Biochemical Investigation of the Substrate Specificity of Protein Methyltransferases and the Identification of Novel Substrates
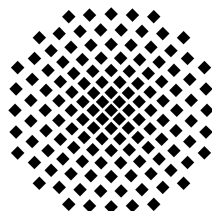
Von der Fakultät 3: Chemie der Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Denis Kušević

geboren am 02.05.1985 in Bietigheim-Bissingen

Hauptberichter: Prof. Dr. Albert Jeltsch
Mitberichter: Prof. Dr. Bernhard Hauer
Tag der mündlichen Prüfung: 12.12.2016

Institut für Biochemie der Universität Stuttgart

2016

Erklärung über die Eigenständigkeit der Dissertation

Ich versichere hiermit, dass ich die vorliegende Arbeit mit dem Titel

Biochemical Investigation of the Substrate Specificity of Protein Methyltransferases and the Identification of Novel Substrates.

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 Investigation of the Substrate Specificity of Protein Methyltransferases and the Identification of Novel Substrates.

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

Denis Kušević

Stuttgart, den 12. Dezember 2016

II

Acknowledgements

IV

List of Publications

D. Kušević, S. Kudithipudi, A. Jeltsch. **Substrate Specificity of the HEMK2 Protein Glutamine Methyltransferase and Identification of Novel Substrates.** *Journal of Biological Chemistry*, vol. 291, no. 12, pp. 6124 - 6133, (2016).

S. Weirich, **D. Kušević**, S. Kudithipudi, A. Jeltsch. **Investigation of the methylation of Numb by the SET8 protein lysine methyltransferase.** *Scientific reports*, vol. 22, no. 5, (2015).

M. K. Schuhmacher, S. Kudithipudi, **D. Kušević**, S. Weirich, A. Jeltsch. **Activity and specificity of the human SUV39H2 protein lysine methyltransferase.** *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1849, no. 1, pp. 55 - 63, (2015).

S. Kudithipudi, **D. Kušević**, S. Weirich, A. Jeltsch. **Specificity analysis of protein lysine methyltransferases using SPOT peptide arrays.** *JoVE (Journal of Visualized Experiments)*, no. 93, e52203, (2014).

S. Kudithipudi, **D. Kušević**, A. Jeltsch. **Non-radioactive protein lysine methyltransferase microplate assay based on reading domains.** *ChemMedChem*, vol. 9, no. 3, pp. 554 - 559, (2014).

Contents

Contents

x

Zusammenfassung

Posttranslationale Proteinmodifikationen (PTMs) sind wichtig, um verschiedene Proteinfunktionen, wie z. B. Lokalisation, Aktivität, Stabilität und Protein-Protein Interaktionen zu regulieren. In Proteinen können viele Aminosäuren methyliert werden, darunter auch Lysin, Arginin und Glutamin. Methylierungen sind auf vielen verschieden Protein zu finden, jedoch sind Histonproteine die bedeutendsten. Die Histonmethylierung beeinflusst die Chromatinstrukur und spielt eine große Rolle in der Regulation der Transkription. Die Enzyme, die für den Transfer von Methylgruppen auf die Proteine zuständig sind, werden Protein Methyltransferasen (PMTs) genannt. Sie sind sehr spezifisch und methylieren immer nur eine Art von Aminosäuren. Dabei zeigt die schnell steigende Anzahl an Berichten über die Methylierung von Proteinen, dass die Methylierung als posttranslationale Modifikation in den letzten Jahren immer mehr an Bedeutung gewinnt.

In dieser Doktorarbeit wurde die Substratspezifität dreier unterschiedlicher Protein Methyltransferasen untersucht, und zwar von HEMK2, einer Glutamin Methyltransferase, sowie von NSD2 und Clr4, zwei Protein Lysin Methyltransferasen (PKMTs).

Die Glutamin Methyltransferase HEMK2 methyliert Q185 des Terminationsfaktors eRF1 (eukaryotic translation release factor 1), der für die Termination der Peptidsynthese und für die Hydrolyse der Polypeptidkette von der tRNA am Ribosom verantwortlich ist. Zur Bestimmung der Substratspezifität von HEMK2 wurde die Aminosäuresequenz von eRF1 als Vorlage verwendet und die erhaltenen Daten zeigen, dass das Substrat für eine Methylierung ein G-Q-X$_3$-K Sequenzmotiv besitzen muss. Eine Suche nach dieser Sequenz in einer Proteindatenbank ergab, dass mehrere humane Proteine dieses Sequenzmotiv besitzen. Von diesen identifizierten Substratkandidaten wurden 125 von HEMK2 auf Peptidebene methyliert. Außerdem konnte gezeigt werden, dass von diesen 125 Kandidaten 16 auf Proteinebene methyliert werden. Zuletzt wurde eine Methylierung der „Chromodomain helicase DNA binding protein 5" (CHD5) und „Nuclear protein in Testis" (NUT) Proteine mit Hilfe eines glutaminspezifischen Antikörpers in menschlichen HEK293 Zellen nachgewiesen.

NSD2 ist ein Mitglied der „nuclear receptor SET domain-containing" Enzymfamilie und dimethyliert Lysin K36 von Histon H3 und Lysin K44 von Histon H4. Es wurde gezeigt, dass eine abnormale Expression von NSD2 zu verschiedenen Arten von Krebs und dem Wolf-Hirschhorn Syndrom führen kann. Die Analyse der Substratspezifität von NSD2 zeigte, dass dieses Enzym die Aminosäuren G33 bis P38 von H3 erkennt. Dabei werden hydrophobe Aminosäuren an den Positionen -1 und +2 (das Ziellysin wird hierbei als Position 0 definiert) bevorzugt. Mit Hilfe des Spezifitätsprofils von NSD2 wurden mehrere humane Proteine identifiziert, die dieses Sequenzmotiv enthalten. Von diesen identifizierten Substratkandidaten wurden 45 durch NSD2

auf Peptidebene methyliert. Des Weiteren wurde gezeigt, dass 3 Kandidaten (ATRX, FANCM und SET8) auf Proteinebene methyliert wurden und zusätzlich konnte die Methylierung von ATRX und FANCM durch NSD2 in HEK293 Zellen nachgewiesen werden. Da die Methylierungen einen erheblichen Einfluss auf die Eigenschaften und Funktionen von Proteinen besitzen, müssen weitere Experimente an den neuen Substraten von HEMK2 (CHD5 und NUT) und NSD2 (ATRX und FANCM) durchgeführt werden, um die Auswirkungen auf die biologischen Funktionen der Methylierung herauszufinden.

Abgesehen von den menschlichen Enzymen, wurden ähnliche Untersuchungen auch an der Histon Lysin Methyltransferase Clr4, einem SUV39H1-Homolog aus *S. pombe*, durchgeführt. Clr4 trimethyliert Lysin K9 des Histonproteins H3. Zur Bestimmung des Spezifitätsprofils von Clr4 wurde die Aminosäuresequenz von H3 (1 - 18) verwendet. Die Ergebnisse zeigten, dass Clr4 spezifisch die Aminosäuren der Positionen -1 bis +3 der Zielsequenz erkennt. Zusätzlich wurden 6 neue Peptidsubstrate aus *S. pombe* identifizieren, die durch Clr4 methyliert wurden.

Um die Detektion von Proteinmethylierungen weiter zu verbessern, wurde eine neue radioaktivitätsfreie, Mikrotiter-Untersuchungsmethode entwickelt, die natürlich vorkommende Lese-Domänen anstelle von methylspezifischen Antikörpern zur Erkennung von Methylierungen auf Histonpeptiden verwendet. Es wurde gezeigt, dass diese Methode erfolgreich die Methyltransferaseaktivität bestimmen und für die Suche nach PKMT Inhibitoren verwendet werden kann.

Abstract

Posttranslational modifications (PTMs) are crucial for the regulation of protein properties, such as localization, activity, stability and protein-protein interactions. One important PTM is protein methylation. This occurs on various amino acids, most frequently at lysine and arginine but also glutamine. Methylation was found on many proteins, though the most prominent group is constituted out of histone proteins. Histone methylation influences the chromatin structure and plays an important role in transcriptional regulation. The enzymes responsible for the transfer of methyl groups are called protein methyltransferases (PMTs) and they are very specific toward the methylated substrate. The rapidly increasing number of reports about protein methylation illustrates that this modification is very frequent and has important roles in various cellular signaling pathways.

In this doctoral thesis, the substrate specificity of three different protein methyltransferases, namely the glutamine methyltransferase HEMK2 and the two protein lysine methyltransferases (PKMTs), NSD2 and Clr4 were investigated. The glutamine methyltransferase HEMK2 has been shown to methylate Q185 of the eukaryotic translation release factor eRF1, which is responsible for termination of peptide synthesis and hydrolysis of the nascent polypeptide from the tRNA at the ribosome. The substrate specificity profile of HEMK2 was determined using the eRF1 sequence as template, the data showed that HEMK2 requires a G-Q-X_3-K motif for methylation activity. Based on the obtained substrate specificity profile, several human proteins containing the corresponding sequence motif were identified and methylation at the peptide level was shown for 125 substrates. Furthermore, the *in vitro* methylation of 16 substrates at the protein level was confirmed. Finally, the cellular methylation could be demonstrated for Chromodomain helicase DNA binding protein 5 (CHD5) and Nuclear protein in Testis (NUT), by using a Qme-specific antibody.

NSD2, a member of the nuclear receptor SET domain-containing enzyme family, was shown to dimethylate K36 of histone H3 and K44 of histone H4. The aberrant expression of NSD2 was reported to be associated with several cancers and the Wolf-Hirschhorn syndrome (WHS). The substrate specificity analysis of NSD2, revealed that the enzyme recognizes the residues between G33 and P38, on the H3 tail. NSD2 prefers hydrophobic residues at the positions -1 and +2, considering the target lysine as position 0. Several human proteins containing the sequence motif of NSD2 were identified and methylation on 45 novel non-histone peptide substrates was observed. For 3 of the substrates (ATRX, FANCM and SET8) methylation could be confirmed at protein level. In addition, the methylation of ATRX and FANCM could be shown in HEK293 cells, upon ectopic expression of NSD2. Since methylation can strongly influence protein properties, further experiments have to be carried out to uncover the biological effects of the novel substrates of HEMK2 (CHD5 and NUT protein) and NSD2 (ATRX and FANCM).

Apart from the human enzymes, similar studies were performed for the histone lysine methyltransferase Clr4, the yeast homolog of the human SUV39H1, which trimethylates K9 of histone H3. The specificity profile of Clr4 was investigated using the H3 (1 - 18) sequence as template. The analysis revealed that the enzyme specifically recognizes the residues from -1 to +3 of the H3 tail. Additionally, it was shown that Clr4 is able to methylate 6 novel *S. pombe* substrate candidates at peptide level.

To facilitate the detection of protein methylation, a new radioactivity free, microplate assay was developed, which employs a natural reading domain instead of methyl specific antibodies for the recognition of methylation on histone peptides. It was demonstrated that this approach can be successfully used to determine the activity of PKMTs as well as screen for PKMT inhibitors in medium or high throughput scale.

List of Abbreviations

53BP1	Tumor suppressor p53-binding protein 1
A-site	Aminoacyl site
aa	Amino acid
ADD	ATRX-DNMT-DNMT3L
ADMA	asymmetric dimethylarginine
ADP	Adenosine diphosphate
AML	Acute myeloid leukemia
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
AWS	Associated with SET domain
BLBC	Basal-like breast cancer
cAMP	Cyclic adenosine monophosphate
CARM1	Histone arginine methyltransferase CARM1
CbiF	Cobalt-precorrin-4 C(11) methyltransferase
CBP/p300	CREB-binding protein/p300 Histone acetyltransferase complex
cDNA	Complementary deoxyribonucleic acid
CHD1	Chromodomain helicase DNA-binding protein 1
CHD5	Chromodomain helicase DNA-binding protein 5
Chromo	**Chr**omatin **o**rganization **mo**difier
Clr4	Histone-lysine N-methyltransferase, H3 lysine-9 specific
ClrC	Clr4 methyltransferase multiprotein complex
CML	Chronic myelogenous leukemia
COSMIC	Catalogue of somatic mutations in cancer
Cul4	Cullin-4
D. melanogaster	Drosophila melanogaster
DNA	Deoxyribonucleic acid
DNMT	DNA Methyltransferase
E. coli	Escherichia coli
EMT	Epithelial-mesenchymal transition
eRF1	Eukaryotic release factor 1
EZH2	Enhancer of Zeste 2, Histone-lysine N-methyltransferase
FBXL11	Lysine-specific demethylase 2A
FGFR3	Fibroblast growth factor receptor 3
G9a	Histone-lysine N-methyltransferase
GDP	Guanosine diphosphate

GST	Glutathione S-transferase
GTP	Guanosine triphosphate
H1	Histone 1
H2A	Histone 2A
H2B	Histone 2B
H3	Histone 3
H4	Histone 4
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HemK	HEMK glutamine methyltransferase family member
HMG	High mobility group
HOX	Homeobox gene
HP1	Heterochromatin protein 1
HT29	Human colorectal adenocarcinoma cell line
IgH	Immunoglobulin heavy chain
IL-6	Interleukin-6
IRX3	Iroquois-class homeodomain protein IRX-3
KDM	Lysine demethylase
MBT	Malignant brain tumor
me1	Monomethylation
me2	Dimethylation
me3	Trimethylation
MetH	Methionine synthase
Mlo3	mRNA export protein Mlo3
MM	Multiple myeloma
MMA	Monomethylarginine
MMSET	Multiple myeloma SET domain-containing protein
MPP8	M-phase phosphoprotein 8
mRNA	Messenger ribonucleic acid
MTase	Methyltransferase
Mtq2	N^5-glutamine methyltransferase MTQ2
Mut	Mutation
N6AMT1	N^6-adenine-specific DNA methyltransferase 1
NAT	N-terminal acetyltransferase
NF-κB	Nuclear factor of kappa light polypeptide gene enhancer in B-cells
Ni-NTA	Nickel-Nitrilotriacetic acid
Nkx2-5	Homeobox protein Nkx-2.5

NSD	Nuclear receptor SET domain-containing protein
Nu	Nucleophile
NUP98	Nuclear pore complex protein 98
OD$_{600}$	Optical density at 600 nm
P-site	Peptidyl site
p53	Cellular tumor antigen p53
PEV	Position effect variegation
PHD	Plant homeodomain
PKA	Protein kinase A
PKMT	Protein lysine methyltransferase
PrmC	**Pr**otein **m**ethyltransferase **C**
PRMT	Protein arginine methyltransferase
PTM	posttranslational modification
PWWP	Proline-tryptophan-tryptophan-proline motif containing domain
Raf1	RAF proto-oncogene serine/threonine-protein kinase
Raf2	Rik1-associated factor 2
RF	Release factor
Rik1	Chromatin modification-related protein Rik1
RNA	Ribonucleic acid
S. cerevisiae	Saccharomyces cerevisiae
S. pombe	Schizosaccharomyces pombe
S-phase	Synthesis phase
S$_N$2	Bimolecular nucleophilic substitution
SAH	S-Adenosyl-L-homocysteine
SAM	S-Adenosyl-L-methionine
SDMA	Symmetric dimethylarginine
SET	**Su**(var)3-9, **E**nhancer of Zeste and **T**rithorax
SET7/9	SET domain-containing protein 7/9, Histone-lysine N-methyltransferase
SET8	Histone-lysine N-methyltransferase
SFRP1	Secreted frizzled-related protein 1
SMN	Survival motor neuron
SMYD2	SET and MYND domain-containing protein 2, Histone-lysine N-methyltransferase
Su(var)3-9	Suppressor of variegation 3-9, Histone-lysine N-methyltransferase
SUV39H1	Human suppressor of variegation 3-9 homolog 1, Histone-lysine N-methyltransferase
SUV39H2	Human suppressor of variegation 3-9 homolog 2, Histone-lysine N-methyltransferase

Swi6	Chromatin-associated protein Swi6
TBL1X	F-box-like/WD repeat-containing protein TBL1X
TNF-α	Tumor necrosis factor
TRM112	tRNA methyltransferase 112 homolog
tRNA	Transfer ribonucleic acid
TWIST1	Twist family bHLH transcription factor 1
WHS	Wolf-Hirschhorn syndrome
WHSC1	Wolf-Hirschhorn syndrome candidate 1 protein
WHSC1L1	Wolf-Hirschhorn syndrome candidate 1-like protein 1
WNT	Wingless/int signaling pathway
WT	Wild-type

1 Introduction

1.1 Posttranslational Modification of Proteins

Proteins harbor several posttranslational modifications (PTMs) that can be categorized into two major classes: enzyme-catalyzed modification and hydrolytic cleavage of proteins. The enzymatic-catalyzed modification reactions need cosubstrates, which provide the activated molecule that is added to the substrate. The second class is the hydrolytic cleavage, where one or more polypeptides are cleaved from proteins by enzymes called proteases. Additionally, the generation of disulfide bonds between two cysteine residues is also considered to be a posttranslational modification. Disulfide bonds are important for the proper folding and stability of many proteins[1].

Most of the covalently added PTMs occur on the side chains of amino acids, where functional groups of the amino acids serve as nucleophiles. These are hydroxyl groups (serine, threonine and tyrosine), carboxylates (aspartate and glutamate), thiolates (cysteine) or the functional groups of lysine, arginine and histidine. Even weaker amide nucleophiles of asparagine and glutamine can be modified in various ways[1]. These modifications regulate many protein properties and functions, such as stability, localization, interaction with other proteins or ligands, or alter the enzymatic activity. PTMs may also act in combination. Different modifications can influence each other, and preventing a certain event by blocking an adjacent residue[2], influence the catalytic activity[3] or change the substrate recognition efficiency[4] of the enzyme that is setting new modifications.

1.1.1 Protein Phosphorylation

The most common and well studied posttranslational modification is phosphorylation. Kinases, are the enzymes responsible for the addition of a phosphoryl group to the side chain of serine, threonine and tyrosine residues[5]. More than 500 enzymes are encoded in the human genome. Kinases use adenosine triphosphate (ATP) or more rarely guanosine triphosphate (GTP) as cosubstrates for the transfer of a phosphoryl group to their target substrates[1]. The phosphorylation of amino acid side chains can be reverted by dephosphorylation, which is catalyzed by phosphatases. Phosphorylation and dephosphorylation are important regulators of cellular processes. The introduction of the bulky and negatively charged phosphoryl group to one or more amino acids of a protein has drastic effects on protein function, conformation and interactions with other proteins[6].

For example, protein kinase A (PKA) is one of the best studied kinases and it serves as a good model enzyme. PKA is activated by increased levels of cyclic adenosine monophosphate

(cAMP), a second messenger, which is produced in the cAMP-dependent pathway by an initial signal transduced through a receptor at the plasma membrane[7]. Glycogen synthase[8] and phosphoryl kinase[9] are two of the many substrates, regulated by PKA. In addition, PKA regulates several other pathways by phosphorylating serine or threonine residues of enzymes, thereby modulating their activities.

1.1.2 Protein Acetylation

Another important posttranslational modification is the acetylation of proteins. Similar to phosphorylation, acetylation is also a very frequent occurring PTM on proteins. Acetyltransferases utilize acetyl-coenzyme A as a cofactor to transfer the acetyl group to the target residue of the protein. Two possible positions can be acetylated, the ε-amino group of a lysine residue or the Nα-terminus of a protein. Nα-terminal acetylation occurs during protein biosynthesis and it is therefore called co-translational modification. The enzyme complexes responsible for these modifications are called N-terminal acetyltransferases (NATs)[10]. The effects of N-terminal acetylation are extensive. This modification can influence protein stability[11], localization[12,13], protein synthesis[14] and is connected to metabolic regulation and apoptosis[15,16]. The second important position for acetylation is the ε-amino group of lysines. Although it is not as frequent as Nα-terminal acetylation, it contributes to many cellular functions[17]. By contrast to Nα-terminal acetylation, the transfer of the acetyl group onto the ε-amino side chain can be removed by deacetylases. Acetylation on a lysine residue neutralizes its positive charge and thereby alters the biochemical properties of the protein. In addition, it can also block other modifications on this lysine residue. This modification regulates transcriptional activity by changing the strength of the interaction between histones and deoxyribonucleic acid (DNA) on chromatin[18], affects protein-protein interactions[19] or influences protein stability by preventing lysine ubiquitination, which could lead to protein degradation[20].

The enzymes responsible for the acetylation of lysine residues on histones are called histone acetyltransferases (HATs) and the removal of this modification is catalyzed by histone deacetylases (HDACs). The ability to set and remove such a functional group on lysine can change the charge on histones and thus, alter the accessibility of the DNA in chromatin. Therefore it has a high impact on the regulation of transcriptional activity. Besides histones some HATs can also acetylate non-histone proteins, like α-tubulin[21] or the transcriptional regulator p53[22]. The effect of acetylation on non-histone proteins often depends on the position of the lysine that is acetylated. For instance, in the transcription factor p53, lysine acetylation next to the sequence-specific DNA binding domain increases the DNA binding[22]. By contrast, lysine (K65) acetylation within the DNA binding domain decrease sequence-specific DNA binding and disrupts enhanceosome[23].

1.1.3 Protein Methylation

Protein methylation has gained more and more interest in the last decades. Although the first methylated protein was already discovered 1959 by Ambler *et al.*[24], the understanding of this modification has begun only in recent years. Protein methylation can occur on several amino acids. Among the best studied are lysine and arginine methylation. Methylation of other residues, such as histidine, cysteine, asparagine or glutamine was also documented[25,26]. Methylation of amino acids has many functions. It can affect protein stability, protein-protein interactions, protein localization and have indirect effects on other posttranslational modifications. Protein methylation can also regulate gene transcription or DNA repair. The most well-studied protein methylation is lysine and arginine methylation on histone proteins. These residues are reported to be modified at numerous sites on N- and C-terminal histone tails. The side chain amino group of lysine can harbor up to three methyl groups and the guanidino group of the arginine side chain can accommodate up to two methyl groups. This makes a determination of the effect more complex. In contrast to phosphorylation and acetylation, the methyl group is relatively small and except for methylation of aspartate and glutamate[27] it does not change the charge of the modified residue. Therefore it is more likely that other effects of this modification control chromatin processes. One way is through recognition and binding of the methylated amino acid residues by other proteins, which further can lead to an activation or repression of gene transcription or initiation of DNA repair[28].

1.1.3.1 Lysine Methylation

Lysine methylation is an ubiquitous modification that occurs on numerous proteins and regulates various important cellular functions. The ε-amino group of the lysine side chain can accommodate up to three methyl groups, resulting in either un-, mono-, di- or trimethylated lysine, as depicted in Figure 1.
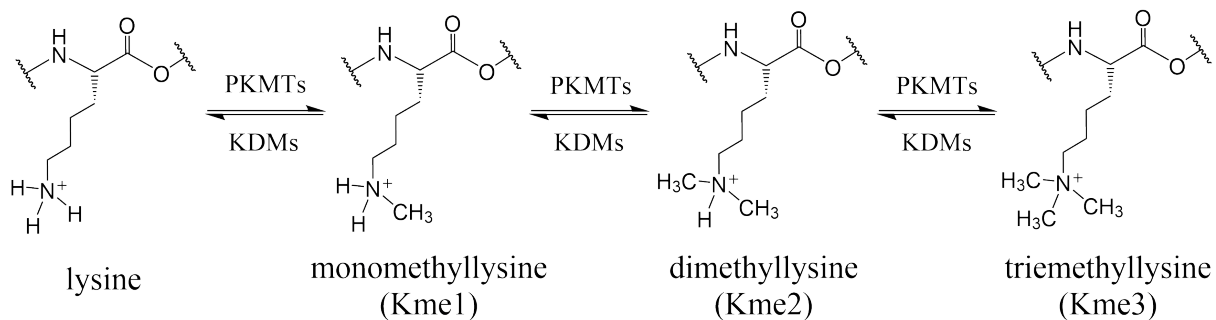
Figure 1: *Methylation states of lysine.* *Protein lysine methyltransferases (PKMTs) catalyzing the methylation of the ε-N atom of lysine. The removal of the methyl groups is catalyzed by lysine demethylases (KDMs). The four different methylation states are: unmethylated, monomethylated, dimethylated and trimethylated lysine.*

The most well characterized lysine methylation occurs on the histone tails of H3 and H4 proteins.

On histone H3, lysine residues at positions, such as 4, 9, 27 or 36 can be methylated, whereas on histone H4 the residues K20 and K44 are methylated. Lysine methylation can lead to different biological effects, based on the position and degree of methylated residues. Trimethylation of H3K4 is associated with active gene transcription[29], whereas methylation on H3K9, H3K27 or H4K20 is connected to heterochromatin formation, and subsequent gene repression[30–32]. Furthermore, different methylation stages lead to different signaling functions. Trimethylation of H4K20 is found at pericentric heterochromatin and is connected to gene repression[32], whereas H4K20 dimethylation is involved in DNA repair[33] and monomethylation of H4K20 oscillates during cell cycle[34].

Lysine methylation marks serve as binding sites for different proteins, which are capable to recognize with conserved functional domains called "reading" domains the methylated residues based on the degree of methylation and the surrounding sequence[35]. One of these binding domains is the chromodomain, which is present in many chromatin proteins binding to different methylated lysines: H3K4, H3K9me2/3, H3K27me2/3, H3K36me3 and H4K20me1[36]. Due to the different specificity of this domain one cannot generalize its effect on transcriptional regulation. The chromodomain of heterochromatin protein 1 (HP1) can bind to di- and trimethylated H3K9[37,38] and mediate transcriptional repression of genes[39], whereas the chromodomain helicase DNA-binding protein 1 (CHD1) from *Saccharomyces cerevisiae* binds trimethylated H3K4, a methylation mark associated with transcriptionally active chromatin[40,41]. In addition to the chromodomain there are several other domains, which can recognize methylated lysines on histones in a degree-specific manner: PHD (plant homeodomain) fingers, MBT (malignant brain tumor) repeats, Tudor and ADD (ATRX-DNMT-DNMT3L) domains[35].

Lysine methylation is not only present on histones, but also on non-histone proteins. Apart from the histone proteins, the effect of lysine methylation are thoroughly investigated in the tumor suppressor protein p53. This protein plays an important role in DNA repair, cell cycle regulation and apoptosis based on various stimuli. p53 is methylated at several lysine residues, such as K370, K372, K382 and K386 by different protein lysine methyltransferases (PKMTs)[42]. For example, SET7/9 and SMYD2 monomethylate p53 at K372 and K370, respectively. Methylation influences the activity of p53 depending on the lysine that is modified and the number of methyl groups added to the corresponding lysine. Methylation of K372 by SET7/9 increases transcription of p21, which further controls cell cycle arrest[43]. However, monomethylation by SMYD2 at K370 suppresses the binding of p53 to the p21 promotor and restrain the transcription. Interestingly, SET7/9 mediated K372 methylation inhibits K370 methylation by SMYD2, suggesting a regulatory crosstalk[44].

1.1.3.2 Arginine Methylation

Similar to lysine methylation, arginine methylation is present in both nuclear and cytoplasmic proteins. Methylation on arginine residues was identified 1967 by Paik and Kim[45]. With the discovery of arginine methylation on histone proteins and its role in various cellular functions, the importance of this PTM has gained significant attention.
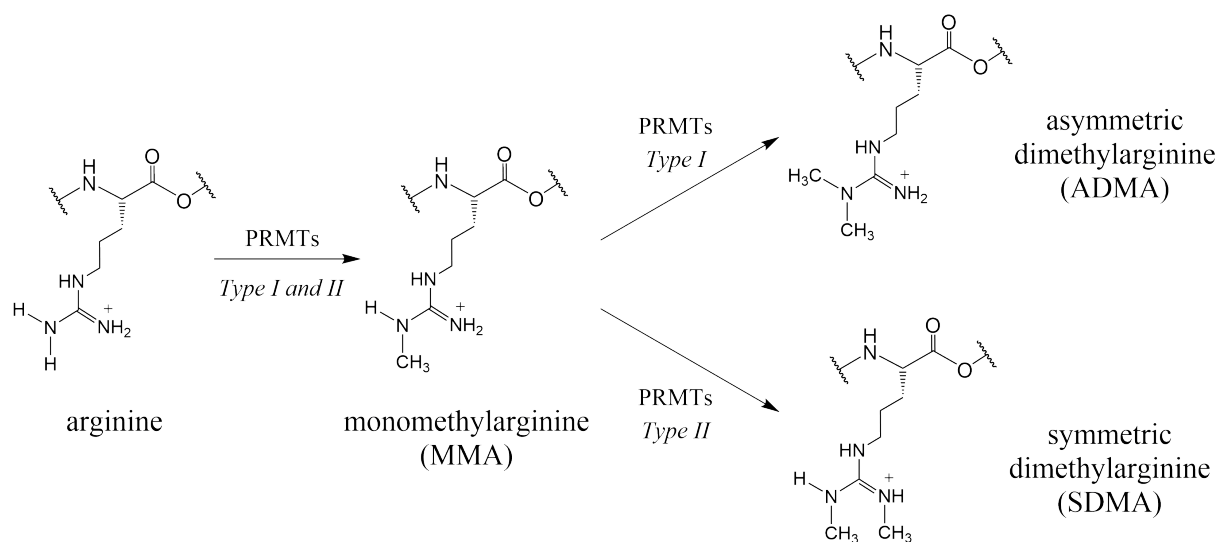
Figure 2: *Methylation states of arginine. The protein methyltransferases (PRTMs) catalyzing the monomethylation of arginine (MMA) on one of the guanidino ω-N atoms. The further methylation to asymmetrical dimethylarginine (ADMA) is catalyzed by type II enzymes and the generation of symmetrical dimethylarginine (SDMA) is catalyzed by type II enzymes.*

Arginine can have two different methylation states at the guanidino group of its side chain, which can be either mono- or dimethylated. The dimethylated guanidino group can be further differentiated based on the position of the methyl groups. It is referred to as symmetrical methylation, when the methyl groups are on different ω-N^G atoms, and as asymmetrical methylation, when both methyl groups are on the same ω-N^G atom (Figure 2). Arginine methylation influences many cellular processes, such as protein sorting[46], protein-protein interaction[47], transcriptional regulation[48,49], RNA processing[50,51], signal transduction[52–54] and DNA repair[55].

Enzymes catalyzing arginine ω-N^G-methylation are called protein arginine methyltransferases (PRMTs) and they can be divided into two types. Type I consists of the enzymes PRMT1, PRMT3, PRMT4 (CARM1) and PRMT6, which generate monomethylarginine and asymmetric dimethylarginine. Type II PRMTs are PRMT5 and PRMT7, which catalyze the formation of monomethylarginine and symmetric dimethylarginine[56]. So far only the Tudor domain, has been reported to interact specifically with methylarginine residues. The survival motor neuron (SMN) protein was one of the first proteins identified to bind to methylarginines via its tudor domain[57].

1.1.3.3 Glutamine Methylation

Glutamine methylation is a very rare modification unlike the lysine and arginine methylations described above. Only a handful of proteins were reported to possess a methylglutamine modification, although the first protein containing a methylated glutamine (the ribosomal protein L3 from *Escherichia coli*), was already found 1977 by Lhoest and Colson[58]. A recent study identified the only known glutamine methylation on histone H2A in yeast and human[59]. This glutamine methylation occurs at position Q105 in yeast and Q104 in human and are catalyzed by the glutamine methyltransferase Nop1 and the human ortholog Fibrillarin. Another important glutamine methylation was discovered at ribosomal polypeptide release factors (RFs)[60]. RFs are important for the termination of the synthesis of polypeptides at the ribosome. They recognize the stop codons within mRNA at the A-site of ribosomes and hydrolyze the ester bond between the nascent polypeptide chain and the peptidyl-tRNA at the P-site[61]. In bacteria two different release factors are necessary to recognize all three stop codons. RF1 recognize the UAA and UAG codons, while RF2 recognize the UAA and UGA codons[62]. In contrast, eukaryotes possess only one release factor, eRF1, which is able to recognize all three stop codons[63]. Though bacterial RFs and eukaryotic eRF1 does not share sequence or structural homology, they have a small universally conserved motif[64]. This motif comprises a glycine-glycine-glutamine (GGQ) tripeptide and it was shown to be involved in the hydrolysis of tRNA bound peptides[65]. Interestingly, the glutamine of the universal conserved GGQ motif is methylated, suggesting that it could affect the hydrolysis of nascent polypeptides. Later Dinçbas-Renqvis *et al.* confirmed that the glutamine methylation at the GGQ motif stimulates the translation termination in *E. coli*[60].

1.2 Protein Methyltransferases

In general, methyltransferases (MTases) catalyzes the transfer of a methyl group from a methyl donor to a substrate. The most commonly used methyl donor is S-Adenosyl-L-methionine (SAM) and the enzymes, utilizing this cosubstrate, are called SAM-dependent methyltransferases[66,67]. MTases catalyze a bimolecular nucleophilic substitution (S_N2) reaction, where the lone pair electrons of a nucleophile (substrate) attack the carbon atom of the methyl group of SAM. This results in the generation of a methylated "nucleophile" and S-Adenosyl-L-homocysteine (SAH)(Figure 3)[68].
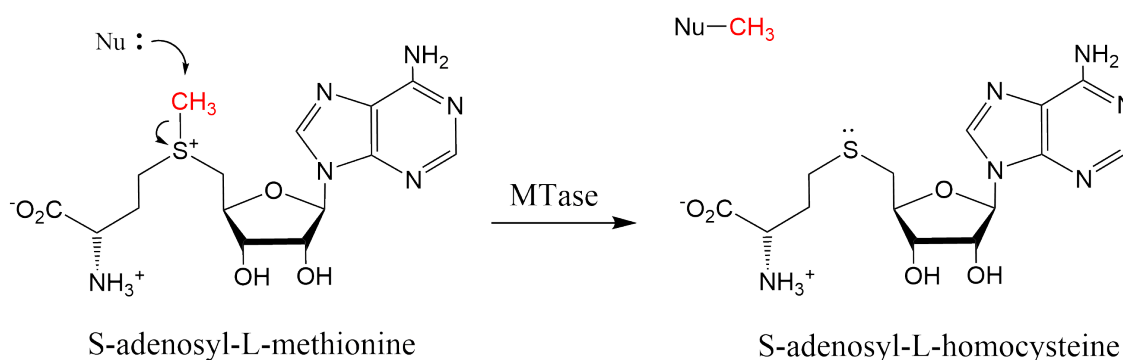
Figure 3: *General scheme of the methyl transfer reaction from methyl donor S-adenosyl-L-methionine to a nucleophile (Nu) catalyzed by methyltransferases (MTases), resulting in the formation of S-adenosyl-L-homocysteine and the methylated nucleophile (Nu-CH₃)*

MTases methylate a great variety of substrates. These can be DNA, RNA, proteins and small molecules. The enzymes are classified into different types depending on the substrates they methylate, like DNA methyltransferase (DNMTs) or protein methyltransferases (PMTs). These MTases are very specific with respect to the substrate, i.e. to a specific nucleobase or to a specific amino acid. DNA methylation can occur at the C^5 and N^4 position of cytosine and the N^6 position of adenine[69–71]. Protein methyltransferases show a much higher level of diversity and complexity than DNA methyltransferases. They can methylate a broad spectrum of amino acids. Several protein methyltransferases have been identified that are specific for lysine, arginine, glutamine, histidine or cysteine residues[26].

SAM-dependent MTases were initially categorized into five classes (I-V) depending on their structures[72]:

- Class I contains the MTases that harbor a Rossmann-like fold. It includes all DNMTs and several PMTs. It is the biggest group of MTases and has a large diversity of substrates. Class I MTases show high structural similarity, even when only little sequence similarity is notable. They are composed of a seven-stranded β-sheet flanked by α-helices. A conserved GxGxG sequence motif at the end of the first β-sheet is responsible for binding to

the nucleotide of SAM. Examples for this class of enzymes are the protein arginine methyl-transferases[73] and the members of the HemK group of glutamine methyltransferase[74,75].

- Class II MTases have a distinct protein structure with eight long antiparallel β-strands forming the core flanked by several α-helices on each side. SAM is bound by a conserved RxxxGY motif in a shallow groove formed by the β-strands. The methionine synthase, MetH, is the only known member of this class of MTases[76].

- The third class of MTases has a homodimeric structure. Similar to class I MTases, these proteins possess a GxGxG motif, but this is not involved in binding SAM. The SAM binding site of class III MTases is located between two βαβ-domains, each consisting of five-stranded β-sheets flanked by four α-helices. CbiF, a cobalt-precorrin-4 MTase is a member of this class[77].

- The class IV consist of the SPOUT family of RNA MTases. Their structure is made of a six-stranded parallel β-sheet flanked by seven α-helices. The first three strands form a half of a Rossmann-fold and part of the C-terminus forms a knot, which creates a binding cleft for the cofactor[78,79].

- Class V MTases are SET domain containing proteins. This includes a large number of enzymes with various substrates. The most prominent members of this class are protein lysine methyltransferases. The SET domain was named after the three proteins, which share this common motif, **S**u(var)3-9, **E**nhancer of Zeste and **T**rithorax. It consists of twelve β-strands forming up to five interwoven sheets flanked by regions called pre- and post-SET domains. These are important for methyltransferase activity and play a role in substrate recognition and specificity[80].

In the following sections of this doctoral thesis, the structural aspects and the known target substrates of three different protein methyltransferases will be described in more detail.

1.2.1 HEMK2

The bacterial N^5-glutamine methyltransferase HemK was first discovered in *E. coli* during a genetic screen for new heme biosynthesis mutants[81]. It was assumed that HemK plays a role in the oxidation of protoporphyrinogen to protoporphyrin IX. However, following knock-out experiments and phenotype analysis did not support this hypothesis[82].
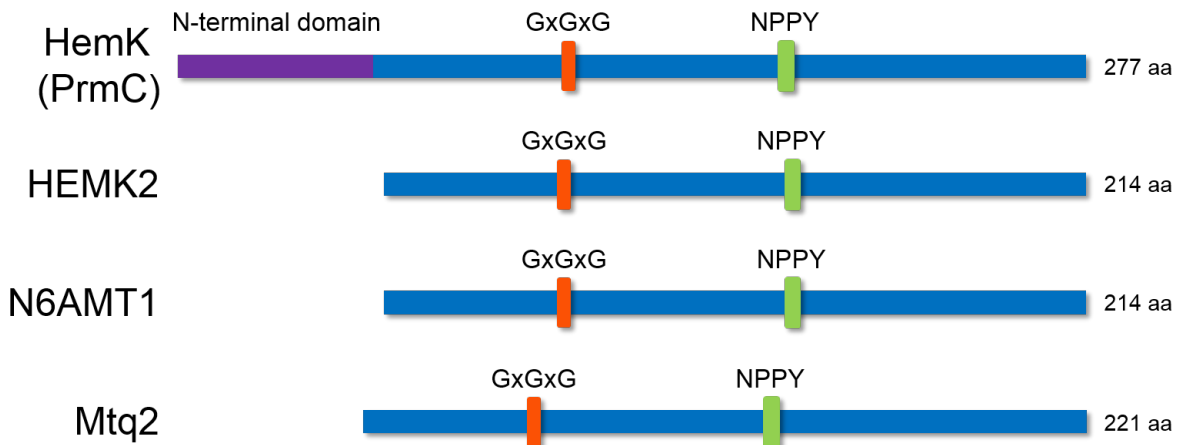
Figure 4: *Schematic representation of functional motifs of the HemK family. E. coli HemK, human HEMK2, mouse N6AMT1 and yeast Mtq2; GxGxG motif responsible for binding of the cofactor SAM, NPPY motif necessary for binding of the glutamine side chain, N-terminal domain, which is missing in eukaryotic family members (purple).*

Sequence alignment studies revealed that besides bacteria, several lower and higher eukaryotes also possess HemK homologs (Figure 4). An analysis of all HemK homolog sequences showed a shared NPPY motif[74]. It was thought that these conserved (D/N/S)PP(Y/F/W) motifs are limited to N^6-adenine and N^4-cytosine DNA MTases[83] and, therefore, HemK enzymes were classified as members of the SAM-dependent DNA MTase group[84]. Based on this finding, HemK was renamed to N^6-adenine-specific DNA methyltransferase (N6AMT). However, subsequent experiments could not show methyltransferase activity toward DNA[85]. Later, the seminal discovery that HemK methylates the glutamine residue in the universal conserved GGQ motif of the ribosomal release factors RF1 and RF2 was reported[86,87]. This finding confirmed the classification of HemK as a SAM-dependent MTase, however the substrate is a protein instead of DNA, as initially predicted. This led to renaming HemK as PrmC (**Pr**otein **m**ethyltransferase **C**). The eukaryotic homologs of the bacterial HemK enzyme are called HEMK2 in human, N6AMT1 or PRED28 in mice and Mtq2p or YDR140w in *S. cerevisiae*. They all methylate the conserved glutamine residue of the corresponding eukaryotic release factor 1 (eRF1)[88–90].

1.2.1.1 Structure and Catalytic Mechanism of HemK

The first crystal structures of bacterial HemK were derived from *Thermotoga maritima*[91] and *E. coli*[74]. Although the sequences of these two enzymes share only 31 % identity and 51 % similarity, the overall structure is very similar. The enzymes consist of two structural domains: a small N-terminal domain with a bundle of α-helices connected via a β-hairpin linker to the larger catalytic C-terminal domain. The C-terminal domain consists of a seven-stranded mixed β-sheet flanked by several α-helices, which is characteristic for class I MTases (Figure 5).
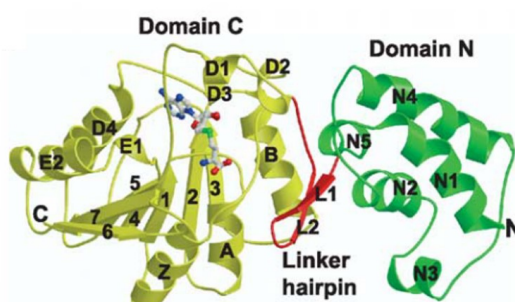
Figure 5: *Ribbon representation of the structure of E. coli HemK-SAM. The N-terminal domain is painted in green, the catalytically active C-terminal domain is painted yellow and a linker connecting these two domain is represented in red. SAM is depicted as a stick model. The picture was adopted from Yang et al.*[74]*.*

In both enzymes, the cofactor SAM is bound by the nucleotide-binding sequence motif GxGxG, placed at the C-terminal end of β-strand 1[74]. Additionally to the GxGxG motif, another conserved motif can be found in all HemK homologs, the NPPY tetrapeptide. This is positioned at the end of β-strand 4 and forms the bottom of the active site pocket of HemK. The NPPY motif is necessary for binding the glutamine side chain[91]. Recently, the crystal structure of *E. coli* HemK in complex with its substrate RF1 and the cofactor S-Adenosyl-L-homocysteine (SAH, methyl donor reaction product) was solved. RF1 is composed of four domains, a compact structurally rigid center formed by the domains 2 and 4, which are flanked by two more flexible domains 1 and 3. In both release factors (RF1 and RF2) the universally conserved GGQ motif is positioned on a flexible loop protruding from domain 3[92]. This explains its ability to enter the peptidyl transferase center (PTC) of the ribosome and promote the hydrolysis of a nascent polypeptide from the tRNA[62]. Apart from this, RFs have an anticodon segment, which is an important part to recognize the stop codons at the A-site of a ribosome. Although both release factors contain a tripeptide as an anticodon segment, the residues of these tripeptides are different. RF1 possess a proline-valine-threonine (PVT) motif, whereas the tripeptide of RF2 consists of serine-proline-phenylalanine (SPF). This explains the different specificity of RF1 and RF2 toward the stop codons[93].

In contrast to the bacterial HemK, the eukaryotic glutamine MTases need a binding partner to methylate the glutamine residue of eRF1. The glutamine MTase interacts with the small

zinc-binding protein called TRM112 (Ynr046w in yeast). TRM112 consist of two domains: a zinc-binding domain composed of N- and C-terminal residues and a central domain (Figure 6A). The zinc atom is coordinated by four cysteine residues, two from the N-terminal part (Cys^{11} and Cys^{16}) and two from the C-terminal section (Cys^{112} and Cys^{115})(Figure 6B)[94].
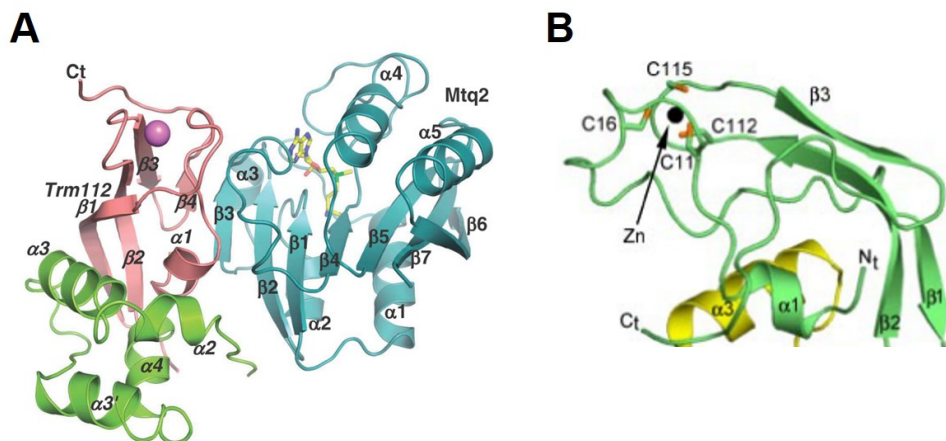
Figure 6: *Ribbon diagrams of the structure of the yeast Mtq2 and TRM112. (A) Representation of the structure of the Mtq2-TRM112 complex. Mtq2 is painted in blue, the TRM112 zinc-binding domain is shown in pink and its central domain is painted green. The zinc atom is represented as a purple sphere. The picture was adopted from Liger et al.[75]. (B) Structure of the yeast TRM112 protein. The TRM112 zinc-binding domain is shown in green and part of the central domain is painted yellow. The zinc atom is represented as a black sphere and the Cys side chains coordinating the zinc atom are shown as sticks. The picture was adopted from Heurgué-Hamard et al.[94].*

The yeast glutamine MTase Mtq2 together with TRM112 forms a heterodimeric complex, which stimulates the activity of Mtq2 and prevents its aggregation. TRM112 masks hydrophobic regions of Mtq2 upon interaction. This enhances the solubility of Mtq2. In addition, TRM112 increases the SAM binding of Mtq2, because the loop connecting the β-strands 3 and 4, which is involved in SAM binding, is stabilized by the TRM112 interaction. Structural comparison of the bacterial and yeast glutamine MTase, showed that the yeast homolog Mtq2 possess only the class I SAM-dependent MTase domain, but not the additional N-terminal domain present in HemK from *E. coli*. However, the superposition of the HemK-RF1 and Mtq2-TRM112 structures clearly revealed that TRM112 is not a substitute for the N-terminal domain[75].

During methylation of the glutamine residue, RF1 fits perfectly onto the concave surface, formed by the two domains of HemK and the GGQ motif is inserted into the active site pocket. The N-terminal domain of HemK contacts the domains 2 and 3 of RF1, whereas the C-terminal part of HemK only binds domain 3 of RF1. Here, the glutamine side chain forms hydrogen bonds with the NPPY motif of HemK, which facilitate the methyl transfer[92]. The hydrogen bonds are formed between the two hydrogens of the N^5-amide of glutamine and the main chain oxygen of proline 198 and the side chain oxygen of asparagine 197 of the NPPY motif. Furthermore, the

side chain oxygen of glutamine interacts with the tyrosine 200 main chain amide via a hydrogen bond. These hydrogen bonds induce a change in hybridization of the amide nitrogen from sp^2 to sp^3, which allows a nucleophilic attack of the lone-pair electrons toward the methyl group of SAM[91]. The (D/N/S)PP(Y/F/W), which is generally referred as DPPY motif is mainly found in N^6-adenine and N^4-cytosine DNA MTases, however it does not exclusively bind nucleotides. It interacts rather with nitrogens associated with a planar system, like the amide in glutamine or nucleotide bases in adenine or cytosine. The hydrogen bond formation between DPPY and substrate is common for MTases with such a motif and was observed in DNMTs, like *TaqI* or PMTs, such as HEMK2[72].

Not much is known about the mechanism of substrate recognition and the interaction between HEMK2 and its substrate in mammals. A detailed crystal structure of HEMK2 in complex with eRF1 could provide more information and reveal the residues involved in the interaction between enzyme and its substrate. However, such a crystal structure is not available yet, and the existing crystal structures of bacterial HemK in complex with its cognate release factor are not helpful, since the *E. coli* and mammalian release factor amino acid sequences differ outside of the conserved GGQ motif.

1.2.1.2 Effects of Glutamine Methylation

After the identification of HemK as the responsible enzyme methylating the bacterial and eukaryotic release factors, many groups determined the outcome of glutamine MTase depletion in different species. In *E. coli*, knock-out of HemK reduced the termination activity of unmethylated RF1 and RF2 by approximately 3- to 4-fold. While this had no major effect on cell growth in rich media, growth was reduced on poor carbon sources[95]. Deletion strains of the yeast homolog Mtq2p showed stronger growth defects and several phenotypes in rich media. However, the deletion strain did not show a significant decrease of translation termination efficiency. The cells displayed cold-sensitivity and they were also sensitive to paromomycin or geneticin, two aminoglycosides affecting protein synthesis by binding to ribosomes. They also revealed increased resistance to the fungicides thiabendazole and benomyl[96]. Compared to bacteria or lower eukaryotes, depletion of the glutamine methyltransferase N6AMT1 in mice has drastic consequences. The knock-out leads to reduced cell proliferation, heavily impaired post-implantation development of mutant embryos and early embryonic lethality[89].

Apart from methylation of the eRF1 protein, not much is known about the cellular functions of HEMK2. The drastic effects of HEMK2 knock-out in mice suggest that it may have a broader role in cellular processes and development. In the recent years several studies showed that many PMTs possess additional unknown substrates[97,98]. Their identification may contribute to a better understanding of the role of the enzyme and its methylated substrates in cells. To

gain deeper insights into the cellular role of HEMK2, it would be helpful to understand the mechanism of how HEMK2 recognize its substrate, eRF1, and find out whether HEMK2 may have additional unknown substrates with other biological functions. In this study, the substrate specificity profile of HEMK2 was determined and used to identify novel HEMK2 substrates *in vivo* and *in vitro*.

1.2.2 The NSD Family

The nuclear receptor SET domain-containing (NSD) enzyme family belongs to the SET-domain containing class V of PMTs. The family consists of NSD1, NSD2 (also MMSET or WHSC1), and NSD3 (WHSC1L1), which all share the same functional domains: PWWP (proline-tryptophan-tryptophan-proline motif) domains, PHD (plant homeodomain) domains and the catalytically active SET domain with an AWS- (associated with SET) and Post-SET domain. The members differ in the overall protein sizes and exact arrangement of the domains[99].
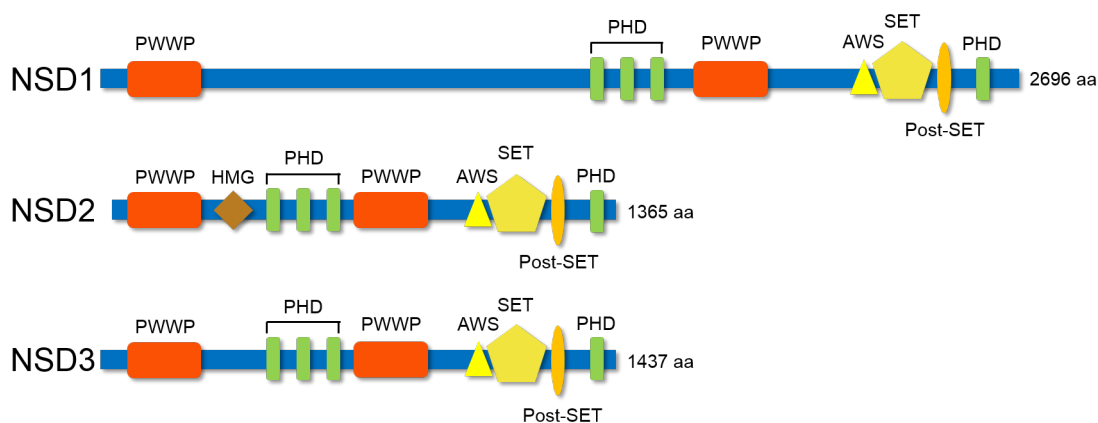
Figure 7: *Schematic representation of functional domains of NSD1, NSD2 and NSD3. PWWP domain; PHD zinc-finger domain; SET lysine methyltransferase (KMT); AWS domain (associated with SET domain); Post-SET domain, HMG box.*

While the precise biological function of the three NSD family members is still not completely understood, several studies showed that NSD1, NSD2 and NSD3 mainly catalyze mono- and dimethylation of H3K36[100]. In addition, multiple other histone lysine residues were also reported to be methylated by these enzymes: H4K20[101], H4K44[102] and H1.5K168[102] for NSD1, H3K4[103], H3K27[104], H4K20[105,106] and H4K44[107] for NSD2, and H3K4 and H3K27 for NSD3[108]. However, some doubts were raised with respect to the methylation activities toward H3K4, H3K27 and H4K20, due to disagreements among published reports[107]. The biological functions of all three NSD family members seems to be very important, since a dysregulation of protein level is involved in many different cancer types and genetic disorders.

The *NSD1* gene is located on chromosome 5q35 and encodes a 2696 aa long protein. Haploinsufficiency caused by either microdeletions or intragenic mutations of the *NSD1* gene leads to the Sotos syndrome[109,110]. This is characterized by prenatal and postnatal overgrowth, characteristic facial appearance, advanced bone age, developmental delay[111] and malignancies[112,113]. A second genetic disorder is the Beckwith-Wiedemann syndrome. This is more rare, and is associated with heterozygous loss-of-function or truncating mutations of NSD1[114]. Besides these two genetic disorders, NSD1 is connected to several cancer types, like breast cancer[115], neuroblastomas and glioblastomas[116], multiple myeloma[117] and acute myeloid leukemia (AML)[118].

Approximately 5 % of all AML patients are diagnosed to contain a t(5;11)(q35;p15.5) translocation, which encodes for a NUP98-NSD1 fusion protein. This fusion protein interacts with CBP/p300 in a complex and exhibits acetyltransferase activity along with the H3K36 methylation activity, which leads to the aberrant expression of HOX genes[119]. Lu *et al.* showed that NSD1 is also able to methylate the non-histone protein NF-κB, which plays a crucial role in innate and adaptive immune responses. Mono- and dimethylation of lysine 218 (K218me1) and lysine 221 (K221me2) activates the protein, while demethylation of the same residues by the protein lysine demethylase FBXL11 inactivates NF-κB. Methylation of K218 and K221 of NF-κB favors cell proliferation, colony formation and gene expression in HT29 cancer cells[120]. While recent studies showed that NF-κB was not methylated by NSD1[102], the regulation of NF-κB in cells through methylation and demethylation at K218 and K221 by other PKMTs cannot be denied. Although many studies suggested that NSD1 is an important oncogene other reports showed that NSD1 can act as a tumor suppressor[121,122]. Taken together, it is possible that NSD1 acts as tumor suppressor or an oncogene depending on the cellular context and already existing variations of other chromatin modifiers.

NSD2, which was investigated in this study, will be described in more detail in section 1.2.2.1.

NSD3 is the third member of the NSD family and consist of 1437 aa. It harbors four zinc-finger PHD domains, two PWWP domains and the catalytically active SET-domain. It is also referred as Wolf-Hirschhorn syndrome candidate 1-like 1 (WHSC1L1), although in contrast to the other two NSD family members, no relevant overgrowth syndromes were connected to defects in the *NSD3* gene. Similar to *NSD1*, the *NSD3* gene undergoes a chromosomal translocation, t(8;11)(p11.2;p15) in AML, which leads to the generation of NUP98-NSD3 fusion protein[123]. Besides AML, NSD3 was also frequently found upregulated in human breast cancer cell lines[124,125], bladder cancer, lung cancer, liver cancer and chronic myelogenous leukemia (CML)[126]. Yang *et al.* demonstrated the differential expression of two transcription factors IRX3 and TBL1X, in cancer cells that overexpress NSD3 and also in cells ectopically expressing NSD3. IRX3 and TBLIX are known to positively regulate WNT-signaling pathway. At the same time SFRP1, a negative regulator of the WNT-signaling pathway, is downregulated by NSD3[127]. This suggests that NSD3 may be a driver of oncogenesis.

1.2.2.1 NSD2

The NSD2 enzyme, also known as Wolf-Hirschhorn syndrome candidate 1 (WHSC1) or multiple myeloma SET domain (MMSET), is the smallest member of the NSD family, with a length of 1365 aa. NSD2 consists of the catalytically active SET domain with its AWS and Post-SET domains, two PWWP domains, four PHD zinc-finger domains and one HMG (high mobility group) box. Several studies reported different substrate lysines on histones H3 and H4 for NSD2.

As such, the dimethylation of K4 and K9 of histone H3[103], trimethylation of H3K27[104], di- and trimethylation of H4K20[105,106], monomethylation of H4K44[107] and di- and trimethylation of H3K36[107,128] were documented.

1.2.2.2 Aberrant NSD2 Expression is Involved in the Wolf-Hirschhorn Syndrome and Various Cancers

Dysregulation of NSD2 causes the Wolf-Hirschhorn syndrome (WHS). This is characterized by developmental defects, like a prominent forehead with widely spaced eyes, divergent strabism, heart and several midline fusion defects, growth retardation and brain anomalies, which lead to mental retardation[129,130]. WHS patients either show a partial or complete deletion of the *NSD2* gene, leading to a haploinsufficiency of NSD2. This suggests that *NSD2* is essential in causing this syndrome[131]. Nimura *et al.* showed NSD2-deficient mice exhibit phenotypes similar to the human WHS, such as growth defects, deficiencies in midline fusion and congenital heart defects. Mice with heterozygous *NSD2* $^{+/-}$ mutation exhibit lower level of the protein than the WT mice, show symptoms as described above, but are viable and fertile. In contrast, homozygous *NSD2* $^{-/-}$ mice show more severe growth defects and die 10 days after birth[128].

Besides the significant role of NSD2 in WHS, many reports also connect NSD2 to different cancer types. Expression profile analysis showed elevated levels of *NSD2* mRNA in bladder, lung, breast, prostate, renal and pancreas cancer lines[132]. An upregulation in protein levels was documented in ganglioneuromas, ganglioneuroblastomas and neuroblastomas[133]. While NSD2 seems not to affect survival, several studies showed a correlation between elevated NSD2 protein levels and progression of cancer, in oligodendroglioma, breast, prostate and head and neck cancers[134]. In endometrial cancer and hepatocellular carcinoma it was reported that increased levels of NSD2 were associated with tumor development, shorter overall survival and disease-free survival[135,136]. NSD2 was mentioned for the first time at the t(4;14)(p16.3;q32.3) translocation in multiple myeloma (MM). This is the second most common translocation occurring in about 20 % of all multiple myeloma patients[137]. Upon translocation, the immunoglobulin heavy chain (IgH) promotor (14q32.3) is connected to the *NSD2* gene (4p16). This results in a chimeric fusion transcript of *IgH-NSD2* and leads to aberrant overexpression of two proteins: the fibroblast growth factor receptor 3 (FGFR3) and NSD2. Initially *FGFR3* was assumed to be the driving oncogene in MM, later it was shown that about 30 % of MM patients lack overexpressed FGFR3, but still have an increased *NSD2* gene product. This suggests a crucial role of NSD2 in multiple myeloma[138–140]. Kuo *et al.* found that the dimethylation of H3K36 is the critical chromatin mark affected in multiple myeloma with t(4;14) chromosomal translocation[141]. They demonstrated that the catalytic activity of NSD2 is responsible for the H3K36 dimethylation and subsequent gene activation in these cell lines. Altering the genome-wide profile of H3K36me2 in MM cell lines leads to upregulation of silenced cancer-associated genes or genes linked to cell

proliferation or survival. In addition to the globally increased level of H3K36 dimethylation, the level of methylated H3K27, a modification associated with gene repression, was significantly reduced. This alteration of histone modifications changed the chromatin structure to a more open state. The genes affected by NSD2 are involved in the regulation of cell death, DNA repair, cell cycle, p53 pathway and integrin-mediated signaling. A depletion of NSD2 in MM cells lead to decreased growth, increased cell adhesion and apoptosis[142].

1.2.2.3 Somatic Cancer Mutations of NSD2

The catalogue of somatic mutations in cancer (COSMIC) database contains approximately 300 varying mutations in NSD2. These were identified by sequencing analysis of numerous different cancer cell lines and patient specimens. Interestingly, among these, some mutation appeared to be more frequent than others. One of these is the exchange of a glutamic acid to lysine at the position 1099 (E1099K). This mutation resides in the catalytic SET domain located in a loop adjacent to the substrate binding pocket, and it was hypothesized that it may alter the methyltransferase activity or the substrate specificity of NSD2[143]. Another cancer database (CCLE = Cancer Cell Line Encyclopedia) shows that the E1099K mutation of NSD2 mostly appears in pediatric lymphoid malignancies, such as hypodiploid acute lymphoid leukemia (ALL), chronic lymphocytic leukemia (CLL), multiple myeloma, lung adenocarcinoma and adenocarcinoma of the stomach[144]. Jaffe *et al.* also observed the recurrent occurrence of NSD2 E1099K mutation in 14 % of pediatric B-cell ALL, but not in adult ALL patients[143]. In both studies the authors could show a higher methyltransferase activity of the NSD2 E1099K mutant. This led to an increased level of H3K36 dimethylation and decreased level of H3K27 trimethylation comparable to the effect in cells with the t(4;14)(p16.3;q32.3) translocation[143,144].

1.2.2.4 Effects of the Aberrant Expressed NSD2 and its Recurrent Somatic Cancer Mutant

NSD2 affects numerous of genes connected with different cancer types. Although often the exact role and mechanism of NSD2 is not enlightened, recent reports provided new insights into the function of this enzyme in various diseases. Ezponda *et al.* revealed the binding of NSD2 to the *TWIST1* gene (twist family bHLH transcription factor 1), which is associated with epithelial-mesenchymal transition (EMT) and invasion in different cancers, such as prostate cancer. The upregulation of TWIST1 is induced by NSD2 mediated H3K36 dimethylation of the *TWIST1* locus[145]. The same effect on TWIST1 was observed by Oyer *et al.* with the hyperactive NSD2 E1099K mutant, leading to an upregulation of about 21-fold compared to wild-type NSD2[144]. The NSD2 protein has also been reported to be overexpressed in 40 % of the primary prostate cancer tumors and its overexpression correlated with the activation of NF-κB in the tumors[146]. NSD2 acts as a coactivator to regulate the NF-κB signaling in castration therapy resistant

prostate cancer. NSD2 interacts with NF-κB and elevates the expression of NF-κB target genes, by di- and trimethylaton of H3K36 in the promotor regions. Interestingly, the NF-κB target genes, inflammatory cytokines IL-6 and TNF-α, are in turn able to stimulate NSD2 expression thereby creating a positive-feedback loop, which plays an important role in tumor growth[146].

Despite the described roles of NSD2 or its hyperactive mutant (E1099K) in the promotion of proliferation, survival and tumorigenicity of multiple myeloma and other cancer types, the cellular function in normal cells was hardly investigated. Considering, the numerous reported target sites of NSD2 on the histone proteins H3 and H4[103,107,128], and the rising number of identified non-histone substrates for various PKMTs in the last years[97,98,102,147] a closer look should be taken, at whether NSD2 can affect cellular processes by methylation of non-histone proteins as well. Without a crystal structure of NSD2 together with its cognate substrate, important information on how NSD2 interacts with its substrates and how the recognition and selection may work are missing.

For this reason, the substrate specificity profile of NSD2 was characterized. Based on this, several substrates candidates were identified, which are methylated by NSD2 at peptide level. Additionally, methylation on three novel substrate was shown at protein level and cellular methylation for two of them was confirmed in HEK293 cells. These information may help to understand more about the cellular functions of NSD2 and could be useful for the treatment of the various cancers in which NSD2 is involved.

1.2.3 The Suv39 Family

The Suv39 protein family, was named after the first member Su(var)3-9, identified in a genetic screening for position effect variegation (PEV) mutations in *Drosophila melanogaster*. It was shown that Su(var)3-9 is a suppressor protein, which is associated with heterochromatin condensation[148]. Su(var)3-9 possesses several eukaryotic homologs, like Clr4 in *Schizosaccharomyces pombe*, Suv39h1 in mice and SUV39H1 in humans. In higher eukaryotes, like mouse and human, an additional homolog, SUV39H2, is present along with the SUV39H1[149].

1.2.3.1 SUV39H1

SUV39H1 consists of 412 amino acids and contains two conserved chromatin-associated domains, which are characteristic of the Suv39 family. These are the C-terminal SET domain, which is the catalytic center, and the N-terminal chromodomain that recognizes and binds methylated lysine residues. In addition, the catalytically active SET domain is flanked by a Pre-SET and Post-SET domain (Figure 8).
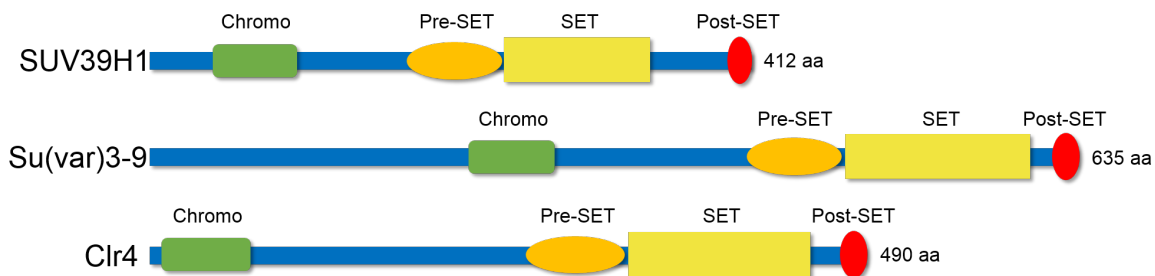
Figure 8: *Schematic representation of functional domains of the Suv39 family. Human SUV39H1, D. melanogaster Su(var)3-9 and S. pombe Clr4; N-terminal Chromo domain; C-terminal Pre-SET, SET lysine methyltransferase and Post-SET domains.*

SUV39H1 was the first identified histone lysine methyltransferase in humans and it was shown to trimethylate lysine 9 on histone H3[150]. The trimethylated H3K9 deposited by SUV39H1 and other enzymes, provides binding sites for HP1 proteins, which are associated with heterochromatin formation, spreading and gene silencing[37]. The process of spreading this mark along chromatin by SUV39H1-dependent methylation utilizes the recognition and binding of a chromodomain at H3K9me3 sites. Besides the ability to bind H3K9me3 marks, the chromodomain of SUV39H1 is important for the catalytic activity as well. Deletion of the N-terminal part (including the chromodomain) or just deletion of the chromodomain, led to a radically reduced methylation activity[151]. The same effect was observed after truncation of the N-terminus of Su(var)3-9 in *D. melanogaster*[152]. Based on structural modeling studies with the HP1 chromodomain, Chin *et al.* could show that the amino acid residues located in the SUV39H1 chromodomain i.e., tryptophan 64 and tyrosine 67 that are part of the aromatic binding pocket, are also necessary for its enzymatic activity. Mutation of one of these amino acids leads to similarly

decreased catalytic activity as the deletion of the entire chromodomain[151]. A dysregulation of SUV39H1 could have an effect on the regulation of its target genes and genomic stability. Peters *et al.* showed that a deletion of Suv39h1 and Suv39h2 in mice led to reduced H3K9 methyl levels in pericentric heterochromatin, followed by growth defects, reduced viability, genomic instability and increased tumorigenesis[153]. An increased expression level of SUV39H1 is observed in basal-like breast cancer (BLBC). This leads to increased H3K9me3 levels and DNA methylation at the promoter of the *E-cadherin* gene[154]. It was also reported that SUV39H1 interacts with several transcriptional factors, which are thought to be oncogenic proteins. This causes transcriptional repression, aberration in bone marrow immortalization and hematopoietic differentiation, and involvement in acute myeloid leukemia[155,156].

1.2.3.2 Clr4

The histone lysine methyltransferase Clr4 is the yeast homolog of Su(var)3-9. Clr4 is a 490 amino acid long protein, with an N-terminal chromodomain and a catalytically active SET domain. In fission yeast the Clr4 multiprotein complex (ClrC), which consists of Clr4, Cul4, Rik1, Raf1 and Raf2, is necessary for heterochromatin formation. Clr4 functions as a reader and writer of H3K9 methylation. It is recruited to chromatin via the RNAi machinery. The chromodomain of Clr4 can bind to H3K9me sites and the SET domain can modify adjacent nucleosomes, thereby providing new binding sites for ClrC. This allows the maintenance and spreading of heterochromatin structures[157]. Additionally, it was shown that Swi6 (the HP1 homolog in *S. pombe*) co-localizes at H3K9me3 sites via its chromodomain. It further interacts with Clr4 and strengthens the binding of Clr4 at heterochromatin[158,159]. Though several studies reported a role for Clr4 in the maintenance and spreading of heterochromatin, there are still discrepancies regarding how specific the enzyme is recruited to the methylation sites. An exact answer is not known yet, but it seems that a difference in the selectivity of the chromodomains of Clr4 and Swi6 avoids competition in binding of methylated H3K9. The chromodomain of Clr4 showed a higher preference for H3K9me3 over H3K9me2 (5- to 6-fold), compared to the Swi6 chromodomain, which displayed only a 1.5- to 2-fold discrimination for H3K9me3 over H3K9me2[160].

The expansion of substrates to non-histone proteins was consistently shown for a lot of protein histone methyltransferases during the last years[26]. The yeast homolog Clr4 revealed methyltransferase activity on the non-histone protein Mlo3 *in vitro* and *in vivo*. Mlo3 is required for nuclear export of RNA and is associated with mRNA quality control. It was found to interact with Clr4 and Rik1, a subunit of ClrC complex. The methylation at lysine 167 of Mlo3 is necessary for the production of centromeric siRNA and suppression of antisense RNA[161]. In this doctoral thesis, the substrate specificity profile of Clr4 was determined. Additionally, it was shown that Clr4 methylates six novel substrate candidates at peptide level.

2 Aims of the Study

In the recent years, the identification of novel substrates for PKMTs gained more and more attention, as it was shown that many enzymes of this type are able to methylate non-histone proteins in addition to their known histone substrates. Methylation of non-histone proteins has several important regulatory roles in cellular processes and may also influence the functions and properties of the methylated protein. It is believed that there are still numerous of non-histone substrates, which remain to be identified.

The primary aim of this doctoral thesis was to characterize the substrate recognition of three different protein methyltransferases and to discover novel protein substrates. HEMK2 is a glutamine methyltransferase, which had been reported to methylate the glutamine residue of the universally conserved GGQ motif of the eukaryotic release factor eRF1. Methylation of eRF1 is crucial for normal translation termination and the hydrolysis of nascent polypeptides from the ribosome. Therefore, it was planned to characterize the substrate specificity profile of HEMK2 and based on this recognition motif to find novel protein substrates, which are methylated at peptide and at protein level. To show the cellular methylation of *in vitro* methylated substrates, it was intended to develop a Qme-specific antibody, able to detect HEMK2-dependent methylation *in vivo*. Furthermore, the biological effects of the methylation of some of the novel substrates were addressed.

The second enzyme studied in this thesis was the NSD2 histone lysine methyltransferase. This was known to dimethylate K36 of histone H3 and play pivotal role in various diseases and cancers. The objective of this part of the present work was to determine the specificity profile of NSD2 and to screen for novel non-histone substrates candidates, which are methylated at peptide and at protein level and in mammalian cells. Additionally, the effects of the NSD2 cancer mutations, occurring within the catalytically active SET domain on the substrate specificity should be analyzed.

Clr4, the histone lysine methyltransferase in *Schizosaccharomyces pombe* was the third enzyme studied in this thesis. Clr4 trimethylates K9 of histone H3, a mark associated with heterochromatin formation, spreading and gene silencing. It was already reported that Clr4 methylates Mlo3, a non-histone protein as well. The goal of this project was to characterize the substrate recognition profile of Clr4 and to investigate whether it may methylate additional reported interaction partners at the peptide level.

The aim of the final part of this thesis was to develop a novel assay, which is able to detect the methyltransferase activity of PKMTs with the help of natural reading domains. Reading domains are natural protein domains able to detect methylation on a specific lysine residue and

the degree of methylation. A reading domain based assay may overcome known disadvantages of conventional PKMT assays, such as the use of radioactively labeled SAM, high costs and batch-to-batch variability associated with the application of methyl-specific antibodies. It was planned to develop a microplate based assay suitable for PKMT inhibitor screening.

3 Results

3.1 Characterization of the Substrate Specificity of the Glutamine Methyltransferase, HEMK2

The glutamine methyltransferase HEMK2 was the second reported MTase, and has glutamine as target. HEMK2 catalyzes the transfer of a methyl group form the cofactor SAM to the glutamine side chain of the eukaryotic release factor eRF1. The target glutamine, which is part of an universal conserved GGQ motif, plays an important role in the hydrolysis of nascent polypeptides from tRNA at the ribosome[65]. Methylation at this residue increases the translational termination efficiency and prevents aberrant read-through. Importantly, deletion of the glutamine methyltransferase in higher eukaryotes, such as mice, led to diminished cell proliferation, strongly impaired embryonic development and early embryonic lethality[89]. Currently, no crystal structures of higher eukaryotic HemK homologs are available, so a more detailed insight into the interaction between mammalian HEMK2 and its eRF1 substrate is not possible. The exact process of substrate recognition and differentiation is still not fully understood. Since the amino acid sequences of the bacterial and mammalian enzymes and their substrates are too different outside of the conserved motifs, the crystal structure of *E. coli* HemK with its release factors is not helpful in addressing these issues.

3.1.1 Purification and Assessment of Methyltransferase Activity

The His_6-tagged mouse HEMK2 enzyme and the mouse complex partner TRM112 were co-expressed in *E. coli* BL21-CodonPlus (DE3) cells and purified by affinity chromatography with good yield and purity (Figure 9). To determine the methyltransferase activity of the recombinant
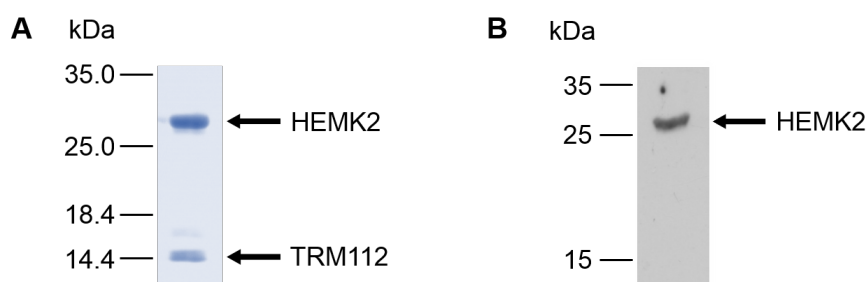
Figure 9: *Quality of the purified His_6-fused HEMK2 enzyme. (A) Coomassie stained SDS-PAGE gel of the co-purification of His_6-tagged HEMK2 and TRM112. (B) Confirmation of the identity of the purified His_6-tagged HEMK2 enzyme by probing with anti-His antibody.*

HEMK2, peptide arrays were synthesized on a cellulose membrane using the peptide SPOT synthesizer. Due to the fact the eukaryotic release factor 1 is the only known substrate of HEMK2, peptides were synthesized with the sequence of eRF1 (178 - 192) harboring the target

glutamine methylation site Q185. Peptides containing the target glutamine Q185 exchanged to alanine were included as negative control. Peptide arrays were incubated with the HEMK2-TRM112 complex and radioactively labeled [methyl-^{3}H]-SAM. The transfer of the methyl groups was detected by autoradiography. A clear methylation signal was observed for the wild-type, but not for the peptides containing the mutations (Figure 10A). Heurgué-Hamard *et al.* reported that GTP is required for the methylation of eRF1 by HEMK2[90], so the activity of HEMK2 was tested by incubation of the above described peptide arrays in methylation buffer with or without GTP.
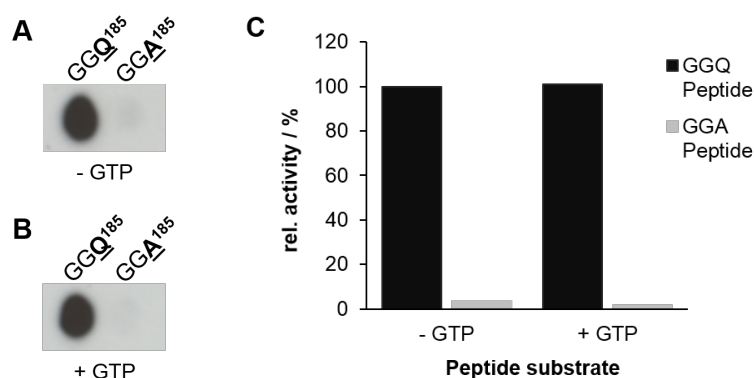
Figure 10: *Determination of the methyltransferase activity of HEMK2-TRM112 complex. The membrane contained eRF1 178 - 192 (KKHGRGG**Q**SALRFAR) and Q185A (KKHGRGG**A**SALRFAR) peptides. Peptide arrays were incubated without (A) or with (B) GTP in methylation buffer. The images were taken from the same autoradiography film. (C) Bar diagram presenting the quantitative analysis of the autoradiography images shown in (A) and (B), activities were normalized to the GGQ peptide of (A).*

Similar results were obtained when GTP was added to the methylation buffer (Figure 10B) and quantification of the methylation reactions demonstrated that the addition of GTP had no significant effect on the methyltransferase activity of the HEMK2-TRM112 complex (Figure 10C). After demonstrating the activity of the HEMK2-TRM112 complex on the peptide substrate, the efficiency of the HEMK2-TRM112 activity on the protein substrates was assessed. For this, the His$_6$-tagged eRF1 protein was expressed in *E. coli* and purified by affinity chromatography with a good yield and purity (Figure 11A). The identity of the purified His$_6$-tagged eRF1 protein was confirmed by western blot with an anti-His antibody (Figure 11B). Methylation reactions were performed by incubation of the purified eRF1 protein with the HEMK2-TRM112 complex in the presence of radioactively labeled [methyl-^{3}H]-SAM. The reaction mixture was separated by SDS-PAGE and the transfer of the radioactive methyl groups to the substrate was detected by autoradiography. Figure 11C clearly shows a methylation signal corresponding to the eRF1 protein size.
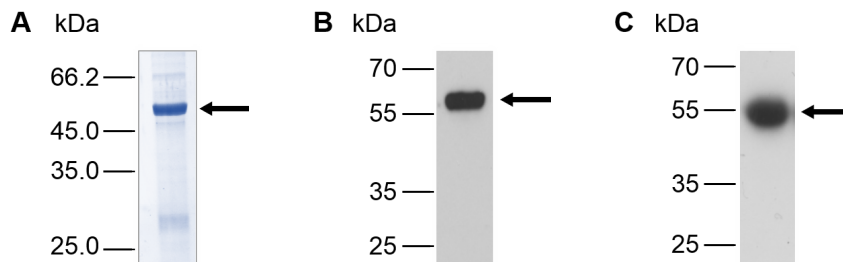
Figure 11: *Methyltransferase activity of HEMK2 on purified His$_6$-eRF1 protein.* (A) Coomassie stained SDS-PAGE gel of the purified eRF1 protein. (B) Confirmation of the identity of the purified eRF1 protein by probing with anti-His antibody. (C) Autoradiography of methylated eRF1 by the HEMK2-TRM112 complex. Bands corresponding to the eRF1 size are marked with an arrow.

3.1.2 Determination of the Specificity Profile of HEMK2

To investigate the substrate recognition of HEMK2, variable scanning peptide SPOT arrays were synthesized using the eRF1 sequence (179-192) as template. These arrays contained peptides in which one amino acid of the original sequence was exchanged by one of the other 19 proteinogenic amino acids, such that all possible single mutant variants of the original sequence are examined. 300 peptides were synthesized in total, including one wild-type peptide at the beginning of each row. The peptide arrays were incubated in methylation buffer containing the HEMK2-TRM112 complex and radioactively labeled [methyl-^{3}H]-SAM. The methylation of the peptides was detected by autoradiography (Figure 12A). The experiment was performed three times and the results of each experiment were quantitatively analyzed using the Phoretix^{TM} Array software. First, the intensity of each spot was normalized as described in section 6.1.2. The normalized spot intensities of each experiment were then averaged and color-coded using Conditional Formatting with dual color (black to light gray) scale in Microsoft Office Excel (Figure 12B). Thereby, the black squares represent a strong methylation of the corresponding peptides and the light gray boxes represent low activity. To evaluate the quality of the results obtained from the specificity array experiments, the standard deviations of the spot methylation intensities (SD) were calculated. For about 70 % of the peptides a SD of less than 10 % and for about 95 % of all peptides a SD of less than 20 % was obtained. This indicates a very good quality of the data (Figure 13A). Subsequently, the discrimination factor for the recognition of each amino acid at each single position was calculated. This factor reveals if the HEMK2-TRM112 complex favors a specific amino acid, over all other amino acids at the particular site for methylation (Figure 13B). The recognition motif of HEMK2 comprises the amino acids from R182 (-3) to R192 (+7) of the eRF1 sequence, the position of the target Q185 is defined as 0. As expected, exchange of the target glutamine Q185 to any other amino acid prevented methylation. In addition to this, HEMK2 showed a strict specificity toward the residues G184 (-1) and R189 (+4). Mutating these amino acids to any other residue, eliminated the methylation activity. Some preferences were observed at residues S186 (+1) for S, R and G, and R192 (+7) for R and

A

B
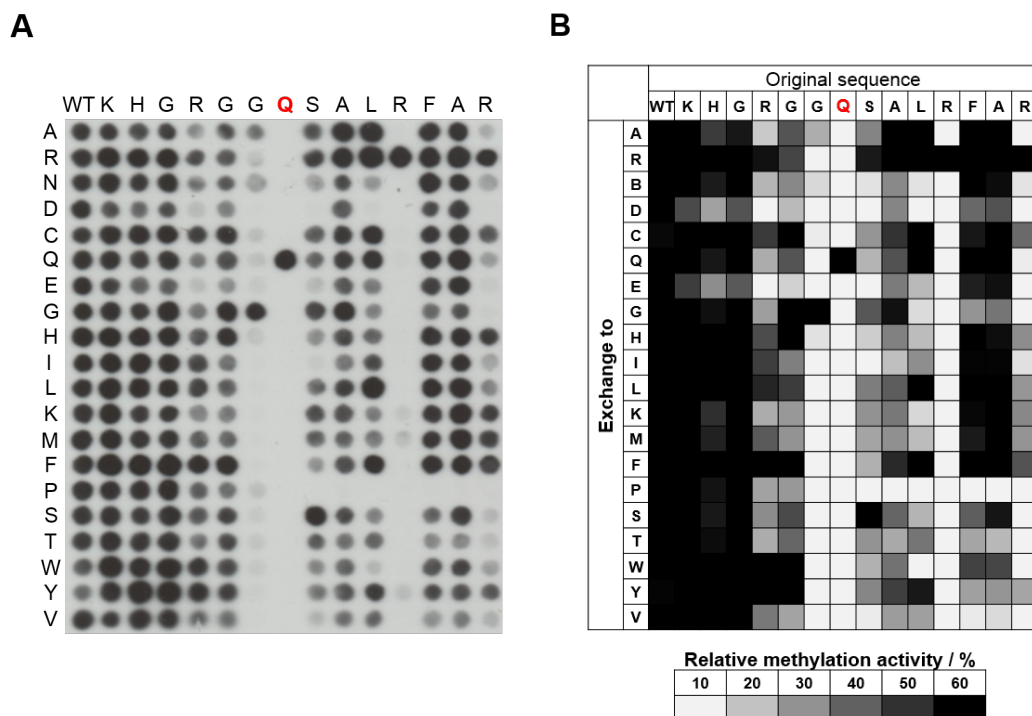
Figure 12: *Substrate specificity profile of HEMK2. (A) Autoradiography image of the specificity profile array based on eRF1 sequence (179 - 192) methylated by HEMK2-TRM112. The horizontal axis represents the original eRF1 sequence and the target glutamine Q185 is highlighted red. The vertical axis shows the residues exchanged at the corresponding position in the original sequence, which provides an array with all possible single amino acid mutations of the eRF1 sequence. The first column contains the wild-type sequence of eRF1 as a control (labeled with WT). (B) Averaged spot intensities of each peptide array experiment of the HEMK2-TRM112 complex. The individual results were normalized and color-coded depending on their methylation activity. Black to light gray represents a strong to weak methylation.*

slightly less strong for F, C and K. Furthermore, some weaker preferences were detected at the amino acids R182 (-3), G183 (-2), A187 (+2) and L188 (+3). Interestingly, HEMK2 did not tolerate an amino acid exchange to proline, glutamic acid, aspartic acid or valine at the positions 184 - 192. Based on these findings, a minimal recognition motif of HEMK2 can be proposed as: G-**Q**-X_3-R.
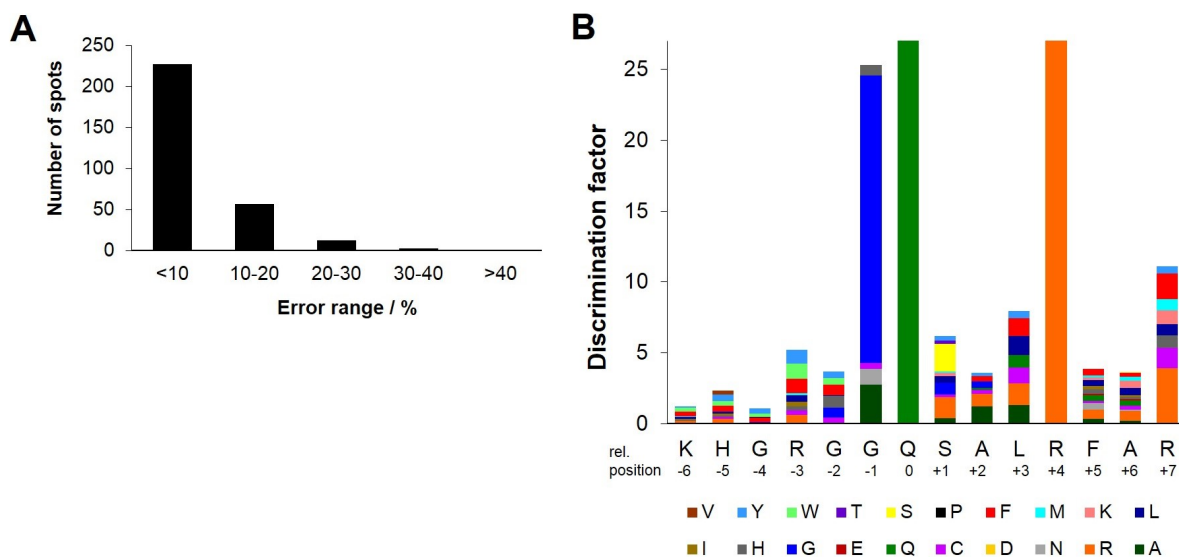
Figure 13: *Evaluation of the specificity profile results. (A) Quality control of the peptide intensities derived from the specificity profile arrays of HEMK2, which is shown in Figure 12B. Standard deviation of the averaged HEMK2 activity on all peptide substrates was calculated. (B) The discrimination factors for the recognition of each amino acid at the corresponding position by HEMK2 is represented in a bar diagram.*

3.1.3 Identification of Putative HEMK2 Peptide Substrates

The substrate specificity profile analysis revealed that HEMK2 prefers or tolerates several other residues apart from the amino acids naturally present in eRF1 at many positions. This suggests that HEMK2 might methylate additional substrates. Therefore, Scansite[162] searches were performed with the substrate specificity profiles shown in Table 1, to retrieve putative novel protein substrates in the human proteome.

Table 1: *Substrate specificity profiles utilized to identify putative novel HEMK2 substrate*

Cognate residue Position	G184 -1	Q185 0	S186 +1	A187 +2	L188 +3	R189 +4
Search profile 1 (stringent)	G	Q	SRYKLG	ARFGL WYCS	LARQ CFYT	R
Search profile 1 (relaxed)	G	Q	SRYKL GAMTC	ARFGLW YCSQKH	LARQC FYTI	R

The first search was performed with a relatively stringent substrate specificity motif (Table 1; Search profile 1), which retrieved 138 putative novel candidates. A second search with a more relaxed specificity profile (Table 1; Search profile 2), discovered 164 additional potential substrates. 15 amino acid long peptides harboring the predicted target glutamine of the 302 identified putative candidates were synthesized on 2 peptide arrays. The protein names and the peptide

sequences of the synthesized substrates are listed in Table 10 and 11 in section 6.1.3. The peptide arrays were methylated with the HEMK2-TRM112 complex as described before, and the transfer of the radioactive methyl groups was detected by autoradiography (Figure 14A and B). eRF1 wild-type and the corresponding Q185A mutant peptides were included as controls (marked by black and grey arrow heads).
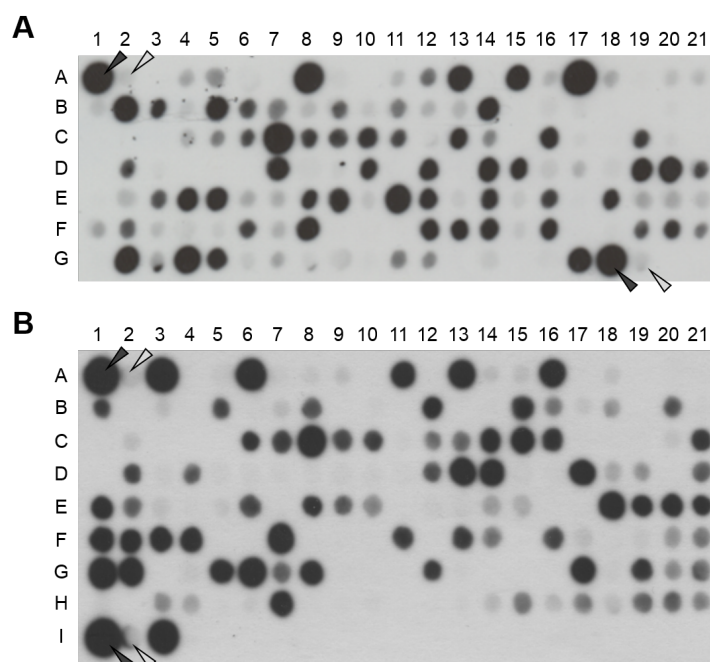
Figure 14: *Methylation of novel peptide substrates by HEMK2. Peptide arrays containing putative novel peptide substrates were methylated by the HEMK2-TRM112 complex. Protein names and peptide sequences are listed in Table 10 and 11 in section 6.1.3. (A) Peptide array containing the putative substrates identified using search profile 1. (B) Peptide array containing the putative candidates identified using search profile 2 (Table 1). Wild-type and Q185A peptides of eRF1 were included as control and are marked by black and grey arrow heads.*

A quantitative analysis of the peptide arrays showed that 49 of the potential substrates were methylated with an equal or slightly weaker preference compared to the eRF1 control peptide. 76 putative candidates showed a reduced, but still detectable methylation signal. The high amount of methylated peptide substrates in these experiments confirms the reliability of the search profiles. Interestingly, 17 of the 42 strongly methylated peptides contain a preferred R at the +7 position (which correspond to the R189 residue in the original sequence) although this was not specified in the search profile. Similarly, S, R or G residues at the +1 position (S186) were observed in 26 of the 42 strongly methylated peptides, although seven other amino acids were allowed in the search at this site. Taking together, it was shown that HEMK2 methylated around 40 % of the 302 putative novel peptide substrates with good activity *in vitro* and the derived substrate specificity profile is in agreement with the sequences of the strong methylated peptides.

3.1.4 *In vitro* Methylation of the Putative Novel Protein Substrates

As described above, HEMK2 catalyzes the methylation of several novel peptide substrates *in vitro*. However, peptide methylation can not be directly correlated to the protein level, as sometimes the target glutamine may not be accessible for the methyltransferase in the 3D context of the folded protein.

Therefore, the methylation of the most promising peptide substrates was investigated also at protein level. 58 strongly methylated peptide substrates were chosen and cloned as protein domains, harboring the target glutamine (Table 7 in section 6.1.1). Putative protein substrates were overexpressed as GST fusion proteins and purified by affinity chromatography. From the 58 selected substrates, 35 were purified with sufficient yield to proceed with further experiments (Figure 15).
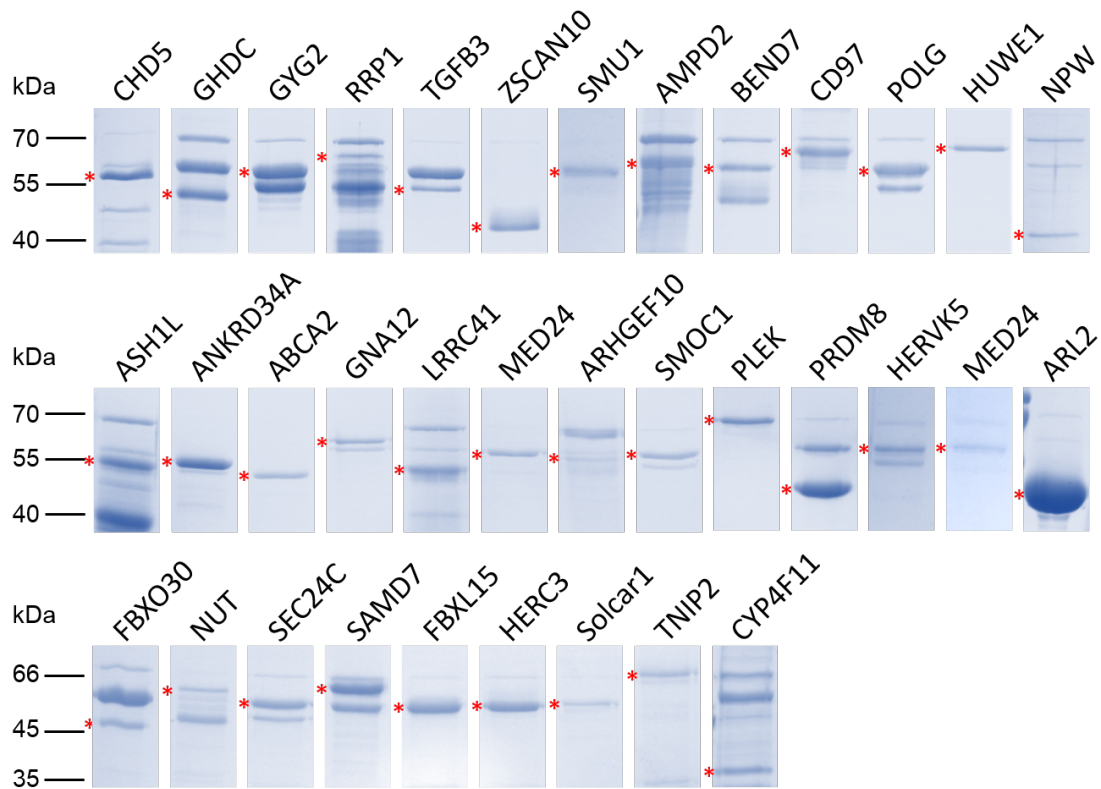
Figure 15: *Quality of the purified GST-fused substrate candidates. Coomassie stained SDS-PAGE gels of the 35 successfully purified GST-fused protein domains. The corresponding bands of the expected size are marked with a red asterisk.*

The other 23 candidates failed during different steps from cloning to the purification stage. Comparable amounts of the 35 proteins were separated by SDS-PAGE and stained with Coomassie to provide an input control of the proteins used for the methylation reactions (Figure 16A). Methylation assays were performed with the substrate candidates by incubation of the target proteins with the HEMK2-TRM112 complex and radioactively labeled [methyl-^{3}H]-SAM in methylation buffer. The samples were separated by SDS-PAGE and the transfer of the radioactive methyl groups to the target proteins was detected by autoradiography. As shown in Figure 16B, suc-
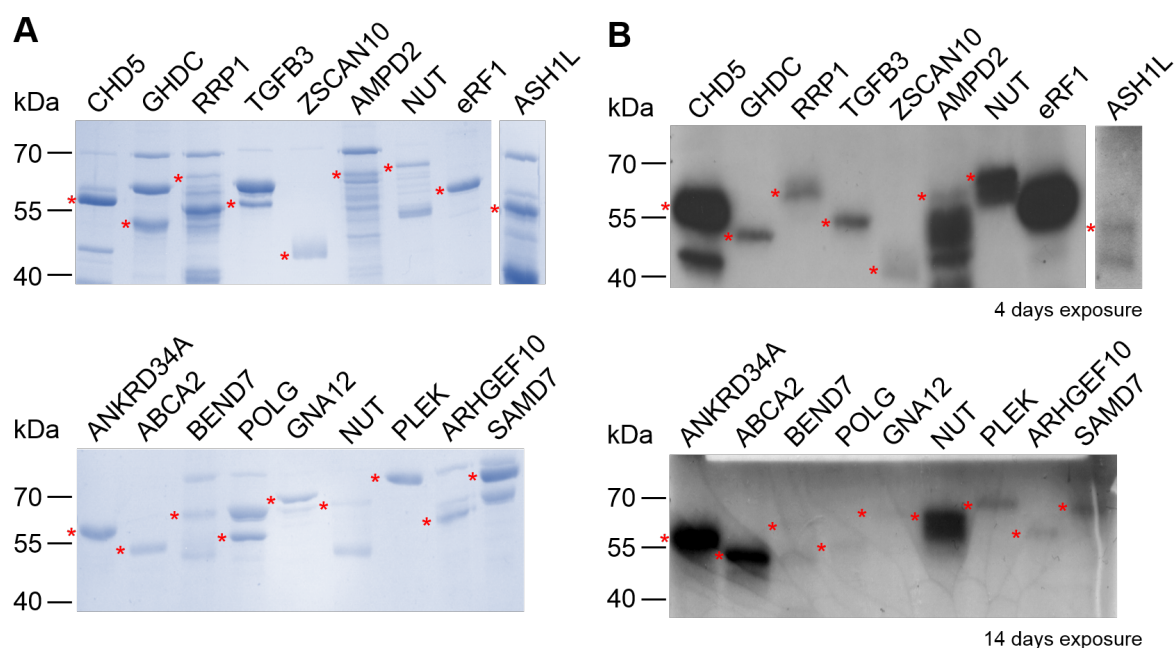
Figure 16: *Example of two methylation assays of the purified putative protein substrates. (A) Coomassie stained SDS-PAGE gels of the protein substrates (left top and bottom pictures). (B) Autoradiography images of the protein substrates methylated by HEMK2-TRM112 complex (right top and bottom pictures). The corresponding bands of the expected size are marked with a red asterisk.*

cessful methylation of 16 protein substrates by HEMK2 was observed. Out of these 5 proteins showed strong methylation (CHD5, AMPD2, NUT, ANKRD34A and ABCA2), 8 substrates exhibited weaker signals (GHDC, RRP1, TGFB3, ZSCAN10, ASH1l, PLEK, ARHGEF10 and SAMD7) and very weak methylation was observed on BEND7, POLG and GNA12 (detailed information about the substrate proteins are listed in Table 2). An expected strong methylation signal was observed for eRF1, which was included as positive control. The other 19 purified proteins did not show any methylation signal in the autoradiography images (data not shown). To confirm that the substrate methylations occurred at the predicted target glutamine residues, site-directed mutagenesis was performed for 11 proteins and the target glutamine was exchanged to an arginine (Table 9 in section 6.1.1). The mutant proteins were overexpressed and purified by affinity chromatography. Comparable amounts of purified wild-type and mutant proteins were used in *in vitro* methylation reactions (Figure 17 upper panels). Methylation assays were
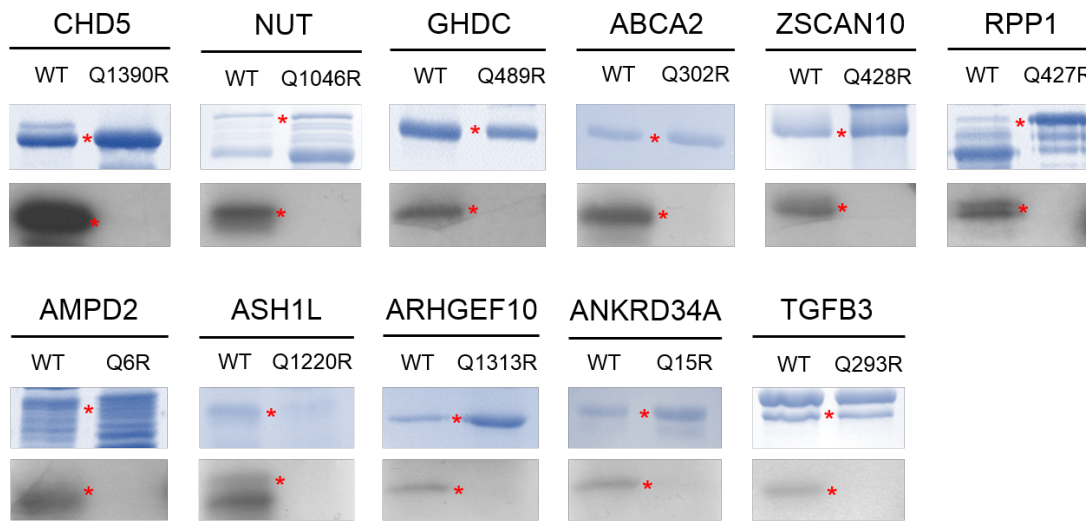
Figure 17: *Confirmation of the target glutamine methylation. Purified wild-type (WT) and mutant (Q to R) protein substrates were methylated by the HEMK2-TRM112 complex in presence of radioactively labeled [methyl-³H]-SAM and separated by SDS-PAGE. The amounts of wild-type and mutant substrates used in the assay were verified by Coomassie staining of the SDS-PAGE gels (upper panels). Detection of the radioactive methyl groups is shown by the autoradiography images on the lower panels. The corresponding bands of the expected size are marked with a red asterisk.*

performed by incubating the wild-type (WT) or mutant (Q to R) protein domains with HEMK2 in the reaction buffer containing radioactively labeled [methyl-³H]-SAM (Figure 17 lower panels). The results revealed that the HEMK2-TRM112 complex methylates the wild-type substrate proteins, but not the corresponding Q to R mutants (Figure 17 lower panels). This confirms that the methylation observed on the substrate proteins takes place at the predicted target glutamine residues and mutation at this position to arginine prevents the methylation of the protein domains by HEMK2.

Table 2: *In vitro methylated protein substrates of HEMK2. Names, abbreviations, boundaries of the protein domains and the position of the predicted target glutamine are provided.*

Name	Abbre-viation	Swiss Prot no.	Domain boundaries (aa)	Target Q Position
AMP deaminase 2	AMPD2	Q01433	2 − 135	6
Ankyrin repeat domain-containing protein 34A	ANKRD34A	Q69YU3	5 − 235	15
ATP-binding cassette sub-family A member 2	ABCA2	Q9BZC7	168 − 403	302
BEN domain-containing protein 7	BEND7	Q8N7W2	9 − 282	78
Chromodomain-helicase-DNA-binding protein 5	CHD5	Q8TDI0	1234 − 1530	1390
DNA polymerase subunit gamma-1	POLG	P54098	154 − 387	330
GH3 domain-containing protein	GHDC	Q8N2G8	325 − 529	489
Guanine nucleotide-binding protein subunit alpha-12	GNA12	Q03113	183 − 320	231
Histone-lysine N-methyltransferase ASH1L	ASH1L	Q9NR48	1119 − 1333	1220
Pleckstrin	PLEK	P08567	2 − 350	107
Protein NUT	NUT	Q86Y26	867 − 1132	1046
Rho guanine nucleotide exchange factor 10	ARHGEF10	O15013	1107 − 1343	1313
Ribosomal RNA processing protein 1 homolog A	RRP1	P56182	219 − 461	427
Sterile alpha motif domain-containing protein 7	SAMD7	Q7Z3H4	71 − 416	179
Transforming growth factor beta-3	TGFB3	P10600	159 − 405	293
Zinc finger and SCAN domain-containing protein 10	ZSCAN10	Q96SZ4	364 − 521	428

3.1.5 Cellular Methylation of the Novel Target Substrates

In vitro protein methylation assays confirmed the methylation of the predicted target glutamine residues on 11 substrates, at protein level. To determine if HEMK2 is able to methylate the substrate proteins in cells, 7 of the strongly methylated targets were selected (CHD5, NUT, AMPD2, TGFB3, GHDC, RRP1 and ZSCAN10) to investigating cellular methylation.

The chosen protein domains were subcloned into mammalian expression vectors. To detect their cellular methylation, a methyl-glutamine antibody was generated (Biotem, France). The antibody was raised against a GQ(me)G tripeptide and was validated by probing against *in vitro* methylated and unmethylated protein domains. GST-fused protein domains were incubated in methylation buffer containing the HEMK2-TRM112 complex and unlabeled SAM. As a negative control, comparable amounts of unmethylated targets were used. The protein samples were separated by SDS-PAGE, transferred onto nitrocellulose membranes and then probed with the methyl-glutamine-specific antibody (Figure 18).
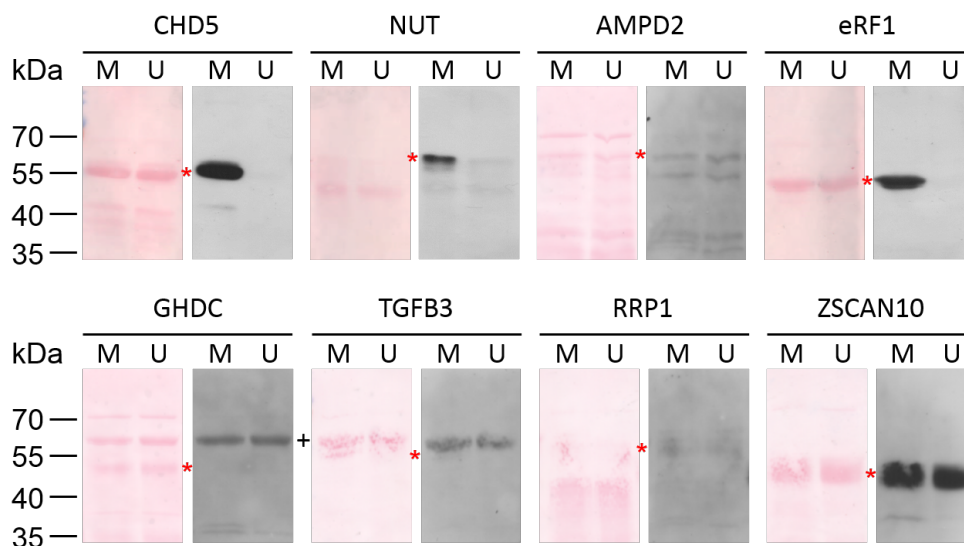
Figure 18: *Validation of methyl-glutamine specific antibody on methylated and unmethylated substrates. Unmethylated and methylated protein substrates were separated by SDS-PAGE, transferred onto nitrocellulose membranes and probed with methyl-glutamine specific antibody. The corresponding bands of the expected size are marked with a red asterisk. The bands marked with a + in the GHDC and TGFB3 samples are not the protein substrates, they represent unspecific binding of the antibody.*

The methyl-glutamine antibody showed a good discrimination between methylated and unmethylated substrates for CHD5 and NUT (Figure 18). For other substrates, the antibody either showed no differences in signal between the methylated and unmethylated proteins (ZSCAN10, AMPD2 or RRP1) or no signal at all (GHDC and TGFB3). In the case of GHDC and TGFB3, the antibody signal was observed at much higher size than the corresponding substrate protein sizes, which might be due to cross reactivity. Since, CHD5 and NUT were the only

substrates that showed a methylation-specific antibody signal and these two substrates exhibited the strongest methylation level *in vitro*, they were chosen for further cellular methylation studies.

For this purpose, expression of the YFP-fused substrates (CHD5, NUT and eRF1), and of the glutamine methyltransferase HEMK2-HA and its complex partner TRM112-myc was tested. HEK293 cells were transfected with one of the mentioned plasmids and harvested three days after transfection. The cells were lysed, and the extracted proteins were separated by SDS-PAGE and transferred onto nitrocellulose membranes. These were probed with the corresponding antibodies to determine the expression levels of the desired proteins (Figure 19).
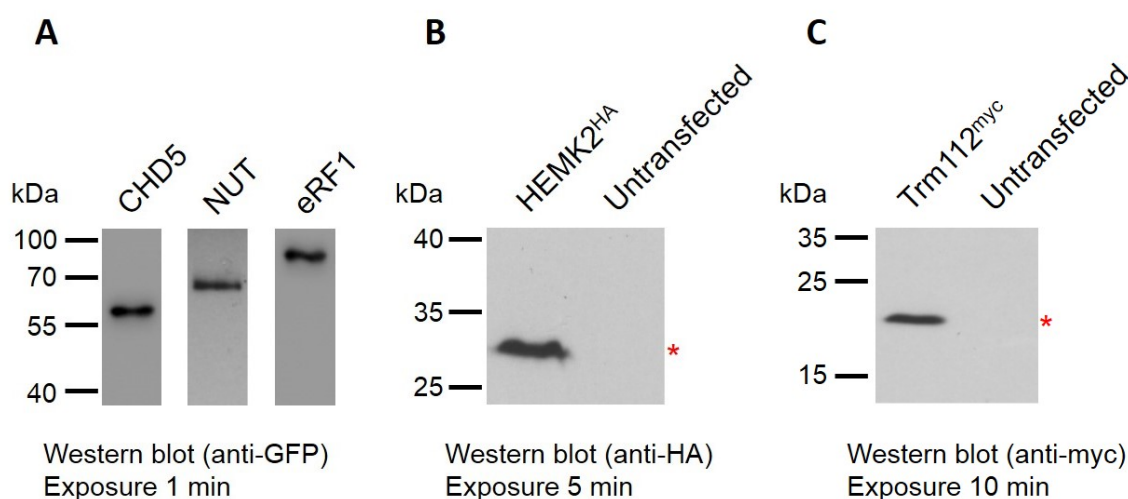
Figure 19: *Detection of protein expression in HEK293 cells. Cells were transfected with the YFP-fused CHD5, NUT or eRF1 proteins (A), HA-tagged HEMK2 (B), and HA-tagged HEMK2 together with its myc-tagged complex partner TRM112 (C). Cells were lysed and the protein samples were separated by SDS-PAGE and transferred onto nitrocellulose membranes. Proteins were detected with the corresponding antibodies. Untransfected cells were used as control. The corresponding bands of the expected size are marked with a red asterisk.*

The results revealed the successful overexpression of the 2 novel HEMK2 substrates (CHD5 and NUT) and of eRF1. Ectopic expression of HA-fused glutamine methyltransferase (HEMK2-HA) and its myc-tagged complex partner TRM112 was also confirmed in HEK293 cells (Figure 19C and B). To examine the cellular methylation of the newly identified substrates by HEMK2 in human cells, the YFP-fused protein domains (CHD5 and NUT) were transiently expressed either with or without the HEMK2-TRM112 complex. Three days after transfection, the YFP-fused protein substrates were purified using GFP-Trap® A beads. Approximately 10 % of the immunoprecipitated protein samples were separated by SDS-PAGE, transferred onto a nitrocellulose membrane and probed with an anti-GFP antibody to provide a loading control (Figure 20).

CHD5-YFP	+	+
HEMK2 + TRM112	+	-

kDa
70 —
55 —

NUT-YFP	+	+
HEMK2 + TRM112	+	-

kDa
70 —
55 —
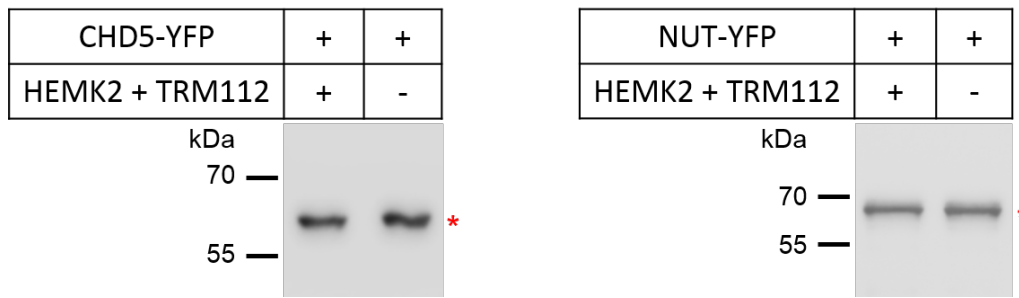
Figure 20: *Immunoblot detection of protein expression in HEK293 cells. The cells were transfected with the YFP-fused protein substrates CHD5 or NUT either with or without the glutamine methyl-transferase HEMK2 and the complex partner TRM112. The cells were harvested, lysed and the substrate proteins were purified by GFP-Trap®. Approximately 10 % of the purified target proteins were separated by SDS-PAGE and transferred onto nitrocellulose membranes. The expression of the target protein was analyzed by probing with anti-GFP antibody to adjust the amounts for further experiments. The corresponding bands of the expected size are marked with a red asterisk.*

To assess the methylation, the purified protein substrates were used for the western blot as described above and probed with the methyl-glutamine antibody. To rule out non-specific binding of the antibody to unmethylated proteins, increasing amounts of the recombinant unmethylated protein substrate were included as negative controls (Figure 21).

A HEK293

CHD5	+	+	recombinant CHD5 Control
HEMK2+TRM112	+	-	

kDa
70 —
55 —

70 —
55 —

B HEK293

NUT	+	+	recombinant NUT Control
HEMK2+TRM112	+	-	

kDa
70 —
55 —

70 —
55 —
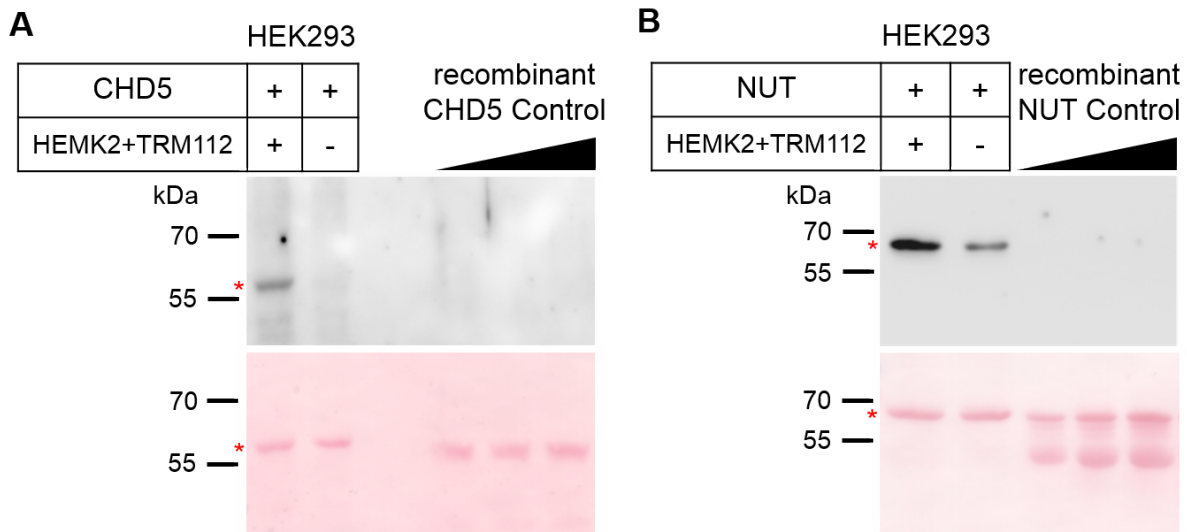
Figure 21: *Detection of glutamine methylation performed by HEMK2 in HEK293 cells. (A) The YFP-fused substrate CHD5 was ectopically expressed with (T) or without (S) HEMK2 and TRM112. CHD5 was purified by GFP-Trap® A shown in Figure 20A, recombinant unmethylated GST-fused CHD5 substrate was utilized as negative control for specificity of the antibody (lower panel). Cellular glutamine methylation was determined by Western blot with the methyl-glutamine specific antibody (upper panel). The Ponceau S staining revealed the loading of proteins. (B) Cellular glutamine methylation of YFP-fused NUT substrate expressed in HEK293 cells. The experiment was conducted as described in A. As a negative control recombinant unmethylated NUT protein domain was used. The corresponding protein bands are marked with a red asterisk.*

The methyl-glutamine antibody bound specifically to substrate proteins purified from HEK293 cells that were coexpressed with HEMK2-TRM112, but not to the corresponding recombinant unmethylated protein domains (Figure 21). No signal was detected with the recombinant unmethylated proteins even with higher concentrations. For the CHD5 substrate a methylation signal was only detected after coexpression with HEMK2 and TRM112, but not when CHD5 was expressed alone, indicating CHD5 is methylated by HEMK2 in human cells (Figure 21A). A similar result was observed for NUT. Although in this case, a weak signal was detected for the NUT substrate expressed in HEK293 cells without coexpression of HEMK2 and TRM112. However, a significantly increased methylation signal was observed when NUT was coexpressed with HEMK2 and TRM112, indicating a HEMK2-dependent methylation *in vivo* (Figure 21B). The weaker methylation signal for NUT in the absence of HEMK2 and TRM112 in HEK293 cells, could be attributed to endogenous HEMK2. Taken together, the HEMK2-dependent methylation of two novel substrate proteins (CHD5 and NUT) in human cells was shown. Additionally, HEMK2 methylated at least 11 new substrates at protein level and approximately 120 further peptides substrates *in vitro*.

3.2 Characterization of the Substrate Specificity of the Histone Lysine Methyltransferase, NSD2

The nuclear receptor SET domain-containing protein 2 (NSD2) is a member of the NSD family of PKMTs. NSD2 is a histone lysine methyltransferase that has been reported to methylate H3K36 as well as other residues, such as K4 and K27 of histone H3 or K20 and K44 of histone H4. Although the methylation of most targets is still debated, the dimethylation of H3K36 has been repeatedly described[107,128].

3.2.1 Purification and Assessment of Methyltransferase Activity

Similar to HEMK2, no crystal structure of the NSD2 SET domain is currently available, which could provide more detailed knowledge about the recognition, discrimination and interaction between the enzyme and its peptide substrates.

Since previous attempts to purify the catalytically active SET domain (without the post-SET domain) of NSD2 (residues 1074 - 1182) failed, three different constructs containing the SET domain flanked by the AWS and post-SET domain with variable domain boundaries were cloned in this study (Figure 22).
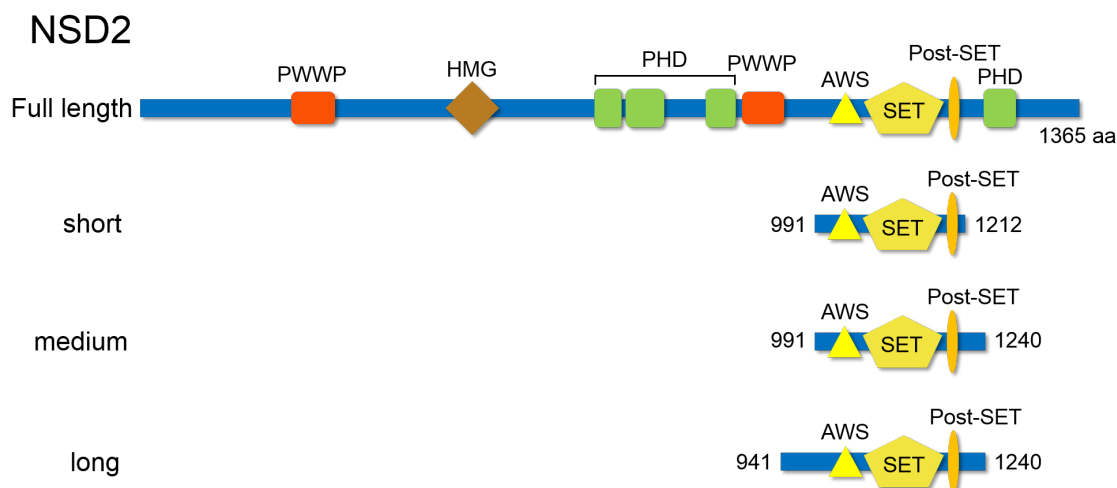
Figure 22: *Schematic representation of the three cloned NSD2 SET domain constructs. Short (amino acids 991 - 1212), medium (amino acids 991 - 1240) and long (amino acids 941 - 1240) constructs containing functional domains: PWWP domain; PHD zinc-finger domain; SET lysine methyltransferase (KMT); AWS domain (associated with SET domain); post-SET domain, HMG box.*

GST-fused NSD2 SET domains were expressed in *E. coli* cells and purified by affinity chromatography with good yield and purity (Figure 23).
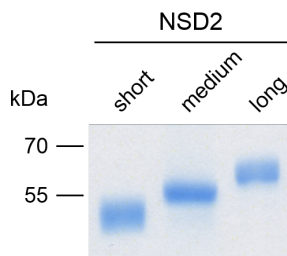
Figure 23: *Quality of the GST-fused NSD2 SET domain constructs. Coomassie stained SDS-PAGE gel of the three purified NSD2 constructs.*

The methyltransferase activity of the three recombinant NSD2 enzymes was tested on peptide SPOT arrays. H3 (29-43) peptides, containing the K36 site, were synthesized on a cellulose membrane using the peptide SPOT synthesizer. Additionally, peptides with a target lysine to alanine mutation were included as negative controls. Methylation reactions were performed by incubating the peptide arrays in methylation buffer containing the purified NSD2 enzymes and radioactively labeled [methyl-^{3}H]-SAM. The transfer of methyl groups was detected by autoradiography (Figure 24).
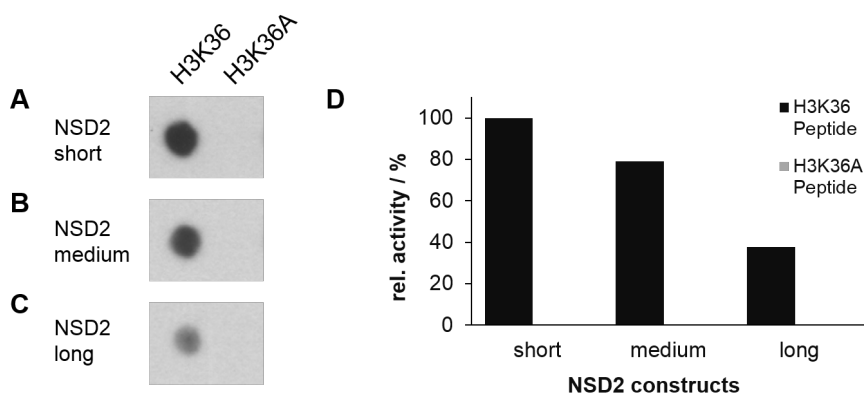
Figure 24: *Determination of methyltransferase activity of the three NSD2 enzyme constructs. The membrane contained histone H3 (APATGGV**K**KPHRYRP) and K36A (APATGGV**A**KPHRYRP) peptides methylated by the short NSD2 construct (A), medium NSD2 construct (B) or the long NSD2 construct (C). The images were taken from the same autoradiography film. (D) Bar diagram represents the quantitative analysis of the methylation images shown in (A), (B) and (C).*

As shown in Figure 24 all three purified NSD2 constructs exhibited methyltransferase activity on the H3 wild-type peptides. This was lost in peptides containing the target K to A mutations. Additionally, the short NSD2 construct seems to be the most active, followed by the medium and long NSD2 construct (Figure 24A, B and C). Though the activity of the medium NSD2 construct was slightly weaker than that of the short one, it could be purified with better yield, so all further experiments were performed with the medium NSD2 construct and from now it is referred as

„NSD2" enzyme. Next, the methyltransferase activity of NSD2 on peptides containing H4K44 (37 - 51) and H1.5K168 (161 - 175) was determined. It was previously shown that these lysine residues are methylated by NSD2[107] (methylation of H1.5K168 peptide by NSD2 was shown in the doctoral thesis of Dr. Qazi M. Raafiq[163]). As before, the peptide array contained H3K36 peptides as positive control and all corresponding target lysine to alanine mutant peptides as negative controls (Figure 25A).
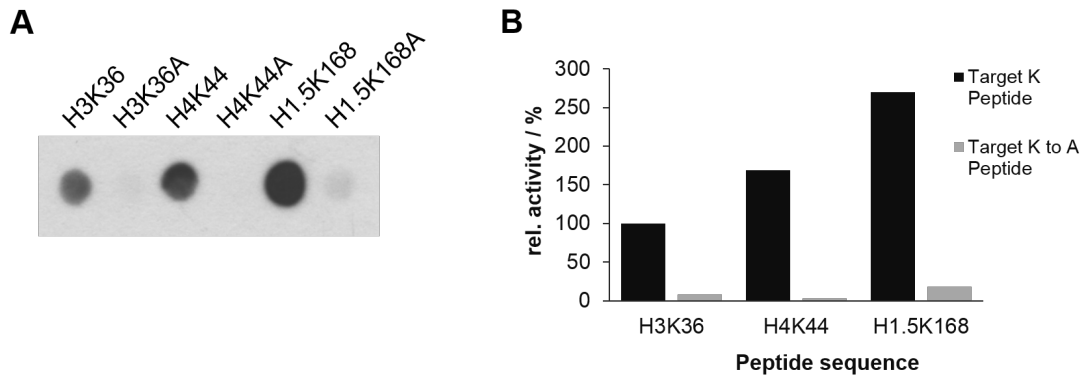
Figure 25: *Methyltransferase activity of NSD2 on peptide array. (A) Autoradiography of peptide array methylation by the medium NSD2 construct. The membrane contained histone H3K36 (APATGGV**K**KPHRYRP), K36A (APATGGV**A**KPHRYRP), H4K44 (LAR-RGGV**K**RISGLIY), K44A (LARRGGV**A**RISGLIY), H1.5K168 (KPAAAGV**K**KVAKSPK) and K168A (KPAAAGV**A**KVAKSPK) peptides. (B) Bar diagram represents the quantitative analysis of the autoradiography shown in (A).*

The results revealed methyltransferase activity of NSD2 toward the wild-type peptides of H3K36, H4K44 and H1.5K168 (Figure 25A). No activity was observed on the corresponding lysine to alanine mutant peptides. Interestingly, quantitative analysis showed a 2.7 and 1.7 times higher activity of NSD2 on the H1.5K168 and H4K44 peptides than on the H3K36 peptides (Figure 25B).

To confirm the methyltransferase activity of NSD2 on histone proteins as well, methylation assays with recombinant histone proteins H3.1 and H4 (NEB) were performed. After methylation the samples were separated by SDS-PAGE and subjected to autoradiography. Similar to the peptide array methylation result, the methylation signal of the recombinant H4 protein was stronger than that of H3.1 (Figure 26).
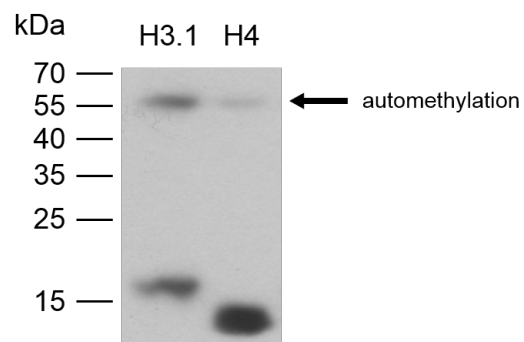
Figure 26: *Methyltransferase activity of NSD2 on histone proteins. Autoradiography of the methylated recombinant histone proteins H3.1 and H4 by NSD2. Bands marked with an arrow show automethylation of NSD2.*

3.2.2 Determination of the Specificity Profile of NSD2

To determine the substrate recognition specificity of histone lysine methyltransferase NSD2, specificity profile arrays were synthesized based on the H3K36 substrate sequence (residues 29‑43) by the peptide SPOT synthesizer. Synthesized peptides contained one single amino acid mutation, where one residue of the original sequence was exchanged by another amino acid. This allows to investigate methylation of all possible single amino acid mutant peptides. In total, 288 peptides were synthesized that probed 18 proteinogenic amino acids (cysteine and tryptophan were not included) at each site. Methylation reactions were performed by incubating the peptide arrays with NSD2 in methylation buffer containing radioactively labeled [methyl-^{3}H]-SAM (Figure 27A). Three independent methylation experiments were performed and the results of each methylation assay were analyzed by the Phoretix^{TM} Array software. Normalization and the color-coding (Figure 27B) was conducted as described in section 3.1.2.
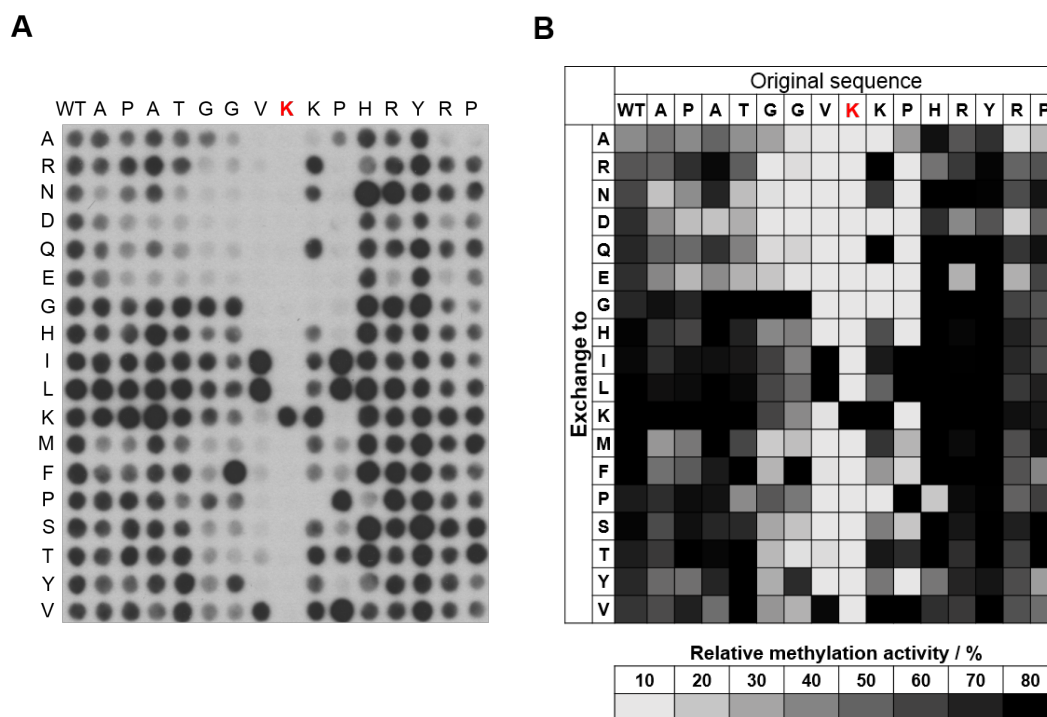
Figure 27: *Substrate specificity profile of NSD2. (A) Autoradiography of a specificity profile array based on H3K36 sequence (29‑43) methylated by NSD2. The horizontal axis represents the original H3 sequence and the target lysine K36 is highlighted red. The vertical axis shows the residues exchanged in the corresponding row. The first column contains the wild-type sequence of H3 used as a control labeled with WT. (B) Data from three experiments were normalized, averaged and the results were color-coded depending on their methylation activity. Black to light gray represents a strong to weak methylation.*

To determine the quality of the results, standard deviations of the averaged spot methylation signals (SD) were calculated. Approximately 60 % of the peptides showed an SD smaller than 10 % and about 90 % of the peptides had an SD less than 20 %, which reveals a good quality of

the data (Figure 28A). In addition, the discrimination factor for the recognition of each amino acid at each single position was calculated (Figure 28B).
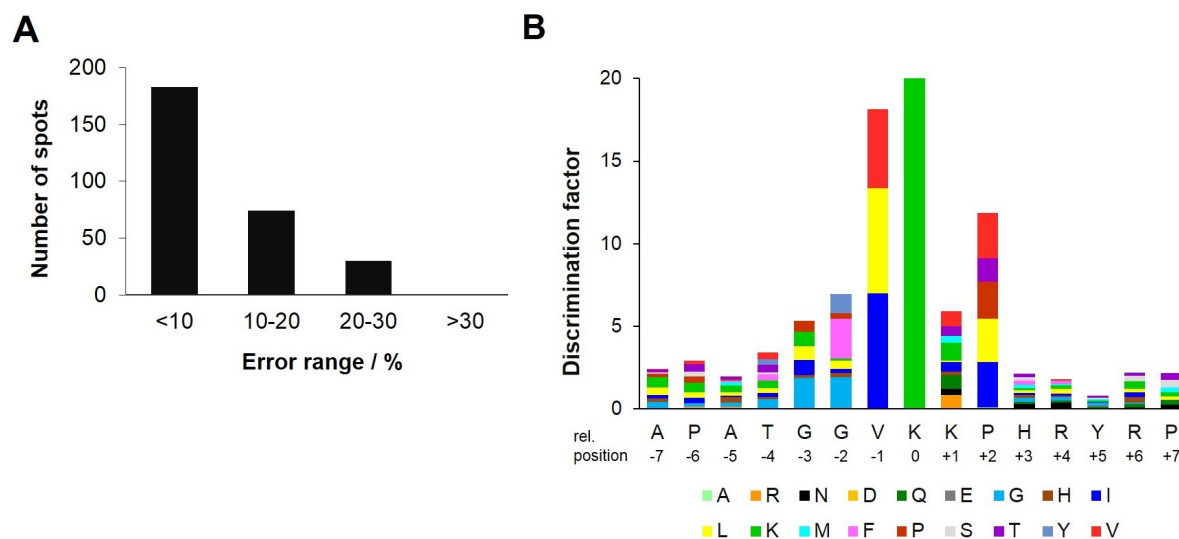
Figure 28: *Evaluation of the specificity profile results. (A) Quality control of the peptide intensities derived from the specificity profile arrays of NSD2 shown in Figure 27B. Standard deviation of the averaged NSD2 activity on all peptide substrates was calculated. (B) The discrimination factors for the recognition of each amino acid at the corresponding position of the H3 substrate by NSD2 represented in a bar diagram.*

The specificity analysis shows that NSD2 has a rather narrow recognition range, which spans the H3 residues from 33 to 38 (Figure 27 and 28). The results reveal a preference for aromatic but also for small amino acids at the -2 position (F>G>Y>I, L, P, H). At the -1 position only large aliphatic amino acids (I>L>V) are allowed, whereas at the +1 site many residues are tolerated. At the +2 position exclusively hydrophobic residues are preferred (V>I>L>P>T). Further to the N-terminus, the glycine at -3 is favored and bulky and acidic amino acids are avoided. Based on this, two substrate recognition motifs of NSD2 were proposed in Table 3.

Table 3: *NSD2 substrate specificity profiles utilized to search for putative novel NSD2 substrates.*

Cognate residue Position	V35 -1	K36 0	K37 +1	P38 +2	H39 +3	R40 +4	Y41 +5
Search profile 1 (stringent)	ILV	K	KR	VILP	NGLSFTM IHQAEK	LNQG HIKMF	QEGHI
Search profile 1 (relaxed)	ILV	K	KR VQN	VILP	NGLSFTM IHQAEK	LNQG HIKMF	QEGHI

3.2.3 Identification of Putative NSD2 Peptide Substrates

The specificity profiles of NSD2, which are described in Table 3, were used to search in the Scansite[162] website for putative novel protein substrates of NSD2 in the human proteome. The search profile 1, which is more stringent at the +1 position, identified 114 potential target lysine sites in 110 proteins. With the more relaxed search profile 2 additional 112 potential target lysine sites in 107 proteins were identified. Since NSD2 is present in the nucleus, the search was restricted to nuclear proteins[164]. A peptide array was synthesized containing 15 amino acid long peptides of the 226 identified putative substrate candidates using the peptide SPOT synthesizer. A detailed list of the protein names and peptide sequences of the putative novel substrates is provided in Table 15 in section 6.2.2. The membrane was incubated with NSD2 to investigate the transfer of radioactively labeled methyl groups by autoradiography. Quantitative analysis
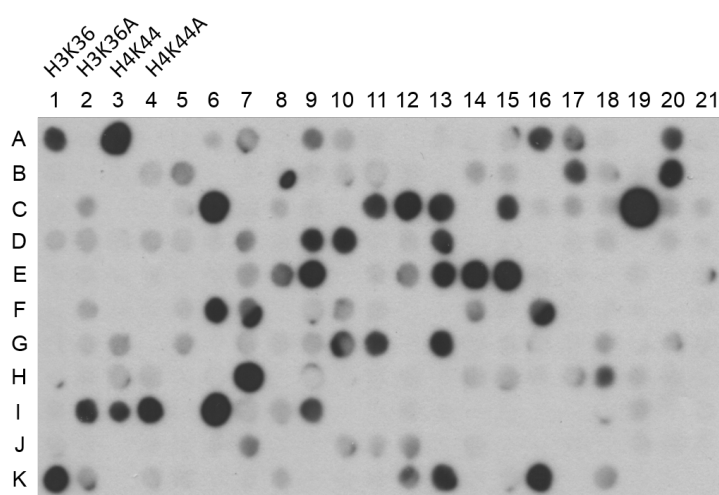
Figure 29: *Methylation of novel peptide substrates by NSD2. The image shows an autoradiography of the methylated novel peptide substrates. 15 amino acid long peptides identified by the search profiles 1 and 2 (Table 3) containing the predicted target lysine in the middle were synthesized on a peptide array. Protein names and peptide sequences are listed in Table 15 in section 6.2.2. H3K36 wild-type and H3K36A mutant peptides as well as H4K44 wild-type and H4K44A mutant peptides were included as controls.*

of the methylated peptide array revealed 45 methylated peptides (in addition to the H3K36 and H4K44 controls), 19 of them were strongly methylated, 15 had approximately the same intensity as the H3K36 control and 13 peptides showed weaker methylation signals (Figure 29). Therefore, NSD2 methylated ~20 % of the identified putative novel peptide substrates with a good activity, which verifies the specificity profile.

3.2.4 *In vitro* Methylation of the Putative Protein Substrates

It was demonstrated in the last section that NSD2 is able to methylate at least 45 novel peptide substrates. Since the 15 amino acid long peptides possess no tertiary structure, a similar methy-

lation of the target lysine cannot be automatically expected to occur in proteins as well. This is because in proteins, the hydrophobic residues of the recognition sequence might be shielded inside the folded structure and not accessible for the enzyme.

To investigate if NSD2 is able to methylate the previously identified peptide substrates also at protein level, 27 of the most strongly methylated peptide substrates were selected and cloned as proteins domains containing the target lysine (Table 15 in section 6.2.2). The domains were overexpressed as GST fusion proteins and purified by affinity chromatography. Out of the 27 chosen putative substrates, 22 could be purified in a sufficient yield (Figure 30A). The identity of the substrate protein candidates was confirmed by immunoblotting with an anti-GST antibody (Figure 30B). The remaining 5 substrates failed during different steps from the cloning to purification stage.
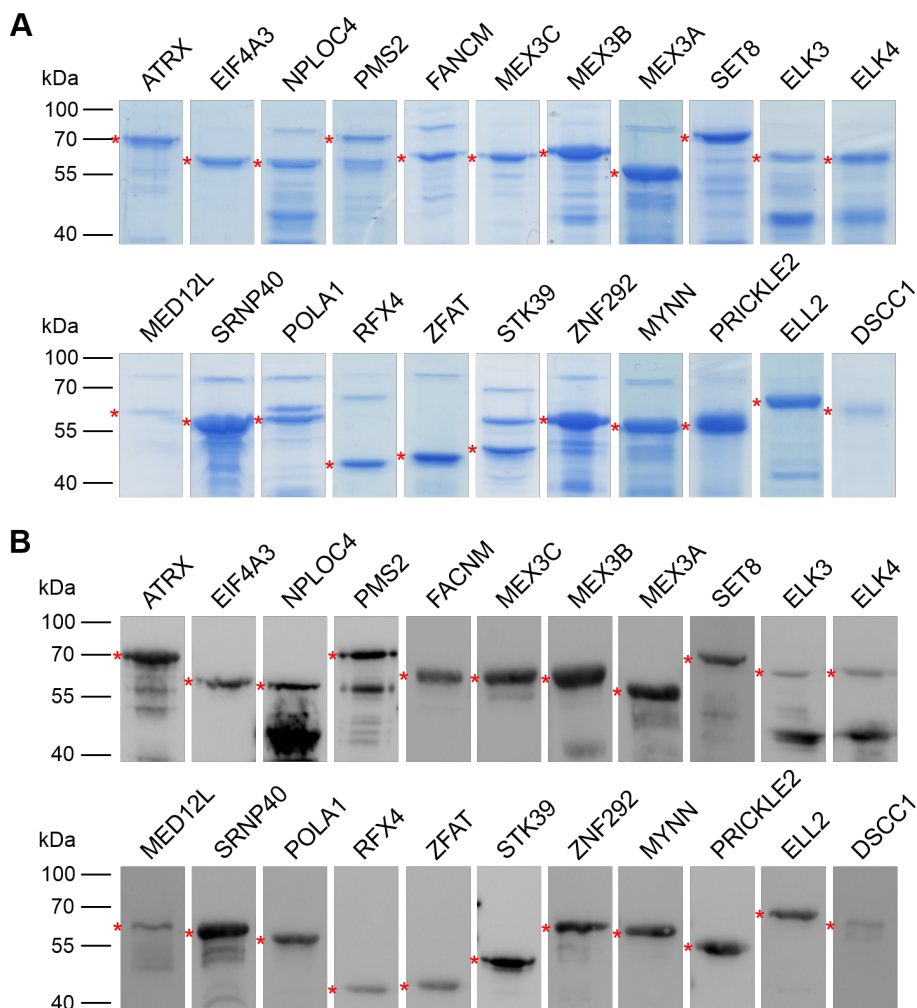
Figure 30: *Purification and confirmation of putative NSD2 substrates. (A) Coomassie stained SDS-PAGE gels of the 22 purified GST-fused substrate candidates. (B) Confirmation of the identity of the purified substrate proteins by western blot with anti-GST antibody. The corresponding bands with the expected size are marked with a red asterisk.*

For the following methylation experiments, the concentration of the 22 purified putative substrate proteins was determined by Coomassie staining of SDS-PAGE gels and equal amounts of proteins were used for methylation reactions (Figure 31A). The methylation assays with the 22 purified putative protein substrates were performed by incubating the proteins with NSD2. GST-fused H3 substrate protein was included as positive control. Additionally, a sample containing only NSD2 was prepared without additional substrate protein. The methylation samples were separated by SDS-PAGE and subjected to autoradiography. The autoradiography images
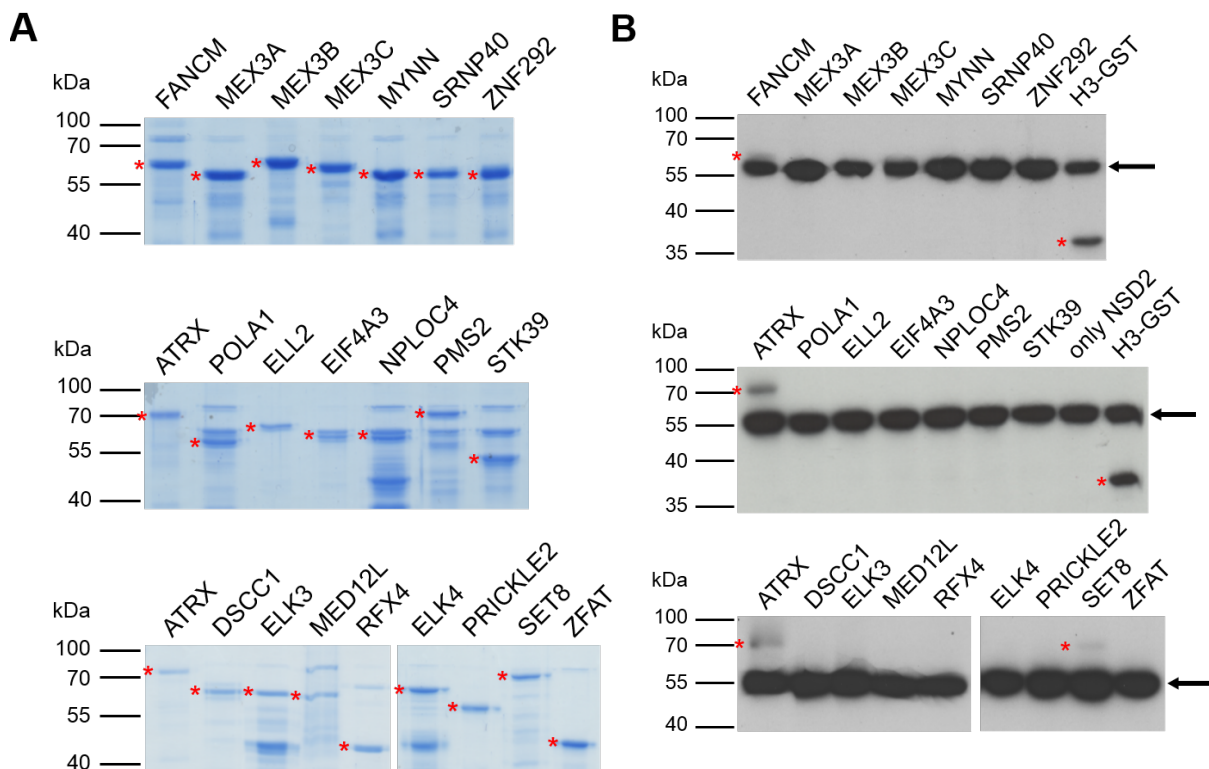
Figure 31: *Methylation assays of the purified putative protein substrates. (A) Coomassie stained SDS-PAGE gels of the substrate candidates (left panels). The same amounts were used for the methylation reactions. (B) Autoradiography images of the potential protein substrates methylated by the NSD2 medium construct (right panels). The corresponding bands with the expected size are marked with a red asterisk. Automethylation of NSD2 is indicated by an arrow.*

showed a clearly visible methylation signal for ATRX and weaker signals for FANCM and SET8 (very weak) (Figure 31B). As expected, the GST-fused H3 protein revealed a strong methylation signal. Interestingly, much stronger bands appear in all methylation assays at the same height. Even the methylated sample containing only NSD2 enzyme revealed a strong methylation activity, suggesting the signal represents an automethylation of NSD2 (marked by arrows). Unfortunately most of the putative protein substrates have a similar size as the NSD2 medium construct, therefore the automethylation bands of NSD2 might cover signals from the methylated protein substrates.

To discriminate the potential methylation signal of substrate proteins from the automethylation of NSD2, the methylation assays were repeated with same substrate candidates and the short NSD2 construct, as described above. A weak but clear methylation signal corresponding to the
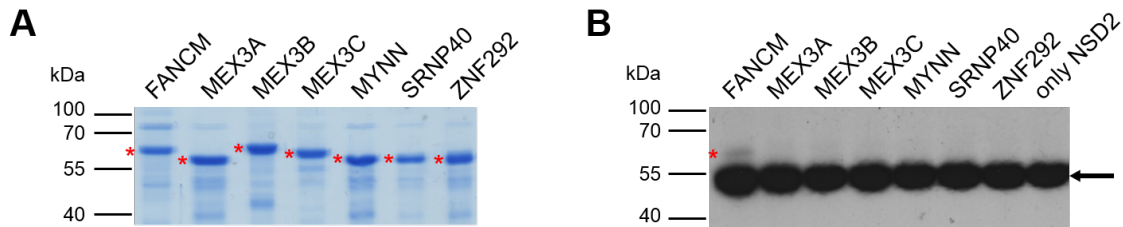
Figure 32: *Methylation assays of the selected putative protein substrates for methylation by the short NSD2 enzyme construct.* *(A) Coomassie stained SDS-PAGE gel of the normalized substrate candidates same as the gel in Figure 31A top. Similar amounts were used for the methylation reactions. (B) Autoradiography of the selected substrate candidates methylated by the short NSD2 construct. The corresponding bands with the expected size are marked with a red asterisk. Automethylation of NSD2 is indicated by an arrow.*

size of the FANCM protein was observed, which is not caused by the automethylation of NSD2 (Figure 32B). However, FANCM was the only band observed in this methylation assay. But, based on the protein sizes it cannot be excluded that the automethylation band from NSD2 still covers potential signals from the methylated target proteins.

To identify the automethylation site of NSD2, a peptide scan array of the NSD2 (941 - 1243) sequence was synthesized with 15 amino acid long peptides always shifted by 5 amino acids. A list of the peptide sequences is given in Table 16 in section 6.2.2. The peptide array was methylated by NSD2 and the transfer of the radioactive methyl groups was detected by autoradiography (Figure 33).
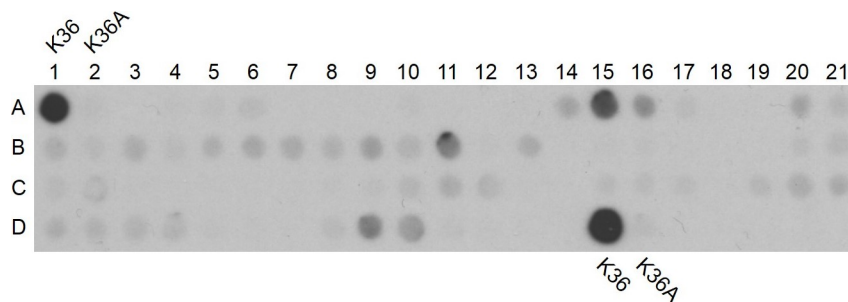
Figure 33: *Methylation of a peptide scan array to identify the NSD2 automethylation site.* *Autoradiography of the peptide scan array methylated by NSD2 containing 15 amino acid long peptides with the sequence of NSD2. The peptide sequences of each spot are listed in Table 16 in section 6.2.2. The H3K36 wild-type and K36A mutant peptides were included as controls.*

H3K36 control peptides (A1 and D15) showed strong methylation signals. Additionally, a weaker but still clearly detectable signal at peptide spot A15 was observed (Figure 33). The methylated NSD2 peptide contained lysine residue K992. Interestingly, the corresponding residue was identified to be automethylated in NSD1 as well (Doctoral thesis of Dr. Srikanth Kudithipudi[165]).

In an attempt to remove the automethylation activity of NSD2, site-directed mutagenesis was performed to exchange lysine 992 to arginine or alanine. The GST-fused NSD2 enzymes containing either the K992R or K992A mutation were overexpressed and purified by affinity chromatography (Figure 34).
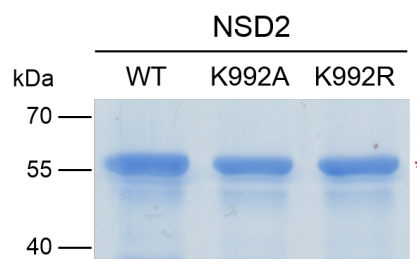
Figure 34: *Quality of the purified NSD2 K992A and K992R mutants. Coomassie staining of the purified NSD2 enzyme mutations. The corresponding bands of the expected size are marked with a red asterisk.*

The purified NSD2 mutant enzymes were used for methylation of ATRX or recombinant H3.1 as previously described. Methylated samples were separated by SDS-PAGE and the transfer of the radioactive labeled methyl groups was detected by autoradiography (Figure 35).
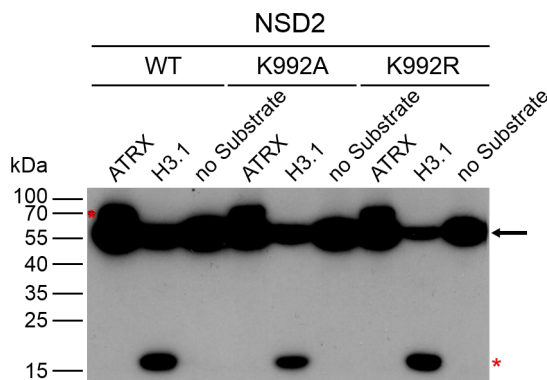
Figure 35: *Methyltransferase activity of the NSD2 K992A and K992R mutants. Methylation of ATRX and H3.1 by NSD2 wild-type, K992A and K992R mutants. The image shows an autoradiography of methylated ATRX and recombinant H3.1 substrates. The corresponding bands of the expected size are marked with a red asterisk. Automethylation of NSD2 is indicated by an arrow.*

As observed in Figure 35, the two NSD2 mutants were able to methylate ATRX and recombinant H3.1, but still showed strong automethylation bands. This was comparable to wild-type NSD2 indicating that lysine K992 is not the true or only automethylation site of NSD2. As it seems, non-specific automethylation on lysine or even other amino acid residues in close proximity to the SAM binding site is not trivial to prevent.

Therefore, NSD2 was subcloned in the pET-28a(+) and the pMAL-c2x expression vectors, which produce NSD2 either with the small His_6-tag or with the larger MBP-tag (maltose-binding protein). The His_6- and MBP-fused NSD2 enzymes were overexpressed and purified by affinity chromatography (Figure 36A). The purified NSD2 enzymes were used in *in vitro* methylation assays with ATRX and recombinant H3.1. The methylation samples were separated by SDS-PAGE and subjected to autoradiography. As expected, strong methylation signals were observed
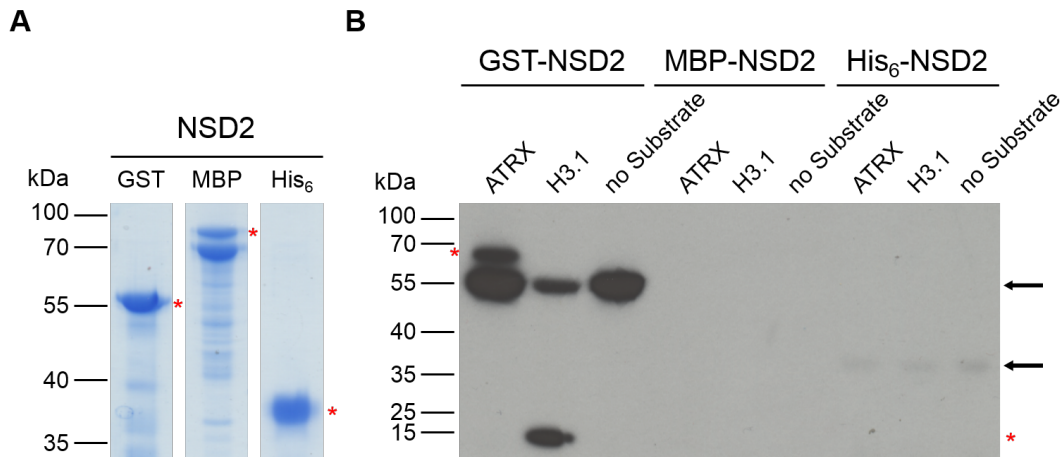
Figure 36: *Methylation of ATRX and H3.1 protein by NSD2 with different affinity tags. (A) Coomassie staining of the purified GST-, MBP- and His$_6$-fused NSD2 enzymes. (B) Autoradiography image of methylated ATRX and recombinant H3.1 substrates by NSD2 enzymes fused either to GST-, MBP- or His$_6$-tag. The corresponding bands of the expected size are marked with a red asterisk. Automethylation of NSD2 is indicated by an arrow.*

for ATRX and H3.1 after methylation by the GST-fused NSD2 enzyme (Figure 36B). Although the MBP- and His$_6$-fused NSD2 variants were purified with a good yield, the enzymes did not show considerable activity. Additionally, an automethylation signal was detected for the His$_6$-tagged NSD2. This was much weaker than the automethylation of the GST-tagged enzyme. Since preventing automethylation of NSD2 is not trivial, in the next steps the potential protein substrates could be subcloned into the pET-28a(+) vector to obtain the substrate proteins and GST-fused enzyme with different sizes and avoid the overlapping of NSD2 automethylation signal with the substrate protein methylation signal.

To confirm the methylation of the three newly methylated substrates on the predicted lysine residues, site-directed mutagenesis was performed to exchange the predicted target lysine to arginine (Table 14 in section 6.2.1). The mutant proteins were overexpressed and purified by affinity chromatography (Figure 37A). The purified wild-type and corresponding mutant proteins were methylated by NSD2. For the methylation reactions of ATRX and SET8 the medium NSD2 construct and for the methylation of FANCM the short NSD2 construct was used. The results showed that the methylation signals on the corresponding mutant proteins of ATRX and FANCM disappeared, whereas the wild-type proteins revealed clear methylation signals (Figure 37B). However, the SET8 mutant protein still revealed a methylation signal, which was even stronger than for the wild-type protein. Since, SET8 is also a PKMT it was checked if this signal was due to its potential automethylation. *In vitro* methylation assays for SET8 were repeated as described above and additionally two control methylation reactions were included: the SET8 wild-type or mutant protein without NSD2 enzyme (Figure 38).
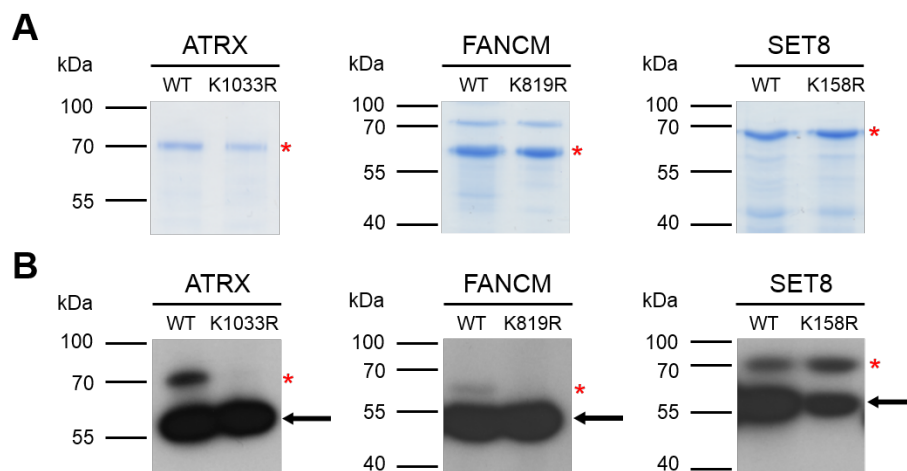
Figure 37: *Confirmation of methylation on the predicted lysine. Purified wild-type (WT) and mutant (K to R) protein substrates of ATRX, FANCM and SET8 were methylated by the NSD2 enzymes. (A) Equal amounts of wild-type and mutant proteins was verified by Coomassie staining of the SDS-PAGE gels (upper panels). (B) Autoradiography images of the methylated protein substrates (lower panels). The corresponding bands of the expected size are marked with a red asterisk. Automethylation of NSD2 is indicated by an arrow.*
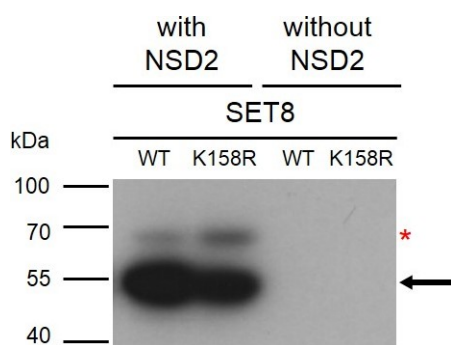
Figure 38: *Methylation of the SET8 wild-type and mutant proteins. Autoradiography image of the methylation of SET8 wild-type and mutant proteins by NSD2. The wild-type and mutant proteins of SET8 were incubated either with or without NSD2 to test for automethylation of SET8. The corresponding bands of the expected size are marked with a red asterisk. Automethylation of NSD2 is indicated by an arrow.*

As previously seen a weaker methylation signal of SET8 wild-type and a stronger methylation signal on SET8 K158R mutant protein was observed when the proteins were incubated with NSD2, but no methylation signal on the SET8 wild-type and mutant proteins was noticed in the absence of NSD2 enzyme (Figure 38). This methylation assay confirmed that the observed bands are not caused by automethylation of SET8, rather that NSD2 methylates another or an additional lysine residue within the SET8 sequence, besides the predicted target lysine K158.

All in all, it was shown that NSD2 specifically methylates two novel substrate (ATRX and FANCM) at protein level. NSD2 might methylate even more of the purified substrate candidates, however, due to its strong automethylation it is not trivial to discern the signals, so investigation of other purified substrate proteins was discontinued in the current study. Additionally, methylation of SET8 was observed, although the target residue could not been identified.

3.2.5 Cellular Methylation of the Novel Target Substrates

It was shown that NSD2 methylates two of the predicted protein substrates *in vitro* (Figure 31). To check if these methylated substrates are also methylated in HEK293 cells, ATRX and FANCM were subcloned into the mammalian expression vector pEYFP-C1. The coding sequence of the full-length (FL) NSD2 enzyme was also subcloned into the mammalian expression vector pECFP-C1.

To develop an appropriate tool to determine the cellular methylation, the specificity of an anti-H3K36me2 antibody (Active Motif, USA; Cat. no.: 39255) was investigated. This was selected, because NSD2 dimethylates H3K36 and the methylated sequences in ATRX and FANCM resemble the H3K36 sequence. To investigate if the anti-H3K36me2 antibody recognizes the *in vitro* methylated substrates, GST-fused histone H3 or recombinant H3.1 proteins were methylated by NSD2 in methylation buffer containing unlabeled SAM. The protein samples were separated by SDS-PAGE and transferred onto nitrocellulose membrane and then probed with the anti-H3K36me2 antibody (1 : 1.000). However, no signal was detected (data not shown). Since the used enzyme construct consists only out of the SET domain and not of the full-length protein, it might be that NSD2 is just able to monomethylate the protein substrates. Methylation assays were repeated as described above and the membrane was probed with an anti-H3K36me1 antibody (1 : 2.000) (Abcam, UK; Cat. no.: ab9048). Figure 39 shows a clear signal for both the GST-fused histone H3 and the recombinant H3.1 proteins, respectively, but no signal for the sample without a protein substrate.
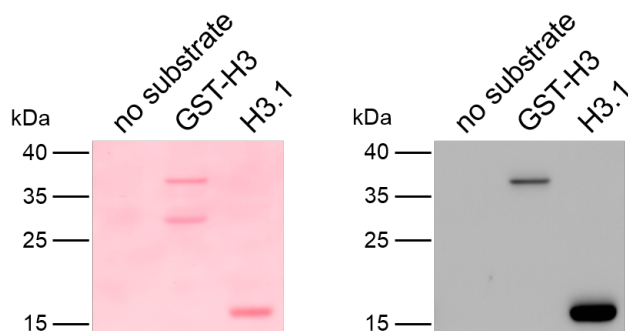
Figure 39: *Validation of the anti-H3K36me1 antibody with methylated histone H3 substrates. Ponceau S staining of the nitrocellulose membrane (left panel). Western blot of the transferred GST-fused H3 and the recombinant H3.1 substrates detected by anti-H3K36me1 antibody (right panel).*

To investigate if the anti-H3K36me1 antibody is also able to detect the methylation of the novel NSD2 substrates ATRX and FANCM, similar *in vitro* methylation reactions were performed by incubating either ATRX or FANCM with or without NSD2 enzyme in methylation buffer containing unlabeled SAM. The methylation samples were separated by SDS-PAGE and transferred onto nitrocellulose membranes. The membrane was probed with the anti-H3K36me1 antibody

$(1 : 2.000)$ (Figure 40).
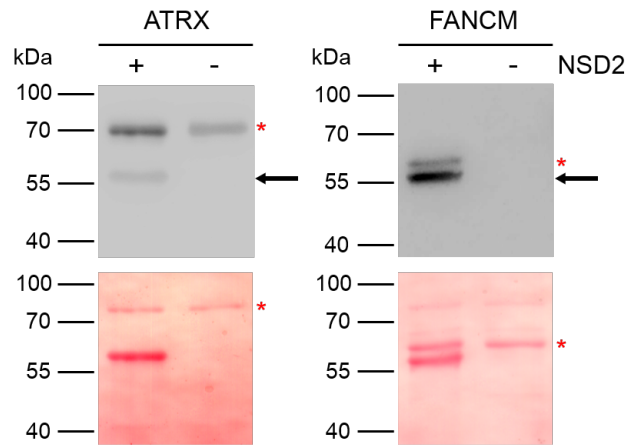
Figure 40: *Validation of the anti-H3K36me1 antibody with methylated and unmethylated ATRX and FANCM proteins. Ponceau S staining of the methylated and unmethylated protein substrates (lower panel). Western blot of the transferred methylated and unmethylated protein substrates probed with the anti-H3K36me1 antibody (upper panel). The corresponding bands of the expected size are marked with a red asterisk.*

A clear signal was observed for the methylated FANCM protein, only when NSD2 was present. For ATRX the H3K36me1 antibody was not able to discriminate fully between methylated and unmethylated proteins (Figure 40). However, the signal for methylated ATRX protein was much stronger than for the unmethylated protein. ATRX and FANCM were selected for further cellular studies, because the antibody revealed a good discrimination between methylated and unmethylated proteins. The expression levels of the CFP-fused full-length NSD2 enzyme (NSD2 FL) and the two YFP-fused substrates ATRX and FANCM was tested. HEK293 cells were individually transfected with each of the three plasmids and harvested three days after transfection. The cells were lysed, the proteins separated by SDS-PAGE and transferred onto nitrocellulose membranes. The membranes were probed with an anti-GFP antibody (Clontech, USA) to determine the expression of the corresponding proteins (Figure 41).
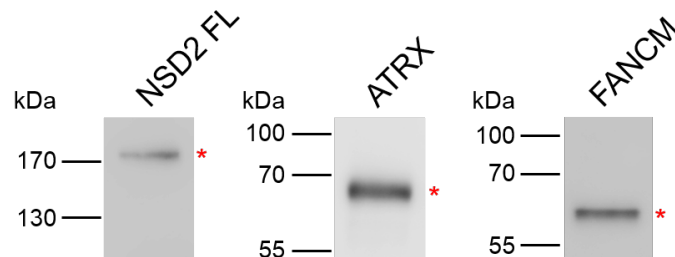
Figure 41: *Immunoblot detection of protein expression in HEK293 cells. The cells were transfected either with NSD2 FL, ATRX or FANCM. Cells were harvested, lysed, separated by SDS-PAGE and transferred onto nitrocellulose membranes. The expressed proteins were detected by probing with an anti-GFP antibody. The corresponding bands of the expected size are marked with a red asterisk.*

Western blot results revealed the appearance of signals for all the three proteins (Figure 41) at the expected size, indicating a successful expression. To investigate if the two protein substrates are methylated by NSD2 *in vivo*, HEK293 cells were transiently transfected with one of the YFP-fused substrate proteins either with or without the full-length NSD2 enzyme. The cells were harvested three days after transfection and the YFP-fused substrate proteins were purified using GFP-Trap® A beads. Approximately 10 % of the purified substrate proteins samples were separated by SDS-PAGE, transferred onto nitrocellulose membranes and probed with an anti-GFP antibody to provide a loading control (Figure 42).

ATRX-YFP	+	+
NSD2 FL	+	-

FANCM-YFP	+	+
NSD2 FL	+	-
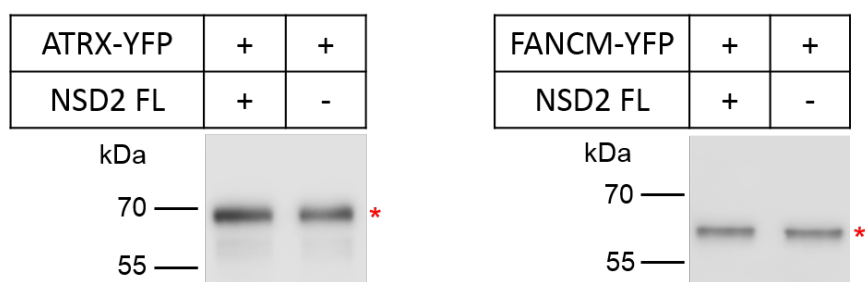
Figure 42: *Immunoblot detection of protein expression in HEK293 cells. The cells were transfected with YFP-fused ATRX or FANCM either with or without NSD2. Cells were harvested, lysed and the substrate proteins were purified by GFP-Trap®. Approximately 10 % of the purified target proteins were separated by SDS-PAGE and transferred onto nitrocellulose membranes. The expression of the protein substrates was analyzed by probing with an anti-GFP antibody. The corresponding bands of the expected size are marked with a red asterisk.*

The adjusted amounts of the target proteins ATRX or FANCM, were loaded on an SDS gel and separated by SDS-PAGE, transferred onto nitrocellulose membranes and detected by western blot with the anti-H3K36me1 antibody (1 : 500). Recombinant GST-fused unmethylated substrate proteins with increased concentration were included to rule out possible unspecific binding of the antibody to unmethylated protein (Figure 43). The results clearly showed that the anti-H3K36me1 antibody binds specifically to the substrate proteins that were expressed together with the lysine methyltransferase NSD2, but not to the recombinant GST-fused unmethylated protein domains (Figure 43). Even with higher concentrations of recombinant unmethylated substrate protein no signal was observed. For both protein substrates, ATRX and FANCM, a strong methylation signal was detected when they were coexpressed with the NSD2 enzyme, indicating a NSD2-dependent methylation in human cells. However, when the substrates were expressed in the absence of NSD2, only a very weak signal was observed. The weaker methylation signal, in absence of ectopically expressed NSD2, could be probably due to the endogenous enzyme. In total, two novel substrate proteins, ATRX and FANCM, were identified, which are methylated in human cells by NSD2. Additionally, it was shown that NSD2 methylates 43 further substrate candidates at peptide level and 3 substrates at protein level.
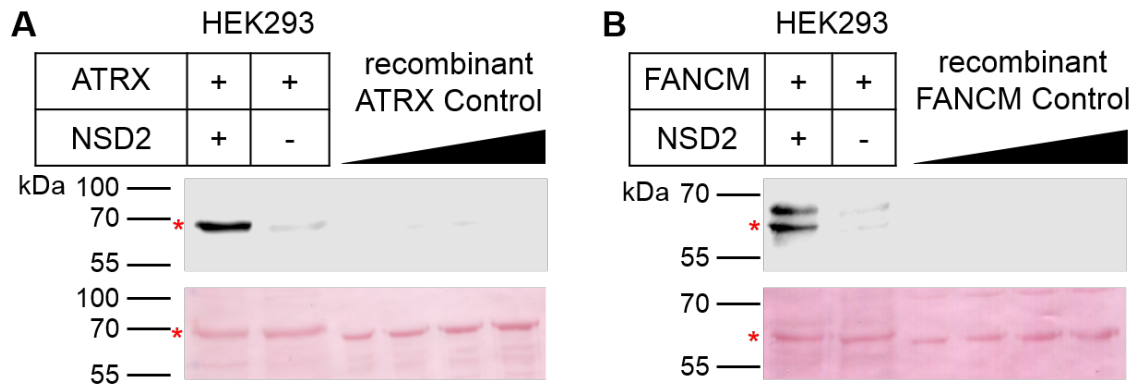
Figure 43: *Immunoblot detection of lysine methylation performed by NSD2 in HEK293 cells. (A) The YFP-fused substrate ATRX was ectopically expressed with or without NSD2. ATRX was purified by GFP-Trap® as shown in Figure 42. Cellular lysine methylation was determined by western blot with the anti-H3K36me1 antibody (upper panel). The Ponceau S staining revealed the loading of recombinant unmethylated GST-fused ATRX substrate as negative control for the specificity of the antibody (lower panel). (B) Cellular lysine methylation of YFP-fused FANCM substrate expressed in HEK293 cells. The experiment was conducted as described in A. As a negative control recombinant unmethylated FANCM protein domain was used. The corresponding bands of the expected size are marked with a red asterisk.*

3.2.6 Somatic Cancer Mutations of NSD2

Jaffe *et al.* identified in a global chromatin profiling analysis, a novel NSD2 mutation (E1099K) that was highly prevalent in ALL cell lines lacking the NSD2 t(4;14) translocation[143]. The glutamic acid E1099 resides within the catalytically active SET domain of NSD2. It was assumed that the E1099K mutation could alter the interaction between NSD2 and its substrate as it is located in a loop close to the substrate binding pocket. It was shown that the recombinant NSD2 E1099K enzyme has higher *in vitro* methylation activity on nucleosomes, compared to the wild-type NSD2. Furthermore, an increased level of dimethylated H3K36 was observed in cell lines containing the E1099K mutant, suggesting a hyperactivity of the corresponding NSD2 mutant in cells[144]. A similar increased methylation activity was also found for D1125N mutation, which is located in the SET domain as well[143]. To study the effects of these mutations on the substrate specificity, they were generated by site-directed mutagenesis (Table 4).

Table 4: *Reported somatic cancer mutants of NSD2*

Name	Abbreviation	Domain boundaries (aa)	Amino acid mutation
Nuclear receptor SET-domain containing protein 2 E1099K Mutant	NSD2 E1099K Mut	991 − 1240	E1099K
Nuclear receptor SET-domain containing protein 2 D1125N Mutant	NSD2 D1125N Mut	991 − 1240	D1125N

The GST-fused NSD2 enzymes containing either the E1099K or D1125N mutation were overexpressed and purified by affinity chromatography with good yield and purity (Figure 44A).

However, the protein band of the NSD2 E1099K mutant ran lower than the NSD2 D1125N mutant, sequencing results showed that both plasmids were correct and did not contain additional mutations. The methyltransferase activity of the two somatic cancer mutants was tested with peptide arrays as described before. The membranes were methylated by one of the NSD2 mutants and subjected to autoradiography. Both enzymes showed a specific methyltransferase activity toward the H3K36 wild-type peptide (Figure 44B).
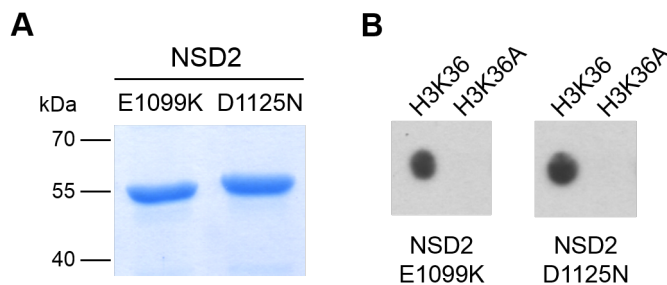
Figure 44: *Methyltransferase activity of the NSD2 somatic cancer mutants on peptides. (A) Coomassie staining of the purified NSD2 E1099K and D1125N mutants. (B) Autoradiography of peptide array methylation by either NSD2 E1099K (upper panel) or D1125N (lower panel) mutants. The arrays contained histone H3K36 (APATGGV**K**KPHRYRP) and K36A (APATGGV**A**KPHRYRP) peptides.*

To confirm the methyltranferase activity at protein level, methylation assays with recombinant histone proteins H3.1 and H4 were performed as well. Coomassie staining of an SDS-PAGE gel with the purified enzymes was used as loading control (Figure 45A). Then, equal amounts of the recombinant histone proteins H3.1 or H4 were incubated with NSD2 wild-type, E1099K or D1125N mutant in methylation buffer containing radioactively labeled [methyl-^{3}H]-SAM and the transfer of the radioactive methyl groups was detected by autoradiography. All three NSD2 enzymes had a good methyltransferase activity toward the recombinant histone proteins, however the H4 proteins showed stronger methylation signals (Figure 45B).
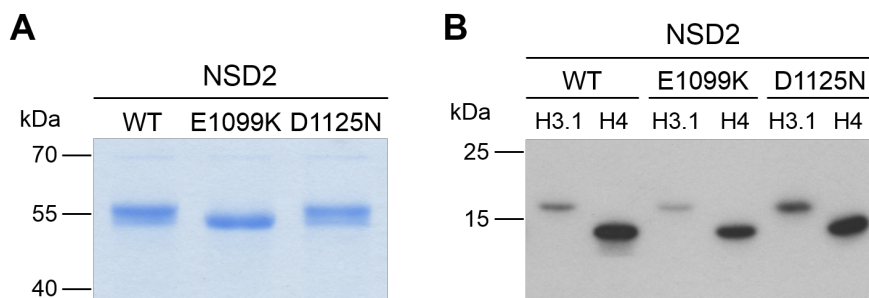
Figure 45: *Methyltransferase activity of the NSD2 somatic cancer mutants on proteins. (A) Coomassie staining of the three enzymes NSD2 WT, E1099K and D1125N mutant. (B) Autoradiography of recombinant histone protein methylation by either NSD2 WT, E1099K or D1125N mutant. The histone proteins H3.1 or H4 were either methylated by the NSD2 WT, E1099K or D1125N mutant enzyme in presence of radioactively labeled [methyl-^{3}H]-SAM.*

3.2.7 Comparison of the Substrate Specificity of the NSD2 Somatic Cancer Mutants to the Wild-Type Protein

To investigate if the substrate specificity of the two NSD2 mutants differs from the wild-type enzyme, methylation of specificity profile arrays were performed as described in section 3.2.2. The membranes were methylated by either the E1099K or D1125N mutant, in presence of radioactively labeled [methyl-^{3}H]-SAM and subjected to autoradiography (Figure 46).
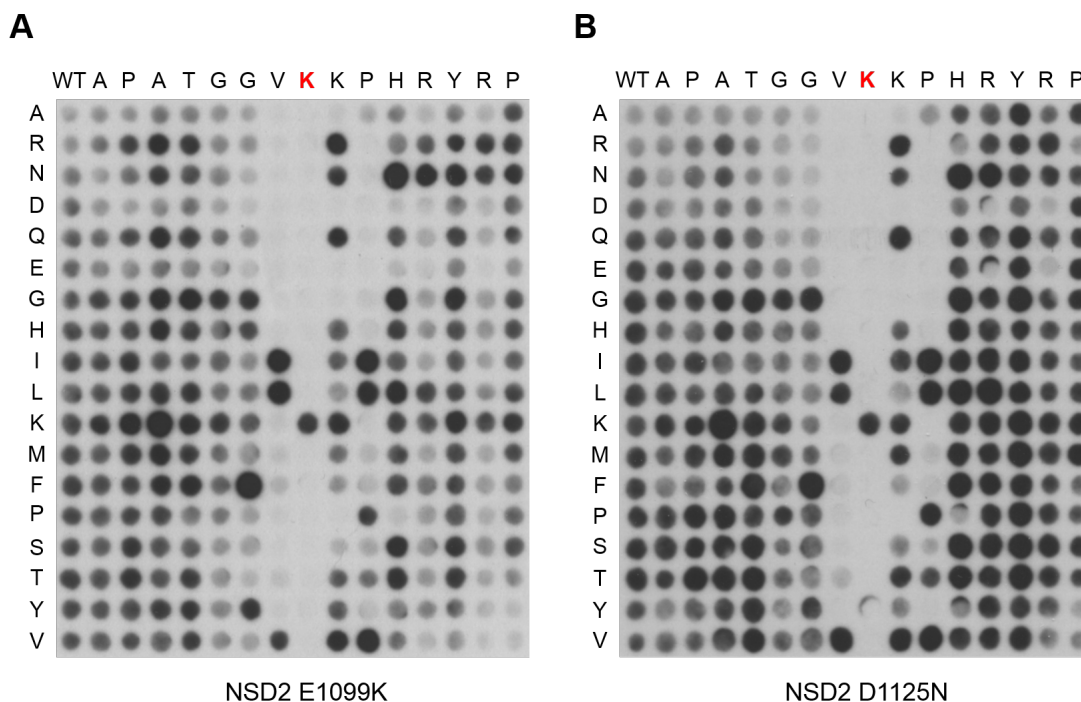
Figure 46: *Substrate specificity profiles of the NSD2 mutants. Autoradiography images of specificity profile arrays based on H3K36 sequence (29 - 43) methylated by NSD2 E1099K (left) and D1125N (right) mutant. The horizontal axis represents the original H3 sequence and the target lysine K36 is highlighted red. The vertical axis shows the residues exchanged in the corresponding row. The first column contains the wild-type sequence of H3 used as a control labeled with WT.*

The signal intensities of the specificity arrays of the two NSD2 mutants were quantified, normalized and the residues from G34 to R40 were compared to the wild-type profile (Figure 47A). Additionally, the discrimination factors at these residues were calculated and the comparison of all three enzymes is shown in Figure 47B. The specificity profile arrays of NSD2 wild-type and the two somatic variants (E1099K and D1125N) showed no large differences in the preference of amino acids between positions -2 (G34) and +2 (P38). Nevertheless, at the positions +3 (H39) and +4 (R40), the E1099K mutant showed a higher specificity than the other two enzymes (Figure 47A and B). All three enzymes showed a strict preference for the same residues at the -1 (I, L and V) and +2 (I, L, V and P) positions. The D1125N mutant showed a higher preference for valine at -1 position than the other two variants. At the -2 and +1 positions the

three NSD2 variants were less strict and tolerated also other amino acids. Taken these results
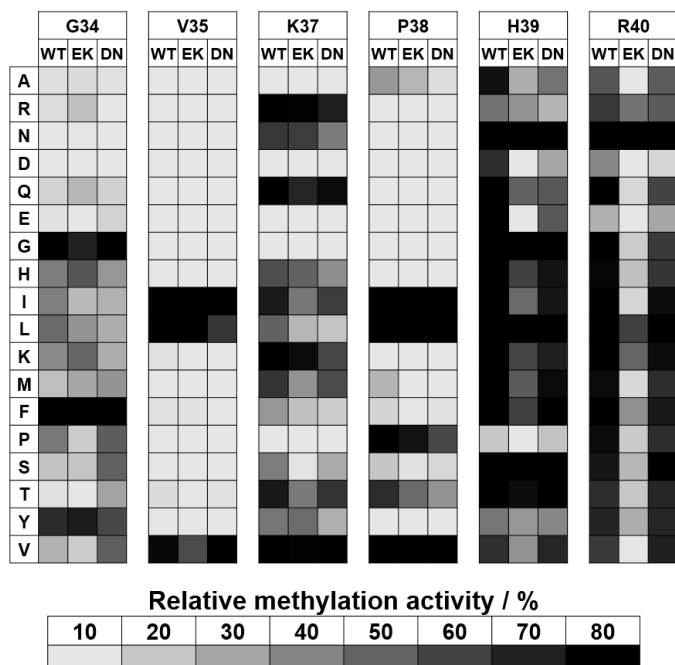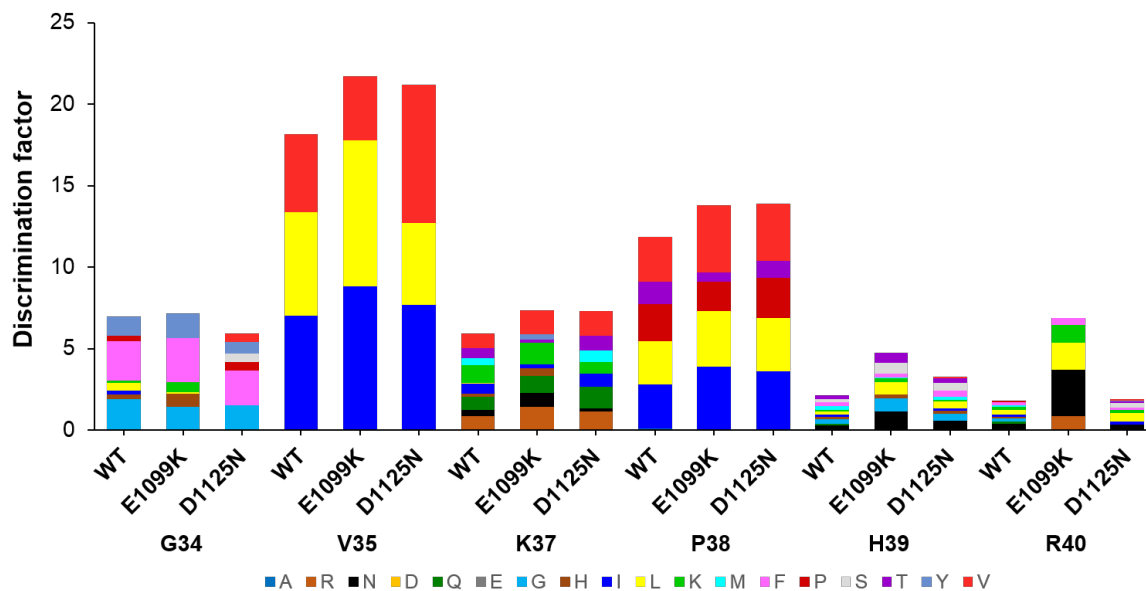
Figure 47: *Specificity profile analysis of NSD2 wild-type, E1099K and D1125N mutant. Comparison of the specificity from residues G34 to R40 of NSD2 WT, E1099K and D1125N specificity profile data. The quantified and normalized results were color-coded depending on their methylation activity. Black to light gray represents a strong to weak methylation.*

together, the substrate recognition profile of the two NSD2 mutants is very similar to the NSD2 wild-type (Table 3). This suggests that the tested mutations within the SET domain of NSD2 do not change the specificity of the enzyme. Next, it was investigated whether the NSD2 E1099K variant methylates additional substrates. A peptide array containing all predicted potential substrates was methylated by NSD2 E1099K and detection was performed as described above (Figure 48).
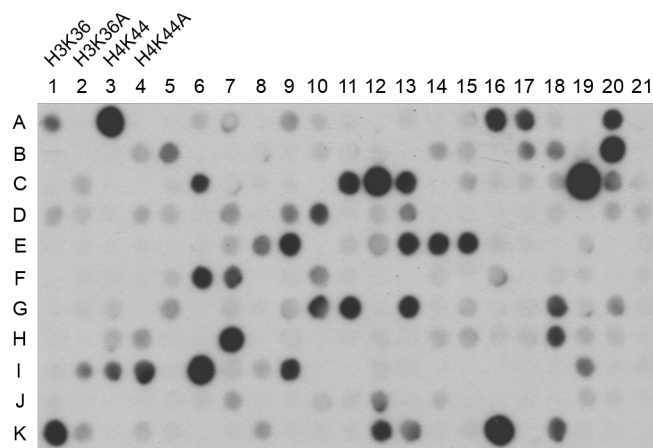
Figure 48: *Methylation of novel peptide substrates by NSD2 E1099K mutant. Autoradiography of the methylated novel peptide substrates by NSD2 E1099K mutant. 15 amino acid long peptides identified by the search profiles 1 and 2 (Table 3) containing the predicted target lysine in the middle were synthesized. Protein names and peptide sequences are listed in Table 15 in section 6.2.2. H3K36 and H3K44 wild-type and the corresponding mutant peptides were included as controls.*

For a better comparison, the intensities were quantified, normalized and the relative activities of the peptide array methylated by NSD2 wild-type were plotted against the activities of the array methylated by NSD2 E1099K (Figure 49).
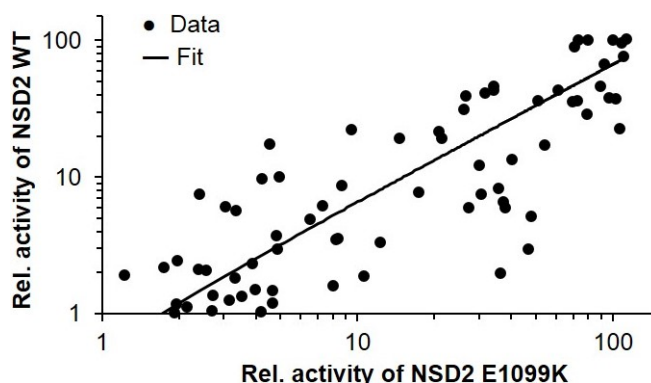
Figure 49: *Scatter plot of the methylated peptide substrate by NSD2 wild-type and E1099K. The intensities of the peptide substrates methylated by NSD2 wild-type were plotted against intensities of the corresponding peptide substrates methylated by NSD2 E1099K. Peptide methylation intensities < 1 % of the H4K44 peptide were excluded from the analysis.*

The intensities for most of the methylated peptides were similar for both methylation assays, but

the scatter plot shows also discrepancies for several peptide substrates, which are indicated by data points lying off an imaginary bisecting diagonal (Figure 49). However, the results showed that the E1099K mutant did not methylate any novel peptides.

It was also tested, if the NSD2 mutants are able to methylate the two novel protein substrates ATRX and FANCM *in vitro*. Corresponding amounts of the protein substrates were subjected to *in vitro* methylation assays with either NSD2 E1099K or D1125N mutant. Figure 50B clearly shows that both NSD2 mutants methylate ATRX. Additionally, the E1099K mutant revealed detectable activity for FANCM. The weak activity of the D1125N mutant on FANCM on the other hand is almost completely covered by the NSD2 automethylation band.
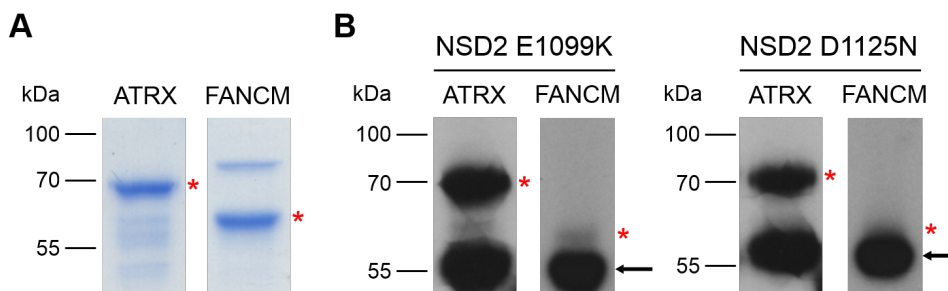
Figure 50: *Methyltransferase activity of NSD2 E0199K and D1125N on protein substrates. (A) Coomassie staining of the three protein substrates ATRX and FANCM. (B) Autoradiography of the protein substrates methylated either by NSD2 E1099K or D1125N mutant. The corresponding bands of the expected size are marked with a red asterisk. Automethylation of the NSD2 mutants is indicated by an arrow.*

To confirm the methyltransferase activity of the two NSD2 cancer mutants in human cells, HEK293 cells were transiently transfected either with the full-length NSD2 WT, E1099K or D1125N variants and one of the YFP-fused protein substrates. As control, the cells were transfected only with the YFP-fused substrates. Harvesting of the cells, purification of the substrates and western blot analysis with H3K36me1 antibody was performed as described above for NSD2 wild-type (Figure 51A and B). The results showed specific binding of the H3K36me1 antibody to the substrate proteins, which were expressed together with the NSD2 variants (Figure 51B). However, no signal was detected when the substrates were expresses alone in HEK293 cells. Even the recombinant GST-fused unmethylated protein domains, which were included as negative controls, did not show antibody binding. Additionally, it was observed that the ATRX substrate, when coexpressed with the NSD2 mutants showed higher methylation signals in comparison to the cotransfection with wild-type NSD2. This agrees with the reported hyperactivity of the NSD2 mutants[143,144]. Interestingly, such an increased methylation activity was not observed for the FANCM substrate when coexpressed with one of the NSD2 variants. The NSD2 D1125N mutant showed even weaker methylation activity on the FANCM substrate than the other two variants, however, the loading control reveals also a weaker Ponceau S staining for this sample.

Taken together, this data show that the E0199K and D1125N mutants have a comparable substrate recognition profile to NSD2 wild-type and that they are able to methylate the substrate proteins, ATRX and FANCM both *in vivo* and *in vitro*.
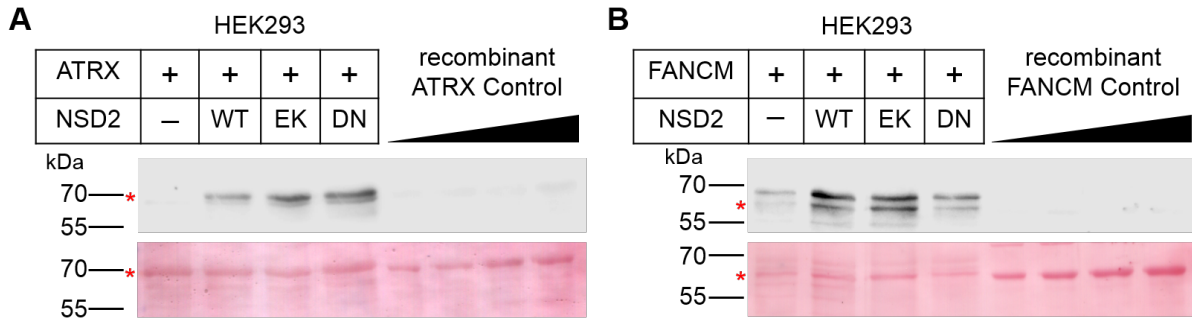
Figure 51: *Detection of cellular methylation of ATRX and FANCM by NSD2 somatic variants.*
(A) The YFP-fused substrate ATRX was ectopically expressed with the described NSD2 enzymes or alone and purified by GFP-Trap® A. Cellular lysine methylation was determined by probing with anti-H3K36me1 antibody (upper panel). Ponceau S staining represents the loading control (lower panel). Recombinant unmethylated GST-fused ATRX substrate was utilized as negative control for the specificity of the antibody. (B) Cellular lysine methylation of YFP-fused FANCM substrate expressed in HEK293 cells. The experiment was conducted as described in A. As a negative control recombinant unmethylated FANCM protein domain was used. The corresponding bands of the right size are marked with a red asterisk.

3.2.8 *In vitro* Methylation of Histone H3 Somatic Cancer Mutations

In the last years an increasing number of studies showed recurrent somatic mutations within the histone H3 genes, which were identified in different cancer types. Some of these mutations encode for lysine to methionine exchanges at position 27 or 36. Other residues surrounding the characterized lysine methylation sites were altered as well [166–168]. The K27M, G34R or G34V mutations within the histone H3.3 gene *H3F3A* were identified in approximately 50 % cases of pediatric high-grade gliomas (pHGGs), a highly malignant type of brain tumor in children [166]. In about 92 % of giant cell tumors of bone, the *H3F3A* gene encoding the histone H3.3 contained the recurrent mutations G34W or G34L [168]. The K36M mutation of histone H3.3, occurring in 95 % all chondroblastomas (a rare and aggressive bone tumor mostly affecting the epiphyses or apophyses of long bones), was found mainly in the *H3F3B* gene but sometimes also in *H3F3A* [168].

Since H3K36 is a known substrate of NSD2, the influence of the H3 somatic cancer mutations on the NSD2 methyltransferase activity was investigated. Peptides with the H3 sequence (29 - 43) containing the H3 missense alterations, which are listed in the COSMIC (Catalogue of Somatic Mutations in Cancer) database were synthesized on membranes. A detailed list of the sequences and the precise mutations of the histone H3 cancer variants is shown in Table 17 in section 6.2.2. The H3K36 and the K36A mutant peptides were included as controls at the beginning and the

end of each peptide array. The peptide arrays were methylated with either NSD1 or NSD2 WT, E1099K or D1125N mutants in methylation buffer containing radioactively labeled [methyl-^{3}H]-SAM (Figure 52A). Methylated peptide arrays were quantitatively analyzed by Phoretix^{TM} Array software and normalized spot intensities were color-coded as described above. The red represent no detectable methylation signal and yellow a strong methylation (Figure 52B).
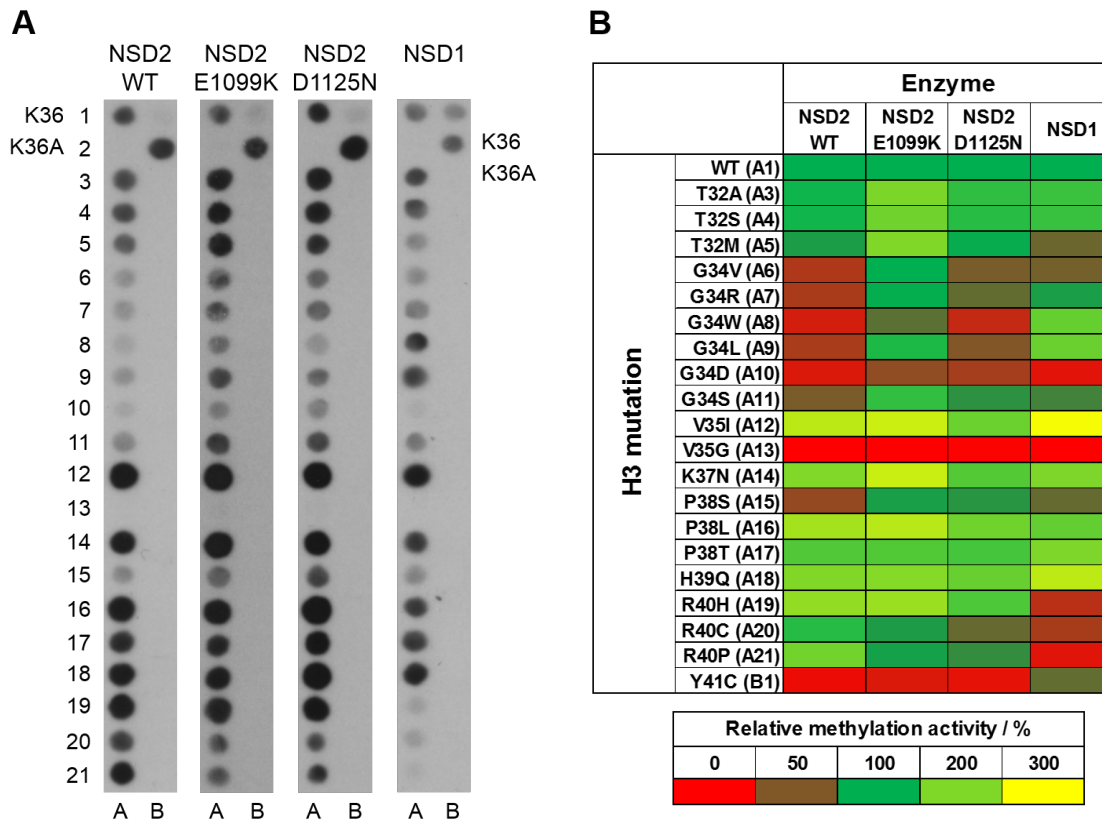
Figure 52: *Methyltransferase activity of NSD2 wild-type, E1099K, D1125N and NSD1 on H3 missense mutation peptides. (A) Autoradiography images of peptide arrays containing the histone missense cancer mutations methylated by NSD2 WT, E1099K, D1125N and NSD1. 15 amino acid long peptides containing the sequence of several histone missense cancer mutations (residues 29 - 43) received from the COSMIC database were synthesized on peptide arrays. The peptide sequences with the corresponding cancer mutations are listed in Table 17 in section 6.2.2. Peptides of H3.3 K36 wild-type and H3.3 K36A mutant were included as controls. (B) The quantified signals were normalized, averaged and color-coded dependent on the methylation intensities. Red to yellow represents a weak to strong methylation.*

Overall, NSD1 and the three NSD2 variants showed a similar methylation pattern of the histone H3 somatic missense mutations (Figure 52). Mutations of T32 (A3 - A5) have little effect on the activity of NSD1 and the NSD2 variants, the intensities of these spots were comparable to the corresponding H3K36 peptides. However, mutations of G34 clearly showed differences of the methylation intensities at peptides A6 to A11. NSD2 wild-type and D1125N mutant showed decreased activity on these peptides, whereas NSD2 E1099K and NSD1 revealed no big changes

at the peptides except at G34D (A10), which reduced the activity as well. Mutation of V35 to I increased the methylation signal, since isoleucine is more preferred than valine at position -1. In contrast, mutation of the same valine to glycine (V35G) completely prevents the activity of all four enzymes. Stronger methylation signals were also observed on the K37N, P38L, P38T, H39Q and R40H mutant-containing peptides. These were either tolerated or even preferred over the native H3 peptide. The peptides containing the P38S mutation, showed a decreased methylation signal, since serine is not preferred at this position. Some differences were also observed for the peptides mutated at position 40 (R40C and R40P), NSD2 wild-type showed an increased activity, whereas, the activities of the NSD2 mutants and NSD1 were weaker. The mutation of Y41 to C abolished or reduced the methyltransferase activity of the NSD2 enzymes and NSD1, respectively, thus cysteine seems to be not tolerated at position +5. The results showed that the H3 missense mutations have either an elevating or suppressing effect on the methyltransferase activity of the four tested NSD enzymes. Additionally, it was observed that the methylation intensities of the H3 mutations fit to the described specificity profiles of the NSD2 variants.

3.2.9 The H3K36M Missense Mutation Inhibits the Methyltransferase Activity of NSD2

Lewis *et al.* showed that the H3K27M mutation inhibits the methyltransferase activity of the Polycomb repressive complex 2 (PRC2)[169]. PRC2 di- and trimethylates K27 and maintains epigenetic gene silencing and X chromosome inactivation. The K27M mutant competes with wild-type substrates for the active site of PRC2 and prevents global methylation at K27 residues. This can alter cellular processes and lead to gliomagenesis.

To test whether the K36M mutant inhibits the methyltransferase activity of NSD2, *in vitro* methylation assays with K36 wild-type and K36M mutant peptides were performed. First, increasing concentrations of K36 peptide was methylated in methylation buffer containing NSD2 and radioactively labeled [methyl-^{3}H]-SAM. The methylation reactions were separated by SDS-PAGE and the transfer of the radioactively labeled methyl groups was detected by autoradiography. The results revealed an increasing methylation activity with increasing concentrations of K36 peptide (Figure 53A). However, above 100 μM of peptide the methyltransferase activity reached saturation and did not rise further.

To investigated the effect of K36M peptide on the methylation activity of NSD2, two different concentrations of the K36 containing peptide were selected and methylation assays were performed in the presence of increasing concentrations of K36M mutant peptides (Figure 53B an C). With increasing concentrations of the K36M peptide, a stronger inhibition of the methyltransferase activity of NSD2 was observed. Quantification of the methylation signals of all three methylation assays are presented next to the corresponding autoradiography images.
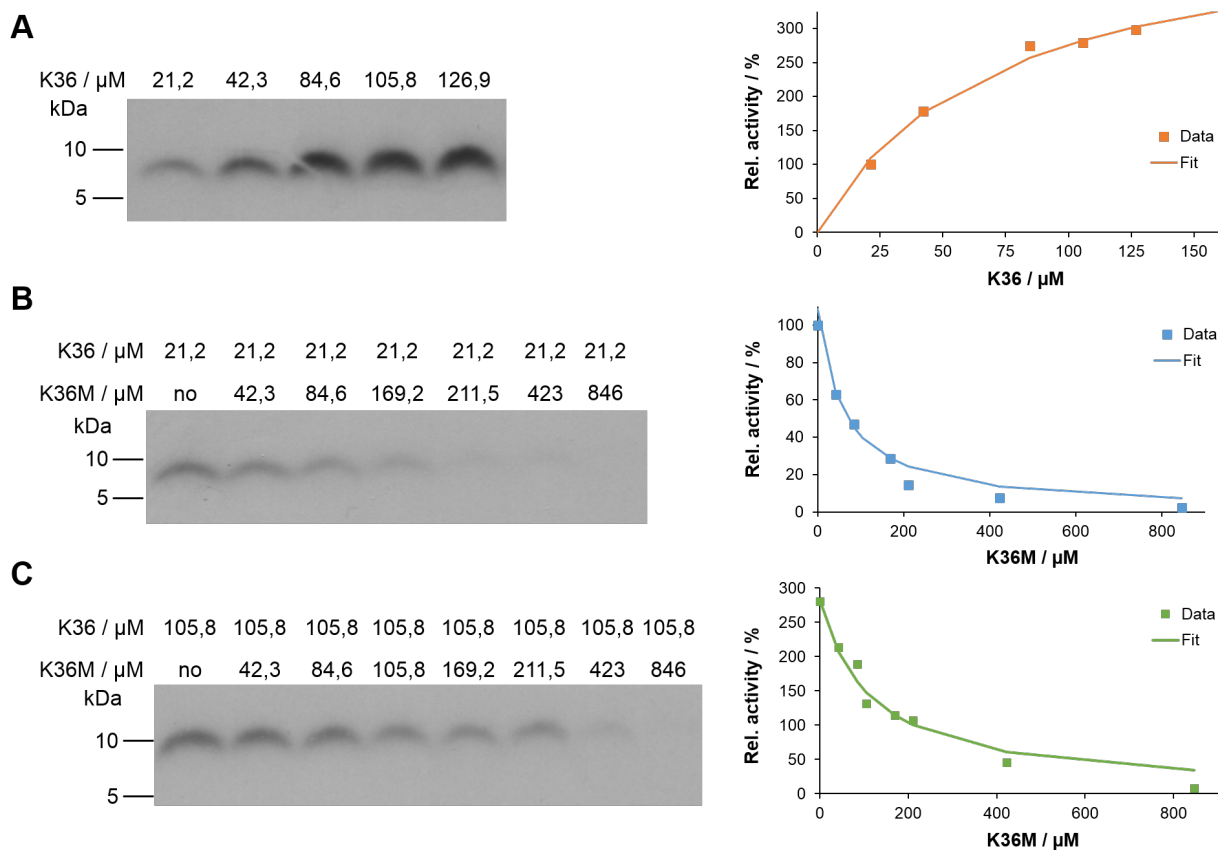
Figure 53: *Analysis of NSD2 enzymatic activity on H3K36 peptide. (A) Methylation of varying concentrations of K36 wild-type peptide. (B) Methylation of varying concentrations of K36M peptide in presence of 21.2 μM K36 peptide. (C) Methylation of varying concentrations of K36M peptide in presence of 105.8 μM K36 peptide. Quantification of the corresponding autoradiography images is represented next to the methylation assays.*

With the received data a Lineweaver-Burk plot was generated to specify the type of inhibition. The plot shows four different regression lines corresponding to the used inhibitor concentrations, which have all the same y-intercept, suggesting K36M is a competitive inhibitor with the substrate peptide (Figure 54).
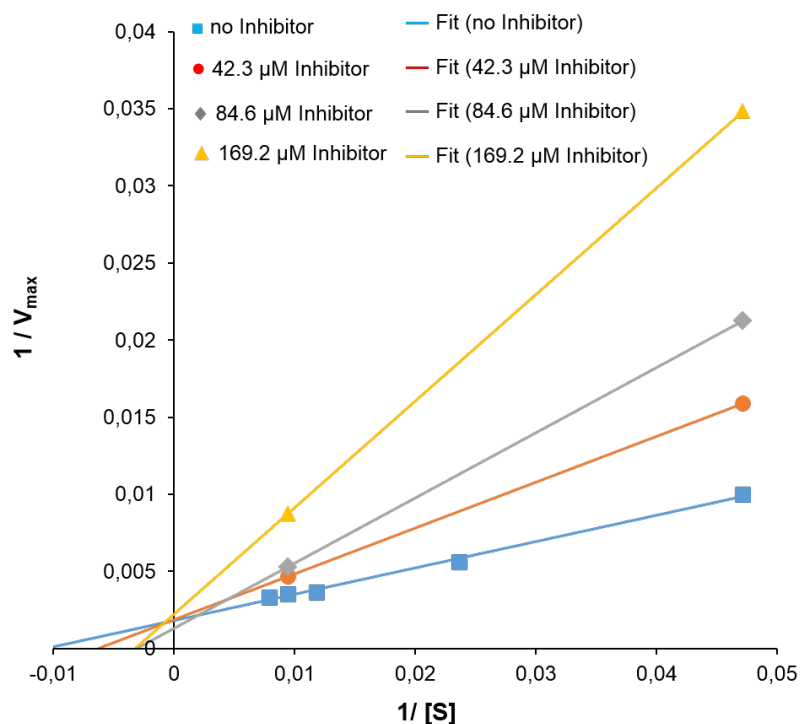
Figure 54: *Lineweaver-Burk plot of the NSD2 peptide methylation assays. Data obtained from the methylation assays shown in Figure 53 were used generate a Lineweaver-Burk plot to specify the type of inhibition of the K36M peptide on the H3K36 peptide methylation by NSD2.*

Fitting of the data to the equation describing a substrate competitive inhibition by least-squares method revealed a K_M value of 71 µM for the methylation reaction of the K36 peptide by NSD2. K_M represents the Michaelis-Menten constant that shows the substrate concentration at which the reaction rate is equal to one half of the maximum rate for the reaction. Additionally, the K_I value of the K36M peptide inhibitor was determined and revealed a K_I of 47 µM for the performed reactions. In summary, it was shown that the somatic cancer mutations of H3 have an effect on the methyltransferase activity of NSD2 and that the K36M peptide is a competitive inhibitor of NSD2 activity on H3K36 peptides.

3.3 Characterization of the Substrate Specificity of the Yeast Histone Lysine Methyltransferase, Clr4

Clr4 is the homolog of the *Schizosaccharomyces pombe* histone lysine methyltransferase (HKMT) Su(var)3-9. Clr4 catalyzes the trimethylation of lysine K9 of histone H3, which is an important mark for the formation of heterochromatin. Trimethylated H3K9 serves as a binding site for the chromodomain of Clr4 but also for the chromodomain of Swi6, an important protein associated with heterochromatin formation, maintaining heterochromatin structure and transcriptional regulation. Additionally Zhang *et al.* reported that Clr4 methylates K167 of Mlo3, a non-histone protein, which is required for nuclear export of RNA[161].

3.3.1 Purification and Assessment of Methyltransferase Activity

To obtain better insights into the substrate recognition pattern of Clr4 and to identify additional potential non-histone substrates, the substrate specificity profile of this methyltransferase was determined. The bacterial expression construct of Clr4 was kindly provided by Prof. Dr. Danesh Moazed. The His$_6$-tagged Clr4 protein was expressed in bacteria and purified by affinity chromatography with sufficient yield (Figure 55A).
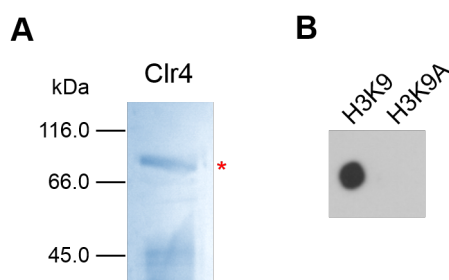
Figure 55: *Methyltransferase activity of Clr4. (A) Coommasie staining of the purified Clr4 enzyme. The band of the expected size is marked with a red asterisk. (B) Autoradiography of peptide array based on histone H3 sequence (ARTKQTAR**K**STGGKAPRKQ) containing H3K9 and corresponding K9A mutant peptides methylated by Clr4.*

To test the methyltransferase activity of Clr4, peptide arrays were methylated by the purified enzyme. The peptide array contained histone H3 (1 - 15) and the corresponding target lysine to alanine (K9A) mutant peptides. The results showed that Clr4 exhibited strong methyltransferase activity toward the H3K9 wild-type, but not against the H3K9A mutant peptide (Figure 55).

3.3.2 Determination of the Specificity Sequence Profile of Clr4

To examine the substrate specificity of the histone lysine methyltransferase Clr4, a mutational scanning peptide SPOT array was synthesized using the histone H3 sequence (residues 1 - 18) as template. The membrane was incubated with Clr4 in methylation buffer containing radioactively labeled [methyl-^{3}H]-SAM and subjected to autoradiography (Figure 56).
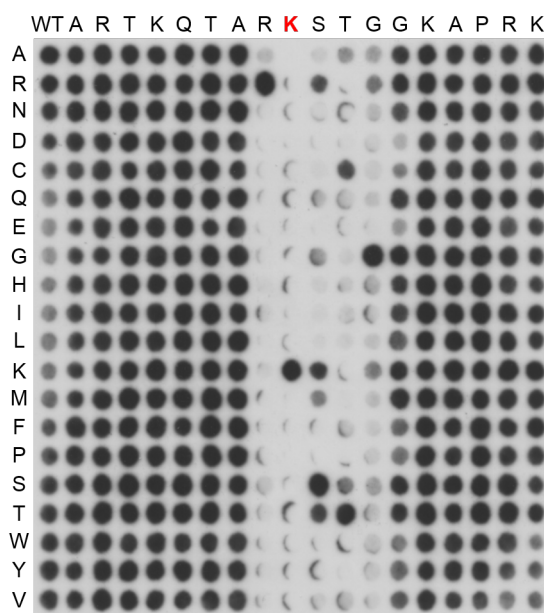
Figure 56: *Substrate specificity profile of Clr4. Autoradiography image of the specificity profile array based on H3K9 sequence (1 - 18) methylated by Clr4. Horizontal axis represents the original H3 sequence and the target lysine K9 is highlighted red. Vertical axis shows the residues exchanged at the corresponding position in the original sequence, which provides an array with all possible single amino acid mutations of the histone H3 sequence. The first column contains the wild-type sequence of H3 as a control labeled with WT.*

The methylated peptide array revealed a recognition motif covering the residues R8 (-1) to G12 (+3) of the H3 sequence, where the target lysine K9 is defined as position 0. Besides the target lysine K9, Clr4 showed very strict specificity at the -1 (R8) and +3 (G12) positions. A mutation of the cognate residues at these positions to any other amino acids abolished the methylation signal. Strong preferences were also observed at the +1 (S10) and +2 (T11) positions. Besides S, at the +1 position R, K, and T were tolerated and at the +2 (T11) position, along with the cognate threonine, cysteine and serine were allowed. Outside of this recognition motif, Clr4 is very unspecific and tolerates almost all other amino acids. The substrate recognition motif is summarized in (Table 5).

Table 5: *Substrate specificity profiles utilized to identify putative novel NSD2 substrate*

Cognate residue	R8	K9	S10	T11	G12
Position	-1	0	+1	+2	+3
Search profile	R	K	SKRT	TCS	G

3.3.3 Identification of Putative Novel Substrates of Clr4

It is known that Clr4 interacts with several non-histone proteins, such as Swi6 and Mlo3 and it was also reported that Clr4 is able to methylate lysine K167 of Mlo3[161]. Therefore, the interactors of Clr4 that contain the substrate sequence motif of Clr4 were further analyzed (Table 5). An alignment of selected K residues of Clr4 interaction partners with its substrate sequence motif is shown in Table 6. This indicates that all the listed proteins may be potential substrates of Clr4. Residues matching with the specificity profile were marked in green, while those which were not matching were marked in red.

Table 6: *Sequence alignment of residues surrounding the target lysine of the proteins interacting with Clr4 and the substrate specificity profile of Clr4. X in the specificity profile stands for all amino acids are allowed. Amino acids in the green box are tolerated and in the red box are not tolerated by Clr4 at the corresponding position. The potential target lysines are shown in brackets.*

Position	-4	-3	-2	-1	0	+1	+2	+3	Position on peptide array	
Specificity Profile	X	X	X	R	**K**	SK RT	TCS	G	Target K	K to A
H3 (K9)	Q	T	A	R	**K**	S	T	G	A1	A2
Mlo3 (K167)	S	S	K	R	**K**	T	T	R	A3	A4
Swi6 (K144)	P	S	K	R	**K**	R	T	A	A5	A6
Spbc28F2.11 (K250)	K	P	K	R	**K**	H	T	R	A7	A8
Spbc28F2.11 (K292)	K	K	R	R	**K**	S	S	M	A9	A10
Hrp3 (K89)	S	K	H	R	**K**	G	T	R	A11	A12
Dbp2 (K165)	R	A	G	A	**K**	G	T	A	A13	A14
Iec3 (K153)	K	Q	K	R	**K**	R	T	S	A15	A16
Mcp1 (K132)	P	P	A	R	**K**	T	T	G	A17	A18
Cbc1 (K11)	T	R	P	R	**K**	R	T	R	A19	A20
Rik1 (K460)	Y	D	S	A	**K**	R	S	R	A21	B1
Rik1 (K502)	E	V	A	R	**K**	V	F	E	B2	B3

To investigate if these proteins are potential substrates for Clr4, peptides of the corresponding sequences containing the putative target lysine were synthesized on cellulose membrane. Protein names with the corresponding peptide sequences are listed in Table 18 in section 6.3.2. The membrane was methylated by Clr4 and the transfer of the radioactive labeled methyl groups

was detected by autoradiography (Figure 57).
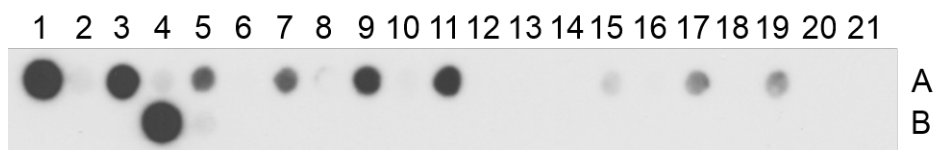
Figure 57: *Methylation of putative novel peptide substrates by Clr4.* Autoradiography of the peptide array with the putative substrate peptides containing the predicted target lysine and the corresponding K to A mutation methylated by Clr4.

The methylation of the array revealed 3 strongly methylated peptides, which belong to Mlo3, Spbc28F2.11 (K292) and Hrp3 (Figure 57). Additionally, 4 peptides showed weak methylation signals: Swi6, Spbc28F2.11 (K250), Mcp1 and Cbc1. However, the highest signals were observed for the native H3K9 peptides, which served as controls. Loss of methylation signal on the target lysine to alanine mutation peptides confirmed the methylation of the predicted target residues. All in all, the substrate specificity profile of the histone lysine methyltransferase Clr4 was determined and it was shown that the enzyme was able to methylate six additional putative substrates at peptide level. Two of them showed a comparable strong methylation as Mlo3 (Figure 57).

3.4 Development of an Advanced Non-radioactive, High-throughput PKMT Activity Assay

Many assays already exist that allow the detection of PKMT activity[170–174]. However, these procedures exhibit a couple of disadvantages, such as the usage of radioactively labeled SAM. Moreover, sometimes, the identification of the methylated target site and determination of the degree of methylation is not possible. Also, the employment of methyl-specific antibodies has disadvantages, the use in high-throughput systems needs large amounts of antibodies. These are expensive, require animals for production and may have batch-to-batch variability.

To overcome the disadvantages of current PKMT assays, we developed a novel sensitive microplate ELISA PKMT assay, which is working with non-labeled SAM and uses methyl-lysine specific reading domains instead of methyl-specific antibodies. These natural reading domains can be recombinantly produced in *E. coli*. They have the important advantages that they are less expensive and exhibit constant quality. Moreover, like antibodies, reading domains are highly specific for the methylation site and the degree of methylation[175]. The procedure of this assay consists of seven main steps, which are schematically illustrated in Figure 58. The biotinylated peptides are methylated by a PKMT in presence of unlabeled SAM and then bound to the avidin coated wells of a microplate. The wells are next incubated with 2 % BSA blocking solution, to prevent unspecific binding of the reading domains or antibodies. Subsequently, the reading domain is added to the plate and incubated to allow binding. Afterwards, the primary antibody, which is either specific for the reading domain or its affinity tag, is added. Finally, the detection and quantification of the luminescence signal is performed with an HRP-coupled secondary antibody.
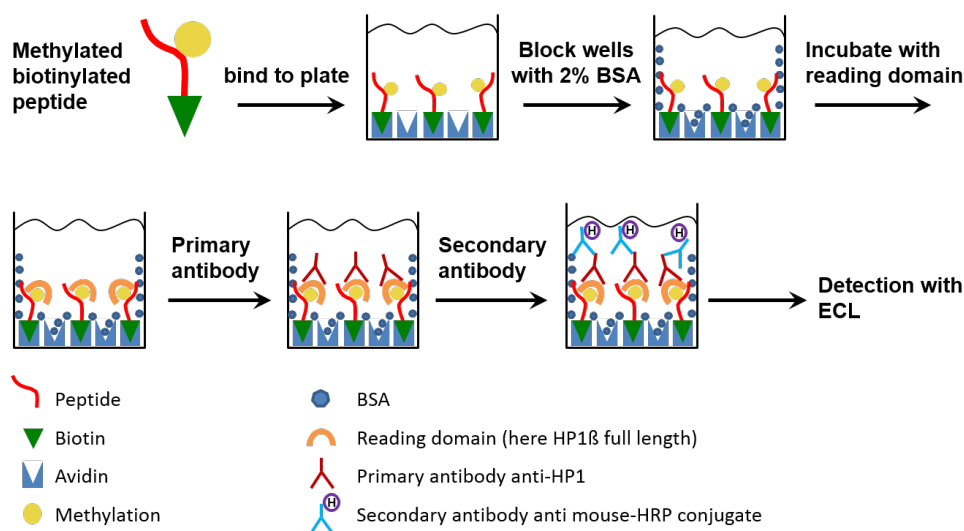
Figure 58: *Schematic illustration of the basic steps of the developed PKMT assay utilizing the reading-domains.*

First, it was investigated whether the reading domain HP1β interacts with the H3K9 methylated peptides in the microplate assay. Biotinylated H3 peptides (residues 1-19) of all K9 methylation states, were bound to the avidin coated wells. Unmethylated H3K9 and trimethylated H3K36 peptides were included as negative controls. The wells containing different peptides were incubated with GST-fused HP1β and binding of the reading domain was analyzed with a GST-specific antibody (Figure 59A).
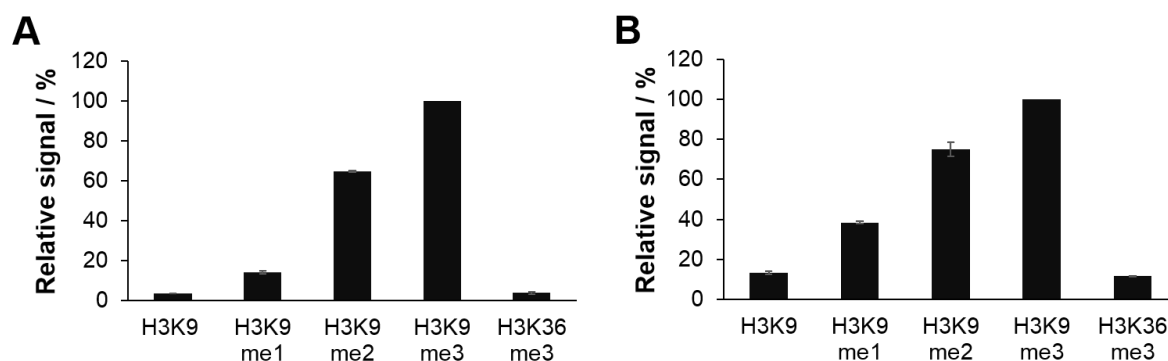
Figure 59: *Application of the HP1β protein as methyl lysine reader. (A) Interaction of GST-tagged HP1β with un-, mono-, di- and trimethylated H3K9 peptides detected with anti-GST antibody. Trimethylated H3K36 peptide included as control. The signals were normalized to the trimethylated H3K9 peptide. (B) Interaction of GST-tagged HP1β protein with the same peptides used in A, detected with HP1β-specific antibody. The signals were normalized to the trimethylated H3K9 peptide.*

Two independent experiments were performed. Both replicates were consistent. In this assay, the HP1β reading domain interacted strongly with the trimethylated H3K9 peptides (Figure 59A). Approximately 1.7-fold weaker binding was observed with the dimethylated and 7.6-fold weaker with the monomethylated peptides. These results reflect the ratios of the binding constants of HP1β, shown in previous reports[176,177]. There, the binding constant of HP1β to H3K9me3 was around 6 μM, to dimethylated H3K9 ~1.8-fold weaker and to monomethylated H3K9 ~7.3-fold weaker. The binding signals to the unmethylated H3K9 and trimethylated H3K36 peptides were both very weak (<5 %). The experiments were repeated under the same conditions as described above, but a primary antibody directed toward HP1β was used, instead of a GST-specific antibody. This was because in some experimental setups, the use of a reading domain-specific antibody might be more advantageous, for example if the PKMT itself is fused to a GST-tag (Figure 59B). The signals of HP1β binding to the different peptides showed a pattern comparable to that observed with the GST-specific antibody. Although the signals were slightly increased, the ratio of monomethylated to unmethylated H3K9 peptides was almost the same as the ratio observed with the GST antibody. However, the ratios of trimethylated to dimethylated H3K9 peptide binding (~1.3-fold) and trimethylated to monomethylated H3K9 peptide binding (~2.6-fold) were lower. Even with various concentrations of HP1β, primary and secondary antibody,

we observed the same relative binding signals, suggesting a different binding specificity of the two primary antibodies. It seems that the HP1β-specific antibody exhibit a higher detection sensitivity towards HP1β bound to H3K9me1 compared to the GST-specific antibody, which is an advantage because it allows the detection of monomethylated H3K9. H3K9me1 is the methylation product of the first cycle of a PKMT, and this increases the sensitivity of the assay. These results confirmed the suitability of reading domains to detect peptide methylation in a microplate assay approach.

To investigate the optimal concentration of the HP1β protein, un- and trimethylated H3K9 peptides were incubated with various concentration of the reading domain (Figure 60). The
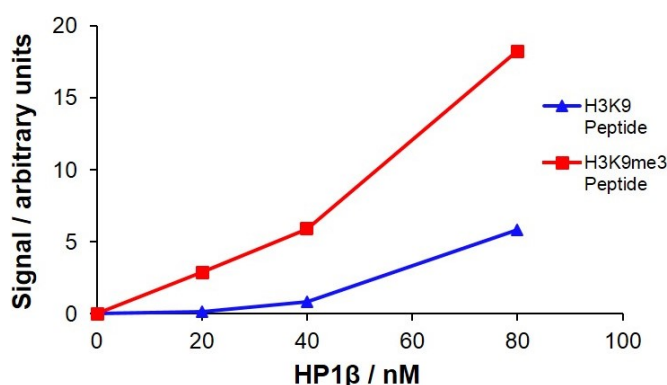
Figure 60: *Determination of optimal HP1β concentration. Interaction of unmethylated (▲) and trimethylated (■) H3K9 peptides with various concentrations of GST-tagged HP1β protein, detected with HP1β-specific antibody.*

increasing concentration of HP1β led to an increase in the luminescence signal for the trimethylated peptide. However, at higher HP1β concentration, increased background (unmethylated H3K9 peptide) signal was noticed as well (Figure 60). All the following experiments were therefore performed with 40 nM of reading domain.

Next, it was investigated if enzymatically methylated substrates could be detected in this approach as well. An *in vitro* methylation of biotinylated H3 (1‑19) peptide was performed with the recombinant SET domain of SUV39H1. The methylation reactions were carried out with 200 nM of SET-SUV39H1 to provide sufficient methylated peptide for the experiment. The unmethylated and trimethylated H3K9 peptides were used as controls and HP1β as reading domain. For detecting the reading domain the HP1β-specific antibody was used (Figure 61). The signal of the SET-SUV39H1 methylated peptide was about 85 % of the synthetic trimethylated H3K9 control peptide and more than 7-fold higher than the signal observed with the unmethylated control peptide (Figure 61). This shows that a high level of methylation was obtained under these conditions and documents a very good dynamic range for the assay. A remarkable signal-to-noise (SN) ratio of 9.5 showed a very good reproducibility of the experiments. To further assess the quality of the assay, its Z-factor was calculated. This is an established statistical
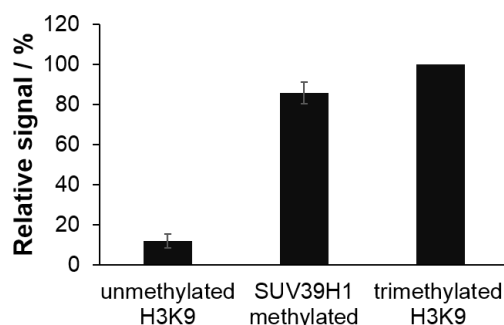
Figure 61: *Analysis of SUV39H1 enzymatic activity using the HP1β protein as methyl reader. Unmethylated H3K9 peptide was methylated by SUV39H1. Unmethylated and trimethylated H3K9 peptides were included as controls. The peptides were incubated with the HP1β protein and the interactions were detected with the HP1β-specific antibody. The signals were normalized to the synthetic H3K9me3 peptide.*

parameter to judge the overall quality of a high-throughput screening assay. With the results observed in the experiments, a Z-factor of 0.65 was received. In larger high-throughput assay systems, Z-factors >0.7 are considered as very good and factors >0.5 are acceptable[178].

To investigate if this assay can be used to screen for PKMT inhibitors, the fungal toxin chaetocin, which was reported to inhibit the methyltransferase activity of SUV39H1, was tested[173]. The SET-SUV39H1 enzyme was pre-incubated with DMSO and different concentrations of chaetocin for 15 min. Afterwards, biotinylated H3 (1 - 19) peptides were methylated with the pre-incubated enzymes (20 nM final concentration of SET-SUV39H1), in methylation buffer containing unlabeled SAM. The methylation reactions were transferred to the avidin coated plate and handled as described above. For detection of the reading domain, the HP1β-specific antibody was used (Figure 62).
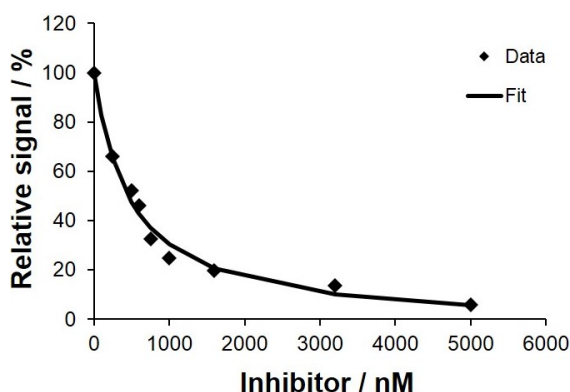
Figure 62: *Analysis of SUV39H1 enzymatic activity in presence of various concentrations of chaetocin using the HP1β protein as methyl reader. The signal of the methylated peptide observed after methylation without inhibitor was set as 100 %. The signal obtained with unmethylatd peptide was considered as background, and the signals of the samples were normalized accordingly. The IC_{50} value was determined by least-squares fit of the data to an equation describing the simple inhibition of an enzyme by binding of an inhibitor.*

With increasing concentration of chaetocin, a stronger inhibition of the methyltransferase activity of SUV39H1 was observed (Figure 62). The data were analyzed by least-squares fitting method to an equation, which described a simple enzyme inhibition reaction. The calculated IC_{50} value of the inhibition of SUV39H1 by chaetocin for our data is 480 nM, which matches with the published IC_{50} of 600 nM[173].

Afterwards, the methylation of these samples was further verified by mass spectrometric. The H3 (1 - 19) peptides were methylated by SUV39H1 with various concentration of chaetocin. As control, we included H3 peptide methylated by SUV39H1 in the absence of inhibitor (Figure 63).
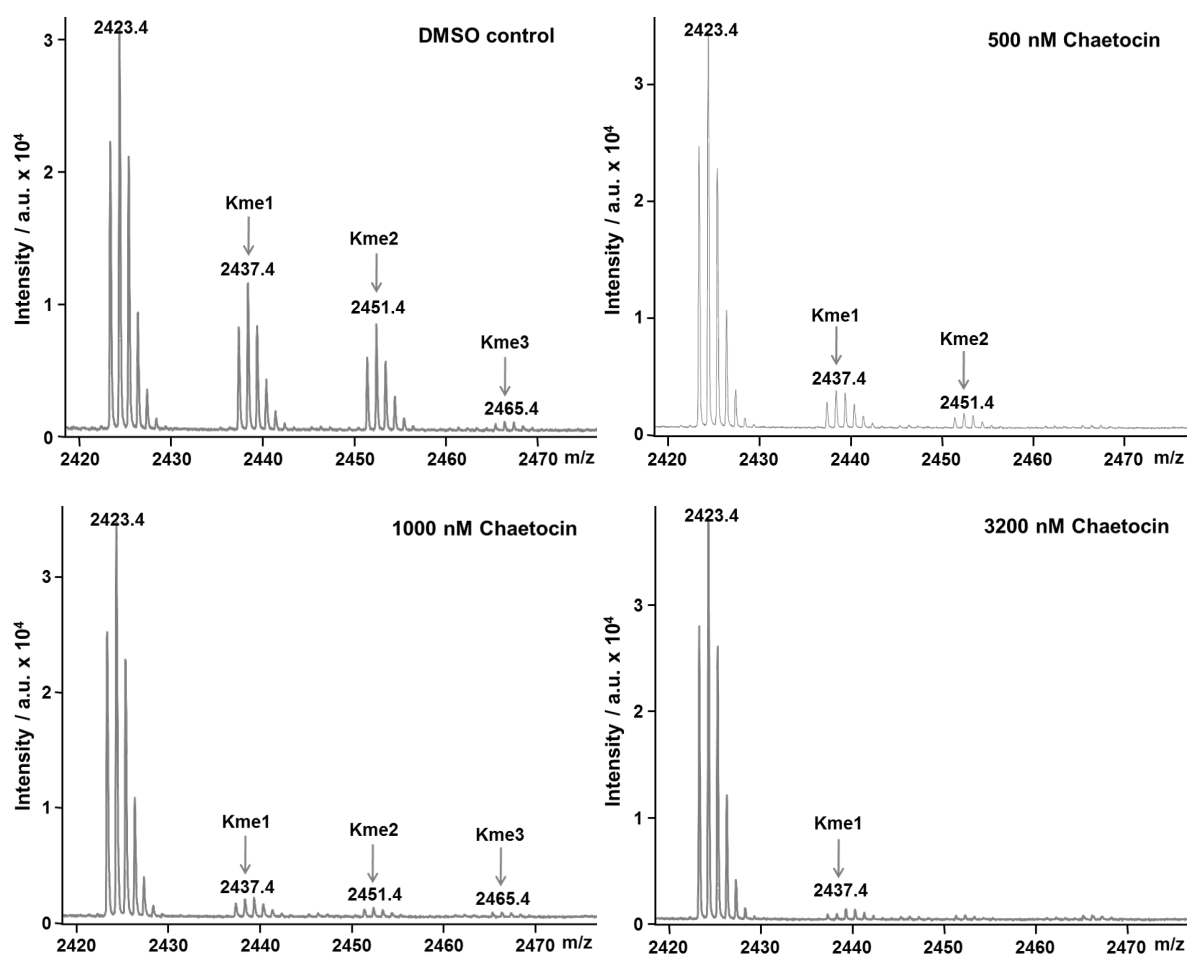
Figure 63: *MALDI MS analysis of the H3K9 peptide methylation samples incubated with SUV39H1 in presence of various concentrations of chaetocin as shown in Figure 62. The peptide sequence is ARTKQTAR**K**STGGKAPRKQ-K(Biot)-NH$_2$ and its theoretical molecular weight in the unmethylated state is 2423 kDa. The theoretical masses of the methylated peptides are indicated.*

The mass spectrum of the control sample indicated all three methylation levels of H3K9, although the highest peak was observed for the unmethylated state (2423.4 kDa). The signals of

mono- (2437.4 kDa) and dimethylated (2451.4 kDa) H3K9 were weaker and the trimethyled form of H3K9 (2465.4 kDa) was very weak, but still detectable (Figure 63). However, with increasing concentrations of chaetocin the MALDI analysis showed strongly decreased levels of H3K9 methylation. This is validating the results of the ELISA assay. In summary, it was shown that reading domains can be used for high-throughput PKMT inhibitor screens instead of antibodies and that the results obtained by the novel assay were confirmed by mass spectrometry.

4 Discussion

PTMs, such as protein methylation, are important regulators of cellular processes. Methylation of histone proteins regulates chromatin structure, thereby affecting functions, such as transcriptional regulation and DNA damage response[30–33]. Additionally, methylation of non-histone proteins controls many other protein functions and properties, such as protein stability, activity, protein-protein interactions and cellular localization[43,44,46,47]. The enzymes responsible for transferring the methyl group to protein substrates play important biological roles. This is indicated by the finding that abnormal expression or aberrant methyltransferase activity is often associated with various diseases and cancer types[129,130,141]. In the recent years, numerous reports discovered many novel protein methyltransferase substrates and the number of new protein methylation sites is growing rapidly[97,102]. The identification of such novel substrates is important for a complete understanding of the function of the methyltransferase enzymes and the role of protein methylation in various signaling functions.

In this work, an established method to characterize the substrate specificity profile of three different protein methyltransferases (PMTs) was used[179,180]. Based on the identified substrate recognition motif of the enzymes, several novel substrates were discovered and their methylation was confirmed *in vivo* and *in vitro*.

4.1 Specificity Analysis of HEMK2 and Identification of Novel Target Substrates

The glutamine methyltransferase HEMK2 is one of few enzymes responsible for the methylation of a glutamine side chain. HEMK2 was reported to modify the eukaryotic release factor (eRF1) within the universally conserved GGQ motif. eRF1 is responsible for the recognition of stop codons and induces the hydrolysis of nascent polypeptides from tRNA. These functions are stimulated by its methylation by HEMK2.

The crystal structure of HEMK2 is not available, thereby the residues involved in the substrate recognition were difficult to pinpoint. To circumvent this limitation, peptide arrays were employed to determine the specificity profile of HEMK2. For these, the known substrate sequence of eRF1 (179 - 192) was used as a template. The experiments revealed a wide substrate recognition motif comprising the positions from -3 (R182) to +7 (R192), when the target Q185 is defined as position 0. The results of the *in vitro* peptide array methylation propound a G-**Q**-X$_3$-R minimal recognition motif for HEMK2. With this specificity profile, 302 proteins that contain the sequence motif could be identified. Peptide array methylation confirmed the methylation at 125 of these predicted peptide substrates. The highly methylated peptides were selected to be further investigated and 35 protein domains were successfully purified. *In vitro* methylation experi-

ments showed that 16 protein substrates were methylated. 5 of them displayed strong methylation signals (CHD5, AMPD2, NUT, ANKRD34A and ABCA2), 8 proteins domains revealed weaker methylation signals (GHDC, RRP1, TGFB3, ZSCAN10, ASH1l, PLEK, ARHGEF10 and SAMD7), and very weak methylation could be detected on 3 candidates (BEND7, POLG and GNA12). Additionally, cellular methylation of the strongest targets, CHD5 and NUT, was confirmed in HEK293 cells. Based on the large set of novel identified peptide and protein substrates, one may speculate that HEMK2 may have more substrates in human cells apart from eRF1. These however, could not be elucidated in this study. Overall, these results show that the impact of glutamine methylation in general is underestimated.

Unlike other posttranslational modification, such as phosphorylation or even methylation at lysine or arginine residues, the glutamine methylation is a subtle PTM. In case of eRF1, the glutamine methylation influences a hydrogen bonding network within the peptidyl transferase center (PTC). The lack of a methyl group at glutamine increases the mobility of the side chain, leading to an increase in the activation energy for hydrolysis of freshly synthesized polypeptides[181]. The glutamine methylation of histone H2A affects protein-protein interactions of the chaperone FACT (Facilitator of transcription)[59]. The two novel strongly methylated substrates, CHD5 and NUT, identified in this study, are known to have important cellular functions:

- CHD5 is a member of the chromodomain-helicase-DNA binding protein family containing several chromodomains and PHD domains. It was reported that decreased levels of CHD5, caused either by deletion of the gene or hyper-methylation of its promoter, lead to the formation and progression of multiple cancer types, such as breast or epithelial ovarian cancer, suggesting a role as tumor suppressor gene[182–184].

- The function of the NUT protein is poorly characterized, however, 75 % of all NUT midline carcinoma (NMC) and other cancers are caused by a genetic translocation of the *NUT* and *BRD4* genes leading to formation of a *BRD4-NUT* oncogene[185]. The BRD4-NUT fusion protein blocks differentiation, and its knock-down in NMC cells results in differentiation and growth arrest[186]. Moreover, the target glutamine in the NUT protein is located adjacent to the nuclear localization signal, so the methylation may change the localization of the protein and its targeting to chromatin.

However, the biological function of the methylation at the glutamine residue of CHD5 and NUT protein is still not elucidated and has to be investigated thoroughly.

4.2 Specificity Analysis of NSD2 and Identification of Novel Protein Substrates

The nuclear receptor SET domain-containing protein 2, NSD2, is a histone lysine methyltransferase, which was shown to mono- and dimethylate lysine K36 of histone H3. Some studies reported that NSD2 methylates additional lysine residues of H3 and H4, such as H4K20, H3K27 or H3K4[107]. These findings could be however not confirmed in follow up biochemical assays. Though, many reports confirmed the methylation of K36 of histone H3 by NSD2, proposing H3K36 as the main substrate of NSD2[107,141,142].

The substrate specificity profile of NSD2 was determined by methylating mutational scanning peptide SPOT arrays prepared using the sequence of histone H3 (29-43) as a template. The obtained data revealed a relatively short recognition motif, starting from residue G33 and reaching to residue P38. It indicates that NSD2 prefers aromatic and small residues at the -2 position (F>G>Y), with hydrophobic amino acids being tolerated as well. The positions -1 (I>L>V) and +2 (V>I>L>P) are more specific and hydrophobic residues are allowed at this sites. The +1 position is not that specific as the surrounding positions, however, basic and uncharged amino acids (K>V>R>Q) are preferred.

To show which of the reported substrates may be methylated by NSD2, a comparison of their sequences with the determined specificity profile could help. Interestingly, the substrate sequence motif of NSD2 is not matching with the K20 of H4, instead it is fitting with the residues surrounding K44. In addition, H1.5K168 matches very good to the specificity motif of NSD2. At position +2, the H4K44 and H1.5K168 substrates possess a more preferred amino acid (I46 for H4 and V170 for H1.5) compared to P38 of histone H3, which would explain the higher methyltransferase activity of NSD2 on H4K44 and H1.5K168 at peptide level (Figure 25). Although, the methylation activity of NSD2 was not tested on H3K4, H3K27 and H4K20 peptides in this study, the alignment of the corresponding sequences to the specificity profile reveals many mismatching residues, which makes a methylation very unlikely.

With the derived substrate specificity profile of NSD2, 217 proteins with 226 potential target lysines were identified and methylation on 45 of these substrates was observed at peptide level. 19 peptides showed stronger methylation than H3K36 and 15 had a comparable methylation intensity to the H3K36 substrate peptide. 13 of the 45 methylated substrate peptides revealed weaker methylation signals than the control peptide. In addition, 3 of the 22 tested protein domains (ATRX, FANCM and SET8) were methylated by NSD2, although the signals were weaker compared to the H3 control protein. Methylation on the target lysine was confirmed for ATRX and FANCM, whereas the mutation of the predicted target lysine of SET8 did not abrogate the methylation signal at the SET8 mutant protein. This suggests that SET8 has

more than one target lysine or even other residues, for example cysteine or arginine, that may be methylated by NSD2. It was not possible to identify the methylated residue(s) in SET8 in this study, however, an automethylation of SET8 was excluded. Furthermore, cellular methylation of the *in vitro* methylated substrate proteins ATRX and FANCM was determined in human HEK293 cells using a lysine methylation antibody. Considering that a substrate relatively weakly methylated *in vitro*, such as ATRX, shows a strong cellular methylation, it might be possible that NSD2 possesses additional substrates among the other 42 methylated peptides.

A look onto the already identified methylated non-histone substrates shows the importance of searching for such substrates and investigating the effect of methylation on protein functions and properties. ATRX and FANCM are known to have very important cellular functions in transcriptional regulation and DNA damage response:

- ATRX is named after its ATR-X syndrome (alpha-thalassemia X-linked mental retardation) characterized by mental retardation, developmental delay and distinctive facial features[187]. ATRX is a ATP-dependent chromatin-remodeling factor and is involved in transcriptional regulation. It has multiple other roles, including control of histone deposition[188]. It was also shown that ATRX interacts with many different other chromatin proteins, such as HP1α, EZH2 and MeCP2[189–191].

- FANCM (Fanconi anemia group M protein) is homologues to the archaeal DNA helicase/nuclease protein HEF and possesses a helicase/ATPase domain and an endonuclease domain. FANCM belongs to the FA (Fanconi anemia) core complex, which is important for monoubiquitination of FANCD2, a key step in the FA DNA damage response pathway[192]. Fanconi anemia is a rare genetic disease characterized by congenital abnormalities, bone marrow failure, genomic instability, and increased risk of cancer development. FA can be caused by mutation in any of the involved genes[193].

It is known that both proteins are posttranslationally phosphorylated, but until now no study showed methylation of ATRX and FANCM. Therefore it is important to further investigate the effects of the lysine methylation on the biological functions of the two novel NSD2 substrates.

Additionally, it was observed that NSD2 is subjected to automethylation. This was already shown for NSD1, where the lysine K1769 was identified as the target residue[165]. A candidate screening approach to identify the amino acid automethylated in NSD2, revealed lysine K992 as a possible target site. However, the generated and purified NSD2 mutants (K992A and K992R) still showed automethylation. This indicates either that lysine 992 was not the methylated residue or that other lysines or even cysteine might be methylated. Nevertheless, it would be interesting and important to identify the automethylation site of NSD2. Automethylation might also occur in cells and may alter some functions or properties of NSD2 or create binding sites for other proteins.

In addition, the substrate specificity profiles of two NSD2 somatic cancer mutations were elucidated. Recent studies identified the NSD2 E1099K and D1125N mutations within the catalytically active SET domain in several tumor samples[143]. The E1099K mutation was found in several different lymphoid malignancies, such as hypodiploid acute lymphoblastic leukemia and chronic lymphocytic leukemia. It was reported that these two mutants exhibit an enhanced histone methyltransferase activity, thereby leading to an increase in global H3K36me2 levels[143,144]. Since these mutations are located within the SET domain, it might be that they change the substrate specificity of the enzymes. However, the determination of the substrate specificity profiles of these two mutants, revealed comparable specificity motifs to the NSD2 wild-type, with only minor differences. Also methylation of the peptide and protein substrates with the cancer mutants did not show any additional candidates. Although it was reported that the NSD2 cancer mutants exhibit an increased MTase activity, the results of this study could not support these findings. Noteworthy, in the performed *in vitro* methylation assays, only a small part of the enzyme containing the catalytically active SET domain was used. It was already shown that the PHD domains of NSD2 are important for its biological function, and truncation of the enzyme can decrease the methyltransferase activity of NSD2[194]. Finally, the methylation of ATRX and FANCM by the NSD2 E1099K and D1125N mutant enzymes could be confirmed in human cells. An elevated methyltransferase activity for the somatic cancer mutants was observed towards ATRX in HEK293 cells. By contrast, all three NSD2 variants displayed similar activity on the FANCM substrate.

Interestingly, during detection of cellular methylation of FANCM with the H3K36me1-specific antibody, a band with higher molecular mass could be detected in addition. The fact that this signal was observed after purification of the YFP-tagged FANCM protein, suggests either this signal comes from a co-purified interaction partner of FANCM or that the used antibody detected a modified FANCM species with a higher molecular weight. In the first case it would be interesting to know why this protein got detected by the H3K36me1-specific antibody. The interaction partner may be methylated by NSD2 and detected by the antibody or it may be unspecific antibody binding. The second possibility is that FANCM was already modified. Such a modification might be ubiquitination or sumoylation, which would increase the weight enough to explain both FANCM species during western blot analysis. Indeed, western blot with anti-ubiquitin antibody showed an ubiquitination of FANCM, which makes co-purification and detection of an interaction partner very unlikely. Additionally, the ubiquitinated FANCM protein must be strongly methylated as well, because the detected anti methyl K antibody signal is much stronger than the corresponding band on the Ponceau S image. Since the second band was also observed in samples, where FANCM was expressed without NSD2, this indicates that endogenous NSD2 enzyme efficiently methylated FANCM in cells.

Additionally, the effect of somatic missense mutations of histone H3 on the NSD2 methyltransferase activity was analyzed. In the past years, recurrent mutations at different positions of the amino-terminal tail of histone H3 were found in pediatric brain and bone malignancies. Since these positions undergo important posttranslational modifications, abnormalities, such as alteration of the target residue, may lead to changed histone modifications, protein-protein interactions, chromatin structure and dysregulation of gene transcription. Peptide arrays covering residues from 32 to 41 of H3 were synthesized. These membranes contain the most prevalent single missense mutations of histone H3. A comparison of the observed methylation intensities between the H3K36 control and the somatic missense mutant peptides showed clear discrepancies. However, the different intensities of each mutant peptide were in agreement with the characterized specificity profile of NSD2. Peptides with a more preferred residue than the native amino acid, revealed stronger activity and peptides with less preferred amino acids were weakly methylated. In addition, these arrays were methylated by NSD1 as well, however, the methylation intensities showed only minor differences compared to the NSD2 variants.

The influence of a K36M mutant peptide on the methyltransferase activity of NSD2 was tested as well. Different concentrations of H3K36 wild-type peptides were methylated by NSD2 in presence of various concentrations of K36M-containing peptide. The assay results were analyzed by the least-squares method. The analysis revealed a competitive inhibition of NSD2 by the K36M peptide on the K36 substrate. Finally, a K_M value of $71\,\mu M$ for methylation of the K36 peptide and a K_I value of $47\,\mu M$ for inhibition of the methylation reaction by the K36M peptide was determined, indicating that the K36M peptide inhibitor binds stronger to NSD2 than to the K36 peptide substrate. One reason for this is the proper fitting of the M36 side chain in the hydrophobic pocket, which is probably formed by residues Y1092, M1119, F1177 and Y1179 (corresponding conserved residues were shown to be responsible for interaction to K/M36 in SETD2[195]). Additionally, for SETD2 it was shown that the side chain of K36M is further stabilized by sulfur-aromatic and CH-π interactions through stacking of the side chain against the aromatic ring of Y1666. The same might hold true for NSD2, since the enzyme contains a conserved tyrosine (Y1179) in the catalytic pocket[195].

4.3 Specificity Analysis of Clr4 and Identification of Novel Peptide Substrates

In this part of the thesis the substrate specificity profile of Clr4, the *Schizosaccharomyces pombe* homolog of the human histone lysine methyltransferase SUV39H1, was characterized. Clr4 catalyzes trimethylation of H3K9, which is important for formation, maintenance and spreading of heterochromatin. Similar to other H3K9 methyltransferases, such as G9a or SUV39H1, the specificity profile of Clr4 showed a strong preference for an RK motif (R8 and K9). Furthermore, Clr4 displayed high specificity at position +3, where it tolerated only glycine (G12).

Next, the interaction partners of Clr4 that contained its substrate sequence motif were selected and tested at the peptide level. The results revealed 3 strongly methylated and 4 weakly methylated substrate peptides. Since Clr4 and its interaction partners are from *S. pombe*, a collaborating laboratory conducted further experiments to investigate the methylation of the substrate candidates at protein level.

4.4 Development of an Advanced Non-radioactive, High-throughput PKMT Activity Assay

In this project, a non-radioactive, high-throughput PKMT activity assay was developed by employing reading domains to recognize methylated lysine residues. This assay overcomes many disadvantages of other PKMT activity assays, such as the usage of radioactivity or methyl-specific antibodies. This microplate based assay can be used to screen for novel PKMT inhibitors.

It was investigated if the reading domain, in this case HP1β, is able to detect H3K9 peptides with different methylation levels. The results showed a good discrimination by the reading domain between unmethylated and trimethylated H3K9 peptides. In addition, the results demonstrated that HP1β detected *in vitro* methylated substrate with a comparable intensity to synthetic trimethylated H3K9 peptide as well. Finally, the results revealed that this assay can be used to screen PKMT inhibitors. Using this experimental setting, chaetocin, an inhibitor of SUV39H1, was tested and an IC_{50} value of 480 nM of the inhibition of SUV39H1 was calculated. This value is close to the published IC_{50} of 600 nM[173]. In conclusion, the newly developed assay is able to successfully detect peptide and protein methylation with a very good dynamic range and high sensitivity by employing natural reading domains. In addition, the microplate format allows a medium- to high-throughput campaign analysis and could therefore be used for PKMT inhibitor screens.

5 Conclusions

Protein methyltransferases play an important role in many biological processes. They are also involved in numerous diseases and cancers. By methylation of target residues the enzymes are able to influence properties, functions, interactions and localization of their targets. The characterization of MTases is important to gain deeper insights about these enzymes and their biological functions. Analysis of the substrate specificity, activity and the identification of novel substrates, help to provide new detailed information about chemical properties of the substrates that are recognized by the enzymes. This knowledge could be used to develop inhibitors to prevent or treat related diseases. Moreover, novel substrates of protein MTases can be identified, which helps to uncover the regulatory roles of these enzymes in cells. Determination of the substrate specificity profile and the identification of novel substrates was already successfully applied for different MTases, such as Dim-5, G9a, SET7/9 and NSD1.

The aim of this study was to characterize the substrate specificity of three important protein methyltransferases and to identify novel substrates and eventually confirm the methylation and its consequences in human cells. Specificity analysis of HEMK2, a protein glutamine methyltransferase, identified a G-**Q**-X$_3$-R recognition motif, which is essential for the methylation activity. Moreover, it was demonstrated that HEMK2 methylates two novel substrates, CHD5 and NUT, in mammalian cells. Both of these targets have important cellular functions. In addition, several other substrate candidates were methylated at protein and peptide level *in vitro*. Furthermore, a Pan-Qme-specific antibody was developed for detecting cellular methylation. This should be further refined for use in proteomic studies, such as to enrich for proteins containing methyl glutamine modification.

Similarly, the substrate specificity of NSD2, a protein lysine methyltransferase which methylates histone H3 at K36, demonstrated that the enzyme recognizes the residues from 33 to 38. Like its family member NSD1, it also prefers hydrophobic residues surrounding the target lysine. With the derived motif, numerous potential substrates were identified and methylation of several of these substrate candidates could be confirmed at peptide level. However, only 3 proteins were shown to be methylated at protein level *in vitro*. One potential reason for this discrepancy in the number of targets between the two assays might be that NSD2 prefers target lysines surrounded by a hydrophobic motif. This may be not accessible for methylation in a folded protein. Moreover, methylation of two novel important protein substrates, ATRX and FANCM were identified in human cells.

Furthermore, the specificity profile analysis of Clr4 identified a crucial „RK" recognition motif, as observed for its human homolog SUV39H1 or other H3K9 methyltransferases, such as G9a. Apart from this RK dipeptide, Clr4 requires a glycine at the position +3 (G12) for proper

MTase activity. Based on this, the methylation of 6 novel substrates by Clr4 could be shown at peptide level. In the future, it should be further investigated if Clr4 is able to methylate the identified peptide substrates at protein level *in vitro* and *in vivo* as well, and to study the effects of methylation on the properties of the putative substrates.

In summary, these novel and important results expand the product portfolio of HEMK2, NSD2 and Clr4. They support the notion that protein methylation is a general PTM both in humans and lower organisms and hints that many more methylation sites need to be identified by similar or proteomic approaches. For the future, the biological role of the novel substrates of HEMK2 and NSD2 should be elucidated. The introduced methylation may change the properties, localization or cellular functions of these proteins.

6 Materials and Methods

6.1 The Glutamine Methyltransferase HEMK2

6.1.1 Cloning, Site-directed Mutagenesis, Expression and Purification

The bacterial expression pRSF-Duet1 vector containing the full-length murine HEMK2 and TRM112 genes, the pRSET vector with the human eRF1 gene and the mammalian expression constructs (pcDNA3-HEMK2 and pcDNA4-TRM112) were kindly provided by Dr. G. L. Xu.

The sequences encoding for the putative human substrate protein domains were amplified from cDNA of HEK293 cells and cloned into the pGEX-6P-2 vector (GE Healthcare) as GST fusion proteins (Table 7).

Table 7: *List with information of putative novel HEMK2 substrate proteins, which were selected for investigation of methylation at protein level.*

Name	Abbreviation	Domain boundaries (aa)	NCBI accession number
5-aminolevulinate synthase, erythroid-specific, mitochondrial	ALAS2	$360 - 533$	NP_000023.2
ADP-ribosylation factor-like protein 2	ARL2	$2 - 184$	NP_001658.2
Histone-lysine N-methyltransferase ASH1L	ASH1L	$1119 - 1333$	NP_060959.2
AMP deaminase 2	AMPD2	$2 - 135$	NP_001244289.1
Ankyrin repeat domain-containing protein 34A	ANKRD34A	$5 - 235$	NP_001034977.1
ATP-binding cassette sub-family A member 2	ABCA2	$168 - 403$	NP_997698.1
BEN domain-containing protein 7	BEND7	$9 - 282$	XP_011517694.1
Cadherin-23	CDH23	$3001 - 3353$	NP_071407.4
CD97 antigen	CD97	$197 - 529$	NP_510966.1
Chromodomain-helicase-DNA-binding protein 5	CHD5	$1234 - 1530$	NP_056372.1
Collagen alpha-1(XIX) chain	COL19A1	$344 - 577$	NP_001849.2
Collagen alpha-6(IV) chain	COL4A6	$1464 - 1690$	NP_001838.2
C-X-C chemokine receptor type 3	CXCR3	$249 - 343$	NP_001136269.1

Cytochrome P450 4F11	CYP4F11	7 – 140	NP_001122404.1
D(3) dopamine receptor	DRD3	11 – 274	NP_000787.2
DNA polymerase subunit gamma-1	POLG	154 – 387	NP_001119603.1
DNA repair protein REV1	REV1	568 – 818	NP_057400.1
E3 ubiquitin-protein ligase HUWE1	HUWE1	3622 – 3945	NP_113584.3
E3 ubiquitin-protein ligase RNF220	RNF220	282 – 440	NP_060620.2
F-box only protein 30	FBXO30	453 – 661	NP_115521.3
F-box/LRR-repeat protein 15	FBXL15	1 – 300	NP_077302.3
Gamma-aminobutyric acid receptor subunit delta	GABRD	279 – 439	NP_000806.2
Gamma-aminobutyric acid receptor subunit rho-2	GABRR2	223 – 465	P28476
GH3 domain-containing protein	GHDC	325 – 529	NP_115873.1
Glucokinase regulatory protein	GCKR	360 – 613	NP_001477.2
Glycogenin-2	GYG2	3 – 239	NP_003909.2
Guanine nucleotide-binding protein subunit alpha-12	GNA12	42 – 366	NP_031379.2
Endogenous retroviral sequence K 6	ERVK6	1202 – 1419	Q9WJR5
HERV-K_5q33.3 provirus ancestral Pol protein	HERVK5	252 – 495	NW_007925255.1
Kinesin-like protein KIF23	KIF23	113 – 370	NP_612565.1
Latent-transforming growth factor beta-binding protein 4	LTBP4	13 – 381	NP_001036009.1
Leucine-rich repeat-containing protein 41	LRRC41	410 – 636	NP_006360.3
Leukotriene-B(4) omega-hydroxylase 1	CYP4F2	121 – 360	NP_001073.3
Leukotriene-B(4) omega-hydroxylase 2	CYP4F3	109 – 343	NP_000887.2
Mediator of RNA polymerase II transcription subunit 24	MED24	730 – 985	NP_055630.2
Neuropeptide W	NPW	17 – 157	NP_001092926.2

Pleckstrin	PLEK	$2-350$	NP_002655.2
PR domain zinc finger protein 8	PRDM8	$12-195$	NP_001092873.1
Probable E3 ubiquitin-protein ligase HERC3	HERC3	$122-460$	NP_055421.1
Protamine-2	PRM2	$1-102$	P04554
Protein NUT	NUT	$867-1132$	NP_786883.1
Protein transport protein Sec24C	SEC24C	$702-970$	NP_004913.2
Ras-related protein Rab-12	RAB12	$1-233$	NP_001020471.2
Rho guanine nucleotide exchange factor 10	ARHGEF10	$1107-1343$	NP_055444.2
Ribosomal RNA processing protein 1 homolog A	RRP1	$219-461$	NP_003674.1
Serine/arginine repetitive matrix protein 4	SRRM4	$89-334$	NP_919262.2
Solute carrier family 25 member 47	Solcar1	$2-299$	NP_997000.2
SPARC-related modular calcium-binding protein 1	SMOC1	$13-217$	NP_001030024.1
Sterile alpha motif domain-containing protein 7	SAMD7	$71-416$	Q7Z3H4
Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	SVEP1	$6-239$	NP_699197.3
Tetratricopeptide repeat protein 9B	TTC9B	$42-229$	NP_689692.2
TNFAIP3-interacting protein 2	TNIP2	$1-417$	NP_077285.3
Transforming growth factor beta-3	TGFB3	$159-405$	NP_003230.1
Unconventional myosin-XVIIIa	MYO18A	$105-355$	NP_510880.2
WD40 repeat-containing protein SMU1	SMU1	$5-511$	NP_060695.2
Zinc finger and SCAN domain-containing protein 10	ZSCAN10	$364-521$	Q96SZ4
Zinc finger protein ZFPM1	ZFPM1	$1-264$	NP_722520.2

For bacterial expression, the plasmids were transformed into *E. coli* BL21-CodonPlus (DE3) cells (Novagen, USA). These were grown in LB medium at $37\,°C$ until an OD_{600} of 0.6 to 0.8 was reached. Protein expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The culture was then either shifted to $30\,°C$ for 4 h or to $20\,°C$ over night (14 to 16 h). Afterwards, the cells were harvested by centrifugation at 4.500 rpm, washed once with STE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA and 100 mM NaCl) and the cell pellet was stored at $-20\,°C$ until purification.

For purification the cell pellet was thawed on ice, resuspended in affinity tag-specific sonication buffer and lysed by sonication. Next, the lysate was centrifuged at 18.000 rpm for 90 min at $4\,°C$ and depending on the affinity tag, the supernatant was passed through Glutathione Sepharose 4B (GE Healthcare) or Nickel-Nitrilotriacetic acid (Ni-NTA; Genaxxon) resin. Afterwards, the bound protein was washed with sonication buffer. For GST-tagged proteins, the beads were additionally washed with washing buffer containing high amounts of salt. The bound protein was eluted with similar buffers containing excess of either glutathione (for GST purification) or imidazole (for Ni-NTA purification), and then dialyzed against low glycerol dialysis buffer 1 for 3 h and afterwards overnight against high glycerol dialysis buffer 2. The composition of the used buffers are shown in Table 8.

Table 8: *Buffers used for GST-fused (left column), His$_6$-fused (middle column) and MBP-fused protein purification (right column).*

GST-Tag Purification	His$_6$-Tag Purification	MBP-Tag Purification
Sonication buffer (pH 7.5)	**Sonication buffer (pH 7.5)**	**Sonication buffer (pH 7.5)**
50 mM Tris	30 mM KPI buffer	30 mM KPI buffer
150 mM NaCl	500 mM KCl	500 mM KCl
1 mM DTT	0.2 mM DTT	0.2 mM DTT
—	1 mM EDTA	1 mM EDTA
5 % Glycerol	10 % Glycerol	10 % Glycerol
Washing buffer (pH 8.0)	**0.5 M KPI buffer (pH 7.2)**	**0.5 M KPI buffer (pH 7.2)**
50 mM Tris	358.5 mL 1 M K$_2$HPO$_4$	358.5 mL 1 M K$_2$HPO$_4$
500 mM NaCl	142.5 mL 1 M KH$_2$PO$_4$	142.5 mL 1 M KH$_2$PO$_4$
1 mM DTT	500 mL H$_2$O	500 mL H$_2$O
5 % Glycerol	—	—
Elution buffer (pH 8.0)	**Elution buffer (pH 7.2)**	**Elution buffer (pH 7.5)**
40 mM reduced Glutathione	220 mM Imidazole	20 mM Maltose Monohydrate
in Washing buffer	in Sonication buffer	in Sonication buffer
Dialysis buffer 1 (pH 7.4)	**Dialysis buffer 1 (pH 7.2)**	**Dialysis buffer 1 (pH 7.5)**
20 mM Tris	20 mM HEPES	20 mM HEPES
100 mM KCl	200 mM KCl	200 mM KCl
0.5 mM DTT	0.2 mM DTT	0.2 mM DTT
—	1 mM EDTA	1 mM EDTA
10 % Glycerol	10 % Glycerol	10 % Glycerol
Dialysis buffer 2 (pH 7.4)	**Dialysis buffer 2 (pH 7.2)**	**Dialysis buffer 2 (pH 7.5)**
20 mM Tris	20 mM HEPES	20 mM HEPES
500 mM KCl	200 mM KCl	200 mM KCl
0.2 mM DTT	0.2 mM DTT	0.2 mM DTT
—	1 mM EDTA	1 mM EDTA
60 % Glycerol	65 % Glycerol	65 % Glycerol

The target glutamine mutation of the target substrates was performed by site-directed mutagenesis using PCR-megaprimers according to the protocol of Jeltsch & Lanio[196]. In addition to the glutamine to arginine mutation, silent mutations were introduced to allow the identification of the plasmids containing the mutated targets by specific restriction sites. The successful mutagenesis of the novel substrates was confirmed by restriction digest and DNA sequencing. In Table 9 all methylated substrates with the target glutamine to arginine mutations are listed.

Table 9: *Target glutamine to arginine mutations of methylated HEMK2 substrates.*

Name	Abbreviation	Target Q mutation
AMP deaminase 2	AMPD2 Mut	Q6R
Ankyrin repeat domain-containing protein 34A	ANKRD34A Mut	Q15R
ATP-binding cassette sub-family A member 2	ABCA2 Mut	Q302R
Chromodomain-helicase-DNA-binding protein 5	CHD5 Mut	Q1390R
GH3 domain-containing protein	GHDC Mut	Q489R
Histone-lysine N-methyltransferase ASH1L	ASH1L Mut	Q1220R
Protein NUT	NUT Mut	Q1046R
Rho guanine nucleotide exchange factor 10	ARHGEF10 Mut	Q1313R
Ribosomal RNA processing protein 1 homolog A	RRP1 Mut	Q427R
Transforming growth factor beta-3	TGFB3 Mut	Q293R
Zinc finger and SCAN domain-containing protein 10	ZSCAN10 Mut	Q428R

6.1.2 Synthesis of Peptide SPOT Arrays

Peptide array were synthesized by an Autospot peptide array synthesizer (Intavis AG, Köln) using the SPOT synthesis method[197]. Each spot had a diameter of 2 mm and contained approximately 9 nmol of peptide (Autospot Reference Handbook, Intavis AG). The successful synthesis of each peptide array was confirmed by bromophenol blue staining of the membranes[179].

6.1.3 *In vitro* Methylation of the Peptide SPOT Arrays

All peptide arrays were washed for 10 min in methylation buffer containing 10 mM Tris (pH 7.6), 50 mM KCl, 10 mM Mg(OAc)$_2$ and 1 mM DTT. Then, the membranes were incubated for 60 min in methylation buffer containing 1.3 µM HEMK2 and 0.76 µM labeled [methyl-^{3}H]-SAM (Perkin Elmer) at 25 °C. Afterwards, the arrays were washed five times with 50 mM NH$_4$HCO$_3$ and 1 % SDS and then incubated for 5 min in Amplify NAMP100V solution (GE Healthcare). The membranes were exposed to Hyperfilm^{TM} high performance autoradiography films (GE Healthcare) in the dark, for several days at −80 °C. Film development was performed on an Optimus TR developing machine. The developed films were scanned and the intensities were analyzed with the Phoretix^{TM} Array software. Background substraction was performed by measuring an area on the array, outside of the spot-containing region. The intensities of the spots were quantified and the normalization was performed in Microsoft Office Excel by presenting all spot intensities relative to the minimum and the maximum intensity spot in each array. The used equation for calculating the normalized intensity of each spot was defined as:

$$\text{Spot intensity (normalized)} = \frac{(\text{Spot intensity (raw)} - \text{Min intensity (raw)})}{(\text{Max intensity (raw)} - \text{Min intensity (raw)})}$$

Information of synthesized peptides on array A identified from search profile 1:

Table 10: *List of putative novel HEMK2 substrates identified in Scansite searches using the stringent specificity profile as shown in Table 1. The predicted target glutamine is printed in bold. The position of the corresponding peptide spots in Figure 14A are indicated.*

Name	Swiss Prot no.	Position on Array	Target Q Position	Sequence
eFR1 WT	P62495	A1	185	KKHGRGG**Q**SALRFAR
eRF1 Mutant		A2	A185	KKHGRGG**A**SALRFAR
B-cell CLL/lymphoma 9-like protein	Q86UU0	A3	755	LSPPMG**Q**SGLREVDP
Coiled-coil alpha-helical rod protein 1	Q8TD31	A4	667	ARKEEG**Q**RLARRLQE
C-C chemokine receptor type 10	P46092	A5	188	DGQREG**Q**RRCRLIFP
B-cell receptor CD22	P20273	A6	116	HLNDSG**Q**LGLRMESK
Uncharacterized protein C7orf63	A5D8W1	A7	255	DPDPSG**Q**LLFRSSEI
Chromodomain-helicase-DNA-binding protein 5	Q8TDI0	A8	1390	EERPEG**Q**SGRRQSRR

Chondroitin sulfate glucuronyltransferase	Q9P2E5	A9	137	LLYFTG**Q**RGARAPAG
UMP-CMP kinase 2, mitochondrial	Q5EBM0	A10	171	EADPRG**Q**LWQRLWEV
Collagen alpha-5(VI) chain	A8TX70	A11	1559	SRGREG**Q**RGLRGVSG
Collagen alpha-1(XIX) chain	Q14993	A12	407	PPGKEG**Q**RGRRGKTG
Cytochrome P450 1B1	Q16678	A13	37	ATVHVG**Q**RLLRQRRR
Carnitine O-palmitoyltransferase 2, mitochondrial	P23786	A14	33	AGSGPG**Q**YLQRSIVP
C-X-C chemokine receptor type 3	P49682	A15	295	LLVSRG**Q**RRLRAMRL
Disks large homolog 5	Q8TDM6	A16	1108	KVDELG**Q**KRRRPKSA
Eukaryotic peptide chain release factor subunit 1	P62495	A17	185	KHGRGG**Q**SALRFARL
Protein FAM113B	Q96HM7	A18	47	RLLTPG**Q**LRARGELN
Protein FAM123C	Q8N944	A19	708	GSGLFG**Q**RWARGPDM
F-box/LRR-repeat protein 2	Q9UKC9	A20	414	TAVAGSG**Q**RLCRCCV
Extracellular matrix protein FRAS1	Q86XX4	A21	265	LRCGKG**Q**SRARRHGQ
FRAS1-related extracellular matrix protein 3	P0C091	B1	1991	RLPVGG**Q**LGARFPTT
Glycogenin-2	O15488	B2	57	GALVLG**Q**SLRRHRLT
Probable E3 ubiquitin-protein ligase HERC6	Q8IVU3	B3	54	GDNSRG**Q**LGRRGAQR
Hermansky-Pudlak syndrome 1 protein	Q92902	B4	686	LVQQAG**Q**LARRLWEA
E3 ubiquitin-protein ligase HUWE1	Q7Z6Z7	B5	3783	QMVREG**Q**RARRQQQA
Intersectin-1	Q15811	B6	886	TVPSAG**Q**LRQRSAFT
Jerky protein homolog	O75564	B7	431	GSSCPG**Q**LRQRQAAS
Uncharacterized protein KIAA1908	Q96PY0	B8	202	GLSHLG**Q**SLCRTVKE
Laminin subunit gamma-3 LAMC3	Q9Y6N6	B9	781	THCPPG**Q**RGRRCEVC
Laminin subunit gamma-3 LAMC3	Q9Y6N6	B10	287	GPDVAG**Q**LACRCQHN

Protein	Accession	ID	Position	Sequence
Mineralocorticoid receptor	P08235	B11	916	CPNNSG**Q**SWQRFYQL
Mediator of RNA polymerase II transcription subunit 12	Q93074	B12	813	LGGEDG**Q**KRRRNRPE
Nucleotide-binding oligomerization domain-containing protein 1	Q9Y239	B13	669	QSQKVG**Q**LAARGICA
Neuropeptide W	Q8N729	B14	137	DFSGAG**Q**RLRRDVSR
Nuclear receptor subfamily 5 group A member 2	O00482	B15	498	QTEKFG**Q**LLLRLPEI
Pleckstrin homology domain-containing family G member 4B	Q96PX9	B16	1002	NLKEQG**Q**LRCRDEFI
PRAME family member 10	O60809	B17	15	LLELAG**Q**SLLRNQFL
PRAME family member 13	Q5VWM6	B18	15	LLELAG**Q**SLLRDQAL
PRAME family member 14	Q5SWL7	B19	15	LLELAG**Q**SLLRDQAL
PRAME family member 16	Q5VWM1	B20	15	LLELAG**Q**SLLRNQFL
PRAME family member 17	Q5VTA0	B21	15	LLELAG**Q**SLLRNQFL
PRAME family member 18	Q5VWM3	C1	15	LLELAG**Q**SLLRDQAL
PRAME family member 1	O95521	C2	15	LLELAG**Q**SLLRDQAL
PRAME family member 2	O60811	C3	15	LLELAG**Q**SLLRDQAL
Patched domain- containing protein 3	Q3KNS1	C4	314	GSLGMG**Q**LLLRAKAM
Probable peptidyl-tRNA hydrolase	Q86Y79	C5	55	GMAVLG**Q**LARRLGVA
E3 ubiquitin-protein ligase RLIM	Q9NVW2	C6	208	VPPTRG**Q**RRARSRSP
Ribosomal RNA processing protein 1 homolog A	P56182	C7	427	QPRGRG**Q**RGARQRRR
RuvB-like 2	Q9Y230	C8	49	SQGMVG**Q**LAARRAAG
Sodium-dependent neutral amino acid transporter B(0)AT1	Q695T7	C9	94	LEFAIG**Q**RLRRGSLG
Protein transport protein Sec24C	P53992	C10	819	YTSCAG**Q**RRLRIHNL
Protein transport protein Sec24D	O94855	C11	757	YTTISG**Q**RRLRIHNL

Scavenger receptor class B member 1	Q8WTV0	C12	472	SQVGAG**Q**RAARADSH
N-lysine methyltransferase SET8	Q9NQR1	C13	28	AAVVAG**Q**RRRRLGRR
Transcription elongation factor SPT6	Q7KZ85	C14	116	VKVKRG**Q**KYRRVK
Probable leucine-tRNA ligase, mitochondrial	Q15031	C15	610	FRLPSG**Q**YLQREEVD
TNFAIP3-interacting protein 2	Q8NFZ5	C16	29	LYHEAG**Q**RLRRLQDQ
Thymidine phosphorylase	P19971	C17	429	LLVDVG**Q**RLRRGTPW
RanBP-type and C3HC4-type zinc finger-containing protein 1	Q9BYM8	C18	105	QQWVIG**Q**RLARDQET
Vacuolar protein sorting-associated protein 72 homolog	Q15906	C19	154	VQERQG**Q**SRRRKGPH
DDB1- and CUL4- associated factor 11	Q8TEB1	C20	119	VELATG**Q**LGLRRAAQ
Y+L amino acid transporter 1	Q9UM01	C21	400	GLSIVG**Q**LYLRWKEP
Y+L amino acid transporter 2	Q92536	D1	408	GLSVVG**Q**LYLRWKEP
Zinc finger homeobox protein 3	Q15911	D2	2946	SGDRPG**Q**KRFRTQMT
Zinc finger matrin-type protein 3	Q9HA38	D3	193	ESSELG**Q**RRARKEGN
free space		D4		
free space		D5		
free space		D6		
AMP deaminase 2	Q01433	D7	6	RG**Q**GLFRLRSRCFLH
Anthrax toxin receptor 1	Q9H6X2	D8	28	VLICAG**Q**GGRREDGG
Apolipoprotein M	O95445	D9	139	MLNETG**Q**GYQRFLLY
ADP-ribosylation factor-like protein 2	P36404	D10	70	IWDVGG**Q**KSLRSYWR
ADP-ribosylation factor-like protein 6	Q9H0F7	D11	73	VFDMSG**Q**GRYRNLWE
BEN domain-containing protein 7	Q8N7W2	D12	78	SGQFSG**Q**YGTRSRTF
Biotin-protein ligase	P50747	D13	417	GEIKSG**Q**LSLRFVSS
Cadherin-23	Q9H251	D14	3263	GDHSPG**Q**GSLRFRHK

CD97 antigen	P48960	D15	376	KNVTMG**Q**SSARMKLN
Cyclin-dependent kinase 10	Q15131	D16	274	KLPLVG**Q**YSLRKQPY
Protein CIP2A	Q8TCG1	D17	39	LEVISG**Q**KLTRLFTS
Uncharacterized protein C12orf56	Q8IXR9	D18	193	KLSLHG**Q**GAFRPLPS
Leukotriene-B(4) omega-hydroxylase 1	P78329	D19	254	YLTPDG**Q**RFRRACRL
Leukotriene-B(4) omega-hydroxylase 2	Q08477	D20	254	YLTPDG**Q**RFRRACRL
Cytochrome P450 4F11	Q9HBI6	D21	254	YLTPDG**Q**RFRRACHL
Putative uncharacterized protein encoded LINC00526	Q96FQ7	E1	29	GGLPPG**Q**YATRMTGQ
Cytoplasmic tRNA 2-thiolation protein 2	Q2VPK5	E2	446	WAQRCG**Q**GACRREDP
DCC-interacting protein 13-alpha	Q9UKG1	E3	474	QAKAFG**Q**GGRRTNPF
DNA polymerase subunit gamma-1	P54098	E4	330	PPTKQG**Q**KSQRKARR
D(3) dopamine receptor	P35462	E5	144	YQHGTG**Q**SSCRRVAL
FERM, RhoGEF and pleckstrin domain-containing protein 1	Q9Y4F1	E6	407	AESPGG**Q**SCRRGKEP
F-box only protein 18	Q8NFZ0	E7	54	KRGSRG**Q**GSQRCIPE
F-box only protein 30	Q8TB52	E8	525	FTFVCG**Q**LFRRKEFS
F-box/LRR-repeat protein 15	Q9H469	E9	126	ALGGCG**Q**LSRRALGA
von Willebrand factor A domain-containing protein 7	Q9Y334	E10	135	DAERLG**Q**GRARLVGA
GH3 domain-containing protein	Q8N2G8	E11	489	RVHLVG**Q**GAFRALRA
Guanine nucleotide-binding protein subunit alpha-12	Q03113	E12	231	MVDVGG**Q**RSQRQKWF
G-protein coupled receptor 98	Q8WXG9	E13	3490	FIWEMG**Q**SSFRYFQS
Solute carrier family 25 member 47	Q6Q0C1	E14	252	QADGQG**Q**RRYRGLLH
E3 ubiquitin-protein ligase HECW1	Q76N89	E15	199	DETVQG**Q**GSRRLISF
Probable E3 ubiquitin-protein ligase HERC3	Q15034	E16	351	WAAHSG**Q**LSARADRF

Iroquois-class homeodomain protein IRX-2	Q9BZI1	E17	405	NAALQG**Q**GLLRYNSA
Serine/arginine repetitive matrix protein 4	A7MD48	E18	215	RSPEEG**Q**KSRRRHSR
Keratin-associated protein 1-1	Q07627	E19	161	RPSYCG**Q**SCCRPVCC
Keratin-associated protein 1-3	Q8IUG1	E20	161	RPSYCG**Q**SCCRPVCC
Keratin-associated protein 1-4	P0C5Y4	E21	105	RPSYCG**Q**SCCRPACC
Keratin-associated protein 1-5	Q9BYS1	F1	161	RPSYCG**Q**SCCRPVCC
Leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 3	P0C6S8	F2	581	PAAAAG**Q**GGARKFNM
Protein LZIC	Q8WZA0	F3	105	AKKQPG**Q**LRTRLAEM
Melanoma-associated antigen 8	P43361	F4	5	G**Q**KSQRYKAEEGLQA
Microtubule-associated serine/threonine-protein kinase 4	O15021	F5	2449	SPSATG**Q**SSFRSTAL
Melanin-concentrating hormone receptor 1	Q99705	F6	40	GACAPG**Q**GGRRWRLP
Netrin-4	Q9HB63	F7	369	QHNTEG**Q**YCQRCKPG
Protein NUT	Q86Y26	F8	1046	HHASGG**Q**GSQRASHL
Otopetrin-3	Q7RTS5	F9	504	SLLELG**Q**GLQRASLA
Serine/threonine-protein phosphatase 2A regulatory subunit B" subunit gamma	Q969Q6	F10	185	LYDVAG**Q**GYLRESDL
Pleckstrin homology-like domain family B member 1	Q86UU1	F11	828	ERELAG**Q**GLLRSKAE
Pleckstrin	P08567	F12	107	KCIEGG**Q**KFARKSTR
PR domain zinc finger protein 8	Q9NQV8	F13	111	AYIKNG**Q**LFYRSLRR
Ras-related protein Rab-12	Q6IQ22	F14	28	PALSGG**Q**GRRRKQPP
E3 ubiquitin-protein ligase RNF25	Q96BH1	F15	439	PRLPRG**Q**GAYRPGTR
Sterile alpha motif domain-containing protein 7	Q7Z3H4	F16	179	FEESWG**Q**RCRRLRKN
Semaphorin-4G	Q9NTN9	F17	607	MGLSDG**Q**GGYRVGVD

Protein	Accession	ID	Position	Sequence
Histone-lysine N-methyltransferase SETDB1	Q15047	F18	415	LEKKQG**Q**LRTRPNMG
SH2B adapter protein 2	O14492	F19	589	PRPVEG**Q**LSARSRSN
Putative E3 ubiquitin-protein ligase SH3RF2	Q8TEC5	F20	506	GPGTLG**Q**GSLRKGRS
Somatoliberin	P01286	F21	47	YRKVLG**Q**LSARKLLQ
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1	Q96GM5	G1	46	MGPAPG**Q**GLYRSPMP
Transforming growth factor beta-3	P10600	G2	293/296	RLDNPG**Q**GG**Q**RKKRA
Transforming growth factor beta-3	P10600	G3	296	RLDNPGAGG**Q**RKKRA
Transforming growth factor beta-3	P10600	G4	293	RLDNPG**Q**GGARKKRA
Tetratricopeptide repeat protein 9B	Q8N6N2	G5	73	AFKAEG**Q**RCYREKKF
E3 ubiquitin-protein ligase UBR5	O95071	G6	1247	KTLIAG**Q**KSARLDLL
Uromodulin	P07911	G7	208	WYRFVG**Q**GGARMAET
USP6 N-terminal-like protein	Q92738	G8	750	RPETQG**Q**SWTRDASR
Putative zinc finger protein 852	Q6ZMS4	G9	122	EGVLKG**Q**KSYRCDEC
Zinc finger B-box domain-containing protein 1	A8MT70	G10	628	RITLAG**Q**KSQRPSTA
Probable palmitoyltransferase ZDHHC1	Q8WTX9	G11	37	SPELQG**Q**RSRRNGWS
Zinc finger homeobox protein 2	Q9C0A1	G12	2065	VPDGMG**Q**RRYRTQMS
Zinc finger protein 142	P52746	G13	143	KAVDKG**Q**GAQRLEGD
Zinc finger protein 167	Q9P0L1	G14	380	EGVLKG**Q**KSYRCDEC
Zinc finger protein 407	Q9C0G0	G15	503	QEAEQG**Q**GSARPPDS
Zinc finger protein 787	Q6DD87	G16	352	VCSSCG**Q**SYYRAGGE
Zinc finger and SCAN domain-containing protein 10	Q96SZ4	G17	428	LCSHCG**Q**SFQRRSSL
eFR1 WT	P62495	G18	185	KKHGRGG**Q**SALRFAR
eRF1 Mutant		G19	A185	KKHGRGG**A**SALRFAR

Information of synthesized peptides on array B identified from search profile 2:

Table 11: *List of putative novel HEMK2 substrates identified in Scansite searches using the relaxed specificity profile as shown in Table 1. The predicted target glutamine is printed in bold. The position of the corresponding peptide spots in Figure 14B are indicated.*

Name	Swiss Prot no.	Position on Array	Target Q Position	Sequence
eRF1 WT	P62495	A1	185	KHGRGG**Q**SALRFARL
eRF1 Mutant		A2	A185	KHGRGG**A**SALRFARL
Chromodomain-helicase-DNA-binding protein 5 WT	Q8TDI0	A3	1390	EERPEG**Q**SGRRQSRR
Chromodomain-helicase-DNA-binding protein 5 Mutant		A4	A1390	EERPEG**A**SGRRQSRR
DNA-3-methyladenine glycosylase	P29372	A5	20	FCRRMG**Q**KKQRPARA
ATP-binding cassette sub-family A member 2	Q9BZC7	A6	302	DAVCSG**Q**AAARARRF
Abhydrolase domain-containing protein 3	Q8WU67	A7	92	CWEGRG**Q**TLLRPFIT
Acyl-CoA synthetase family member 3, mitochondrial	Q4G176	A8	465	VVFKDG**Q**YWIRGRTS
Disintegrin and metalloproteinase domain-containing protein 15	Q13444	A9	199	PEHPLG**Q**RHIRRRRD
Neuroblast differentiation-associated protein AHNAK	Q09666	A10	98	RSPEPG**Q**TWTREVFS
Ankyrin repeat domain-containing protein 34A	Q69YU3	A11	15	LLRAVG**Q**GKLRLARL
Adenomatous polyposis coli protein 2	O95996	A12	814	SPFLQG**Q**ALARTPPT
Rho guanine nucleotide exchange factor 10	O15013	A13	1313	LVVCGG**Q**GHRRVHRK
ADP-ribosylation factor-like protein 1	P40616	A14	71	VWDLGG**Q**TSIRPYWR
ADP-ribosylation factor-like protein 3	P36405	A15	71	VWDIGG**Q**RKIRPYWK
Histone-lysine N-methyltransferase ASH1L	Q9NR48	A16	1220	AEKFCG**Q**KKRRHSFE
Ancient ubiquitous protein 1	Q9Y679	A17	320	VAKELG**Q**TGTRLTPA
Brain-specific angiogenesis inhibitor 1	O14514	A18	1002	ALILIG**Q**TQTRNKVV
Protein BTG1	P62324	A19	82	MDPLIG**Q**AAQRIGLS

Complement C1q tumor necrosis factor-related protein 2	Q9BXJ5	A20	262	GLVHNG**Q**YRIRTFDA
Uncharacterized protein C1orf170	Q5SV97	A21	426	NKPGSG**Q**ASARPSAP
Voltage-dependent T-type calcium channel subunit alpha-1H	O95180	B1	2253	KGERWG**Q**ASCRAEHL
Calcium-binding and coiled-coil domain-containing protein 1	Q9P1Z2	B2	330	MKDTLG**Q**AQQRVAEL
Cathepsin Z	Q9UBR2	B3	31	LYFRRG**Q**TCYRPLRG
Uncharacterized protein C2orf50	Q96LR7	B4	137	LDTPLG**Q**TLIRMDFF
Uncharacterized protein C3orf30	Q96M34	B5	86	GHSTPG**Q**AGRRASNP
Coiled-coil domain-containing protein 157	Q569K6	B6	494	SEREQG**Q**CQLRAQQE
Glucose-fructose oxidoreductase domain-containing protein 1	Q9NXC2	B7	92	WGVRKG**Q**RHIRYGMC
Uncharacterized protein C9orf84	Q5VXU9	B8	1399	KVPGRVDG**Q**TRLRFF
UPF0614 protein C14orf102	Q9H7Z3	B9	515	LFDDIG**Q**SLIRLSSH
Cyclic nucleotide-gated cation channel beta-1	Q14028	B10	878	GAATAG**Q**TYYRSCMD
Collagen alpha-2(I) chain	P08123	B11	212	ENGTPG**Q**TGARGLPG
Collagen alpha-6(IV) chain	Q14031	B12	1676	ETLKAG**Q**LHTRVSRC
Collagen alpha-1(IX) chain	P20849	B13	812	ENGFPG**Q**MGIRGLPG
Collagen alpha-1(XVI) chain	Q07092	B14	854	RDGQQG**Q**TGLRGTPG
Collagen alpha-1(XIX) chain	Q14993	B15	407	PPGKEG**Q**RGRRGKTG
Copine-1	Q99829	B16	479	LHTRSG**Q**AAARDIVQ
Putative uncharacterized protein encoded LINC00526	Q96FQ7	B17	29	GGLPPG**Q**YATRMTGQ
Cone-rod homeobox protein	O43186	B18	111	QQPPGG**Q**AKARPAKR
Caskin-2	Q8WXE0	B19	1943	GTVGPG**Q**AQQRLEQT
Protein FAM207A	Q9NSI2	B20	216	PLVAIG**Q**TLARQMQL
Extracellular matrix protein 1	Q16610	B21	30	GFTATG**Q**RQLRPEHF

Histone-lysine N-methyltransferase EHMT2	Q96KQ7	C1	989	SNCLCG**Q**LSIRCWYD
Echinoderm microtubule-associated protein-like 3	Q32P44	C2	298	GPGGGG**Q**RHYRGHTD
Histone acetyltransferase p300	Q09472	C3	233	SPQMGG**Q**TGLRGPQP
DNA excision repair protein ERCC-6	Q03468	C4	1444	QAHTDG**Q**ASTREILQ
Protein FAM150B	Q6UX46	C5	35	REPADG**Q**ALLRLVVE
F-box/WD repeat-containing protein 7	Q969H0	C6	218	LRAANG**Q**GQQRRRIT
Zinc finger protein ZFPM1	Q8IX07	C7	100	PVVQDG**Q**RRIRARLS
Gamma-aminobutyric acid receptor subunit delta	O14764	C8	412	AARSGG**Q**GGIRARLR
Gamma-aminobutyric acid receptor subunit rho-2	P28476	C9	454	KGLLKG**Q**TGFRIFQN
Glucokinase regulatory protein	Q14397	C10	529	LQRFSG**Q**SKARCIES
Glial cell line-derived neurotrophic factor	P39905	C11	176	VSDKVG**Q**ACCRPIAF
Golgin subfamily A member 2-like protein 2	Q9H5Y0	C12	85	STSARG**Q**CQRRSTGR
Golgin subfamily A member 2-like protein 3	Q8NCE8	C13	85	STSARG**Q**CQRRSTGR
Solute carrier family 25 member 47	Q6Q0C1	C14	252	QADGQG**Q**RRYRGLLH
5-aminolevulinate synthase, erythroid-specific, mitochondrial	P22557	C15	444	LKGEEG**Q**ALRRAHQR
Hepatocyte growth factor	P14210	C16	32	IPYAEG**Q**RKRRNTIH
Integrator complex subunit 1	Q8N201	C17	1334	PEQPIG**Q**GRIRVGTQ
IQ domain-containing protein C	Q4KMZ1	C18	426	EPSHEG**Q**KKQRTIPW
Integrin alpha-5	P08648	C19	257	INLVQG**Q**LQTRQASS
Uncharacterized protein KIAA1908	Q96PY0	C20	202	GLSHLG**Q**SLCRTVKE
Kinesin-like protein KIF23	Q02241	C21	284	EVFWRG**Q**KKRRIANT
Kinesin-like protein KIF7	Q2M1P5	D1	763	AELSEG**Q**RQLRELEG
Kinesin-like protein KIF7	Q2M1P5	D2	837	MRQQQG**Q**LQRRLREE

Laminin subunit alpha-5	O15230	D3	559	CDPDTG**Q**CRCRVGFE
Laminin subunit alpha-5	O15230	D4	695	CDPRSG**Q**CSCRPRVT
Laminin subunit alpha-5	O15230	D5	1549	CDTDSG**Q**CKCRPNVT
Laminin subunit alpha-5	O15230	D6	2087	CHPQSG**Q**CHCRPGTM
Laminin subunit beta-1	P07942	D7	528	CFAESG**Q**CSCRPHMI
Laminin subunit beta-1	P07942	D8	791	CDPNGG**Q**CQCRPNVV
Laminin subunit beta-2	P55268	D9	540	CDEGTG**Q**CHCRQHMV
Laminin subunit beta-2	P55268	D10	849	CEKTSG**Q**CLCRTGAF
Laminin subunit beta-2	P55268	D11	1113	CNEFTG**Q**CHCRAGFG
La-related protein 4B	Q92615	D12	459	QTRQAG**Q**TRTRIQNP
Leucine-rich repeat-containing protein 41	Q15345	D13	517	LRALSG**Q**AGCRLRAL
Latent-transforming growth factor beta-binding protein 4	Q8N2S1	D14	59	CRCCPG**Q**TSRRSRCI
Nuclear body protein SP140	Q13342	D15	851	KYKDFG**Q**MGFRLEAE
Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 2	Q86UL8	D16	1305	ELSACG**Q**KKQRLGEQ
Mediator of RNA polymerase II transcription subunit 24	O75448	D17	827	YSSHKG**Q**ASTRQKKR
Midnolin	Q504T8	D18	367	ASLLQG**Q**SQIRMCKP
Histone-lysine N-methyltransferase MLL2	O14686	D19	4588	GCPVNG**Q**SQLRGAFG
Putative helicase MOV-10	Q9HCE1	D20	138	HEARDG**Q**LLIRLDLN
M-phase inducer phosphatase 3	P30307	D21	249	KKYFSG**Q**GKLRKGLC
Unconventional myosin-XVIIIa	Q92614	E1	248	DRGPEG**Q**ACRRVVHF
Myosin light chain kinase 2, skeletal/cardiac muscle	Q9H1R3	E2	136	KKAAEG**Q**AAARRGSP
Endonuclease 8-like 3	Q8TAT5	E3	256	KRPNCG**Q**CHCRITVC
Ras GTPase-activating protein nGAP	Q9UJF2	E4	867	RQNSTG**Q**AQIRKVDQ

C2 calcium-dependent domain-containing protein 4C	Q8TF44	E5	319	AEYEAG**Q**ARLRVHLL
Obscurin-like protein 1	O75147	E6	1327	GCRMCG**Q**RKARTCVS
Occludin/ELL domain-containing protein 1	Q9H607	E7	23	ELQTLG**Q**AARRPPPP
Olfactory receptor 4K17	Q8NGC6	E8	236	NHSPTG**Q**SKARSTLT
Polyadenylate-binding protein 5	Q96DU9	E9	378	SKPLHVTLG**Q**ARRRC
Paralemmin-3	A6NDB9	E10	76	PQSPEG**Q**AQARIRNL
Protocadherin-16	Q96JQ0	E11	1784	DVGANG**Q**LQYRILDG
Protocadherin-17	O14917	E12	386	DSGKNG**Q**LQCRVLGG
Protocadherin gamma-C5	Q9Y5F6	E13	729	GDGGGG**Q**CCRRQDSP
Peroxisomal biogenesis factor 3	P56589	E14	34	ILGKYG**Q**KKIREIQE
Phosphoinositide 3-kinase regulatory subunit 5	Q8WYR1	E15	419	GHRRPG**Q**KFIRIYKL
Phosphatidylinositol N-acetylglucosaminyl- transferase subunit Q	Q9BRB3	E16	558	LEAERG**Q**AGLRELLA
Pleckstrin homology domain-containing family G member 5	O94827	E17	725	TFQASG**Q**ALCRGWVD
HERV-K_5q33.3 provirus ancestral Pol protein	P10266	E18	385	IATLIG**Q**TRLRITKL
HERV-K_1q22 provirus ancestral Pol protein	P63135	E19	385	IATLIG**Q**TRLRIIKL
HERV-K_11q22.1 provirus ancestral Pol protein	P63136	E20	385	IATLIG**Q**TRLRIIKL
HERV-K_3q27.3 provirus ancestral Pol protein	Q9UQG0	E21	385	IATLIG**Q**TRLRIIKL
HERV-K_7p22.1 provirus ancestral Pol protein	Q9BXR3	F1	385	IATLIG**Q**TRLRIIKL
HERV-K_19q12 provirus ancestral Pol protein	Q9WJR5	F2	1308	IATLIG**Q**TRLRIIKL
HERV-K_19p13.11 provirus ancestral Pol protein	P63132	F3	385	IATLIG**Q**TRLRIIKL
HERV-K_8p23.1 provirus ancestral Pol protein	P63133	F4	385	IATLIG**Q**TRLRIIKL
PRAME family member 1	O95521	F5	15	LLELAG**Q**SLLRDQAL

PRAME family member 2	O60811	F6	15	LLELAG**Q**SLLRDQAL
Protamine-2	P04554	F7	50	YERTHG**Q**SHYRRRHC
PH and SEC7 domain-containing protein 4	Q8NDX1	F8	223	EPEGEG**Q**AWLREGTP
Adenylosuccinate lyase	P30566	F9	241	AFIITG**Q**TYTRKVDI
Glutamine-rich protein 1	Q2TAL8	F10	752	SREQMG**Q**MLTRILVI
DNA repair protein REV1	Q9UBZ9	F11	689	FGPKTG**Q**MLYRFCRG
E3 ubiquitin-protein ligase RNF133	Q8WVZ7	F12	236	LQNTFG**Q**LQLRVVKE
E3 ubiquitin-protein ligase RNF220	Q5VTB9	F13	361	EYEWCG**Q**KRIRATTL
Ribosome-recycling factor, mitochondrial	Q96E11	F14	67	KAKGKG**Q**SQTRVNIN
Solute carrier family 22 member 12	Q96S37	F15	473	TAVGLG**Q**MAARGGAI
Solute carrier family 4 member 8	Q2Y0W8	F16	69	HHRTHG**Q**KHRRRGRG
Protein transport protein Sec23A	Q15436	F17	491	YQHSSG**Q**RRIRVTTI
Serine/arginine-rich splicing factor 9	Q13242	F18	79	NGYDYG**Q**CRLRVEFP
Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 1	Q92835	F19	431	WFLSKG**Q**GKTRDDSA
Neutral and basic amino acid transport protein rBAT	Q07837	F20	79	LFQFSG**Q**ARYRIPRE
Mothers against decapentaplegic homolog 7	O15105	F21	405	FVKGWG**Q**CYTRQFIS
SPARC-related modular calcium-binding protein 1	Q9H4F8	G1	92	RCKDAG**Q**SKCRLERA
WD40 repeat-containing protein SMU1	Q2TAY7	G2	297	WKIQSG**Q**CLRRFERA
Suppressor of cytokine signaling 4	Q8WXH5	G3	175	TELRDG**Q**LKRRNMEE
Stabilin-1	Q9NY15	G4	1994	GMSGSG**Q**CLCRSGFA
Sushi domain-containing protein 5	O60279	G5	614	YKLNVG**Q**RQARHYHQ
Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	Q4LDE5	G6	66	RVERLG**Q**AFRRRVRL

TP53-target gene 1 protein	Q9Y2A0	G7	18	SRRHSG**Q**AALRPRRY
Transcription factor IIIB 90 kDa subunit	Q92994	G8	298	PSYTAG**Q**RKLRMKQL
Transferrin receptor protein 2	Q9UP52	G9	628	VAQLAG**Q**LLIRLSHD
Transmembrane protein 50A	O95807	G10	90	SEGCLG**Q**TGARIWLF
Transmembrane and coiled-coil domain-containing protein 6	Q96DC7	G11	330	VETVGG**Q**MQLRDERV
Tumor necrosis factor ligand superfamily member 6	P48023	G12	237	SYCTTG**Q**MWARSSYL
Transformation/ transcription domain-associated protein	Q9Y4A5	G13	612	QIAGNG**Q**TYIRVANC
Tryptase alpha/beta-1	Q15661	G14	24	AAPAPG**Q**ALQRVGIV
Tryptase beta-2	P20231	G15	24	AAPAPG**Q**ALQRVGIV
Testis-specific Y-encoded-like protein 5	Q86VY4	G16	99	AAGDHG**Q**AAARPGPG
Tubulin polyglutamylase TTLL13	A6NNM8	G17	736	RLTSQG**Q**ASRRLEAI
Protein shisa-7	A6NL88	G18	16	LASSAG**Q**ARARPSNA
Ubiquitin carboxyl-terminal hydrolase 42	Q9H9J4	G19	1314	RLFEYG**Q**GKRRYLEL
Unhealthy ribosome biogenesis protein 2 homolog	Q14146	G20	222	TWTQAG**Q**GQLRQVLS
VEGF co-regulated chemokine 1	Q6UXB2	G21	37	GHRDRG**Q**ASRRWLQE
Voltage-dependent anion-selective channel protein 2	P45880	H1	6	MATHG**Q**TCARPMCIP
Vinculin	P18206	H2	516	DDRGVG**Q**AAIRGLVA
DDB1- and CUL4-associated factor 11	Q8TEB1	H3	119	VELATG**Q**LGLRRAAQ
Wee1-like protein kinase	P30291	H4	250	LLHSSG**Q**CRRRKRTY
Putative UPF0607 protein ENSP00000381514	A8MUA0	H5	292	EALLVG**Q**ASQREGRL
Nuclease-sensitive element-binding protein 1	P67809	H6	181	ESAPEG**Q**AQQRRPYR
Nuclear transcriptional regulator 1-like protein	A6NF83	H7	80	QKLLNG**Q**RKRRQRQL
Putative UPF0607 protein ENSP00000382826	A8MV72	H8	262	EALLVG**Q**ASQREGRL

Putative UPF0607 protein FLJ37424	Q8N9G6	H9	292	EALLVG**Q**ASQREGRL
Putative zinc finger protein 852	Q6ZMS4	H10	122	EGVLKG**Q**KSYRCDEC
Putative UPF0607 protein ENSP00000383144	A8MX80	H11	292	EALLVG**Q**ASQREGHL
Putative UPF0607 protein ENSP00000381418	A8MU76	H12	292	EALLVG**Q**ASQREGRL
Zinc finger BED domain-containing protein 2	Q9BTP6	H13	117	EKSGHG**Q**AGQRQDPR
Zinc finger protein 256	Q9Y2P7	H14	143	QKQHVG**Q**KHFRSNGG
Zinc finger protein 479	Q96JC4	H15	276	TCEECG**Q**AFRRSSAL
Zinc finger protein 517	Q6ZMY9	H16	324	RCLRCG**Q**RFIRGSSL
Zinc finger protein 648	Q5T619	H17	483	PCTQCG**Q**AFARSSTL
Zinc finger protein 648	H18	539	483	QCEDCG**Q**AFTRSNHL
Putative zinc finger protein 735	P0CB33	H19	276	ACEECG**Q**AFRRSSTL
Zinc finger protein 785	A8K8V0	H20	204	RPFSCG**Q**CQARFSQR
Tight junction protein ZO-3	O95049	H21	528	LHPGPG**Q**SHARGGHW
eRF1 WT	P62495	I1	185	KHGRGG**Q**SALRFARL
eRF1 Mutant		I2	A185	KHGRGG**A**SALRFARL
Chromodomain-helicase-DNA-binding protein 5 WT	Q8TDI0	I3	1390	EERPEG**Q**SGRRQSRR
Chromodomain-helicase-DNA-binding protein 5 Mutant		I4	A1390	EERPEG**A**SGRRQSRR

6.1.4 *In vitro* Methylation of the Protein Domains

The methylation of protein domains was performed overnight at $25\,^{\circ}$C in $40\,\mu$L methylation buffer ($10\,$mM Tris pH 7.6, $50\,$mM KCl, $10\,$mM Mg(OAc)$_2$, $1\,$mM DTT) containing $0.76\,\mu$M labeled [methyl-^{3}H]-SAM ($2.7\,$Tbq/mmol; PerkinElmer) and $4.5\,\mu$M HEMK2. Equal loading of target protein amounts were confirmed by Coomassie Brilliant Blue staining. The methylation reaction was stopped by addition of an appropriate volume of $5\,$x SDS-loading buffer, subsequently boiled at $95\,^{\circ}$C for $5\,$min and separated by loading on a $12\,\%$ SDS-PAGE gel. Afterwards, the gel was dried with vacuum at $65\,^{\circ}$C and then incubated with Hyperfilm$^{\text{TM}}$ high performance autoradiography films at $-80\,^{\circ}$C, in the dark for varying time.

6.1.5 Cell culture, Transfection and Immunoprecipitation

For mammalian expression of the full length proteins, the coding sequences of Chromodomain-helicase-DNA-binding protein 5 (CHD5) (kindly provided by Dr. A. A. Mills), Protein NUT (kindly provided by Dr. C. A. French) and eRF1 were subcloned into the pEYFP-C1 vector (Clontech, USA). The corresponding target glutamine to arginine mutants of CHD5 (Q1390R) and Protein NUT (Q1046) were subcloned into the pECFP-C1 vector (Clontech, USA).

The transfection of the plasmids into eukaryotic cell lines was performed using Polyethylenimine (PEI; MAX 40000; Polyscience, USA). The transfection of HEK293 cells was routinely performed done at a confluency of approximately $80\,\%$. $2\,$h before transfection the DMEM medium supplemented with $10\,\%$ Fetal Bovin Serum (Sigma-Aldrich, USA), 100 units Penicillin and $100\,\mu$g per mL Streptomycin (Sigma-Aldrich, USA) and $2\,\mu$M L-Glutamine (Sigma-Aldrich, USA) was removed and replaced with fresh pre-warmed growth medium. For transfection, a ratio of DNA:PEI of 1:3 was used. The DNA ($0.8\,\mu$g per mL of total culture volume) was added to $5\,\%$ Serum-free media (SFM; of total transfection volume) and mixed by pipetting. In a separate tube, three times more PEI was added to $5\,\%$ Serum-free media and mixed as well. Then the DNA/SFM mixture was added to the PEI/SFM mixture, gently mixed by pipetting and incubated at room temperature for $20\,$min. Afterwards, the DNA-PEI solution was added drop-wise to the flask containing the cells and incubated at $37\,^{\circ}$C and $5\,\%$ CO$_2$. $72\,$h after transfection, the cells were washed with PBS buffer and harvested by centrifugation at $500\,$g for $5\,$min. The YFP-fused CHD5 and NUT substrate proteins were immunoprecipitated from cell extract using GFP-Trap$^{\circledR}$ A (Chromotek). The composition of the buffers used for the GFP-Trap$^{\circledR}$ A are shown in Table 12.

The cell pellet was thawed on ice, resuspended with 150 µL lysis buffer using a syringe with a 26 gauge needle and mixed by vortexing. This resuspention was incubated end-over-end for 30 min at 8 ˚C with vortexing every 10 min. After that, the cell lysate was centrifuged for 5 min at 13.000 rpm at 4 ˚C and the supernatant was transferred to a pre-cooled tube. Then, dilution buffer (4 times the volume of lysis buffer) was added to the supernatant and the diluted lysate was added to 20 µL pre-equilibrated GFP-Trap® A magnetic beads and rolled end-over-end for 2 h at 8 ˚C. The beads were magnetically separated until the supernatant was clear and the latter was discarded. The magnetic beads were washed three times with 300 µL lysis buffer. Finally, the beads were resuspended in 30 µL lysis buffer and 30 µL 2 x SDS-loading buffer was added to release the bound proteins.

Table 12: *Composition of the buffers used for GFP-Trap® A purification.*

Lysis buffer (pH 7.5)	Dilution buffer (pH 7.5)
10 mM Tris-HCl	10 mM Tris-HCl
300 mM NaCl	300 mM NaCl
0.5 mM EDTA	0.5 mM EDTA
0.5 % NP 40	—
1 x mammalian	1 mM PMSF
Inhibitor Cocktail	

6.2 The Histone Lysine Methyltransferase NSD2

6.2.1 Cloning, Site-directed Mutagenesis, Expression and Purification

The sequence encoding for the human NSD2 enzyme and the putative human substrate protein domains were amplified from cDNA of HEK293 cells. NSD2 was amplified in three different variants, which differed in size, but all possessing the AWS, the catalytic active SET domain and the Post-SET domain (Figure 22).

The amplified NSD2 inserts were subcloned into pGEX-6P-2 (fused to GST), pET-28a(+) (fused to His$_6$) and pMAL-c2x (fused to MBP) vectors. The putative substrates were cloned as GST fusion proteins into the pGEX-6P-2 vector (GE Healthcare)(Table 13).

Table 13: *List with information of putative novel NSD2 substrate proteins, which were selected for methylation at protein level.*

Name	Abbreviation	Domain boundaries (aa)	NCBI accession number
Abnormal spindle-like microcephaly-associated protein	ASPM	2112 – 2361	NP_060606.3
Bromodomain adjacent to zinc finger domain protein 2B	BAZ2B	807 – 1092	NP_038478.2
DNA polymerase alpha catalytic subunit	POLA1	803 – 1074	NP_058633.2
ETS domain-containing protein Elk-3	ELK3	1 – 272	NP_005221.2
ETS domain-containing protein Elk-4	ELK4	1 – 275	NP_001964.2
Eukaryotic initiation factor 4A-III	EIF4A3	2 – 280	NP_055555.1
Fanconi anemia group M protein	FANCM	723 – 933	NP_065988.1
Mediator of RNA polymerase II transcription subunit 12-like protein	MED12L	1486 – 1749	NP_443728.3
Mismatch repair endonuclease PMS2	PMS2	446 – 750	NP_000526.1
Myoneurin	MYNN	372 – 603	NP_001172047.1
N-lysine methyltransferase SET8	SET8	1 – 346	NP_065115.3
Nuclear protein localization protein 4 homolog	NPLOC4	1 – 289	NP_060391.2
Prickle-like protein 2	PRICKLE2	1 – 276	NP_942559.1

Probable U3 small nucleolar RNA-associated protein 11	UTP11L	$3 - 253$	NP_057121.2
Putative homeodomain transcription factor 2	PHTF2	$456 - 719$	NP_001120829.1
RNA polymerase II elongation factor ELL2	ELL2	$379 - 640$	NP_036213.2
RNA-binding E3 ubiquitin-protein ligase MEX3C	MEX3C	$224 - 478$	NP_057710.3
RNA-binding protein MEX3A	MEX3A	$134 - 371$	NP_001087194.1
RNA-binding protein MEX3B	MEX3B	$29 - 303$	NP_115622.2
Sister chromatid cohesion protein DCC1	DSCC1	$13 - 278$	NP_076999.2
STE20/SPS1-related proline-alanine-rich protein kinase	STK39	$66 - 296$	NP_037365.2
Transcription factor RFX4	RFX4	$336 - 493$	NP_998759.1
Transcription factor SOX-17	SOX17	$41 - 287$	NP_071899.1
Transcriptional regulator ATRX	ATRX	$893 - 1188$	NP_000480.3
U5 small nuclear ribonucleoprotein 40 kDa protein	SRNP40	$4 - 293$	NP_004805.2
Zinc finger protein 292	ZNF292	$2415 - 2655$	NP_055836.1
Zinc finger protein ZFAT	ZFAT	$232 - 496$	NP_065914.2

For bacterial expression, the plasmids were transformed into *E. coli* BL21-CodonPlus (DE3) cells (Novagen, USA). These were grown in LB medium at $37\,^{\circ}$C until an OD_{600} of 0.6 to 0.8 was reached. Protein expression was induced with $1\,\text{mM}$ isopropyl-β-D-thiogalactopyranoside (IPTG). The culture was then shifted to $20\,^{\circ}$C overnight (14 to 16 h). Afterwards the cells were collected by centrifugation at $4.500\,\text{rpm}$, washed once with STE buffer ($10\,\text{mM}$ Tris-HCl pH 8.0, $1\,\text{mM}$ EDTA and $100\,\text{mM}$ NaCl) and the cell pellet was stored at $-20\,^{\circ}$C until purification.

For purification the cell pellet was thawed on ice, resuspended in sonication buffer and lysed by ultra sound. Then the samples were centrifuged at $18.000\,\text{rpm}$ for $90\,\text{min}$ and the supernatants were passed through Glutathione Sepharose 4B resin (GE Healthcare). Afterwards, the beads were washed once with sonication buffer and twice with washing buffer. Subsequently, the bound proteins were eluted with elution buffer containing excess of glutathione and then dialyzed against low glycerol dialysis buffer 1 for 3 h and afterwards over night against high glycerol dialysis buffer 2. The composition of the used buffers is shown in Table 8 in section 6.1.1.

The mutation of the target lysine of the methylated substrates was performed by site-directed mutagenesis using PCR-megaprimers according to the protocol of Jeltsch & Lanio[196]. In addition to the needed lysine to arginine mutations, silent mutations were introduced to allow the identification of the plasmids containing the mutated targets by specific restriction sites. The successful mutagenesis of the novel substrates was confirmed by restriction digest and DNA sequencing (Table 14).

Table 14: *Target lysine mutations of methylated substrate proteins of NSD2*

Name	Abbreviation	Target K mutation
Transcriptional regulator ATRX	ATRX Mut	K1033R
Fanconi anemia group M protein	FANCM Mut	K819R
N-lysine methyltransferase SET8	SET8 Mut	K158R

6.2.2 *In vitro* Methylation of the Peptide SPOT Arrays

Synthesis of the peptide arrays was performed according to the description in section 6.1.2. All peptide arrays were washed for 10 min with methylation buffer containing 50 mM Tris (pH 8.5), 50 mM NaCl and 0.5 mM DTT. Then the peptide SPOT membranes were incubated for 60 min in methylation buffer containing 3 µM NSD2 and 0.76 µM labeled [methyl-^{3}H]-SAM (Perkin Elmer) at 23 °C. Afterwards, the arrays were washed five times with 50 mM NH_4HCO_3 and 1 % SDS and then incubated for 5 min in Amplify NAMP100V solution (GE Healthcare). The membranes were exposed on Hyperfilm^{TM} high performance autoradiography films (GE Healthcare) in the dark, for several days at −80 °C. Film development was performed on an Optimus TR developing machine. Quantification and analysis of the developed films were performed as described in section 6.1.3.

Information of peptide array containing the synthesized peptides harboring the predicted target lysine residues derived from Scansite searches using the two specificity profiles from NSD2:

Table 15: *List of putative novel NSD2 substrates identified in Scansite searches using a stringent and a relaxed specificity profile as shown in Table 3. The predicted target lysine is printed in bold. The position of the corresponding peptide spots in Figure 29 are indicated.*

Name	Swiss Prot no.	Position on Array	Target K Position	Sequence
Histone H3 K36	P68431	A1	36	APATGGV**K**KPHRYRP

Histone H3 K36A Mut		A2	36	APATGGV**A**KPHRYRP
H4K44 WT	P62805	A3	44	LARRGGV**K**RISGLIY
H4K44 Mut		A4	A44	LARRGGV**A**RISGLIY
Apoptotic chromatin condensation inducer in the nucleus	Q9UKV3	A5	969	PPAEHEV**K**KVTLGDT
Apoptotic chromatin condensation inducer in the nucleus	Q9UKV3	A6	548	GITEECL**K**QPSLEQK
Afadin- and alpha-actinin-binding protein	Q9Y2D8	A7	278	LMENAEL**K**KVLQQMK
Adipocyte enhancer-binding protein 1	Q8IUX7	A8	340	DEEKEEL**K**KPKKEDS
Amino-terminal enhancer of split	Q08117	A9	83	HKQAEIV**K**RLNGICA
Protein arginine N-methyltransferase 7	Q9NVM4	A10	117	NGFSDKI**K**VINKHST
Ankyrin repeat domain-containing protein 23	Q86SG2	A11	194	GGHLVIL**K**QLLNQGA
Poly(ADP-ribose) glycohydrolase ARH3	Q9NX46	A12	213	SSSEHFL**K**QLLGHME
AT-rich interactive domain-containing protein 4A	P29374	A13	39	VKRLVKV**K**VLLKQDN
Aryl hydrocarbon receptor nuclear translocator 2	Q9HBZ2	A14	439	ICTNTNV**K**QLQQQQA
Abnormal spindle-like microcephaly-associated protein	Q8IZT6	A15	888	ALSKFTL**K**KLLLLVC
Abnormal spindle-like microcephaly-associated protein	Q8IZT6	A16	2213	QTYFNKL**K**KITKTVQ
Transcriptional regulator ATRX	P46100	A17	1033	CHFPKGI**K**QIKNGTT
Ataxin-1	P54253	A18	688	VCISLTL**K**NLKNGSV
Protein BANP	Q8N9N5	A19	165	RQNTIVV**K**VPGQEDS
Bromodomain adjacent to zinc finger domain protein 2B	Q9UIF8	A20	948	MKQQEKI**K**RIQQIRM
BRCA2 and CDKN1A-interacting protein	Q9P287	A21	75	DNDYDGI**K**KLLQQLF
Ribosome biogenesis protein BOP1	Q14137	B1	708	NPLLVPV**K**VLKGHVL
Bromodomain-containing protein 8	Q9H0E9	B2	109	AERVEEL**K**KVIKETQ
E3 ubiquitin-protein ligase BRE1A	Q5VTR2	B3	627	KKEAEII**K**QLKIELK

Transcription factor BTF3 homolog 3	Q13892	B4	79	KKLQFSL**K**KLQVNNI
Mitotic checkpoint serine/threonine-protein kinase BUB1	O43683	B5	1055	DLLRQKL**K**KVFQQHY
Coiled-coil domain-containing protein 110	Q8TBZ0	B6	57	IQPQSAL**K**VLQQQLE
Coiled-coil alpha-helical rod protein 1	Q8TD31	B7	360	LEHSDSV**K**QLKGQVA
PITSLRE serine/threonine-protein kinase CDC2L1	P21127	B8	467	TDEIVAL**K**RLKMEKE
PITSLRE serine/threonine-protein kinase CDC2L2	Q9UQ88	B9	455	TDEIVAL**K**RLKMEKE
Cell division protein kinase 7	P50613	B10	41	TNQIVAI**K**KIKLGHR
Cell division protein kinase 9	P50750	B11	48	TGQKVAL**K**KVLMENE
Centrosomal protein of 290 kDa	O15078	B12	1645	SSLLVKL**K**KVSQDLE
Centrosomal protein of 290 kDa	O15078	B13	1681	ENHEDEV**K**KVKAEVE
E3 ubiquitin-protein ligase CHFR	Q96EP1	B14	88	GTVINKL**K**VVKKQTC
Cirhin	Q969X6	B15	514	GVHVYNV**K**QLKLHCT
Condensin-2 complex subunit G2	Q86XI2	B16	421	TILIDLL**K**KVTGELA
Cleavage and polyadenylation specificity factor subunit 2	Q9P2I0	B17	550	RSDGDSI**K**KIINQMK
Catenin delta-1	O60716	B18	433	LGACGAL**K**NISFGRD
Dachshund homolog 1	Q9UI36	B19	347	IAEAMKV**K**KIKLEAM
Sister chromatid cohesion protein DCC1	Q9BVC3	B20	139	RPKLKKL**K**KLLMENP
Deoxycytidine kinase	P27707	B21	22	SSEGTRI**K**KISIEGN
Probable dimethyladenosine transferase	Q9UNQ2	C1	40	GIGQHIL**K**NPLIINS
H/ACA ribonucleoprotein complex subunit 4	O60832	C2	367	HGIVAKI**K**RVIMERD
Dentin matrix acidic phosphoprotein 1	Q13316	C3	482	SEEDGQL**K**NIEIESR
Diphthamide biosynthesis protein 1	Q9BZG8	C4	43	QIPPEIL**K**NPQLQAA
DNA polymerase delta subunit 3	Q15054	C5	292	SKKAEPV**K**VLQKEKK

DNA polymerase alpha catalytic subunit	P09884	C6	926	VERRKQV**K**QLMKQQD
Dual specificity tyrosine-phosphorylation-regulated kinase 2	Q92630	C7	226	AYRYEVL**K**VIGKGSF
Erythroid differentiation-related factor 1	Q3B7T1	C8	131	VSDSENI**K**KLLKIPY
EF-hand calcium-binding domain-containing protein 6	Q5THR3	C9	61	TLSSLDV**K**RILFQKI
Eukaryotic translation initiation factor 3 subunit E	P60228	C10	279	QVLKDLV**K**VIQQESY
ETS domain-containing protein Elk-3	P41970	C11	73	YYDKNII**K**KVIGQKF
ETS domain-containing protein Elk-4	P28324	C12	73	YYVKNII**K**KVNGQKF
RNA polymerase II elongation factor ELL2	O00472	C13	625	HNKLAHI**K**RLIGEFD
Alpha-enolase	P60228	C14	279	QVLKDLV**K**VIQQESY
Separin	Q14674	C15	515	RLQVESL**K**KLGKQAQ
Separin	Q14674	C16	1075	TQHLDSV**K**KVHLQKG
Ecotropic virus integration site 1 protein homolog	Q03112	C17	543	PQSPGEV**K**KLQKGSS
Exonuclease 1	Q9UQ84	C18	252	ANNPDIV**K**VIKKIGH
Fanconi anemia group M protein	Q8IYD8	C19	819	HKKSSFI**K**NINQGSS
FK506-binding protein 5	Q13451	C20	38	RGVLKIV**K**RVGNGEE
FYN-binding protein	O15117	C21	683	KTEEKDL**K**KLKKQEK
Glucocorticoid receptor	P04150	D1	770	KYSNGNI**K**KLLFHQK
Vasculin-like protein 1	Q9HC44	D2	199	PSKMLVI**K**KVSKEDP
General transcription factor II-I	P78347	D3	185	AGISFII**K**RPFLEPK
Histone H2B type F-M	P0C1H6	D4	69	PYFPRVL**K**QVHQGLS
Histone H2B type W-T	Q7Z2G1	D5	90	TYFRRVL**K**QVHQGLS
Histone deacetylase 6	Q9UBN7	D6	1199	HQALLDV**K**NIAHQNK
Homeodomain-interacting protein kinase 2	Q9H2X6	D7	29	SSAFCSV**K**KLKIEPS

Homeobox protein HMX3	A6NHT5	D8	34	KESPFSI**K**NLLNGDH
Heterochromatin protein 1-binding protein 3	Q5SSJ5	D9	517	RPSSTVI**K**KPSGGSS
Eukaryotic initiation factor 4A-III	P38919	D10	70	AIQQRAI**K**QIIKGRD
Zinc finger protein Aiolos	Q9UKT9	D11	440	ICPRDSV**K**VINKEGE
Integrator complex subunit 6	Q9UL03	D12	457	YSVISYL**K**KLSQQAK
Integrator complex subunit 7	Q9NVH2	D13	269	NDPRKAV**K**RLAIQDL
Iroquois-class homeodomain protein IRX-2	Q9BZI1	D14	341	EIATSDL**K**QPSLGPG
Lysine-specific demethylase 2B	Q8NHM5	D15	549	QALLEGV**K**NVLKEHA
Lysine-specific demethylase 4A	O75164	D16	468	EVKFEEL**K**NVKLEEE
Kinesin-like protein KIF20B	Q96Q89	D17	1300	TDAKKQI**K**QVQKEVS
Antigen KI-67	P46013	D18	548	MHTPPVL**K**KIIKEQP
Chromosome-associated kinesin KIF4B	Q2VIQ3	D19	562	FQYQDNI**K**NLELEVI
Kinetochore-associated protein KNL-2 homolog	Q6P0N0	D20	752	TRLLPKL**K**KIENQVA
Kinetochore-associated protein 1	P50748	D21	1737	EKAEALL**K**KLHIQYR
Ribosomal protein S6 kinase alpha-2	Q15349	E1	63	PSQFELL**K**VLGQGSY
Putative leucine-twenty homeobox	A8MZ59	E2	92	LREPSGI**K**NPGGASA
Protein lin-54 homolog	Q6MZP7	E3	184	KLPPQQI**K**VVTIGGR
Leucine zipper protein 1	Q86V48	E4	766	VIVDKDV**K**KIMGGSG
Leucine zipper protein 1	Q86V48	E5	265	KGGLDYL**K**QVENETR
Mitogen-activated protein kinase kinase kinase 2	Q9Y2U5	E6	411	ECEIQLL**K**NLLHERI
Metastasis-associated in colon cancer protein 1	Q6ZN28	E7	615	LVHCKNV**K**VISKEQV
MAD protein	Q05195	E8	207	GYSSTSI**K**RIKLQDS
Mediator of RNA polymerase II transcription subunit 12-like protein	Q86YW9	E9	1604	RAYMNLV**K**KLKKELG

Midasin	Q9NU22	E10	1622	MGEEAAL**K**RPEIIST
Midasin	Q9NU22	E11	1670	ECLKFLI**K**RLAKIVR
Merlin	P35240	E12	578	SSKHNTI**K**KLTLQSA
RNA-binding protein MEX3A	A1L020	E13	247	GPKGATI**K**RIQQQTN
RNA-binding protein MEX3B	Q6ZN04	E14	184	GPKGATI**K**RIQQQTH
RNA-binding protein MEX3C	Q5U5Q3	E15	350	GPKGATI**K**RIQQQTH
MAX gene-associated protein	Q8IWI9	E16	29	PTFFVIL**K**QPGNGKT
Methylated-DNA-protein-cysteine methyltransferase	P16455	E17	104	QVLWKLL**K**VVKFGEV
Msx2-interacting protein	Q96T58	E18	2930	VTQGGTV**K**VLTQGIN
Histone-lysine N-methyltransferase MLL2	O14686	E19	4553	RASEALL**K**QLKQELS
Histone-lysine N-methyltransferase MLL2	O14686	E20	5244	GRPEFVI**K**VIEQGLE
Mitogen-activated protein kinase kinase kinase MLT	Q9NYL2	E21	45	QDKEVAV**K**KLLKIEK
Myeloid cell nuclear differentiation antigen	P41218	F1	102	TQEKAPV**K**KINQEEV
M-phase phosphoprotein 8	Q99549	F2	228	VKETKEL**K**KVKKGEI
Protein maestro	Q9BYG7	F3	47	FQKREPL**K**NVFFILA
Myb proto-oncogene protein	P10242	F4	524	ENGPPLL**K**KIKQEVE
Myc proto-oncogene protein	P01106	F5	126	PDDETFI**K**NIIIQDC
Myoneurin	Q9NPC7	F6	512	GNSYTDI**K**NLKKHKT
Histone acetyltransferase MYST3	Q92794	F7	18	EWILEAI**K**KVKKQKQ
Myelin transcription factor 1	Q01538	F8	1040	SSMEKNL**K**NIEEENK
NGFI-A-binding protein 2	Q15742	F9	376	ELGGPPL**K**KLKQEVG
NACHT, LRR and PYD domains-containing protein 1	Q9C000	F10	765	IKFSRHV**K**KLQLIEG
Nuclear cap-binding protein subunit 2	P52298	F11	67	FSKSGDI**K**KIIMGLD

Nipped-B-like protein	Q6KC79	F12	1639	GSIERIL**K**QVSGGED
RNA-binding protein NOB1	Q9ULX3	F13	215	WITPSNI**K**QIQQELE
Nucleolar complex protein 3 homolog	Q8WTT2	F14	386	EMCCEAV**K**KLFKQDK
Nucleolar protein 14	P78316	F15	772	LFTPRLV**K**VLEFGRK
Nuclear protein localization protein 4 homolog	Q8TAT6	F16	31	ETAATFL**K**KVAKEFG
Nuclear receptor subfamily 0 group B member 2	Q15466	F17	119	APVPSIL**K**KILLEEP
Nuclear receptor subfamily 1 group I member 2	O75469	F18	331	LKFHYML**K**KLQLHEE
Nucleolin	P19338	F19	513	FEKATFI**K**VPQNQNG
Origin recognition complex subunit 2	Q13416	F20	288	PSFSAEL**K**QLNQQYE
Serine/threonine-protein kinase PAK 2	Q13177	F21	278	LGQEVAI**K**QINLQKQ
Poly [ADP-ribose] polymerase 15	Q460N3	G1	444	TPSLKTV**K**VVIFQPE
Phosphorylated CTD-interacting factor 1	Q9H4Z3	G2	126	QPSGNGV**K**KPKIEIP
Periplakin	O60437	G3	1099	SFLQDKL**K**RLEKERA
Periplakin	O60437	G4	550	AERAKDL**K**NITNELL
Period circadian protein homolog 3	P56645	G5	470	YASVNKI**K**NLGQQLY
PHD finger protein 21A	Q96BD5	G6	214	ATPPQPI**K**VPQFIPP
PH domain leucine-rich repeat-containing protein phosphatase 1	O60346	G7	1317	MSCEEEL**K**RIKQHKA
PH domain leucine-rich repeat-containing protein phosphatase 1	O60346	G8	246	HKGGGVV**K**VLGQGPG
Putative homeodomain transcription factor 1	Q9UMS5	G9	585	EIPHFRL**K**KVENIKI
Putative homeodomain transcription factor 1	Q9UMS5	G10	210	TIFGNRI**K**RVKLISN
Putative homeodomain transcription factor 2	Q8N3S3	G11	608	EVPHFRL**K**KVQNIKM
Pinin	Q9H307	G12	108	DPEDDDV**K**KPALQSS
Mismatch repair endonuclease PMS2	P54278	G13	630	SSLAKRI**K**QLHHEAQ

POU domain, class 6, transcription factor 2	Q12972	G14	81	LVYHKHL**K**RVFLIDL
DNA polymerase kappa	Q9UBT6	G15	461	RTVTIKL**K**NVNFEVK
Protein phosphatase 1 regulatory subunit 7	Q15435	G16	287	DIASNRI**KK**IENISH
Nuclear inhibitor of protein phosphatase 1	Q12972	G17	81	LVYHKHL**K**RVFLIDL
Prickle-like protein 2	Q7Z3G6	G18	74	PGEKLRI**K**QLLHQLP
Proteasome subunit beta type-4	P28070	G19	109	YADFQYL**K**QVLGQMV
Proteasome activator complex subunit 3	P61289	G20	237	TLHDMIL**K**NIEKIKR
Protein QN1 homolog	Q5TB80	G21	704	PVTGEKL**K**QIQKEIQ
RB1-inducible coiled-coil protein 1	Q8TDY2	H1	893	EENENKI**K**KLKGELV
Probable RNA-binding protein 19	Q9Y4C8	H2	792	EQAQKAL**K**QLQGHVV
RNA-binding protein 40	Q96LT9	H3	425	PNCRIYV**K**NLAKHVQ
E3 SUMO-protein ligase RanBP2	P49792	H4	458	PGIRKWL**K**QLFHHLP
Regulator of nonsense transcripts 3A	Q9H1J1	H5	285	EVRIKLL**K**KPEKGEE
Regulator of nonsense transcripts 3B	Q9BZI7	H6	285	VNQKNLL**K**KPEKGDE
Transcription factor RFX4	Q33E94	H7	421	AKRQGSL**K**KVAQQFL
RANBP2-like and GRIP domain-containing protein 8	O14715	H8	1713	AANLEYL**K**NVLLQFI
Telomere-associated protein RIF1	Q5UIP0	H9	290	RSGAPMI**KK**IAFIAW
RecQ-mediated genome instability protein 1	Q9H9A7	H10	188	LLKPENV**K**VLGGEVD
Nuclear receptor ROR-beta	Q92753	H11	187	GLDMTGI**K**QIKQEPI
DNA-directed RNA polymerase II subunit RPB1	P24928	H12	19	ACPLRTI**K**RVQFGVL
Ribosomal RNA processing protein 1 homolog A	P56182	H13	131	MVLNESL**K**VLKMQGW
SAM domain and HD domain-containing protein 1	Q9Y3Z3	H14	148	FQRLRYI**K**QLGGGYY
Sex comb on midleg-like protein 1	Q9UN30	H15	115	KHSYRLV**K**KLKLQKM

Serologically defined colon cancer antigen 1	O60524	H16	578	GATSCVI**K**NPTGEPI
Septin-10	Q9P0V9	H17	396	QAKFEHL**K**RLHQEER
Histone-lysine N-methyltransferase SET8	Q9NQR1	H18	199	AIAKQAL**K**KPIKGKQ
Splicing factor 1	Q15637	H19	165	GPRGNTL**K**NIEKECN
Serine/threonine-protein kinase Sgk1	O00141	H20	102	PSDFHFL**K**VIGKGSF
Protein SGT1	O95905	H21	96	WFIVYVI**K**QITKEFP
Paired amphipathic helix protein Sin3a	Q96ST3	I1	813	KEDKYKI**K**QIMHHFI
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1	Q96GM5	I2	173	LDIQEAL**K**RPIKQKR
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 3	Q6STE5	I3	148	VDIQEAL**K**RPMKQKR
U5 small nuclear ribonucleoprotein 40 kDa protein	Q96DI7	I4	145	SETGERV**K**RLKGHTS
SOSS complex subunit B1	Q9BQ15	I5	15	KDIKPGL**K**NLNLIFI
Transcription factor SOX-17	Q9H6I2	I6	149	RKQVKRL**K**RVEGGFL
Transcription elongation factor SPT5	O00267	I7	1042	PTKNNKV**K**VILGEDR
Serine/threonine-protein kinase SRPK1	Q96SB4	I8	190	GLPLPCV**K**KIIQQVL
STE20/SPS1-related proline-alanine-rich protein kinase	Q9UEW8	I9	92	RQERVAI**K**RINLEKC
Serine/threonine-protein kinase 3	Q13188	I10	441	DGDFDFL**K**NLSLEEL
ATP-dependent RNA helicase SUPV3L1, mitochondrial	Q8IYB8	I11	749	LLTPDML**K**QLEKEWM
Polycomb protein SUZ12	Q15022	I12	72	GAAVLPV**K**KPKMEHV
Synaptonemal complex protein 2-like	Q5T4T6	I13	370	KIFIIYL**K**KPMIISY
Synaptonemal complex protein 1	Q15431	I14	437	EVELEEL**K**KVLGEKE
Nesprin-1	Q8NF91	I15	4833	QDSGIVL**K**RVTIHLE
Nesprin-2	Q8WXH0	I16	2601	KLLESQI**K**QLEHGWE

Nesprin-2	Q8WXH0	I17	3992	QEQNELL**K**VVIKQTN
General transcription factor IIF subunit 2	P13984	I18	128	SENYMRL**K**RLQIEES
TATA box-binding protein-associated factor RNA polymerase I subunit A	Q15573	I19	357	KYLAKYL**K**NILMGNH
Transcription initiation factor TFIID subunit 2	Q6P1X5	I20	142	WKHVDEL**K**VLKIHIN
Transcription initiation factor TFIID subunit 4B	Q92750	I21	179	KVAVTPV**K**KLAQIGT
TATA box-binding protein-like protein 2	Q6SJ96	J1	218	LACKLDL**K**KIALHAK
Transcription elongation regulator 1	O14776	J2	981	TSTWKEV**K**KIIKEDP
G/T mismatch-specific thymine DNA glycosylase	Q13569	J3	248	EVFGVKV**K**NLEFGLQ
Methylcytosine dioxygenase TET1	Q8NFU7	J4	50	TLSPGKL**K**QLIQERD
General transcription factor 3C polypeptide 3	Q9Y5Q9	J5	728	FCLRLML**K**NPENHAL
Transcription factor AP-4	Q01664	J6	189	HMYPEKL**K**VIAQQVQ
Transducin-like enhancer protein 2	Q04725	J7	82	HKQAEIV**K**RLSGICA
Serine/threonine-protein kinase tousled-like 1	Q9UKI8	J8	436	NLHIREL**K**RINNEDN
Serine/threonine-protein kinase tousled-like 2	Q86UE8	J9	442	NLHIREL**K**RIHNEDN
DNA topoisomerase 2-beta	Q02880	J10	1226	KVGKPKV**K**KLQLEET
Targeting protein for Xklp2	Q9ULW0	J11	585	NLPEKKVKNVTQIEP
tRNA pseudouridine synthase A	Q9Y606	J12	184	HIRILGL**K**RVTGGFN
Transcription termination factor 2	Q9UNY4	J13	1023	SQWTNML**K**VVALHLK
U5 small nuclear ribonucleoprotein 200 kDa helicase	O75643	J14	2080	SNSLISI**K**RLTLQQK
Ubiquitin-conjugating enzyme E2 variant 1	Q13404	J15	118	WQNSYSI**K**VVLQELR
SUMO-conjugating enzyme UBC9	P63279	J16	110	WRPAITI**K**QILLGIQ
Ubiquitin carboxyl-terminal hydrolase 21	Q9UK80	J17	64	GLPDERL**K**KLELGRG
Ubiquitin carboxyl-terminal hydrolase 36	Q9P275	J18	408	LVHSSNV**K**VVLNQQA

Uridine-cytidine kinase-like 1	Q9NWZ5	J19	170	DLIISTL**K**KLKQGKS
Protein unc-84 homolog A	O94901	J20	374	KPTTSRL**K**QPLQGDS
Uracil-DNA glycosylase	P13051	J21	147	MCDIKDV**K**VVILGQD
Probable U3 small nucleolar RNA-associated protein 11	Q9Y3A2	K1	189	VTNQTGL**K**RIAKERQ
Small subunit processome component 20 homolog	O75691	K2	1591	HRRARAL**K**KLAKQLM
Small subunit processome component 20 homolog	O75691	K3	2690	SEQDPLL**K**NLSQEII
Vitamin D3 receptor	P11473	K4	321	IKFQVGL**K**KLNLHEE
Vezatin	Q9HBM0	K5	525	HCTVVPL**K**QPTLHIA
WW domain-binding protein 4	O75554	K6	81	KAYQEDL**K**RLGLESE
YEATS domain-containing protein 2	Q9ULM3	K7	900	AQGQQTL**K**VISGQKT
Zinc finger protein 280A	P59817	K8	305	FKCLSCV**K**VLKNIKF
Zinc finger protein 585A	Q6P3V2	K9	150	SQLKVHL**K**VLAGEKL
Zinc finger protein 585B	Q52M93	K10	150	SQFKVHL**K**VPTGEKL
Zinc finger and BTB domain-containing protein 4	Q9P1Z0	K11	301	GGPEHVV**K**VVGGHVL
Zinc finger protein ZFAT	Q9P243	K12	367	KKKYSDV**K**NLIKHIR
Zinc finger homeobox protein 4	Q86UP3	K13	1525	VSHLHKL**K**KVLQEAS
Zinc finger MYM-type protein 5	Q9UJ78	K14	454	GSSNTLL**K**KIEGIPE
Zinc finger protein 251	Q9BRH9	K15	663	KRYFIHI**K**KIFQERH
Zinc finger protein 292	O60281	K16	2531	RQKASNL**K**RVNKEKN
Zinc finger protein 509	Q6ZSB9	K17	70	DVFHLDV**K**NVSGIGQ
Zinc finger protein 644	Q9H582	K18	809	DHRRVAV**K**RVIKESK
Zinc finger protein 41	P51814	K19	201	NNLLSHV**K**VLIKERG
Zinc finger protein 8	P17098	K20	157	LKEQNNL**K**QLEFGLK

Information of peptide array containing the synthesized peptides for the search for a potential automethylation site in NSD2:

Table 16: *List of synthesized peptides for the NSD2 automethylation scanning array.*

Name	Residues	Position on Array	Sequence
Histone H3 K36 WT	29 – 43	A1	APATGGVKKPHRYRP
Histone H3 K36A Mut	29 – 43	A2	APATGGVAKPHRYRP
free space		A3	
Nuclear receptor SET domain-containing protein 2 NSD2	941 – 955	A4	EGDRGSRYQGVRGIG
Nuclear receptor SET domain-containing protein 2 NSD2	945 – 959	A5	GSRYQGVRGIGRVFK
Nuclear receptor SET domain-containing protein 2 NSD2	949 – 963	A6	QGVRGIGRVFKNALQ
Nuclear receptor SET domain-containing protein 2 NSD2	953 – 967	A7	GIGRVFKNALQEAEA
Nuclear receptor SET domain-containing protein 2 NSD2	957 – 971	A8	VFKNALQEAEARFRE
Nuclear receptor SET domain-containing protein 2 NSD2	961 – 975	A9	ALQEAEARFREIKLQ
Nuclear receptor SET domain-containing protein 2 NSD2	965 – 979	A10	AEARFREIKLQREAR
Nuclear receptor SET domain-containing protein 2 NSD2	969 – 983	A11	FREIKLQREARETQE
Nuclear receptor SET domain-containing protein 2 NSD2	973 – 987	A12	KLQREARETQESERK
Nuclear receptor SET domain-containing protein 2 NSD2	977 – 991	A13	EARETQESERKPPPY
Nuclear receptor SET domain-containing protein 2 NSD2	981 – 995	A14	TQESERKPPPYKHIK
Nuclear receptor SET domain-containing protein 2 NSD2	985 – 999	A15	ERKPPPYKHIKVNKP
Nuclear receptor SET domain-containing protein 2 NSD2	989 – 1003	A16	PPYKHIKVNKPYGKV
Nuclear receptor SET domain-containing protein 2 NSD2	993 – 1007	A17	HIKVNKPYGKVQIYT
Nuclear receptor SET domain-containing protein 2 NSD2	997 – 1011	A18	NKPYGKVQIYTADIS
Nuclear receptor SET domain-containing protein 2 NSD2	1001 – 1015	A19	GKVQIYTADISEIPK

Nuclear receptor SET domain-containing protein 2 NSD2	1005 − 1019	A20	IYTADISEIPKCNCK
Nuclear receptor SET domain-containing protein 2 NSD2	1009 − 1023	A21	DISEIPKCNCKPTDE
Nuclear receptor SET domain-containing protein 2 NSD2	1013 − 1027	B1	IPKCNCKPTDENPCG
Nuclear receptor SET domain-containing protein 2 NSD2	1017 − 1031	B2	NCKPTDENPCGFDSE
Nuclear receptor SET domain-containing protein 2 NSD2	1021 − 1035	B3	TDENPCGFDSECLNR
Nuclear receptor SET domain-containing protein 2 NSD2	1025 − 1039	B4	PCGFDSECLNRMLMF
Nuclear receptor SET domain-containing protein 2 NSD2	1029 − 1043	B5	DSECLNRMLMFECHP
Nuclear receptor SET domain-containing protein 2 NSD2	1033 − 1047	B6	LNRMLMFECHPQVCP
Nuclear receptor SET domain-containing protein 2 NSD2	1037 − 1051	B7	LMFECHPQVCPAGEF
Nuclear receptor SET domain-containing protein 2 NSD2	1041 − 1055	B8	CHPQVCPAGEFCQNQ
Nuclear receptor SET domain-containing protein 2 NSD2	1045 − 1059	B9	VCPAGEFCQNQCFTK
Nuclear receptor SET domain-containing protein 2 NSD2	1049 − 1063	B10	GEFCQNQCFTKRQYP
Nuclear receptor SET domain-containing protein 2 NSD2	1053 − 1067	B11	QNQCFTKRQYPETKI
Nuclear receptor SET domain-containing protein 2 NSD2	1057 − 1071	B12	FTKRQYPETKIIKTD
Nuclear receptor SET domain-containing protein 2 NSD2	1061 − 1075	B13	QYPETKIIKTDGKGW
Nuclear receptor SET domain-containing protein 2 NSD2	1065 − 1079	B14	TKIIKTDGKGWGLVA
Nuclear receptor SET domain-containing protein 2 NSD2	1069 − 1083	B15	KTDGKGWGLVAKRDI
Nuclear receptor SET domain-containing protein 2 NSD2	1073 − 1087	B16	KGWGLVAKRDIRKGE
Nuclear receptor SET domain-containing protein 2 NSD2	1077 − 1091	B17	LVAKRDIRKGEFVNE
Nuclear receptor SET domain-containing protein 2 NSD2	1081 − 1095	B18	RDIRKGEFVNEYVGE
Nuclear receptor SET domain-containing protein 2 NSD2	1085 − 1099	B19	KGEFVNEYVGELIDE

Nuclear receptor SET domain-containing protein 2 NSD2	1089 − 1103	B20	VNEYVGELIDEEECM
Nuclear receptor SET domain-containing protein 2 NSD2	1093 − 1107	B21	VGELIDEEECMARIK
Nuclear receptor SET domain-containing protein 2 NSD2	1097 − 1111	C1	IDEEECMARIKHAHE
Nuclear receptor SET domain-containing protein 2 NSD2	1101 − 1115	C2	ECMARIKHAHENDIT
Nuclear receptor SET domain-containing protein 2 NSD2	1105 − 1119	C3	RIKHAHENDITHFYM
Nuclear receptor SET domain-containing protein 2 NSD2	1109 − 1123	C4	AHENDITHFYMLTID
Nuclear receptor SET domain-containing protein 2 NSD2	1113 − 1127	C5	DITHFYMLTIDKDRI
Nuclear receptor SET domain-containing protein 2 NSD2	1117 − 1131	C6	FYMLTIDKDRIIDAG
Nuclear receptor SET domain-containing protein 2 NSD2	1121 − 1135	C7	TIDKDRIIDAGPKGN
Nuclear receptor SET domain-containing protein 2 NSD2	1125 − 1139	C8	DRIIDAGPKGNYSRF
Nuclear receptor SET domain-containing protein 2 NSD2	1129 − 1143	C9	DAGPKGNYSRFMNHS
Nuclear receptor SET domain-containing protein 2 NSD2	1133 − 1147	C10	KGNYSRFMNHSCQPN
Nuclear receptor SET domain-containing protein 2 NSD2	1137 − 1151	C11	SRFMNHSCQPNCETL
Nuclear receptor SET domain-containing protein 2 NSD2	1141 − 1155	C12	NHSCQPNCETLKWTV
Nuclear receptor SET domain-containing protein 2 NSD2	1145 − 1159	C13	QPNCETLKWTVNGDT
Nuclear receptor SET domain-containing protein 2 NSD2	1149 − 1163	C14	ETLKWTVNGDTRVGL
Nuclear receptor SET domain-containing protein 2 NSD2	1153 − 1167	C15	WTVNGDTRVGLFAVC
Nuclear receptor SET domain-containing protein 2 NSD2	1157 − 1171	C16	GDTRVGLFAVCDIPA
Nuclear receptor SET domain-containing protein 2 NSD2	1161 − 1175	C17	VGLFAVCDIPAGTEL
Nuclear receptor SET domain-containing protein 2 NSD2	1165 − 1179	C18	AVCDIPAGTELTFNY
Nuclear receptor SET domain-containing protein 2 NSD2	1169 − 1183	C19	IPAGTELTFNYNLDC

Nuclear receptor SET domain-containing protein 2 NSD2	1173 − 1187	C20	TELTFNYNLDCLGNE
Nuclear receptor SET domain-containing protein 2 NSD2	1177 − 1191	C21	FNYNLDCLGNEKTVC
Nuclear receptor SET domain-containing protein 2 NSD2	1181 − 1195	D1	LDCLGNEKTVCRCGA
Nuclear receptor SET domain-containing protein 2 NSD2	1185 − 1199	D2	GNEKTVCRCGASNCS
Nuclear receptor SET domain-containing protein 2 NSD2	1189 − 1203	D3	TVCRCGASNCSGFLG
Nuclear receptor SET domain-containing protein 2 NSD2	1193 − 1207	D4	CGASNCSGFLGDRPK
Nuclear receptor SET domain-containing protein 2 NSD2	1197 − 1211	D5	NCSGFLGDRPKTSTT
Nuclear receptor SET domain-containing protein 2 NSD2	1201 − 1215	D6	FLGDRPKTSTTLSSE
Nuclear receptor SET domain-containing protein 2 NSD2	1205 − 1219	D7	RPKTSTTLSSEEKGK
Nuclear receptor SET domain-containing protein 2 NSD2	1209 − 1223	D8	STTLSSEEKGKKTKK
Nuclear receptor SET domain-containing protein 2 NSD2	1213 − 1227	D9	SSEEKGKKTKKKTRR
Nuclear receptor SET domain-containing protein 2 NSD2	1217 − 1231	D10	KGKKTKKKTRRRRAK
Nuclear receptor SET domain-containing protein 2 NSD2	1221 − 1235	D11	TKKKTRRRRAKGEGK
Nuclear receptor SET domain-containing protein 2 NSD2	1225 − 1239	D12	TRRRRAKGEGKRQSE
Nuclear receptor SET domain-containing protein 2 NSD2	1229 − 1243	D13	RRRRAKGEGKRQSED
free space		D14	
Histone H3 K36 WT	29 − 43	D15	APATGGVKKPHRYRP
Histone H3 K36A Mut	29 − 43	D16	APATGGVAKPHRYRP

Information of peptide array containing the synthesized peptides of the somatic missense mutations in histone H3 variants:

Table 17: *List of somatic missense mutations in histone H3 variants. The position of the corresponding peptide spots in Figure 52 are indicated. Target lysines (K36) are printed in bold and the corresponding mutations are highlighted red.*

Histone H3 variant	Cancer Mutation	Position on Array	Sequence
Histone H3.3 WT	K36	A1	APATGGV**K**KPHRYRP
Histone H3.3 K36A Mut	K36A	A2	APATGGV**A**KPHRYRP
Histone H3.1	T32A	A3	APA**A**GGV**K**KPHRYRP
Histone H3.1	T32S	A4	APA**S**GGV**K**KPHRYRP
Histone H3.1	T32M	A5	APA**M**GGV**K**KPHRYRP
Histone H3.1	G34V	A6	APATG**V**V**K**KPHRYRP
Histone H3.1	G34R	A7	APATG**R**V**K**KPHRYRP
Histone H3.1	G34W	A8	APATG**W**V**K**KPHRYRP
Histone H3.1	G34L	A9	APATG**L**V**K**KPHRYRP
Histone H3.1	G34D	A10	APATG**D**V**K**KPHRYRP
Histone H3.3	G34S	A11	APATG**S**V**K**KPHRYRP
Histone H3.1	V35I	A12	APATGG**I**KKPHRYRP
Histone H3.2	V35G	A13	APATGG**G**KKPHRYRP
Histone H3.1	K37N	A14	APATGGV**K**NPHRYRP
Histone H3.1	P38S	A15	APATGGV**K**K**S**HRYRP
Histone H3.3	P38L	A16	APATGGV**K**K**L**HRYRP
Histone H3.1	P38T	A17	APATGGV**K**K**T**HRYRP
Histone H3.2	H39Q	A18	APATGGV**K**KP**Q**RYRP
Histone H3.1	R40H	A19	APATGGV**K**KPH**H**YRP
Histone H3.1	R40C	A20	APATGGV**K**KPH**C**YRP

Histone H3.1	R40P	A21	APATGGV**K**KPH**P**YRP
Histone H3.1	Y41C	B1	APATGGV**K**KPHR**C**RP
Histone H3.3 WT	K36	B2	APATGGV**K**KPHRYRP
Histone H3.3 K36A Mut	K36A	B3	APATGGV**A**KPHRYRP

6.2.3 *In vitro* Methylation of the Protein Domains

Protein domain methylation was performed in $40\,\mu$L methylation buffer containing $50\,$mM Tris pH 8.5, $50\,$mM NaCl and $0.5\,$mM DTT supplemented with $0.76\,\mu$M labeled [methyl-^{3}H]-SAM ($2.7\,$Tbq/mmol; PerkinElmer) and $3\,\mu$M NSD2 at $23\,^\circ$C for $4\,$h. Equal loading of target protein amounts were confirmed by Coomassie Brilliant Blue staining. The methylation was stopped by adding $5\,\mu$L of $5\,$x SDS-loading buffer. The proteins were subsequently boiled at $95\,^\circ$C for $5\,$min and separated by loading on a $12\,$% SDS-PAGE gel. Afterwards, the gel was dried with vacuum at $65\,^\circ$C and then incubated with Hyperfilm^{TM} high performance autoradiography films at $-80\,^\circ$C, in the dark for several days.

6.2.4 *In vitro* Methylation of the Histone H3 Peptides

Methylation of histone H3K36 and the K36M missense peptides was performed in $20\,\mu$L methylation buffer containing $50\,$mM Tris pH 8.5, $50\,$mM NaCl and $0.5\,$mM DTT supplemented with $0.76\,\mu$M labeled [methyl-^{3}H]-SAM ($2.7\,$Tbq/mmol; PerkinElmer) and $1.5\,\mu$M NSD2 at $23\,^\circ$C for $4\,$h. The methylation was stopped by adding $20\,\mu$L of $2\,$x Tricin-SDS-loading buffer. The reaction was subsequently boiled at $95\,^\circ$C for $5\,$min and separated by loading on a $16\,$% Tricine-SDS-PAGE gel. Afterwards, the gel was dried with vacuum at $55\,^\circ$C and then incubated with Hyperfilm^{TM} high performance autoradiography films at $-80\,^\circ$C, in the dark for several days.

6.2.5 Cell culture, Transfection and Immunoprecipitation

For mammalian expression, the full-length sequence encoding for the histone lysine methyltransferase NSD2 was cloned into the pECFP-C1 (Clontech, USA) and the substrate protein domains of ATRX and FANCM were subcloned into the pEYFP-C1 vector (Clontech, USA). Transfection and purification of the YFP-fused protein substrates was performed according to the protocol described in section 6.1.5.

6.3 The Histone Lysine Methyltransferase Clr4

6.3.1 Protein Expression and Purification

For bacterial expression, the plasmid encoding for Clr4 was transformed into *E.coli* BL21-CodonPlus (DE3) cells (Novagen, USA). These were grown in LB medium at 37 °C until an OD_{600} of 0.6 to 0.8 was reached. Protein expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and shifted to 30 °C for 4 h. Afterwards, the cells were harvested by centrifugation at 4.500 rpm, washed once with STE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA and 100 mM NaCl) and the cell pellet was stored at −20 °C until purification.

For purification the cell pellet was thawed on ice, resuspended in sonication buffer and lysed by ultra sound. The lysed cells were centrifuged at 18.000 rpm for 90 min at 4 °C and the supernatant was passed through Nickel-Nitrilotriacetic acid (Ni-NTA; Genaxxon) resin. Afterwards, the beads were washed twice with sonication buffer. The bound proteins were eluted with elution buffer containing excess of imidazole and then dialyzed against low glycerol dialysis buffer 1 for 3 h and afterwards over night against high glycerol dialysis buffer 2. The composition of the used buffers are shown in Table 8 (His_6-tag purification).

6.3.2 *In vitro* Methylation of the Peptide SPOT Arrays

Synthesis of the peptide arrays was performed according to the description in section 6.1.2. All peptide arrays were washed for 10 min with methylation buffer containing 50 mM Tris (pH 8.0), 20 mM KCl, 500 mM $MgCl_2$ and 1 mM DTT. Then the peptide SPOT membranes were incubated for 60 min in methylation buffer containing 0.5 μM Clr4 and 0.76 μM labeled [methyl-^{3}H]-SAM (Perkin Elmer) at 23 °C. Afterwards, the arrays were washed five times with 50 mM NH_4HCO_3 and 1 % SDS and then incubated for 5 min in Amplify NAMP100V solution (GE Healthcare). The membranes were exposed on Hyperfilm^{TM} high performance autoradiography films (GE Healthcare) in the dark for several days at −80 °C. Film development was performed on an Optimus TR developing machine. Quantification and analysis of the developed films was performed as described in section 6.1.3.

Information of peptide array containing the synthesized peptides of the putative non-histone substrates of Clr4:

Table 18: *List of interaction partners of Clr4, which are putative substrate of Clr4. The position of the corresponding peptide spots in Figure 57 are indicated. Target lysine residues or the corresponding lysine to alanine mutants are printed in bold.*

Name	Target Lysine	Position on Array	Sequence
Histone H3 WT	K9	A1	ARTKQTAR**K**STGGKAPRK
Histone H3 Mut	K9A	A2	ARTKQTAR**A**STGGKAPRK
mRNA export protein Mlo3 WT	K167	A3	NGAKSSKR**K**TTRRRRTPN
mRNA export protein Mlo3 Mut	K167A	A4	NGAKSSKR**A**TTRRRRTPN
Chromatin-associated protein Swi6 WT	K144	A5	GRPEPSKR**K**RTARPKKPE
Chromatin-associated protein Swi6 Mut	K144A	A6	GRPEPSKR**A**RTARPKKPE
HMG box-containing protein Spbc28F2.11 WT	K250	A7	QHAKKPKR**K**HTRSTVPTS
HMG box-containing protein Spbc28F2.11 Mut	K250A	A8	QHAKKPKR**A**HTRSTVPTS
HMG box-containing protein Spbc28F2.11 WT	K292	A9	KREKKKRR**K**SSMSSSITT
HMG box-containing protein Spbc28F2.11 Mut	K292A	A10	KREKKKRR**A**SSMSSSITT
Chromodomain helicase Hrp3 WT	K189	A11	DVFPSKHR**K**GTRNGSSFS
Chromodomain helicase Hrp3 Mut	K189A	A12	DVFPSKHR**A**GTRNGSSFS
ATP-dependent RNA helicase Dbp2 WT	K165	A13	GRTGRAGA**K**GTAYTYFTS
ATP-dependent RNA helicase Dbp2 Mut	K165A	A14	GRTGRAGA**A**GTAYTYFTS
Iec3 WT	K153	A15	SSSRKQKR**K**RTSEGPSER
Iec3 Mut	K153A	A16	SSSRKQKR**A**RTSEGPSER
Meiotic coiled-coil protein 1 Mcp1 WT	K132	A17	PESSPPAR**K**TTGKIENKK
Meiotic coiled-coil protein 1 Mcp1 Mut	K132A	A18	PESSPPAR**A**TTGKIENKK
Nuclear cap-binding protein subunit 1 Cbc1 WT	K11	A19	YRGSTRPR**K**RTREGENYG

Nuclear cap-binding protein subunit 1 Cbc1 Mut	K11A	A20	YRGSTRPR**A**RTREGENYG
Chromatin modification-related protein Rik1 WT	K460	A21	FLCIYDSA**K**RSRLVYIEK
Chromatin modification-related protein Rik1 Mut	K460A	B1	FLCIYDSA**A**RSRLVYIEK
Chromatin modification-related protein Rik1 WT	K502	B2	KKDTEVAR**K**VFESEISCL
Chromatin modification-related protein Rik1 Mut	K502A	B3	KKDTEVAR**A**VFESEISCL
Histone H3 WT	K9	B4	ARTKQTAR**K**STGGKAPRK
Histone H3 Mut	K9A	B5	ARTKQTAR**A**STGGKAPRK

6.4 Development of an Advanced Non-radioactive, High-throughput PKMT Activity Assay

6.4.1 Protein Expression and Purification

The DNA sequence encoding the SET domain of human SUV39H1 (residues 81 - 412) and human HP1β protein (residues 1 - 185) was amplified from cDNA isolated from human HEK293 cells and cloned as GST fusion protein into pGEX-6P-2 vector. GST-tagged SUV39H1 was cloned and purified by Dr. S. Kudithipudi and the GST-fused HP1β protein was cloned and purified by Dr. A. Dhayalan. Purification was conducted as described in section 6.1.1.

6.4.2 Reading Domain PKMT Assay

Each well of a 96-well plate was coated with $100\,\mu$L avidin solution ($1\,\mu$g avidin in $100\,\mu$L of $100\,$mM NaHCO$_3$ pH 9.6) overnight at $4\,°$C. The wells were washed three times for 5 min with $200\,\mu$L of PBST buffer ($140\,$mM NaCl, $2.7\,$mM KCl, $4.3\,$mM Na$_2$HPO$_4$, $1.4\,$mM K$_2$HPO$_4$, $0.05\,\%$ (v/v) Tween 50, pH 7.2) + $500\,$mM NaCl and subsequently once for 5 min with $200\,\mu$L of PBST buffer. The biotinylated peptides (Intavis AG, Germany) ($100\,$nM) were then added to the avidin-coated wells and incubated for 30 min with continuous shaking. Afterwards, the unbound peptides were removed by washing the plate three times for 5 min with $200\,\mu$L of PBST buffer. The wells with the bound peptides were then blocked with BSA blocking solution ($2\,\%$ bovine serum albumin in PBST buffer) for $2\,$h at room temperature. After blocking, the plate was washed three times for 5 min with $200\,\mu$L of PBST buffer. In the meantime, HP1β protein was diluted in interaction buffer ($100\,$mM KCl, $20\,$mM HEPES, $1\,$mM EDTA, $0.1\,$mM DTT and $10\,\%$ glycerol pH 7.5) and $50\,\mu$L of the solution was added to the wells, followed by incubation for $1\,$h at room temperature, with continuous shaking. The wells were washed three times for 5 min with $200\,\mu$L of PBST buffer and then incubated with $50\,\mu$L HP1β monoclonal antibody (Active Motif,#39979; $0.37\,\mu$g*mL^{-1}) or GST-specific antibody (GE Healthcare, #27-4577; 1:6000 dilution) for $1\,$h at room temperature with continuous shaking. Subsequently, the plate was washed and incubated with $50\,\mu$L of the respective HRP-conjugated secondary antibody for $1\,$h at room temperature, with continuous shaking. Alternatively, the HRP-conjugated anti-GST antibody (GE Healthcare, RPN1236) was used. Finally, the wells were washed five times for 5 min with $200\,\mu$L of PBST buffer and $50\,\mu$L of an enhanced chemiluminescent substrate for detection of HRP (Thermo Scientific, #32106) was added to the plate, and luminescence signal was detected with an Enspire microplate reader (PerkinElmer).

The used equation for calculating the signal-to-noise (SN) ratio was defined as:

$$\text{SN} = \frac{\mu_{\text{Signal}} - \mu_{\text{Control}}}{\text{SD}_{\text{Signal}}}$$

in which μ_{Signal} and μ_{Control} are the received average values of signal and background control, and $\text{SD}_{\text{Signal}}$ is the standard deviation of the signal.

The Z-factor was defined as:

$$Z = 1 - \frac{3 \cdot (\text{SD}_{\text{Signal}} + \text{SD}_{\text{Control}})}{\mu_{\text{Signal}} + \mu_{\text{Control}}}$$

in which $\text{SD}_{\text{Signal}}$ and $\text{SD}_{\text{Control}}$ are the calculated standard deviations of signal and background controls, respectively, and μ_{Signal} and μ_{Control} are the means of corresponding signals.

6.4.3 *In vitro* Methylation of Peptides and MALDI Analysis

In vitro peptide methylation was performed by incubation of 100 nM biotinylated H3 (residues 1 - 19) peptide in methylation buffer (50 mM Tris-HCl, 5 mM MgCl$_2$, 4 mM DTT pH 9.0) containing 200 nM SET-SUV39H1 enzyme and unlabeled SAM for 3 h at 25 °C. 50 µL of the methylation reactions were added to the avidin-coated wells.

For MALDI analysis, 1 µL of the methylation reaction was diluted with 9 µL of 0.1 % TFA. 10 % of this mixture was spotted on a pre-spotted Anchor chip (PAC) HCCA plate (Bruker Daltonics, #227463), and the methylation of the peptides was assessed by mass spectrometry using Bruker Autoflex Speed MALDI-TOF system (Brucker Daltonics). The spectra were collected in the mass-to-charge ratio 500 – 3.500 Da range in reflector mode. The system was calibrated with a peptide calibration standard (Bruker Daltonics), with peptides covering masses of 700 – 3.200 Da. The spectra were collected using flexControl and flexAnalysis software (Bruker Daltonics).

7 Bibliography

[1] C. Walsh, **Posttranslational modification of proteins: expanding nature's inventory**. *Roberts and Company Publishers*, (2006).

[2] W. Fischle, B. S. Tseng, H. L. Dormann, B. M. Ueberheide, B. A. Garcia, J. Shabanowitz, D. F. Hunt, H. Funabiki, and C. D. Allis, **Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation**. *Nature*, vol. 438, no. 7071, pp. 1116–1122, (2005).

[3] C. J. Nelson, H. Santos-Rosa, and T. Kouzarides, **Proline isomerization of histone H3 regulates lysine methylation and gene expression**. *Cell*, vol. 126, no. 5, pp. 905–916, (2006).

[4] A. Clements, A. N. Poux, W.-S. Lo, L. Pillus, S. L. Berger, and R. Marmorstein, **Structural basis for histone and phosphohistone binding by the GCN5 histone acetyltransferase**. *Molecular cell*, vol. 12, no. 2, pp. 461–473, (2003).

[5] G. A. Khoury, R. C. Baliban, and C. A. Floudas, **Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database**. *Scientific reports*, vol. 1, no. 90, (2011).

[6] L. N. Johnson and R. J. Lewis, **Structural basis for control by phosphorylation**. *Chemical reviews*, vol. 101, no. 8, pp. 2209–2242, (2001).

[7] P. Sassone-Corsi, **The cyclic AMP pathway**. *Cold Spring Harbor perspectives in biology*, vol. 4, no. 12, p. a011148, (2012).

[8] H. Nimmo and P. Cohen, **Glycogen synthetase kinase 2 (GSK 2); the identification of a new protein kinase in skeletal muscle**. *FEBS letters*, vol. 47, no. 1, pp. 162–166, (1974).

[9] R. J. Brushia and D. A. Walsh, **Phosphorylase kinase: the complexity of its regulation is reflected in the complexity of its structure**. *Front Biosci*, vol. 4, pp. D618–D641, (1999).

[10] J. Hollebeke, P. Van Damme, and K. Gevaert, **N-terminal acetylation and other functions of Nα-acetyltransferases**. *Biological chemistry*, vol. 393, no. 4, pp. 291–298, (2012).

[11] L. Guo, H. Münzberg, R. C. Stuart, E. A. Nillni, and C. Bjørbæk, **N-acetylation of hypothalamic α-melanocyte-stimulating hormone and regulation by leptin**. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 32, pp. 11797–11802, (2004).

[12] R. Behnia, B. Panic, J. R. Whyte, and S. Munro, **Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p**. *Nature cell biology*, vol. 6, no. 5, pp. 405–413, (2004).

[13] K. K. Starheim, D. Gromyko, R. Evjenth, A. Ryningen, J. E. Varhaug, J. R. Lillehaug, and T. Arnesen, **Knockdown of human Nα-terminal acetyltransferase complex C leads to p53-dependent apoptosis and aberrant human Arl8b localization**. *Molecular and cellular biology*, vol. 29, no. 13, pp. 3569–3581, (2009).

[14] M. Kamita, Y. Kimura, Y. Ino, R. M. Kamp, B. Polevoda, F. Sherman, and H. Hirano, **Nα-Acetylation of yeast ribosomal proteins and its effect on protein synthesis**. *Journal of proteomics*, vol. 74, no. 4, pp. 431–441, (2011).

[15] D. Gromyko, T. Arnesen, A. Ryningen, J. E. Varhaug, and J. R. Lillehaug, **Depletion of the human Nα-terminal acetyltransferase A induces p53-dependent apoptosis and p53-independent growth inhibition**. *International Journal of Cancer*, vol. 127, no. 12, pp. 2777–2789, (2010).

[16] C. Yi, H. Pan, J. Seebacher, I.-H. Jang, S. Hyberts, G. Heffron, M. G. Vander Heiden, R. Yang, F. Li, J. Locasale, H. Sharfi, B. Zhai, R. Rodriguez-Mias, H. Luithardt, L. Cantley, G. Daley, J. Asara, S. Gygi, G. Wagner, C.-F. Liu, and J. Yuan, **Metabolic Regulation of Protein N-alpha-acetylation by Bcl-xL Promotes Cell Survival**. *Cell*, vol. 146, no. 4, pp. 607 – 620, (2011).

[17] T. Kouzarides, **Acetylation: a regulatory modification to rival phosphorylation?**. *The EMBO journal*, vol. 19, no. 6, pp. 1176–1179, (2000).

[18] V. Allfrey, R. Faulkner, and A. Mirsky, **Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis**. *Proceedings of the National Academy of Sciences*, vol. 51, no. 5, pp. 786–794, (1964).

[19] H. Chen, R. J. Lin, W. Xie, D. Wilpitz, and R. M. Evans, **Regulation of hormone-induced histone hyper-acetylation and gene activation via acetylation of an acetylase**. *Cell*, vol. 98, no. 5, pp. 675–686, (1999).

[20] Y.-H. Jin, E.-J. Jeon, Q.-L. Li, Y. H. Lee, J.-K. Choi, W.-J. Kim, K.-Y. Lee, and S.-C. Bae, **Transforming growth factor-β stimulates p300-dependent RUNX3 acetylation, which inhibits ubiquitination-mediated degradation**. *Journal of Biological Chemistry*, vol. 279, no. 28, pp. 29409–29417, (2004).

[21] S. W. L'Hernault and J. L. Rosenbaum, **Chlamydomonas α-tubulin is posttranslationally modified by acetylation on the ε-amino group of a lysine**. *Biochemistry*, vol. 24, no. 2, pp. 473–478, (1985).

[22] W. Gu and R. G. Roeder, **Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain**. *Cell*, vol. 90, no. 4, pp. 595–606, (1997).

[23] N. Munshi, M. Merika, J. Yie, K. Senger, G. Chen, and D. Thanos, **Acetylation of HMG I (Y) by CBP turns off IFNβ expression by disrupting the enhanceosome**. *Molecular cell*, vol. 2, no. 4, pp. 457–467, (1998).

[24] R. Ambler, **ε-N-Methyl-lysine in bacterial flagellar protein**. *Nature*, vol. 184, pp. 56–57, (1959).

[25] S. G. Clarke, **Protein methylation**. *Current opinion in cell biology*, vol. 5, no. 6, pp. 977–983, (1993).

[26] S. G. Clarke, **Protein methylation at the surface and buried deep: thinking outside the histone box**. *Trends in biochemical sciences*, vol. 38, no. 5, pp. 243–252, (2013).

[27] R. Sprung, Y. Chen, K. Zhang, D. Cheng, T. Zhang, J. Peng, and Y. Zhao, **Identification and validation of eukaryotic aspartate and glutamate methylation in proteins**. *Journal of proteome research*, vol. 7, no. 3, pp. 1001–1006, (2008).

[28] S. L. Sanders, M. Portoso, J. Mata, J. Bähler, R. C. Allshire, and T. Kouzarides, **Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage**. *Cell*, vol. 119, no. 5, pp. 603–614, (2004).

[29] S. Pérez-Lluch, E. Blanco, A. Carbonell, D. Raha, M. Snyder, F. Serras, and M. Corominas, **Genome-wide chromatin occupancy analysis reveals a role for ASH2 in transcriptional pausing**. *Nucleic acids research*, vol. 39, no. 11, pp. 4628–4639, (2011).

[30] J. C. Rice, S. D. Briggs, B. Ueberheide, C. M. Barber, J. Shabanowitz, D. F. Hunt, Y. Shinkai, and C. D. Allis, **Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains**. *Molecular cell*, vol. 12, no. 6, pp. 1591–1598, (2003).

[31] P. Völkel and P.-O. Angrand, **The control of histone lysine methylation in epigenetic regulation**. *Biochimie*, vol. 89, no. 1, pp. 1–20, (2007).

[32] G. Schotta, M. Lachner, K. Sarma, A. Ebert, R. Sengupta, G. Reuter, D. Reinberg, and T. Jenuwein, **A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin**. *Genes & development*, vol. 18, no. 11, pp. 1251–1262, (2004).

[33] M. V. Botuyan, J. Lee, I. M. Ward, J.-E. Kim, J. R. Thompson, J. Chen, and G. Mer, **Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair**. *Cell*, vol. 127, no. 7, pp. 1361–1373, (2006).

[34] J. C. Rice, K. Nishioka, K. Sarma, R. Steward, D. Reinberg, and C. D. Allis, **Mitotic-specific methylation of histone H4 Lys 20 follows increased PR-Set7 expression and its localization to mitotic chromosomes**. *Genes & development*, vol. 16, no. 17, pp. 2225–2230, (2002).

[35] S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, and D. J. Patel, **How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers**. *Nature structural & molecular biology*, vol. 14, no. 11, pp. 1025–1040, (2007).

[36] J. C. Eissenberg, **Structural biology of the chromodomain: form and function**. *Gene*, vol. 496, no. 2, pp. 69–78, (2012).

[37] M. Lachner, D. O'Carroll, S. Rea, K. Mechtler, and T. Jenuwein, **Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins**. *Nature*, vol. 410, no. 6824, pp. 116–120, (2001).

[38] A. J. Bannister, P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire, and T. Kouzarides, **Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain**. *Nature*, vol. 410, no. 6824, pp. 120–124, (2001).

[39] S. J. Nielsen, R. Schneider, U.-M. Bauer, A. J. Bannister, A. Morrison, D. O'Carroll, R. Firestein, M. Cleary, T. Jenuwein, R. E. Herrera, and T. Kouzarides, **Rb targets histone H3 methylation and HP1 to promoters**. *Nature*, vol. 412, no. 6846, pp. 561–565, (2001).

[40] R. J. Sims, C.-F. Chen, H. Santos-Rosa, T. Kouzarides, S. S. Patel, and D. Reinberg, **Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains**. *Journal of Biological Chemistry*, vol. 280, no. 51, pp. 41789–41792, (2005).

[41] M. G. Pray-Grant, J. A. Daniel, D. Schieltz, J. R. Yates, and P. A. Grant, **Chd1 chromodomain links histone H3 methylation with SAGA-and SLIK-dependent acetylation**. *Nature*, vol. 433, no. 7024, pp. 434–438, (2005).

[42] K. Egorova, O. Olenkina, and L. Olenina, **Lysine methylation of nonhistone proteins is a way to regulate their stability and function**. *Biochemistry (Moscow)*, vol. 75, no. 5, pp. 535–548, (2010).

[43] S. Chuikov, J. K. Kurash, J. R. Wilson, B. Xiao, N. Justin, G. S. Ivanov, K. McKinney, P. Tempst, C. Prives, S. J. Gamblin, N. A. Barlev, and D. Reinberg, **Regulation of p53 activity through lysine methylation**. *Nature*, vol. 432, no. 7015, pp. 353–360, (2004).

[44] J. Huang, L. Perez-Burgos, B. J. Placek, R. Sengupta, M. Richter, J. A. Dorsey, S. Kubicek, S. Opravil, T. Jenuwein, and S. L. Berger, **Repression of p53 activity by Smyd2-mediated methylation**. *Nature*, vol. 444, no. 7119, pp. 629–632, (2006).

[45] W. K. Paik and S. Kim, **Enzymatic methylation of protein fractions from calf thymus nuclei**. *Biochemical and biophysical research communications*, vol. 29, no. 1, pp. 14–20, (1967).

[46] E. C. Shen, M. F. Henry, V. H. Weiss, S. R. Valentini, P. A. Silver, and M. S. Lee, **Arginine methylation facilitates the nuclear export of hnRNP proteins**. *Genes & development*, vol. 12, no. 5, pp. 679–691, (1998).

[47] M. T. Bedford, A. Frankel, M. B. Yaffe, S. Clarke, P. Leder, and S. Richard, **Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains**. *Journal of Biological Chemistry*, vol. 275, no. 21, pp. 16030–16036, (2000).

[48] D. Chen, H. Ma, H. Hong, S. S. Koh, S.-M. Huang, B. T. Schurter, D. W. Aswad, and M. R. Stallcup, **Regulation of transcription by a protein methyltransferase**. *Science*, vol. 284, no. 5423, pp. 2174–2177, (1999).

[49] S. Pal, S. N. Vishwanath, H. Erdjument-Bromage, P. Tempst, and S. Sif, **Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes**. *Molecular and cellular biology*, vol. 24, no. 21, pp. 9630–9645, (2004).

[50] J. Coté, F.-M. Boisvert, M.-C. Boulanger, M. T. Bedford, and S. Richard, **Sam68 RNA binding protein is an in vivo substrate for protein arginine N-methyltransferase 1**. *Molecular biology of the cell*, vol. 14, no. 1, pp. 274–287, (2003).

[51] F. Herrmann, M. Bossert, A. Schwander, E. Akgün, and F. O. Fackelmayer, **Arginine methylation of scaffold attachment factor A by heterogeneous nuclear ribonucleoprotein particle-associated PRMT1**. *Journal of Biological Chemistry*, vol. 279, no. 47, pp. 48774–48779, (2004).

[52] C. Abramovich, B. Yakobson, J. Chebath, and M. Revel, **A protein-arginine methyltransferase binds to the intracytoplasmic domain of the IFNAR1 chain in the type I interferon receptor**. *The EMBO journal*, vol. 16, no. 2, pp. 260–266, (1997).

[53] B. P. Pollack, S. V. Kotenko, W. He, L. S. Izotova, B. L. Barnoski, and S. Pestka, **The human homologue of the yeast proteins Skb1 and Hsl7p interacts with Jak kinases and contains protein methyltransferase activity**. *Journal of Biological Chemistry*, vol. 274, no. 44, pp. 31531–31542, (1999).

[54] K. A. Mowen, B. T. Schurter, J. W. Fathman, M. David, and L. H. Glimcher, **Arginine methylation of NIP45 modulates cytokine gene expression in effector T lymphocytes**. *Molecular cell*, vol. 15, no. 4, pp. 559–571, (2004).

[55] F.-M. Boisvert, U. Déry, J.-Y. Masson, and S. Richard, **Arginine methylation of MRE11 by PRMT1 is required for DNA damage checkpoint control**. *Genes & development*, vol. 19, no. 6, pp. 671–676, (2005).

[56] M. T. Bedford and S. Richard, **Arginine Methylation: An Emerging Regulatorof Protein Function**. *Molecular Cell*, vol. 18, no. 3, pp. 263 – 272, (2005).

[57] W. J. Friesen, S. Massenet, S. Paushkin, A. Wyce, and G. Dreyfuss, **SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets**. *Molecular cell*, vol. 7, no. 5, pp. 1111–1117, (2001).

[58] J. Lhoest and C. Colson, **Genetics of ribosomal protein methylation in *Escherichia coli*. II. A mutant lacking a new type of methylated amino acid, N5-methylglutamine, in protein L3.**. *Molecular & general genetics: MGG*, vol. 154, no. 2, p. 175, (1977).

[59] P. Tessarz, H. Santos-Rosa, S. C. Robson, K. B. Sylvestersen, C. J. Nelson, M. L. Nielsen, and T. Kouzarides, **Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification**. *Nature*, vol. 505, no. 7484, pp. 564–568, (2014).

[60] V. Dinçbas-Renqvist, Å. Engström, L. Mora, V. Heurgué-Hamard, R. Buckingham, and M. Ehrenberg, **A post-translational modification in the GGQ motif of RF2 from *Escherichia coli* stimulates termination of translation**. *The EMBO journal*, vol. 19, no. 24, pp. 6900–6907, (2000).

[61] H. Song, P. Mugnier, A. K. Das, H. M. Webb, D. R. Evans, M. F. Tuite, B. A. Hemmings, and D. Barford, **The crystal structure of human eukaryotic release factor eRF1-mechanism of stop codon recognition and peptidyl-tRNA hydrolysis**. *Cell*, vol. 100, no. 3, pp. 311–321, (2000).

[62] E. Scolnick, R. Tompkins, T. Caskey, and M. Nirenberg, **Release factors differing in specificity for terminator codons**. *Proceedings of the National Academy of Sciences*, vol. 61, no. 2, pp. 768–774, (1968).

[63] L. Frolova, X. Le Goff, H. H. Rasmussen, S. Cheperegin, G. Drugeon, M. Kress, I. Arman, A.-L. Haenni, J. E. Celis, M. Phllippe, J. Justesen, and L. Kisselev, **A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor**. *Nature*, vol. 372, pp. 701 – 703, (1994).

[64] B. Vestergaard, L. B. Van, G. R. Andersen, J. Nyborg, R. H. Buckingham, and M. Kjeldgaard, **Bacterial polypeptide release factor RF2 is structurally distinct from eukaryotic eRF1**. *Molecular cell*, vol. 8, no. 6, pp. 1375–1382, (2001).

[65] L. Y. Frolova, R. Y. Tsivkovskii, G. F. Sivolobova, N. Y. Oparina, O. I. Serpinsky, V. M. Blinov, S. I. Tatkov, and L. L. Kisselev, **Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis**. *Rna*, vol. 5, no. 08, pp. 1014–1020, (1999).

[66] G. L. Cantoni, **Biological methylation: selected aspects**. *Annual review of biochemistry*, vol. 44, no. 1, pp. 435–451, (1975).

[67] P. Chiang, R. K. Gordon, J. Tal, G. Zeng, B. Doctor, K. Pardhasaradhi, and P. P. McCann, **S-Adenosylmethionine and methylation.**. *The FASEB Journal*, vol. 10, no. 4, pp. 471–480, (1996).

[68] B. C. Smith and J. M. Denu, **Chemical mechanisms of histone lysine and arginine modifications**. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1789, no. 1, pp. 45–57, (2009).

[69] X. Cheng, **DNA modification by methyltransferases**. *Current opinion in structural biology*, vol. 5, no. 1, pp. 4–10, (1995).

[70] A. Jeltsch, **Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases**. *Chembiochem*, vol. 3, no. 4, pp. 274–293, (2002).

[71] A. Bird, **DNA methylation patterns and epigenetic memory**. *Genes & development*, vol. 16, no. 1, pp. 6–21, (2002).

[72] H. L. Schubert, R. M. Blumenthal, and X. Cheng, **Many paths to methyltransfer: a chronicle of convergence**. *Trends in biochemical sciences*, vol. 28, no. 6, pp. 329–335, (2003).

[73] X. Zhang, L. Zhou, and X. Cheng, **Crystal structure of the conserved core of protein arginine methyltransferase PRMT3**. *The EMBO journal*, vol. 19, no. 14, pp. 3509–3519, (2000).

[74] Z. Yang, L. Shipman, M. Zhang, B. P. Anton, R. J. Roberts, and X. Cheng, **Structural Characterization and Comparative Phylogenetic Analysis of *Escherichia coli* HemK, a Protein (N5)-glutamine Methyltransferase**. *Journal of molecular biology*, vol. 340, no. 4, pp. 695–706, (2004).

[75] D. Liger, L. Mora, N. Lazar, S. Figaro, J. Henri, N. Scrima, R. H. Buckingham, H. van Tilbeurgh, V. Heurgué-Hamard, and M. Graille, **Mechanism of activation of methyltransferases involved in translation by the Trm112 'hub' protein**. *Nucleic Acids Research*, vol. 39, no. 14, pp. 6249–6259, (2011).

[76] M. M. Dixon, S. Huang, R. G. Matthews, and M. Ludwig, **The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B12**. *Structure*, vol. 4, no. 11, pp. 1263–1275, (1996).

[77] H. L. Schubert, K. S. Wilson, E. Raux, S. C. Woodcock, and M. J. Warren, **The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase**. *Nature Structural & Molecular Biology*, vol. 5, no. 7, pp. 585–592, (1998).

[78] G. Michel, V. Sauvé, R. Larocque, Y. Li, A. Matte, and M. Cygler, **The structure of the RlmB 23S rRNA methyltransferase reveals a new methyltransferase fold with a unique knot**. *Structure*, vol. 10, no. 10, pp. 1303–1315, (2002).

[79] K. Lim, H. Zhang, A. Tempczyk, W. Krajewski, N. Bonander, J. Toedt, A. Howard, E. Eisenstein, and O. Herzberg, **Structure of the YibK methyltransferase from Haemophilus influenzae (HI0766): a cofactor bound at a site formed by a knot**. *Proteins: Structure, Function, and Bioinformatics*, vol. 51, no. 1, pp. 56–67, (2003).

[80] R. C. Trievel, B. M. Beach, L. M. Dirk, R. L. Houtz, and J. H. Hurley, **Structure and catalytic mechanism of a SET domain protein methyltransferase**. *Cell*, vol. 111, no. 1, pp. 91–103, (2002).

[81] T. Nakayashiki, K. Nishimura, and H. Inokuchi, **Cloning and sequencing of a previously unidentified gene that is involved in the biosynthesis of heme in *Escherichia coli***. *Gene*, vol. 153, no. 1, pp. 67–70, (1995).

[82] L. Le Guen, R. Santos, and J.-M. Camadro, **Functional analysis of the hemK gene product involvement in protoporphyrinogen oxidase activity in yeast**. *FEMS microbiology letters*, vol. 173, no. 1, pp. 175–182, (1999).

[83] T. Malone, R. M. Blumenthal, and X. Cheng, **Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyl-transferases, and suggests a catalytic mechanism for these enzymes**. *Journal of molecular biology*, vol. 253, no. 4, pp. 618–632, (1995).

[84] J. M. Bujnicki and M. Radlinska, **Is the HemK Family of Putative S-Adenosylmethionine-Dependent Methyltransferases a "Missing" ζ Subfamily of Adenine Methyltransferases? A Hypothesis**. *IUBMB life*, vol. 48, no. 3, pp. 247–249, (1999).

[85] D. Ratel, J.-L. Ravanat, M.-P. Charles, N. Platet, L. Breuillaud, J. Lunardi, F. Berger, and D. Wion, **Undetectable levels of N6-methyl adenine in mouse DNA: cloning and analysis of PRED28, a gene coding for a putative mammalian DNA adenine methyltransferase**. *FEBS letters*, vol. 580, no. 13, pp. 3179–3184, (2006).

[86] K. Nakahigashi, N. Kubo, S.-i. Narita, T. Shimaoka, S. Goto, T. Oshima, H. Mori, M. Maeda, C. Wada, and H. Inokuchi, **HemK, a class of protein methyl transferase with similarity to DNA methyl transferases, methylates polypeptide chain release factors, and hemK knockout induces defects in translational termination**. *Proceedings of the National Academy of Sciences*, vol. 99, no. 3, pp. 1473–1478, (2002).

[87] V. Heurgué-Hamard, S. Champ, Å. Engström, M. Ehrenberg, and R. H. Buckingham, **The hemK gene in Escherichia coli encodes the N5-glutamine methyltransferase that modifies peptide release factors**. *The EMBO journal*, vol. 21, no. 4, pp. 769–778, (2002).

[88] S. Figaro, N. Scrima, R. H. Buckingham, and V. Heurgué-Hamard, **HemK2 protein, encoded on human chromosome 21, methylates translation termination factor eRF1**. *{FEBS} Letters*, vol. 582, no. 16, pp. 2352 – 2356, (2008).

[89] P. Liu, S. Nie, B. Li, Z.-Q. Yang, Z.-M. Xu, J. Fei, C. Lin, R. Zeng, and G.-L. Xu, **Deficiency in a Glutamine-Specific Methyltransferase for Release Factor Causes Mouse Embryonic Lethality**. *Molecular and Cellular Biology*, vol. 30, no. 17, pp. 4245–4253, (2010).

[90] V. Heurgué-Hamard, S. Champ, L. Mora, T. Merkoulova-Rainon, L. L. Kisselev, and R. H. Buckingham, **The glutamine residue of the conserved GGQ motif in *Saccharomyces cerevisiae* release factor eRF1 is methylated by the product of the YDR140w gene**. *Journal of Biological Chemistry*, vol. 280, no. 4, pp. 2439–2445, (2005).

[91] H. L. Schubert, J. D. Phillips, and C. P. Hill, **Structures along the Catalytic Pathway of PrmC/HemK, an N5-Glutamine AdoMet-Dependent Methyltransferase**. *Biochemistry*, vol. 42, no. 19, pp. 5592–5599, (2003). PMID: 12741815.

[92] M. Graille, V. Heurgué-Hamard, S. Champ, L. Mora, N. Scrima, N. Ulryck, H. van Tilbeurgh, and R. H. Buckingham, **Molecular basis for bacterial class I release factor methylation by PrmC**. *Molecular cell*, vol. 20, no. 6, pp. 917–927, (2005).

[93] K. Ito, M. Uno, and Y. Nakamura, **A tripeptide anticodon deciphers stop codons in messenger RNA**. *Nature*, vol. 403, no. 6770, pp. 680–684, (2000).

[94] V. Heurgué-Hamard, M. Graille, N. Scrima, N. Ulryck, S. Champ, H. van Tilbeurgh, and R. H. Buckingham, **The Zinc Finger Protein Ynr046w Is Plurifunctional and a Component of the eRF1 Methyltransferase in Yeast**. *Journal of Biological Chemistry*, vol. 281, no. 47, pp. 36140–36148, (2006).

[95] L. Mora, V. Heurgué-Hamard, M. de Zamaroczy, S. Kervestin, and R. H. Buckingham, **Methylation of Bacterial Release Factors RF1 and RF2 Is Required for Normal Translation Termination *in Vivo***. *Journal of Biological Chemistry*, vol. 282, no. 49, pp. 35638–35645, (2007).

[96] B. Polevoda, L. Span, and F. Sherman, **The yeast translation release factors Mrf1p and Sup45p (eRF1) are methylated, respectively, by the methyltransferases Mtq1p and Mtq2p**. *Journal of Biological Chemistry*, vol. 281, no. 5, pp. 2562–2571, (2006).

[97] P. Rathert, A. Dhayalan, M. Murakami, X. Zhang, R. Tamas, R. Jurkowska, Y. Komatsu, Y. Shinkai, X. Cheng, and A. Jeltsch, **Protein lysine methyltransferase G9a acts on non-histone targets**. *Nature chemical biology*, vol. 4, no. 6, pp. 344–346, (2008).

[98] A. Dhayalan, S. Kudithipudi, P. Rathert, and A. Jeltsch, **Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase**. *Chemistry & biology*, vol. 18, no. 1, pp. 111–120, (2011).

[99] C. He, F. Li, J. Zhang, J. Wu, and Y. Shi, **The methyltransferase NSD3 has chromatin-binding motifs, PHD5-C5HCH, that are distinct from other NSD (nuclear receptor SET domain) family members in their histone H3 recognition**. *Journal of Biological Chemistry*, vol. 288, no. 7, pp. 4692–4703, (2013).

[100] M. Morishita and E. di Luccio, **Cancers and the NSD family of histone lysine methyltransferases**. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1816, no. 2, pp. 158 – 163, (2011).

[101] N. Huang, E. vom Baur, J.-M. Garnier, T. Lerouge, J.-L. Vonesch, Y. Lutz, P. Chambon, and R. Losson, **Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators**. *The EMBO Journal*, vol. 17, no. 12, pp. 3398–3412, (1998).

[102] S. Kudithipudi, C. Lungu, P. Rathert, N. Happel, and A. Jeltsch, **Substrate specificity analysis and novel substrates of the protein lysine methyltransferase NSD1**. *Chemistry & biology*, vol. 21, no. 2, pp. 226–237, (2014).

[103] H.-B. Kang, Y. Choi, J. M. Lee, K.-C. Choi, H.-C. Kim, J.-Y. Yoo, Y.-H. Lee, and H.-G. Yoon, **The histone methyltransferase, NSD2, enhances androgen receptor-mediated transcription**. *FEBS Letters*, vol. 583, no. 12, pp. 1880 – 1886, (2009).

[104] J.-Y. Kim, H. J. Kee, N.-W. Choe, S.-M. Kim, G.-H. Eom, H. J. Baek, H. Kook, H. Kook, and S.-B. Seo, **Multiple myeloma-related WHSC1/MMSET isoform RE-IIBP is a histone methyltransferase with transcriptional repression activity**. *Molecular and cellular biology*, vol. 28, no. 6, pp. 2023–2034, (2008).

[105] J. Marango, M. Shimoyama, H. Nishio, J. A. Meyer, D.-J. Min, A. Sirulnik, Y. Martinez-Martinez, M. Chesi, P. L. Bergsagel, M.-M. Zhou, S. Waxman, B. A. Leibovitch, M. J. Walsh, and J. D. Licht, **The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor**. *Blood*, vol. 111, no. 6, pp. 3145–3154, (2008).

[106] H. Pei, L. Zhang, K. Luo, Y. Qin, M. Chesi, F. Fei, P. L. Bergsagel, L. Wang, Z. You, and Z. Lou, **MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites**. *Nature*, vol. 470, no. 7332, pp. 124–128, (2011).

[107] Y. Li, P. Trojer, C.-F. Xu, P. Cheung, A. Kuo, W. J. Drury, Q. Qiao, T. A. Neubert, R.-M. Xu, O. Gozani, and D. Reinberg, **The Target of the NSD Family of Histone Lysine Methyltransferases Depends on the Nature of the Substrate**. *Journal of Biological Chemistry*, vol. 284, no. 49, pp. 34283–34295, (2009).

[108] S. M. Kim, H. J. Kee, G. H. Eom, N. W. Choe, J. Y. Kim, Y. S. Kim, S. K. Kim, H. Kook, H. Kook, and S. B. Seo, **Characterization of a novel WHSC1-associated SET domain protein with H3K4 and H3K27 methyltransferase activity**. *Biochemical and biophysical research communications*, vol. 345, no. 1, pp. 318–323, (2006).

[109] J. Douglas, S. Hanks, I. K. Temple, S. Davies, A. Murray, M. Upadhyaya, S. Tomkins, H. E. Hughes, R. T. Cole, and N. Rahman, **NSD1 mutations are the major cause of Sotos syndrome and occur in some cases of Weaver syndrome but are rare in other overgrowth phenotypes**. *The American Journal of Human Genetics*, vol. 72, no. 1, pp. 132–143, (2003).

[110] N. Kurotaki, N. Harada, O. Shimokawa, N. Miyake, H. Kawame, K. Uetake, Y. Makita, T. Kondoh, T. Ogata, T. Hasegawa, T. Nagai, T. Ozaki, M. Touyama, R. Shenhav, H. Ohashi, L. Medne, T. Shiihara, S. Ohtsu, Z.-i. Kato, N. Okamoto, J. Nishimoto, D. Lev, Y. Miyoshi, S. Ishikiriyama, T. Sonoda, S. Sakazume, Y. Fukushima, K. Kurosawa,

J.-F. Cheng, K.-i. Yoshiura, T. Ohta, T. Kishino, N. Niikawa, and N. Matsumoto, **Fifty microdeletions among 112 cases of Sotos syndrome: Low copy repeats possibly mediate the common deletion**. *Human Mutation*, vol. 22, no. 5, pp. 378–387, (2003).

[111] J. F. Sotos, P. R. Dodge, D. Muirhead, J. D. Crawford, and N. B. Talbot, **Cerebral gigantism in childhood: a syndrome of excessively rapid growth with acromegalic features and a nonprogressive neurologic disorder**. *New England Journal of Medicine*, vol. 271, no. 3, pp. 109–116, (1964).

[112] J. H. Hersh, T. Cole, A. S. Bloom, S. J. Bertolone, and H. E. Hughes, **Risk of malignancy in Sotos syndrome**. *The Journal of pediatrics*, vol. 120, no. 4, pp. 572–574, (1992).

[113] M. M. Cohen, **Tumors and nontumors in Sotos syndrome**. *American journal of medical genetics*, vol. 84, no. 2, pp. 173–175, (1999).

[114] G. Baujat, M. Rio, S. Rossignol, D. Sanlaville, S. Lyonnet, M. Le Merrer, A. Munnich, C. Gicquel, V. Cormier-Daire, and L. Colleaux, **Paradoxical NSD1 mutations in Beckwith-Wiedemann syndrome and 11p15 anomalies in Sotos syndrome**. *The American Journal of Human Genetics*, vol. 74, no. 4, pp. 715–720, (2004).

[115] Q. Zhao, O. L. Caballero, S. Levy, B. J. Stevenson, C. Iseli, S. J. De Souza, P. A. Galante, D. Busam, M. A. Leversha, K. Chadalavada, Y.-H. Rogersa, J. C. Ventera, A. J. G. Simpsonb, and R. L. Strausberg, **Transcriptome-guided characterization of genomic rearrangements in a breast cancer cell line**. *Proceedings of the National Academy of Sciences*, vol. 106, no. 6, pp. 1886–1891, (2009).

[116] M. Berdasco, S. Ropero, F. Setien, M. F. Fraga, P. Lapunzina, R. Losson, M. Alaminos, N.-K. Cheung, N. Rahman, and M. Esteller, **Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma**. *Proceedings of the National Academy of Sciences*, vol. 106, no. 51, pp. 21830–21835, (2009).

[117] F. Zhao, Y. Chen, L. Zeng, R. Li, R. Zeng, L. Wen, Y. Liu, and C. Zhang, **Role of triptolide in cell proliferation, cell cycle arrest, apoptosis and histone methylation in multiple myeloma U266 cells**. *European journal of pharmacology*, vol. 646, no. 1, pp. 1–11, (2010).

[118] R. J. Jaju, C. Fidler, O. A. Haas, A. J. Strickson, F. Watkins, K. Clark, N. C. P. Cross, J.-F. Cheng, P. D. Aplan, L. Kearney, J. Boultwood, and J. S. Wainscoat, **A novel gene, NSD1, is fused to NUP98 in the t (5; 11)(q35; p15. 5) in de novo childhood acute myeloid leukemia**. *Blood*, vol. 98, no. 4, pp. 1264–1267, (2001).

[119] L. H. Kasper, P. K. Brindle, C. A. Schnabel, C. E. Pritchard, M. L. Cleary, and J. M. van Deursen, **CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98-HOXA9 oncogenicity**. *Molecular and cellular biology*, vol. 19, no. 1, pp. 764–776, (1999).

[120] T. Lu, M. W. Jackson, B. Wang, M. Yang, M. R. Chance, M. Miyagi, A. V. Gudkov, and G. R. Stark, **Regulation of NF-κB by NSD1/FBXL11-dependent reversible lysine methylation of p65**. *Proceedings of the National Academy of Sciences*, vol. 107, no. 1, pp. 46–51, (2010).

[121] Cancer Genome Atlas Network, **Comprehensive genomic characterization of head and neck squamous cell carcinomas**. *Nature*, vol. 517, no. 7536, pp. 576–582, (2015).

[122] M. K. Keck, Z. Zuo, A. Khattri, T. P. Stricker, C. D. Brown, M. Imanguli, D. Rieke, K. Endhardt, P. Fang, J. Brägelmann, R. DeBoer, M. El-Dinali, S. Aktolga, Z. Lei, P. Tan, S. G. Rozen, R. Salgia, R. R. Weichselbaum, M. W. Lingen, M. D. Story, K. K. Ang, E. E. Cohen, K. P. White, E. E. Vokes, and T. Y. Seiwert, **Integrative analysis of head and neck cancer identifies two biologically distinct HPV and three non-HPV subtypes**. *Clinical Cancer Research*, vol. 21, no. 4, pp. 870–881, (2015).

[123] R. Rosati, R. La Starza, A. Veronese, A. Aventin, C. Schwienbacher, T. Vallespi, M. Negrini, M. F. Martelli, and C. Mecucci, **NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t (8; 11)(p11. 2; p15)**. *Blood*, vol. 99, no. 10, pp. 3857–3860, (2002).

[124] P.-O. Angrand, F. Apiou, A. F. Stewart, B. Dutrillaux, R. Losson, and P. Chambon, **NSD3, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines**. *Genomics*, vol. 74, no. 1, pp. 79–88, (2001).

[125] Z. Q. Yang, K. L. Streicher, M. E. Ray, J. Abrams, and S. P. Ethier, **Multiple interacting oncogenes on the 8p11-p12 amplicon in human breast cancer**. *Cancer research*, vol. 66, no. 24, pp. 11632–11643, (2006).

[126] D. Kang, H.-S. Cho, G. Toyokawa, M. Kogure, Y. Yamane, Y. Iwai, S. Hayami, T. Tsunoda, H. I. Field, K. Matsuda, D. E. Neal, B. A. J. Ponder, Y. Maehara, Y. Nakamura, and R. Hamamoto, **The histone methyltransferase Wolf-Hirschhorn syndrome candidate 1-like 1 (WHSC1L1) is involved in human carcinogenesis**. *Genes, Chromosomes and Cancer*, vol. 52, no. 2, pp. 126–139, (2013).

[127] Z.-Q. Yang, G. Liu, A. Bollig-Fischer, C. N. Giroux, and S. P. Ethier, **Transforming properties of 8p11-12 amplified genes in human breast cancer**. *Cancer research*, vol. 70, no. 21, pp. 8487–8497, (2010).

[128] K. Nimura, K. Ura, H. Shiratori, M. Ikawa, M. Okabe, R. J. Schwartz, and Y. Kaneda, **A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf–Hirschhorn syndrome**. *Nature*, vol. 460, no. 7252, pp. 287–291, (2009).

[129] U. Wolf, H. Reinwein, R. Porsch, R. Schröter, and H. Baitsch, **Defizienz an den kurzen Armen eines Chromosomes Nr. 4**. *Human Genetics*, vol. 1, no. 5, pp. 397–413, (1965).

[130] K. Hirschhorn, H. L. Cooper, and I. L. Firschein, **Deletion of short arms of chromosome 4–5 in a child with defects of midline fusion**. *Human Genetics*, vol. 1, no. 5, pp. 479–482, (1965).

[131] A. D. Bergemann, F. Cole, and K. Hirschhorn, **The etiology of Wolf–Hirschhorn syndrome**. *Trends in genetics*, vol. 21, no. 3, pp. 188–195, (2005).

[132] G. Toyokawa, H.-S. Cho, K. Masuda, Y. Yamane, M. Yoshimatsu, S. Hayami, M. Takawa, Y. Iwai, Y. Daigo, E. Tsuchiya, T. Tsunoda, H. I. Field, J. D. Kelly, D. E. Neal, Y. Maehara, B. A. Ponder, Y. Nakamura, and R. Hamamoto, **Histone lysine methyltransferase Wolf-Hirschhorn syndrome candidate 1 is involved in human carcinogenesis through regulation of the Wnt pathway**. *Neoplasia*, vol. 13, no. 10, pp. 887–IN11, (2011).

[133] H. R. Hudlebusch, J. Skotte, E. Santoni-Rugiu, Z. G. Zimling, M. J. Lees, R. Simon, G. Sauter, R. Rota, M. A. De Ioris, M. Quarto, J. V. Johansen, M. Jørgensen, C. Rechnitzer, L. L. Maroun, H. Schrøder, B. L. Petersen, and K. Helin, **MMSET Is Highly Expressed and Associated with Aggressiveness in Neuroblastoma**. *Cancer Research*, vol. 71, no. 12, pp. 4226–4235, (2011).

[134] V. Saloura, H.-S. Cho, K. Kiyotani, H. Alachkar, Z. Zuo, M. Nakakido, T. Tsunoda, T. Seiwert, M. Lingen, J. Licht, Y. Nakamura, and R. Hamamoto, **WHSC1 Promotes Oncogenesis through Regulation of NIMA-Related Kinase-7 in Squamous Cell Carcinoma of the Head and Neck**. *Molecular Cancer Research*, vol. 13, no. 2, pp. 293–304, (2015).

[135] M. Xiao, S. Yang, J. Chen, X. Ning, L. Guo, K. Huang, and L. Sui, **Overexpression of MMSET in endometrial cancer: a clinicopathologic study**. *Journal of surgical oncology*, vol. 107, no. 4, pp. 428–432, (2013).

[136] P. Zhou, L.-L. Wu, K.-M. Wu, W. Jiang, J.-d. Li, L.-d. Zhou, X.-Y. Li, S. Chang, Y. Huang, H. Tan, G.-W. Zhang, F. He, and Z.-M. Wang, **Overexpression of MMSET is Correlation with Poor Prognosis in Hepatocellular Carcinoma**. *Pathology & Oncology Research*, vol. 19, no. 2, pp. 303–309, (2012).

[137] M. Chesi, E. Nardini, R. S. Lim, K. D. Smith, W. M. Kuehl, and P. L. Bergsagel, **The t (4; 14) Translocation in Myeloma Dysregulates Both FGFR3and a Novel Gene, MMSET, Resulting in IgH/MMSET Hybrid Transcripts**. *Blood*, vol. 92, no. 9, pp. 3025–3034, (1998).

[138] M. Chesi, E. Nardini, L. A. Brents, E. Schröck, T. Ried, W. M. Kuehl, and P. L. Bergsagel, **Frequent translocation t (4; 14)(p16. 3; q32. 3) in multiple myeloma is associated with increased expression and activating mutations of fibroblast growth factor receptor 3**. *Nature genetics*, vol. 16, no. 3, p. 260, (1997).

[139] J. J. Keats, T. Reiman, C. A. Maxwell, B. J. Taylor, L. M. Larratt, M. J. Mant, A. R. Belch, and L. M. Pilarski, **In multiple myeloma, t (4; 14)(p16; q32) is an adverse prognostic factor irrespective of FGFR3 expression**. *Blood*, vol. 101, no. 4, pp. 1520–1529, (2003).

[140] M. Santra, F. Zhan, E. Tian, B. Barlogie, and J. Shaughnessy, **A subset of multiple myeloma harboring the t (4; 14)(p16; q32) translocation lacks FGFR3 expression but maintains an IGH/MMSET fusion transcript**. *Blood*, vol. 101, no. 6, pp. 2374–2376, (2003).

[141] A. Kuo, P. Cheung, K. Chen, B. Zee, M. Kioi, J. Lauring, Y. Xi, B. Park, X. Shi, B. Garcia, W. Li, and O. Gozani, **NSD2 Links Dimethylation of Histone H3 at Lysine 36 to Oncogenic Programming**. *Molecular Cell*, vol. 44, no. 4, pp. 609 – 620, (2011).

[142] E. Martinez-Garcia, R. Popovic, D.-J. Min, S. M. M. Sweet, P. M. Thomas, L. Zamdborg, A. Heffner, C. Will, L. Lamy, L. M. Staudt, D. L. Levens, N. L. Kelleher, and J. D. Licht, **The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t (4; 14) multiple myeloma cells**. *Blood*, vol. 117, no. 1, pp. 211–220, (2011).

[143] J. D. Jaffe, Y. Wang, H. M. Chan, J. Zhang, R. Huether, G. V. Kryukov, H. C. Bhang, J. E. Taylor, M. Hu, N. P. Englund, F. Yan, Z. Wang, E. R. McDonald III, L. Wei, J. Ma, J. Easton, Z. Yu, R. deBeaumount, V. Gibaja, K. Venkatesan, R. Schlegel, W. R. Sellers, N. Keen, J. Liu, G. Caponigro, J. Barretina, V. G. Cooke, C. Mullighan, S. A. Carr, J. R. Downing, L. A. Garraway, and F. T. Stegmeier, **Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia**. *Nature genetics*, vol. 45, no. 11, pp. 1386–1391, (2013).

[144] J. A. Oyer, X. Huang, Y. Zheng, J. Shim, T. Ezponda, Z. Carpenter, M. Allegretta, C. I. Okot-Kotber, J. P. Patel, A. Melnick, R. L. Levine, A. Ferrando, A. D. MacKerell Jr., N. L. Kelleher, J. D. Licht, and R. Popovic, **Point mutation E1099K in MMSET/NSD2 enhances its methyltranferase activity and leads to altered global chromatin methylation in lymphoid malignancies**. *Leukemia*, vol. 28, no. 1, pp. 198–201, (2014).

[145] T. Ezponda, R. Popovic, M. Y. Shah, E. Martinez-Garcia, Y. Zheng, D.-J. Min, C. Will, A. Neri, N. L. Kelleher, J. Yu, and J. D. Licht, **The histone methyltransferase MMSET/WHSC1 activates TWIST1 to promote an epithelial–mesenchymal transition and invasive properties of prostate cancer**. *Oncogene*, vol. 32, no. 23, pp. 2882–2890, (2013).

[146] P. Yang, L. Guo, Z. J. Duan, C. G. Tepper, L. Xue, X. Chen, H.-J. Kung, A. C. Gao, J. X. Zou, and H.-W. Chen, **Histone Methyltransferase NSD2/MMSET Mediates Constitutive NF-κB Signaling for Cancer Cell Proliferation, Survival, and Tumor Growth via a Feed-Forward Loop**. *Molecular and Cellular Biology*, vol. 32, no. 15, pp. 3121–3131, (2012).

[147] P. Rathert, X. Zhang, C. Freund, X. Cheng, and A. Jeltsch, **Analysis of the substrate specificity of the Dim-5 histone lysine methyltransferase using peptide arrays**. *Chemistry & biology*, vol. 15, no. 1, pp. 5–11, (2008).

[148] B. Tschiersch, A. Hofmann, V. Krauss, R. Dorn, G. Korge, and G. Reuter, **The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes.**. *The EMBO journal*, vol. 13, no. 16, p. 3822, (1994).

[149] D. O'Carroll, H. Scherthan, A. H. Peters, S. Opravil, A. R. Haynes, G. Laible, S. Rea, M. Schmid, A. Lebersorger, M. Jerratsch, L. Sattler, M. G. Mattei, P. Denny, S. D. M. Brown, D. Schweizer, and T. Jenuwein, **Isolation and Characterization of Suv39h2, a Second Histone H3 Methyltransferase Gene That Displays Testis-Specific Expression**. *Molecular and cellular biology*, vol. 20, no. 24, pp. 9423–9433, (2000).

[150] S. Rea, F. Eisenhaber, D. O'Carroll, B. D. Strahl, Z.-W. Sun, M. Schmid, S. Opravil, K. Mechtler, C. P. Ponting, C. D. Allis, and T. Jenuwein, **Regulation of chromatin structure by site-specific histone H3 methyltransferases**. *Nature*, vol. 406, no. 6796, pp. 593–599, (2000).

[151] H. G. Chin, D. Patnaik, P.-O. Esteve, S. E. Jacobsen, and S. Pradhan, **Catalytic properties and kinetic mechanism of human recombinant Lys-9 histone H3 methyltransferase SUV39H1: participation of the chromodomain in enzymatic catalysis**. *Biochemistry*, vol. 45, no. 10, pp. 3272–3284, (2006).

[152] R. Eskeland, B. Czermin, J. Boeke, T. Bonaldi, J. T. Regula, and A. Imhof, **The N-terminus of Drosophila SU(VAR) 3-9 mediates dimerization and regulates its methyltransferase activity**. *Biochemistry*, vol. 43, no. 12, pp. 3740–3749, (2004).

[153] A. H. Peters, D. O'Carroll, H. Scherthan, K. Mechtler, S. Sauer, C. Schöfer, K. Weipoltshammer, M. Pagani, M. Lachner, A. Kohlmaier, S. Opravil, M. Doyle, M. Sibilia, and T. Jenuwein, **Loss of the *Suv39h* Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability**. *Cell*, vol. 107, no. 3, pp. 323 – 337, (2001).

[154] C. Dong, Y. Wu, Y. Wang, C. Wang, T. Kang, P. G. Rychahou, Y.-I. Chi, B. M. Evers, and B. P. Zhou, **Interaction with Suv39H1 is critical for Snail-mediated E-cadherin repression in breast cancer**. *Oncogene*, vol. 32, no. 11, pp. 1351–1362, (2013).

[155] S. Goyama, E. Nitta, T. Yoshino, S. Kako, N. Watanabe-Okochi, M. Shimabe, Y. Imai, K. Takahashi, and M. Kurokawa, **EVI-1 interacts with histone methyltransferases SUV39H1 and G9a for transcriptional repression and bone marrow immortalization**. *Leukemia*, vol. 24, no. 1, pp. 81–88, (2010).

[156] H. Chaib, A. Nebbioso, T. Prebet, R. Castellano, S. Garbit, A. Restouin, N. Vey, L. Altucci, and Y. Collette, **Anti-leukemia activity of chaetocin via death receptor-dependent apoptosis and dual modulation of the histone methyl-transferase SUV39H1**. *Leukemia*, vol. 26, no. 4, pp. 662–674, (2012).

[157] K. Zhang, K. Mosch, W. Fischle, and S. I. Grewal, **Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin**. *Nature structural & molecular biology*, vol. 15, no. 4, pp. 381–388, (2008).

[158] K. Yamamoto and M. Sonoda, **Self-interaction of heterochromatin protein 1 is required for direct binding to histone methyltransferase, SUV39H1**. *Biochemical and biophysical research communications*, vol. 301, no. 2, pp. 287–292, (2003).

[159] S. Haldar, A. Saini, J. S. Nanda, S. Saini, and J. Singh, **Role of Swi6/HP1 self-association-mediated recruitment of Clr4/Suv39 in establishment and maintenance of heterochromatin in fission yeast**. *Journal of Biological Chemistry*, vol. 286, no. 11, pp. 9308–9320, (2011).

[160] B. Al-Sady, H. D. Madhani, and G. J. Narlikar, **Division of labor between the chromodomains of HP1 and Suv39 methylase enables coordination of heterochromatin spread**. *Molecular cell*, vol. 51, no. 1, pp. 80–91, (2013).

[161] K. Zhang, T. Fischer, R. L. Porter, J. Dhakshnamoorthy, M. Zofall, M. Zhou, T. Veenstra, and S. I. Grewal, **Clr4/Suv39 and RNA quality control factors cooperate to trigger RNAi and suppress antisense RNA**. *Science*, vol. 331, no. 6024, pp. 1624–1627, (2011).

[162] J. C. Obenauer, L. C. Cantley, and M. B. Yaffe, **Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs**. *Nucleic acids research*, vol. 31, no. 13, pp. 3635–3641, (2003).

[163] Q. M. Raafiq, **Investigation of the specificity of protein lysine methyltransferases**. PhD thesis, Jacobs Univeristy, (2013).

[164] J. J. Keats, C. A. Maxwell, B. J. Taylor, M. J. Hendzel, M. Chesi, P. L. Bergsagel, L. M. Larratt, M. J. Mant, T. Reiman, A. R. Belch, and L. M. Pilarski, **Overexpression of transcripts originating from the MMSET locus characterizes all t (4; 14)(p16; q32)-positive multiple myeloma patients**. *Blood*, vol. 105, no. 10, pp. 4060–4069, (2005).

[165] S. Kudithipudi, **Identifying Novel Substrates by Specificity Profile Analysis of Protein Lysine Methyltransferases**. PhD thesis, Jacobs Univeristy, (2011).

[166] J. Schwartzentruber, A. Korshunov, X.-Y. Liu, D. T. Jones, E. Pfaff, K. Jacob, D. Sturm, A. M. Fontebasso, D.-A. K. Quang, M. Tönjes, V. Hovestadt, S. Albrecht, M. Kool, A. Nantel, C. Konermann, A. Lindroth, N. Jager, T. Rausch, M. Ryzhova, J. O. Korbel, T. Hielscher, P. Hauser, M. Garami, A. Klekner, L. Bognar, M. Ebinger, M. U. Schuhmann, W. Scheurlen, A. Pekrun, M. C. Fruhwald, W. Roggendorf, C. Kramm, M. Durken, J. Atkinson, P. Lepage, A. Montpetit, M. Zakrzewska, K. Zakrzewski, P. P. Liberski, Z. Dong, P. Siegel, A. E. Kulozik, M. Zapatka, A. Guha, D. Malkin, J. Felsberg, G. Reifenberger, A. von Deimling, K. Ichimura, V. P. Collins, H. Witt, T. Milde, O. Witt, C. Zhang, P. Castelo-Branco, P. Lichter, D. Faury, U. Tabori, C. Plass, J. Majewski, S. M. Pfister, and N. Jabado, **Driver mutations in histone H3. 3 and chromatin remodelling genes in paediatric glioblastoma**. *Nature*, vol. 482, no. 7384, pp. 226–231, (2012).

[167] G. Wu, A. Broniscer, T. A. McEachron, C. Lu, B. S. Paugh, J. Becksfort, C. Qu, L. Ding, R. Huether, M. Parker, J. Zhang, A. Gajjar, M. A. Dyer, C. G. Mullighan, R. J. Gilbertson, E. R. Mardis, R. K. Wilson, J. R. Downing, D. W. Ellison, J. Zhang, and S. J. Baker, **Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas**. *Nature genetics*, vol. 44, no. 3, pp. 251–253, (2012).

[168] S. Behjati, P. S. Tarpey, N. Presneau, S. Scheipl, N. Pillay, P. Van Loo, D. C. Wedge, S. L. Cooke, G. Gundem, H. Davies, S. Nik-Zainal, S. Martin, S. McLaren, V. Goody, B. Robinson, A. Butler, J. W. Teague, D. Halai, B. Khatri, O. Myklebost, D. Baumhoer, G. Jundt, R. Hamoudi, R. Tirabosco, M. F. Amary, P. A. Futreal, M. R. Stratton, P. J. Campbell, and A. M. Flanagan, **Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone**. *Nature genetics*, vol. 45, no. 12, pp. 1479–1482, (2013).

[169] P. W. Lewis, M. M. Müller, M. S. Koletsky, F. Cordero, S. Lin, L. A. Banaszynski, B. A. Garcia, T. W. Muir, O. J. Becher, and C. D. Allis, **Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma**. *Science*, vol. 340, no. 6134, pp. 857–861, (2013).

[170] C. L. Hendricks, J. R. Ross, E. Pichersky, J. P. Noel, and Z. S. Zhou, **An enzyme-coupled colorimetric assay for S-adenosylmethionine-dependent methyltransferases**. *Analytical biochemistry*, vol. 326, no. 1, pp. 100–105, (2004).

[171] E. Collazo, J.-F. Couture, S. Bulfer, and R. C. Trievel, **A coupled fluorescent assay for histone methyltransferases**. *Analytical biochemistry*, vol. 342, no. 1, pp. 86–92, (2005).

[172] K. M. Dorgan, W. L. Wooderchak, D. P. Wynn, E. L. Karschner, J. F. Alfaro, Y. Cui, Z. S. Zhou, and J. M. Hevel, **An enzyme-coupled continuous spectrophotometric assay for S-adenosylmethionine-dependent methyltransferases**. *Analytical biochemistry*, vol. 350, no. 2, pp. 249–255, (2006).

[173] D. Greiner, T. Bonaldi, R. Eskeland, E. Roemer, and A. Imhof, **Identification of a specific inhibitor of the histone methyltransferase SU (VAR) 3-9**. *Nature chemical biology*, vol. 1, no. 3, pp. 143–145, (2005).

[174] H. Gowher, X. Zhang, X. Cheng, and A. Jeltsch, **Avidin plate assay system for enzymatic characterization of a histone lysine methyltransferase**. *Analytical biochemistry*, vol. 342, no. 2, pp. 287–291, (2005).

[175] I. Bock, S. Kudithipudi, R. Tamas, G. Kungulovski, A. Dhayalan, and A. Jeltsch, **Application of Celluspots peptide arrays for the analysis of the binding specificity of epigenetic reading domains to modified histone tails**. *BMC biochemistry*, vol. 12, no. 1, p. 48, (2011).

[176] R. M. Hughes, K. R. Wiggins, S. Khorasanizadeh, and M. L. Waters, **Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect**. *Proceedings of the National Academy of Sciences*, vol. 104, no. 27, pp. 11184–11188, (2007).

[177] W. Fischle, Y. Wang, S. A. Jacobs, Y. Kim, C. D. Allis, and S. Khorasanizadeh, **Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains**. *Genes & development*, vol. 17, no. 15, pp. 1870–1881, (2003).

[178] J.-H. Zhang, T. D. Chung, and K. R. Oldenburg, **A simple statistical parameter for use in evaluation and validation of high throughput screening assays**. *Journal of biomolecular screening*, vol. 4, no. 2, pp. 67–73, (1999).

[179] S. Kudithipudi, D. Kusevic, S. Weirich, and A. Jeltsch, **Specificity analysis of protein lysine methyltransferases using SPOT peptide arrays**. *JoVE (Journal of Visualized Experiments)*, no. 93, pp. e52203–e52203, (2014).

[180] S. Kudithipudi and A. Jeltsch, **Approaches and Guidelines for the Identification of Novel Substrates of Protein Lysine Methyltransferases**. *Cell Chemical Biology*, vol. 23, no. 9, pp. 1049–1055, (2016).

[181] S. Trobro and J. Åqvist, **A model for how ribosomal release factors induce peptidyl-tRNA cleavage in termination of protein synthesis**. *Molecular cell*, vol. 27, no. 5, pp. 758–766, (2007).

[182] R. R. Y. Wong, L. K. Y. Chan, T. P. T. Tsang, C. W. S. Lee, T. H. Cheung, S. F. Yim, N. S. S. Siu, S. N. C. Lee, M. Y. Yu, S. S. C. Chim, Y. F. Wong, and T. K. H. Chung, **CHD5 downregulation associated with poor prognosis in epithelial ovarian cancer**. *Gynecologic and obstetric investigation*, vol. 72, no. 3, pp. 203–207, (2011).

[183] X. Wu, Z. Zhu, W. Li, X. Fu, D. Su, L. Fu, Z. Zhang, A. Luo, X. Sun, L. Fu, and J.-T. Dong, **Chromodomain helicase DNA binding protein 5 plays a tumor suppressor role in human breast cancer**. *Breast Cancer Research*, vol. 14, no. 3, p. 1, (2012).

[184] V. Kolla, T. Zhuang, M. Higashi, K. Naraparaju, and G. M. Brodeur, **Role of CHD5 in human cancers: 10 years later**. *Cancer research*, vol. 74, no. 3, pp. 652–658, (2014).

[185] C. A. French, I. Miyoshi, I. Kubonishi, H. E. Grier, A. R. Perez-Atayde, and J. A. Fletcher, **BRD4-NUT fusion oncogene a novel mechanism in aggressive carcinoma**. *Cancer research*, vol. 63, no. 2, pp. 304–307, (2003).

[186] C. French, C. Ramirez, J. Kolmakova, T. Hickman, M. Cameron, M. Thyne, J. Kutok, J. Toretsky, A. Tadavarthy, U. Kees, J. A. Fletcher, and J. C. Aster, **BRD–NUT oncoproteins: a family of closely related nuclear proteins that block epithelial differentiation and maintain the growth of carcinoma cells**. *Oncogene*, vol. 27, no. 15, pp. 2237–2242, (2007).

[187] R. J. Gibbons, D. J. Picketts, L. Villard, and D. R. Higgs, **Mutations in a putative global transcriptional regulator cause X-linked mental retardation with α-thalassemia (ATR-X syndrome)**. *Cell*, vol. 80, no. 6, pp. 837–845, (1995).

[188] P. Drané, K. Ouararhni, A. Depaux, M. Shuaib, and A. Hamiche, **The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3. 3**. *Genes & development*, vol. 24, no. 12, pp. 1253–1265, (2010).

[189] M. S. Lechner, D. C. Schultz, D. Negorev, G. G. Maul, and F. J. Rauscher, **The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain**. *Biochemical and biophysical research communications*, vol. 331, no. 4, pp. 929–937, (2005).

[190] C. Cardoso, S. Timsit, L. Villard, M. Khrestchatisky, M. Fontès, and L. Colleaux, **Specific Interaction between the XNP ATR-X Gene Product and the SET Domain of the Human EZH2 Protein**. *Human molecular genetics*, vol. 7, no. 4, pp. 679–684, (1998).

[191] X. Nan, J. Hou, A. Maclean, J. Nasir, M. J. Lafuente, X. Shu, S. Kriaucionis, and A. Bird, **Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation**. *Proceedings of the National Academy of Sciences*, vol. 104, no. 8, pp. 2709–2714, (2007).

[192] I. Garcia-Higuera, T. Taniguchi, S. Ganesan, M. S. Meyn, C. Timmers, J. Hejna, M. Grompe, and A. D. D'Andrea, **Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway**. *Molecular cell*, vol. 7, no. 2, pp. 249–262, (2001).

[193] Z. Yan, M. Delannoy, C. Ling, D. Daee, F. Osman, P. A. Muniandy, X. Shen, A. B. Oostra, H. Du, J. Steltenpool, T. Lin, B. Schuster, C. Decaillet, A. Stasiak, A. Z. Stasiak, S. Stone, M. E. Hoatlin, D. Schindler, C. L. Woodcock, H. Joenje, R. Sen, J. P. de Winter, L. Li, M. M. Seidman, M. C. Whitby, K. Myung, A. Constantinou, and W. Wang, **A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability**. *Molecular cell*, vol. 37, no. 6, pp. 865–878, (2010).

[194] Z. Huang, H. Wu, S. Chuai, F. Xu, F. Yan, N. Englund, Z. Wang, H. Zhang, M. Fang, Y. Wang, J. Gu, M. Zhang, T. Yang, K. Zhao, Y. Yu, J. Dai, W. Yi, S. Zhou, Q. Li, J. Wu, J. Liu, X. Wu, H. Chan, C. Lu, P. Atadja, E. Li, Y. Wang, and M. Hu, **NSD2 is recruited through its PHD domain to oncogenic gene loci to drive multiple myeloma**. *Cancer research*, vol. 73, no. 20, pp. 6277–6288, (2013).

[195] S. Yang, X. Zheng, C. Lu, G.-M. Li, C. D. Allis, and H. Li, **Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase**. *Genes & Development*, vol. 30, no. 14, pp. 1611–1616, (2016).

[196] A. Jeltsch and T. Lanio, **Site-directed mutagenesis by polymerase chain reaction**. *In Vitro Mutagenesis Protocols*, vol. 182, pp. 85–94, (2002).

[197] R. Frank, **The SPOT-synthesis technique: synthetic peptide arrays on membrane supports-principles and applications**. *Journal of immunological methods*, vol. 267, no. 1, pp. 13–26, (2002).