Elucidating the biochemical and structural features required for SMYD5 mediated methylation of histone H4 and other potential substrates

Thesis Submitted to the Faculty of Graduate and Postdoctoral Studies

University of Ottawa

In Partial Fulfillment of the Requirements for the Degree of Master of Science

Department of Biochemistry, Microbiology and Immunology

Faculty of Medicine

By Vanessa Mongeon

Abstract

Lysine methylation modulates diverse biological processes and is catalyzed by SET domain methyltransferases such as the SMYDs (SMYD1-5), which possess a SET domain split by a MYND motif. Through association with NCoR, the H4 Lys20 methyltransferase activity of SMYD5 represses inflammation by restricting TLR-4 mediated expression in macrophages, yet biochemical and structural features required for SMYD5 methylation activity remain elusive. To determine how SMYD5 catalyses methylation, crystallization screens were conducted with SMYD5 in complex with the co-factor AdoMet and histone H4. Screens yielded lead conditions but no crystals. To determine the motif recognized by SMYD5 and decipher its methylome, peptide arrays were conducted to produce a methylation motif used to identify putative substrates. Surprisingly, arrays revealed that substitution of Lys16, not Lys20, is detrimental to SMYD5 activity. Further enzymatic assays are required to determine if SMYD5 methylates residues other than Lys20 on the H4 tail, or if structural determinants or interacting partners restrict methylation of target lysines.

Acknowledgements

Firstly, I would like to thank Dr. Jean-François Couture for giving me the opportunity to be a member of his lab. During the course of the last two years, I have learned under his guidance not only about lysine methylation and structural biology, but how to tackle the metaphorical wall and attack a problem at different angles. I give thanks for his support and guidance, as well as the opportunity to learn more about science and about myself.

I would also like to thank members of the Couture lab, both past and present, for the great atmosphere in the lab as well as out. Thank you to Veronique for instructing me during the first few months in the lab and for being there when I needed advice. Thanks to Sylvain for fruitful discussion regarding the SMYDs and science in general, as well as getting me involved in a supporting role for the BMIGSA. Thanks to Elisa, especially for the work with the nucleosomes and helping me get through the last hurdles. And of course, thanks to Myriam, Pam and Sabina for the help in the lab and the stimulating lunchtimes.

Finally, I would like to thank my friends and family. Their support and encouragement helped me get through the joys and the hardships both in life and during the course of my studies.

Statement of Contribution

All experiments were conducted by Vanessa Mongeon (VM), with the exception of the limited proteolysis and ESI-MS experiments in Figure 12a-b, which were conducted by Véronique Tremblay. Experimental design was conducted by Jean-François Couture (JFC) and VM, with the exception of the design of the peptide arrays, which were designed by JFC. Data analysis for the peptide arrays was conducted by VM with the advice and mentoring of Sylvain Lanouette.

Table of Contents

List of Figures

List of Tables

Abbreviations

AcOH - Acetic Acid

AdoMet - S-adenosyl-L-methionine

Ala - Alanine

Arg - Arginine

ATP - Adenosine Triphosphate

BME - β-mercaptoethanol

bp - Base Pairs

BSA - Bovine Serum Albumin

CHD - Chromodomain, Helicase, DNA Binding

D. melanogaster - *Drosophila melanogaster*

Da - Daltons

DMSO - Dimethyl Sulfoxide

DNA - Deoxyribonucleic Acid

DTT - Dithiothreitol

EDTA - Ethylenediaminetetraacetic Acid

EtOH - Ethanol

G. gallus - *Gallus gallus*

Gln - Glutamine

Gly - Glycine

H. sapiens - *Homo sapiens*

His - Histidine

IPTG - Isopropyl-β-D-1-Thiogalactopyranoside

ISWI - Imitation Switch

Lys - Lysine

MYND - Myeloid translocation protein 8, Nervy and DEAF-1

NaCl - Sodium Chloride

$NaHCO_3$ - Sodium Bicarbonate

NaPi - Sodium Phosphate

NMR - Nuclear Magnetic Resonance

NuRD - Nucleosome Remodeling and Deacetylation

PBS - Phosphate Buffered Saline

PKMT - Protein Lysine Methyltransferase

PTM - Post-Translational Modification

SDS - Sodium Dodecyl Sulfate

SET - Suppressor of Variegation, Enhancer of Zeste and Trithorax

SMYD - SET and MYND Domain Protein

SWI/SNF - Switching Defective/Sucrose Nonfermenting

TBS - Tris Buffered Saline

TOF - Time of Flight Mass Analyzer

TOTD - Tioxatridecanediamine

TPR - Tetratrico Peptide Repeat

$ZnCl_2$ - Zinc Chloride

Introduction

1. Chromatin

1.1 The compaction of genetic material

Two decades following the observation that genetic material and its double helix structure (Watson and Crick, 1953) was in reality encircling a core histone octamer, the nucleosome was described as the fundamental unit of genomic compaction (Kornberg, 1974; Kornberg and Thomas, 1974). The octamer is comprised of small, evolutionary conserved basic proteins called histones, and is organized in two histone dimers of H2A-H2B and one tetramer of H3-H4 (Kelley, 1973; Roark et al., 1974) **(Fig. 1A)**. The histones provide a scaffold via ionic interactions between the negatively charged phosphates in the deoxyribonucleic acid (DNA) backbone and the positively charged lysine (Lys) and arginine (Arg) histone residues (Luger et al., 1997). Each octamer is enwrapped with 146 base pairs (bp) of DNA (Noll and Kornberg, 1977) forming a 1 3/4 left handed superhelical coil around the octamer core (Finch et al., 1977).

In parallel, electron microscopy revealed that the assembly of nucleosomes spaced with DNA form a beaded fiber (Olins and Olins, 1974; Woodcock et al., 1976). This beaded fiber can be condensed to form the 30 nm filament with the help of the histone H1 linker (Thoma et al., 1979). Further levels of compaction produce the mitotic chromosome seen during metaphase **(Fig. 1B)**. Chromatin can exist in two forms, depending on the level of compaction of the nucleosomes.

Figure 1.

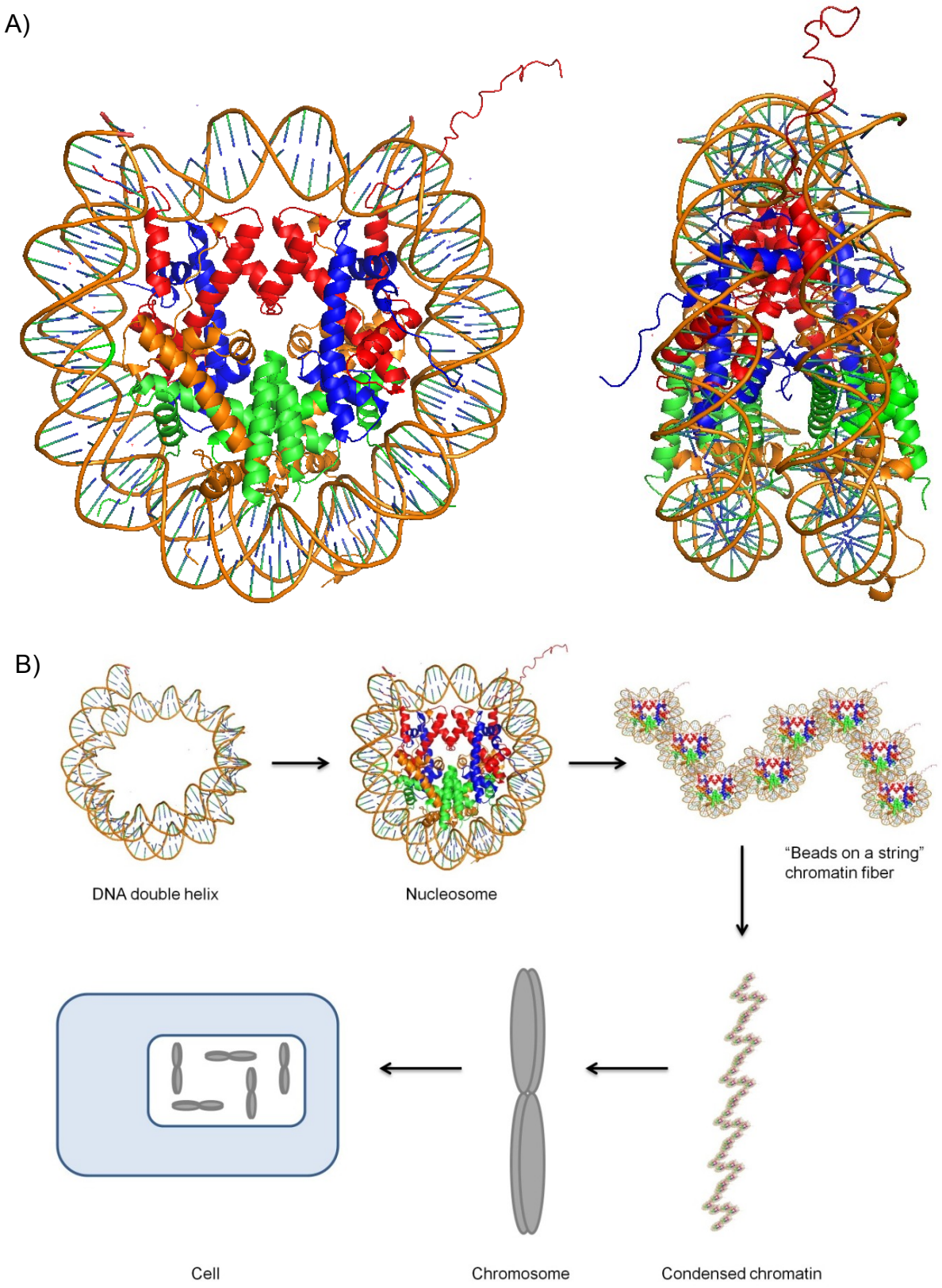

A)
B)
DNA double helix
Nucleosome
"Beads on a string" chromatin fiber
Cell
Chromosome
Condensed chromatin

Figure 1: Levels of compaction of genetic material in the form of DNA.
A) Ribbon structure of the nucleosome adapted with permission from NATURE PUBLISHING GROUP [Nature] (Luger et al., 1997), copyright 1997, reveals the scaffolding of DNA by the histone octamer core, with front and side views. DNA encircles histone H3 (red), H4 (blue), H2A (orange) and H2B (green). B) Different levels of condensation of DNA adapted with permission from NATURE PUBLISHING GROUP [Nature Reviews Nephrology] (Tonna et al., 2010), copyright 2010.

Heterochromatin represents the most condensed form, encompassing regions that displays little transcriptional activity, while euchromatin contains de-condensed DNA, rendering it available for the transcription machinery and gene expression (Woodcock and Ghosh, 2010). The histones possess amino-terminal tails (Kornberg and Lorch, 1999) available for reversible post-translational modifications which affect chromatin fiber formation and access to the transcription machinery (Hansen et al., 1998). To render genes and the genetic information either silent or available for transcription, chromatin has to be dynamic in response to numerous biological processes.

1.2 Chromatin dynamics

Chromatin configuration changes in response to developmental and environmental cues, playing a central role in the biology of a cell. Alterations in chromatin structure ease or restrict access to DNA for biological processes such as replication, gene expression and DNA repair, while also conferring stability and proper chromosome segregation during the cell cycle (Ehrenhofer-Murray, 2004). Errors and mis-regulation of the remodeling of chromatin can have adverse effects, producing mismatches during chromosome recombination or resulting in disease states such as cancer and genetic abnormalities (Bhaumik et al., 2007; Cho et al., 2004; Hock et al., 2007). Factors contributing to chromatin dynamics include adenosine triphosphate (ATP) -dependant chromatin remodeling complexes, histone chaperones, histone variants and covalent post-translational modifications of histones.

1.2.1 ATP- dependant chromatin remodeling complexes

Re-structuring of nucleosomes by ATP dependant chromatin remodeling complexes modulates the accessibility of DNA to non-histone proteins. This energy-dependant process is enabled through the hydrolysis of ATP by the complexes (Varga-Weisz, 2014), allowing the complex to slide or evict the nucleosomes, or exchange for histone variants (Wang et al., 2007) **(Fig. 2)**.

Members of different families of complexes possess different chromatin remodeling abilities. The SWI/SNF remodeling family (switching defective/sucrose nonfermenting) is required for transcription of different genes. These complexes render DNA available by the translocation of nucleosomes, they remodel nucleosomes by creating transient DNA bulges on the surface without disassembly, and also have the ability to disassemble nucleosomes (Dechassa et al., 2010). The ISWI family (imitation switch) slide the nucleosome (Ito et al., 1997; Varga-Weisz et al., 1997) and is involved in transcription activation and repression (Corona and Tamkun, 2004). The NuRD (nucleosome remodeling and deacetylation) /Mi-2/CHD (chromodmain, helicase, DNA binding) family is linked to the regulation of DNA damage responses and the reparation of double strand breaks (Smeenk et al., 2010). This family plays a role in gene transcription, DNA damage repair, the maintenance of genome stability, and chromatin assembly (Allen et al., 2013). The NuRD family of complexes also possess dual activity, with both an ATPase and histone deacetylase amongst its subunits (Allen et al., 2013). The INO80 (inositol requiring 80) family promote transcriptional activation and DNA repair, while family

Figure 2.

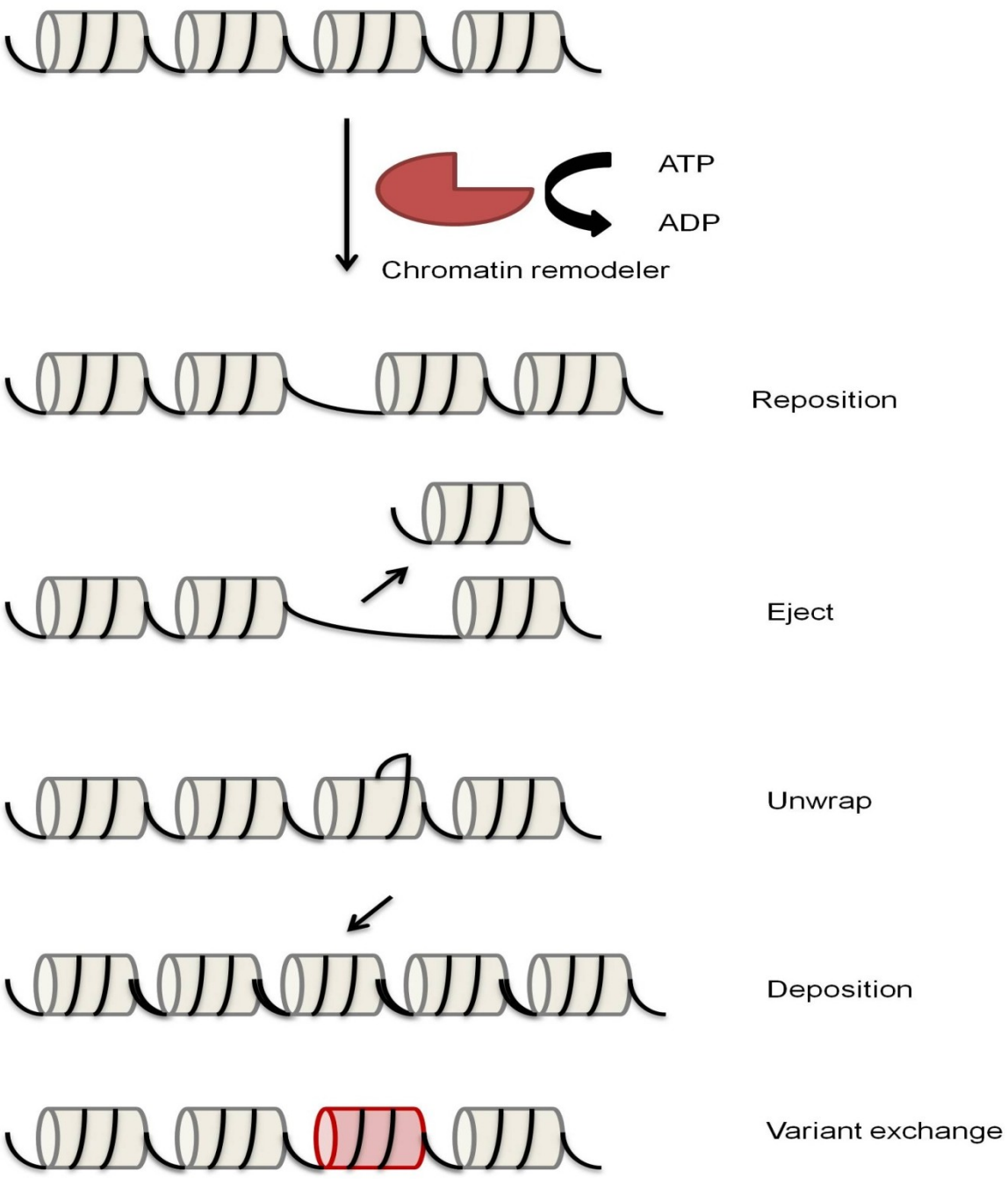

ATP
ADP
Chromatin remodeler
Reposition
Eject
Unwrap
Deposition
Variant exchange

Figure 2: Mechanisms of ATP-dependant remodeling complexes.
Adapted from ANNUAL REVIEWS [Annual Review of Biochemistry] (Clapier and Cairns, 2009), copyright 2009, this schematic demonstrates the remodeling complex rendering DNA accessible or inaccessible to DNA binding proteins. This can be done by exposing the site through repositioning, ejection, unwrapping of DNA, depositing core histones, or by changing the composition of the nucleosomes by exchanging variants of histones.

member SWR1 has the ability to replace H2A-H2B dimers with H2AZ-H2B dimers (Corona and Tamkun, 2004).

1.2.2 Histone chaperones

Histone chaperones do not directly modify chromatin structure but they play a key role in chromatin dynamics nonetheless, and are responsible for binding histones when they are not associated to DNA. Among their roles, they prevent aggregation and degradation of histones, transfer histones from one chaperone to another, stock histones prior to their use, and stabilize them during exchanges between canonical histones and their variants. Chaperones can function alone or in complex with each other or other enzymes, can shuttle histones between the cytoplasm and nucleus, and can be associated to chromatin remodeling complexes. The specific roles of the histone chaperone families and their members are extensively reviewed (De Koning et al., 2007; Loyola and Almouzni, 2004; Ray-Gallet and Almouzni, 2010).

1.2.3 Histone variants

While nucleosomes formed during replication are mainly assembled with the canonical histones H2A, H2B, H3, H4 and H1 expressed during S-phase, there also exist histone variants incorporated independently of replication. These variants possess different sequences, functions and methods of incorporation, allocating a specific structure and function to the chromatin with which it associates. They are deposited in specific regions of the genome and play important parts in cellular

processes ranging from transcription regulation, to DNA repair and heterochromatin formation. They are also responsible for the formation of specific structures, such as the variant CENP-A in the centromeres, and can have tissue specific functions, such as H3t and H3.5 located in the testis. The histone variants and their role, particularly in transcription, are comprehensively reviewed (Law and Cheung, 2013).

1.2.4 Post-translational modifications (PTMs) of histones

Post-translational modifications (PTMs) are reversible chemical modifications mostly deposited on the N- and C-terminal tails of histones. These marks, dynamically regulated, recruit non-histone proteins which interact with the chromatin and mediate biological processes including gene expression, the DNA damage response and cell survival (Karch et al., 2013). Their role in the regulation of transcription was first described with the identification of the acetylation and methylation marks by Allfrey and colleagues in the 1960's (Allfrey et al., 1964). Since, there has been interest in how these modifications affect DNA scaffolding, as well as which enzymes add, remove and recognize these marks. Although more information is available on acetylation, phosphorylation and methylation, there are a number of other PTMs that are being discovered and characterized, including ubiquitylation, sumoylation, ADP ribosylation, deimination, proline isomerization (Kouzarides, 2007) among others. Advances in proteomics allow the discovery and identification of novel PTMs, as well as new sites for these PTMs (Arnaudo and Garcia, 2013).

2. Lysine methylation

2.1 Lysine methylation of histones

Lysine methylation is a covalent PTM that modulates a myriad of biological processes. The presence of this PTM alters chromatin structure and function, which regulates the DNA damage responses, transcriptional activity and cell cycle progression (Kouzarides, 2007). Difficult to characterize in itself, its complexity also arises from the potential addition of one, two or three methyl moieties to the ε-amine (**Fig. 3**), producing multiple states for the single modification (Couture and Trievel, 2006).

This PTM was identified to be deposited on histone proteins by two groups in 1964 (Allfrey et al., 1964; Murray, 1964). While methylation sites on histones were identified, the biological functions of this PTM remained elusive. The biological consequence of the methylation of lysine residues on histones and the effects on genetic expression were first observed through the formation of heterochromatin mediated by the recruitment of HP1 to the H3K9 methyl mark (Bannister et al., 2001; Lachner et al., 2001), heightening interest in the ramifications of lysine methylation on epigenetic regulation.

2.2 Lysine methylation of non-histone proteins

Lysine methylation also occurs on non-histone proteins, controlling the stability, cellular localization, and function of the modified protein (Ferguson et al., 2011). Lysine methylation is found throughout all domains of life on nuclear and

Figure 3.

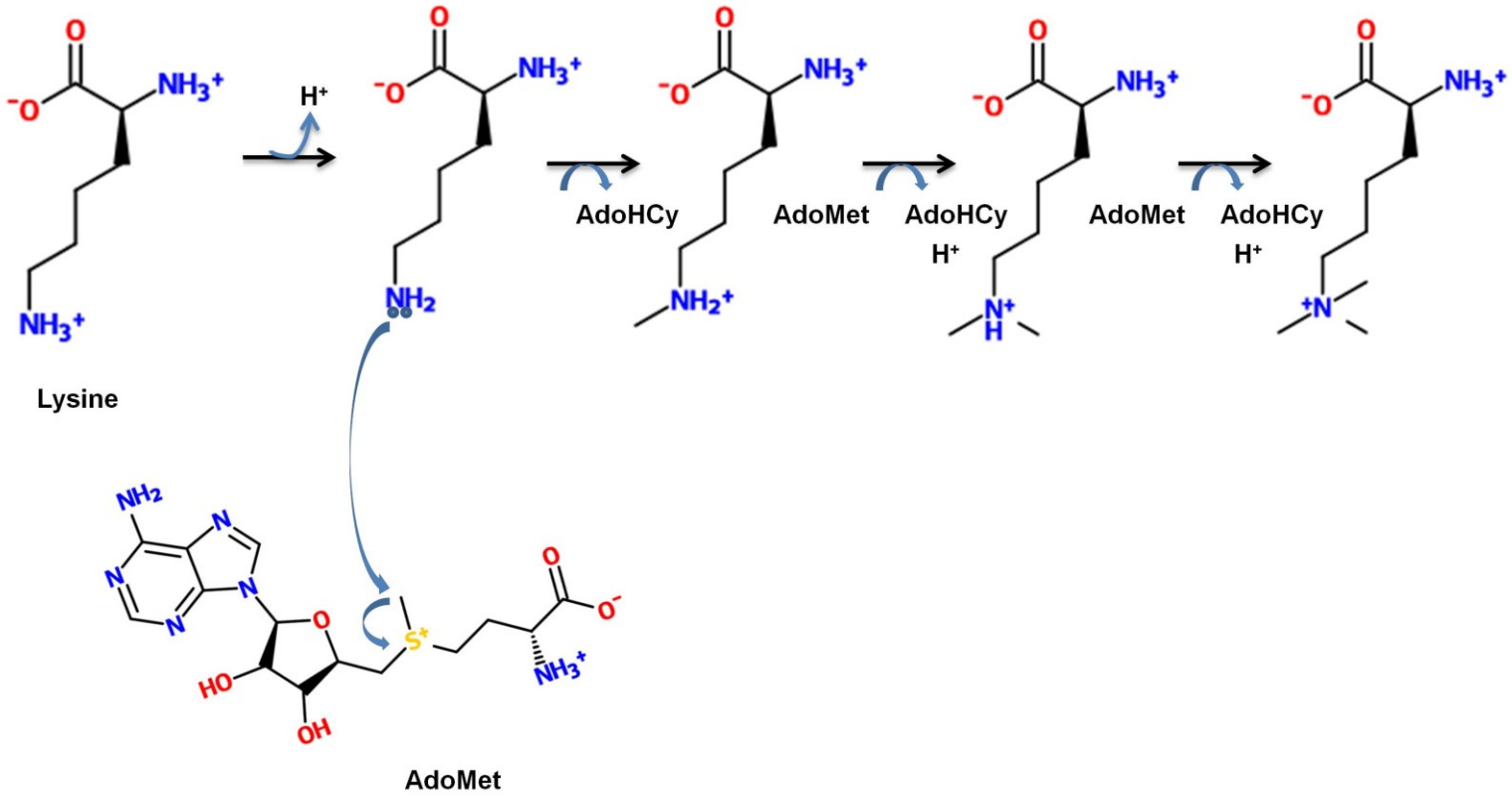


Figure 3: The catalysis of the methyl transfer from co-factor S-Adenosyl-L-methionine (AdoMet) to the ε-amine of the lysine residue
Adapted from PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA [PNAS] (Zhang and Bruice, 2008), copyright 2008. The methyl transfer follows de-protonation of the amine and results in the formation of the by-product S-Adenosyl-L-homocysteine (AdoHcy). In addition, the reaction can continue following subsequent deprotonation and methyl transfer to produce di-methylated and tri-methylated lysine.

cytoplasmic proteins, and was initially identified on the flagellin proteins of *Salmonella typhimurium* by Ambler and Rees (Ambler and Rees, 1959).

The biological role of the methylation of non-histone proteins also long remained a mystery. The methylation of the elongation factor EF-Tu, for example, was first identified in 1979 (Ames and Niakido, 1979; L'Italien and Laursen, 1979) but the first hint of a biological role emerged over a decade later, namely that EF-Tu methylation occurred during nutrient deprivation suggesting that environmental factors regulate this PTM (Young et al., 1990; Young and Bernlohr, 1991). A direct biological consequence of lysine methylation was finally ascribed following the discovery that methylation of calmodulin K115 decreases its ability to stimulate the activity of NAD kinase but does not prevent activation of its other targets (Molla et al., 1981; Roberts et al., 1986), demonstrating that this modification regulates only a subset of the enzyme's methylation activity on its substrates.

3. Protein lysine methyltransferases (PKMT)

The transfer of the methyl group from the methyl donor S-Adenosyl-L-methionine (AdoMet) to the ε-amine of the target lysine can be catalyzed by two groups of protein lysine methyltransferases (PKMTs): the seven β-strand (class I) methyltransferases and the SET (class V) domain lysine methyltransferases (Schubert et al., 2003) **(Fig. 4)**. The seven β-strand methyltransferases, characterized by the Rossmann-fold surrounding the central β-sheet structure harboring its conserved catalytic motifs **(Fig. 4A)**, methylate DNA, RNA or other amino acids including glutamine, aspartate, arginine, and histidine

Figure 4.

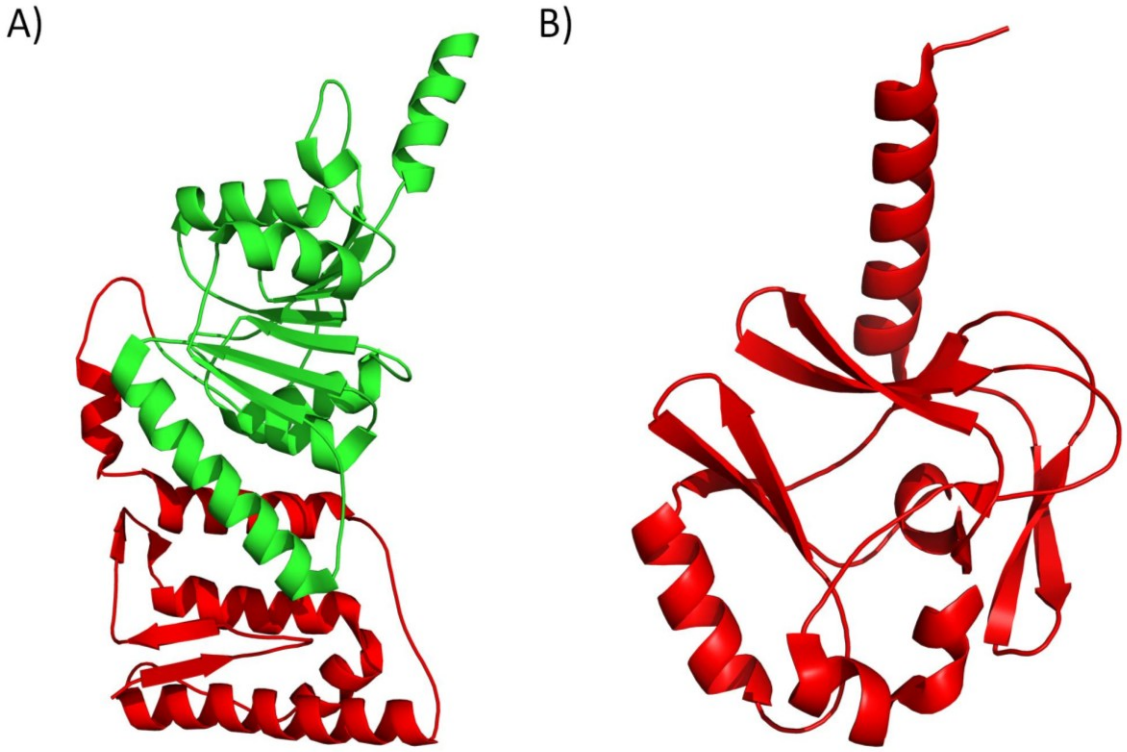


Figure 4: Examples of protein folds of lysine methyltransferase domains
Cartoon structure of A) Dot1L with the Rossmann fold highlighted in green (PDB ID 1NW3) adapted with permission from ELSEVIER [Cell] (Min et al., 2003), copyright 2003 and B) the SET8 catalytic domain (PDB ID 1ZKK) adapted from CSH PRESS [Genes and Development] (Couture et al., 2005), copyright 2005.

(Martin and McMillan, 2002; Schubert et al., 2003). Members of the seven β-strand methyltransferases also methylate lysine residues. Tri-methylation of histone H3 Lys79, for example, is catalyzed by Dot1 in eukaryotes (Nguyen and Zhang, 2011).

The second group of protein lysine methyltransferase possess a catalytic SET domain named after the common motif first identified on *Drosophila* regulator genes Suppressor of variegation (Tschiersch et al., 1994), Enhancer of Zeste (Jones and Gelbart, 1993) and Trithorax (Stassen et al., 1995). The SET domain comprises three evolutionary conserved regions playing pivotal roles in substrate recognition as well as the binding of the substrate and cofactor in the active site for catalysis of the methyl transfer (Couture and Trievel, 2006). Combined with the other structure elements, these domains fold into a β-sheet knot-like structure, orienting the conserved motifs GXG, YXG, NHXCXPN and ELXFDY into the active site next to the substrate binding cleft and the co-factor binding pocket (Cheng and Zhang, 2007; Couture and Trievel, 2006; Dillon et al., 2005b; Qian and Zhou, 2006) **(Fig. 4B)**. The SET domain lysine methyltransferases can be separated into seven families based on similarities in the sequences encompassing the catalytic SET domain: EZ, SET1, SET2, SMYD, SUV39, SUV4-20 and RIZ (Dillon et al., 2005a). In addition, two members of the SET domain methyltransferases outside of the seven families, SET7/9 and SET8, are reported to methylate a number of substrates including both histones and non-histones (Chuikov et al., 2004; Dillon et al., 2005a; Ea and Baltimore, 2009; Kouskouti et al., 2004; Kurash et al., 2008; Munro et al., 2010; Pagans et al., 2010; Shi et al., 2007; Subramanian et al., 2008; Wang et al., 2009; West et al., 2010)

4. Characterizing lysine methylation

High-throughput studies uncovered the implication of acetylation (Choudhary et al., 2009; Henriksen et al., 2012; Weinert et al., 2011) and phosphorylation (Ptacek et al., 2005; Sopko and Andrews, 2008) in a number of biological processes, yet lysine methylation has not been studied in depth. This is due to a number of challenges in detection, owing partially to the small molecular weight of 14 Daltons (Da) relative to other PTMs, as well as the lack of a difference in charge between methylated and un-methylated residues, rendering lysine methylation difficult to characterize using direct physico-chemical methods. From targeted discovery of specific methylation sites to high-throughput discovery of methylation sites and prediction based discovery using known substrates of PKMTs, various approaches were considered to identify novel methylation targets.

Single sites of lysine methylation can be detected using one of three methods: amino acid sequencing, radio-labelled enzymatic assays and immunoblotting. Sequencing by Edman degradation, is a precise and reliable method of detection of methyllysine (Bloxham et al., 1981), but is not amenable for high-throughput identification of lysine methylation sites as it is time consuming and requires a lot of target protein (Ong et al., 2004). An alternative is using radioactively labeled methyl donors in media or cell lysate followed by separation by 2D SDS-PAGE or liquid chromatography (Wang and Lazarides, 1984; Wang et al., 1982; Wang et al., 1992). In addition to working with large amounts of radioactivity, a drawback of this method is that it does not allow the detection of specific methylation sites and does not discriminate between the type of residue that is methylated,

whether a lysine, arginine, histidine, aspartate or glutamate or even the amino-terminus of the protein (Petrossian and Clarke, 2011; Sprung et al., 2008; Stock et al., 1987; Webb et al., 2010). Immunoblotting provides an alternative method to identifying methylated proteins (Iwabata et al., 2005). The pan-methyllysine antibodies prove to be a poor detection method due to their low sensitivity, reproducibility and specificity between batches and producers. However, antibodies produced against specific sites can be useful in the identification and confirmation *in vivo* of methylated protein (Guo et al., 2014; Levy et al., 2011).

These techniques allow for the detection of single methylation sites, but they cannot be applied for high-throughput identification of methylated proteins. The method of choice is mass spectrometry, which is sensitive enough to detect the 14 Da shift of a methylated protein and also identify the residue on the peptide that is methylated (Afjehi-Sadat and Garcia, 2013). Advances in instruments such as Orbitrap and triple TOF increased precision of the detection and render identification easier (Chu et al., 2012; Huq et al., 2009). In the past, pan-acetyllysine antibodies were used to enrich acetylated proteins for mass spectrometry identification, but the low specificity and sensitivity of the pan-methyllysine antibody render this approach less effective (Choudhary et al., 2009). Alternative methods to enrich methylated proteins include the use of an antibody cocktail (Guo et al., 2013; Ong et al., 2004), SILAC labeling (Ong et al., 2004) and the use of a known methyl lysine binding domain, such as was done with the HP1β chromodomain, as a bait to capture methylated proteins prior to mass spectrometry analysis (Liu et al., 2013).

While these methods allow the detection of methylated proteins and determine which sites are methylated, they do not identify the PKMT associated with the methylated site. The substrate recognition of a PKMT can be probed using a peptide library encompassing the methylated lysine with systematic mutations to identify sequences favoring methylation, thus establishing a methylation motif. This method, along with structural studies, elucidates the substrate specificity of a given enzyme and enables the prediction new substrates. This method has been applied to the study of G9a (Rathert et al., 2008), SETD6 (Levy et al., 2011), SET 7/9 (Couture et al., 2008; Dhayalan et al., 2011) and SET8 (Kudithipudi et al., 2012). New substrates have been identified in some cases, although with SET8 no new substrates were identified (Kudithipudi et al., 2012). An alternative to using peptides is full length protein arrays. A group recently used such approach to assign PLK1 and PAK4 as new substrates for SETD6 (Levy et al., 2011). Progress has been made in the characterization of lysine methylation sites, yet efforts are now focusing on linking PKMT with their substrates, validation of the sites and their function.

5. The SMYD family of PKMT

Among the seven families of SET domain lysine methyltransferases, the SMYD family consists of 5 paralogues (SMYD1-5). This family is characterized by a SET domain split by a zinc chelating MYND motif (Myeloid translocation protein 8, Nervy and DEAF-1) (**Fig. 5**) (Liu et al., 2007; Sims et al., 2002). With the exception of SMYD5, SMYD proteins possess a tetratrico peptide repeat (TPR). In other

Figure 5.

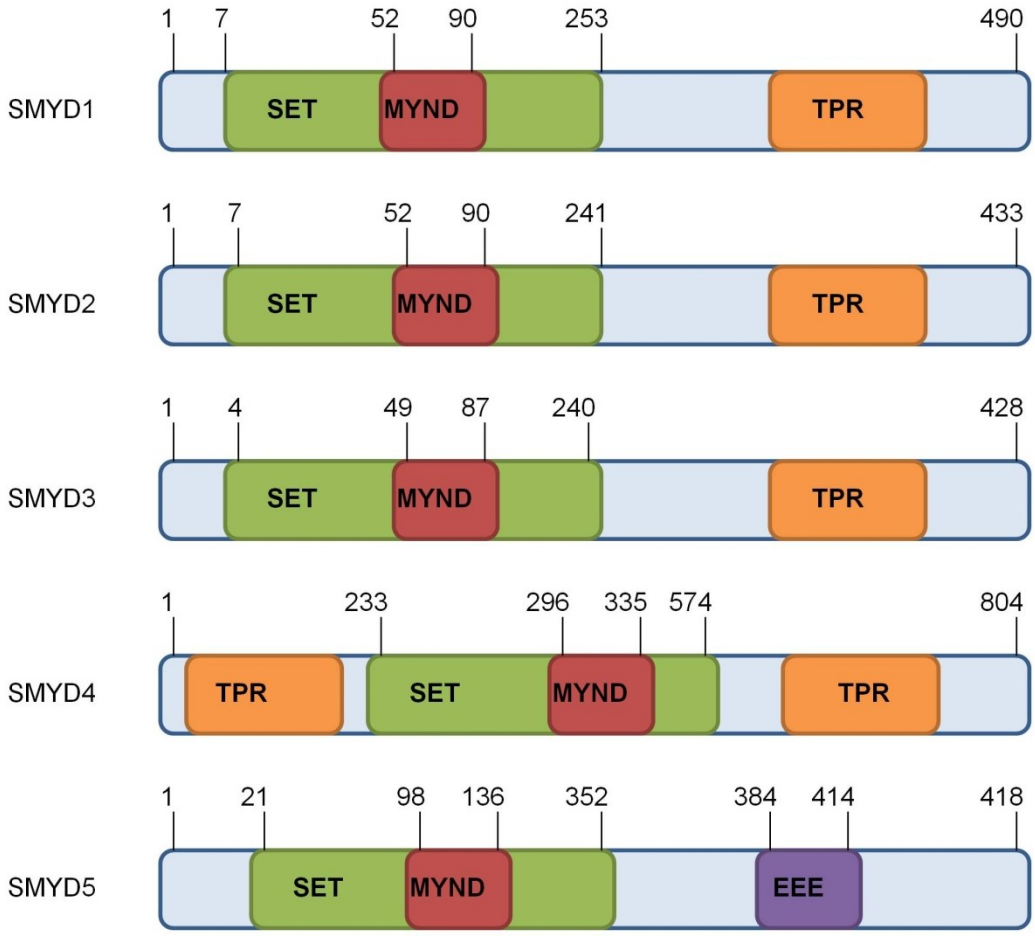

SMYD1
1
7
52
90
253
490
SET
MYND
TPR
SMYD2
1
7
52
90
241
433
SET
MYND
TPR
SMYD3
1
4
49
87
240
428
SET
MYND
TPR
SMYD4
1
233
296
335
574
804
TPR
SET
MYND
TPR
SMYD5
1
21
98
136
352
384
414
418
SET
MYND
EEE

Figure 5: Organization of the catalytic and protein-interacting domains of the SMYD family members.
Individual domains are delimited by colors, and their known positions within the primary amino acid sequence: SET domain (green), MYND domain (red), TPR domain (orange), poly-glutamate stretch (purple).

proteins, TPR domains are important for the modulation of protein-protein interactions, and SMYD4 contains an additional TPR domain at its amino-terminus (Abu-Farha et al., 2011). Structural analysis of SMYD1-3 have provided key insights on catalysis and the modulation of protein-protein interactions. Comparative analysis between the crystal structures of SMYD1-3 reveal a similar overall fold of the protein, but that the arrangement of the C-terminal and N-terminal lobes differ, conferring differences in shape and size of the peptide binding grove (Ferguson et al., 2011; Sirinupong et al., 2011; Sirinupong et al., 2010). Additionally, product inhibition and steady state kinetics analysis of SMYD2 revealed the catalytic rapid equilibrium Bi-Bi mechanism (Wu et al., 2011).

5.1 SMYD methylation targets

Substrates of SMYD family members include histone and non-histone proteins, summarized in **Table 1**, and the consequences of methylation of some of these proteins are known. Mono-methylation of p53 on Lys370 by SMYD2 limits its DNA binding capacity, which can be reversed through de-methylation by LSD1 (Huang et al., 2006; Huang et al., 2007). SMYD2 mediated methylation of Lys810 on pRb enhances the phosphorylation at S807/S811 and accelerates cell cycle progression through E2F activity (Cho et al., 2012), while monomethylation by SMYD2 at Lys860 on pRb provides a binding site for the transcriptional repressor L3MBTL1 (Saddic et al., 2010). Another non-histone protein, VEGFR1, is methylated by SMYD3 at Lys831, which enhances its kinase activity

Table 1.

Enzyme	Substrate (lysine)	Methylation State	Reference
SMYD1	H3 (K4)	?	(Tan et al., 2006)
SMYD2	H3 (K4)	?	(Abu-Farha et al., 2011)
	H3 (K36)	me2	(Brown et al., 2006)
	p53 (K370)	me1	(Huang et al., 2006)
	pRb (K810)	me1	(Cho et al., 2012)
	pRB (K860)	me1	(Saddic et al., 2010)
	HSP90 (K209)	me1	(Abu-Farha et al., 2011)
	HSP90 (K615)	me1	(Abu-Farha et al., 2011)
SMYD3	H4 (K20)	?	(Foreman et al., 2011)
	H4 (K5)	me1	(Van Aller et al., 2012)
	VEGFR1 (K831)	?	(Kunizaki et al., 2007)
SMYD5	H4 (K20)	me3	(Stender et al., 2012)

Table 1: Summary of known substrates for the SMYD family of methyltransferases.
Methylated residues are identified in parenthesis and methylation states if determined are indicated.

(Kunizaki et al., 2007). Lysine methylation by the SMYDs can also be regulated. SMYD2 monomethylation of HSP90 Lys615, for example, can be prevented by the association of HSP90 to its co-chaperone HOP, and the methylation can also be reversed by the amine oxidase LSD1 (Abu-Farha et al., 2011).

5.2 The role of the SMYDs in development

The SMYD family members SMYD1-4 are involved in important developmental processes, particularly related to cardiac and muscle development. SMYD1 is necessary for cardiac development through cardiomyocyte differentiation and maturation (Gottlieb et al., 2002) and its expression in the heart is directed by MEF2 (Phan et al., 2005). SMYD1 is expressed during development in zebrafish and is required for proper skeletal muscle development, myofibril organization and muscle contraction (Tan et al., 2006), as well as sarcomerogenesis (Just et al., 2011). In addition to SMYD1, SMYD2 is also expressed in cardiac muscle during cardiogenesis (Diehl et al., 2010) and both enzymes are expressed during muscle development in *X. laevis* (Kawamura et al., 2008). In zebrafish, SMYD3 is also important in heart and trunk muscle development. Interestingly, knockdown of SMYD3 results in abnormal expression of heart-chamber markers and myogenic regulatory factors (Fujii et al., 2011). Knockdown of SMYD4 is lethal in embryos of Drosophila, while knockdown of SMYD4 expression in muscles specifically results in eclosion failure, suggesting that this PKMT is a transcriptional modulator involved in muscle development (Thompson and Travers, 2008).

5.3 The role of SMYDs in cancer

While necessary for specific developmental processes, misregulation of the SMYDs can also result in cancer pathology. The increased expression of SMYD3 promotes proliferation of colorectal carcinoma, hepatocellular carcinoma(Chen et al., 2007; Hamamoto et al., 2004) and breast cancer (Hamamoto et al., 2006). It also promotes cell growth and invasiveness of cervical carcinoma (Wang et al., 2008b). SMYD3 increases the proliferation, adhesion and viability in NIH3T3 (Luo et al., 2007; Luo et al., 2009) and HepG2 (Yang et al., 2009), as well as migration of MC-7 breast cancer cells (Luo et al., 2014). This proliferation phenotype depends on the catalytic activity of SMYD3 (Van Aller et al., 2012). SMYD3 enables the expression of factors which lead to tumorigenesis, such as c-myc expression in hepatoma (Yang et al., 2009), c-MET and the promotion of tumour invasiveness (Zou et al., 2009). It also promotes the expression of metalloproteinase MMP-9, leading to cancer cell migration and correlates with H3 Lys4 trimethylation levels (Cock-Rada et al., 2012). Variations in the number of tandem repeats in the SMYD3 promoter are also known to correlate with the risk of esophageal squamous cell carcinoma in smokers (Wang et al., 2008a). Overexpression of SMYD2 has been reported in esophageal squamous cell carcinoma and correlates with decrease in survival rate and the promotion of proliferation of esophageal squamous carcinoma cells (Komatsu et al., 2009).

Expression of SMYD4 in mammary epithelial cells results in the suppression of the growth of tumour cells and the deletion of one allele of SMYD4 results in tumorigenesis, suggesting that SMYD4 plays a role in tumour suppression in breast

carcinogenesis (Hu et al., 2009). In addition, SMYD4 is often deleted or downregulated in medulloblastoma (Northcott et al., 2009).

6. SMYD5

SMYD5 was first identified as a retinoic responsive gene in murine F9 embryonal carcinoma cells (Shago and Giguere, 1996). Since its initial discovery, few reports elaborated on the possible role of SMYD5. A gain-of-function screen of the SMYD5 homologue CG3353 in *Drosophila* revealed defects in neuron number in the mushroom bodies, substructures of the *Drosophila* brain essential for memory (Nicolai et al., 2003). In a genetic screen to determine regulators of the gene *eyes absent (eya)*, the founding member of the Eya family of proteins, SMYD5 was found to be important for the correct expression of *eya* during development, and that disruption of the SMYD5 homologue resulted in decreased *eya* expression levels in the visual primordium and optic lobes (Salzer et al., 2010), suggesting that SMYD5 could play a role in embryonic development and differentiation.

The role of SMYD5 seems to extend beyond development and differentiation. According to a recent report, in association with the NCoR chromatin remodeling complex, SMYD5 methylation of histone H4 on Lys20 represses the inflammatory response through restriction of toll-like receptor 4 (TLR-4) mediated expression in macrophages (Stender et al., 2012). Increased bacterial count in cecael content and increased colon weight in non-conditional knockout mice after infection with citrobacter (Sanger Institute, http://*www.sanger.ca.uk/mouseportal/*) (Ayadi et al., 2012) as well as high mRNA expression levels in the hemolymphoid system of

E14.5 mouse embryos detected through high-throughput *in situ* expression profiling (Diez-Roux et al., 2011) also provided evidence for a role in the regulation of the immune response by SMYD5.

Additional roles potentially exist, as the same expression profiling reveals significant SMYD5 mRNA levels in the brain, digestive system, salivary glands and skeletal muscles (Diez-Roux et al., 2011). Furthermore, immunohistology of 80 normal cell types revealed strong staining in squamous epithelia in the cytoplasm and membrane, moderate staining in the kidney tubuli and gall bladder as well as moderate staining in luminal membranes of intestinal cells (Uhlen et al., 2010). Behavioral and metabolic roles may exist, as conditional knockout mice with reporter tag insertion demonstrate hyperactivity, increased oxygen uptake, increased carbon dioxide expenditure and increased energy expenditure, observed only in female mice (Sanger Institute, http://*www.sanger.ca.uk/mouseportal/*) (Ayadi et al., 2012).

7. Exploring protein function through the macromolecular structure: X-ray crystallography

Understanding the biochemical mechanisms, involvement in disease pathways and intermolecular interactions of a protein is obtained through its macromolecular structure (Garman, 2014). This requires a resolution at the atomic level, which can be obtained by x-ray crystallography and nuclear magnetic resonance (NMR). While NMR can determine the conformation of proteins with low molecular weight, its use is limited in the determination of the structure of high molecular weight proteins (Feng et al., 2011). X-ray diffraction patterns of a crystal

can be obtained for proteins and macromolecular complexes of a greater range of sizes, but the critical step in obtaining a structure is the crystallization itself because many proteins are difficult to crystallize (Feng et al., 2011). Atomic resolution (<1.0 Å) is required for a high precision structure that reveals details at the atomic level, therefore crystals need to diffract at high resolution (Acharya and Lloyd, 2005).

Conformational variations of a protein exist, and while a crystal structure represents a snapshot of the protein within a given crystal lattice, obtaining crystals of the protein in different conformations can reveal much about the movement of regions important for function (Garman, 2014). Advances in technologies enabled high-throughput determination of structures, particularly in expression and purification procedures, as well as the development of powerful synchrotron radiation sources and crystallization robots (Acharya and Lloyd, 2005).

Crystallization of proteins is an energetically unfavorable process, requiring dissolved and freely diffusing molecules in an aqueous solution to arrange themselves in an ordered three-dimensional crystal lattice (Bukowska and Grutter, 2013). Various methods to crystallize a protein exist, such as batch under oil and dialysis method although the most common is by vapor diffusion (Giege, 2013). Vapor diffusion occurs in a closed system, where the protein is diluted by the reservoir solution twofold in a drop, and a slow gradient of increasing protein concentration occurs as the solution evaporates from the drop towards the reservoir (Takeda, 2006). The reservoir solution can contain a number of chemical compounds to facilitate crystallization, including salts, polymers, reducing agents, organic molecules, detergents, divalent cations, precipitating agents and buffers to

modulate different pH (Takeda, 2006). Reducing flexibility and increasing the stability of a single conformation can also encourage crystallization. The protein can be stabilized by adding inhibitors, other macromolecules, additives or chaperones as well as through protein engineering of constructs of a stable domain, of homologues in different organisms, of truncations, by incorporating mutations of surface residues, or trimming parts of the polypeptide by controlled proteolysis (Bukowska and Grutter, 2013). The temperature, concentration, purity and conformational homogeneity of the protein affect solubility and crystal growth mechanisms (Giege, 2013). Once crystals are obtained, further optimization of crystallization conditions is often required, using a grid screen modifying two parameters at a time to obtain large enough crystals of diffraction quality (Takeda, 2006). Careful observation of the outcome of different crystallization experiments is required to determine parameters optimal for crystallization (Benvenuti and Mangani, 2007) **(Fig. 6)**.

When diffraction data can be obtained from a crystal, the correct space group must be assigned and multiple reflections must be measured at the highest possible resolution to increase the completeness and accuracy of the data (Acharya and Lloyd, 2005). A model is then produced by fitting structural components into the experimentally obtained electron density map followed by refinement and the modeling of ligands, metals and water molecules (Acharya and Lloyd, 2005). Crystallographic software such as Collaborative Computational Project Number 4 (CCP4) (Winn et al., 2011), PHENIX (Afonine et al., 2012), Phaser (McCoy et al., 2007) facilitate structure determination through phasing, electron density map

calculation and interpretation, structure refinement and model building (Garman, 2014).

Figure 6.

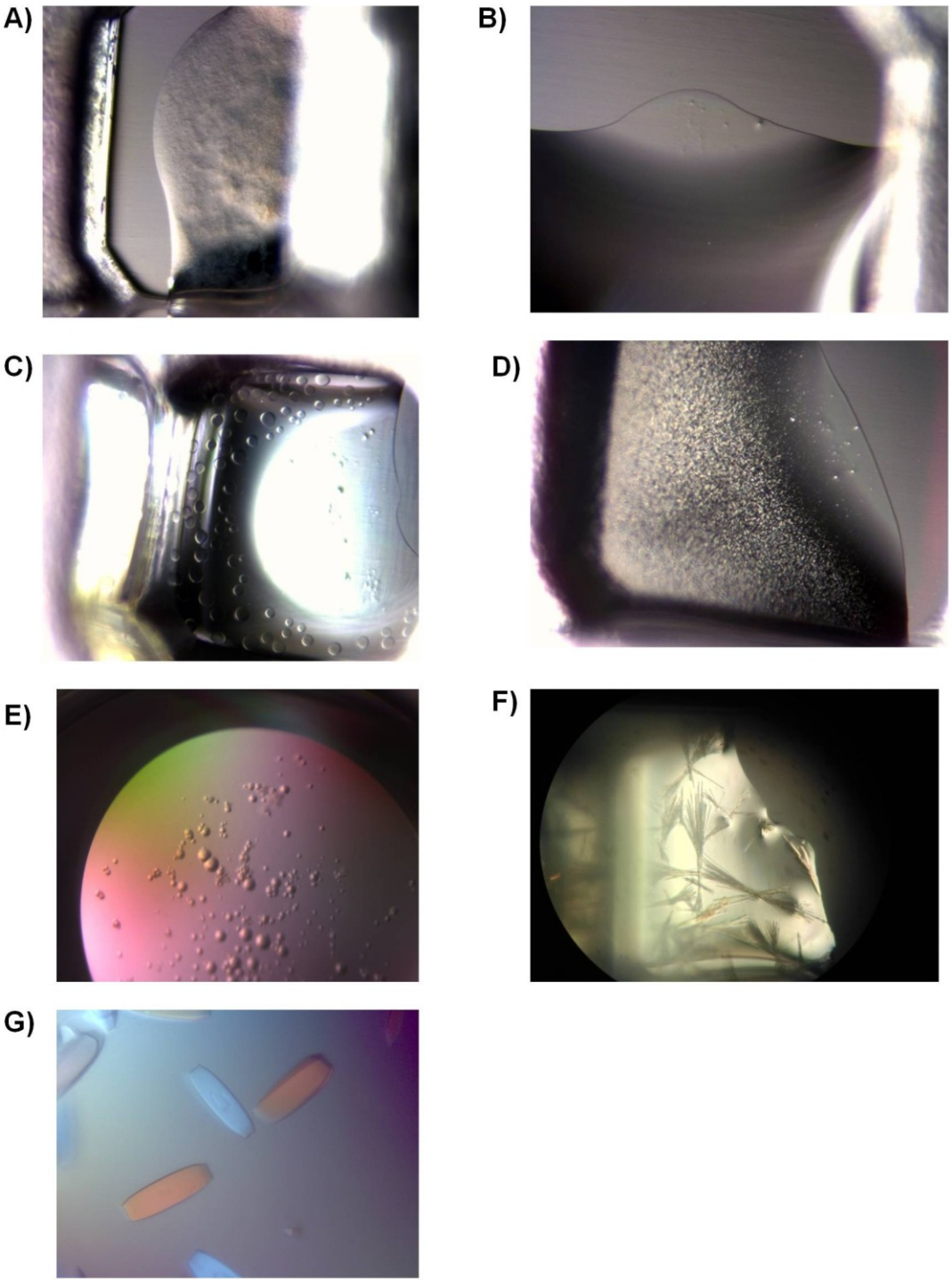


Figure 6: Outcomes of crystallization experiments and possible observations in the protein drop

Adapted with permission from NATURE PUBLISHING GROUP [Nature] (Benvenuti and Mangani, 2007), copyright 2007. Different results obtained in crystallization experiments reveal how the protein behaves in different conditions. The protein in solution can form A) heavy brown precipitate, B) clear drop, C) phase separation, D) granular precipitate, E) spherulites, F) needle shaped crystals G) well shaped single crystals.

Statement of hypothesis

Biological roles of SMYD5 have started to emerge, but the biochemical and structural features needed to facilitate SMYD5 mediated methylation of histone H4, as well as other potential substrates, has not been explored. Using radio-labeled enzymatic assays, an in-house peptide assay and protein structure determination using x-ray crystallography, insights can be gained as to how SMYD5 catalyses the methyl transfer from the co-factor to its substrate.

Hypothesis - SMYD5 is identified as an H4 Lys20 methyltransferase, but the roles of the other SMYD family members extend beyond histone methylation and the modulation of chromatin structure and function. We hypothesize that, by extension of this logic, SMYD5 will also play a larger role by methylating substrates other than histones and that methylation of its specific subset of substrates is mediated by its unique structure.

Aims:

1. Obtain a crystal structure of SMYD5 to gain insight on specific features of the enzyme at the atomic level that play a role in catalysis.

2. Establish the peptide motif of SMYD5 substrates to determine the primary amino acid sequence selectively methylated by SMYD5 and identify novel substrates.

3. Explore substrate specificity, whether SMYD5 only requires the primary amino acid sequence, or as for other SET domain enzymes, other structural determinants are required.

Materials and Methods

1. Plasmid Constructs

The DNA sequence corresponding to full-length *Homo sapiens* (*H. sapiens*) SMYD5 (**Appendix 1**) was purchased from GenScript with the codons optimized for expression in *Escherichia coli (E. coli).* The cDNA was subcloned, using BamHI and XhoI restriction sites, into the parallel expression vector pHIS2 (Sheffield et al., 1999) in frame with a sequence encoding a N-terminal hexahistidine (HIS) tag and an intervening TEV protease cleavage site. The DNA sequences of the *Gallus gallus (G. gallus)* (**Appendix 2**) and *Drosophila melanogaster* (*D. melanogaster*) (**Appendix 3**) SMYD5 orthologues were also purchased from GenScript and subcloned, using BamHI and XhoI restriction sites, into the pSMT3 vector; a home-made vector allowing the expression of proteins in fusion with a His-SUMO tag.

Truncated constructs were prepared from the full length *H. sapiens* SMYD5 (**Fig. 7**) using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) following manufacturer's instructions using the primers listed and the vectors indicated in **Table 2** and verified by sequencing.

2. Protein Expression and Purification

Plasmids were transformed in *E. coli* strain BL-21 Rosetta (Novagen) and overexpressed as fusion proteins with their respective tags by induction with 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 0.4-0.5 at 18 °C during 16 h.

Figure 7.

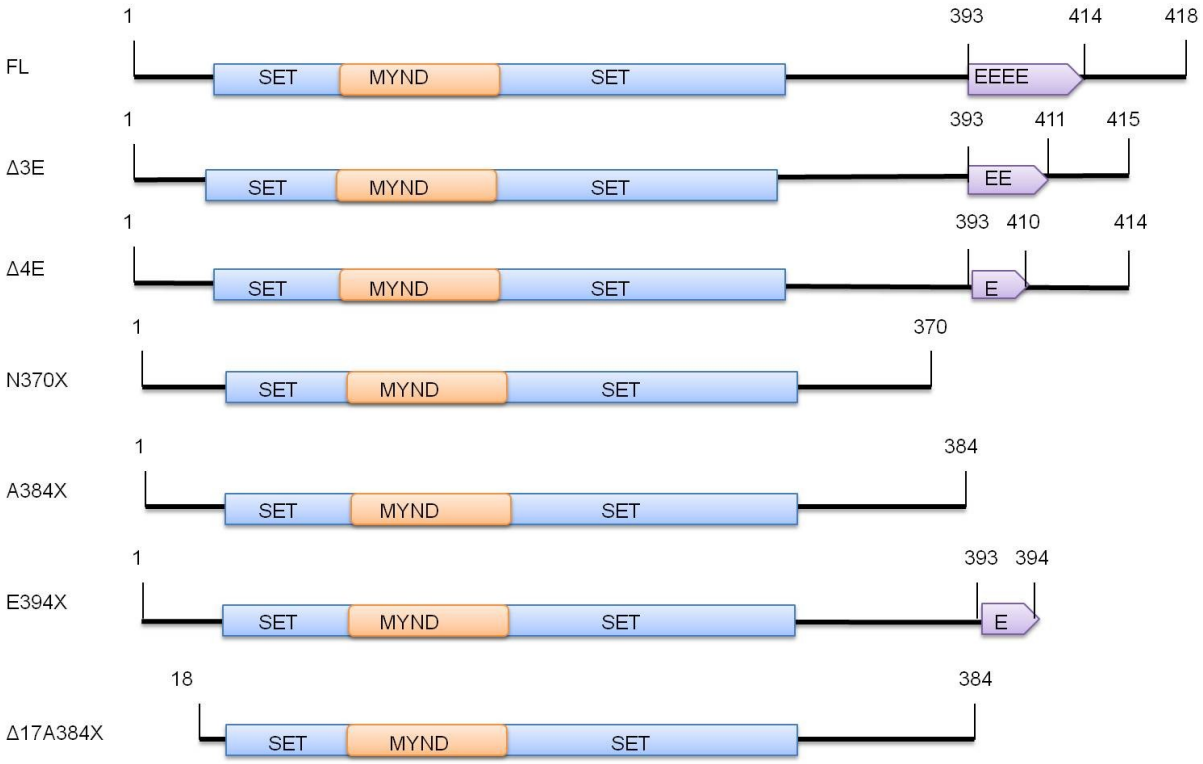

FL
1
393
414
418
SET
MYND
SET
EEEE
Δ3E
1
393
411
415
SET
MYND
SET
EE
Δ4E
1
393
410
414
SET
MYND
SET
E
N370X
1
370
SET
MYND
SET
A384X
1
384
SET
MYND
SET
E394X
1
393
394
SET
MYND
SET
E
Δ17A384X
18
384
SET
MYND
SET

Figure 7: SMYD5 constructs based on the full length H. sapiens sequence (FL). Names of each construct are indicated to the left, with the length in amino acids indicated above each section. The SET and MYND domain, as well as the length of the poly-glutamate stretch, are illustrated as colored rectangles.

Table 2.

Construct	Vector	Strand	Sequence
A384X	pGST2	Forward	TCATGTCCGAAATGTCTGGCAGAATAAGATGAACCGAATGTGACCTCGGAA
		Reverse	TTCCGAGGTCACATTCGGTTCATCTTATTCTGCCAGACATTTCGGACATGA
E394X	pGST2	Forward	GAACCGAATGTGACCTCGGAAGAATGAGAAGAAGAAGAAGAAGAAGAAGGCGAA
		Reverse	TTCGCCTTCTTCTTCTTCTTCTTCTTCTCATTCTTCCGAGGTCACATTCGGTTC
N370X	pSMT3	Forward	TCCCGCCACAAAATCCTGCGTGAATAATACCTGTTTGTCTGCTCATGTCCG
		Reverse	CGGACATGAGCAGACAAACAGGTATTATTCACGCAGGATTTTGTGGCGGGA
Δ17A384X	pSMT3	Forward	GAGTGGATCCCGTGCTCGTGTCTCGGTG
		Reverse	GAGTCTCGAGTTATTCTGCCAGACATTT
Δ3E	pHIS2	Forward	ACCGAATGTGACCTCGGAAGAAGAAGAAGAAGAAGAAGAAGGCGAACCGAAA
		Reverse	TTTCGGTTCGCCTTCTTCTTCTTCTTCTTCTTCTTCCGAGGTCACATTCGGT
Δ4E	pHIS2	Forward	GTGACCTCGGAAGAAGAAGAAGAAGAAGAAGGCGAATGGGAAGACGCAGAACTGGGCGACGAA
		Reverse	TTCGTCGCCCAGTTCTGCGTCTTCCCATTCGCCTTCTTCTTCTTCTTCTTCTTCCGAGGTCAC

Table 2: Oligonucleotide sequences used for mutagenesis of the H. sapiens SMYD5 full-length template to produce truncated constructs cloned in the indicated vectors.

Pellets were collected following centrifugation at 1 270 g for 20 minutes. Cells expressing HIS or SUMO tag constructs were lysed by sonication during three intervals of one minute (Misonix Sonicator 3000) in 50 mM sodium phosphate pH 7.0, 150 mM sodium chloride, 5 mM β-mercaptoethanol (50 mM NaPi pH 7.0, 150 mM NaCl, 5 mM BME - solubility buffer). Lysates were cleared by centrifugation at 30 900 g during 30 minutes (Beckman Avanti J-25 centrifuge, JA-25.5 rotor) followed by filtration with a 0.45 µm filter. The supernatant was applied by gravity to a Talon Co^{2+} affinity chromatography column (Clontech) and the protein was eluted with the solubility buffer supplemented with 500 mM imidazole (pH 7.0). The affinity tags were cleaved using TEV protease (HIS) or ULP1 (SUMO) in the presence of 50 µM zinc chloride ($ZnCl_2$) during 16 hours dialysis in 50 mM NaPi pH 7.0, 150 mM NaCl, 5 mM BME at 4 °C. The protein was re-applied onto Talon Co^{2+} resin for a one hour batch binding, then the protein was collected by centrifugation for two 5 minute intervals at 1 392 g.

Cells of constructs with GST tags were lysed, clarified and filtered as previously mentioned with the exception that the pellets were harvested in phosphate buffered saline (PBS). The supernatant was applied to Glutathione Sepharose 4B resin (GE Healthcare) during 1 h batch binding at 4 °C, and washed three times with 10 bead volumes of PBS. The GST tag was cleaved on beads with TEV protease in 20 mM Tris pH 8.0, 500 mM NaCl, 5 mM BME and 50 µM $ZnCl_2$ at 4 °C during 16 hours. The resin was centrifuged for two 5 minute intervals at 1 392 g and the supernatant, which contains SMYD5, was conserved for additional purification steps.

SMYD5 was concentrated using a 30K Amicon Ultra-15 Centrifugal Filter (Millipore) and further purified by size-exclusion chromatography on a Superdex 200

column (AKTA FPLC, GE Healthcare) in 20 mM Tris-HCl pH 7.0, 100 mM NaCl, 5 mM BME and 1 µM $ZnCl_2$ for all constructs except E394X and A384X, which were purified by size exclusion chromatography with 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM BME and 1 µM $ZnCl_2$. Fractions corresponding to SMYD5 were collected and concentrated in a 30k Amicon Ultra-4 Centrifugal Filter. Protein was centrifuged at 17 000 g and the concentration was measured by UV-visible absorbance (280 nm) in 6 M guanidine hydrochloride (Nanodrop ND-1000 Spectrophotometer) and calculated using the theoretical molecular weight and extinction coefficient (web.expasy.org/protparam/) (Gasteiger et al., 2005).

3. Limited Proteolysis

Limited proteolysis was conducted on purified full length *H. sapiens* SMYD5 to obtain a protein fragment suitable for crystallization. Endoproteinase Glu-C (Sigma) was incubated for one hour at room temperature with purified recombinant SMYD5 in a 1:25 molar ratio in 100 mM NaCl and 20 mM Tris-HCl pH 7.0. The reaction in absence of Glu-C was used as a negative control. The reaction was stopped with the addition of 0.1% formic acid. Proteolysis products and the uncleaved sample were analyzed by ESI-MS. The sample was also separated by SDS-PAGE, transferred on a PVDF membrane and analyzed by Edman sequencing.

Tests for *in situ* limited proteolysis conditions were conducted on the recombinant purified *D. melanogaster* and *G. gallus* constructs using 10 µM protein in 20 mM Tris pH 7.0, 100 mM NaCl, 5 mM BME with a 1:10, 1:25 and 1:50 molar ratio of Endoproteinase Glu-C (Sigma) as well as 1:50 and 1:100 weight/weight ratio (w/w) trypsin for one hour

at room temperature. Reactions were stopped with the addition of Laemmli buffer, boiled and analyzed by SDS-PAGE. For in *situ proteolysis*, the 1:100 molar ratio Glu-C and 1:100 weight/weight (w/w) trypsin were used during crystallization experiments.

4. Synthetic peptides

Synthetic peptides purchased from GenScript of different lengths of the histone H4 amino terminal tail encompassing Lys20 as illustrated in **Fig. 8** were suspended in Milli-Q ultrapure water to roughly 100 mM and rotated at 4°C until completely solubilized. Peptides were centrifuged at 17 000 g and the concentration was measured by UV-visible absorbance (280 nm) in 6 M guanidine hydrochloride (Nanodrop ND-1000 Spectrophotometer) and calculated using the theoretical molecular weight and extinction coefficient (http://web.expasy.org/protparam/) (Gasteiger et al., 2005).

5. Crystallization Screening

To identify conditions yielding diffraction quality crystals of SMYD5, crystallization screens were carried out with the purified SMYD5 constructs and various sparse matrix screens such as the HT HR2-134 Index Screen (Hampton Research), JCSG Core Suite Screen (Qiagen), the series MCSG Screen (Microlytic), Midas screen (Molecular Dimensions), and HT HR2-130 (Hampton Research). To test also the impact of vapor diffusion rates, these crystallization trials were performed at 4 °C, 18 °C and 21 °C. Recombinant proteins were incubated with co-factor S-adenosyl-L-methionine (AdoMet) (Sigma) in the presence or absence of synthetic histone H4 Lys20 peptides (GenScript) (**Fig. 8**) suspended in ultrapure water in equimolar to 10-fold the molar ratio of protein.

Figure 8.

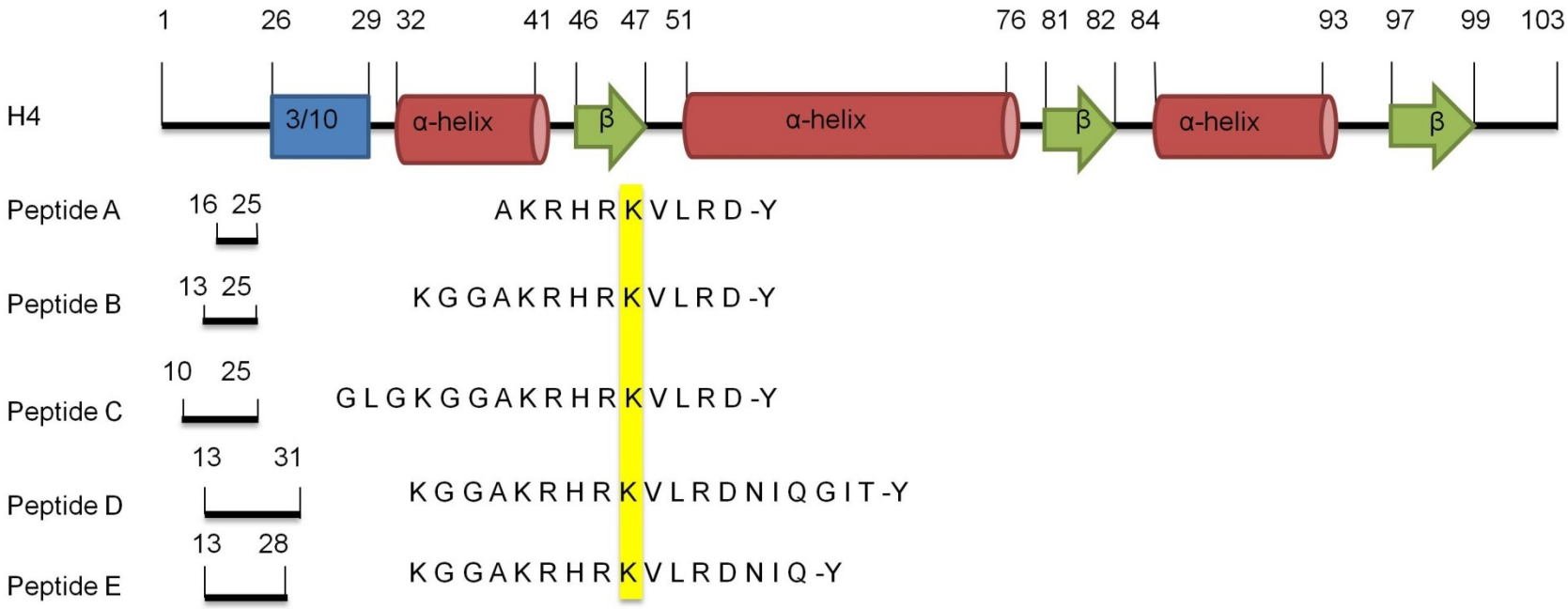

1 26 29 32 41 46 47 51 76 81 82 84 93 97 99 103
H4
3/10
α-helix
β
α-helix
β
α-helix
β
Peptide A
16 25
A K R H R K V L R D -Y
Peptide B
13 25
K G G A K R H R K V L R D -Y
Peptide C
10 25
G L G K G G A K R H R K V L R D -Y
Peptide D
13 31
K G G A K R H R K V L R D N I Q G I T -Y
Peptide E
13 28
K G G A K R H R K V L R D N I Q -Y

Figure 8: Cartoon representation of the secondary structures of histone H4 and the peptides used for crystallization and methylation assays.
The sequences of each peptide are indicated to the right of their respective cartoon, with Lys20 highlighted in yellow. A tyrosine was added during synthesis for quantification purposes.

6. Purification of histones H3 and H4

The pET-3a vectors containing the full length sequence of *Xenopus laevis* histone H3 or H4 were transformed in BL21-RIL *E. coli* strains and overexpressed by induction with 0.1 mM IPTG at an OD_{600} of 0.4 at 25°C for 3 h. Pellets were collected following centrifugation at 1 270 g for 20 minutes.

Cells were lysed by three passes in a high-pressure homogenizer (EmulsiFlex C3, Avestin) in TW buffer (50 mM Tris pH 7.6, 100 mM NaCl, 1 mM EDTA, 5 mM BME, 1% Triton X-100, protease inhibitors 1X). Lysates were cleared by centrifugation at 30 900 g during 30 minutes (Beckman Avanti J-25 centrifuge, JA-25.5 rotor), and the pellet was re-suspended in TW buffer. The pellet was collected following centrifugation at 24 000 g during 10 minutes (Beckman Avanti J-25 centrifuge, JA-25.5 rotor). The pellet was washed with wash buffer (50 mM Tris pH 7.6, 100 mM NaCl, 1 mM EDTA, 5 mM BME) and collected by centrifugation at 24 000 g during 10 minutes (Beckman Avanti J-25 centrifuge, JA-25.5 rotor) twice.

Inclusion bodies were solubilized following re-suspension of the pellet in dimethyl sulfoxide (DMSO) at ambient temperature for 30 minutes, followed by 20 passes in a dounce in unfolding buffer (20 mM Tris-HCl pH 7.6, 7 M guandidine-HCl, 5 mM BME) and rotation for 1 h at ambient temperature. Insoluble material was removed by centrifugation at 24 000 g during 20 minutes (Beckman Avanti J-25 centrifuge, JA-25.5 rotor), and the supernatant was dialyzed in three changes of urea buffer (7 M urea, 10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 5 mM BME), twice 1 h and once for 16 h.

Inclusion bodies were centrifuged during 10 minutes at 24 000 g (Beckman Avanti J-25 centrifuge, JA-25.5 rotor) and the histones were purified by ion exchange chromatography. The soluble portion was applied first on Q-sepharose for a 1 h rotation at ambient temperature, then the supernatant is applied on SP-sepharose for a 1 h rotation at ambient temperature. Resin is washed with 7 M urea, 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 5 mM BME, and histones are eluted with 7 M urea, 10 mM Tris-HCl pH 8.0, 600 mM NaCl, 1 mM EDTA, 5 mM BME. Histones are dialysed twice 1 h and once for 16 h in Milli-Q ultrapure water. Histones are lyophilized and stored at -80 °C.

Lyophilized histones are re-suspended in Milli-Q ultrapure water for 1h at ambient temperature, centrifuged at 17 000 g and the concentration was measured by UV-visible absorbance (280 nm) (Nanodrop ND-1000 Spectrophotometer) and calculated using the theoretical molecular weight and extinction coefficient (http://web.expasy.org/protparam/) (Gasteiger et al., 2005).

7. In vitro Methyltransferase Assays with Histone H4

The methyltransferase activity of SMYD5 was characterized. Using 7 µM recombinant full length *H. sapiens* SMYD5, methylation activity was tested *in vitro* using 0.3 µM Set7/9 as a positive control against 2.9 µM recombinant histones H3 and H4 purified in-house. Reactions were carried out in 50 mM Tris HCl pH 8.0, 50 mM NaCl, 2 mM DTT, with protease inhibitors with 1 µCi radio-labeled methyl donor [^{3}H]-AdoMet (Perkin Elmer) at 30 °C for 1 h. Reactions were stopped with the addition of SDS-PAGE loading buffer and boiled for 5 minutes, then were resolved by SDS-PAGE on a 15% gel

at 100 V and revealed by Coomassie staining. The gel was incubated in DMSO (3X), followed by incubation during 1 h in a 20% solution (w/v) of 2,5-diphenyloxazole in DMSO. The gel was then incubated in water, dried and exposed on an autoradiography film for 7 days at -80 °C.

Reaction conditions were tested with a molar excess of histone H4 versus SMYD5 to determine conditions with the strongest methylation signal. Reaction buffer conditions were tested using 1 µM recombinant full length *H. sapiens* SMYD5 with 2.9 µM histone H4 in 50 mM buffer (acetate pH 4.0, citrate pH 5.0, NaPi pH 7.0, HEPES pH 7.0, tris pH 7.0, 8.0 and 9.0, bicine pH 9.0) with 0.5 mM DTT and 0.01% w/v BSA. Reactions occurred at 30 °C over 16 h and were stopped by spotting on Whatman P-81filter papers. Filter papers were washed to remove excess free [^{3}H]-AdoMet in 50 mM sodium bicarbonate ($NaHCO_3$) pH 9.0 and activity was quantified by scintillation counting (Perkin Elmer Tri-Carb 2910 TR LSC).

The concentration curve of the methylation of histone H4 by full length recombinant *H. sapiens* SMYD5 was determined by incubation of 0.2, 0.8, 1, 2, 5, and 8 µM SMYD5 with 10 µM Peptide B in 50 mM Tris pH 9.0, 0.5 mM DTT, 0.01% (w/v) BSA at 30 °C for 16 h. Reactions were stopped by spotting on filter paper, then treated and quantified by scintillation counts as previously mentioned.

8. Peptide Arrays

Degenerate peptide arrays of the sequence flanking Lys20 of histone H4 (AKRHRKVLRD) were synthesized on cellulose Tioxatridecanediamine (TOTD) ether-

type modified membranes (Kinexus) with 200-300 nmol peptide/spot. Membranes were washed with 95% ethanol, followed by three washes of two minutes in TBS-Tween 0.01% and a final wash in 50 mM Bicine pH 9.0, 50 mM Nacl, 0.5 mM DTT (reaction buffer). The array was incubated during 2 hours at 30 °C in reaction buffer containing 1% BSA, 1 µM freshly purified full length SMYD5 from *H. sapiens* with 10 µM total AdoMet comprised of 2.5 µM tritium labeled [^{3}H]-AdoMet (Perkin Elmer) and 7.5 µM un-labelled AdoMet (Sigma). Membranes were then washed four times during five minutes in 8 M urea with 1% SDS and 0.5% BME, three times five minutes in 50% ethanol (EtOH) with 10% acetic acid (AcOH) and three times two minutes in 95% EtOH. Membranes were dried overnight then exposed to a tritium Storage Phosphor Screen (GE Healthcare) in a cassette for 8h. Signal from the radioactive decay were detected using the Typhoon Trio Phosphorimager (GE Healthcare) and spots were quantified using ImageQuant TL software (GE Healthcare).

Peptide array methylation assays were conducted in duplicate. Quantification for each spot was normalized against one wild type spot on the same array, then averages for each mutation of each residue is calculated. The methylation factor is calculated using the formula $(I_i-I_j^*)/I_j^*$, where I_i represents the value of the mutated residue analyzed, and I_j^* represents the average of all the residues for that position excluding the residue analyzed. Residues with methylation factors greater than or equal to the methylation factor of the wild type residue were incorporated in the motif.

Putative substrates were subsequently identified by conducting a search with the motif using ScanProSite (http://prosite.expasy.org/scanprosite/) (de Castro et al., 2006; Sigrist et al., 2002; Sigrist et al., 2013; Sigrist et al., 2005), and classified using the

Panther Classification System (http://pantherdb.org/) (Mi et al., 2013a; Mi et al., 2013b; Mi and Thomas, 2009). Proteins from the categories molecular functions, biological processes, protein classes and pathways with GO term related to SMYD5 (transporter activity, translation regulator activity, protein binding transcription factor activity, receptor activity, nucleic acid binding transcription factor activity, cellular process, apoptotic process, developmental process, immune system process, nucleic acid binding, defense/immunity protein, transcription factor, chaperone, p53 pathway, Wnt signaling pathway, Toll receptor signaling pathway, inflammation mediated by chemokine and cytokine signaling pathway). Putative substrates from this list were manually filtered, and proteins containing a minimum of two GO terms related to SMYD5 were considered for future analysis. The conservation of the putative target residue of each protein was verified by aligning the sequences from *Homo sapiens, Mus musculus, Gallus gallus, Xenopus laevis* and *Drosophila melanogaster*. Proteins with the target lysine conserved in all organisms were added to the list of putative substrates.

9. Testing substrate specificity using peptide strips

Peptide strips with the sequence flanking Lys20 of histone H4 (AKRHRKVLRD), a K20R mutation and the K16R/K20R double mutant, as well as the sequences of known SMYD2 methylated peptides of p53 (HSSHLKSKKGQ), H3 (HAARTKQTARQ), HSP90 (HERIMKAQALQ) with the p53 peptide with all lysine residues mutated to arginine were synthesized on the same cellulose TOTD ether-type modified membranes (Kinexus) as for the degenerate peptide arrays. Additionally, a longer histone H4 amino-terminal peptide (KGGKGLGKGGAKRHRKVLRD) with systematic and

combinatorial lysine to arginine mutations was synthesized. Reactions were conducted the same as with the peptide arrays, except the smaller Lys20 H4 peptides and SMYD2 substrate peptides were also tested with 0.1%, 0.5% or 1% BSA in their reaction conditions. Membranes were dried overnight then exposed to a tritium Storage Phosphor Screen (GE Healthcare) in a cassette for 4h. Signal from the radioactive decay were detected using the Typhoon Trio Phosphorimager (GE Healthcare) and spots were quantified using ImageQuant TL software (GE Healthcare).

Results

1. SMYD5 is expressed in bacteria and purified by immobilized affinity chromatograhy

The human SMYD family proteins all possess a SET and MYND domain, as illustrated in **Fig.5** (Liu et al., 2007; Sims et al., 2002). SMYD1-4 harbor a TPR protein interacting domain at their carboxyl terminal and SMYD4 also has an amino-terminal TPR domain (Abu-Farha et al., 2011). SMYD5 differs from the other SMYD family members in that it does not have a TPR domain, but does possess a poly-glutamate stretch at its carboxyl terminal (**Fig.5, Appendix 4**). Additionally, there are notable differences between SMYD5 and the other SMYD family members at the level of the amino acid sequence (**Appendix 4**). In light of these differences, it is of interest to determine the mechanistic basis underlying the SMYD5 catalytic properties including cofactor and substrate binding. However, to perform these studies, SMYD5 must first be expressed and purified.

A cDNA corresponding to the SMYD5 gene, with codon optimized for overexpression in *E. coli,* was cloned in pET-based vector enabling protein overexpression in fusion with a TEV-cleavable hexahistidine tag. Following several rounds of optimization of protein overexpression and solubility, we found that while the presence of the overexpressed protein in the whole cell lysate is difficult to observe (**Fig. 9A, lane 1**), a significant enrichment of a protein migrating between the 50kDa and 75kDa molecular weight marker was detected following immobilized affinity

Figure 9.

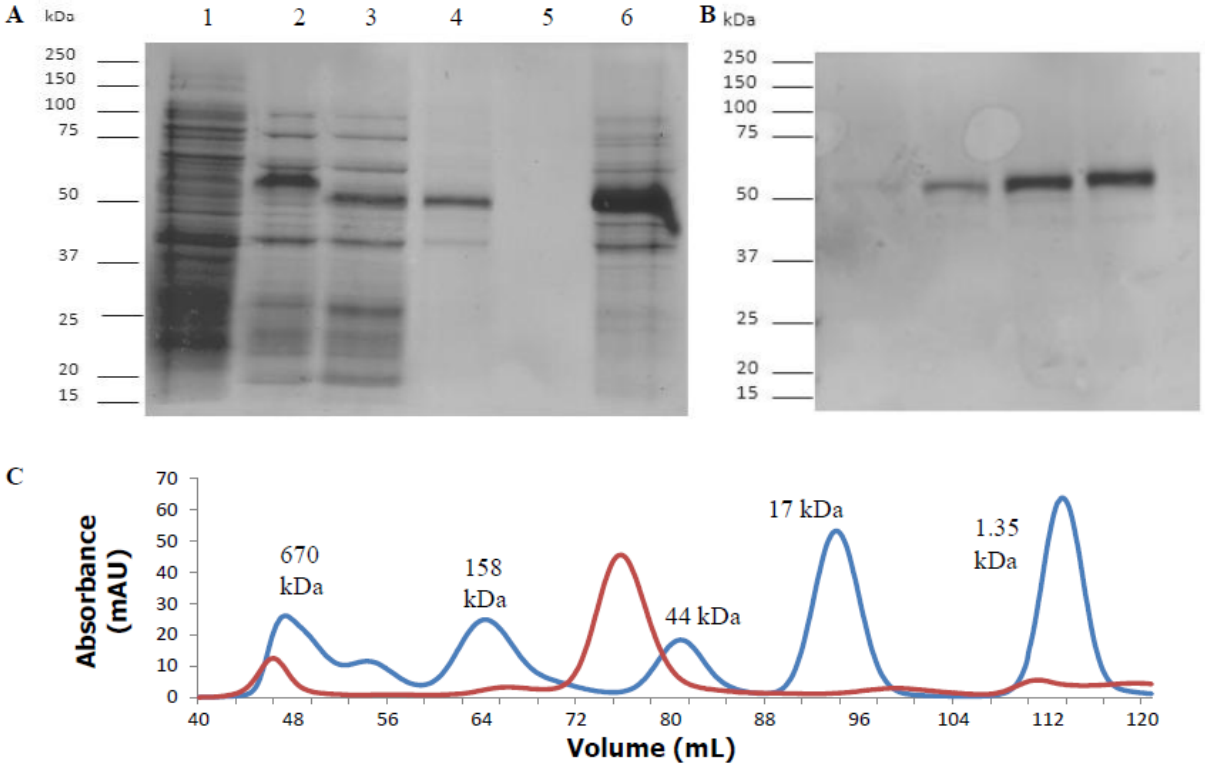

A
kDa
1
2
3
4
5
6
250
150
100
75
50
37
25
20
15
B
kDa
250
150
100
75
50
37
25
20
15
C
Absorbance (mAU)
70
60
50
40
30
20
10
0
670 kDa
158 kDa
44 kDa
17 kDa
1.35 kDa
40
48
56
64
72
80
88
96
104
112
120
Volume (mL)

Figure 9: The purification of recombinant H. sapiens SMYD5 yields a pure monomer.
SDS-PAGE of SMYD5 purification by immobilized metal affinity (IMAC) **(A)** and size exclusion **(B)** chromatographies. Molecular weight markers are shown on the left and each lane is labeled as follows: Soluble proteins after sonication of Rosetta expressing His-SMYD5 (1), purification by IMAC on Co^{2+} resin (2), TEV cleavage (3), second IMAC (4) and the input for gel filtration after concentration (6). **C)** SMYD5 is a monomer in solution. The chromatogram of SMYD5 purification on a Superdex-S200 column (red) is overlaid with gel filtration standards (blue). SMYD5 migrates at approximately 50kDa, which is consistent with the calculated molecular weight of the native protein (47.3kDa).

chromatography (IMAC) (**Fig. 9A, lane 2**). Following TEV cleavage, we observed that the apparent molecular weight of this predominant band decreased of few kilodaltons (**Fig. 9A, lane 3**), suggesting that this enriched protein is selectively cleaved with the TEV protease. Consistent with the hypothesis that this protein is SMYD5, the cleaved protein could no longer bind to TALON beads while the cobalt beads retained the remaining contaminants (**Fig. 9A, lane 4**) yielding a pure preparation of protein.

2. SMYD5 is a monomer in solution.

To perform biochemical and structural studies, the protein preparation must be pure and homogeneous (Garman, 2014; Giege, 2013). To evaluate the homogeneity of the purified protein, we performed size-exclusion chromatography using a Superdex S200. As shown in **Fig. 9C**, the cleaved protein eluted as two distinct peaks. The first peak was observed at approximately ~42ml while the second and predominant peak eluted between 72-80ml. Overlay of the chromatogram with the one corresponding to known molecular weight standards revealed that the predominant form of the cleaved protein eluted near the 50 kDa mark. Considering that the theoretical molecular weight of SMYD5 is 48 kDa (http://web.expasy.org/protparam/) (Gasteiger et al., 2005), the SDS-PAGE gel (**Fig. 9B**) of the predominant peak obtained by gel filtration strongly suggests that the cleaved protein is SMYD5 and that the methyltransferase elutes as a monomer.

3. Recombinant SMYD5 methylates histone H4

While the initial purification of the ~50kDa protein, both by IMAC and size-exclusion chromatography, strongly suggested that the purified polypeptide is SMYD5, we sought to biochemically characterized this polypeptide. SMYD5 has recently been shown to methylate histone H4 *in vitro* (Stender et al., 2012). In agreement with preliminary methylation experiments (**Fig. 10**) and in contrast to the histone H3 lysine methyltransferase SET7/9, the purified protein showed methyltransferase activity toward histone H4 specifically by autoradiography (**Fig. 10A**). These observations strongly indicate that the purified protein is SMYD5 and that the protein is amenable to biochemical characterization.

To identify optimal reaction conditions that favor histone H4 methylation, a concentration curve with different quantities of SMYD5 was determined. The allosteric sigmoidal curve of methylation of histone H4 by different quantities of SMYD5 reveals a linear range of methylation when reactions contain 1 μM of SMYD5 with 2.9 μM of histone H4 (**Fig. 10B**). Buffers at different pH were also tested to determine which conditions favor methylation. Consistent with other SET domain methyltransferases, methylation of histone H4 by SMYD5 is more prominent at a basic pH with maximal activity in bicine pH 9.0 (**Fig. 10C**). These experiments further confirm that SMYD5 is histone H4 methyltransferase and also demonstrate that the purified recombinant SMYD5 is active.

Figure 10.

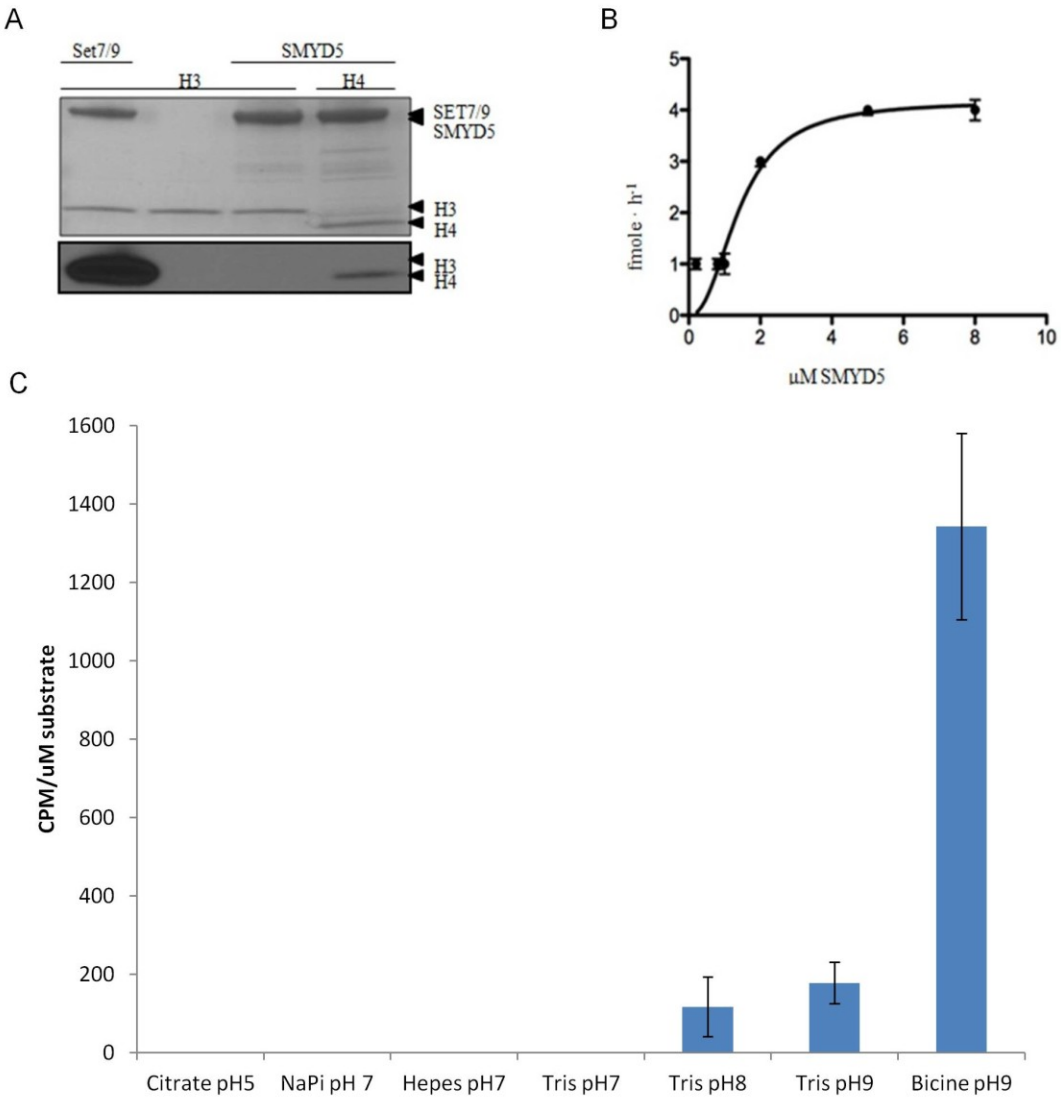

A
Set7/9
SMYD5
H3
H4
SET7/9
SMYD5
H3
H4
H3
H4
B
fmole · h-1
µM SMYD5
C
CPM/uM substrate
Citrate pH5
NaPi pH 7
Hepes pH7
Tris pH7
Tris pH8
Tris pH9
Bicine pH9

Figure 10: Establishing in vitro conditions for the methylation of H4 by SMYD5. SMYD5 methylates histone H4 *in vitro*. **A)** Purified SMYD5, H3, H4 and Set7/9 were incubated with ^{3}H-AdoMet for 1 hour at 30°C, followed by separation on 4-20% (w/v) gradient SDS-PAGE gel and revealed by Coomassie staining (upper panel). Methylated substrates were detected by auto-radiography (lower panel). **B)** Allosteric sigmoidal curve calculated using Prism of the methylation of full length histone H4 in fmole/h using different concentrations of SMYD5. Each point represents one concentration of SMYD5 and error bars represent standard deviation (n=3). **C)** Enzymatic assays using full length histone H4 testing different buffers at different pH with 1 μM SMYD5, with counts represented as CPM/μM substrate. Error bars represent the standard deviation (n=3).

.

4. Crystallization trial of SMYD5

Sequence alignment of SMYD5 with the other members of the SMYD family **(Appendix 4)** revealed several structural differences including different lengths of sequences, large variations in amino acid sequence at the amino and carboxyl termini, as well as additional stretches of amino acids within the SET and MYND regions of SMYD5 not found in the sequences of the other SMYDs. These observations suggest that SMYD5 recognizes its substrate employing divergent structural determinants in comparison to SMYD1-3, for which crystallographic structures are available (Ferguson et al., 2011; Foreman et al., 2011; Jiang et al., 2011; Jiang et al., 2014; Sirinupong et al., 2011; Sirinupong et al., 2010; Xu et al., 2011a; Xu et al., 2011b) whereby providing a strong rationale to undertake structural studies of SMYD5 at the molecular level using X-ray crystallography. Obtaining a crystal structure of SMYD5 in complex with histone H4 would reveal how these differences translate into histone H4 specificity. However, before obtaining the crystal structure of a protein, we must be able to grow diffraction quality crystals; a key but often rate limiting step (Feng et al., 2011). Different approaches were employed, as outlined below, to obtain crystals of SMYD5.

4.1 Crystallization conditions

Vapor diffusion is the most common method currently used for the screening of crystallization conditions (Benvenuti and Mangani, 2007; Giege, 2013). Vapor diffusion by sitting drop method was used for initial 96-well screens, as well as in grid screens to optimize lead conditions. To test as many different reservoir solutions as possible to

determine which chemical compounds facilitate crystallization, commercially available sparse matrix crystallization screens were used, as listed in Section 5 of Materials and Methods. These screens were performed at 4°C, 18°C and 21°C to test the impact of different vapor diffusion rates.

Stabilizing a single conformation of a protein through the addition of macromolecules, inhibitors, additives or chaperones is another way to favor crystallization (Benvenuti and Mangani, 2007; Bukowska and Grutter, 2013; Giege, 2013; Wlodawer et al., 2013). The structures of SMYD1-3 were obtained by crystallization of the protein with the co-factor, co-factor product, inhibitor as well as the addition of a peptide of the known substrate (Ferguson et al., 2011; Foreman et al., 2011; Jiang et al., 2011; Jiang et al., 2014; Sirinupong et al., 2011; Sirinupong et al., 2010; Xu et al., 2011a; Xu et al., 2011b). Using this approach of co-crystallization to stabilize the enzyme, crystallization trials of SMYD5 were conducted with co-factor S-adenosyl-L-methionine (AdoMet) in presence or absence of synthetic histone H4 Lys20 peptides. In summary, approximately 1500 conditions were tested. However, little to no lead conditions were obtained suggesting that the repertoire of tested conditions were insufficient or full-length SMYD5 is not amenable for crystallization.

4.2 Protein engineering: homologues and truncations

To improve the chances of obtaining protein crystals, the polypeptide can also be stabilized through protein engineering either by the introduction of truncations in the protein constructs to obtain a more stable domain, the crystallization of protein

homologs, or incorporating mutations of surface residues. (Benvenuti and Mangani, 2007; Bukowska and Grutter, 2013; Giege, 2013; Wlodawer et al., 2013). Alignment of SMYD5 homologues amongst higher eukaryotes revealed high homology of the amino acid sequence (**Appendix 5**). However, upon closer inspection, differences were identified between the homologues. For example, SMYD5 has a carboxyl terminal poly-glutamate stretch, which could prevent the arrangement of the molecules of SMYD5 into an ordered crystal lattice, thus inhibiting crystallization. The homologue *Pan troglodytes (PtSMYD5)*, for example, is highly similar to *Homo sapiens* SMYD5, with the exception for a deletion of three glutamates within the stretch. This homologue inspired the design of the Δ3E and Δ4E constructs. In addition, a large proportion of the variation found within the homologues is located at the amino and carboxyl termini outside of the conserved SET and MYND domain. The sequences of the homologues *Gallus gallus* (GgSMYD5) and *Drosophila melanogaster* (DmSMYD5) both show variations in length and amino acid sequence of both their termini, while the highly conserved SET and MYND domain remain intact (**Appendix 5**). Similar to Δ3E and Δ4E constructs, cDNAs of these homologues were cloned, expressed and purified (**Fig.11A**) as for full-length *H. sapiens* SMYD5 and screened to obtain diffraction quality crystals. Unfortunately, after screening more than 3000 conditions, no crystals were obtained.

Figure 11.

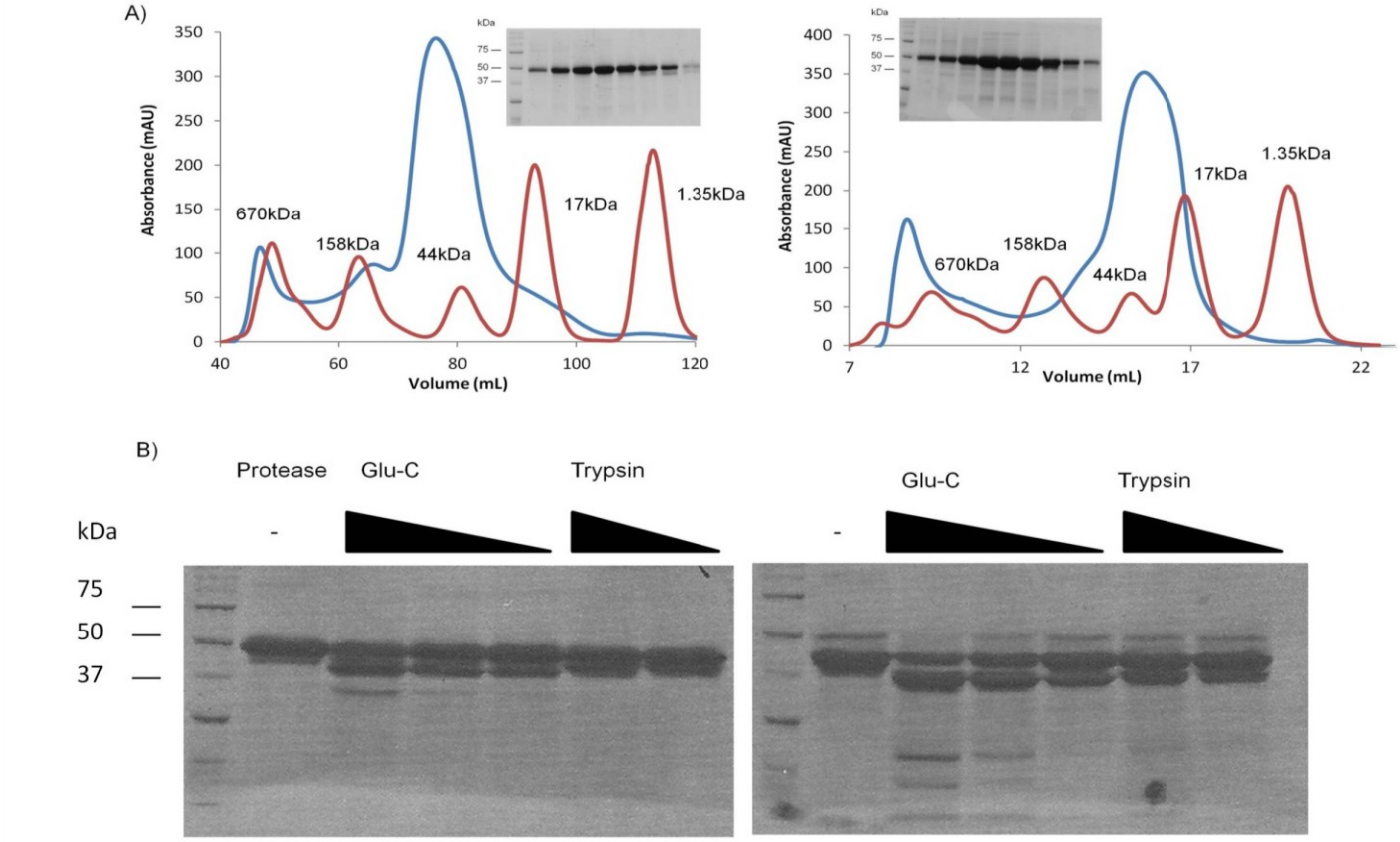

A)
Absorbance (mAU)
Volume (mL)
670kDa
158kDa
44kDa
17kDa
1.35kDa
kDa
75
50
37
B)
Protease
Glu-C
Trypsin
-
kDa
75
50
37

Figure 11: Purification of Drosophila melanogaster and Gallus gallus SMYD5 homologues and limited proteolysis for screening of crystallization conditions.
A) Size exclusion chromatogram of *Drosophila* (left) and *Gallus* (right) with SDS-PAGE of the fractions for each protein as inlay. Molecular weights are indicated to each standard peak (red). B) SDS-PAGE of the stable fragments produced by limited proteolysis of SMYD5 *Drosophila* (left) and *Gallus* (right) homologues using Endoproteinase Glu-C (1:10, 1:25, 1:50 molar ratio Glu-C relative to SMYD5) and trypsin (1:50 and 1:100 w/w ratio trypsin relative to SMYD5) at decreasing concentrations. These proteases were then used for *in situ* proteolysis within the protein drop during crystallization screens.

4.3 Limited proteolysis of SMYD5

Limited proteolysis of protein regions that are predicted to fold as loops has been shown to increase protein crystallizability. These observations are explained by the presence of the loops hindering the assembly of the crystal lattice (Benvenuti and Mangani, 2007; Bukowska and Grutter, 2013; Giege, 2013; Wlodawer et al., 2013). To test whether of regions of SMYD5 could be removed, DmSMYD5 and GgSMYD5 full length constructs were incubated with varying concentrations of endoproteinase Glu-C and trypsin (**Fig. 11B**). As shown in **Fig.11B**, limited proteolysis yielded stable fragments that were used for subsequent crystallization screening.

Detecting a proteolysis-resistant fragments of a given protein, likely to present a more compact fold and thus more amenable to crystallization, can also be used to design constructs. An endoproteinase specific to glutamate (Glu-C) was selected, as the region immediately following the SMYD5 predicted catalytic site is not as conserved evolutionarily and is rich in glutamate residues. The endoproteinase Glu-C cleaves at the carboxyl terminal peptide bond of glutamate and aspartate residues, with a preference for glutamate in certain conditions (Houmard and Drapeau, 1972).

Following the incubation of a 1:25 molar ratio Glu-C to full length *Homo sapiens* SMYD5, a stable fragment was produced (**Fig.12a**). The proteolytic product was analyzed by ESI-MS to obtain a precise measurement of the molecular weight. The stable fragment yielded an intense peak corresponding to 42.3kDa (**Fig. 12b**). As it is not clear where the endoproteinase cleaved based on this molecular weight due to the many glutamate residues found in the sequence, this fragment was further analyzed by

Figure 12.

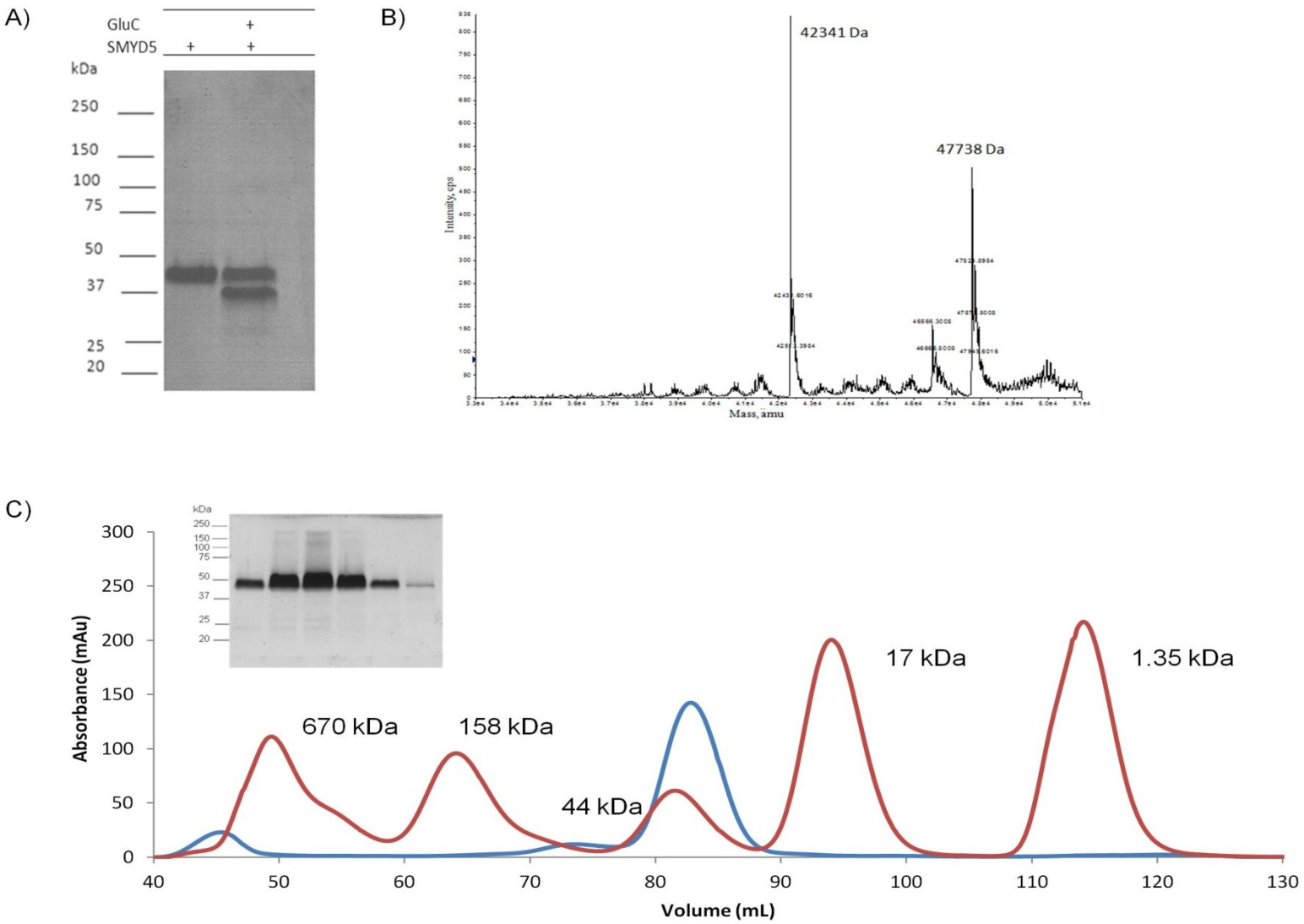

A)
GluC
SMYD5
kDa
250
150
100
75
50
37
25
20
B)
42341 Da
47738 Da
Intensity, cps
Mass, amu
C)
kDa
250
150
100
75
50
37
25
20
Absorbance (mAu)
300
250
200
150
100
50
0
670 kDa
158 kDa
44 kDa
17 kDa
1.35 kDa
40
50
60
70
80
90
100
110
120
130
Volume (mL)

Figure 12: Limited proteolysis of full length Homo sapiens SMYD5 yields a stable fragment.
A) SDS-PAGE of the fragment produced by limited proteolysis of SMYD5. Purified SMYD5 was incubated with a 1:25 molar ratio of Glu-C endoproteinase (30kDa) relative to SMYD5 for 1 hour at room temperature. SMYD5 without Glu-C was used as a control. The reaction was stopped with 0.1% formic acid and analyzed by ESI-MS and N-terminal sequencing. B) ESI-MS results of the stable fragment from limited proteolysis. ESI mass spectrometry analysis of the stable fragment produced from the limited proteolysis of SMYD5 with Glu-C. The peak at 42.3 kDa corresponds to the fragment while the peak at 47.7 kDa corresponds to the full length SMYD5. C) Size exclusion chromatogram with SDS-PAGE gel inlay of the fractions corresponding to the N370X construct designed from the limited proteolysis of Homo sapiens SMYD5. Molecular weights are indicated next to standard peaks (red).

Edman sequencing. Sequencing results revealed that the N-terminus of SMYD5 remains intact following limited proteolysis, indicating that the C-terminus is the only portion of the protein that is removed by the protease. Based on these results, the fragment named N370X was engineered, cloned, expressed and purified (**Fig. 12c**) for additional screening of potential crystallization conditions.

4.4 SMYD5 has yet to be crystallized

Crystallization screening of SMYD5 yielded a number of results, including drops with heavy brown precipitate, clear drops, phase separation, granular precipitate and spherulites (**Fig. 13**). Given that spherulites are considered crystals with high solvent content, additional conditions with similar precipitant, salt and buffer conditions were tested. Unfortunately, these optimization screens failed to improve the quality of the crystals (**Fig. 13B**). Despite these efforts and lead conditions that were optimized, more than 10 000 conditions were screened and none of them have produced a SMYD5 crystal for suitable for structural studies.

5. Establishment of the motif recognized by SMYD5

Recent reports have shown that the use of degenerate peptide arrays is a genuine approach to define the motif recognized by a protein lysine methyltransferase (Couture et al., 2008; Dhayalan et al., 2011; Kudithipudi et al., 2012; Levy et al., 2011; Rathert et al., 2008) (**Fig. 14**). With the identification that SMYD5 methylates Lys20 of histone H4 (Stender et al., 2012), we sought to use this substrate as a starting point to

Figure 13.

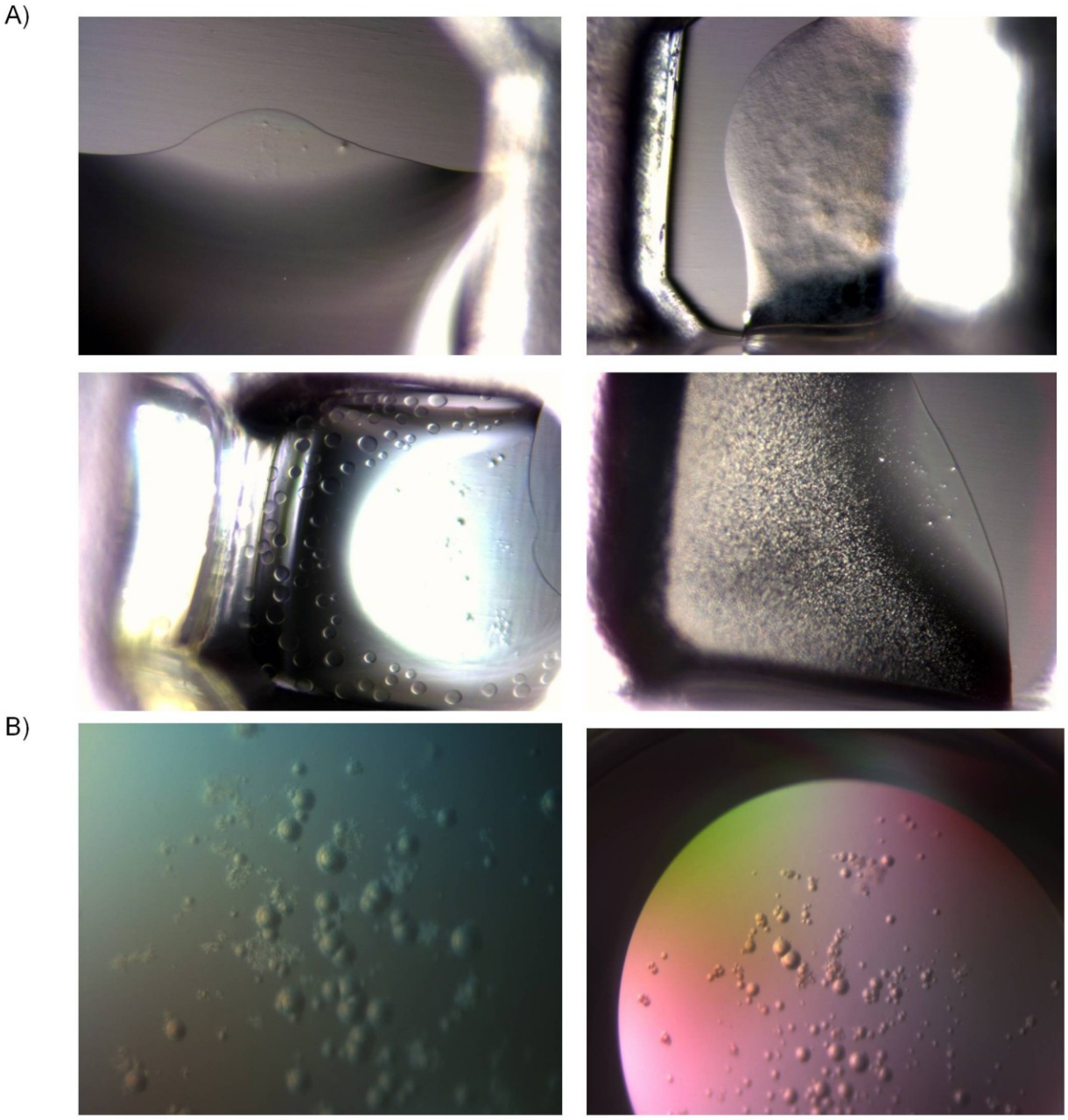


Figure 13: Results of crystallization trials vary, but yield no crystal for discerning a diffraction pattern.
A) Crystallization dynamics vary depending on conditions and often produce one of the following: clear drops (top left), brown heavy precipitate (top right), phase separation (bottom left) and granular precipitate (bottom right). B) Lead conditions in the form of spherulites have been obtained with 26μM SMYD5 N370X with equimolar AdoMet in 20% PEG 400, 0.1M HEPES: NaOH pH 7.5, 0.2M $MgCl_2$ yielded no crystals after optimization.

Figure 14.

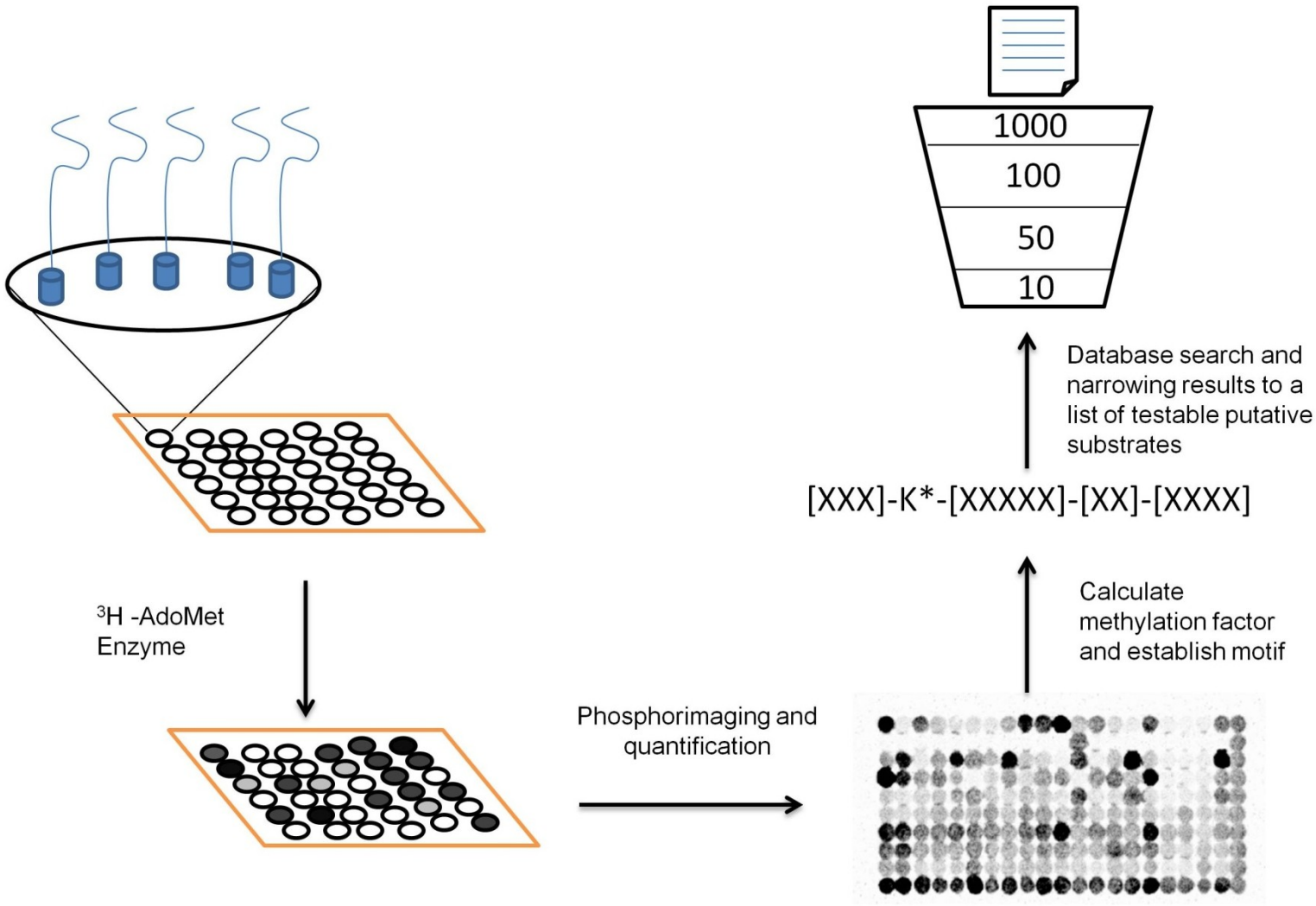

1000
100
50
10
Database search and narrowing results to a list of testable putative substrates
[XXX]-K*-[XXXXX]-[XX]-[XXXX]
Calculate methylation factor and establish motif
Phosphorimaging and quantification
3H -AdoMet
Enzyme

Figure 14: Flow diagram of the steps required to determine the methylation motif and list of putative substrates for further tests.

Cellulose membranes with peptides spotted in an array, where each position of the peptide is mutated to the 20 proteinogenic amino acids. The membrane is incubated with the enzyme and tritiated AdoMet, followed by subsequent washes. The intensity of methylation at each spot is measured by phosphorimaging and quantified. Methylation factors for each spot are calculated and a methylation motif is established encompassing the target lysine (K*). A list of proteins is identified by searching the motif in ScanProsite, followed by database searches, narrowing down the list of putative substrates to test by selecting annotations related to what is available in the literature of the enzyme of interest and identifying if the target lysine of each peptide is conserved evolutionarily.

identify the key residues important for histone H4 recognition and define a motif recognized by the methyltransferase.

Enzymatic assays were performed with purified full-length *Homo sapiens* SMYD5 and an array of peptide derivatives cross-linked to cellulose where each residue of histone H4 (A15-D24) was replaced by all 20 amino acids. To detect methylation, enzymatic assays were performed with tritiated AdoMet, and methylation of peptides was detected and quantified by phosphorimaging (**Fig.15A**). Once the methylation factor was calculated, all residues with a methylation score equal or higher to the wild type residue at the same position were incorporated in the motif (**Fig. 15B**).

Surprisingly, upon analysis of the array data, there was little difference in methylation levels when the Lys20 was mutated to any of the other natural 20 amino acids. However, when Lys16 was mutated to the other amino acids, methylation activity was completely lost. When looking at the effects of substitutions on the non-lysine residues, certain patterns become apparent. In general, polar negatively charged residues are disfavored at any position. In fact, at the Ala15 and His18 positions, any polar residues are generally disfavored, except for the Ala16 to Gln substitution. At the Ala15 position, non-polar aliphatic and aromatic residues tend to increase methylation.

The other position next to Lys16, at Arg17, positive polar, aromatic and the smallest residues (Ala and Gly) favor methylation. Continuing further towards the carboxyl terminus, at the His18 position, aromatic residues exclusively favor methylation. At Arg19 methylation signals are highest when substituted to the smallest residues (Ala and Gly) or to positive polar residues. From the Lys20 position and

Figure 15.

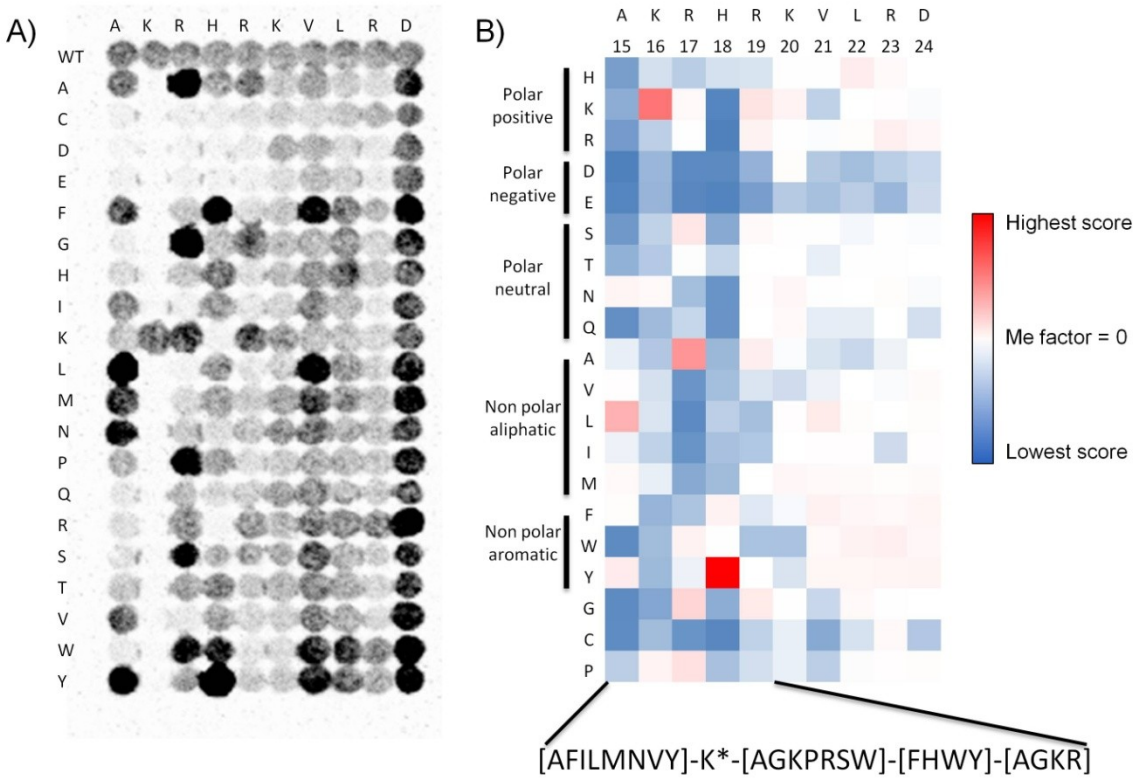

A)
A K R H R K V L R D
WT
A
C
D
E
F
G
H
I
K
L
M
N
P
Q
R
S
T
V
W
Y
B)
A K R H R K V L R D
15 16 17 18 19 20 21 22 23 24
Polar positive
Polar negative
Polar neutral
Non polar aliphatic
Non polar aromatic
H
K
R
D
E
S
T
N
Q
A
V
L
I
M
F
W
Y
G
C
P
Highest score
Me factor = 0
Lowest score
[AFILMNVY]-K*-[AGKPRSW]-[FHWY]-[AGKR]

Figure 15: Establishing a methylation motif for SMYD5 using a peptide encompassing H4 Lys20 reveals Lys16 as the target lysine, used to determine the methylation motif.

A) Phosphorimaging of methylation intensity of each peptide spot, where the sequence on the top of the image corresponds to the 10 amino acid peptide. Under each amino acid, spots represent the wild type (WT) and a mutation of that amino acid to each of the 20 proteinogenic amino acids. B) Heat map of the methylation scores from the array. The position of each amino acid in the peptide is numbered on top, and the mutations to each amino acid are grouped by physiochemical properties of the amino acid. The scale is centralized at a methylation factor of zero (Me factor = 0), with red coloration indication a methylation score >0 and blue coloration indicating a methylation score < 0. Amino acids with methylation scores greater than or equal to the wild type residue at that position were incorporated in the methylation motif indicated below the heat map, with Lys16 as the target lysine (K*).

onwards, although there is some effect of substitution, they are not as pronounced as they are for residues Ala15 to Arg19. Given that the mutation of Lys16 to any other residue eliminates methylation and that the effects of substitutions surrounding Lys16 are more pronounced, while mutating Lys20 does not prevent methylation of the peptide, the peptide motif was designed around Lys16 and the effects of the surrounding substitutions.

Using the Lys16 as the methylated lysine for the motif, a partial motif could be established of [AFILMNVY]-K-[AGKPRSW]-[FHWY]-[AGKR] which resulted in 2393 hits following a motif search on ScanProSite (http://prosite.expasy.org/scanprosite/) (de Castro et al., 2006; Sigrist et al., 2002; Sigrist et al., 2013; Sigrist et al., 2005). Using the Panther Classification System (http://pantherdb.org/) (Mi et al., 2013a; Mi et al., 2013b; Mi and Thomas, 2009), the ScanProsite hits were categorized and annotated.

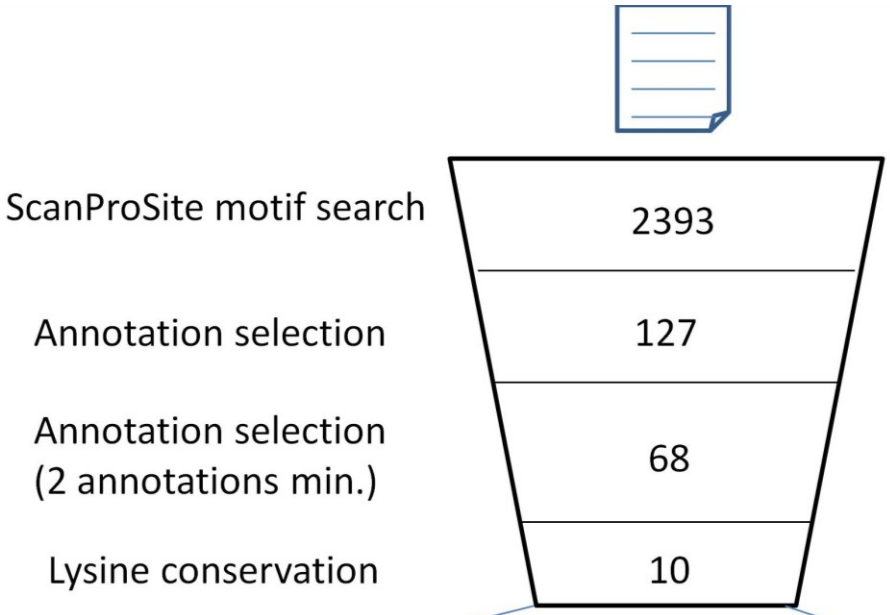
Categories with terms pertaining to SMYD5 were selected: molecular function - transporter activity, translation regulator activity, protein binding transcription factor activity receptor activity, nucleic acid binding transcription factor activity; biological process - cellular process, apoptotic process, developmental process, immune system process; protein class - nucleic acid binding, defense/immunity protein, transcription factor, chaperone; pathway - p53 pathway, Wnt signaling pathway, inflammation mediated by chemokine and cytokine signaling pathway.

Targets within the lists underwent further selection by keeping only the proteins with at least two annotations related to SMYD5, leaving a list of 68 hits. The conservation of the target lysine found in the motif was verified from a ClustalW2 alignment of the *H. sapiens* sequence and homologues from *P. troglodytes, M. musculus, G. gallus, X. laevis* and *D. melanogaster*. Proteins with the target lysine conserved in all the organisms form a list of 10 putative substrates, revealing proteins with various functions (**Fig. 16**).

6. SMYD5 methylates K5/K8/K12/K16/K20

The observation that the substitution of Lys16, and not Lys20, results in a complete loss of activity when mutated to any other amino acid is puzzling given published data identifying SMYD5 as a histone H4 Lys20 methyltransferase (Stender et al., 2012). Tests on peptide strips were conducted to determine the substrate specificity of SMYD5 **(Fig. 17)**. Consistent with our results obtained with the systematic array, SMYD5 methylates peptides harboring a point mutation at Lys20 (**Fig. 17A**). To determine whether SMYD5 is specific to histone H4 or if it will methylate any peptide with a lysine, assays using peptide strips with known SMYD2 methylated peptides from p53, histone H3 and HSP90 were conducted. Methylation is observed on the p53 peptides, but H3 and HSP90 were not methylated by SMYD5 **(Fig. 17B)** suggesting that SMYD5 displays specificity in our experimental conditions.

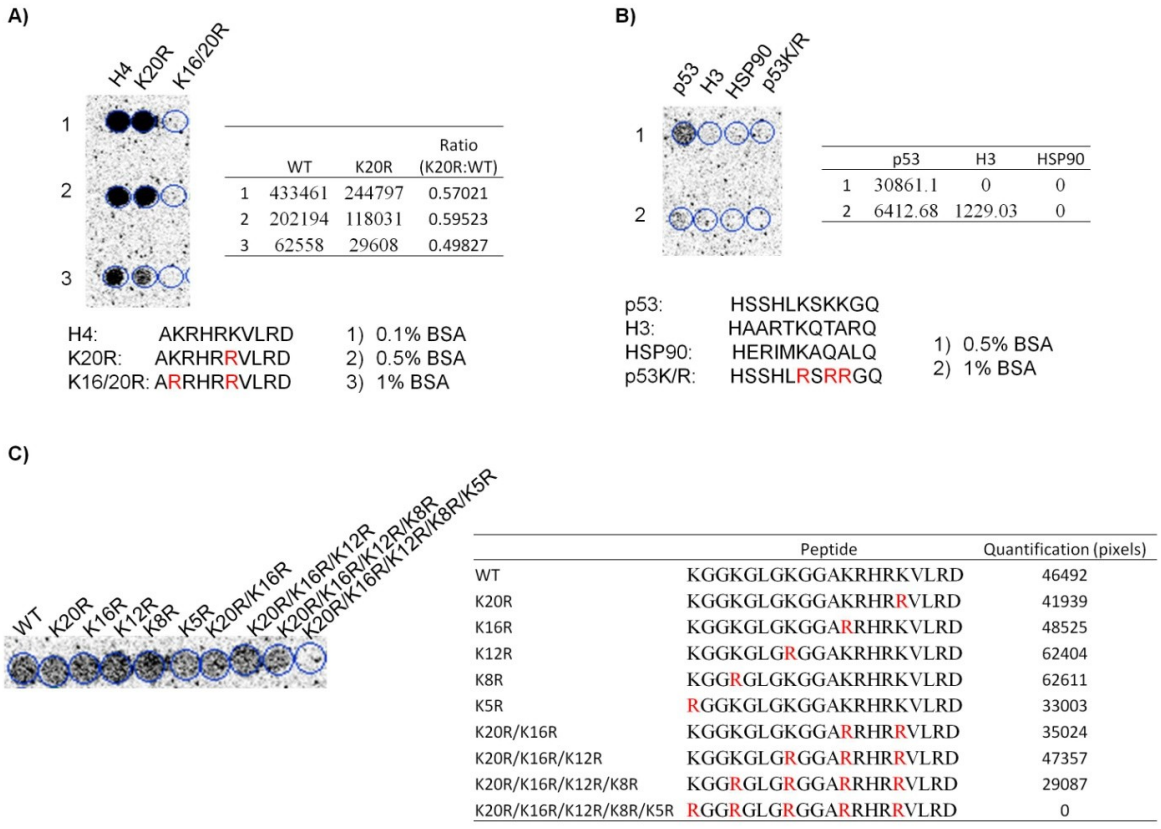
Figure 16.



UniProt Accession	Protein Name	Methylation Motif
P07550	Beta-2 adrenergic receptor;ADRB2	LK*AYG
Q13237	cGMP-dependent protein kinase 2;PRKG2	IK*KHR
Q13547	Histone deacetylase 1;HDAC1	MK*PHR, IK*GHA
Q92769	Histone deacetylase 2;HDAC2	MK*PHR
O15379	Histone deacetylase 3;HDAC3	MK*PHR
P78426	Homeobox protein Nkx-6.1;NKX6-1	LK*KHK
Q9C056	Homeobox protein Nkx-6.2;NKX6-2	LK*KHK
Q969G2	LIM/homeobox protein Lhx4;LHX4	FK*RFG
Q08397	Lysyl oxidase homolog 1;LOXL1	LK*RYA
O60902	Short stature homeobox protein 2;SHOX2	AK*KHA

Figure 16: Diagram illustrating the selection of putative methylation targets from the hits obtained by ScanProSite of the methylation motif, with the final list of testable putative methylation targets.
Each selection step is indicated to the right of the funnel, and the number of hits obtained at each stage. The table underneath indicates the final list of putative methylation targets with their respective accession numbers and the motif found within each protein. An asterisk (*) to the right of a lysine indicates the target lysine verified for conservation within the motif. The conservation of the lysine was verified by a ClustalW2 alignment of the sequences from *H. sapiens, P. troglodytes, M. musculus, G. gallus, X. laevis* and *D. melanogaster*. Proteins identified have the target lysine conserved in all the organisms listed.

Figure 17.



	WT	K20R	Ratio (K20R:WT)
1	433461	244797	0.57021
2	202194	118031	0.59523
3	62558	29608	0.49827

	p53	H3	HSP90
1	30861.1	0	0
2	6412.68	1229.03	0

	Peptide	Quantification (pixels)
WT	KGGKGLGKGGAKRHRKVLRD	46492
K20R	KGGKGLGKGGAKRHRRVLRD	41939
K16R	KGGKGLGKGGARRHRKVLRD	48525
K12R	KGGKGLGRGGAKRHRKVLRD	62404
K8R	KGGRGLGKGGAKRHRKVLRD	62611
K5R	RGGKGLGKGGAKRHRKVLRD	33003
K20R/K16R	KGGKGLGKGGARRHRRVLRD	35024
K20R/K16R/K12R	KGGKGLGRGGARRHRRVLRD	47357
K20R/K16R/K12R/K8R	KGGRGLGRGGARRHRRVLRD	29087
K20R/K16R/K12R/K8R/K5R	RGGRGLGRGGARRHRRVLRD	0

Figure 17: Peptide strips tested to determine substrate specificity of SMYD5. Enzymatic assays using peptide strips from Kinexus were conducted during 2 hours at 30°C with 1µM SMYD5, followed by treating of the membrane and detection on a phosphoscreen using the GE Typhoon Trio Phosphorimager. Images were analyzed using Image Quant software. Far left panels are the images detected while peptide sequences and quantification can be found in centre and reaction conditions tested on the right. **A)** Testing BSA concentrations in 50 mM Bicine pH 9.0, 0.5 mM DTT reactions using the H4 Lys20 peptide and single and double arginine mutants, **B)** Testing peptides known to be methylated by SMYD2 from p53, histone H3 and HSP90 using the same conditions to determine whether SMYD5 methylates specifically histone H4. **C)** Array strip test of single and combinatorial mutations of lysines on the histone H4 tail. Conditions are as mentioned above, with 1% BSA.

Similar to our findings, Gozani and colleagues recently found that SET5, a possible SMYD5 homolog in budding yeast, is able to methylate histone H4 on several lysine residues including K5, K8, K16 and K20 (Green et al., 2012a; Green et al., 2012b). To test whether SMYD5 is functionally analogous to SET5, assays were repeated with peptides harboring systematic and combinatorial mutations on histone H4 tails. As shown in Figure 17, SMYD5 methylates several residues on histone H4 and only the K5R/K8R/K16R/K20R peptide could no longer be methylated (**Fig.17C**).

Discussion

The biological roles of SMYD5 are being elucidated, but the biochemical and structural features which facilitate SMYD5 mediated methylation of its substrates has not been explored. SMYD5 is a histone H4 Lys20 methyltransferase (Stender et al., 2012), but its role likely extends beyond histone methylation as is the case with the other SMYD family members (summarized in **Table 1**). SMYD5 possesses a unique structure in the SMYD family, namely the lack of the protein binding TPR domain (**Fig. 5**) and differences in the amino acid sequence (**Apepndix 4**), which would mediate methylation of its subset of substrates. These include different lengths of sequences, large variations in amino acid sequence at the amino and carboxyl termini, as well as additional stretches of amino acids within the SET and MYND regions of SMYD5 not found in the sequences of the other SMYDs. Exploration of a crystal structure and determining the methylation motif would reveal features at the atomic level found within the enzyme and substrate which facilitate SMYD5 methylation, and furthermore identify novel substrates.

1. Crystallization trials of SMYD5

The *H. sapiens* SMYD5, purified as a pure monomer in solution (**Fig.9**), showed specific activity for histone H4 but not H3 (**Fig. 10**) which support previous studies (Stender et al., 2012). This purified, active recombinant protein was used in a number of crystallization trials. As crystallization is an energetically unfavorable process, it requires the stabilization of a free diffusing molecule in solution into an ordered crystal lattice (Bukowska and Grutter, 2013). Multiple approaches were

used to obtain crystals, including co-crystallization with co-factor AdoMet and a library of H4 Lys20 encompassing peptides (**Fig. 8**). Crystallization conditions with different sparse matrix screens were tested, as well as the impact of vapor diffusion rates by crystallizing at 4°C, 18°C and 21° C. Using limited proteolysis either *in situ* (**Fig. 11**) or to identify a stable fragment which was then expressed and purified as a recombinant protein fragment (**Fig. 12**), stable fragments of SMYD5 were also tested in crystallization studies. Trials of the *H. sapiens* SMYD5 along with a number of homologues and truncated constructs (**Fig. 7**) yielded no crystals after testing over 10 000 conditions. Lead conditions were obtained (**Fig. 13**), yet even after optimization, crystals could not be produced.

Crystallization conditions specific to each protein cannot be predicted, requiring coarse screening of chemical conditions (Garman, 2014) and obtaining a crystal is often the rate limiting step in elucidating a crystal structure (Feng et al., 2011). Different possibilities exist as to why SMYD5 has not yet crystallized. Systematically altering a number of variables including protein concentration, different constructs, crystallization methods (we've tried vapor diffusion but other methods exist including dialysis, free interface diffusion, batch under oil with their variants, listed in (Giege, 2013)) and the physico-chemical crystallization parameters (Benvenuti and Mangani, 2007) yielded different results and possible leads, but the combinations are extensive. The possibility exists that the right combination of all these has not been explored. The right conditions must be identified and it also possible that an additive necessary for crystallization is missing.

As SMYD1-3 have been co-crystallized with co-factor AdoMet, inhibitor or co-factor product with a peptide of a known substrate (Ferguson et al., 2011; Foreman et al., 2011; Jiang et al., 2011; Jiang et al., 2014; Sirinupong et al., 2011; Sirinupong et al., 2010; Xu et al., 2011a; Xu et al., 2011b), SMYD5 likely needs to be co-crystallized with the appropriate peptide substrate. Previous work has lend credence to this idea as the histone H4 Lys20 methyltransferase SET8 cannot be crystallized in absence of peptide (Couture et al., 2005). While a library of peptides encompassing histone H4 Lys20 of various lengths were used, it has not proved to aid in the crystallization of SMYD5. With the peptide array experiments, methylation of the Lys16 residue was lost when it was mutated to any other residue but not Lys20 (**Fig. 15**), suggesting that at least at the peptide level Lys16 is favorite site for SMYD5. Consequently, testing peptides of various lengths centered around Lys16 could represent a better option for co-crystallization experiments. Another option would be to test peptides from the array that have the highest methylation signals.

Chemical characteristics of the surface of the enzyme and conformational heterogeneity can prevent the assembly of the crystal lattice (Bukowska and Grutter, 2013). Prediction software exist to determine physico-chemical properties of the enzyme based on the primary amino acid sequence such as XtalPred-RF (http://ffas.burnham.org/XtalPred-cgi/xtal.pl) (Slabinski et al., 2007a; Slabinski et al., 2007b) and could shed insight into other possibilities of the rational design of constructs more amenable to crystallization. Other approaches exist that we have not tested, including the addition of chemical modifications and mutations on SMYD5 to facilitate crystallization (Benvenuti and Mangani, 2007) to increase

stability or decrease flexibility (Bukowska and Grutter, 2013). Alternatively, we could use natural or non-natural binding partners to stabilize a single conformation (Bukowska and Grutter, 2013).

2. SMYD5: A Lys20 or Lys16 methyltransferase?

Degenerate peptide arrays are increasingly used to define the motif recognized by a protein lysine methyltransferase (Couture et al., 2008; Dhayalan et al., 2011; Kudithipudi et al., 2012; Levy et al., 2011; Rathert et al., 2008) (**Fig. 14**). SMYD5 was demonstrated to methylate histone H4 Lys20 (Stender et al., 2012), therefore a 10 residue peptide encompassing Lys20 was used to create a degenerate peptide array. Each residue was systematically mutated to the 20 natural amino acids, and following the methylation of the peptides and detection, surprisingly Lys16 and not Lys20 was the targeted lysine (**Fig. 15**).

Mutation of Lys16 to any other residue eliminated methylation of the peptide by SMYD5, while mutation of Lys20 still resulted in methylation of the peptide. It is possible that because Lys20 methylation was detected using a polyclonal antibody against H4 Lys20 trimethylation (Stender et al., 2012), there could be cross-reactivity between the antibody and other methylation sites. The authors also showed that SMYD5 is unable to methylate histone H4 when Lys20 is tri-methylated or substituted to alanine (Stender et al., 2012). The bulky trimethyl group could be blocking the enzyme from methylating Lys16 but also from accessing surrounding residues, as Lys16 is close to Lys20 within the primary amino acid sequence. Mutating Lys20 to an alanine also abrogated methylation by SMYD5, yet changing a

larger positively charged residue to a small non-polar aliphatic residue could hinder binding or recognition of the sequence by SMYD5, thus preventing methylation. Accordingly, we observed a detectable decrease of SMYD5 activity on K20A peptide in our peptide array.

To determine which residue is methylated, a methylation assay with SMYD5 and either recombinant histone H4 or an H4 tail peptide followed by identification by mass spectrometry would precisely identify how many lysine residues are methylated. Alternatively, knock down of SMYD5 followed by acid extraction of histones and proteomics would enable the identification of the genuine methylation site in vivo. Finally, given that Lys16 methylation has been found in humans through high-throughput mass spectrometry studies (Jufvas et al., 2011; Tan et al., 2011) the proposed experiments would link the results obtained in peptide arrays and the validation of the SMYD5 functional roles *in vivo*.

3. The Lys16 methylation motif for SMYD5

The methylation motif was thus established using Lys16 as the methylated lysine (**Fig. 15**), although this would only be a partial motif because of the limited amount of residues at the amino terminus. The motif extends to the Arg19 position, as including Lys20 greatly reduces the number of hits following the motif search by ScanProSite. The residue at position 15 is preferentially a non polar aliphatic or aromatic residue. At position 18, residues with aromatic groups are favored and at position 19 either small or positively charged residues are permitted (**Fig. 15**).

Following the identification of proteins harboring the methylation motif and selecting proteins with annotations pertaining to SMYD5 which possess the conserved lysine in higher eukaryotes, 10 proteins have been identified as putative substrates (**Fig. 16**). Although these proteins were identified using the partial motif, they represent interesting candidates as putative SMYD5 substrates. Identifying components of corepressor complexes are of interest, as SMYD5 was demonstrated to co-elute with NCoR and HDAC3 as well as Co-IP with NCoR (Stender et al., 2012). The list of putative substrates for SMYD5 obtained from the arrays reveal HDAC3 as a possible substrate, as well as HDAC1 and HDAC2, also known to associate to co-repressor complexes including Sin3, NuRD and CoREST (Kelly and Cowley, 2013). HDAC1/2 have also been associated with development of tissues and cell types such as brain, immune cells and heart (Kelly and Cowley, 2013), which correspond with the possible role of SMYD5 in neural structure development (Nicolai et al., 2003) as well as the developmental roles associated with the other SMYD family members, particularly in cardiac and muscle development (Diehl et al., 2010; Fujii et al., 2011; Gottlieb et al., 2002; Just et al., 2011; Kawamura et al., 2008; Phan et al., 2005; Tan et al., 2006; Thompson and Travers, 2008).

Given the roles in development of the SMYD family members mentioned previously, the number of homeobox proteins identified following data analysis of the arrays is of interest. Homeobox gene SHOX2 is implicated in brain, heart and limbs development (Blaschke et al., 1998). The expression profiles of transcription factors NKx6.1 and 6.2 in progenitor cells during development of the spinal cord is

regulated by sonic hedgehog signaling (Ribes and Briscoe, 2009) and Nkx-6.1 is expressed in the neural plate during the central nervous system development (Qiu et al., 1998). Homeobox protein NKx6.1 might be important for the development of islet in the pancreas or regulation of the synthesis of insulin (Inoue et al., 1997), while NKx6.2 may play a role in oligodendrocyte formation and regulation of myelin gene expression, as it is highly expressed in the brain may be a candidate tumor suppressor gene for brain tumors (Lee et al., 2001). The homeobox Lhx4 gene is shown to be deregulated in leukemic cells (Kawamata et al., 2002), while studies of orthologs in flies and mice reveal that this homeobox transcription factor is essential in development of the nervous system and pituitary gland, which controls growth, reproduction and stress response, and this gene has been shown to regulate proliferation and differentiation of pituitary lineages (Machinis et al., 2001; Pfaeffle et al., 2008).

A rapid method to validate if these hits represent substrates of SMYD5 consists of *in vitro* methyltransferase assays conducted with each hit as a recombinant protein and purified recombinant SMYD5. Upon detection of methylation by autoradiography, enzymatic assays can be conducted for further analysis by mass spectrometry to determine which residue is methylated and whether or not it corresponds to the target lysine identified by the motif. Additional validation could then be conducted *in vivo* and *in vitro* to determine the effect of this SMYD5 mediated methylation on the substrate. If upon validation these proteins prove to be substrates of SMYD5, this would link SMYD5 to a number of developmental roles, similar to other SMYD family members.

4. Exploring SMYD5 substrate specificity

To explore the substrate specificity of SMYD5 and whether it methylates histone H4 Lys20, enzymatic assays using peptide strips were conducted with a K20R mutation and a combinatorial K16R K20R mutated peptide (**Fig. 17**). Only when both lysine residues were mutated did the activity disappear. SMYD5 has the ability to methylate Lys16 on the peptide. Next, known substrates for SMYD2 were tested to see if SMYD5 is H4 specific or will methylate other lysines known to be methylated. Aside from methylation detected on p53, SMYD5 does not methylate histone H3 and HSP90 demonstrating that there is some specificity in methylation of lysine residues by the enzyme.

Similar to our findings, Gozani and colleagues recently found that SET5, a possible SMYD5 homolog in budding yeast, is able to methylate histone H4 on several lysine residues including K5, K8, K16 and K20 (Green et al., 2012a; Green et al., 2012b). To test whether SMYD5 is functionally analogous to SET5, assays were repeated with peptides harboring systematic and combinatorial mutations on histone H4 tails. Consistent with recent observations on SET5, SMYD5 methylates several residues on histone H4 and only the K5R/K8R/K16R/K20R peptide could no longer be methylated by SMYD5 (**Fig.17C**).

Conclusion

SMYD5, identified as an H4 Lys20 methyltransferase, may have a role outside of histone methylation and modulating chromatin structure and function. To explore this role, a crystal structure would shed insight as to the specific biochemical and structural features within SMYD5 mediate its particular activity. In addition, an in-house peptide array would reveal which determinants of the substrate facilitate methylation, and would also allow to identify additional substrates.

After over 10 000 crystallization experiments with SMYD5, a crystal amenable for structure determination was not obtained. With the discovery that peptides with the Lys16 and not the Lys20 residue on the array lost all activity when mutated, this reveals the possibility that it is in fact Lys16 and not Lys20 that is the substrate of SMYD5. This means that for future crystallization trials, a library of Lys16 peptides might prove to aid in the stabilization of SMYD5 sufficiently for its organization in the crystal lattice. However, prior to screening the Lys16 peptide library, it would be prudent to conduct enzymatic assays with SMYD5 and a Lys16 peptide followed by mass spectrometry analysis to confirm that it is indeed the lysine methylated by SMYD5. In addition, the peptides with stronger methylation signal can also be used in co-crystallization experiments with SMYD5 as the stronger signal suggests that these peptides may provide better substrates for SMYD5.

The peptide arrays not only revealed that the Lys16 and not the Lys20 seems to be the lysine targeted by SMYD5, but it also revealed a number of

mutations which favor or diminish methylation by SMYD5. Polar negative residues are generally disfavored throughout. At the Ala15 position, non polar aliphatic and aromatic residues are favored, while at the Arg17 polar positive, aromatic and small residues are generally favored. Continuing further towards the carboxyl terminus at His18, only aromatics strongly favor methylation, followed by either polar positive or small residues at the Arg19 position. Conducting arrays with more residues at the amino terminus would shed light on the impact of additional residues, and may provide a complete methylation motif. Using the current motif with Lys16, however, already reveals interesting putative substrates that associate with chromatin remodeling complexes or have a role in development.

To determine which residue on the histone H4 tail is methylated by SMYD5 and whether it is functionally analogous to the yeast SET domain methyltransferase SET5, we conducted enzymatic assays on peptide strips with sequential and combinatorial mutations of all the lysine residues, including a peptide with all lysine mutated to arginine. We found that SMYD5 methylates several residues on the histone H4 tail, raising a number of questions as to what confers specificity in methylation by SMYD5. One option is to look at the substrates and whether their three dimensional structure affect SMYD5 methylation. By conducting enzymatic assays with histone H4, the tetramer, octamer and nucleosome would provide information on the potential role of the impact of the 3D structure of the substrate on the activity of SMYD5. Combined enzymatic assays performed with these substrates in which the target lysines are substituted to arginine residues would help identify SMYD5 methylation sites on histone H4.

Another option is to observe SMYD5 in the context of its interacting partners and determine if the proteins associated with it are responsible for its selectivity. Conducting enzymatic assays with the SMYD5 complexes either with wild type or mutant substrates would help us identify if associating subunits restrict/change SMYD5 activity.

While a crystal structure of SMYD5 is still pending, the results of the arrays present exciting new prospects as to possible roles of SMYD5 outside of chromatin. In addition, these results are revealing what characteristics of the substrate at the primary amino acid sequence level enable methylation of the target lysine by SMYD5.

References

Abu-Farha, M., Lanouette, S., Elisma, F., Tremblay, V., Butson, J., Figeys, D., and Couture, J.F. (2011). Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2. J Mol Cell Biol *3*, 301-308.

Acharya, K.R., and Lloyd, M.D. (2005). The advantages and limitations of protein crystal structures. Trends Pharmacol Sci *26*, 10-14.

Afjehi-Sadat, L., and Garcia, B.A. (2013). Comprehending dynamic protein methylation with mass spectrometry. Curr Opin Chem Biol *17*, 12-19.

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr *68*, 352-367.

Allen, H.F., Wade, P.A., and Kutateladze, T.G. (2013). The NuRD architecture. Cell Mol Life Sci *70*, 3513-3524.

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. Proc Natl Acad Sci U S A *51*, 786-794.

Ambler, R.P., and Rees, M.W. (1959). Epsilon-N-Methyl-lysine in bacterial flagellar protein. Nature *184*, 56-57.

Ames, G.F., and Niakido, K. (1979). In vivo methylation of prokaryotic elongation factor Tu. J Biol Chem *254*, 9947-9950.

Arnaudo, A.M., and Garcia, B.A. (2013). Proteomic characterization of novel histone post-translational modifications. Epigenetics Chromatin *6*, 24.

Ayadi, A., Birling, M.C., Bottomley, J., Bussell, J., Fuchs, H., Fray, M., Gailus-Durner, V., Greenaway, S., Houghton, R., Karp, N.*, et al.* (2012). Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. Mamm Genome *23*, 600-610.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Benvenuti, M., and Mangani, S. (2007). Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. Nat Protoc *2*, 1633-1651.

Bhaumik, S.R., Smith, E., and Shilatifard, A. (2007). Covalent modifications of histones during development and disease pathogenesis. Nat Struct Mol Biol *14*, 1008-1016.

Blaschke, R.J., Monaghan, A.P., Schiller, S., Schechinger, B., Rao, E., Padilla-Nash, H., Ried, T., and Rappold, G.A. (1998). SHOT, a SHOX-related homeobox gene, is implicated in craniofacial, brain, heart, and limb development. Proc Natl Acad Sci U S A *95*, 2406-2411.

Bloxham, D.P., Parmelee, D.C., Kumar, S., Wade, R.D., Ericsson, L.H., Neurath, H., Walsh, K.A., and Titani, K. (1981). Primary structure of porcine heart citrate synthase. Proc Natl Acad Sci U S A *78*, 5381-5385.

Brown, M.A., Sims, R.J., 3rd, Gottlieb, P.D., and Tucker, P.W. (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.

Bukowska, M.A., and Grutter, M.G. (2013). New concepts and aids to facilitate crystallization. Curr Opin Struct Biol *23*, 409-416.

Chen, L.B., Xu, J.Y., Yang, Z., and Wang, G.B. (2007). Silencing SMYD3 in hepatoma demethylates RIZI promoter induces apoptosis and inhibits cell proliferation and migration. World J Gastroenterol *13*, 5718-5724.

Cheng, X., and Zhang, X. (2007). Structural dynamics of protein lysine methylation and demethylation. Mutation research *618*, 102-115.

Cho, H.S., Hayami, S., Toyokawa, G., Maejima, K., Yamane, Y., Suzuki, T., Dohmae, N., Kogure, M., Kang, D., Neal, D.E.*, et al.* (2012). RB1 methylation by SMYD2 enhances cell cycle progression through an increase of RB1 phosphorylation. Neoplasia *14*, 476-486.

Cho, K.S., Elizondo, L.I., and Boerkoel, C.F. (2004). Advances in chromatin remodeling and human disease. Curr Opin Genet Dev *14*, 308-315.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science *325*, 834-840.

Chu, Y., Zhang, Z., Wang, Q., Luo, Y., and Huang, L. (2012). Identification and characterization of a highly conserved crenarchaeal protein lysine methyltransferase with broad substrate specificity. J Bacteriol *194*, 6917-6926.

Chuikov, S., Kurash, J.K., Wilson, J.R., Xiao, B., Justin, N., Ivanov, G.S., McKinney, K., Tempst, P., Prives, C., Gamblin, S.J.*, et al.* (2004). Regulation of p53 activity through lysine methylation. Nature *432*, 353-360.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu Rev Biochem *78*, 273-304.

Cock-Rada, A.M., Medjkane, S., Janski, N., Yousfi, N., Perichon, M., Chaussepied, M., Chluba, J., Langsley, G., and Weitzman, J.B. (2012). SMYD3 promotes cancer

invasion by epigenetic upregulation of the metalloproteinase MMP-9. Cancer Res *72*, 810-820.

Corona, D.F., and Tamkun, J.W. (2004). Multiple roles for ISWI in transcription, chromosome organization and DNA replication. Biochim Biophys Acta *1677*, 113-119.

Couture, J.F., Collazo, E., Brunzelle, J.S., and Trievel, R.C. (2005). Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. Genes Dev *19*, 1455-1465.

Couture, J.F., Dirk, L.M., Brunzelle, J.S., Houtz, R.L., and Trievel, R.C. (2008). Structural origins for the product specificity of SET domain protein methyltransferases. Proc Natl Acad Sci U S A *105*, 20659-20664.

Couture, J.F., and Trievel, R.C. (2006). Histone-modifying enzymes: encrypting an enigmatic epigenetic code. Curr Opin Struct Biol *16*, 753-760.

de Castro, E., Sigrist, C.J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A., and Hulo, N. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res *34*, W362-365.

De Koning, L., Corpet, A., Haber, J.E., and Almouzni, G. (2007). Histone chaperones: an escort network regulating histone traffic. Nat Struct Mol Biol *14*, 997-1007.

Dechassa, M.L., Sabri, A., Pondugula, S., Kassabov, S.R., Chatterjee, N., Kladde, M.P., and Bartholomew, B. (2010). SWI/SNF has intrinsic nucleosome disassembly activity that is dependent on adjacent nucleosomes. Mol Cell *38*, 590-602.

Dhayalan, A., Kudithipudi, S., Rathert, P., and Jeltsch, A. (2011). Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. Chem Biol *18*, 111-120.

Diehl, F., Brown, M.A., van Amerongen, M.J., Novoyatleva, T., Wietelmann, A., Harriss, J., Ferrazzi, F., Bottger, T., Harvey, R.P., Tucker, P.W.*, et al.* (2010). Cardiac deletion of Smyd2 is dispensable for mouse heart development. PLoS One *5*, e9748.

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I.*, et al.* (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS Biol *9*, e1000582.

Dillon, S.C., Zhang, X., Trievel, R.C., and Cheng, X. (2005a). The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol *6*, 227.

Dillon, S.C., Zhang, X., Trievel, R.C., and Cheng, X. (2005b). The SET-domain protein superfamily: protein lysine methyltransferases. Genome biology *6*, 227.

Ea, C.K., and Baltimore, D. (2009). Regulation of NF-kappaB activity through lysine monomethylation of p65. Proc Natl Acad Sci U S A *106*, 18972-18977.

Ehrenhofer-Murray, A.E. (2004). Chromatin dynamics at DNA replication, transcription and repair. Eur J Biochem *271*, 2335-2349.

Feng, W., Pan, L., and Zhang, M. (2011). Combination of NMR spectroscopy and X-ray crystallography offers unique advantages for elucidation of the structural basis of protein complex assembly. Sci China Life Sci *54*, 101-111.

Ferguson, A.D., Larsen, N.A., Howard, T., Pollard, H., Green, I., Grande, C., Cheung, T., Garcia-Arenas, R., Cowen, S., Wu, J.*, et al.* (2011). Structural basis of substrate methylation and inhibition of SMYD2. Structure *19*, 1262-1273.

Finch, J.T., Lutter, L.C., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M., and Klug, A. (1977). Structure of nucleosome core particles of chromatin. Nature *269*, 29-36.

Foreman, K.W., Brown, M., Park, F., Emtage, S., Harriss, J., Das, C., Zhu, L., Crew, A., Arnold, L., Shaaban, S.*, et al.* (2011). Structural and functional profiling of the human histone methyltransferase SMYD3. PLoS One *6*, e22290.

Fujii, T., Tsunesumi, S., Yamaguchi, K., Watanabe, S., and Furukawa, Y. (2011). Smyd3 is required for the development of cardiac and skeletal muscle in zebrafish. PLoS One *6*, e23491.

Garman, E.F. (2014). Developments in x-ray crystallographic structure determination of biological macromolecules. Science *343*, 1102-1108.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.e., Wilkins, M., Appel, R., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook, 571-607.

Giege, R. (2013). A historical perspective on protein crystallization from 1840 to the present day. FEBS J *280*, 6456-6497.

Gottlieb, P.D., Pierce, S.A., Sims, R.J., Yamagishi, H., Weihe, E.K., Harriss, J.V., Maika, S.D., Kuziel, W.A., King, H.L., Olson, E.N.*, et al.* (2002). Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis. Nat Genet *31*, 25-32.

Green, E.M., Mas, G., Young, N.L., Garcia, B.A., and Gozani, O. (2012a). Methylation of H4 lysines 5, 8 and 12 by yeast Set5 calibrates chromatin stress responses. Nat Struct Mol Biol *19*, 361-363.

Green, E.M., Morrison, A.J., and Gozani, O. (2012b). New marks on the block: Set5 methylates H4 lysines 5, 8 and 12. Nucleus *3*, 335-339.

Guo, A., Gu, H., Zhou, J., Mulhern, D., Wang, Y., Lee, K.A., Yang, V., Aguiar, M., Kornhauser, J., Jia, X.*, et al.* (2013). Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. Mol Cell Proteomics *13*, 372-387.

Guo, A., Gu, H., Zhou, J., Mulhern, D., Wang, Y., Lee, K.A., Yang, V., Aguiar, M., Kornhauser, J., Jia, X.*, et al.* (2014). Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. Mol Cell Proteomics *13*, 372-387.

Hamamoto, R., Furukawa, Y., Morita, M., Iimura, Y., Silva, F.P., Li, M., Yagyu, R., and Nakamura, Y. (2004). SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. Nat Cell Biol *6*, 731-740.

Hamamoto, R., Silva, F.P., Tsuge, M., Nishidate, T., Katagiri, T., Nakamura, Y., and Furukawa, Y. (2006). Enhanced SMYD3 expression is essential for the growth of breast cancer cells. Cancer Sci *97*, 113-118.

Hansen, J.C., Tse, C., and Wolffe, A.P. (1998). Structure and function of the core histone N-termini: more than meets the eye. Biochemistry *37*, 17637-17641.

Henriksen, P., Wagner, S.A., Weinert, B.T., Sharma, S., Bacinskaja, G., Rehman, M., Juffer, A.H., Walther, T.C., Lisby, M., and Choudhary, C. (2012). Proteome-wide analysis of lysine acetylation suggests its broad regulatory scope in Saccharomyces cerevisiae. Mol Cell Proteomics *11*, 1510-1522.

Hock, R., Furusawa, T., Ueda, T., and Bustin, M. (2007). HMG chromosomal proteins in development and disease. Trends Cell Biol *17*, 72-79.

Houmard, J., and Drapeau, G.R. (1972). Staphylococcal protease: a proteolytic enzyme specific for glutamoyl bonds. Proc Natl Acad Sci U S A *69*, 3506-3509.

Hu, L., Zhu, Y.T., Qi, C., and Zhu, Y.J. (2009). Identification of Smyd4 as a potential tumor suppressor gene involved in breast cancer development. Cancer Res *69*, 4067-4072.

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

Huang, J., Sengupta, R., Espejo, A.B., Lee, M.G., Dorsey, J.A., Richter, M., Opravil, S., Shiekhattar, R., Bedford, M.T., Jenuwein, T.*, et al.* (2007). p53 is regulated by the lysine demethylase LSD1. Nature *449*, 105-108.

Huq, M.D., Ha, S.G., Barcelona, H., and Wei, L.N. (2009). Lysine methylation of nuclear co-repressor receptor interacting protein 140. J Proteome Res *8*, 1156-1167.

Inoue, H., Rudnick, A., German, M.S., Veile, R., Donis-Keller, H., and Permutt, M.A. (1997). Isolation, characterization, and chromosomal mapping of the human Nkx6.1 gene (NKX6A), a new pancreatic islet homeobox gene. Genomics *40*, 367-370.

Ito, T., Bulger, M., Pazin, M.J., Kobayashi, R., and Kadonaga, J.T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. Cell *90*, 145-155.

Iwabata, H., Yoshida, M., and Komatsu, Y. (2005). Proteomic analysis of organ-specific post-translational lysine-acetylation and -methylation in mice by use of anti-acetyllysine and -methyllysine mouse monoclonal antibodies. Proteomics *5*, 4653-4664.

Jiang, Y., Sirinupong, N., Brunzelle, J., and Yang, Z. (2011). Crystal structures of histone and p53 methyltransferase SmyD2 reveal a conformational flexibility of the autoinhibitory C-terminal domain. PLoS One *6*, e21640.

Jiang, Y., Trescott, L., Holcomb, J., Zhang, X., Brunzelle, J., Sirinupong, N., Shi, X., and Yang, Z. (2014). Structural Insights into Estrogen Receptor alpha Methylation by Histone Methyltransferase SMYD2, a Cellular Event Implicated in Estrogen Signaling Regulation. J Mol Biol.

Jones, R.S., and Gelbart, W.M. (1993). The Drosophila Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax. Mol Cell Biol *13*, 6357-6366.

Jufvas, A., Stralfors, P., and Vener, A.V. (2011). Histone variants and their post-translational modifications in primary human fat cells. PLoS One *6*, e15960.

Just, S., Meder, B., Berger, I.M., Etard, C., Trano, N., Patzel, E., Hassel, D., Marquart, S., Dahme, T., Vogel, B.*, et al.* (2011). The myosin-interacting protein SMYD1 is essential for sarcomere organization. J Cell Sci *124*, 3127-3136.

Karch, K.R., Denizio, J.E., Black, B.E., and Garcia, B.A. (2013). Identification and interrogation of combinatorial histone modifications. Front Genet *4*, 264.

Kawamata, N., Sakajiri, S., Sugimoto, K.J., Isobe, Y., Kobayashi, H., and Oshimi, K. (2002). A novel chromosomal translocation t(1;14)(q25;q32) in pre-B acute lymphoblastic leukemia involves the LIM homeodomain protein gene, Lhx4. Oncogene *21*, 4983-4991.

Kawamura, S., Yoshigai, E., Kuhara, S., and Tashiro, K. (2008). smyd1 and smyd2 are expressed in muscle tissue in Xenopus laevis. Cytotechnology *57*, 161-168.

Kelley, R.I. (1973). Isolation of a histone IIb1-IIb2 complex. Biochem Biophys Res Commun *54*, 1588-1594.

Kelly, R.D., and Cowley, S.M. (2013). The physiological roles of histone deacetylase (HDAC) 1 and 2: complex co-stars with multiple leading parts. Biochem Soc Trans *41*, 741-749.

Komatsu, S., Imoto, I., Tsuda, H., Kozaki, K.I., Muramatsu, T., Shimada, Y., Aiko, S., Yoshizumi, Y., Ichikawa, D., Otsuji, E.*, et al.* (2009). Overexpression of SMYD2 relates to tumor cell proliferation and malignant outcome of esophageal squamous cell carcinoma. Carcinogenesis *30*, 1139-1146.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285-294.

Kornberg, R.D., and Thomas, J.O. (1974). Chromatin structure; oligomers of the histones. Science *184*, 865-868.

Kouskouti, A., Scheer, E., Staub, A., Tora, L., and Talianidis, I. (2004). Gene-specific modulation of TAF10 function by SET9-mediated methylation. Mol Cell *14*, 175-182.

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

Kudithipudi, S., Dhayalan, A., Kebede, A.F., and Jeltsch, A. (2012). The SET8 H4K20 protein lysine methyltransferase has a long recognition sequence covering seven amino acid residues. Biochimie *94*, 2212-2218.

Kunizaki, M., Hamamoto, R., Silva, F.P., Yamaguchi, K., Nagayasu, T., Shibuya, M., Nakamura, Y., and Furukawa, Y. (2007). The lysine 831 of vascular endothelial growth factor receptor 1 is a novel target of methylation by SMYD3. Cancer Res *67*, 10759-10765.

Kurash, J.K., Lei, H., Shen, Q., Marston, W.L., Granda, B.W., Fan, H., Wall, D., Li, E., and Gaudet, F. (2008). Methylation of p53 by Set7/9 mediates p53 acetylation and activity in vivo. Mol Cell *29*, 392-400.

L'Italien, J.J., and Laursen, R.A. (1979). Location of the site of methylation in elongation factor Tu. FEBS Lett *107*, 359-362.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Law, C., and Cheung, P. (2013). Histone variants and transcription regulation. Subcell Biochem *61*, 319-341.

Lee, S.H., Davison, J.A., Vidal, S.M., and Belouchi, A. (2001). Cloning, expression and chromosomal location of NKX6B TO 10Q26, a region frequently deleted in brain tumors. Mamm Genome *12*, 157-162.

Levy, D., Liu, C.L., Yang, Z., Newman, A.M., Alizadeh, A.A., Utz, P.J., and Gozani, O. (2011). A proteomic approach for the identification of novel lysine methyltransferase substrates. Epigenetics Chromatin *4*, 19.

Liu, H., Galka, M., Mori, E., Liu, X., Lin, Y.F., Wei, R., Pittock, P., Voss, C., Dhami, G., Li, X.*, et al.* (2013). A method for systematic mapping of protein lysine methylation identifies functions for HP1beta in DNA damage response. Mol Cell *50*, 723-735.

Liu, Y., Chen, W., Gaudet, J., Cheney, M.D., Roudaia, L., Cierpicki, T., Klet, R.C., Hartman, K., Laue, T.M., Speck, N.A.*, et al.* (2007). Structural basis for recognition of SMRT/N-CoR by the MYND domain and its contribution to AML1/ETO's activity. Cancer Cell *11*, 483-497.

Loyola, A., and Almouzni, G. (2004). Histone chaperones, a supporting role in the limelight. Biochim Biophys Acta *1677*, 3-11.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luo, X.G., Ding, Y., Zhou, Q.F., Ye, L., Wang, S.Z., and Xi, T. (2007). SET and MYND domain-containing protein 3 decreases sensitivity to dexamethasone and stimulates cell adhesion and migration in NIH3T3 cells. J Biosci Bioeng *103*, 444-450.

Luo, X.G., Xi, T., Guo, S., Liu, Z.P., Wang, N., Jiang, Y., and Zhang, T.C. (2009). Effects of SMYD3 overexpression on transformation, serum dependence, and apoptosis sensitivity in NIH3T3 cells. IUBMB Life *61*, 679-684.

Luo, X.G., Zhang, C.L., Zhao, W.W., Liu, Z.P., Liu, L., Mu, A., Guo, S., Wang, N., Zhou, H., and Zhang, T.C. (2014). Histone methyltransferase SMYD3 promotes MRTF-A-mediated transactivation of MYL9 and migration of MCF-7 breast cancer cells. Cancer Lett *344*, 129-137.

Machinis, K., Pantel, J., Netchine, I., Leger, J., Camand, O.J., Sobrier, M.L., Dastot-Le Moal, F., Duquesnoy, P., Abitbol, M., Czernichow, P.*, et al.* (2001). Syndromic short stature in patients with a germline mutation in the LIM homeobox LHX4. Am J Hum Genet *69*, 961-968.

Martin, J.L., and McMillan, F.M. (2002). SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Curr Opin Struct Biol *12*, 783-793.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013a). Large-scale gene function analysis with the PANTHER classification system. Nat Protoc *8*, 1551-1566.

Mi, H., Muruganujan, A., and Thomas, P.D. (2013b). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res *41*, D377-386.

Mi, H., and Thomas, P. (2009). PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. Methods Mol Biol *563*, 123-140.

Min, J., Feng, Q., Li, Z., Zhang, Y., and Xu, R.M. (2003). Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. Cell *112*, 711-723.

Molla, A., Kilhoffer, M.C., Ferraz, C., Audemard, E., Walsh, M.P., and Demaille, J.G. (1981). Octopus calmodulin. The trimethyllysyl residue is not required for myosin light chain kinase activation. J Biol Chem *256*, 15-18.

Munro, S., Khaire, N., Inche, A., Carr, S., and La Thangue, N.B. (2010). Lysine methylation regulates the pRb tumour suppressor protein. Oncogene *29*, 2357-2367.

Murray, K. (1964). THE OCCURRENCE OF EPSILON-N-METHYL LYSINE IN HISTONES. Biochemistry *3*, 10-15.

Nguyen, A.T., and Zhang, Y. (2011). The diverse functions of Dot1 and H3K79 methylation. Genes Dev *25*, 1345-1358.

Nicolai, M., Lasbleiz, C., and Dura, J.M. (2003). Gain-of-function screen identifies a role of the Src64 oncogene in Drosophila mushroom body development. J Neurobiol *57*, 291-302.

Noll, M., and Kornberg, R.D. (1977). Action of micrococcal nuclease on chromatin and the location of histone H1. J Mol Biol *109*, 393-404.

Northcott, P.A., Nakahara, Y., Wu, X., Feuk, L., Ellison, D.W., Croul, S., Mack, S., Kongkham, P.N., Peacock, J., Dubuc, A.*, et al.* (2009). Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. Nat Genet *41*, 465-472.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science *183*, 330-332.

Ong, S.E., Mittler, G., and Mann, M. (2004). Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. Nat Methods *1*, 119-126.

Pagans, S., Kauder, S.E., Kaehlcke, K., Sakane, N., Schroeder, S., Dormeyer, W., Trievel, R.C., Verdin, E., Schnolzer, M., and Ott, M. (2010). The Cellular lysine methyltransferase Set7/9-KMT7 binds HIV-1 TAR RNA, monomethylates the viral transactivator Tat, and enhances HIV transcription. Cell Host Microbe *7*, 234-244.

Petrossian, T.C., and Clarke, S.G. (2011). Uncovering the human methyltransferasome. Mol Cell Proteomics *10*, M110 000976.

Pfaeffle, R.W., Hunter, C.S., Savage, J.J., Duran-Prado, M., Mullen, R.D., Neeb, Z.P., Eiholzer, U., Hesse, V., Haddad, N.G., Stobbe, H.M.*, et al.* (2008). Three novel missense mutations within the LHX4 gene are associated with variable pituitary hormone deficiencies. J Clin Endocrinol Metab *93*, 1062-1071.

Phan, D., Rasmussen, T.L., Nakagawa, O., McAnally, J., Gottlieb, P.D., Tucker, P.W., Richardson, J.A., Bassel-Duby, R., and Olson, E.N. (2005). BOP, a regulator of right ventricular heart development, is a direct transcriptional target of MEF2C in the developing heart. Development *132*, 2669-2678.

Ptacek, J., Devgan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R.*, et al.* (2005). Global analysis of protein phosphorylation in yeast. Nature *438*, 679-684.

Qian, C., and Zhou, M.-M. (2006). SET domain protein lysine methyltransferases: Structure, specificity and catalysis. Cellular and molecular life sciences : CMLS *63*, 2755-2763.

Qiu, M., Shimamura, K., Sussel, L., Chen, S., and Rubenstein, J.L. (1998). Control of anteroposterior and dorsoventral domains of Nkx-6.1 gene expression relative to other Nkx genes during vertebrate CNS development. Mech Dev *72*, 77-88.

Rathert, P., Dhayalan, A., Murakami, M., Zhang, X., Tamas, R., Jurkowska, R., Komatsu, Y., Shinkai, Y., Cheng, X., and Jeltsch, A. (2008). Protein lysine methyltransferase G9a acts on non-histone targets. Nat Chem Biol *4*, 344-346.

Ray-Gallet, D., and Almouzni, G. (2010). Nucleosome dynamics and histone variants. Essays Biochem *48*, 75-87.

Ribes, V., and Briscoe, J. (2009). Establishing and interpreting graded Sonic Hedgehog signaling during vertebrate neural tube patterning: the role of negative feedback. Cold Spring Harb Perspect Biol *1*, a002014.

Roark, D.E., Geoghegan, T.E., and Keller, G.H. (1974). A two-subunit histone complex from calf thymus. Biochem Biophys Res Commun *59*, 542-547.

Roberts, D.M., Rowe, P.M., Siegel, F.L., Lukas, T.J., and Watterson, D.M. (1986). Trimethyllysine and protein function. Effect of methylation and mutagenesis of lysine 115 of calmodulin on NAD kinase activation. J Biol Chem *261*, 1491-1494.

Saddic, L.A., West, L.E., Aslanian, A., Yates, J.R., 3rd, Rubin, S.M., Gozani, O., and Sage, J. (2010). Methylation of the retinoblastoma tumor suppressor by SMYD2. J Biol Chem *285*, 37733-37740.

Salzer, C.L., Elias, Y., and Kumar, J.P. (2010). The retinal determination gene eyes absent is regulated by the EGF receptor pathway throughout development in Drosophila. Genetics *184*, 185-197.

Schubert, H.L., Blumenthal, R.M., and Cheng, X. (2003). Many paths to methyltransfer: a chronicle of convergence. Trends Biochem Sci *28*, 329-335.

Shago, M., and Giguere, V. (1996). Isolation of a novel retinoic acid-responsive gene by selection of genomic fragments derived from CpG-island-enriched DNA. Mol Cell Biol *16*, 4337-4348.

Sheffield, P., Garrard, S., and Derewenda, Z. (1999). Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors. Protein Expr Purif *15*, 34-39.

Shi, X., Kachirskaia, I., Yamaguchi, H., West, L.E., Wen, H., Wang, E.W., Dutta, S., Appella, E., and Gozani, O. (2007). Modulation of p53 function by SET8-mediated methylation at lysine 382. Mol Cell *27*, 636-646.

Sigrist, C.J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A., and Bucher, P. (2002). PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform *3*, 265-274.

Sigrist, C.J., de Castro, E., Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2013). New and continuing developments at PROSITE. Nucleic Acids Res *41*, D344-347.

Sigrist, C.J., De Castro, E., Langendijk-Genevaux, P.S., Le Saux, V., Bairoch, A., and Hulo, N. (2005). ProRule: a new database containing functional and structural information on PROSITE profiles. Bioinformatics *21*, 4060-4066.

Sims, R.J., 3rd, Weihe, E.K., Zhu, L., O'Malley, S., Harriss, J.V., and Gottlieb, P.D. (2002). m-Bop, a repressor protein essential for cardiogenesis, interacts with skNAC, a heart- and muscle-specific transcription factor. J Biol Chem *277*, 26524-26529.

Sirinupong, N., Brunzelle, J., Doko, E., and Yang, Z. (2011). Structural insights into the autoinhibition and posttranslational activation of histone methyltransferase SmyD3. J Mol Biol *406*, 149-159.

Sirinupong, N., Brunzelle, J., Ye, J., Pirzada, A., Nico, L., and Yang, Z. (2010). Crystal structure of cardiac-specific histone methyltransferase SmyD1 reveals unusual active site architecture. J Biol Chem *285*, 40635-40644.

Slabinski, L., Jaroszewski, L., Rodrigues, A.P., Rychlewski, L., Wilson, I.A., Lesley, S.A., and Godzik, A. (2007a). The challenge of protein structure determination--lessons from structural genomics. Protein Sci *16*, 2472-2482.

Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I.A., Lesley, S.A., and Godzik, A. (2007b). XtalPred: a web server for prediction of protein crystallizability. Bioinformatics *23*, 3403-3405.

Smeenk, G., Wiegant, W.W., Vrolijk, H., Solari, A.P., Pastink, A., and van Attikum, H. (2010). The NuRD chromatin-remodeling complex regulates signaling and repair of DNA damage. J Cell Biol *190*, 741-749.

Sopko, R., and Andrews, B.J. (2008). Linking the kinome and phosphorylome--a comprehensive review of approaches to find kinase targets. Mol Biosyst *4*, 920-933.

Sprung, R., Chen, Y., Zhang, K., Cheng, D., Zhang, T., Peng, J., and Zhao, Y. (2008). Identification and validation of eukaryotic aspartate and glutamate methylation in proteins. J Proteome Res *7*, 1001-1006.

Stassen, M.J., Bailey, D., Nelson, S., Chinwalla, V., and Harte, P.J. (1995). The Drosophila trithorax proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. Mech Dev *52*, 209-223.

Stender, J.D., Pascual, G., Liu, W., Kaikkonen, M.U., Do, K., Spann, N.J., Boutros, M., Perrimon, N., Rosenfeld, M.G., and Glass, C.K. (2012). Control of proinflammatory gene programs by regulated trimethylation and demethylation of histone H4K20. Mol Cell *48*, 28-38.

Stock, A., Clarke, S., Clarke, C., and Stock, J. (1987). N-terminal methylation of proteins: structure, function and specificity. FEBS Lett *220*, 8-14.

Subramanian, K., Jia, D., Kapoor-Vazirani, P., Powell, D.R., Collins, R.E., Sharma, D., Peng, J., Cheng, X., and Vertino, P.M. (2008). Regulation of estrogen receptor alpha by the SET7 lysine methyltransferase. Mol Cell *30*, 336-347.

Takeda, S. (2006). Preparation of protein crystals for X-ray structural study. Methods Mol Med *129*, 291-303.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N.*, et al.* (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell *146*, 1016-1028.

Tan, X., Rotllant, J., Li, H., De Deyne, P., and Du, S.J. (2006). SmyD1, a histone methyltransferase, is required for myofibril organization and muscle contraction in zebrafish embryos. Proc Natl Acad Sci U S A *103*, 2713-2718.

Thoma, F., Koller, T., and Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J Cell Biol *83*, 403-427.

Thompson, E.C., and Travers, A.A. (2008). A Drosophila Smyd4 homologue is a muscle-specific transcriptional modulator involved in development. PLoS One *3*, e3008.

Tonna, S., El-Osta, A., Cooper, M.E., and Tikellis, C. (2010). Metabolic memory and diabetic nephropathy: potential role for epigenetic mechanisms. Nat Rev Nephrol *6*, 332-341.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J *13*, 3822-3831.

Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S.*, et al.* (2010). Towards a knowledge-based Human Protein Atlas. Nat Biotechnol *28*, 1248-1250.

Van Aller, G.S., Reynoird, N., Barbash, O., Huddleston, M., Liu, S., Zmoos, A.F., McDevitt, P., Sinnamon, R., Le, B., Mas, G.*, et al.* (2012). Smyd3 regulates cancer cell phenotypes and catalyzes histone H4 lysine 5 methylation. Epigenetics *7*, 340-343.

Varga-Weisz, P.D. (2014). Chromatin remodeling: a collaborative effort. Nat Struct Mol Biol *21*, 14-16.

Varga-Weisz, P.D., Wilm, M., Bonte, E., Dumas, K., Mann, M., and Becker, P.B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. Nature *388*, 598-602.

Wang, C., and Lazarides, E. (1984). Arsenite-induced changes in methylation of the 70,000 dalton heat shock proteins in chicken embryo fibroblasts. Biochem Biophys Res Commun *119*, 735-743.

Wang, C., Lazarides, E., O'Connor, C.M., and Clarke, S. (1982). Methylation of chicken fibroblast heat shock proteins at lysyl and arginyl residues. J Biol Chem *257*, 8356-8362.

Wang, C., Lin, J.M., and Lazarides, E. (1992). Methylations of 70,000-Da heat shock proteins in 3T3 cells: alterations by arsenite treatment, by different stages of growth and by virus transformation. Arch Biochem Biophys *297*, 169-175.

Wang, G.G., Allis, C.D., and Chi, P. (2007). Chromatin remodeling and cancer, Part II: ATP-dependent chromatin remodeling. Trends Mol Med *13*, 373-380.

Wang, H., Liu, Y., Tan, W., Zhang, Y., Zhao, N., Jiang, Y., Lin, C., Hao, B., Zhao, D., Qian, J.*, et al.* (2008a). Association of the variable number of tandem repeats polymorphism in the promoter region of the SMYD3 gene with risk of esophageal squamous cell carcinoma in relation to tobacco smoking. Cancer Sci *99*, 787-791.

Wang, J., Hevi, S., Kurash, J.K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G.*, et al.* (2009). The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. Nat Genet *41*, 125-129.

Wang, S.Z., Luo, X.G., Shen, J., Zou, J.N., Lu, Y.H., and Xi, T. (2008b). Knockdown of SMYD3 by RNA interference inhibits cervical carcinoma cell growth and invasion in vitro. BMB Rep *41*, 294-299.

Watson, J.D., and Crick, F.H. (1953). The structure of DNA. Cold Spring Harb Symp Quant Biol *18*, 123-131.

Webb, K.J., Lipson, R.S., Al-Hadid, Q., Whitelegge, J.P., and Clarke, S.G. (2010). Identification of protein N-terminal methyltransferases in yeast and humans. Biochemistry *49*, 5225-5235.

Weinert, B.T., Wagner, S.A., Horn, H., Henriksen, P., Liu, W.R., Olsen, J.V., Jensen, L.J., and Choudhary, C. (2011). Proteome-wide mapping of the Drosophila acetylome demonstrates a high degree of conservation of lysine acetylation. Sci Signal *4*, ra48.

West, L.E., Roy, S., Lachmi-Weiner, K., Hayashi, R., Shi, X., Appella, E., Kutateladze, T.G., and Gozani, O. (2010). The MBT repeats of L3MBTL1 link SET8-mediated p53 methylation at lysine 382 to target gene repression. J Biol Chem *285*, 37725-37732.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A.*, et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr *67*, 235-242.

Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2013). Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. FEBS J *280*, 5705-5736.

Woodcock, C.L., Frado, L.L., Hatch, C.L., and Ricciardiello, L. (1976). Fine structure of active ribosomal genes. Chromosoma *58*, 33-39.

Woodcock, C.L., and Ghosh, R.P. (2010). Chromatin higher-order structure and dynamics. Cold Spring Harb Perspect Biol *2*, a000596.

Wu, J., Cheung, T., Grande, C., Ferguson, A.D., Zhu, X., Theriault, K., Code, E., Birr, C., Keen, N., and Chen, H. (2011). Biochemical characterization of human SET and MYND domain-containing protein 2 methyltransferase. Biochemistry *50*, 6488-6497.

Xu, S., Wu, J., Sun, B., Zhong, C., and Ding, J. (2011a). Structural and biochemical studies of human lysine methyltransferase Smyd3 reveal the important functional roles of its post-SET and TPR domains and the regulation of its activity by DNA binding. Nucleic Acids Res *39*, 4438-4449.

Xu, S., Zhong, C., Zhang, T., and Ding, J. (2011b). Structure of human lysine methyltransferase Smyd2 reveals insights into the substrate divergence in Smyd proteins. J Mol Cell Biol *3*, 293-300.

Yang, L., He, J., Chen, L., and Wang, G. (2009). Hepatitis B virus X protein upregulates expression of SMYD3 and C-MYC in HepG2 cells. Med Oncol *26*, 445-451.

Young, C.C., Alvarez, J.D., and Bernlohr, R.W. (1990). Nutrient-dependent methylation of a membrane-associated protein of Escherichia coli. J Bacteriol *172*, 5147-5153.

Young, C.C., and Bernlohr, R.W. (1991). Elongation factor Tu is methylated in response to nutrient deprivation in Escherichia coli. J Bacteriol *173*, 3096-3100.

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

Zou, J.N., Wang, S.Z., Yang, J.S., Luo, X.G., Xie, J.H., and Xi, T. (2009). Knockdown of SMYD3 by RNA interference down-regulates c-Met expression and inhibits cells migration and invasion induced by HGF. Cancer Lett *280*, 78-85.

Appendix

1. DNA sequence encoding codon optimized full length *H. sapiens* SMYD5 for *E. coli* with BamHI restriction site in frame with ATG (GGATCCATG) and the XhoI restriction site placed after the stop codon TGA (TGACTCGAG).

GGATCCATGGCAGCGTCAATGTGCGATGTGTTCTCGTTCTGTGTTGGTGTGGC
AGGTCGTGCTCGTGTCTCGGTGGAAGTGCGTTTCGTTAGCAGTGCCAAAGGC
AAAGGTCTGTTTGCAACCCAGCTGATTCGCAAAGGTGAAACGATCTTTGTCGA
ACGTCCGCTGGTTGCGGCCCAATTCCTGTGGAACGCTCTGTATCGTTACCGCG
CGTGCGATCATTGTCTGCGCGCGCTGGAAAAAGCCGAAGAAAATGCACAGCG
TCTGACCGGTAAACCGGGCCAAGTTCTGCCGCACCCGGAACTGTGCACGGTC
CGTAAAGATCTGCATCAGAACTGCCCGCACTGTCAAGTTATGTATTGCAGCGC
CGAATGTCGTCTGGCAGCTACCGAACAGTACCATCAAGTCCTGTGTCCGGGTC
CGTCTCAGGATGACCCGCTGCATCCGCTGAATAAACTGCAAGAAGCTTGGCG
CAGCATTCACTATCCGCCGGAAACCGCGTCTATCATGCTGATGGCTCGTATGG
TGGCGACGGTTAAACAGGCCAAAGATAAAGACCGCTGGATTCGTCTGTTTTCT
CAATTCTGCAACAAAACCGCAAACGAAGAAGAAGAAATCGTTCATAAACTGCTG
GGTGATAAATTTAAAGGCCAGCTGGAACTGCTGCGTCGCCTGTTCACCGAAGC
CCTGTATGAAGAAGCAGTTAGTCAGTGGTTTACGCCGGACGGTTTTCGCTCCC
TGTTCGCGCTGGTCGGCACCAACGGTCAGGGTATTGGCACGAGCTCTCTGAG
CCAATGGGTGCACGCTTGTGATACCCTGGAACTGAAACCGCAGGACCGTGAA
CAACTGGATGCGTTCATTGACCAGCTGTACAAAGATATCGAAGCGGCCACGGG
TGAATTTCTGAATTGCGAAGGTAGTGGCCTGTTCGTGCTGCAGTCATGCTGTA
ACCATTCGTGTGTTCCGAATGCCGAAACCTCCTTTCCGGAAAACAATTTCCTGC
TGCACGTGACGGCACTGGAAGATATTAAACCGGGCGAAGAAATTTGCATCAGC
TATCTGGACTGCTGTCAGCGTGAACGCAGTCGTCATTCCCGCCACAAAATCCT
GCGTGAAAACTACCTGTTTGTCTGCTCATGTCCGAAATGTCTGGCAGAAGCTG
ATGAACCGAATGTGACCTCGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA
GGCGAACCGGAAGACGCAGAACTGGGCGACGAAATGACCGATGTTTGACTCG
AG

2. DNA sequence encoding codon optimized full length *D.melanogaster* SMYD5 for *E. coli* with BamHI restriction site in frame with ATG (GGATCCATG) and the XhoI restriction site placed after the stop codon TGA (TGACTCGAG).

GGATCCATGAACAACTTTGAAATCCGTGAACTGCCGGGTAAAGGCCGTGCTAT
GATTGCTACGAAAAACTTTGCTAAAGACGAAGTGATCTTTGAAGAAGAACCGTT
TGTTAGCCGTCAGTTCTCTTGGAACGTCGCATATGGCTACGCGGCCTGCGATC
ATTGTATGCGCCCGCTGGAAACCGTGCTGGAAAATGTTCGTCGCCTGGCTAGC
GATCCGAAAGTCGAAGTGCCGCTGCTGCAGCACGACCCGACCGCACAATGGG
TCGCTCAGTTTACGCAATGCCCGCGTTGTAAAGTGCGCTATTGCAGTGAAGAT
TGTCTGATGGAAGCACAGAAACGTTACCATCGCGTTGCGTGCATGGGCGCCTT
CCATTCTGATGACACCCACCCGATTAACGTCCTGAATGAAACGTGGAAGAAAA
TGCACTATCCGCCGGAAACCGGTTCAATTATGCTGATCGTGCGTCTGATGGCC
CTGTACCAGCAATCGACGAAAAAAGAAGAATTTCTGGAACAGCTGCAAAGCTT
CCAGTCTCTGATTGTTAACCGCGAACAAAAAATCTACCATAAAATGCTGGGCGA
AAATTTTGAACAGCAAATGGAACAGCTGTACCTGGCATTTTGTAACGCTTTCAC
CGGTGAAGAATTTAGCATCTTCAAAACCCCGGATGCGTTCAAAACGCTGATGG
CCATTCTGGGCACCAATTCTCAGGGTATCGCGACGAGTGTGCTGTCCCAGTG
GGTTGCCAAAGTCAGTGATCTGCCGCTGACCGACTCCGAAAAAGAACAGCTG
GATACGGTGATTGACGGCCTGTATGCAAAAGTTGGCGAATTTGCTGGTGAATT
CCTGAACAATGAAGGCAGCGGTCTGTACCTGCTGCAGAGTAAAATCAACCACT
CCTGCGTTCCGAATGCGTGTTCAACCTTTCCGTATTCGAACGATATTGTGGTTC
TGAAAGCACTGGCTCCGATCCAGCAAGGTGAAGAAATTTGCATCTCCTATCTG
GACGAATGTATGCTGGAACGTTCACGCCATTCGCGTCACAAAGTCCTGCGCGA
AAATTACGTGTTTATTTGCCAGTGTCCGAAATGCCGTGCGCAAGCCAGCGATC
CGGACGAAACGTCTGAAGACGACGATGATGACGACGAAATGGACGACTATGA
TGATGATGATGATATGAACTGACTCGAG

3. DNA sequence encoding codon optimized full length *G. gallus* SMYD5 for *E. coli* with BamHI restriction site in frame with ATG (GGATCCATG) and the XhoI restriction site placed after the stop codon TGA (TGACTCGAG).

GGATCCATGGCGGCGGCGGCGGGTGATGTTCGTGGTGCAGCCCTGGGTGCT
CGTCCGGGTCTGGGTGCGGCAGCGGCGGCGGCAGCGGCGGCGGAAGCTCG
CTTTATTAGCTCTGCGAAAGGCAAAGGTCTGTTCGCTACCCGCAGTATCCGTA
AAGGTGAAGCGGTCTTTGTGGAAAAACCGGTTGTTAGTTCCCAGTTCCTGTGG
AACGCCCTGTATAATTACCGCGCATGCGATCATTGTCTGCGTGCACTGGAAAC
CGCTGAAGAAAACGCACAGCGTCTGCTGGGCCGTTCATCGCTGGTGCTGCCG
CATCCGGAACAATGCTCAATTCGCAAAGATCTGCACCAGCAATGCCCGCGTTG
TCAGGTCACCTATTGCTCGGCAGAATGTCGTCAAGCAGCACTGGAACAGTACC
ACCAAGTTCTGTGTCTGGGTCCGAGCCGCGATGACCCGACGCATCCGCTGAA
CAAACTGCAGGAAGCATGGCGTAATATGCACTATCCGCCGGAAACCAGCTCTA
TCATGCTGATGGCCCGTATGGTTGCAACGGTCAAACAGGCCAAAGATAAAGAC
TGGTGGATTAAAGCATTTAGCCAATTCTGCTCTAAAACCGCCAATGAAGAAGAA
GAAATCGCACATAAACTGCTGGGTGATAAATTTAAAGGCCAGCTGGAACTGCT
GCGCCTGCTGTTCACCGAAGCTCTGTACGACGAACAACTGAGCCGCTGGTTTA
CGCCGGAAGGTTTTCGTTCTCTGTTCGCGCTGGTTGGCACCAACGGCCAGGG
TATTGGCACGAGTTCCCTGAGCCAATGGGTTCACGCGTGTGATGCCCTGGAC
CTGCCGATGCTGCAGCGTGAAGAACTGGATGCCTTCATCGACCAACTGTACAA
AGATATCGAAAAAGAAAGCGGTGAATTCCTGAATTGCGAAGGTTCGGGCCTGT
ACATGCTGCAGAGTTGCTGTAACCATTCCTGTATTCCGAATGCGGAAACCTCAT
TTCCGGATAACAATTTCCTGCTGTATCTGACGGCTCTGGAAGACATCGAAGCG
GGCGAAGAAATTTGCATCTCTTACCTGGATTGCTGTCAGCGTGAACGCTCACG
TCACTCGCGCAACAAAATCCTGCGTGAAAATTACCTGTTTACCTGCAGTTGTCC
GAAATGTCTGGCACAGGCTGATGACCCGGATGTTACGTCCGACGAAGAAGAA
GAAGCCGAAGGCGAAACCGACGACGCTGAACTGGAAGACGAAATGACC
GATGTGTGACTCGAG

4. MUSCLE alignment of the amino acid sequences of *H. sapiens* SMYD1-5 reveal differences in amino acid sequence between SMYD5 and the other SMYD family members. Asterisks (*) indicate residues that are exactly conserved throughout all the SMYD family proteins, whereas dots indicate conservation of residues with similar properties. A colon (:) represents strongly similar properties and a period (.) indicates conservation between residues of weakly similar properties.

```
SMYD1      ------------------------------------------------------------
SMYD2      ------------------------------------------------------------
SMYD3      ------------------------------------------------------------
SMYD4      MDLPVDEWKSYLLQKWASLPTSVQVTISTAETLRDIFLHSSSLLQPEDELFLKRLSKGYL
SMYD5      ------------------------------------------------------------

SMYD1      -------------------------------------------------MTIGRMENVE
SMYD2      -------------------------------------------------MRAEGLGGLE
SMYD3      -----------------------------------------------------MEPLKVE
SMYD4      VGKDSDAPLFYREEGNKKFQEKDYTGAAVLYSKGVSHSRPNTEDMSLCHANRSAALFHLG
SMYD5      ----------------------------------MAASMCDVFSFCVGVAGRARVSVE
                                                                    :

SMYD1      -VFTA--------------------------------------------------------
SMYD2      -RFCS--------------------------------------------------------
SMYD3      -KFAT--------------------------------------------------------
SMYD4      -QYETCLKDINRAQTHGYPERLQPKIMLRKAECLVALGRLQEASQTISDLERNFTATPAL
SMYD5      VRFVS---------------------------------------------------------
             : :

SMYD1      ------------------------------------------------------------
SMYD2      ------------------------------------------------------------
SMYD3      ------------------------------------------------------------
SMYD4      ADVLPQTLQRNLHRLKMKMQEKDSLTESFPAALAKTLEDAALREENEQLSNASSSIGLCV
SMYD5      ------------------------------------------------------------

SMYD1      -EGKGRGLKATKEFWAADIIFAERA-----------YSAVVFDSLV--------------
SMYD2      -PGKGRGLRALQPFQVGDLLFSCPA-----------YAYVLTVNER--------------
SMYD3      -ANRGNGLRAVTPLRPGELLFRSDP-----------LAYTVCKGSR--------------
SMYD4      DPLKGRCLVATKDILPGELLVQEDA-----------FVSVLNPGELPPPHHGLDSKWDTR
SMYD5      -SAKGKGLFATQLIRKGETIFVERPLVAAQFLWNALYRYRACDHCLRALEKAEENAQRLT
              .*. * *   :  .: :.   .

SMYD1      ----NFVCHT--CFKRQEKLHRCGQCKFAHYCDRTCQKDAW-LNHKNECS---------A
SMYD2      ----GNHCEY--CFTRKEGLSKCGRCKQAFYCNVECQKEDW-PMHKLECS---------P
SMYD3      ----GVVCDR--CLLGKEKLMRCSQCRVAKYCSAKCQKKAW-PDHKRECK---------C
SMYD4      VTNGDLYCHR--CLKHTLATVPCDGCSYAKYCSQECLQQAWELYHRTECPLGGLLLTLGV
SMYD5      GKPGQVLPHPELCTVRKDLHQNCPHCQV-MYCSAECRLAATEQYHQVLCPGPSQDDPLHP
                       *         *  *    **.  *        *.  *

SMYD1      IKRYGK---------------------------VPNEN--IRLAARIMWRVEREG----
SMYD2      MVVFGENW-------------------------NPSET--VRLTARILAKQKIHP----
SMYD3      LKSCKPR--------------------------YPPDS--VRLLGRVVFKLM-DG----
SMYD4      FCHIALRLTLLVGFEDVRKIITKLCDKISNKDICLPESNNQVKTLNYGLGESEKNGNIVE
SMYD5      LNKLQEAW---------------------RSIHYPPETASIMLMARMVATVK-------
```

```
          :                                  * ..  :      :

SMYD1     --------TGLTEGCLVSVDDLQNHVEHFGEEEQKDL--------RV-DVDTFLQYWPPQ
SMYD2     --------ERTPSEKLLAVKEFESHLDKLDNEKKDLI--------QS-DIAALHHFYSKH
SMYD3     --------APSESEKLYSFYDLESNINKLTEDKKEGL--------RQ-LVMTFQHFMREE
SMYD4     TPIPGCDINGKYENNYNAVFNLLPHTENHSPEHKFLCALCVSALCRQLEAASLQAIPTER
SMYD5     --------QAKDKDRWIRLFS--QFCNKTANEEEEIVHKLLGDKFKG-QLELLRRLFTEA
                       .     . .     ::   : :           .      :

SMYD1     SQ-----------QFSMQY-ISHIFGV-----------INCN-------------GFT--
SMYD2     L------------GFPDNDSLVVLFAQ-----------VNCN-------------GFT--
SMYD3     IQDAS--------QLPPAFDLFEAFAK-----------VICN-------------SFT--
SMYD4     IVNSSQLKAAVTPELCPDVTIWGVAML------RHMLQLQCNAQAMTTIQHTGPKGSI--
SMYD5     LYEEAVSQWFTPDGFRSLFALVGTNGQGIGTSSLSQWVHACDTLELKPQDREQLDAFIDQ
                        :     :                  *:            .

SMYD1     ------LSDQRGLQAVGVGIFPNLGLVNHDCWPNCTVIFNNGNHEAVKSMFHTQMRIELR
SMYD2     ------IEDEE-LSHLGSAIFPDVALMNHSCCPNVIVTY-KGTL------------AEVR
SMYD3     ------ICNAE-MQEVGVGLYPSISLLNHSCDPNCSIVF-NGPH------------LLLR
SMYD4     ------VTDSR-QVRLATGIFPVISLLNHSCSPNTSVSF-ISTV------------ATIR
SMYD5     LYKDIEAATGEFLNCEGSGLFVLQSCCNHSCVPNAETSFPENNF-----------LLHVT
                         . .::   .  **.* **    :  .               :

SMYD1     ALGKISEGEELTVSY--IDFLNVSEERKRQ---LKKQYYFDCTCEHCQKKLK--------
SMYD2     AVQEIKPGEEVFTSY--IDLLYPTEDRNDR---LRDSYFFTCECQECTTKDK--------
SMYD3     AVRDIEVGEELTICY--LDMLMTSEERRKQ---LRDQYCFECDCFRCQTQDK--------
SMYD4     ASQRIRKGQEILHCYGPHKSRMGVAERQQK---LRSQYFFDCACPACQTEAHRMAAGPRW
SMYD5     ALEDIKPGEEICISY--LDCCQRERSRHSRHKILRENYLFVCSCPKCLAEAD--------
          *   *  *:*:  .*   .      .*. .   *...* * * *  *  :

SMYD1     ------------D-DLFLGVKDNPKPSQEVVKE-MIQFSKDTLEKID---KARSEGLYHE
SMYD2     ------------DKAKVEIRKLSDPPKAEAIRD-MVRYARNVIEEFR---RAKHYKSPSE
SMYD3     ------------DADMLTG-------DEQVWKE-----VQESLKKIE---ELKAHWKWEQ
SMYD4     EAFCCNSCGAPMQGDDVLRCGSRSCAESAVSRDHLVSRLQDLQQQVRVAQKLLRDGELER
SMYD5     -------------------------EPNVTSE---------EEEEE---EEEEEGEPED
                                .  .  :          ::

SMYD1     VVKLCRECLEKQEPVFADTNIYMLRMLSIVSEVLSYLQAFEEASFYARRMVDGYMKLYHP
SMYD2     LLEICELSQEKMSSVFEDSNVYMLHMMYQAMGVCLYMQDWEGALQYGQKIIKPYSKHYPL
SMYD3     VLAMCQAIISSNSERLPDINIYQLKVLDCAMDACINLGLLEEALFYGTRTMEPYRIFFPG
SMYD4     AVQRLSGCQRDAESFLWAEHAVVGEIADGLARACAALGDWQKSATHLQRSLYVVEVRHGP
SMYD5     ------------------------------------------------------------


SMYD1     NNAQLGMAVMRAGLTNWHAGNIEVGHGMICKAYAILLVTHGPSHPITKDLEAMRVQTEME
SMYD2     YSLNVASMWLKLGRLYMGLEHKAAGEKALKKAIAIMEVAHGKDHPYISEIKQEIESH---
SMYD3     SHPVRGVQVMKVGKLQLHQGMFPQAMKNLRLAFDIMRVTHGREHSLIEDLILLLEECDAN
SMYD4     SSVEMGHELFKLAQIFFNGFAVPEALSTIQKAEEVLSLHCGPWDDEIQELQKMK-SCLLD
SMYD5     --AELGDEMTDV--------------------------------------------------
               .

SMYD1     LRMFRQNEFMYYKMREAALNNQPMQVMAEPSNEPSPALFHKKQ
SMYD2     ----------------------------------------
SMYD3     IRAS------------------------------------
SMYD4     LPPTPVGPAL------------------------------
SMYD5     ----------------------------------------
```

5. ClustalW2 alignment of the amino acid sequences of SMYD5 homologues reveals that SMYD5 is highly conserved amongst higher eukaryotes. Asterisks (*) indicate residues that are exactly conserved throughout all the SMYD family proteins, whereas dots indicate conservation of residues with similar properties. A colon (:) represents strongly similar properties and a period (.) indicates conservation between residues of weakly similar properties.

```
H. sapiens          --MAASMCDVFSFCVGVAG------RARVSVEVRFVSSAKGKGLFATQLI 42
P. troglodytes      --MAASMCDVFSFCVGVAG------RARVSVEVRFVSSAKGKGLFATQLI 42
C. procellus        --MAASMCEVFSFCVGVAG------SAGIAVEVRFVSSAKGKGLFATQLI 42
M. musculus         --MAASMCDVFSFCVGVAD------RARGSVEVRYVDSIKGKGLFATQLI 42
G. gallus           --MAAAAGDVRGAALGARPGLGAAAAAAAAAAEARFISSAKGKGLFATRSI 48
X. laevis           --MAASMCDVFAFCAEQE-------PARRTVEIRFVSSGKGKGLFAIRTI 41
D. rerio            --MAAPVDDMFSRCVDSAK-------ASNCVDVRFINNVKGKGLFAKKPF 41
D. melanogaster     --------------------------MNNFEIRELPG-KGRAMIATKNF 22
S. cerevisiae       MTLTIKIGTLNDSDQSAVHN-----GTENGSDFRKITPTEEEICDDVVLL 45
                                 : * :   : .       :

H. sapiens          RKGE---------------TIFVERPLVAAQFLWNALYRYRACDHCLRAL 77
P. troglodytes      RKGE---------------TIFVEQPLVAAQFLWNALYRYRACDHCLRAL 77
C. procellus        RKGE---------------TIFVERPLVAAQFLWNALYRYRACDHCLRAL 77
M. musculus         RKGE---------------TIFIERPLVAAQFLWNALYQYRACDHCLRAL 77
G. gallus           RKGE---------------AVFVEKPVVSSQFLWNALYNYRACDHCLRAL 83
X. laevis           RKGE---------------TIFQEKPLVSSQFQWNALYRYRACDHCLRSL 76
D. rerio            KKGD---------------TIFIERPLVSSQFLWNALYKYRACEYCLRAL 76
D. melanogaster     AKDE---------------VIFEEEPFVSRQFSWNVAYGYAACDHCMRPL 57
S. cerevisiae       WKEEPGTEDATIQHLYDRITERNQSWKLSASRFRKILNEHHLYDTDLETV 95
                     * :               .   :   :: .   :    :   :  :..:

H. sapiens         EKAEENAQR---LTGKPGQVLP--HPELCTVRKDLHQNCPHCQVMYCSAE 122
P. troglodytes      EKAEENAQR---LTGKPGQVLP--HPELCTVRKDLHQNCPHCQVMYCSAE 122
C. procellus        EKAEENAQR---LTGKPGQVLP--HPELCTVRKDLHQNCPHCQVMYCSAE 122
M. musculus         EKAEENAQR---LTGKPSQILP--HPELCSVRKDLHQNCPHCQVMYCSAE 122
G. gallus           ETAEENAQR---LLGRSSLVLP--HPEQCSIRKDLHQQCPRCQVTYCSAE 128
X. laevis           ETAEENAQR---LSGNAHVLLP--YPELCTVRNGLHQQCPRCQVTYCSAE 121
D. rerio            ETAEENARR---LSGLPALILP--HPELCKVRPDRHQACPQCQVMYCSSE 121
D. melanogaster     ETVLENVRR---LASDPKVEVPLLQHDPTAQWVAQFTQCPRCKVRYCSED 104
S. cerevisiae       SLYKDKIHFPKALDSDAKVEVKFIDDEHGRGLFAKRDFSKGQIILKENKP 145
                    .   :: :    * . .   :     :           .    :   .

H. sapiens          CRLAATEQYHQVLCPGPSQDDPLHPLNKLQEAWRSIHYPPETASIMLMAR 172
P. troglodytes      CRLAATEQYHQVLCPGPSQDDPLHPLNKLQEAWRSIHYPPETASIMLMAR 172
C. procellus        CRLAAAEQYHRVLCPGPSQDDPLHPLNKLQEAWRSIHYPPETASIMLMAR 172
M. musculus         CRLAAAEQYHQILCPGPS-HDPRHPLNKLQEAWRSVHYPPETASIMLMAR 171
G. gallus           CRQAALEQYHQVLCLGPSRDDPTHPLNKLQEAWRNMHYPPETSSIMLMAR 178
X. laevis           CLKAAAEQYHQILCLETSRDNPAHPLNKLEEAWRNMHYPPETASIMLMAR 171
D. rerio            CRQAAMDQYHKILCLGPSNDDPDHPVNKLQDAWRSVHFPPETSSVMILAK 171
D. melanogaster     CLMEAQKRYHRVACMGAFHSDDTHPINVLNETWKKMHYPPETGSIMLIVR 154
S. cerevisiae       IVYIPPLDKLFFISNGKACARCXKALYDLTQHKIMVHYLDCEVCKAIWCS 195
                        .      . .         :.:  * :    :*:     .  :

H. sapiens          MVATVKQAKDKDRWIR----------------LFSQFCNKTANEEEEIVH 206
P. troglodytes      MVATVKQAKDKDRWIR----------------LFSQFCNKTANEEEEIVH 206
C. procellus        MVATVKQAKDKDRWVR----------------LFSQFCNKTANEEEEIVH 206
M. musculus         MVATVKQAKDKDHWVR----------------LFNHFCSRTANQEQAIVH 205
G. gallus           MVATVKQAKDKDWWIK----------------AFSQFCSKTANEEEEIAH 212
X. laevis           MVGTIKQAQDKDWWLH----------------LFSQFCNKTANEEEEIVH 205
D. rerio            MVATIKQTQDKERWQR----------------LFTNFCSRTANEEEEIVH 205
```

```
D. melanogaster   LMALYQQSTKKEEFLE----------------QLQSFQSLIVNREQKIYH 188
S. cerevisiae     EKCKKAHASLHELLYHSWRSNRIDILHAGNWKRFVNYCEKYCFTAAFSVG 245
                        ::  ::   .                :  : .

H. sapiens        KLLGDKFKGQLELLRRLFT--EALYEEAVSQWFTPDGFRSLFALVGTNGQ 254
P. troglodytes    KLLGDKFKGQLELLRRLFT--EALYEEAVSQWFTPDGFRSLFALVGTNGQ 254
C. procellus      KLLGDKFKGELELLRRLFT--EALYEEALSQWFTPDGFRSLFALVGTNGQ 254
M. musculus       KLLKGKFKDQLELLLGLFK--EALYEEALSLWFTPEGFRSLFALVGTNGQ 253
G. gallus         KLLGDKFKGQLELLRLLFT--EALYDEQLSRWFTPEGFRSLFALVGTNGQ 260
X. laevis         KLLGEKFKGQLDQLRRLFV--DALYEERMSRWFTPEGFRSLFALVGTNGQ 253
D. rerio          KLLGEKFQGQLGLLRNLFT--TALYEDRLSQWFTPEGFRSLFSLVGTNGQ 253
D. melanogaster   KMLGENFEQQMEQLYLAFC--NAFTGEEFSIFKTPDAFKTLMAILGTNSQ 236
S. cerevisiae     LIYGSMLLDTTGEVKEQWQKLASISQRERIKLRDASGIGSTFSLLNGTTV 295
                   :    :      :   :    ::          ...: : ::::. .

H. sapiens        GIGTSSLSQWVHACDTLELKPQDREQLDAFIDQLYKDIEAAT--GEFLNC 302
P. troglodytes    GIGTSSLSQWVHACDTLELKPQDREQLDTFIDQLYKDIEAAT--GEFLNC 302
C. procellus      GIGTSSLSQWVHACDALELKPQDREQLDAFIDQLYKDIEAAT--GEFLNC 302
M. musculus       GIGTSSLSQWVHACDALELTPQDREQLDTFIDQLYKDIEAAT--GEFLNC 301
G. gallus         GIGTSSLSQWVHACDALDLPMLQREELDAFIDQLYKDIEKES--GEFLNC 308
X. laevis         GIGTSSLSQWVHACDALELPPRDREKLDALIDQLYKDIEKVT--GEFLNC 301
D. rerio          GIGTSSLSQWVHACDALELPRQQREQLDAFIDQLYKDIDKET--GDFLNC 301
D. melanogaster   GIATSVLSQWVAKVSDLPLTDSEKEQLDTVIDGLYAKVGEFA--GEFLNN 284
S. cerevisiae     HTEEESDNGTKKGVEKNIDDETVWEKCYELFCGAFPKASEEIDFEKFLTM 345
                      .  .      .         *:   .:   : .        .**.

H. sapiens        EGSG--------LFVLQSCCNHSCVPNAETSFPENNFLLHVTALEDIKPG 344
P. troglodytes    EGSG--------LFVLQSCCNHSCVPNAETSFPENNFLLHVTALEDIKPG 344
C. procellus      EGSG--------LFVLQSCCNHSCVPNAETSFPENNFLLHVTALEDISPG 344
M. musculus       EGSG--------LFVLQSCCNHSCVPNAETSFPENNFVLHVTALEDIKPG 343
G. gallus         EGSG--------LYMLQSCCNHSCIPNAETSFPDNNFLLYLTALEDIEAG 350
X. laevis         EGSG--------LYLLQSCCNHSCVPNAEASFPDNNFILHLTALEDIQPG 343
D. rerio          EGSG--------LFLLQSSCNHSCVPNAEASFPENNFLLHLTALGDIGPG 343
D. melanogaster   EGSG--------LYLLQSKINHSCVPNACSTFPYSNDIVVLKALAPIQQG 326
S. cerevisiae     IGTFNINQYNGQVYHWISFINHDCEPNAYIEQVEEHEELRLHARKPIKKG 395
                   *:         ::   *  **.* ***      .:  : : *   *  *

H. sapiens        EEICISYLDCCQRERSRHSRHKILRENYLFVCSCPKCLAEADEPNVTSEE 394
P. troglodytes    EEICISYLDCCQRERSRHSRHKILRENYLFVCSCPKCLAEADEPNVTSEE 394
C. procellus      EEICISYLDCCQRERSRHSRHKILRENYLFVCSCPKCLAEADEPNVTSEE 394
M. musculus       EEICISYLDCCQRERSRHSRHKILRENYLFNCSCPKCLAEADDPNVTSEE 393
G. gallus         EEICISYLDCCQRERSRHSRNKILRENYLFTCSCPKCLAQADDPDVTSDE 400
X. laevis         EEICISYLDCCQRDRSRHSRQKILRENYLFMCSCPKCLAQADEPDITSEE 393
D. rerio          EEICISYLDCCQRDRSRHSRHKILRENYLFICSCQKCLSQMDDADMTSED 393
D. melanogaster   EEICISYLDECMLERSRHSRHKVLRENYVFICQCPKCRAQASDPDETSED 376
S. cerevisiae     EQIRITYVNPLHGVRLRR---RELRVNWGFLCQCDRCQNELSTFERVPNL 442
                  *:* *:*::     * *:   : ** *: * *.* :*  : .  : ..:

H. sapiens        EEEEEEEE----EGEPEDAELGDEMTDV---------------------- 418
P. troglodytes    EEEEEE-------GEPEDAELGDEMTDV---------------------- 415
C. procellus      EEDEEE------EGEPEDAELGDEMTDV---------------------- 416
M. musculus       EEEEDEE-----EGEPEDAELGDEMTDV---------------------- 416
G. gallus         EEEAEG--------ETDDAELEDEMTDV---------------------- 420
X. laevis         EEEEEEEDDAELEGEPEDAELEDEMTDV---------------------- 421
D. rerio          EEEVEG------EGETEGEDMEDEMTDV---------------------- 415
D. melanogaster   DDDDDE---------MDDYDDDDDMN----------------------- 393
S. cerevisiae     EKKNADANLGVEKIDSNDSSEDGSKKSTGNRKSSMREAQPDLKEILKNGK 492
                  :..             :. .  .. .


H. sapiens        --------------------------------
P. troglodytes    --------------------------------
C. procellus      --------------------------------
```

```
M. musculus       --------------------------------
G. gallus         --------------------------------
X. laevis         --------------------------------
D. rerio          --------------------------------
D. melanogaster   --------------------------------
S. cerevisiae     EFELDIPETVDTQGNVRKTYSQIRFKRFGRSG 524
```

Journal Permissions

Figure 1A

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Figure 1B

Tonna, S., El-Osta, A., Cooper, M.E., and Tikellis, C. (2010). Metabolic memory and diabetic nephropathy: potential role for epigenetic mechanisms. Nat Rev Nephrol *6*, 332-341.

Figure 2

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu Rev Biochem *78*, 273-304.

Figure 3

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

Rights and Permissions

Figure 4A

Min, J., Feng, Q., Li, Z., Zhang, Y., and Xu, R.M. (2003). Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. Cell *112*, 711-723.

Are you the author of this Elsevier article?	No
Will you be translating?	No
Title of your thesis/dissertation	Elucidating the biochemical and structural features required for SMYD5 mediated methylation of histone H4 and other potential substrates
Expected completion date	Jun 2014
Estimated size (number of pages)	100
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD

Figure 4B

Couture, J.F., Collazo, E., Brunzelle, J.S., and Trievel, R.C. (2005). Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. Genes Dev *19*, 1455-1465.

"Permissions

1. Articles not designated as Open Access are distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see **Terms** for complete details). After six months, they are available under a Creative Commons License **(Attribution-NonCommercial 4.0 International License)**.

Authors of these non-Open Access articles retain copyright in the articles but grant Cold Spring Harbor Laboratory Press exclusive publishing rights for six months following full-issue publication. This grant of rights includes the rights to publish, reproduce, distribute, display, and store the article in all formats; to translate the article into other languages; to create adaptations, summaries, extracts, or derivations of the article; and to license others to do any or all of the above.

2. Articles that carry the Open Access designation are immediately distributed under one of two Creative Commons Licenses: (i) **Creative Commons Attribution-NonCommercial 4.0 International License (CC-BY-NC)** or (ii) **Creative Commons Attribution 4.0 International License (CC-BY)**. The CC-BY license permits commercial use, including reproduction, adaptation, and distribution of the article provided the original author and source are credited. *Please note specific licensing information within article of interest.*

3. To request permission to reproduce/adapt artwork from *Genes & Development* elsewhere (e.g., in other publications) during the first six months after full-issue publication, **click here**.

4. Please contact ***Copyright Clearance Center*** to request permission to photocopy articles or for use in a coursepack during the first six months after full-issue publication.

5. To request permission for any other use, including for commercial purposes, **click here**."

Figure 6

Benvenuti, M., and Mangani, S. (2007). Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. Nat Protoc *2*, 1633-1651.

Vanessa Mongeon

SUMMARY OF QUALIFICATIONS

- M.Sc. Biochemistry - predicted thesis defense summer 2014
- 4+ years of laboratory research experience, specialized in protein characterization
  - 2 years M.Sc.
  - 1 year B.Sc. Honours Research Project
  - 1.5 years CO-OP internships
- Communication of research findings through publications, posters and Powerpoint presentations
- Bilingual English and French

EDUCATION

M.Sc. Biochemistry *May 2012 - Summer 2014*
University of Ottawa, Ontario
B.Sc. Honours Biopharmaceutical Sciences, Specialized in Genomic *Fall 2007 - June 2012*
University of Ottawa, Ontario

Relevant Employment

Research Assistant *May - August 2011; September - December 2010*
Agriculture and Agri-Food Canada, Saint Hyacinthe, Québec

- Production of protein allergen reference material
- Assisted in data assembly for an FAO (World Health Organization) database on protein digestibility

Research Assistant *January - August 2010; January-April 2011*
Agriculture and Agri-Food Canada, Ottawa, Ontario

- Developed a selection system for low cadmium accumulating soybean varieties using hydroponic culture

Resource Conservation Monitoring Assistant *May - August 2009*
Parks Canada - St Lawrence Islands National Park, Mallorytown Landing, Ontario

- Environmental health monitoring at St Lawrence Islands National Park

Field Technician Assistant *May - August 2008*
Raisin Region Conservation Authority, Cornwall, Ontario

- Environmental health monitoring in the Raisin River Watershed

Research Assistant *July - August 2007*
University of Ottawa, Ottawa, Ontario

- Assisted in research on methylmercury in sediment from the St Lawrence River

PUBLICATIONS

Jacob Y*, Bergamin E*, Donoghue MTA, **Mongeon V**, LeBlanc C, Voigt P, Underwood CJ, Brunzelle J, Micheals SD, Reinberg D, Couture JF, Martienssen RA. (2014) Selective methylation of histone H3 variant H3.1 regulates heterochromatin replication. *Science;* **343** (6176): 1249-1253

Lanouette S, **Mongeon V**, Figeys D, Couture JF. (2014) The Functional diversity of protein lysine methylation. *Molecular Systems Biology;* **10:** 724

Azarnia S, Boye JI, **Mongeon V**, Sabik H. (2013) Detection of ovalbumin in egg white, whole egg and incurred past using LC-ESI-MS/MS and ELISA. *Food Research International;* **52**: 526-534

Alvarez PA, **Mongeon VJ**, Boye JI. (2013) Characterization of a gluten reference material: Wheat-contaminated oats. *Journal of Cereal Science*; **57**: 418-423

*These authors contributed equally to this work

CONFERENCE PRESENTATIONS - INTERNATIONAL MEETINGS

Mongeon V*, Bergamin E*, Brunzelle J, Couture J-F (November 2013) Structural insights into cis-regulatory mechanisms mediating histone H3 Lys-27 mono-methylation. Poster presented at the Chromatin: Structure and Function Conference hosted by Abcam, Grand Cayman, Cayman Islands (MSc work)

Mongeon V*, Tremblay V*, Couture J-F (June 2012) Proteomic and Functional Studies of SMYD5. Poster presented at the 2nd China-Canada Systems Biology and 19th Methods in Protein Structure Analysis Joint Conference, Ottawa, Ontario, Canada (MSc work)

Xing L*, **Mongeon V*** (2012) SILAC heavy labeled Neuro 2A cells: applications in standardization of mouse SCN tissue for circadian cycle and neurodegenerative studies. Poster presented at the 2nd China-Canada Systems Biology and 19th Methods in Protein Structure Analysis Joint Conference, Ottawa, Ontario, Canada (as part of the Second Annual Summer School in the Systems Biology of Neurodegenerative Disease)

*These authors contributed equally to this work

CONFERENCE PRESENTATIONS - NATIONAL MEETINGS

Mongeon V*, Bergamin E*, Brunzelle J, Couture J-F (November 2013) Structural insights into cis-regulatory mechanisms mediating histone H3 Lys-27 mono-methylation. Poster presented at GRASP, Montreal, Quebec, Canada (MSc work)

Mongeon V, Tremblay V, Couture J-F (May 2013) Structural and Functional Insights into SMYD5. Poster presented at PROTEO, Laval, Quebec, Canada (MSc work)

Mongeon V*, Tremblay V*, Couture J-F (November 2012) Structural and Functional Study in the Role of SMYD5 in Centrosome Dynamics. Poster presented at GRASP, Montreal, Quebec, Canada (MSc work)

*These authors contributed equally to this work

ACADEMIC AWARDS AND ACHIEVEMENTS

May 2013 - First prize poster presentation award at the 13th annual PROTEO symposium, $200
June 2012 - CIHR Training Program in Neurodegenerative Lipidomics (CTPNL) Conference Travel Award, $460
May 2012 – April 2013 Admission Scholarship, University of Ottawa (M.Sc.), $ 5 000/session
July-Aug 2007 Undergraduate Research Scholarship, University of Ottawa, $ 3 750
Aug 2007 - April 2008 Admission Scholarship, University of Ottawa (B.Sc.), $ 3 000
Sept 2007 - Aug 2008 Irving Tissue "Royale" Scholarship, $ 2 000
Sept 2007 - Aug 2008 Fellowship to study in French, $ 1 500

COMMUNITY INVOLVEMENT

Faculty of Medicine - BMI Graduate Student Recruitment Day Volunteer *February 21, 2014*
University of Ottawa, Ottawa, Ontario

OISB Representative - BMIGSA Student Council *2012-2013*
University of Ottawa, Ottawa, Ontario

University Student Liaison *Winter 2008 and 2009*
École secondaire catholique Nouvelle-Alliance, Barrie, Ontario

President of the prom committee *2006-2007*
École secondaire catholique Nouvelle-Alliance, Barrie, Ontario