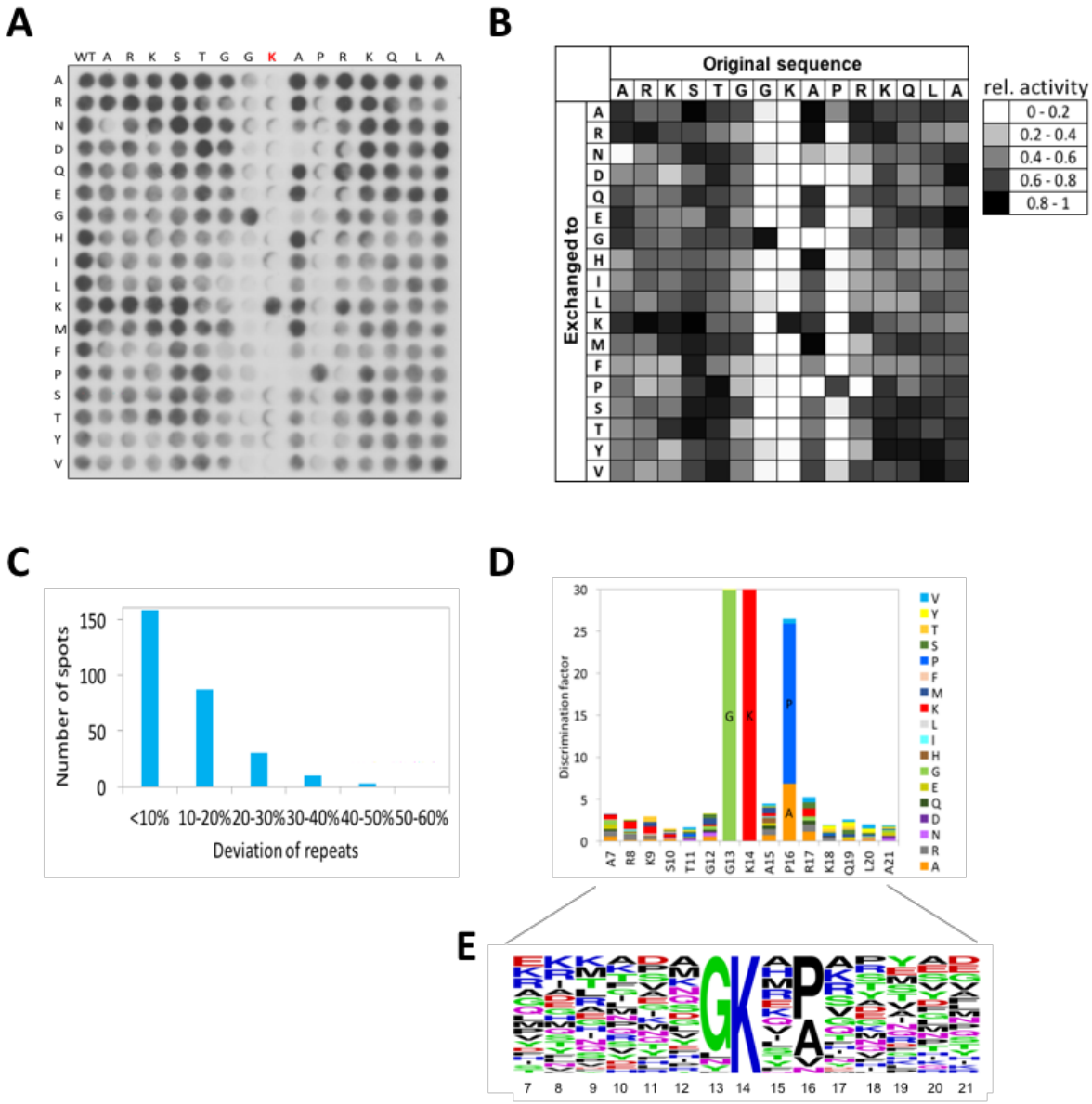


Figure 36: Determination of the substrate specificity profile of RomA. A) Peptide array methylation with RomA and radioactively labeled AdoMet. The vertical axis represents the 18 amino acids which were used to create single amino acid mutations of the H3 (7-21) template sequence shown on the horizontal axis. The red labeled K indicates K14. B) Average of data obtained after quantification of two independent substrate specificity peptide arrays of RomA. The signal intensity is indicated in a greyscale on the right side. C) Standard deviations of relative methylation activities of individual peptides in both arrays. D) Discrimination factors of the substrate specificity of RomA. E) Weblogo of the substrate specificity profile. The figure is taken from (manuscript 5, Figure 2).

The obtained substrate specificity revealed a recognition motif of RomA ranging from the -2 to +3 positions. RomA showed a very specific readout at the position -1 (K14 is annotated as position zero), where only G was accepted, and at the +2 position, in which the natural P and to a much lesser extent A were tolerated. In summary, the substrate specificity of RomA was

defined as follows (the natural amino acids are underlined and the target lysine 14 is printed in red):

(A, N, G, K, M, V, S) - G - K - A (not D, G, P) - (P>A) - (A, R, K>V, S, Q, G)

Next, a Scansite search was performed to identify potential protein substrates in the human proteome. In order to ensure that all possible non-histone targets are included, a relaxed substrate specificity profile was used:

(A, N, G, K, M, V) - G - K - X - P - (A, R, K, V, S)

This search retrieved 93 putative targets, which were synthesized as fifteen amino acid long peptides on a membrane using the SPOT synthesis method. A positive control (H3K14) was included. In addition, for each peptide a negative control was included, in which the predicted target lysine was mutated to alanine. The membrane was processed as described previously showing that 34 putative non-histone target peptides were methylated with similar activity as H3K14 or a little weaker (manuscript 5, Figure 3 A). In addition, 15 peptides were weaker methylated than H3K14 and therefore not further investigated. A deeper look into the sequence of the methylated non-histone target peptides revealed that they showed a strong overrepresentation of positively charged residues (manuscript 5, Figure 3 B) which is in agreement with the structure of a closely related PKMT (Son et al. 2015) showing many acidic residues in the peptide binding site. Among the 34 identified peptide targets, 30 were investigated at protein level *in vitro*. To this end, the putative substrate proteins were cloned as domain or full-length constructs C-terminal to the GST tag, purified by affinity chromatography and methylated with RomA in the presence of radioactively labeled AdoMet (Figure 37 A). After separation by SDS-PAGE, the signal was visualized by autoradiography (Figure 37 B).

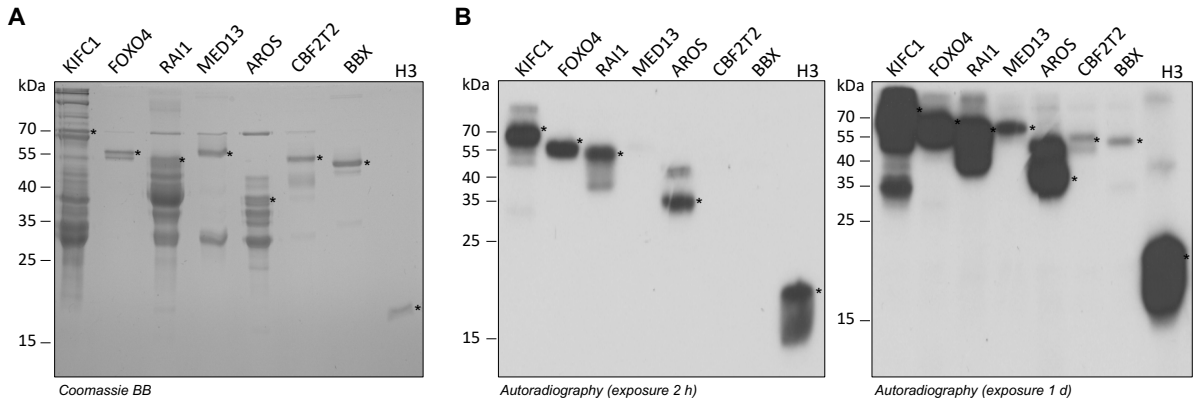


Figure 37: Protein analysis and methylation of non-histone targets of RomA *in vitro*. A) Coomassie BB stained SDS-gel showing the comparable loading of the investigated substrate proteins with recombinant H3 (positive control). The size of the target proteins are marked by asterisks. B) The purified proteins were incubated with methylation buffer containing the RomA enzyme and radioactively labeled AdoMet. The methylation signal was visualized by autoradiography. Two different exposure times are shown to better visualize the weaker methylation signals. The figure was adopted from (manuscript 5, Figure 4).

Among the selected 30 putative targets, nine proteins could be purified and investigated. Four of them were strongly methylated (KIFC1, FOXO4, AROS, RAI1), one moderately (MED13), two weakly (CBFA2T2, BBX) and only two showed no methylation signal (KIF14, TCO1) (manuscript 5, Suppl. Figure 1). To confirm the target lysine methylation, mutations of the target lysine to arginine were created using site-directed mutagenesis (manuscript 5, Figure 5). The mutant proteins were expressed and purified in comparable amount and quality as the wildtype proteins (manuscript 5, Figure 5). For six of the non-histone protein substrates, the target lysine methylation was confirmed. This indicates that these human proteins are targets of RomA and are methylated at the predicted target lysine. For the FOXO4 K189R target lysine mutant, the methylation signal was weaker but still present and this will be addressed in the following paragraph.

In the case of FOXO4, a weaker methylation signal in the mutant protein K189R was observed. This indicates that not only the predicted target lysine is methylated but that there are other lysine residues methylated by RomA as well. A close inspection of the amino acid sequence of FOXO4 revealed that three other sites are possible methylation sites as well. Based on this, three additional mutant proteins were cloned containing double and triple mutants. After expression and purification of the mutant proteins, methylation reactions were performed as

previously described (Figure 38 A). It was observed that K210 is the main methylation target of RomA in FOXO4, which was followed by K189, K186 and K174 (Figure 38; manuscript 5, Figure 6).

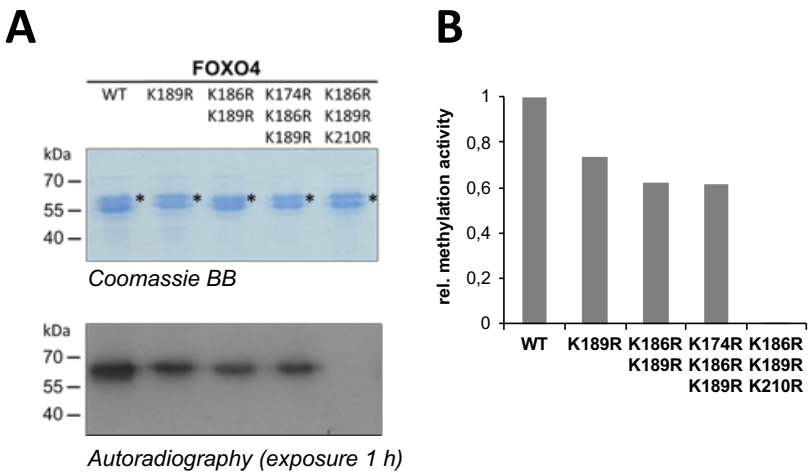


Figure 38: Identification of the target lysine methylation site of FOXO4 by RomA. A) After purification, the GST- tagged wildtype and mutant proteins of FOXO4 were analyzed by SDS-PAGE (Coomassie BB) and afterwards methylated by RomA in the presence of radioactive AdoMet. The transfer of radioactive methyl groups was detected by autoradiography. The asterisks indicate the size of the FOXO4 WT and mutant proteins. B) Quantification of the methylation experiment. Averaging of two experiments gave an error range of $\pm$ 9 %. The figure was adopted from (manuscript 5, Figure 6).

In the following experiments, the target lysine methylation was validated by Western Blot using two different anti-pan di/trimethyl-lysine antibodies (manuscript 5, Figure 7). For these experiments, the substrates were methylated with RomA using unlabeled AdoMet. As negative controls, corresponding samples without RomA were prepared. By Western Blot the methylation of FOXO4, AROS and KIFC1 was confirmed (manuscript 5, Figure 7). The missing antibody signal for BBX is in agreement with a weak methylation signal of this substrate in the radioactive protein methylation experiment (Figure 37). For RAI1, a signal for methylated and unmethylated substrate was observed, which indicates that the antibody is not able to discriminate between both samples. This antibody validation was important for further experiments, which will be described in the next paragraph.

As next step, cellular studies were performed to investigate the methylation of the non-histone protein substrates of RomA in cell lines stably expressing RomA. Additionally

non-histone target methylation experiments were performed during infection with *L. pneumophila* (manuscript 5, Figure 8). Using immunoprecipitation experiments, performed by our collaborators, it was shown that RomA stably expressed in HEK293T cells methylated the AROS substrate (manuscript 5, Figure 8 A). This observation was confirmed in a HEK293T cell line, which expressed a catalytically inactive mutant of RomA. Immunoprecipitation experiments with this cell line showed no AROS methylation. Another step was the investigation of AROS methylation by RomA during infection with *L. pneumophila* in HEK293T cells. Our collaborators observed as well that AROS was methylated by RomA during infection with *L. pneumophila*, whereas no methylation signal was observed for the uninfected cells and cells infected with the RomA deletion *L. pneumophila* strain (manuscript 5, Figure 8 B). These findings were supported by co-transfection experiments of RomA and AROS in Hela cells (manuscript 5, Figure 8 C-E). In these experiments, it was demonstrated that both proteins co-localize in the cell nucleus. In summary, these data indicate that AROS is a target of RomA during infection of human cells by *L. pneumophila*.

3.4.1 Discussion

In this part of the thesis, the substrate specificity profile of RomA was determined. It was shown that not only the motif is important for a binding of RomA but also the surrounding amino acids which should be basic. This suggests that the surrounding amino acids are also contacted by RomA. It was shown that RomA has human non-histone protein substrates, apart from H3K14. Methylation of seven proteins (FOXO4, AROS, RAI1, KIFC1, MED13, BBX and CBFA2T2) by RomA was detected and confirmed by lysine to arginine mutation. In addition, for three targets the methylation was proved by antibody detection. Many other proteins were not analyzed, because of technical problems. These data demonstrated for the first time that a bacterial PKMT methylates human non-histone targets. Based on the successful validation of methylation of many peptide substrates at protein level, it can be proposed that other non-histone target peptides are most likely methylated by RomA at protein level as well. For one of the identified protein substrates (AROS), our collaborators showed that it is a target of RomA during infection with *L. pneumophila* and that it co-localizes with RomA in Hela cell nucleus. Unfortunately, no good antibodies for other substrates were available to perform *in vivo* studies during infection with *L. pneumophila*.

FOXO4 is a member of the mammalian forkhead box O transcription factors which have an influence on stress tolerance, aging and metabolism (van der Horst et al. 2004). In this thesis FOXO4 was identified to be methylated by RomA at four lysine residues, K210 is the most preferred methylation site which is followed by K189, K186 and K174. It is interesting that K186 and K189 are methylated by RomA, because previously it was shown that these two lysine residues and additionally K408 are acetylated by the CREB-binding protein upon peroxidase stress. The acetylation of FOXO4 leads to an inactivation of its transcriptional activity (Fukuoka et al. 2003; van der Horst et al. 2004). The methylation of FOXO4 may block the acetylation keeping FOXO4 in a more active state. Another interesting connection is that the deacteylase SIRT1 is responsible for the deacetylation of the acetylated lysine residues in FOXO4 (van der Horst et al. 2004). The AROS protein is a non-histone target of RomA as well but the influence of its methylation has still to be investigated. Previously it was shown that AROS is an interactor of SIRT1 which may suggest an interplay between these proteins by the deacetylation of FOXO4 (Kim et al. 2007). Another study showed that AROS stimulates the deacetylation of p53 by SIRT1 and it has an influence on the p53-mediated transcriptional activity (Kim et al. 2007).

Other examples of proteins which were methylated by RomA are KIFC1, RAI1, CBFA2T2, BBX and MED13. Their cellular function will be explained in the next sections

- KIFC1: The KIFC1 is a member of the kinesin-14 family and is important for spindle formation (Zou et al. 2014). It also plays a role in the induction of centrosome clustering in cancer cells (Pawar et al. 2014; Yang et al. 2014). It was also shown that it is overexpressed in human ovarian tumors and by this it was correlated with tumor aggressiveness (Pawar et al. 2014). In other cancer types like for example estrogen receptor (ER)-positive breast cancer it was shown that KIFC1 expression is induced by estrogen (Zou et al. 2014). Another study showed that KIFC1 can be inhibited by AZ82 (Yang et al. 2014).

- RAI1: The RAI1 protein is involved in many developmental diseases like for example Smith-Magenis and Potocki-Lupski Syndromes (Carmona-Mora et al. 2010). In this study, they also showed that the N-terminal part of RAI1 is responsible for its

transactivational activity and its C-terminal part has the ability for nuclear localization. The methylated lysine residue of RAI1 is located within this nuclear localization domain and by this has maybe an influence in the nuclear localization of RAI1. Thereafter it was demonstrated that RAI1 is also involved in the regulation of the circadian clock components (Williams et al. 2012).

- CBFA2T2: It is a Myeloid Translocation Gene (MTG) (Barrett et al. 2011). Other members of the MTG proteins are involved in acute myeloid leukemia (AML). It was shown that CBFA2T2 is able to regulate pluripotency and germline specification in mice (Tu et al. 2016). In mice it is capable to form a complex with PRDM14. By oligomerization of CBFA2T2 the PRDM14 and OCT4 are stabilized on chromatin (Tu et al. 2016).

- BBX: This protein belongs to the high mobility group (HMG)-box proteins. These proteins are involved in DNA repair or replication (Chen et al. 2014). The Bobby sox (BBX) is a potential sequence-specific transcription factor (Chen et al. 2014).

- MED13: Cdk8 and cycline C (CycC) are able to perform gene regulation. The MED13 together with MED12 are subunits of Cdk8 that are important for the linkage of Cdk8 and CycC (Kuuluvainen et al. 2014). Another aspect is that the ubiquitin ligase SCF-FBW7 is able to targets MED13 and MED13L for proteasomal degradation (Davis et al. 2013). To form this interaction the T326 of MED13 is important (Davis et al. 2013). MED13 haploinsufficiency is linked to a disease which is characterized by cataract, hearing loss and semicircular canal dysplasia (Boutry-Kryza et al. 2012).

However, the influence of the methylation of these proteins on their function is not known and needs further investigation.

4. Conclusions and outlook

After the identification of the first PKMT in 2000, several additional PKMTs were determined. Within this emerging field, the structure and basic mechanisms of PKMTs were studied but the focus was also turned to the investigation of their role in the human regulatory mechanisms. To analyze the functional role of a PKMT, it is important to know its complete substrate spectrum. Today different methods have been established to search for novel PKMT substrates, one of which is based on the substrate specificity profiles. In general PKMTs methylate lysine residues embedded in a defined amino acid sequence. However, the different positions of the target sequence are recognized with different stringencies and preferences. The preferences can be most efficiently studied by peptide array methylation experiments which allow to investigate the methylation of many (at least 320) peptides in one reaction. Hence methylation of all possible single amino acid mutants of a template sequence can be studied at the same time providing a detailed information about the recognition of each amino acid.

An interesting aspect of the substrate specificity is that it can be used for the discrimination between two PKMTs, which are able to methylate the same histone target. In the present thesis this was demonstrated for the SUV39H1 and SUV39H2 enzyme which have different substrate specificity motifs and based on this non-overlapping non-histone substrates. It is essential to know the substrate recognition discrepancies of enzymes which methylate the same histone target, in order to understand their different biological roles better. Similar results have been obtained in the past, when comparing SUV4-20H1 and H2 (Weirich et al. 2016), or a larger panel of H3K9 specific PKMTs including the SUV39H2 and SUV39H1 data from this study but also Clr4 (from *S. pombe*), Dim-5 (form *N. crassa*), and human G9a (Kusevic et al. 2017). This analysis has shown that the substrate specificity profiles of these enzymes can be clustered and grouped, showing for example that the profile of SUV39H2 is most similar to Clr4 while SUV39H1 is most similar to G9a (Figure 39).

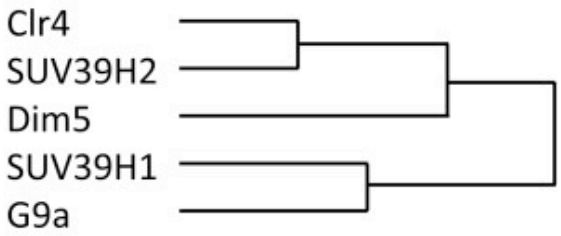


Figure 39: Cluster of the similarity between the substrate specificity profiles of different H3K9 specific methyltransferases. The figure was adopted from (Kusevic et al. 2017).

The knowledge of the substrate specificity motif can guide the search for additional targets of the investigated PKMTs. Based on the sequence profile a preselection of the most promising substrate candidates can be performed such that it is not necessary to analyze the whole proteome for putative targets. As shown in this thesis in most cases the knowledge of the substrate specificity motif has led to the determination of additional peptide targets, and it also allowed the recognition of potential wrong assignments of non-histone substrates to a PKMT as shown here for the H2AX non-histone target and SUV39H2 enzyme. The identification of wrong assignments of PKMT substrates is important to prevent misinterpretations in follow up studies, which then can lead to connections between processes that do not correlate with each other.

However, the determined non-histone targets based on peptide array experiments do not always reflect the situation at protein level as shown in this thesis for the SETD2 enzyme. One reason for this could be that the specificity profile of SETD2 is rich in hydrophobic amino acids, which usually are buried in folded proteins and therefore they are not available as interaction partners for the PKMT. In the case of these substrates one could envisage that the PKMT may methylate fully or partially unfolded proteins or proteins *in statu nascendi* immediately following protein biosynthesis, but this has not yet been systematically studied for any enzyme of this type. For this it would be an interesting question to study if PKMTs interact with chaperones directly acting on newly synthesized proteins. Another promising research approach which takes up this observation could be to search for PKMT target site in natively unfolded protein regions. These regions can be predicted (Li et al. 2015) and this prediction could be combined with the sequence based search for new PKMT substrates.

Another type of substrate recognition of PKMTs is not based on a stringent readout of the sequence surrounding the target lysine, but on protein/protein interaction of the substrate protein and PKMT. By this, one lysine residue of the substrate protein can be brought in close proximity to the PKMT active site, leading to its methylation. This type of substrate interaction cannot be predicted by a specificity profile analysis representing one limitation of this approach. This aspect has to be considered in the interpretation of the obtained results, suggesting that PKMTs may have a second class of substrates based on protein/protein interaction. In this analysis, the available information about PKMT interacting proteins are usually included and known interactors are inspected carefully for potential methylation sites. Protein/protein interaction maps are still incomplete, and it cannot be excluded that in the context of a strong interaction, peptide sequence requirements for methylation may be relaxed. This assumption is supported by the findings that many PKMTs catalyze automethylation (in this thesis seen for SETD2), although they do not contain *bona fide* substrate sequences.

As demonstrated in the present thesis for SETD2 the substrate specificity analysis can also be used for the generation of a “super-substrate”, in particular if it reveals preference for amino acids that differ from the known best substrate. In this case it is possible to create a substrate which only contains the most preferred amino acids at each sequence position. As shown here for the “super-substrate” peptide of SETD2 this design was very successful as this super-substrate is much better methylated than H3K36, as indicated by the observation, that the super-substrate was methylated at protein level, where the natural H3K36 substrate is not methylated (only in nucleosomal context). Similarly, at peptide level it was observed that the super-substrate is methylated 50 times stronger than H3K36 substrate. In the case of SETD2 no human protein exists which contains this sequence, leading to the puzzling question of why SETD2 has been developed in evolution to methylate a sequence best that does not exist.

In summary, it is crucial to know the substrate specificity profile of PKMTs, because based on this, new substrates can be recognized as demonstrated here for SUV39H1, SETD2 and RomA. The description of new substrates of PKMTs leads then to the understanding of other processes in more detail and prevent wrong assignments of targets to a PKMT. The discovery of novel non-histone substrates of PKMTs is particularly important to gain more insights into

their biological role, because the primary effect of PKMTs is the methylation of substrate proteins. This methylation causes then downstream effects, like changing of the protein stability or modulation of protein/protein interaction, which then finally lead to biological outcomes.

5. Material and Methods

The Material and Methods section describes the methods, which were not explained in detail in the published papers (Appendix II).

5.1 Cloning

The cloning of the domain or full-length non-histone substrates were performed by PCR using cDNA isolated from HEK293 cells as template. The non-histone targets were cloned into pGEX-6p2 vector (GE Healthcare) for bacterial expression. Using Scooby domain prediction tool (George et al. 2005) the domains of the non-histone targets were predicted. Information about the cloned domains of the SETD2 non-histone targets are provided in the results part in Table 5.

Using site-directed mutagenesis the lysine to arginine mutations were introduced into the corresponding non-histone WT plasmid (Jeltsch and Lanio 2002). The same method was used to introduce the N324K mutation into SUV39H2 and to create the GST-H3K36M (29-43) super-substrate mutant or the GST-super-substrate protein. The introduction of the sequence or the mutation was confirmed by sequencing.

5.2 Protein overexpression

The His_6-tagged SETD2 construct was kindly provided by Dr. Masoud Vedadi. The SETD2 (amino acids 1347-1711, Uniprot No: Q9BYW2) was cloned into pET28-MHL bacterial expression vector and transformed into BL21 DE3 Codon Plus cells (Novagen) for protein expression. The overexpression of His_6-SETD2 was performed in Luria-Bertani medium supplemented with kanamycin. The pre-culture was grown for 7 hours at 37 °C. The following main culture was grown at 37 °C till an OD_{600} between 0.6 and 0.8 was reached. Afterwards the protein expression was induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and protein expression was performed in a shaker at 20 °C over night. At the next day, the cells were harvested at 4000 g and 4 °C for 20 minutes and resuspended in 1x STE buffer (100 mM NaCl, 10 mM Tris/HCl pH 8, 0.1 mM EDTA). Another centrifugation step at 4500 g, 4 °C and 25 minutes was performed which was followed by the storage of the obtained cell pellet at -20 °C.

5.3 Peptide array methylation

The peptide arrays were produced by different members of the lab, using the SPOT synthesis method and an Autospot peptide array synthesizer (Intavis AG, Köln) as described previously (Frank 2002; Kudithipudi et al. 2014a). The peptide arrays contain fifteen amino acid long peptides which were immobilized on a cellulose membrane. Each membrane contains at least one positive and one negative control to show the successful synthesis of the peptides and the technical quality of the experiment.

The cellulose membrane was pre-incubated for 5 minutes at 25 °C in methylation buffer without enzyme and radioactively labeled AdoMet (Table 7). Afterwards the membrane was incubated for 60 minutes in methylation buffer supplemented with 0.76 µM radioactively labeled AdoMet (Perkin Elmer) and the respective enzyme. Thereafter the membrane was washed for 5 times 5 minutes in wash buffer (100 mM NaH_4CO_3, 1 % SDS) and incubated for 5 minutes in Amplify NAMP100V (GE Healthcare). The transferred radioactively labeled methyl groups were detected by hyperfilm™ high performance autoradiography film (GE Healthcare). Thereafter, the film was developed using an Optimax Typ TR machine after different periods of time.

The autoradiography images were scanned and the signal intensity was measured by Phoretix™ array analyzing software. Afterwards the data was transferred to Microsoft Office Excel and analyzed. First the obtained data was normalized and then the individual experiments were averaged. Based on the averaged data of at least two duplicates, the standard errors of the methylation levels of individual peptides were determined. Thereafter for the substrate specificity motif, the discrimination factors were calculated as described previously (Kudithipudi et al. 2014a). These factors show the preference of the enzyme for different amino acids at specific positions. A Weblogo based on the averaged data, can be used as another step for the visualization of the substrate specificity motif.

Table 7: Composition of the different methylation buffers for the different PKMTs. This table shows the composition of the methylation buffer for both SUV39H enzymes, SETD2 and the bacterial methyltransferase RomA.

Component	SUV39H1	SUV39H2	SETD2	RomA
Tris/HCl pH 9 [mM]	50	50	20	50
$MgCl_2$ [mM]	5	5	1.5	2.5
Triton X-100 [%]	-	-	0.01	-
DTT [mM]	4	4	10	4
enzyme [µM]	0.05	0.05	3.0	0.0075 - 0.06

5.4 Protein methylation assay

The methylation reaction of the purified non-histone target proteins, H2AX and recombinant H3 (NEW ENGLAND BioLabs (NEB)) was performed in the enzyme specific methylation buffer (see Table 7) supplemented with 0.76 µM radioactive labeled AdoMet and the respective enzyme (Table 8) for 3 hours at 25 °C. Afterwards the reaction was halted by the addition of 5x SDS loading buffer and incubation at 95 °C for 5 minutes. Then the samples were loaded onto a 16 % SDS-gel and after separation the gel was incubated for 45 minutes in Amplify NAMP100V (GE Healthcare) on a shaker. Thereafter the SDS-gel was dried under vacuum at 65 °C for 90 minutes. This step was followed by the visualization of the transferred radioactive methyl groups using a hyperfilm™ high performance autoradiography film (GE Healthcare). The film was developed with an Optimax Typ TR machine after different periods of time.

The autoradiography images were scanned and the signal intensity was measured using the IMAGE STUDIO (LI-COR) software. Thereafter, the background was subtracted and the obtained values were normalized. The normalization was followed by averaging of the individual experiments. Based on the averaged values, the standard error of mean or the standard deviation was calculated.

Table 8: Enzyme concentrations of both SUV39H enzymes, SETD2 and RomA. Used concentrations of the different PKMTs in protein methylation experiments.

Enzyme	SUV39H1	SUV39H2	SETD2	RomA
concentration [µM]	1.4	0.2 - 0.5	6	0.9

5.5 Tricine-SDS-PAGE

The Tricine-SDS-PAGE is used to separate small molecules, e.g. peptides. The Tricine-gel consists of three parts. The 4 % stacking gel (4 % acrylamide (29:1), 1x gel buffer (Table 9)), a 10 % spacer gel (10 % acrylamide (29:1), 1x gel buffer, 10 % glycerol) and a 16 % resolving gel (16 % acrylamide (19:1), 1x gel buffer, 10 % glycerol). First the 16 % resolving gel has to be prepared, polymerized using 40 % APS and TEMED for 20 minutes and overlaid with isopropanol. After polymerization, the isopropanol is removed and the 10 % spacer gel solution supplemented with 40 % APS and TEMED is added and overlaid with isopropanol. Another 20 minutes later isopropanol is removed and the 4 % stacking gel is casted for 20 minutes. The samples were supplemented with 2x Tricine-SDS loading buffer (50 mM Tris/HCl pH 6.8, 4 % SDS, 12 % glycerol, 2 % β-mercaptoethanol, 0.01 % bromophenol blue) and afterwards incubated at 95 °C for 5 minutes. As running buffers an 1x anode and 1x cathode buffer were used (Table 9). The separated peptides can be visualized by Coomassie BB staining or by autoradiography in case of radioactive samples.

Table 9: Buffer composition of used buffers in Tricine-SDS-PAGE.

Ingredients	Anode buffer (10x)	Cathode buffer (10x)	Gel buffer (3x)
Tris [M]	1	1	3
Tricine [M]	-	1	-
HCl [M]	0.225	-	1
SDS [%]	-	1	0.3
pH	8.9	8.25	8.45

References

Allali-Hassani, A. et al. (2012). "Fluorescence-based methods for screening writers and readers of histone methyl marks." J Biomol Screen **17**(1): 71-84.

Allfrey, V. G. and Mirsky, A. E. (1964). "Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis." Science **144**(3618): 559.

Baker, M. (2011). "Making sense of chromatin states." Nat Methods **8**(9): 717-722.

Balakrishnan, L. and Milavetz, B. (2010). "Decoding the histone H4 lysine 20 methylation mark." Crit Rev Biochem Mol Biol **45**(5): 440-452.

Bannister, A. J. and Kouzarides, T. (2005). "Reversing histone methylation." Nature **436**(7054): 1103-1106.

Bannister, A. J. et al. (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain." Nature **410**(6824): 120-124.

Barrett, C. W. et al. (2011). "MTGR1 is required for tumorigenesis in the murine AOM/DSS colitis-associated carcinoma model." Cancer Res **71**(4): 1302-1312.

Behjati, S. et al. (2013). "Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone." Nat Genet **45**(12): 1479-1482.

Berger, K. H. and Isberg, R. R. (1993). "Two distinct defects in intracellular growth complemented by a single genetic locus in Legionella pneumophila." Mol Microbiol **7**(1): 7-19.

Boutry-Kryza, N. et al. (2012). "An 800 kb deletion at 17q23.2 including the MED13 (THRAP1) gene, revealed by aCGH in a patient with a SMC 17p." Am J Med Genet A **158A**(2): 400-405.

Carmona-Mora, P. et al. (2010). "Functional and cellular characterization of human Retinoic Acid Induced 1 (RAI1) mutations associated with Smith-Magenis Syndrome." BMC Mol Biol **11**: 63.

Carvalho, S. et al. (2014). "SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint." Elife **3**: e02482.

Chen, P. et al. (2013). "Histone variants in development and diseases." J Genet Genomics **40**(7): 355-365.

Chen, T. et al. (2014). "Characterization of Bbx, a member of a novel subfamily of the HMG-box superfamily together with Cic." Dev Genes Evol **224**(4-6): 261-268.

Cheng, X. et al. (2005). "Structural and sequence motifs of protein (histone) methylation enzymes." Annu Rev Biophys Biomol Struct **34**: 267-294.

Chuikov, S. et al. (2004). "Regulation of p53 activity through lysine methylation." Nature **432**(7015): 353-360.

Copeland, R. A. et al. (2009). "Protein methyltransferases as a target class for drug discovery." Nat Rev Drug Discov **8**(9): 724-732.

Crooks, G. E. et al. (2004). "WebLogo: a sequence logo generator." Genome Res **14**(6): 1188-1190.

Davis, M. A. et al. (2013). "The SCF-Fbw7 ubiquitin ligase degrades MED13 and MED13L and regulates CDK8 module association with Mediator." Genes Dev **27**(2): 151-156.

Dhayalan, A. et al. (2011). "Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase." Chem Biol **18**(1): 111-120.

Dhayalan, A. et al. (2010). "The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation." J Biol Chem **285**(34): 26114-26120.

Dietz, H. C. et al. (1991). "Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene." Nature **352**(6333): 337-339.

Dillon, S. C. et al. (2005). "The SET-domain protein superfamily: protein lysine methyltransferases." Genome Biol **6**(8): 227.

Duns, G. et al. (2010). "Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma." Cancer Res **70**(11): 4287-4291.

Dupont, C. et al. (2009). "Epigenetics: definition, mechanisms and clinical perspective." Semin Reprod Med **27**(5): 351-357.

Edmunds, J. W. et al. (2008). "Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation." EMBO J **27**(2): 406-420.

Egorova, K. S. et al. (2010). "Lysine methylation of nonhistone proteins is a way to regulate their stability and function." Biochemistry (Mosc) **75**(5): 535-548.

Eram, M. S. et al. (2015). "Kinetic characterization of human histone H3 lysine 36 methyltransferases, ASH1L and SETD2." Biochim Biophys Acta **1850**(9): 1842-1848.

Falnes, P. O. et al. (2016). "Protein lysine methylation by seven-beta-strand methyltransferases." Biochem J **473**(14): 1995-2009.

Fischle, W. et al. (2003). "Histone and chromatin cross-talk." Curr Opin Cell Biol **15**(2): 172-183.

Flemming, W. (1882). Zellsubstanz, Kern und Zellteilung, Leipzig, F.C.W. Vogel.

Fnu, S. et al. (2011). "Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining." Proc Natl Acad Sci U S A **108**(2): 540-545.

Fontebasso, A. M. et al. (2013). "Mutations in SETD2 and genes affecting histone H3K36 methylation target hemispheric high-grade gliomas." Acta Neuropathol **125**(5): 659-669.

Francastel, C. et al. (2000). "Nuclear compartmentalization and gene activity." Nat Rev Mol Cell Biol **1**(2): 137-143.

Frank, R. (2002). "The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports--principles and applications." J Immunol Methods **267**(1): 13-26.

Fraser, D. W. et al. (1977). "Legionnaires' disease: description of an epidemic of pneumonia." N Engl J Med **297**(22): 1189-1197.

Fukuda, A. et al. (2014). "The role of maternal-specific H3K9me3 modification in establishing imprinted X-chromosome inactivation and embryogenesis in mice." Nat Commun **5**: 5464.

Fukuoka, M. et al. (2003). "Negative regulation of forkhead transcription factor AFX (Foxo4) by CBP-induced acetylation." Int J Mol Med **12**(4): 503-508.

George, R. A. et al. (2005). "Scooby-domain: prediction of globular domains in protein sequence." Nucleic Acids Res **33**(Web Server issue): W160-163.

Goldberg, A. D. et al. (2007). "Epigenetics: a landscape takes shape." Cell **128**(4): 635-638.

Grant, P. A. et al. (1999). "Expanded lysine acetylation specificity of Gcn5 in native complexes." J Biol Chem **274**(9): 5895-5900.

Guo, A. et al. (2014). "Immunoaffinity enrichment and mass spectrometry analysis of protein methylation." Mol Cell Proteomics **13**(1): 372-387.

Hacker, K. E. et al. (2016). "Structure/Function Analysis of Recurrent Mutations in SETD2 Protein Reveals a Critical and Conserved Role for a SET Domain Residue in Maintaining Protein Stability and Histone H3 Lys-36 Trimethylation." J Biol Chem **291**(40): 21283-21295.

Huang, J. et al. (2010). "G9a and Glp methylate lysine 373 in the tumor suppressor p53." J Biol Chem **285**(13): 9636-9641.

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

Hyun, K. et al. (2017). "Writing, erasing and reading histone lysine methylations." Exp Mol Med **49**(4): e324.

Iwabata, H. et al. (2005). "Proteomic analysis of organ-specific post-translational lysine-acetylation and -methylation in mice by use of anti-acetyllysine and -methyllysine mouse monoclonal antibodies." Proteomics **5**(18): 4653-4664.

Jagannathan, V. et al. (2013). "A mutation in the SUV39H2 gene in Labrador Retrievers with hereditary nasal parakeratosis (HNPK) provides insights into the epigenetics of keratinocyte differentiation." PLoS Genet **9**(10): e1003848.

Jeltsch, A. and Lanio, T. (2002). "Site-directed mutagenesis by polymerase chain reaction." Methods Mol Biol **182**: 85-94.

Jensen, S. A. and Handford, P. A. (2016). "New insights into the structure, assembly and biological roles of 10-12 nm connective tissue microfibrils from fibrillin-1 studies." Biochem J **473**(7): 827-838.

Jenuwein, T. et al. (1998). "SET domain proteins modulate chromatin domains in eu- and heterochromatin." Cell Mol Life Sci **54**(1): 80-93.

Johnson, W. L. et al. (2017). "RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin." Elife **6**.

Jorgensen, S. et al. (2013). "Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity." Nucleic Acids Res **41**(5): 2797-2806.

Justin, N. et al. (2016). "Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2." Nat Commun **7**: 11316.

Karmodiya, K. et al. (2012). "H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells." BMC Genomics **13**: 424.

Kim, E. J. et al. (2007). "Active regulator of SIRT1 cooperates with SIRT1 and facilitates suppression of p53 activity." Mol Cell **28**(2): 277-290.

Kizer, K. O. et al. (2005). "A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation." Mol Cell Biol **25**(8): 3305-3316.

Kouzarides, T. (2002). "Histone methylation in transcriptional control." Curr Opin Genet Dev **12**(2): 198-209.

Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.

Kudithipudi, S. et al. (2012). "The SET8 H4K20 protein lysine methyltransferase has a long recognition sequence covering seven amino acid residues." Biochimie **94**(11): 2212-2218.

Kudithipudi, S. and Jeltsch, A. (2016). "Approaches and Guidelines for the Identification of Novel Substrates of Protein Lysine Methyltransferases." Cell Chem Biol **23**(9): 1049-1055.

Kudithipudi, S. et al. (2014a). "Specificity analysis of protein lysine methyltransferases using SPOT peptide arrays." J Vis Exp(93): e52203.

Kudithipudi, S. et al. (2014b). "Substrate specificity analysis and novel substrates of the protein lysine methyltransferase NSD1." Chem Biol **21**(2): 226-237.

Kusevic, D. et al. (2017). "Clr4 specificity and catalytic activity beyond H3K9 methylation." Biochimie **135**: 83-88.

Kuuluvainen, E. et al. (2014). "Cyclin-dependent kinase 8 module expression profiling reveals requirement of mediator subunits 12 and 13 for transcription of Serpent-dependent innate immunity genes in Drosophila." J Biol Chem **289**(23): 16252-16261.

Lachner, M. et al. (2003). "An epigenetic road map for histone lysine methylation." J Cell Sci **116**(Pt 11): 2117-2124.

Lanouette, S. et al. (2014). "The functional diversity of protein lysine methylation." Mol Syst Biol **10**: 724.

Leventopoulos, G. et al. (2009). "A clinical study of Sotos syndrome patients with review of the literature." Pediatr Neurol **40**(5): 357-364.

Li, F. et al. (2013a). "The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha." Cell **153**(3): 590-600.

Li, J. et al. (2015). "An Overview of Predictors for Intrinsically Disordered Proteins over 2010-2014." Int J Mol Sci **16**(10): 23446-23462.

Li, T. et al. (2013b). "SET-domain bacterial effectors target heterochromatin protein 1 to activate host rDNA transcription." EMBO Rep **14**(8): 733-740.

Lu, C. et al. (2016). "Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape." Science **352**(6287): 844-849.

Lu, T. et al. (2010). "Regulation of NF-kappaB by NSD1/FBXL11-dependent reversible lysine methylation of p65." Proc Natl Acad Sci U S A **107**(1): 46-51.

Lucio-Eterovic, A. K. et al. (2010). "Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function." Proc Natl Acad Sci U S A **107**(39): 16952-16957.

Luger, K. (2001). Nucleosome: Structure and Function, Nature Publishing Group.

Luger, K. et al. (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-260.

Margueron, R. et al. (2009). "Role of the polycomb protein EED in the propagation of repressive histone marks." Nature **461**(7265): 762-767.

Margueron, R. et al. (2005). "The key to development: interpreting the histone code?" Curr Opin Genet Dev **15**(2): 163-176.

Marmorstein, R. (2001). "Structure and function of histone acetyltransferases." Cell Mol Life Sci **58**(5-6): 693-703.

Marra, A. et al. (1992). "Identification of a Legionella pneumophila locus required for intracellular multiplication in human macrophages." Proc Natl Acad Sci U S A **89**(20): 9607-9611.

Melcher, M. et al. (2000). "Structure-function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation, and mitotic progression." Mol Cell Biol **20**(10): 3728-3741.

Min, J. et al. (2003). "Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase." Cell **112**(5): 711-723.

Muller, M. M. et al. (2016). "A two-state activation mechanism controls the histone methyltransferase Suv39h1." Nat Chem Biol **12**(3): 188-193.

Murray, K. (1964). "The Occurrence of Epsilon-N-Methyl Lysine in Histones." Biochemistry **3**: 10-15.

Mutonga, M. et al. (2015). "Targeting Suppressor of Variegation 3-9 Homologue 2 (SUV39H2) in Acute Lymphoblastic Leukemia (ALL)." Transl Oncol **8**(5): 368-375.

Nakayama, J. et al. (2001). "Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly." Science **292**(5514): 110-113.

O'Carroll, D. et al. (2000). "Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression." Mol Cell Biol **20**(24): 9423-9433.

Obenauer, J. C. et al. (2003). "Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs." Nucleic Acids Res **31**(13): 3635-3641.

Olins, D. E. and Olins, A. L. (2003). "Chromatin history: our view from the bridge." Nat Rev Mol Cell Biol **4**(10): 809-814.

Pal, S. et al. (2004). "Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes." Mol Cell Biol **24**(21): 9630-9645.

Park, I. Y. et al. (2016). "Dual Chromatin and Cytoskeletal Remodeling by SETD2." Cell **166**(4): 950-962.

Pasillas, M. P. et al. (2011). "NSD1 PHD domains bind methylated H3K4 and H3K9 using interactions disrupted by point mutations in human sotos syndrome." Hum Mutat **32**(3): 292-298.

Pawar, S. et al. (2014). "KIFCI, a novel putative prognostic biomarker for ovarian adenocarcinomas: delineating protein interaction networks and signaling circuitries." J Ovarian Res **7**: 53.

Peters, A. H. et al. (2003). "Partitioning and plasticity of repressive histone methylation states in mammalian chromatin." Mol Cell **12**(6): 1577-1589.

Peters, A. H. et al. (2001). "Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability." Cell **107**(3): 323-337.

Piao, L. et al. (2015). "SUV39H2 methylates and stabilizes LSD1 by inhibiting polyubiquitination in human cancer cells." Oncotarget **6**(19): 16939-16950.

Pinheiro, I. et al. (2012). "Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity." Cell **150**(5): 948-960.

Qian, C. and Zhou, M. M. (2006). "SET domain protein lysine methyltransferases: Structure, specificity and catalysis." Cell Mol Life Sci **63**(23): 2755-2763.

Qiao, Q. et al. (2011). "The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation." J Biol Chem **286**(10): 8361-8368.

Rao, V. K. et al. (2017). "A drive in SUVs: From development to disease." Epigenetics **12**(3): 177-186.

Rathert, P. et al. (2008a). "Protein lysine methyltransferase G9a acts on non-histone targets." Nat Chem Biol **4**(6): 344-346.

Rathert, P. et al. (2008b). "Analysis of the substrate specificity of the Dim-5 histone lysine methyltransferase using peptide arrays." Chem Biol **15**(1): 5-11.

Rayasam, G. V. et al. (2003). "NSD1 is essential for early post-implantation development and has a catalytically active SET domain." EMBO J **22**(12): 3153-3163.

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

Redon, C. et al. (2002). "Histone H2A variants H2AX and H2AZ." Curr Opin Genet Dev **12**(2): 162-169.

Rega, S. et al. (2001). "Identification of the full-length huntingtin- interacting protein p231HBP/HYPB as a DNA-binding factor." Mol Cell Neurosci **18**(1): 68-79.

Rice, J. C. et al. (2003). "Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains." Mol Cell **12**(6): 1591-1598.

Rolando, M. and Buchrieser, C. (2014). "Legionella pneumophila type IV effectors hijack the transcription and translation machinery of the host cell." Trends Cell Biol **24**(12): 771-778.

Rolando, M. et al. (2013). "Legionella pneumophila effector RomA uniquely modifies host chromatin to repress gene expression and promote intracellular bacterial replication." Cell Host Microbe **13**(4): 395-405.

Sakai, L. Y. et al. (1986). "Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils." J Cell Biol **103**(6 Pt 1): 2499-2509.

Schotta, G. et al. (2004). "A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin." Genes Dev **18**(11): 1251-1262.

Segal, E. et al. (2006). "A genomic code for nucleosome positioning." Nature **442**(7104): 772-778.

Segal, G. et al. (1998). "Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the Legionella pneumophila genome." Proc Natl Acad Sci U S A **95**(4): 1669-1674.

Shi, X. et al. (2007). "Modulation of p53 function by SET8-mediated methylation at lysine 382." Mol Cell **27**(4): 636-646.

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

Shirai, A. et al. (2017). "Impact of nucleic acid and methylated H3K9 binding activities of Suv39h1 on its heterochromatin assembly." Elife **6**.

Son, J. et al. (2015). "Crystal structure of Legionella pneumophila type IV secretion system effector LegAS4." Biochem Biophys Res Commun **465**(4): 817-824.

Sone, K. et al. (2014). "Critical role of lysine 134 methylation on histone H2AX for gamma-H2AX production and DNA repair." Nat Commun **5**: 5691.

Sudol, M. and Hunter, T. (2000). "NeW wrinkles for an old domain." Cell **103**(7): 1001-1004.

Sun, X. J. et al. (2005). "Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase." J Biol Chem **280**(42): 35261-35271.

Tachibana, M. et al. (2007). "Functional dynamics of H3K9 methylation during meiotic prophase progression." EMBO J **26**(14): 3346-3359.

Tamaru, H. and Selker, E. U. (2001). "A histone H3 methyltransferase controls DNA methylation in Neurospora crassa." Nature **414**(6861): 277-283.

Tamaru, H. et al. (2003). "Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in Neurospora crassa." Nat Genet **34**(1): 75-79.

Trewick, S. C. et al. (2005). "Methylation: lost in hydroxylation?" EMBO Rep **6**(4): 315-320.

Tu, S. et al. (2016). "Co-repressor CBFA2T2 regulates pluripotency and germline development." Nature **534**(7607): 387-390.

van der Horst, A. et al. (2004). "FOXO4 is acetylated upon peroxide stress and deacetylated by the longevity protein hSir2(SIRT1)." J Biol Chem **279**(28): 28873-28879.

Vaquero, A. et al. (2007). "SIRT1 regulates the histone methyl-transferase SUV39H1 during heterochromatin formation." Nature **450**(7168): 440-444.

Waddington, C. H. (1968). Towards a Theoretical Biology.

Wang, D. et al. (2013). "Methylation of SUV39H1 by SET7/9 results in heterochromatin relaxation and genome instability." Proc Natl Acad Sci U S A **110**(14): 5516-5521.

Wang, T. et al. (2012). "Crystal structure of the human SUV39H1 chromodomain and its recognition of histone H3K9me2/3." PLoS One **7**(12): e52977.

Weirich, S. et al. (2016). "Specificity of the SUV4-20H1 and SUV4-20H2 protein lysine methyltransferases and methylation of novel substrates." J Mol Biol **428**(11): 2344-2358.

Weirich, S. et al. (2015). "Investigation of the methylation of Numb by the SET8 protein lysine methyltransferase." Sci Rep **5**: 13813.

West, L. E. and Gozani, O. (2011). "Regulation of p53 function by lysine methylation." Epigenomics **3**(3): 361-369.

Williams, S. R. et al. (2012). "Smith-Magenis syndrome results in disruption of CLOCK gene transcription and reveals an integral role for RAI1 in the maintenance of circadian rhythmicity." Am J Hum Genet **90**(6): 941-949.

Wu, C. and Morris, J. R. (2001). "Genes, genetics, and epigenetics: a correspondence." Science **293**(5532): 1103-1105.

Wu, H. et al. (2010). "Structural biology of human H3K9 methyltransferases." PLoS One **5**(1): e8570.

Xiao, B. et al. (2003). "Structure and catalytic mechanism of the human histone methyltransferase SET7/9." Nature **421**(6923): 652-656.

Xie, P. et al. (2008). "Histone methyltransferase protein SETD2 interacts with p53 and selectively regulates its downstream genes." Cell Signal **20**(9): 1671-1678.

Yang, B. et al. (2014). "Discovery of potent KIFC1 inhibitors using a method of integrated high-throughput synthesis and screening." J Med Chem **57**(23): 9958-9970.

Yang, H. et al. (2008). "Preferential dimethylation of histone H4 lysine 20 by Suv4-20." J Biol Chem **283**(18): 12085-12092.

Yang, S. et al. (2016). "Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase." Genes Dev **30**(14): 1611-1616.

Yang, Y. and Bedford, M. T. (2013). "Protein arginine methyltransferases and cancer." Nat Rev Cancer **13**(1): 37-50.

Yun, M. et al. (2011). "Readers of histone modifications." Cell Res **21**(4): 564-578.

Zhang, X. and Bruice, T. C. (2008). "Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases." Proc Natl Acad Sci U S A **105**(15): 5728-5732.

Zhang, X. et al. (2002). "Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase." Cell **111**(1): 117-127.

Zou, J. X. et al. (2014). "Kinesin family deregulation coordinated by bromodomain protein ANCCA and histone methyltransferase MLL for breast cancer cell growth, survival, and tamoxifen resistance." Mol Cancer Res **12**(4): 539-549.

Appendix I

Table S1:

List of putative non-histone peptide substrates of SETD2 based on the super-substrate sequence.

Table S2:

List of selected putative non-histone peptide substrates of SETD2 based on the super-substrate sequence including the lysine to alanine mutation.

Appendix II (not included in the published thesis)

Manuscript 1:

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.*

Manuscript 2:

Schuhmacher MK, Kudithipudi S, Jeltsch A. (2016). **Investigation of H2AX methylation by the SUV39H2 protein lysine methyltransferase.** *FEBS Letters, 2016 Jun;590(12):1713-9. doi: 10.1002/1873-3468.12216. Epub 2016 May 30.*

Manuscript 3:

Kudithipudi S, **Schuhmacher MK**, Kebede AF, Jeltsch A. (2017). **The SUV39H1 Protein Lysine Methyltransferase Methylates Chromatin Proteins Involved in Heterochromatin Formation and VDJ Recombination.** *ACS Chemical Biology, 2017 Apr 21;12(4):958-968. doi: 10.1021/acschembio.6b01076. Epub 2017 Feb 16.*

Manuscript 4:

Schuhmacher MK, Kusevic D, Kudithipudi S, Jeltsch A. (2017). **Kinetic Analysis of the Inhibition of the NSD1, NSD2 and SETD2 Protein Lysine Methyltransferases by a K36M Oncohistone Peptide**. *Chemistry Select, 2017 Oct 18;2(29):9532-9536 doi: 10.1002/slct.201701940. Epub 2017 Oct 18.*

Manuscript 5:

Schuhmacher MK, Rolando M, Bröhm A, Weirich S, Kudithipudi S, Buchrieser C, Jeltsch A. (2018). **The *Legionella pneumophila* methyltransferase RomA methylates also non-histone proteins during infection.** *Journal of Molecular Biology, 2018 June 22;430(13):1912-1925 doi: 10.1016/j.jmb.2018.04.032. Epub 2018 May 04.*

Table S1: Sequences of the investigated non-histone targets of SETD2 shown in Figure 29. The non-histone substrates were identified by a Scansite search with the super-substrate specificity motif of SETD2. The target lysine is labeled in red, the lysine to alanine in blue and amino acids labeled in purple represents small differences to the used scan motif. The spot no., the protein name, the abbreviation (name), the Uniprot No., the target lysine position and the synthesized fifteen amino acid long peptide sequences are provided.

Spot No.	Protein name	Name	Uniprot No.	Target lysine	Sequence
A 1	Histone 3.1	H3K36	P68431	36	APATGGVKKPHRYRP
A 2	Histone 3.1-K36A	H3K36A		36	APATGGVAKPHRYRP
A 3	H3K36 super-substrate	super-substrate		36	APRFGGVKRPNRYRP
A 4	H3K36 super-substrate-A	super-substrate-A		36	APRFGGVARPNRYRP
A 5	ATP-binding cassette sub-family C member 8	ABCC8_HUMAN	Q09428	1312	ELQLGAVKRIHGLLK
A 6	Acetyl-CoA carboxylase 2	ACACB_HUMAN	O00763	823	ELIYGGVKYILKVAR
A 7	Neuronal acetylcholine receptor subunit alpha-5	ACHA5_HUMAN	P30532	119	PDDYGGIKVIRVPSD
A 8	Acetylcholine receptor subunit alpha	ACHA_HUMAN	P02708	121	PDDYGGVKKIHIPSE
A 9	Activated CDC42 kinase 1	ACK1_HUMAN	Q07912	514	PQHLGGVKKPTYDPV
A 10	A-kinase anchor protein 1, mitochondrial	AKAP1_HUMAN	Q92667	800	YVDYGGYKRVKVDVL
A 11	Retinal dehydrogenase 1	AL1A1_HUMAN	P00352	470	QCPFGGFKMSGNGRE
A 12	Retinal dehydrogenase 2	AL1A2_HUMAN	O94788	487	QSPFGGFKMSGNGRE
A 13	Aldehyde dehydrogenase family 1 member A3	AL1A3_HUMAN	P47895	481	QAPFGGFKMSGNGRE
A 14	Mitochondrial 10-formyltetrahydrofolate dehydrogenase	AL1L2_HUMAN	Q3SY69	897	AAPFGGVKQSGFGKD
A 15	4-trimethylaminobutyraldehyde dehydrogenase	AL9A1_HUMAN	P49189	461	ELPFGGYKKSGFGRE
A 16		AL9A1_HUMAN	P49189	344	ERVLGFVKVAKEQGA
A 17	Aldehyde dehydrogenase, mitochondrial	ALDH2_HUMAN	P05091	486	QSPFGGYKMSGSGRE
A 18	Ankyrin and armadillo repeat-containing protein	ANKAR_HUMAN	Q7Z5J8	317	RRGIGYLKLICFLIP
A 19	Progressive ankylosis protein homolog	ANKH_HUMAN	Q9HCJ1	214	TLCLGYYKNIHDIIP
A 20	Protein arginine N-methyltransferase 3	ANM3_HUMAN	O60678	91	LEFYGYIKLINFIRL
B 1	MICOS complex subunit MIC27	MIC27_HUMAN	Q6UXV4	180	QQIFGAVKSLWTKSS
B 2	Advillin	AVIL_HUMAN	O75366	109	DTFRGYFKQGIIYKQ
B 3	BAH and coiled-coil domain-containing protein 1	BAHC1_HUMAN	Q9P281	659	LVGLGGLKASCIQQE
B 4	Bromodomain adjacent to zinc finger domain protein 1A	BAZ1A_HUMAN	Q9NRL2	999	QGTLGAIKVTDRHIW
B 5	Betaine-homocysteine S-methyltransferase 1	BHMT1_HUMAN	Q93088	40	LEKRGYVKAGPWTPE
B 6	S-methylmethionine--homocysteine S-methyltransferase BHMT2	BHMT2_HUMAN	Q9H2M3	40	LEKRGYVKAGLWTPE

B 7	Ankyrin repeat and BTB/POZ domain-containing protein BTBD11	BTBDB_HUMAN	A6QL63	969	CIEIGYVKYSIFQLV
B 8	MHC class II transactivator	C2TA_HUMAN	P33076	1034	ITDLGAYKLAEALPS
B 9	SprT-like domain-containing protein Spartan	SPRTN_HUMAN	Q9H040	184	PPYYGYVKRATNREP
B 10	Voltage-dependent T-type calcium channel subunit alpha-1G	CAC1G_HUMAN	O43497	804	YGPFGYIKNPYNIFD
B 11	UPF0565 protein C2orf69	CB069_HUMAN	Q8N8R5	185	NTDFGAFKHLYMLLV
B 12	Cyclin-dependent kinase 3	CDK3_HUMAN	Q00526	142	INELGAIKLADFGLA
B 13	Carcinoembryonic antigen-related cell adhesion molecule 7	CEAM7_HUMAN	Q14002	84	YRIIGYVKNISQENA
B 14	Centromere protein W	CENPW_HUMAN	Q5EE01	23	KAPRGFLKRVFKRKK
B 15	Alpha-tubulin N-acetyltransferase 1	ATAT_HUMAN	Q5SQI0	98	GAIIGFIKVGYKKLF
B 16	UPF0553 protein C9orf64	CI064_HUMAN	Q5T6V5	253	LAHLGALKYSDDLLK
B 17	Cilia- and flagella-associated protein 77	CFA77_HUMAN	Q6ZQR2	165	AMNRGAVKAGLVTAR
B 18	Uncharacterized protein C10orf67, mitochondrial	CJ067_HUMAN	Q8IYJ2	123	QVDFGFLKQLLQLKF
B 19	Uncharacterized protein C11orf65	CK065_HUMAN	Q8NCR3	78	RFRLGGVKFPPDIYY
B 20	Transmembrane protein 260	TM260_HUMAN	Q9NX78	692	ADILGALKHLRKELQ
C 1	Collagen alpha-3(VI) chain	CO6A3_HUMAN	P12111	310	AQVLGAVKALGFAGG
C 2	Collagen alpha-5(VI) chain	CO6A5_HUMAN	A8TX70	857	RVQFGALKYSDQPNI
C 3	Collagen alpha-6(VI) chain	CO6A6_HUMAN	A6NMZ7	852	QVRFGALKYADDPEV
C 4	Collagen alpha-1(XIV) chain	COEA1_HUMAN	Q05707	61	RGKFGGYKLLVTPTS
C 5	Conserved oligomeric Golgi complex subunit 2	COG2_HUMAN	Q14746	584	DSCFGFLKSALEVPR
C 6	Collagen alpha-1(XXII) chain	COMA1_HUMAN	Q8NFW1	472	SEQIGFLKTINCSCP
C 7	Protein FAM189B	F189B_HUMAN	P81408	171	DEARGALKNLLFSVC
C 8	Cytochrome P450 4B1	CP4B1_HUMAN	P13584	28	ILVLGFLKLIHLLLR
C 9	Uncharacterized protein KIAA0930	K0930_HUMAN	Q6ICG6	236	KMSFGFYKYSNMEFV
C 10	ATP-dependent RNA helicase DDX54	DDX54_HUMAN	Q8TDD1	848	FLQRGGLKQLSARNR
C 11	Deleted in lung and esophageal cancer protein 1	DLEC1_HUMAN	Q9Y238	494	YCLIGGVKMTRFICK
C 12	Dual specificity phosphatase DUPD1	DUPD1_HUMAN	Q68J44	191	LPNRGFLKQLRELDK
C 13	Integrator complex subunit 6-like	INTL6_HUMAN	Q5JSJ4	373	GYPFGYLKASTTLTC
C 14	Dynein heavy chain 1, axonemal	DYH1_HUMAN	Q9P2D7	2332	RKIIGAFKNLVDINF
C 15	Dysferlin	DYSF_HUMAN	O75923	338	AGARGYLKTSLCVLG
C 16	Endonuclease/exonuclease/phosphatase family domain-containing protein 1	EEPD1_HUMAN	Q7L9B9	75	REYIGGFKKVEDLAL
C 17	Ephrin type-A receptor 4	EPHA4_HUMAN	P54764	282	ACKIGYYKALSTDAT

C 18	Ephrin type-A receptor 7	EPHA7_HUMAN	Q15375	285	PCGRGFYKSSSQDLQ
C 19	Ephrin type-A receptor 8	EPHA8_HUMAN	P29322	282	ACELGFYKSAPGDQL
C 20	FAST kinase domain-containing protein 5, mitochondrial	FAKD5_HUMAN	Q7L8L6	327	TICLGFFKSSTNLSE
D 1	Fibrillin-1	FBN1_HUMAN	P35555	666	STCYGGYKRGQCIKP
D 2	Fibrillin-2	FBN2_HUMAN	P35556	711	STCYGGIKKGVCVRP
D 3	Fidgetin-like protein 1	FIGL1_HUMAN	Q6PIW4	325	GSSYGGVKKSLGASR
D 4	Cytosolic 10-formyltetrahydrofolate dehydrogenase	AL1L1_HUMAN	O75891	876	AAPFGGFKQSGFGKD
D 5	FYN-binding protein	FYB_HUMAN	O15117	563	RGSYGYIKTTAVEID
D 6	Endogenous retrovirus group K member 10 Gag polyprotein	GAK10_HUMAN	P87889	29	LLKRGGVKVSTKNLI
D 7	Endogenous retrovirus group K member 5 Gag polyprotein	GAK5_HUMAN	Q9HDB9	555	CGQIGHLKRSCPVLN
D 8	Polypeptide N-acetylgalactosaminyltransferase 6	GALT6_HUMAN	Q8NCL4	502	PTFYGAIKNLGTNQC
D 9	Growth arrest-specific protein 6	GAS6_HUMAN	Q14393	269	CDGRGGLKLSQDMDT
D 10	eIF-2-alpha kinase activator GCN1	GCN1L_HUMAN	Q92616	599	LSSLGGFKLAHGLLE
D 11	Gelsolin	GELS_HUMAN	P06396	162	ATFLGYFKSGLKYKK
D 12	Glucosamine 6-phosphate N-acetyltransferase	GNA1_HUMAN	Q96EK6	55	DLNRGFFKVLGQLTE
D 13	Solute carrier family 2, facilitated glucose transporter member 10	GTR10_HUMAN	O95528	280	SVGLGAVKVAATLTA
D 14	Histone H4	H4_HUMAN	P62805	45	LARRGGVKRISGLIY
D 15	Hyaluronan-binding protein 2	HABP2_HUMAN	Q14520	319	KRIYGGFKSTAGKHP
D 16	Histone deacetylase 9	HDAC9_HUMAN	Q9UKV0	927	TPPLGGYKVTAKCFG
D 17	Hemicentin-1	HMCN1_HUMAN	Q96RW7	127	EMSIGAIKIALEISL
D 18	Hemicentin-1	HMCN1_HUMAN	Q96RW7	5253	KNTRGGYKCIDLCPN
D 19	Insulin receptor-related protein	INSRR_HUMAN	P14616	397	LVSLGFFKNLKLIRG
D 20	Integrator complex subunit 6	INT6_HUMAN	Q9UL03	369	GHPFGYLKASTALNC
E 1	Integrator complex subunit 6	INT6_HUMAN	Q9UL03	428	PYYLGPLKKAVRMMG
E 2	IQ domain-containing protein G	IQCG_HUMAN	Q9H095	417	RREIGGFKMPKDKVD
E 3	Integrin alpha-1	ITA1_HUMAN	P56199	1170	LWKIGFFKRPLKKKM
E 4	Integrin alpha-5	ITA5_HUMAN	P08648	1027	LYKLGFFKRSLPYGT
E 5	Integrin alpha-9	ITA9_HUMAN	Q13797	653	YLALGAVKNISLNIS
E 6	Tectonin beta-propeller repeat-containing protein 2	TCPR2_HUMAN	O15040	1217	LSQLGAVKLTSLACG
E 7	TELO2-interacting protein 1 homolog	TTI1_HUMAN	O43156	409	SLLLGYLKLLGPKIN
E 8	Little elongation complex subunit 1	ICE1_HUMAN	Q9Y2F5	787	KSGLGFVKSTSWHHS
E 9	Protein kinase C delta type	KPCD_HUMAN	Q05655	149	MNRRGAIKQAKIHYI

E 10	Protein kinase C theta type	KPCT_HUMAN	Q04759	150	HQRRGAIKQAKVHHV
E 11	Pyruvate kinase PKM	KPYM_HUMAN	P14618	504	GKARGFFKKGDVVIV
E 12	LIM and senescent cell antigen-like-containing domain protein 1	LIMS1_HUMAN	P48059	111	LADIGFVKNAGRHLC
E 13	LIM and senescent cell antigen-like-containing domain protein 2	LIMS2_HUMAN	Q7Z4I7	116	LADLGFVKNAGRHLC
E 14	Leiomodin-1	LMOD1_HUMAN	P29536	126	EPKRGGLKKSFSRDR
E 15	Protein phosphatase 1 regulatory subunit 37	PPR37_HUMAN	O75864	81	DEVIGAYKQACQKLN
E 16	Leucine-rich repeats and immunoglobulin-like domains protein 3	LRIG3_HUMAN	Q6UXM1	242	FQGLGALKSLKMQRN
E 17	Low-density lipoprotein receptor-related protein 2	LRP2_HUMAN	P98164	4116	GSRFGAIKRAYIPNF
E 18	MAM domain-containing protein 2	MAMC2_HUMAN	Q7Z304	423	YAIYGFLKMSDTLAV
E 19	Multiple epidermal growth factor-like domains protein 8	MEGF8_HUMAN	Q7Z7M0	941	GRGRGALKSPEECPP
E 20	Mitoferrin-1	MFRN1_HUMAN	Q9NYZ2	92	TSIYGALKKIMRTEG
F 1	Putative helicase MOV-10	MOV10_HUMAN	Q9HCE1	900	DFNLGFLKNPKRFNV
F 2	Metastasis-associated protein MTA1	MTA1_HUMAN	Q13330	431	WKKYGGLKMPTRLDG
F 3	Metastasis-associated protein MTA2	MTA2_HUMAN	O94776	405	WKKYGGLKTPTQLEG
F 4	Metastasis-associated protein MTA3	MTA3_HUMAN	Q9BTC8	417	WKKYGGLKMPTQSEE
F 5	Myocardin	MYCD_HUMAN	Q8IZQ8	644	HSPLGAVKSPQHISL
F 6	N-acetyl-D-glucosamine kinase	NAGK_HUMAN	Q9UJ70	185	PHDIGYVKQAMFHYF
F 7	Neuroblastoma-amplified sequence	NBAS_HUMAN	A2RRP1	490	ARYFGYIKQGLYLVT
F 8	Nuclear factor of activated T-cells, cytoplasmic 1	NFAC1_HUMAN	O95644	452	EGSRGAVKASAGGHP
F 9	Nuclear factor of activated T-cells, cytoplasmic 2	NFAC2_HUMAN	Q13469	434	EGSRGAVKAPTGGHP
F 10	Nuclear factor of activated T-cells, cytoplasmic 3	NFAC3_HUMAN	Q12968	457	EGSRGAVKASTGGHP
F 11	Nuclear factor of activated T-cells, cytoplasmic 4	NFAC4_HUMAN	Q14934	443	EGSRGAVKAAPGGHP
F 12	Nucleotide-binding oligomerization domain-containing protein 1	NOD1_HUMAN	Q9Y239	633	SSLRGYLKSLPRVQV
F 13	Olfactory receptor 2T11	O2T11_HUMAN	Q8NH01	299	KDVIGAFKKVFACCS
F 14	Obscurin	OBSCN_HUMAN	Q5VST9	6483	RGVFGFVKRVQHKGN
F 15	Olfactory receptor 2J1	OR2J1_HUMAN	Q9GZK6	303	KDVRGAVKRLMGWEW
F 16	Olfactory receptor 2J3	OR2J3_HUMAN	O76001	305	NKVVRGAVKRLMGWE
F 17	Tumor protein p73	P73_HUMAN	O15350	532	IEDLGALKIPEQYRM
F 18	Procollagen C-endopeptidase enhancer 1	PCOC1_HUMAN	Q15113	365	VSLIGAYKTGGLDLP
F 19	Pecanex-like protein 3	PCX3_HUMAN	Q9H6A9	973	GFCLGAIKTPWPEQH

F 20	High affinity cGMP-specific 3',5'-cyclic phosphodiesterase 9A	PDE9A_HUMAN	O76083	519	TAQIGFIKFVLIPMF
G 1	PDZ domain-containing protein 8	PDZD8_HUMAN	Q8NEN9	423	LIAIGGVKITSTLQV
G 2	Phosphate-regulating neutral endopeptidase	PHEX_HUMAN	P78562	319	FDWLGYIKKVIDTRL
G 3	Piwi-like protein 2	PIWL2_HUMAN	Q8TC59	685	DDLYGAIKKLCCVQS
G 4	Plakophilin-4	PKP4_HUMAN	Q99569	573	VCRLGGIKHLVDLLD
G 5	Putative PRAME family member 13	PRA13_HUMAN	Q5VWM6	323	YPSLGYLKHLNLSYV
G 6	PRAME family member 2	PRAM2_HUMAN	O60811	323	FPSLGYLKHLNLSYV
G 7	Receptor-type tyrosine-protein phosphatase gamma	PTPRG_HUMAN	P23470	1093	VNVLGFLKHIRTQRN
G 8	Receptor-type tyrosine-protein phosphatase zeta	PTPRZ_HUMAN	P23471	1966	VNIFGFLKHIRSQRN
G 9	DNA repair and recombination protein RAD54B	RA54B_HUMAN	Q9Y620	589	LICIGALKKLCNHPC
G 10	Ras-related C3 botulinum toxin substrate 1	RAC1_HUMAN	P63000	153	AKEIGAVKYLECSAL
G 11	E3 ubiquitin-protein ligase RBBP6	RBBP6_HUMAN	Q7Z6E9	81	RIPIGGVKSTSKTYV
G 12	Rho GTPase-activating protein 1	RHG01_HUMAN	Q07960	109	SKLLGYLKHTLDQYV
G 13	Rho GTPase-activating protein 6	RHG06_HUMAN	O43182	132	PLGRGGLKKSMAWDL
G 14	RING finger protein 214	RN214_HUMAN	Q8ND24	541	PPGLGGVKASAETPR
G 15	Oxygen-regulated protein 1	RP1_HUMAN	P56715	1629	EYNIGFVKRAIEKLY
G 16	RRP12-like protein	RRP12_HUMAN	Q5JTH9	977	KSALGFIKVAVTVMD
G 17	Protein RRP5 homolog	RRP5_HUMAN	Q14690	737	MLLIGFVKSIKDYGV
G 18	Sodium-dependent multivitamin transporter	SC5A6_HUMAN	Q9Y289	197	YTALGGLKAVIWTDV
G 19	Sodium-coupled monocarboxylate transporter 1	SC5A8_HUMAN	Q8N695	183	YCTLGGLKAVIWTDV
G 20	Sodium-coupled monocarboxylate transporter 2	SC5AC_HUMAN	Q1EHB4	179	YCTLGGLKAVVWTDA
H 1	Sodium channel protein type 1 subunit alpha	SCN1A_HUMAN	P35498	1313	YSELGAIKSLRTLRA
H 2	Sodium channel protein type 4 subunit alpha	SCN4A_HUMAN	P35499	126	VVRRGAIKVLIHALF
H 3	Sodium channel protein type 8 subunit alpha	SCN8A_HUMAN	Q9UQD0	1258	WTAYGFVKFFTNAWC
H 4	Splicing factor 3B subunit 1	SF3B1_HUMAN	O75533	998	GSILGALKAIVNVIG
H 5	Epsilon-sarcoglycan	SGCE_HUMAN	O43556	188	GDFLGAVKNVWQPER
H 6	CMP-N-acetylneuraminate-poly-alpha-2,8-sialyltransferase	SIA8D_HUMAN	Q92187	350	LHNRGALKLTTGKCV
H 7	Succinate-semialdehyde dehydrogenase, mitochondrial	SSDH_HUMAN	P51649	508	ECPFGGVKQSGLGRE

H 8	Symplekin	SYMPK_HUMAN	Q92797	567	AMKLGAVKRILRAEK
H 9	Synaptophysin	SYPH_HUMAN	P08247	30	KEPLGFVKVLQWVFA
H 10	Synaptophysin-like protein 1	SYPL1_HUMAN	Q16563	37	KEPLGFIKVLEWIAS
H 11	Synaptophysin-like protein 2	SYPL2_HUMAN	Q5VXT5	39	EEPLGFIKVLQWLFA
H 12	Threonine-tRNA ligase, mitochondrial	SYTM_HUMAN	Q9BW92	255	TGQIGGLKLLSNSSS
H 13	Taste receptor type 2 member 40	T2R40_HUMAN	P59535	247	KAHIGAIKATSYFLI
H 14	Putative ATP-dependent RNA helicase TDRD12	TDR12_HUMAN	Q587J7	884	LPSFGYIKIIPFYIL
H 15	Transmembrane protein 14C	TM14C_HUMAN	Q9P0S9	31	GGIIGYVKAGSVPSL
H 16	Tumor necrosis factor ligand superfamily member 13B	TN13B_HUMAN	Q9Y275	283	SLDGDVTFFGALKLL
H 17	Tumor protein D55	TPD55_HUMAN	Q96J77	112	CRKLGGVKKSATFRS
H 18	tRNA selenocysteine 1-associated protein 1	TSAP1_HUMAN	Q9NX07	143	SKGYGFVKFTDELEQ
H 19	UDP-N-acetylhexosamine pyrophosphorylase-like protein 1	UAP1L_HUMAN	Q3KQV9	342	FFTRGFLKAVTREFE
H 20	Ubiquitin-like modifier-activating enzyme 1	UBA1_HUMAN	P22314	97	NIILGGVKAVTLHDQ
I 1	Ubiquitin-like protein 3	UBL3_HUMAN	O95164	77	NVTLGALKLPFGKTT
I 2	Ubiquitin carboxyl-terminal hydrolase 40	UBP40_HUMAN	Q9NVE5	933	LPPLGFLKVPIWWYQ
I 3	Ubiquitin carboxyl-terminal hydrolase 47	UBP47_HUMAN	Q96K76	688	GLLLGGVKSTYMFDL
I 4	Villin-1	VILI_HUMAN	P09327	112	EAFRGYFKQGLVIRK
I 5	Serine/threonine-protein kinase WNK3	WNK3_HUMAN	Q9BYP7	159	ELGRGAFKTVYKGLD
I 6	Zinc finger CCCH domain-containing protein 7B	Z3H7B_HUMAN	Q9UGR2	549	FDPLGGVKRGSLTIA
I 7	Palmitoyltransferase ZDHHC15	ZDH15_HUMAN	Q96MV8	268	GFNLGFIKNIQQVFG
I 8	Palmitoyltransferase ZDHHC5	ZDHC5_HUMAN	Q9C0B5	701	SPTRGGVKKVSGVGG
I 9	Histone 3.1	H3K36	P68431	36	APTAGGVKKPHRYRP
I 10	Histone 3.1-K36A	H3K36A		36	APATGGVAKPHRYRP
I 11	H3K36 super-substrate	super-substrate		36	APRFGGVKRPNRYRP
I 12	H3K36 super-substrate-A	super-substrate-A		36	APRFGGVARPNRYRP

Table S2: Sequences of the investigated non-histone targets of SETD2 shown in Figure 30. The target lysine is labeled in red and its alanine mutant in blue. The spot no., the protein name, the abbreviation (name), the Uniprot No., the target lysine position and the synthesized fifteen amino acid long peptide sequences are provided.

Spot No.	Protein name	Name	Uniprot No.	Target lysine	Sequence
A 1	Histone H3.1	H3K36	P68431	36	APATGGVKKPHRYRP
A 2					APATGGVAKPHRYRP
A 3	H3K36 super-substrate	super-substrate		36	APRFGGVKRPNRYRP
A 4					APRFGGVARPNRYRP
A 5	Ankyrin and armadillo repeat-containing protein	ANKAR_HUMAN	Q7Z5J8	317	RRGIGYLKLICFLIP
A 6					RRGIGYLALICFLIP
A 7	Voltage-dependent T-type calcium channel subunit alpha-1G	CAC1G_HUMAN	O43497	804	YGPFGYIKNPYNIFD
A 8					YGPFGYIANPYNIFD
A 9	Centromere protein W	CENPW_HUMAN	Q5EE01	23	KAPRGFLKRVFKRKK
A 10					KAPRGFLARVFKRKK
A 11	Alpha-tubulin N-acetyltransferase 1	ATAT_HUMAN	Q5SQI0	98	GAIIGFIKVGYKKLF
A 12					GAIIGFIAVGYKKLF
A 13	Collagen alpha-1(XXII) chain	COMA1_HUMAN	Q8NFW1	472	SEQIGFLKTINCSCP
A 14					SEQIGFLATINCSCP
A 15	Deleted in lung and esophageal cancer protein 1	DLEC1_HUMAN	Q9Y238	494	YCLIGGVKMTRFICK
A 16					YCLIGGVAMTRFICK
A 17	Dysferlin	DYSF_HUMAN	O75923	338	AGARGYLKTSLCVLG
A 18					AGARGYLATSLCVLG
A 19	Fibrillin-1	FBN1_HUMAN	P35555	666	STCYGGYKRGQCIKP
A 20					STCYGGYARGQCIKP
B 1	Fibrillin-2	FBN2_HUMAN	P35556	711	STCYGGIKKGVCVRP
B 2					STCYGGIAKGVCVRP
B 3	Gelsolin	GELS_HUMAN	P06396	162	ATFLGYFKSGLKYKK
B 4					ATFLGYFASGLKYKK
B 5	Hyaluronan-binding protein 2	HABP2_HUMAN	Q14520	319	KRIYGGFKSTAGKHP
B 6					KRIYGGFASTAGKHP
B 7	Histone deacetylase 9	HDAC9_HUMAN	Q9UKV0	927	TPPLGGYKVTAKCFG
B 8					TPPLGGYAVTAKCFG
B 9	Hemicentin-1	HMCN1_HUMAN	Q96RW7	127	EMSIGAIKIALEISL

B 10					EMSIGAIAIALEISL
B 11	Integrator complex subunit 6	INT6_HUMAN	Q9UL03	369	GHPFGYLKASTALNC
B 12					GHPFGYLAASTALNC
B 13	Integrin alpha-1	ITA1_HUMAN	P56199	1170	LWKIGFFKRPLKKKM
B 14					LWKIGFFARPLKKKM
B 15	Olfactory receptor 2T11	O2T11_HUMAN	Q8NH01	299	KDVIGAFKKVFACCS
B 16					KDVIGAFAKVFACCS
B 17	Piwi-like protein 2	PIWL2_HUMAN	Q8TC59	685	DDLYGAIKKLCCVQS
B 18					DDLYGAIAKLCCVQS
B 19	DNA repair and recombination protein RAD54B	RA54B_HUMAN	Q9Y620	589	LICIGALKKLCNHPC
B 20					LICIGALAKLCNHPC
C 1	CMP-N-acetylneuraminate-poly-alpha-2,8-sialyltransferase	SIA8D_HUMAN	Q92187	350	LHNRGALKLTTGKCV
C 2					LHNRGALALTTGKCV
C 3	Tumor protein D55	TPD55_HUMAN	Q96J77	112	CRKLGGVKKSATFRS
C 4					CRKLGGVAKSATFRS
C 5	Histone H3.1	H3K36	P68431	36	APATGGVKKPHRYRP
C 6					APATGGVAKPHRYRP
C 7	H3K36 super-substrate	super-substrate		36	APRFGGVKRPNRYRP
C 8					APRFGGVARPNRYRP