uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Mohamed M.S. Abu-Farha
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Biochemistry)
GRADE / DEGREE

Department of Biochemistry, Microbiology and Immunology
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Identification of SMYD2 Interactions and Substrates by Immunoprecipitation Coupled to Mass Spectrometry

TITRE DE LA THÈSE / TITLE OF THESIS

Daniel Figeys
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

Alexandre Blais

Marjorie Brand

François Robert
IRCM – Université de Montréal

Laura Trinkle-Mulcahy

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

IDENTIFICATION OF SMYD2 INTERACTIONS AND SUBSTRATES BY IMMUNOPRECIPITATION COUPLED TO MASS SPECTROMETRY

Mohamed Abu-Farha

Thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Biochemistry in the Graduate School of The University of Ottawa

Thesis Supervisor: Professor Daniel Figeys

Faculty of Medicine

Department of Biochemistry, Microbiology and Immunology

Your file Votre référence
ISBN: 978-0-494-79723-5
Our file Notre référence
ISBN: 978-0-494-79723-5

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ChIP-HTMS:	Chromatin ImmunoPrecipitation and High throughput MS
DAPI:	4, 6-DiAmidino-2-phenylIndole
DHB:	2, 5-DihydroxyBenzoic acid
Dot1:	Disruptor of Telomeric Silencing-1
ESI:	Electrospray Ionization
FT:	Fourier Transform
H3:	Histone 3
HA:	Hemagglutinin
HSP:	Heat Shock Protein
IMAC:	Immobilized Metal Affinity Chromatography
IP:	ImmunoPrecipitation
IP-HTMS:	ImmunoPrecipitation and High Throughput MS
K36:	Lysine 36
K4:	Lysine 4
LTQ:	Linear Ion Trap
m/z:	Mass to Charge Ratio
MALDI:	Matrix assisted laser Desorption Ionization
MS:	Mass Spectrometry
MTase:	Methyltransferase
MYND:	Myeloid, Nervy and DEAF-1
PPI:	Protein-Protein Interactions
PTM:	Post Translational Modification

SAM:	S-Adenosyl Methionine
SET:	Su (var) 3-9, Enhancer of zeste, and Trithorax
SILAC:	Stable Isotope Labeling with Amino Acids in Cell Culture
SMN:	Survival of Motor Neuron
SMYD2:	SET and MYND containing protein 2
TACC2:	Transforming, Acidic Coiled-coil Containing protein 2
TAP:	Tandem Affinity Purification
TOF:	Time-of-Flight
Y2H:	Yeast Two Hybrid

ACKNOWLEDGMENTS

First, I would like to start by thanking everyone in the Figeys lab for their help and support in making this thesis deserving a special mention. It is a pleasure to convey my gratitude to them all in my humble acknowledgment. I would also like to record my gratitude to Dr. Daniel Figeys for his supervision, advice, and guidance throughout my time working in his lab. With his experience and scientific intuition Daniel was my oasis of ideas and passions in science, that inspired and enriched my growth as a student, a researcher and a scientist want to be. I am indebted to him more than he knows.

I would like to thank all my colleagues who assisted me throughout my Ph.D. I would like to give special recognition to my experienced thesis advisory committee, Drs. Jocelyn Côté, Ilona Skerjanc and David Picketts. They were of great help to me. I would also like to thank the defense committee for taking time from their busy schedule to contribute to the making of my dream of becoming a scientist.

I am happy to be on the path to fulfillment of my dream, but more importantly, my father's. I want to thank my father and mother and my siblings Maysa, Izzo, Abdo, Deya, Alaa, Belal, Salah and Sudqi. I would also like to give special appreciation to my grandmother فاطمة. She would have been so proud of me has she lived enough to see this day.

Finally, I would like to thank my wife Abeer for all of her patience during all those years of graduate studies. I would also like to mention my pretty daughters Jenna and Aya. Even though they were making every effort to make me stay home, they were able to motivate me and inspire me to go back to the lab and do my best.

ABSTRACT

Very little is known about SMYD2 (SET and MYND containing protein 2), which is a member of the SMYD protein family. However, the interest in better understanding the roles of SMYD2 has grown due to recent reports indicating that SMYD2 methylates p53 and histone H3. I report the cytosolic and nuclear interactome of SMYD2 using a combination of IP-HTMS, ChIP-HTMS and co-immunoprecipitation methods. In particular, I report that SMYD2 interacts with HSP90α independently of the SET and MYND domain. The interaction of SMYD2 with HSP90α enhances SMYD2 histone methyltransferase activity and specificity for histone H3 at lysine 4 (H3K4) *in vitro*. Interestingly, histone H3K36 methyltransferase activity is independent of its interaction with HSP90α similar to LSD1 dependency on the androgen receptor. I also show that HSP90α is methylated by SMYD2 at lysine 209 and 565. Methylation of HSP90α is removed by LSD1. We have demonstrated that SMYD2 is being automethylated. Automethylation of SMYD2 is inhibited by the increasing concentration of recombinant SMYD2 *in vitro*. Even though the function of SMYD2 is still not fully understood, it is clear that it acts as an activator of gene expression as shown by microarray analysis. In this experiment, I have showed using a cDNA microarray that SMYD2 gain of function is correlated with the up-regulation of 37 and down-regulation of 4 genes. The majority of these genes are involved in the cell cycle, chromatin remodeling, and transcriptional regulation. TACC2 is one of the genes up regulated as a result of SMYD2 gain of function. Up-regulation of TACC2 by SMYD2 occurs as a result of SMYD2 binding to the TACC2 promoter where it methylates H3K4. When combined together, the interaction and the gene expression data confirm the suggested role of SMYD2 in DNA damage response.

1.0 INTRODUCTION

The past decade has witnessed a huge leap in the amount of data generated from different areas of life sciences. One of these advancements was the completion of the human genome project in 2003 (Collins et al., 2003). The new challenge was to understand the function of all the genes, shifting the focus from DNA to proteins. Since proteins compose most of the functional units in the cell, the complete understanding of their role in the cell is very critical to fully grasp how the cell functions. Proteomics focus on understanding the many aspects of proteins that can involve their structures, modifications, localization and their protein-protein interactions (de Hoog and Mann, 2004). Proteomics can be subdivided into expression and functional proteomics. Expression proteomics looks at changes in protein expression under different conditions compared to normal cells. These include changes in protein expression between cells exposed to different drugs, different types of stress and disease states and normal cells. This field of study has evolved from the traditional one protein scale (Western blot analysis) to two dimensional gels (Lopez, 2007) and finally to large scale analysis of changes in protein expression using techniques like isotope labeling (Gygi et al., 1999) such as Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC) (Ong et al., 2003).

The second major area in proteomics is functional proteomics which looks at understanding protein functions and elucidating their role in the cell. Identification of protein-protein interaction has emerged as one of the most important ways to understanding the functions of different proteins. This stems from the fact that most proteins are not "islands" (Gavin et al., 2002) and function by forming different complexes under different

conditions (Figeys, 2002). Hence, understanding protein-protein interactions in the cell offers an invaluable tool to understanding the functions of many unknown proteins (Figeys, 2002).

A key advancement that revolutionized the field of proteomics is the use of mass spectrometry (MS) in protein identification. Use of MS in biomolecule analysis had been extremely inefficient due to the fact that biomolecules are large and polar ions making their transfer to the gas phase very hard. Use of MS in biomolecule analysis had to wait the development of ionization methods that can solve these problems. It was only possible to readily analyze these large molecules in the past two decades as a result of the development of electrospray ionization (Fenn et al., 1989) and Matrix-Assisted Laser Desorption/Ionization (MALDI) (Karas and Hillenkamp, 1988) (Koichi Tanaka, 1988). The second key development was the efforts in sequencing and annotating genomes, which lead to a wealth of sequences database that became the foundation of high-throughput bioinformatics for protein and peptide analysis by mass spectrometry. Also, the development of suites of separation techniques and data analysis reinforced the utility of these approaches. Together these advancements were very important in the increasing interest in MS and its development to become the key tool in proteomic research (Mann et al., 2001). This section will focus on the use of MS coupled techniques to identify protein-protein interactions. I will focus on the process of identifying protein interactors by affinity purification and their identification by MS. But first, I will give a brief description about MS and its components.

1.1 Mass Spectrometry

MS is an analytical technique that dates back to the beginning of the twentieth century. It is used to measure the mass to charge ratio (m/z) of charged gaseous ions. The development of ionization methods were followed by progress in sample preparation as well

as data analysis. Together, these advancements were very important in the increasing interest in MS and its development to become a key tool in proteomic research (Mann et al., 2001). In the next part I will explain ionization by MALDI and ESI

1.2 Ion Source

MS instruments can be divided into three components, ion source, mass analyzer and a detector. In the first stage ions are being produced from the sample using ESI or MALDI ionization. Ions produced in this stage are analyzed by a mass analyzer. Different mass analysis methods can be used such as Time-of-Flight (TOF) (Wolff et al., 2001), quadrupole (March, 2000), ion traps (Schwartz et al., 2002), and Fourier transform ion cyclotron resonance (Marshall et al., 1998). Another mass analyzer that has been introduced recently is the Orbitrap (Hu et al., 2005). Mass analyzers can be coupled to either method of ionization separately or in combination with other mass analyzers. Finally, ions are detected by the detector to record their m/z ratio.

1.2.1 MALDI Ionization

Ionization and generation of gas phase molecules in MALDI is facilitated by a matrix that is mixed at a high ratio with the sample being analyzed (analyte). The matrix is usually made of small organic molecules. The most commonly used matrix for biomolecules are 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), α-cyano-4-hydroxycinnamic acid (alpha-cyano or alpha-matrix) and 2,5-dihydroxybenzoic acid (DHB) (Karas and Kruger, 2003). To generate protonated gas phase molecules the matrix is mixed with the analyte at a high ratio and spotted into a metal substrate. The dried crystals are irradiated by a laser beam which will ionize the matrix. Ionization of the analyte is believed to occur through the matrix as it transfers part of it to the analyte. At the same time the matrix also offers protection to

the analyte from the disruptive energy of the laser beam. In general, most ions generated by MALDI ionization are singly-charged ions, but multiply charged ions can also be observed (Karas et al., 2000). Ionization and generation of gas phase molecules in MALDI is facilitated by a matrix that is mixed at a high ratio with the sample being analyzed (analyte) (Koichi Tanaka, 1988). Figure 1 gives an overview of protein identification by affinity purification coupled to MALDI ionization.

1.2.2 Electrospray Ionization

Electrospray is another soft ionization method that is used to generate gas phase protonated molecules. In this process, the analyte is dissolved at low concentration in a volatile solvent (Mann et al., 2001). The solvent containing the analyte is pumped through a hypodermic needle at a low flow rate at high voltage to electrostatically disperse, or electrospray small, micrometer-sized droplets. These droplets rapidly evaporate imparting their charge onto the analyte molecules (Mann et al., 2001). Electrospray ionization occurs under atmospheric pressure preserving the structure of the sample being analyzed (Mann et al., 2001). One method to stabilize the spray is to use nebulizer gas. Molecules are then transferred into MS with high efficiency for analysis (Mann et al., 2001). An overview of affinity purification coupled to electrospray ionization is shown in Figure 2.

Figure 1: Protein identification by MALDI ionization coupled to MS. The Ionization and generation of gas phase molecules in MALDI is achieved by the use of a special matrix that is mixed at a high ratio with the sample being analyzed. This process leads to the generation of protonated gas phase molecules when the dried crystals are irradiated by a laser beam which will ionize the matrix. Ionization of the analyte is believed to occur through the matrix as it transfers part of its charge to the analyte.

Identification of Proteins with MALDI-TOF

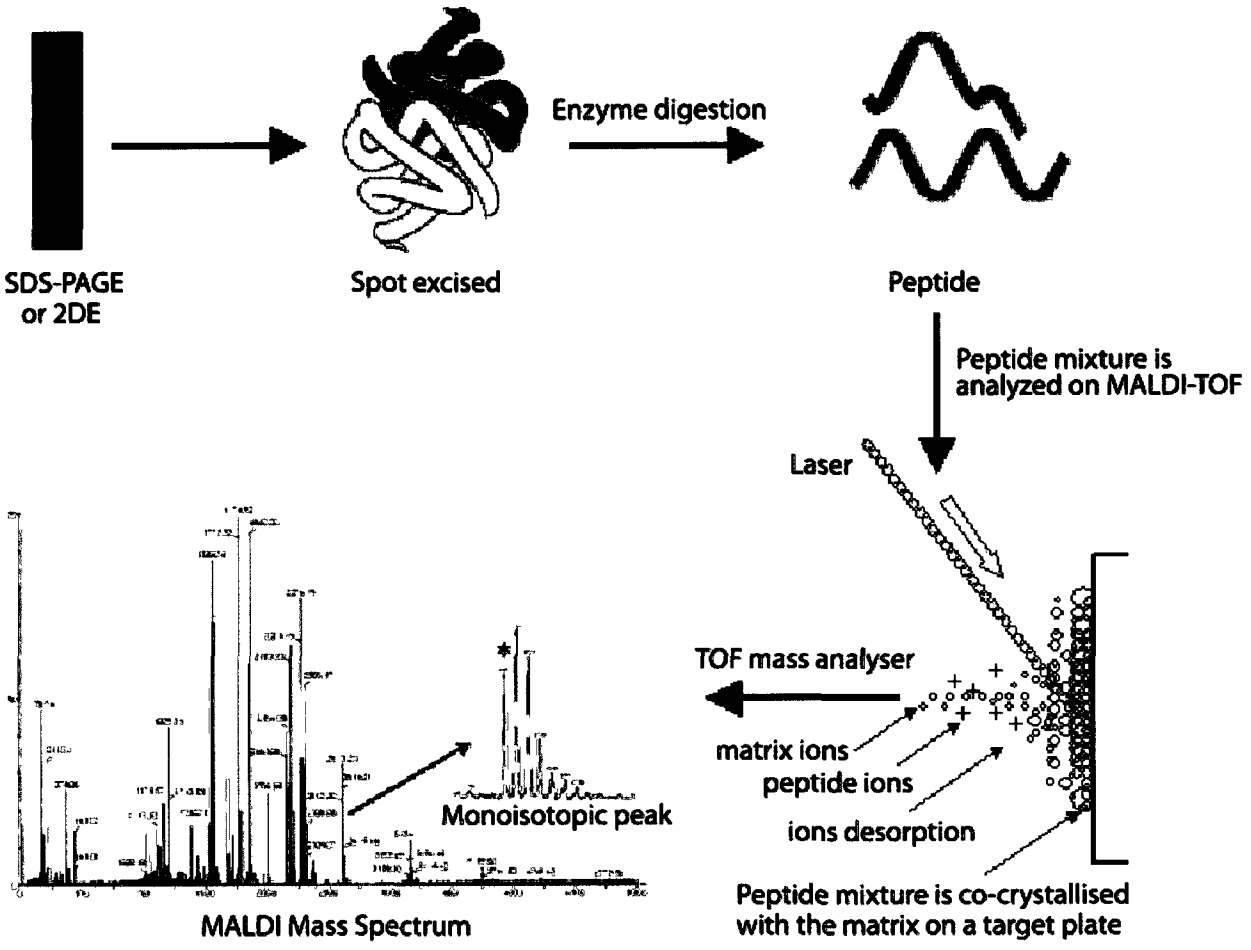
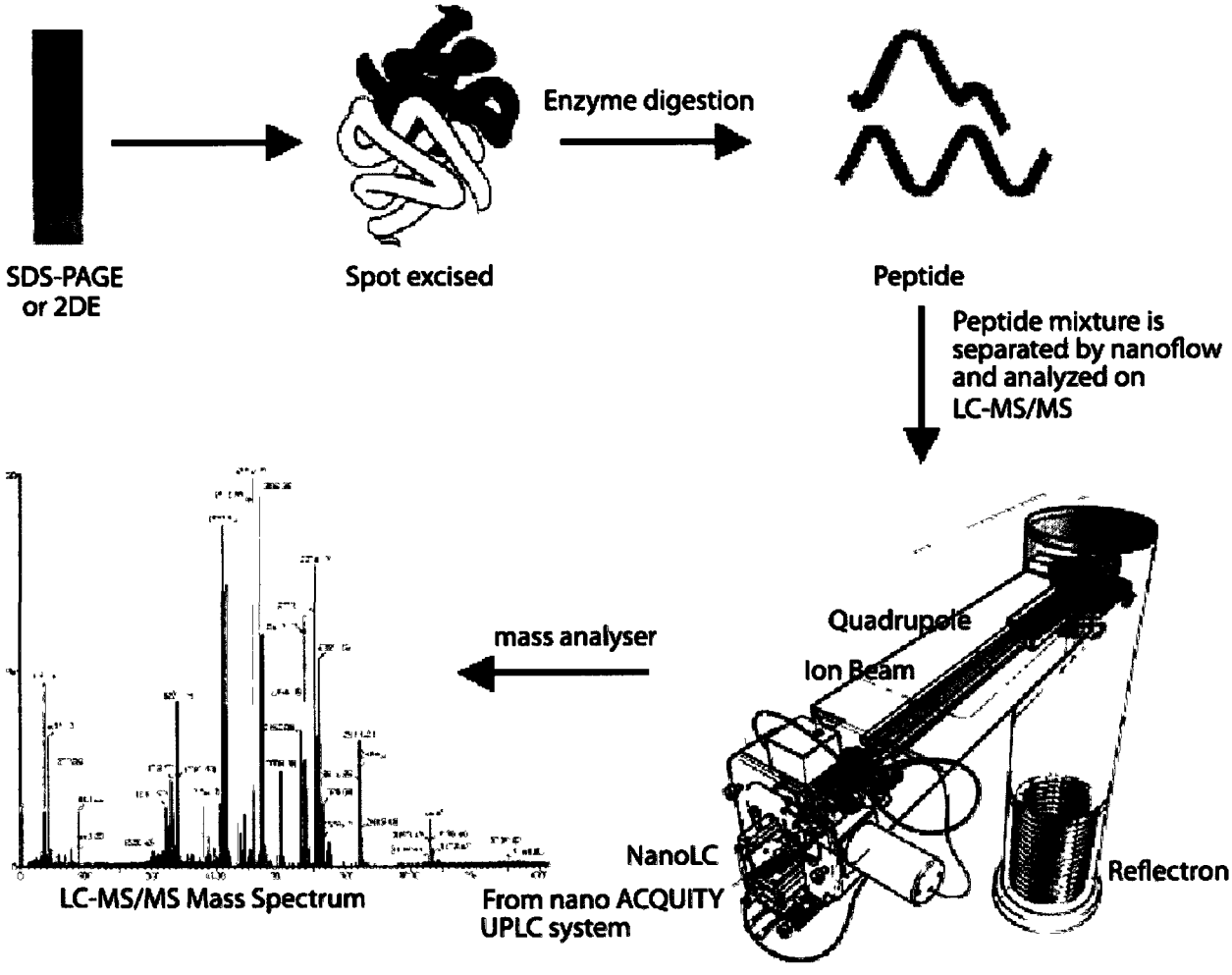


Figure 2: Protein identification by ESI ionization coupled to MS. In this process, digested peptides or whole proteins are dissolved at low concentration in a volatile solvent. The solvent containing the analyte is pumped through a hypodermic needle at a low flow rate at high voltage to electrostatically disperse, or electrospray, small droplets. These droplets rapidly evaporate imparting their charge onto the analyte molecules that are transferred into MS with high efficiency for analysis.

Identification of Proteins with ESI LC-MS/MS



1.3 Mass Analyzers

In all of the commonly used mass analyzers, electric and magnetic fields are used to apply a force on charged particles separating them according to their mass-to-charge ratio (m/z). A few examples of such analyzers are given below with a special focus on analyzers used in the thesis work. The first example is TOF analyzer; in this method, ions generated by different ionizations techniques are accelerated through an electric field with a constant kinetic energy down a flight tube. The speed of ions is inversely related to its size making smaller size ions reach the detector faster than large ions. The different ions detected will collectively form the TOF spectrum (Mann et al., 2001).

The two types of analyzers that were used extensively in my thesis work are the Linear Ion Trap (LTQ) and the LTQ Orbitrap. LTQ is a quadrupole ion trap that traps ions in a two dimensional quadrupole field for analysis, and then sequentially ejects them (James, 2002). The quadrupole mass analyzer acts as a mass filter that applies oscillating electrical fields to ions passing through the four rods. Certain masses are selected to pass through to the detector while other masses are destabilized, preventing them from reaching the detector. In order to generate a full spectrum of the ions analyzed, the amplitude of the electric field is scanned and ions are detected (Mann et al., 2001).

The Orbitrap is one of the most recently introduced mass analyzers. In this analyzer, ions are electrostatically trapped in an orbit around a central, spindle-shaped electrode (Hu et al., 2005). The Orbitrap is similar to Fourier Transform MS (FT-MS) as they both have a high mass accuracy, high sensitivity and a good dynamic range (Hu et al., 2005). FT-MS measures ion masses by detecting the image current produced by a cyclotron (charge particle accelerator) in the presence of a magnetic field (Marshall et al., 1998). Other than measuring

the deflection of ions with a detector such as an electron multiplier, the ions are injected into a Penning trap (a static electric/magnetic ion trap) where they effectively form part of a circuit (Comisarow and Marshall, 1996).

1.4 Mapping Protein-Protein Interaction

The less than anticipated number of genes identified by the human genome has further enforced the idea that proteins can have many functions in the cell. Although one should not discount the importance of non-coding RNA as regulatory elements and functional elements (Shamovsky et al., 2006) (Chen et al., 2005), proteins are the fundamental units of the cell; they participate in most biological processes. Their involvement in different processes is mostly the result of direct or indirect interaction between different proteins in a complex or pathway. As a result, mapping protein-protein interactions (PPI) has taken center stage in our understanding of different biological processes over the past decade. Technological advancements in MS and its coupling to affinity purification (AP) as well as the development of other techniques to study PPI, like yeast two hybrids (Y2H), enabled large scale analysis of PPIs. Y2H identifies binary protein interactions based on the fact that in yeast transcription factors require a DNA binding domain (DBD) and an activation domain (AD) to activate gene expression. In Y2H, the protein of interest is fused to the DBD while the prey protein is fused to the AD. Interaction of the bait and prey will reunite the DBD and AD causing them to activate transcription of a reporter gene (Figure 3A) (Fields and Song, 1989). This method can also be scaled up to screen a whole organism's cDNA library or a small part of a library (Suter et al., 2008). Y2H has been modified to create a more natural setting for the bait and prey interaction including adopting it to a mammalian system (Luo et al., 1997). These modifications include

membrane (Paumi et al., 2007), cytosolic (Mockli et al., 2007) and other compartment-specific Y2H methods that address caveats in the original method. The numerous variants of the Y2H method enhance its potential as a method of choice for mapping the human interactome at a high throughput level. Y2H still faces some challenges, such as its limited ability to identify protein complexes since it looks at binary interactions, and the issue of protein tagging and over-expression. Another example of a high throughput technique that has been used to identify PPI networks is the Protein-fragment Complementation Assay (PCA) (Tarassov et al., 2008). The principle of this assay is based on protein folding. In this process, the bait protein is fused to a fragment of a reporter protein while the prey is fused to the other fragment. Upon interaction between the bait and prey, the reporter protein is reconstituted, giving a measurable readout (Michnick et al., 2007). A number of reporter proteins have been developed that can be used to evaluate the interaction, such as murine dihydrofolate reductase and GFP.

The other most commonly used technique in identifying PPI is AP coupled to MS. In this method, a protein of interest is tagged with a specific tag and then the tag is used in the purification of protein complex using an anti-tag system immobilized on a solid support. Protein sample is then separated using one of multiple separation methods and then digested into small peptides that are identified by MS. This strategy offers the advantage of using anti-tag systems that are highly specific and commercially available in different formats (Dordick, 1991). This technique has also been made easier with the advancement in molecular biology techniques that make the tagging process very robust and simple. The use of different tags in the affinity purification of protein complexes and their identification by MS will be discussed in the following section.

1.3 Affinity purification of Protein Complexes

In recent years a number of affinity-based protein purification methods have been used to identify protein-protein interactions. These methods typically depend on the expression of a protein of interest with an affinity tag. These tags are generally made of short hydrophilic peptides such as the FLAG, hemagglutinin (HA) or poly-His tags. Other tags are small proteins like GST, thioredoxin or GFP tag (Bauer and Kuster, 2003). Table 1 gives a list of commonly used affinity tags and their sequences as well as the most common antibodies or anti-tag methods used for their enrichments. Many of these tags can be used in combination with MS to identify protein-protein interactions (Veraksa et al., 2005). Figure 3B shows the outline of the affinity purification process. In this section I will focus on large scale immunoprecipitation (IP) coupled to MS strategies for the identification of protein interactions.

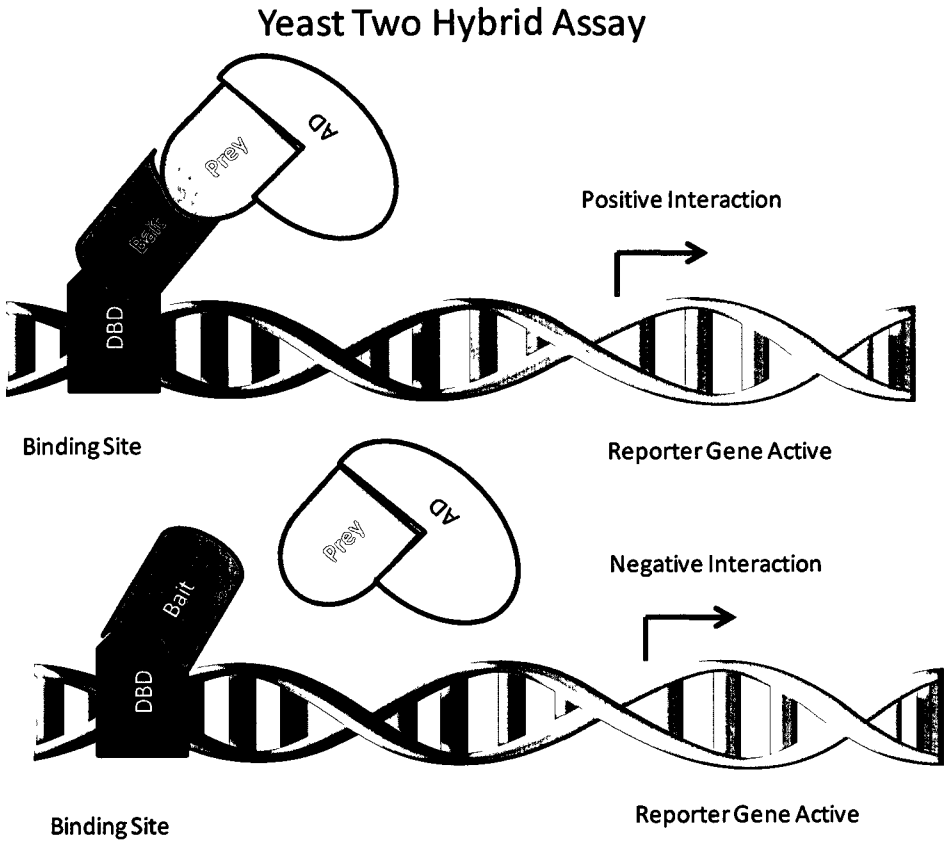
1.3.1 Tandem Affinity Purification

Tandem affinity purification (TAP) was developed as a method to purify protein complexes expressed at physiological levels under normal conditions (Rigaut et al., 1999). This method relies on the use of two tags, as the name implies. In their original paper, Rigaut et al., 1999 tested a number of tags including FLAG tag, two IgG-binding units of protein A of *Staphylococcus aureus* (ProtA), the Strep tag, the His tag, the calmodulin-binding peptide (CBP) and the chitin-binding domain (CBD) (Rigaut et al., 1999). Although none of the tags interfered with the protein function, ProtA and the CBP gave the highest recovery efficiency (Rigaut et al., 1999). The two tags are spaced by a tobacco etch virus (TEV) protease recognition site (Rigaut et al., 1999). Gavin et al., 2002 used the TAP method coupled to MS to identify the interaction partners of 589 yeast proteins. This study resulted in the

identification of 232 multi-protein complexes (Gavin et al., 2002). The quest to identify the rest of the yeast interactome using the TAP purification techniques was achieved by two other studies that looked at all the yeast 6466 ORFs (Gavin et al., 2006) (Krogan et al., 2006). In these studies, proteins of interest were fused to the TAP tag by homologous recombination. This process allows the expression of these proteins under the control of their endogenous promoters offering physiological levels of tagged protein expression. Cellular lysates containing the tagged proteins were applied to IgG-sepharose where the tagged protein bound to the IgG-sepharose through its ProtA tag. The tagged proteins along with theirs binding partners were washed to reduce the level of contamination. To further reduce the level of contamination the immobilized protein complex was incubated with TEV protease to release the bait of interest as well as its interactors. Then a second purification step is performed using the calmodulin-sepharose which binds to the CBP tag on the protein of interest in the presence of calcium. After washing, the protein complex was eluted with EGTA (Rigaut et al., 1999). Eluted protein complexes can then be resolved using different methods. In the case of Gavin et al., 2002 it was a 1-D SDS-PAGE gel. The gel was then stained and bands of interest were proteolytically digested and analyzed by MALDI-MS (Gavin et al., 2002).

Figure 3: Identification of Protein-protein interaction by Y2H and affinity purification. **A:** Y2H identifies binary protein interactions based on the fact that in yeast transcription, factors require DBD and an AD to activate gene expression. In Y2H, the protein of interest is fused to the DBD while the prey protein is fused to the AD. Interaction of the bait and prey will reunite the DBD and AD causing them to activate transcription of a reporter gene. **B:** Affinity purification depends on the generation of a tagged protein using various kinds of tags that can be added to the sequence of the protein of interest in an expression vector. Proteins of interest are expressed and then purified using antibodies or other molecules that recognizes the sequence of interest. After purification, proteins of interest are washed multiple times to remove non-specific proteins. Proteins are then eluted and separated on an SDS-PAGE gel before analyzing them on MS for protein identification.

A


Yeast Two Hybrid Assay
AD
Prey
Bait
DBD
Positive Interaction
Binding Site
Reporter Gene Active
AD
Prey
Bait
DBD
Negative Interaction
Binding Site
Reporter Gene Active

B

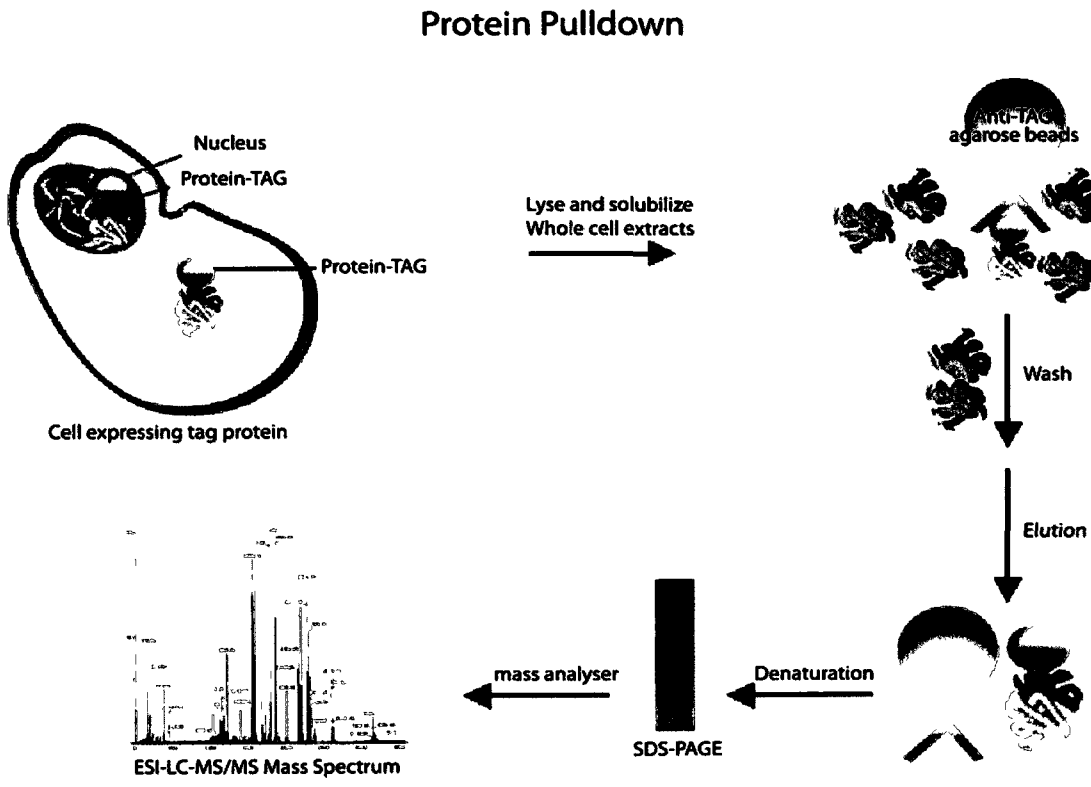

Protein Pulldown
Nucleus
Protein-TAG
Protein-TAG
Cell expressing tag protein
Lyse and solubilize
Whole cell extracts
Anti-TAG
agarose beads
Wash
Elution
mass analyser
Denaturation
SDS-PAGE
ESI-LC-MS/MS Mass Spectrum

Table 1: A list of affinity purification tags used to purify protein complexes. The table shows the sequence of these peptides or the number of amino acids in the case of small proteins. It also shows the most commonly used antibody or anti-tag molecule for their purifications.

Tag	Amino Acid Sequence	Molecular Recognition
FLAG	DYKDDDDK	M1,M2, M5 antibodies
6 × His	HHHHHH	Anti-His
HA	YPYDVPDYA	12CA5 antibody
c-myc	EQKLISEEDL	9E10 Antibody
GST	220 amino Acids	Anti-GST Glutathione
Protein A	IgG-binding domain	IgG
Strep-tag	WSAPQFEK	Strep-Tactin
CBP	CBP peptide	Anti-CBP Calmodulin
TAP	Calmodulin- and IgG-binding domains	Anti-CBP Calmodulin and IgG

The use of yeast as a model system offers the advantages of using homologues recombination to tag the protein of interest and have it expressed under the control of its own promoter (Gingras et al., 2005). This process has the advantage of eliminating untagged proteins from the cell as well as expressing the protein at its normal level. Different methods have been used in mammalian systems to overcome the problem of not having homologues recombination. These include the use of transient transfection, stable cell lines and the use of inducible promoters (Gingras et al., 2005). Due to these obstacles, identification of protein-protein interaction in mammalian cells has been more limited to smaller scale studies. One of the early large scale studies looked at positive TNF-alpha/NF-kappa B signal transduction pathway. This study uses the TAP tag approach to look at the interaction of 32 known and candidate NF-alpha/ NF-kappa B pathway components (Bouwmeester et al., 2004). TNF-alpha-responsive HEK293T cells were stably transfected with the different tagged proteins. Protein complexes were purified from the non-induced and TNF-alpha-induced cells (Bouwmeester et al., 2004). This study had the advantage of identifying protein interactions under different conditions.

1.3.2 One Tag Immunoprecipitation, a Special Look at FLAG Tag

The large scale analysis of yeast protein interaction in 2002 by Gavin et al. was also paralleled by another study performed by Ho et al. (Ho et al., 2002). This study looked at the interaction of 725 proteins detecting 3617 interactions between 1578 unique proteins covering about 25% of the yeast genome. IPs were performed using anti-FLAG antibodies. Briefly, cellular lysates from cells transfected with the FLAG tagged protein of interest were immunoprecipitated using anti-FLAG-antibodies conjugated to sepharose beads. Protein complexes were then eluted and analyzed by 1-D SDS-PAGE and stained with colloidal

Coomassie stain. Bands were then excised from polyacrylamide gels, reduced and S-alkylated, and then subjected to trypsin hydrolysis. Digested peptides were then analyzed by electrospray ionization (ESI) coupled to LC-MS/MS.

FLAG immunoprecipitation is a more simple but robust technique for identifying protein interactors. FLAG is an acidic peptide made of the following amino acid sequence DYKDDDDK. The FLAG tag is recognized by three highly specific anti-FLAG antibodies (M1, M2 and M5) (Einhauer and Jungbauer, 2001). With M1 being a calcium dependent antibody and the M5 requiring a methionine amino acid on the n-terminus, M2 has a wider range of binding with the least requirements. M2 antibody was the antibody used in all IP experiments performed in this thesis (Einhauer and Jungbauer, 2001). Another advantage about using FLAG is the size. FLAG is a small hydrophilic peptide (~1KDa) compared to the large TAP tag (original TAP is ~20KDa). It was reported that 18% of C-terminus TAP tagged essential yeast proteins gave rise to unviable strains (Gavin et al., 2002). This high percentage of unviable strains shows the great advantage FLAG has over the TAP tag and the need to use smaller size TAP tags. On the other hand, FLAG-IP suffers from a higher false positive protein-interaction identification rate compared to the TAP tag (von Mering et al., 2002). The high false positive rate in protein-protein interaction studies creates the need for cross-validation of reported interactions.

A more recent large scale study of 338 human protein-protein interactions was published by Ewing et al., 2007. This study used the FLAG-IP system coupled to ESI- LC-MS/MS to look at protein interaction in HEK293T cells (see Figure 3 for an outline of this process). Analysis of these protein interactions resulted in the identification of 24,540 potential protein interactions that was further validated to generate 6463 interactions between

2235 unique proteins. Data set generated using this method was validated using different methods generating high confidence rate in the quality of this data (Ewing et al., 2007).

In other attempts to study protein-protein interactions in lower organisms, Arifuzzaman et al., 2006 performed a large scale pull-down study using 4339 His-tagged Escherichia coli ORFs. Unlike FLAG-IP which utilizes anti-FLAG antibodies conjugated to sepharose beads, the His tag is purified on a nickel column (Arifuzzaman et al., 2006). Purified proteins were then identified using MALDI-TOF MS (Arifuzzaman et al., 2006).

1.4 Quantitative Protein-Protein Interaction

Studying protein interactions by AP-MS has the great advantage of identifying protein complexes. Identification of protein complexes rather than binary interactions allows for the placement of proteins within their biologically relevant settings. AP-MS is directly affected by advances in mass spectrometry, sample preparation, and bioinformatics (Vermeulen et al., 2008). However, as MS development allows for higher sensitivity in complex identification, the number of contaminant proteins identified will also inevitably increase. This will lead to a large false positive rate unless more stringent conditions are used to reduce contaminants. Yet, as stringency increases, so does the loss of the weak but biologically relevant interactors. To address this issue of high false positive identification rate, quantitative proteomics has been used. In these studies, mild conditions were used to preserve weak interactions (Vermeulen et al., 2008).

Many studies use quantitative proteomics to study protein interactions. Examples of quantitative techniques in protein interactions include the use of ICAT (Hara et al., 2007) (Himeda et al., 2008) (Jasavala et al., 2007), ITRAQ (Bai et al., 2008) and SILAC (Dobreva

et al., 2008). One exemple of using SILAC to identify weak protein-protein interactions was published by Trinkle-Mulcahy *et al.* (Trinkle-Mulcahy et al., 2008). They used GFP as an affinity purification tag. This tag uses the newly derived GFP binder protein which was derived from lama heavy chain antibody (Rothbauer et al., 2008). GFP binder has a high affinity and specificity to GFP (Trinkle-Mulcahy et al., 2008). By using this affinity tag in combination with three different matrices (Sepharose, agarose and magnetic beads), Trinkle-Mulcahy *et al* identified the matrices' most common contaminants in SILAC labeled mammalian cells. Identification of these proteins from either whole cell, cytoplasmic, or nuclear extracts composes what is called a "bead proteome" (Trinkle-Mulcahy et al., 2008). Such studies will help establish lists of common non-specific contaminants that are commonly seen in affinity purification. To validate their method, they used SMN (Survival of Motor Neuron) as an example. GFP affinity purification of this protein against a non-specific control identified most of SMN's known interactors. However, challenges still remain. For example, this immunoprecipitation method identified a number of validated SMN binding proteins with a ratio similar to those of non-specific interactors. Some of these specific interactors even had a ratio of less than one, like SmD1 and SmD2, which are known interactors of SMN (Gubitz et al., 2002) (Little and Jurica, 2008) (Friesen et al., 2001).

Another example of quantitative protein-protein interaction can be seen in the new method developed by Wepf et al. (Wepf et al., 2009). In this method, they integrate a small peptide as part of the affinity tag in the construct which, upon digestion, can act as a reference peptide. Similar to AQUA (Gerber et al., 2007), absolute quantification of the reference peptide can be performed. The peptide used is called SH-quant (AADITSLYK). After affinity purification, a certain concentration of the heavy form of the peptide is added.

This peptide, SH-quant*, contains a heavy isotope-labeled lysine form. The amount of bait in different affinity purifications is calculated as a ratio of the precursor-ion signals of SH-quant* and SH-quant respectively. Another form of the SH-quant peptide "SH-quant**, which contains heavy isotope-labeled lysine and leucine, was used as a correction factor for sample loss during processing (Wepf et al., 2009). The SH-quant** peptide was added just before LC-MS/MS analysis, and the correction factor was calculated as a ratio between the precursor ion signals for SH-quant* and SH-quant**. This method will offer a great advantage in following the changes in protein complexes under changing environments such as differentiation or a drug treatment (Wepf et al., 2009).

1.5 Quality of Protein Interaction Data

Generation of large protein interaction data sets is a great resource to understand the functions of many previously uncharacterized proteins. Nonetheless, validation of these data sets is a very daunting task. As the availability of large scale data sets increases, it gives researchers the chance to compare data generated by different methods. For example, comparison of data generated by Gavin et al., 2002, Ho et al., 2002 and the two other Y2H studies (Uetz et al., 2000) (Ito et al., 2001) shows a false positive rate as high as 80%. This high false positive rate can be due to many reasons including the techniques. For example, as I mentioned earlier Y2H suffers from major issues that increase its false positive rate.

One of the methods used to validate protein-protein interaction is co-immunoprecipitation. The high number of interactions in the large scale studies makes the validation process near impossible. Alternatively, bioinformatics tools can be used to give higher confidence levels to large scale data sets. For example, in the large scale experiment performed by Ewing et al., 2007 the authors used various methods to ensure the quality of

their data. Data generated from each IP was scored according to six different parameters (Ewing et al., 2007). This method generated a confidence score for each protein-protein interaction enabling the authors to judge its validity and setting a cutoff point to accept generated data. Using these criteria the number of interactions was reduced from 6463 protein-protein interactions to 2251 interactions that had high confidence scores (Ewing et al., 2007). Another example is the use of an unsupervised probabilistic scoring scheme developed by Hart et al., 2007 (Hart et al., 2007). This approach consists of giving a confidence score to each interaction that was generated by the matrix method interpretation (technique used for the creation of the interactions data sets that will include all the prey-prey interactions from given bait pull-down). This method not only increases recall and/or precision over other methods like the standard spoke model interpretation (only bait-prey interactions) but can be used to integrate data sets from other sources. The authors used this scoring scheme to combine the data generated by Gavin et al., 2002, Krogan et al., 2006 and Ho et al 2002. The results show that the scoring metric is more accurate than the filtering schemes used by the other groups.

1.6 Chromatin Modifications

Histones are basic proteins that are involved in the folding of DNA into an organized structure called chromatin. The basic unit compromising chromatin is the nucleosome which is a DNA/protein complex made of 147 bp wrapped around a histone octomer composed of two copies of the core histone proteins H2A, H2B, H3 and H4 (Couture and Trievel, 2006). These histones play an important role in controlling access to the DNA. Histone modifications have been identified as an important mechanism affecting gene regulation. Histones are subject to a wide range of post-translational modifications, mainly on their N-

termini, leading to control of accessibility to underlying DNA (Jenuwein and Allis, 2001). Known posttranslational modifications of histones include acetylation, phosphorylation, methylation, ubiquitination and SUMOylation as well as ADP ribosylation (Santos-Rosa and Caldas, 2005). These modifications of histone proteins regulate different cellular processes including transcription where histone modifications change their affinity to DNA hence control gene expression (Jenuwein and Allis, 2001). They are also involved in regulating the cellular DNA damage response as well as DNA replication (Chi et al., 2010). On the other hand their miss regulation of histone modifications will lead to diseases like cancer (Lin et al., 2010). The next section focuses on lysine methylation and demethylation of histone and non histone proteins.

1.7 Histone Methylation

Histone methylation is the result of enzyme activity of histone-lysine N-methyltransferase and histone-arginine N-methyltransferase. These enzymes add one, two or three methyl groups to lysine residues while adding one or two methyl groups on arginine residues of histone proteins (Figure 5) (Sims et al., 2003). The methyl groups transferred to the lysine or the arginine residues on histone proteins are donated by S-AdenosylMethionine (SAM) (Sims et al., 2003). Most histone lysine methyltransferases contain the SET domain, one exception is the recently discovered, Disruptor of Telomeric Silencing-1 Dot1 which lacks the characteristic SET domain (Shilatifard, 2006). Most lysine methylation sites are found on histone 3. The most commonly studied methylation sites are K4, K9, K27, K36 and K79. Another common site of histone lysine methylation is K20 on histone 4 (Ng et al., 2009). Methylation of different lysine residues on histone proteins correlates with sites of active or inactive gene expression. For example, methylation of K4 on histone 3 can be

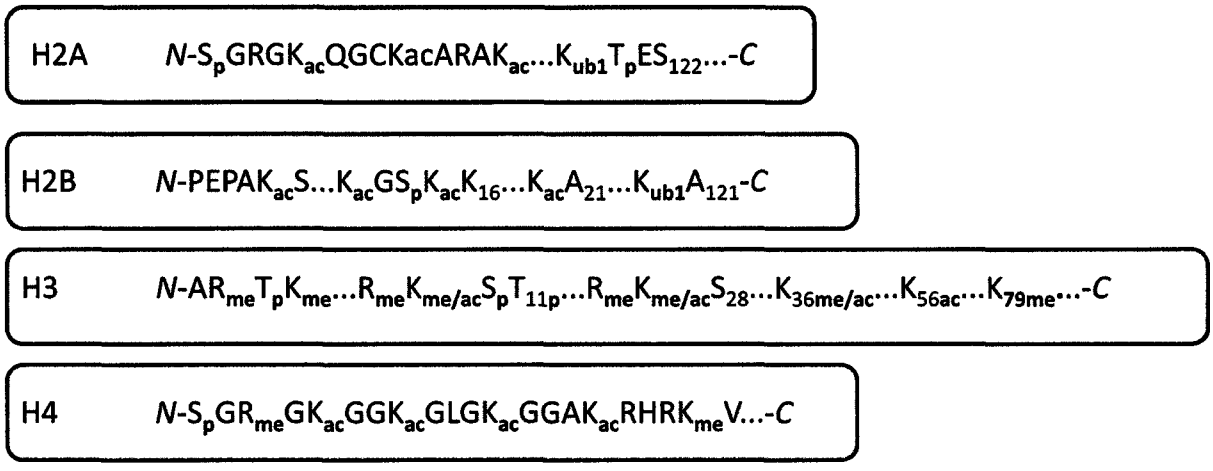
associated with active or repressed transcriptional sites (Malik and Bhaumik, 2010) while methylation of K9 on the same histone is the hallmark of repressed genes (Ng et al., 2009). Their ability to control gene expression revealed their involvement in various cellular processes and implicated them in various types of cancers (Malik and Bhaumik, 2010).

A number of proteins have been identified as methyltransferases based on the presence of a conserved SET domain in their active site. The SET domain is made of 130-140 amino acids, evolutionary conserved sequence motif. The name of this domain is derived from the three proteins it was initially characterized in, Su(var)3-9, Enhancer-of-zeste and Trithorax (Jenuwein et al., 1998)(Sims et al., 2003). Even though the SET domain is required for lysine methylation, it is not sufficient on its own. Methylation can only occur after the addition of two cystine-rich pre and post SET sequences (Kouzarides, 2002).

Known methyltransferases containing the SET domain are classified into seven different families the SUV39, SET1, SET2, EZ, RIZ, SUV4-20, and SMYD families. Other unique proteins that are not included in these families also exist such as SET7/9 and SET8 (Dillon et al., 2005). The SMYD family of proteins is largely uncharacterized. Members of this family are grouped by bioinformatics' assignment based on the presence of the two-conserved SET and MYND domain. The MYND domain in the SMYD proteins is a zinc finger motif that is involved in protein-protein interaction. The MYND domain is named after (myeloid, Nervy, and DEAF-1) the best three-characterized members of the MYND containing proteins (Gross and McGinnis, 1996).

The SMYD protein family consists of five proteins (SMYD 1 to 5) which are not fully characterized. These enzymes add methyl groups to lysine residues of histone H3 using SAM as a methyl donor substrate (Hamamoto et al., 2004)(Tan et al., 2006). Interest in the

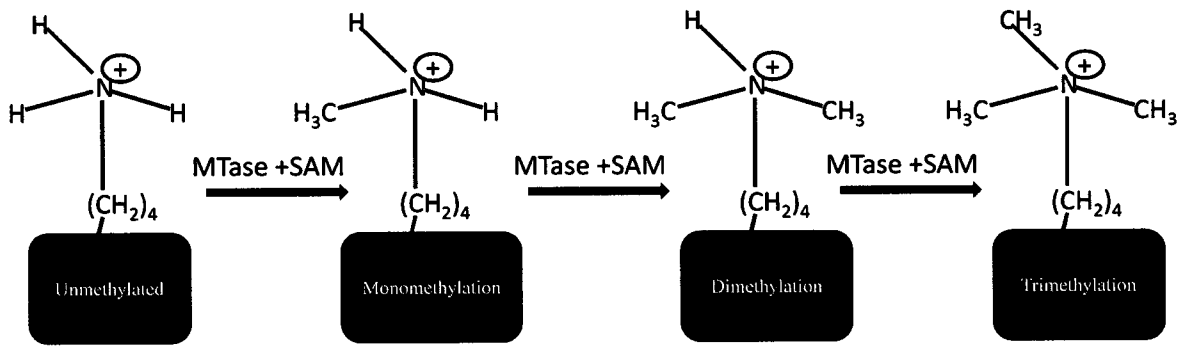
Figure 4: Various modifications on histone proteins. A diagram showing the four different histones, H2A, H2B, H3 and H4 with various modifications like acetylation, methylation and phosphorylation on various amino acid residues.



P:Phosphorylation, ac: acetylation, me: methylation, ub1: ubiquitination

SMYD family of proteins has grown significantly due to recent reports indicating that SMYD 1, 2, and 3 control gene-expression through histone methylation (Hamamoto et al., 2004)(Tan et al., 2006)(Brown et al., 2006). In addition, *in vitro* studies have shown that both SMYD1 and SMYD3 specifically methylates histone (H3) at K4 in the presence of HSP90α (Hamamoto et al., 2004)(Tan et al., 2006). In contrary to reports on SMYD1 and 3, Brown *et al.* (Brown et al., 2006) have shown that SMYD2 dimethylates H3K36 *in vitro* in the absence of HSP90α (Brown et al., 2006).

Figure 5: Methylation of lysine residues by Methyltransferase (MTase) enzymes. This diagram is showing various steps in the generation of a mono, di and trimethylated lysine residue on a substrate protein by an MTase. These steps can be carried by a combination of different enzymes that transfers the methyl group from SAM which acts as a methyl donor to the substrate lysine residue on the target protein.

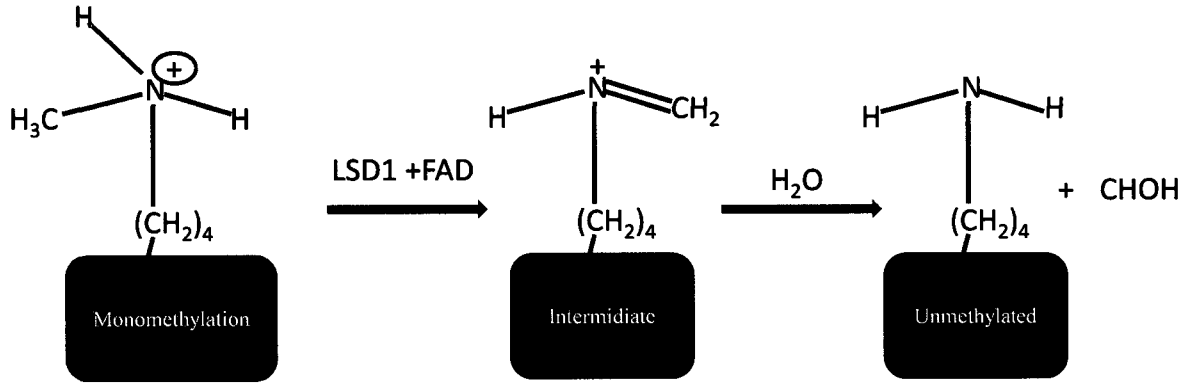

MTase +SAM
MTase +SAM
MTase +SAM
Unmethylated
Monomethylation
Dimethylation
Trimethylation

1.8 Lysine Demethylase Enzymes

Lysine methylation was considered permanent until the discovery of demethylase enzymes that are capable of removing this modification. Demethylation of histones has been shown to be mediated by proteins that have domains responsible for amine oxidation such as LSD1 and hydroxylation by JmjC-domain containing proteins (Shi et al., 2004) (Tsukada et al., 2006). Lysine-specific demethylase 1 (LSD1) was the first demethylase enzyme found to demethylate mono and dimethylated histone K4 sites on H3. LSD1 contains an FAD-dependent amine oxidase domain that acts as the active site in the amine oxidation process that leads to the removal of mono and di methylation marks from histones. LSD1 has been shown to be active on non-histone proteins such as p53 (Huang et al., 2007). Figure 6 shows the demethylation of lysine residue by LSD1. LSD1 activity is limited to mono and di methylated lysine residues due to the fact that the flavin-catalyzed amine oxidation reaction requires a lone pair of electrons on the lysine amino group (Forneris et al., 2009).

Demethylation of trimethylated histones is catalyzed by another group of enzymes that contains a Jumonji C domain. In addition to having the specific Jmjc domain the demethylation reaction requires Fe(II) and α-ketoglutarate as cofactors (Metzger and Schule, 2007). This reaction leads to the removal of mono, di and trimethyl groups and the generation of formaldehyde and succinate (Metzger and Schule, 2007).

Figure 6: Lysine demethylation of a monomethylated lysine by LSD1. Demethylation by LSD proceeds via the amine oxidation reaction that uses FAD as a cofactor. In this process an unstable intermediate carbinolamine is degraded to give the unmethylated lysine in addition to formaldehyde.


H
H3C
N
H
(CH2)4
Monomethylation
LSD1 +FAD
H
N
CH2
(CH2)4
Intermidiate
H2O
H
N
H
(CH2)4
Unmethylated
+ CHOH

1.9 Protein Methylation

Beyond their function in histone methylation, proteins containing the SET domain were also shown to methylate non-histone proteins. Examples of this activity are the methylation of p53 (Chuikov et al., 2004) and the TBP-associated factor TAF10 by human SE7/9 (Kouskouti et al., 2004). Human SET7/9 stabilizes p53 through K372 methylation (Chuikov et al., 2004). Similarly, SMYD2 regulates p53 activity through methylation at K370 which represses p53-mediated transcriptional regulation (Huang et al., 2006). These data suggest that p53, like histones, can be activated or repressed through methylation. It also points to a potential role for SMYD2 in tumorgenesis through its direct methylation activity that regulates p53, and potentially other proteins (Huang et al., 2006).

Lysine methylation has emerged as an important post translational modification due to its role in chromatin modifications. Unlike other charged modifications such as phosphorylation; methylation is an inert modification that does not change the charge of the modified protein. This inert nature of methylation made it difficult to identify methylated proteins (Lan and Shi, 2009). The breakthrough for histones was the development of antibodies that recognizes methylated histones (Perez-Burgos et al., 2004). But this approach requires developing antibody for each site of methylation and requires the prior knowledge that this site will be modified. Alternatively, high resolution MS and other proteomic methods are being used in the unbiased identification of methylation sites. A few methylation sites have been identified on non-histone proteins leading to an increased attention to this approach (Ng et al., 2009). The best example of such protein is p53, which has been shown to be methylated by a number of histone methyltransferases including SET7 and SMYD2. This methylation of p53 plays an important role in regulating p53 response to

DNA damage. P53 has been shown to be methylated at lysine 370, 372 and 382 by SMYD2, SET9 and SET8 respectively (Huang et al., 2006) (Chuikov et al., 2004) (Shi et al., 2007). In addition to p53 other non-histone proteins has been shown to be methylated by other methyltransferases such as VEGFR1 that is methylated by SMYD3 (Kunizaki et al., 2007). HSP90 is a Heat Shock Protein that acts as a molecular chaperone. It is one of the most expressed proteins in the cell (Pearl and Prodromou, 2001). The function of Hsp90 includes assisting in protein folding, cell signaling, and tumor repression. Two isoforms of HSP90 exist in the cell HSP90α and HSP90β. HSP90α is divided into three domains, N-Terminal (NTD), Middle Domain (MD) and a C-Terminal Domain (CTD) (Prodromou and Pearl, 2003) (Pearl and Prodromou, 2001). The NTD is involved in the ATP binding and hydrolysis, whereas the MD is known to interact with HSP90α client proteins (Pearl and Prodromou, 2001). The CTD is a dimerization domain that contains the tetratricopeptide repeat (TPR) motif recognition site (Mayer et al., 2009). This site is a conserved pentapeptide sequence made of MEEVD that mediates HSP90α interaction with co-factors such as the immunophilins FKBP51 and FKBP52 as well as other proteins (Young et al., 1998).

Huang and Berger made the argument based on the number of sites found on histone proteins that there will be around 40,000 sites of lysine methylation in the genome (Huang and Berger, 2008). This becomes even more complex when we take into account the fact that methylation can occur in mono, di and tri as well as the existence of various demethylases that are capable of removing the methyl mark (Bannister and Kouzarides, 2005). Taken together, this shows the need to identify sites of lysine methylation and to understand their biological functions. This thesis deals with the characterization of SMYD2 and identification

of a subset of its substrates. The next couple of pages describe into more details the rationale behind my study and my hypothesis as well as the objectives I have set for myself to achieve in this thesis.

1.10 Rationale

Numerous lines of evidence suggest SMYD2, a protein containing a SET (methyltransferase) (Marmorstein, 2003) and MYND (protein interaction) domain (Spadaccini et al., 2006), plays a pivotal role in cancer. I also obtained preliminary gene expression data on human tissues indicating that SMYD2 mRNA expression levels changes significantly for certain cancers. These observations are reinforced by SAGE public data (Lash et al., 2000) which indicates that SMYD2 mRNA is up regulated in breast carcinoma and fibroadenoma. Furthermore, SMYD2 was identified as having a genetic variant SNP (G->E at positon 165) associated with cancer (Aouacheria et al., 2005). As well, other members of the SMYD family, SMYD3 in particular have been implicated in cancer (Hamamoto et al., 2004). Despite its potentially important role in the cell, very little was known about the biological function of SMYD2 at the time this project started. These reasons inspired me to start my project on SMYD2 in an attempt to discern its biological function in the cell which also led to my following hypothesis.

1.11 Hypothesis

SMYD2 is a novel protein that is involved in gene expression regulation through its activity as a histone lysine methyltransferase. Histone methylation activity of SMYD2 is responsible for the activation of downstream genes. SMYD2 activity might require other protein cofactors that regulate its activity and specificity toward its substrates. SMYD2 is also localized in the cytoplasm where it could methylate non histone proteins.

1.12 Objectives

In order to further understand the function of SMYD2 in the cell and test my hypothesis I set forward a list of objectives. These objectives are the following:

1. My first objective was to identify and validate SMYD2 interaction partners.
2. I was also interested in the role played by SMYD2's SET and MYND domain in these interactions. This was crucial in understanding the nature of interaction between SMDY2 and its interactors.
3. Identify its histone substrates and the sites that are methylated by SMYD2 under different conditions.
4. Study the effects of SMYD2 on gene expression and its potential mechanism of action.
5. Investigate the role SMYD2 plays in the methylation of non-histone protein.
6. Identify methylation sites on a non-histone protein and validate these sites of methylation *in vitro* and *in vivo.*
7. Identify the demethylation enzyme that targets this site of methylation.
8. Extend protein interaction studies to other SMYD protein family members and identify their substrate specificity.

2.0 MATERIALS AND METHODS

2.1 Cell Culture

HeLa, NIH3T3 and HEK293T cells were cultured in high glucose Dulbecco's Modified Eagle medium (DMEM) at 37°C in a 5% CO_2 humidified incubator. DMEM medium was supplemented with 10% fetal bovine serum, 100 μg/ml of penicillin, 100 μg/ml streptomycin, and 50 μg/ml antimycotic.

2.2 FLAG-SMYD2 Cloning

Using the Promega Access RT-PCR, a full length SMYD2 cDNA was amplified from a total RNA which was extracted from HEK293T cells using the RNeasy RNA extraction kit according to the manufacture's protocol (Qiagen, Mississauga, ON). SMYD2 cDNA was then cloned into a directional Gateway Topo pENTR vector (Invitrogen, Carlsbad, CA). In order to transfer the SMYD2 cDNA into a mammalian expression vector, the GATEWAY™ cloning system was used. The pENTR vector contains attL1 and attL2 recombination sites on its 5' and 3' ends. These two sites were then utilized to transfer the SMYD2 clone into the mammalian expression vector containing FLAG tag at its c-terminus (ATCC). Small batches of transient transfection of HEK293T cells were used to verify the FLAG-SMYD2 expression.

2.3 Site-Directed Mutagenesis

The SET domain has two conserved sequences that are known to bind to SAM. These two conserved sequences in the SET domain, termed ΔNHSC (aa206-209) and ΔGEE (aa233-235), as well as the MYND domain (aa51-90) were each deleted using the Quick-Change site-directed mutagenesis Kit (Stratagene, La Jolla, CA) according to the manufacturer's protocol. Mutagenesis primers are as follows: ΔNHSC F 5'cct gat gtt gca ttg

atg tgc ccc aat gtc att gtg acc 3', ΔNHSC R 5' ggt cac aat gac att ggg gca cat caa tgc aac atc agg 5', ΔGEE F 5'gct gta cag gaa atc aag ccg gtt ttt acc agc tat att gat ctc c 3', ΔGEE R 5'gga gat caa tat agc tgg taa aaa ccg gct tga ttt cct gta cag c 3', MYND F 5'cgg tca acg agc ggg gca acc acc cca tgg ttg ttt ttg ggg 3, MYND R 3'ccc caa aaa caa cca tgg ggt ggt tgc ccc gct cgt tga ccg 3'. The resulting clones were screened for the presence of the mutation by sequencing.

2.4 IP-HTMS

IP-HTMS experiments for SMYD2 were performed as part of a large scale report previously published (Ewing et al., 2007). Briefly, transient transfections of SMYD2 wild type and mutant constructs into HEK293T, NIH3T3, and HeLa cells were done using Lipofectamine 2000 as instructed by the manufacture (Invitrogen, Carlsbad, CA). Cells were plated into 75 cm^2 flasks the day before transfection in an antibiotic free media to produce 80% confluent monolayer on the day of transfection. Following transfection, cells were split into two new flasks and left for 48 hours, reaching about full confluence on the day of harvest. After 48 hours of transfection, the cells were washed twice with ice cold PBS and then harvested. The cells were then frozen in liquid nitrogen and stored at -80°C until ready to be used. Immunoprecipitation experiments were performed from $2x10^7$ cells. The cells were lysed in 1 ml of lysis buffer (50 mM Tris HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100) containing a cocktail of protease inhibitors Mini Complete (Roche Diagnostics, Laval, Quebec) for 30 min at 4°C. Cell debris was removed by centrifugation at 14,000 rpm for 10 minutes at 4°C. The supernatant was then pre-cleared with agarose beads for two hours with shaking at 4°C. Agarose beads were removed and replaced with pre-washed anti-FLAG M2 affinity gel and incubated in a shaker over night at 4°C. The beads

were then washed five times with washing buffer (50 mM Tris HCl pH 7.4 and 150 mM NaCl). Bound proteins were eluted by incubation with 100 μL of 3xFLAG peptide (150ng/μL) for 30 min at 4°C in 1x wash buffer. The proteins were analyzed by mass spectrometry as outlined below.

2.5 ChIP-HTMS

For the modified IP protocol, the HEK293T cells which were transfected with SMYD2-FLAG or the control vector (empty vector) were cross-linked in 1% formaldehyde for 10 min at 37°C. The cross-linking reaction was then stopped by adding glycine to a final concentration of 125 mM. The cells were lysed in 1 ml of lysis buffer (50 mM Tris HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100) containing a cocktail of protease inhibitors Mini Complete and centrifuged at 5,000 rpm for 5 min at 4°C to remove cellular debris. The pellet was washed three times in 1X lysis buffer and then resuspended in RIPA buffer (150 mM NaCl,1% Nonidet P-40, 0.5% deoxycholate, 0.1% SDS, 50 mM Tris, pH 8.0, 1 mM NaF and 1 mM sodium orthovanadate) with a Mini Complete protease-inhibitor cocktail. The protein pellet was sonicated three times for 20 seconds with at least 1 minute on ice between each pulse. The lysate was then centrifuged at 13,000 rpm for 10 min. The supernatant was then immunoprecipitated using anti-FLAG antibodies and then analyzed by mass spectrometry as outlined below.

2.6 Dynabeads Immunoprecipitation

Cells expressing SMYD2, SMYD3 and SMYD5 were harvested and lysed according to the previous protocol. The supernatant of these cells was then incubated with either an IgG control or anti-HSP90α antibody while shaking at 4°C overnight. Protein G coupled Dynabeads (Invitrogen) were washed with PBS twice and added to the reaction mixture and

incubated on a shaker for two hours at 4°C. The beads were then washed 4 times with washing buffer (50 mM Tris HCl pH 7.4 and 150 mM NaCl). Bound proteins were eluted by incubation with 50 μL of 1X SDS sample buffer without DTT at 95°C for 5 minutes. Eluted proteins were then analyzed by SDS-PAGE and visualized by silver stain or blotted on a nitrocellulose membrane and probed with anti-FLAG or anti-HSP90α antibodies.

2.7 MS Analysis

Gel bands were excised with a scalpel and placed in fresh microtubes. The gel pieces were briefly washed with 50mM ammonium bicarbonate solution and subsequently shrunk in a solution of 50% acetonitrile (V/V) and 25mM ammonium bicarbonate at room temperature for 15 min. The supernatants were discarded, and the gel pieces were dried for ~10 min at low temperature in a speed-vac. Protein disulfide bonds were reduced by swelling the gel pieces in 10mM DTT and 50mM ammonium bicarbonate at 56°C for 15 min. The free sulfhydryl groups were then alkylated with 100mM iodoacetamide in 50mM ammonium bicarbonate for 15 min at room temperature. The gel pieces were washed with 50mM ammonium bicarbonate and shrunk in a solution of 50% acetonitrile (V/V) and 25mM ammonium bicarbonate at room temperature for 15 min. The pieces were dried for ~10 min. Trypsin in 50mM ammonium bicarbonate was added to each tube and the samples were placed at 37°C overnight (12 to 16 hours) for digestion. Peptides were extracted from the gel pieces with 5% formic acid (V/V) and 50% acetonitrile (V/V). The peptide solutions were then speed-vac to dryness and stored at -20°C until mass spectrometric analysis.

Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) was performed by dissolving the peptide samples in 5% formic acid. The samples were injected on a 200μm x 5 cm pre-column packed in-house with 5 μm YMC ODS-A C18 beads

(Waters, Milford, MA) using a micro Agilent 1100 HPLC system (Agilent Technologies, Santa Clara , CA). The peptides were desalted on-line with 95% water, 5% acetonitrile and 0.1% formic acid (V/V) for 10 min at 10μL/min. The flow rate was then split before the pre-column to produce a flow rate of approximately 200μl/min at the column. Following their elution from the pre-column, the peptides were directed to a 75μm x 5 cm analytical column packed with 5 μm YMC ODS-A C18 beads. The peptides were eluted using a 1 hour gradient (5 to 80% acetonitrile with 0.1% formic acid) into a LTQ linear ion-trap mass spectrometer (Thermo-Electron, Waltham, MA). MS/MS spectra were acquired in a data-dependent acquisition mode that automatically selected and fragmented the five most intense peaks from each MS spectrum generated.

Peak lists were generated from the MS/MS .raw file using Mascot Distiller 2.0.0.0 (Matrix Science, UK) to produce .mgf files. For each MS/MS individual peak lists were generated assuming a 2+ and a 3+ charge. The .mgf files were then analyzed and matched to the 248060 Homo sapiens protein sequences in a concatenated database containing the NCBI forward and reverse database (released January 2007) using the Mascot database search engine version 2.1.0.4 (Matrix Science, Boston, MA) with Trypsin as digestion enzyme, carbamidomethyl of cysteine as a fixed modification and methionine oxidation as a variable modification. The searches were restricted to the Homo sapiens protein sequence because the experiments were performed using a human cell line. Peptide and MS/MS mass tolerances were set at ± 2 Da and 0.8 Da, respectively, with 1 miss-cleavage allowed. An ion score cut-off of 30 was used for acceptance of individual MS/MS spectra. This threshold was chosen based on the report by Elias et al. (Elias et al., 2005) to ensure a false positive rate of less than 1% in our MS data. Protein hit required at least one "bold red peptide", i.e. the most

logical assignment of the peptide in the database selected, to be considered. Furthermore when peptides matched to more than one database entry, only the highest scoring protein was considered. All of the proteins observed in the negative control ChIP-HTMS were subtracted from the SMYD2 ChIP-HTMS.

2.8 Immunocytochemistry

HeLa cells grown on cover slips, at about 50% confluence, were washed twice with PBS and then fixed using a freshly prepared mixture of methanol: acetone (1:1) for one minute and then washed four times with PBS. Cells were incubated for 1 hour with FITC-conjugated FLAG antibody (5 μg/ml) and then washed twice with PBS before adding the (4',6-diamidino-2-phenylindole) DAPI stain for 15 minutes.

2.9 Cellular Protein Extraction and Western Blots

Total cellular proteins were prepared by lysing cells in RIPA buffer containing a Mini Complete protease-inhibitor cocktail. Proteins were separated on 4–12% SDS-PAGE gel (NuPAGE, NOVEX, San Diego) and then transferred to nitrocellulose membranes. A western blot analysis for SMYD2 and its mutants was performed using monoclonal horseradish peroxidase-conjugated mouse anti-FLAG M2 antibodies (Sigma, St. Louis, MO). H3 methylated lysines were detected by rabbit polyclonal antibodies for Mono, Di, and Tri-methylated K4 (Upstate Cell Signaling Solutions, Lake Placid, NY) . HSP90α protein was detected using mouse monoclonal antibodies purchased from (Upstate Cell Signaling Solutions, Lake Placid, NY). P53 and EBP41L3 were detected using mouse monoclonal antibodies purchased from (Calbiochem, Darmstadt) and (Abnova, Jhongli City, Taoyuan) respectively. Rabbit IgG was used as a negative control and were purchased from (Sigma, St.

Louis, MO). Membranes were then probed with horseradish peroxidase-labeled goat anti-rabbit IgG or goat anti-mouse secondary antibody (Dako Cytomation Inc., Mississauga, ON).

2.10 *In vitro* Methylation assay

In vitro methylation assay was performed on extracts from HEK293T cells transfected with plasmids expressing the mock vector (pcDNA3.1), wild type SMYD2 (pcDNA3.1-FLAG-SMYD2) or one of the SMYD2 mutants (pcDNA3.1FLAG-SMYD2 ΔGEE, pCDNA3.1 FLAG-SMYD2 ΔNHSC or pcDNA3.1-FLAG-SMYD2 ΔMYND). FLAG tagged proteins were then immunoprecipitated according to the previously described protocol using Anti-FLAG antibodies conjugated to agarose beads. The *in vitro* histone methyltransferase assay was performed using an *in vitro* methylation kit purchased from Upstate Cell Signaling Solutions (Lake Placid, NY). *In vitro* methylation assay was done on the beads before the FLAG protein elution. After incubating the cell lysate with the anti-FLAG M2 affinity gel, the beads were washed to remove nonspecific binding proteins. The beads containing FLAG tagged proteins were mixed with 1.65 μCi of ^{3}H labeled S-adenosyl-L- [methyl-^{3}H] methionine SAM (PerkinElmer life sciences), 4 μg of chicken core histone, 1.5 μg of HSP90α (Calbiochem, La Jolla, CA) in 50 mM Tris, pH 9.0 and 0.5 mM DTT. The reaction mixture was incubated for 3 hours at 30°C with gentle mixing. The methyltransferase activity was measured by scintillation counting. In order to measure the H3-K4 specific activity, SMYD2 was incubated with 15 μM unlabelled SAM and 4 μg of recombinant histone H3. H3-K4 methylation was detected using an anti Mono, Di and Tri H3-K4 antibody or di-methylated H3-K36 (Upstate Cell Signaling Solutions, Lake Placid, NY).

2.11 Gene Expression Analysis by Microarray

Changes in gene expression resulting from over-expression of SMYD2 were measured using cDNA microarray. Human HEK293T cells of about 80 % confluent were transfected with either SMYD2 or a mock vector. After 24 hours, cells were harvested and frozen in liquid nitrogen. Total RNA was extracted from HEK293T cells using RNeasy kit according to the manufacturer's protocol (Qiagen, Missisauga, ON). Total RNA samples were treated with DNAse to remove potential genomic DNA contamination. Transfection efficiency was determined to be higher than 95 % using immunocytochemistry with anti-FLAG antibodies conjugated to FITC. Samples were sent for analysis at the University of Toronto Microarray Center. The two different mRNA samples (sample versus mock control) were labeled with Cy3 or Cy5 dye and subjected to co-hybridization onto a single-spotted human 19K Array (H19K) containing 19,008 characterized and unknown human ESTs. The gene expression analysis was done in triplicate. The data generated were analyzed by the SAM software to generate statistically significant hits. Reported genes were shown by the SAM software to be statistically significant and were filtered to eliminate all genes with absolute value of $\log_2$ ratio > 0.6 in less than two observations.

2.12 Chromatin ImmunoPrecipitation Assays

Chip assay was performed using a chromatin IP assay kit from (Upstate Cell Signaling Solutions, Lake Placid, NY). HEK293T cells transfected with SMYD2-FLAG or the control vector were cross-linked in 1% formaldehyde for 10 min. The cross-linking reaction was then stopped by adding 125 mM glycine. The samples were then immunoprecipitated using the anti-FLAG M2 affinity gel. The following primers were used to examine whether the promoter sequence of the DNA of interest was pulled down. The

primers used for the PCR amplification are, Tacc2 chip1 F 5'-aaccaatcagcggcactatc, Tacc2 chip1 R 5'-ttcttcatagcacttacaatgacc, Tacc2 chip2 F 5'- ggctaaagggataggtgga, Tacc2 chip2 R 5'-gcacagtgactcatgcctgt, Tacc2 chip3 F 5'-cccgacctgaaagttgctat, Tacc2 chip3 R 5'-cttatcaggaagcggctgac, Tacc2 chip4 F 5'- gatctcaggagttgggcaag, Tacc2 chip4 R 5'-agcgggtgacttgagaaatc

2.13 Semi-Quantitative RT-PCR

Human HEK293T cells at about 80 % confluency were transfected with a mock control, wild type SMYD2, ΔNHSC or ΔGEE for 24 hours. The cells were harvested and frozen in liquid nitrogen. Total RNA was extracted using an RNeasy kit according to the manufacture's protocol (Qiagen, Missisauga, ON). Total RNA samples were treated with DNAse to remove potential genomic DNA contamination. The gene expression level was measured using semi-quantitaive RT-PCR using the iScript One Step RT-PCR kit (Biorad, Hercules, CA). The primers used were: TACC2 F 5'-cacccactgaggagttggat, TACC2 R 5'-acttgacagggctctcctga, SMYD2 F 5'- tgcaagcaggcattttactg, SMYD2 R 5'- tttctgtttggccagaatcc and GAPDH F 5'-tgatgacatcaagaaggtggtgaag, GAPDH R 5'-tccttggaggccatgtgggccat.

Microarray data for five genes was validated using the previous method with the following primer pairs: Smarca2 F 5'-tccgaggcaaaatcagtcaag, Smarca2 R 5'-ttcctcgatttggccttttct, Wdr9 F 5'-ttccagagctcgtgaatcct, Wdr9 R 5'-cgggaccagtttttctttga, Bat1 F 5'-acagctctggctttcgtgac Bat1 R 5'-gcactcatgctggacttctg, Chd9 F 5'-catctggaaaccggacagat, Chd9 R 5'-ctgccactgatggcctctat, Akap13 F 5'-agttctcttccgctccaaca, Akap13 R 5'-actcaccaaaccctgaagga.

2.14 Combining SMYD2 Interactions and Gene Expression Data

The Unigene list generated by the microarray data and the names of proteins generated by the interaction mapping were imported into Pathway Studio 5.0. This text-mining software uses a database of molecular network that was assembled from scientific abstracts and a manually curated dictionary of synonyms to recognize biological terms (Nikitin et al., 2003). The combined interaction network was created by searching the database for direct interactions between the imported gene/protein and for other proteins that are one edge neighbor from the imported list of gene/protein. The protein-protein interactions are represented as a network graph where the vertices are the proteins and the edges are the references (Figure 24).

3.0 RESULTS

SMYD2, a 49.7 kDa protein, is composed of 433 amino acids and is a member of the SMYD family. Sequence alignment of SMYD2 with other SMYD family members revealed regions that are known to possess methyltransferase activity (Rea et al., 2000). SMYD2 shares 31% sequence identity and 52% sequence similarity with SMYD3 which is another member of the SMYD family. The sequence homology of SMYD2 and SMYD3 is primarily due to the MYND and SET domains. Figure 18A shows the alignment of the SET and MYND domain of the five different proteins of the SMYD family. The alignment and the phylogeny tree in figure 18B shows the close relationship between SMYD1, 2 and 3 while the other two SMYDs are more distant.

3.1 Cellular Localization of SMYD2 and its Oncogenic Activity

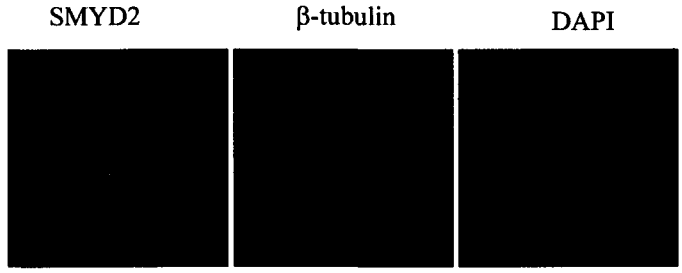
Immunocytochemical staining of SMYD2 in HeLa cells demonstrated that SMYD2 was localized in both the cytoplasm and in the nucleus (Figure 7A). SMYD2 localization in the nucleus was in accordance with its histone methyltransferase activity; however, the cytoplasmic localization of SMYD2 suggested that it had a structural role as observed in (Figure 7A). In general, cells over-expressing SMYD2 tend to have a rounded shape with a lower microtubule network.

SMYD2 was identified by Aouacheria *et al.*, 2005 as a gene with a genetic variant (SNP) that is associated with cancer (G165E). It has also been recently reported that SMYD2 functions as a proto-oncogene (Huang et al., 2006). The hallmark of oncogenes is their abilities to enhance growth and proliferation. In order to test oncogenic activity of SMYD2, we performed both colony formation and proliferation assays. NIH3T3 cells over expressing wild type SMYD2 showed 50 folds increase in the number of colonies as compared to mock

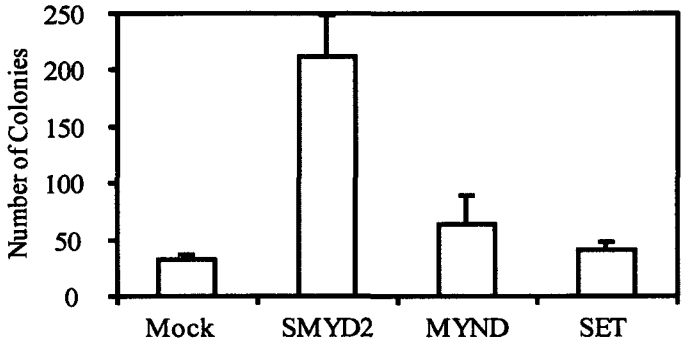
control (Figure 7B). In contrast, cells transfected with SMYD2 constructs lacking either MYND or SET domain showed comparable colony numbers to that of the mock transfected cells. These results were further supported by proliferation assays. NIH3T3 cells expressing wild type SMYD2 showed two folds increase in cell numbers as compare to mock transfected cells (Figure 7C).

Figure 7: Cellular Localization of SMYD2 and its Oncogenic Activity. ***(A)*** Immunocytochemistry staining of HeLa cells transfected with FLAG tag SMYD2 and immunostained with anti-FLAG FITC-conjugated antibodies. SMYD2 is localized to both the nuclear and the cytoplasm of cells. The cells were also stained with anti-β-Tubulin-Cy3 conjugated antibodies indicating a change in the skeletal structure of SMYD2 transfected cells. DAPI was used to stain nuclei. ***(B)*** Colony formation assay in NIH3T3 cells showing the effects of the wild type SMYD2 and its deletion mutants (MYND and the SET domain) on the number of colonies formed in agar media n=3 ± SD. ***(C)*** Increase in growth rate of NIH3T3-SMYD2 expressing cells compared to cells transfected with a mock vector. Cells were grown in 6-well plates and counted by Trypan blue staining n=3 ± SD.

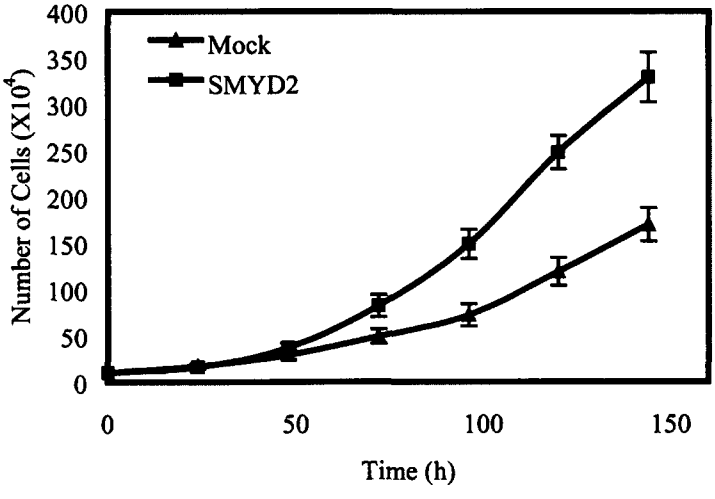
A


SMYD2
β-tubulin
DAPI

B


Number of Colonies
250
200
150
100
50
0
Mock
SMYD2
MYND
SET

C


Number of Cells (X10⁴)
400
350
300
250
200
150
100
50
0
Mock
SMYD2
0
50
100
150
Time (h)

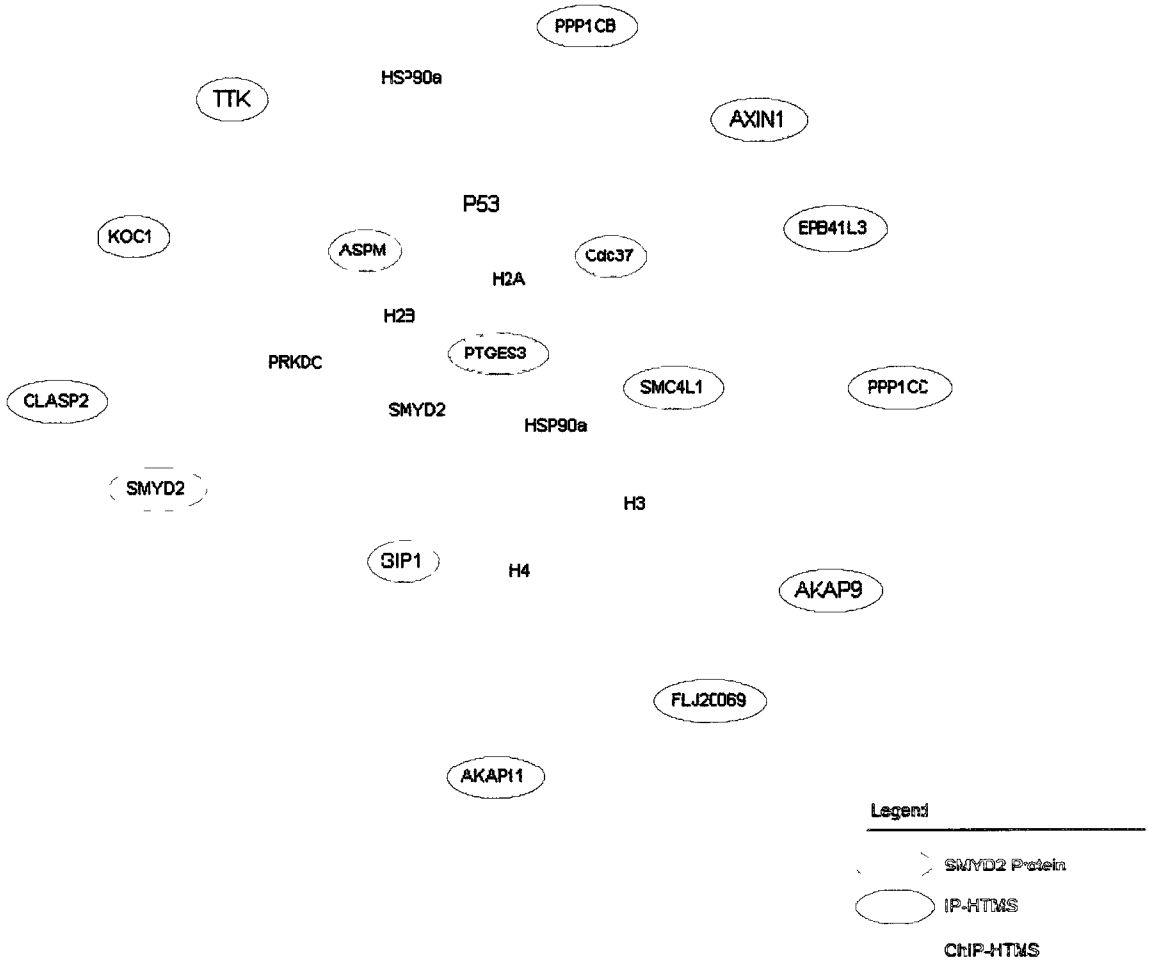
3.2 Identification of SMYD2 Protein Interactions.

In order to understand the role of SMYD2 in the cell, we mapped its interactome using two approaches. The first approach, termed IP-HTMS, looked at the global interactors of SMYD2 (Ewing et al., 2007). In this approach, FLAG tagged SMYD2 was transiently transfected in HEK293T cells 48 hours after the transfection, the FLAG tagged SMYD2 was Immunoprecipitated, and its protein interactors identified by mass spectrometry. The results from the global IP-HTMS method are represented in figure 1 and reference 1. One limitation of the IP-HTMS method is a bias against nuclear proteins. Therefore, a second approach, termed ChIP-HTMS, was used. The approach consists of a modified chromatin IP protocol (see experimental section) coupled to HTMS to enrich SMYD2 interaction partners from the nucleus and from SMYD2 bound to DNA. The ChIP-HTMS method enlarged the list of SMYD2 nuclear proteins interactors (Figure 8). In total, 21 proteins were found to interact with SMYD2.

Proteins identified as SMYD2 interaction partners were divided into two groups according to their functions: histone regulation and microtubule dynamics. The former group includes histone H3, histone H4, and Epb4113 (Regulator of arginine-N-methyltransferase specifically PRMT3 and 5) (Singh et al., 2004) (Jiang et al., 2005). Histone H3 is a known substrate for SMYD2. Another protein that is known to interact with SMYD2 is p53, which was recently shown to be methylated by SMYD2 (Huang et al., 2006). The second group includes Clasp2, a protein that binds to the end of growing microtubules in association with cytoplasmic linker proteins (clips) (Lansbergen et al., 2006). Other microtubule related proteins were observed including Axin and TTK which are known to play a role in microtubule dynamics. TTK is localized to the centrosome and is

known to phosphorylate TACC2 (Dou et al., 2004). We have shown that TACC2 is up-regulated as a result of SMYD2 over-expression (see below). Co-immunoprecipitation was performed to validate the interactions of SMYD2 with HSP90α, EPB41L3, and p53. The interactions with these proteins are explored in more details below.

Figure 8: A diagram showing the SMYD2 protein-protein interaction partners with their cellular localization as identified by Ewing et al (Ewing et al., 2007) using IP-HTMS and with our modified ChIP-HTMS method as described in the experimental section. The SMYD2 interaction partners identified by IP-HTMS were located outside the nucleus while the ChIP-HTMS interactors were located in the nucleus.


PPP1CB
HSP90a
TTK
AXIN1
P53
KOC1
ASPM
Cdc37
EPB41L3
H2A
H2B
PRKDC
PTGES3
SMC4L1
PPP1CC
CLASP2
SMYD2
HSP90a
SMYD2
H3
SIP1
H4
AKAP9
FLJ20069
AKAP11
Legend
SMYD2 Protein
IP-HTMS
ChIP-HTMS

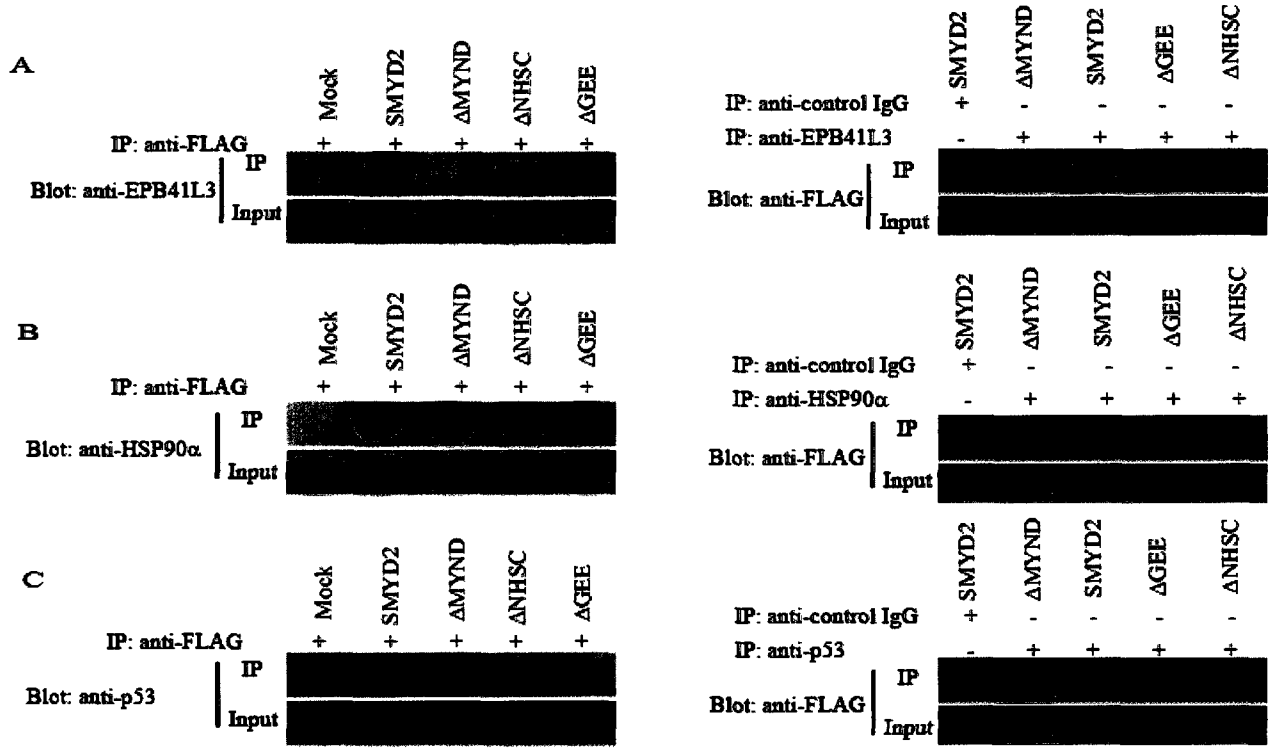
3.3 Involvement of the SET and MYND domains of SMYD2 in its Interactions.

The MYND domain mediates protein-protein interactions by binding to the conserved PXLXP motif (Ansieau and Leutz, 2002). Interestingly, five out of the 21 proteins found to interact with SMYD2 possess a PXLXP motif, including EPB41L3. EPB41L3 was chosen for further validation because of its role in regulating arginine methyltransferase 3 and 5 (Singh et al., 2004) (Jiang et al., 2005). To test whether the MYND domain of SMYD2 was required for its interaction with EPB41L3, co-IP assays were performed between EPB41L3 and wild type SMYD2 or the ΔMYND and the ΔSET mutants. As shown in (Figure 9A), EPB41L3 co-immunoprecipitated with the wild type SMYD2 but not with the MYND mutant. Other proteins that did not contain the PXLXP motif such as HSP90α and p53 interacted with SMYD2 independently of the MYND domain (Figure 9B and 9C). EPB41L3 interaction with SMYD2 was not affected by deletions of the SET domain consensus sequences (NHSC and GEE). This clearly indicates that the interaction between SMYD2 and EPB41L3 is dependent on the MYND domain and that the PXLXP motif in interactors is a good predictor that interactions occur with the MYND domain.

Other interactions depend on the activity of the SET domain. For example, the interaction between p53 and SMYD2 was abolished by the deletion of the two SET domain consensus sequences (Figure 9C). This correlates with the nature of interaction between p53 and SMYD2 in which p53 acts as a substrate for SMYD2.

The interaction between SMYD2 and HSP90α was not affected by any of the deletions which raises the possibility that HSP90α interacts with another part of SMYD2 or indirectly through another protein that is also not affected by the deletion of the MYND and SET domains (Figure 9B).

Figure 9: Validation of some physical interactions of SMYD2 by Co-IP. **(A)** Western blots showing Co-IP assays. Transiently expressed wild type FLAG-SMYD2 and its mutants, endogenous EPB41L3, and control IgG were immunoprecipitated from HEK293T cell extracts, and precipitates were probed for the immunoprecipitants (IP; top panel). The bottom panel shows the western blot analysis of input whole cell lysate. SMYD2 interaction with EPB41L3 requires a functional MYND domain as shown by the inability of the MYND domain deletion to immunoprecipitate EPB41L3 and vice versa. **(B)** Similar experiments performed for FLAG-SMYD2, endogenous HSP90α, and control IgG. Interactions between HSP90α and SMYD2 is not mediated through the MYND domain as shown by the ability of the MYND deleted form of SMYD2 to immunoprecipitate HSP90α and vice versa. **(C)** Similar experiments performed for FLAG-SMYD2, endogenous p53 and control IgG. Interaction between p53 and SMYD2 is not mediated through the MYND domain as shown by the ability of the MYND deleted form of SMYD2 to immunoprecipitate p53 protein and vice versa.


A
Mock
SMYD2
ΔMYND
ΔNHSC
ΔGEE
IP: anti-FLAG
Blot: anti-EPB41L3
IP
Input
IP: anti-control IgG
IP: anti-EPB41L3
Blot: anti-FLAG
B
IP: anti-HSP90α
C
IP: anti-p53

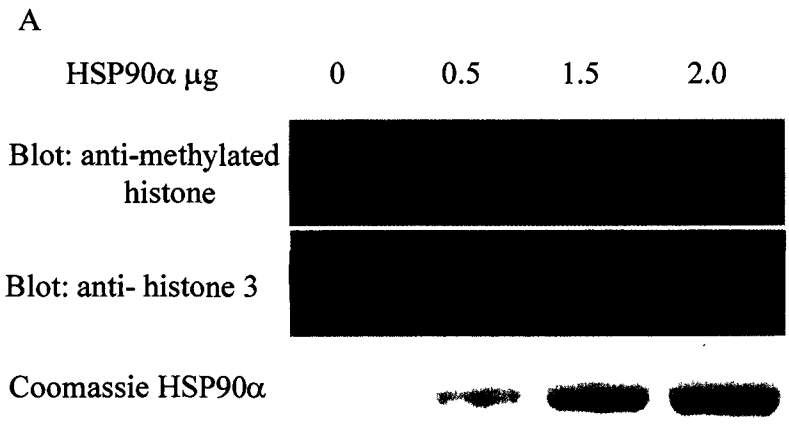
3.4 The Interaction between HSP90α and SMYD2 Changes SMYD2 Methyltransferase Activity and Specificity.

Our protein-protein interaction data showed that HSP90α interacts with SMYD2 which was further confirmed by co-IP experiments (Figure 8 and 9B). Previous studies have shown that SMYD1 and 3 can methylate H3K4 in the presence of HSP90α (Hamamoto et al., 2004)(Tan et al., 2006). However, a recent report published by Brown (Brown et al., 2006) suggested that SMYD2 dimethylates K36 of H3 and not K4, although their *in vitro* experiments did not include HSP90α.

The interaction of SMYD2 with HSP90α suggests that HSP90α plays an important role in the functions of SMYD2. Hence, we performed *in vitro* methylation assays to verify if HSP90α affects the activity and specificity of SMYD2 as a methyltransferase. Briefly, SMYD2 was immunoprecipitated using anti-FLAG antibodies, and the immunoprecipitate was incubated with recombinant histone H3, SAM with or without HSP90α. Proteins were fractionated on SDS-PAGE and immunoblotted with anti-histone H3 antibodies (Figure 10A). Clearly, the *in vitro* methyltransferase activity of SMYD2 towards histone H3 was dependent on the presence of HSP90α. A second set of experiments was performed to determine the specificity of the methylation site using antibodies directed against mono-, di- or tri- methylated H3K4. In short exposure experiments, no methylation activity towards K4 or K36 were observed without HSP90α (Figure 11). Methylation activity of SMYD2 towards H3 at K4 was enhanced by the addition of recombinant HSP90α. H3K4 methylation was readily detectable after a short exposure of the autoradiography film (Figure 10B and Figure 11). Moreover, methylation at K36 was only weakly observable in the absence of

recombinant HSP90α and following much longer exposure of the autoradiography film (Figure 11).

Figure 10: HSP90α Enhances SMYD2 Methyltransferase Activity. **(A)** *In vitro* histone methylation assay using an increasing amount of HSP90α recombinant protein. This shows that SMYD2 requires HSP90α for its activity. **(B)** *In vitro* methylation assay showing SMYD2's ability to methylate histone H3 at K4 and not K36. In the top panel, mock control and immunoprecipitated SMYD2 from HEK293T cells were added to the *in vitro* assay and methylation identified on western blot using specific antibodies. FLAG IP refers to immunoprecipitated proteins using anti-FLAG antibodies from HEK293T cells. The bottom western blot analysis of input whole cell lysate (input).


A
HSP90α μg
0
0.5
1.5
2.0
Blot: anti-methylated histone
Blot: anti- histone 3
Coomassie HSP90α

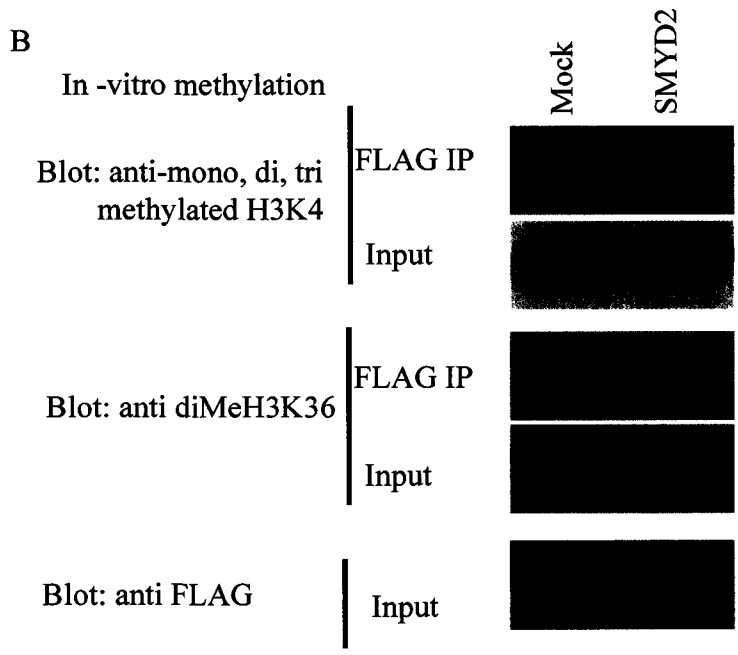
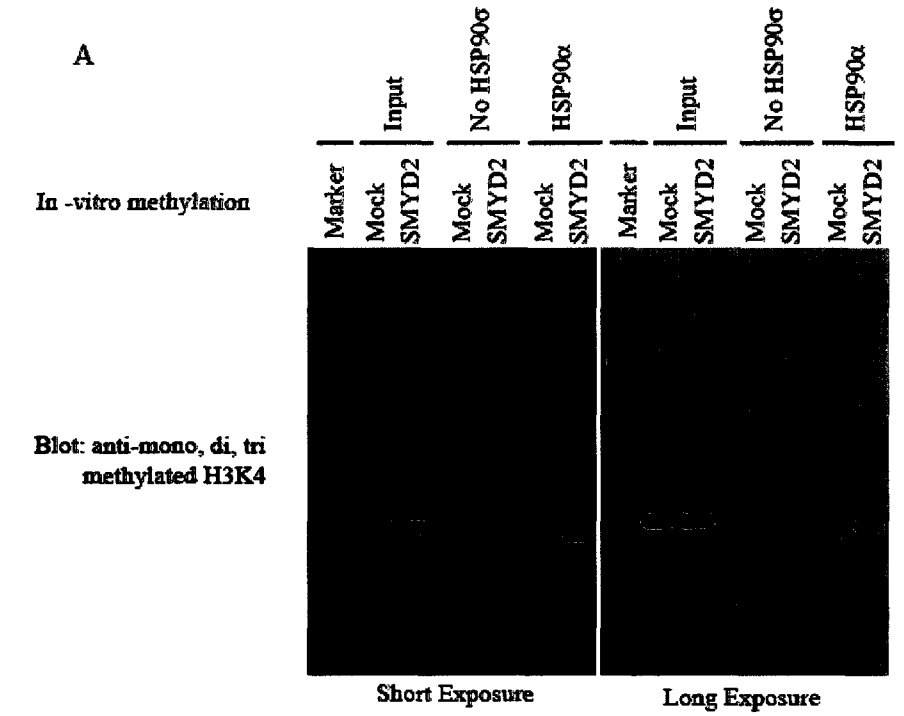
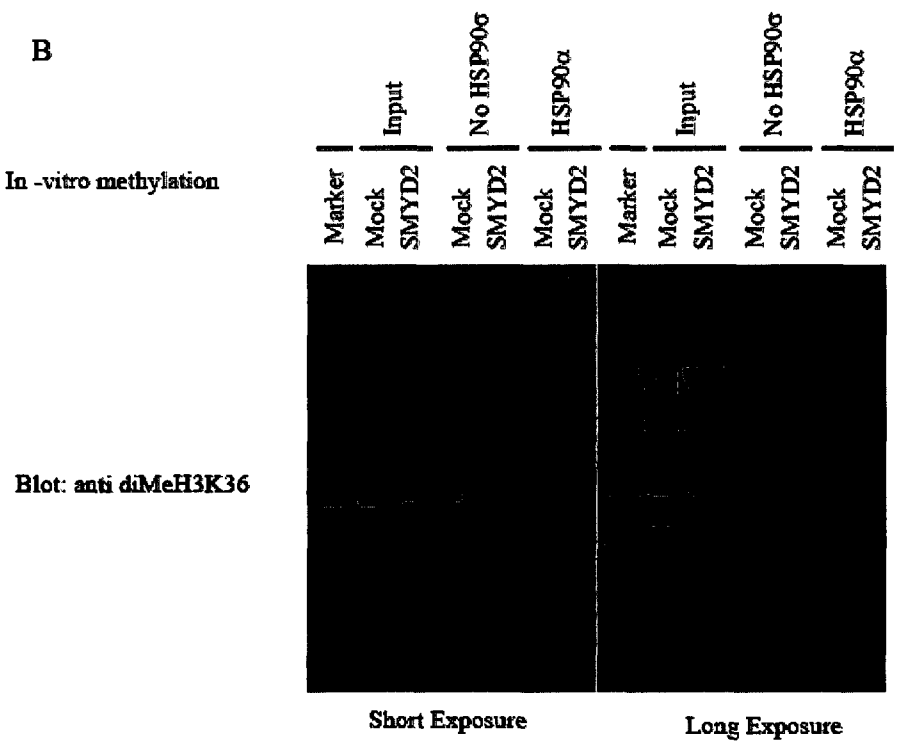

B
In -vitro methylation
Mock
SMYD2
Blot: anti-mono, di, tri methylated H3K4
FLAG IP
Input
Blot: anti diMeH3K36
FLAG IP
Input
Blot: anti FLAG
Input

Figure 11: Effect of HSP90α on SMYD2 methylation specificity. **A:** *In vitro* methylation assay showing that SMYD2 methylation of H3K4 is dependent on the presence of HSP90α. In the absence of HSP90α, no K4 methylation is observed even after long exposure of the film. **B:** SMYD2 methylates H3K36 *in vitro* very weakly in the absence of HSP90α. This activity is weak and can only be detected after long exposure of the film. No methylation of H3K36 is observed in the presence of HSP90α even after long exposure of the film.


A
In -vitro methylation
Blot: anti-mono, di, tri methylated H3K4
Marker
Input
Mock
SMYD2
No HSP90α
HSP90α
Short Exposure
Long Exposure

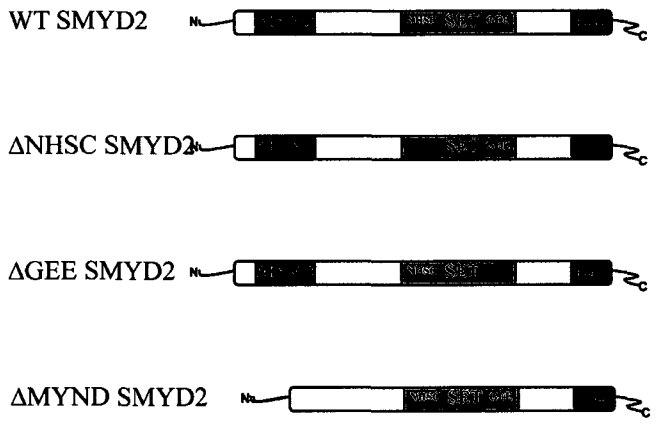

B
In -vitro methylation
Blot: anti diMeH3K36
Marker
Input
Mock
SMYD2
No HSP90α
HSP90α
Short Exposure
Long Exposure

3.5 Role of SMYD2 SET Domain in its Methyltransferase Activity.

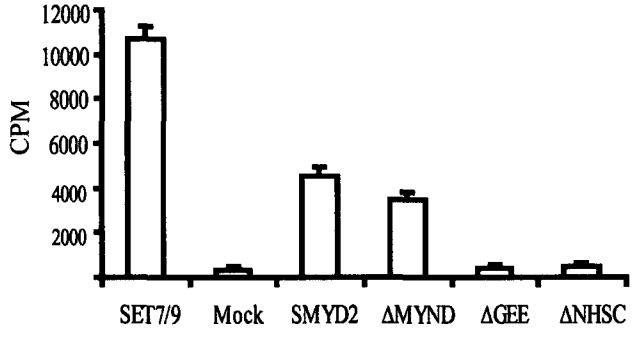
The SET domain is a known histone lysine methyltransferase domain. The presence of this domain in SMYD2 indicates a potential role for this protein as a histone methyltransferase. Two homologues of SMYD2 (SMYD3 and SMYD1) have methyltransferase activity for H3K4 (Hamamoto et al., 2004)(Tan et al., 2006). SMYD3 and SUV39H1 have two conserved regions in its SET domain ΔNHSCXPN and ΔGEEXXXXY to which SAM binds (Hamamoto et al., 2004). These two regions are also present in the SET domain of SMYD2. In order to investigate the role of the SET domain and the SAM binding motif in the SET domain, these two sites (ΔNHSC and ΔGEE) were mutated (Figure 12A). The wild type SMYD2 and its mutants were transfected into HEK293T cells, immunoprecipitated using anti-FLAG antibodies, and used for methyltransferase assays in the presence of tritium ^{3}H-SAM and HSP90α. The methylation activity was measured using a scintillation counter as previously described (Hamamoto et al., 2004). Wild type and the MYND deletion mutant showed respectively a 10 fold and a 7 fold increase in methylation activities compared to the mock control. Purified human SET7/9 protein was used as a positive control (Figure 12B). Furthermore, the deletion of either the SAM binding consensus sequences ΔGEE or ΔNHSC resulted in a sharp reduction of SMYD2 methylation activity which indicated that both SAM binding domain sequences were critical for SMYD2 methyltransferase activity (Figure 12B). These results were further confirmed following an *in vitro* methylation assay using mono-, di-, tri-methylated H3K4 antibodies. Wild type and a MYND deletion mutant of SMYD2 were found to methylate H3K4 while deleting GEE or NHSC sequences rendered SMYD2 inactive.

Figure 12: Role of SMYD2 SET domain in its methyltransferase activity. **A:** Diagram showing the structure of the wild type form of SMYD2 as well as the mutants introduced to its MYND domain and to the two conserved sequences in the SET domain ΔNHSC and ΔGEE. **B:** *In vitro* methylation activity of SMYD2 and its mutants showing the decrease in its activity after point mutations in the two conserved sequences of the SET domain. Deletion of the MYND domain does not seem to affect the histone methylation activity of SMYD2. **C:** Methylation of H3K4 by wild type SMYD2 as detected by mono-di and tri-methylation H3K4 specific antibodies. SET domain mutations obliterate the methylation activity of SMYD2. The MYND domain deletion does not have any effect on the methylation activity of SMYD2. Histone H3 is used as a loading control. Coomassie stain SDS-PAGE gel showing equal amount of HSP90α used for the different methylation reactions except for the positive control (Set7/9).

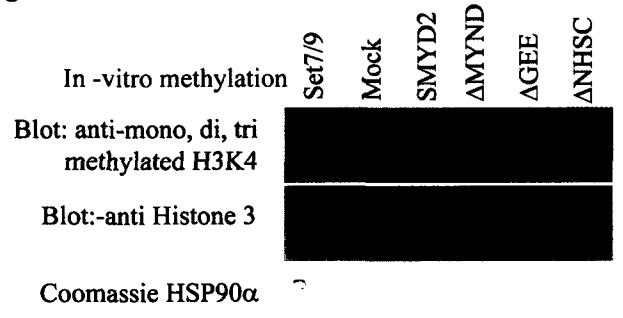
A


WT SMYD2
ΔNHSC SMYD2
ΔGEE SMYD2
ΔMYND SMYD2

B


12000
10000
8000
6000
4000
2000
CPM
SET7/9
Mock
SMYD2
ΔMYND
ΔGEE
ΔNHSC

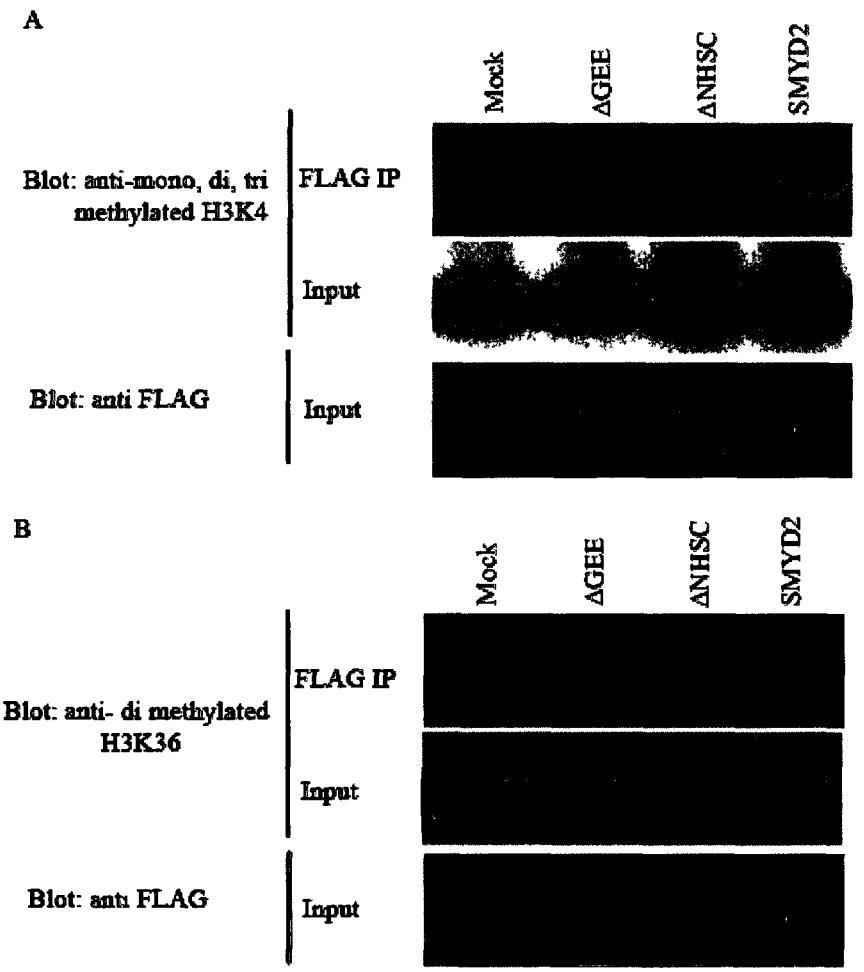
C


In -vitro methylation
Set7/9
Mock
SMYD2
ΔMYND
ΔGEE
ΔNHSC
Blot: anti-mono, di, tri methylated H3K4
Blot:-anti Histone 3
Coomassie HSP90α

3.6 Modified Chromatin IP Reveals Increased H3K4 Methylation *in vivo* in the Vicinity of SMYD2 Binding to DNA.

In order to further validate our *in vitro* methylation assay, we performed a chromatin IP experiment to look at histone modifications in the vicinity of DNA bound by SMYD2. However, instead of following the chromatin IP by gene chips, we performed gel electrophoresis and western blot analysis to look for the enrichment of lysine methylation on histone H3. Briefly, Wild type SMYD2, ΔNHSC, and ΔGEE deletion mutants were individually over-expressed in HEK293T cells, a cross-linker was used to fix the interactions, the cell lysate was sonicated to reduce the size of the DNA, and anti-FLAG antibodies was used to purify SMYD2 and its associated proteins. Immunoblot analyses using antibodies directed against either mono-, di- tri-methylated H3K4 or di-methylated H3K36 revealed that chromatin IP against wild type SMYD2 specifically enriches methylated H3K4 but not dimethylated H3K36 (Figure 13A& B). In contrast, methylated H3K36 was not detected in SMYD2 methylation even after long exposure. Furthermore, chromatin IP performed with the two SET domain deletion mutants showed no enrichment for H3K4 (Figure 13A).

Figure 13: Enrichment of H3K4 methylation as a result of SMYD2 over-expression by Chromatin IP. *In vivo* methylation activity was assessed by performing chromatin IP of over-expressed SMYD2, SMYD2 mutants, and a mock control with a subsequent analysis by western blot to measure the level of enrichment of **(A)** H3K4 methylation and **(B)** H3K36 methylation **top panel**. The **middle panels** in A and B are the western blot of the input whole cell lysate probed for **(A)** H3K4 methylation and **(B)** H3K36 methylation. FLAG IP refers to immunoprecipitated proteins using anti-FLAG antibodies from HEK293T cells. The bottom panels are western blot analysis of input whole cell lysate (input) for the presence of FLAG-tag proteins.


A
Mock
ΔGEE
ΔNHSC
SMYD2
Blot: anti-mono, di, tri methylated H3K4
FLAG IP
Input
Blot: anti FLAG
Input
B
Mock
ΔGEE
ΔNHSC
SMYD2
Blot: anti- di methylated H3K36
FLAG IP
Input
Blot: anti FLAG
Input

3.7 ***In vitro* and *In vivo* HSP90α Methylation by SMYD2**

Identification of lysine methylation sites on non-histone proteins has become more common in the last few years with an increasing number of proteins shown to be modified by methylation. A recent example is a paper published on the methylation of E2F1 by Set9 which was shown to play a role in regulating DNA damage response in p53 deficient cells (Kontaki and Talianidis, 2010). This methylation site was also removed by LSD1, which is important for E2F1 stabilization and apoptotic function (Kontaki and Talianidis, 2010). Even though HSP90α interaction with SMYD2 was not mediated by the two SET domain consensus sequences, I was interested in investigating whether HSP90α was as a substrate for SMYD2. This was possible with the use of MS by simply configuring the search engine to look for mono, di and tri-methylation on lysine residues. Two sites were identified as potential methylation sites by MS analysis. These two sites were identified from *in vivo* and *in vitro* assay. The *in vivo* assay involves the purification of FLAG SMYD2 protein from HEK293T cells. This was followed by separation on an SDS-gel and in gel digestion before analysis on MS. MS data was then searched by Mascot with methylation as a variable modification. On the other hand, the *in vitro* assay involves performing a methylation assay using recombinant SMYD2 and HSP90α in the presence of SAM and then trypsin digestion followed by MS analysis. Two sites were identified from both experiments, those two sites were K209 and K294 (Figure 14). Methylation of HSP90α was further validated by probing wild type SMYD2 IP and one of the SET domain inactive mutants for methylation using an antibody that recognizes mono, di and tri-methylated lysine residues. As shown in Figure 14B HSP90α was methylated in the wild type SMYD2 but not the mutant, indicating that

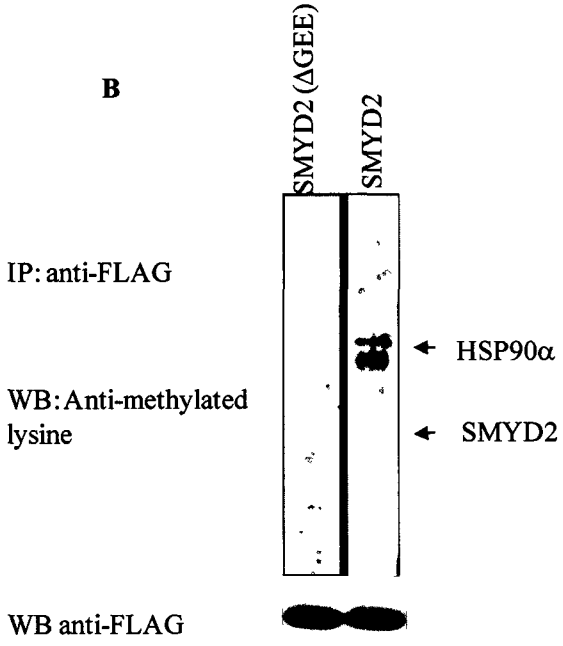
SMYD2 was responsible for this methylation event. This figure also shows other potential methylation sites on SMYD2 itself as well as other proteins.

Figure 14: HSP90α methylation. **A**: Identification of HSP90α methylation sites as identified by MS from *in vitro* and *in vivo* experiments. **B:** A western blot showing an increase in the methylation of HSP90α in the wild type anti-FLAG -SMYD2 immunoprecipitation compared to the SET domain ΔGEE mutant.

A

Protein	Tryptic Peptide	Methylation	Mascot Score
HSP90α	K.T<u>K</u>PIWTRNPDDITNEEYGEFYK.S	Trimethylation	49
	K.<u>K</u>HSQFIGYPITLFVEK.E	Mono, di and trimethylated	48

B



3.8 Confirmation of HSP90α Methylation Sites

Methylation of K209 was confirmed by site directed mutagenesis of the lysine residue to arginine. Mutagenesis of this site showed a reduction in the level of HSP90α methylation. On the other hand, mutation of the other lysine residue on K294 did not decrease the level of methylation of HSP90α. Many attempts were made to identify the other sites on HSP90α including the use of a computer model that looks at the structure of HSP90α and the availability of certain lysine residues to be methylated. One of the sites that were predicted to be methylated from the structure was K565. Methylation of this site was confirmed by synthesizing a peptide corresponding to that site on HSP90α. When used in an in vitro assay, this peptide was found to be methylated by MS and later confirmed by site directed mutagenesis to be the second site of SMYD2 lysine methylation on HSP90α. Figure 15 shows the effects of mutating both single sites on HSP90α as well as the affect of mutating both sites on HSP90α.

3.9 Demethylation of HSP90α by LSD1

The discovery of lysine demethylation enzymes has shed the light on the dynamic nature of methylation events. One of those demethylation enzymes is LSD1. LSD1 was shown to demethylate p53 at K370, a site that is mono-methylated by SMYD2 (Huang et al., 2007). To test for the ability of LSD1 to demethylate HSP90α, HSP90α was first methylated by SMYD2 in presence of radio-labeled SAM for two hours. The methylated protein was incubated with an increasing concentration of LSD1 as shown in Figure 16. LSD1 was able to demethylate HSP90α in a concentration dependent fashion.

Figure 15: HSP90α methylation sites. Methylation assay using radioactive SAM as substrate for the SMYD2 methylation reaction in presence of wild type HSP90α and the two single mutants of lysine residues 209 and 565 that were changed to arginine. A double mutant was also generated and used in the methylation reaction. Deletion of both sites reduced methylation activity to a background level. Experiment was performed by Sylvain Lanouette from Dr. Jean-Francois Couture's laboratory.

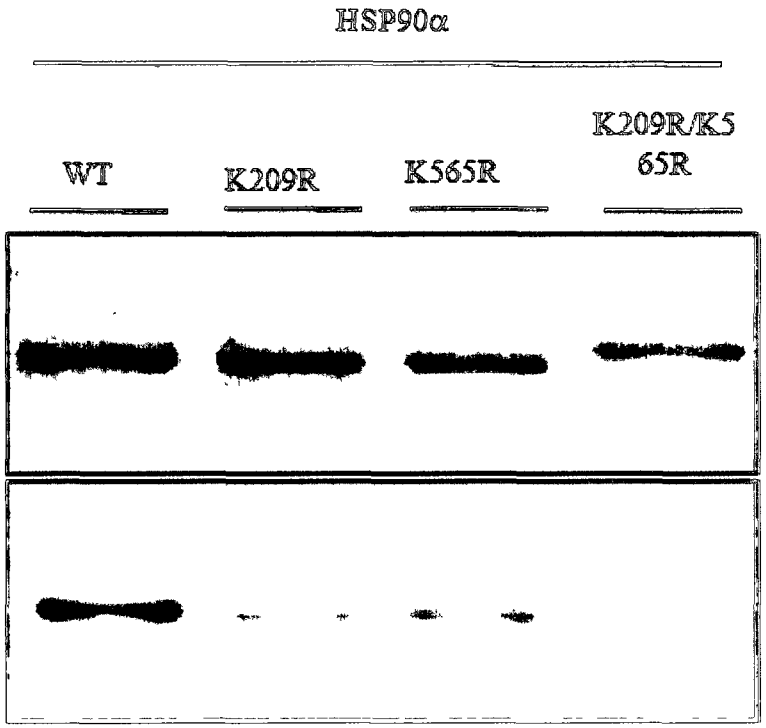
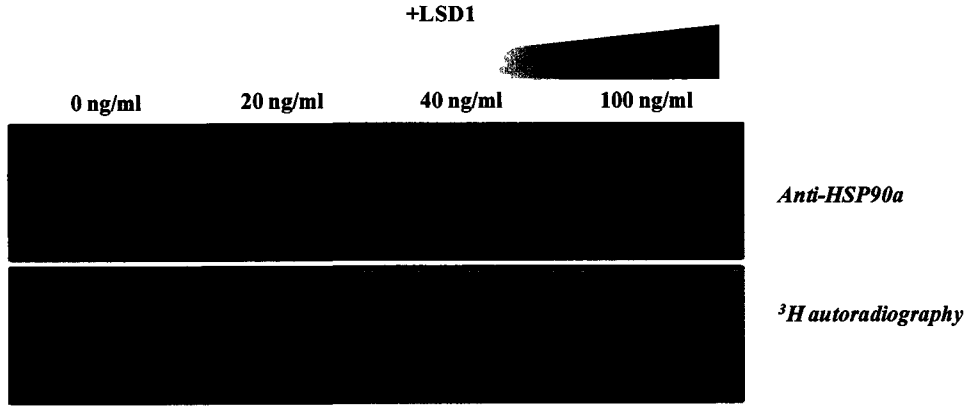

HSP90α
WT
K209R
K565R
K209R/K565R

Figure 16: Demethylation of HSP90α by LSD1. In this experiment HSP90α was methylated *in vitro* using radioactive SAM as a substrate for the SMYD2 methylation reaction. Methylated HSP90α was incubated with an increasing amount of LSD1 to test its ability to reduce the methylation on HSP90α. LSD1 was found to diminish the radioactive signal produced by methylated HSP90α. Reactions were normalized to the amount of HSP90α loaded in each well by plotting against HSP90α. Experiment was performed by Sylvain Lanouette from Dr. Jean-Francois Couture's laboratory.

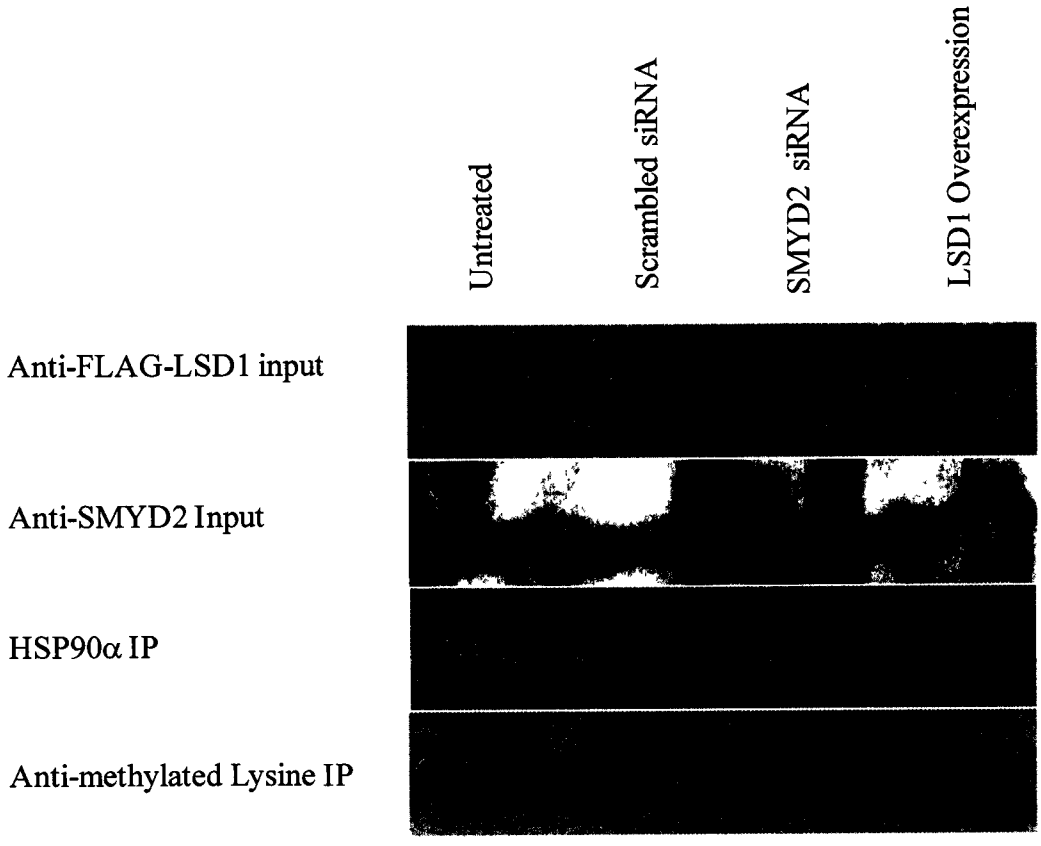

+LSD1
0 ng/ml
20 ng/ml
40 ng/ml
100 ng/ml
Anti-HSP90a
^{3}H autoradiography

3.10 *In vivo* Analysis of SMYD2 Knockdown on HSP90α Methylation

In order to confirm the responsibility of SMYD2 for methylation of HSP90α, an *in vivo* experiment that involves knocking down SMYD2 was performed. The cell model used for this experiment was HepG2 cells, which has a high SMYD2 expression level. HepG2 cells were transfected with siRNA against SMYD2. To check for the efficiency of SMYD2 siRNA transfection SMYD2 level was tested using an anti-SMYD2 polyclonal antibody. SMYD2 level was reduced to about 50% of the untreated sample. Scrambled siRNA was used as a negative control to account for the affect of the RNA on the level of HSP90α methylation. Level of methylation on HSP90α was assessed by immunoprecipitating HSP90α and measuring changes in methylation by an antibody that recognizes mono, di and tri-methylated lysine. Level of HSP90α methylation was reduced as shown in Figure 17 as a result of SMYD2 siRNA indicating that SMYD2 is responsible for the methylation of HSP90α in HepG2 cells. Level of methylation was not affected in the untreated and the negative control. HSP90α level was used as a loading control to make sure that the same amount was loaded in each lane.

A similar approach was used to verify the role of LSD1 in demethylating HSP90α *in vivo*. In this approach, LSD1 was over expressed in HepG2 cells and its effects on HSP90α methylation was assessed using the same approach mentioned above. Figure 17 shows the reduction in methylation of HSP90α as a result of LSD1 over expression. This indicates that LSD1 can reduce methylation on HSP90α but not remove all sites of methylation. The fact that LSD1 does not remove all methylation on HSP90α suggests that HSP90α might be methylated by another enzyme since LSD1 was capable of removing all methylation sites on HSP90α *in vitro*.

Figure 17: SMYD2 knockdown in HepG2 cells reduces the level of methylation on HSP90α. Methylation levels of HSP90α under different conditions were assessed by immunoprecipitating HSP90α under different conditions and then probing for lysine methylation. Cells were treated with the siRNA against SMYD2 and a scrambled siRNA and transfection reagent alone as negative controls. SiRNA of SMYD2 led to a decrease in the level of methylation on HSP90α as compared to the negative controls. Methylation level of HSP90a in the presence of LSD1 was assessed in a similar manner. Over expression of LSD1 has led to the reduction of methylation level on HSP90α methylation.


Untreated
Scrambled siRNA
SMYD2 siRNA
LSD1 Overexpression
Anti-FLAG-LSD1 input
Anti-SMYD2 Input
HSP90α IP
Anti-methylated Lysine IP

3.11 SMYDs Phylogenic Tree and Sequence Alignment

In order to investigate the ability of other SMYD proteins to methylate HSP90α, sequence alignment and evolutionary history of the different SMYDs was investigated. Sequence alignment of SMYD2 with other SMYD family members revealed that SMYD2 shares 31% sequence identity and 52% sequence similarity with SMYD3 which is another member of the SMYD family. The sequence homology of SMYD2 and SMYD3 is primarily due to the MYND and SET domains. Figure 18A shows the alignment of the SET and MYND domain of the five different proteins of the SMYD family. The alignment and the phylogeny tree in figure 18B shows the close relationship between SMYD1, 2 and 3 while the other two SMYDs are more distant.

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The evolutionary distances were computed using the Poisson correction method (Tamura et al., 2007) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 387 positions in the final dataset. Phylogenic analyses were conducted in MEGA4 (Tamura et al., 2007). The phylogenic tree in Figure 18 shows that the most related SMYDs are SMYD2 and SMYD3 followed by SMYD1. SMYD5 is most distant family member. This tree was used as a guide to understand the evolutionary history of different SMYDs.

Figure 18: Sequence alignment and the evolutionary relationship between the different SMYD family members. **A:** ClustalW alignment of the MYND and SET domain residues of the SMYD proteins. **B:** This phylogeny tree shows a closer relation between SMYD1, 2 and 3. SMYD5 is the most distant member of the family.

A

MYND Domain

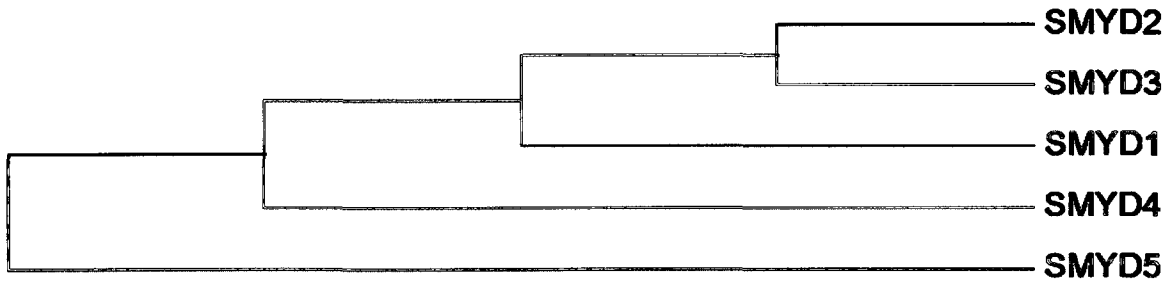
```
                    10        20        30
SMYD1 1 CHTCFKRQEKLHRCGQCKFAHYCDRTCQKDAWLN-HKNEC  39
SMYD2 1 CEYCFTRKEGLSKCGRCKQAFYCNVECQKEDWPM-HKLEC  39
SMYD3 1 CDRCLLQKEKLMRCSQCRVAKYCSAKCQKKAWPD-HKREC  39
SMYD4 1 CHRCLKHTLATVPCDGCSYAKYCSQECLQQAWELYHRTEC  40
SMYD5 1 PELCTVRKDLHQNCPHCQ-VMYCSAECRLAATEQYHQVLC  39
```

SET Domain

```
                    10        20        30        40        50        60
SMYD1  1 -LVSVDDLQNHVEHFGEEEQKDLRVDVDTFLQYWPPQSQQFSMQY----ISHIFGVINCNGFTLSDQRGL 65
SMYD3  1 ----FYDLESNINKLTEDKKEGLRQLVMTFQHFMREEIQDASQLPPAFDLFEAFAKVICNSFTICN-AEM 65
SMYD2  1 KLLAVKEFESHLDKLDNEKKDLIQSDIAALHHFYSKHLGFPDNDS----LVVLFAQVNCNGFTIED-EEL 65
SMYD5  1 -WFTPDGFRSLFALVGTNGQGIGTSSLSQWVHACDTLELKPQDRE----QLDAFIDQLYKDIEAATGEFL 65
SMYD4  1 -QAIPTERIVNSSQLKAAVTPELCPDVTIWGVAMLRHMLQLQCNAQ---AMTTIQHTGPKGSIVTD-SRQ 65

                    80        90        100       110       120       130
SMYD1 66 QAVGVGIFPNLGLVNHDCWPNCTVIFNNGNHEAVKSMFHTQMRIELRALGKISEGEELTVSYIDFLNVS 134
SMYD3 66 QEVGVGLYPSISLLNHSCDPNCSIVFNG----------PHLLLRAVRDIEVGEELTICYLDMLMTS 121
SMYD2 66 SHLGSAIFPDVALMNHSCCPNVIVTYKG----------TLAEVRAVQEIKPGEEVFTSYIDLLYPT 121
SMYD5 66 NCEGSGLFVLQSCCNHSCVPNAETSFPENN----------FLLHVTALEDIKPGEEICISYLDCCQ-- 121
SMYD4 66 VRLATGIFPVISLLNHSCSPNTSVSFIS----------TVATIRASQRIRKGQEILHCYGPHKS-- 119
```

B



3.12 HSP90α Interaction with SMYDs

Using anti-FLAG agarose beads SMYD2, SMYD3 and SMYD5 were immunoprecipitated and visualized by silver staining. Figure 19A shows the IP data of the three SMYDs compared to the mock vector control. After in gel digestion and identification of proteins interacting with the various SMYDs, HSP90α was shown to interact with SMYD2 and SMYD3. No interaction was observed with SMYD5. These data were analyzed by MS at an n=3. Validation of these interactions was done by co-IP of various SMYDs and then HSP90α was detected with an anti-HSP90α antibody. Figure 19B shows interaction between HSP90α and SMYD2 and 3 but not SMYD5, which confirms the AP-MS data. The reciprocal immunoprecipitation with HSP90α antibodies was performed to validate these interactions. When probed with anti-FLAG antibodies only SMYD2 and SMYD3 were detected. No signal was observed for SMYD5 confirming the results obtained earlier. In conclusion, Figure 19 shows the interaction between HSP90α and SMYD2 and SMYD3 but not SMYD5. GAPDH was used to ensure equal loading of different samples.

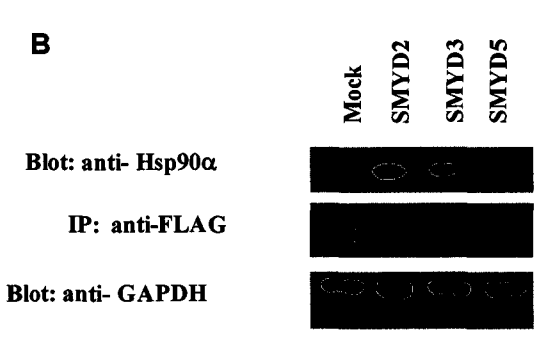
Since HSP90α was shown to interact with SMYD2 as well as SMYD3, I was interested in finding whether SMYD3 has the ability to methylate HSP90α. In order to do that, its methylation status was assessed both *in vivo* and *in vitro*. As discussed earlier the *in vivo* assay involves the IP of FLAG SMYD2 proteins from HEK293T cells and then search for methylation on HSP90α. On the other hand the *in vivo* assay involves performing a methylation assay using recombinant SMYD2 and HSP90α. Our data shows that HSP90α was not methylated by SMYD3 and SMYD5. Only SMYD2 was able to methylate HSP90α *in vivo* and *in vitro* data not shown.

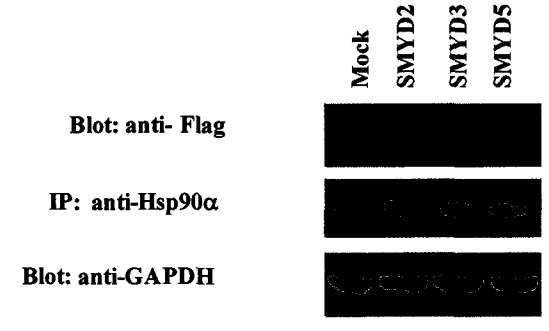
Figure 19: HSP90α interaction with various SMYD proteins. **A**: MS data on the interaction between SMYD2, SMYD3 and SMYD5 showing that only SMYD2 and 3 interacts with HSP90α. **B**: The reciprocal IP using HSP90α antibodies was used to validate the interaction. When probed with anti-FLAG antibodies only SMYD2 and SMYD3 were detected. No signal was observed for SMYD5 confirming that data obtained earlier. GAPDH was used to ensure equal loading of different samples.

A

	Repeats	HSP90 α Mascot Score	HSP90 α # Peptide	HSP90 α % Coverage
SMYD3 IP	N1	2109	32	40
	N2	1950	37	48
	N3	2485	42	51
SMYD2 IP	N1	2709	39	48
	N2	3102	43	53
	N3	2505	35	42
SMYD5 IP	N1	NA	NA	NA
	N2	NA	NA	NA
	N3	NA	NA	NA

B



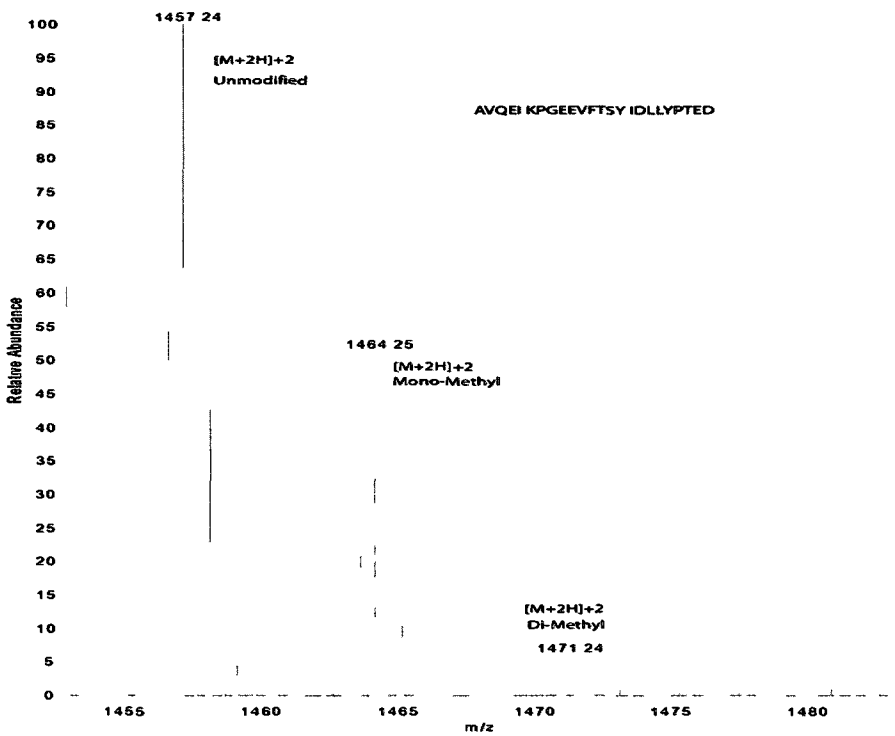


3.13 SMYD2 Automethylation

Expanding the substrate specificity of methyltransferases beyond histone proteins has led to believe that these proteins might be capable of methylating themselves. One such example of automethylation is G9a which was known to methylate H3 at K9 (Tachibana et al., 2002) (Sampath et al., 2007). G9a was found to be capable of automethylating an ARKT motif in its sequence similar to its ARKS recognition sequence on H3 (Chin et al., 2007). Automethylation of G9a changes its binding affinity to HP1 class proteins creating a binding site for these proteins. On the other hand mutation of this site leads to loss of interaction between HP1 and G9a (Chin et al., 2007). In an attempt to identify methylation sites on SMYD2 MS data of the IP experiments were searched. Searching the MS data lead to the identification of a methylation site on SMYD2 itself. This methylation site is located close to the conserved GEE site that is part of the SET domain. MS data showed that this lysine methylation is mostly mono and dimethylated. Figure 20A shows the precursor ion scan of the three different forms of the SMYD2 peptide that is methylated. Figure 20B shows the methylated peptides with their mascot score and level of methylation. This methylation event is believed to be an automethylation of SMYD2. In order to confirm this hypothesis an *in vitro* experiment was performed to look at SMYD2's ability to methylate itself. *In vitro* methylation data confirmed that SMYD2 was automethylated in a concentration dependent manner. A higher concentration of SMYD2 leads to inhibition of this automethylation reaction. This indicated that this automethylation event might act as a regulator of SMYD2 methylation activity. This can be very important for the function of SMYD2 since it's over expression of SMYD2 might inhibit its ability to methylate key sites on HSP90α or other proteins that are important in the cell.

Figure 20: Identification of SMYD2 automethylation sites by MS. **A:** an MS spectrum showing unmethylated and methylated SMYD2 peptide. This peptide was identified after digesting SMYD2 protein purified from HEK293T cells transfected with a FLAG tagged SMYD2. The SMYD2 protein was purified by agarose beads and the separated on SDS-PAGE gel before its processing for MS analysis. Gel band corresponding to SMYD2 were digested with trypsin and analyzed on the Orbitrap to identify sites of methylation *in vitro*. MS spectra of SMYD2 peptides showing the methylation sites on SMYD2 that are mono and di-methylated. **B**: The sequence of the SMYD2 peptide that is methylated by SMYD2 with its level of methylation and the mascot score.

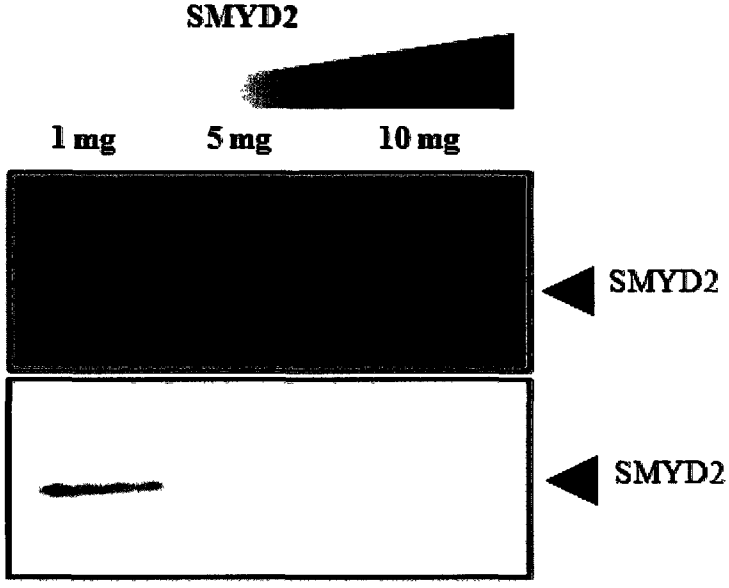
A



B

Protein	Peptide	PTM	Residue	Score
SMYD2	AVQEIKPGEEVFTSYIDLLYPTEDRNDR	Mono-Me	226-250	60
SMYD2	AVQEIKPGEEVFTSYIDLLYPTEDRNDR	Di-Me	226-250	40,35

Figure 21: SMYD2 Automethylation. An *in vitro* methylation assay using an increasing concentration of recombinant SMYD2 that was incubated with radioactive SAM. This experiment shows the automethylation activity of SMYD2 in the presence of a methyl donor. Increasing the concentration of recombinant SMYD2 has led to a reduction in the level of SMYD2 automethylation. This indicates that SMYD2 automethylation is auto-inhibited by SMYD2. Experiment was performed by Sylvain Lanouette from Dr. Jean-Francois Couture's laboratory.


SMYD2
1 mg
5 mg
10 mg
SMYD2
SMYD2

3.14 SMYD2 Role in Gene Expression Regulation

Histone methylation has been associated with the regulation of gene expression. In particular, methylation at H3K4 is generally associated with the activation of gene expression (Lachner and Jenuwein, 2002) (Sims et al., 2003). We have studied the impact of the over-expression of SMYD2 on gene expression levels. Briefly, wild type SMYD2 was over-expressed in HEK293T cells, and the total RNA extracted was compared to RNA extracted from cells transfected with a mock vector using microarrays (see experimental section). A total of 41 genes showed statistically significant changes in expression according to the SAM software (Tusher et al., 2001); 37 genes were up-regulated, while only 4 genes were down regulated following the over expression of SMYD2 (Table 2). Validation of microarray data was done by semi-quantitative RT-PCR for six of the differentially expressed genes TACC2, WDR9, AKAP13, Bat1, CHD9 and SMARCA2 (Figure 23). As shown in Supplementary Table 3, some of these genes function in chromatin remodeling as well as in the cell cycle and transcription regulation. Our data suggests that SMYD2 over-expression will lead to the overall activation of gene expression in agreement with the observed H3K4 methylation activity of SMYD2.

We proposed that SMYD2, similar to other SMYD family members, regulates gene expression by directly binding to specific DNA consensus sites and/or indirectly through protein-protein interactions with co-factors. In order to investigate the binding of SMYD2 to regulatory regions of its target genes, the promoter region of TACC2 was used for chromatin IP experiments. Sequence analyses revealed that the promoter region of TACC2 contains the CCCTCC motif, which is a binding site for SMYD3 and is localized at the chip 3 region (Figure 22A). The promoter of the TACC2 gene spanning the region -1279 to +130, with

respect to the transcription start site, was divided into 4 regions: chip1 (-1279 to -949 nt), chip 2 (-949 to -629 nt), chip 3 (-629 to -200 nt) and chip 4 (-200 to +130 nt). Specific primers for each region were used to identify the region to which SMYD2 specifically binds (Figure 22B). Chromatin IP followed by PCR amplification demonstrated that SMYD2 binds to the TACC2 promoter sequence in the region spanned by chip 4 (Figure 22C). IP with a mock antibody (mouse IgG) was used as a negative control to show that the PCR product obtained in this experiment was specific to the presence of SMYD2. Our data indicates that SMYD2 binds to the promoter region of TACC2 at a site that is different from the site recognized by SMYD3. We then tested if SMYD2 regulated TACC2 activity by its histone methyltransferase activity or indirectly through a different mechanism. Briefly, HEK293T cells were transfected with wild type SMYD2, ΔGEE, or ΔNHSC deletion mutants as well as with the mock control. Gene expression of TACC2 was measured by semi-quantitative RT-PCR on RNA extracted from the transfected cells. As shown in Figure 22D, only the wild type form had the ability to increase TACC2 gene expression. The two histone methyltransferase-inactive mutants and the mock control did not increase the gene expression of TACC2. This data indicates that the methylation activity of SMYD2 affected gene expression of TACC2.

Figure 22: SMYD2 affects gene expression and binds the promoter of TACC2. **A:** Pie chart showing the classification of the genes with altered expression following over-expression of SMYD2 in HEK293T cells after 24 h of transient transfection compared to cells transfected with a mock vector. Genes were classified into four categories according to their function: transcription, chromatin remodeling, cell cycle, and other. **B:** Promoter sequence of TACC2 gene showing the location of primers used in the ChIP assay and the TSS (transcription start site). **C:** PCR results of the chromatin IP showing a PCR product for Chip 4 primers only when anti-FLAG-SMYD2 antibodies were used, indicating that SMYD2 can bind to the TACC2 promoter. The last panel shows PCR amplification products for the input DNA as a PCR amplification control for the primers. **D:** Semi-quantitative RT-PCR analysis of the level of expression of TACC2 transcripts following the over-expression of the wild type SMYD2, its SET domain deletion mutants, and a mock control in HEK293T cells. The middle panel shows the level of expression of wild type SMYD2 and its SET domain deletion mutants by Semi-quantitative RT-PCR. GAPDH was used as an internal control.

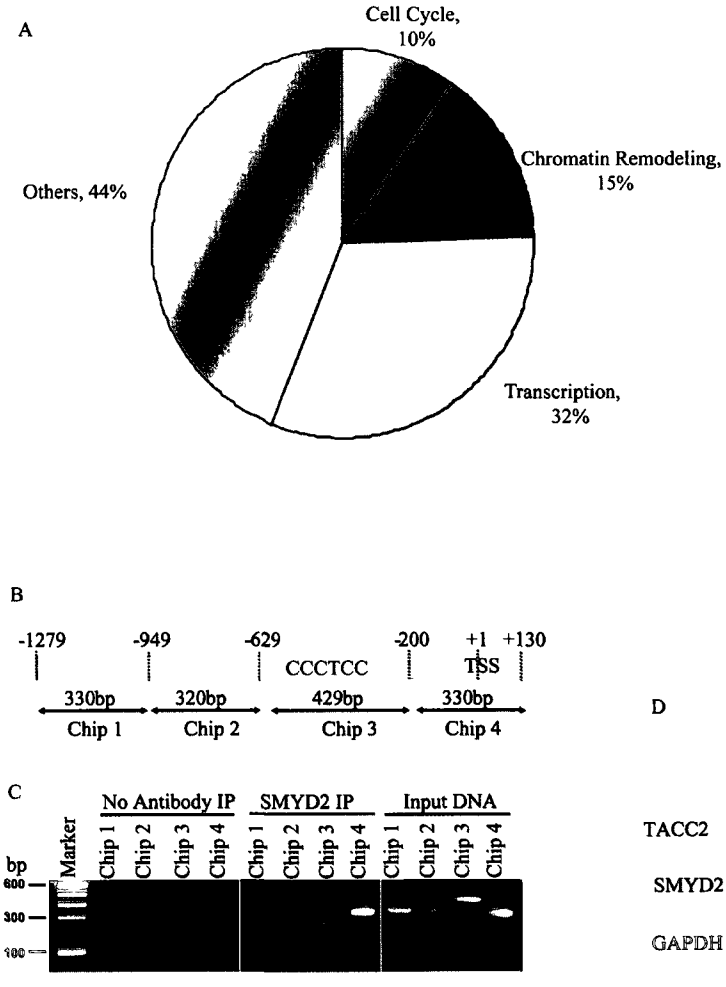

A
Cell Cycle, 10%
Chromatin Remodeling, 15%
Transcription, 32%
Others, 44%
B
-1279
-949
-629
CCCTCC
-200
+1
TSS
+130
330bp
320bp
429bp
330bp
Chip 1
Chip 2
Chip 3
Chip 4
C
No Antibody IP
SMYD2 IP
Input DNA
Marker
Chip 1
Chip 2
Chip 3
Chip 4
bp
600
300
100
D
Mock
SMYD2
NHSC
GEE
TACC2
SMYD2
GAPDH

Table 2: List of proteins differentially regulated by SMYD2 over-expression in HEK293T cells as measured by cDNA microarray.

Upregulated Genes

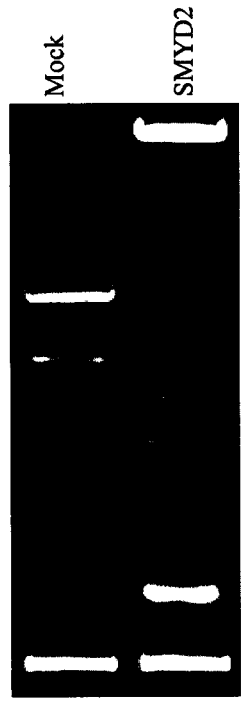
UGCluster	function	Name
Hs.100426	Cell cycle	Breast cancer metastasis suppressor 1
Hs.183994	Cell Cycle	Protein phosphatase 1, catalytic subunit, alpha isoform
Hs.145925	Cell cycle	Testis specific A2 homolog (mouse)
Hs.459211	cell cycle	A kinase (PRKA) anchor protein 13
Hs.59159	Chromatin Remodeling	Chromodomain helicase DNA binding protein 9
Hs.298990	Chromatin Remodeling	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2
Hs.444445	Chromatin Remodeling	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3
Hs.463010	Chromatin Remodeling	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1
Hs.501252	Chromatin Remodeling	Transforming, acidic coiled-coil containing protein 2
Hs.491494	Others	Chaperonin containing TCP1, subunit 3 (gamma)
Hs.339809	Others	Copine II
Hs.183850	Others	DCMP deaminase
Hs.302754	Others	EF-hand calcium binding protein 1
Hs.436298	Others	Epithelial membrane protein 1
Hs.41296	Others	Fibronectin leucine rich transmembrane protein 3
Hs.148266	Others	Glycerol-3-phosphate dehydrogenase 2 (mitochondrial)
Hs.567242	Others	Immunoglobulin lambda joining 3
Hs.413513	Others	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta
Hs.274313	Others	Insulin-like growth factor binding protein 6
Hs.507971	Others	Leucine-rich repeats and calponin homology (CH) domain containing 1
Hs.2256	Others	Matrix metallopeptidase 7 (matrilysin, uterine)
Hs.513071	Others	Mesoderm development candidate 1
Hs.235935	Others	Nephroblastoma overexpressed gene
Hs.379970	Others	Ras association (RalGDS/AF-6) domain family 2
Hs.49407	Others	Regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)
Hs.532768	Others	Serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1
Hs.162241	Others	Ubiquitin carboxyl-terminal esterase L3 (ubiquitin thiolesterase)
Hs.22109	Transcription	Bromo adjacent homology domain containing 1
Hs.547443	Transcription	Chromosome 15 open reading frame 20
Hs.388297	Transcription	Chromosome 8 open reading frame 36
Hs.118651	Transcription	Hematopoietically expressed homeobox
Hs.444409	Transcription	MADS box transcription enhancer factor 2, polypeptide C (myocyte enhancer factor 2C)
Hs.418198	Transcription	PAP associated domain containing 4
Hs.558398	Transcription	Sine oculis homeobox homolog 1 (Drosophila)
Hs.369063	Transcription	Zic family member 2 (odd-paired homolog, Drosophila)
Hs.469601	Transcription	Zinc finger protein 586
Hs.521064	Transcription	Zinc finger protein 655
Hs.513292	Transcription	Zinc finger protein 75a

Downregulated Genes

Hs.525324	Cell Cycle	Cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)
Hs.254042	Transcription	HLA-B associated transcript 1
Hs.534052	Transcription	Zinc finger protein 36, C3H type, homolog (mouse)
Hs.314338	Cghromatin Remodeling	Bromodomain and WD repeat domain containing 1

Figure 23: Validation of microarray data of six selected genes from the list of genes differentially regulated by SMYD2 over-expression in HEK293T cells. Gene expression is measured by semi-quantitative RT-PCR. Gene expression is normalized to GAPDH gene expression.

UniGene ID	Gene Name
Hs. 66170	SMYD2
Hs. 66170	TACC2
Hs.314338	WDR9
Hs.254042	Bat1
Hs.59159	CHD9
Hs.459211	AKAP13
Hs.298990	SMARCA2
Hs.169474	GAPDH



4.0 DISCUSSION

4.1 Interaction Mapping of SMYD2

Our knowledge about the SMYD protein family and SMYD2 in particular has been limited. At the beginning of my thesis there was no paper published on SMYD2 and the only information known about SMYD2 was the fact that it has two evolutionary conserved domains (MYND and SET domains) as well as a potential role in cancer. As a result, I have initiated a proteomics and genomics study of SMYD2 to enhance our understanding of its molecular roles. Our mapping of the interactome of SMYD2 clearly points to different molecular roles for SMYD2 and that the MYND and SET domains are critical to many of these interactions. In particular, we have found that five of the interaction partners of SMYD2 contain the PXLXP motif. It was previously demonstrated that the MYND domain of the human protein BS69 binds to proteins that contain the motif PXLXP (Ansieau and Leutz, 2002). This is also the case for SMYD1 where the PXLXP motif in skNAC is required for their interaction (Sims et al., 2002). Hence, we hypothesized that SMYD2 interactors that contain the PXLXP motif, such as EBP41L3, are likely to interact with the MYND domain of SMYD2. EBP41L3, also known as DAL-1, has been implicated in the molecular pathogenesis of breast, lung, and brain cancers and it is thought to be a tumor suppressor (Singh et al., 2004). Our co-IP data supports the view that the interaction between SMYD2 and EBP41L3 is mediated by the MYND domain through the PXLXP motif. It is likely that the seven other proteins that have the PXLXP motif also interact with the MYND domain of SMYD2. In addition to the MYND domain, the SET domain appeared also important in mediating some of the protein-protein interactions. We showed that the two

conserved regions in the SET domain were important for its interaction with p53, which is a substrate of SMYD2 (Huang et al., 2006).

HSP90α interaction with SMYD2 was mediated through a region other than the MYND and SET domains, as shown by its ability to interact with SMYD2 deletion mutants. I find that interaction between SMYD2 and HSP90α is important for SMYD2's histone methyltransferase activity in agreement with previous results for SMYD1 and 3 (Hamamoto et al., 2004)(Tan et al., 2006). The role HSP90α plays in the methylation activity of SMYD2 is unclear. However, interactions between HSP90α and its binding partners are mediated through a tetratricopeptide repeat (TPR) domain; for example, its interaction with protein phosphatase 5 (Ppp5) (Silverstein et al., 1997) (Chen et al., 1996) (Yang et al., 2005). HSP90 acts as a mediator for the association between PP5 and glucocorticoid receptor (GR) hetero-complexes. An expressed Ppp5 TPR domain acts as a dominant negative mutant which strongly inhibits GR-mediated transcriptional activation (Chen et al., 1996). The crystal structure of auto-inhibited Ppp5 reveals an extensive interface between the TPR domain and the phosphatase catalytic subunit (Yang et al., 2005). This interface includes the region of the TPR domain that interacts with HSP90α which led the authors to conclude that the binding between the TPR domain and HSP90α activates Ppp5 by dissociating the TPR domain from the phosphatase domain (Yang et al., 2005)(Chen and Cohen, 1997). SMYD2 has a TPR-like domain predicted by the InterPro database. We hypothesize that SMYD2 activation by HSP90α may be mediated through its TPR domain. This TPR-like domain could act as an inhibitor of SMYD2 activity by binding to its SET domain. Similar to Ppp5, SMYD2 interaction with HSP90α might be occurring through the TPR domain leading to the

enhancement of SMYD2 methylation activity. At this stage, further work is required to validate this hypothesis.

4.2 Methyltransferase Activity of SMYD2

Histone methylation of different lysine residues correlates with either the activation or the repression of gene expression (Lachner and Jenuwein, 2002) (Holbert and Marmorstein, 2005). For example, methylation of H3K4 is associated with gene activation while methylation of K9 on the same histone is associated with gene repression (Lachner and Jenuwein, 2002). Our data clearly shows that SMYD2 specifically methylates H3K4 in the presence of HSP90α, similar to SMYD1 and SMYD3. Furthermore, our *in vitro* data shows that in the absence of HSP90α, SMYD2 has some weak activity for dimethylating K36. The K36 methylation is in agreement with the data reported by Brown (Brown et al., 2006). However, our *in vivo* experiments strongly suggest that H3K4 is the predominant site of methylation. We did not observe any *in vivo* methylation at K36 above the basal level. Our *in vitro* experiments also suggest that K36 methylation does not occur in the presence of HSP90α. This dual specificity has been observed for other proteins such as LSD1. LSD1 is a histone demethylase that interacts with the androgen receptor leading to demethylation of the repressive histone marks mono- and dimethyl H3K9 (Metzger et al., 2005). When not bound by the androgen receptor, LSD1 demethylates the activating histone marks mono- and di-methyl on K4 of H3 (Shi et al., 2004). Hence, we suggest that SMYD2 when bound to HSP90α specifically methylates H3K4 and predominantly leads to the activation of gene expression. We also suggest that in the absence of HSP90α, SMYD2 has some level of activity to methylate H3K36 which could lead to gene repression.

4.3 Microarray Analysis of SMYD2 Over-Expression

The over-expression of SMYD2 causes changes in expression of genes associated with chromatin remodeling, cell cycle, and transcription regulation. The majority of genes regulated by SMYD2 were up-regulated in agreement with H3K4 methylation. Examples of proteins involved in chromatin remodeling that were up-regulated by the over-expression of SMYD2 are the three proteins from the SWI/SNF complex (Roberts and Orkin, 2004) (Lee et al., 2005). The SWI/SNF complex is recruited to chromatin to remodel nucleosomes using ATP hydrolysis as a source of energy (Johnson et al., 2005). The SWI/SNF complex has been implicated in cancer through different mechanisms (Gibbons, 2005) (Roberts and Orkin, 2004). Furthermore, we have observed a down-regulation of the protein WRD9 (bromodomain and WD repeat domain containing 1) which is known to interact with SMARCA4 (SWI2-related gene 1) (Huang et al., 2003). Bromodomains are known to bind to acetylated histone H3 (Yang, 2004). As well, we have observed the up-regulation of CHD9 (chromodomain helicase DNA binding protein 9) also named the chromatin-related mesenchymal modulator which is known to bind to methylated histones H3. Another gene that was up-regulated by SMYD2 over-expression is MEF2C which has transcription factor activity, and interestingly, regulates SMYD1 expression (Phan et al., 2005). Finally, many zinc-finger proteins which are known to bind DNA were up-regulated by SMYD2. The up-regulation of members of the SWI/SNF complex points to a chromatin remodeling effect on the cell. As well, the down-regulation of a bromodomain containing protein and the up-regulation of a chromo-domain containing protein point to potential feedback mechanisms for histone acetylation/methylation.

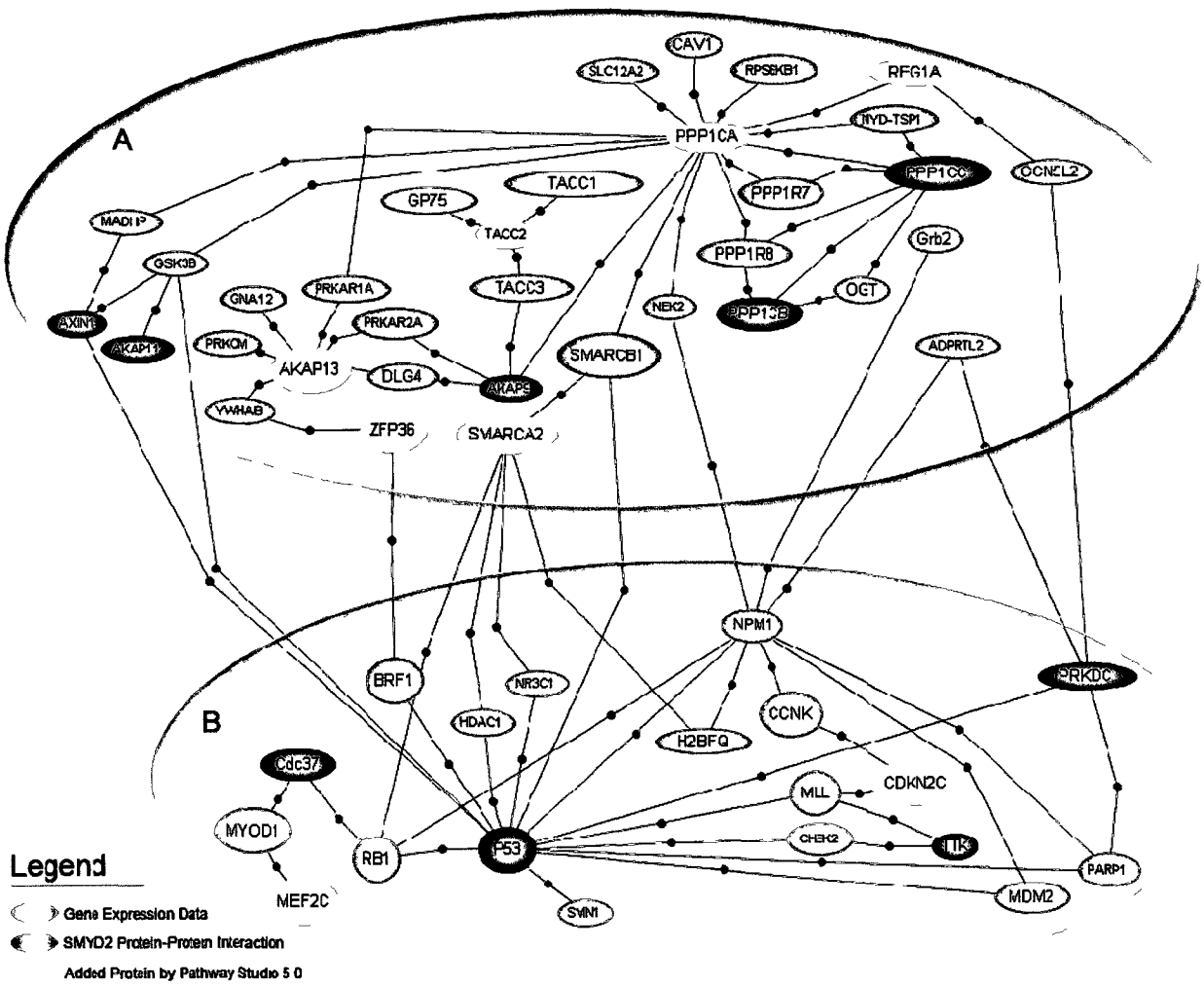
4.4 Bioinformatics Analysis of Proteomic and Genomic Data

In order to gain a global understanding of how SMYD2 protein-protein interactions affect gene expression, we combined the protein interaction data with the protein products of the gene up-regulated by SMYD2 with a bioinformatics tool called Pathway Studio. Combining the data in Pathway Studio allowed us to search for the interacting partners of the combined data in the network and to find the direct links and the shortest paths between entities (Nikitin et al., 2003). As shown in Figure 24, linkages were identified between the list of SMYD2 protein interactors and the protein products of the genes up-regulated by SMYD2. These proteins were divided for simplicity into two groups (Figure 24). The color code shows the source of entries in the list. The first group (A) includes a set of proteins that are associated with the centrosome and MTs. An interesting part of group A is the interaction of SMYD2 with members of the Protein phosphatase 1 (PP1) complex which is essential for cell division (Trinkle-Mulcahy et al., 2006). PPP1CC which interacts with SMYD2 forms an essential complex with another protein called Repo-Man that is required for the recruitment of PP1 to chromatin (Trinkle-Mulcahy et al., 2006). This complex plays important roles at multiple stages of the cell cycle, including both interphase and mitosis (Trinkle-Mulcahy et al., 2006). Inhibition of PPP1CC results in an aberrant chromatin condensation (Swain et al., 2007). AKAP9, another SMYD2 interaction protein, is a scaffolding protein that assembles several protein kinases and phosphatases, including PPP1CC, at the centrosome throughout the cell cycle (Takahashi et al., 2002)(Witczak et al., 1999). AKAP9 interacts with members of the TACC protein family (Steadman et al., 2002), one of which (TACC2) was up-regulated by SMYD2 over-expression. TACC2 is particularly interesting because it is phosphorylated by TTK (Dou et al., 2004) which we found interacting with SMYD2. Other

members of the AKAP and PP1 families were also found to associate with SMYD2 as shown in Figure 24. These include PPP1CA and B as well as AKAP11 and 13. The link to the centrosome is further supported by the up-regulation of PPPICA by SMYD2. PPP1CA has been shown to form a complex with Nek2, a serine threonine kinase that is localized to the centrosome (Helps et al., 2000). These protein-protein interaction data and the gene expression data show that SMYD2 might have an important role in the centrosomal MT dynamics. Other proteins that were also observed to associate with SMYD2 suggest, as well, a role for SMYD2 in microtubule dynamics. In particular, CLASP2 (CLIP-associating protein) is a mammalian MT plus-end binding protein, and it binds to EB1 and to MTs. It mediates interactions between MT plus-ends and the cell cortex (Mimori-Kiyosue et al., 2005).

The second group (B) includes genes such as p53, and PRKDC. The link between these proteins is probably centered on the role of SMYD2 in its regulation of p53. SMYD2 methylates p53 which leads to its dissociation from promoter regions of its target genes (Huang et al., 2006). In response to DNA damage, equilibrium is shifted towards SET7/9-mediated methylation of p53 which blocks SMYD2 methylation (Huang et al., 2006). Finally, the other protein that interacts with SMYD2 is PRKDC which is a Serine/threonine-protein kinase that acts as a molecular sensor for DNA damage. PRKDC also acts as a scaffold protein to aid the localization of DNA repair proteins to the site of damage (Cahill et al., 2006). Although it requires further experimental validations, combining our proteomic and genomic data using new bioinformatics tools demonstrates clear links to different cellular processes that support existing data on SMYD2 and provide new venues for understanding the function of SMYD2.

Figure 24: Combination of SMYD2 protein-protein interaction data with microarray data obtained from SMYD2 over-expression in HEK293T cells regrouped in two major networks. **A:** Enrichment of SMYD2 protein interaction partners to the centrosome; proteins anchored to the centrosome might have a role in centrosomal microtubule dynamics and the regulation of cell cycle. **B:** Potential role of SMYD2 as a DNA damage response gene through its interaction with genes like p53, and PRKDC.


A
B
CAV1
SLC12A2
RPS6KB1
RFG1A
NYD-TSP1
PPP1CA
PPP1CC
OCNEL2
TACC1
GP75
TACC2
PPP1R7
MADIP
GSK3B
PPP1R8
Grb2
TACC3
GNA12
PRKAR1A
NEK2
PPP1CB
OGT
AXIN1
AKAP11
PRKAR2A
PRKCM
AKAP13
DLG4
AKAP9
SMARCB1
ADPRTL2
YWHAB
ZFP36
SMARCA2
NPM1
BRF1
NR3C1
PRKDC
HDAC1
H2BFQ
CCNK
Cdc37
MLL
CDKN2C
MYOD1
RB1
P53
CHEK2
ITK
PARP1
MDM2
MEF2C
SMN1
Legend
Gene Expression Data
SMYD2 Protein-Protein Interaction
Added Protein by Pathway Studio 5 0

4.5 HSP90α Methylation by SMYD2

Proteins are subject to various types of PTMs that modulate their functions under various cellular stimuli (Ong et al., 2004) (Deribe et al., 2010). Identification of PTMs and their role in regulating protein functions has been very crucial to our understanding of cellular machinery (Eisenhaber and Eisenhaber, 2010). In the past decade many methods has been developed to study these PTMs. Such methods take advantage of certain chemical properties of PTMs to facilitate their identification. For example, Immobilized Metal Affinity Chromatography (IMAC) was developed to enrich phosphorylated peptides (Block et al., 2009). The IMAC method takes advantage of the negatively charged phosphate groups which allows the phosphate group to interact with the positively charged metal ions (Fe 3+, Ga 3+ , and Al 3+) (Wang et al., 2003). This interaction between the phosphate group and the metal ions makes it possible to enrich phosphorylated peptides from very complex peptide mixtures (Corthals et al., 2005). Unlike phosphorylation, methylation is an inert chemical group with a small mass that makes its detection very hard (Ellis et al., 2009). But in the last few years the development of various proteomic techniques has allowed for the identification of more methylation sites on arginine or lysine residues (Ong et al., 2004). Advancement in technology and the unraveling of important biological functions of lysine and arginine methylation has increased the interest in this PTM.

In my attempt to identify the interaction partners of SMYD2, I have probed SMYD2 IPs for protein methylation using a pan antibody that recognizes methylated lysine residues. This data showed that a protein with a similar size to HSP90α was possibly methylated. Using a combination of MS analysis and radioactive methylation assays, two sites were confirmed to be methylated by SMYD2. Those two sites are K209 and K565. These two sites

are located in the N- and C-terminal domain of HSP90α respectively. HSP90α is divided into three domains, N-Terminal (NTD), Middle Domain (MD) and a C-Terminal Domain (CTD) (Prodromou and Pearl, 2003) (Pearl and Prodromou, 2001). The NTD is involved in the ATP binding and hydrolysis, while the MD is known to interact with HSP90α client proteins (Pearl and Prodromou, 2001). The CTD is a dimerization domain that contains the tetratricopeptide repeat (TPR) motif recognition site (Mayer et al., 2009). This site is a conserved pentapeptide sequence made of MEEVD that mediates HSP90α interaction with co-factors such as the immunophilins FKBP51 and FKBP52 as well as other proteins (Young et al., 1998).

The methylation of HSP90α by SMYD2 raises a few questions in regard to the mechanism of action and the importance of this methylation of HSP90α. This is particularly interesting due to the role HSP90α in regulating methylation activity of SMYD2 as well as SMYD 1 and 3. The first question is whether this methylation of HSP90α is induced by a certain stimuli to change the activity of SMYD2 toward its substrates like H3, p53 or even HSP90α itself. It would also be important to study whether the methylation of HSP90α is compartmentalized. Since HSP90α is translocated to the nucleus upon certain stresses such as DNA damage, it would be worthwhile to investigate its localization upon such a treatment. It will also be important to see if there is interplay between the different modifications. For example, it is known that methylation of p53 by SMYD2 is inhibited by DNA damage. This is due in part to the fact that p53 is methylated on K372 that inhibits the binding between SMYD2 and p53 (Huang et al., 2006). But knowing the importance of p53 and its role as a tumor suppressor, it's believed that multiple levels of regulations might exist. This indeed is observed with the ubiquitination of p53 by Mdm2 which was believed to

be the sole E3 ubiquitin ligase involved in the regulation of p53 (Lee and Gu, 2010). However, some recent studies have showed that other E3 ubiquitin ligases might be involved in the regulation of p53 in an Mdm2-independent regulation (Chen et al., 2005). As a result it will be worthwhile to check the level of methylation on these different proteins that are methylated by SMYD2 and involved in DNA damage and dissect their role in responding to DNA damage. So HSP90α methylation might be one of those factors that regulate the activity of SMYD2. It is possible that the lack of methylation or even presence of methylation will lead to an increase or decrease of SMYD2 activity toward p53. This will give an added level of protection at the cellular level to ensure that p53 is activated at the right time and under the right conditions. It will also be important to keep looking for other substrates of SMYD2 which will enhance our understanding of this protein and its role in various cellular processes.

5.0 CONCLUSION AND FUTURE DIRECTIONS

In conclusion, I have established, through a combination of proteomics and genomics methods, some molecular roles of SMYD2. Our interactome mapping identified 21 protein interactors of SMYD2. I demonstrated that the MYND domain present in SMYD2 interacts with proteins that have the PXLXP motif, such as EPB41L3. Five of the SMYD2 interactors contain the PXLXP motif. I also established that SMYD2 interacts with HSP90α. I demonstrated that SMYD2, through its SET domain, acts as a methyltransferase that specifically methylates H3K4 in the presence of HSP90α and methylates H3K36 in the absence of HSP90α *in vitro*. I also showed that *in vivo* SMYD2 predominantly methylates H3K4. I established that SMYD2 interaction with its substrates like p53 is dependent on the SET domain.

I also discovered methylation sites on non-histone proteins including HSP90α and SMYD2 itself. The two sites that are methylated on HSP90α are K209 and K565 which are located in the NTD and the CTD respectively. I also showed that SMYD2 can automethylate itself in a concentration dependent manner where increasing the concentration of recombinant SMYD2 inhibits the methylation reaction in presence of radioactive SAM.

Microarray data indicates that SMYD2 over-expression leads to the up-regulation of 37 genes involved in chromatin remodeling, cell cycle as well as transcription regulation such as TACC2. I established that the regulation of downstream target genes like TACC2 is mediated through SMYD2's ability to associate with the promoter region of target genes and methylating histones in the promoter region. Furthermore, I demonstrated that SMYD2 and SMYD3 do not bind to the same promoter regions of TACC2. The combination of proteomics and genomics data indicates possible feedback mechanisms and combined

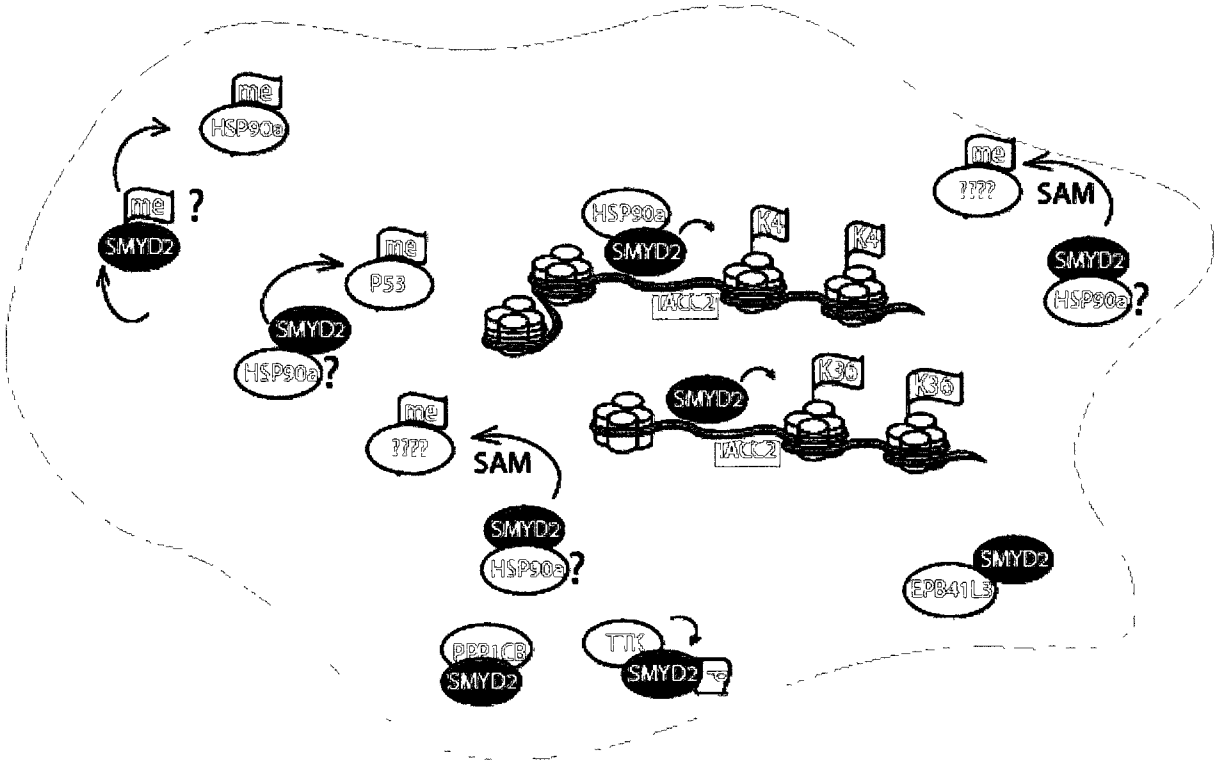
regulation at the interaction and expression levels. Finally, data provided here will greatly contribute to our understanding of SMYD2 and the SMYD family of proteins. Figure 25 gives a summary of the techniques used in our quest to understand the molecular role of SMYD2 and their value to our understanding of SMYD2 function.

5.1 Future Directions

SMYD2 is a member of the SMYD family which is getting an increasing amount of attention due to its role in diseases like cancer. I have started working on the SMYD2 project in 2005, at that time there was very little information know about the SMYD family and SMYD2 in particular. Since then, more data has been generated on these proteins. Nonetheless, this family is still not fully understood. As a result more work is still required to get a better picture of these proteins especially SMYD2 which has been shown to regulate important pathways in the cell. One of the first things that will require follow up from the data I generated is the methylation of HSP90α. In this work I have identified two sites of methylation on this protein but their function is still unknown. It would be very important to mutate these two sites and study their affect on the function of SMYD2. For example, they might affect its interaction with other proteins. This can be simply investigated with AP-IP coupled to MS. In performing these experiments, one should keep in mind that these methylation sites might be important under some type of a stress. So their importance might not be easily revealed under normal conditions. HSP90α methylation might also affect the function of SMYD2 itself. This can be investigated in an in vitro methylation assay to measure its methylation activity. SMYD2 substrates like p53 can be used to see if the mutants change its level of methylation. These mutants can also affect the automethylation of SMYD2 which should also be investigated.

Figure 25: A model for SMYD2 function in the cell. This model depicts the role of SMYD2 in methylating H3K4 and K36 and their role in regulating gene expression and other functions of SMYD2I have obtained data indicating that SMYD2 as a methyltransferase for non-histone proteins such as HSP90α. Others have showed that SMYD2 can methylate p53. The interplay between these methylation events is not fully understood at this stage which requires further investigation. Another factor that might play a crucial role in the function of SMYD2 is its association with HSP90α, which as I showed for H3 protein can change its specificity. It is also clear that the automethylation of SMYD2 plays an important role in regulating its activity. This model offers a simplistic view of the role SMYD2 can play in the cell. It also clearly shows that more work is required before the full function of SMYD2 in the cell is revealed.

Model of SMYD2 Function



One of the advantages of using MS for the identification of PTMs is its ability to perform unbiased large scale identification of these sites. This technique when coupled to other techniques like SILAC can lead to identification of high number of substrates for a certain enzyme. Identification of substrates of the SMYD protein family can lead to better understanding of their function in the cell. Identification of enzyme specific substrates can be using a SILAC reagent that labels methylated proteins with a methyl group containing Deuterium (Ong et al., 2004). Deuterium will be incorporated from D3-L methionine which will be used in media instead of the normal methionine. Labeled methionine will be used to synthesize SAM the methyl donor for the histone methyltransferase. SAM is synthesized endogenously from ATP and methionine by methionine adenosyltransferase. This is the main and only pathway in the cell to synthesize SAM. Growing the cells in D3-L-methionine media will replace all unlabelled SAM with labeled SAM if cells are grown in this media for about six doublings. L-methionine is a perfect choice for SILAC labeling since it is an essential amino acid and cannot be synthesized by the cell. Since methylation occurs in a mono, di and tri methylation, addition of a methyl group that is D3 labeled will result in a shift in mass of labeled peptide compared to unlabeled peptides. The shift will have values of 3, 6 or 9 Da depending on the degree of methylation. This shift can be detected by the mass spectrometer and with help of special software that is commercially available, we will be able to compare spectra from different treatments and determine fold changes in methylation of certain lysine residues. It will also allow us to determine the degree of methylation which is very important in understanding control of gene expression. Using this method combined with use of mass spectrometry and special software we will be able to identify methylation sites of the different SMYDs

Microarray data has shown that a number of cell cycle related genes were affected by SMYD2 overexpression. This as well as the fact that SMYD3 has been shown to be expressed in a cell cycle dependent manner may indicate that SMYD2 might also be expressed in a similar manner. Determining the localization of SMYD2 during the cell cycle is important to understand its function in regulating other genes in the cell during throughout this process. This process can be investigated by immunohistochemistry in cells synchronized at G1/S phase by agents such as Aphidicolin, which is a reversible inhibitor of eukaryotic nuclear DNA replication (Dhillon et al., 2003). Preliminary data on SMYD2 localization are given in Figure S7 Appendix I. Localization of SMYD2 during cell cycle or other processes is very important to understanding its substrate specificity. For example, it would be expected that cellular localization of SMYD2 in the nucleus will increase histone methylation. Identification of conditions that affect SMYD2 localization will also help advance our understanding of its function.

Role of histone modifications in muscle developments has been given a greater amount of attention. Recent reports show that Suv39h1 association with MyoD as well as its methyltransferase activity is important for repression of muscle related genes (Mal, 2006). Suv39h1 methylates lysine 9 on histone 3 on promoter regions of muscle related genes leading to repression of their gene expression (Mal, 2006). This finding was supported by an earlier report showing that the polycomb Ezh2 methyltransferase acts as a repressor of muscle related genes in a process that is dependent on its methyltransferase activity (Caretti et al., 2004). These separate reports indicate the important role histone methylation plays in regulating muscle differentiation and hence the need to study this process systematically. SMYD2 and SMYD1 are highly expressed in muscle tissues, where SMYD2 is expressed in the cardiac muscle and SMYD1 is expressed skeletal muscles. SMYD2 might be involved in

regulating the function of important genes in muscle development. SMYD2 overexpression has led to increase Mef2C expression level, which is an important transcription factor in the muscle development. I also obtained data which is given in Figure S10 Appendix 1 that shows SMYD2 binds to Mef2C promoter. These data indicate that SMYD2 might be an important regulator of muscle development. Our knowledge about SMYD2 can be used to establish its role in this process, where proteomic and genomic techniques can be used. SILAC based techniques can be used to study its methylation activity during the differentiation of cell models like P19 cells, which can differentiate into cardiomyocytes. On the other hand genomic techniques like ChIP-Chip can be used to identify the promoter regions SMYD2 binds. This technique combines the use of chromatin immunoprecipitation and DNA microarray where the DNA of different genes promoter regions will be plotted into the array and probed with DNA isolated from the chromatin immunoprecipitation assay (Wu et al., 2006). This can be done over the course of muscle differentiation to monitor the change in the binding of SMYD2 to different promoter sequences. This experiment combined with the previous experiments will allow us to determine genes that are regulated by SMYD2 and how this regulation is changing in accordance to the change in methylation status of histones as well as the change of the DNA binding patterns to different promoter regions during the differentiation process.

6.0 REFERENCES

1. Ansieau, S., and Leutz, A. (2002). The conserved Mynd domain of BS69 binds cellular and oncoviral proteins through a common PXLXP motif. The Journal of biological chemistry *277*, 4906-4910.
2. Aouacheria, A., Navratil, V., Wen, W., Jiang, M., Mouchiroud, D., Gautier, C., Gouy, M., and Zhang, M. (2005). In silico whole-genome scanning of cancer-associated nonsynonymous SNPs and molecular characterization of a dynein light chain tumour variant. Oncogene *24*, 6133-6142.
3. Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., Ara, T., Nakahigashi, K., Huang, H.C., Hirai, A., *et al.* (2006). Large-scale identification of protein-protein interaction of Escherichia coli K-12. Genome research *16*, 686-691.
4. Bai, Y., Markham, K., Chen, F., Weerasekera, R., Watts, J., Horne, P., Wakutani, Y., Bagshaw, R., Mathews, P.M., Fraser, P.E., *et al.* (2008). The in vivo brain interactome of the amyloid precursor protein. Mol Cell Proteomics *7*, 15-34.
5. Bannister, A.J., and Kouzarides, T. (2005). Reversing histone methylation. Nature *436*, 1103-1106.
6. Bauer, A., and Kuster, B. (2003). Affinity purification-mass spectrometry. Powerful tools for the characterization of protein complexes. European journal of biochemistry FEBS *270*, 570-578.
7. Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J., and Schafer, F. (2009). Immobilized-metal affinity chromatography (IMAC): a review. Methods in enzymology *463*, 439-473.

8. Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P.O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S., *et al.* (2004). A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat Cell Biol *6*, 97-105.

9. 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.

10. Cahill, D., Connor, B., and Carney, J.P. (2006). Mechanisms of eukaryotic DNA double strand break repair. Front Biosci *11*, 1958-1976.

11. Caretti, G., Di Padova, M., Micales, B., Lyons, G.E., and Sartorelli, V. (2004). The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. Genes & development *18*, 2627-2638.

12. Chen, D., Kon, N., Li, M., Zhang, W., Qin, J., and Gu, W. (2005). ARF-BP1/Mule is a critical mediator of the ARF tumor suppressor. Cell *121*, 1071-1083.

13. Chen, M.S., Silverstein, A.M., Pratt, W.B., and Chinkers, M. (1996). The tetratricopeptide repeat domain of protein phosphatase 5 mediates binding to glucocorticoid receptor heterocomplexes and acts as a dominant negative mutant. The Journal of biological chemistry *271*, 32315-32320.

14. Chen, M.X., and Cohen, P.T. (1997). Activation of protein phosphatase 5 by limited proteolysis or the binding of polyunsaturated fatty acids to the TPR domain. FEBS Lett *400*, 136-140.

15. Chi, P., Allis, C.D., and Wang, G.G. (2010). Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. Nat Rev Cancer *10*, 457-469.

16. Chin, H.G., Esteve, P.O., Pradhan, M., Benner, J., Patnaik, D., Carey, M.F., and Pradhan, S. (2007). Automethylation of G9a and its implication in wider substrate specificity and HP1 binding. Nucleic acids research *35*, 7313-7323.

17. 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.

18. Collins, F.S., Green, E.D., Guttmacher, A.E., and Guyer, M.S. (2003). A vision for the future of genomics research. Nature *422*, 835-847.

19. Comisarow, M.B., and Marshall, A.G. (1996). The early development of Fourier transform ion cyclotron resonance (FT-ICR) spectroscopy. J Mass Spectrom *31*, 581-585.

20. Corthals, G.L., Aebersold, R., and Goodlett, D.R. (2005). Identification of phosphorylation sites using microimmobilized metal affinity chromatography. Methods in enzymology *405*, 66-81.

21. Couture, J.F., and Trievel, R.C. (2006). Histone-modifying enzymes: encrypting an enigmatic epigenetic code. Current opinion in structural biology *16*, 753-760.

22. de Hoog, C.L., and Mann, M. (2004). Proteomics. Annu Rev Genomics Hum Genet *5*, 267-293.

23. Deribe, Y.L., Pawson, T., and Dikic, I. (2010) Post-translational modifications in signal integration. Nature structural & molecular biology *17*, 666-672.

24. Dhillon, V.S., Husain, S.A., and Ray, G.N. (2003). Expression of aphidicolin-induced fragile sites and their relationship between genetic susceptibility in breast cancer, ovarian cancer, and non-small-cell lung cancer patients. Teratogenesis, carcinogenesis, and mutagenesis *Suppl 1*, 35-45.

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

26. Dobreva, I., Fielding, A., Foster, L.J., and Dedhar, S. (2008). Mapping the integrin-linked kinase interactome using SILAC. Journal of proteome research *7*, 1740-1749. Dordick, J.S. (1991). Affinity-based separations and purifications. Patents and literature. Applied biochemistry and biotechnology *27*, 93-109.

27. Dou, Z., Ding, X., Zereshki, A., Zhang, Y., Zhang, J., Wang, F., Sun, J., Huang, H., and Yao, X. (2004). TTK kinase is essential for the centrosomal localization of TACC2. FEBS Lett *572*, 51-56.

28. Einhauer, A., and Jungbauer, A. (2001). The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. Journal of biochemical and biophysical methods *49*, 455-465.

29. Eisenhaber, B., and Eisenhaber, F. (2010). Prediction of posttranslational modification of proteins from their amino acid sequence. Methods in molecular biology (Clifton, NJ *609*, 365-384.

30. Elias, J.E., Haas, W., Faherty, B.K., and Gygi, S.P. (2005). Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nature methods *2*, 667-675.

31. Ellis, L., Atadja, P.W., and Johnstone, R.W. (2009). Epigenetics in cancer: targeting chromatin modifications. Molecular cancer therapeutics *8*, 1409-1420.

32. Ewing, R.M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M.D., O'Connor, L., Li, M., *et al.* (2007). Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol *3*, 89.

33. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science (New York, NY *246*, 64-71.

34. Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-246.

35. Figeys, D. (2002). Functional proteomics: mapping protein-protein interactions and pathways. Curr Opin Mol Ther *4*, 210-215.

36. Forneris, F., Battaglioli, E., Mattevi, A., and Binda, C. (2009). New roles of flavoproteins in molecular cell biology: histone demethylase LSD1 and chromatin. The FEBS journal *276*, 4304-4312.

37. Friesen, W.J., Massenet, S., Paushkin, S., Wyce, A., and Dreyfuss, G. (2001). SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. Molecular cell *7*, 1111-1117.

38. Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., *et al.* (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631-636.

39. Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., *et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.

40. Gerber, S.A., Kettenbach, A.N., Rush, J., and Gygi, S.P. (2007). The absolute quantification strategy: application to phosphorylation profiling of human separase serine 1126. Methods in molecular biology (Clifton, NJ *359*, 71-86.

41. Gibbons, R.J. (2005). Histone modifying and chromatin remodelling enzymes in cancer and dysplastic syndromes. Hum Mol Genet *14 Spec No 1*, R85-92.

42. Gingras, A.C., Aebersold, R., and Raught, B. (2005). Advances in protein complex analysis using mass spectrometry. J Physiol *563*, 11-21.

43. Gross, C.T., and McGinnis, W. (1996). DEAF-1, a novel protein that binds an essential region in a Deformed response element. The EMBO journal *15*, 1961-1970.

44. Gubitz, A.K., Mourelatos, Z., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). Gemin5, a novel WD repeat protein component of the SMN complex that binds Sm proteins. The Journal of biological chemistry *277*, 5631-5636.

45. Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol *17*, 994-999.

46. 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.

47. Hara, T., Honda, K., Shitashige, M., Ono, M., Matsuyama, H., Naito, K., Hirohashi, S., and Yamada, T. (2007). Mass spectrometry analysis of the native protein complex containing actinin-4 in prostate cancer cells. Mol Cell Proteomics *6*, 479-491.

48. Hart, G.T., Lee, I., and Marcotte, E.R. (2007). A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality. BMC Bioinformatics *8*, 236.

49. Helps, N.R., Luo, X., Barker, H.M., and Cohen, P.T. (2000). NIMA-related kinase 2 (Nek2), a cell-cycle-regulated protein kinase localized to centrosomes, is complexed to protein phosphatase 1. Biochem J *349*, 509-518.

50. Himeda, C.L., Ranish, J.A., and Hauschka, S.D. (2008). Quantitative proteomic identification of MAZ as a transcriptional regulator of muscle-specific genes in skeletal and cardiac myocytes. Molecular and cellular biology *28*, 6521-6535.

51. Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., *et al.* (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature *415*, 180-183.

52. Holbert, M.A., and Marmorstein, R. (2005). Structure and activity of enzymes that remove histone modifications. Current opinion in structural biology *15*, 673-680.

53. Hu, Q., Noll, R.J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R. (2005). The Orbitrap: a new mass spectrometer. J Mass Spectrom *40*, 430-443.

54. Huang, H., Rambaldi, I., Daniels, E., and Featherstone, M. (2003). Expression of the Wdr9 gene and protein products during mouse development. Dev Dyn *227*, 608-614.

55. Huang, J., and Berger, S.L. (2008). The emerging field of dynamic lysine methylation of non-histone proteins. Current opinion in genetics & development *18*, 152-158.

56. 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.

57. 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.

58. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A *98*, 4569-4574.

59. James, W.H. (2002). A new linear ion trap mass spectrometer. Rapid Communications in Mass Spectrometry *16*, 512-526.

60. Jasavala, R., Martinez, H., Thumar, J., Andaya, A., Gingras, A.C., Eng, J.K., Aebersold, R., Han, D.K., and Wright, M.E. (2007). Identification of putative androgen receptor interaction protein modules: cytoskeleton and endosomes modulate androgen receptor signaling in prostate cancer cells. Mol Cell Proteomics *6*, 252-271.

61. Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science (New York, NY *293*, 1074-1080.

62. Jenuwein, T., Laible, G., Dorn, R., and Reuter, G. (1998). SET domain proteins modulate chromatin domains in eu- and heterochromatin. Cell Mol Life Sci *54*, 80-93.

63. Jiang, W., Roemer, M.E., and Newsham, I.F. (2005). The tumor suppressor DAL-1/4.1B modulates protein arginine N-methyltransferase 5 activity in a substrate-specific manner. Biochem Biophys Res Commun *329*, 522-530.

64. Johnson, C.N., Adkins, N.L., and Georgel, P. (2005). Chromatin remodeling complexes: ATP-dependent machines in action. Biochem Cell Biol *83*, 405-417.

65. Karas, M., Gluckmann, M., and Schafer, J. (2000). Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. J Mass Spectrom *35*, 1-12.

66. Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem *60*, 2299-2301.

67. Karas, M., and Kruger, R. (2003). Ion formation in MALDI: the cluster ionization mechanism. Chemical reviews *103*, 427-440.

68. Koichi Tanaka, H.W.Y.I.S.A.Y.Y.T.Y.T.M. (1988). Protein and polymer analyses up to <I>m/z</I> 100 000 by laser ionization time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry *2*, 151-153.

69. Kontaki, H., and Talianidis, I. (2010). Lysine methylation regulates E2F1-induced cell death. Molecular cell *39*, 152-160.

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

71. Kouzarides, T. (2002). Histone methylation in transcriptional control. Current opinion in genetics & development *12*, 198-209.

72. Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., *et al.* (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637-643.

73. 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 research *67*, 10759-10765.

74. Lachner, M., and Jenuwein, T. (2002). The many faces of histone lysine methylation. Curr Opin Cell Biol *14*, 286-298.

75. Lan, F., and Shi, Y. (2009). Epigenetic regulation: methylation of histone and non-histone proteins. Science in China *52*, 311-322.

76. Lansbergen, G., Grigoriev, I., Mimori-Kiyosue, Y., Ohtsuka, T., Higa, S., Kitajima, I., Demmers, J., Galjart, N., Houtsmuller, A.B., Grosveld, F., *et al.* (2006). CLASPs attach microtubule plus ends to the cell cortex through a complex with LL5beta. Dev Cell *11*, 21-32.

77. Lash, A.E., Tolstoshev, C.M., Wagner, L., Schuler, G.D., Strausberg, R.L., Riggins, G.J., and Altschul, S.F. (2000). SAGEmap: a public gene expression resource. Genome research *10*, 1051-1060.

78. Lee, D.Y., Teyssier, C., Strahl, B.D., and Stallcup, M.R. (2005). Role of protein methylation in regulation of transcription. Endocr Rev *26*, 147-170.

79. Lee, J.T., and Gu, W. (2010). The multiple levels of regulation by p53 ubiquitination. Cell death and differentiation *17*, 86-92.

80. Lin, C., Smith, E.R., Takahashi, H., Lai, K.C., Martin-Brown, S., Florens, L., Washburn, M.P., Conaway, J.W., Conaway, R.C., and Shilatifard, A. (2010). AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. Molecular cell *37*, 429-437.

81. Little, J.T., and Jurica, M.S. (2008). Splicing factor SPF30 bridges an interaction between the prespliceosome protein U2AF35 and tri-small nuclear ribonucleoprotein protein hPrp3. The Journal of biological chemistry *283*, 8145-8152.

82. Lopez, J.L. (2007). Two-dimensional electrophoresis in proteome expression analysis. J Chromatogr B Analyt Technol Biomed Life Sci *849*, 190-202.

83. Luo, Y., Batalao, A., Zhou, H., and Zhu, L. (1997). Mammalian two-hybrid system: a complementary approach to the yeast two-hybrid system. BioTechniques *22*, 350-352.

84. Mal, A.K. (2006). Histone methyltransferase Suv39h1 represses MyoD-stimulated myogenic differentiation. The EMBO journal *25*, 3323-3334.

85. Malik, S., and Bhaumik, S.R. (2010). Mixed lineage leukemia: histone H3 lysine 4 methyltransferases from yeast to human. The FEBS journal *277*, 1805-1821.

86. Mann, M., Hendrickson, R.C., and Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. Annu Rev Biochem *70*, 437-473.

87. March, R.E. (2000). Quadrupole ion trap mass spectrometry: a view at the turn of the century. International Journal of Mass Spectrometry *200*, 285-312.

88. Marmorstein, R. (2003). Structure of SET domain proteins: a new twist on histone methylation. Trends in biochemical sciences *28*, 59-62.

89. Marshall, A.G., Hendrickson, C.L., and Jackson, G.S. (1998). Fourier transform ion cyclotron resonance mass spectrometry: a primer. Mass Spectrom Rev *17*, 1-35.

90. Mayer, M.P., Prodromou, C., and Frydman, J. (2009). The Hsp90 mosaic: a picture emerges. Nature structural & molecular biology *16*, 2-6.

91. Metzger, E., and Schule, R. (2007). The expanding world of histone lysine demethylases. Nature structural & molecular biology *14*, 252-254.

92. Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., Peters, A.H., Gunther, T., Buettner, R., and Schule, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature *437*, 436-439.

93. Michnick, S.W., Ear, P.H., Manderson, E.N., Remy, I., and Stefan, E. (2007). Universal strategies in research and drug discovery based on protein-fragment complementation assays. Nature reviews *6*, 569-582.

94. Mimori-Kiyosue, Y., Grigoriev, I., Lansbergen, G., Sasaki, H., Matsui, C., Severin, F., Galjart, N., Grosveld, F., Vorobjev, I., Tsukita, S., *et al.* (2005). CLASP1 and CLASP2 bind to EB1 and regulate microtubule plus-end dynamics at the cell cortex. The Journal of cell biology *168*, 141-153.

95. Mockli, N., Deplazes, A., Hassa, P.O., Zhang, Z., Peter, M., Hottiger, M.O., Stagljar, I., and Auerbach, D. (2007). Yeast split-ubiquitin-based cytosolic screening system to detect interactions between transcriptionally active proteins. BioTechniques *42*, 725-730.

96. Ng, S.S., Yue, W.W., Oppermann, U., and Klose, R.J. (2009). Dynamic protein methylation in chromatin biology. Cell Mol Life Sci *66*, 407-422.

97. Nikitin, A., Egorov, S., Daraselia, N., and Mazo, I. (2003). Pathway studio--the analysis and navigation of molecular networks. Bioinformatics *19*, 2155-2157.

98. Ong, S.E., Foster, L.J., and Mann, M. (2003). Mass spectrometric-based approaches in quantitative proteomics. Methods *29*, 124-130.

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

100. Paumi, C.M., Menendez, J., Arnoldo, A., Engels, K., Iyer, K.R., Thaminy, S., Georgiev, O., Barral, Y., Michaelis, S., and Stagljar, I. (2007). Mapping protein-protein interactions for the yeast ABC transporter Ycf1p by integrated split-ubiquitin membrane yeast two-hybrid analysis. Molecular cell *26*, 15-25.

101. Pearl, L.H., and Prodromou, C. (2001). Structure, function, and mechanism of the Hsp90 molecular chaperone. Advances in protein chemistry *59*, 157-186.

102. Perez-Burgos, L., Peters, A.H., Opravil, S., Kauer, M., Mechtler, K., and Jenuwein, T. (2004). Generation and characterization of methyl-lysine histone antibodies. Methods in enzymology *376*, 234-254.

103. 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.

104. Prodromou, C., and Pearl, L.H. (2003). Structure and functional relationships of Hsp90. Current cancer drug targets *3*, 301-323.

105. Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., *et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

106. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol *17*, 1030-1032.

107. Roberts, C.W., and Orkin, S.H. (2004). The SWI/SNF complex--chromatin and cancer. Nat Rev Cancer *4*, 133-142.

108. Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M.C., and Leonhardt, H. (2008). A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. Mol Cell Proteomics *7*, 282-289.

109. Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution *4*, 406-425.

110. Sampath, S.C., Marazzi, I., Yap, K.L., Sampath, S.C., Krutchinsky, A.N., Mecklenbrauker, I., Viale, A., Rudensky, E., Zhou, M.M., Chait, B.T., *et al.* (2007). Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly. Molecular cell *27*, 596-608.

111. Santos-Rosa, H., and Caldas, C. (2005). Chromatin modifier enzymes, the histone code and cancer. Eur J Cancer *41*, 2381-2402.

112. Schwartz, J.C., Senko, M.W., and Syka, J.E. (2002). A two-dimensional quadrupole ion trap mass spectrometer. Journal of the American Society for Mass Spectrometry *13*, 659-669.

113. Shamovsky, I., Ivannikov, M., Kandel, E.S., Gershon, D., and Nudler, E. (2006). RNA-mediated response to heat shock in mammalian cells. Nature *440*, 556-560.

114. 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. Molecular cell *27*, 636-646.

115. Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

116. Shilatifard, A. (2006). Chromatin Modifications by Methylation and Ubiquitination: Implications in the Regulation of Gene Expression. Annu Rev Biochem.

117. Silverstein, A.M., Galigniana, M.D., Chen, M.S., Owens-Grillo, J.K., Chinkers, M., and Pratt, W.B. (1997). Protein phosphatase 5 is a major component of glucocorticoid receptor.hsp90 complexes with properties of an FK506-binding immunophilin. The Journal of biological chemistry *272*, 16224-16230.

118. Sims, R.J., 3rd, Nishioka, K., and Reinberg, D. (2003). Histone lysine methylation: a signature for chromatin function. Trends Genet *19*, 629-639.

119. 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. The Journal of biological chemistry *277*, 26524-26529.

120. Singh, V., Miranda, T.B., Jiang, W., Frankel, A., Roemer, M.E., Robb, V.A., Gutmann, D.H., Herschman, H.R., Clarke, S., and Newsham, I.F. (2004). DAL-1/4.1B tumor suppressor interacts with protein arginine N-methyltransferase 3 (PRMT3) and inhibits its ability to methylate substrates in vitro and in vivo. Oncogene *23*, 7761-7771.

121. Spadaccini, R., Perrin, H., Bottomley, M.J., Ansieau, S., and Sattler, M. (2006). Structure and functional analysis of the MYND domain. Journal of molecular biology *358*, 498-508.

122. Steadman, B.T., Schmidt, P.H., Shanks, R.A., Lapierre, L.A., and Goldenring, J.R. (2002). Transforming acidic coiled-coil-containing protein 4 interacts with centrosomal AKAP350 and the mitotic spindle apparatus. The Journal of biological chemistry *277*, 30165-30176.

123. Suter, B., Kittanakom, S., and Stagljar, I. (2008). Interactive proteomics: what lies ahead? BioTechniques *44*, 681-691.

124. Swain, J.E., Ding, J., Brautigan, D.L., Villa-Moruzzi, E., and Smith, G.D. (2007). Proper chromatin condensation and maintenance of histone H3 phosphorylation during mouse oocyte meiosis requires protein phosphatase activity. Biol Reprod *76*, 628-638.

125. Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., *et al.* (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes & development *16*, 1779-1791.

126. Takahashi, M., Yamagiwa, A., Nishimura, T., Mukai, H., and Ono, Y. (2002). Centrosomal proteins CG-NAP and kendrin provide microtubule nucleation sites by anchoring gamma-tubulin ring complex. Mol Biol Cell *13*, 3235-3245.

127. Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular biology and evolution *24*, 1596-1599.

128. 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.

129. Tarassov, K., Messier, V., Landry, C.R., Radinovic, S., Serna Molina, M.M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S.W. (2008). An in vivo map of the yeast protein interactome. Science (New York, NY *320*, 1465-1470.

130. Trinkle-Mulcahy, L., Andersen, J., Lam, Y.W., Moorhead, G., Mann, M., and Lamond, A.I. (2006). Repo-Man recruits PP1 gamma to chromatin and is essential for cell viability. The Journal of cell biology *172*, 679-692.

131. Trinkle-Mulcahy, L., Boulon, S., Lam, Y.W., Urcia, R., Boisvert, F.M., Vandermoere, F., Morrice, N.A., Swift, S., Rothbauer, U., Leonhardt, H., *et al.* (2008). Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. The Journal of cell biology *183*, 223-239.

132. Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.

133. Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A *98*, 5116-5121.

134. Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., *et al.* (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623-627.

135. Veraksa, A., Bauer, A., and Artavanis-Tsakonas, S. (2005). Analyzing protein complexes in Drosophila with tandem affinity purification-mass spectrometry. Dev Dyn *232*, 827-834.

136. Vermeulen, M., Hubner, N.C., and Mann, M. (2008). High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. Current opinion in biotechnology *19*, 331-337.

137. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., and Bork, P. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. Nature *417*, 399-403.

138. Wang, J.L., Zhang, Y.J., Cai, Y., Li, L., Liu, S.Y., Yang, H.Y., and Qian, X.H. (2003). [Analysis of protein phosphorylation by combination of IMAC, phosphatase with biological mass spectrometry]. Sheng wu hua xue yu sheng wu wu li xue bao Acta biochimica et biophysica Sinica *35*, 459-466.

139. Wepf, A., Glatter, T., Schmidt, A., Aebersold, R., and Gstaiger, M. (2009). Quantitative interaction proteomics using mass spectrometry. Nature methods *6*, 203-205.

140. Witczak, O., Skalhegg, B.S., Keryer, G., Bornens, M., Tasken, K., Jahnsen, T., and Orstavik, S. (1999). Cloning and characterization of a cDNA encoding an A-kinase anchoring protein located in the centrosome, AKAP450. The EMBO journal *18*, 1858-1868.

141. Wolff, J.C., Eckers, C., Sage, A.B., Giles, K., and Bateman, R. (2001). Accurate mass liquid chromatography/mass spectrometry on quadrupole orthogonal acceleration time-of-flight mass analyzers using switching between separate sample

and reference sprays. 2. Applications using the dual-electrospray ion source. Anal Chem *73*, 2605-2612.

142. Wu, J., Smith, L.T., Plass, C., and Huang, T.H. (2006). ChIP-chip comes of age for genome-wide functional analysis. Cancer research *66*, 6899-6902.

143. Yang, J., Roe, S.M., Cliff, M.J., Williams, M.A., Ladbury, J.E., Cohen, P.T., and Barford, D. (2005). Molecular basis for TPR domain-mediated regulation of protein phosphatase 5. The EMBO journal *24*, 1-10.

144. Yang, X.J. (2004). Lysine acetylation and the bromodomain: a new partnership for signaling. Bioessays *26*, 1076-1087.

145. Young, J.C., Obermann, W.M., and Hartl, F.U. (1998). Specific binding of tetratricopeptide repeat proteins to the C-terminal 12-kDa domain of hsp90. The Journal of biological chemistry *273*, 18007-18010.

7.0 APPENDIX I: Additional Data

Figure S1: A silver stain showing proteins that are pulled down in an anti-FLAG IP of SMYD2, SMYD3, SMYD4 and SMYD5 compared to the empty vector control. Experiment was performed at an n=3.

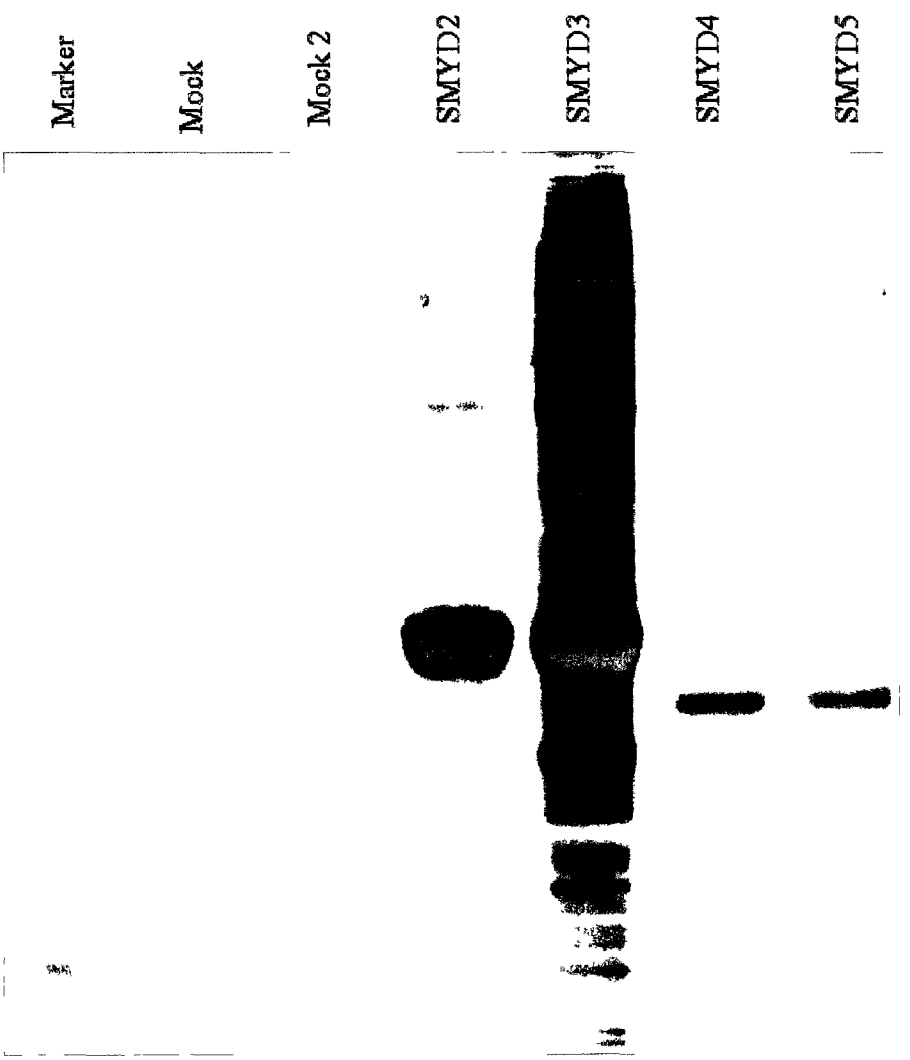

Marker
Mock
Mock 2
SMYD2
SMYD3
SMYD4
SMYD5

Figure S2: An illustration of the list of SMYD3 protein interactors pulled down with anti-FLAG agarose conjugated antibodies and identified by MS.

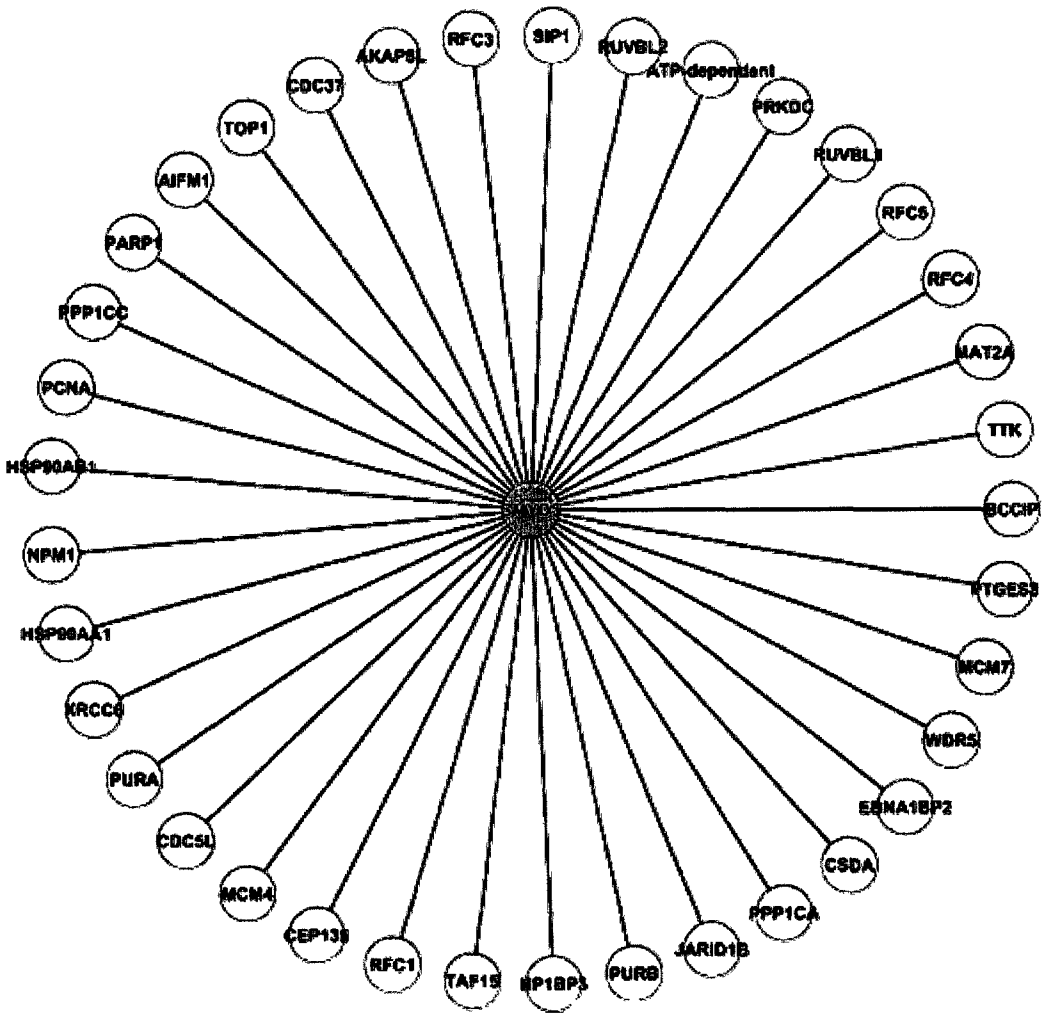

RFC3
SIP1
RUVBL2
ATP-dependent
PRKDC
RUVBL1
RFC5
RFC4
MAT2A
TTK
BCCIP
PTGES3
MCM7
WDR5
EBNA1BP2
CSDA
PPP1CA
JARID1B
PURB
TAF15
RFC1
CEP135
MCM4
CDC5L
PURA
XRCC6
HSP90AA1
NPM1
HSP90AB1
PCNA
PPP1CC
PARP1
AIFM1
TOP1
CDC37
AKAP8L

Figure S3: An illustration of the list of SMYD4 protein interactors pulled down with anti-FLAG agarose conjugated antibodies and identified by MS.

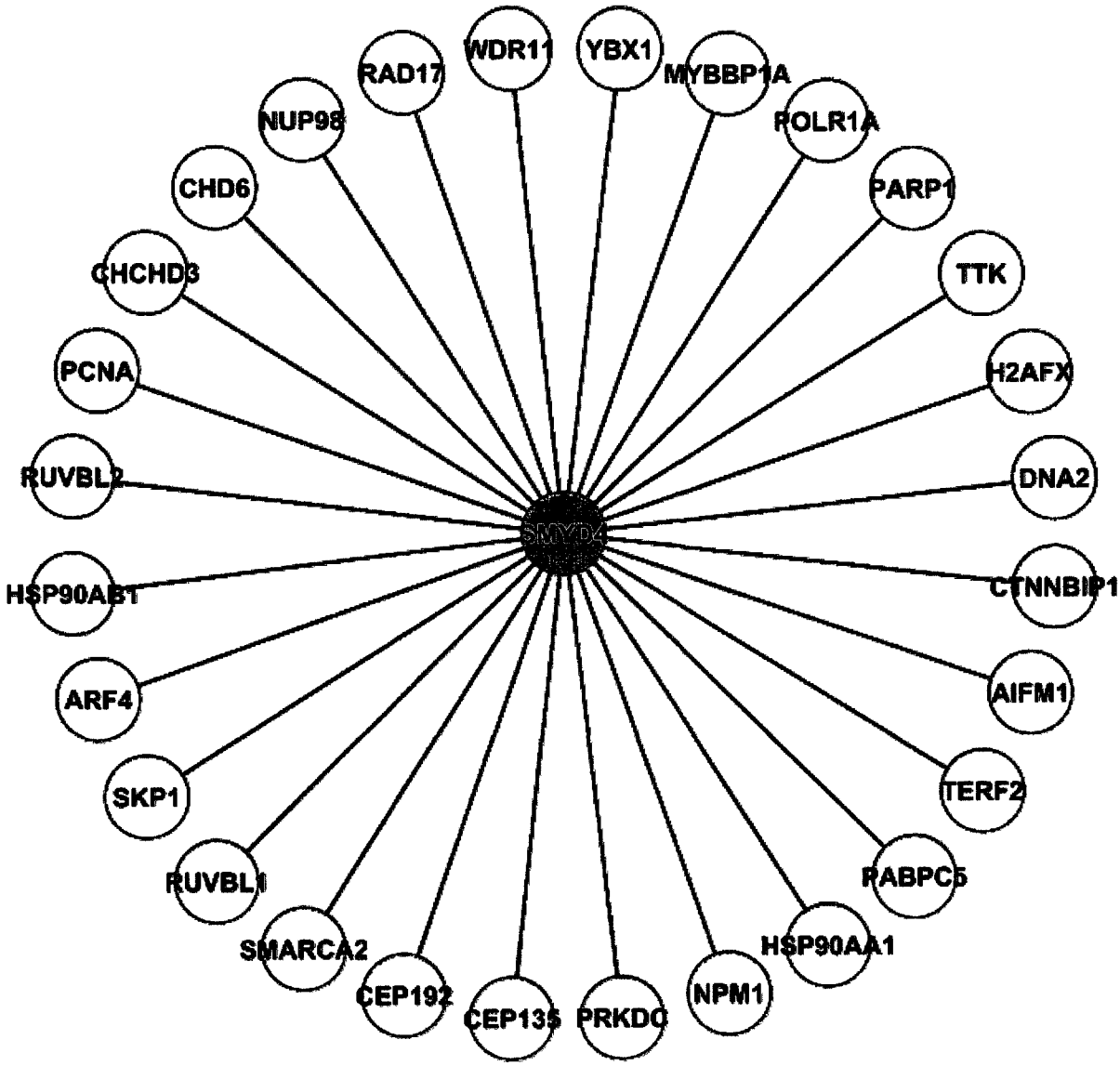

WDR11
YBX1
RAD17
MYBBP1A
NUP98
POLR1A
CHD6
PARP1
CHCHD3
TTK
PCNA
H2AFX
RUVBL2
DNA2
SMYD4
HSP90AB1
CTNNBIP1
ARF4
AIFM1
SKP1
TERF2
RUVBL1
PABPC5
SMARCA2
HSP90AA1
CEP192
CEP135
PRKDC
NPM1

Figure S4: An illustration of the list of SMYD5 protein interactors pulled down with anti-FLAG agarose conjugated antibodies and identified by MS.

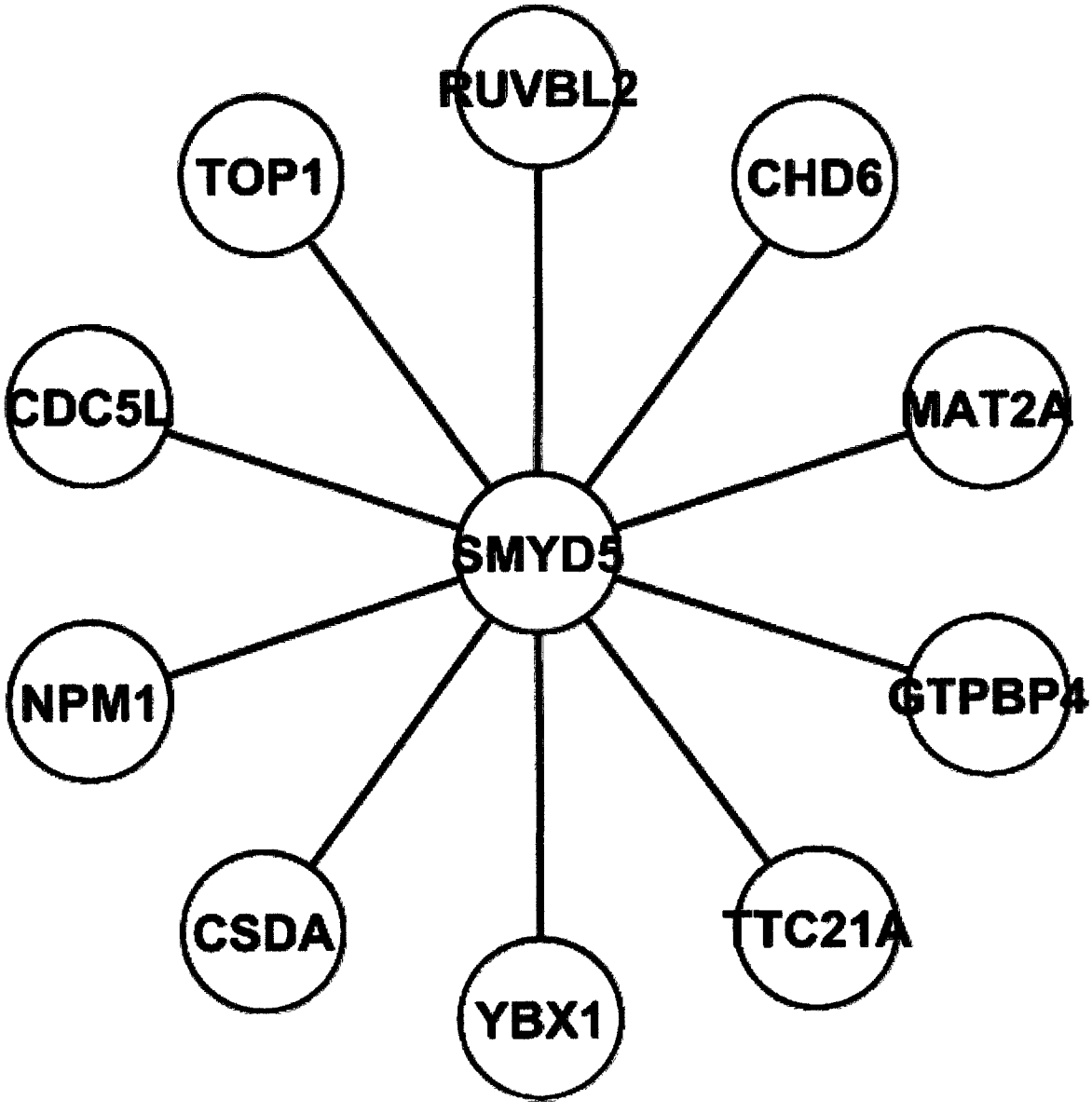
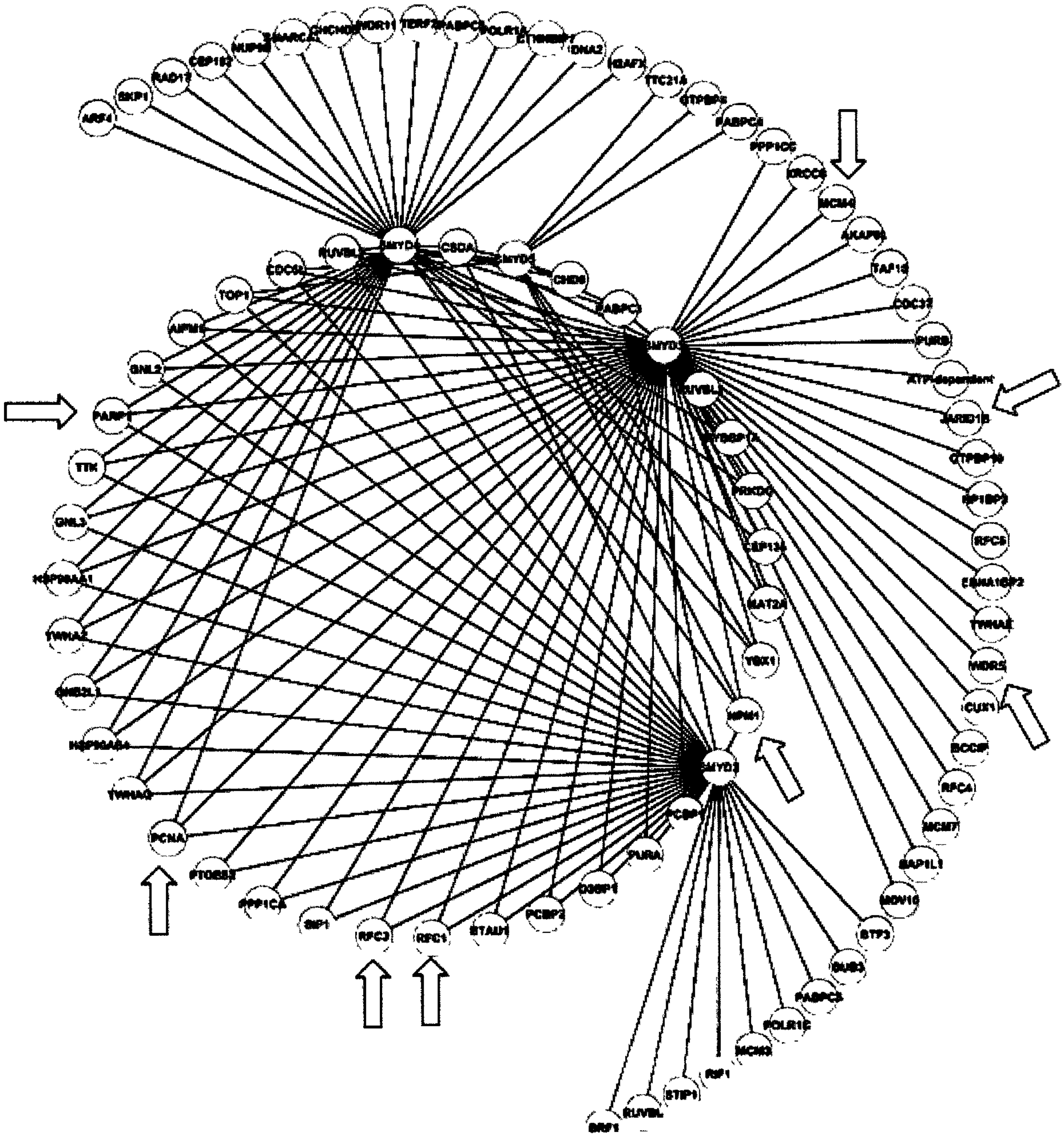

RUVBL2
TOP1
CHD6
CDC5L
MAT2A
SMYD5
NPM1
GTPBP4
CSDA
TTC21A
YBX1

Figure S6: An illustration of the combined interaction data form SMYD protein family members that was identified in pulled down with anti-FLAG agarose conjugated antibodies coupled to MS. The arrows indicate proteins that are common between SMYDs and play a role in DNA replication and DNA damage.



ARF4
SKP1
RAD17
WDR11
TERF2
DNA2
KRCC5
MCM4
CSDA
CDC5L
TOP1
PARP1
TTK
GNL3
GNL2
PCNA
PPP1CA
RFC3
RFC1
STAU1
PCBP2
PURA
PRKDC
MAT2A
YBX1
RFC5
WDR5
CUX1
BCCIP
RFC4
MCM7
MDV10
BUB3
POLR1C
MCM3
STIP1
RUVBL
PURB
CDC37
JARID1B
HSP90AA1
PABPC3

Figure S7: Cell cycle localization of SMYD2 after cell cycle arrest of HeLa cells at the G1/S phase. The cells are transfected withFLAG-SMYD2 for 24 hours and then plated with media containing 5 mg/ml of the cell blocker Amphidicolin. After 36 hour of incubation with the drug, cells were washed with PBS and grown in normal growth media. Cells were then harvested at 0,4,8 and 12 hours after the release. Cells are stained with anti-FLAG-FITC conjugated antibodies

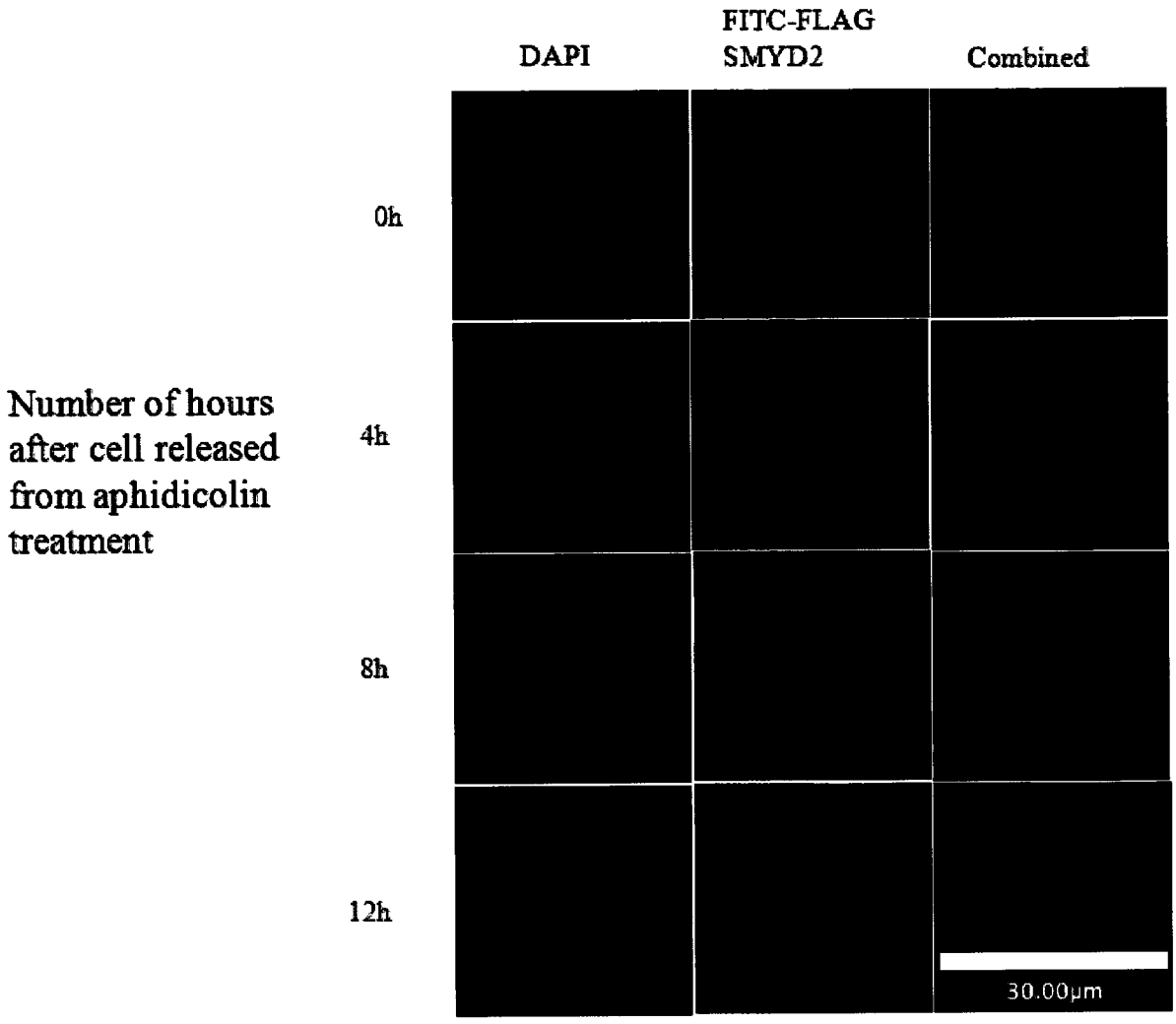
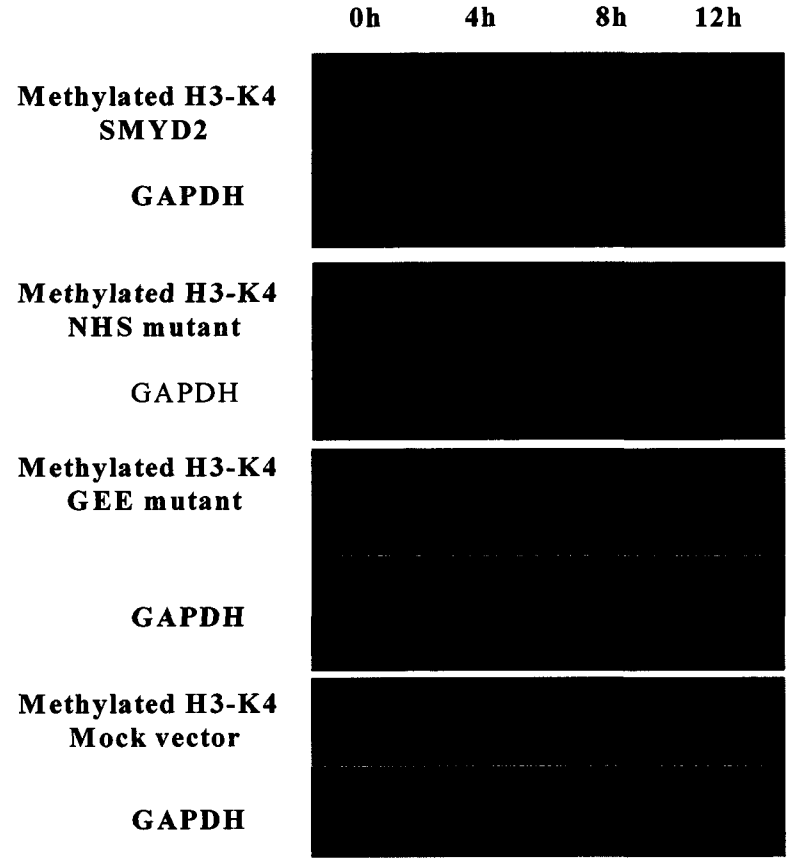

DAPI
FITC-FLAG SMYD2
Combined
0h
4h
8h
12h
Number of hours after cell released from aphidicolin treatment
30.00µm

Figure S8: *In vivo* histone methylation. **A**: Western blot showing the increase of H3K4 methylation in cells over-expressing SMYD2 after their release from the G1/S phase arrest. It shows the increase in methylation of wild type expressing HeLa cells compared to mock and the SET domain deletion mutants. GAPDH is used as a loading control. **B**: The spot densitometry graph of date in part **A** normalized to GAPDH expression showing the increase in histone methylation due to SMYD2 over-expression after the cells are released.

A


0h
4h
8h
12h
Methylated H3-K4
SMYD2
GAPDH
Methylated H3-K4
NHS mutant
GAPDH
Methylated H3-K4
GEE mutant
GAPDH
Methylated H3-K4
Mock vector
GAPDH

B

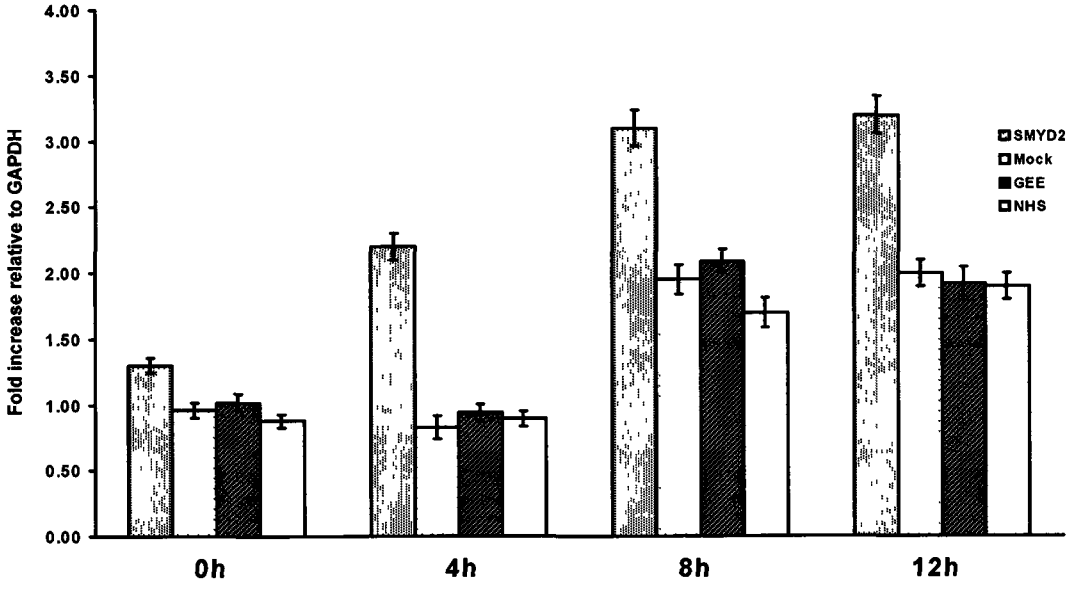
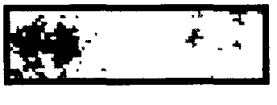
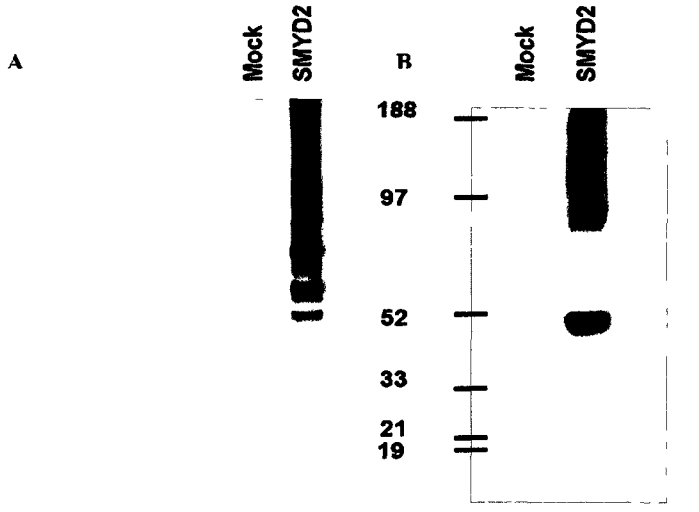

Fold increase relative to GAPDH
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
0h
4h
8h
12h
SMYD2
Mock
GEE
NHS

Table S1: RNA expression level in various cancer tissues compaired to normal tissues. SMYD gene expression was normalized to the gene expression level of actin.

Sample	SMYD2	Actin	Change
Adrenal Tumor			4.7
Adipose Tumor			-17.6
Brain Tumor			-8.8
Duodenum Tumor			-5.6
Esophagus Tumor Adenocarcinoma			15.1
Kidney Tumor Clear Cell Carcinoma			-6.4
Liver Tumor Hepatocellullar Carcinoma Moderately Differentiated			3.7
Lung Tumor Alveolar Carcinoma			-4.2
Lymphoma			-5.1
Ovary Thecoma			2.3
Prostate Tumor			-3.2
Rectum Adenocarcinoma Poorly Differentiated			6.3

Small intestine Tumor		**3.7**
Soft tissue tumor		**16.9**
Stomach tumor Squamous Cell Carcinoma		**7.8**
Stomach tumor Adenocarcinoma Moderately Differentiated		**-9.6**
Testis Tumor		**2.1**
Thyroid Tumor Adenocarcinoma		**13.7**
Uterus Tumor Adenocarcinoma		**2.3**

Figure S9: FLAG SMYD2 immunoprecipitation using anti-FLAG antibodies. A: silver stain gel of mock and SMYD2 transfected HEK293T cells using anti-FLAG antibodies of formaldehyde crosslinked SMYD2/DNA complex. B: Immunoprecipitated samples from A probed with HRP conjugated anti-FLAG antibodies showing SMYD2 protein and its binding partners at higher molecular weight. C: Comparison between different immunoprecipitation samples. D: Anti-histone 3 western blot showing the IP of H3 using the formaldehyde crosslinking method.


A
Mock
SMYD2
B
Mock
SMYD2
188
97
52
33
21
19

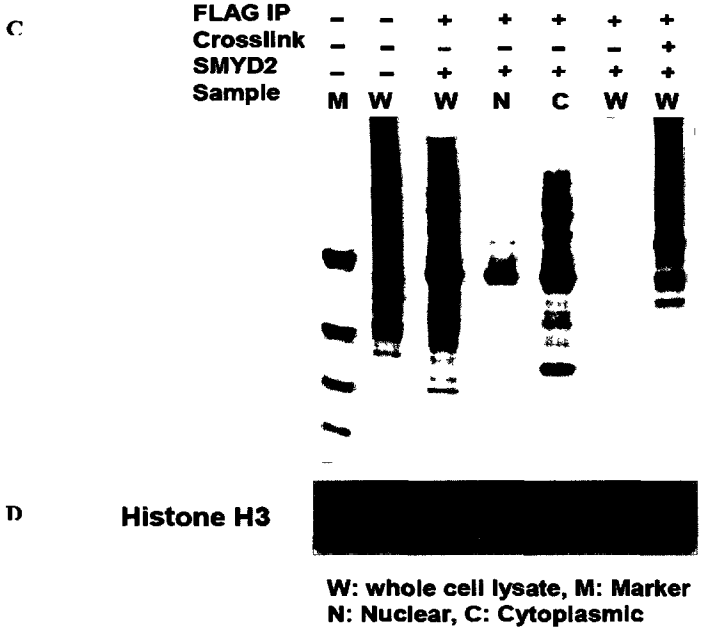
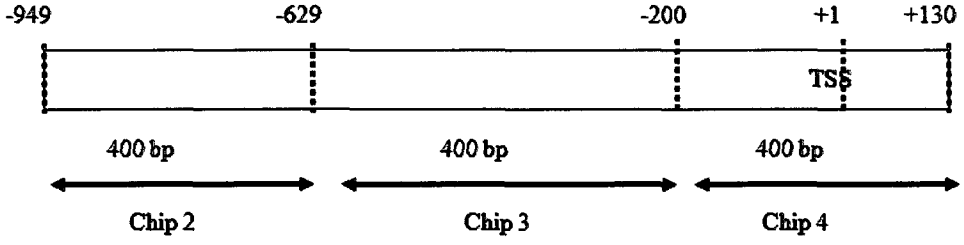
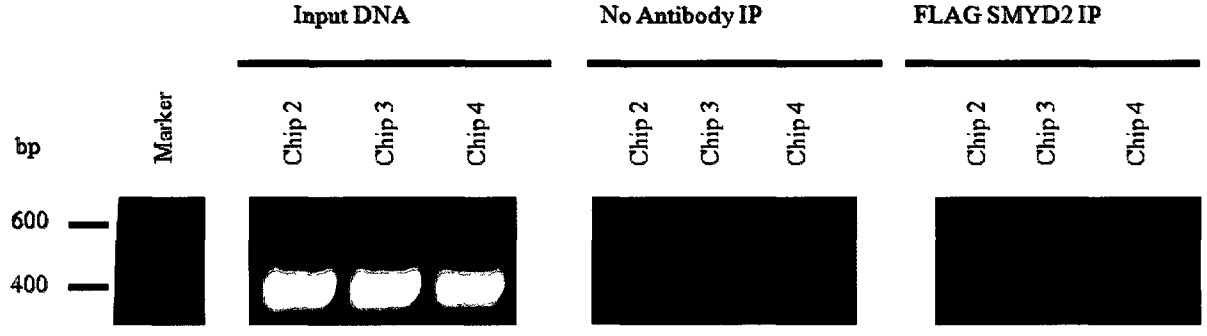

C
FLAG IP
Crosslink
SMYD2
Sample
M W W N C W W
D
Histone H3
W: whole cell lysate, M: Marker
N: Nuclear, C: Cytoplasmic

Figure S10: Promoter sequence of Mef2C gene showing the location of primers used in the ChIP assay and the TSS (transcription start site). PCR results of the chromatin IP showing a PCR product for Chip 4 primers only when anti-FLAG-SMYD2 antibodies were used, indicating that SMYD2 can bind to the Mef2C promoter. The last panel shows PCR amplification products for the input DNA as a PCR amplification control for the primers.


-949
-629
-200
+1
+130
TSS
400 bp
400 bp
400 bp
Chip 2
Chip 3
Chip 4


Input DNA
No Antibody IP
FLAG SMYD2 IP
bp
Marker
Chip 2
Chip 3
Chip 4
Chip 2
Chip 3
Chip 4
Chip 2
Chip 3
Chip 4
600
400

8.0 APPENDIX II: Curriculum Vitae

Mohamed Abu-Farha

CAREER OBJECTIVE

Seeking a post-doctoral position that allows me to build on my molecular biology, cell biology and proteomics skills and apply them in a clinical setting to answer disease related questions

EDUCATION

2005–Present	**Doctor of Philosophy.**
2003–2005	**Master's of Science**. Thesis title "Regulation of Erythroid-Specific 5-Aminolevulinate Synthase (ALAS2) by Hypoxia" Carleton University, Ottawa, Canada. Recipient of Dean's Scholarship, 2003-2005
1998–2003	**Bachelor of Science Highest Honors**, Biochemistry and Biotechnology. Fourth year project title "Hydroxilation And Protein Stability Under Hypoxia" Carleton University, Ottawa, Canada. Recipient of Dean's Scholarship, 1999-2003

SKILLS PROFILE

Proteomics Skills, Experienced in:

- Protein tagging and Immunoprecipitation followed by in gel or in solution enzyme digestion.
- Protein identification by mass spectrometry ESI-LCMS/MS machines such as LTQ and LTQ-OrbiTrap from Thermo scientific
- Quantitative proteomic techniques such as SILAC
- Identification of post translational protein modifications by a combination of mass spectrometry and data base searching algorithms. These modifications can include phosphorylation, methylation and acetylation
- Database searching algorithms such as Mascot and other protein identification and quantification softwares such as MSquant and Maxquant

Molecular Biology Skills, Excellent Knowledge in:

- DNA and RNA isolation, RT-PCR, Real Time PCR, northern blotting, siRNA and site directed mutagenesiss
- Using different DNA cloning techniques like Gateway, TA and TOPO cloning and epitope tagging in mammalian and bacterial expression systems
- Protein expression and protein purification using affinity chromatography
- 2D gels, different protein staining techniques like (silver, Comassie blue and radioactive labeling)

Cell Biology Skills, Exceptional in:

- Tissue culturing of various carcinoma cell lines, making stable cell lines, making growth media, making cell stocks and cell splitting.
- Bacterial culturing and manipulations
- Use of different kinds of microscopes like phase contrasts and epifluorescence microscopes
- Immunostaining using different fluorescence dyes

Biochemistry Skills, Proficient in:

- Experimental biochemistry. Proteins quantitation by the use of spectrophotometer and colorimetric assays, Differential centrifugation and solubility fractionation techniques for isolating the different organelles
- Enzyme kinetics analysis according to the Michalis-Menten equation
- Radioactive labeling of proteins and protein synthesis inhibition studies
- Certified to use radio isotopes as well as ability to use scintillation counter

WORK EXPERIENCE

2005-Present **PhD Student, University of Ottawa**

My PhD work involves the characterization of a novel histone lysine methyltransferase protein called SMYD2. I use both genomic and proteomic approaches to understand the molecular roles of SMYD2. Using interactome mapping methods, I identified novel SMYD2 interaction partners and demonstrated the role of SMYD2's two domains (MYND and SET) in mediating some of these interactions, in particular the interactions with EBP41L3, HSP90α, and p53. Furthermore, I showed that SMYD2 acts as a H3K4 methyltransferase in the presence of HSP90α and acts as a H3K36 methyltransferase in the absence of HSP90α. Gene expression studies indicated that the SMYD2 gain of function in 293T cells leads predominantly to up-regulation of gene expression. ChIP data indicated that activation of downstream genes, such as TACC2 (Transforming, Acidic Coiled-coil Containing protein 2), occurs as a result of SMYD2's ability to methylate H3K4 in promoter regions of its target genes. I also showed that SMYD2 binds to p53 which was recently found to be methylated by SMYD2.

2003–2005 **Master's Student,** Carleton University

Studying protein modification via reactive oxygen species (ROS) and post-translational modification of proteins. Research is performed on various carcinoma cell lines like HeLa, K562, HepG2 and Cos7. The project involves the

cloning of certain genes into a mammalian expression vector with a FLAG tag. These proteins are then transfected into mammalian systems to follow the fate of the protein. We also performed coimmunoprecipitation studies to look at the interaction between protein of interest and other proteins like vHL, which is responsible for binding hydroxylated proteins leading them to degradation by the proteasome.

2002–2003 **Lab Technologist,** Carleton University
During this time, I worked on projects that involved identifying key proteins that may be regulated, at the level of the protein, by oxygen tension. The work required the maintenance of mammalian (human and otherwise) cell lines in culture. Tasks included selecting candidate proteins from a genomic screen that could be potentially regulated by oxygen, testing protein activity (enzymatic, DNA-binding, etc.) under different oxygen regimes (high, normal atmospheric, low), and cloning key candidates for expression and *in vitro* testing of function.

April-Sept 2001 **Laboratory Chemical Technologist,** Drain All Ltd
Lab Technologist at an environmental company (Drain-All Ltd) as a Co-op student. Involved in performing various wet analytical tests. Control the quality of the arriving and departing chemical wastes as well as classifying wastes, as given by the municipality, according to their chemical contents. Tests performed in the lab were PCB test using GC, flash point, pH, metals test using AA spectrometer, percent of water in oils, percent of halogens, oxidizer test, and density.

LIST OF SCHOLARSHIPS AND AWARDS

Name of Award	*Value*	*Period Held*
NSERC Post-doctoral Fellowship	$80,000.00	01/11-12/12
NSERC postgraduate Scholarships PGSD)	$63,000.00	05/06-05/09
HUPO Young Investigator Award	$600.00	09/09
University of Ottawa Excellence Scholarship	$6,000.00	05/06-05-09
University of Ottawa Admission Scholarship	$5,280.00	09/05-05/09
Biochemistry, Microbiology and Immunology Travel Award	$800.00	05/08
Ontario Graduate Scholarship	$15,000.00	05/06-05-07*
Biochemistry, Microbiology and Immunology Travel Award	$1000.00	05/07
Faculty of graduate and post-graduate studies travel award	$600.00	05/07
Canadian Society of Biochemistry and Molecular and Cellular Biology travel award	$600.00	05/04
Carleton University graduate studies student travel award	$224.00	08/04
Dean's Graduate scholarship	$3,231.00	09/03-09/05
Clarence C. Gibson Scholarship	$750.00	09/01-04/02
Murdoch Maxwell Macodrum Scholarship	$750.00	09/00-04/01
University Part Time Scholarship	$350.00	09/99-04/00

* Scholarship declined because I also received NSERC scholarship.

PUBLICATIONS AND PRESENTATIONS

Refereed Journal Papers

- **Abu-Farha, M.**, Lambert, J.P., Al-Madhoun, A.S., Elisma, F., Skerjanc, I.S., Figeys, D. (2008). The tale of two domains: proteomics and genomics analysis of SMYD2, a new histone methyltransferase. Mol Cell Proteomics. **7**(3):560-72.
- Ewing, R.M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M.D., O'Connor, L., Li, M., Taylor, R., Dharsee, M., Ho, Y., Heilbut, A., Moore, L., Zhang, S., Ornatsky, O., Bukhman, Y.V., Ethier, M., Sheng, Y., Vasilescu, J., **Abu-Farha, M.**, Lambert, J.P., Duewel, H.S., Stewart, I.I., Kuehl, B., Hogue, K., Colwill, K., Gladwish, K., Muskat, B., Kinach, R., Adams, S.L., Moran, M.F., Morin, G.B., Topaloglou, T., Figeys, D. (2007). Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol. **3**:89.
- **Abu-Farha, M.**, Nile, J., and Willmore, W.G. (2005). Erythroid-Specific 5-Aminolevulinate Synthase Protein s Stabilized By Low Oxygen And Proteosomal Inhibition. Bioch and Cell Biol. **83**: 620-630.
- **Abu-Farha, M.**, and Willmore, W.G. (2004). Hypoxic Stabilization And Proteolytic Degradation Of Erythroid-Specific 5-Aminolevulinate Synthase. Int Cong Ser. **1275**:71-78.

Referred Book Chapters

- **Abu-Farha, M.**, Elisma, F., Figeys, D. (2008). Identification of protein-protein interaction by MS-coupled techniques. Adv Biochem Eng Biotechnol. **110**:67-80.

Referred Review Articles

- **Abu-Farha, M.**, Elisma, F., Zhou, H., Tian, R., Zhou, H., Asmer, M.S., Figeys, D. (2009). Proteomics: From Technology Developments to Biological Applications. Anal Chem. **81**:4585-99.

Referred Conference Abstracts

- **Abu-Farha, M.**, Lambert, J.P., Elisma, F., Figeys, D. (2009). Proteomic Analysis of a Novel HKMT and its Role in Regulating Gene Expression. HUPO 8th Annual World Congress (In Press)
- **Abu-Farha, M.**, Lambert, J.P., Al-Madhoun, A.S., Elisma, F., Skerjanc, I.S., Figeys, D. (2009). A Combined Proteomic and Genomic Approach to Characterize SMYD2 a Histone Lysine Methyltransferase. Canadian Society of Biochemistry, Molecular & Cellular Biology (CSBMCB) 51st Annual Meeting. Biochemistry and Cell Biology. **87**:349-373.
- **Abu-Farha, M.**, and Willmore W.G. (2004). Post-Translational Modification and Protein Stabilization of ALAS2 under Hypoxia. Canadian Society of Biochemistry, Molecular & Cellular Biology (CSBMCB) 47th Annual Meeting. Biochemistry and Cell Biology **82** (6): 755.

Oral Presentation

- **Abu-Farha, M.**, and Figeys, D. (2008). Identification of the SMYDs Protein Interaction Network and their Implication in Cancer. Biochemistry, Microbiology & Immunology Research Presentations. University of Ottawa, Canada.
- **Abu-Farha, M.**, and Figeys, D. (2006). The Role of Histone Methylation in Microtubule Dynamics. Biochemistry, Microbiology & Immunology. Ottawa, Canada.
- **Abu-Farha, M.**, and Willmore, W. G. (2005). Post-Translational Modification and Protein Stabilization of ALAS2 under Hypoxia. Biology department, Carleton University, Ottawa. Canada.

Non-Referred Conferences Abstracts

- **Abu-Farha, M**, Lambert, J.P., Al-Madhoun, A.S., Elisma, F., Skerjanc, I.S., Figeys, D. (2007). Characterization of SMYD2 and its Role in Cell Proliferation. EMBO Chromatin and Epigenetic. Heidelberg, Germany.
- **Abu-Farha, M.**, Lambert J.P., Figeys, D. (2006). Analysis of Protein-Protein Interactions of a Novel Histone Methyltransferase. Progress in Systems Biology. Ottawa Institute of Systems Biology, Ottawa, Canada. **Abu-Farha, M.**, and Figeys, D. (2006). Characterization of a New Histone Lysine Methyltransferase and its Role in Microtubule Dynamics. Biochemistry, Microbiology & Immunology Research Day. Ottawa, Canada.
- **Abu-Farha, M.**, Sheng, Y., Figeys, D. (2005). The Role of SMYD2 in Cancer. Progress in Systems Biology. OISB, Ottawa, Canada.
- **Abu-Farha, M.**, and Willmore, W.G. (2005). Erythroid-Specific 5-Aminolevulinate Synthase Protein is Stabilized by Low Oxygen and Proteosomal Inhibition. Second Northern Lights Summer Conference, Canadian Federation of Biological Sciences 48th Annual Meeting. Guelph, Canada.
- **Abu-Farha, M.**, and Willmore, W.G. (2004). Post-Translational Modification and Protein Stabilization of ALAS2 under Hypoxia. Canadian Proteomic Initiative. Montreal, Canada.
- **Abu-Farha, M.**, and Willmore, W.G. (2003). Regulation of Erythroid-Specific 5-Aminolevulinate Synthase (ALAS2) By Hypoxia. Ottawa Carleton Chemistry Institute. Ottawa, Canada.
- **Abu-Farha, M.**, and Willmore, W.G. (2003). Hydroxilation and Protein Stability under Hypoxia. BioNorth. Ottawa, Canada.
- **Abu-Farha, M.**, and Willmore, W.G. (2003). Hypoxia Upregulates eALAS In K562 Cells. College of Natural Science. Ottawa, Canada.

RELEVANT ACTIVITIES

- Awarded various prestigious national and international awards such as NSERC post-doctoral and post-graduate scholarship, OGS and HUPO Young investigator award.
- Nominated for Carleton University Senator's Medal for Master's of science graduates.
- Finished Master's degree in biology and passed with distinction
- University teaching experience through working as a teacher's assistant for different biology courses at Carleton University and the University of Ottawa. I started working as a teacher's assistant during my first year of graduate study until the present day. I was always keen on teaching different courses, which will diversify and enhance my knowledge in the different field of biology.
- Mini-enrichment course instructor. I was responsible for giving this course to high school students. The course was designed to give high school students a general idea about advances in the field of biotechnology and genetic engineering. This was a challenging project for me where I had to make sure I explain biological concepts to students in a plain language that they can understand. 2005. Carleton University, Ottawa, Canada.
- As a graduate student I was involved in supervising fourth year project students. I had the chance to be responsible for two students in my study. This involved helping them in their literature search, setting up experiments and guiding them in the lab. This gave me the chance to be responsible for directing student's research projects and helping in finding solutions to the problems they face.
- Elected as the research vice president to Carleton University Biochemistry Society, 2002/2003.
- Named to Dean's Honor List 2000-2003.

REFERENCES

- Available upon request

9.0 APPENDIX III: Published Papers

1) **Abu-Farha M***, Lanouette S*, Couture JF, Figeys D. HSP90a a new target of SMYD2 methylation. (In preparation).

2) **Abu-Farha M***, Lanouette S*, Couture JF, Figeys D. The interactome of the SMYD protein family (In preparation).

3) **Abu-Farha M**, Lambert JP, Al-Madhoun AS, Elisma F, Skerjanc IS, Figeys D. The tale of two domains: proteomics and genomics analysis of SMYD2, a new histone methyltransferase. Mol Cell Proteomics. 2008 Mar;7(3):560-72.

4) **Abu-Farha M**, Elisma F, Figeys D. Identification of protein-protein interactions by mass spectrometry coupled techniques. Adv Biochem Eng Biotechnol. 2008;110:67-80.

5) **Abu-Farha M**, Elisma F, Zhou H, Tian R, Zhou H, Asmer MS, Figeys D. Proteomics: from technology developments to biological applications. Anal Chem. 2009 Jun 15;81(12):4585-99.

6) Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, Taylor R, Dharsee M, Ho Y, Heilbut A, Moore L, Zhang S, Ornatsky O, Bukhman YV, Ethier M, Sheng Y, Vasilescu J, **Abu-Farha M**, Lambert JP, Duewel HS, Stewart II, Kuehl B, Hogue K, Colwill K, Gladwish K, Muskat B, Kinach R, Adams SL, Moran MF, Morin GB, Topaloglou T, Figeys D. Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol. 2007;3:89.