Identifying Novel Substrates by Specificity Profile Analysis of Protein Lysine Methyltransferases

by

Srikanth Kudithipudi

A Thesis submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy in Biochemistry

Approved, Thesis Committee

Prof. Dr. Albert Jeltsch
(Prof. of Biochemistry, Jacobs University
Bremen, Germany)

Prof. Dr. Mike Schutkowski
(Department of Enzymology, Martin- Luther
Universitat Halle-Wittenberg, Germany)

Prof. Dr. Sebastian Springer
(Associate Prof. of Biochemistry andCellBiology,
Jacobs University Bremen, Germany)

Date of Defense: August 31, 2011

School of Engineering and Science

Jacobs University Bremen, Germany

Acknowledgements

This thesis would not have been possible without the guidance and support of several people who directly and indirectly contributed their valuable assistance in the preparation of this study.

First and foremost, I would like to thank my supervisor Prof. Dr. Albert Jeltsch for providing me the opportunity to work in his laboratory and for being a great mentor. His commitment, joy and enthusiasm for research was always been a great motivation to me. I appreciate all his knowledge, contribution of ideas and time for my thesis work. I am indebted to him more than he knows.

I would like to thank Prof. Dr. Mike Schutkowski and Prof. Dr. Sebastian Springer for being co-referees of my PhD thesis.

I express my sincere gratitude to Dr. Arun Kumar Dhayalan and Dr. Philip Rathert for introducing me to all the techniques necessary for my thesis work and for extending their unconditional support even after leaving Prof. Jeltsch's lab.

I would like to thank Adam F. Kebede and Cristiana Lungu for their valuable contributions to my work.

I am also grateful to the following colleagues whose suggestions and discussions were invaluable to my work; Arumugam rajavelu, Ina Bock, Dr. Sergey Ragozine, Raluca Tamas, Qazi Raafiq.

Many thanks to my co-workers Martin, Renata and Tomek Jurkowski, Pavel, Razvan for creating friendly atmosphere in the lab. And special thanks to Sandra Becker for patiently organising the lab and for supporting us with all the ordering jobs.

My time at Jacobs was made more enjoyable and memorable due to my friends, I am grateful to my flatmate Binit Lukose with whom I enjoyed talking about everything and seeking suggestions. Special thanks to Arumugam, Mahendran, Rajesh, Mahesh and Rami for supporting me always with their words and actions. I also wish to thank Sunitha, Raghu, Saini, Tripti for hosting weekend get together parties with delicious Indian food and for giving me the feeling of home.

I would like to thank Hans-Dieter and Ingvield for the great time I had with them and who always reminds me about life out side the lab.

Finally, I want to express my appreciation to my parents Ravindra Mohan Rao, Balamani and my brothers Raghu Ram, Raghavendra for their patience, understanding, love and encouragement they have been providing. And last but not least, I thank my fiancee Neelima for being very understanding and supportive during the writing part of my thesis.

Table of contents

1. Abstract

In the cell nucleus the DNA binds to histone proteins and forms a compact structure called chromatin. Both the components of chromatin are subjected to several post-translational modifications which regulate the gene expression. Enzymes (histone acetyltransferases and histone lysine methyltransferases) known to methylate histone protein have also been shown to act on non-histone proteins and methylation and acetylation of non-histone proteins carries many important biological signals, but not many non-histone methylation substrates of protein lysine methyltransferases are known. In this study we have characterised the substrate specificity of histone lysine methyltransferases and based on the specificity data, we identified several novel histone and non-histone substrates.

The NSD1 enzyme is a histone lysine methyltransferase enzyme. Mutations of this protein cause the Sotos syndrome. We studied the substrate specificity of NSD1 using the H3 (30-50) sequence as a template. With the obtained consensus sequence motif we identified several novel histone and non-histone NSD1 substrates. We showed that NSD1 could not methylate H4K20, instead it methylates K44 in H4 protein, which is in agreement with our specificity profile. For the first time we showed NSD1 methylates H1 proteins in a variant specific manner; NSD1 methylates K168 in H1.2, H1.3 and H1.5 proteins but not in H1.4. Apart from the novel histone substrates, we also identified several non-histone proteins containing the NSD1 consensus sequence motif and confirmed methylation of 45 novel non-histone peptides and of the (ATRX and Probable U3 Small Nucleolar RNA-associated Protein) proteins. Based on the candidate screening approach, we also identified an automethylation site in NSD1 and confirmed the loss of methylation signal with the corresponding predicted lysine; NSD1-K1769R mutant protein. We also show that the NSD1 Sotos SET domain mutants impair its methyltransferase activity and thus establish a possible deregulation of signalling networks in Sotos patients.

SUV39H1 is a H3K9 methyltransferase enzyme which plays a vital role in the formation of heterochromatin. We derived the specificity profile of this enzyme and showed that it mainly recognises an ‘RK’ motif corresponding to R^8 and K^9 in the H3 tail. In addition, lysine 4 of the H3 tail is very important for substrate recognition. With the derived specificity profile of SUV39H1 we identified several novel non-histone peptide substrates and confirmed methylation of RAG2, SET8, Jumonji and Sex comb on midleg protein 2 proteins at the protein level, albeit methylation on Jumonji and Sex comb on midleg protein 2 were weak. Similar to

the K4 recognition on H3 tail, we have also observed lysine at -5 position with respect to the target lysine is important for SUV39H1 to methylate the newly identified targets RAG2 and SET8. We have shown that methylation of RAG2 alters its sub-nuclear localization and found that the JMJD2A tandem tudor domain interacts with the newly identified targets in a methyl specific manner.

SET8 is a H4K20 specific mono-methyl transferase which acts preferentially on H4 integrated into nucleosomes. By employing peptide arrays we have shown that it has long recognition sequence motif covering 7 amino acids ($R^{17}H^{18}R^{19}K^{20}V^{21}L^{22}R^{23}$). Based on the derived specificity profile, we identified only 4 potential non-histone substrate proteins. But after relaxing the specificity profile we identified several proteins and showed methylation of 22 non-histone peptides. However, apart from p53 and H4 proteins, none of the identified targets were methylated at the protein level. Celluspot analysis revealed that symmetric and asymmetric methylation on R^{17} of H4 tail further inhibits methylation on H4K20, while other modifications on K^{16} and R^{19} affected H4K20 methylation partially. In summary, our specificity analysis results and methylation assays demonstrate that SET8 as a highly specific histone H4 methyltransferase enzyme.

The SMYD family of protein methyltransferases is a group of enzymes which are unique for having a characteristic MYND domain inserted into the catalytic SET domain. The SMYD proteins have roles in the regulation of the cell cycle and important development pathways such as heart and muscle differentiation. A member of this family, SMYD2, is an uncharacterised histone lysine methyltransferase enzyme, which has been shown to methylate both H3K4 and H3K36. In addition, it was also found to methylate one non-histone substrate (p53) and, thereby, repress its activity. Here we applied peptide arrays and derive a specificity profile for SMYD2 via two approaches: a “best target” approach using a p53 peptide as template and an unbiased random approach. Results revealed that SMYD2 possesses a strong preference for a ‘LK’ or ‘FK’ motif. With the derived sequence motif, we have identified 40 novel peptide substrates from human proteins and for 8 proteins we showed methylation at the protein level and confirmed the predicted target lysine by mutagenesis. Experiments to show cellular methylation and to understand the possible downstream consequences of methylation of some of the identified non-histone proteins are in progress.

2. List of publications

1) Arunkumar Dhayalan, **Srikanth Kudithipudi**, Philipp Rathert and Albert Jeltsch.
Specificity Analysis-Based Identification of New Methylation Targets of the SET7/9 Protein Lysine Methyltransferase.
Chemistry and Biology. 18: 111-120

2) Ina Bock, Arunkumar Dhayalan, **Srikanth Kudithipudi**, Ole Brandt, Philipp Rathert and Albert Jeltsch.
Detailed specificity analysis of antibodies binding to modified histone tails with peptide arrays
Epigenetics. 6(2): 256-263.

3) Arunkumar Dhayalan, Raluca Tamas, Ina Bock, Anna Tattermusch, Emilia Dimitrova, **Srikanth Kudithipudi**, Sergey Ragozin and Albert Jeltsch.
The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9.
Human Molecular Genetics. 20(11): 2195-2203

4) Ina Bock, **Srikanth Kudithipudi**, Raluca Tamas, Goran Kungolovski, Arunkumar Dhayalan, and Albert Jeltsch.
Application of celluspot peptide arrays for the analysis of the binding specificity of epigenetic reading domains to modified histone tails
Submitted to BMC Biochemistry

3. Introduction

“Epigenetics is the study of mitotically and or meiotically heritable changes in the gene function that cannot be explained by alteration in the DNA sequence” (Feil 2008). The term epigenetics was first coined by C.H. Waddington in the year 1940. He derived it from the word epigenesis, the theory which proposed that the adult form developed by successive differentiation form the embryo, as opposed to being fully formed in the zygote. (Holliday, 1994 and Bonasio et al., 2010).

Though the cells in a multicellular organism carry the same genetic information they develop different terminal phenotypes, which suggest that the genes are differently regulated in different cells at appropriate time during development. The term epigenetics is used to classify those process that ensure the inheritance of variations above and beyond the changes in the DNA sequence. To put it in more simple terms, if the genetic code is the hardware of life, the epigenetic code is software that determines how the hardware behaves-and as such it can be rewritten (Brower, 2011). The three fundamental criteria to call it as an epigenetic mechanism are that it should be a heritable, self-perpetuating and reversible process (Bonasio et al., 2010). Epigenetic phenomena include DNA methylation, post-translational modification of histone proteins and small RNA molecules. With the recent findings and information from various studies, the scientific community appreciated the epigenetic system as an important contributor to process from development to metabolism to oncogenesis (Kaufman et al., 2010)

3.1 Chromatin

Eukaryotic genomes are organised into a nucleoprotein complex known as chromatin. The fundamental unit of chromatin is the nucleosome which consists of 146 base pairs of DNA wrapped around histone octamer fomed by 4 histone proteins, an H3-H4 tetramer is assembled with 2 H2A-H2B dimers (figure1) (Jenuwein et al., 2001). The individual nucleosomes pack against each other with the help of H1 proteins and attain a higher order chromatin structure which regulates the accessibility of chromatin for transcription factors.

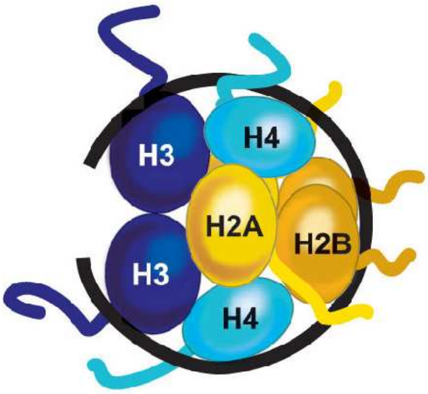


Figure1: Schematic representation of histone organization within the octamer core around which the DNA is wrapped (figure copied from Allis et al., 2007).

The chromatin structure is highly conserved from yeast to humans but mammalian chromatin appears to be more complex than that of lower organisms mainly due to several additional histone modifications and additional histone isoforms (Rando et al., 2009). Histone proteins are small proteins (11-17 kDa) with highly basic charge (either basic proteins or positive charge) that have high affinity for negatively charged DNA. Histone proteins constitute globular domains which are mainly responsible for the nucleosome core formation and unstructured N-terminal tails which are subjected to several post-translational modifications. Covalent post-translational modifications of histones include phosphorylation (of S and T residues), acetylation (K), methylation (K and R), Ubiquitination (K), and Sumoylation (K). These modifications alter the structure and function of chromatin by modifying the interactions between these proteins and DNA and also by recruiting other proteins which are specific to the corresponding mark (Rivera et al., 2010) (Margueron et al., 2005). Although functional consequences of most of the modifications are yet to be discovered, phosphorylation, acetylation and methylation are well studied histone post translational modifications.

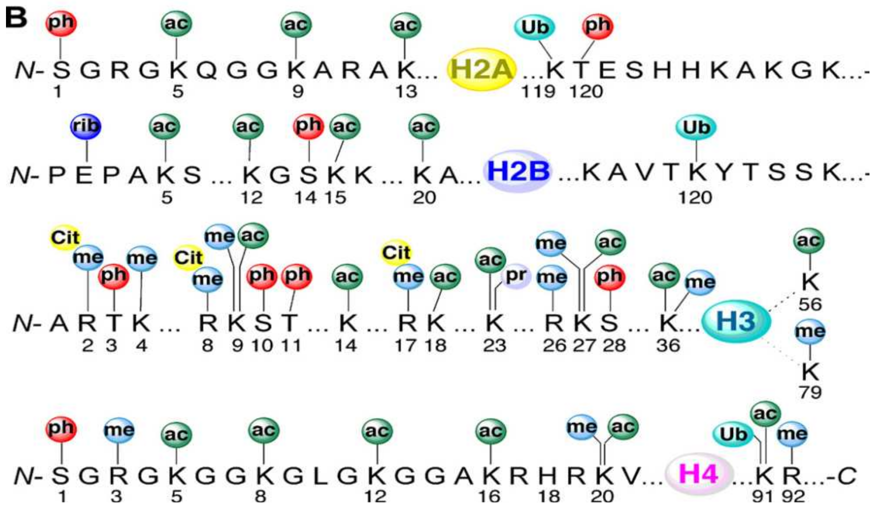

B
ph ac ac ac Ub ph
N- S G R G K Q G G K A R A K... H2A ...K T E S H H K A K G K...-
1 5 9 13 119 120
rib ac ac ph ac ac Ub
N- P E P A K S ... K G S K K ... K A... H2B ...K A V T K Y T S S K...-
5 12 14 15 20 120
Cit me ph me Cit me ac me ph ph ac Cit me ac ac pr me me ac ph ac me ac me
N- A R T K ... R K S T ... K ... R K ... K ... R K S ... K... H3
2 3 4 8 9 10 11 14 17 18 23 26 27 28 36
K 56
K 79
ph me ac ac ac ac me ac Ub ac me
N- S G R G K G G K G L G K G G A K R H R K V... H4 ...K R...-C
1 3 5 8 12 16 18 20 91 92

Figure 2: Schematic representation of histone tails and their post-translational modifications. Groups are indicated ad follows; ac is acetylation, Cit is citrullyl, me is methyl, ph is phosphoryl, pr is propionyl, rib is ADP ribosyl and Ub is Ubiquityl (figure adopted from Bhaumik et al., 2007)

3.2 Histone Acetylation

Histone acetylation is an extensively studied epigenetic mark of histone proteins. Histone acetylation was first discovered by Allfrey et al., in 1964 (Kimura et al., 2005). Histone acetylation is catalysed by a class of enzymes known as histone acetyl transferases (HATs), which catalyse the transfer of acetyl group from acetyl coenzyme A to the ε-amino group of lysine residues. HATs in cells mainly operate as multimeric complexes, these complexes are typically more active than the individual catalytic subunits (Verdone et al., 2006). Histone acetylation neutralises the positive charge on the lysine residues and results in decrease of electrostatic interaction between DNA and histones and thus alters the chromatin structure, which facilitates the interaction of transcription machinery to DNA. Histone acetylation also provides a signal for protein binding, acetylation on lysine residues creates a docking site for protein modules known as bromodomains and few chromodomain containing proteins.

Bromodomains has been the first identified reader proteins which could specifically identify the covalent modification on histone tails. Bromodomains are the major acetyl specific readers, these domains were found to be present in transcription and chromatin regulator proteins which explicitly hint their role in the involvement of regulating the chromatin structure and transcription.

Histone acetylation is rapid and reversible, the turnover of histone acetyaltion is as short as few minutes. Enzymes that counteract histone acetyltransferases are histone deacetylases, these enzymes also mainly present in multi-subunit complexes which are known as histone deacetylase complexes (HDAC). Majorly transcriptional activation is correlated with histone acetylation and transcriptional repression with histone deacetylation, but with the recent findings lysine acetylation emerges as key regulator in different cellular process like DNA repair and cell cycle progression (Verdone et al., 2006).

3.3 Histone methylation

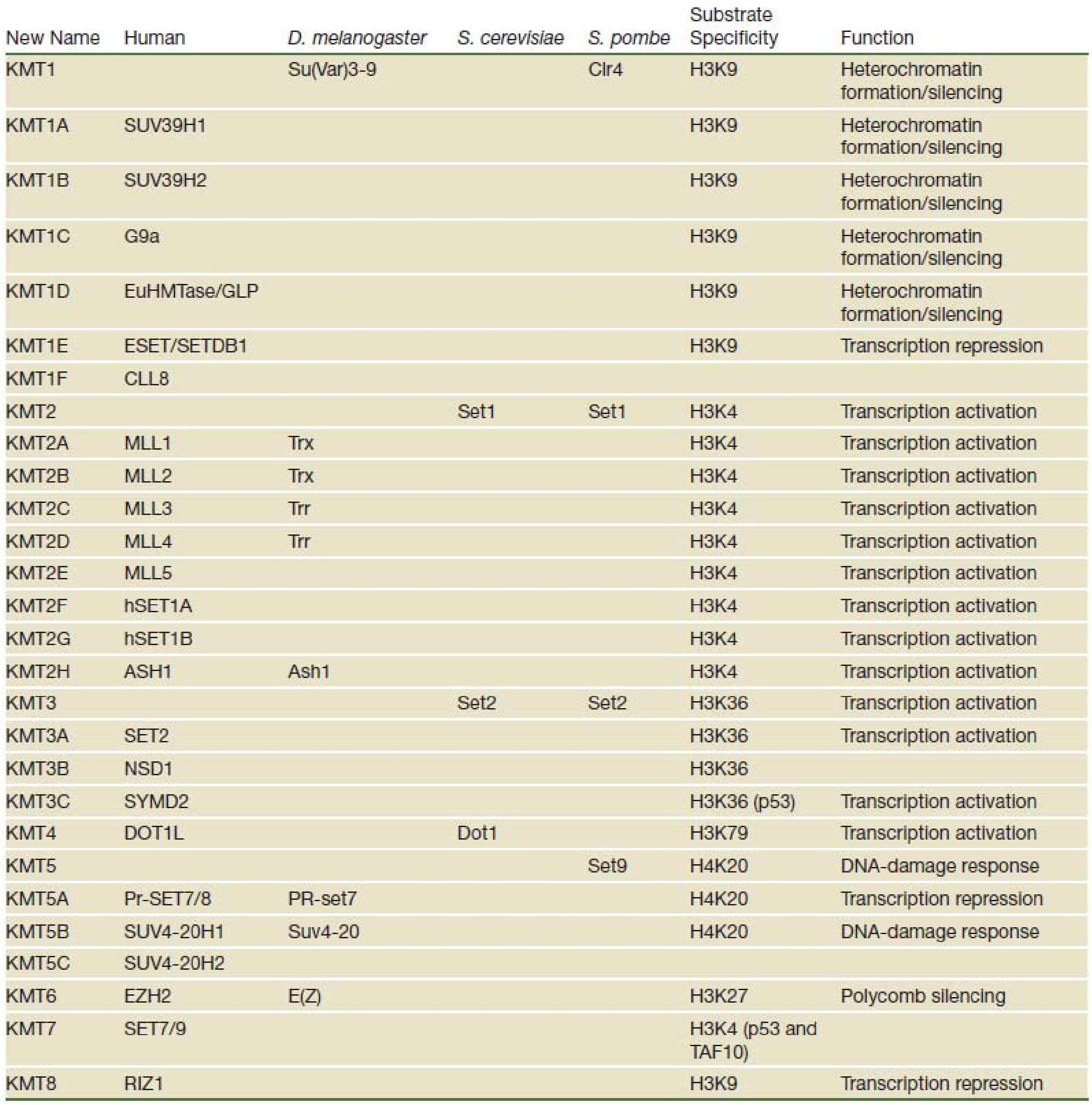
In addition to acetylation, histones proteins can also undergo methylation. The major methylation sites within histone proteins are basic amino acid side chains of lysine and arginine residues and it is catalysed by two distinct classes of enzymes known as PRMT's (protein arginine methyl transferases) family proteins are responsible for arginine methylation and PKMT's (protein lysine methyl transferases) are responsible for lysine methylation. Lysine residues can undergo mono-, di- and tri methylations on their amine groups whereas arginine residues can be mono and dimethylated (which can be asymmetric or symmetric) on their guanidinyl group but here we majorly focus on the lysine methylation. All the known histone methyltransferases uses S-adenosyl-L-methionine (Adomet) as the methyl donor (Andrew J et al., 2002). The extensively studied histone methylation marks include five major lysine (K) residues located within the amino-terminal histone tails of H3 (K4, K9, K36) and H4 (K20) and also at H3K79 in the globular core domain (Ciccone et al., 2009). With the advancement in the field of mass spectrometry, several lysine residues on H1, H2A and H2B proteins were also found to be methylated in vivo but their functional consequences are yet to understand. Methylation on histone proteins is much more complex than the acetylation. Unlike acetylation, methylation on lysine residues does not alter their charge to influence the chromatin structure but it influences the chromatin structure by altering the hydrophobic and steric properties directly and indirectly by recruiting effector proteins to the specific methylated lysine residues. (Martin et al., 2005). While acetylation on histone proteins majorly coincides with the transcriptionally active chromatin state as mentioned above, lysine methylation is associated with either chromatin compaction or decondensation based on the site of methylation and also on the methylation state (mono-, di- and tri-). In general methylation of H3K9, H3K27 and H4K20 is associated with condensed and repressed chromatin whereas H3K36, H3K4 and H3K79 methylation associated with open and transcriptionally active chromatin (Ciccone et al., 2009). Aberrant methylation of histone lysines has been to shown to involve in various diseases like cancers and X-linked mental retardation (Upadhyay et al., 2011)

3.4 SET Domain proteins

Although histone methylation was reported 4 decades ago, the first family of mammalian protein lysine methyltransferases was discovered only in the year 2000 by Jenuwein and colleagues (Rea et al., 2000). The first HKMT (histone lysine methyltransferases) discovered was SUV39H1 which is responsible for H3K9 methylation. Later many SET domain proteins have been shown to possess histone methylation activity towards specific lysine residues. Till now more than 50 SET domain containing proteins have been identified. Some of them were shown to be active in histone methylation, others possess dual substrates activity on histone and non-histone proteins (G9a, NSD1) and for many of them the specific substrate still has not been identified (like for example SMYD4 and SMYD5, PRDM's and ASH Set domain proteins)

3.5 Nomenclature of Histone Lysine Methyltransferases

Since 2000 several families of enzymes responsible for histone lysine methylation have been identified and many more can come in the future but this in turn has led to non-coherent nomenclature that is inconsistent between species. For instance SET7/9 is a human H3K4 methyltransferase while Set9 is yeast H4K20 methyltransferase. To avoid this confusion, recently Allis et al., proposed a new nomenclature for all the characterised members of the families of lysine demethylases, acetyltransferases and lysine methytransferases (table1) (Allis et al., 2007). The new nomenclature is based on the close relationship in sequence and domain structure, second consideration is the substrate specificity. The related enzymes from a single species have been given the same name but with the capital letter as a distinguished suffix (e.g., A or B). Similarly, enzymes from different species have been given an identical name but with different prefix to denote species of origin (e.g., h= Human, d= Drosophila, Sc= Saccharomyces cerevisiae). The first three numbers in the nomenclature were assigned according to the order of discovery.



New Name	Human	*D. melanogaster*	*S. cerevisiae*	*S. pombe*	Substrate Specificity	Function
KMT1		Su(Var)3-9		Clr4	H3K9	Heterochromatin formation/silencing
KMT1A	SUV39H1				H3K9	Heterochromatin formation/silencing
KMT1B	SUV39H2				H3K9	Heterochromatin formation/silencing
KMT1C	G9a				H3K9	Heterochromatin formation/silencing
KMT1D	EuHMTase/GLP				H3K9	Heterochromatin formation/silencing
KMT1E	ESET/SETDB1				H3K9	Transcription repression
KMT1F	CLL8					
KMT2			Set1	Set1	H3K4	Transcription activation
KMT2A	MLL1	Trx			H3K4	Transcription activation
KMT2B	MLL2	Trx			H3K4	Transcription activation
KMT2C	MLL3	Trr			H3K4	Transcription activation
KMT2D	MLL4	Trr			H3K4	Transcription activation
KMT2E	MLL5				H3K4	Transcription activation
KMT2F	hSET1A				H3K4	Transcription activation
KMT2G	hSET1B				H3K4	Transcription activation
KMT2H	ASH1	Ash1			H3K4	Transcription activation
KMT3			Set2	Set2	H3K36	Transcription activation
KMT3A	SET2				H3K36	Transcription activation
KMT3B	NSD1				H3K36	
KMT3C	SYMD2				H3K36 (p53)	Transcription activation
KMT4	DOT1L		Dot1		H3K79	Transcription activation
KMT5				Set9	H4K20	DNA-damage response
KMT5A	Pr-SET7/8	PR-set7			H4K20	Transcription repression
KMT5B	SUV4-20H1	Suv4-20			H4K20	DNA-damage response
KMT5C	SUV4-20H2					
KMT6	EZH2	E(Z)			H3K27	Polycomb silencing
KMT7	SET7/9				H3K4 (p53 and TAF10)	
KMT8	RIZ1				H3K9	Transcription repression

Table1: New nomenclature for Lysine methyltransferases (KMT's)

Structures of different SET domain proteins have been solved either in the free form or in combination with bound substrate and methyl donor (Adomet) or reaction product (S-adenosyl-L-homocysteine, AdoHcy). These structures reveal that the conserved SET domain has a unique fold that is different from the other methyltransferases like DNA methyltransferases and protein arginine methyltransferases, that also use the cofactor S-adenosyl-L-methionine (SAM) as the methyl donar. The majority of HKMTs posses a conserved 130-residues SET domain flanked by preSET (N-terminal) and postSET (C-terminal) domains (figure 2). The preSET domain helps to keep the structure stability by interacting with different surfaces of core SET domain. The SET domain adopts a unique structure formed by a series of β-strands folded into three sheets surrounded by postSEt domain. The postSET domain forms a knot like structure to support the formation of active site in core SET domain, this knot like structure brings the two conserved sequence motifs (RFINHXCXPN and ELX(F/Y)DY) of the SET domain, in close proximity to the cofactor binding region and substrate binding pocket and thus construct a hydrophobic channel (Qian et al., 2006 and Upadhyay et al., 2011). Another intriguing feature of the SET domain methyltransferases is an inserted region called i-SET, amino acid residues in the i-SET domain have been observed to interact with the substrate peptide in three dimensional structures of different SET domain proteins. This domain varies considerable in length and sequence is not conserved among different SET domain proteins. The i-SET region plays a major role in discriminating between their different substrate targets, for instance SET7/9 and MLL1 both have identical substrate specificity but very different i-SET region (Xiao et al., 2003). Though these two enzymes share the same primary substrate they interact with different residues in the substrate through the i-SET domain, hence studying the specificity profile of each enzyme is crucial to understand its specificity towards substrates.

The enzyme active site in SET domain proteins is majorly formed by hydrophobic amino acids, they constitute a narrow hydrophobic channel that links the cofactor binding site on one surface with the substrate binding site on the opposite surface of the domain. The cofactor and substrate bind in two different grooves located on the opposite surfaces of SET domain. The geometry, shape and type of amino acids that comprise this lysine access channel are responsible for determining how many methyl groups that Set domain protein can add (Xiao et al., 2003 and Qian et al., 2006). Recent biochemical studies performed with F/Y mutants of the conserved ELx(F/Y)DY motif of lysine access channel in DIM5 (F281Y), G9a(F1205Y), SET8(Y334F) and SET7/9(Y305F) showed that the F/Y switch regulates the product specificity (mono-, di- or tri-methylations) of SET domain proteins (Upadhyay et al., 2011).

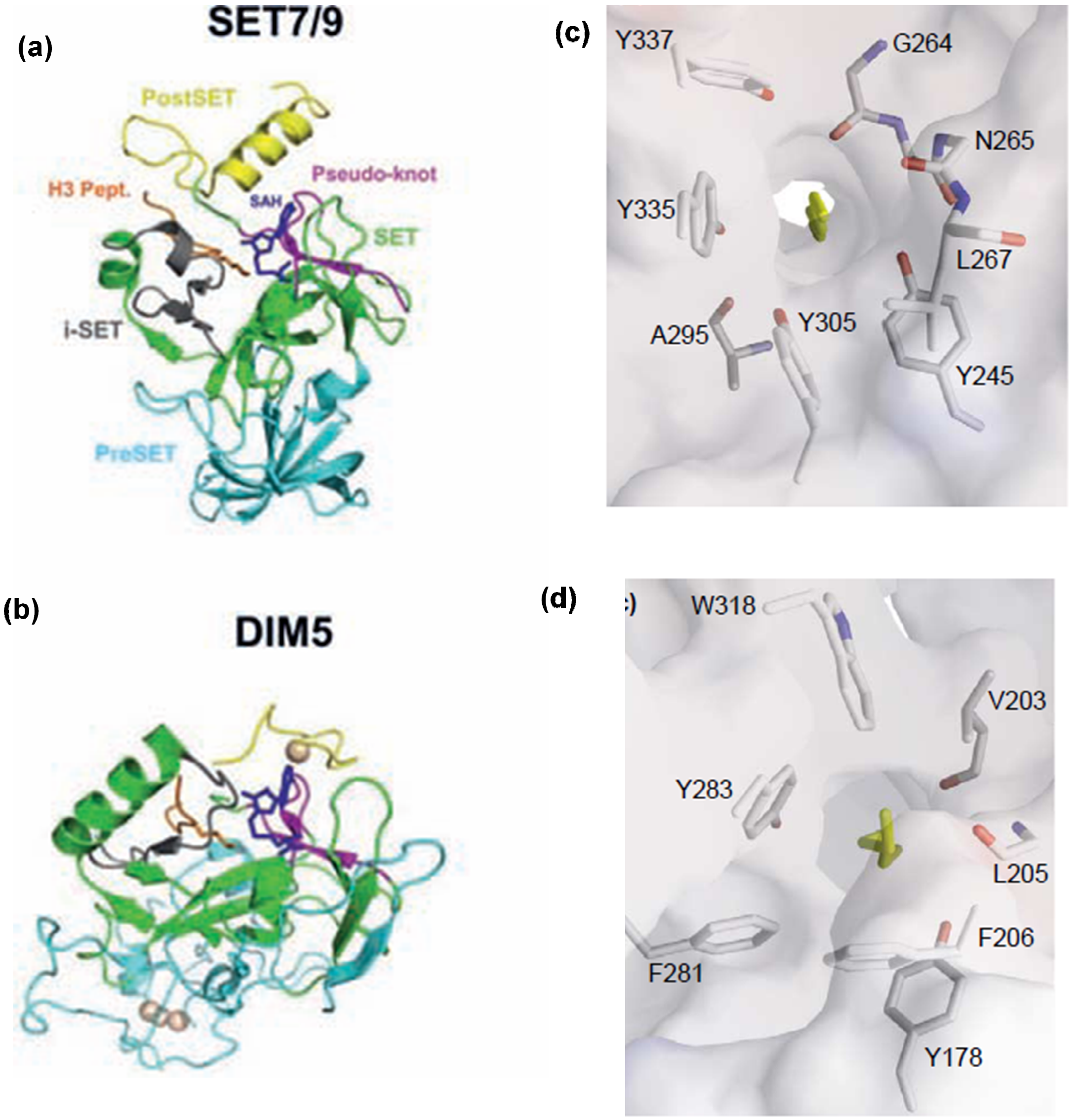


Figure 2: (a, b) Three-dimensional structures of SET domain proteins. Proteins, preSET, i-SET, SET and postSET regions are depicted in cyan, light gray, green and yellow; the pseudo knot, cofactor product SAH and substrate peptides are shown in magneta, blue and orange (Qian et al., 2006). (c,d) surface representation of lysine access channel viewed from the peptide binding site: c- SET7/9 and d- DIM5 (Xiao et al., 2003)

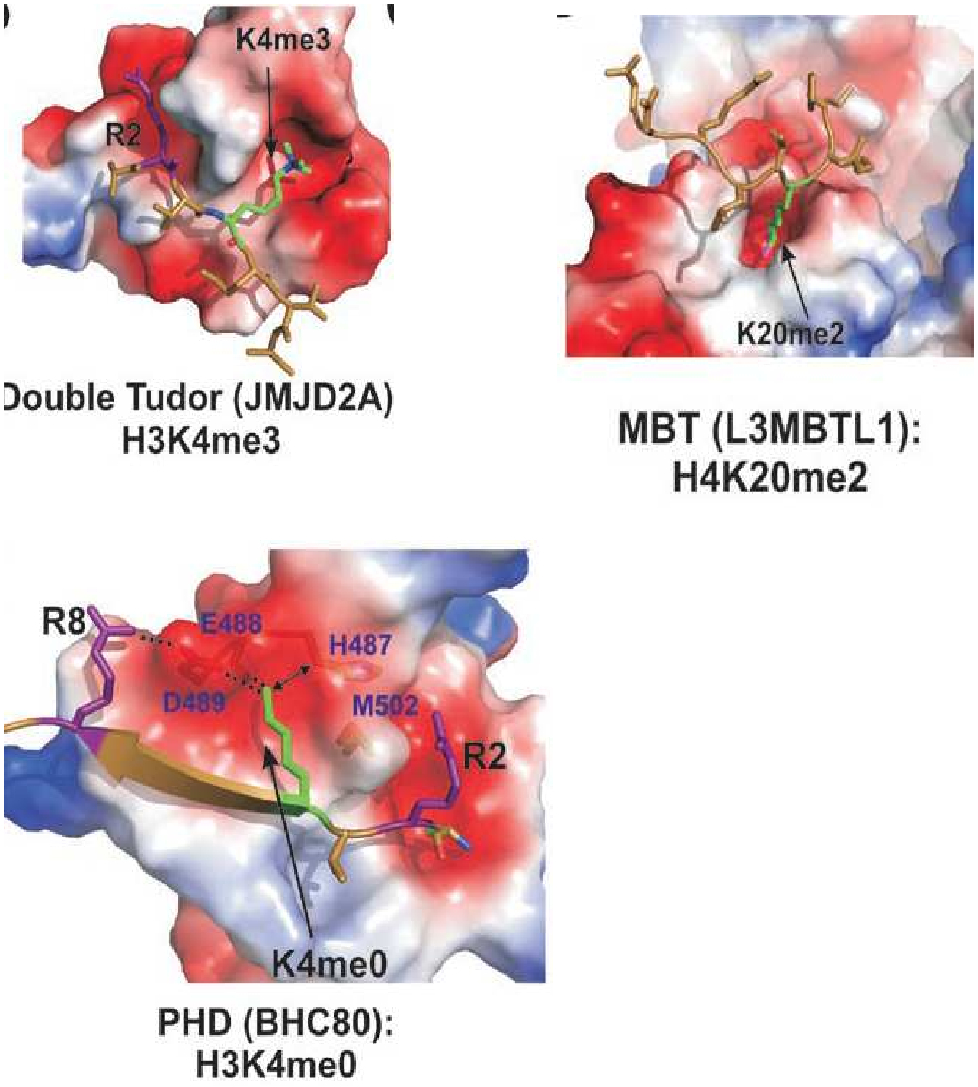
3.6. Reading Domains

Modifications on histone proteins can directly influence chromatin structure. For instance acetylation of histone lysine residues majorly mediates their effects on chromatin organisation through altering the charge properties of the modified residue. In contrast, methylation of lysine residue is relatively inert which excludes any direct influence on chromatin structure (Volkel et al., 2007). However, the diverse chemical moieties involved in the specific histone modifications transmit their biological signals through recruiting effector proteins that recognize distinct modification on specific residue. Acetylated lysines residues can be recognised by bromodomain and PHD domain (Yun et al., 2011) containing proteins while methylation on lysine residues is recongnised by chromodomain, PHD finger, Tudor domain, Ankyrin repeats, PWWP domaisn and MBT domain containing proteins. Chromodomain of heterochromatin protein (HP1) recognises the trimethylation mark on H3K9 and facilitate the formation of heterochromation and maintainenece of gene repression. Chromodomain of polycomb protein in PRC1 complex recognises the H3K27me3 trimethylation mark which is also majorly associated with gene inactivation (Daniel et al., 2005).

Compared with acetylation, signalling on methylation is more complex because lysines can present four types of signals: unmethylated, as well as mono-, di- and tri- methylation (Bottomley 2004). Unmodified lysine is included in the methyl-lysine (MeK) signalling because most of the me0 readers are sensitive to the addition of the methyl group on the lysine (Yun et al., 2011). Instead of categorising the methyl lysine readers on their function of gene activation or repression we categorise and discuss them based on their ability to recognize the state of methylation. Readers typically provide the accessible surface (groove) to accommodate modified lysine residue based on the state of methylation, MeK readers also interact with the flanking sequence of the modified amino acid in order to distinguish sequence context but the MeK readers which do not make extensive contacts with flanking sequence show a promiscuous methyl recognition pattern.

3.6.1.Binding pockets

MeK binder’s forms an aromatic pocket to accommodate the MeK, primary function of these pockets are to discriminate different methylation states. Mono- and dimethyl binders tend to have small key hole like cavity which limits the access of large trimethyl group while the di- and trimethyl binders often use a wider and more accessible surface groove as binding pocket (figure 3). Mono and dimethyl readers possess partial aromatic pocket with acidic residues, the acidic residue interacts with methyl ammonium group sterically constricting the cavity and precludes the recognition of me3 methylation state while the me3 binders possess fully aromatic pocket (Yue et al., 2009). Unmethylated lysines (UmK) binders do not have apparent pocket, unmethylated lysine is stabilised by hydrogen bond interactions upon binding with the reader and however addition of methyl groups will disturb the binding surface (Yun et al., 2011).



Double Tudor (JMJD2A) H3K4me3

MBT (L3MBTL1): H4K20me2

PHD (BHC80): H3K4me0

Figure 3: Recognition of H3K4me3 by the double-Tudor domain of JMJD2A (PDB 2GFA). L3MBTL1 MBT bound to H4K20me2 (Yun et al., 2011).

Unmethylated lysine binders: The ADD domain of DNMT3a and DNMT3L, the PHD domains of AIRE and BHC80, WD40 of WDR5 and WDR9 specifically recognise unmethylated H3K4 (Zhang et al., 2010 and Yun et al., 2011)

Mono- and Dimethyl lysine binders: Several domains are known to interact with mono and dimethylated lysines of histone proteins but here we list out only the well studied domains through structural and biophysical experiments. Ankyrin repeats of KMT1C and KMT1C like protein preferentially bind to H3K9me1/me2 marks. Tandem tudor domain of 53BP1 protein selectively recognises H4K20me1/me2 marks, malignant brain tumour like protein1 (L3MBTL1) binds to various me1/me2 marks (Ng et al., 2009)

Trimethyl binders: Trimethyl-lysine marks are the most stable marks on histone proteins and majorly involve in the regulating the gene expression. Several protein domains are known to interact with the trimethyl marks. Chromodomain of HP1 protein specifically recognises H3K9me3, choromodomain of polycomb protein in PRC1 complex recognises H3K27me3 mark (Daniel et al., 2005). PWWP domain of DNMT3a recognises H3K36me3 mark, ADD domain of ATRX protein binds to H3K9me3 mark (Dhayalan et al., 2009 and 2011). RAG2 PHD finger of VDJ protein binds to H3K4me3 mark (Ng et al., 2009), EED protein of PRC2 complex specifically binds to H3K27me3 mark (Margueron et al., 2009). Double tudor domain of JMJD2A binds with H3K4me3 and H4K20me3 histone marks (Huang et al., 2006)

Earlier MeK readers were thought to be only specific for the histone proteins but the recent findings suggests that these readers can also interact with the methylated lysines on non histone proteins based on the state of methyaltion. For instance the MBT domain of L3MBTL1 recognises the p53 K382me1 mark (West et al., 2010) and lysine 860 K860me1 in retinoblastoma protein (Saddic et al., 2010). Ankyrin repeats of KMT1C like protein recognises the K310me1 on ReIA protein (Chang et al., 2011). 53BP1 protein specifically recognises the dimethylation marks on K372 and K382 of p53 protein and thus positively regulates the transcription of its target genes (Kachirskaia et al., 2008 and Huang et al., 2007). Recognition of methyl lysine marks in non-histone proteins by the MeK binders suggest that the methylation on non histone proteins could also leads to the same biological signalling effects of histone lysine methylation like gene activation or repression.

3.7. Non-histone protein methylation

Cellular proteins undergo various post-translational modifications which usually transmit various regulatory signals from protein to protein. Covalent modifications of a protein could lead to protein to protein or protein to nucleic acid interaction, regulate protein stability or enzyme activity and alter the sub cellular localisation (Morgunkova et al., 2006). Protein phosphorylation on serine/threonine and tyrosine are the most intensively studied covalent modification on different proteins and it has been shown to involve in cell cycle regulation and in regulating several other cellular functions (Huang and Berger 2008). Recently lysine methylation and -acetylation on non-histone proteins emerged as the potential modification and increasing number of reports have been shown that these modifications are involved in regulating various cellular processes like phosphorylation. Most of our understanding of lysine methylation comes from the histone proteins, methylation on non-histone proteins also can be seen in the similar lines and also an anology can be made to other covalent modifications like acetylation and phosphorylation.

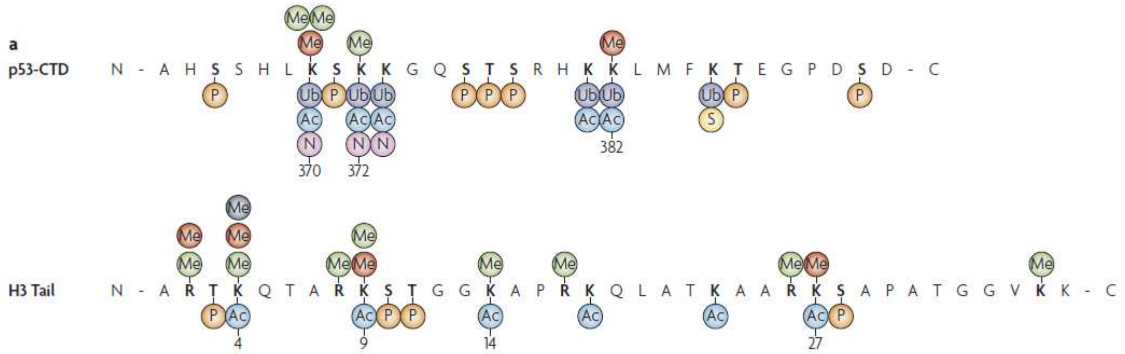
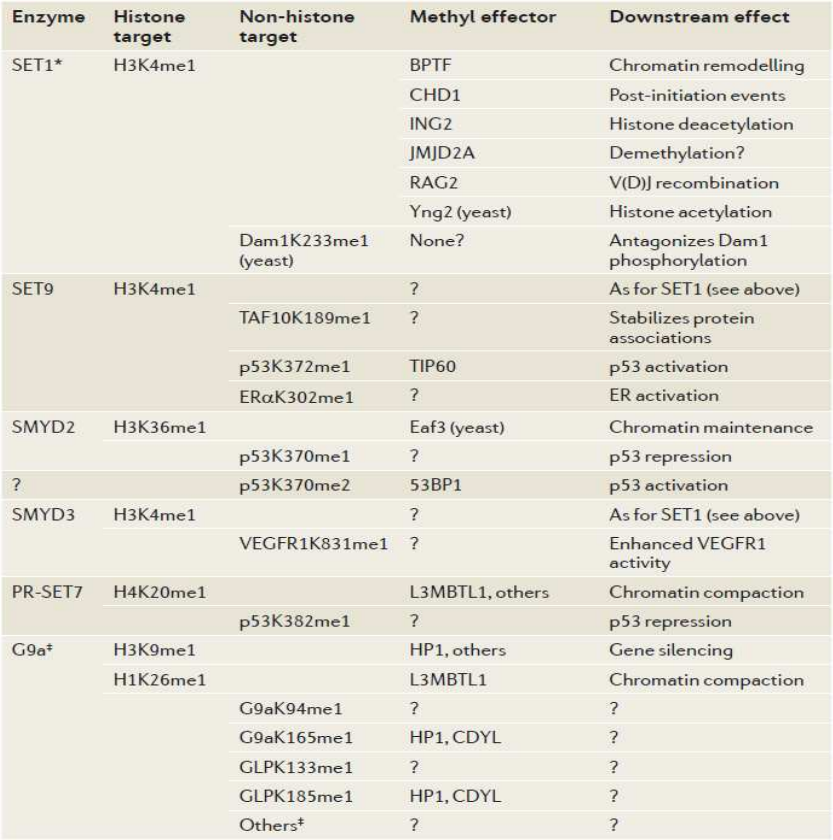

a
p53-CTD
N - A H S S H L K S K K G Q S T S R H K K L M F K T E G P D S D - C
Me
P
Ub
Ac
N
S
370
372
382
H3 Tail
N - A R T K Q T A R K S T G G K A P R K Q L A T K A A R K S A P A T G G V K K - C
4
9
14
27

Figure 4: Post-translational modifications on p53 and histone H3 protein. The different modifications indicated as P-phosphorylation, Ub-Ubiquitylation, Ac-acetyaltion, S-sumoylation (Sims et al., 2008)

SET7/9 was the founding member of non-histone protein lysine methyltransferases (PKMT's) in 2004 when it was identified that SET7/9 methylates the TAF10 protein at K189 position and showen that the specific modification positively influences the transcription of some TAF10 dependent genes (Kouskouti et al., 2004). Soon other group also showed that SET7/9 monomethylates the p53 protein at the K372 position and it enhances its stability (Chuikov et al., 2004). From then on several non-histone proteins have been showed as substrates for the SET7/9 enzyme with distinct functions specific to different substrates. Recently from our lab, we have also identified several non-histone proteins as potential substrates for the G9a and SET7/9 enzyme (Rathert et al., 2008 and Dhayalan et al., 2011), with all these novel non-histone substrates, SET7/9 enzyme evolved as a protein lysine methyltransferase from a canonic histone lysine methyltransferase. The results of SET7/9 intrigued scientific community to search novel non-histone targets for other histone lysine methyltransferases (table 2), p53 protein is methylated at different lysine residues on c-terminal end by distinct protein lysine methyltransferases, SET8 (k372me1), G9a (K373me1), SMYD2 (K370me1).



Enzyme	Histone target	Non-histone target	Methyl effector	Downstream effect
SET1*	H3K4me1		BPTF	Chromatin remodelling
			CHD1	Post-initiation events
			ING2	Histone deacetylation
			JMJD2A	Demethylation?
			RAG2	V(D)J recombination
			Yng2 (yeast)	Histone acetylation
		Dam1K233me1 (yeast)	None?	Antagonizes Dam1 phosphorylation
SET9	H3K4me1		?	As for SET1 (see above)
		TAF10K189me1	?	Stabilizes protein associations
		p53K372me1	TIP60	p53 activation
		ERαK302me1	?	ER activation
SMYD2	H3K36me1		Eaf3 (yeast)	Chromatin maintenance
		p53K370me1	?	p53 repression
?		p53K370me2	53BP1	p53 activation
SMYD3	H3K4me1		?	As for SET1 (see above)
		VEGFR1K831me1	?	Enhanced VEGFR1 activity
PR-SET7	H4K20me1		L3MBTL1, others	Chromatin compaction
		p53K382me1	?	p53 repression
G9a‡	H3K9me1		HP1, others	Gene silencing
	H1K26me1		L3MBTL1	Chromatin compaction
		G9aK94me1	?	?
		G9aK165me1	HP1, CDYL	?
		GLPK133me1	?	?
		GLPK185me1	HP1, CDYL	?
		Others‡	?	?

Table 2: Non-histone targets of various protein lysine methyltransferases (J. Sims et al.,2008)

3.8. p53 as a model for Non-histone protein methylation

The p53 protein is the most commonly mutated gene in all forms of cancer and is known to be regulated via several posttranslational modifications on both N- and C-terminal ends. The C-terminal domain of p53 proteins is also known as basic domain (BD) (residues: 363-393). It contains 6 lysine residues and out of which 4 lysines (K370, K372, K373 and K382) were known to be methylated by distinct protein methyltransferases with a specific biological signal (Scoumanne and Chen 2008) and another lysine K386 was identified to be mono and dimethylated in cells but the specific enzyme responsible for the modification is not yet known (Kachirskaia et al., 2008). The first histone lysine methyltransferase shown to methylate p53 protein was SET7/9. Mono-methylation of p53 protein at K372 by SET7/9 increases its stability which further positively regulates the transcription of p53 target genes but however this signalling pathway via methylation is yet to be understand (Chuikov et al., 2004). SMYD2 an uncharacterised histone lysine methyltransferase was also shown to mono-methylate K370 in p53 protein. SMYD2 methylation on p53 protein inhibits its binding to DNA and thus negatively regulates the expression of target genes. Similar to SMYD2, SET8 also mono-methylates p53 protein at K382 and negatively regulates the transcription of its target genes but how this signalling is mediated was not known then (Shi et al., 2007). Later, it was shown that SET8 mediated methylation of p53 at K382me1 promotes interaction between L3MBTL1 protein and p53 protein, under basal conditions L3MBTL1 interacts with p53 in a methyl (K382me1) specific manner and repress its target genes. In response to DNA damage, p53K382me1 level decreased, resulting in the release of L3MBTL1 from p53 target genes and thus promotion of transcription (West et al., 2010). G9a/GLP di-methylates the p53 protein at K373 and like K370me1 and K382me1, it also helps to maintain p53 in an inactive state (Huang et al., 2010).

p53 also undergoes di-methylation at K370 and K382 but the enzymes responsible for this modification are not known yet. Both the dimethylation signals are specifically recognised by the tandem tudor domains of 53BP1 protein and positively regulates the p53 target gene expressions (Huang et al., 2007 and Kachirskaia et al., 2008). Interestingly, cross talk exists between the different modifications of p53 protein like the histone proteins, SET7/9 methylation on K372 inhibits the methylation of K370 mediated by SMYD2 which is also in accord with their opposite biological outcomes of the corresponding lysine methylations (Huang et al., 2006).

Taken together, this information illustrates that the methylation signalling on p53 protein is analogues to that of histone methylation, indeed we see the cross talk between K370 and K372 methylation similar to cross talk between K4 and K9 in histone H3 and each modification on a specific residue leads to distinct biological outcome. Since p53 is one of the most highly investigated proteins due to its biological importance, it is more likely to observe the possible modifications on the protein. As mentioned above, the p53 protein has 6 lysines on the C-terminal end out of which 5 were shown to be modified by different enzymes. This observation suggests the existence of many more lysine methylation sites within the 20,000 proteins in the human proteome.

3.9. Aims of the present study

Histone lysine methyltransferases has very important role in the epigenetic signalling, lysine methylation on histone proteins alters the chromatin sturucture and thus regulates the expression of target genes. Till now only 5 lysine methylation sites (H3K4, H3K9, H3K27, H3K36, H3K79 and H4K20) have been well characterised on histone proteins. With the advancement in the mass spectrometry applications in the proteomics field novel lysine methylation sites on histone proteins were identified for instance H3K18, H3K23 (Garcia et al., 2006) and several other sites on H1 and H2 proteins (Wisniewski et al., 2007). But for most of these novel lysine methylations, the enzyme(s) responsible for the methylation events and their biological consequence(s) are not known. Apart from the histone lysine methylation, non-histone protein methylation has been emerging as a major post translation modification from the past couple of years. Till now only a few of the non-histone proteins were identified as the substrates for the methylatransferases and thousands of potential targets waiting to be identified.

The main objective of our study is to characterise the specificity of protein lysine methyltransferases and to screen for the specific novel substrates in histone and non-histone proteins. We employed peptide arrays (SPOT synthesis) to determine the specificity profile for the enzymes. Based on the derived specific sequence motif, we identified the proteins in human proteome possessing the target sequence motif and then confirmed the site specific methylation at both peptide and protein level. With this strategy we identified several target lysines in histone and non-histone proteins as substrates for NSD1, SMYD2, SUV39H1 and SET8 proteins. Understanding the biological signalling of the corresponding methylation on non-histone proteins is not trivial, since each protein needs a different experimental setup like knockdown and knockout of corresponding proteins and enzymes. Nevertheless we studied the downstream effects of methylation on the non-histone proteins of SUV39H1 and the experiments are in progress for NSD1 and SMYD2 target proteins.

4. Results

4.1. Specificity analysis of NSD1

4.1.1. Scientific background of NSD1

The Nuclear receptor binding SET Domain containing protein1 , NSD1 (KMT3a), is a 2588 amino acid long protein with a conserved SET domain and other functional domains including PHD and PWWP domains (Huang et al., 1998). The SET domain of NSD1 was reported to methylate H3K36 and H4K20 (Rayasam et al., 2003) and the PHD domains has been shown to recognise methylated H3K4 and H3K9 (Pasillas et al., 2011). NSD1 belongs to a family of proteins including NSD2 (WHSC1/MMSET) and NSD3 (WHSC1L/MMSETL). The SET domain of NSD1 shares sequence similarity with SET2, the sole H3K36 methyltransferase in Saccharomyces Cerevisiae (Li et al., 2009). NSD1 is responsible for post-implantation development, mice deficient in NSD1 exhibits embryonic lethality (Rayasam et al., 2003). On average 5% of human acute myeloid leukemia is caused by the translocation of chromosome 5 which generates NUP98-NSD1, a chimeric gene comprising of encoding the FG-repeat domain of NUP98 fused to the carboxy terminal 60% of NSD1 which contains all the vital domain for transcriptional regulation like PHD, PWWP and SET domain (Wang et al., 2006). NSD1 has been shown to positively regulate the transcription of Hox genes via H3K36 methylation and also the transcription of bone morphogenetic protein 4 (BMP4) and zinc finger protein 36 C3H type-like 1 (ZFP36L1/TPP) (Wang et al., 2006, Lucio-Eterovic et al., 2010). Epigenetic inactivation of the NSD1 promoter through CpG hypermethylation has been shown to be involved in neuroblastomas and glioblastomas. The epigenetic inactivation of NSD1 is associated with global diminished levels of H3K36 and H4K20 trimethylations (Berdasco et al., 2009). Mutations in the NSD1 protein are also responsible for the Sotos syndrome; characterised by facial features like a high anterior hairline, frontal bossing, downslanting palpebral fissures and prominent mandible (FARAVELLI et al., 2005). Recently it has been shown that NSD1 also could methylate proteins other than histones, it was shown to mono and dimethylate p65 protein at K218 and K221 (Lu et al., 2010).

4.1.2. Substrate specificity of NSD1

To analyse the substrate specificty of NSD1, we used the catalytic SET domain of NSD1 coupled to GST, expressed in bacteria and purified. The NSD1 protein had been reported to strongly methylate H3K36 and weakly on H4K20 (Rayasam et al., 2003). Since we aimed to identify the best substrate to derive the specificity profile of enzyme, we proceeded with

H3K36. To confirm the specificity of the NSD1 enzyme, we synthesised peptides of 20 amino acids length with the sequence of histone H3 from 31 to 50 and a peptide with K36A variant in which target lysine had been replaced by alanine on the cellulose membrane by SPOT synthesis. The methylation of the respective substrates was analysed by following the enzymatic transfer of radioactively labelled methyl groups from the coenzyme radio labelled Adomet to the immobilised peptides. After incubation with the NSD1 enzyme, a clear methylation signal was observed at wild (H3K36) peptides and no methylation signal was detected on H3K36A peptides (figure 1).

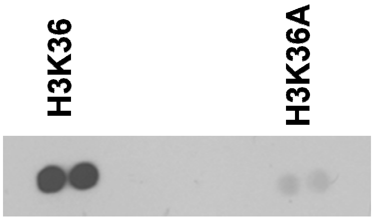


Figure 1: Specificity of NSD1: HKMT assay was performed on H3K36 wild type and mutant peptides to confirm the specificity of NSD1, autoradiography represents the deposition of radio labeled methyl groups on H3K36 peptide and no methylation on peptides with lysine exchanged to alanine

We then performed an alanine scan experiment to understand the importance of each residue on peptide recognition by the NSD1 enzyme. We synthesised an alanine scan array of the H3K36 sequence comprising 21 peptides each carrying an exchange of a single residue against alanine. Methylation was observed in all mutant peptides at similar level as in the wild type peptide except in the case of the V35A, K36A and K37A mutants. The reduced methylation of peptides carrying the substitutions at positions 35 to 37 demonstrated the importance of the corresponding residues in the peptide recognition by NSD1 (figure 2).

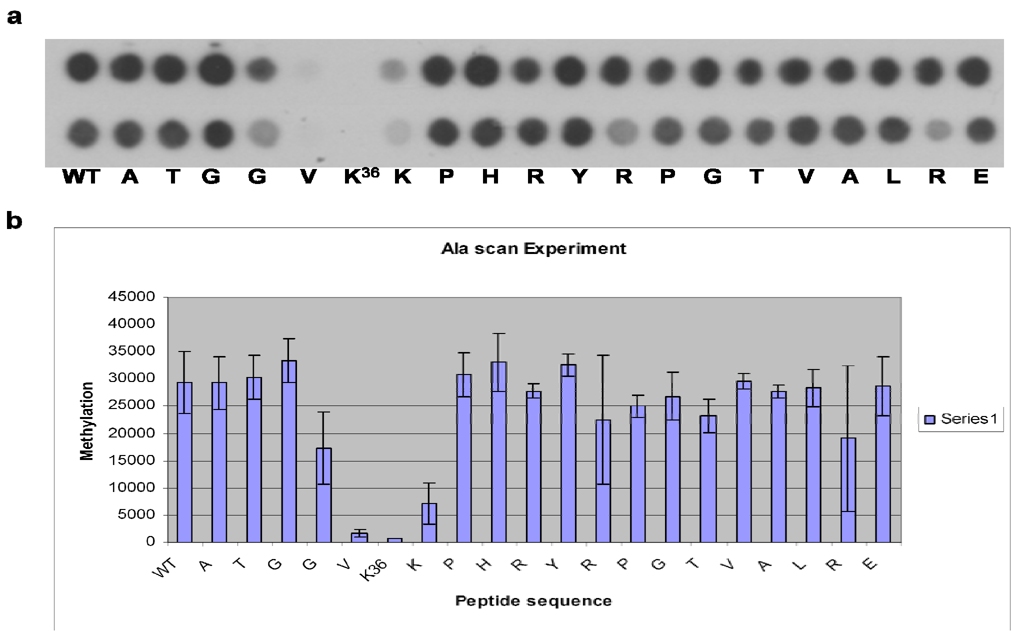


Figure. 2. Alanine scan of H3 tail methylation by NSD1. a) Autoradiography images from the two independent experiments of alanine scan of H3(31-50) peptides. In this assay, all 20 positions of H3 tail were exchanged individually against alanine. The spot labelled with WT contains the wild-type H3 tail sequence. b) Quantitative analysis of the results indicating the average activity and standard error of each target peptide.

Next, we determined the influence of each possible amino acid exchange at each position of the peptide substrate on the activity of NSD1. For this purpose we synthesised a complete array of histone H3 (31 - 50) peptides comprising 420 individual spots, in which each peptide contains an exchange of one amino acid of the wild type H3 tail sequence against each of the 20 natural amino acids. Then, the complete membrane was subjected to methylation by incubation with the NSD1 enzyme and radiolabelled Adomet. The same experiment was repeated three times. After normalisation, the results showed excellent reproducibility, as indicated by the distribution of standard deviations shown in the figure 3c. We calculated the contribution of each amino acid to the recognition of the substrate by the NSD1 enzyme. The relative contribution of each amino acid i at position x for peptide recognition was calculated by discrimination factor D.

$$D = \frac{V_{j\neq i}}{V_i} -1$$

Where as V_i is the rate of modification of peptide carrying amino acid i and $V_{j\neq i}$ is the average rate of methylation of all 19 peptides carrying a different amino acid $j\neq i$ at position x (including wild type sequence) (Rathert et al., 2008).

For example in figure 3d, the discrimination factor of 22 for Isoleucine at position 35 indicates that the peptide with isoleucine at that position is methylated 22 times faster than the average of all the peptides carrying any of the other amino acids at that site. Since the detection limit of the experiment was at about 3% of the full activity, the discrimination factor for K36, which could not be replaced by any other residue, was 35. The results showed that the residues 34 to 38 of histone H3 are important for the substrate recognition of the NSD1 enzyme (figure 3a and 3b). The role of P^{38} was not detected in the alanine scan, because alanine was one of the amino acid residue that could replace proline at position 38. This observation exemplifies the advantage of a complete specificity analysis over just an alanine scan. NSD1 prefers mostly hydrophobic residues at position 34, 35 and 38 in histone H3. Specifically at -1 position to target lysine it accepts only hydrophobic amino acids, apart from valine, the natural amino-acid of the H3 tail, it accepts only leucine and isoleucine, exchange of any other amino acid at -1 position led to the complete loss of the methylation of the corresponding peptide. At the +1 position to the target lysine, NSD1 exhibited a similar activity when lysine was exchanged to arginine and moderate activity when it was exchanged to aspargine, glutamine and methionine. Thus NSD1 accepts positively charged or uncharged polar residues at the +1 position and showed loss of activity when negatively charged or aromatic amino acids are placed there.

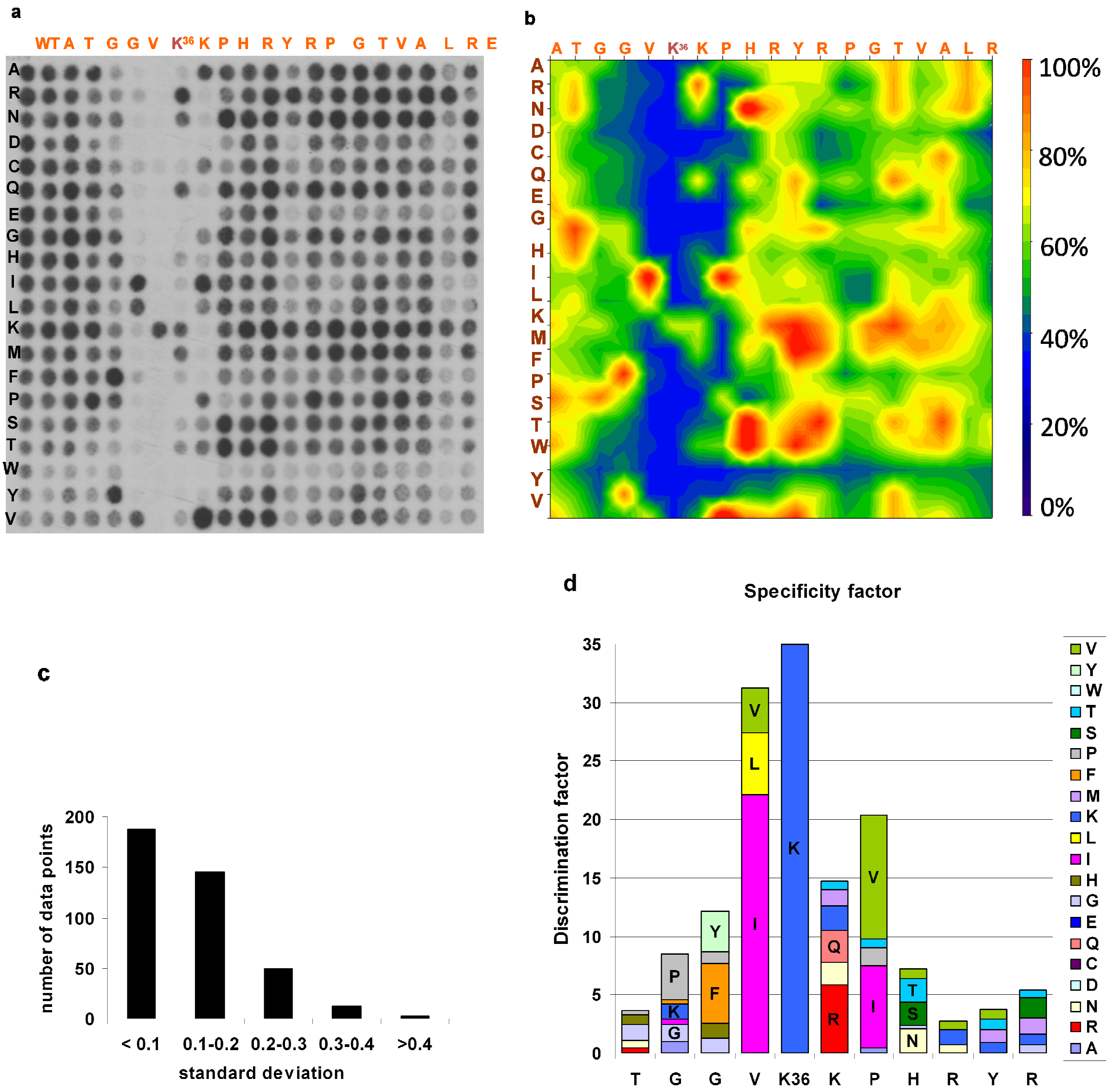


Figure 3: Specificity of peptide methylation by NSD1. a) Example of a complete H3 (31-50) peptide tail array. The horizontal axis represents the sequence of H3 tail. Each residue was exchanged against all 20 natural amino acid residues (represents vertical axis) and the relative efficiency of methylation by NSD1 was analysed. b) Compilation of the results of peptide scan experiments with NSD1. Data are averaged numbers from three experiments after normalizing full activity to 1. c) Distribution of standard deviation of the three experiments compiled. d) Bar diagram showing the discrimination factors of NSD1 at the positions tested.

4.1.3. Identification of non-histone NSD1 target peptides

The specificity profile data obtained from the peptide array experiments, indicated that NSD1 recognises the H3 tail sequence from positions 34 to 38 (-2 to +2 with respect to the target K36). We observed in several cases that exchanges with amino acids different from the natural one in the H3 tail led to an increase in the methylation, for instance peptide exchanged with isoleucine at -1 position exhibits higher activity than with the native sequence peptide, which suggest that

NSD1 might prefer other substrates in the cell. We performed a scansite (http://scansite.mit.edu) search with the NSD1 substrate specificity profile [(YF**G**) (**V**LI) (**K**) (QR**K**NM) (IV**P**)], which resulted in 315 human proteins containing such potential target sites. Based on the localization and function of the NSD1 protein, we narrowed down the search only to the nuclear localized proteins. Among the identified potential targets of NSD1, three proteins were particularly striking, of which one is the well know target of NSD1 protein i.e, H3K36 and the other two are novel histone targets; H4K44 and H1.5K168. As a preliminary screening, we synthesized peptides of all the 48 nuclear proteins encompassing the predicted target lysine in duplicates on a cellulose membrane. The membrane was subjected to methylation by NSD1 and out of the 48 potential peptide targets, 28 got methylated to the same degree or more intensely than histone H3K36 peptide. Along with the other non-histone proteins, we also observed methylation on the two other histone proteins H4 and H1.5 (figure 4).

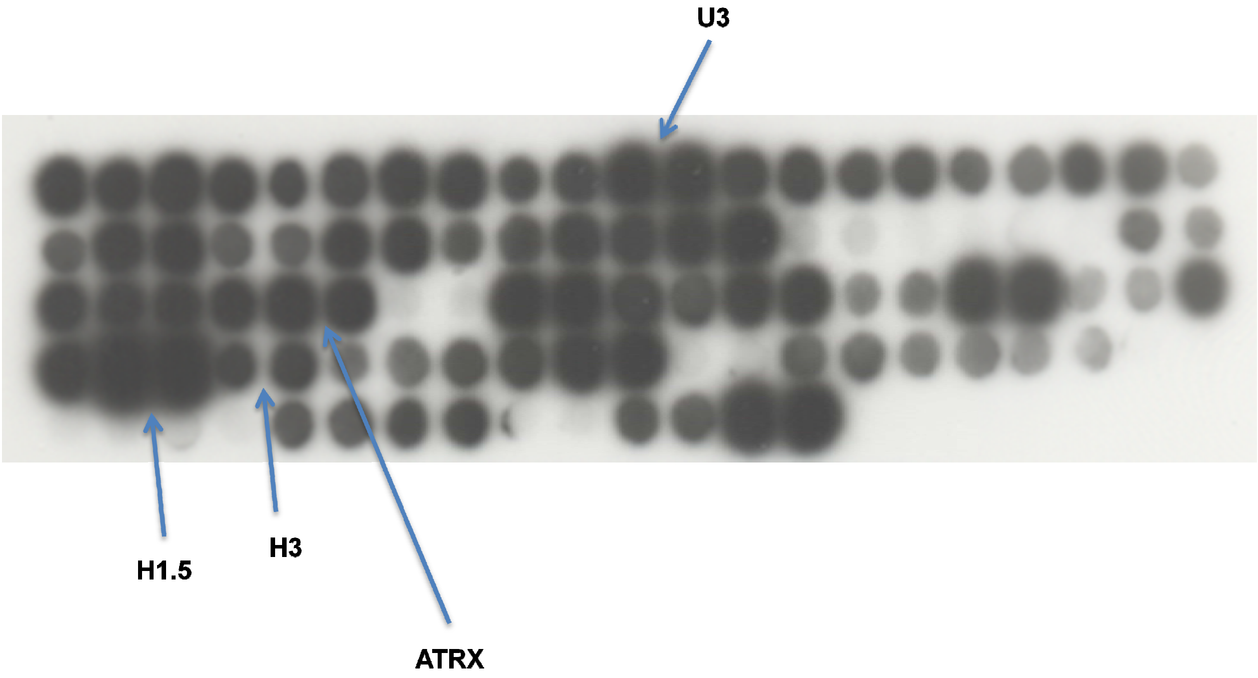


Figure4: Methylation of non-histone targets at peptide level: 20 amino acid length peptides are synthesized in duplicates for non-histone proteins which were identified by scan site search based on the specificity profile of NSD1. Then the membrane was subjected to methylation with NSD1 enzyme, dark spots represent the methylation signal

4.1.4. H4K44 methylation by NSD1 protein

NSD1 had been described to methylate H3K36 and H4K20 (Rayasam et al., 2003). However, the results of our specificity profile experiments were not fitting to the H4K20 sequence, because NSD1 accepts only hydrophobic residues (VLI) at -1 position and charged residues at +1 position. The H4K20 sequence does not contain either hydrophobic residue at -1 position or charged amino acid at +1 position. This motivated us further to investigate the H4 methylation by NSD1. However, when we performed the scansite search with the derived sequence motif of NSD1, we identified another lysine residue in the H4 protein (H4K44). The sequence surrounding the K44 in H4 protein is aptly fitting to the specificity profile of NSD1 (figure 5a). So we speculated that instead of H4K20, NSD1 protein would methylate K44 in the H4 protein. To examine this we synthesised 20 amino acid length peptides containing the target lysines in H3K36 (30-50), H4K20 (10-40), H4K44 (35-55) and also mutant peptides with the putative target lysines exchanged to alanine. The membrane was then incubated with NSD1 to observe the transfer of radio-labelled Adomet by autoradiography. As expected, we did not observe any signal on H4K20 peptide, but we observed a significant deposition of radio labelled methyl groups on H3K36, H4K44 peptides and no methylation on their corresponding lysine mutatant peptide (figure 5b). This confirms that the NSD1 does not methylate K20 in H4 but instead it methylates K44. This also shows the strength of our approach in determining the substrate specificity profiles for the SET domain proteins.

After confirming the target lysine of NSD1 in H4 by the peptide array, we next sought to further examine the H4K44 methyaltion by NSD1 protein via in-solution experiment by MALDI analysis. For this we synthesised H3 (29-44), H4 (37-52) and H4 (12-26) peptides which contain the target lysines, and then methylated the peptides with NSD1 protein in presence of unlabelled Adomet and subjected to MALDI analysis. With H3K36 and H4K44 peptide, we observed a mono-methylation peak at +14 Da in addition to the un-methylated peptide peak (figure 6). However, in accordance with the peptide array experiments we did not observe any methylated peak with H4K20 peptide. Together with the peptide array, mass spectrometry analysis further confirms the methylation on H4K44 peptide.

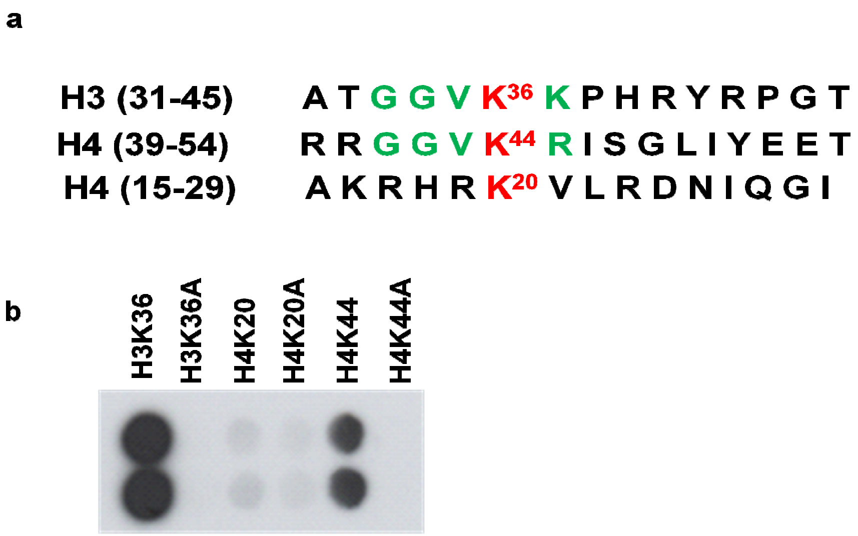

a
H3 (31-45) A T G G V K36 K P H R Y R P G T
H4 (39-54) R R G G V K44 R I S G L I Y E E T
H4 (15-29) A K R H R K20 V L R D N I Q G I
b
H3K36
H3K36A
H4K20
H4K20A
H4K44
H4K44A

Figure 5: Identification of the target lysine of NSD1 in H4: a) Sequence alignment of H3 and H4 sequences encompassing K36 in H3, K20 and K44 in H4. b) Autoradiography image of peptides synthesised with target lysines and the corresponding mutants of H3 and H4 and methylated with NSD1 in presence of radio labelled AdoMet.

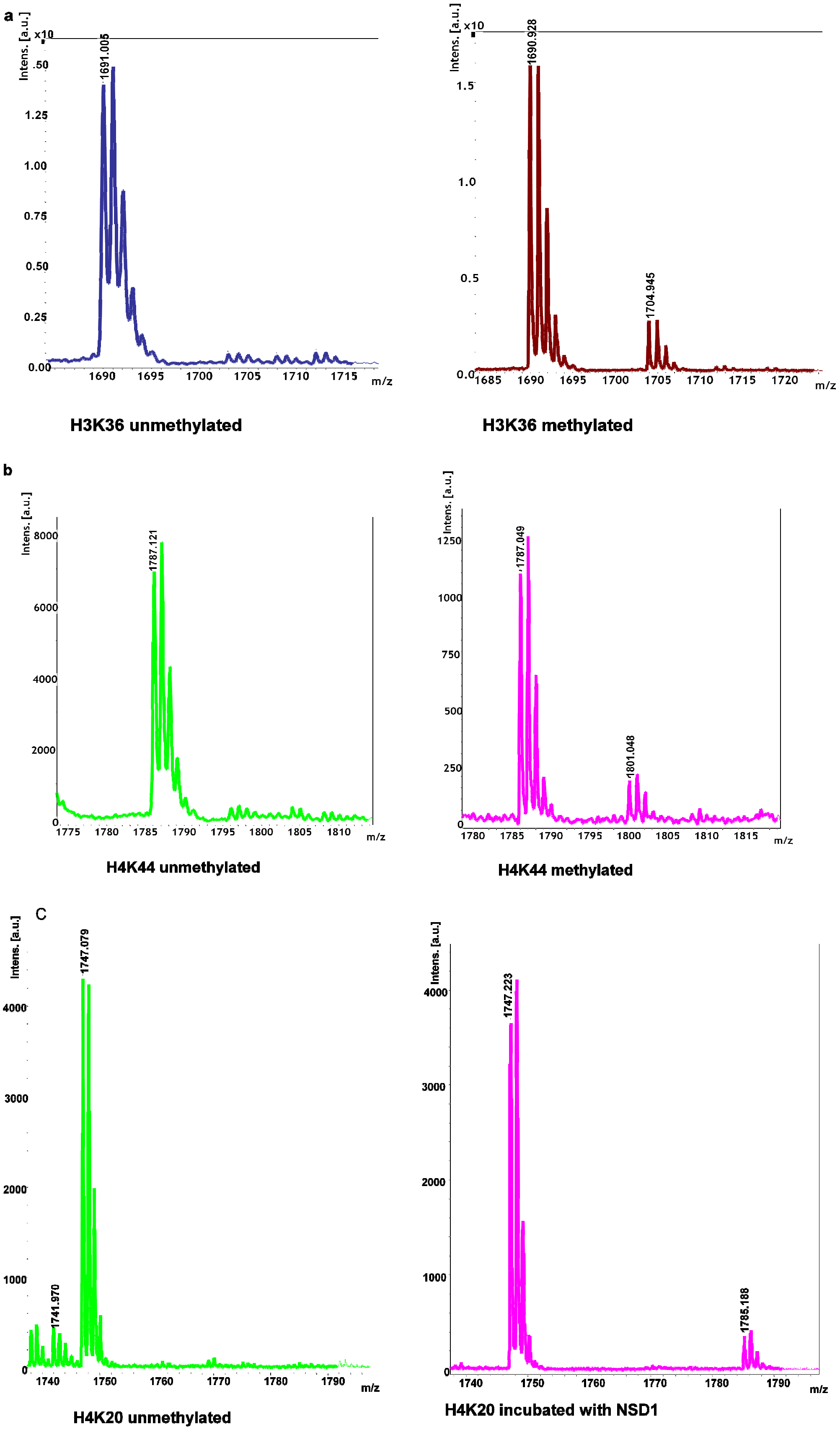


Figure 6: Methylation of H3K36, H4K44 and H4K20 by NSD1. a) MALDI analysis of H3K36 (APATGGVKKPHRYRPG) peptide before and after methylation with NSD1 in presence of unlabelled Adomet. b) MALDI analysis of H4K44 (LARRGGVKRISGLIYE) peptide before and after methylation with NSD1 in presence of unlabelled Adomet. c) MALDI analysis of H4K20 (KGGAKRHRKVLRDNI) peptide before and after methylation with NSD1 in presence of unlabelled Adomet.

4.1.5. H1.5K168 methylation

The H1 linker histones generally participate in the establishment of the chromatin structure and are also involved in the regulation of gene expression. In humans, 11 H1 variants exist, these variants exhibit cell type and tissue specific expression pattern, H1.1 to H1.5 histones express replication dependently and they are present in all somatic cells (Happel et al., 2009). Similarly as the core histones the H1 proteins are subjected to several post translational modifications. The phosophorylation on H1 proteins has been studied and it has been shown to be cell cycle dependent. Analysis with the modern mass spectrometry has recently revealed several post translational modifications on H1 histones including methylation (Wisniewski et al., 2007) (Zougman et al., 2009) but the enzymes responsible for methylaiton are not known. The first methylation site identified on H1 proteins was H1K26 methylation, it was shown to be methylated by Ezh2 (Kuzmichev et al., 2004) and lately G9a was shown to methylate H1.4K26 (Rathert et al., 2008, Trojer et al., 2009) and H1.2K187 (Weiss et al., 2010).

With the target sequence motif of NSD1 via scansite search (http://scansite.mit.edu/), we identified that K168 in H1.5 protein could be a substrate for NSD1 and further observed its methylation at the peptide level (figure 4). Recently by mass spectrometry analysis it has been shown that the H1.5 and H1.3 proteins are indeed methyated either at the K168 or K169 positions (Wisniewski et al., in 2007). The sites could not be differentiated but was shown that one of them exists is methylated in vivo.

In the light of these observations, we were stimulated to examine the NSD1 protein methylation site on the H1.5 protein and to confirm the methylation site both at the peptide and protein level. We did a sequence alignment of all the identified novel histone sites (H4K44 and H15K168) with the previously identified H3K36 (figure 7a). H3K36, H4K44 and H1.5K168 show high sequence similarity, H4K44 and H3K36 shares same residues on -1,-2 and -3 positions and differ only at the +1 side, but the change is in agreement with the NSD1 specificity profile. Similarly H1.5K168 shares same residues with H3K36 at -1 and -2 position and also on +1 position to target lysine, but at +2 position to the target lysine in H1.5 it differs from H3K36 but in line with the specificity profile of NSD1 protein.

To confirm the methylation on the predicted lysines, we synthesised 15 amino acid length peptides for H3K36 (30-44), H4K44 (38-52), H1.5K168 (161-175) and also the mutant peptides with the corresponding target lysine exchanged to alanine. Though the sequence alignment of H1.5 protein with H3K36 suggests K168 as target lysine to further confirm it, we mutated both the lysines at the K168 and K169 positions. We then incubated the membrane with the NSD1 protein and subjected to autoradiography (figure 7b). As expected strong methylation signals were observed on H3K36, H4K44 and H1.5K168 wild type peptides and no signal was observed on the corresponding mutated peptides. The methylation signal on H1.5K169A mutant peptide is in par with the wild type H1.5 peptide and no methylation signal was detected with H1.5K168 mutated peptide, which further confirms that NSD1 protein methylates H1.5 peptide on K168. It also showed that H1.5 was methylated better than H3, which is in accordance with our specificity profile data, because a strong methylation signal was observed in the specificity profile H3(31-50) peptide array when proline at +2 was exchanged with valine (figure 3a). To examine, if the methylation of these novel targets was specific to NSD1, we studied their methylation with another H3K36 methyltransferase HYPB (SETD2). With HYPB we observed strong methylation on H3K36, very weak methylation on H1.5K168 and H4K44 peptides and no methylation on the corresponding lysine mutants (figure 7c). This suggests that the novel targets are specific to NSD1.

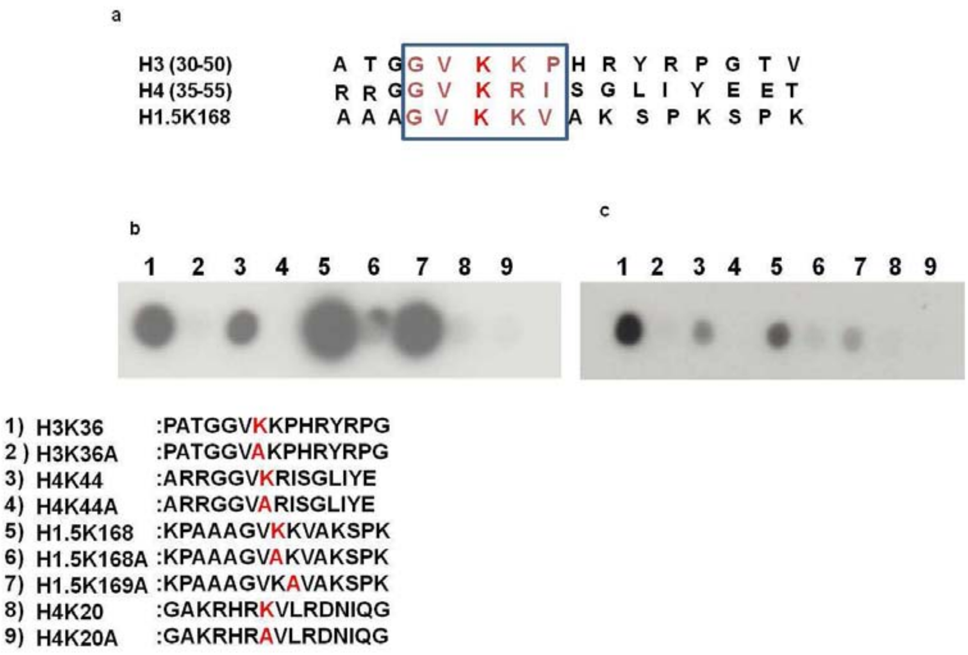
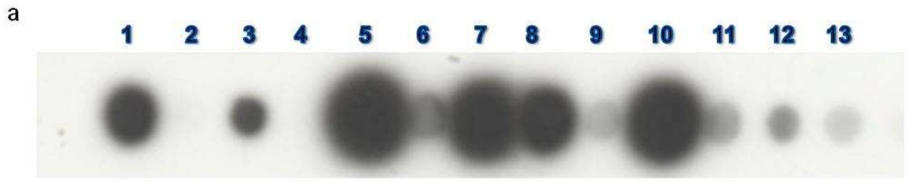

a
H3 (30-50) A T G G V K K P H R Y R P G T V
H4 (35-55) R R G G V K R I S G L I Y E E T
H1.5K168 A A A G V K K V A K S P K S P K
b
1 2 3 4 5 6 7 8 9
c
1 2 3 4 5 6 7 8 9
1) H3K36 :PATGGVKKPHRYRPG
2) H3K36A :PATGGVAKPHRYRPG
3) H4K44 :ARRGGVKRISGLIYE
4) H4K44A :ARRGGVARISGLIYE
5) H1.5K168 :KPAAAGVKKVAKSPK
6) H1.5K168A :KPAAAGVAKVAKSPK
7) H1.5K169A :KPAAAGVKAVAKSPK
8) H4K20 :GAKRHRKVLRDNIQG
9) H4K20A :GAKRHRAVLRDNIQG

Figure7: a) Alignment of identified histone target sequenced with H3(31-50) and H4(35-55) and target lysine highlighted in red b) Autoradiography of peptides synthesised with the wild type sequences and mutated peptides, with predicted target lysine exchanged to alanine and incubated with NSD1 in presence of radio labelled Adomet c) Autoradiography of peptides synthesised with the wild type sequences and mutated peptides, with predicted target lysine exchanged to alanine and incubated with HYPB (SETD2) in presence of radio labelled Adomet

4.1.6. Variant specific H1 methylation

Unlike core histone proteins, H1 protein exists in several subtypes. H1.1 to H1.5 proteins exists in all the somatic cells whereas others are expressed tissue specific and only in germ cells (Nicole et al., 2009). Since K168 was also shown to be methylated in H1.2 (Lu et al., 2009), H1.3 and H1.4 (Wisniewski et al., 2007) in mammalian cells, we proceeded to investigate whether NSD1 could methylate other H1 proteins or not. To examine this we synthesised 15 amino acid length peptides of H3, H4 and variants of H1 from H1.2 to H1.5 along with the predicted target lysine mutants. Upon incubation with the NSD1 protein in presence of radio labelled Adomet, strong methylation signals were observed on H1.3 and H1.5 peptides, moderate signals were detected with H3 and H1.2 peptides and methylation of H4 protein was weak. Loss of methylation signal was observed with all the corresponding target lysine mutants. Though we see strong methylation on H1.2, H1.3 and H1.5, no methylation was detected on the H1.4 peptide (figure 8a and b). NSD1 accepts majorly isoleucine, valine or proline to -1 position to target lysine (H3K36), where as H1.4 protein has alanine (A167) at -1 position. In the specificity profile experiment (Fig. 3a) we see the complete lost of NSD1 activity on the peptide when valine at -1 position was exchanged to alanine, which is in accordance with the results of no methylation of H1.4 protein by NSD1 enzyme. This suggests that the NSD1 protein exhibits H1 variant specific methylation.



b Peptide sequences

1) H3K36	: PATGGVKKPHRYRP	8)H1.2	:KPAAATVTKKVAKSP
2) H3K36A	: PATGGVAKPHRYRP	9) H1.2A	: KPAAATVTAKVAKSP
3) H4K44	: ARRGGVKRISGLIYE	10) H1.3	: KPATAAGTKKVAKSA
4) H4K44A	: ARRGGVARISGLIYE	11)H1.3A	: KPATAAGTAKVAKSA
5) H1K168	: PAAAGVKKVAKSPK	12)H1.4	: KPAAAAGAKKAKSPK
6) H1.5K168A	: PAAAGVAKVAKSPK	13)H1.4A	: KPAAAAGAAKAKSPK
7)H1K169A	: PAAAGVKAVAKSPK		

Figure 8: Autoradiography of NSD1 novel targets a) Peptides of the predicted histone substrates of NSD1 protein were synthesised with the target lysine at the centre along with the mutant peptides in which target lysine exchanged to alanine and subjected to methylation with NSD1 in presence of radio labelled Adomet. b) Sequences used for the peptide synthesis and predicted lysine and the mutated amino acids were highlighted in red colour.

After confirming the methylation of histone H1 at the peptide level, we proceeded further to study the methylation of H1 variants at the protein level. Since the target lysines in H1 proteins is located in the C-terminal domain, they might be involved in the folding and not accessible for methylation. So, it is important to confirm the methylation at the protein level. We used the recombinant untagged H1 proteins to perform in vitro methylation assays, using the 4 H1 variants (H1.2, H1.3, H1.4 and H1.5) for which we had methylation data from the peptide arrays. The methylation of H1 variants by NSD1 was analysed by incubating the proteins with NSD1 in reaction buffer containing radio labelled Adomet. The proteins were then separated by SDS-PAGE and the transfer of methyl groups was analysed by autoradiography (Figure 9a and 9b). As expected we detected a significant incorporation of radio active methyl groups on H1.5 protein. Relatively less methyation signal was detected on H1.2 and H1.3 protein but no signal was observed on H1.4 protein. The lower bands visible in the H1.2, H1.3 and H1.5 lines, likely resulted from the truncations of these proteins. Out of the 4 variants of H1, strong methylation was detected on H1.5 protein which is evident from both the peptide and protein methylation and no methylation on H1.4 protein.

The H1 proteins are highly basic and contain several lysine residues, to confirm the methylation is occurring on the target lysine and to exclude the possible methylation signal from other lysine residues in the protein, we exchanged the target lysines in all three H1 variants (H1.2, H1.3 and H1.5) to arginine. To perform this we cloned the H1 proteins in GST expression vector and exchanged the target lysine to arginine by site directed mutagenesis, subsequently the wild type and mutated proteins were overexpressed and purified by affinity chromatography. We used NSD1 enzyme to perform in vitro methylation assays with either wild type or its KtoR mutant proteins as substrate. The autoradigraphy results (figure 9c) showed a significant incorporation of radioactive methyl groups on all the three wild type H1 proteins but no signal was detected on their corresponding target lysine mutants. This further confirms that the H1 proteins were methylated by NSD1 enzyme on the predicted lysine (K168). These results demonstrates that the NSD1 enzyme catalyses specific methylation on K168 in H1.2, H1.3 and H1.5 proteins and not on H1.4 protein. However K168 methyaltion on H1.4 was also observed in cells along with the other H1 variants, but it might be catalysed by other enzymes which are yet to be identified.

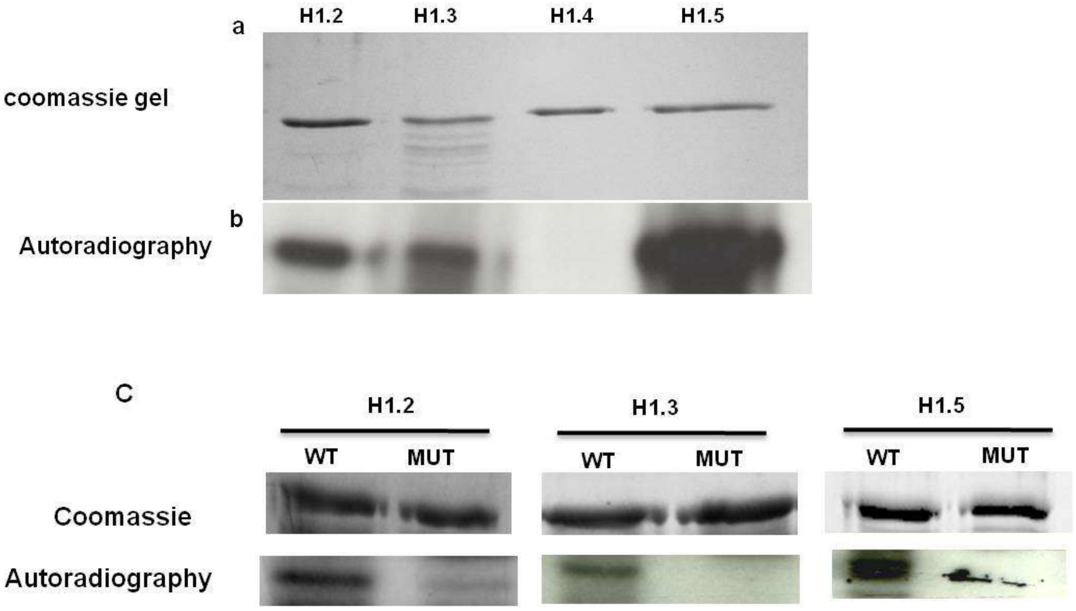


Figure 9: In vitro methylation of H1 proteins: a) Untagged H1 proteins stained with coomasie to show the equal amount used for methylation assays. b) Autoradiography of H1 proteins to assess the transfer of radio labelled methyl groups from radio lableed AdoMet to H1 proteins after incubated with NSD1 enzyme. c) Confirmation of predicted target lysine: The GST tagged wild type H1 target proteins and the mutant proteins in which the target lysine was mutated to arginine were methylated with NSD1 in presence of radio labelled AdoMet.

4.1.7. Product Specificity on H1.5K168

The K36 residue in histone proteins is subjected to mono, di- and trimethyaltion. The SET2 protein was shown to carry out all the three possible methylations on H3K36 in vitro and trimethylation in vivo (Yuan et al., 2009) while NSD2 and its homologs are reported to do di-methylation on H3K36 in vivo (Li et al., 2009). Li et al., 2009 showed that NSD2 protein dimethylates H3K36 and H4K44. However, their mass spectrometry analysis with the H3 protein showed equal mono- and dimethyation peaks but with H4 protein they showed strong mono-methylated peak and little or no dimethylation peak. However, we observed only mono-methylation on both H3K36 and H4K44.

After confirming the target lysine on H1.5 protein, we next sought to study the degree of methylation on H1.5K168. We performed an HKMT assay by incubating the NSD1 protein, unlabelled Adomet and peptide and followed by analysis on mass spectrometry. The mass spectrometry analysis with the methylated sample revealed a monomethylated peak at 1326.831 Da along with the unmethylated peptide peak 1312.8 Da, while we did not see any peak at 1326.8 in unmethylated sample. This suggests that the NSD1 protein monomethylates K168 in H1.5 protein (figure 10).

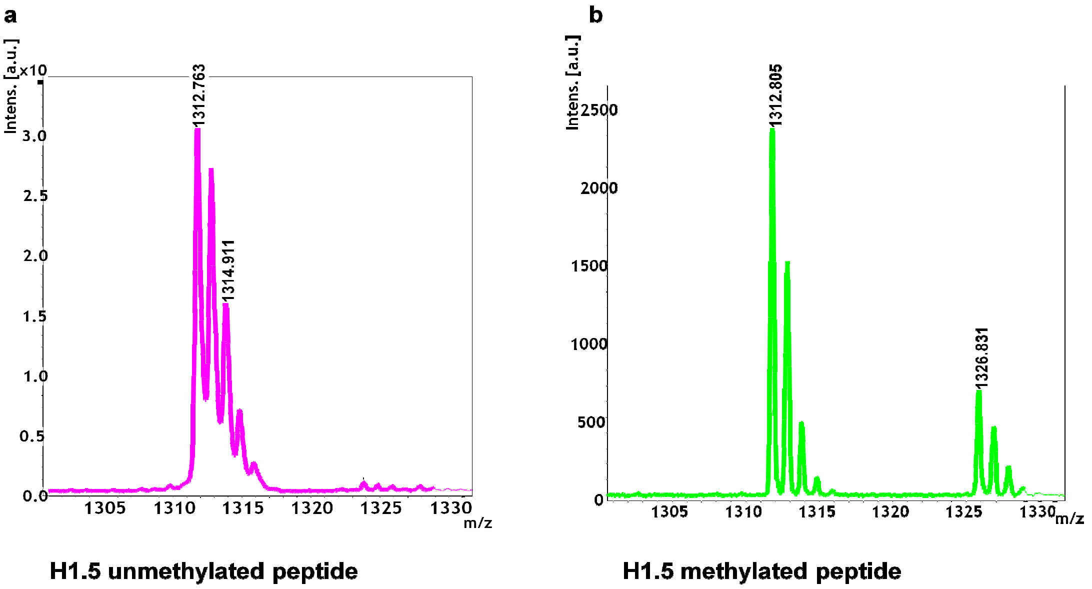


Figure10: In vitro methylation analysis of H1.5 peptide: H1.5 peptide was incubated with NSD1 and SAM in reaction mixture for 4h at 37C and subjected to analysis by mass spectrometry. a) Unmethylated sample, 1312.8 corresponds to unmethylated peak. b) Methylated sample; additional peak was detected at 1326.8 which corresponds to the mono-methylated peptide along with the unmethylated peptide peak (1312.8).

4.1.8. H1.5 K168 methylation Vs H3K36 methylation

In all our peptide array experiments, we observed the methylation on H1.5K168 was much stronger than methylation on H3K36 and H4K44. Indeed, the high methylation activity of NSD1 on H1.5K168 can also be explained from our specificity profile array results, the exchange of proline to valine at +2 postion in H3(31-50) peptide exhibited strongest methylation among the other 420 peptides synthesised and methylated in parallel (figure 3a). In agreement with this, histone H1.5 protein also has a valine (170) at +2 position to the target lysine K168.To directly compare the efficiency of NSD1 mono methylation on H1.5 K168 and H3K36, we used catalytic domain of NSD1 to perform in vitro methyltransferase assay with the GST tagged H1.5 and H3 protein. After methylation we separated the samples by SDS-PAGE and subjected them to autoradiography. We detected a significant incorporation of radio labelled methyl groups on both H1.5 protein and H3 protein, however methylation on H1.5 protein was much stronger than the methylation on H3 protein. From this we can conclude that methyaltion on H1.5 is stronger than H3 protein (figure11).

To further study the efficiency of methylation on H1.5 protein over H3, we did competitive methylation of both the peptides encompassing K168 of H1.5 and K36 of H3 with catalytic domain of NSD1 protein and analysed by MALDI. To perform this we incubated both peptides in the same reaction tube in the presence of unlabelled Adomet and catalytic domain of NSD1 for 4 hrs at 37^{o}c and then analysed the conversion of the unmethylated peptide to methylated peptide. Mass spectrometry analysis revealed that the methylation on H1.5 peptide is three times faster than on the H3 peptide which is in accordance with out peptide array and protein methylation results. All this strongly support that the H1.5K168 is the more preferred substrate for NSD1 than H3K36. Moreover our data suggest that NSD1 shares the H3K36 substrate with other methyltransferases (like SMYD2, SETD2, ASH1) but H1.5K168 is more specific for NSD1.

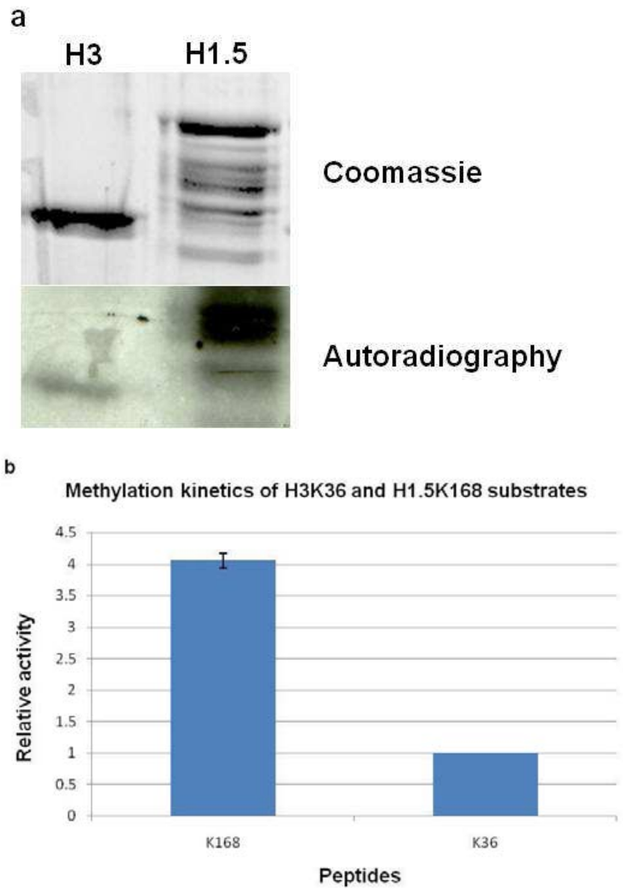


Figure11: H3K36 methylation vs H1.5K168 methylation. a) GST- H3 and GST-H1.5 proteins were incubated with NSD1 protein in presence of radio labelled Adomet and transfer of methyl groups analysed by autoradiography. b) H1.5K168 and H3K36 peptides were methylated in competition and the products were analysed by MALDI, relative areas of unmethylated and mono- methylated peaks were calculated and plotted.

4.1.9. NSD1 methylation on Non-histone targets

Since we identified and showed methylation at peptide level for several other non-histone targets of NSD1 along with the H3, H4 and H1 proteins, we sought to analyze whether NSD1 could methylate these target proteins at protein level or not. We selected 23 proteins which were equally or stronger methylated in comparison to the H3 peptide. Some of the putative target proteins are of more than 1000 amino acids, so we decided to clone the domains for these proteins containing the target site instead of full length proteins. Domain boundaries were predicted by the dompred site http://bioinf.cs.ucl.ac.uk/dompred. We amplified the DNA of the corresponding protein form cDNA prepared from HEK293 cells. Out of 23 proteins, we could amplify the PCR products for 19 proteins (table 1) and successfully cloned these into pGEX6p2 vector.

We proceeded further to express and purify the GST-fusion proteins, however, some of the proteins were expressed as inclusion bodies (IB), few proteins did not express at all and only 8 proteins expressed in soluble form. As the target sequence motif of NSD1 protein is hydrophobic, we initially suspected that the target proteins will be more hydrophobic and may express as inclusion bodies or the target lysine would be folded inside and not available for methylation. The 8 protein domains which were expressing in soluble form were purified and incubated with the NSD1 SET domain protein and the transfer of radio-labelled methyl group was detected by autoradiography. A strong methylation signal was observed only with the Transcriptional regulator ATRX and the Probable U3 small nucleolar RNA-associated protein 11 in comparison with histone 3. To confirm the methylation is occurring at the predicted lysine in these 2 non histone target protein domains, we mutated the predicted lysine into arginine by site directed mutagenesis and purified the corresponding proteins. The two identified wild type non-histone protein domains and the corresponding target lysine mutants were methylated with the NSD1 SET domain protein. We observed a clear methylation signal on both the wild type proteins and the loss of signal on the corresponding lysine mutants (figure 12). This confirms that the methylation is happening in the identified non-histone protein domains on the predicted lysine.

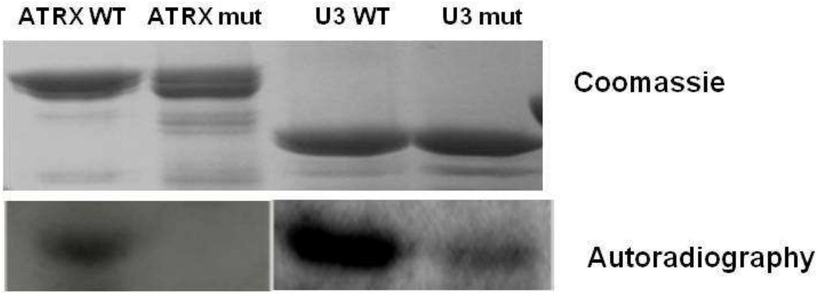


Figure12: Methylation of the non-histone targets of NSD1: Coomassie stained gel of the purified ATRX and Probable U3 small nucleolar RNA-associated protein 11 to show the equal loading of wild type and mutant proteins for methylation. Purified GST-tagged ATX and Probable U3 small nucleolar RNA-associated protein 11 wild type and lysine mutant proteins were incubated with NSD1 in presence of radio labelled Adomet and methylation of proteins was analysed by autoradiography.

S.No.	Protein Name	Expression
1	tRNA pseudouridine synthase A	IB
2	Activating signal cointegrator 1	soluble
3	RNApolymeraseIIIpolypeptideA(RPC1Human)	IB
4	Histone lysine N methyltransferase MLL4	No
5	Heat repeat containing protein 1	IB
6	Heterogenous ribonuclear protein L	Soluble
7	DualspecificityproteinphosphataseCDC14B	IB
8	Cullin 3	IB
9	M-phase inducer phosphatase3	No
10	U6snRNA-associatedSm-likeproteinLSm6	Soluble
11	Ran-binding protein 17	IB
12	RNA binding protein 12	Soluble
13	Transcription elongation factor SPT6	No
14	Transcriptional regulator ATRX	Soluble
15	DNA-dependent protein kinase catalytic subunit	No
16	Pre-mRNA-splicing regulator WTAP	No
17	Zinc finger 331	Soluble
18	r RNA protein EBP2	Soluble
19	Probable U3 small nucleolar RNA-associated protein 11	Soluble

Table1: Proteins selected for cloning and the details of expression

4.1.10. Sotos Mutants

Mutations in the NSD1 protein were reported to cause Sotos syndrome (Kamimura et al., 2003), Sotos mutations are present in all domains of NSD1 protein including the PHD and SET domains (Tatton –Brown et al., 2005). Sotos mutations in the PHD domains of NSD1 disrupt its binding to the methylated tails of histone H3 and also its interaction with the transcription cofactor Nizp1 and thus interfere with the transcriptional regulation of target genes (Pasillas et al., 2011). We sought to determine the influence of Sotos mutants of SET domain on NSD1 activity. We specifically selected 3 basic amino acids (arginine) and one hydrophobic amino acid (tyrosine) of all the Sotos mutation in SET doamin, speculating that they might be involved in the recognition of substrate side chains and in the binding of Adomet. We mutated the selected amino acids to those present in the Sotos patients and tested for their methylation activity on the H3 protein. Our data showed that all the four mutants completely lost their activity on H3 protein (figure 13). When we are preparing this manuscript similar results were published along with the crystal structure of NSD1 (Qiao et al., 2011) and showed in that R1952 and R1984 are involved in interactions with negatively charged Asp and Glu residues in the NSD1-CD (catalytic domain). R2017 plays a vital role in stabilising the confirmation of the 3 aromatic residues Y1870, Y1977 and F2018. However, Qiao et al., (2011) did not analyse all the Sotos mutants, here we have included the additional mutant Y1197C and showed that it also lost the methylation activity. Our data along with the others strongly support that the Sotos mutants of SET domain impair the activity on histone proteins and also binding to methylated histone tails and thus hints the probable epigenetic mechanism changes involved in the Sotos syndrome patients.

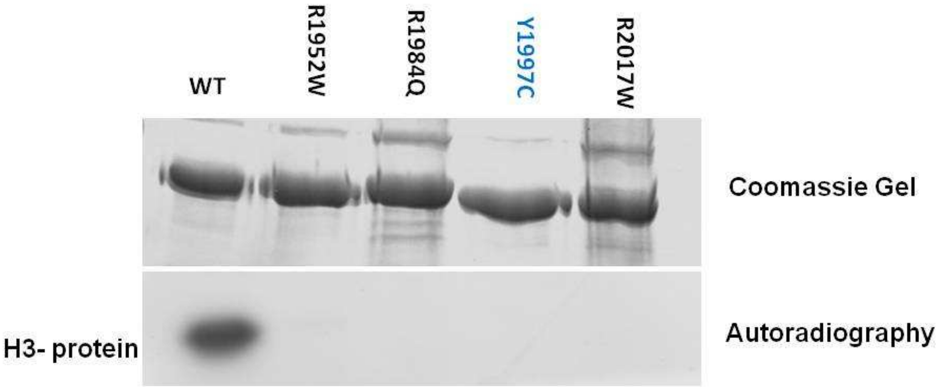


Figure13: Sotos mutants: Individual amino acids were exchanged to those present in the SOTOS mutant proteins. coomassie staining gel of Sotos mutated proteins along with wild type to show equal loading.
In-vitro methylation: Sotos mutant proteins were incubated with H3 protein in presence of radio labelled adomet and observed the transfer of radio labelled methyl groups by autoradiography.

4.1.11. Automethylation of NSD1

When we incubated the NSD1 protein with radio labelled Adomet for in vitro methylation assays we observed three radioactive bands appearing in autoradiography, one strong radioactive band corresponding to the GST-NSD1 protein size and the two weak bands corresponding to degradation products of GST-NSD1. This could suggest that the NSD1 protein is getting either automethylated like G9a (Chin et al., 2007 and Rathert et al., 2008)) and PRMT6 (Frankel et al., 2002) or it could bind to Adomet so strongly that the cofactor is not released during the SDS-PAGE.

To identify the target lysine of potential auto-methylation, we used a candidate screen approach and synthesised the entire catalytic domain of NSD1 protein as 20 amino acid length peptides and methylated this array with NSD1 protein. We observed weak methylation on few peptides and strong methylation on one peptide (figure 14a). We further selected all the methyalted peptides and made new peptides of 15 amino acid lengths by individually exchanged all lysines to alanine and also included H3K36 peptide as control. Upon incubation with NSD1 and radio labelled Adomet, we observed strong methylation of the H3K36 peptide and two other peptides of NSD1. Incorporation of radio labelled methyl groups was observed on peptide with the 1766-1780 sequence and no apparent signal was detected when K1769 was exchanged to alanine, which shows that the NSD1 protein is getting methylated at K1769 (fig. 14b). By the candidate screening approach we successfully identified the lysine getting methylated in NSD1 protein.

After obtaining this preliminary result from the peptide arrays methylation, we next sought to confirm this at the protein level. For this, we generated the NSD1-K1769R mutant protein by site directed mutagenesis and subsequently expressed and purified the corresponding protein. To confirm the loss of K1769 methylation at the protein level, we incubated both the wild type protein and the NSD1-K1769R mutant protein with radio labelled Adomet for 4 hours, followed by SDS-PAGE and autoradiography to analyse the incorporation of radio labelled methyl groups. As expected, we observed three bands with the wild type protein and major loss of the methylation signal was observed with NSD1 (K1769R) mutant protein. However, a minor signal was still detected with the mutant protein, this might be due to methylation of other lysines with lower efficiency (figure 14c). We also performed an in-vitro methyltransferase assay with wild type NSD1 and NSD1-K1769R mutant along with recombinant H3 protein in presence of radio labelled Adomet to assess whether the methylated lysine has any regulatory role in the activity of enzyme. The result shows that both the proteins methylate H3 protein to same extent, suggesting that the mutation of K1769 does not influence on the activity of enzyme. Since the methylated lysine is located on the N-terminal part of the pre SET domain, it may not influence the activity by the SET domain. However, we can not rule out the possible interaction with methyl specific binding proteins which may allosterically stimulating the activity of the NSD1 enzyme in cells.

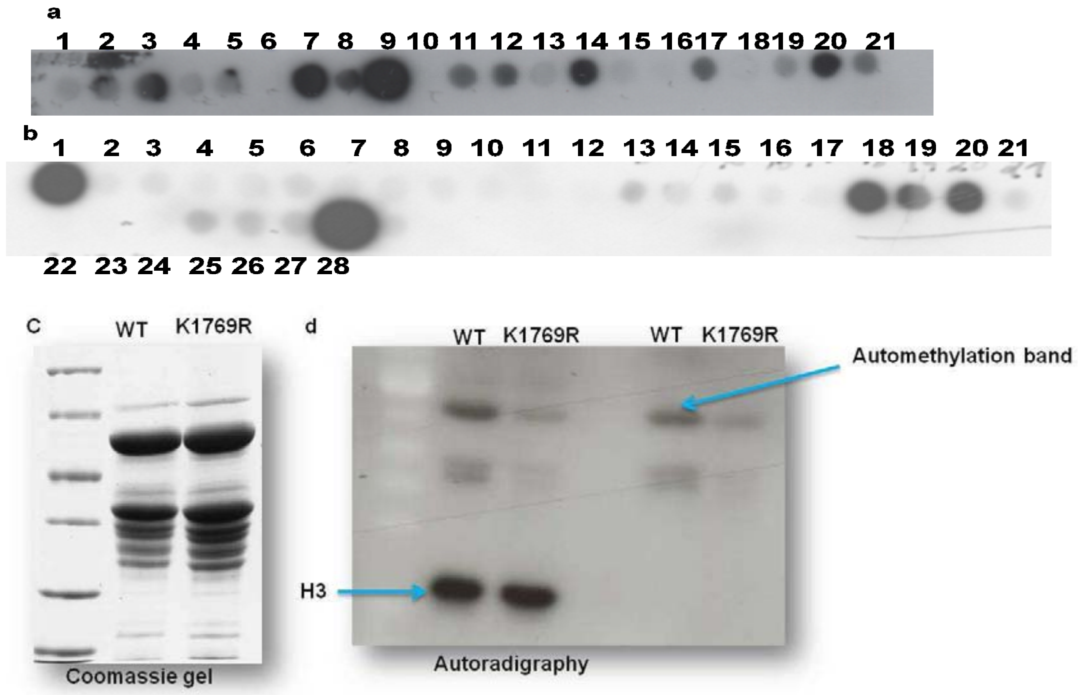


Figure14: Screening for the automethylation site of NSD1 a) Made a peptide scan of NSD1 catalytic domain by synthesising 20 amino acid length peptides and incubated with NSD1 protein in presence of radio labelled Adomet. Find peptide details below b) Based on initial peptide scan experiment, 15 amino acid length peptides were synthesised by individually exchanging each lysine to alanine to identify the target lysine. c) Coomassie stain gel of the wild-type protein and the mutant protein in which target lysine residue was exchanged to arginine. d) Autoradiography to observe the incorporation of radio labelled methyl groups in the
wild-type and mutant protein.

Peptide sequences of Figure 14a:

S.No	Peptide sequence	Methylation
1	R-N-H-E-H-V-N-V-S-W-C-F-V-C-S-E-G-G-S-L	
2	L-L-C-C-D-S-C-P-A-A-F-H-R-E-C-L-N-I-D-I	
3	I-P-E-G-N-W-Y-C-N-D-C-K-A-G-K-K-P-H-Y-R	
4	R-E-I-V-W-V-K-V-G-R-Y-R-W-W-P-A-E-I-C-H	
5	H-P-R-A-V-P-S-N-I-D-K-M-R-H-V-G-E-F-P-V	
6	V-L-F-F-D-Y-L-W-T-H-Q-A-R-V-F-P-Y-M-E-G	
7	G-D-V-S-S-K-D-K-M-G-K-G-V-D-G-T-Y-K-K-A	Yes
8	A-L-Q-E-A-A-A-R-F-E-E-L-K-A-R-K-E-L-R-Q	
9	Q-L-Q-E-D-R-K-N-D-K-K-P-P-P-Y-K-H-I-K-V	Yes
10	V-N-R-P-I-G-R-V-Q-I-F-T-A-D-L-S-E-I-P-R	
11	R-C-N-C-K-A-T-D-E-N-P-C-G-I-D-S-E-C-I-N	
12	N-R-M-L-L-Y-E-C-H-P-T-V-C-P-A-G-V-R-C-Q	
13	Q-N-Q-C-F-S-K-R-Q-Y-P-D-V-E-I-F-R-T-L-Q	
14	Q-R-G-W-G-L-R-T-K-T-D-I-K-K-G-E-F-V-N-E	Yes
15	E-Y-V-G-E-L-I-D-E-E-E-C-R-A-R-I-R-Y-A-Q	
16	Q-E-H-D-I-T-N-F-Y-M-L-T-L-D-K-D-R-I-I-D	
17	D-A-G-P-K-G-N-Y-A-R-F-M-N-H-C-C-Q-P-N-C	
18	C-E-T-Q-K-W-S-V-N-G-D-T-R-V-G-L-F-A-L-S	
19	S-D-I-K-A-G-T-E-L-T-F-N-Y-N-L-E-C-L-G-N	
20	N-G-K-T-V-C-K-C-G-A-P-N-C-S-G-F-L-G-V-R	Yes
21	G-A-P-N-C-S-G-F-L-G-V-R-P-K-N-Q-P-I-V-T	

Peptide sequences of figure 14b

S.No	Peptide sequences	Methylation
1	P A T G G V K K P H R Y R P G (H3K36)	Yes
2	P A T G G V A K P H R Y R P G (H3K36A)	No
3	N D C K A G K K P H Y R E I V	
4	N D C K A G A K P H Y R E I V	
5	N D C K A G K A P H Y R E I V	
6	G D V S S K D K M G K G V D G	
7	G D V S S A D K M G K G V D G	
8	G D V S S K D A M G K G V D G	
9	G D V S S K D K M G A G V D G	
10	G V D G T Y K K A L Q E A A A	
11	G V D G T Y A K A L Q E A A A	
12	G V D G T Y K A A L Q E A A A	
13	Q E D R K N D K K P P P Y K H	
14	Q E D R A N D K K P P P Y K H	
15	Q E D R K N D A K P P P Y K H	
16	Q E D R K N D K A P P P Y K H	
17	Q E D R K N D K K P P P Y A H	
18	P P Y K H I K V N R P I G R V	Yes
19	**P P Y A H I K V N R P I G R V**	No
20	P P Y K H I A V N R P I G R V	Yes
21	L R T K T D I K K G E F V N E	
22	L R T A T D I K K G E F V N E	
23	L R T K T D I A K G E F V N E	
24	L R T K T D I K A G E F V N E	
25	C L G N G K T V C K C G A P N	
26	C L G N G A T V C K C G A P N	
27	C L G N G K T V C A C G A P N	
28	P A T G G V K K P H R Y R P G (H3K36)	Yes
29	P A T G G V A K P H R Y R P G (H3K36A)	No

4.2. Specificity analysis of SUV39H1

4.2.1. Scientific background of SUV39H1

The SUV39H1 and SUV39H2 proteins are the mammalian homologues of Drosophila SU(VAR)3-9, responsible for suppressors of position effect variegation (PEV) in Drosophila and S. pombe. SUV39H1 was identified as a first mammalian histone lysine methyltransferase. It specifically methylates lysine 9 of histone3 (H3K9) and prefers mono- or dimethylated H3K9 as substrates. SUV39H1 trimethylates H3K9 and proved to be essential in the establishment of constitutive heterochromatin at pericentromeric and telomeric regions in the human genome. It consists of two vital evolutionary conserved domains of chromatin regulators, a chromo and a SET domain (fig1) (51, Rea et al., 2000), both domains are needed for its heterochromatic localisation and the SET domain for its methyltransferase activity.

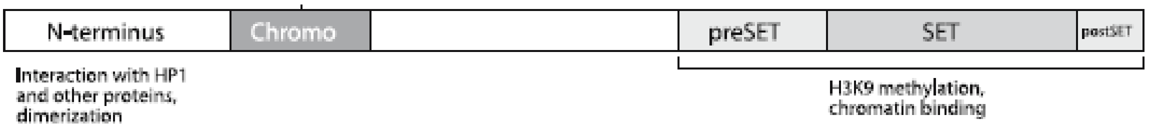


Fig.1: Functional motifs of SUV39H1 protein (adapted from Krauss 2008)

SUV39H1 methylation of H3K9 creates a binding site for HP1 (heterochromatin protein 1) – a family of adaptor molecules shown to be important for heterochromatic maintainence (Bannister et al., 2001). SUV39H1 also interacts with DNA methyltransferases (DNMT1 and DNMT3) (Fuks et al., 2003) and thus plays a functional role specific for constitutive heterochromatin both by enzymatic activity and also as a structural component (Schotta et al., 2003). SUV39H1 exists as a mega multimeric complex with other H3K9 methyltransferases (G9a/GLP, SETDB1) and also involved in regulating G9a target genes (Fritsch et al., 2009). Here, for the first time we showed, SUV39H1 could also methylates non-histone target proteins apart from Histone H3K9.

4.2.2. Specificity analysis of SUV39H1

To determine the target specificity of SUV39H1, peptide arrays synthesized on cellulose membranes by employing SPOT synthesis were utilized. Since SUV39h1 methylates lysine 9 on histone H3, the first 20 amino acids of histone H3 were used as a template to prepare the arrays in which each residue was replaced with each of the 20 amino acids. Thus, a total of 389 peptides were synthesized on a single membrane and incubated with the enzyme in the presence of radioactively labeled Adomet and the transfer of methyl groups to the immobilized peptides were detected by autoradiography. The results show that SUV39H1 has a distinct profile to its counterpart euchromatic H3K9 methylatransferase G9a (Rathert et al., 2008). like G9a it also showed high specificity to arginine at 8th position and lysine at 9th position, replacing any other amino acid at these positions completely abolished the activity of enzyme on H3 (1-20) peptide substrate. Apart from arginine and lysine at 8th and 9th position, lysine at 4th position is an important specificity determinant for the SUV39H1. Any other amino acid substituted at that position completely abolished the activity of SUV39H1 on peptide substrate. This specific recognition of lysine at 4th position explains why SUV39H1 is specific for H3K9 and could not methylate H3K27 which also has an 'ARK' unlike G9a. This result further suggests that both SUV39H1 and G9a could have distinct non-histone substrate proteins. Threonine at 6th position was important but it could be substituted with other amino acids like serine, phenylalanine, isoleucine and alanine as well. Serine and threonine at 10th and 11th position correspondingly were important but they could be substituted with several other amino acids with out loss of activity. All other adjacent residues on H3 are not important for the specificity of SUV39H1 as they could be exchanged to almost any other amino acid. In summary, SUV39H1 specifically recognised an Arg-Lys dipeptide together with lysine at 4th position. It also recognised 6th, 10th and 11th positions in H3 tail sequence, but at these positions they could tolerate few other amino acids. This further suggested that SUV39H1 might methylate other non-histone substrate proteins.

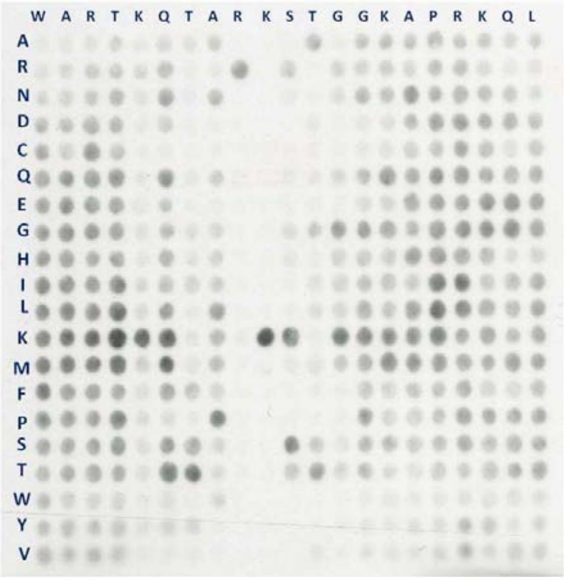
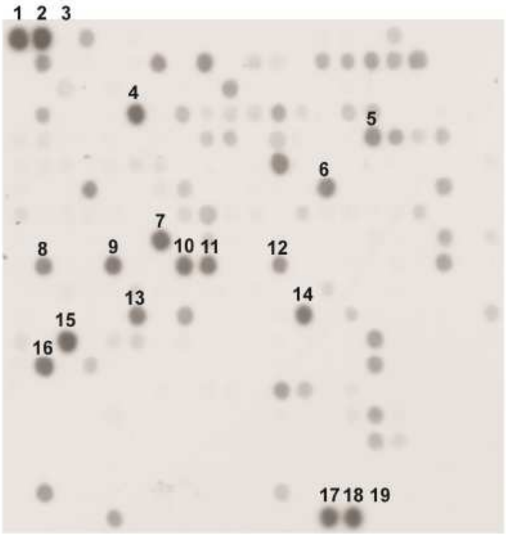


Figure 2: Example of specificity analysis of SUV39H1: Peptide arrays were synthesised with H3(1-20) sequence as a template and individually exchanged each amino acid at each place with all the naturally available amino acids and then incubated the membrane with SUV39H1 in presence of radioactively labeled Adomet and the transfer of methyl groups was measure by autoradiography.

With the identified 'RK' motif, we did a scansite search (http://scansite.mit.edu/) and looked for non-histone proteins with nuclear localisation. It revealed a large number of proteins containing the target sequence motif. To narrow it down, we retrieved the known interaction partners of SUV39H1 from the Human Protein Reference Database (http://www.hprd.org/). Currently, with the in vitro and in vivo studies there are about 40 known interaction partners for SUV39H1 including the DNMT1, DNMT3a, Histone deacetlyase and many other interesting proteins. Furthermore, we have also looked at the interaction partners of SUV39H1 interactors which might form complex with SUV39H1 indirectly. Altogether, we identified 276 target proteins containing Arg-Lys sites, some proteins possess more than one site. We altogether synthesized 415 peptides of 20 amino acid length with target sequence motif along with H3 (1-20) and H3K9A (target lysine exchanged to alanine) mutant peptides as a control on cellulose membrane and tested for methylation by SUV39H1. Out of the 415 potential target peptides, 13 peptides got methylated in par with the H3 tail and 27 peptides were methylated weakly in comparison to H3 (1-20) peptide (figure 3).



1. H3 - ARTKQTARKSTGGKAPRKQ
2. H3 - ARTKQTARKSTGGKAPRKQ
3. H3K9A - ARTKQTARASTGGKAPRKQ

4. Deoxyribonuclease gamma - AFTNSKKSVTLRKKTKSKRS
5. Jumonji protein - SKPTPKRGPRKRATVDVPPS
6. Serine/Threonine protein kinase Nek3 - IKNSKHNTPRKKTNPSRIRI
7. Recombination Activating Gene 2 (Rag2) - LKKPPMKSLRKKGSGKILTP
8. Sex comb on midleg-like protein 2 (Sex) - SSSVKNITPRKKGPNSGKKE
9. SET8 - KPIKGKQAPRKKAQGKTQQN
10. Splicing factor arginine/serine-rich 8 - RVSRSPGASRKRTRSRSPHE
11. Structural maintenance of chr-1-like protein - ETSDEFEAARKRAKKAKQAF
12. Paired amphipathic helix protein Sin3b - EHDKTPEHSRKRSRPSLLRP
13. U3 small RNA-interacting protein 2 - MSATAAARKRGKPASGAGAG
14. YL-1 protein - MSLAGGRAPRKTAGNRLSGL
15. Zinc finger protein 330 - MPKKKTGARKKAENRREREK
16. Zinc finger and homeobox protein 1 - SSMNGLSSLRKRGRGRPKGR

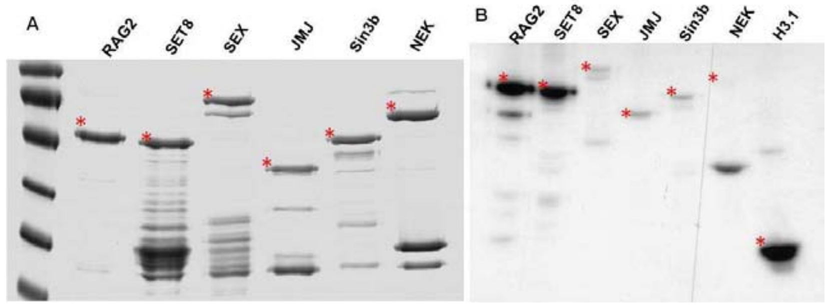
17. H3 - ARTKQTARKSTGGKAPRKQ
18. H3 - ARTKQTARKSTGGKAPRKQ
19. H3K9A - ARTKQTARASTGGKAPRKQ

Figure 3: Methylation of potential non-histone target peptides of Suv39h1 identified through the Human Protein Reference Database protein interaction data. Numbers on top of individual spots indicate the strongly methylated peptides plus control peptides. A legend is given on the right pane with the protein name and sequence of the specific peptide at each numbered spot. The expected target lysines (always next to an arginine) are highlighted in red.

4.2.3. Methylation of Potential Non-histone substrates at protein level

After confirming methylation of potential targets at the peptide level, we proceeded further to show the methylation at protein level which is more important physiologically. Nine potential target proteins were selected based on their high methylation at the peptide level. We identified the domain boundaries for the corresponding protein via domain prediction web programme (http://bioinf.cs.ucl.ac.uk/dompred). We successfully cloned 6 out of 9 protein target domain into GST expressing vector. The six protein domains in fusion with GST were successfully overexpressed and purified by affinity chromatography. The methylation at the protein level was analysed by incubating the purified target protein domains with SUV39H1 in methylation buffer with radioactively labeled Adomet. The transfer of radioactive methyl groups to target proteins were analysed by separating the proteins on SDS-PAGE and then subjecting to autoradiography. Out of the 6 protein domains, 2 proteins – VDJ (RAG2) and SET8 proteins got strongly methylated while Jumonji and Sex comb on midleg protein 2 got weakly methylated. No methylation signal was detected on the Paired amphipathic helix protein (Sin3b) and Serine/Threonine protein kinase Nek3 (figure 4).

S.No	Protein name	Domain boundaries	Target lysine
1	RAG2	311-520	K507
2	SET8	1-228	K169
3	Sex comb on midleg protein 2	240-618	K308
4	Jumonji protein	1075-1245	K1222
5	Paired amphipathic helix protein (Sin3b)	111-324	K268
6	Serine/Threonine protein kinase Nek3	254-500	K291



Coomassie stain Autoradiography

Figure 4: In vitro methylation of non-histone target proteins by SUV39H1: Purified protein domains were incubated with SUV39H1 in presence of radioactively labeled Adomet and separated on SDS-PAGE, and the transfer of radio labelled methyl groups were observed by autoradiography. Asterisk mark indicates the target protein size

To confirm the identified non-histone target proteins were getting methylated at the predicted lysine, we exchanged the predicted target lysine to arginine by site direct mutagenesis. Subsequently, we incubated the target lysine mutated protein and the wild type protein with SUV39H1 in presence of radio labelled Adomet to analyse the methylation (figure 5). Indeed, the methylation signal was completely lost for each of the four non-histone protein domains when the target lysine was mutated to argnine suggesting that the predicted lysine always in the context of an R-K motif were specifically getting methylated. Methylation on the RAG2 protein was also confirmed by Mass specteometry, MALDI analyses confirms the methylation on the predicted target lysine and further, it also shows that the addition of 3 methyl groups on the target lysine (figure 7).

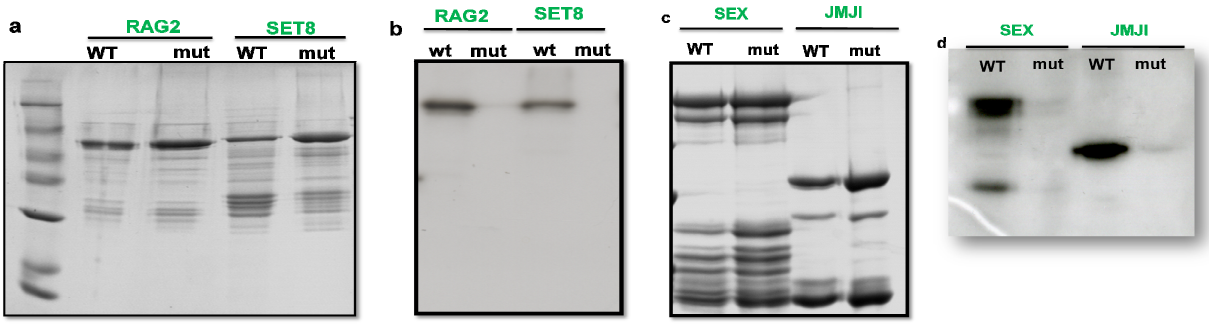


Figure 5: Methylation of SUV39H1 targets in vitro and identification of target lysine methylation site. a) and c) are the coomasie staining of GST-SUV39H1 wild type targets and the mutant proteins in which the target lysine exchanged to arginine, coomasie gel represents the loading control of the methylated proteins. b) and d): In vitro methylation, GST-SUV39H1 was incubated with wild type target proteins and mutated proteins in presence of radioactively labeled Adomet and the transfer of radio labelled methyl groups was detected by autoradiography

4.2.4. Degree of Methylation

Next we sought to determine the degree of methyaltion of the prominently methylated non-histone targets RAG2 and SET8. SUV39H1 is known to trimethylate H3K9. RAG2 and SET8 were subjected to in vitro methylation reaction by SUV39H1 followed by detection using a pan methyl lysine specific antibody which had been validate with cellulospot peptide arrays as trimethyl lysine specific antibody (data not shown). The results showed that both RAG2 and SET8 were trimethylated by SUV39H1 (figure 6), methylation signal was more on the RAG2 protein than SET8 which is in consistent with the radioactive methylation experiments (figure 4). In an independent experiment by mass spectrometry, we also showed that RAG2 protein is getting both di- and tri-methylated at the target lysine by SUV39H1 (Figure 7), though the major peak is of tri-methylation. However, in MALDI we could not detect the peptide peak containing the target lysine with both methylated and unmethylated samples of SET8. Collectively results from the two experiments showed that the SUV39H1 is adding three methyl groups to the target lysines in non-histone target proteins.

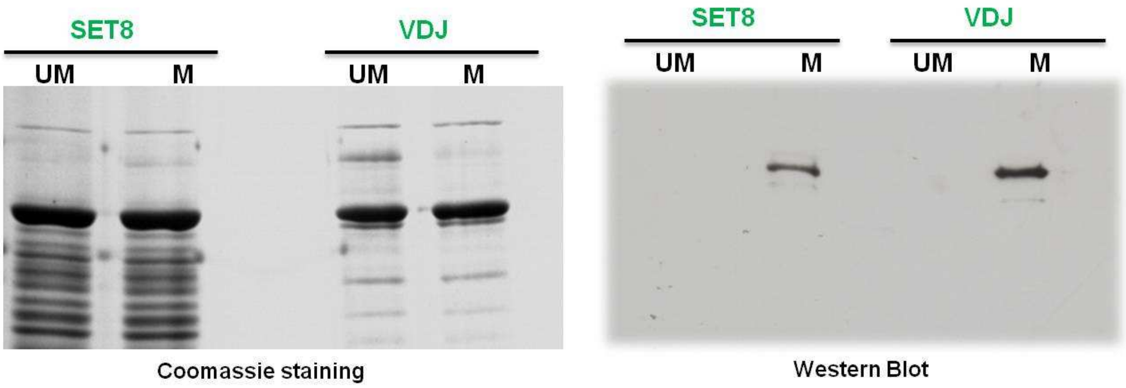


Figure 6: Detection of methylated proteins by PAN methyl antibody: PAN methyl antibody specifically recognised in vitro methylated SET8 and RAG2 proteins. Coomassie gel shows the loading controls of methylated and unmethylated proteins.

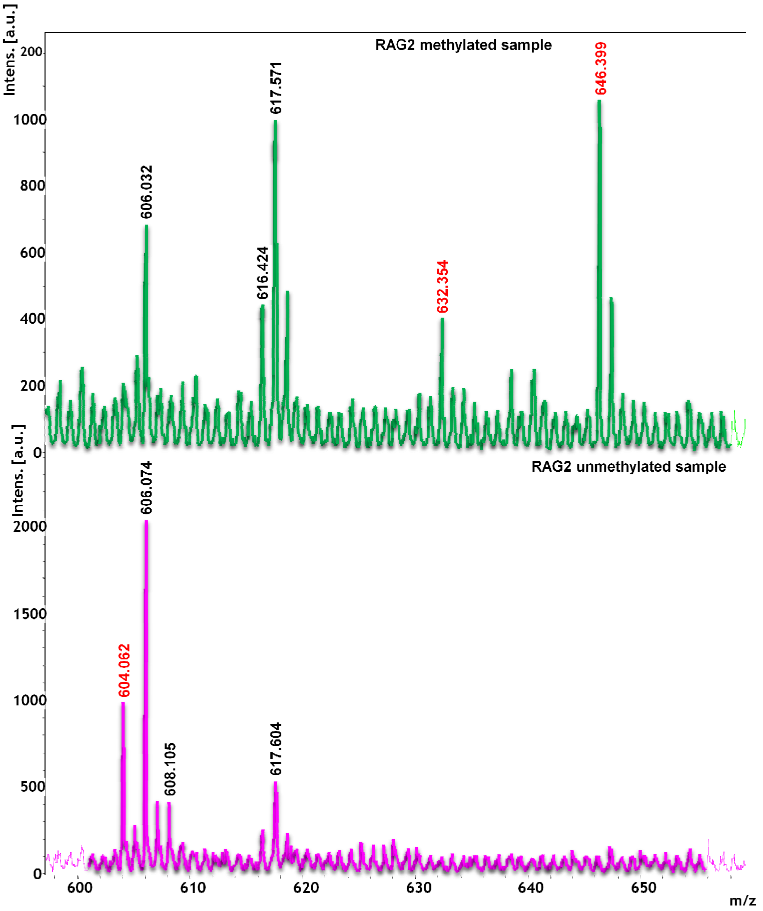


Figure 7: SUV39H1 trimethylates RAG2 at K507 in vitro: RAG2 protein was incubated with and without unlabelled SAM in presence of SUV39H1 and subjected to MALDI analysis after in gel trypsin digestion. Upper panel represents trypsin digested peaks from the methylated sample, lower panel represents peaks from un-methylated sample. Numbers in red indicates the mass of interested peptide peaks. Calculated masses of unmethylated peptide (KKGSGK)- 604.362 , di-methylated peptide (632.409) and tri-methylated peptide (646.425)

4.2.5. Specific recognition of Lysine at -4 position

According to the histone code hypothesis, distinct histone modifications on one or more tails act sequentially or in combination to form a histone code (Strahl and Allis, 2000). It has been shown that the SET7/9 methylation at H3K4 leads to transcriptional activation, while SUV39H1 and G9a methylation at H3K9 leads to transcriptional repression. This suggests that there exists interplay between H3K4 and H3K9 methylation. Nishioka et al (2002) showed that methylated H3K4 drastically decreased the ability of SUV39H1 to methylate K9, whereas H3K4 methylation did not influence the activity of G9a on H3K9. Our specificity profile analysis also shows that lysine (H3K4) at -5 position is very important for SUV39H1 to recognise the substrate, substitution of any amino acid at this position completely abolished the activity of the enzyme. This intrigued us to check whether the same phenomenon could be applied to the non-histone targets of SUV39H1.The newly identified SUV39H1 targets which got strongly methylated; SET8 and RAG2 also contains lysine at -5 (K164) and -4 (K503) positions with respect to target lysine respectively (figure 8a), the Jumonji and Sex comb on midleg protein 2 which got weakly methylated also has lysine at -5 (K1217) and -6 (K302) positions respectively (figure 6a).

To confirm the importance of lysine at -5 or -4 position, we selected the two strongly methylated target proteins and mutated the K164 in SET8 and K503 in RAG2 to alanine by site directed mutagenesis while keeping the target lysine unchanged. The -5K or -4K mutated proteins were purified and analysed for methylation with SUV39H1 along with wild type and target lysine mutated proteins. As shown before we did not see incorporation of radioactivity on the target lysine mutant but in addition, we observed 70 to 80% loss of methylation signal on the -5K or -4K mutated proteins (figure 8b). Taken together these results strongly suggest that the like K4 in histone H3, K164 of SET8 and and K503 of RAG2 are important for the SUV39H1 to act on the target lysine.

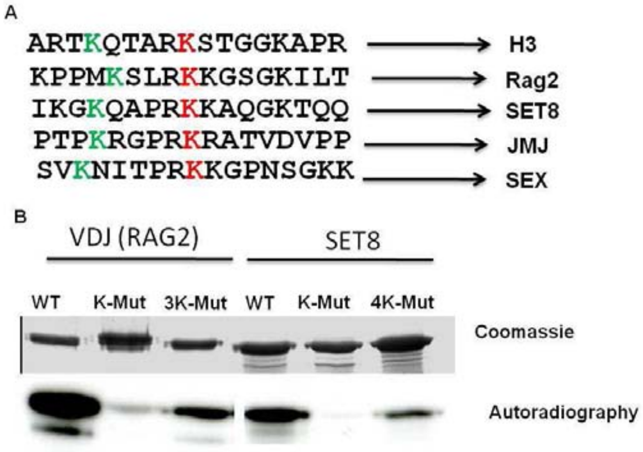


Figure 8: Recognition of -4 or -5 lysine in SUV39H1 target proteins. a) Alignment of identified SUV39H1 targets RAG2 and SET8 with histone 3, Lysine in red represents the target lysine site for methylation and lysine in green represents the -5 or -4 lysine in SUV39H1 substrates.
b) In vitro methylation of RAG2 and SET8 with SUV39H1 protein in presence of radio labelled Adomet. Coomassie- stained control gel of the proteins illustrates equal loading on the left side and right side is the autoradigraphy image to show the incorporation of radio labelled methyl groups into proteins.

4.2.6. Cellular methylation of SUV target proteins

After confirming the methylation of SUV39H1 non-histone targets in vitro, we sought to check the methylation of these target proteins in the cells, for this we narrowed down to two target proteins (Rag2 and SET8) which got highly methylated in vitro. For cellular experiments we attempted to clone both the target proteins in full length. Eventually we were only successful with the SET8 protein. The RAG2 protein has NLS at the C-terminal end and the previous studies from other groups have also failed in expressing the full length proteins since it is subjected to heavy degradation, nevertheless they also showed that a RAG2 C-terminal domain which includes the target lysine is sufficient for its nuclear localisation and activity (Grundy et al., 2010).

To assess if the RAG2 and SET8 are getting methylated in vivo, we co-transfected each target protein and the corresponding target lysine mutants individually in HEK293 cells together with the SUV39H1. After 48h of transfection, purified the target proteins by GFP trap (Chromtek) followed by western analysis using anti-pan-methyl-lysine antibody. The results showed that the pan-methyl specific antibody recognised wild type SET8 protein overexpressed with SUV39H1 but not the corresponding target lysine mutant (figure 8). This evidence supports that the SET8 protein was getting methylated in cells at the predicted target lysine by SUV39H1. But we did not see any signal on either wild type RAG2 protein or its target lysine mutant, this could be attributed to the poor expression of RAG2 protein in cells.

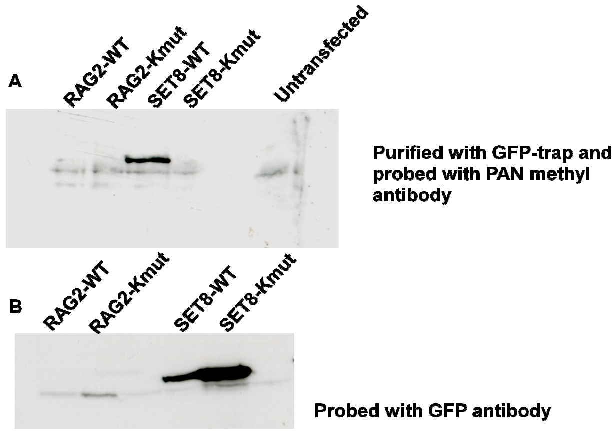


Figure 8: In vivo methylation of novel targets at the predicted lysine by SUV39H1: a) GFP tagged wild type and mutant proteins were over expressed together with SUV39H1 protein in HEK293 cells. GFP tagged proteins were purified by GFP-trap and then subjected to western blot analysis. To detect lysine methylation PAN methy specific antibody was used. Methylation signal was detected on SET8 wild type protein and loss of signal on the corresponding lysine mutant. b) Purified proteins via GFP trap were probed with GFP antibodies to show the loading control.

4.2.7. Sub-nuclear localization of RAG2 protein

Basic amino acid residues play a vital role in the nuclear localisation of proteins, either mutations or modifications on these residues could influence the sub-cellular localization of proteins (Corneo et al., 2002). For RAG2 protein, the lysine (K507) residue which was getting methylated by SUV39H1 is present at its nuclear localisation signal. Therefore we speculated that the trimethyl modification on the corresponding lysine might change the localization pattern of RAG2 protein. To examine this we cloned both the wild type and K507 mutant of RAG2 protein domain into YFP (pEYFP-C1) and CFP (pECFP-C1) vectors correspondingly and also the SUV39H1 full length protein was cloned into both CFP and YFP vectors. NIH3T3 cells were treated with wild type and mutant RAG2 protein domains separately, showing that the wild type RAG2 protein is localised in the nucleus with a speckled distribution, while the K507- RAG2 mutant unlike wild type protein was showed uniform distribution in the nucleus (figure 9a). To further understand the influence of methylation by SUV39H1 on K507 of RAG2 protein we co-expressed both the SUV39H1 and RAG2 wild type protein in NIH3T3 cells. The results demonstrate that the wild type RAG2 protein instead of spotty appearance showed uniform distribution in the nucleus after co-expression with SUV39H1 and we also observed little amount of RAG2 protein in the cytoplasm. However, we did not observed any changes in the localization pattern of RAG2K507 mutant when co-expressed with SUV39H1 (figure 9b). Collectively these results show that the K507 is vital for the nuclear localisation of the RAG2 protein. Either mutation of this residue or additional modifications on this residue could severely impair the localization of RAG2 protein at nuclear spots and thus might influence its functional role in cells.

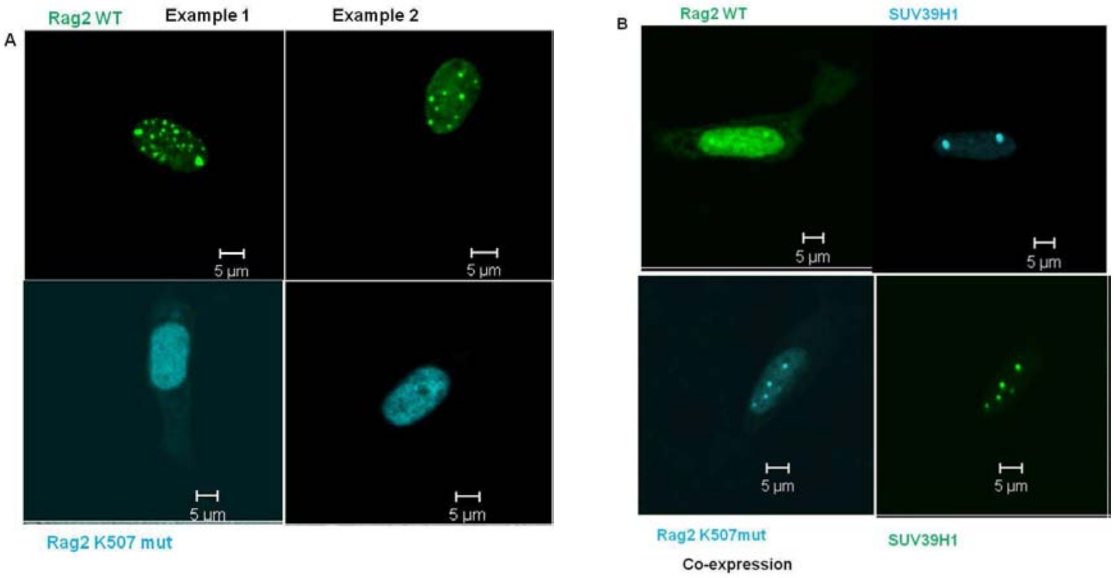


Figure 9: Sub-cellular localization studies of RAG2 protein in NIH3T3 cells: a) RAG2 protein (green) and Rag2K507 mutant protein (blue) was transfected individually in NIH3T3 cells, RAG2 wild type protein exhibits spotty appearance and the Rag2K507 mutant show diffused appearance in nucleus. b) RAG2 protein co-expressed with SUV39H1; When wild type Rag2 protein (Green) co-expressed with SUV39H1 (blue) shows diffused localisation, while SUV39H1 goes to heterochromatic spots. Where as RAG2K507 mutant protein (Blue) shows no changes in the localisation pattern when co-expressed with SUV39H1 (Green)

4.2.8. JMJD2A tandem Tudor Domain Binding

To understand the functional role of lysine methylation on novel targets, we sought to screen the binders that could specifically identify trimethylation marks on these targets. For this we synthesised different unmethylated and tri-methyl lysine anlalog peptides of H3 (1-15), RAG2 and SET8 on cellulose membrane and probed with several known GST-tagged tri-methyl lysine reading domains. Given the similarity of the residues surrounding H3K9 and the newly identified SUV39H1 targets and the degree of methylation we screened with HP1 and ATRX ADD domain domain containing proteins which were shown to interact specifically with H3K9 trimethylation marks. In both the cases we observed specific interaction to tri-methyl H3K9 peptides, but did not observe any binding with the novel tri-methyl target peptides.

The JMJD2A tandem tudor domain which previously shown to interact with H3K4, H3K9 and H4K20 trimethylation marks (Kim et al., 2007), recognises trimethyl marks in completely different flanking sequences suggesting that the JMJD2A tandem tudor domain is a promiscuous tri-methyl binder on histone proteins, but its methyl sepecific interaction on non-histone proteins is yet to explore. To examine if trimethylation on novel targets is able to mediate this binding process, we probed the methylated and unmethylated peptides of

SUV39H1 target proteins and H3 protein. As expected JMJD2A interacted strongly with the tri-methyl peptides of H3K9, RAG2 and SET8 and apparently no binding was observed to the corresponding unmethylated peptides (figure 10). Since JMJD2A is a histone demethylase and exists in complex with histone deacetylase complex (Zhang et al., 2005 and Gray et al., 2005) and functions as a transcription repressor, it is further interesting to study the function of methyl specific interaction on novel SUV39H1 targets, however, we did not pursue further in this study.

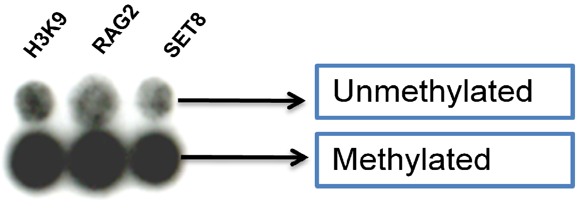


Figure 10: JMJD2A tandem Tudor Domain binding to unmethyalted and methyalted peptide analysed using peptide arrays. The array was incubated with JMJD2A and probed with GST antibody and the signal was detected by ECL method. JMJD2A specifically binds to trimethylated peptides.

4.3. Specificity analysis of SET8

4.3.1. Scientific Background of SET8

Modification of Histone tail peptides by lysine methylation is an important signal involved in gene regulation, chromatin structure and cell cycle (Chi, 2010). The SET8/Pr-Set7/KMT5a protein lysine methyltransferase (PKMT) enzyme is responsible for the methylation of lysine 20 in histone H4 (H4K20). It was one of the first histone lysine methyltransferase to be biochemically purified and identified (Nishioka, 2002; Fang, 2002). Like most HKMTs (Cheng, 2005), SET8 contains a SET domain which harbours the active center and is active in isolated form [Yin, 2005]. SET8 has been shown to interact with 4-5 amino acid residues on either side of K20 (Yin, 2005) including side chain contacts to R17, H18, R19, L22 and R23 seen in structural analysis of enzyme-peptide complexes (Cotoure, 2005; Xiao, 2005). Different studies demonstrated that it specifically monomethylates K20 (Yin, 2005; Xiao, 2005; Cotoure, 2005; Couture, 2008].

While trimethylation of H4K20 is related to heterochromatin formation and gene repression (Schotta, 2004), SET8-dependent H4K20 monomethylation plays a role in cell cycle progression, with increasing prevalence in late S-phase and highest during mitosis (Rice, 2002; Houston, 2008) and it stimulates the assembly of pre-replication complexes on origins during late M- and G1-phases (Tardat, 2010). Lack of SET8 methyltransferase activity leads to cell cycle arrest in G2 and but also loss of genomic stability including global centromere condensation failure and DNA damage (Houston, 2008; Tardat, 2008). SET8 is an essential gene in Drosophila and mice, deletion of which causes early embryonic lethality (Nishioka, 2002; Oda, 2009). Deletion of H4K20me1 also led to reduction of H4K20me2 and me3, modifications which are introduced by Suv4-20h1 and h2 (Schotta, 2005), suggesting that H4K20me1 is the preferred substrate for them, which may explain the effect of SET8 on genome condensation.

Recent studies have shown that SET enzyme including Set7/9, G9a and SET8 also methylate non-histone proteins (see for example: Chuikov, 2004; Kouskouti, 2004; Couture, 2006; Rathert, 2008; Dhayalan, 2010, and references therein), suggesting that lysine methylation is a common and widespread post translational modification with variable biological roles (Huang, 2008; Rathert, 2009). SET8 was shown to methylate K382 in C-terminal domain of the tumor suppressor p53 (Shi, 2007). We were interested to explore the possiblitiy that SET8 might methylate additional proteins. Following a strategy that we successfully employed previously (Rathert, 2008; Dhayalan, 2010), we determined the specificity profile of SET8 by methylation of many different peptides synthesized by the SPOT method. We show that SET8 is a very specific enzyme which recognizes the longest peptide motif identified so far for PKMTs (G9a, SET7/9). Based on this profile, methylation substrates were prediceted and their target peptides synthesized and methylated. 22 peptides were shown to be strongly methylated. The corresponding protein domains were cloned, expressed and purified and afterwards subjected to methylation by SET8. We found none of the identified targets got methylated at the protein level. However, we see the methylation of p53 protein by SET8 reported by others (Shi et al., 2007), which was weaker than methylation of H4K20. We conclude that the long substrate binding cleft of SET8 makes it difficult to methylate a folded protein suggesting that Histone H4 is the main celluar substrate of SET8.

4.3.2. Specificity profile of SET8

We cloned the catalytic SET domain of SET8 from the cDNA derived from human HEK293 cells as GST fusion construct and verified its sequence. The enzyme was overexpressed in *E. coli* and it could be purified with very good yield (figure 1A). The SET8 methyltransferase activity and specificity was investigated by methylating histone tail peptides synthesised on cellulose membranes. The membranes were then incubated with SET8 in the presence of radioactively labelled [methyl-^{3}H]-*S*-adenosyl-L-methionine (AdoMet), and the transfer of methyl groups to the immobilized peptides was detected by autoradiography. Among all Histone tails, SET8 specifically methylated K20 on the H4(10-30) peptide (figure 1B).

To study the influence of each residue on peptide recognition by SET8, an alanine scanning experiment was performed by synthesizing a small H4 (10-30) array of 21 peptides each of them carrying an exchange of a single residue against alanine (figure 1C). The reduced methylation of peptides carrying substitutions at positions 17-23 demonstrated an important role of the corresponding residues in the peptide recognition by SET8. D24 also had a somehow weaker effect similar to R19 which also plays a less important role in peptide recognition by SET8. The result that the SET8 has a strong specificity for H4K20 and it contacts several residues in the substrate peptide is in accordance with the crystal structure analyses of SET8 with H4 peptide which showed that the substrate is inserted into a deep binding cleft with close interactions of the enzyme with residues from R17 to R23 [Cotoure, 2005; Xiao, 2005].

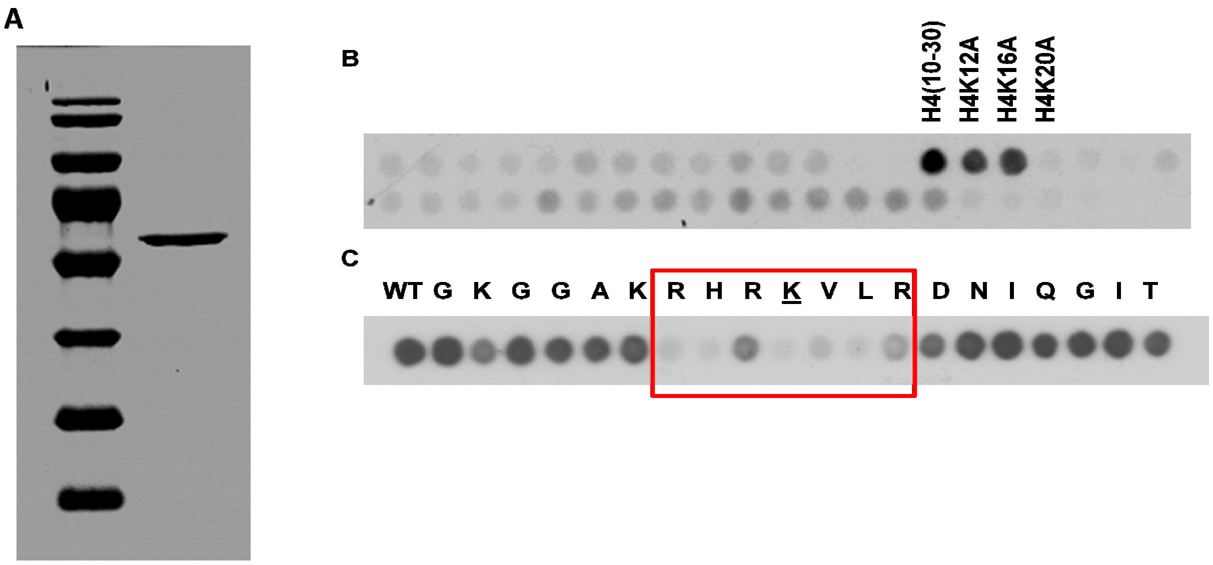


Figure 1: Purification and specificity analysis of SET8: a) GST-SET8 (114-352 AA) protein was expressed and purified. b) SET8 specificity was examined on all the characterized lysine sites on histone proteins, autoradiography image shows SET8 was very specific towards H4K20. c) Alanine scan of H4(10-30) methylation by SET8: here, all the 20 amino acids of H4 tail were individually exchanged to alanine and the first peptide labelled with WT was with the native H4 (10-30) sequence

To investigate the peptide recognition of SET8 in more details, we investigated the recognition of each amino acid residue in the substrate peptide by methylation of an H4 (10-30) tail array comprising 420 individual peptides in which each peptide contained an exchange of one amino acid of the wild type H4 tail sequence against any of the 20 natural amino acids (figure 2). Three independent membrane arrays were synthesized and methylated yielding basically very similar results. SET8 interacts with H4 residues from K16 to D24 with strongest specificity towards R17, H18, L22 and R23. The exact peptide motif recognized by SET8 can be defined as:

Position	-3 R17	-2 H18	-1 R19	0 K20	+1 V21	+2 L22	+3 R23
Preference	R	H	RK>other residues	K	I>V>YFL	L>FY	R>other residues

In addition to K20 (the target methylation site), R17 and H18 are very important specificity determinants for SET8 for substrate recognition. Any other amino acid introduced at that positions completely abolished the activity of SET8. SET8 equally accepts K and R at position 19, followed Y and other hydrophobic amino acids like L, H and I. Residues on the C-terminal side of K20 also play important roles in the specificity of SET8. The enzyme accepts majorly I at the +1 position followed by V, the natural amino acid at this position. Other hydrophobic amino acids like F, Y and L gave a weaker methylation signal. SET8 exhibits strong specificity on L22 and R23, enzyme showed residual methylation activity when these amino acids were exchanged with the hydrophobic amino acids, but the major activity was showed only with the native amino acids of H4.

In general, these findings are in nice agreement with the structural data: the guanidino group of R17 forms several hydrogen bonds to the enzyme, the side chain of H18 is contacted by the 3' hydroxyl group of the cofactor. R19 is contacted by a salt bridge to Glu259, which may also meditate an interaction with a K introduced at position 19. L22 is positioned in a hydrophobic pocket, which explains the specific readout at this position. V21 is in hydrophobic contact to F275 and R23 forms a hydrogen bond to the backbone of G337.

4.3.3. SET8 methylation on Celluspot arrays

Recently it has been showed by our lab that celluspot peptide arrays can be effectively used to characterize antibodies (Bock et al., 2011) and also reading domains (Dhayalan et al., 2011). In this study we employed the same celluspot peptide arrays comprising 384 peptides from 8 different regions of the N-terminal histone tails, viz. H3 1-19, 7-26, 16-35 and 26-45, H4 1-19 and 11-30, H2A 1-19 and H2B 1-19, featuring 59 post-translational modifications (most of them identified, some of them hypothetical) in many different combinations, which are commercially available from Active Motif. Results of SET8 methylation on celluspot arrays demonstrates that the only H4 (1-19) peptides were methylated (figure 3). The strongest methylation signal was observed with unmodified H4K20 and on K12acetylated H4 (1-19) peptide. In our specificity analysis (figure 2) SET8 exhibited strong specificity towards R^{17}, in coherent to this, either symmetric or asymmetric methylation on R^{17} completely inhibited the methylation on H4K20 whereas the modifications on K^{16}, R^{19} and R^{23} partially reduced the methylation on H4K20. The results of celluspot arrays were in strong agreement with our peptide array specificity analysis. Collectively, the data suggests that either exchange of amino acids adjacent to target lysine or their posttranslational modification impairs the activity of SET8 on H4K20.

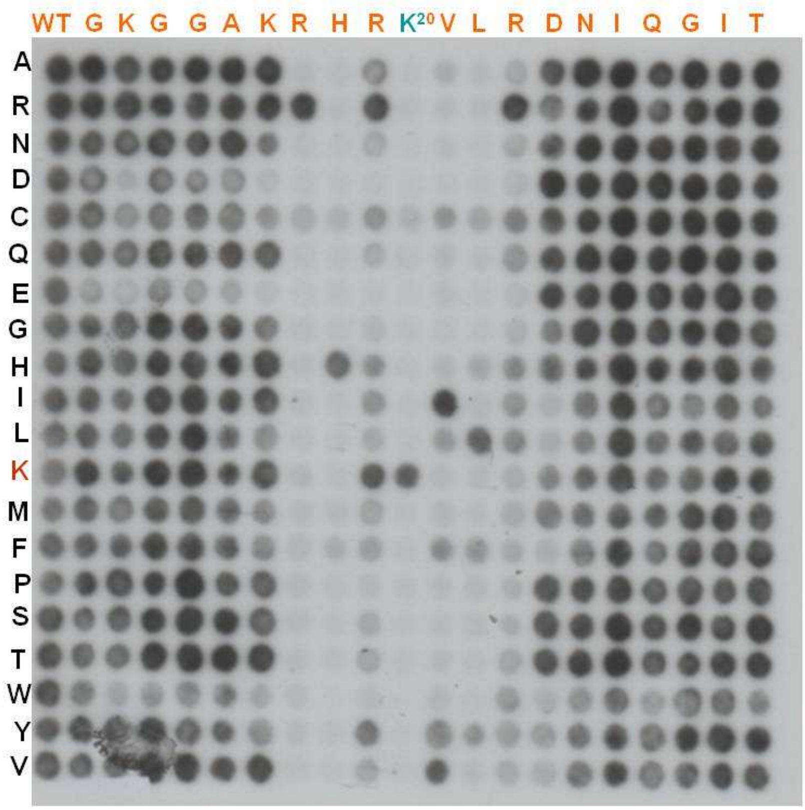


Figure 2: Specificity of Peptide Methylation by SET8: a) Example of one full H4 peptide tail array. The sequence of the H4 tail is given on the horizontal axis. Each residue was exchanged against all 20 natural amino acid residues (as indicated on the vertical axis) and the relative efficiency of methylation by SET8 analyzed

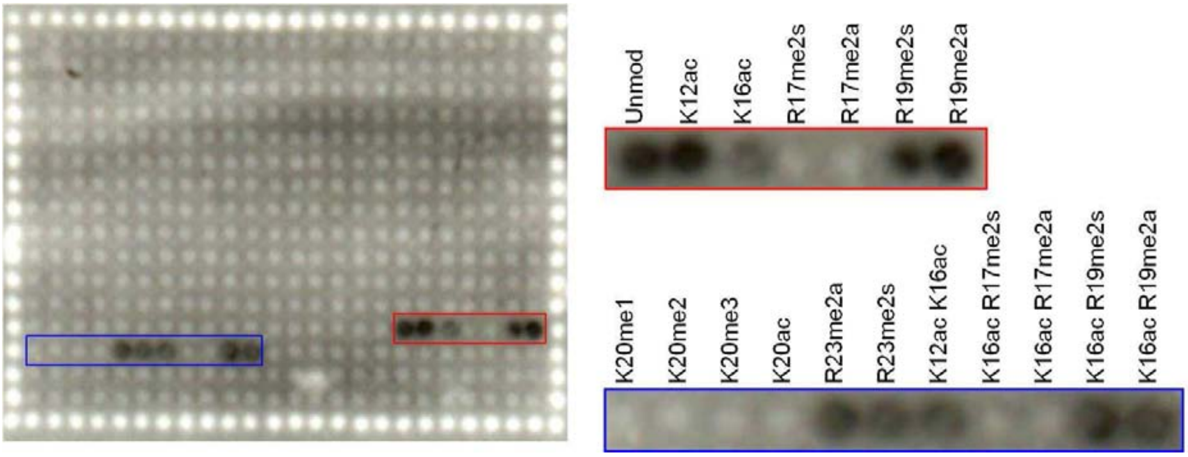


Figure 3: SET8 methylation on celluspot arrays: celluspot arrays were incubated with SET8 protein in the presence of radiolabelled Adomet and the transfer of methyl groups was observed by autoradiography. Right side is the autoradiography image of the complete array and left side is the blow up of methylated peptides

4.3.4. Methylation of Potential Non-histone substrates

We then performed a Scansite search [Obenauer, 2003] with the SET8 specificity profile (R^{17},H^{18}, (RKY), K^{20}, (V^{21}ILFY), (L^{22}FY)) to identify other proteins carrying this motif. With the derived specificity profile we could find only 4 proteins carrying the above sequence motif. As we observed weak methylation when the residues on either side of the target lysine [-1(R^{19})

+1(V^{21})] were exchanged with several amino acids, we repeated the scansite search with relaxed specificity in -1, +1 and +2 positions keeping the other preferences same as above. With this profile we identified 59 proteins carrying the sequence motif with $(KR)^{-3}$, H^{-2}, X^{-1}, K^{0} (X- Any amino acid).

For all these proteins, we synthesised 20 amino acid long peptides in duplicates on cellulose membrane and tested for the methylation by the SET8 enzyme. Out of the 59 potential targets 22 got methylated at peptide level (table 1), 14 of them to equal or stronger than H4, 8 were less than H4 (figure 4). The background activity at H4K20A peptides and also on other peptides most likely is due to SET8 protein being bound at the spots, which carry tightly bound Adomet.

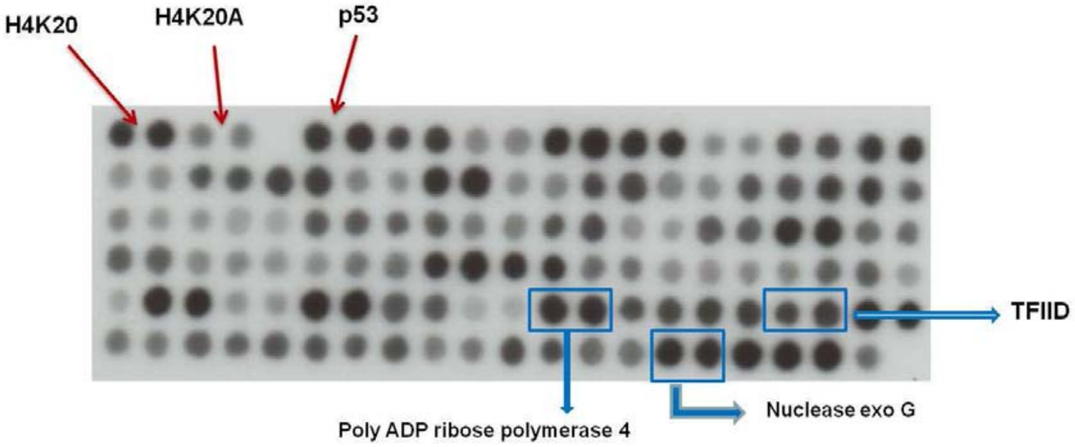


Figure 4: Peptide methyaltion of potential non-histone targets by SET8: Non-histone targets identified with SET8 specificity profile were synthesised in 20 aminoacid length encompassing the target lysine and methylated with SET8

Table-1: List of potential targets methylated at the peptide level

	Protein Name	**Peptide Sequence**	K position	Swissprot No.
1	Adenylate cyclase type 9	KINPKQLSSN SHPKHC**K**YSI	K44	O60503
2	Kinetochore-associated protein	DASMDSAKRR HP**K**LLAKALE	K1493	NP_055523.1
3	TBC1 domain family member 10A	NNWDKWMAKK HK**K**IRLRCQK	K103	Q9BXI6
4	Zinc finger protein 471	SFSKN SMVIKHK**K**VY VGKKLF	K198	Q9BX82
5	Endonuclease G like 1	SKIMGDADRK HC**K**FKPDPNI	K113	Q9Y2C4
6	Hypermethylated in cancer 1	LVALCKKRLK RHG**K**YCHLRG	K154	Q14526
7	ADP-ribosyltransferase like 1	PELRLSKRKH R**K**IPFSKRKM	K1242	Q9UKK3
8	Endonuclease VIII-like 3	RKAGLALSKHY**K**VYKRPNCG	K247	Q8TAT5
9	Phosphoinositide 3-Kinase-C2-beta	FLCRHE**K**IFHPNKGYIYVVK	K1377	O00750
10	P53_HUMAN	SKKGQSTSRH K**K**LMFKTEGP	K382	
11	PR domain zinc finger protein 16	SKLDLRRHK**K** YTCGSVGAAL	K250	Q9HAZ2
12	BMP-2-inducible protein kinase	TYRTPERARRHK**K**VGRRDSQ	K1023	Q9NSY1
13	H4K20	GGAKRHR**K**VLRNDIQ	K20	
14	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform	LCDRHS**K**LFE EPEDPSNRSF	K267	P63151
15	HERV-K_5q33.3 provirus ancestral Pol protein	GENQLPVWLP TRHL**K**FYNEP	No	
16	PR domain zinc finger protein 5	VQVVHERHK**K** YRCELCNKAF	K460	Q9NQX1
17	Exportin-7	EINQADTTHP LTKHR**K**IASS	K185	Q9UIA9
18	DNA-directed RNA polymerase I subunit RPA2	APGIADSLRH F**K**VLREKRIP	K582	Q9H9Y6
19	F-box only protein 11	IRTNS CPIVRHN**K**IHDGQH	K504	Q86XK2
20	Transcription initiation factor TFIID subunit 11	VRRLKSKGQI PNSKHK**K**IIF	K207	Q15544
21	Zinc finger protein Helios	QKGNLLRHI**K** LHSGEKPFKC	K160	Q9UKS7
22	Zinc finger protein 505	SSTLIKHK**K**I HTREKPYKCE	K529	P35789

To check the methylation of these potential substrates at the protein level, we have selected 15 protein domains (table 2) containing the new SET8 target sites and the p53 domain, identified previously as SET8 substrate. The selected protein domains were cloned as GST-fusion proteins**.** Out of 16 proteins, we succeeded in getting the clones for 11 proteins. The candidate non-histone target protein domains were over-expressed and purified by affinity chromatography. Out of the 11 protein domains, 6 got expressed well and could be purified (figure 5)

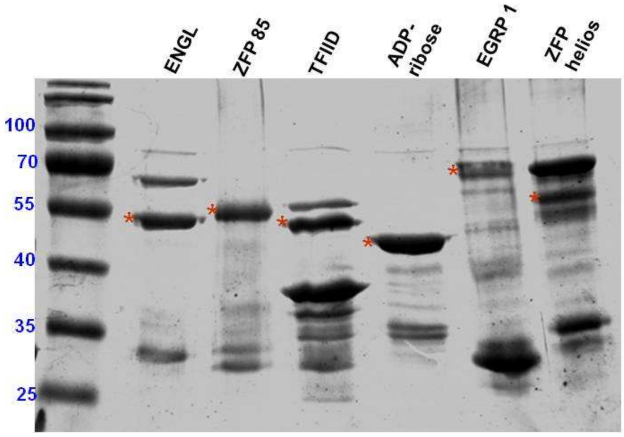


Figure 5: Non histone protein substrates of SET8: Non-histone proteins were expressed and purified by GST-affinity chromatography and purified protein samples were anlaysed on SDS-PAGE. Asterisk marks represents the expected protein size

The methylation of the the target domains was analysed by incubating the purified target proteins with SET8 in a reaction mixture containing radio-labelled Adomet. The methylation activity of SET8 on these protein domains was measured by the transfer of radio-labelled methyl groups by autoradiography. We did not observed the incorporation of radio labelled methyl groups on any of the newly identified substrates, though we see the strong methylation on H4 protein. However on long exposure we observed faint signal on 3 proteins (data not shown). Furthermore, we incubated those three proteins individually with SET8 in presence of radio-labelled methyl Adomet and also included p53 and H4 as controls, autoradiography result demonstrated the incorporation of methylgroups only on H4 and p53 protein and no signal was observed on newly identified SET8 substrates (figure 6).

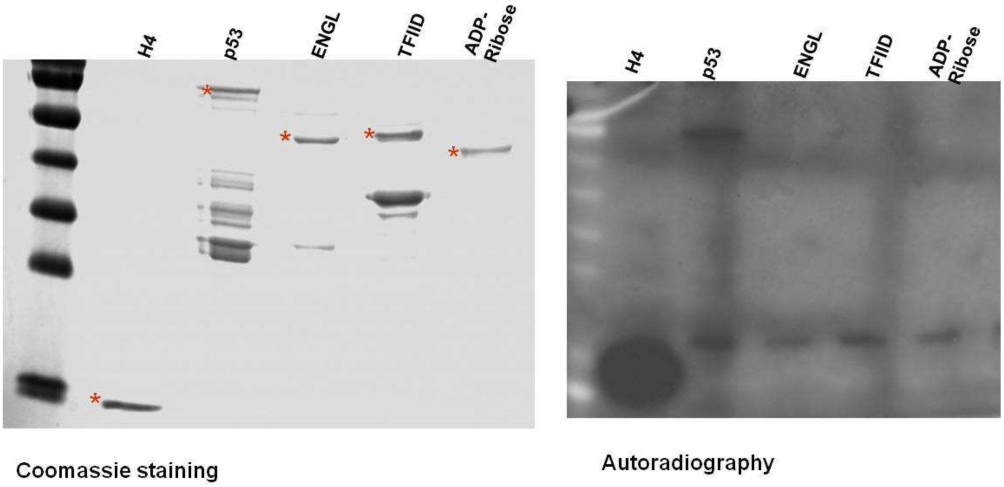


Figure 6: In vitro methylation of identified targets at protein level: Three potential targets which were very faintly methylated were loaded together with p53 and H4 proteins and then incubated with SET8 in presence of radio-labelled Adomet. . Asterisk marks represents the expected protein size

Table-2: Proteins selected for the study of methylation at protein level

Protein Name	**K position and Doamin boundaries**	**Swiss Prot No.**
Adenylate cyclase	K44 (1-365)	O60503
Endonuclease G like 1 (ENGL)	K113(50-273)	Q9Y2C4
Zinc finger protein 505	K529 (388-612)	P35789
Zinc finger protein 85 (ZFP 85)	K530(352-584)	Q03923
Zinc finger protein 43	K502(422-633)	P17038
TBC1 domain family member 10A	K103(1-308)	Q9BXI6
TAF11 RNA polymerase II TATA box binding protein TBP-associated factor 28kDa (TFIID)	K207(1-211) FL cloned	Q15544
ADP-ribosyltransferase like 1 (ADP-ribose)	K1242(1175-1314)	Q9UKK3
Hypermethylated in cancer 1	K154(100-219)	Q14526
Phosphoinositide 3-Kinase-C2-beta	K1377(1291-1599)	O00750
Early growth response protein 1 (EGRP 1)	K416(157-465)	P18146
Zinc finger protein Helios (ZFP helios)	K160(19-298)	Q9UKS7
Zinc finger protein 471	K198(124-352)	Q9BX82
Inositol polyphosphate multikinase	K323 (282-403)	Q8NFU5
Endonuclease VIII-like 3	K247 (1-293)	Q8TAT5

4.4. Specificty analysis of SMYD2

4.4.1. Scientific Background of SMYD2

Within the SET domain family of proteins, SMYD proteins share unique domain architecture. The SET domain in these proteins is split into two segments by the insertion of MYND domain (myeloid-Nervy-DEAF-1), which constitute SET and MYND domain (SMYD) containing protein family (Gottlieb et al., 2002) (Xu et al., 2011). The SMYD family proteins consist of 5 proteins (SMYD1-5) that are not fully characterised and they are grouped based on the presence of two conserved SET and MYND domains. The MYND domains of these proteins are responsible for protein-protein interactions and the SET domain is for methyltransferasse activity like other SET domain containing proteins (Abu-Farha et al., 2008). Unlike NSD family of proteins SMYD family of proteins possess distinct specificities towards histones; SMYD1 and SMYD3 have been shown to methylate H3K4, while SMYD2 is shown to dimethylate H3K36 and it might also methylate H3K4 in presence of Hsp90α (Abu-Farha et al., 2008) (Xu et al., 2011) (Brown et al., 2006) whereas SMYD4 and SMYD5 proteins are not well studied and there is no evidence showing that SMYD4 and SMYD5 possess histone methylation activity.

Recent studies have shown that the SMYD2 gene is amplified in various human solid tumours and overexpression of SMYD2 was able to drive proliferation of esophageal squamous cell carcinoma (ESCC) and predict the bad outcome in ESCC patients (Komatsu et al., 2009). Smyd3 plays an important role in transcriptional regulation of oncogenes and cell cycle regulation-related genes through its intrinsic H3K4-specific methyltransferase activity – its up-regulation linked to the development of certain cancers (Hamamoto et al., 2004).

Interestingly, SMYD2 has been implicated in having both H3K4 and H3K36 methylating activity although there is no *in vivo* evidence for the latter (Brown et al., 2006) (Abu-Farha et al., 2008). In addition, SMYD2 is also involved in the repression of tumor suppressor p53 by methylating lysine 370 (K370) in its C-terminal regulatory domain (Huang et al., 2006). SMYD2 has been shown to methylate retinoblastoma protein at K860, a highly conserved residue in RB protein and further establishing its role in cell cycle regulation (Saddic et al., 2010).

In this study, we have sought to decipher the substrate specificity of SMYD2 through the use of SPOT peptide arrays. We apply both a ‘‘best-target’ and randomized approach to derive the consensus sequence motif, with which we further identifed non-histone targets of SMYD2 and showed methylation at the peptide and protein level, we also confirmed the methylation on the specific lysine by site site directed mutagenesis. We have shown that the identified non-histone proteins are strongly methylated than the p53 protein which validated this approach.

4.4.2. Screening of histone substrates for SMYD2

In its identification in 2006, SMYD2 was characterized as a histone lysine methyltransferase dimethylating H3K36 (Brown et al., 2006). Later it has been shown that SMYD2 prefers to methylate H3K4 in the presence of HSP90α. Moreover, it was also demonstrated that the enrichment of H3K4 methylation is as a result of SMYD2 overexpresssion (Abu-Farha et al., 2008). Both these studies were, however, based on the use of antibodies to specifically detect both site and degree of methylation on the histone tail. Hence, we first attempted to identify which lysine residue(s) could get methylated by SMYD2 using peptides synthesized on cellulose membranes where enzymatic activity could be compared in one experiment with all peptides in competition. For this, we included N- and C- terminal tails of all the histones in 20 amino acids segments and carried out an *in vitro* methylation reaction. We saw that H3 (1-20), H4 C-terminus (81-100) and H2B (1-20) were the strongest methylated peptides. With the exception of the H3 (67-87) peptide (which had the lysine 79 site), all other peptides were weakly methylated by SMYD2, indicating that SMYD2 has a weak specificity on histone tails (figure 1a).

To further dissect which lysine residues within the histone tails were getting specifically methylated, we prepared a second membrane where we included peptides that had the already known target lysines mutated to alanines. In addition, the p53 peptide (366-384) was added since SMYD2 is known to specifically monomethylate p53 to allow a comparison of enzymatic activity. The result clearly showed that p53K370 was the most strongly methylated peptide as compared to all the six described methylation sites on H3 and H4 (H3K4, H3K9, H3K27, H3K36 and H4K20). It is also worth noting that, although SMYD2 has not been shown to methylate any other site on p53, the K370A mutation did not completely abolish methylation of the peptide (figure 1b, right most spot). We observed that the H3K4A mutation induced reduction on the methylation of the H3 (1-20) peptide than the H3K9A mutation suggesting that H3K4 was preferentially methylated (figure 1b; compare spots 2 and 3 from left). H3K36A also reduced methylation of H3 (28-48) significantly but even the wild type peptide was only weakly methylated. However, we did not observed any changes in the methylation of H3K9A and H3K27A peptides.

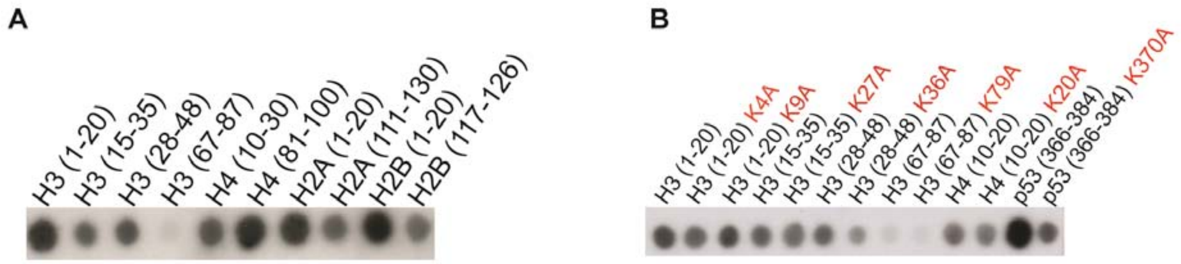


Figure 1: (a) Methylation of N-terminal and C-terminal tails of all histones by Smyd2. Since most of the lysine methylation sites discovered are on H3, several peptides were synthesized to represent each site. (b) Methylation of all known sites of lysine methylation on histones H3 and H4. Target lysines were mutated to alanines to assess if Smyd2 specifically methylates any of the already known sites. The p53 peptide with the corresponding K370A mutation was also included for comparison of activity.

Under the given experimental conditions, we observed that SMYD2 methylated K370 in the p53 peptide much stronger than any of the known target lysines on H3 and H4. The result indicates that the lysine residue(s) on histone proteins are not the preferred substrates for SMYD2.

Consequently, we used the p53 peptide as template to derive the specificity profile of SMYD2 using a 420-aminoacid peptide array. Each residue in p53 (360-380) was exchanged against each of the 20 natural amino acids and all resulting peptides incubated with SMYD2 to determine the critical residues indispensable for the successful transfer of radioactively-labeled SAM. The result showed that leucine at position -1 of the target lysine was the most important specificity determinant with only phenylalanine being the only other accepted residue (figure 2). Positions +1, +2, and +3 accepted most polar uncharged and basic residues but exchanges to acidic residues (aspartate and glutamate), cysteine and large hydrophobic as well as aromatic residues were not tolerated. SMYD2 did not exhibit any specificity to amino acids N-terminal to the leucine at position -1 (figure 2). Interestingly, some exchanges, such as lysine at +3 to serine, brought about a higher activity. The obtained specificity profile was confirmed by three independent experiments to rule out peptide synthesis problems.

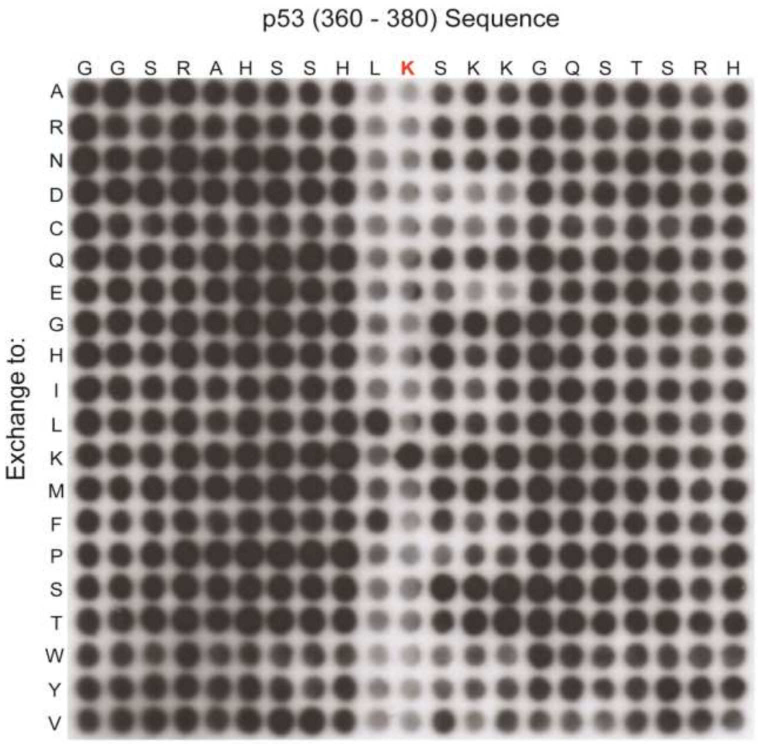


Figure 2: Specificity analysis of Smyd2. The specificity of Smyd2 was studied using a 21x20 peptide array using residues 360-380 of p53 as template. Each residue was exchanged against all 20 aminoacids and methylation activity tested. Result was confirmed by three independent experiments.

4.4.3. Randomization arrays to determine the specificity of SMYD2

Thus far, SMYD2 seemed to have the main specificity of **[F/L] [K]** where K is the target lysine. We aimed next to independently confirm this specificity profile using a randomized peptide array approach. For the first randomized peptide array, the target lysine was place in the center of a 15-mer peptide. Subsequently, the residues immediately adjacent to it (at +1 and -1 position) were substituted by each of 17 aminoacids (Cysteine, Methionine and Tryptophan were excluded) including every possible permutation. This was a total of 17X17 = 289 peptides. The remaining 12 residues (6 on each side) were randomly assigned such that there was a statistical representation of each amino acid at each position in every possible permutation resulting in 289 peptides. Methylation of this peptide array revealed that the strongest methylated peptides shared lysine-leucine-lysine (KLK) motif (figure 3). The darkest spot on the array was of a peptide that had two KLK sites. The next strongest spot was a peptide with one KLK site. Interestingly, this was in agreement with the specificity profile obtained using p53 (360-380) as template where exchange of histidine at -2 to lysine showed a higher activity (figure 2). Some of the other strongly methylated peptides from the randomized array contained Phenylalanine-Lysine (FK) motifs which also corresponded to the p53-based specificity profile. However, it is important to stress that there were highly methylated peptides which matched neither the ‘KLK’ nor the ‘FK’ specificity profile. Prominent examples are the 6th most strongly methylated peptide (TEGKSAGKIVRSHIR) and the 9th most strongly methylated peptide (ATKQGIKKIYKDRYP). Our analysis was also made difficult by the fact that other lysines were not excluded the flanking the central XKY sequence within the randomized peptides. Thus it was hard to tell which lysines were getting methylated for a given peptide. Moreover, the presence of an LK or FK site did not always correlate with strong or even any methylation in some cases .This suggested other residues are still read as was also shown in figure 2.

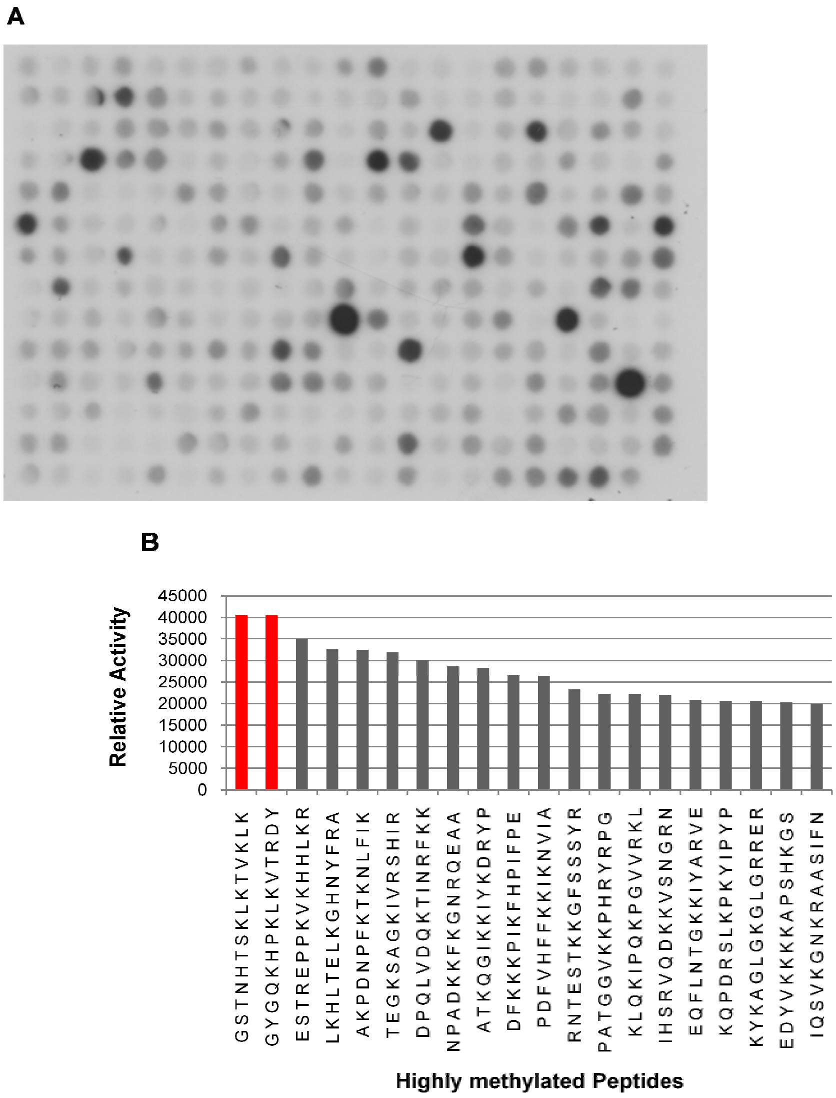


Figure 3: SMYD2 methylation on first generation random peptide array: a) Autoradiography image of the randome peptide arraz: First generation randome peptides were designed by keeping lysine at the centre and randomly substituted other amino acids excluding methionine, cysteine and tryptophan in all the possible combinations and then subjected to methylation with SMYD2 protein. b) Narrowing down the specificity of SMYD2 using a randomized peptide array. The 20 most strongly methylated peptide spots were quantified and plotted. The sequences of the peptides together with the relative activity are shown. The activity bars for two most strongly methylated peptides bearing the ‘KLK’ motif are colored red.

Nevertheless, the two best hits of the first randomized array methylation by SMYD2 had ‘KLK’ sites and we sought to use this ‘KLK’ motif and exchange adjacent residues for a second randomized peptide array. In this second randomization, we also made sure that there were no other lysines in the peptides except the lysines in the ‘KLK’. In the second randomization array we kept ‘KLK’ at the centre and randomized 16 natural amino acid residues (K, M, C, W were excluded) on either sides which resulted in total of 256 peptides, we also included p53 wild type and K370 mutant peptides as control.

Very few peptides got methylated in this approach and the two most highly methylated ones had an arg-threonine (RT) or an arg-serine (RS) next to the ‘KLK’ motif whereas residues preceding the ‘KLK’ motif did not seem to be important (figure 4). Interestingly these two peptides were also methylated stronger than the p53 suggesting that another randomization might still need to find a better peptide substrate sequence for SMYD2.

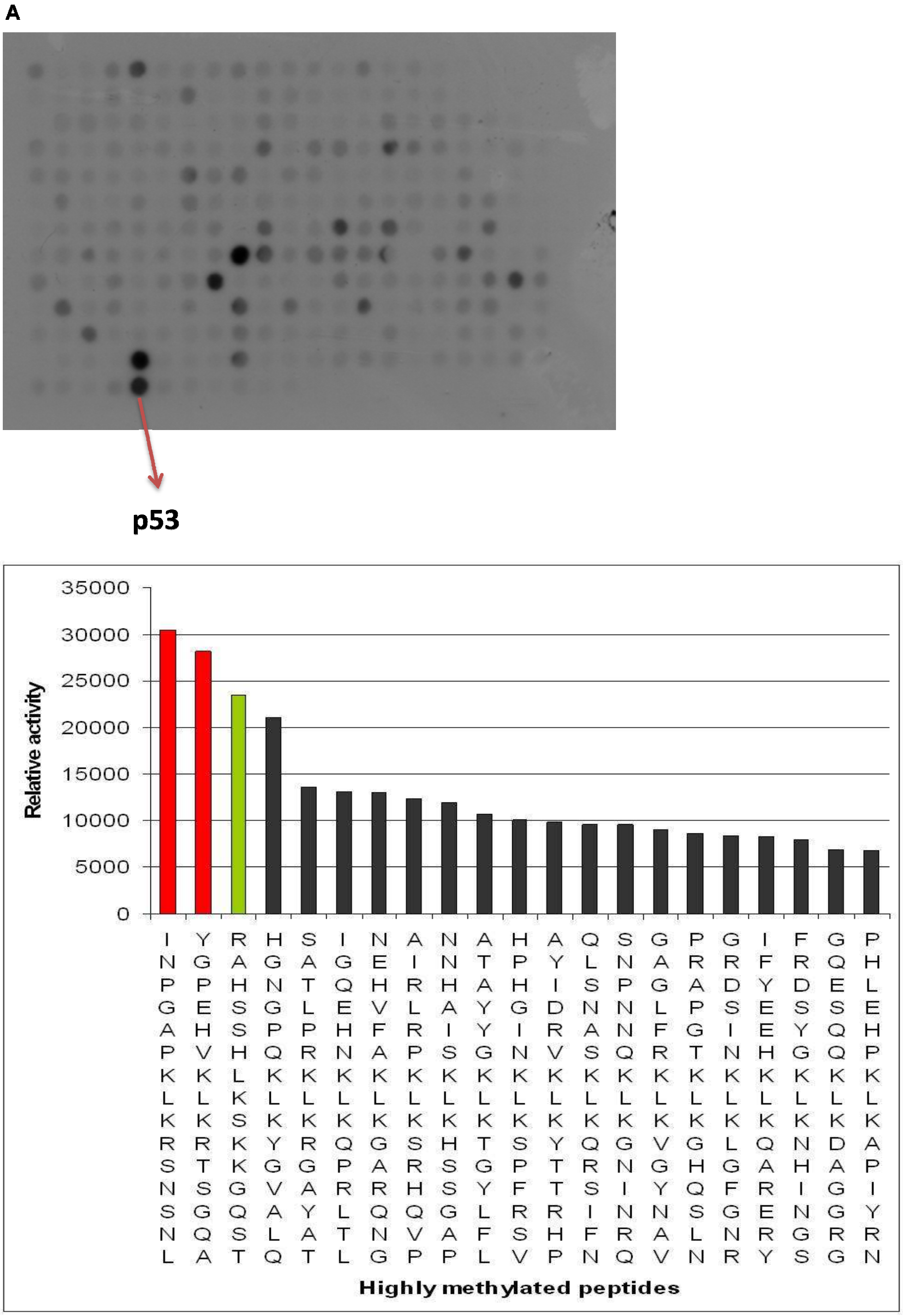


Figure 4: Second generation of randomization array: a) Autoradiography of second randomization array: With the results from the first array, 'KLK' kept at the centre and randomly arranged amino acids on either side in all the possible combinations. The 20 most strongly methylated peptide spots were quantified and plotted. The sequences of the peptides together with the relative activity are shown. Red bars indicate the highly methylated peptides than p53, green bar represents the p53 peptide.

Stimulated with the results of second randomization, we designed another 15 amino acid length peptide array based on the results of the second randomization experiment. Again we kept 'KLK' at the centre but substituted amino acid residues from -4 to -1 preceding 'KLK' and +1 to +4 after 'KLK' with the amino acids observed in the highly methylated peptides of second randomization corresponding to that sequence and (-5 and +5 to +7 to ‘KLK’) the additional positions were completely randomized and synthesized a total of 212 peptides including p53 wild type and mutant peptides. Upon methylation with SMYD2 protein we observed several peptides got methylated and some of them were stronger than p53 and many as good as p53. From the results of these experiments it was obvious that we succeeded in an objective to develop the better substrate sequences for SMYD2. Since we have several peptides which got strongly methylated than p53, we derived the quantitative information of all the peptides and calculated the probable appearance of each amino acid at each specific position in all the highly methylated peptides. At each position we have selected the frequently appeared amino acid in the highly methylated peptides, for instance at +1 position we observed the appearance of ‘R’ in the highly methylated peptides than other amino acids and hence we selected ‘R’ at that position. Similarly we screened by the quantitative analysis in other positions and derived a hypothetical peptide sequence for further study.

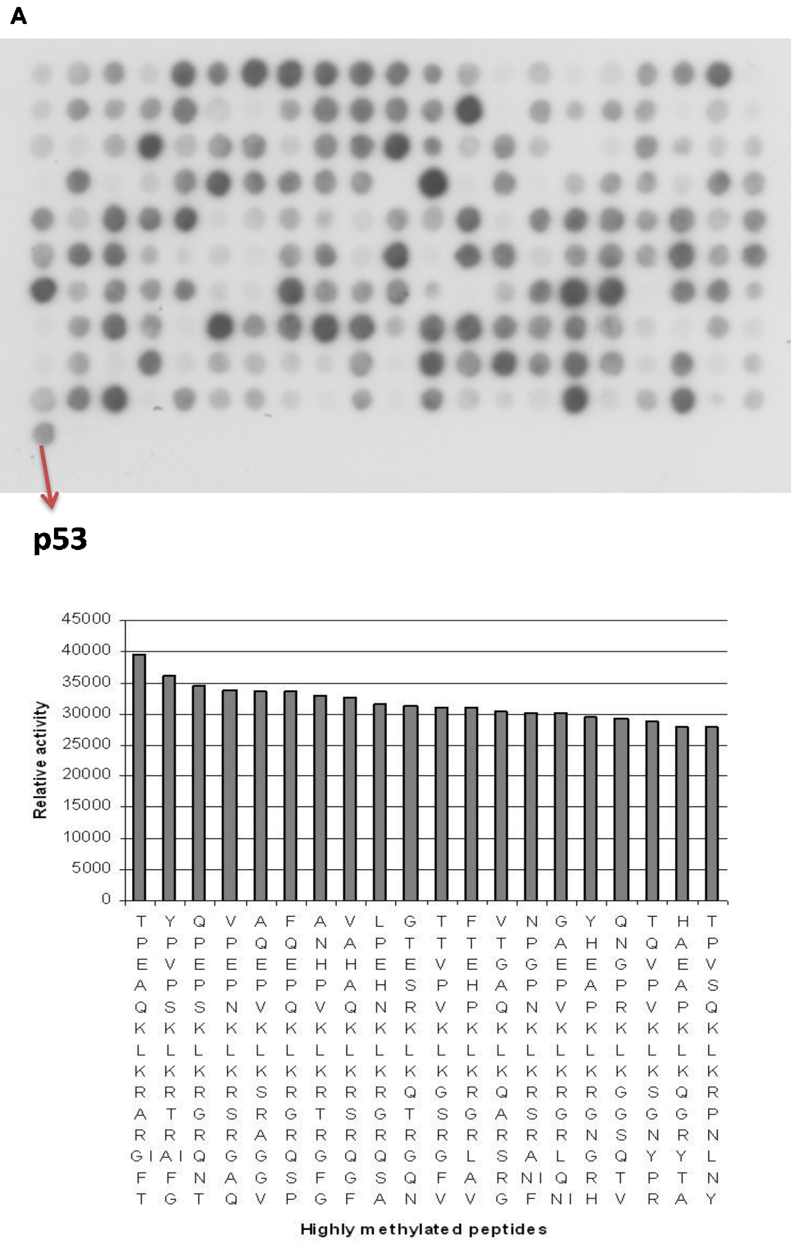


Figure 5: Third randomization array: a) Autoradiography of peptide array: Designed third randomization array based on the results of second randomization experiment and then subjected to methylation with SMYD2. The 20 most strongly methylated peptide spots were quantified and plotted. The sequences of the peptides together with the relative activity are shown

4.4.4. Specificity analysis with the Hypothetical Peptide Sequence

After obtaining several better substrate sequences than p53 with our randomization experiments, we sought to study the specificity profile by using the hypothetical peptide sequence (RNEPPKL**K**RSRGAFT). We used the hypothetical peptide sequence as a template and each residue in that was exchanged against each of the 20 natural amino acids and all resulting peptides incubated with SMYD2. The result of experiment (figure 6A) indicates that the enzyme is specific towards ‘LK’. As expected exchange of the target lysine with other amino acids completely abolished the activity of SMYD2 on peptides. For the -1 position SMYD2 showed preference for leucine, it tolerated only phenylalanine in the place of leucine and exhibited very weak activity when it was exchanged to methionine and glycine. Though SMYD2 majorly recognises ‘LK’ residues, it also has some preferences towards C-terminal residues to target lysine. Activity of SMYD2 was severely impaired when arginine, serine, arginine at +1 to +3 to target lysine were replaced with acidic residues (aspartic and glutamic acids). At +2 (serine) position to target lysine, several hydrophilic residues are accepted and loss of activity was seen when serine exchanged to aromatic or charged residues.

It is interesting to note that the consensus sequence motif of SMYD2 derived from the hypothetical peptide sequence is almost similar to what we derived with the p53 sequence (figure 6). With both the sequences it was clear that enzyme specifically recognises either ‘LK’ or ‘FK’ motif in substrates and also SMYD2 can not tolerate charged residues at +1 to +3 positions to the target lysine with either sequences. This shows the reliability of our approach in studying the specificity profile of histone lysine methyltransferases. Here for the first time we derived the consensus sequence motif for an enzyme from two different back bones, one is the well characterised SMYD2 substrate; p53 and the other is completely a hypothetical backbone derived by randomized method. These results show that the target sequence motif for the SMYD2 is conserved irrespective of its backbone, however, such conclusion need to be confirmed for every enzyme with different sequences.

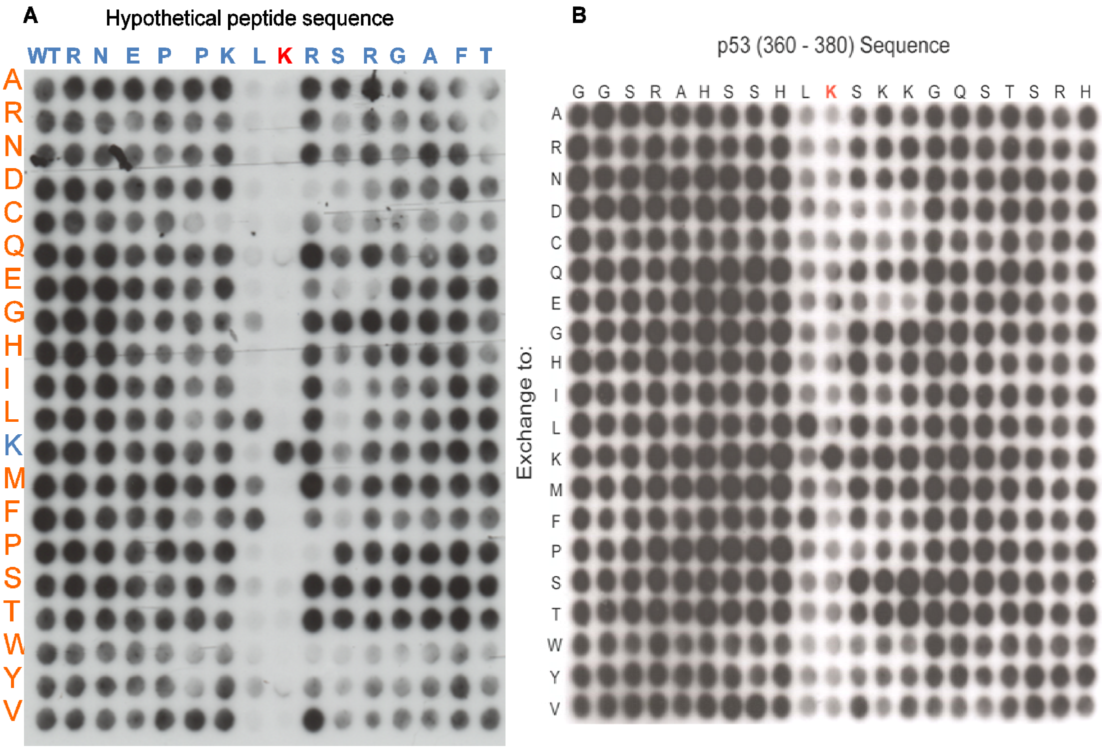

A
Hypothetical peptide sequence
WT R N E P P K L K R S R G A F T
A R N D C Q E G H I L K M F P S T W Y V
B
p53 (360 - 380) Sequence
G G S R A H S S H L K S K K G Q S T S R H
Exchange to:
A R N D C Q E G H I L K M F P S T W Y V

Figure 6: Specificity analysis of SMYD2: a) The specificity of SMYD2 was studied using a 16x20 peptide array using hypothetical sequence derived by randomization experiments as a template. Each residue was exchanged against all 20 aminoacids and methylation activity tested b) The specificity of SMYD2 was studied using a 21x20 peptide array using residues 360-380 of p53 as template.

4.4.5. In vitro peptide methylation of SMYD2 non-histone targets

As shown above the consensus sequence motif of SMYD2 is not matching with the H3 tail, but it is in agreement with its non-histone substrates: p53 (SSHLK^{370}SKK) and retinoblastoma protein (RVLK^{860}RSAE). The results stimulated us to further screen for the non-histone targets containing the SMYD2 specificity sequence motif. Though SMYD2 protein contains a MYND domain which is known to interact with the proteins, but the interactors for this protein are not well characterised enzyme, only p53 has been shown to interact with p53 in HPRD database. We blasted the scanstite search (http://scansite.mit.edu/) with the specificity profile of SMYD2 protein which retrieved several non-histone targets containing potential target sites. Of all the identified targets, 125 potential targets with known or predicted nuclear localisation were selected for further analysis. As expected, with the SMYD2 specificity profile we could not retrieve any lysine residues in H3 or H4 proteins, however, couple of lysine residues in H1 proteins were identified as potential substrates. We synthesised peptides encompassing the predicted target lysine for all the identified potential substrates of SMYD2 on cellulose membrane including the hypothetical peptide and p53 as controls. It was observed that 40 peptides got methylated in par with the hypothetical peptide or p53 peptide and other peptides are either weakly or not methylated at all. However, methylations on the histone peptides were very weak in agreement with our previous results.

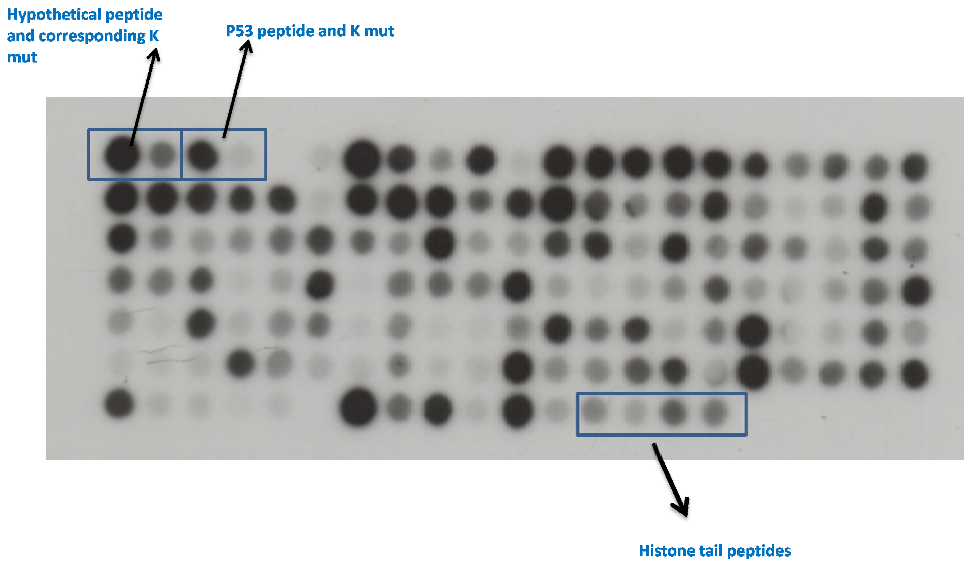


Figure 7: Mehtylation of potentioal non-histone target petides by SMYD2: Identified non-histone proteins of SMYd2 were synthesised as 15 amino acid length peptides encompassing the target lysine and subjected to methylation with SMYD2 and observed the transfer of radio labelled methyl groups by autoradiography.

4.4.6. In vitro protein methylation of SMYD2 non-histone targets

After confirming the methylation on potential non-histone targets of SMYD2 at the peptide level on cellulose membrane, next we sought to check at the protein level, where the target lysine may not be accessible because of the folding of the protein. Of the 40 proteins which got methylated at the peptide level, we have selected 17 proteins (table 1) based on their high intensity of methylation. The protein domains containing the SMYD2 target sites were cloned as GST fusion proteins. For few protein domains we could not amplify the PCR product from cDNA and few protein domains (Cullin3, Negative elongation factor E, Structural maintenance chromosome 3, INO80 complex homolog 1) failed in expression. Eventually we could express and purify 14 protein domains. Though we could purify the NFKB like protein but the sequencing results showed that it has few mutations, hence, we excluded it for further analysis. The methylation of purified protein domains was analysed by incubating the protein domains with SMYD2 protein in presence of radio labelled Adomet. Out of 13 proteins the autoradiography results show a significant deposition of radio labelled methyl groups on 8 proteins, 6 protein domains: MLL2, CPW, Ph3KE, E3UBQ, TFIID, PHF-20 were methylated much stronger than p53 protein and the other 2 protein domains: CHDBP3 and UHRF2 were in par with the p53. The strong methylation on the identified non-histone targets and the high yield of successful prediction of target proteins illustrates the efficiency of our approach in determining the consensus sequence motif for SMYD2.

Table 1: Proteins selected for cloning to check methylation at protein level.

S.No.	Protein name	Domain Boundaries	Target lysine
1	NFKB like protein (NFKBL)	231-462 AA	K454
2	Negative elongation factor E	6-266 AA	K87
3	MLL2	668-916 AA	K883
4	Centromere Protein W (CPW)	9-88 AA	K84
5	Trans membrane Ubiquitin like domain containing protein (TMUDCP)	9-212 AA	K129
6	Chromodomain helicase DNA binding protein (CHDBP3)	358-518 AA	K407
7	Phosphoinositide 3 kinase enhancer (Ph3KE)	274-550 AA	K329
8	PHF-2	688-906 AA	K847
9	Structural maintenance chromosome 3	670-753 AA	K729
10	UHRF2	81-350 AA	K166
11	E3 UBQ ligase RAD 18 (E3UBQ)	64-232 AA	K127
12	RNAPOL1RRN3 (RRN3)	399-634 AA	K567
13	Cullin 3	398-567 AA	K458
14	Cullin 3 N-terminal domain (N-Cullin)	113-405 AA	K396
15	Transcription initiation factor TFIID subunit 1 (TFIID)	413-664 AA	K556
16	INO80 complex homolog 1	13-282 AA	K119
17	PHF-20	266-451 AA	K298
18	Zinc phosphodiesterase ELAC protein 1 (ELAC)	1-320 AA	K50

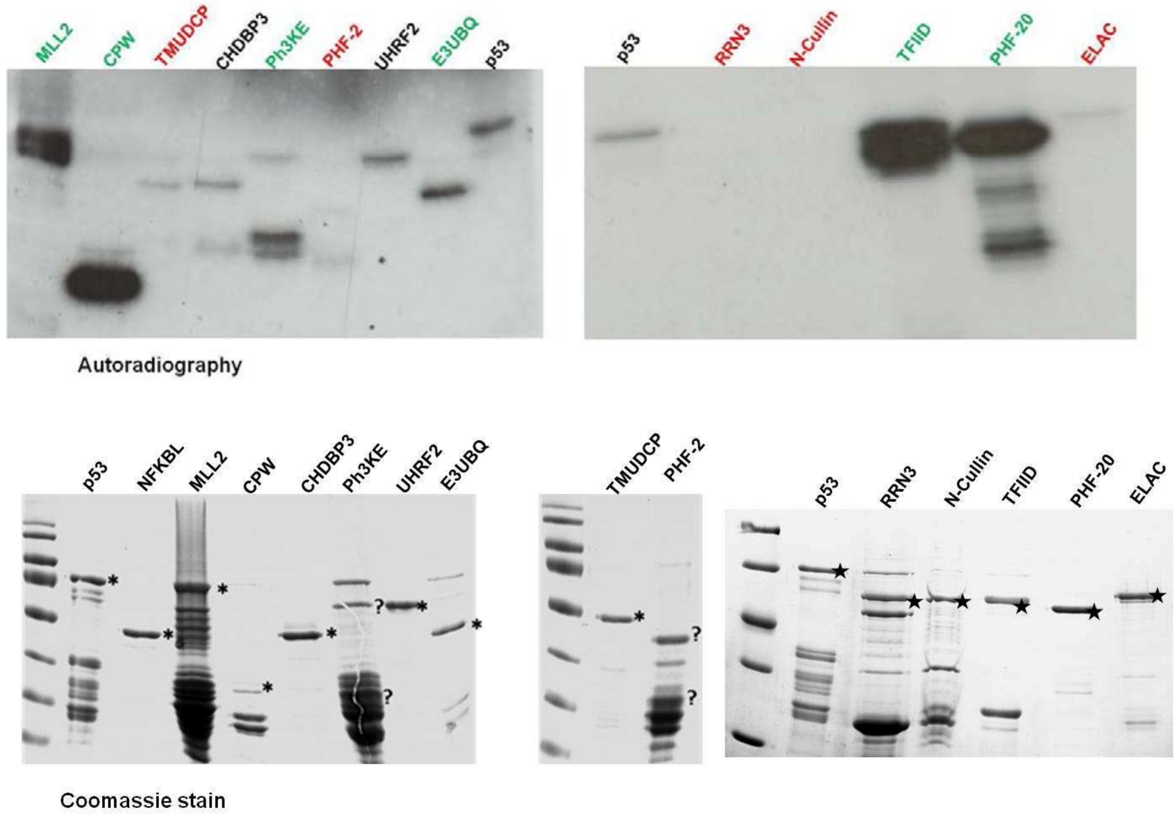


Figure 8: in vitro methylation of newly identified proteins: a) Novel potential target proteins were cloned in GST fusion and subsequently expressed and purified. Purified protein domains were analysed on the SDS-PAGE to compare loading. b) Purified protein domains were incubated with SMYD2 in presence of radio labelled Adomet and transfer of radio labelled methyl groups were analysed by autoradiography. Red color indicates no methylation on the corresponding proteins and green color indicates the strongly methylated substrates than p53.Asterisk marks represent the expected protein band.

To determine, if the methylation on non-histone proteins by SMYD2 is happening at the predicted target lysine, we performed site directed mutagenesis in all the methylated protein domains. The predicted target lysine was exchanged to arginine, the resultant mutated protein domains were expressed, purified and examined again for methylation with SMYD2 (figure 9). CHDBP3 mutant protein band was shorter than the wild type protein, sequencing results showed that it contains 50 base pair deletion mutation towards C-terminal end, however, it still contains the target lysine mutation to arginine. With all the mutant protein domains we observed either no methylation or very weak methylation in compared to their corresponding wild type proteins, this suggest that the target protein domains were methylated at the target lysine as predicted by the SMYD2 specificity profile. However, we still see a very weak methylation signal on the mutant proteins of CPW, CHDBP3, Ph3KE and PHF20 this might be due to residual methylation from other lysine residues in the proteins. But the loss of strong methylation signal in compared to their wild type protein suggests that the methylation is majorly happening on the predicted lysine.

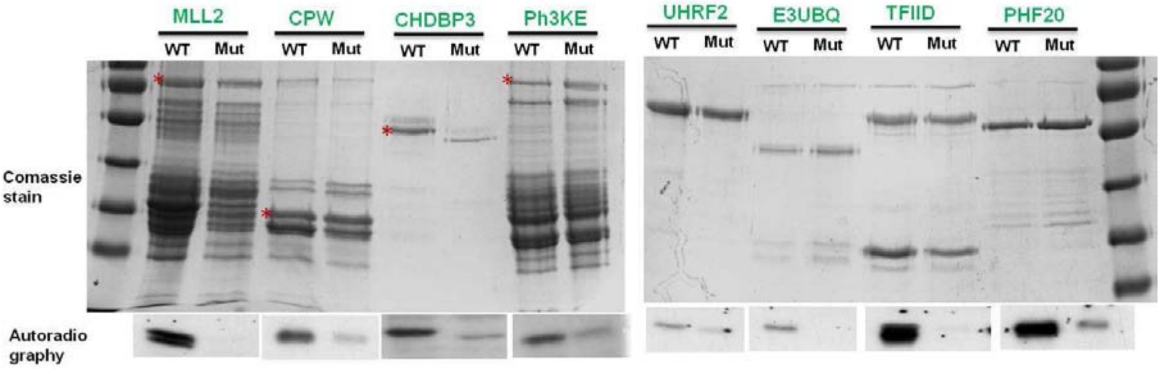


Figure 9: Identification of target lysine: Wild type proteins and the mutated proteins in which target lysine was exchanged to arginine were incubated with SMYD2 in presence of radiolabelled Adomet and the transfer of radiolabelled methyl groups were assessed by autoradiography. Asterisk marks represents the target protein size

4.5. Specificity Analysis-Based Identification of New Methylation Targets of the SET7/9 Protein Lysine Methyltransferase

We applied peptide array methylation to determine an optimized target sequence for the SET7/9 (KMT7) protein lysine methyltransferase. Based on this, we identified 91 new peptide substrates from human proteins, many of them better than known substrates. We confirmed methylation of corresponding protein domains in vitro and in cells with a high success rate for strongly methylated peptides and showed methylation of nine nonhistone proteins (AKA6, CENPC1, MeCP2, MINT, PPARBP, ZDH8, Cullin1, IRF1, and [weakly] TTK) and of H2A and H2B, which more than doubles the number of known SET7/9 targets. SET7/9 is inhibited by phosphorylation of histone and nonhistone substrate proteins. One lysine in the MINT protein is dimethylated in vitro and in vivo demonstrating that the product pattern created by SET7/9 depends on the amino acid sequence context of the target site.

These results were published in Chemistry and Biology (Dhayalan et al., 2011). Annex contains further details.

Contribution

SK (Srikanth Kudithipudi) has contributed to synthesise peptides for several experiments and confirmed the cellular methylation of identified targets by mass spectrometry. I also synthesised peptides for in-solution experiments, did the competitive methylation kinetics with H3K4 and MINT peptide and showed that the number of methyl groups introduced by the enzyme depends on the sequence of the substrate. I have synthesised peptide arrays with H2A and H2B sequences and showed that SET7/9 enzyme could methylate multiple lysines on H2A and H2B proteins, and furthermore confirmed this at the protein level. I have also synthesised the peptide arrays with the several known and identified non-histone targets of SET7/9 enzyme and observed that the targets with KSK motif got strongly methylated.

Since, H3 tails are subjected to several post translational modifications and it has been shown that a significant cross talk exists between different modifications. To investigate the interference of different modifications on methylation activity of SET7/9, a modified CelluSpot array was methylated by SET7/9 enzyme and the results demonstrate that phosphorylation on T3, S10, T11 inhibits methylation on H3K4.

To confirm the cellular methyation of identified novel targets, transfected the protein domain constructs in HEK293 cells together with the SET7/9 enzyme and after two days purified the proteins by immuno-precipitation and subjected them to MALDI analysis. Mass spectrometry analysis reveals that SET7/9 specifically methylates the 5 protein domains at the predicted lysine site in cells.

4.6. Application of celluspot peptide arrays for the analysis of the binding specificity of epigenetic reading domains to modified histone tails

Epigenetic reading domains are involved in the regulation of gene expression and chromatin state by interacting with post-translational modifications of histones. A detailed knowledge of target modifications including enhancing and inhibiting secondary modifications will lead to a better understanding of a specific readout by reading domains. We describe celluspot peptide arrays as a relatively inexpensive and fast method for initial screening for specific interactions of reading domains with modified histone peptides. We tested nine epigenetic reading domains with known histone tail modification targets on celluspot peptide arrays. In general the results agree with literature data with respect to the primary specificty, but in almost all cases we obtained additional new information concerning the influence of secondary modifications surrounding the target modification. We showed that celluspot peptide arrays are a powerful screening tool for the specificity of putative reading domain binding to modified histone peptides.

Results were described in the attached manuscript, it was submitted to BMC Biochemistry
(Bock et al., 2011). Annex contains further details.

Contribution

SK has contributed to clone the RAG2-PHD domain and subsequently expressed and purified the protein. I tested it on CelluSpot arrays and analyzed the interference of secondary modifications in recognizing the primary target H3K4me^3.

4.7. Detailed specificity analysis of antibodies binding to modified Histone tails with peptide arrays

Chromatin structure is greatly influenced by histone tail post-translational modifications (PTM), which also play a central role in epigenetic processes. Antibodies against modified histone tails are central research reagents in chromatin biology and molecular epigenetics. We applied Celluspots peptide arrays for the specificity analysis of 36 commercial antibodies from different suppliers which are directed towards modified histone tails. The arrays contained 384 peptides from eight different regions of the N-terminal tails of histones, viz. H3 1–19, 7–26, 16–35 and 26–45, H4 1–19 and 11–30, H2A 1–19 and H2B 1–19, featuring 59 post-translational modifications in many different combinations. Using various controls we document the reliability of the method. Our analysis revealed previously undocumented details in the specificity profiles of the tested antibodies. Most of the antibodies bound well to the PTM they have been raised for, but some failed. In addition, some antibodies showed high cross-reactivity and most antibodies were inhibited by specific additional PTMs close to the primary one. Furthermore, specificity profiles for antibodies directed toward the same modification sometimes were very different. The specificity of antibodies used in epigenetic research is an important issue. We provide a catalog of antibody specificity profiles for 36 widely used commercial histone tail PTM antibodies. Better knowledge about the specificity profiles of antibodies will enable researchers to implement necessary control experiments in biological studies and allow more reliable interpretation of biological experiments using these antibodies.

These results are published in Epigenetics (Bock et al., 2011). Annex contains further details.

Contribution

SK has contributed in performing the quality analysis of the arrays. Since, we observed several discrepancies from the documented information of the antibodies, to strengthen our results and approach, we cleaved the peptides from the Cellulose membrane and SK did quality analysis on the MALDI.

4.8. The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9

Abstract

Mutations in the ATRX protein are associated with the alpha-thalassemia and mental retardation X-linked syndrome (ATR-X). Almost half of the disease-causing mutations occur in its ATRX-Dnmt3-Dnmt3L (ADD) domain. By employing peptide arrays, chromatin pull-down and peptide binding assays, we show specific binding of the ADD domain to H3 histone tail peptides containing H3K9me3. Peptide binding was disrupted by the presence of the H3K4me3 and H3K4me2 modification marks indicating that the ATRX-ADD domain has a combined readout of these two important marks (absence of H3K4me2 and H3K4me3 and presence of H3K9me3). Disease-causing mutations reduced ATRX-ADD binding to H3 tail peptides. ATRX variants, which fail in the H3K9me3 interaction, show a loss of heterochromatic localization in cells, which indicates the chromatin targeting function of the ADD domain of ATRX. Disruption of H3K9me3 binding may be a general pathogenicity pathway of ATRX mutations in the ADD domain which may explain the clustering of disease mutations in this part of the ATRX protein.

Results of this study were published in Human Molecular Genetics (Dhayalan et al., 2011). Annex contains further details.

Contribution

SK has synthesized peptides and coupled it with fluorescent tags to perform the fluorescence depolarization studies.

5. Discussion

Histone lysine methyltransferases transfers methyl groups from Adomet to the lysine residues in histone protein and plays a vital role in chromatin biology. Since the discovery of first non-histone protein methyaltion substrate (TAF10), tremendous progress has been made in this field. Several non-histone substrates were identified for HKMT's via candidate screening approach and by mass spectrometry analysis. Here, we have characterised the specificity profile of enzymes. The specificity study analysis of histone lysine methyltransferases revealed important similarities and differences by which these enzymes recognise the substrates. Based on the specificity profile we identified several non-histone substrates.

5.1. Histone H1 variant specific methylation by NSD1

We have characterised the substrate specificity of the NSD1 protein by employing peptide arrays. In the past we have successfully characterised the specificity of the G9a and SET7/9 histone methyltransferases by using the same methodology and identified several novel non-histone substrates. The NSD1 protein has been reported to methylate H3K36 and H4K20 (Rayasam et al., 2003), epigenetic regulation of the NSD promoter via CpG methylation leads to diminished levels of H3K36 and H4K20 trimethylations (Berdasco et al., 2009). It has been shown recently that the down regulation of NSD2 significantly lowers the levels of H4K20 (Pei et al., 2011).

Since the sequence environment of H3K36 and H4K20 is entirely different from each other. This intrigued us to investigate the substrate specificity of NSD1. We studied the specificity profile of NSD1 using H3 (31-50) as a template and derived the consensus sequence motif. The scansite search with the derived target consensus sequence motif of NSD1 retrieved another lysine on H4 protein i.e, H4K44. Here we showed that NSD1 protein can not methylate K20 on H4 and showed that instead it methyaltes K44 in H4 protein. We confirmed methylation on H4K44 both by a peptide array experiment and also in solution experiment by MALDI analysis. These contradicting results with the literature intrigued us to investigate the data on H4K20 methylation by NSD1. Rayasam et al (2003) did in vitro methylation of H4 protein with NSD1 and then probed with H4K20 di-methyl antibody and they observed a signal on H4 protein incubated with NSD1 when compared with control. Since, there was only the H4K20 characterised methylation mark on H4 protein, they speculated that methylation signal could be from K20. Recognition of K44 methylation mark by H4K20 antibody could be attributed to the poor quality of antibody, similar cases were thoroughly discussed in our recent publication

(Bock et al., 2011). However, later on mass spectrometry analysis revealed another methylation mark on H4 protein: H4K44 (Zhang et al., 2004). Recently it has been shown that the NSD2 protein also methylates H4K44 on histone octamerers and showed DNA inhibits NSD2 methyaltion on H4K44 (Li et al., 2009). Collectively these results suggest that NSD family of proteins methylate H4K44 but not H4K20, although it is not yet clear whether NSD family of proteins are responsible for H4K44 methylation in vivo. Interestingly, knockdown of NSD2 in cells led to diminished H4K20 methylation which is contradicting with the in-vitro data. Further investigation need to be done to examine the validity of these in vivo studies and possible indirect effects of NSD family of enzymes for H4K20 methylation.

Our data shows that the NSD1 protein also methylates lysine-168 in H1 proteins, H1.5, H1.2 and H1.3 which are the new substrates for NSD1 protein. We showed that NSD1 methylates K168 in H1 proteins by in vitro experiments, as confirmed by peptide array, mutational analysis at the protein level and also the degree of methylation by MALDI analysis. To our knowledge, this is the first characterisation of K168 in (H1.5, H1.2 and H1.3) methylation by a histone methyltransferase. Both NSD1 and NSD2 (data not shown) proteins exhibit methylation on H1 protein in a variant specific manner. Among all the identified targets of NSD1 by scansite search, H1.5K168 has been shown to be the best substrate in our peptide array methylation analysis of the putative non-histone targets. Subsequently we also showed that methylation by NSD1 on H1.5 protein is 3 times faster than its primary substrate H3K36. Our experiments to shown that NSD1 is responsible for in vivo methylation of H1.5K168 are still in progress. It will be interesting in the future to determine the biological functions and localisation pattern of H1 modifications and also to check the probable cross talk between H1.5K168 methylation and H3K36 and vice versa.

The NSD1 enzyme has been described as di-methyltransferase enzyme, but apparently our MALDI analysis showed only mono-methylated products with all the substrates. This might be due to the low activity of the enzyme. In all the methylated samples we observed only the partial conversion of un-methylated peptide to mono-methylated peptide (~25%) and since, majorly SET domains introduce methyl groups in a distributive manner, we assume a little of mono-methyl peptide would have been converted to di-methyl peptide, which we could not detect under the given conditions. However, recently it has been shown that NSD1 methylates p65 protein at K218 and K221, Mass spectrometry analysis revealed that NSD1 mono methylates K218 and dimethylates K221 of p65 (Lu et al., 2010). This might suggest that the number of methyl groups introduced by NSD1 also depends on the sequence of the substrate similar as we showed with SET7/9 (Dhayalan et al., 2011).

Apart from the histone proteins we also showed NSD1 could methylate non-histone proteins as well. NSD1 methylates ATRX and Probable U3 small nucleolar RNA-associated protein specifically at the predicted lysine in vitro. Since methyaltion on non-histone proteins was weaker when compared to H3 protein, we did not proceeded further with the non-histone targets. The consensus sequence motif of NSD1 is hydrophobic, perhaps the predicted lysine for most of the targets are involved in the folding of proteins and thus not accessible for NSD1 methylation.

Finally we also showed that the NSD1 protein is subjected to automethylation. Such phenomenon is not unique to NSD1, earlier it was also shown for murine G9a and PRMT6 (Chin et al., 2007, Rathert et al., 2008 and Frankel et al., 2002). The biological significance of PRMT6 automethylation is not known, but G9a automethylation recruits heterochromatic protein (HP1) and perhaps plays a role in heterochromatin formation. With an unbiased candidate screening approach we identified the lysine responsible for automethylation in NSD1, and also showed the loss of automethylation after mutating the predicted lysine (K1769) at the protein level. The methylation activity of NSD1 was not altered by the exchange of K1769, suggesting that mutation of K1769 does not influence the enzyme activity. Since NSD1is a nuclear receptor protein and involved in several diseases, it is further interesting to see whether it could recruit any proteins specific to the K1769 methylation and thus play a role in the transcriptional regulation or not. In this study we also showed that the Sotos mutations of the NSD1 catalytic domain led to a complete loss of the activity of NSD1 on H3K36. Thus the consequence of Sotos syndrome mutations appear to be a loss of methylation activity of NSD1. Whether loss of methylaiton on H3K36, HK168 or loss of methylation of any of the weaker in vitro substrates (H4K44, ATRX, Probable U3 Small Nucleolar RNA) is most important remains to be studied.

5.2. Epigenetic substrates of SUV39H1

In this study we have derived the consensus peptide sequence motif for lysine methylation by the SUV39H1 enyzme by employing peptide arrays. Previously we did similar studies with other H3K9 methyltransferase and showed that arginine immediately fallowed by lysine, so called ‘RK’ motif is very important to methylate novel substrates (Rathert et al., 2008).

Similarly for SUV39H1 we have shown that ‘RK’ motif is a central recognition sequence, in addition to it lysine at -5 position to the N-terminal of target lysine is very important to methylate the novel substrates. This specific recognition of -5K explains the specificity of SUV39H1 only towards H3K9 unlike G9a which methylates both H3K9 and H3K27. With the derived specificity profile we have predicted several potential substrated and showed methylation for 16 proteins at the peptide level and to 5 proteins at the protein level. With both the RAG2 and SET8 protein we have showed that the lysine residue at -4 and -5 position with respect to target lysine were important to methylate the substrates, however, it would be further interesting to study whether any of the known methyltransferase could methylate the K164 in SET8 and K503 in RAG2 and then to investigate the interplay between these two modifications.

The RAG2 protein and SET8 were strongly methylated by SUV39H1 in vitro, however, so far we could confirm methylation only for the SET8 protein in cells. Though methylation on RAG2 protein is stronger than SET8 but we could not see any methylation signal on RAG2 protein with the PAN methyl specific antibody, but this might be due to the poor expression of RAG2 protein in cells. Experiments are in progress to enhance the transfection efficiency of RAG2 into mammalian cells and also to optimise expression in different cell lines to get the better yield.

The RAG2 protein together with the RAG1 is responsible for the VDJ recombination activity **(**McBlane et al., 1995), intitally the C-terminal part RAG2 protein was considered to be dispensable for the activity but later on it has been shown to be important for the localization and also to stabilize the RAG1/RAG2 heteromeric complex **(**Grundy et al., 2010, Spanopoulou et al., 1995, Akamatsu et al., 2003). Moreover, recently it has been shown that the PHD finger present in the C-terminal part of RAG2 binds to H3K4 tri-methylated lysine and might be involved in epigenetic mechanisms (Matthews et al., 2007). Collectively these results suggest that the C-terminal of RAG2 protein has a vital role in the regulation of RAG2 protein. The SUV39H1 target lysine K507 of RAG2 is located in the extreme C-terminal region and also part of the NLS. In our sub cellular localization studies we have observed the changes in the sub-nuclear localisation pattern of RAG2 protein when co-expressed with SUV39H1. We have also observed the methyl specific interaction of JMJD2A tandem tudor domain to the RAG2 and SET8 proteins in vitro at the peptide level, which also suggests that either JMJD2A or some other methyl specific binding proteins might interacted with the RAG2 and altered its sub-nuclear localisation pattern. The RAG2 protein has been shown to be regulated by several post-translational modifications, serine365 phosphorylation enhances the activity of RAG2, while RAG2 protein also subjected to degradation via ubiquitylation (Jiang et al., 2005). Similarly it would be further interesting to investigate the downstream effects of methylation on K507.

The SET8 protein is an H4K20 specific mono-methyltransferase and was shown to involve in the cell cycle regulation. Here, for the first time we have showed a HKMT methylating an other HKMT. The SET8 enzyme interacts with PCNA (Proliferating cell nuclear antigen) via PIP2 (PCNA-interacting protein) box. SET8 interaction with PCNA enhances the degradation by polyubiquitylation during S phase. A PIP2 mutant was shown to be more stable than wild type SET8 protein (Jørgensen et al., 2011) and (Oda et al., 2010). SUV39H1 tri-methylates K169 in SET8 protein, which is in close proximity to the PIP2 (178-185 AA) box. It would be further interesting to study whether methylation at K169 would interfere with PCNA interaction or not and further investigate to study the stability of the SET8 protein. SET8 protein stability and localisations are cell cycle regulated, SET8 has been shown to be shift from the nucleus to cytoplasm during the different phases of cell cycle (Yin et al., 2008). In our sub cellular localisation experiments with SET8 we have seen it accumulated in cytoplasm in some cells and in nucleus in few cells, however we could not derive any conclusions about the effects of methyaltion from these experiments. Perhaps, we should repeat the co-expression experiments by arresting the cells at different phases to see the localisation differences of SET8 in presence and absence of SUV39H1. Both RAG2 and SET8 proteins stability varies with the cell cycle progression, it would be exciting to study whether the methylation on the target lysine’s on these proteins are cell cycle regulated or not. Since, SUV39H1, RAG2 and SET8 are part of epigenetic machinery it would be also interesting to study the consequence of methylation.

5.3. SET8 – A very specific protein lysine methyltransferase

In this study we have derived the substrate specificity profile for the SET8 protein lysine methyltransferase enzyme by employing peptide arrays, which were successfully used in our lab to investigate the substrate recognition motif of histone lysine methyltransferases. Unlike other SET domain proteins (G9a/GLP, SUV39H1), SET8 has notable distinction between its activity towards nucleosomal H4 and octamer H4, under optimal conditions it has been shown that SET8 most efficiently methylates histone 4 incorporated into nucleosomes (Fang et al., 2002). With the peptide arrays we have shown that SET8 has long substrate recognition motif, the results of alanine scan experiment show that the residues form R^{16} to R^{23} of H4 are important for the SET8 to methylate K20, this was in agreement with the previous findings about the SET8 recognition sequence (Yin et al., 2005).

Crystallographic structure analysis of SET8 bound with the H4 peptide reveals that the substrate pocket of SET8 is much deeper and more pronounced than pockets of other enzymes. The side chain of the residues in the N-terminal part of the H4K20 (R^{17}, H^{18} and R^{19}) has been shown to be involved in several interactions with the residues in the binding pocket of SET8, in contrast to this, residues on C-terminal to K20 of H4 are solvent exposed and do not engage in significant interactions with enzyme (Couture et al., 2005). The results of our complete peptide array specificity analysis, in which each amino acid at each position of H4 (10-30) backbone were exchanged with the all possible naturally available amino acids also show that SET8 has specific recognition from R^{16} to R^{23} of H4 tail. We observed SET8 is very specific towards R^{17} and H^{18} along with the K^{20}, exchange of any amino acids at corresponding position leads to the complete loss of activity. However we observed some discrepancies with the structural data regarding the residues to the C-terminus to K20. It has been identified that L22 of H4 binds in a shallow hydrophobic pocket formed by aliphatic side chains of Thr-307, Leu-309 and Leu-318 in enzyme. In agreement with this we also observed that SET8 could tolerate only exchange of hydrophobic residues phenylalanine and tyrosine, however the activity was still less in relative to the native sequence. Unlike other enzymes (G9a/GLP, SET7/9), SET8 has long and very defined recognition sequence motif, it specifically interacts with 6 residues of H4 and thus represents a very specific histone methyltransferase enzyme.

However, recently it has been shown that SET8 could also methylate lys-382 in p53 protein. An inspection of the p53 sequence (figure 7) shows that the amino acid residues (-3,-2-1) N-terminal to the target lys-382 are exactly matching with the histone 4 sequence but residues on C-terminal side of the target lysine were different from H4. Still, the p53 contains hydrophobic amino acids which could still bind in a shallow hydrophobic cavity formed by the residues of SET8 as indicated in the SET8-H4 peptide structure analysis.

H4 :GKGGAKRHRKVLRDNIQG
p53: KGQSTSRHKKLMFKTEGP

Figure : Sequence alignment of p53 K382 with the H4K20 sequence

We did the scan site search with the obtained specificity profile (R^{17}K, H^{18}, R^{19}KY, K^{20}, V^{21}ILFY, L^{22}FM) to identify the potential targets, however, with this sequence motif we found only 4 proteins. And then with the relaxed specificity profile we identified several potential targets and observed methylation of 22 proteins at the peptide level but surprisingly we did not observe methylation on any of the identified novel targets at the protein level. Nevertheless, we observed methylation signal on p53 protein but relatively weaker than H4 protein.

With SET7/9 and G9a HKMT's we have shown 60 to 70% success rate of methylation with the identified substrates at the peptide level and further at the protein level. In contrast to this SET8 protein methylated newly identified substrates at the peptide level and no methylation was observed with the corresponding protein domains at the protein level. This could be attributed to the long recognition motif of SET8 protein, a minimum of 6 to 7 amino acids must be available for the SET8 protein to interact and methylate the target lysine in the substrates. The peptides are unfolded, hence the target lysine and the adjacent amino acids were accessible to the SET8 protein, however, at the protein level either the target lysine or any of the adjacent vital residues might be involved in the folding of protein and further unavailable for the SET8 protein to act on those substrates. Where as in the p53 protein the target lysine resided in the unstructured C-terminal tail similar to the unstructured histone tails which can be easily access by the SET8 protein.

And in addition to that, we synthesised peptides on cellulose surface and the efficiency of enzymes to act on immobilised substrates is much better than in solution experiments. Substrates which gets methylated at peptide level may not be methylated on protein level but would be surprised to see the vice versa. Moreover, the substrate pocket of SET8 is much deeper than other SET domain proteins which also might hamper to interact with the folded protein domains. Moreover SET8 enzyme prefers to methylate H4 protein incorporated in nucleosomes than H4 in octamers which further suggest this enzyme was specifically designed for chromatin functions unlike SET7/9 enzyme which can not methlyate nucleosomes and has been shown as a protein lysine methyltransferase (Dhayalan et al., 2011).

In the light of the above events and the efficiency of enzyme to prefer nucleosomes as substrates tempted us to speculate that the SET8 is a H4K20 specific methyltransferase. However, in the future we can not completely rule out the identification of novel substrates similar to p53, which has target lysine on the extreme C-terminal end. SMYD2 methylates

H3K36 in native conditions but in presence of HS90α it methylates H3K4, similarly the conformation of the SET8 enzyme might change in the cells depending upon the interaction partners and could methylate additional proteins.

5.4. SMYD2 - A Non-histone protein lysine methyltransferase

Here we have characterised the substrate specificity of SMYD2 enzyme, which has been described as H3K4 and H3K36 methyltransferase (Brown et al., 2006) (Abu-Farha et al., 2008). Recently it has been reported that SMYD2 could also methylate p53 protein. Surprisingly these target sites are very different from each other with respect to the amino acid sequence context. To clear the discrepancies associated with SMYD2 we examined specificities on all the known histone substrates and on also on p53 and showed that p53 was the most preferred substrate over the histone substrates.

We have also managed in this study to apply both a "best-target" and a randomized peptide array approach to reach at a specificity profile for SMYD2. The best-target approach is very useful in that it starts from an already known biological target and establishes which residues are critical for methylation activity and which are dispensable. A similar approach has been used before to derive the specificity profile of G9a and SET7/9 and to identify other targets (Rather et al., 2008 and Dhayalan et al., 2011). Using this approach we have shown that SMYD2 is highly specific for an LK or FK motif in target peptides with some readout of adjacent residues.

We have also derived the best substrate peptide sequence for SMYD2 by employing a random array approach. However, with both the best substrate specificity profile and random approach apart from minor deviations, we ended up in having the same consensus sequence motif for SMYD2. This shows the reliability of our peptide array approach in determining the specificity profile for SET domain proteins. In an unbiased approach we have showed that the SMYD2 recognises either 'LK' or 'FK' motifs in the substrates, the preference for 'LK' or 'FK' suggests that the SMYD2 specificity is different from the known H3 substrates; H3K4 and H3K36. SMYD2 did not prefer either threonine or valine in the place of leucine, so the activity observed, on H3K4 or H3K36 is only the residual activity of enzyme, the true substrate for this enzyme is yet to explore. Recently it has been shown that SMYD2 methylates K860 in retinoblastoma protein which was also in agreement with our derived consensus sequence motif. In the light of these results, it tempted us to speculate that SMYD2 may not be a specific histone methyltransferase enzyme unlike SET8 or SUV39H1 or the appropriate substrate is yet to be identified for this enzyme on the histone proteins. Perhaps SMYD2 can also be considered as a non-histone protein methyltransferase enzyme similar to SET7/9, which actually prefers non-histone protein substrates (Dhayalan et al., 2011).

With the derived specificity profile we identified several non-histone target proteins which had the consensus sequence motif of SMYD2 and confirmed methylation for 40 proteins at the peptide level. Further, we confirmed the methylation for selected protein domains at the protein level. Altogether we showed the specific methylation at the predicted lysine for 8 proteins and six out of the eight identified targets were more strongly methylated than the p53 protein. However, protein methylation on the non-histone targets was shown only in vitro and experiments are in progress to check whether the methylation is occurring in cells as well. After this we plan to search for methyl specific interactors for the identified targets. Additional experiments will address the downstream effects of methylation in particular of the non-histone substrates.

Since, SMYD2 has been shown to di-methylate H3K36 in vitro (Brown et al., 2006) and mono methylate p53 (Huang et al., 2006), the degree of methylation for this enzyme is yet to be characterised. Our autoradiography experiments only confirm the methylation on the proteins but can not specify the degree of methyaltion. Since, we do not have well characterised antibodies to check the degree of methylation, we are planning to do the in vitro methylation of the protein domains by SMYD2 with unlabelled Adomet and followed by mass spectrometry.

In summary our results shown that the SMYD2 appears as a highly active protein lysine methyltransferase with relatively little sequence specificity. Considering its cellular functions and poor activity on histone proteins, it is more likely that this enzyme has many cellular targets, some of which we identified here and many are yet to identify and the methylation of which may explain the important biological role of this enzyme.

6. Conclusion

We set out in this study to characterise the specificity and mechanism of histone lysine methyltransferases. Since some of these enzymes recently were shown to methylate non-histone substrates as well, the name protein lysine methyltransferase (PKMT) is more appropriate. Based on the specificity data we aimed to identify novel substrates for the enzymes. Here we have showed some enzymes are specific towards histone substrates like SET8 and few enzymes are potential non-histone protein methyltransferases like SET7/9 and SMYD2. Each enzyme has distinct specificity towards substrates, for instance SET8 requires long a motif to recognise the target lysine in the substrate. SET8 recognises 6 to 7 residues to methylate H4K20, which is unique among protein lysine methyltransferases. Our specificity analysis and protein methylation results demonstrated that SET8 could be considered as a specific histone methylatransferase. Other histone HKMT's like SMYD2 and SET7/9 evolved more as specific protein methyltransferases unlike SET8. SET7/9 and SMYD2 exhibits weak activity on the histone substrates. For these enzymes we employed randomized peptide arrays to screen the best sequence motif. We identified 'KSK' motif as the consensus sequence motif for the SET7/9 and showed that the novel substrates identified with this motif are stronger methylated than the histones (H3K4) and previously identified targets. Similarly with the SMYD2, we have shown that histones are not the primary substrates because the non-histone protein p53 is preferred. We used peptide array approaches to identify the consensus sequence motif (LK or FK) for SMYD2, and identified novel substrates which were even better than p53.

It is to be noted that an important biochemical difference exists between SET8 and SET7/9. SET8, which is a H4K20 specific enzyme preferred to methylate H4 in associated with DNA (in nucleosomes) rather than free histones. In contrast to this SET7/9 prefers to methylate H3 in octamers not incorporated into nucleosomes, from which it is clearly evident that SET7/9 was designed to methylate non-histone proteins and SET8 for the DNA associated histones. Specificity of SMYD2 on histone octamers and nucelosomes is yet to be investigated.

NSD1 is another H3K36 methyltransferase similar to SMYD2, however, these have completely different specificity profiles. NSD1 recognises four residues (-2 to +1) in the H3 tail unlike SMYD2 which has preference either to LK or FK motif. This suggests that, though both the enzymes have been discovered as H3K36 methyltransferases they could eventually have different non-histone substrates. Based on the NSD1 specificity profile we identified K168 a novel lysine methylation site in H1 proteins. SMYD2 which prefers either L or F at -1 position can not methylate H1K168 substrates.

SUV39H1 is also a heterochromatic protein like SET8. However, SUV39H1 recognises mainly an RK motif corresponding to R^8 and K^9 in the H3 tail unlike SET8 which has long recognition motif. In the past, we have shown that G9a (an other H3K9 methyltransferase) also recognises an RK motif but SUV39H1 in addition to RK, also recognises K at -4 position. Our data illustrates that irrespective of identified substrates the specificities of protein lysine methyltransferases should be studied to identify potential substrates and to understand their cellular mechanisms via protein methylation.

7. Materials and Methods:

Cloning, expression and Purification

The sequences encoding the protein domains were amplified from the cDNA derived from HEK293 cells. Protein domains were cloned into the corresponding restriction sites of the pGEX-6p2 vector (GE Health care) to express as a fusion protein with GST. Domain boundaries of each protein were listed in the results part of the corresponding projects. Mutagenesis was introduced using PCR-megaprimer mutagenesis method. Mutagenesis was confirmed by restriction marker site analysis and followed by DNA sequencing.

SET domain constructs; GST-Suv39h1 SET domain (82-412, also includes pre- and post-SET regions) construct in pGEX-2T was obtained from our collaborator Dr. Xaiodong Cheng. GST-mNSD1 SET domain (1700-1987) construct in pGEX-2T was obtained from our collaborator Dr. Lerouge Thierry. GST-SET8 (190-352) and full length SMYD2 proteins were cloned into pGEX-6p2 vector.

For mammalian expression, an oligonucleotide coding for the nuclear localization signal of simian virus large T-antigen was cloned in frame with YFP protein in pEYFP-C1 vector (Clontech) by using BspEI/XhoI sites to generate the pEYFP-C1-AJ-NLS construct. The wild type RAG2 protein domain and SET were subcloned into pEYFP-C1 and the mutant proteins were cloned into pEYFP-C1. SUV39H1 protein was cloned both into pEYFP-C1and pEYFP-C1. Mouse NSD1 SET protein domain (1700-1987) was cloned into pEYFP-C1-AJ-NLS construct.

For expression of each target protein or enzyme, the *E.coli* BL21 cells (Novagen) transformed with the corresponding plasmids were grown in Luria-Bertani medium at 37 ^{0}C to an $OD_{600} \approx$ 0.6 to 0.7, then shifted to 22 ^{0}C for 20 minutes then induced overnight with 1 mM Isopropyl β-D-thiogalactoside (IPTG).

Harvested cells were resuspended in 25 mL sonication buffer (50 mM Tris, 150 mM NaCl, 1mM DTT and, 5% Glycerol pH 7.4) and disrupted by. The lysates were spun down at 20000 rpm (Avanti Ultracentrifuge) for 1 h and 20 minutes and the supernatants were then loaded onto pre equilibrated Glutathione-Sepharose beads (GE Health Care). Columns were washed 1 time with Sonication buffer and 2 times with HGCB buffer (50 mM Tris, 500 mM NaCl, 1mM DTT and, 5% Glycerol, pH 8) followed by elution of bound proteins with 40 mM Glutathione pH 7.4). Eluted proteins were dialyzed first against 20 mM Tris, 100 mM KCl, 0.5 mM DTT and 10% glycerol for 3 h, and overnight against the same buffer with 60% glycerol.

Methylation of Purified Protein Domains and mutants

Methylation of protein domains and the corresponding mutants was performed by incubating with the HKMT's in the corresponding methylation buffer (check below for the buffer details) with 16 - 25 nM enzyme and 3.5 – 5 µM of target protein. Reactions were started by addition of 0.35 µM labeled [methyl-^{3}H]-AdoMet (Perkin Elmer) and incubated for 3to 8 h to at 37 ^{0}C.

SUV methylation buffer:

Protein domain methylation reactions were performed in methylation buffer [50 mM Tris (pH8.5), 5 mM Mgcl2, and 4 mM DTT] supplemented with 0.76 µM tritium labelled Adomet in presence of SUV39H1 enzyme at 25^{0}C for 5 to 6 h. The reactions were stopped by the addition of SDS loading dye and followed by boiling at 95^{0}C for 5 minutes.

NSD1 methylation buffer

Protein domain methylation reactions were performed in methylation buffer [50 mM Tris (pH 9.0), 5 mM Mgcl2, and 1 mM DTT] supplemented with 0.76 µM tritium labelled Adomet in presence of NSD1 enzyme at 37^{0}C for 5 to 6 h. The reactions were stopped by the addition of SDS loading dye and followed by boiling at 95^{0}C for 5 minutes.

SET8 methylation buffer

Protein domain methylation reactions were performed in methylation buffer [50 mM Tris (pH 9.0), 100 mM Nacl, and 5 mM DTT] supplemented with 0.76 µM tritium labelled Adomet in presence of SET8 enzyme at 37^{0}C for 5 to 6 h. The reactions were stopped by the addition of SDS loading dye and followed by boiling at 95^{0}C for 5 minutes.

SMYD2 methylation buffer

Protein domain methylation reactions were performed in methylation buffer [50 mM Tris (pH 9.0) 5 mM DTT and 100 mM NaCl] supplemented with 0.76 µM tritium labelled Adomet in presence of SMYD2 enzyme at 37^{0}C for 5 to 6 h. The reactions were stopped by the addition of SDS loading dye and followed by boiling at 95^{0}C for 5 minutes.

SDS-PAGE and Autoradiography

Methylated reaction products were separated on a 16% SDS-PAGE gel, the gel then washed with Amplify NAMP100V solution (GE Healthcare) for 45 min and dried on Whatman paper (Whatman GmbH) before being incubated with Hyperfilm^{TM} high performance autoradiography films (GE Healthcare) in the dark at -80 ^{0}C for 1-4 days. The films were developed using AGFA Curix 60 developing machine (Agfa Deutschland Vertriebsgesellschaft mbH & Co. KG) (Rathert et al., 2008).

Synthesis and Methylation of Peptide Arrays

Peptide arrays were synthesized as described using the SPOT synthesis method and using Multipep RS^{TM} (Intavis AG) (Frank et al., 2002 and Wenschuh et al., 2000) . Methylation of the arrays containing peptides was also carried out as descibed (Rather et al., 2008). Briefly, membranes containing the peptide spots were incubated in a methylation buffer for 5 min. Membranes were then incubated in the same buffer with 0.35 µM labeled [methyl-^{3}H]-AdoMet and with corresponding enzyme for 1-1.5 h. Afterwards, the membranes were washed with 100 mM NH_4HCO_3 with 0.1% SDS five times for 5 min each and incubated for 10 min in Amplify NAMP100V solution (GE Healthcare) before exposure to Hyperfilm^{TM}autoradiography films (GE Healthcare). Development was done similarly as for the SDS-PAGE gels.

Strategy for generation of randomized peptide arrays

For the first randomized peptide array, the target lysine was place in the center of a 15-mer peptide. Subsequently, the residues immediately adjacent to it (at +1 and -1 position) were substituted by each of 17 aminoacids (Cysteine, Methionine and Tryptophan were excluded) including every possible permutation. This was a total of 17X17 = 289 peptides. The remaining 12 residues (6 on each side) were randomly assigned such that there was a statistical representation of each amino acid at each position. For the second randomized peptide array, a Lysine-Leucine-Lysine central tri-peptide remained unchanged while the residues immediately adjacent to it were permutated in the same way as for the first randomized array. The remaining 10 aminoacids (5 at each terminus) were picked at random in a similar manner as for the first randomization.

Solid-phase peptide synthesis:

Peptides were synthesised by following Fmoc solid-phase peptide synthesis using Multipep RS^{TM} (Intavis AG) on resin. After synthesis, quality of the peptides were analysed by MALDI.

Binding of protein domains to peptide arrays

Cellulose membrane containing peptide arrays was blocked in TTBS buffer containing 5% milk powder [10 mM Tris (pH 8.5), 0.05% tween-20 and 150 mM Nacl] overnight. The membrane was then washed with 2 times in TTBS buffer and incubated with the interested purified protein domain (10 to 50 nM) at room temperatue for 1 hour in interaction buffer [100 mM KCl, 20 mM HEPES (pH 7.5), 1 mM EDTA, 0.1 mM DTT and 10% glycerol]. Mmebrane was washed again with TTBS buffer for 2 times followed by incubation with anti- GST antibody (GE Healthcare #27- 4577-01, at 1:5000 dilution in TTBS) for 1 hour at room temperature. Then membrane was washed with TTBS buffer for 2 times and incubated with secondary antiobody [horseradish peroxidise conjugated anti-Goat antibody (Invitrogen#81-1620)]. Finally wash the membrane with TTBS buffer for 4 to 5 times and detected the signal by ECL developing solution (GE Healthcare) and image was captured in X-ray film.

Cell cuture transfection:

HEK 293 or HELA cells were seeded in T25 flasks, after attaining 60% confluence the cells were transfected with the target protein domain construct together with or without HKMT's using the transfection reagent fugene6 (Roche) according to the manufacturer's protocol. Two days after transfection, the cells were harvested and nuclear extracts were prepared as described (Andrews and Jones, 1991). Immunoprecipitation of the target protein domains were carried out using the GFP-Trap (Chromotek^{TM}-gtm-100) according to the manufacturer's instructions.

Similarly for the sub-nuclear localisation studies NIH-3T3 cells were seeded on the coverslips and transfected with the RAG2 or SET8 proteins together with or with out SUV39H1 protein using the the transfection reagent fugene6 (Roche). 30 hours after transfection cells were washed with PBS buffer and fixed with paraformaldehyde. Finally the cells were embeded with Mowiol (Carl Roth) using nail lock. Confocal images were taken using a Zeiss LSM 510 Meta (software version 3.0) and oil immersion objectives.

Histones Isolation:

H1 histone proteins were extracted with perchloric acid. Hela cells or HEK293 cells were seeded in T25 flasks. After 48 hours of transfection with H1.5 protein the cells were harvested and washed with PBS (phosphate buffer saline). Then cell pellet was dissolved in 0.83 M perchloric acid and cells were lysed and extracted by incubation for 1 hour on ice. Samples were centrifuged (10 min, 4^{0}C, 13500 RPM). Acid soluble proteins in the supernatant were precipitated with 20% TCA (w/v) for 1 hour. After centrifugation pellet was collected and washed with ice cold acetone and dried. Finally the dried histones were dissolved in 30 mM HCL (Nicole et al., 2005). H3 and H4 histones from the cell lines were isolated by acid extraction method as described (Shechter et al., 2007).

Mass Spectrometry Analysis

The immunoprecipitated target protein domains were separated on 12% SDSPAGE and the bands of expected size were excised and further processed for MALDI mass spectrometric analysis using an Autoflex II device (Bruker Daltonics) as described (Shevchenko et al., 2006). After overnight in gel digestion of the target protein with 1 mg of Trypsin Gold (Promega) in a reaction volume of 50 ml, the peptides samples were diluted 1:5 with 0.1% TFA and subjected to MALDI analysis as described below. In case of competitive methylation of H3 and H1.5 peptides, 1 ml of the reaction mixtures were diluted 1:10 in 0.1% TFA and subjected to MALDI analysis as described below. One microliter of the peptide dilution was applied to one spot on a prespotted Anchorchip (PAC) HCCA Plate (Cat. No. 227463, Bruker Daltonics) and washed with 10 mM sodium phosphate / 0.1% TFA solution. Spectra were recorded using an AutoFlex II (Bruker Daltonics, Bremen, Germany) with default settings using peptide calibration standard mixture with the mass range of 1000 to 4000 Da (Cat. No. 206195, Bruker Daltonics) and processed using the FlexAnalysis software (Bruker Daltonics). Protein methylation and phosphorylation was investigated using the Biotools program (Bruker Daltonics).

8. References

Abu-Farha M, Lambert JP, Al-Madhoun AS, Elisma F, Skerjanc IS, Figeys D. (2008) The Tale of Two Domains PROTEOMICS AND GENOMICS ANALYSIS OF SMYD2, A NEW HISTONE METHYLTRANSFERASE. Molecular & Cellular Proteomics. 7(3): 560-572.

Akamatsu Y, Monroe R, Dudley DD, Elkin SK, Gartner F, Talukder SR, Takahama Y, Alt FW, Bassing CH, Oettinger MA. (2003) Deletion of the RAG2 C terminus leads to impaired lymphoid development in mice. PNAS. 100(3): 1209-1214.

Allis CD, Berger SL, Cote J, Dent S, Jenuwien T, Kouzarides T, Pillus L, Reinberg D, Shi Y, Shiekhattar R, Shilatifard A, Workman J, Zhang Y.(2007) New nomenclature for Chromatin-Modifying Enzymes. Cell. 131:633-636.

Bannister AJ, Schneider R, Kouzarides T. (2002) Histone Methylation: Dynamic or Static?. Cell. 109, 801–806.

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature. 410: 120-124.

Berdasco M, Ropero S, Setien F, Fraga MF, Lapunzina P, Losson R, Alaminos M, Cheung NK, Rahman N, Esteller M. (2009) Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. PNAS. 106(51): 21830-21835.

Bhaumik SR, Smith E, Shilatifard A. (2007) Covalent modifications of histones during development and disease pathogenesis. NATURE STRUCTURAL & MOLECULAR BIOLOGY. 14: 1008-1016.

Bonasio R, Tu S, Reinberg D. (2010) Molecular Signals of Epigenetic States. Science. 330: 612-616.

Bottomley MJ. (2004) Structures of protein domains that create or recognize histone modifications. EMBO reports. 5: 464-469.

Brower V. (2011) Unravelling the cancer code. Nature. 471: S12-13.

Brown MA, Sims RJ 3rd, Gottlieb PD, Tucker PW. (2006) Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. Mol. Cancer. 5(26): 1-11.

Chang Y, Levy D, Horton JR, Peng J, Zhang X, Gozani O, Cheng X. (2011) Structural basis of SETD6-mediated regulation of the NF-kB network via methyl-lysine signalling . Nucleic Acids Research. 1-10.

Chin HG, Estève PO, Pradhan M, Benner J, Patnaik D, Carey MF, Pradhan S. (2007) Automethylation of G9a and its implication in wider substrate specificity and HP1 binding. Nucleic Acids Research. 35(21): 7313–7323.

Chuikov S, Kurash JK, Wilson JR, Xiao B, Justin N, Ivanov GS, McKinney K, Tempst P, Prives C, Gamblin SJ, Barlev NA, Reinberg D. (2004) Regulation of p53 activity through lysine methylation. Nature. 432: 353-360.

Ciccone DN, Chen T. (2009) Histone lysine methylation in genomic imprinting. Epigenetics. 4: 216-220.

Corneo B, Benmerah A, Villartay JP. (2002) A short peptide at the C terminus is responsible for the nuclear localization of RAG2. Eur. J. immunology. 32: 2068-2073.

Couture JF, Collazo E, Brunzelle JS, Trievel RC. (2005) Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. Genes and development. 19(12):1455-65.

Couture JF, Trievel RC. (2006) Histone-modifying enzymes: encrypting an enigmatic epigenetic code. Current Opinion in Structural Biology. 16: 753–760.

Daniel JA, Pray-Grant MG, Grant PA. (2005) Effector Proteins for Methylated Histones, An expanding family. Cell cycle 4(7): 919-926.

Dhayalan A, Tamas R, Bock I, Tattermusch A, Dimitrova E, Kudithipudi S, Ragozin S, Jeltsch A. (2011) The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9. Hum Mol Genet. 20(11): 2195-2203.

Dhayalan A, Kudithipudi S, Rathert P, Jeltsch A. (2011) Specificity Analysis-Based Identification of New Methylation Targets of the SET7/9 Protein Lysine Methyltransferase. Chemistry & Biology. 18: 1–10.

Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A. (2009) The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation 2009. JBC. 285(34): 114–120.

Fang J, Feng Q, Ketel CS, Wang H, Cao R, Xia L, Erdjument-Bromage H, Tempst P, Simon JA, Zhang Y. (2002) Purification and Functional Characterization of SET8, a Nucleosomal Histone H4-Lysine 20-Specific Methyltransferase. Current Biology. 12(13):1086-99

Faravelli F. (2005) NSD1 Mutations in Sotos Syndrome. American Journal of Medical Genetics Part C (Semin. Med. Genet.) 137C: 24–31 .

Feil R. (2008) Epigenetics, an emerging discipline with broad implications. CR biologies. 331(11): 837-843.

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

Frankel A, Yadav N, Lee J, Branscombe TL, Clarke S, Bedford MT. (2002) The Novel Human Protein Arginine N-Methyltransferase PRMT6 Is a Nuclear Enzyme Displaying Unique Substrate Specificity. JBC. 277(5): 3537–3543.

Fritsch L, Robin P, Mathieu JR, Souidi M, Hinaux H, Rougeulle C, Harel-Bellan A, Ameyar-Zazoua M, Ait-Si-Ali S. (2009) A Subset of the Histone H3 Lysine 9 Methyltransferases Suv39h1, G9a, GLP, and SETDB1 Participate in a Multimeric omplex. Molecular Cell 37: 46–56.

Fuks F, Hurd PJ, Deplus R, Kouzarides T. (2003) The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Research. 31(9): 2305-2312

Garcia BA, Hake SB, Diaz RL, Kauer M, Morris SA, Recht J, Shabanowitz J, Mishra N, Strahl BD, Allis CD, Hunt DF. (2006) Organismal Differences in Post-translational Modifications in Histones H3 and H4. JBC. 282(10): 7641–7655.

Gottlieb PD, Pierce SA, Sims RJ, Yamagishi H, Weihe EK, Harriss JV, Maika SD, Kuziel WA, King HL, Olson EN, Nakagawa O, Srivastava D. (2002) Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis. Nature Genetics. 31: 25-32.

Gray SG, Iglesias AH, Lizcano F, Villanueva R, Camelo S, Jingu H, Teh BT, Koibuchi N, Chin WW, Kokkotou E, Dangond F. (2005) Functional Characterization of JMJD2A, a Histone Deacetylase- and Retinoblastoma-binding Protein. JBC. 281 (31): 28507-28518.

Grundy GJ, Yang W, Gellert M. (2010) Autoinhibition of DNA cleavage mediated by RAG1 and RAG2 is overcome by an epigenetic signal in V(D)J recombination. 107(52): 22487-22492.

Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, Yagyu R, Nakamura Y. (2004) SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. NATURE CELL BIOLOGY VOLUME. 6(8): 731-740.

Happel N, Doenecke D. (2009) Histone H1 and its isoforms: Contribution to chromatin structure and function. Gene. 431: 1-12.

Happel N, Warneboldt J, Hänecke K, Haller F, Doenecke D. (2009) H1 subtype expression during cell proliferation and growth arrest. Cell Cycle 8:14, 2226-2232.

Holliday R. (1994) Developmental genetics. 15: 453-457.

Houston SI, McManus KJ, Adams MM, Sims JK, Carpenter PB, Hendzel MJ, Rice JC. (2008) Catalytic Function of the PR-Set7 Histone H4 Lysine 20 Monomethyltransferase Is Essential for Mitotic Entry and Genomic Stability. JBC. 283(28):19478-88.

Huang J, Berger SL. (2008) The emerging field of dynamic lysine methylation of non-histone proteins. Current Opinion in Genetics & Development. 18:152–158.

Huang J, Dorsey J, Chuikov S, Pérez-Burgos L, Zhang X, Jenuwein T, Reinberg D, Berger SL. (2010) G9a and Glp Methylate Lysine 373 in the Tumor Suppressor p53. JBC. 285(13): 9636–9641.

Huang J, Perez-Burgos L, Placek BJ, Sengupta R, Richter M, Dorsey JA, Kubicek S, Opravil S, Jenuwein T, Berger SL. (2006) Repression of p53 activity by Smyd2-mediated methylation. Nature. 444: 629-632.

Huang J, Sengupta R, Espejo AB, Lee MG, Dorsey JA, Richter M, Opravil S, Shiekhattar R, Bedford MT, Jenuwein T, Berger SL. (2007) p53 is regulated by the lysine demethylase LSD1. Nature. 449: 105-109.

Huang N, vom Baur E, Garnier JM, Lerouge T, Vonesch JL, Lutz Y, Chambon P, Losson R. (1998) Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. The EMBO Journal. 17(12):3398–3412.

Huang Y, Fang J, Bedford MT, Zhang Y, Xu RM. (2006) Recognition of Histone H3 Lysine-4 Methylation by the Double Tudor Domain of JMJD2A. Science. 312: 748-751.

Jenuwein T, Allis CD. (2001) Translating the Histone Code. Science. 293: 1074-1080.

Jiang H, Chang FC, Ross AE, Lee J, Nakayama K, Nakayama K, Desiderio S. (2005) Ubiquitylation of RAG-2 by Skp2-SCF Links Destruction of the V(D)J Recombinase to the Cell Cycle. Molecular Cell. 18: 699–709.

Jørgensen S, Eskildsen M, Fugger K, Hansen L, Larsen MS, Kousholt AN, Syljuåsen RG, Trelle MB, Jensen ON, Helin K, Sørensen CS. (2011) SET8 is degraded via PCNA-coupled CRL4(CDT2) ubiquitylation in S phase and after UV irradiation. Journal of cell biology. 192: 43-54.

Kachirskaia I, Shi X, Yamaguchi H, Tanoue K, Wen H, Wang EW, Appella E, Gozani O. (2008) Role for 53BP1 Tudor Domain Recognition of p53 Dimethylated at Lysine 382 in DNA Damage Signaling. JBC. 283(50): 34660–34666.

Kamimura J, Endo Y, Kurotaki N, Kinoshita A, Miyake N, Shimokawa O, Harada N, Visser R, Ohashi H, Miyakawa K, Gerritsen J, Innes AM, Lagace L, Frydman M, Okamoto N,Puttinger R, Raskin S, Resic B, Culic V, Yoshiura K, Ohta T, Kishino T, Ishikawa M, Niikawa N, Matsumoto N. (2003) Identification of eight novel NSD1 mutations in Sotos syndrome. J Med Genet. 40: 1-3.

Kaufman PD, Rando OJ. (2010) Chromatin as a potential carrier of heritable information. Current Opinion in Cell Biology. 22: 284–290.

Kim J, Daniel J, Espejo A, Lake A, Krishna M, Xia L, Zhang Y, Bedford MT. (2006) Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO reports. 7(4): 397-403.

Kimura A, Matsubara K, Horikoshi M. (2005) A Decade of Histone Acetylation: Marking EukaryoticChromosomes with Specific Codes. J. Biochem. 138, 647–662.

Komatsu S, Imoto I, Tsuda H, Kozaki KI, Muramatsu T, Shimada Y, Aiko S, Yoshizumi Y, Ichikawa D, Otsuji E, Inazawa J. (2009) Overexpression of SMYD2 relates to tumor cell proliferation and malignant outcome of esophageal squamous cell carcinoma. Carcinogenesis. 30(7):1139–1146.

Kouskouti A, Scheer E, Staub A, Tora L, Talianidis I. (2004) Gene-Specific Modulation of TAF10 Function by SET9-Mediated Methylation. Molecular Cell. 14: 175–182.

Krauss V. (2008) Glimpses of evolution: heterochromatic histone H3K9 methyltransferases left its marks behind. Genetica. 133:93–106.

Kuzmichev A, Jenuwein T, Tempst P, Reinberg D. (2004) Different Ezh2-Containing Complexes Target Methylation of Histone H1 or Nucleosomal Histone H3. Molecular Cell. 14:183–193.

Li Y, Trojer P, Xu CF, Cheung P, Kuo A, Drury WJ 3rd, Qiao Q, Neubert TA, Xu RM, Gozani O, Reinberg D. (2009) The Target of the NSD Family of Histone Lysine Methyltransferases Depends on the Nature of the Substrate. JBC. 284(49): 34283–34295.

Lu A, Zougman A, Pudełko M, Bebenek M, Ziółkowski P, Mann M, Wiśniewski JR. (2009) Mapping of Lysine Monomethylation of Linker Histones in Human Breast and Its Cancer. Journal of Proteome Research. 8: 4207–4215.

Lu T, Jackson MW, Wang B, Yang M, Chance MR, Miyagi M, Gudkov AV, Stark GR. (2010) Regulation of NF-κB by NSD1/FBXL11-dependent reversible lysine methylation of p65. PNAS. 107: 46-51.

Lucio-Eterovic AK, Singh MM, Gardner JE, Veerappan CS, Rice JC, Carpenter PB. (2010) Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3with RNApolymerase II function. PNAS. 107(39): 16952-16957.

Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ 3rd, Voigt P, Martin SR, Taylor WR, De Marco V, Pirrotta V, Reinberg D, Gamblin SJ. (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. Nature. 461: 762-769.

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

Martin C, Zhang Y. (2005) THE DIVERSE FUNCTIONS OF HISTONE LYSINE METHYLATION. Nature reviews. 6: 838-849.

Matthews AG, Kuo AJ, Ramón-Maiques S, Han S, Champagne KS, Ivanov D, Gallardo M, Carney D, Cheung P, Ciccone DN, Walter KL, Utz PJ, Shi Y, Kutateladze TG, Yang W,Gozani O, Oettinger MA. (2007) RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. Nature. 450: 1106-1111.

McBlane JF, van Gent DC, Ramsden DA, Romeo C, Cuomo CA, Gellert M, Oettinger MA. (1995) Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. Cell. 83; 387-395.

Morgunkova A, Barlev NA. (2006) Lysine Methylation Goes Global. Cell Cycle. 5(12): 1308-1312.

Ng SS, Yue WW, Oppermann U, Klose RJ. (2009) Dynamic protein methylation in chromatin biology. Cell. Mol. Life Sci. 66: 407 – 422.

Nishioka K, Chuikov S, Sarma K, Erdjument-Bromage H, Allis CD, Tempst P, Reinberg D. (2002) Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev. 16: 479-489.

Oda H, Hübner MR, Beck DB, Vermeulen M, Hurwitz J, Spector DL, Reinberg D. (2010) Regulation of the Histone H4 Monomethylase PR-Set7 by CRL4Cdt2-Mediated PCNA-Dependent Degradation during DNA Damage. Molecular Cell. 40: 364-376.

Pasillas MP, Shah M, Kamps MP. (2011) NSD1 PHD domains bind methylated H3K4 and H3K9 using interactions disrupted by point mutations in human sotos syndrome. Human Mutation. 32(3): 292-298.

Pei H, Zhang L, Luo K, Qin Y, Chesi M, Fei F, Bergsagel PL, Wang L, You Z, Lou Z. (2011) MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. Nature. 470: 125-129.

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

Qiao Q, Li Y, Chen Z, Wang M, Reinberg D, Xu RM. (2011) The Structure of NSD1 Reveals an Autoregulatory Mechanism Underlying Histone H3K36 Methylation. JBC. 286(10): 8361–8368.

Rando OJ, Chang HY. (2009) Genome- Wide Views of Chromatin Structure. Annual review Biochemistry. 78:245-71.

Rathert P, Dhayalan A, Murakami M, Zhang X, Tamas R, Jurkowska R, Komatsu Y, Shinkai Y, Cheng X, Jeltsch A. (2008) Protein lysine methyltransferase G9a acts on non-histone targets. Nature Chemical Biology. 4: 344-346.

Rayasam GV, Wendling O, Angrand PO, Mark M, Niederreither K, Song L, Lerouge T, Hager GL, Chambon P, Losson R. (2003) NSD1 is essential for early post-implantation development and has a catalytically active SET domain. The EMBO Journal. 22(12): 3153-3163.

Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, Jenuwein T. (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature. 406: 593-599.

Rivera RM, Bennett LB. (2010) Epigenetics in humans: an overview. Current Opinion in Endocrinology, Diabetes & Obesity. 17:493–499.

Saddic LA, West LE, Aslanian A, Yates JR 3rd, Rubin SM, Gozani O, Sage J. (2010) METHYLATION OF THE RETINOBLASTOMA TUMOR SUPPRESSOR BY SMYD2. JBC. 285(48) 37733-37740.

Schotta G, Ebert A, Reuter G. (2003) SU(VAR)3-9 is a conserved key function in heterochromatic gene silencing. Genetica. 117: 149–158.

Scoumanne A, Chen X. (2008) Protein methylation: a new mechanism of the p53 tumor suppressor. Histol Histopathol. 23(9): 1143–1149.

Shechter D, Dormann HL, Allis CD, Hake SB. (2007) Extraction, purification and analysis of histones. Nature protocols. 2(6): 1445-1457.

Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nature. 6: 2856-2860.

Shi X, Kachirskaia I, Yamaguchi H, West LE, Wen H, Wang EW, Dutta S, Appella E, Gozani O. (2007) Modulation of p53 Function by SET8-Mediated Methylation at Lysine 382. Molecular cell. 27(4):636-46.

Sims RJ 3rd, Nishioka K, Reinberg D. (2003) Histone lysine methylation: a signature for chromatin function. TRENDS in genetics. 19: 629-639.

Sims RJ 3rd, Reinberg D. (2008) Is there a code embedded in proteins that is based on post-translational modifications?. Nature reviews. 9: 1-6.

Spanopoulou E, Cortes P, Shih C, Huang CM, Silver DP, Svec P, Baltimore D. (1995) Localization, interaction, and RNA binding properties of the V(D)J recombination-activating proteins RAG1 and RAG2. Immunity. 3: 715-726.

Strahl BD, Allis CD. (2000) The language of covalent histone modifications. Nature. 403: 41-45.

Tanaka Y, Katagiri Z, Kawahashi K, Kioussis D, Kitajima S. (2007) Trithorax-group protein ASH1 methylates histone H3 lysine 36. Gene 397: 161–168.

Tatton-Brown K, Douglas J, Coleman K, Baujat G, Cole TR, Das S, Horn D, Hughes HE, Temple IK, Faravelli F, Waggoner D, Turkmen S, Cormier-Daire V, Irrthum A, Rahman N. (2005) Genotype-Phenotype Associations in Sotos Syndrome: An Analysis of 266 Individuals with NSD1 Aberrations. Am. J. Hum. Genet. 77:193–204.

Trojer P, Zhang J, Yonezawa M, Schmidt A, Zheng H, Jenuwein T, Reinberg D. (2009) Dynamic Histone H1 Isotype 4 Methylation and Demethylation by Histone Lysine Methyltransferase G9a/KMT1C and the Jumonji Domain-containing JMJD2/KDM4 Proteins. JBC. 284(13):8395–8405.

Upadhyay AK, Cheng X. (2011) Dynamics of Histone Lysine Methylation: Structures of Methyl Writers and Eraser. Prog Drug Res. 67: 107–124.

Verdone L, Agricola E, Caserta M, Di Mauro E. (2006) Histone acetylation in gene regulation. BRIEFINGS IN FUNCTIONAL GENOMICS AND PROTEOMICS. 3: 209-221.

Verdone L, Caserta M, Di Mauro E. (2006) Role of histone acetylation in the control of gene expression. Biochem. Cell Biol. 83: 344–353.

Völkel P, Angrand PO. (2007) The control of histone lysine methylation in epigenetic regulation. Biochimie. 89 :1-20.

Wang GG, Cai L, Pasillas MP, Kamps MP. (2006) NUP98–NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. Nature cell biology. 9(7): 804-812.

Weiss T, Hergeth S, Zeissler U, Izzo A, Tropberger P, Zee BM, Dundr M, Garcia BA, Daujat S, Schneider R. (2010) Histone H1 variant-specific lysine methylation by G9a/KMT1C and Glp1/KMT1D. Epigenetics & Chromatin. 3(7): 1-13.

Wenschuh H, Volkmer-Engert R, Schmidt M, Schulz M, Schneider-Mergener J, Reineke U. (2000) Coherent membrane supports for parallel microsynthesis and screening of bioactive peptides. Biopolymers. 55(3): p. 188-206.

West LE, Roy S, Lachmi-Weiner K, Hayashi R, Shi X, Appella E, Kutateladze TG, Gozani O. (2010) The MBT Repeats of L3MBTL1 Link SET8-mediated p53 Methylation at Lysine 382 to Target Gene Repression. JBC. 285: 37725-37732.

Wisniewski JR, Zougman A, Krüger S, Mann M. (2007) Mass Spectrometric Mapping of Linker Histone H1 Variants Reveals Multiple Acetylations, Methylations, and Phosphorylation as Well as Differences between Cell Culture and Tissue. Molecular & Cellular Proteomics. 6(1): 72-87.

Xiao B, Wilson JR, Gamblin SJ. (2003) SET domains and histone methylation. Current Opinion in Structural Biology. 13:699–705.

Xu S, Zhong C, Zhang T, Ding J. (2011) Structure of human lysine methyltransferase Smyd2 reveals insights into the substrate divergence in Smyd proteins. Journal of Molecular Cell Biology. 0:1–8.

Yin Y, Liu C, Tsai SN, Zhou B, Ngai SM, Zhu G. (2005) SET8 Recognizes the Sequence RHRK20VLRDN within the N Terminus of Histone H4 and Mono-methylates Lysine 20. JBC. 280(34):30025-31.

Yin Y, Yu VC, Zhu G, Chang DC. (2008) SET8 plays a role in controlling G1/S transition by blocking lysine acetylation in histone through binding to H4 N-terminal tail. Cell Cycle 7(10): 1423-1432.

Yuan W, Xie J, Long C, Erdjument-Bromage H, Ding X, Zheng Y, Tempst P, Chen S, Zhu B, Reinberg D. (2009) Heterogeneous Nuclear Ribonucleoprotein L Is a Subunit of Human KMT3a/Set2 Complex Required for H3 Lys-36 Trimethylation Activity in Vivo. 2009. JBC. 284(23): 15701–15707.

Yun M, Wu J, Workman JL, Li B. (2011) Readers of histone modifications. Cell Research. 21:564-578.

Zhang D, Yoon HG, Wong J. (2005) JMJD2A Is a Novel N-CoR-Interacting Protein and Is Involved in Repression of the Human Transcription Factor Achaete Scute-Like Homologue 2 (ASCL2/Hash2). Molecular and Cellular Biology. 25 (15): 6404-6414.

Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, Rathert P, Brandt O, Reinhardt R, Fischle W, Jeltsch A. (2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Research. 38(13): 4246-4253.

Zhang, L., and Freitas, M. A. (2004) Comparison of peptide mass mapping and electron capture dissociation as assays for histone posttranslational modifications. International Journal of Mass Spectrometry. 234: 213–225.

Publications have been removed from the thesis due to copyrights issue