ALLELE SPECIFIC APPROACH TO STUDY HISTONE DEMETHYLATION USING ENGINEERED KDM4A- (2) KETOGLUTARATE PAIRS

by

Megan Marie Breski

BS in Biology, Shippensburg University, 2013

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Megan Breski

It was defended on

April 1, 2015

and approved by

Candace Kammerer, PhD
Associate Professor
Department of Human Genetics
Graduate School of Public Health
University of Pittsburgh

Zsolt Urban, PhD
Associate Professor
Department of Human Genetics
Graduate School of Public Health
University of Pittsburgh

Committee Chair:
Kabirul Islam, PhD
Assistant Professor
Department of Chemistry
Dietrich School of Arts and Sciences
University of Pittsburgh

Kabirul Islam, PhD

ALLELE SPECIFIC APPROACH TO STUDY HISTONE DEMETHYLATION USING ENGINEERED KDM4A- (2) KETOGLUTARATE PAIRS

Megan Marie Breski, M.S.

University of Pittsburgh, 2015

ABSTRACT

KDM4 proteins are epigenetic modifiers that demethylate histone lysine residues to regulate chromatin structure and gene expression. The KDM4 subfamily contributes to the larger JmjC domain containing demethylase family, which all require 2-ketoglutarate as a cofactor for catalytic activity. KDM4 members have been shown to be relevant in biological processes, such as cellular differentiation and DNA damage repair. The amplification of KDM4 proteins has been observed in various cancers and disrupts the cell cycle, normal cellular proliferation and apoptosis. Being able to monitor the members of this protein family independently will elucidate their specific functions and the genes they regulate. Because this family of proteins is involved in various mechanisms related to pathogenesis it is of significant public health importance.

The Bump-Hole approach was used to develop an engineered KDM4A-2-Ketoglutarate system that would be independent of the native demethylase system. In total, 24 KDM4A mutants and 12 2-ketoglutarate analogs were generated, providing 288 engineered systems to be screened for activity. Half of the KDM4A mutants showed activity with 2-KG compounds and three of them provided a bioorthologous enzyme-cofactor system. The most significant engineered KDM4A-2-KG pairs were tested in KDM4 isoforms B and D and gave similar

results. Engineering the enzyme did not disrupt substrate specificity and the engineered systems were active following methylation of histone peptides.

In conclusion, half of the KDM4A mutants rendered the enzyme inactive when paired with the 2-KG compounds obtained (including the native compound). Half of the KDM4A mutants showed activity with at least one of the 2-KG compounds, where three of them resulted in a bioorthologous enzyme-cofactor system. The observations were alike in the KDM4 isoforms and substrate specificity was maintained. These results allow for the KDM4A protein to be monitored independently of the KDM4 isoforms and other JmjC domain containing proteins.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to thank Dr. Kabirul Islam for mentoring me and giving me the opportunity to work in his lab. I have learned several techniques and concepts that will benefit me in my future career as a scientist. I would also like to thank Debasis Dey and Babu Sudhamalla for teaching me in HPLC, FPLC and various other techniques that I was not familiar with. I give tremendous thanks to the entire Islam group for supporting and encouraging me in this process.

Thank you to my friends and family for their lasting support, encouragement and prayers throughout my graduate school journey. I am blessed to have met so many inspiring people to keep me determined and focused to reach my goals, so I give many thanks to all of them.

1.0 BACKGROUND AND SIGNIFICANCE

1.1 INTRODUCTION TO EPIGENETICS

Epigenetics is the study of gene regulation, chromatin organization and phenotypic changes that do not involve changes in the DNA sequence. The human genome is compacted into the structure of chromatin and epigenetic mechanisms regulate the transition between euchromatin (de-condensed chromatin that permits access of transcriptional machinery to DNA) and heterochromatin (tightly condensed chromatin keeping transcriptional machinery from accessing DNA). Nucleosomes are the principal components of chromatin and they consist of 146-147 bp of DNA wrapped around a histone octamer core. The octamer core includes two copies of the histones H2A, H2B, H3 and H4 (Figure 1). Linker DNA and histone H1 compact the core further to become a 30nm fiber [1-2].

Epigenetic post-translational modifications (PTMs) are temporary covalent chemical changes such as acetylation, methylation and phosphorylation etc. with DNA, RNA or histones as substrates. The general mechanism for PTMs to occur is the addition of the modification to the substrate (‘writing’), the processing and recognition of the modification (‘reading’) and the removal of the modification from the substrate (‘erasing’). This cycle of ‘writing’, ‘reading’ and ‘erasing’ requires various families of proteins and differs mechanistically depending on the substrate and modification [3-4].

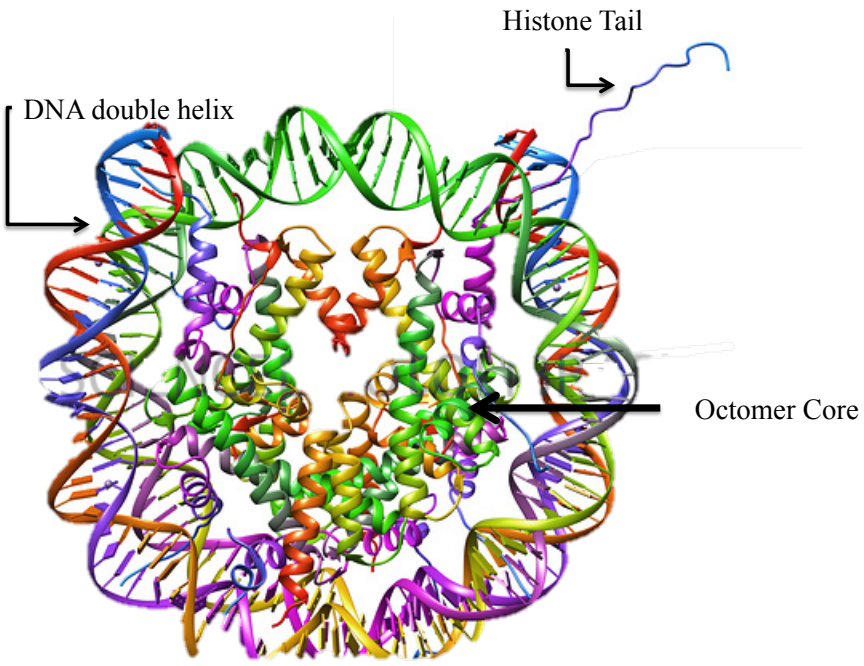


Figure 1: Nucleosome Core Crystal Structure

An image showing the DNA double helix, histone octamer core and a histone tail exposed to post translational modifications such as methylation and demethylation. The image was taken from SciencePhoto.com

The biological impact of PTMs is being explored, but is convoluted by the vast array of protein families involved in the diverse mechanisms. Hundreds of proteins have been identified to contribute to the dynamic process of PTMs and there is evidence that these proteins are implicated in diverse biological processes such as development, differentiation, imprinting and homeostasis. Several of the enzymes involved in epigenetic mechanisms have altered expression in various pathologies and disorders. For instance, KDM4 proteins (histone demethylases) are overexpressed in many types of cancers and FTO (RNA demethylase) is associated with obesity.

Because of necessity of these enzymes in the regulation of biological pathways and the implications of them in various diseases and public health in general, the understanding of their function and interactions can elucidate therapeutic targets.

1.2 SUBSTRATES OF POST TRANSLATIONAL MODIFICATIONS

PTMs occur on DNA, RNA and proteins and are regulated by specific enzymes for proper cell functioning. The various families of enzymes that are involved in PTMs have different mechanisms and levels of expression, but may have similar cofactors. A brief overview of PTMs on DNA and RNA will be discussed, but the focus will be on histone methylation.

DNA methylation mainly occurs at CpG islands where cytosine bases are converted to 5-methylcytosine (5mC) by DNA methyltransferases (DNMTs) and is associated with gene repression. DNMTs either act as *de novo* DNMTs, which bind unmethylated DNA and generate the initial methylation pattern or act as maintenance DNMTs, which use an existing DNA strand as a template to copy the methylation pattern after DNA replication occurs. Promoters and regulatory components are the most common region of the genome that becomes methylated, which may be due to a protective mechanism. About 70% of the promoters in the human genome are considered CpG island promoters, but in most cell types they are active and unmethylated. Transcription factors and other proteins with certain domains (e.g. CXXC zinc finger domain) bind unmethylated CpGs to keep regions from becoming methylated which can act as protection for promoters or as regulation for other regions of the genome. DNA demethylation can be characterized as either passive or active. Passive demethylation is replication dependent where newly synthesized DNA does not have the methylation pattern because PTMs are not replicated.

Active demethylation is replication-independent and is seen throughout different cell types and loci. The process of removing the methyl groups from cytosine nucleotides continues to be explored and is not fully understood. Methylated DNA binding domain-containing protein MBD2 and human ten-eleven-translocation (TET) proteins have been shown to act as a methyl 'erasers' on DNA by converting 5mC to 5-hemimethylcytosine (5hmC) requiring 2-ketoglutarate as a cofactor [5-7].

The stages of the DNA methylation process can be easily depicted through imprinting mechanisms. DNA methylation patterns are lost when fertilization occurs to regain the pluripotency status, but the imprinting-related DNA methylation is re-established in primordial germ cell stage. The 5mC maternal genome is lost after fertilization via passive demethylation whereas the 5mC paternal genome is lost in a matter of hours suggesting a globally active demethylation pathway. The DNA methylation pathway is essential for development, differentiation and regulation of gene expression as well as maintaining cellular identity. Methylation patterns and the activity of the proteins involved in DNA methylation are tissue and developmental stage specific [8].

Modifications in coding and non-coding RNA have been reported, but the function of the various modifications has yet to be elucidated. N^6-methyladenosine (m^6A) characterizes 0.1-0.4% total adenosine residues in the transcriptome and is found in mRNA, snRNA, tRNA and rRNA. In mRNA, there is approxiamately 1 m^6A every 2000 ribonucleotides and is sequence specific, although the consensus sequence for the m^6A modification varies depending on the type of RNA substrate. It is predominantly found in mature mRNA in coding sequences and near stop codons in the 3'UTRs, but not on the poly(A) tails. This modification is not randomly dispersed throughout the mRNA transcript, but it is commonly found in adjacent regions supporting the

idea that there is functional significance. There is also evidence suggesting that highly expressed microRNAs (miRNAs) may regulate methylation of their targets because they target transcripts with m^6A in the 3'UTR [8]. MicroRNAs (miRNAs) have also been shown to have methylation modifications that prevent miRNA maturation. The modifications on mRNA result in different phenotypes depending on the specific modification site. The m^6A mark at one site may lead to the recruitment of m^6A binding proteins to induce mRNA processing, mRNA trafficking or translation, but the modification at a different site could prevent the accessibility of RNA binding proteins for translating and this could explain the inverse correlation between mRNA copy number and protein abundance/expression [9].

Multi-protein complexes that harbor methyltransferase like 3 (METTL3), SAM (S-adenosylmethionine-binding domain), accomplish methylation of RNA. METTL3 has two functional domains, consensus methylation motif I (CM I) and consensus methylation motif II (CM II) which are the SAM binding domain and the catalytic domain respectively. FTO and ALKBH5 are known RNA demethylases and belong to the AlkB family. Characteristics of the AlkB family are the iron binding motif and the 2-KG interaction domain. The m^6A modification is known to regulate meiosis, metabolism and transcription making it an important biological mechanism [9].

The N-terminal histone tails extend from the nucleosomal core allowing for post-translational modifications (PTMs) to occur. With histones being an integral part of chromatin it is important to understand how they are being modified epigenetically and what it results it phenotypically [3-4].

Protein methyltransferases (PMTs) transfer methyl groups on to and off of various proteins and make up >50% of the chromatin-modifying enzymes in humans. The main class of

PMTs includes two families: protein lysine methyltransferases (PKMTs) and protein arginine methyltransferases (PRMTs) [10].

PMT activity on histone lysine residues can result in varying degrees of chemical states: mono-, di-, or tri- methylation and they correspond to unique phenotypic effects within the cells. For example, the phenotypic outcome correlated with trimethylation of H3K9 is a constitutively condensed heterochromatin state and trimethylation of H3K27 is a euchromatin state with gene silencing. PMTs catalyze the reaction of a methyl group transfer from SAM, the universal methyl donor, to the nitrogen acceptor of amino acids, predominantly lysine or arginine. Recognition of methylation states can be performed by PMTs that contain 'reader' domains, which can subsequently trigger downstream processes. There is a considerable amount of structural diversity within the substrate-binding pocket and substrate specificity among PMTs, which promotes the development of inhibitors to target individual PMTs specifically [10-11].

Methyltransferase G9a transfers methyl groups from SAM onto histone 3 (H3) lysine 9 (K9) and H3K27 through its SET (Su (var), enhancer of zeste, trithorax domain) domain. Specific amino acid residues within the histone substrates are essential for G9a activity. For instance, the arginine (R) at position 8 on histone 3 is adjacent to the lysine residue that is methylated by G9a and substituting it for any other amino acid completely abolishes the methylation of H3K9 by G9a. This shows that methyltransferases have site specificity for a specific sequence motif similar to DNMT and METTL3. G9a activity is also impaired by the phosphorylation of amino acid residues near H3K9. The motif recognized by G9a is found in several non-histone proteins and can be methylated. In fact, G9a is automethylated at its N-terminus resembling tri-methylated H3K9 (H3K9me3). Other non-histone protein substrates for G9a include DNMT1, HDAC1 (Histone deacetylase 1), p53 and Reptin. The phenotypic impact

of methylation on non-histone proteins has not been understood for all cases, but has been unmasked for some. For instance, at specific lysine sites reptin methylation induces a mechanism to cope with hypoxia and methylated p53 causes the protein to be transcriptionally inactive. Depletion or overexpression of G9a results in dysregulated differentiation of various cell types affecting brain, skeletal muscle and adipocyte development etc. and is observed in many different types of cancers [12-13].

Methylation of histones plays a role in diverse biological mechanisms. Histone methylation can regulate gene expression by antagonizing the effects that other histone modifications (e.g. histone acetylation) have on gene expression through the binding of chromodomains [14].

Histone demethylases remove methyl groups in a site and state specific manner. Many histone demethylases harbor 'reader' domains, such as the Tudor domain, to recognize the modifications and act as effector proteins. Histone demethylation is an important biological phenomenon and will be the focus of the document and expanded upon in the upcoming chapters [4].

Epigenetic modification patterns vary depending on developmental stage, cell type and tissue type. Several proteins are involved in the dynamic processes that regulate the addition and removal of the covalent epigenetic modifications and these proteins interact with each other and influence downstream pathways such as transcription, replication, and expression on the gene and protein level. For each substrate there are specific chemical mechanisms to explain how these modifications occur and why they occur where they do, but they are not fully understood. Many of the proteins involved have structural similarities with some specific differences and this allows for co-factors required for activity to be redundantly used to interact with catalytic

domains and for site specific modification. For example, 2-KG is a co-factor that can be used by several families of methyltransferases and demethylases [10-13].

1.3 DYNAMICS OF HISTONE METHYLATION AND DEMETHYLATION

Several histone modifications collectively regulate and play a role various biological processes such as transcription, DNA replication and DNA repair. Understanding the effects of specific histone methylation modifications on chromatin regulation and function has become possible with the characterization of various histone methyltransferases. As the comprehension of histone methylation expands, generalizing the correlation of specific histone marks with certain phenotypic outcomes (e.g. transcriptional silencing vs. transcriptional activation) becomes more complex. Generally, H3K9 and H3K27 methylation is correlated with transcriptional silencing, but both have been observed in some regions of transcriptionally active chromatin. Also, methylation of H3K4 and H3K36 are associated with transcriptional activation, but can be seen in chromatin regions of transcriptional silencing. As stated earlier, there are various methylation states that histone residues undertake: mono- (me1), di- (me2) and tri- (me3) and are most abundant on histone H3 and H4 lysine or arginine residues. Methylated lysine residues have also been observed on linker histone H1. These modification states dictate which effector proteins recognize the modification and vary throughout development, tissue type and pathologies. Table 1 shows a comprehensive list of general associations between histone methylation marks and biological function. Methylation modification state can increase (via methyltransferases activity) and decrease (via demethylase activity) providing a dynamic chemical process essential for various biological pathways [15].

Table 1. Methylation State Functions of Histone Lysine Residues

Methylation Site	Methylation State	Associated Functions
H3K4	me1	transcriptional activation in enhancer regions
	me2	Gene activation in promoter regions
	me3	Gene activation in promoter regions
H3K9	me1	Gene activation [18]
	me2	Transcriptional repression [18]
	me3	Transcriptional repression [18]
H3K27	me1	Gene activation [18]
	me2	Transcriptional repression
	me3	Transcriptional repression [18]
H3K36	me1, me2, me3	Transcriptional elongation, association with RNA Pol (II) and histone deacetylation (18,24)
H3K79	me1	Gene activation [19]
	me2	Gene activation [25]
	me3	Transcriptional repression [18-19]
H4K20	me1	Gene activation, DNA replication, DNA repair, Mitotic Regulation [18]
	me2	Gene activation, DNA replication, DNA repair [19]
	me3	Transcriptional repression [18-19]
H1.4K26	me1	Gene activation
	me2	Gene activation
	me3	transcriptional repression; heterochromatin formation [19]

Three classes of enzymes that remove histone methylation have been characterized: peptidylarginine deiminases (PAD), flavin-dependent lysine specific demethylases (LSD) and JmjC domain containing demethylases (JMJ). Each class requires specific cofactors and target residues in a site and state specific manner [15].

The PAD family targets methylated arginine residues on histone 3 and histone 4. They are not necessarily considered histone demethylases because they convert methylated histone arginine to citrulline (citrullination) rather than reversing arginine methylation directly. PADs require calcium and dithiothreitol (DTT) to perform citrullination where the arginine side chain

guanidine group is converted to a ureido group. There are five isoforms in the PAD family (PAD1-4 and PAD6) and they each have tissue specific expression patterns. PADs are involved in myelin formation and hair growth and have been implicated in various pathologies such as cancer, rheumatoid arthritis, Alzheimer's and multiple sclerosis [16-17]

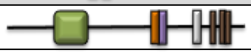
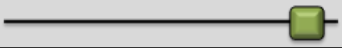
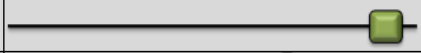
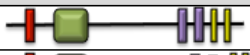
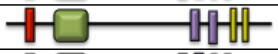
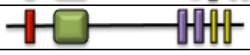
The discovery of the first lysine demethylase (KDM), termed Lysine Specific Demethylase 1 (LSD1) in 2004 eliminated the idea that lysine methylation is irreversible. LSD1 catalyzes the removal of methyl groups from H3K4me1 and H3K4me2 as well as some non-histone proteins. LSD1 is a flavin dependent amine oxidase and consists of an amine oxidase catalytic domain, an N-terminal SWIRM domain involved in recognition of the specific target and an N-terminal extension. LSD1 can be thought of as an inducer of transcriptional silencing because it specifically targets H3K4me1/2, which is known to be associated with transcriptional activation, and has been shown to be involved in protein-protein interactions with transcriptional repressive complexes (e.g. CoREST and HDAC1/2) [23].

1.4 JMJC DOMAIN CONTAINING DEMETHYLASES

The human genome encodes 32 JmjC-domain containing genes with 24 of them showing demethylase activity. There are various combinations of domains (e.g. PHD, Tudor, ARID, JmjN) incorporated in JmjC-domain containing proteins, which define their function. The 24 JmjC domain containing demethylases are categorized into seven subfamilies based on structure and sequence homology and each target specific methylated lysine residues based on sequence and methylation state. A comprehensive list is shown in **Table 2** [26-27].

There are five functional KDM4 genes: *KDM4A-E*, which are localized to various chromosomes. Interestingly, *KDM4D* is an intronless gene that clusters with *KDM4E* and *KDM4F*. *KDM4F* is considered a pseudogene as was *KDM4E*, but it was recently suggested that it is a functional gene by evidence for its expression. Orthologs of KDM4A-C are seen in all vertebrates whereas orthologs of KDM4D are seen only in placental mammals providing support that the five membered KDM4 subfamily is highly conserved [26-28].

Table 2. Human demethylase chromosomal location, protein structure and histone substrates.

Demethylase	Synonyms	Gene loci	Domains	Histone Targets
KDM1A	LSD1, KDM1, AOF2, BHC110	1p36.12		H3K4me1/me2, H3K9me1/me2
KDM1B	LSD2, AOF1, C6orf193	6p22.3		H3K4me1/me2
KDM2A	JHDM1A, FBXL11, CXXC8, FBL11, FBL7	11q13.2		H3K36me1/me2
KDM2B	JHDM1B, FBXL10, PCCX2 CXXC2	12q24.31		H3K4me3, H3K36me1/me2
KDM3A	JHDM2A, JMJD1A, TSGA	2p11.2		H3K9me1/me2
KDM3B	JMJD1B, NET22	5q31		H3K9me1/me2
KDM3C	KDM3C, TRIP8	10q21.3		H3K9me1/me2
KDM4A	JMJD2A, JHDM3A, JMJD2	1p34.1		H3K9me2/me3, H3K36me2/me3, H1.4K26/me2/me3
KDM4B	JMJD2B	19p13.3		H3K9me2/me3, H3K36me2/me3, H1.4K26/me2/me3
KDM4C	JMJD2C, JHDM3C, GASC1	9p24.2		H3K9me2/me3, H3K36me2/me3, H1.4K26/me2/me3
KDM4D	JMJD2D	11q21		H3K9me2/3, H1.4K26me2/me3
KDM4E	JMJD2E, KDM4DL	11q21		H3K9me2/me3, H3K56me3
KDM5A	JARID1A, RBP2	12p11		H3K4me2/me3
KDM5B	JARID1B, CT31, RBBP2H1A	1q32.1		H3K4me2/me3
KDM5C	JARID1C, SMCX, MRXJ	Xp11.22		H3K4me2/me3
KDM5D	JARID1D, HY, SMCY	Yq11		H3K4me2/me3
KDM6A	UTX, KABUK2	Xp11.2		H3K27me2/me3
KDM6B	JMJD3	17p13.1		H3K27me2/me3
KDM7A	JHDM1D	7q34		H3K9me1/me2, H3K27me1/me2
KDM7B	PHF8, JHDM1F, MRXSSD, ZNF422	Xp11.22		H3K9me1/me2, H4K20me1
KDM7C	PHF2, JHDM1E, CRC5	9q22.31		H3K9me2
MINA	MDIG, FLJ14393	3q11.2		H3K9me3
NO66	MAPJD	14q24.3		H3K4me1/me3, H3K36me2
KDM8	JMJD5	16p12.1		H3K36me2

SWIRM domain: pink; Amine oxidase domain: dark green; CW-wild type zinc-finger domain: light pink; JmjC domain: light green; CXXC zinc-finger domain: orange; PHD plant homeodomain: purple; F-box domain (FBOX): white; Leu-rich repeat domain (LRR): brown; JmjN domain: red; Tudor domain: yellow; AT-rich interacting domain (ARID): Blue; C5HC2 zinc-finger domain: grey; Tetratricopeptide domain (TPR): aqua.

1.5 KDM4 PROTEIN MEMBER STRUCTURE AND SPECIFICITY

KDM4A-C are the larger proteins within this subfamily and share greater than 50% sequence identity. The domains within each member play essential roles for correct protein function. KDM4A-C isoforms contain N-terminal JmjN and JmjC domains and on the C-terminal contain two PHD and two Tudor domains. KDM4D and KDM4E lack the PHD and Tudor domains. The JmjC domain is the catalytic domain and is therefore required for demethylase activity in this KDM4 subfamily. Structural integrity of the protein is provided by the JmjN domain, which also interacts with the catalytic domain. The PHD and Tudor domains act as reader domains and recognize specific methylated lysines. The KDM4A Tudor domains recognize and bind H3K4me2/me3 and H4K20me2/me3 whereas the PHD domains bind methylated and unmodified residues on many histone tails, providing a more flexible recognition pathway of histone modifications. KDM4A-C are known to demethylate H3K9me2/me3 and H3K36me2/me3, where KDM4D and –E demethylate only H3K9me2/me3. Crystallography has shown differences in the surface residues of the peptide binding pocket between KDM4A-C and KDM4D-E. The structural variation within the peptide binding pocket of KDM4 isoforms alters the electrostatic properties causing interactions with peptides to be disrupted. The variation of peptide-enzyme interactions may explain the difference in activity toward H3K36 [26].

KDM4A demethylates trimethylated lysine residues more efficiently than dimethylated residues and has approximately 5-fold higher activity towards H3K9me3 than H3K36me3. X-ray crystallography analysis of KDM4A has shown that the electrostatic environment in the methyl group-binding pocket does not allow productive positioning toward the iron-containing catalytic domain of the methyl group on monomethylated lysine residues. This further supports the methylation state specificity of these KDM4 proteins [27-29].

1.6 KDM4 DEMETHYLATION MECHANISM

The KDM4 proteins are JmjC domain containing demethylases requiring Fe(II), 2-KG and molecular oxygen (O_2) for the catalysis of a dioxygenase reaction that results in demethylated lysine residues on histones. In the active site, first a His-Glu/Asp-His triad positions the Fe(II) and then 2-KG and oxygen bind forming oxoferryl (IV) which is highly reactive. 2-KG is decarboxylated and the result is succinate and CO_2. The methyllysine target is hydroxylated resulting in an unstable intermediate (hydroxymethyl moiety), which spontaneously decomposes to formaldehyde and the demethylated product in either the di- or mono- methylated state. The released formaldehyde is oxidized by FDH, which also reduces NAD^+ to NADH [27-29]. Figure 2 depicts the JmjC domain containing demethylation mechanism.

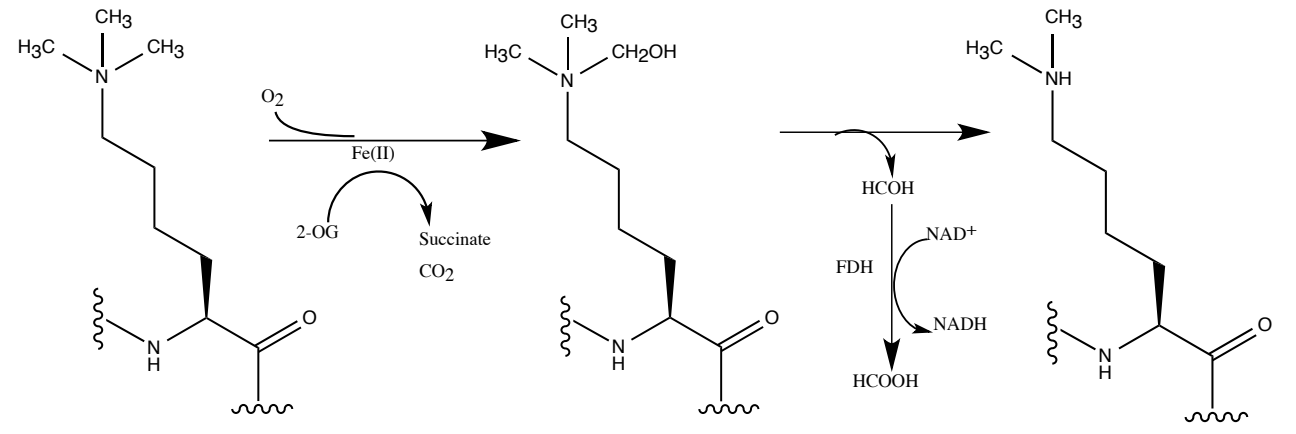


Figure 2: JmjC domain containing demethylase mechanism.

1.7 KDM4 EXPRESSION

The normal expression patterns of the *KDM4* genes suggest that the associated proteins have non-overlapping biological functions in diverse cell types and are regulated by distinct pathways. *KDM4A-C* are ubiquitously expressed in human tissues with the highest expression being in the spleen, ovaries and colon. *KDM4A* and *–C* are expressed 3-6 fold greater than *KDM4B*. *KDM4D* and *KDM4E* are mainly expressed in the testes, but *KDM4E* expression is low relative to the other *KDM4* genes. To elucidate physiological functions of this subfamily, knockout and transgenic models have been established. Homozygous mutants of *KDM4A* and *B* in *Drosophila* are not viable and die. Heart-specific knockout of *KDM4A* and transgenic models in mice results in cardiac hypertrophy. Conditional knockout of *KDM4B* in specific cell types results in delayed tissue development (e.g. mammary epithelial cells). Interestingly, knockout of *KDM4B* or *KDM4D* in mouse models are compatible with life and do not show extreme abnormalities [27-29].

The KDM4 family is known to have important functions during development. Specifically, they maintain the state of euchromatin necessary for efficient proliferation and differentiation in embryonic stem cells. During development, the specific functions of KDM4 proteins are widespread with the promotion of pluripotency and differentiation both observed to be regulated by these proteins. Various tissues rely on the proper functioning of the KDM4 proteins. For instance KDM4A directs embryonic skeletal cell differentiation, KDM4B directs bone cell differentiation and KDM4C directs fat cell differentiation. A pluripotent state can be achieved through KDM4 proteins interacting with or promoting the expression of pluripotency factors such as c-Myc, Oct4 and Sox2. Oct4 and Sox2 are involved in an expression feedback loop that is required for the de-differentiation of adult stem cells. The signaling involved within

the feedback loop is eliminated when KDM4C expression is absent and Oct4 expression is decreased when the cells lack KDM4A [27-29].

1.8 REGULATION OF KDM4 PROTEINS

Cells have various mechanisms to control the KDM4 protein activity, expression and localization, which is not surprising based on the family having significant biological impact. There are pathways, transcription factors, hormones and protein complexes that are involved in the regulation of the KDM4 proteins. The ubiquitination pathway can regulate KDM4A and KDM4B during cell cycle progression in response to DNA damage. Two SCF (Skp, Cullin, F-box containing) complexes: SKP1-CU11-FBox and FBXO22 regulate KDM4A during cell cycle progression by controlling its ubquitination and inducing protein degradation. RNF8 and RNF168 complexes of the ubiquitination pathway are involved in the regulation of KDM4A-B when DNA damage has been detected. PARP-1 can act on the KDM4A-D affecting the H3K9 demethylation function by poly(ADP-riboxyl)ating the glutamic acid residues within the proteins. The demethylation activity of KDM4C is also regulated by IP6K1, which induces the dissociation of KDM4C [30-33].

Transcription factors, such as HIF1, can also regulate KDM4B and C. HIF1 is a protein that responds to hypoxia to regulate cellular and systemic homeostasis. In breast cancer, HIF1-alpha interacts with KDM4C resulting in the recruitment of HIF1-alpha target gene response elements. Androgens can regulate KDM4B activity, which will affect androgen receptor transcriptional activity through demethylation or androgen receptor ubiquitination [34-35].

Homodimers and heterodimers of the KDM4 protein members can form resulting in another regulatory mechanism. KDM4A-C can become incorporated into multiprotein complexes with the SWI/SNF chromatin-remodeling complex and can interact with inhibitory complexes such as histone deacetylase complexes (HDACs). The formation of dimers and incorporation into functionally diverse multiprotein complexes supports that KDM4 proteins have a significant role in directing gene expression in development, homeostasis and pathology [36-38].

1.9 KDM4 MEMBERS AND CANCER

The overexpression or underexpression on the gene and protein level of the KDM4 subfamily has been implicated in various cancers. There have been large analyses of cancer copy number profiles and findings by independent research groups studying the proteins to support the involvement of this family in cancer development or progression.

KDM4A overexpression and amplification is seen in several cancers such as ovarian, breast, squamous cell carcinomas, prostate, lung and colon cancer [39-43]. KDM4A overexpression in tumors triggers localized chromosomal instability, specifically 1q12, 1q21 and Xq13.1 copy gains, which are rich with oncogenes. Overexpression of KDM4A can lead to these localized copy gains by extending S phase via reducing HP1**γ** binding to chromatin. When HP1**γ** cannot bind to chromatin, it results in an open, accessible chromatin state allowing for the re-replication of oncogenic regions. HP1**γ** bound to chromatin hinders KDM4A from being able to modify histones. The antagonistic behavior between KDM4A and HP1**γ** may contribute to the explanation as to how KDM4A encourages DNA replication [39]. Overexpression of KDM4

proteins also disrupts the DNA mismatch repair (MMR) pathway resulting in poor chromatin integrity. Mismatch recognition proteins bind H3K36me3 to ensure proper DNA MMR occurs. When KDM4C is overexpressed, H3K36me3 is less abundant hindering mismatch recognition proteins from binding appropriate regions. The dysregulated pathway allows for mutations to go overlooked and promote genome instability. Proper DNA MMR can be restored through the downregulation of KDM4C. Because KDM4A and KDM4B demethylate H3K36me3, similar results are expected with the overexpression of these proteins [40].

KDM4A overexpression can also allow cells to evade apoptosis or decreases apoptosis all together by extending the G2/M phase. The activating protein 1(AP1) transcription factor is a heterodimer of the JUN and FOS proteins, which controls apoptosis, cell proliferation and differentiation. The activity of AP1 is induced by a positive feedback loop where KDM4A is a priming factor. KDM4A can indirectly induce the expression of AP1 genes because histone demethylation facilitates AP1 binding to *JUN* and *FOSL1* promoters, which stimulates the feedback loop. The overexpression of KDM4A can result in the overexpression of AP1 genes, which will promote cell growth and metastasis [41].

KDM4A can also repress the expression of tumor suppressors, such as CHD5, by impeding senescence. KDM4A can directly bind with tumor suppressors (e.g. p53) rendering them inactive and can form complexes with the androgen and estrogen receptors making them more sensitive resulting in the overexpression of oncogenes. ER-positive breast cancer cells with depleted KDM4A showed decreased expression of oncogenes c-Jun and cyclin D1. Proliferation inhibition of ER-negative and ER-positive breast cancer cells was observed with KDM4A knockdown. These results show that KDM4A is necessary for ER-positive and –negative breast cancer and supports the idea that there are multiple pathways that KDM4A is involved in, which

can influence breast cancer development and/or progression. Contrasting with the results for breast cancer, the overexpression of KDM4A in HeLa cervical carcinoma did not result in tumor growth. With opposing results for the overexpression of KDM4A in different cancer cell lines, general conclusions cannot be made based on the expression of this protein, but rather it is cancer and tissue specific [28, 38-43].

KDM4B plays a significant role in nuclear receptor responsiveness in breast and prostate cancer. KDM4B is highly expressed in estrogen receptor (ER)-positive breast cancer subtypes, can bind to the ER, act on the repressive mark of H3K9me3, and recruit chromatin remodeling complexes. The activity of KDM4B when bound with the ER can induce oncogenic expression (e.g. MYB and MYC) resulting in proliferation. KDM4B expression level is positively correlated with prostate cancer severity, potentially due to it interactions with the androgen receptor (AR). AR transcription is upregulated and the AR is stabilized when cooperating with KDM4B via degradation inhibition. KDM4B is involved in metastasis and hypoxia and can promote cell survival [35, 44-46].

The overexpression of KDM4C has been observed in breast and prostate cancer, esophageal squamous cell carcinoma, metastatic lung sarcomatoid carcinoma, non-solid tumors and medulloblastoma. KDM4C can promote tumorigenesis by activating oncogenes (e.g. MDM2), binding the AR nuclear receptor and establishing stem cell-like phenotypic characteristics. Overexpression of KDM4C resulted in hypomethylation of H3K9 and knockdown studies in breast cancer cells results in the inhibition of proliferation and metastasis [47-49].

It has become clear that the KDM4 subfamily members play various and diverse roles in cancer development and progression. The dysregulation of KDM4 proteins have been correlated

to other diseases due to the disruption of cellular functions. Upregulation of KDM4A has been associated with cardiac anomalies and the progression of viral infections and SNPs within KDM4C genes are correlated with autism. The functions of the KDM4 proteins are widespread and not fully understood, but continue to be elucidated [50-52].

1.10 KDM4A INHIBITORS AND THERAPEUTICS

Many histone demethylase inhibitors have been designed and are classified into five groups: 2-ketoglutarate analogs, iron chelators, zinc chelators, catalytic domain inhibitors, and prodrugs. N-oxalylglycine (NOG) and its derivatives are analogs of 2KG and are competitive inhibitors of the KDM4 family. Another competitive 2KG analog inhibitor is 2-hydroxyglutarate (2-HG), which is an oncometabolite. The reported therapeutic inhibitors lack selectivity and specificity towards histone demethylases, which becomes problematic regarding nondesirable targets. Therapeutic inhibitors for histone demethylases remain in the preclinical phase [53-55].

1.11 KDM4 AND PUBLIC HEALTH

With epigenetics rapidly expanding and implicated in critical biological mechanisms it is important to understand the enzymes involved. The JmjC domain-containing proteins include several subfamilies each containing multiple isoforms. The KDM4 subfamily has been and continues to be extensively studied because its members have been observed in various pathologies and involved in many mechanisms such as the regulation of DNA replication, gene

expression, transcription and cell cycle progression. Because KDM4A overexpression is observed in several types of cancer and results in various phenotypes, increased understanding of this enzyme will contribute to improving public health.

1.12 ALLELE SPECIFIC APPROACH

Approximately 30,000 protein-coding genes give rise to massive diversity among proteins, which are responsible for the complex biochemical processes that occur within the human body. Many of these proteins share similar structures and domains allowing them to use the same cofactor while having a different function. This idea is true for many above-mentioned demethylases that require 2-ketoglutarate as a cofactor. Families of proteins often contain the same domains, which allow the human body to be more efficient, but make studying the function, localization and impact of a specific protein increasingly difficult.

Various techniques and approaches have been applied to understand the role of specific proteins such as introducing single nucleotide mutations (for gain- or loss-of function) and knockout/knockdown and observing the phenotypic effects. These techniques have been beneficial, but the conclusions are limited based on the observations. It is known that many protein-protein complexes form to carry out many processes, there are compensatory mechanisms in place for adaptability and biological processes often occur on the millisecond to minute time scale (e.g. PTMs). With these characteristics and biological mechanisms in place, it is challenging to determine if the phenotypic effects are due to the under- or overexpression of the targeted protein or if there is involvement from other proteins that is not being monitored.

Therefore, it is necessary to use approaches that can be more specific, controlled and timely to monitor the function of a protein.

Allele-specific chemical genetics (ASCG) incorporates different design strategies to develop systems to study widespread proteins and associated ligands. The approaches most commonly used are electrostatic, steric, covalent and unnatural amino acid mutagenesis and all of them involve incorporating novel interactions between the protein and ligand that has been engineered.

The approach used in this study is steric complementation, aka Bump-Hole approach. The principle of this study design is to engineer the protein of interest and its associated ligand to create a complementary shape between them. A bulky group is introduced onto the ligand (Bump) and is accommodated by the space (Hole) created through site directed mutagenesis on the enzyme. The native enzyme due to steric hindrance will not accept the bulky ligand and the engineered enzyme will not accept the native ligand because of lack of shape/ size complementarity. This creates an allele specific activator or inhibitor for the modified enzyme that can be used to monitor isoforms within the same family in a completely independent fashion (Figure 3) [54].

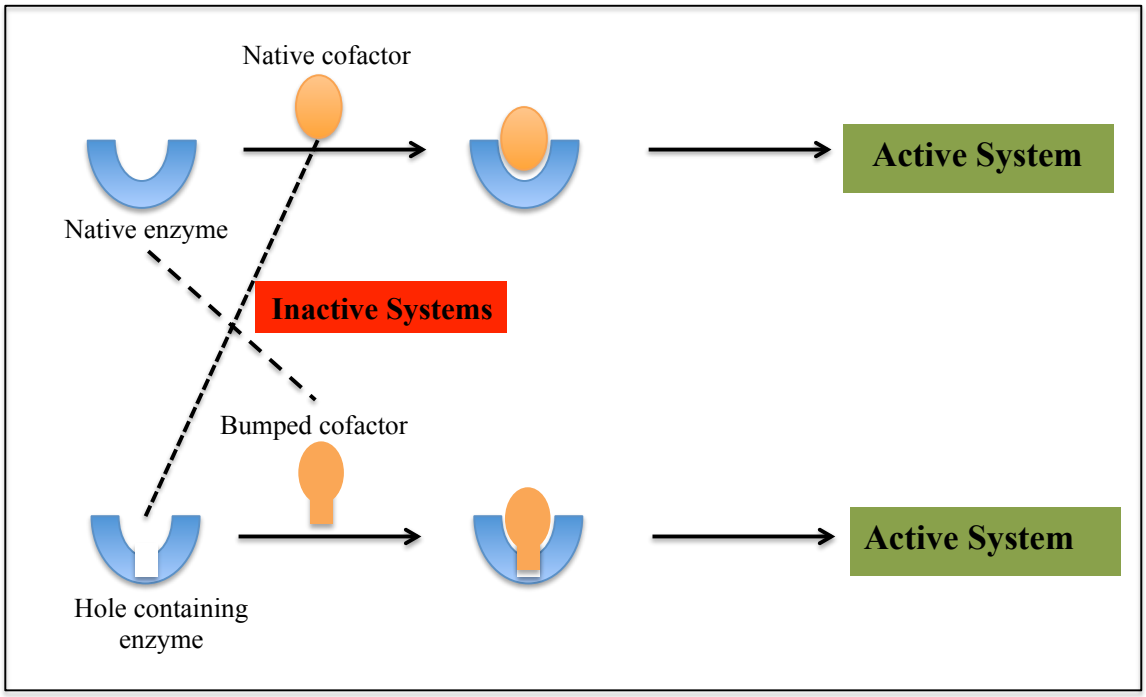


Figure 3. Allele Specific Bump-Hole Approach Scheme

Figure 3 shows the bump-hole approach where the native enzyme is mutated forming a 'hole' in the binding pocket and the cofactor is modified to incorporate a 'bump' that will be complementary to the hole containing enzyme.

1.13 OBJECTIVES

It is essential to understand the mechanism and be able to monitor each KDM4 isoform independently due to their differential expression, non-redundant functions, disease-specific activity and temporal activity. The allele-specific approach is more appropriate than previous methods for studying protein isoforms in the same family because of its rapid effect. For example, RNAi takes more than 10 hours for the method to take effect, whereas the allele-specific approach is on the nanosecond scale. This is relevant because many biological

mechanisms are time-dependent and takes only seconds (e.g. PTMs) or minutes (e.g. mitosis) to occur.

The main objective of this project is to develop an engineered system for KDM4A *in vitro* that is allele specifically activated. The system will involve engineering the catalytic domain of KDM4A and the fourth carbon of 2-KG, which will act independently of and not influence the native KDM4A-2KG pair. The engineered pair should be site and state specific, similar to the native pair. This engineered system should be able to mimic the native KDM4A system and should give consistent results in the KDM4 isoforms. First, the Bump-Hole approach will be used to create the engineered system and all engineered proteins and 2-KG analogs will be screened for enzymatic activity in the presence of trimethylated histone peptides. The substrate specificity will be tested with the engineered system(s) that show similar activity to the native system. The K_m for each engineered pair with substrate specificity will be compared to the K_m of the native system. Next, the KDM4A-mutant-2KG analog pair with similar activity to the native system, but independent of it, will be applied to an *in vitro* coupled assay mimicking the reading, writing and erasing cycle. Finally, the engineered mutant with the most similar activity to the native will be applied to other members within the KDM4 subfamily, which will be tested with significant 2-KG analogues *in vitro*. Below is a list of objectives to fulfill the main goal of developing an allele specific engineered KDM4A-2-KG pair:

1. Generate engineered KDM4A catalytic domain and 2-ketoglutarate analogues
2. Perform screening of engineered KDM4A-2KG pairs for bioorthologous system
3. Test bioorthologous system in KDM4 isoforms for conservation of results
4. Determine if KDM4A substrate specificity remains intact for engineered system(s)

5. Determine if engineered KDM4A-2KG pair(s) are catalytically competent

2.0 MATERIALS AND METHODS

2.1 PLASMIDS

The catalytic domain of the human KDM4A plasmid was obtained from addgene (cat# 38846) and includes an N-terminal hexahistidine tag (HHHHHH) and kanamycin resistance (the native KDM4A catalytic domain will be referred to as KDM4A-WT). The catalytic domains KDM4A, B and D were obtained from Raymond Trievel's group and include an N-terminal strep tag (WSHPQFEK) and ampicillin resistance. The plasmids were purified as described above and verified through sequencing. All plasmids were purified using Fisher Scientific GeneJet Plasmid Miniprep Kit and the concentrations were measured using Eppendorf™ BioSpectrometer. The purified plasmids were sent for sequencing through the University of Pittsburgh's Genomics facility or GeneWiz. The nucleotide sequencing data was translated to the corresponding amino acid sequence and aligned with the associated NCBI FASTA sequence to verify that the sequence is correct.

2.2 SITE DIRECTED MUTAGENESIS

The primers (Table 3) for site directed mutagenesis were designed using Agilent Technologies QuikChange Primer Design Tool (http://www.genomics.agilent.com/primerDesignProgram) and the Integrated DNA Technologies OligoAnalyzer 3.1 (http://www.idtdna.com/calc/analyzer).

Site directed mutagenesis (SDM) was done using the QuikChange Lightning Kit (Table 4). The PCR products were treated with DPNI to digest the parent plasmid and then transformed in XL10-Gold ultracompetent cells with **β** –mercaptoethanol (BME) to enhance transformation efficiency. The transformed plasmids were then inoculated in LB-broth and grown overnight at 37°C while shaking at 225 RPM. KDM4A mutants will be referred to as 'KDM4A-mutation' (e.g. KDM4A-Y132A). To verify the mutation was successfully incorporated, the plasmid was sent for sequencing through GeneWiz.

Table 3. Site Directed Mutagenesis Primers

MUTATION	FORWARD PRIMER
KDM4A-Y132A	5'-CATTCAATCCTCCAATC**GCC**GGTGCAGATGTGAATGG-3'
KDM4A-Y132G	5'-CATTCAATCCTCCAATC**GGC**GGTGCAGATGTGAATGG-3'
KDM4A-Y175A	5'-GAGGGTGTGAACACCCCA**GCC**CTGTACTTTGGCATG-3'
KDM4A-Y175G	5'-GGTGTGAACACCCCA**GGC**CTGTACTTTGGCATG-3'
KDM4A-Y177A	5'-GTGAACACCCATACCTG**GCC**TTTGGCATGTGGAAGAC-3'
KDM4A-Y177G	5'-GTGAACACCCCATACCTG**GGC**TTTGGCATGTGGAAGAC-3'
KDM4A-F185A	5'-GGCATGTGGAAGACATCC**GCC**GCTTGGCACACTGAAGA-3'
KDM4A-F185G	5'-GGCATGTGGAAGACATCC**GGC**GCTTGGCACACTGAAG-3'
KDM4A-F185I	5'-GGCATGTGGAAGACATCC**ATC**GCTTGGCACACTGAAG-3'
KDM4A-F185T	5'-GGCATGTGGAAGACATCC**ACC**GCTTGGCACACTGAAG-3'
KDM4A F185V	5'-GGCATGTGGAAGACATCC**GTG**GCTTGGCACACTGAAG-3'
KDM4A-N198A	5'-GACCTCTACAGCATC**GCC**TACCTGCACTTTGG-3'
KDM4A-N198G	5'-GGACCTCTACAGCATC**GGC**TACCTGCACTTTGGAG-3'
KDM4A-K206A	5'-CTGCACTTTGGAGAACCA**GCC**TCCTGGTACTCTGTTCC-3'
KDM4A-K206G	5'-CTGCACTTTGGAGAACCA**GGC**TCCTGGTACTCTGTTCC-3'
KDM4A-W208A	5'-CTTTGGAGAACCAAAGTCC**GCC**TACTCTGTTCCACCTGAG-3'
KDM4A-W208F	5'-GGAGAACCAAAGTCC**TTC**TACTCTGTTCCACCTGAG-3'
KDM4A-W208G	5'TGGAGAACCAAAGTCC**GGC**TACTCTGTTCCACC-3'
KDM4A-W208V	5'GGAGAACCAAAGTCC**GTG**TACTCTGTTCCACCTG-3'
KDM4A-A286G	5'-CCATGGTTTTAACTGT**GGC**GAGTCTACCAATTTTGC-3'
KDM4A-S288A	5'-GGTTTTAACTGTGCGGAG**GCC**ACCAATTTTGCTACCC-3'
KDM4A-S288G	5-GGTTTTAACTGTGCGGAG**GGC**ACCAATTTTGCTACCCG-3'
KDM4B-F186G	5'-CATGTGGAAGACCACC**GGC**GCCTGGCACAC-3'
KDM4D-F189G	5'- GCATGTGGAAAACCACG**GGC**GCTTGGCATACAGAG-3'

Above is a list of primers designed for Site Directed Mutagenesis. Reverse primers used are the reverse complement to the forward primer. Mutations coding for the amino acid change are bold and underlined.

Table 4. Cycling parameters for Site Directed Mutagenesis

	Site Directed Mutagenesis		
	Step	**Temperature (°C)**	**Time**
	Initial denature	95	2 minutes
18 cycles	Denaturing	95	20 seconds
	Annealing	60	10 seconds
	Elongation	68	4.25 minutes
	Final Elongation	68	5 minutes
	Storage	4	∞

The parameters were chosen based on the QuikChange Lightning Site Directed Mutagenesis kit. The Elongation period was calculated by giving it 30 seconds/kb of plasmid. PCR products were treated with DPNI and then transformed into XL10 Gold Ultracompetent cells or stored in -20°C until further use.

2.3 PROTEIN EXPRESSION AND PURIFICATION

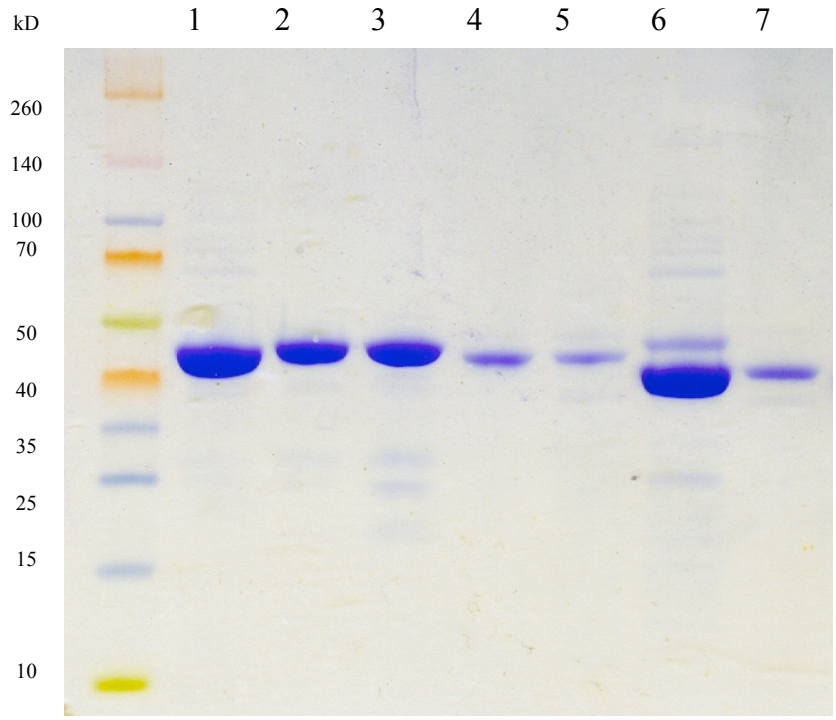
All plasmids were transformed in a strain of chemically competent *E. coli* (Table 5) for protein expression once the sequencing results verified that the protein sequence was correct and that the mutation (if relevant) was in place without disrupting other areas of the protein. The transformants were inoculated in 10 ml LB broth and grown overnight at 37°C while shaking at 225 RPM. The inoculations were then grown in two liters of LB-broth at 37°C to reach an optical density (OD_{600}) of 0.6-0.8 (~4 hours) and protein expression was induced with IPTG (0.3-1mM). Proteins were expressed overnight at 17°C.

The KDM4A native and mutant proteins containing the hexahistidine tag or the strep-tag were first purified through Nickel Affinity Chromatography or StrepTactin Affinity Chromatography, respectively, using gravity flow columns. Then the fractions containing protein observed through SDS-PAGE were purified through Size Exclusion Chromatography using Superdex-200 in FPLC. The fractions containing protein supported by SDS-PAGE (Figure 4) were concentrated using Millipore centrifugal filter units and the concentration was determined using Bradford solution. The proteins were stored in -80°C until further use.

Table 5. *E. coli* competent cells used for protein expression

Protein	Expression Cell line
KDM4A-WT	BL21 [de3] pLysS
KDM4A-Y132A	BL21 [de3] pLysS
KDM4A-Y132G	BL21 [de3] Codon plus
KDM4A-Y175A	BL21 [de3] pLysS
KDM4A-Y175G	BL21 [de3] Codon plus
KDM4A-Y177A	BL21 [de3] pLysS
KDM4A-Y177G	BL21 [de3] Codon plus
KDM4A-F185A	BL21 [de3] pLysS
KDM4A-F185G	BL21 [de3] Codon plus
KDM4A-F185I	BL21 [de3] Arctic Express
KDM4A-F185T	BL21 [de3] Arctic Express
KDM4A-F185V	BL21 [de3] Arctic Express
KDM4A-N198A	BL21 [de3] Codon plus
KDM4A-N198G	BL21 [de3] Arctic Express
KDM4A-K206A	BL21 [de3] pLysS
KDM4A-K206G	BL21 [de3] Arctic Express
KDM4A-W208A	BL21 [de3] pLysS
KDM4A-W208G	BL21 [de3] Codon plus
KDM4A-W208V	BL21 [de3] Arctic Express
KDM4A-W208F	BL21 [de3] Codon plus
KDM4A-A286G	BL21 [de3] Arctic Express
KDM4A-S288A	BL21 [de3] Codon plus
KDM4A-S288G	BL21 [de3] Codon plus
KDM4A-F185G/Y132A	BL21 [de3] Arctic Express
KDM4A-F185G/Y132G	BL21 [de3] Arctic Express
KDM4B-WT	BL21 [de3] Codon plus
KDM4B-F186G	BL21 [de3] Star
KDM4D-WT	Rosetta [de3] pLysS
KDM4D-F189G	BL21 [de3] Codon plus
PHF8-WT	BL21 [de3] Star
JMJD6-WT	Rosetta [de3] pLysS
JMJD3-WT	Rosetta [de3] pLysS

Above are the various *E. coli* competent cells used to express each protein state. At least two liters of protein were express and then purified.



Protein	MW (kD)
KDM4A-WT	44.270
KDM4A-F185A	44.190
KDM4A-F185G	44.180
KDM4B-WT	43.340
KDM4B-F186G	43.250
KDM4D-WT	41.160
KDM4D-F189G	41.070

Figure 4. Purified Proteins

Below shows the SDS-PAGE gel of purified proteins. Lanes 1➔7 consists of KDM4A-WT, KDM4A-F185A, KDM4A-F185G, KDM4B-WT, KDM4B-F186G, KDM4D-WT, KDM4D-F189G. All proteins used in biochemical assay were at least 90% pure.

2.4 HISTONE PEPTIDE PREPARATION

The University of Pittsburgh's Peptide Synthesis Facility synthesized the histone peptides (Table 6). Crude histone peptides were purified through HPLC (Table 7) and concentrated by speedvac and lyophilyzation. The peptides were dissolved in H_2O + 0.1% TFA and stored in -80°C. The peptide concentration was determined using Eppendorf™ Biospectrometer at 205nm and verified through MALDI using CHCA matrix. All peptides were stored in -80°C until used for downstream biochemical assays.

Table 6. Histone Peptides

Peptide	**SEQUENCE**	**MW (Da)**
H3K4me3	H2N-ARTK(Me3)QTARKSTGGKAPRKQLK(biotin)-CO2H	2692
H3K9me3	H2N-ARTKQTARK(Me3)STGGKAPRKQLK(biotin)-CONH2	2693
H3K36me3	H2N-KSAPSTGGVK(Me3)KPHRYKPGTGK(biotin)-CONH2	2562
H3K9	TO BE ADDED	2723
H3K27me3	H2N-APRKQLATKAARK(Me3)SAPATGGVK(biotin)-CONH2	2475
H3R2me2	H2N-AR(Me2)TKQTARKSTGGKAPRK(biotin)-CONH2	2195

The histone peptides used as substrates for the assays performed are shown below with their associated amino acid sequence and molecular weight (daltons).

Table 7. Histone peptide purification

Peptide Purification HPLC conditions		
Time	**% Water + 0.1% TFA**	**% ACN**
Initial	100	0
15 min	91	9
18 min	100	0

The histone peptides listed in Table 6 were purified through HPLC using the conditions shown below. Peptides Eluted between 7 and 10 minutes.

2.5 2-KETOGLUTARATE AND 2-KETOGLUTARATE ANALOG PREPARATION

Various 2-ketoglutarate analogs, including the native molecule, were purified through HPLC after being synthesized by colleagues. The HPLC fractions were concentrated by speedvac and lyophilization and then suspending the solid in H_2O + 0.1% TFA to make 25mM concentration. The 2-KG analogs were verified through LC/MS (ESI-negative) and stored in -20°C.

2.6 MALDI DEMETHYLASE ASSAY

To monitor enzymatic activity, a demethylase assay was optimized and then observed through MALDI. All contents except for 2-KG were added together in a microcentrifuge tube and briefly centrifuged. The 2-KG was then added to the microcentrifuge tube and briefly centrifuged again. The MALDI demethylase assays were performed at 37°C over a three hour time period. The Iron (II) ammonium sulfate and L-ascorbic acid were made fresh every two hours. The modification state of the peptide was observed using MALDI with CHCA matrix. The MALDI plate was analyzed using Voyager on the Reflector Positive mode. The negative control included all components of the assay except for the demethylating enzyme. A schematic of the MALDI demethylase assay is shown in Figure 5 and the components of the assay are listed in Table 8.

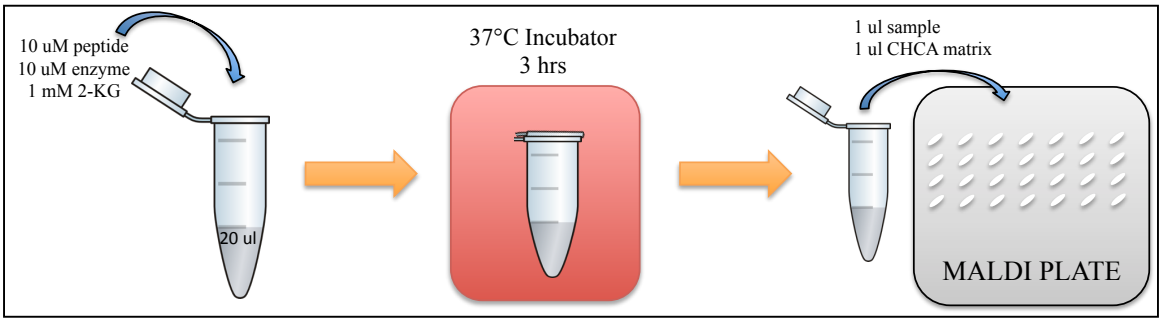


Figure 5. MALDI Demethylase Assay Scheme

Below is the MALDI demethylase assay scheme showing the steps to perform the assay. The compilation of all assay components in a microcentrifuge to a total volume of 20 ul was centrifuged and then incubated for three hours in 37°C. After the assay was completed, a sample was transferred onto the MALDI plate with CHCA matrix and analyzed.

Table 8. MALDI Demethylase Assay Components

MALDI Demethylase Assay	
Tris-HCl pH8	50 mM
$(NH_4)_2SO_4$ $FeSO_4$	50 uM
2-KG	1 mM
L-ascorbic acid	2 mM
Histone peptide	10 uM
KDM4A enzyme	10 uM
ddH20 to vol	to volume
Total vol: 20 ul	

A complete list of the MALDI demethylase assay components are shown below. The Iron (II) sulfate and L-ascorbic acid were made fresh (every two hours) to optimize results. The 2-KG was added to the cocktail last.

2.7 COUPLED METHYLATING-DEMETHYLATING ASSAY WITH G9A

This assay involves, first, methylating a histone peptide using G9A methyltransferase and SAM (universal methyl donor) and, second, demethylating the newly methylated histone peptide via KDM4A-2KG system (native or engineered). H3K9-biotin peptide was incubated with G9A and SAM for 1 hour at room temperature (Table 9). A sample for MALDI analysis was taken to verify that trimethylation occurred. Streptavidin beads were added to the methylation assay to pull down the methylated histone peptide and washed three times (50 mM Tris, pH 8). Following centrifugation and removal of supernatant, the demethylase assay was applied using the same conditions stated earlier, but with a 50 ul demethylase assay volume and a four hour incubation period. After completion of the demethylase assay, the reaction was centrifuged and supernatant was removed. The beads were washed as previously described and the peptide was eluted with ACN + 0.1% TFA. The methylation status of the peptide was observed through MALDI (Figure 6).

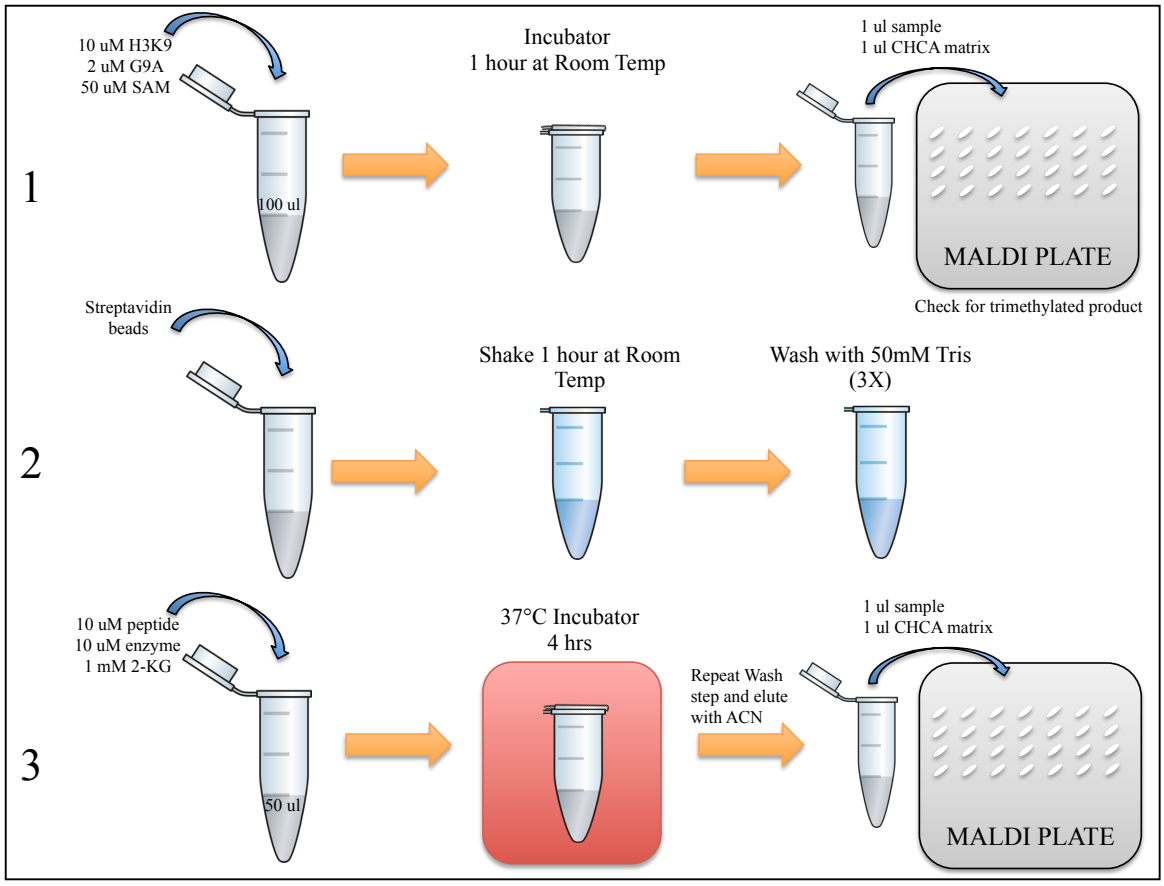


Figure 6. Coupled Methylating-Demethylating Assay Scheme

Coupled methylating-demethylating assay schematic showing the flow of the reaction is below. **1** represents G9A methylating H3K9, **2** is the peptide pull-down and wash step to prepare for stage **3,** which is the demethylating stage applied as described earlier, but with an assay volume of 50 ul and a 4 hour incubation period.

Table 9. Components of the Methylation Assay

Methylation Assay	
Tris HCl pH 8	50mM
ddH2O	to 100 ul
H3K9-biotin	10uM
SAM	50uM
G9A	2uM

A list of components for the methylation portion used in the coupled methylating-demethylating assay is shown below.

2.8 DETERMINING K_M OF 2-OG_3 AND 2-OG_7

The catalytic efficiency of engineered KDM4A-2KG systems were compared to the native KDM4A-2KG by varying the concentration of 2KG in the demethylase assay described above. The concentration of 2KG ranged from 10 uM to 1mM. All other conditions of the MALDI demethylase assay remained the same, but the incubation period was 1.5 hours to eliminate the confounding effect of saturation. The K_m was determined using PRISM. Assays were done in duplicates.

3.0 RESULTS

3.1 ALLELE SPECIFIC ACTIVATION OF KDM4A

The bump-hole technique requires engineering a 'hole' in the enzyme, KDM4A, and a 'bump' in the 2KG cofactor. Several of the amino acids within the catalytic domain of KDM4A are bulky, such as tryptophan and tyrosine and make good targets to mutate to make a hole suitable for the 'bump' incorporated onto the 2KG cofactor. When several mutants have been created, providing the 'hole', and many cofactor analogues have been synthesized, providing the 'bump', each system can be screened for enzymatic activity. Each KDM4A protein expressed (native and mutants) was screened for catalytic activity with each 2-KG molecule (native and analogs) using MALDI.

3.2 GENERATION OF KDM4A VARIANTS

The amino acids to target for mutagenesis were chosen based on the KDM4A structure reported on the protein database (PDB code: 2YBS). When looking at the crystal structure (Figure 7), it is evident that the 2-KG cofactor and the histone peptide are in close proximity within the binding pocket. It was important to target amino acids within the binding pocket that would not disturb peptide binding, but would be complementary to an engineered 2-KG compound. The mutants

designed included a single amino acid change from the native to an alanine, glycine or valine because these residues are relatively small, which will incorporate a ‘hole’ in the enzyme. There were 24 KDM4A mutants generated which were all verified through sequencing.

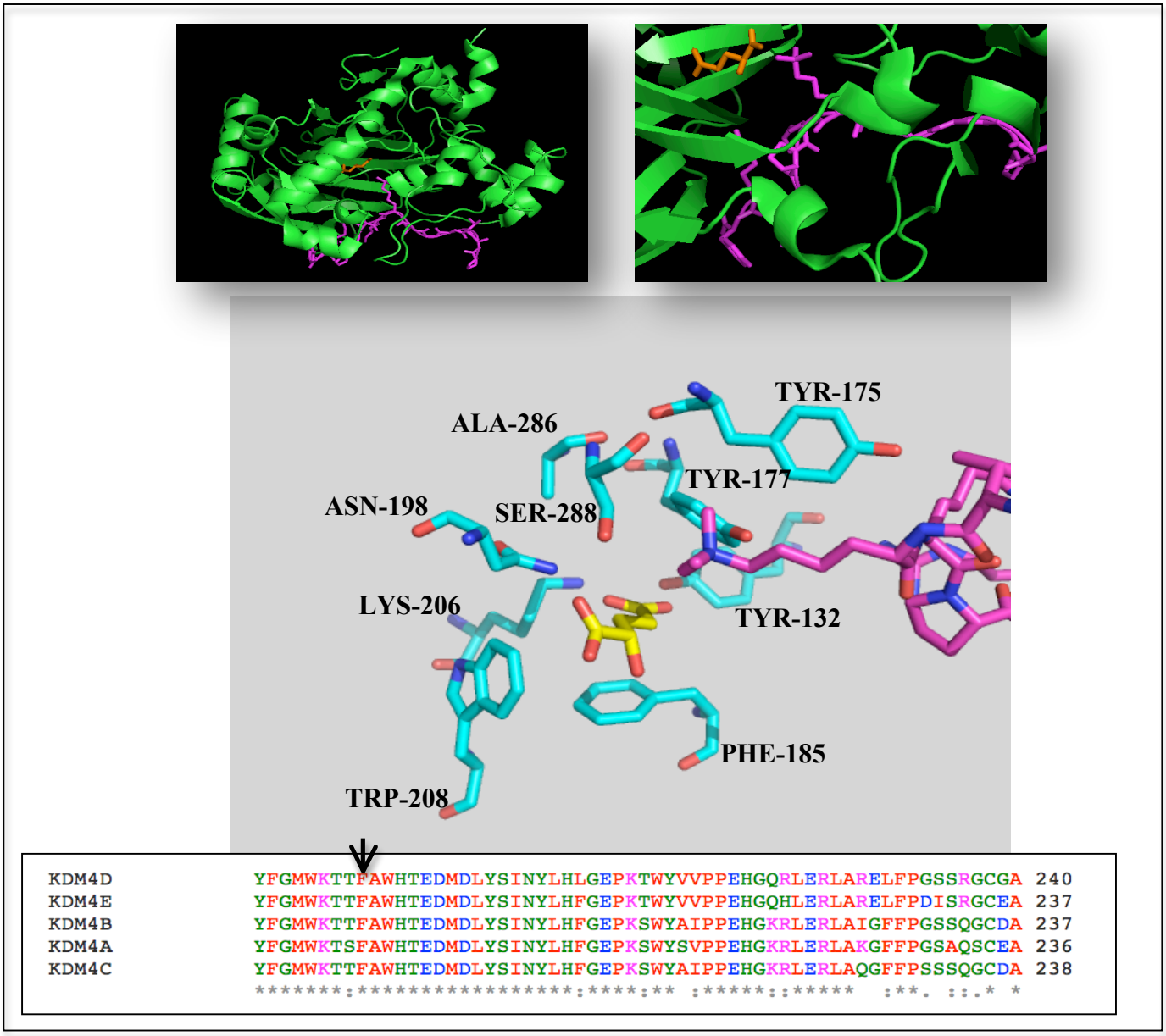


Figure 7. Structure and Sequence of KDM4 proteins

Top: Crystal Structure of 2-Ketoglutarate and methylated lysine binding pocket of the catalytic domain for KDM4A (PDB: 2YBS). The left image shows the entire KDM4A catalytic domain and the right image is a zoomed in image of the binding pocket within the catalytic domain. Middle: Amino acid targets for mutagenesis. Aqua: Amino acid residues to mutate, Yellow: 2KG, Purple: trimethylated lysine residue.

Bottom: Sequence alignment comparing a segment of the catalytic domain of KDM4 isoforms using ClustalW-Omega. Arrow indicates sequence homology at specific amino acid residue: Phe-185.

3.3 GENERATION OF 2-KG ANALOGUES

The cofactor for JmjC domain containing proteins is 2-ketoglutarate and was modified to incorporate a ‘bump’ that would be complementary to the hole generated in the enzyme. In total, there were 12 2-KG analogues synthesized and purified (Figure 8). The bumped cofactors will all be screened with the native KDM4A enzyme and with all the hole-containing KDM4A mutants to determine if there is complementarity for demethylation activity.

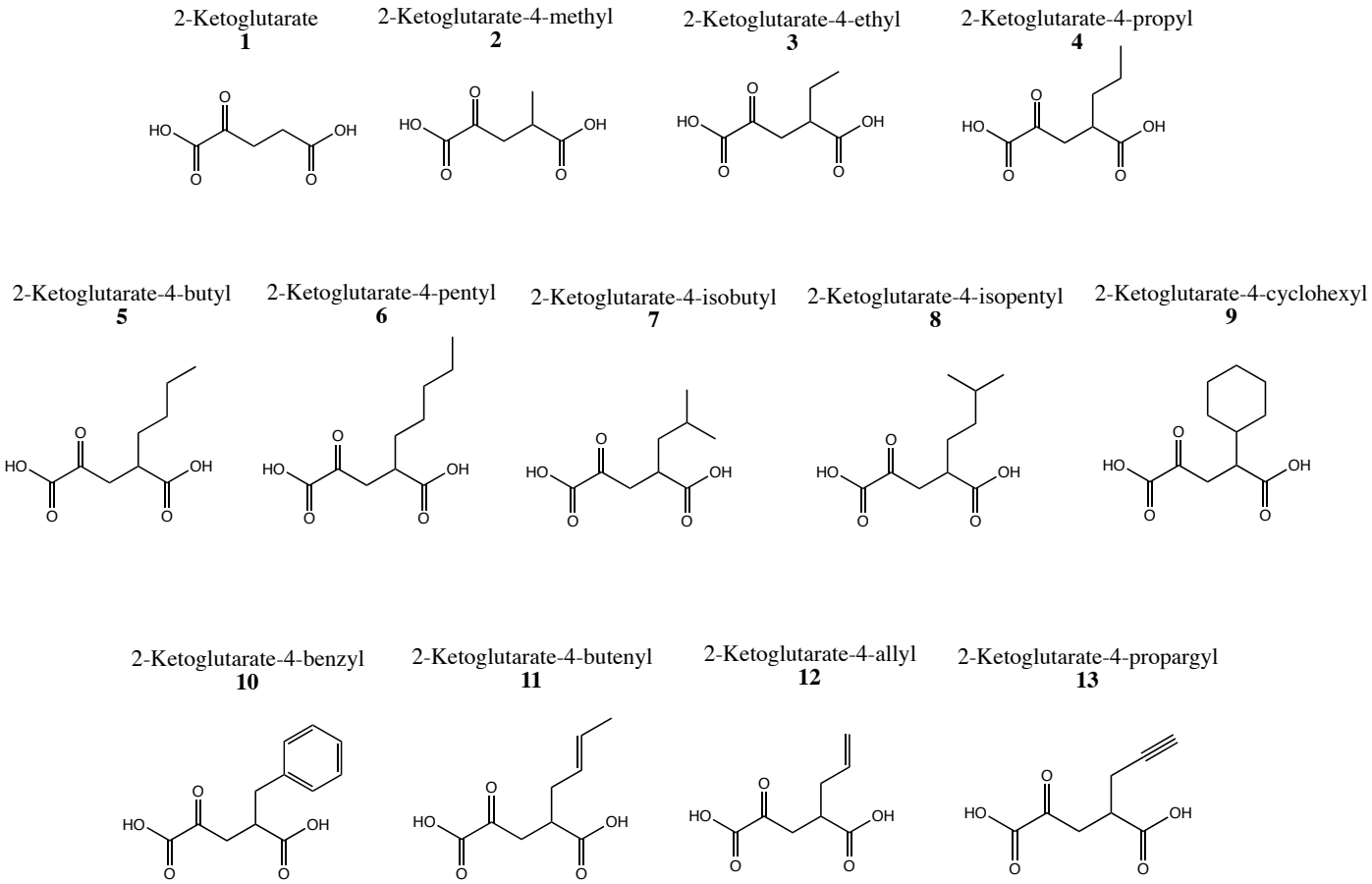


Figure 8: 2-Ketoglutarate Analogues

The 2-ketoglutarate analogues were synthesized, purified by HPLC and verified using LCMS. Bumped analogues have the modification located on the 4-carbon. Each residue will be referenced throughout the document using the bold number indicated above the structure.

3.4 SCREENING OF KDM4A MUTANTS AND 2KG ANALOGUES

With 24 KDM4A mutants and 12 cofactor analogues generated, there were 288 engineered enzyme-cofactor pairs screened for complementary bump-hole configuration. All 24 KDM4A mutants were paired with the native 2-KG compound and KDM4A-WT was paired with all 12 2-

KG analogues to screen for a bioorthogonal system using MALDI (Figure 9). KDM4A-WT did not show catalytic activity with any of the 2-KG analogs, whereas, 9 of the KDM4A-mutants: F185A, F185I, N198A, N198G, W208F, W208G, W208V, A286G and S288A showed catalytic activity with **1**. Half of the incorporated mutations resulted in no catalytic activity when paired with any of the 2-KG compounds. Three KDM4A mutations, all at the same amino acid: Phe-185, provided bioorthogonal systems. These mutations include KDM4A-F185G, F185T and F185V. KDM4A-F185G and F185T are catalytically active with 11 of the 2-KG analogues, whereas KDM4A-F185V is active with only five and at a significantly lesser amount. Peptide modification status for all KDM4A-F185 mutants that showed activity with any 2-KG analog were characterized and compared to the native KDM4A-2KG system (Figures 10 and 11). KDM4A-F185G and F185T were of the high interest because they were not active with the native 2-KG compound. When paired with **3** and **7,** KDM4A-F185G demethylation activity resulted in the modification status most similar to the native system and will be used for future experiments.

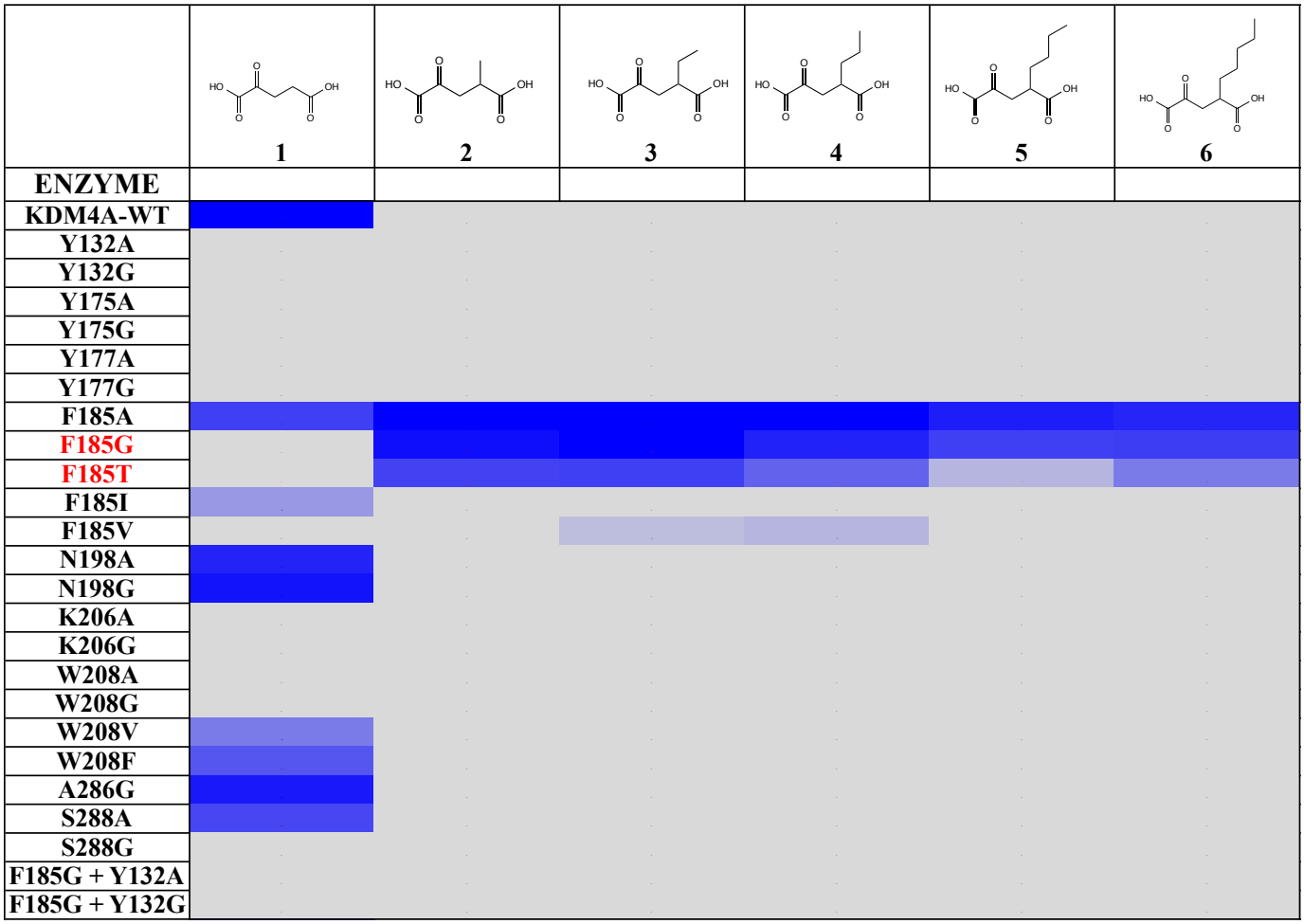


Figure 9a. Heat map of the Screening results for Engineered KDM4A-2-KG (**1-6**) pairs

Above is showing the catalytic activity of each KDM4A-2KG pair towards H3K9me3-biotin peptide. Active pairs are highlighted with a blue spectrum of light to dark correlating will least % modified (>0%) to greatest percent modified (100%), respectively. Pairs showing no activity are shaded grey. Mutants highlighted in red provide bioorthologout systems.

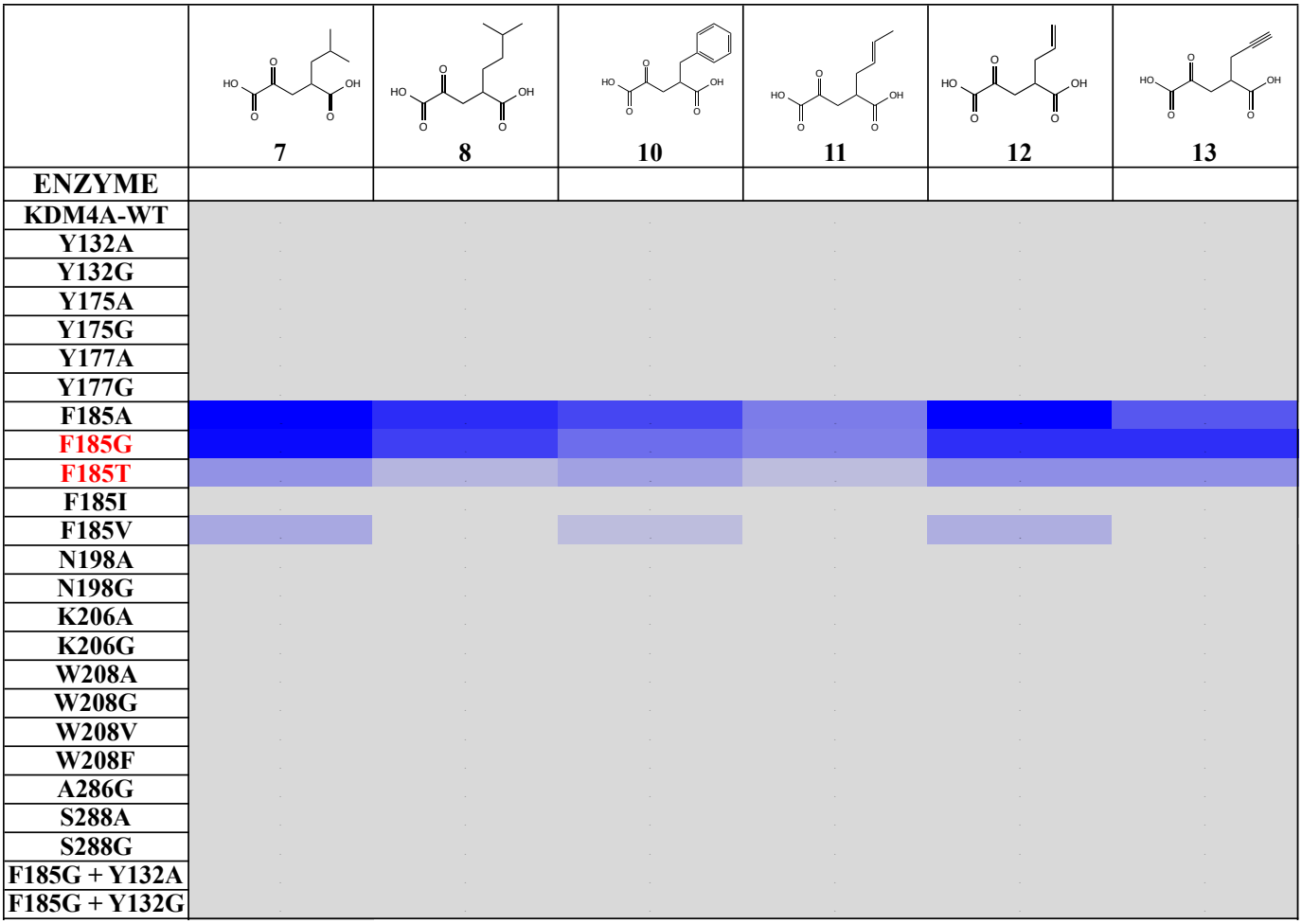


Figure 9b: Heat map for Screening of Engineered KDM4A-**7-13** pairs.

Above is showing the catalytic activity of each KDM4A-2KG pair towards H3K9me3-biotin peptide. Active pairs are highlighted with a blue spectrum of light to dark correlating will least % modified (>0%) to greatest percent modified (100%), respectively. Pairs showing no activity are shaded grey. No activity was observed for any pairs involving 9 (data not shown). Mutants highlighted in red provide bioorthologout systems.

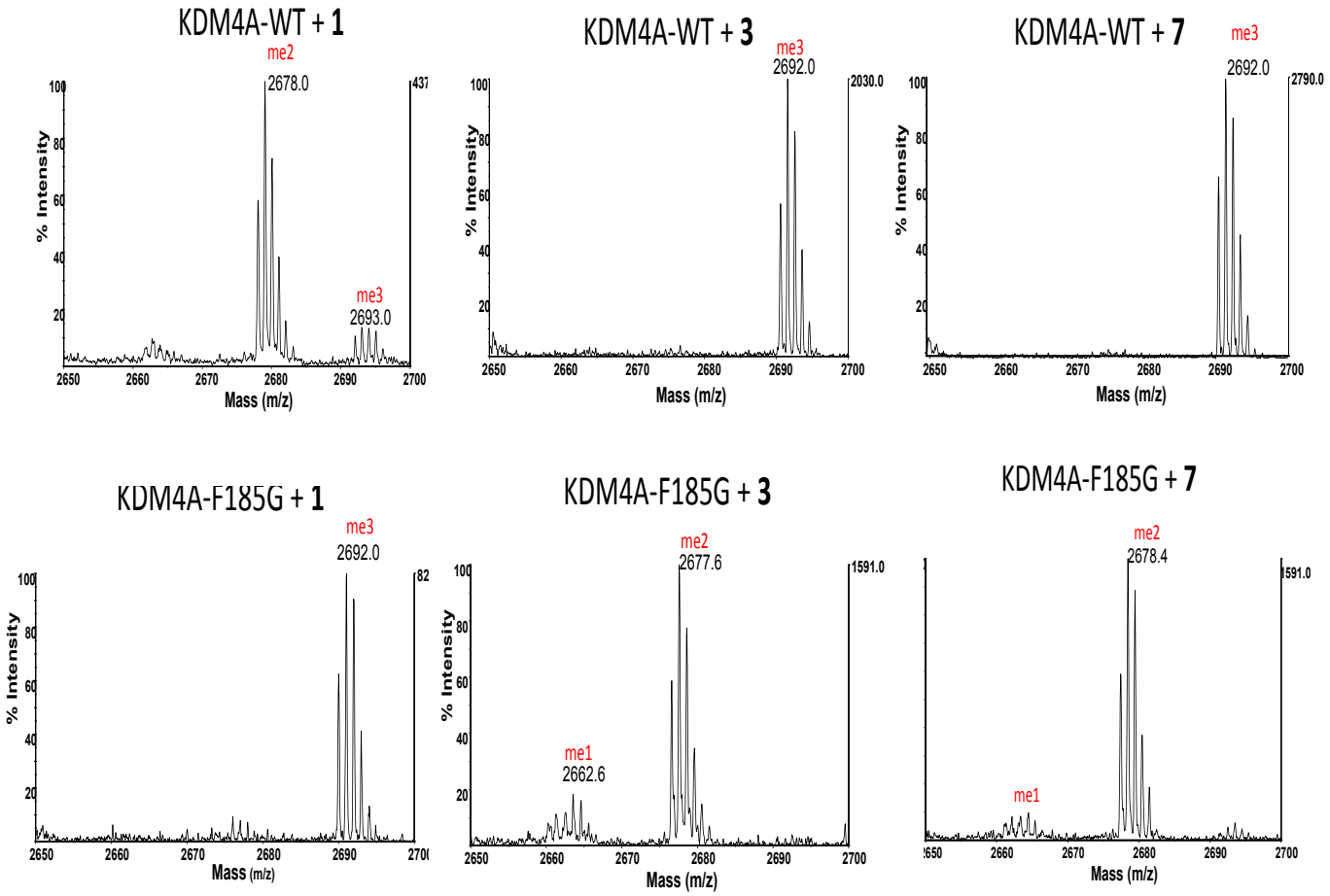


Figure 10: MALDI Spectrum of KDM4AWT/F185G Demethylation Activity

MALDI spectrum KDM4A-WT and KDM4A-F185G with 2-KG analogues: **1, 3** and **7**. Enzyme-cofactor pairs were screened for activity towards H3K9me3-biotin (2692 Da). Demethylated products will show a mass change of -14D (me2 = 2678, me1 = 2664).

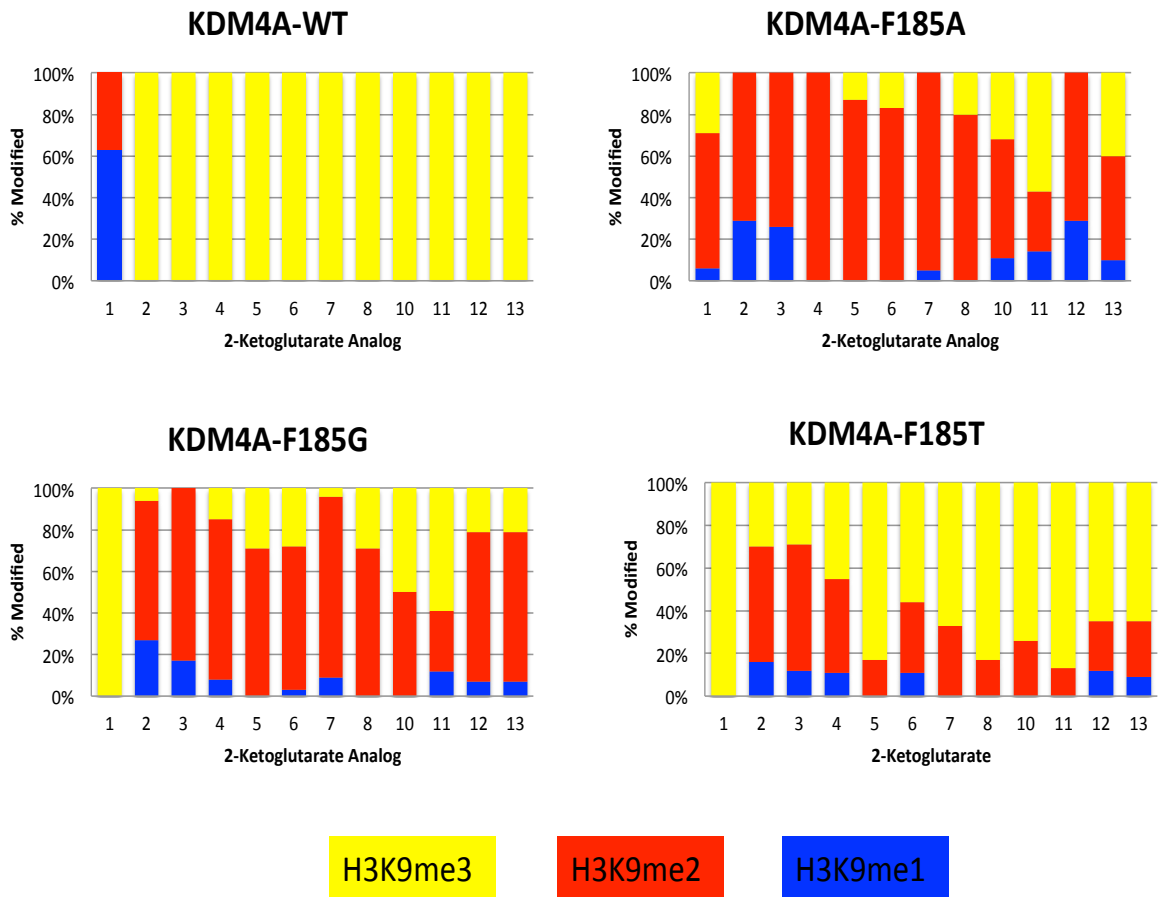


Figure 11: H3K9 Methylation State Profiling Post KDM4A-F185 Mutant Activity

Bar graphs representing the modification states, mono- (blue), di- (red) and tri-methylated (blue), of H3K9. KDM4A-WT and KDM4A-F185-mutants were paired with each 2-KG analog and screened for activity towards H3K9me3-biotin peptide. Activity was measured using MALDI.

3.5 SCREENING OF KDM4B AND -D MUTANTS AND 2KG ANALOGUES

The discovery of an engineered KDM4A-2-KG pair that mimics the native pathway is exciting and leads to the generation of the same mutant in the KDM4 isoforms to determine if the effect will be conserved. The phenylalanine at the 185 position in KDM4A align with the Phe-186 and Phe-189 in KDM4B and KDM4D, respectively (Figure 7). The corresponding phenylalanine residues were mutated to glycine (KDM4B-F186G and KDM4D-F189G). The native and mutant isoforms were screened for catalytic activity on H3K9me3-biotin with **1**, **3** and **7**. The 2-KG analogues were chosen based on the results from KDM4A screening. The native isoforms, KDM4B-WT and KDM4D-WT gave similar results to KDM4A-WT, being active with **1** (86% and 76% respectively) and inactive with **3** and **7**. The mutant residues gave interesting results. KDM4B-F185G was not active when paired with **1** and showed activity with **3** (74%) and **7** (67%). KDM4D-F189G was slightly active with **1** (23%) and highly active with **3** (100%) and **7** (73%) **(Figures 12 and 13)**. These results may be due to dissimilarities of amino acid residues within close proximity to the mutant. For example, there is a serine residue preceding Phe-185 in KDM4A, but there is a threonine residue preceding the Phe-186 and Phe-189 in KDM4B and KDM4D.

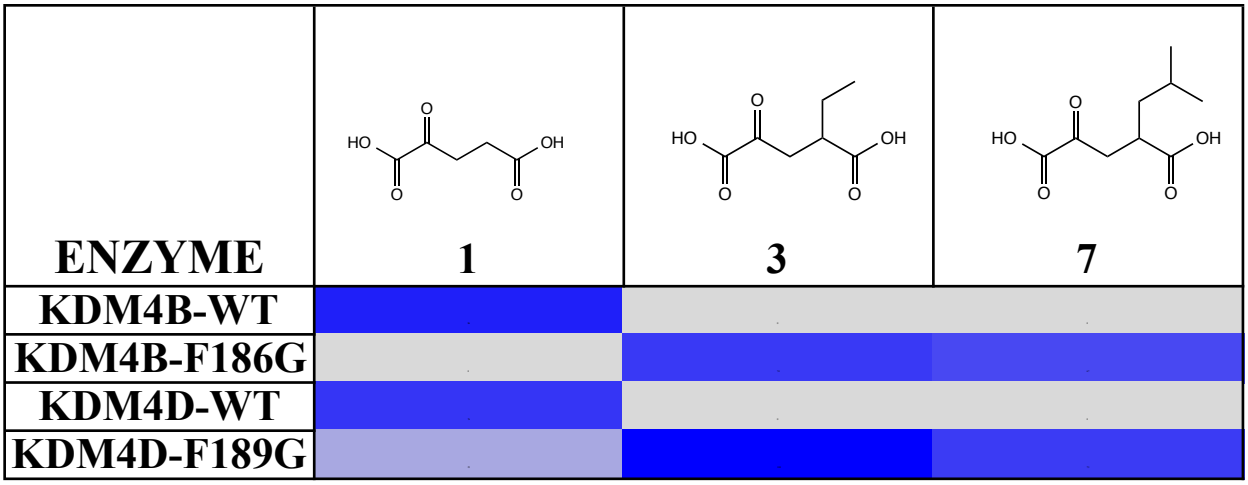


Figure 12. Heat Map of KDM4 isoform Activity with 2-KG Analogues

Below is a summary of the catalytic activity of each KDM4-2KG pair towards H3K9me3-biotin peptide. The values within each well represent the percentage of peptide modified (demethylated) measured via MALDI. Active pairs are highlighted with a green spectrum of light to dark correlating

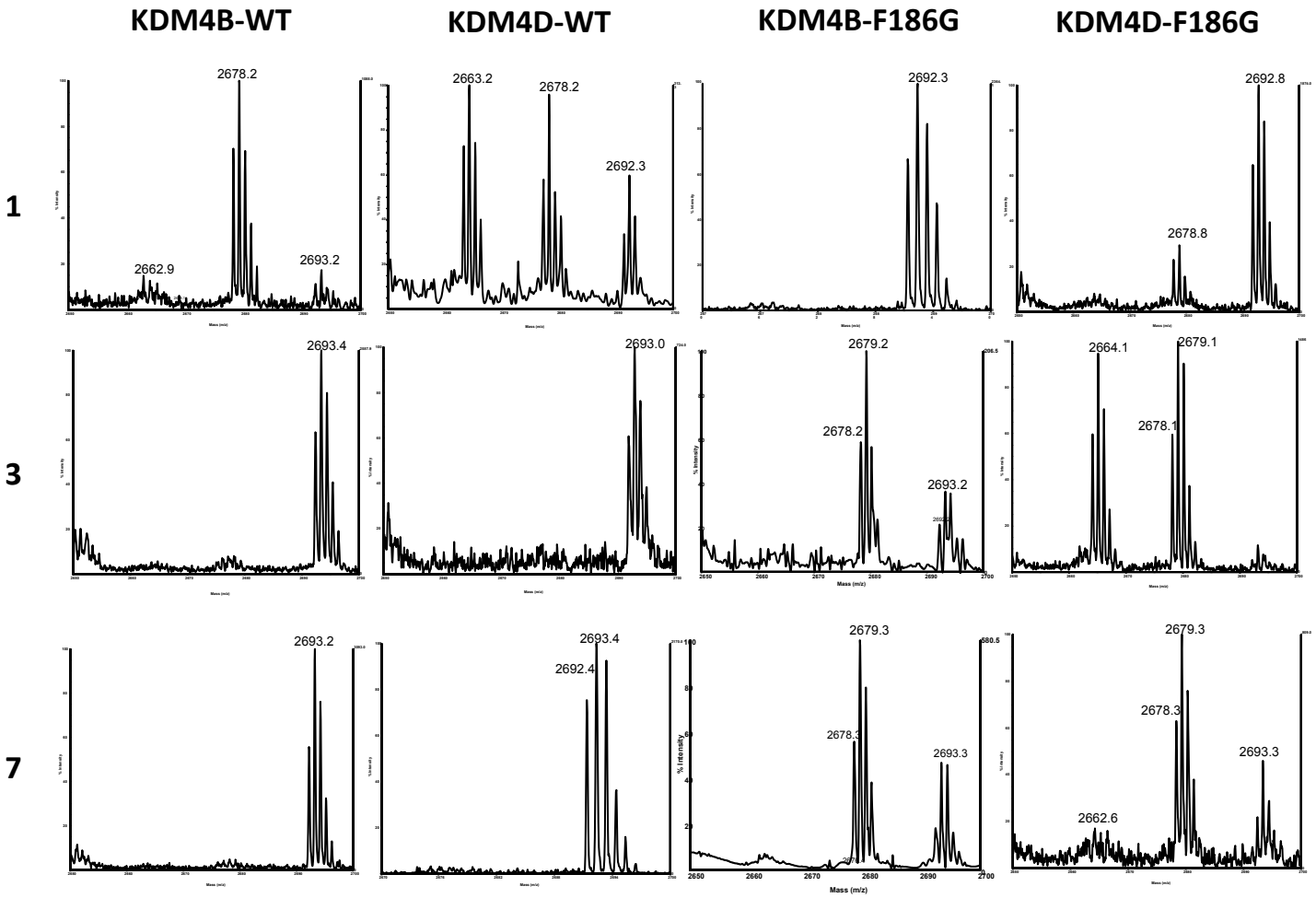


Figure 13: MALDI Spectrums of KDM4 Isoform Activity with 2-KG Analogues

MALDI spectrum for catalytic domains of KDM4 isoforms, B and D, paired with 2-KG analogues **1**, **3** and **7**. Enzyme-2-KG pairs were screened for demethylase activity towards H3K9me3-biotin and measured using MALDI.

3.6 SUBSTRATE SPECIFICITY OF KDM4A-F185G-2KG$_{3/7}$

Demethylase substrate specificity is of great importance. This characteristic may influence gene regulation, protein regulation and development and cell differentiation. Because of the roles methylated histones play in various biological mechanisms, it is important to maintain substrate specificity when generating a bioorthologous pathway. Substrate specificity was tested to observe if the incorporated mutations (Phe>Ala/Gly) affected this characteristic. KDM4A specifically targets H3K9me2/3, H3K36me2/3. Both mutations, F185A and F185G, did not show any catalytic activity towards H3K4me3, H3K27me3 or H3R2me2 peptides. These mutants showed activity towards H3K36me3-biotin peptide, but in a lesser fashion than for H3K9me3-biotin peptide (data not shown). Also, there is less monomethylated lysine on the peptides (H3K9 and H3K36) than dimethylated lysine suggesting that the pathway is more efficiently active on the trimethylated substrate (Figure 14).

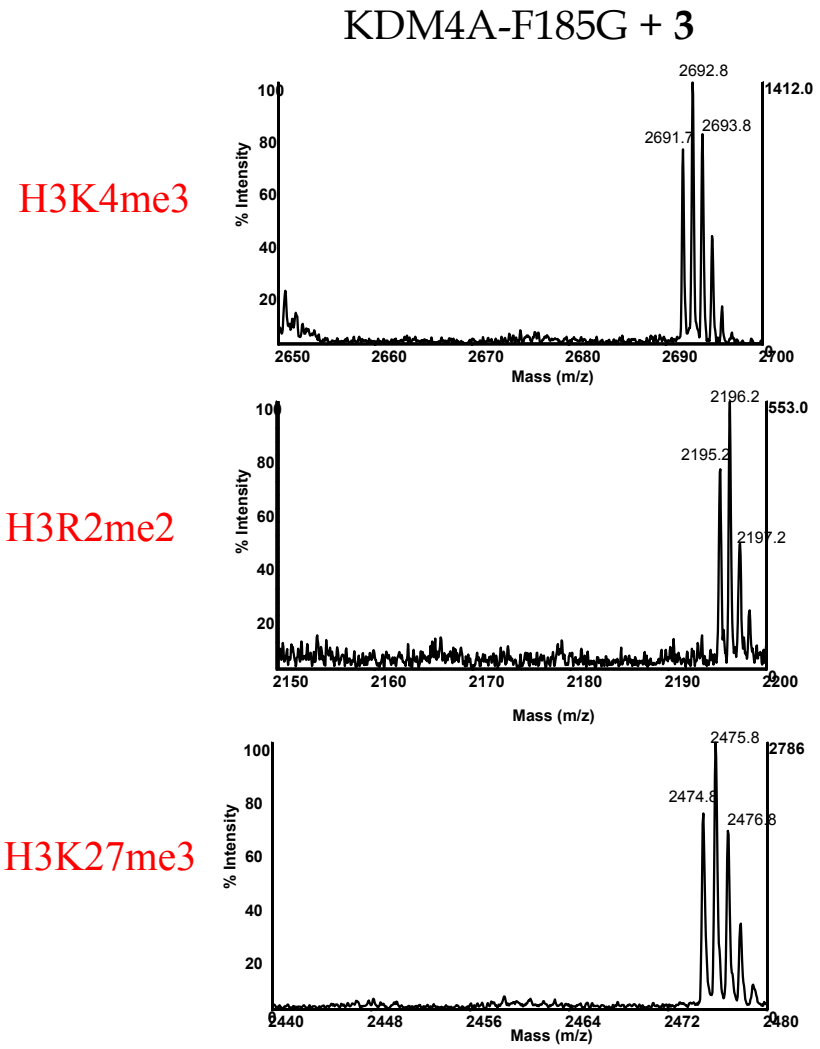


Figure 14: KDM4A-F185G substrate specificity

Below are the MALDI spectrums for non-KDM4A histone substrates when incubated with KDM4A-F185G + **3**. Engineered enzyme-cofactor pair was tested with H3K4me3, H3K27me3 and H3R2me2.

3.7 *IN VITRO* H3K9 METHYLATION BY G9A AND DEMETHYLATION BY ENGINEERED ENZYME-COFACTOR PAIR

Histone lysine residues undergo methylation by various methyltransferases (e.g. G9A) and then get demethylated by KDMs. The engineered enzyme-cofactor pairs were tested to see if they performed in a similar manner after an unmodified H3K9 histone peptide becomes trimethylated by G9A. KDM4A-F185G + **3** and KDM4A-F185G + **7** showed activity towards the newly modified H3K9 (Figure 15). The unmodified peptide has a molecular weight of 2723 kD and after incubation with G9A and SAM the molecular weight is 2765 kD representing H3K9me3. The newly modified peptide was used in the demethylase assay for the engineered pairs and the resulting molecular weights were decreased 2765 kD, highest intensity peak at 2751 kD (-me2) and small peaks at 2737 kD (-me1). The demethylation activity using **3** and **7** is less than what was observed in the previous demethylation assay using the H3K9me3 peptide, which may be due to the demethylation cocktail being applied to peptide attached to streptavidin beads.

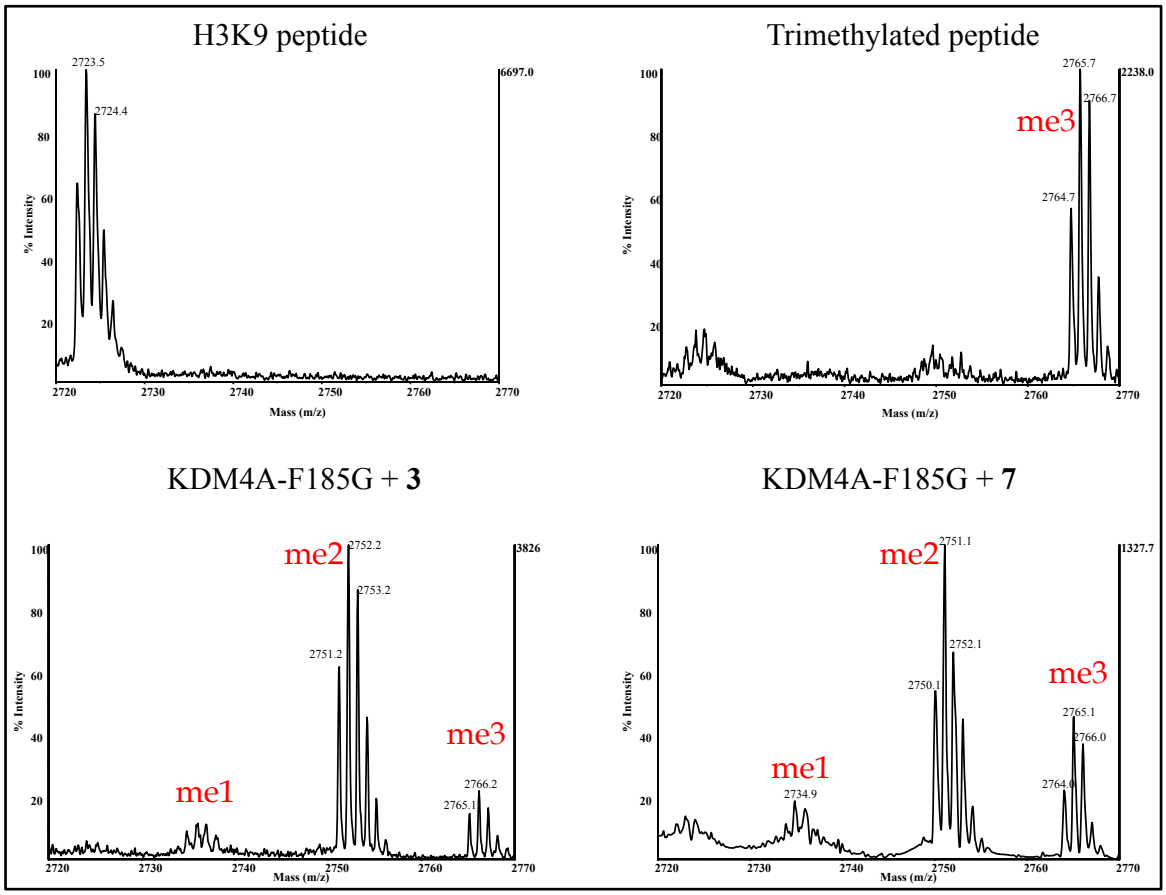
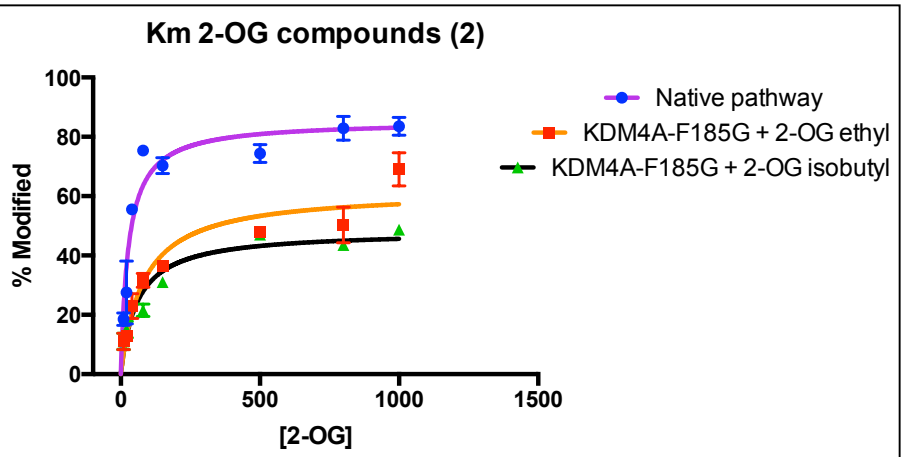


Figure 15. MALDI spectrum of Coupled methylating-demethylating assay

Spectrums show H3K9 peptide prior to modification (top left), H3K9me3 peptide product after G9A methylating activity (top right), demethylation of H3K9me3 to mono- or di-methylated peptide via KDM4A-F185G activity using 2-KG-**3** (bottom left) or using 2-KG-**7** (bottom right)

3.8 DETERMINATION OF K_M FOR 2-OG COMPOUNDS

The K_m for the engineered KDM4A-2KG pairs (KDM4A-F185G + **3** or **7**) was determined and compared to the native system. The same demethylase cocktail was used for the K_m determination experiment, but the concentration of 2-KG compound ranged from 10 uM to 1 mM and the incubation time was shortened to 1.5 hours to avoid artificial results. The K_m for KDM4A-F185G + **7** (58.68 ± 11.65) is closer to the K_m of the native pathway (27.5 ± 4.823) when compared to the K_m for KDM4A-F185G + **3** (80.03 ± 18.44) (Figure 16).



	Native pathway	KDM4A-F185G + 2-OG ethyl	KDM4A-F185G + 2-OG isobutyl
Michaelis-Menten			
Best-fit values			
Vmax	85.37	61.86	48.28
Km	27.50	80.03	58.68
Std. Error			
Vmax	3.209	3.736	2.532
Km	4.823	18.44	11.65

Figure 16. Michaelis-Menten Plot for Engineered KDM4A-F185G + 3/7

Michaelis-Menten plot and data for KDM4A-WT + 2-KG pair (Native pathway), KDM4A-F185G + **3** pair and KDM4A-F185G + **7** pair. The chart plots the % modified histone peptide (considering monodemethylated and didemethylated peptide) vs. the concentration of the 2-KG compound. PRISM was used for the analysis of data.

4.0 CONCLUSION

There are 24 JmjC domain containing proteins that are known to have demethylating activity. Each of these enzymes demethylate specific substrates at distinct periods of development and in certain tissues. The KDM4 subfamily includes KDM4A-E with KDM4F being a pseudogene. Each of the KDM4 proteins has pivotal roles in the regulation of gene transcription and expression, protein expression and cellular development. The functions of this family have not yet been fully understood and the previous approaches to elucidate their functions are limited. The KDM4 proteins have widespread and potentially compensatory functions. Many of the biological processes these proteins regulate are time dependent and occur in seconds or minutes, such as the cell cycle, and to determine the specific roles each KDM4 member plays, an approach that can work in this time dependent manner is necessary.

The allele specific bump-hole method has been applied to the KDM4A pathway. There were 24 KDM4A mutants and 12 2-ketoglutarate compounds generated resulting in 288 enzyme-cofactor pairs screened for a bioorthologous pathway. The engineered KDM4A-F185G + **3** and **7** have given results that are similar to the native system. The mutation did not disturb the substrate specificity seen by its activity with H3K9me2/me3 and its inactivity with H3K4me3, H3K27me3 and H3R2me2. The results were conserved when the mutation was applied to KDM4B and KDM4D except for minimal activity of KDM4D-F189G paired with the native 2-KG compound.

Both KDM4A-F185G + **3** and KDM4A-F185G + **7** were active in a system where H3K9 was first methylated by G9A and have a K_m within a factor of the native system.

In future experiments, the engineered KDM4A-2-KG pairs should be tested on full histone substrates and should be screened with inhibitors. The engineered pair should be tested in cell lysates and in live cell culture. This system will lead to the monitoring of KDM4A *in vivo* and elucidate specific functions and genes regulated by this enzyme.

APPENDIX: ABBREVIATIONS

2-KG: (2)-Ketoglutarate

5hmC: 5-hemimethylcytosine

5mC: 5-methylcytosine

AP: Activating Protein

AR: Androgen receptor

CM: consensus methylation motif

DNMT: DNA methyltransferase

DTT: dithiothreitol

FDH: 10-formyltetrahydrofolate dehydrogenase

FPLC: Fast protein liquid chromatography

HIF: Hypoxia inducible factor

HPLC: High performance liquid chromatography

IP6K: Inositol hexakisphosphate kinase

JMJ: JmjC domain containing demethylases

KDM: Lysine demethylase

LSD: flavin-dependent lysine specific demethylases

m^{6}A: N^6-methyladenosine

MBD2: Methylated DNA binding domain-containing protein

METTL3: methyltransferase like 3 domain

MMR: Mismatch repair

MT: methyltransferase

NADH: Nicotinamide adenine dinucleotide

PAD: peptidylarginine deiminases

PKMT: Protein lysine methyltransferases

PMT: Protein methyltransferases

PRMT: Protein arginine methyltransferases

PTM: Post translational modification

RNF: Ring finger protein

SAM: S-adenosylmethionine- binding domain

SCF: Skp, Cullin, F-box containing complex

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SET: Su (var), enhancer of zeste, trithorax domain

TET: human ten-eleven-translocation domain

BIBLIOGRAPHY

1. Dekker, J., Two ways to fold the genome during the cell cycle: insights obtained with chromosome conformation capture. Epigenetics Chromatin, 2014. 7(1): p. 25.

2. Marino-Ramirez, L., Kann, M., and Landsman, D., Histone structure and nucleosome stability. Expert Rev Proteomics, 2005. 2(5): p. 719–729.

3. Copeland, R.A., Moyer, M.P., and Richon, V.M., Targeting genetic alterations in protein methyltransferases for personalized cancer therapeutics. Oncogene, 2013. 32(8): p. 939–946.

4. Upadhyay, A.K. and Cheng, X., Dynamics of Histone Lysine Methylation: Structures of Methyl Writers and Eraser. Prog Drug Res, 2011. 67: p. 107–124.

5. Blattler A. and Farnham P.J., Cross-talk between Site-specific Transcription Factors and DNA Methylation States. The Journal of Biological Chemistry, 2013. 288(48): p. 34287-34294.

6. Phillips, T., The role of methylation in gene expression. Nature Education, 2008. 1(1): p. 116

7. Guo, J.U., Su, Y., et al., Emerging roles of TET proteins and 5-Hydroxymethylcytosines in active DNA demethylation and beyond. Cell Cycle, 2011. 10 (16): p. 2662-2668.

8. Chuan He, Y.F., Nucleic acid modifications with epigenetic significance. Current Opinion in Chemical Biology, 2012. 16(5-6): p. 516–524.

9. Xhemalce, B., Robson, S.C., Kouzarides, T., Human RNA methyltransferase BCDIN3D regulates microRNA processing. Cell, 2012. 151(2): p. 278-88.

10. Dhayalan, A., et al., A Continuous Protein Methyltransferase (G9a) Assay for Enzyme Activity Measurement and Inhibitor Screening. Journal of Biomolecular Screening, 2009. 14 (9): p. 1129-1133.

11. Islam, K., et al., Bioorthogonal profiling of protein methylation using azido derivative of S-adenosyl-L-methionine. JACS, 2012. 134(13): p. 5909-15.

12. Bharathy, N., et al., G9a, a multipotent regulator of gene expression. Epigenetics, 2013. 8(1): p. 16-22.

13. Lehnertz B, Northrop JP, Antignano F, Burrows K, Hadidi S, Mullaly SC, et al., Activating and inhibitory functions for the histone lysine methyltransferase G9a in T helper cell differentiation and function. J Exp Med, 2010. 207: p. 915–22.

14. Chervona Y. and Costa M., The control of histone methylation and gene expression by oxidative stess, hypoxia and metals. Free radical biology and medicine, 2012. 53(5): p. 1041-1047.

15. Klose, R.J., Kallin, E.M., and Zhang, Y., JmjC-domain containing proteins and histone demethylation. Nature Reviews Genetics, 2006. 7: p. 715-727.

16. Wang, Y., Wysocka, J., et al., Human PAD4 Regulates Histone Arginine Methylation Levels via Demethylimination. Science, 2004. 306 (5694): p. 279-283.

17. Hensen, S., and Pruijn, G., Methods for the Detection of Peptidylarginine Deiminase (PAD) Activity and Protein Citrullination. Molecular and Cellular Proteomics, 2014. 13: p. 388-396.

18. Wang, Y., Xiaoman, L., Hu, H., H3K4me2 reliably defines transcription factor binding regions in different cells. Genomics, 2014. 103(2-3): p. 222-228.

19. Margaritis, T., et al., Two Distinct Repressive Mechanisms for Histone 3 Lysine 4 Methylation through Promoting 3′-End Antisense Transcription. Plos Genetics, 2012. 8(9).

20. Kidder, B., Hu, G., and Zhao, K., KDM5B focuses H3K4 methylation near promoters and enhancers during embryonic stem cell self-renewal and differentiation. Genome Biology, 2014. 15(2): p. R32.

21. Barski, A., Cuddapah, S., Cui, K., et al., High-resolution profiling of histone methylations in the human genome. Cell, 2007. 129(4): p. 823-37.

22. Southall. S., Cronin, N., Wilson, J., A novel route to product specificity in the Suv4-20 family of histone H4K20 methyltransferases. Nucleic Acids research, 2014. 42(1): p. 661-671.

23. Culhane, J. and Cole, P., LSD1 and the Chemistry of Histone Demethylation. Current Opinion in Chemical Biology, 2007. 11(5): p. 561-568.

24. Lee, J.S., and Shilatifard A., A site to remember: H3K36 methylation a mark for histone deacetylation. Mutation Research, 2007. 618(1-2): p. 130-134.

25. Cattaneo P., Kunderfranco P., Greco C., Condorelli G., et al., DOT1L-mediated H3K79me2 modification critically regulates gene expression during cardiomyocyte differentiation. Cell Death and Differentiation, 2014.

26. Hillringhaus, L., et al., Structural and Evolutionary Basis for the Dual Substrate Selectivity of Human KDM4 Histone Demethylase Family.The Journal of Biological Chemistry, 2011. 286: p. 41616–41625.

27. Labbe, R.M., Holowatyj, A., Yang, Z., Histone lysine demethylase (KDM) subfamily 4: structures, functions and therapeutic potential. American Journal of Translational Research, 2014. 6(1): p. 1-15

28. Klose, R.J., Kallin, E.M., and Zhang, Y., JmjC-domain containing proteins and histone demethylation. Nature Reviews Genetics, 2006. 7: p. 715-727.

29. Berry, W. and Janknecht, R., KDM4/JMJD2 Histone Demethylases: Epigenetic Regulators in Cancer Cells. The Journal of Cancer Research, 2013. 73: p. 2936.

30. Burton, A., Azevedo, C., Andreassi, C., Riccio, A., Saiardi, A., Inositol pyrophosphates regulate JMJD2C-dependent histone demethylation. Proc Natl Acad Sci U S A, 2013.110: p. 18970–18975.

31. Mallette, F.A., Mattiroli, F., Cui, G., Young, L.C., Hendzel, M.J., Mer, G., Sixma, T.K., Richard, S., RNF8- and RNF168-dependent degradation of KDM4A/JMJD2A triggers 53BP1 recruitment to DNA damage sites. EMBO J, 2012. 31: p. 1865–1878.

32. Van Rechem, C., et al., The SKP1-Cul1-F-box and leucine-rich repeat protein 4 (SCF-FbxL4) ubiquitin ligase regulates lysine demethylase 4A (KDM4A)/Jumonji domain-containing 2A (JMJD2A) protein. J Biol Chem, 2011. 286: p. 30462–30470.

33. Tan, M.K., Lim, H.J., Harper, J.W., SCF(FBXO22) regulates histone H3 lysine 9 and 36 methylation levels by targeting histone demethylase KDM4A for ubiquitin-mediated proteasomal degradation. Molecular Cell Biology, 2011. 31: p. 3687–3699.

34. Luo, W., Chang, R., Zhong, J., Pandey, A., Semenza, G.L., Histone demethylase JMJD2C is a coactivator for hypoxia-inducible factor 1 that is required for breast cancer progression. Proc Natl Acad Sci U S A, 2012. 109: p. E3367–3376.

35. Coffey, K., Rogerson, L., Ryan-Munden, C., Alkharaif, D., Stockley, J., Heer, R., Sahadevan, K., O'Neill, D., Jones, D., Darby, S., Staller, P., Mantilla, A.,

Gaughan, L., Robson, C.N., The lysine demethylase, KDM4B, is a key molecule in androgen receptor signalling and turnover. Nucleic Acids Res, 2013. 41: p. 4433–4446.

36. Shin, S., Janknecht, R., Diversity within the JMJD2 histone demethylase family. Biochem Biophys Res Commun, 2007. 353: p. 973–977.

37. Kawazu, M., Saso, K., Tong, K.I., McQuire, T., Goto, K., Son, D.O., Wakeham, A., Miyagishi, M., Mak, T.W., Okada, H., Histone demethylase JMJD2B functions as a co-factor of estrogen receptor in breast cancer proliferation and mammary gland development. PLoS One, 2011. 6: p. 17830.

38. Gray, S.G., Iglesias, A.H., Lizcano, F., Villanueva, R., Camelo, S., Jingu, H., The, B.T., Koibuchi, N., Chin, W.W., Kokkotou, E., Dangond, F., Functional characterization of JMJD2A, a histone deacetylase- and retinoblastoma-binding protein. J Biol Chem, 2005. 280: p. 28507–28518.

39. Black, J.C., Manning, A.L., Whetstine, J.R., KDM4A lysine demethylase induces site-specific copy gain and rereplication of regions amplified in tumors. Cell, 2013. 154(3): p. 541-55

40. Awwad, S., Ayoub, N., Overexpression of KDM4 lysine demethylases disrupts the integrity of the DNA mismatch repair pathway. Biology Open, 2015. 4(3).

41. Ding, X., Pan, H., Li, J., Zhong, Q., Chen, X., Dry, S.M., Wang, C.Y., Epigenetic activation of AP1 promotes squamous cell carcinoma metastasis. Sci Signal, 2013. 6:ra28, 1–13, S0–15.

42. Mallette, F.A., Richard, S., JMJD2A promotes cellular transformation by blocking cellular senescence through transcriptional repression of the tumor suppressor CHD5.Cell Rep, 2012. 2: p. 1233–1243.

43. Guerra-Calderas, L., Gonzalez-Barrios, R., Herrera, L., Cantu de Leon, D., Soto-Reyes, E., The role of the histone demethylase KDM4A in cancer. Cancer Genetics, 2014.

44. Liu, Y., An epigenetic role for PRL-3 as a regulator of H3K9 methylation in colorectal cancer. Gut, 2013. 62: p. 571–581.

45. Kawazu, M., Saso, K., Tong, K.I., McQuire, T., Goto, K., Son, D.O., Wakeham, A., Miyagishi, M., Mak, T.W., Okada, H., Histone demethylase JMJD2B functions as a co-factor of estrogen receptor in breast cancer proliferation and mammary gland development. PLoS One, 2011. 6:e17830.

46. Yang, J., et al., The histone demethylase JMJD2B is regulated by estrogen receptor alpha and hypoxia, and is a key mediator of estrogen induced growth. Cancer Res, 2010. 70: p. 6456–6466.

47. Wissmann, M., Yin, N., Muller, J.M., Greschik, H., Fodor, B.D., Jenuwein, T., Vogler, C., Schneider, R., Gunther, T., Buettner, R., Metzger, E., Schule, R., Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. Nat Cell Biol, 2007. 9: p. 347–353.

48. Ishimura, A., Terashima, M., Kimura, H., Akagi, K., Suzuki, Y., Sugano, S., Suzuki, T., Jmjd2c histone demethylase enhances the expression of Mdm2 oncogene. Biochem Biophys Res Commun, 2009. 389: p. 366–371.

49. Liu, G., Bollig-Fischer, A., Kreike, B., van de Vijver, M.J., Abrams, J., Ethier, S.P., Yang, Z.Q., Genomic amplification and oncogenic properties of the GASC1 histone demethylase gene in breast cancer. Oncogene, 2009. 28: p. 4491–4500.

50. Zhang, Q.J., Chen, H.Z., Wang, L., Liu, D.P., Hill, J.A., Liu, Z.P., The histone trimethyllysine demethylase JMJD2A promotes cardiac hypertrophy in response to hypertrophic stimuli in mice. J Clin Invest, 2011. 121: p. 2447–2456.

51. Liang, Y., Vogel, J.L., Arbuckle, J.H., Rai, G., Jadhav, A., Simeonov, A., Maloney, D.J., Kristie, T.M., Targeting the JMJD2 histone demethylases to epigenetically control herpesvirus infection and reactivation from latency. Sci Transl Med, 2013. 5: p. 167ra5

52. Kantojarvi, K., Onkamo, P., Vanhala, R., Alen, R., Hedman, M., Sajantila, A., Nieminen-von Wendt, T., Jarvela, I., Analysis of 9p24 and 11p12-13 regions in autism spectrum disorders: rs1340513 in the JMJD2C gene is associated with ASDs in Finnish sample.Psychiatr Genet, 2010. 20: p. 102–108.

53. Xu, W., Yang, H., Xiong, Y., Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of α-Ketoglutarate-Dependent Dioxygenases. Cancer Cell, 2011. 19(1): p. 17-30

54. Islam, K., Allele-Specific Chemical Genetics: Concept, Strategies, and Applications. American Chemical Society, 2015. 10 (2): p. 343-363.

55. Stockwell, B., et al., Chemical genetics: ligand-based discovery of gene function. Nature Reviews Genetics, 2000. 1: p. 116-125.