Investigating the inhibitor and substrate diversity of the JmjC histone demethylases

A thesis submitted to the Board of the Faculty of Physical Sciences at the University of Oxford in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Rachel Shamo Schiller

Lincoln College

Michaelmas Term 2016

Rachel Shamo Schiller
Lincoln College

DPhil Thesis
Michaelmas Term, 2016

Investigating the inhibitor and substrate diversity of the JmjC histone demethylases

Abstract

Epigenetic control of gene expression by histone post-translational modifications (PTMs) is a complex process regulated by proteins that can 'read', 'write' or 'erase' these PTMs. The histone lysine demethylase (KDM) family of epigenetic enzymes remove methyl modifications from lysines on histone tails. The Jumonji C domain (JmjC) family is the largest family of KDMs. Investigating the scope and mechanisms of the JmjC KDMs is of interest for understanding the diverse functions of the JmjC KDMs *in vivo*, as well as for the application of the basic science to medicinal chemistry design.

The work described in this thesis aimed to biochemically investigate the inhibitor and substrate diversity of the JmjC KDMs, it led to the identification of new inhibitors and substrates and revealed a potential combinatorial dependence between adjacent histone PTMs. Structure–activity relationship studies gave rise to an *n*-octyl ester form of IOX1 with improved cellular potency and selectivity towards the KDM4 subfamily. This compound should find utility as a basis for the development of JmjC inhibitors and as a tool compound for biological studies. The rest of this thesis focused on the biochemical investigations of potential substrates and inhibitors for KDM3A, a JmjC demethylase with varied physiological functions. Kinetic characterisation of reported KDM3A substrates was used as the basis for evaluations of novel substrates and inhibitors. Further studies found TCA cycle intermediates to be moderate co-substrate competitive inhibitors of KDM3A. Biochemical investigations were carried out to study potential protein-protein interactions of KDM3A with intraflagellar transport proteins (IFTs), non-histone proteins involved in the formation of sperm flagellum. Work then addressed the exploration of novel *in vitro* substrates for KDM3 (KDM3A and JMJD1C) mediated catalysis, including: methylated arginines, lysine analogues, acetylated and formylated lysines. KDM3A, and other JmjC KDMs, were found to catalyse novel arginine demethylation reaction *in vitro*. Knowledge gained from studies with unnatural lysine analogues was utilised to search for additional novel PTM substrates for KDM3A. These results constitute the first evidence of JmjC KDM catalysed hydroxylation of an N^{ε}-acetyllysine residue. The H3 K4me3 position seems to be required for acetyllysine substrate recognition, implying a combinatorial effect between PTMs. Preliminary results provide evidence that JMJD1C, a KDM3 protein previously reported to be inactive, may catalyse deacetylation *in vitro*. An additional novel reaction, observed with both KDM3A and JMJD1C, is deformylation of N^{ε}-formyllysine residues on histone H3 fragment peptides. Interestingly, H3 K4 methylation was also observed to enhance the apparent deformylation of both KDM3A and JMJD1C catalysed reactions.

Overall, findings in this thesis suggest that the catalytic activity of JmjC KDMs extends beyond lysine demethylation. In a cellular context, members of the KDM3 subfamily might provide a regulatory link between methylation and acylation marks. Such a link will further highlight the complex relationships between histone PTMs and the epigenetic enzymes that regulate them. The observed dependency of H3 K9 catalysis on H3 K4 methylation adds another layer of complexity to the epigenetic regulation by histone PTMs.

Dedicated to my beloved family

You are my sunshine

Acknowledgments

Firstly I would like to thank my supervisors, Prof. Chris Schofield and Dr. Akane Kawamura, for providing me with an environment of scientific freedom, plenty of opportunities for collaborations and endless ideas, enthusiasm and inspiration. I am especially grateful for their support following my maternity leave. Their encouragement and consideration has had a tremendous influence on my motivation. I would also like to thank my industrial supervisor Dr. Richard Morphy for unfailing patience, eye-opening insights and assistance along the DPhil road.

I would like to acknowledge the Clarendon fund and Eli Lilly for granting me the scholarships that allowed my studies in Oxford University.

Work in this thesis would not have been possible without contributions from many colleagues both in the CJS group and outside it. I am particularly thankful to Dr. Richard Hopkinson for his guidance, fruitful discussions and advice throughout my work. I would also like to thank my collaborators Dr. Catrine Johansson, Dr. Patricia Yeyati, Dr. Holger Kramer and Dr. Louise Walport for their significant input into my research. Thank you to everyone who answered my endless questions: Dr. Emily Flashman and Dr. Hanna Tarhonskaya for advice on kinetics, Dr. James Dunford, Carina Giladi and Joanna Bonnici for guidance on protein production, Dr. Jacob Bush for the introduction to organic synthesis and Dr. James Wickens for his continuous help with mass spectrometry. I am indebted to my colleagues who kindly provided me with enzymes, peptides and modified amino acids: Dr. Richard Hopkinson, Roman Belle, Gareth Langley, Dr. Louise Walport, Rebecca Hancock and Kate Dunne. Thank you to Wendy Sobey and Adam Hardy for their devoted management of the lab. I would like to express my deep gratitude to everyone who proof-read my thesis: Barak Schiller, Dr. Richard Hopkinson, Dr. Hanna Tarhonskaya, Dr. John-Paul Bukowski, Dr. Chris Lohans, Dr. Catrine Johansson, Dr. Bhaskar Bhushan, Dr. Joan Gannon and Dr. Darragh Ennis.

I would also like to thank other members of the CJS group, past and present, for good advice and for providing me with an enjoyable environment in which to work. In particular, thanks to Shifali Shishodia, Dr. Jurgen Brem, Dr. Hanna Tarhonskaya, Dr. Cyrille Thinnes, Dr. Chris Lohans, Dr. Joan Gannon, Joanna Bonnici, Yijia Zhang, Rebecca Hancock, Kate Dunne and Dr. Teresa Borrello.

Most importantly, my heartfelt thanks to my husband Barak Schiller for his continued support and encouragement throughout my DPhil. Thank you for your love, for waking up at night so I could rest, and for sharing the journey with me.

Table of contents

List of publications

1. R. Schiller, G. Scozzafava, A. Tumber, J. R. Wickens, J. T. Bush, G. Rai, C. Lejeune, H. Choi, T.-L. Yeh, M. C. Chan, B. T. Mott, J. S. O. McCullagh, D. J. Maloney, C. J. Schofield, and A. Kawamura. A cell-permeable ester derivative of the JmjC histone demethylase inhibitor IOX1. *ChemMedChem* 9, 566–71 (2014).
   The work is described in Chapter 2.

2. L. J. Walport, R. J. Hopkinson, R. Chowdhury, R. Schiller, W. Ge, A. Kawamura, and C. J. Schofield. Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat. Commun*. 7, 11974 (2016).
   The work is described in Chapter 6.

3. P. L. Yeyati, R. Schiller, G. Mali, I. Kasioulis, A. Kawamura, I. R. Adams, C. Playfoot, N. Gilbert, V. van Heyningen, J. Wills, A. von Kriegsheim, A. Finch, J. Sakai, C. J. Schofield. The JmjC protein KDM3A coordinates actin dynamics with intraflagellar transport modulating cilia number and length. *J. Cell Biol.* (2017). (*Accepted)*
   The work is described in Chapter 5.

Abbreviations

2,4-PDCA	Pyridine-2,4-dicarboxylic acid
2OG	2-Oxoglutarate
6×His-tag	Hexa-histidine tag
Ab	Antibody
Alpha	Amplified luminescent proximity homogeneous assay
BRD	Bromodomain
CC_{50}	Half maximal cytotoxic concentration
ChIP-seq	Chromatin immunoprecipitation-sequencing
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats protein-9 nuclease
C-terminus	Carboxy terminus
CV	Column volumes
Da	Dalton
DNA	Deoxyribosenucleic acid
DSBH	Double-stranded β-helix
DSF	Differential scanning fluorimetry
EC_{50}	Half maximal effective concentration
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
FDH	Formaldehyde dehydrogenase
FL	Full length
h	Hour
H1	Histone H1
H2A	Histone H2A
H2B	Histone H2B
H3	Histone H3
H4	Histone H4

HDAC	Histone deacetylase
IC_{50}	Half maximal effective concentration
IFT	Intraflagellar transport protein
IOX1	5-Carboxy-8-hydroxyquinoline
JmjC	Jumonji-C
JMJD	Jumonji-C domain containing protein
KAc	*N*ε-Acetyllysine
KAcOH	*N*ε-Hydroxy-acetyllysine
k_{cat}	Turnover number
KDM	Lysine demethylase
Kfor	*N*ε-Formyllysine
K_M	Michaelis constant
Kme	*N*ε-Methyllysine
Kme2	*N*ε -Dimethyllysine
Kme3	*N*ε -Trimethyllysine
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
MALDI	Matrix-assisted laser desorption / ionisation
MS	Mass spectrometry
NOG	*N*-Oxalylglycine
N-terminus	Amino terminal
PDB	RCSB protein data bank
PHD	Plant homeodomain
PTM	Post-translational modification
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
Rme	*N*ω-Monomethylarginine
Rme2a	Asymmetric *N*ω -dimethylarginine
Rme2s	Symmetric *N*ω -dimethylarginine
RNA	Ribonucleic acid
SILAC	Stable isotope labelling by amino acids in cell culture

TCA	Tricarboxylic acid cycle
TCAI	Tricarboxylic acid cycle intermediates
T_m	Melting temperature
TOF	Time-of-flight
WT	Wildtype

Chapter 1
Introduction

1.1 Epigenetics

Genetics and developmental biology were almost completely separate fields until the second half of the twentieth century.[1] The term "epigenetics" was first named by Waddington in 1942, it links these two fields and provides a bridge between a genotype and a phenotype.[2] One definition of molecular epigenetics is that it concerns with how changes to the phenotype can occur without alternations in the 'underlying' DNA nucleotide sequence, including by modulation of gene expression levels.[3] More simply, epigenetics can be viewed as the studies of the detailed chemistry underlying genetics. Epigenetics research aims to explain why the cells in our body may share nearly the same DNA sequence (i.e. have the same genotype), but different tissues are distinct in their gene expression profiles and functions (i.e. have different phenotypes). It has been suggested that if the DNA sequence is viewed as an instruction manual text, epigenetic marks can be analogues to highlighters, which emphasise relevant sections in the text and discard others.[4]

In the past few decades the molecular mechanisms behind the epigenetic phenomena have begun to be unravel. These mechanisms include non-coding regulatory RNA, DNA methylation, and modifications to histone proteins. The latter are carried out by 'epigenetic enzymes'.[5–7] Epigenetic proteins can serve as mediators, they read the genetic information and regulate gene expression to ultimately determine the phenotype. Thus, epigenetics is an amended version to the classical view of the central dogma of biology.[8]

The ability of epigenetics to fill in some of the gaps in our scientific understanding has stimulated a great interest in related research, and the knowledge in this field is expanding rapidly. On-going research is frequently uncovering the molecular mechanisms underpinning the role of epigenetics in development and disease.[9,10]

1.2 Packaging of DNA into a chromatin structure

Each human cell contains approximately two meters of linear DNA.[11] This DNA is extensively compacted by several levels of packaging into the cell's nucleus to a volume of ~1000 μm^3. This compaction is achieved by the formation of nucleo-protein complexes in which the protein mass is nearly double that of the DNA.[12] In the first level of packaging the DNA is wrapped around histone proteins to form a nucleosome.

Histones are basic and relatively small (100-150 residues) proteins. Their conserved structure comprises a *C*-terminal globular core domain containing three α-helices separated by two loops, and an unstructured *N*-terminal tail that protrudes out of the nucleosome and is not involved in DNA binding. The standard nucleosome contains ~146 base pairs of DNA wrapped around a histone octamer complex. In animals, a typical octamer comprises of two copies of each of histone 3 (H3), histone 4 (H4), histone 2A (H2A) and histone 2B (H2B) (**Figure 1.1**).[13]

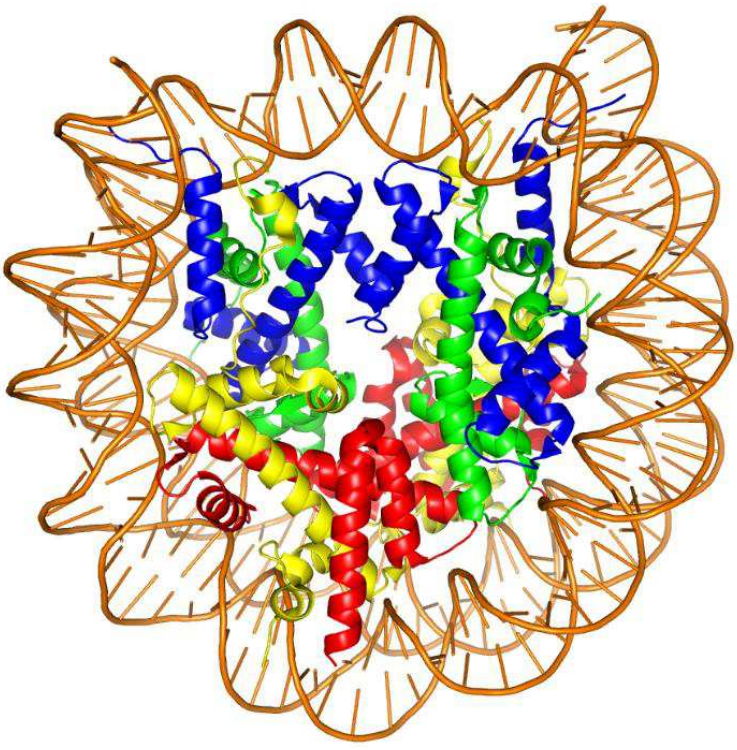


Figure 1.1 View from the crystal structure of a nucleosome (PDB ID 1KX3). View showing how the DNA superhelix (orange) coils around the histone octamer. The histone octamer comprises copies of each of histone 3 (blue), histone 4 (green), histone 2A (yellow) and histone 2B (red). Histone tails are missing because they are intrinsically unstructured in solution and no electron density was observed for them. [14]

The wrapping of the DNA around the histone complex is enabled by electrostatic interactions between the negatively charged DNA and the positively charged histones. A string of nucleosomes forms a 10 nm fibre, sometimes known as a "beads on a string" structure (**Figure 1.2**).[15,16]

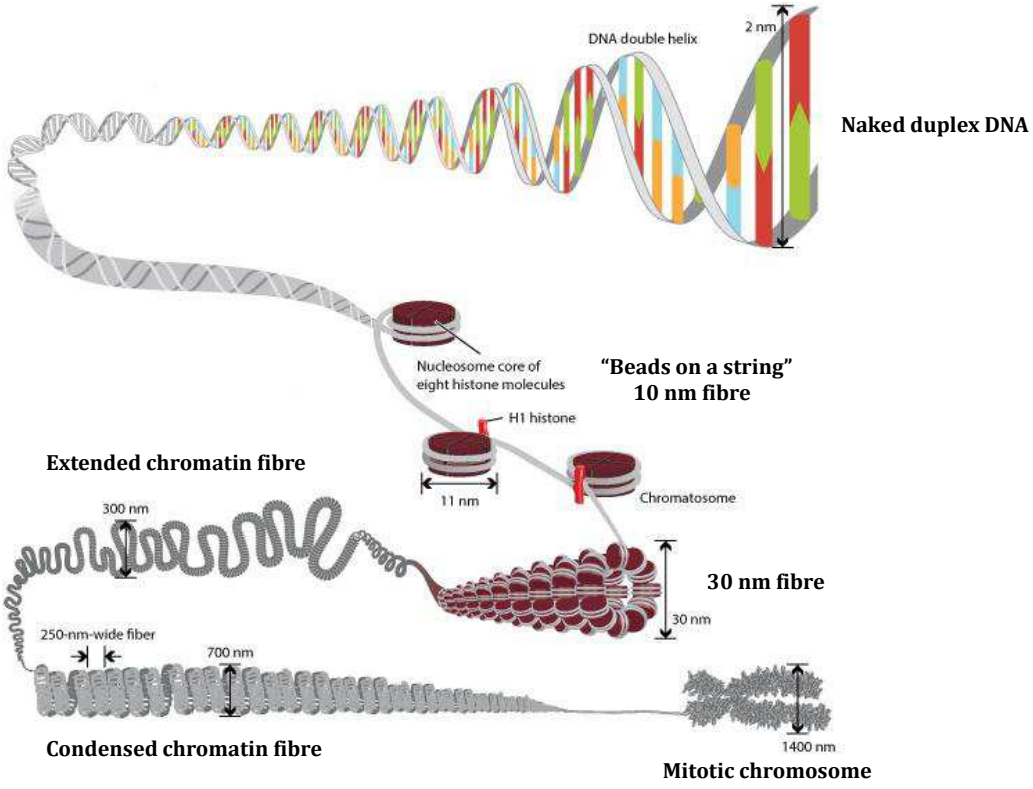


Figure 1.2 Illustration of the different levels of DNA packaging. In the first level of packaging the DNA is wrapped around histones to form the nucleosome complex. A string of nucleosomes forms a ~10 nm fibre known as "beads on a string". In the next level, H1 histones further fold up the nucleosomes to form a ~30 nm chromatin fibre. Additional folding and compression forms the chromatin fibre. During cell division the chromatin is tightly packed to form the mitotic chromosome and facilitate segregation.[16–18]

In the next level of packaging one molecule of histone H1 binds on top of the nucleosome where two turns of DNA occur, sealing the DNA to the nucleosome. In addition to binding to the nucleosome, H1 also binds to the "linker DNA" (~20-80 nucleotides in length) region between nucleosomes, thus allowing the packaging of the nucleosomes against each other to form a tight zig-zagged 30 nm fibre.[19] This fibre is likely to be the template for most nuclear processes.

Additional levels of compaction enable these fibres to be packaged into the small volume of the nucleus (~1000 μm^3).[11] The next high-order structure is the chromatin fibre. Compaction of the DNA into this structure protects the DNA from damage and allows regulation of the gene transcription and DNA replication processes.[20] The tightness of the packaging depends on the region along the DNA strand as well as on the cell cycle stage. Actively transcribed genes are more loosely packed than inactive genes to allow access to transcription related proteins. These lightly packed regions of the chromatin are referred as euchromatin.[21] Conversely, inactive chromatin regions are more condensed and are referred as heterochromatin. During cell division the chromatin is tightly packed to form the mitotic chromosome and facilitate segregation.[17]

The dynamic nature of chromatin packaging is achieved in three main ways: chromatin-remodelling complexes, histone variants and post-transcriptional modifications. Chromatin-remodelling complexes are reported to unpackage chromatin by a number of mechanisms including sliding nucleosomes, restructuring nucleosomes or replacing canonical histones with their variants.[22,23] Chromatin-remodelling complexes are divided into four families which differ in their specific functions: the SWI/SNF (switching defective/sucrose nonfermenting) family, the ISWI (imitation switch) family, the INO80 (inositol requiring 80) family and the CHD (chromodomain, helicase, DNA binding) family.[24–27] All chromatin-remodelling protein families have a DNA-dependent ATPase domain, their mechanisms are currently unknown but it is hypothesised that they move DNA by binding to the nucleosome and then pumping the DNA around it.[28,29]

The second way cells control chromatin packaging is by the incorporation of histone variants. Although histones are among the most conserved of all studied proteins, the canonical histones have variants which share the same histone core structure, but which exhibit variations in their primary sequence.[30] When histone variants replace the canonical histones in the nucleosome they can alter the structure of the nucleosome, the capacity of DNA packaging and the post-translational modifications they present (see Section 1.3).[31] Unlike canonical histones, the amount of the histone variants is not cell cycle-dependent, but is associated with a specific biological context. For instance, histone H3.3 replaces the canonical histone in active chromatin regions.[32] **Table 1.1** summarises some of the reported histone variants and their function.

Table 1.1 Histone variants and their function. Some of the reported histone variants that replace the canonical histones, their origin, chromatin effect and function.[11]

Histone	Variant	Species	Chromatin effect	Function
H1	SpH1	Sea urchin	Chromatin condensation	Chromatin packaging
	H1t	Mouse	Open chromatin	Histone exchange
H2A	MacroH2A	Vertebrate	Condensed chromatin	X-chromosome inactivation
	H2ABbd	Vertebrate	Open chromatin	Transcription activation
	H2A.X	Ubiquitous	Condensed chromatin	DNA repair/recombination /transcription repression
	H2A.Z	Ubiquitous	Open/closed chromatin	Transcription activation/repression, chromosome segregation
H2B	SpH2B	Sea urchin	Chromatin condensation	Chromatin packaging
H3	CenH3	Ubiquitous		Kinetochore formation/function
	H3.3	Ubiquitous	Open chromatin	Transcription

A third way to regulate chromatin packaging is by post-translational modifications on histone tails (see next section).

1.3 Histone post-translational modifications

DNA is tightly folded around histones to form the nucleosomal fibre. The key question of how the DNA is transcribed under such tight packaging has become vital to our understanding of epigenetics as a whole. Extensive research has revealed that chromatin structure is dynamically modulated by interactions between epigenetic enzymes and post-translational modifications (PTMs) present on the histone 'tails'.[33]

The *N*-terminal region of histones protrudes from the compact nucleosome and is largely disordered in solution. Histone tails are amongst the most highly conserved sequences in eukaryotes. Although these tails are relatively short (for H4 and H3 the length is 19 and 26 residues, respectively), they are also amongst the most highly post-translationally modified

proteins within the eukaryotic proteome. Of the modified residues along the histone tail, lysine has been reported to be the most modified residue.[34]

PTMs on histone tails include acetylation of lysines, methylation of lysines and arginines, phosphorylation of threonines and serines, and ubiquitination of lysines.[35] Of these modifications, lysine acetylation and methylation are the most extensively studied.[36] Over the past decade, more than twenty novel histone PTMs have been discovered, including lysine acylation PTMs such as formylation, crotonylation and 2-hydroxyisobutyrylation.[37–40] These discoveries were made possible by recent advances in proteomic and the development of selective antibodies.[37–39,41] **Figure 1.3** outlines the main array of currently reported PTMs on histones.

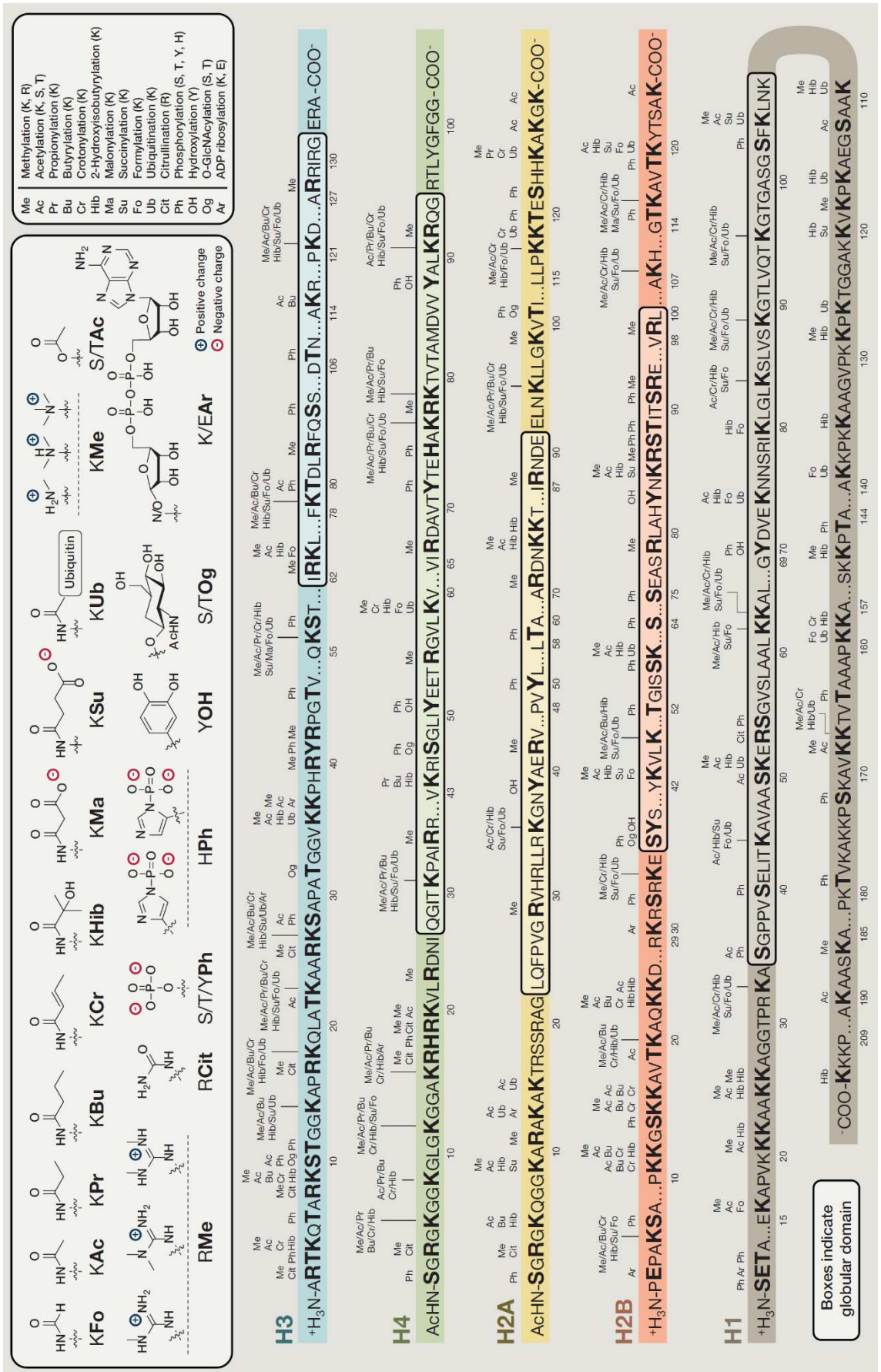

KFo
KAc
KPr
KBu
KCr
KHib
KMa
KSu
KUb
KMe
S/TAc
RMe
RCit
S/T/YPh
HPh
YOH
S/TOg
K/EAr
Ubiquitin
Positive charge
Negative charge
Me Methylation (K, R)
Ac Acetylation (K, S, T)
Pr Propionylation (K)
Bu Butyrylation (K)
Cr Crotonylation (K)
Hib 2-Hydroxyisobutyrylation (K)
Ma Malonylation (K)
Su Succinylation (K)
Fo Formylation (K)
Ub Ubiquitination (K)
Cit Citrullination (R)
Ph Phosphorylation (S, T, Y, H)
OH Hydroxylation (Y)
Og O-GlcNAcylation (S, T)
Ar ADP ribosylation (K, E)
H3
+H3N-ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTV...QKST...IRKL...FKTDLRFQSS...DTN...AKR...PKD...ARRIRG ERA-COO-
H4
AcHN-SGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRR...VKRISGLIYEETRGVLKV...VIRDAVTYTEHAKRKTVTAMDVV YALKRQG RTLYGFGG-COO-
H2A
AcHN-SGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYAERV...PVYL...LTA...ARDNKKT...IRNDE ELNKLLGKVTI...LLPKKTESHHKAKGK-COO-
H2B
+H3N-PEPAKSA...PKKGSKKAVTKAQKKD...RKRSRKESYS...YKVLK...TGISSK...S...SEASRLAHYNKRSTITSRE...VRL ...AKH...GTKAVTKYTSAK-COO-
H1
+H3N-SETA...EKAPVKKKAAKKAGGTPRKASGPPVSELITKAVAASKERSGVSLAALKKALK...GYDVEKNNSRIKLGLKSLVSKGTLVQTKGTGASGSFKLNK
-COO-KKKP...AKAASKA...PKTVKAKKPSKAVKKTVTAAAPKKA...SKKPTA...AKKPKKAAGVPKKPKTGGAKKVKPKAEGSAAK
Boxes indicate globular domain

Figure 1.3 Summary of the reported histone marks of H3, H4, H2A, H2B and H1 histones, including the recently identified lysine acylation marks. Figure reproduced with permission.[35]

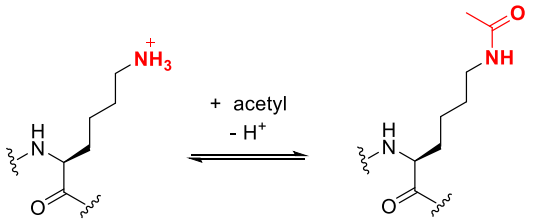
Some histone tail PTMs correlate with transcriptional activation while others correlate with transcriptional repression, depending on the types and the position of the PTMs on the histone tails.[42,43] The cross-talk between different modifications is also likely to play a role in transcription regulation. These PTMs serve as 'marks', and together they represent the 'histone code' hypothesis of epigenetic regulation.[44] In this theory, these marks can dynamically modulate the chromatin structure to either induce or repress gene transcription. Histone marks can regulate chromatin packaging via two mechanisms:[42,45–47] first, PTMs can alter the chemical properties of their residues and thus directly affect the DNA-histone interactions. Second, these histone marks can recruit chromatin-modifying enzymes through binding. Recent studies indicate that histone modifications have context dependent effects, making their interplay more like a language within the chromatin signalling pathway than a code.[47]

Histone marks are regulated by epigenetic enzymes. PTMs are added by 'writers', removed by 'erasers' and identified by 'readers'.[48] Reversible changes to the 'histone code' results in dynamic modulation of the chromatin structure that will ultimately define the accessibility of genes for transcription and will translate into biological outcomes.[49]

1.4 Histone acetylation

Acetylation on lysyl-residues of histone tails is a well-studied epigenetic modification. Studies during the nineties have shown that acetylated chromatin has an increased sensitivity to DNase I due to increased accessibility to interacting proteins.[50,51] These studies were later corroborated by chromatin-immunoprecipitation (ChIP) studies in yeast.[52] The conclusion from such studies was that acetylation neutralises the positive charge on the lysyl-residue, and thereby weakens the interaction between the histones and the negatively charged DNA (**Scheme 1.1**). The biological implication of loosening chromatin structure is that the DNA becomes more accessible, so that the transcription machinery can reach the promoter regions and initiate transcription. Therefore, histone acetylation is associated with euchromatin and enhanced gene transcription. Lysine acetylation is the most studied modification in a growing list of lysine acylation marks which

includes crotonylation, malonylation, succinylation, formylation, and the recently discovered 2-hydroxyisobutyrylation.[37–40]



Scheme 1.1 Acetylation neutralises the positive charge on the lysyl-residue.

1.5 Histone acetyllysine readers

In addition to direct weakening of histone – DNA interactions by the removal of the positive charge, acetylation also affects the chromatin structure by being recognised by a number of histone reader domains in chromatin-associating proteins.

Most histone acetyllysine readers belong to the bromodomain containing family of proteins which can selectively interact with acetylated lysines. In humans there are over 60 bromodomain-containing proteins.[53] Most of the proteins in this family have multiple domains and are chromatin-associated. Bromodomain modules share a conserved fold that consist of a left-handed bundle of four α-helices that are linked by diverse loop regions which surround the acetylated lysine (**Figure 1.4**).[54]

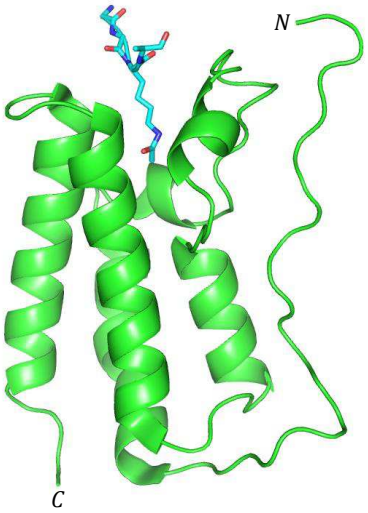


Figure 1.4 Representative structure of a bromodomain acetyllysine reader. View from the crystal structure of bromodomain 1 of mouse Brd4 (green) in complex with histone H3 K14Ac (cyan), showing the classical bromodomains structure that consists of a left-handed 4 α-helix bundle with a binding cavity surrounding the acetyl moiety (PDB ID 3JVK).[55]

The plant homeodomains (PHDs) are another important class of histone acetyllysine reader domains. Although PHDs are mainly reported to recognise histone tail methylation, some PHDs have been found to associate with acetyllysine residues (see Section 1.9 for methyllysine recognition).[56,57] All of the PHD modules share a typical conserved zinc finger topology (see Section 1.9 for structural information). The ability of PHDs to bind different substrates depends on variations in the loops that connect the zinc fingers, similarly to the mechanism of recognition of various substrates by bromodomains. [58,59]

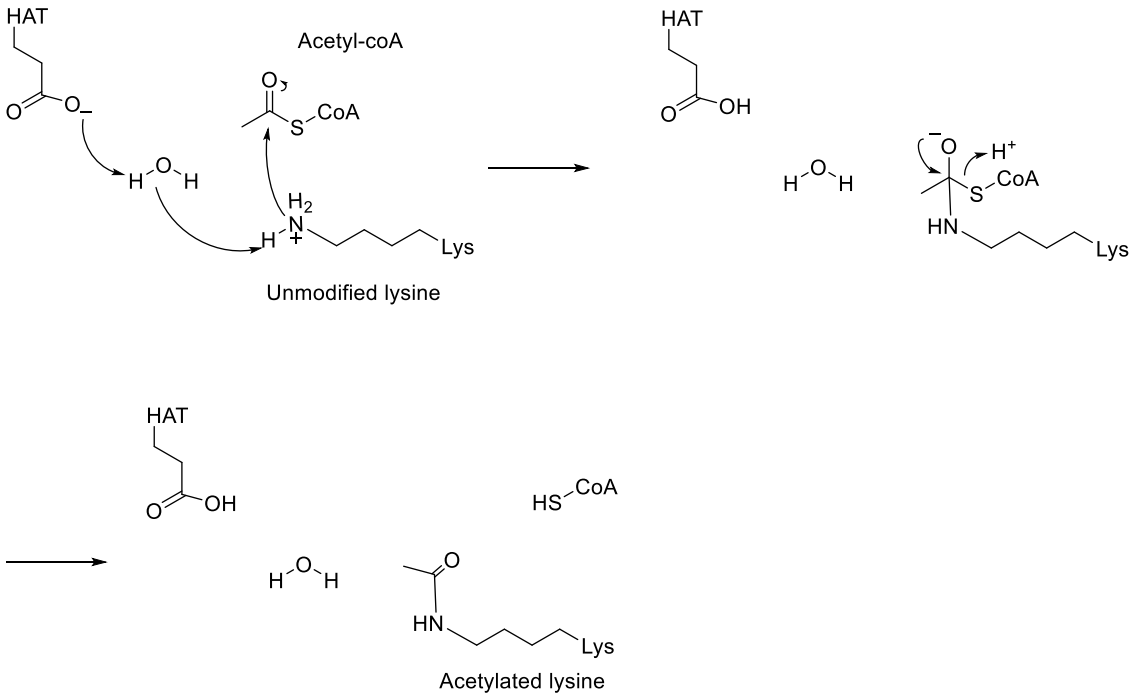
The pleckstrin homology (PH) domain is also reported as an histone acetyllysine reader, acetylation on histone H3 lysine 56 (H3 K56Ac) has been reported to be recognised by the double PH domain of the histone chaperone Rtt106.[60] However, PH domains are more commonly known to bind phosphatidylinositol lipids within biological membranes.

1.6 Histone acetyl-transferases

Acetyl groups can be added to the *N*ε-amino group of lysine residues by histone acetyl-transferases (HATs). HATs utilise acetyl-coenzyme A (acetyl-coA) as their sole donor of acetyl groups and therefore acetyl-coA levels in cells affect histone acetylation.[61] Since

acetyl-coA is a key intermediate in the tricarboxylic acid (TCA) cycle, HATs directly connect energy homeostasis and gene expression.[62]

HATs are divided into four main subfamilies depending on their sequence similarity: the Gcn5-related *N*-acetyltransferases (GNATs), named after their similarity to the Gnc5 HAT. The MYST named for their founding members MOZ, Ybf2, Sas2, and Tip60. The additional HAT subfamilies are smaller and include the p300/CBP and Rtt109 subfamilies.[63] HATs differ in their selectivity for various histone positions. Some HATs have also been reported to catalyse acetylation of non-histone substrates.[64] Different families of HATs employ different catalytic mechanisms to transfer an acetyl group from acetyl-CoA to the N^{ε}-amino group of a target lysine side-chain.[65] The superfamily of GNAT HATs employ a sequential mechanism in which acetyl-CoA and the histone substrate form a ternary complex with the enzyme. A conserved glutamate residue in the active site activates a water molecule for the removal of a proton from the amine group on the target lysine residue (**Scheme 1.2**). The nucleophilic deprotonated lysine then attacks the electrophilic carbonyl carbon of the enzyme-bound acetyl-CoA and the acetylated lysine product is formed, with concomitant release of CoA.[63]

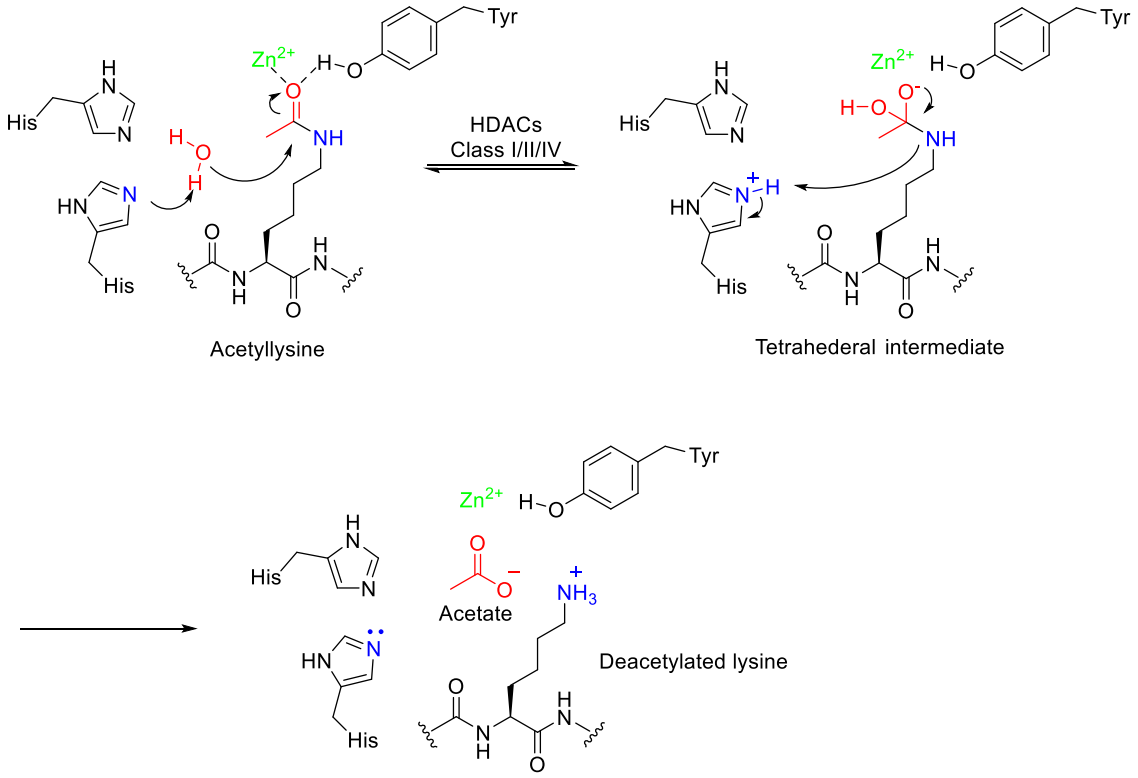


Scheme 1.2 Proposed outline catalytic mechanism for the GNAT family of histone acetyl-transferases (HATs). A conserved glutamate residue in the active site activates a water molecule for the removal of a proton from the amine group on the target lysine residue. The nucleophilic deprotonated lysine then attacks the electrophilic carbonyl carbon of the enzyme-bound acetyl-CoA and the acetylated lysine product is formed with concomitant formation of CoA.[65]

1.7 Histone deacetylases

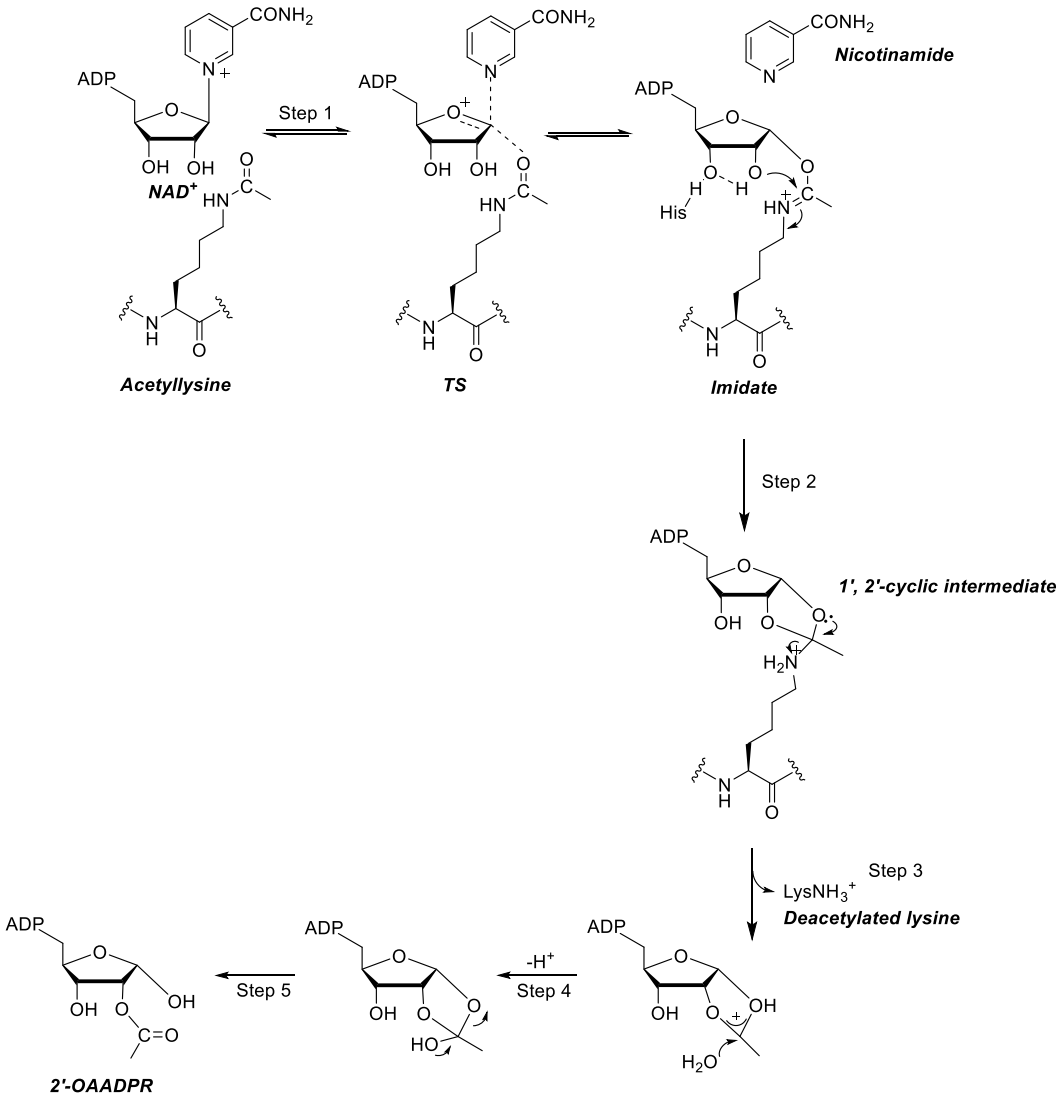
Histone deacetylases (HDACs) remove an acetyl group from the N^{ε}-amino groups of lysine residues on histone tails, thereby promoting a more condensed packaging of the chromatin and repressing gene expression.[66] Mammals have 18 reported HDACs which are divided into four classes based on their sequence homology.[67] Similarly to HATs, HDACs differ in their selectivity for various histone positions and are also reported to accept non-histone substrates. HDACs are therefore involved in a variety of biological processes.[67–69] Class I, II and IV HDACs are proposed to share a conserved zinc-dependent catalytic mechanism, whereas Class III enzymes (also known as sirtuins) are nicotinamide adenine dinucleotide (NAD^+)-dependent. The Class III HDACs are thus important in the regulation of both metabolism and transcription.[70,71]

The proposed mechanism of class I/II/IV zinc-dependent HDACs is outlined in **Scheme 1.3**. In this mechanism a conserved tyrosine residue at the active site and the catalytic zinc polarise the carbonyl group of the acetyl to make it more electrophilic.[72] Two histidine residues at the active site are proposed to activate a water molecule for nucleophilic attack of the acetyl carbonyl carbon. An unstable tetrahedral intermediate is formed, and spontaneously fragments to form the deacetylated lysine with concomitant release of an acetate.



Scheme 1.3 Proposed catalytic mechanism of class I/II/IV zinc-dependent histone deacetylases (HDACs). A conserved tyrosine residue at the active site and the catalytic zinc polarise the carbonyl group to make it more electrophilic. Two histidine residues at the active site activate a water molecule for nucleophilic attack of the acetyl carbonyl carbon. An unstable tetrahedral intermediate is formed which spontaneously fragments to the deacetylated lysine with concomitant release of an acetate ion.[72]

Scheme 1.4 shows the proposed mechanism of class III NAD^+-dependent HDACs. The first step of the enzymatic catalysis involves cleavage of NAD^+ by reversible adenosine diphosphate ribose (ADP)-ribosylation of the acetyl carbonyl to form nicotinamide and a peptidylimidate intermediate, in which the acetate group resides on the ribose ring where the cleavage occurred.[73] This reactive intermediate then undergoes an intramolecular attack by the 2'-OH group to yield 2'-*O*-acetyl-ADP-ribosyl-(2'-*O*AADPR) and the deacetylated lysine (Steps 2-5).



Scheme 1.4 Proposed outline catalytic mechanism for the class III sirtuin catalysed NAD⁺-dependent histone deacetylases (HDACs). NAD⁺ is cleaved by reversible adenosine diphosphate ribose (ADP)-ribosylation of the acetyl carbonyl to form nicotinamide and a peptidylimidate intermediate. This reactive intermediate then undergoes an intramolecular attack by the 2'-OH group to yield 2'-O-acetyl-ADP-ribosyl-(2'-*O*AADPR) and the deacetylated lysine.[73]

1.8 Histone methylation

Methylation is one of the most important modifications to proteins and nucleic acids.[74] Histone methylation has been reported on the side-chains of both arginines and lysines (arginine methylation is described in Chapter 6). Lysine methylation has been reported on all histone tails and there is also one example for methylation on a core histone residue (H3 K79).[75,76]

Unlike acetylation, methylation is considered a subtle modification as it does not significantly alter the charge or the polarity of the residue. Therefore, methylation can signal for either transcriptional activation or repression, depending on the position of the mark on the histone and its methylation state.[77] The mechanisms by which methylation results in activation or repression are unclear, though evidence suggests that methylation can recruit chromatin-modifying enzymes through their methyl-binding domains. For example, H3 K4me3 is mainly observed on promoter regions associated with transcriptionally active genes.[78–81] The H3 K4me3 mark can mediate transcriptional activation by association with the MYST family of HATs via their PHD finger domain. These HATs acetylate histone tails and subsequently induce transcriptional activation (see Section 1.9).[82,83] By contrast, the H3 K9me3 mark is mainly reported to be associated with silent genes and the centromere region which is largely repressed.[84] H3 K9me3 represses gene transcription by recruiting heterochromatin protein 1 (HP1), a chromodomain containing protein that facilitates the formation of heterochromatin leading to repressed gene transcription (see Section 1.9 for chromodomain).[85]

1.9 Histone methyllysine readers

PHDs can bind a number of histone modifications including acetylated and methylated lysine residues. Their substrate selectivity is conferred by variations amongst the loops in the PHD fingers (see Section 1.5 for acetyl recognition).[59] Like other histone readers, the PHDs are commonly found in chromatin remodelling proteins.[59] All PHDs share a conserved structure with a Cys_4-His-Cys_3 ligand motif made up of around 60 amino acids that coordinates two zinc ions (**Figure 1.5**).[86] The PHDs are mainly reported to recognise unmodified or methylated H3 K4 mark. The catalytic activity of enzymes containing PHD fingers can be altered by binding of the PHD finger to the substrate.[87] For instance, KDM7B demethylates H3 K9me1/2, and its catalytic activity is stimulated by binding to H3 K4me3 via its PHD domain.[88] However, there is also evidence that PHDs can recognise other methylated positions such as H3 K9 which is apparently recognised by the PHD containing enzyme KDM5C.[89] The PHDs can also recognise the absence of demethylation in the H3 K4 position. For instance, H3 K4me0, but not methylated H3 K4, is recognised by the PHD domain protein BHC80 which then facilitates binding of the LSD1 histone demethylase to the chromatin (see Section 1.11 for histone demethylase).[90]

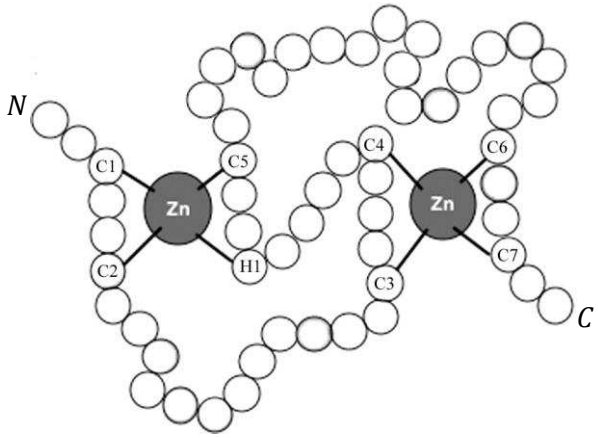


Figure 1.5 Cross-brace zinc ligation is a conserved pattern in PHD domains. All PHDs share a conserved structure with a Cys4-His-Cys3 ligand motif (grey circles) made up of around 60 amino acids that coordinates two zinc ions. Figure adapted with permission.[91]

Many chromatin remodelling proteins contain a chromodomain. Chromodomains bind methyllysine residues on histone tails in heterochromatin regions.[92,93] A typical chromodomain has 50 amino acids that fold into 3 conserved anti-parallel β-sheets and a single α-helix. The histone methyllysine substrate is bound in a cleft formed between a segment of the β-sheets and the loop which connects them to the α-helix (**Figure 1.6**).[94]

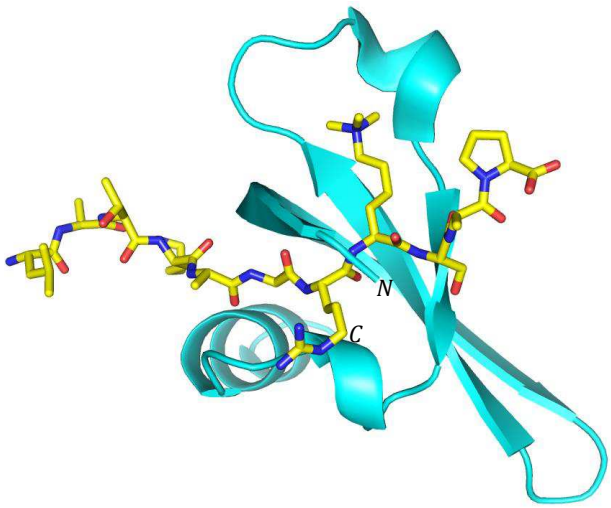


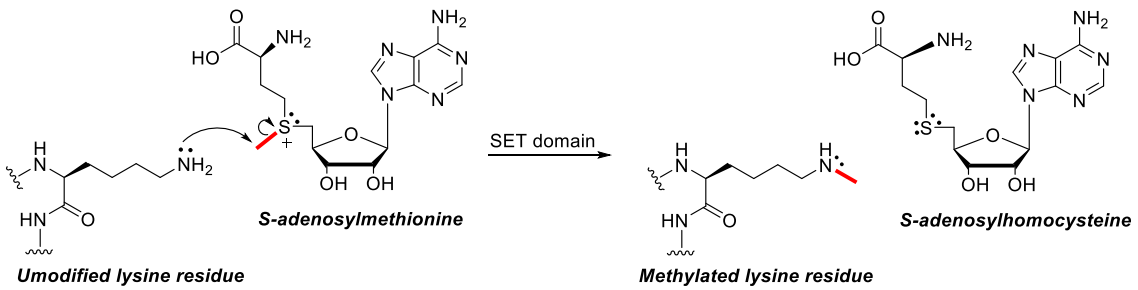
Figure 1.6 Representative view from a structure of a chromodomain methyllysine reader. View from a crystal structure of the *Drosophila* polycomb chromodomain (cyan) in complex with H3 K27me3 peptide (yellow) showing a typical chromodomain with 50 amino acids that fold into 3 conserved anti-parallel β-sheets and a single α-helix. The histone methyllysine substrate is bound in a cleft formed between a segment of the β-sheets and the loop which connects them with the α-helix (PDB ID 1PFB).[94]

Other methyllysine readers include: the tudor domains, the MBTs, the PWWP and the WD.[95] The tudor domains bind di-and tri- methyllysine residues, and can repeat twice in the same protein sequence (tandem tudor domains). The MBTs bind mono- and di- methylated lysines only when their sequence repeats 2-4 times. The PWWP domain has an aromatic methyllysine binding pocket composed of the PWWP residues, and the WD has single or multiple repeats of a propeller-type structure with 7 "vanes" to form a methyllysine binding domain.

1.10 Histone lysine methyltransferases

Methylation of histone lysine and arginine residues is catalysed by histone methyltransferases (HMTs) which utilise S-adenosyl methionine (SAM) as their methyl source. Histone lysine methylation is catalysed by lysine methyltransferases (KMTs) which contain the SET catalytic domain (with the exception of KMT4).[96,97] The KMTs have varied substrates, and their selectivity is dependent on both the sequence and the methylation state of the lysine, KMTs are also reported to catalyse methylation of non-histone substrates.[98,99]

For KMT methylation to occur, SAM and the target histone lysine residue need to bind to the catalytic core of the SET domain in the right orientation.[100] The consensus methylation mechanism is proposed to be S_N2 substitution involving nucleophilic attack of the lysine residue on the electrophilic methyl group on the sulphur atom of the SAM molecule, resulting in a methyl transfer to the lysine residue (**Scheme 1.5**).

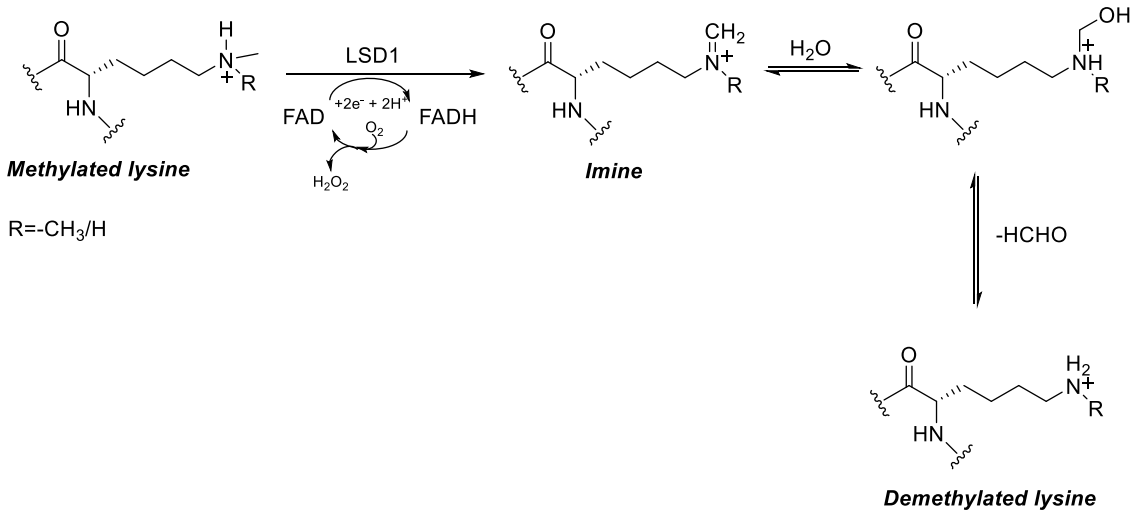


Scheme 1.5 Proposed catalytic mechanism for histone lysine methyltransferases (KMTs). Methyl group is shown in red. S-adenosyl methionine (SAM) and the histone lysine residue bind to the catalytic core of the SET domain. The N^{ε}-amino group of the lysine residue is deprotonated by an active site tyrosine residue. S_N2 substitution by a nucleophilic attack of the deprotonated lysine residue on the electrophilic methyl group on the sulphur atom of the SAM molecule results in a methyl transfer to the lysine residue.[100]

1.11 Histone lysine demethylation

Lysine methylation was considered irreversible until the discovery of the first histone demethylase in 2004.[101] The first lysine demethylase to be identified was the Lysine Specific Demethylase 1 (LSD1, also known as KDM1A), which is a flavin adenine dinucleotide (FAD) dependent oxygenase.[101]

As a typical FAD dependent oxygenase, the demethylation mechanism of LSD1 involves oxidation of the methylated lysine histone substrate to imine, with concomitant reduction of FAD to $FADH_2$ and reduction of molecular oxygen to hydrogen peroxide to regenerate FAD. The resulting imine intermediate is the hydrolysed non-enzymatically to produce the demethylated lysine with release of formaldehyde (**Scheme 1.6**).[102]



Scheme 1.6 Proposed catalytic mechanism for LSD1 histone lysine demethylase. Oxidation of the methylated lysine histone substrate to imine, with concomitant reduction of FAD to $FADH_2$, and reduction of molecular oxygen to hydrogen peroxide to regenerate FAD. The resulting imine intermediate is hydrolysed non-enzymatically to produce the demethylated lysine with release of formaldehyde.[101]

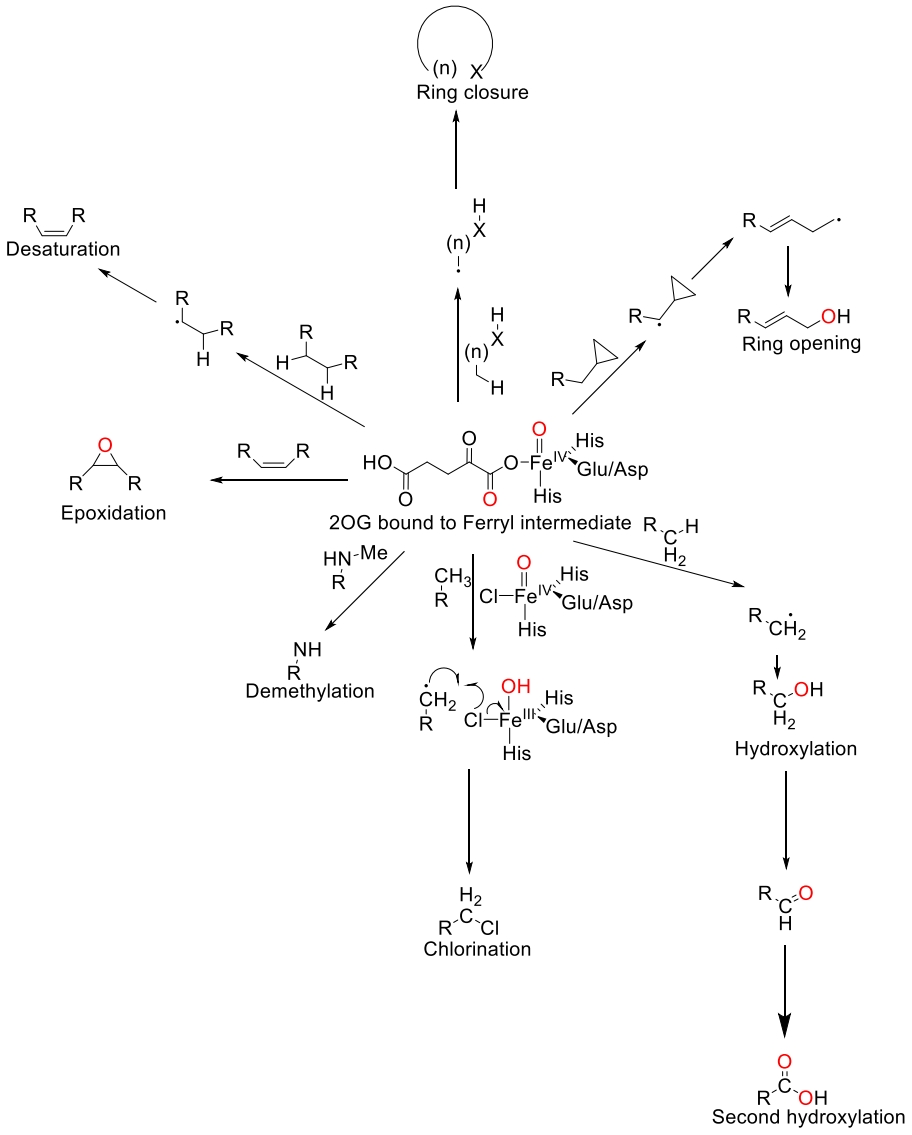
Since an imine is formed during the catalysis mechanism, and imines cannot be readily formed from the quaternary ammonium species, it is likely that LSD1 can only accept mono- and dimethylated lysines as substrates but not trimethyllysine residues. The reported substrates for LSD1 include H3 K4me1/me2 and H3 K9me1/me2 (in a content-dependent manner) as well as a number of non-histone substrates such as the DNA methyltransferases DNMT1 and the transcription factor p53.[103,104]

Subsequently, another member of this family, LSD2 (also known as KDM1B) was identified as an H3 K4me1/me2 demethylase.[105] More recently LSD2 has been reported as an E3 ubiquitin ligase that directly ubiquitylates and thus promotes the proteasome-dependent degradation of O-GlcNAc transferase (OGT).[106]

Histone lysine demethylases can be divided into two classes depending on their catalytic mechanism. The second and larger class of histone lysine demethylases are the Jumonji-C histone lysine demethylases (JmjC KDMs) which belong to the superfamily of 2-oxoglutarate dependent dioxygenases (2OG oxygenases). 2OG oxygenases and JmjC KDMs are described in the next sections (Sections 1.12 - 1.13).

1.12 2-Oxoglutarate dependent dioxygenases

The second class of histone lysine demethylases belong to the superfamily of 2-oxoglutarte dependent dioxygenases (2OG oxygenases).[107] 2OG dependent oxygenases are non-haem enzymes that utilise 2OG and molecular oxygen to catalyse a wide variety of oxidation reactions including hydroxylation and demethylation (**Scheme 1.7**).[108]



Scheme 1.7 The variety of oxidation reactions catalysed by eukaryotic and prokaryotic 2OG oxygenases. Oxygen atoms derived from molecular oxygen are shown in red. Figure adapted with permission.[108]

The 2OG dependent oxygenase superfamily of enzymes is present in most aerobic organisms, with varied distribution depending on the selectivity of the catalysed reaction. The oxidation substrates of 2OG oxygenases can be divided to six major categories:[107] 2OG oxygenases can act on proteins to regulate various processes such as oxygen sensing and histone modifications.[109,110] They can also catalyse polynucleotide modifications and are involved in DNA and RNA repair and transcriptional regulation.[111,112] Some 2OG oxygenases catalyse lipid oxidation and are reported to play a role in metabolism.[113] In plants, some 2OG oxygenases specialise in catalysing modifications of plant-specific metabolites such as flavonoids and gibberellins.[114,115] 2OG oxygenases can also act on small molecules including free amino acids and nucleosides.[116,117] Finally, 2OG oxygenases are also utilised extensively for antibiotic biosynthesis.[118]

Over 60 members of 2OG oxygenases have been identified in humans and other animals. Known catalysis products of these oxygenases are currently limited to hydroxylation and demethylation products. The most studied reactions by human enzymes are those involved in collagen biosynthesis (hydroxylation of prolyl- and lysyl- residues in collagen), the hypoxic response (hydroxylation of hypoxia inducible factor-1α, HIF-1α) and epigenetic regulation (demethylation of histones).[109,110,119]

Despite the diverse array of reactions catalysed by 2OG oxygenases, this superfamily of enzymes shares a similar catalytic domain and all of their reactions are catalysed by the same conserved reaction mechanism. Crystallographic studies on 2OG oxygenases have revealed a conserved double-stranded β-helix (DSBH) or a jelly-roll core fold.[120] Most of the 2OG oxygenases coordinate the active site Fe^{II} at one end of this double-stranded β-helix by using a His-X-Asp/Glu-X_N-His (2-His-1-carboxylate facial triad) motif.[121] The 2OG co-substrate binds within the active site core and the target substrate binds to less conserved regions which enable substrate selectivity (**Figure 1.7**).[122]

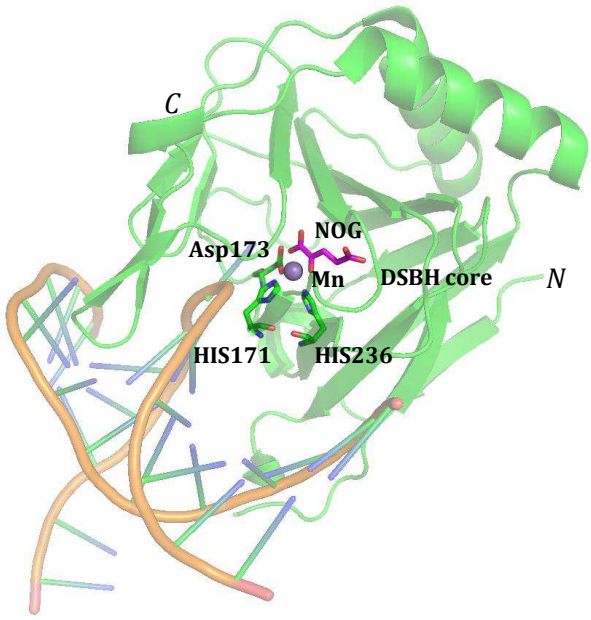


Figure 1.7 Representative structure of a 2OG oxygenase. View from a crystal structure of ABH2 (an AlkB homolog, in green) in complex with its DNA substrate (orange) showing a typical double-stranded β-helix (DSBH) core fold, with a His-X-Asp/Glu-XN-His catalytic trade. *N*-oxalyglycine (NOG) is used as a 2OG mimic and is shown in pink, Manganese (Mn) is shown in grey (PDB ID 3BUC).[122]

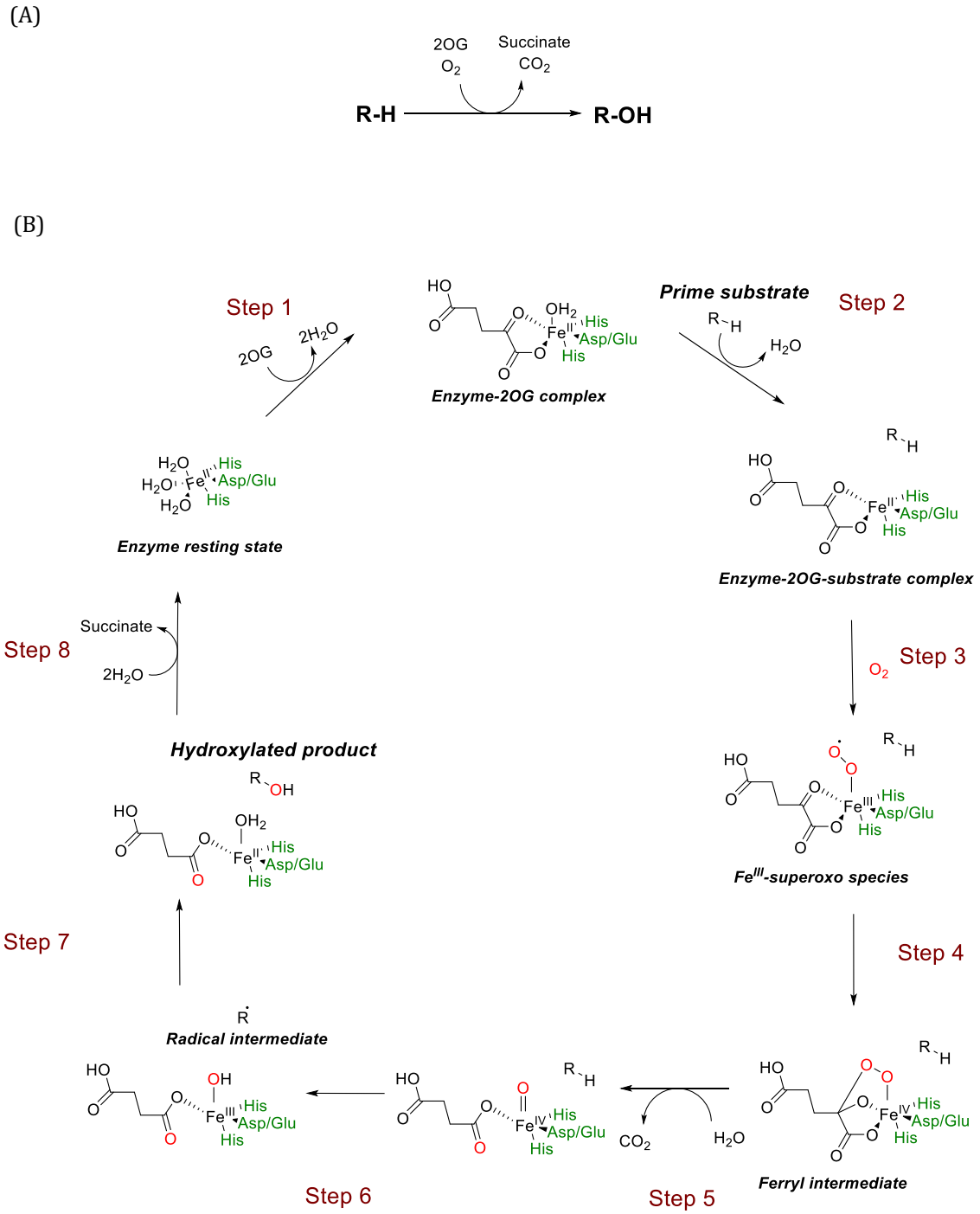
1.12.1 Mechanism of oxidation by 2OG dependent oxygenases

The available evidence is that 2OG oxygenases employ a consensus oxidation mechanism. The first comprehensive proposal of the 2OG oxygenase catalysed oxidation mechanism was suggested in 1982 based on the limited experimental data that were available at that time.[123] The validity of this mechanism was later largely confirmed by experimental studies on 2OG oxygenases, including taurine dioxygenases (TauD).[124]

The consensus mechanism by which 2OG dependent oxygenases are proposed to catalyse hydroxylation is outlined in **Scheme 1.8**, and the individual steps are as follows:[107,125]

1) The co-substrate 2OG binds to the active site Fe^{II} in a bidentate manner via its C-1 carboxylate and C-2 oxo groups, this binding displaces two water molecules that were bound the Fe^{II} in its resting state.
2) The prime substrate then binds to the active site. This leads to weakening of the binding between the remaining water molecule and the Fe^{II}. Displacement of this water molecule causes the iron to react with the oxygen molecule.
3) The metal binds molecular oxygen and a Fe^{III}-superoxo species is generated.

4) The molecular oxygen also binds to 2OG and a Fe^{IV}-oxo species, also known as a ferryl intermediate, is formed.
5) Oxidative decarboxylation produces succinate and CO_2.
6) The ferryl intermediate is a strong oxidant and it extracts a hydrogen atom from the prime substrate to generate a radical and a Fe^{III}-hydroxide.
7) A radical rebound process hydroxylates the radical substrate, to form the hydroxylated product and recycle the enzyme into its resting Fe^{II} state.
8) The hydroxylated product and the succinate leave and two water molecules bind to Fe^{II} in preparation for the next catalytic cycle.



Scheme 1.8 Proposed catalytic mechanism for hydroxylation by 2OG dependent oxygenases. (A) Overall stoichiometry for a 2OG oxygenase catalysed hydroxylation reaction. (B) Outline of the consensus steps in the hydroxylation mechanism as described in text. Oxygen derived from molecular oxygen is shown in red, the 2-His-1-carboxylate facial triad is shown in green. Subsequent spontaneous reactions, such as demethylation, may result from initial substrate hydroxylation.[107,125]

To conclude, the main features of this mechanism are: sequential binding of the co-substrate and the prime substrate leads to a formation of a ternary enzyme – Fe^{II} – 2OG – substrate complex. Oxidation of the catalytic iron proceeds via concomitant reduction of molecular oxygen to generate a reactive ferryl intermediate, and is coupled to conversion of 2OG to succinate. The reactive ferryl intermediate then hydroxylates the substrate by insertion of the second atom of the molecular oxygen into a C-H bond by a radical mechanism, hence the name 2-oxoglutarate dependent dioxygenases: these enzymes are dependent on the tricarboxylic acid (TCA) cycle intermediate 2-oxoglutarate for enzymatic catalysis, and they also incorporate both atoms of the molecular oxygen into substrates, with one atom assimilated in succinate and the other in the hydroxylated product.[126] Their co-substrate dependence is proposed to enable 2OG oxygenases to interfere with both the TCA cycle and oxygen sensing as well as additional physiological functions.[127–129]

1.13 The JmjC histone demethylases

The name Jumonji originates from Japanese meaning 'cruciform'. It originates from the first report (1995) on the Jumonji family that described an abnormal 'cross-like' appearance of neural grooves in *Jmj* knockout mice corresponding to JARID2.[130]

The Jumonji C (JmjC) proteins are a subfamily of the 2-oxoglutarate (2OG)-dependent oxygenases. As members of the 2OG oxygenase superfamily the catalytic oxidation mechanism of the JmjC proteins is dependent on 2OG as a co-substrate and Fe^{II} as a co-factor. Although the 2OG oxygenases are present in most aerobic organisms and catalyse a range of reactions, the JmjC enzymes are only present in animals and their catalysis is currently reported to be limited to hydroxylation and demethylation.[131]

Like all 2OG oxygenases, the JmjC enzymes contain a conserved jelly-roll motif and the metal is coordinated by a His-X-Asp/Glu-X_N-His catalytic triad (see Section 1.12).[74,132] However, the JmjC domain is distinct from the classical 2OG domain in that the basic 2OG binding residue is located in the 4th β-strand of the double-stranded β-helix (DSBH) fold rather than on the 8th strand, a difference that might affect substrate selectivity.[133] Exceptions to this include KDM3B, in which the 2OG binding residue is on the 8th β-strand (PDB ID: 4LXL). Another difference between JmjC and many other 2OG oxygenases is that the basic 2OG binding residue of the JmjC enzymes is a lysine, rather than the arginyl-residue found in many 2OG oxygenases.[110]

In humans, over 30 JmjC proteins have been identified, the majority of which (> 20) have been validated as demethylases with the remainder being hydroxylases (i.e. giving stable alcohol products) or have unidentified catalytic activities.[134,135]

Hydroxylation by JmjC hydroxylases has been observed on asparginyl, histidyl and lysyl side-chains in proteins, as well as on tRNA to give stable hydroxylated products.[135–138] Prior to this study, demethylation by JmjC demethylases had only been observed with N^{ε}-methylated lysyl-residues, although studies with unnatural lysine analogues bearing a larger *N*-substituents such as ethyl and isopropyl groups had also been shown to be catalysed by recombinant KDMs.[139] Subsequent studies described in this thesis suggest that catalysis by JmjC KDMs might not be limited to lysine demethylation (see Chapters 6-9).

JmjC demethylases differ from JmjC hydroxylases in both their reported biological roles and in their structure.[140] JmjC demethylases mostly act on histones and are known as epigenetic regulators. JmjC hydroxylases, on the other hand, act on non-histone substrates and therefore have other biological functions. One example is FIH which hydroxylates the hypoxia inducible factor (HIF) transcription factor.[137] Although the JmjC demethylases are mainly reported to act on histones, emerging data suggest they can also catalyse oxidation of non-histone substrates and thus are likely to have additional physiological roles.[141,142] There are a number of structural differences between JmjC demethylases and hydroxylases: first, demethylases are typically 2-3 fold larger than hydroxylases as they possess non-catalytic domains, in addition to the catalytic JmjC domain, within their structure.[140] Many of these domains are involved in chromatin binding and therefore aid in the epigenetic regulation performed by these enzymes. Second, JmjC demethylases are less prone to dimerization than the JmjC hydroxylases, such as FIH and TYW5, which contain a dimerization domain and require dimerization to catalytically function.[143,144] Interestingly, biochemical analyses suggest that KDM3A requires a homodimer to be catalytically active (reviewed in Chapter 3, Section 3.1).[145] Lastly, JmjC demethylases structurally differ from JmjC hydroxylases in their substrate binding site. To accommodate N^{ε}-methyllysine, JmjC demethylases tend to have a deeper substrate binding site than hydroxylases.[140] The demethylases also have an extended flexible loop at the *N*-terminus of their DSBH which is involved in methylated substrate binding and is not normally present in hydroxylases.[146] The interaction of the demethylases with their substrates is typically more hydrophobic than the interaction of hydroxylases with their substrates.[147]

To date, several JmjC hydroxylases have been reported to catalyse both demethylation and hydroxylation, although these assignments require validation.[135,148–152] The results

presented in this thesis involve a JmjC demethylase that can also catalyse hydroxylation (see Chapter 7).

Demethylation by JmjC demethylases is thought to occur in two stages: first, hydroxylation of the *N*ε-methyl groups occurs via the conserved 2OG oxygenases radical-based mechanism, which results in a hydroxymethyl hemiaminal intermediate (see Section 1.12.1). Second, the unstable hemiaminal spontaneously fragments to release formaldehyde and the demethylated product (**Figure 1.8**).[153] Unlike the LSDs, the intermediate for JmjC enzymes is hemiaminal, where two methyl groups or a protons can bond with the nitrogen, therefore the JmjC KDMs can demethylate tri-methylated lysines in addition to mono- and di-methylated lysines.

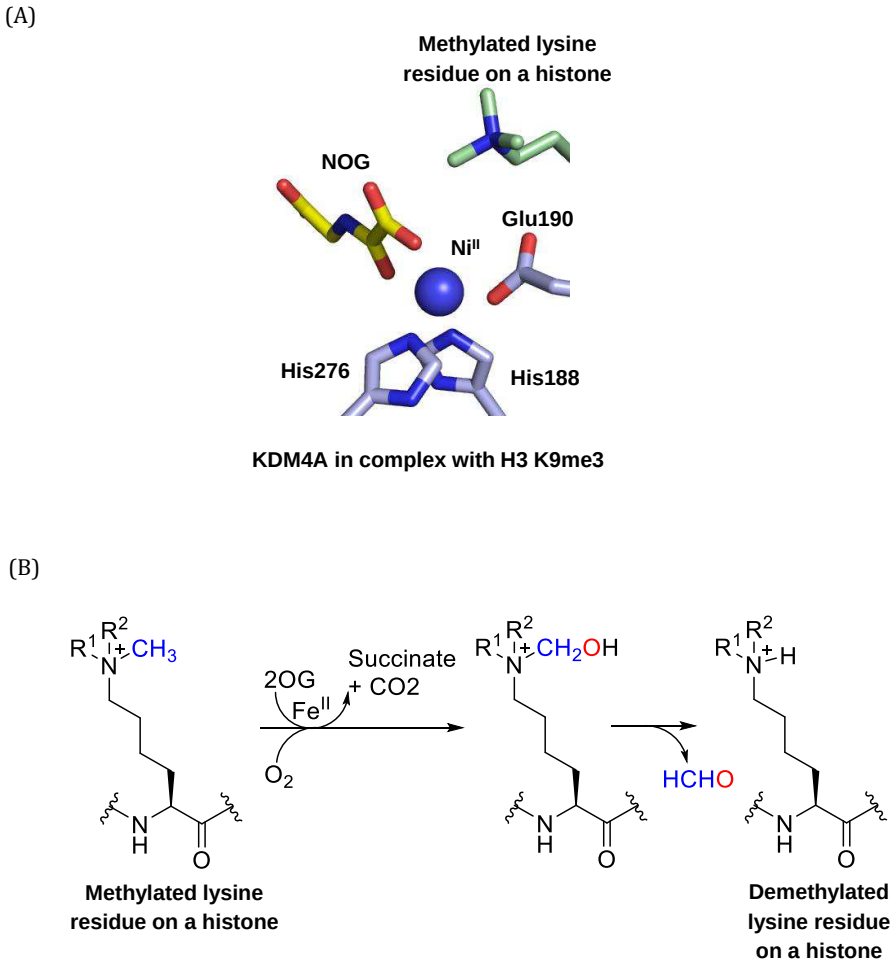


Figure 1.8 **JmjC histone lysine demethylation**. (A) View from the KDM4A crystal structure shows the orientation of the co-substrate mimic *N*-oxalylglycine (NOG, yellow) and the H3 K9me3 substrate peptide (green) with KDM4A active site (grey), Ni^{II} (blue) replaces Fe^{II} in the crystal structure (PDB ID 2OQ6).[153,154] (B) Proposed catalytic mechanism for the JmjC KDMs. The mechanism for demethylation is proposed to occur in two stages. First, hydroxylation of the N^{ε}-methyl group of the methylated lysyl-reside on a histone tail occurs via the conserved 2OG oxygenases radical-based mechanism. The resultant hydroxymethyl hemiaminal intermediate is unstable and the second step involves its spontaneous fragmentation to release formaldehyde and form the demethylated product. Oxygen atoms originating from molecular oxygen are labelled in red, removed methyl group is labelled in blue. Figure was adapted with permission from Schiller *et al.*[155]

The JmjC demethylases family of the 2OG-dependent oxygenase superfamily are the largest known family of KDMs, with members that can remove most of the reported key methyl marks on histone tails.[156] Phylogenetic analyses divide the human JmjC family members into 6 subfamilies (KDM2-7) that tend to share common substrates (**Figure 1.9** and **Table 1.2**).[131] With a highly conserved catalytic domain, sequence and methylation state selectivity of JmjC KDMs is regulated at least three main levels: (i) the cellular protein

complex in which the JmjC KDM resides. (ii) the 'helper' non-catalytic domains which mediate localisation to the target histone sequence, and (iii) subtle variations in the topology of the substrate binding site, which recognises the target methylated lysine residue and additional substrate residues that bind along the enzyme binding site surface.[157]

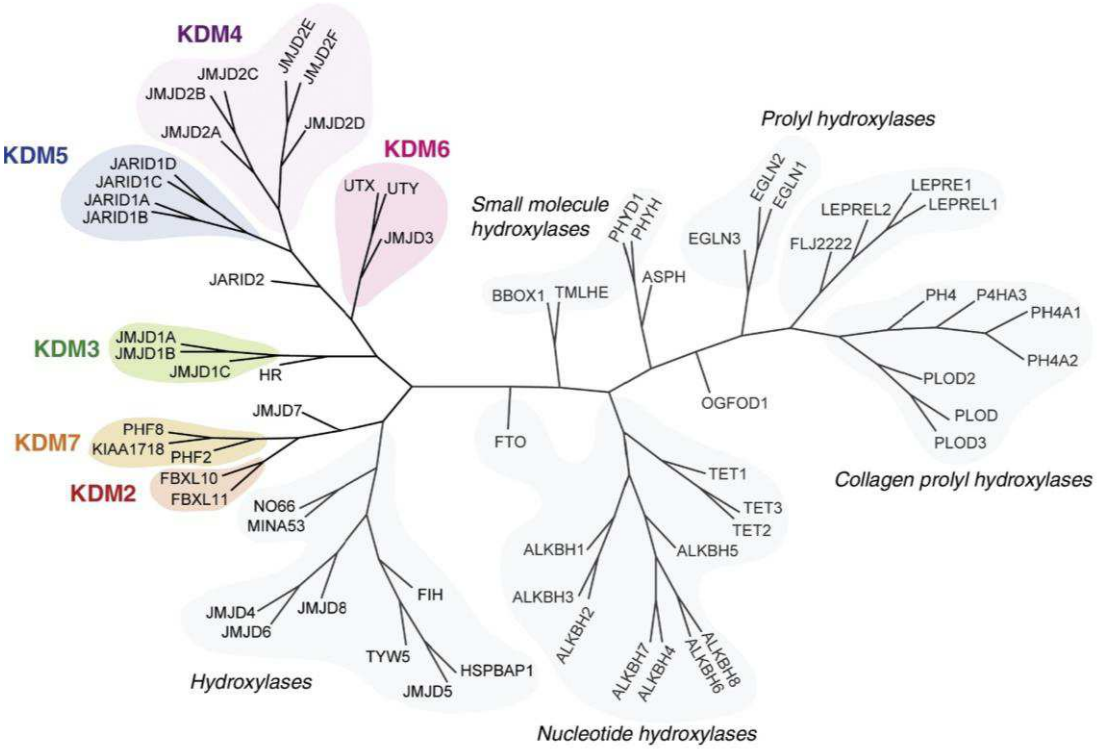


Figure 1.9 Phylogenetic tree of the 2-oxoglutarate oxygenase superfamily with various identified subfamilies (KDM2–7) of histone demethylases. Figure adapted with permission.[158]

Table 1.2 Summary of the subfamily division, reported substrates and domain architecture for human JmjC demethylases. Abbreviations (alphabetical): ARID, AT-rich interaction domain; FBXL, F-box and leucine-rich repeat protein; GATAL, GATA-like domain; JARID, Jumonji/AT-rich interaction domain-containing protein; JmjC, Jumonji-C; JMJD, Jumonji domain-containing; JmjN, Jumonji-N; KDM, lysine-specific demethylase; LRR, leucine rich repeat; MDC1, mediator of DNA-damage checkpoint 1; PHD, plant homeobox domain; PHF2/8, plant homeobox domain finger protein 2/8; TPR, tetratricopeptide repeat; UTX/Y, ubiquitously transcribed tetratricopeptide repeat, X/Y chromosome; Zf-C2HC4, zinc finger domain.[140,157,159] Subfamily domain architecture representation is based on the enzyme labelled with *. Controversial assignments are labelled with **.

Subfamily	Enzyme	Demethylation modifications	Histone substrates	Non-histone substrates	Domain architecture
KDM2 (FBXL)	KDM2A*	Kme2/1	H3 K36	Nuclear factor NF-kappa-B K218, K221	
	KDM2B	Kme2/1	H3 K36		
KDM3 (JMJD1)	KDM3A	Kme2/1	H3 K9	p53-K372	
	KDM3B*	Kme2/1	H3 K9		
	JMJD1C	Kme2/1	Unknown**	MDC1 K45	
	Hairless		Unknown**		
KDM4 (JMJD2D)	KDM4A*	Kme3/2/1	H3 K9, H3 K36, H1.4 K26		
	KDM4B	Kme3/2/1	H3 K9, H3 K36, H1.4 K26		
	KDM4C	Kme3/2/1	H3 K9, H3 K36, H1.4 K26	Pc2 K191	
	KDM4D*	Kme3/2/1	H3 K9, H3 K36, H1.4 K26		
	KDM4E	Kme3/2/1	H3 K9		
KDM5 (JARID)	KDM5A*	Kme3/2/1	H3 K4		
	KDM5B				
	KDM5C				
	KDM5D				
	KDM5E				
KDM6	KDM6A (UTX)*	Kme3/2/1	H3 K27		
	KDM6B (JMJD3)				
	KDM6C (UTY)				
KDM7 (KIAA1718)	KDM7A	Kme3/2/1	H3 K9, H3 K27		
	KDM7B* (PHF8)	Kme3/2/1	H3 K9, H4 K20, H3 K27		
	KDM7C (PHF2)	Kme3/2/1	H3 K9	ARID5B K36	

JmjN JmjC Zf-C2HC4 GATAL PHD Tudor ARID Fbox+LRR TPR

1.14 The KDM3 subfamily

The KDM3 (also known as JMJD1) subfamily of JmjC KDMs is evolutionarily conserved, in mammals this subfamily has four members. KDM3A and KDM3B are H3 K9me1/me2 demethylases while the putative enzymatic activity of JMJD1C and Hairless remains controversal.[156]

All KDM3 members have the same domain structure, which comprises of a CH2C4 zinc finger domain followed by the Leu-Xaa-Xaa-Leu-Leu (LXXLL) motif, a binding sequence common in transcriptional regulation, and a JmjC domain at the *C*-terminal region (**Figure 1.10**).[160]

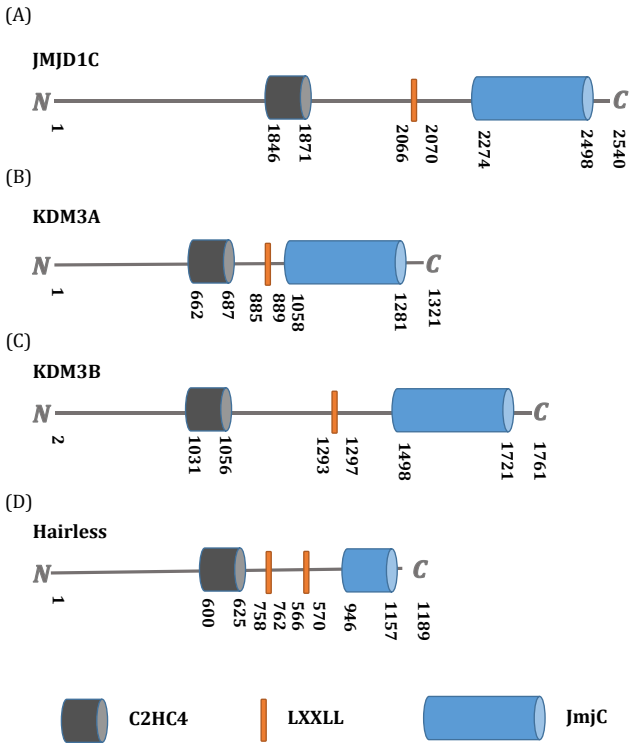


Figure 1.10 Comparison of domains and motifs between JMJD1C and other members of the human KDM3 subfamily. C2HC4 zinc finger domain (grey) followed by the Leu-Xaa-Xaa-Leu-Leu (LXXLL) motif (orange) and the enzymatic JmjC domain at the *C*-terminal region (blue). Amino acid residue positions refer to that of human KDM3s. (A) JMJD1C (UniProtKB - Q15652). (B) KDM3B (UniProtKB – Q7LBC6). (C) KDM3A (UniProtKB - Q9Y4C1). (D) Hairless (UniProtKB – O43593).

** Note regarding formatting of genes and proteins names: gene symbols are italicized, regardless of the type of organism. Protein symbols are not italicized and all letters are in*

upper-case (e.g., GFAP*), regardless of the type of organism. Mice and rats: for gene symbols, only the first letter in upper-case (e.g., Gfap).*[161]

Pairwise JmjC domain comparisons indicate that KDM3A and KDM3B are the most similar (83% protein sequence identity), and both contain the Fe^{II}-binding motif (His-X-Asp/Glu-X_N-His) and the 2OG-binding residues (**Figure 1.11**).

(A)

(B)

	Hairless	JMJD1C	KDM3A	KDM3B
Hairless		38%	35%	36%
JMJD1C			56%	58%
KDM3A				83%

Figure 1.11 JmjC domain comparisons for human KDM3 subfamily members. (A) Sequence alignment of JmjC domains of KDM3 subfamily members. The Fe^{II}-binding motif - His-X-Asp/Glu-X_N-His (blue) and the 2OG-binding residues (green) are conserved amongst KDM3 subfamily members, with the exception of Hairless and JMJD1C.[110,162] (B) Percentage identity between JmjC domains of KDM3 subfamily members based on sequence alignment. UniProtKB accession numbers for amino acids sequences: Hairless - O43593, JMJD1C Q15652, KDM3A - Q9Y4C1 and KDM3B - Q7LBC6. Sequence alignment and percentage identity were analysed using Clustal Omega.[163]

1.14.1 KDM3A

Mammalian KDM3A (also known as JMJD1A, Jhdm2a and TSGA) is the best characterized paralog of the KDM3 subfamily. KDM3A was identified in 2006 by Yamane *et al.,* and reported to be a selective H3 K9me1/me2 demethylase.[164] The 'corresponding' histone methyltransferase G9a/GLP is proposed to catalyse methylation of H3 K27 in addition to that of the H3 K9 position, though reports are controversial.[165–167] On the other hand, KDM3A does not catalyse demethylation in the H3 K27 position, and currently H3 K9 is the only reported KDM3A substrate. KDM3A is also selective for the methylation state of the substrate, it accepts dimethyllysine and monomethyllysine residues as substrates but does not accept trimethyllysine residues.[162]

Although there is not yet a reported crystal structure for KDM3A, a number of epigenetic pathways regulated by KDM3A have been reported. Initial studies by Yamane *et al*. in 2006, found KDM3A to physically interact with the androgen receptor (AR) in a manner that facilitates transcriptional activation of AR target genes via demethylating transcriptionally repressive marks H3 K9me1/me2.[164] A year later, in 2007, KDM3A has been reported to be involved in the epigenetic regulation of self-renewal in embryonic stem (ES) cells in mice. KDM3A was found to upregulate the expression of genes associated with pluripotency by maintaining their H3 K9 in a demethylated state.[168] However, this report shows that *Kdm3a* depletion caused differentiation of ES cells in culture, and this is contradictory to other reports showing that *Kdm3a* knockout mice are viable.[169–172] Since knockout mice formation requires self-renewing ES cells, the observed differentiation could be a result of RNAi off-target effects.

In the same year (2007), Okada and coworkers have reported that *Kdm3a knockout* mice displayed male infertility. KDM3A was first found to be essential for the packaging of sperm chromatin in mice during spermatogenesis through upregulating the expression of associated genes and maintaining their H3 K9 in a demethylated state.[173,174] This phenotype was not unexpected due to the high levels of *Kdm3a* in rat testis observed by Höög *et al* (1991).[175] More recently, in 2013, *Kdm3a* was found to be essential for sex determination: mice lacking *Kdm3a* were observed to undergo a male-to-female sex reversal.[169] KDM3A was observed to demethylate the transcriptionally repressive marks H3 K9me1/me2 of the Y chromosome *Sry* gene, which results in upregulation of *Sry* expression. Interestingly, ChIP analyses in this study have repeatedly found coordinated H3 K4 methylation / H3 K9 demethylation in KDM3A target genes. KDM3A deficiency perturbed H3 K4 methylation on the locus of KDM3A target genes. These results indicate a possible cross-talk between these

two PTMs and suggest that H3 K9 demethylation might be required for subsequent H3 K4 methylation, which is a hallmark for gene activation (see Section 1.8). In 2014, KDM3A was also found to be essential for the proper development of the cytoplasmic structures required to elongate the sperm head in mice.[170] Remarkably in this study, KDM3A was found to be localised in the cytoplasm of maturing spermatids during spermatogenesis. KDM3A was also recently found to exit the nucleus in carcinoma cells exposed to normal extracellular matrix (ECM) resulting in growth restriction of these cells.[176] These two reports of cytoplasmic localised KDM3A suggest that a biological function using non-histone substrates might exist, in addition to the known role of KDM3A in epigenetic regulation.

In recent years, non-histone and non-catalytic roles for KDM3A have been identified. KDM3A has been reported to bind to the transcription factor Gli1 and protect it from proteosomal degradation[177]. Gli1 is reported to activate the transcription of hedgehog responsive genes, reduced hedgehog signalling is associated with Down syndrome.[178–180] Another reported non-histone and non-catalytic regulation by KDM3A is its interaction with the SWI/SNF nucleosome remodelling complex to regulate acute chromatin dinamics.[181] Interestingly, a KDM3A non-histone substrate has been recently identified. KDM3A is reported to promote breast cancer invasion by demethylation of protein P53 (p53-K372me1) in addition to H3 K9 methylation.[159] KDM3A involvement in breast cancer promotion is further highlighted by modulation of ER target gene expression in the absence of oestrogen in breast cancer cells.[182] In this study KDM3A was observed to be phosphorylated by tyrosine kinase 2 (TNK2). KDM3A has also been found to be serine phosphorylated by the mitogen- and stress-activated protein kinase 1 (MSK1) in heat-shocked cells.[183] In addition to promoting breast cancer, KDM3A is reported to trigger the growth of a number of cancerous cell lines including multiple myeloma, Ewing sarcoma, and ovarian cancer cells.[184–187]

Other studies on the regulation of energy homeostasis and oxygen level by KDM3A have found that KDM3A upregulates the expression of genes associated with fat storage and glucose transport by H3 K9me1/2 demethylation.[171] Loss of KDM3A function was observed to disrupt oxygen consumption in brown fat, resulting in decreased fat oxidation. Mice with *Kdm3a* knockout suffer from obesity.[172]

KDM3A has been reported to induce tumour growth in renal cell and colon carcinoma cell lines by enhancing hypoxic gene expression. Regulation of *KDM3A* expression in these cell lines was shown to be regulated by Hypoxia-Inducible Factor 1α (HIF-1α) transcription factor.[188] KDM3A regulation under hypoxia was later found to be independent of the cell

type.[189] KDM3A was also found to be recruited to the locus of genes that are associated with glycolysis during hypoxia in an HIF1-dependent manner.[189] KDM3A demethylates the H3 K9me2 repressive marks in these loci, and upregulates glycolysis-related gene expression.[189] KDM3A was also reported to induce the growth of bladder cancer cells through activation of the transcription of the oncogene *HOXA1.*[190]

To conclude, the mechanism by which KDM3A is suggested to carry out epigenetic regulation is by demethylation of the repressive marks H3 K9me1/me2, which facilitate the transcriptional upregulation of KDM3A target genes.[164] KDM3A reported target genes are associated with a number of physiological processes including embryonic stem cell self-renewal, spermatogenesis, sex determination, tumour growth promotion and skeletal muscle and brown fat metabolism.[159,168–174,182,184–187] In addition to epigenetic regulation, KDM3A has been recently observed to upregulate gene expression in a non-histone dependent manner via interaction with Gli1 and demethylation of P53.[159,177] Regulation of KDM3A appears to be mediated by different factors such as interactions with HIF-1α, by phosphorylation on at least two sites as well as by factors found in the intra cellular matrix.[176,183,188,189]

1.14.2 KDM3B

KDM3B (also known as JMJD1B, Jhdm2b and 5qNCA) is reported to have the same substrate preference as KDM3A, i.e. catalysing the demethylation of H3 K9me1/2 marks.[162] The physiological functions of KDM3B are less well understood than that of some other KDMs. Although KDM3A has been reported to promote tumor growth, KDM3B has been reported to act either as a tumor suppressor or as an oncogene. Two reports from 2012 have shown that KDM3B is frequently upregulated in prostate cancer cells and in leukemia cells.[191,192] On the other hand, it is reported that KDM3B can act as a tumor suppressor in leukemia, colorectal and breast cancers, as well as being an angiogenesis supressor.[193–196]

Similarly to KDM3A, KDM3B is reported to have a regulatory role in stem cell formation. KDM3B is involved in the induction of somatic cells reprogramming leading to pluripotent stem cells (iPSCs).[197] Another similarity to KDM3A, is that KDM3B is also required for normal mouse reproductive function, KDM3B was found to be essential for fertility in female mice and for normal spermatogenesis in male mice.[198,199]

Interestingly, a recent study has been reported on the association of KDM3B with histone H3 Lys9-to-Met variants (H3 K9M).[200] Histone lysine-to-methionine mutation occur in a subtype of aggressive paediatric brain cancers at position H3 K27M. Since both the H3 K9 and the H3 K27 positions are associated with transcriptionally silenced chromatin, H3 K9M mutation may be biologically relevant.[201,202] In this study, KDM3B was found to be recruited to H3 K9M-containing nucleosomes, where it demethylates wildtype H3 K9 and promotes decompression of the chromatin. This study suggests that recognition of mutations at the H3 K9 position can result in aberrant recruitment, and affect chromatin accessibility and therefore gene expression. KDM3B is likely to recognise other amino acid residues in addition to lysine, positioned instead of the H3 K9, and that such recognition might be biologically relevant in the case of histone mutations in cancers. The study did not test H3 K9M as a substrate of KDM3B.

1.14.3 JMJD1C

The roles of JMJD1C (also known as KDM3C and Thyroid Receptor-Interacting Protein 8, TRIP8) are even less defined than those of KDM3A and KDM3B. The reported H3 K9me1/me2 demethylation by JMJD1C is controversial and requires validation. A study published in 2010 reported on JMJD1C-catalysed H3 K9me1/me2 demethylation *in vitro* (based on MALDI-TOF MS analyses of histone fragment peptides) and in cells (based on immunofluorescence experiments using over-expressed JMJD1C).[203] However, subsequent studies reported of an absence of demethylation activity as determined based on *Jmjd1c*-deficient-mice, cellular assays and mass spectrometric measurements.[162,204,205] In agreement with these latter studies, a report by Watanabe *et al.* stated that JMJD1C is not a histone demethylase, but instead it catalyses demethylation of Mediator of DNA damage Checkpoint protein 1 (MDC1) at Lys45, thereby promoting response to DNA double-strand breaks.[206,207] This demethylase activity was also observed in cells using proteomic analysis; however, no direct *in vitro* validation using isolated JMJD1C and MDC1 fragment peptides was performed. MDC1 demethylation catalysis by JMJD1C was also highlighted in another study, patients with Rett syndrome and autism were shown to harbor critical mutations in the *N*-terminal region of JMJD1C and have reduced MDC1 demethylation.[208] *JMJD1C* was observed to be highly expressed in the brain, seven critical mutations were identified in these patients and most are located in the *N*-terminal region of JMJD1C. However, the laboratory that published this study is under investigation for scientific fraud.[209]

Like other KDM3 members, JMJD1C inhibits stem cell differentiation. JMJD1C has been reported to block neural differentiation, providing an additional indication for its physiological function in the brain.[210] An additional similarity of JMJD1C to its subfamily members is a requirement for its expression during spermatogenesis.[204] A recent study supports this involvement and suggests that JMJD1C may participate in the maintenance of spermatogonial stem cell self-renewal by up-regulating the expression of Oct4 transcription factor, a marker for undifferentiated cells.[211] Like KDM3A, JMJD1C is reported to interact with the androgen receptor and promote its transcription.[195,212] Similar to KDM3B, JMJD1C is reported to act as an oncogene as well as a tumour suppressor: JMJD1C is reported to be required for leukemia maintenance in mice and human cells, and has been suggested as a putative tumor suppressor of breast cancer based on qRT-PCR of healthy and diseased human tissues.[205,212,213]

1.14.4 Hairless

Hairless is the fourth member of the KDM3 subfamily. Hairless was first identified in 1994 and is mainly expressed in the brain and skin, where it was found to be involved in hereditary hair loss (alopecias) in mice and humans.[214,215]

Similarly to that of JMJD1C, demethylation by Hairless is controversial and requires validation. Hairless was described as an H3 K9me1/me2 demethylase based on results from *in vitro* demethylation assays and immunofluorescence staining.[216] Analyses of data from both methods were dependent on the selectivity of anti-H3 K9me1 and anti-H3 K9me2 antibodies. However, loss of conserved Fe^{II}-binding site and 2OG-binding site indicate that Hairless is likely to be enzymatically inactive (see Section 1.14).[110]

Hairless has been reported to act as a co-repressor for a number of nuclear receptors, including the thyroid hormone receptor, the retinoic acid receptor-related orphan receptors and the vitamin D receptor, which play a critical role in the maintenance of hair growth.[217–224]

Like other KDM3 members, Hairless has been reported to be involved in stem cell differentiation. Hairless was observed to control the timing of Wnt signalling required for hair cycling.[225,226] Similarly to KDM3A, Hairless was observed to regulate energy storage as fat and maintenance of energy homeostasis.[227]

1.15 Objectives

The introduction presented in this chapter aimed to summarise the current knowledge on epigenetic regulation with a focus on the KDM3 subfamily of JmjC histone lysine demethylases. The work reported in the subsequent chapters of this thesis describes investigations on the inhibitor and substrate diversity of the JmjC histone demethylases, with the aim of identifying novel inhibitors and substrates.

The specific aims of the research described in this thesis were to-

1) Explore inhibitor structures of JmjC KDMs, and improve the cellular permeability of the broad-spectrum inhibitor IOX1 (Chapter 2),
2) Biochemically and kinetically characterise KDM3A, including the characterisation of a KDM3A selective IOX1 derivative inhibitor (Chapter 3),
3) Evaluate TCA cycle intermediates as inhibitors for KDM3A (Chapter 4),
4) Analyse KDM3A putative interaction with intraflagellar transport proteins as a potential role of KDM3A in non-epigenetic regulation (Chapter 5),
5) Investigate novel substrates for KDM3 (KDM3A and JMJD1C) mediated catalysis, including: methylated arginines, lysine analogues, acetylated and formylated lysines (Chapters 6-9).

1.16 References

1. Holliday, R. Epigenetics: a historical overview. *Epigenetics* **1,** 76–80 (2006).

2. Waddington, C. H. & others. An introduction to modern genetics. *An Introd. to Mod. Genet.* (1939).

3. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: A Landscape Takes Shape. *Cell* **128,** 635–638 (2007).

4. Epigenetics 101: a beginner's guide to explaining everything | Cath Ennis | Science | The Guardian. Available at: https://www.theguardian.com/science/occams-corner/2014/apr/25/epigenetics-beginners-guide-to-everything.

5. Jones, P. A. & Takai, D. The Role of DNA Methylation in Mammalian Epigenetics. *Science (80-.).* **293,** (2001).

6. Lee, J. T. Epigenetic Regulation by Long Noncoding RNAs. *Science (80-.).* **338,** (2012).

7. Cosgrove, M. S. & Wolberger, C. How does the histone code work? *Biochem. Cell Biol.* **83,** 468–476 (2005).

8. Crick, F. H. C. ON PROTEIN SYNTHESIS. (1956).

9. Egger, G., Liang, G., Aparicio, A. & Jones, P. A. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429,** 457–463 (2004).

10. Cavalli, G. Chromatin and epigenetics in development: blending cellular memory with cell fate plasticity. *Development* **133,** (2006).

11. Kamakaka, R. T. & Biggins, S. Histone variants: deviants? *Genes Dev.* **19,** 295–310 (2005).

12. Butler, P. J. G. & Crothers, D. M. The Folding of Chromatin. *Crit. Rev. Biochem.* **15,** 57–91 (1983).

13. Richmond, T. J., Luger, K., Mäder, A. W., Richmond, R. K. & Sargent, D. F. Crystal structure of the nucleosome core particle at 2.8|[thinsp]||[Aring]| resolution. *Nature* **389,** 251–260 (1997).

14. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution. *J. Mol. Biol.* **319,** 1097–1113 (2002).

15. Greulich, K. O., Wachtel, E., Ausio, J., Seger, D. & Eisenberg, H. Transition of chromatin from the "10 nm" lower order structure, to the "30 nm" higher order structure as followed by small angle X-ray scattering. *J. Mol. Biol.* **193,** 709–21 (1987).

16. Pierce, B. A. & Preceded by: Pierce, B. A. *Genetics : a conceptual approach*.

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

18. Maeshima, K., Imai, R., Tamura, S. & Nozaki, T. Chromatin as dynamic 10-nm fibers.

Chromosoma **123,** 225–237 (2014).

19. Berezney, R. & Jeon, K. W. *International review of cytology : a survey of cell biology. Volume 162A, Structural and functional organization of the nuclear matrix*. (Academic Press, 1995).

20. Annunziato, A. DNA Packaging: Nucleosomes and Chromatin. *Nat. Educ.* **1,** 26 (2008).

21. Dame, R. T. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.* **56,** 858–870 (2005).

22. Tsukiyama, T. The in vivo functions of atp-dependent chromatin-remodelling factors. *Nat. Rev. Mol. Cell Biol.* **3,** 422–429 (2002).

23. Längst, G. & Becker, P. B. Nucleosome remodeling: one mechanism, many phenomena? *Biochim. Biophys. Acta - Gene Struct. Expr.* **1677,** 58–63 (2004).

24. Tang, L., Nogales, E. & Ciferri, C. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. *Prog. Biophys. Mol. Biol.* **102,** 122–128 (2010).

25. Erdel, F. & Rippe, K. Chromatin remodelling in mammalian cells by ISWI-type complexes - where, when and why? *FEBS J.* **278,** 3608–3618 (2011).

26. Bao, Y. INO80 subfamily of chromatin remodeling complexes. *Mutat. Res. Mol. Mech. Mutagen.* **618,** 18–29 (2007).

27. Marfella, C. G. A. & Imbalzano, A. N. The Chd family of chromatin remodelers. *Mutat. Res. Mol. Mech. Mutagen.* **618,** 30–40 (2007).

28. Zhang, Y. *et al.* DNA Translocation and Loop Formation Mechanism of Chromatin Remodeling by SWI/SNF and RSC. *Mol. Cell* **24,** 559–568 (2006).

29. Strohner, R. *et al.* A 'loop recapture' mechanism for ACF-dependent nucleosome remodeling. *Nat. Struct. Mol. Biol.* **12,** 683–690 (2005).

30. Arents, G. & Moudrianakis, E. N. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 11170–4 (1995).

31. Talbert, P. B. & Henikoff, S. Histone variants--ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* **11,** 264–75 (2010).

32. Ahmad, K. & Henikoff, S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* **9,** 1191–200 (2002).

33. Fletcher, T. M. & Hansen, J. C. The Nucleosomal Array: Structure/Function Relationships. *Crit. Rev. Eukaryot. Gene Expr.* **6,** 149–188 (1996).

34. Azevedo, C. & Saiardi, A. Why always lysine? The ongoing tale of one of the most modified amino acids. *Adv. Biol. Regul.* **60,** 144–50 (2016).

35. Huang, H., Sabari, B. R., Garcia, B. A., Allis, C. D. & Zhao, Y. SnapShot: histone modifications. *Cell* **159,** 458–458.e1 (2014).

36. Huang, H. *et al.* SnapShot: Histone Modifi cations. *Cell* **159,** 458–458.e1 (2014).

37. Huang, H., Lin, S., Garcia, B. A. & Zhao, Y. Quantitative proteomic analysis of histone modifications. *Chem. Rev.* **115,** 2376–418 (2015).

38. Tan, M. *et al.* Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* **146,** 1016–1028 (2011).

39. Dai, L. *et al.* Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nat. Chem. Biol.* **10,** 365–70 (2014).

40. Jiang, T., Zhou, X., Taghizadeh, K., Dong, M. & Dedon, P. C. N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 60–5 (2007).

41. Wisniewski, J. R., Zougman, A., Krüger, S. & Mann, M. Mass spectrmetric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. *Mol. Cell. Proteomics* **6,** 72–87 (2007).

42. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21,** 381–95 (2011).

43. Stadler, S. C. & Allis, C. D. Linking epithelial-to-mesenchymal-transition and epigenetic modifications. *Semin. Cancer Biol.* **22,** 404–10 (2012).

44. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403,** 41–5 (2000).

45. Gardner, K. E., Allis, C. D. & Strahl, B. D. OPERating ON Chromatin, a Colorful Language where Context Matters. *J. Mol. Biol.* **409,** 36–46 (2011).

46. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293,** 1074–80 (2001).

47. Lee, J.-S., Smith, E. & Shilatifard, A. The language of histone crosstalk. *Cell* **142,** 682–5 (2010).

48. Gillette, T. G. & Hill, J. A. Readers, writers, and erasers: chromatin as the whiteboard of heart disease. *Circ. Res.* **116,** 1245–53 (2015).

49. Ruthenburg, A. J., Li, H., Patel, D. J. & Allis, C. D. Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* **8,** 983–94 (2007).

50. Hebbes, T. R., Clayton, A. L., Thorne, A. W. & Crane-Robinson, C. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO J.* **13,** 1823–30 (1994).

51. Krajewski, W. A. & Becker, P. B. Reconstitution of hyperacetylated, DNase I-sensitive chromatin characterized by high conformational flexibility of nucleosomal DNA. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 1540–5 (1998).

52. Suka, N., Suka, Y., Carmen, A. A., Wu, J. & Grunstein, M. Highly Specific Antibodies Determine Histone Acetylation Site Usage in Yeast Heterochromatin and Euchromatin. *Mol. Cell* **8,** 473–479 (2001).

53. Filippakopoulos, P. & Knapp, S. Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat. Rev. Drug Discov.* **13,** 337–356 (2014).

54. Filippakopoulos, P. *et al.* Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **149,** 214–231 (2012).

55. Vollmuth, F., Blankenfeldt, W. & Geyer, M. Structures of the dual bromodomains of the P-TEFb-activating protein Brd4 at atomic resolution. *J. Biol. Chem.* **284,** 36547–56 (2009).

56. Lange, M. *et al.* Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes Dev.* **22,** 2370–84 (2008).

57. Zeng, L. *et al.* Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* **466,** 258–262 (2010).

58. Musselman, C. A., Lalonde, M.-E., Côté, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* **19,** 1218–1227 (2012).

59. Garske, A. L. *et al.* Combinatorial profiling of chromatin binding modules reveals multisite discrimination. *Nat. Chem. Biol.* **6,** 283–290 (2010).

60. Su, D. *et al.* Structural basis for recognition of H3K56-acetylated histone H3–H4 by the chaperone Rtt106. *Nature* **483,** 104–107 (2012).

61. Hasan, S. & Hottiger, M. O. Histone acetyl transferases: a role in DNA repair and DNA replication. *J. Mol. Med.* **80,** 463–474 (2002).

62. Lin, H., Su, X. & He, B. Protein Lysine Acylation and Cysteine Succination by Intermediates of Energy Metabolism. *ACS Chem. Biol.* **7,** (2012).

63. Roth, S. Y., Denu, J. M. & Allis, C. D. Histone Acetyltransferases. *Annu. Rev. Biochem.* **70,** 81–120 (2001).

64. Marmorstein, R. & Trievel, R. C. Histone modifying enzymes: Structures, mechanisms, and specificities. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1789,** 58–68 (2009).

65. Yuan, H. & Marmorstein, R. Histone acetyltransferases: Rising ancient counterparts to protein kinases. *Biopolymers* **99,** 98–111 (2013).

66. Seto, E. & Yoshida, M. Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harb. Perspect. Biol.* **6,** a018713 (2014).

67. de Ruijter, A. J. M., van Gennip, A. H., Caron, H. N., Kemp, S. & van Kuilenburg, A. B. P. Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem. J.* **370,** 737–49 (2003).

68. Longworth, M. S. & Laimins, L. A. Histone deacetylase 3 localizes to the plasma membrane and is a substrate of Src. *Oncogene* **25,** 4495–4500 (2006).

69. Valenzuela-Fernández, A. *et al.* HDAC6: a key regulator of cytoskeleton, cell migration and cell-cell interactions. *Trends Cell Biol.* **18,** 291–7 (2008).

70. Guarente, L., Imai, S., Armstrong, C. M. & Kaeberlein, M. Transcriptional silencing and longevity protein Sir2 is an NAD-dependenthistone deacetylase. *Nature* **403,** 795–800 (2000).

71. Blander, G. & Guarente, L. The Sir2 Family of Protein Deacetylases. *Annu. Rev.*

Biochem. **73,** 417–435 (2004).

72. Smith, B. C. & Denu, J. M. Chemical mechanisms of histone lysine and arginine modifications. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1789,** 45–57 (2009).

73. Sauve, A. A. Sirtuin chemical mechanisms. *Biochim. Biophys. Acta* **1804,** 1591–603 (2010).

74. Walport, L. J., Hopkinson, R. J. & Schofield, C. J. Mechanisms of human histone and nucleic acid demethylases. *Curr. Opin. Chem. Biol.* **16,** 525–34 (2012).

75. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–37 (2007).

76. Feng, Q. *et al. Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases without a SET Domain*. *Current Biology* **12,** (2002).

77. Lachner, M., O'Sullivan, R. J. & Jenuwein, T. An epigenetic road map for histone lysine methylation. *J. Cell Sci.* **116,** (2003).

78. Hung, T. *et al. ING4 Mediates Crosstalk between Histone H3 K4 Trimethylation and H3 Acetylation to Attenuate Cellular Transformation*. *Molecular Cell* **33,** (2009).

79. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442,** 86–90 (2006).

80. Wang, G. G., Allis, C. D. & Chi, P. Chromatin remodeling and cancer, part I: covalent histone modifications. *Trends Mol. Med.* **13,** 363–372 (2007).

81. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).

82. Taverna, S. D. *et al.* Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. *Mol. Cell* **24,** 785–96 (2006).

83. Peña, P. V *et al.* Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442,** 100–3 (2006).

84. Lee, B. M. & Mahadevan, L. C. Stability of histone modifications across mammalian genomes: Implications for â€˜epigeneticâ€™ marking. *J. Cell. Biochem.* **108,** 22–34 (2009).

85. Stewart, M. D., Li, J. & Wong, J. Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment. *Mol. Cell. Biol.* **25,** 2525–38 (2005).

86. Pascual, J., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. Structure of the PHD Zinc Finger from Human Williams-Beuren Syndrome Transcription Factor. *J. Mol. Biol.* **304,** 723–729 (2000).

87. Sanchez, R. & Zhou, M.-M. The PHD finger: a versatile epigenome reader. *Trends Biochem. Sci.* **36,** 364–372 (2011).

88. Feng, W., Yonezawa, M., Ye, J., Jenuwein, T. & Grummt, I. PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation. *Nat.*

Struct. Mol. Biol. **17,** 445–50 (2010).

89. Iwase, S. *et al.* The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell* **128,** 1077–88 (2007).

90. Lan, F. *et al.* Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448,** 718–22 (2007).

91. Townsley, F. M., Thompson, B. & Bienz, M. Pygopus residues required for its binding to Legless are critical for transcription and development. *J. Biol. Chem.* **279,** 5177–83 (2004).

92. Eissenberg, J. C. Molecular biology of the chromo domain: an ancient chromatin module comes of age. *Gene* **275,** 19–29 (2001).

93. Jones, D. O., Cowell, I. G. & Singh, P. B. Mammalian chromodomain proteins: their role in genome organisation and expression. *BioEssays* **22,** 124–137 (2000).

94. Min, J., Zhang, Y. & Xu, R.-M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **17,** 1823–8 (2003).

95. Adams-Cioaba, M. A. & Min, J. Structure and function of histone methylation binding proteins. *Biochem. Cell Biol.* **87,** 93–105 (2009).

96. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406,** 593–9 (2000).

97. Min, J., Feng, Q., Li, Z., Zhang, Y. & Xu, R.-M. Structure of the Catalytic Domain of Human DOT1L, a Non-SET Domain Nucleosomal Histone Methyltransferase. *Cell* **112,** 711–723 (2003).

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

99. Huang, J. & Berger, S. L. The emerging field of dynamic lysine methylation of non-histone proteins. *Curr. Opin. Genet. Dev.* **18,** 152–158 (2008).

100. Trievel, R. C. *et al.* Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell* **111,** 91–103 (2002).

101. Shi, Y. Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119,** 941–53 (2004).

102. Forneris, F., Binda, C., Vanoni, M. A., Mattevi, A. & Battaglioli, E. *Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process*. *FEBS Letters* **579,** (2005).

103. Huang, J. *et al.* p53 is regulated by the lysine demethylase LSD1. *Nature* **449,** 105–108 (2007).

104. Wang, J. *et al.* The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat. Genet.* **41,** 125–9 (2009).

105. Karytinos, A. *et al.* A novel mammalian flavin-dependent histone demethylase. *J. Biol. Chem.* **284,** 17775–82 (2009).

106. Yang, Y., Yin, X., Yang, H. & Xu, Y. Histone Demethylase LSD2 Acts as an E3 Ubiquitin Ligase and Inhibits Cancer Cell Growth through Promoting Proteasomal Degradation

of OGT. *Mol. Cell* **58,** 47–59 (2015).

107. Schofield, Christopher JHausinger, R. P. *2-Oxoglutarate-Dependent Oxygenases*. (The Royal Society of Chemistry , 2015). doi:10.1039/9781782621959

108. Clifton, I. J. *et al.* Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J. Inorg. Biochem.* **100,** 644–69 (2006).

109. Jaakkola, P. *et al.* Targeting of HIF-α to the von Hippel-Lindau Ubiquitylation Complex by O2-Regulated Prolyl Hydroxylation. *Science (80-.).* **292,** (2001).

110. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* **7,** 715–27 (2006).

111. Aas, P. A. *et al.* Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **421,** 859–63 (2003).

112. Delaney, J. C. & Essigmann, J. M. Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 14051–6 (2004).

113. Vaz, F. M., Ofman, R., Westinga, K., Back, J. W. & Wanders, R. J. Molecular and Biochemical Characterization of Rat epsilon -N-Trimethyllysine Hydroxylase, the First Enzyme of Carnitine Biosynthesis. *J. Biol. Chem.* **276,** 33512–7 (2001).

114. Britsch, L. & Grisebach, H. Purification and characterization of (2S)-flavanone 3-hydroxylase from Petunia hybrida. *Eur. J. Biochem.* **156,** 569–77 (1986).

115. Lange, T. Purification and partial amino-acid sequence of gibberellin 20-oxidase from Cucurbita maxima L. endosperm. *Planta* **195,** 108–15 (1994).

116. Baldwin, J. E., Field, R. A., Lawrence, C. C., Merritt, K. D. & Schofield, C. J. Proline 4-hydroxylase: Stereochemical course of the reaction. *Tetrahedron Lett.* **34,** 7489–7492 (1993).

117. Holme, E., Lindstedt, G., Lindstedt, S. & Tofft, M. 18-O studies of the 2-ketoglutarate-dependent sequential oxygenation of thymine to 5-carboxyuracil. *J. Biol. Chem.* **246,** 3314–9 (1971).

118. Kershaw, N. J. *et al.* The enzymology of clavam and carbapenem biosynthesis. *Chem. Commun.* **3713,** 4251 (2005).

119. Rhoads, R. E. & Udenfriendt, S. Decarboxylation of a-ketoglutarate coupled to collagen proline hydroxylase. *Proc Natl Acad Sci U S A.* **60,** 1473–1478 (1968).

120. Roach, P. L. *et al.* Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **387,** 827–30 (1997).

121. Hegg, E. L. & Jr, L. Q. The 2-His-1-Carboxylate Facial Triad - An Emerging Structural Motif in Mononuclear Non-Heme Iron(II) Enzymes. *Eur. J. Biochem.* **250,** 625–629 (1997).

122. Yang, C.-G. *et al.* Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452,** 961–965 (2008).

123. Hanauske-Abel, H. M. & Günzler, V. A stereochemical concept for the catalytic

mechanism of prolylhydroxylase: applicability to classification and design of inhibitors. *J. Theor. Biol.* **94,** 421–55 (1982).

124. Bollinger, J. M., Price, J. C., Hoffart, L. M., Barr, E. W. & Krebs, C. Mechanism of Taurine: α-Ketoglutarate Dioxygenase (TauD) fromEscherichia coli. *Eur. J. Inorg. Chem.* **2005,** 4245–4254 (2005).

125. Rose, N. R., McDonough, M. A., King, O. N. F., Kawamura, A. & Schofield, C. J. Inhibition of 2-oxoglutarate dependent oxygenases. *Chem. Soc. Rev.* **40,** 4364–97 (2011).

126. Welford, R. W. D. D. *et al.* Incorporation of oxygen into the succinate co-product of iron(II) and 2-oxoglutarate dependent oxygenases from bacteria, plants and humans. *FEBS Lett.* **579,** 5170–4 (2005).

127. Jaakkola, P. *et al.* Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science* **292,** 468–72 (2001).

128. Loenarz, C. *et al.* Evidence for a Stereoelectronic Effect in Human Oxygen Sensing. *Angew. Chemie Int. Ed.* **48,** 1784–1787 (2009).

129. Teperino, R., Schoonjans, K. & Auwerx, J. Histone Methyl Transferases and Demethylases; Can They Link Metabolism and Transcription? *Cell Metabolism* **12,** 321–327 (2010).

130. Takeuchi, T. *et al.* Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes Dev.* **9,** 1211–22 (1995).

131. Zhou, X. *et al.* Evolutionary history of histone demethylase families: distinct evolutionary patterns suggest functional divergence. *BMC Evol. Biol.* **8,** 294 (2008).

132. Aik, W., McDonough, M. A., Thalhammer, A., Chowdhury, R. & Schofield, C. J. Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr. Opin. Struct. Biol.* **22,** 691–700 (2012).

133. McDonough, M. A., Loenarz, C., Chowdhury, R., Clifton, I. J. & Schofield, C. J. Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr. Opin. Struct. Biol.* **20,** 659–672 (2010).

134. Loenarz, C. & Schofield, C. J. Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. Chem. Biol.* **4,** 152–6 (2008).

135. Ge, W. *et al.* Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat. Chem. Biol.* **8,** 960–2 (2012).

136. Noma, A. *et al.* Expanding role of the jumonji C domain as an RNA hydroxylase. *J. Biol. Chem.* **285,** 34503–7 (2010).

137. Elkins, J. M. *et al.* Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J. Biol. Chem.* **278,** 1802–6 (2003).

138. Webby, C. J. *et al.* Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science* **325,** 90–3 (2009).

139. Hopkinson, R. J. *et al.* Is JmjC oxygenase catalysis limited to demethylation? *Angew. Chem. Int. Ed. Engl.* **52,** 7709–13 (2013).

140. Markolovic, S. *et al.* Structure-function relationships of human JmjC oxygenases-demethylases versus hydroxylases. *Curr. Opin. Struct. Biol.* **41,** 62–72 (2016).

141. Lu, T. *et al.* Regulation of NF-kappaB by NSD1/FBXL11-dependent reversible lysine methylation of p65. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 46–51 (2010).

142. Ponnaluri, V. K. C., Vavilala, D. T., Putty, S., Gutheil, W. G. & Mukherji, M. Identification of non-histone substrates for JMJD2A-C histone demethylases. *Biochem. Biophys. Res. Commun.* **390,** 280–4 (2009).

143. Lancaster, D. E. *et al.* Disruption of dimerization and substrate phosphorylation inhibit factor inhibiting hypoxia-inducible factor (FIH) activity. *Biochem. J.* **383,** (2004).

144. Kato, M. *et al.* Crystal structure of a novel JmjC-domain-containing protein, TYW5, involved in tRNA modification. *Nucleic Acids Res.* **39,** 1576–85 (2011).

145. Goda, S., Isagawa, T., Chikaoka, Y., Kawamura, T. & Aburatani, H. Control of histone H3 lysine 9 (H3K9) methylation state via cooperative two-step demethylation by Jumonji domain containing 1A (JMJD1A) homodimer. *J. Biol. Chem.* **288,** 36948–56 (2013).

146. Cheng, Z. *et al.* A molecular threading mechanism underlies Jumonji lysine demethylase KDM2A regulation of methylated H3K36. *Genes Dev.* **28,** 1758–71 (2014).

147. Horton, J. R. *et al.* Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat. Struct. Mol. Biol.* **17,** 38–43 (2010).

148. Böttger, A., Islam, M. S., Chowdhury, R., Schofield, C. J. & Wolf, A. The oxygenase Jmjd6--a case study in conflicting assignments. *Biochem. J.* **468,** 191–202 (2015).

149. Lu, Y. *et al.* Lung cancer-associated JmjC domain protein mdig suppresses formation of tri-methyl lysine 9 of histone H3. *Cell Cycle* **8,** 2101–2109 (2009).

150. Sinha, K. M. *et al.* Regulation of the osteoblast-specific transcription factor Osterix by NO66, a Jumonji family histone demethylase. *EMBO J.* **29,** 68–79 (2010).

151. Hsia, D. A. *et al.* KDM8, a H3K36me2 histone demethylase that acts in the cyclin A1 coding region to regulate cancer cell proliferation. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 9671–6 (2010).

152. Youn, M.-Y. *et al.* JMJD5, a Jumonji C (JmjC) domain-containing protein, negatively regulates osteoclastogenesis by facilitating NFATc1 protein degradation. *J. Biol. Chem.* **287,** 12994–3004 (2012).

153. Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439,** 811–816 (2006).

154. Ng, S. S. *et al.* Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **448,** 87–91 (2007).

155. Schiller, R. *et al.* A cell-permeable ester derivative of the JmjC histone demethylase inhibitor IOX1. *ChemMedChem* **9,** 566–71 (2014).

156. Kooistra, S. M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. *Nat. Rev. Mol. Cell Biol.* **13,** 297–311 (2012).

157. Pilka, E. S., James, T. & Lisztwan, J. H. Structural definitions of Jumonji family demethylase selectivity. *Drug Discov. Today* **20,** 743–749 (2015).

158. Nowak, R. P. *et al.* Advances and challenges in understanding histone demethylase biology. *Curr. Opin. Chem. Biol.* **33,** 151–159 (2016).

159. Ramadoss, S., Guo, G. & Wang, C.-Y. Lysine demethylase KDM3A regulates breast cancer cell invasion and apoptosis by targeting histone and the non-histone protein p53. *Oncogene* (2016). doi:10.1038/onc.2016.174

160. Plevin, M. J., Mills, M. M. & Ikura, M. The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends Biochem. Sci.* **30,** 66–9 (2005).

161. Guidelines for Formatting Gene and Protein Names. Available at: http://www.biosciencewriters.com/Guidelines-for-Formatting-Gene-and-Protein-Names.aspx.

162. Brauchle, M. *et al.* Protein complex interactor analysis and differential activity of KDM3 subfamily members towards H3K9 methylation. *PLoS One* **8,** e60549 (2013).

163. Li, W. *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **43,** W580-4 (2015).

164. Yamane, K. *et al.* JHDM2A, a JmjC-Containing H3K9 Demethylase, Facilitates Transcription Activation by Androgen Receptor. *Cell* **125,** 483–495 (2006).

165. Tachibana, M. *et al.* Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev.* **19,** 815–26 (2005).

166. Tachibana, M. *et al.* G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* **16,** 1779–91 (2002).

167. Maze, I. *et al.* Essential Role of the Histone Methyltransferase G9a in Cocaine-Induced Plasticity. *Science (80-.).* **327,** (2010).

168. Loh, Y.-H., Zhang, W., Chen, X., George, J. & Ng, H.-H. Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev.* **21,** 2545–57 (2007).

169. Kuroki, S. *et al.* Epigenetic regulation of mouse sex determination by the histone demethylase Jmjd1a. *Science* **341,** 1106–9 (2013).

170. Kasioulis, I. *et al.* Kdm3a lysine demethylase is an Hsp90 client required for cytoskeletal rearrangements during spermatogenesis. *Mol. Biol. Cell* **25,** 1216–1233 (2014).

171. Tateishi, K., Okada, Y., Kallin, E. M. & Zhang, Y. Role of Jhdm2a in regulating metabolic gene expression and obesity resistance. *Nature* **458,** 757–761 (2009).

172. Inagaki, T. *et al.* Obesity and metabolic syndrome in histone demethylase JHDM2a-deficient mice. *Genes Cells* **14,** 991–1001 (2009).

173. Okada, Y., Tateishi, K. & Zhang, Y. Histone demethylase JHDM2A is involved in male infertility and obesity. *J. Androl.* **31,** 75–8

174. Okada, Y., Scott, G., Ray, M. K., Mishina, Y. & Zhang, Y. Histone demethylase JHDM2A is critical for Tnp1 and Prm1 transcription and spermatogenesis. *Nature* **450,** 119–23 (2007).

175. Höög, C., Schalling, M., Grunder-Brundell, E. & Daneholt, B. Analysis of a murine male germ cell-specific transcript that encodes a putative zinc finger protein. *Mol. Reprod. Dev.* **30,** 173–81 (1991).

176. Kaukonen, R. *et al.* Normal stroma suppresses cancer cell proliferation via mechanosensitive regulation of JMJD1a-mediated transcription. *Nat. Commun.* **7,** 12237 (2016).

177. Schneider, P. *et al.* Identification of a novel actin-dependent signal transducing module allows for the targeted degradation of GLI1. *Nat. Commun.* **6,** 8023 (2015).

178. Roper, R. J. *et al.* Defective cerebellar response to mitogenic Hedgehog signaling in Down [corrected] syndrome mice. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 1452–6 (2006).

179. Das, I. *et al.* Hedgehog Agonist Therapy Corrects Structural and Cognitive Deficits in a Down Syndrome Mouse Model. *Sci. Transl. Med.* **5,** 201ra120-201ra120 (2013).

180. Mao, J. *et al.* Regulation of Gli1 Transcriptional Activity in the Nucleus by Dyrk1. *J. Biol. Chem.* **277,** 35156–35161 (2002).

181. Abe, Y. *et al.* JMJD1A is a signal-sensing scaffold that regulates acute chromatin dynamics via SWI/SNF association for thermogenesis. *Nat. Commun.* **6,** 7052 (2015).

182. Mahajan, K., Bandyopadhyay, S. & Mahajan, N. KDM3A tyrosine phosphorylation by Ack1 promotes tamoxifen-resistance in breast cancer. *Cancer Res.* **74,** (2014).

183. Cheng, M. *et al.* Specific Phosphorylation of Histone Demethylase KDM3A Determines Target Gene Expression in Response to Heat Shock. *PLoS Biol.* **12,** e1002026 (2014).

184. Ohguchi, H. *et al.* The KDM3A–KLF2–IRF4 axis maintains myeloma cell survival. *Nat. Commun.* **7,** 10258 (2016).

185. Sechler, M., Parrish, J. & Jedlicka, P. The histone demethylase KDM3A in Ewing sarcoma metastasis. *Cancer Res.* **76,** (2016).

186. Ramadoss, S., Sen, S., Chaudhuri, G. & Farias-Eisner, R. KDM3A promotes cell growth and cisplatin-resistance in ovarian cancer cells. *Cancer Res.* **74,** (2014).

187. Parrish, J. K., Sechler, M., Winn, R. A. & Jedlicka, P. The histone demethylase KDM3A is a microRNA-22-regulated tumor promoter in Ewing Sarcoma. *Oncogene* **34,** 257–262 (2015).

188. Krieg, A. J. *et al.* Regulation of the histone demethylase JMJD1A by hypoxia-inducible factor 1 alpha enhances hypoxic gene expression and tumor growth. *Mol. Cell. Biol.* **30,** 344–53 (2010).

189. Mimura, I. *et al.* Dynamic change of chromatin conformation in response to hypoxia enhances the expression of GLUT3 (SLC2A3) by cooperative interaction of hypoxia-inducible factor 1 and KDM3A. *Mol. Cell. Biol.* **32,** 3018–32 (2012).

190. Cho, H.-S. *et al.* The JmjC domain-containing histone demethylase KDM3A is a positive regulator of the G1/S transition in cancer cells via transcriptional regulation of the HOXA1 gene. *Int. J. Cancer* **131,** E179–E189 (2012).

191. Kim, J.-Y. *et al.* KDM3B is the H3K9 demethylase involved in transcriptional activation of lmo2 in leukemia. *Mol. Cell. Biol.* **32,** 2917–33 (2012).

192. Björkman, M. *et al.* Systematic knockdown of epigenetic enzymes identifies a novel histone demethylase PHF8 overexpressed in prostate cancer with an impact on cell proliferation, migration and invasion. *Oncogene* **31,** 3444–3456 (2012).

193. Liu, Y. *et al.* An epigenetic role for PRL-3 as a regulator of H3K9 methylation in colorectal cancer. *Gut* **62,** 571–81 (2013).

194. Paolicchi, E., Crea, F., Farrar, W. L., Green, J. E. & Danesi, R. Histone lysine demethylases in breast cancer. *Crit. Rev. Oncol. Hematol.* **86,** 97–103 (2013).

195. Hu, Z. *et al.* A novel nuclear protein, 5qNCA (LOC51780) is a candidate for the myeloid leukemia tumor suppressor gene on chromosome 5 band q31. *Oncogene* **20,** 6946–6954 (2001).

196. Han, A. *et al.* Transcriptional repression of ANGPT1 by histone H3K9 demethylase KDM3B. *BMB Rep.* **48,** 401–6 (2015).

197. Chen, J. *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* **45,** 34–42 (2012).

198. Liu, Z. *et al.* The Histone H3K9 Demethylase Kdm3b Is Required for Somatic Growth and Female Reproductive Function. *Int. J. Biol. Sci.* **11,** 494–507 (2015).

199. Liu, Z. *et al.* Knockout of the Histone Demethylase Kdm3b Decreases Spermatogenesis and Impairs Male Sexual Behaviors. *Int. J. Biol. Sci.* **11,** 1447–57 (2015).

200. Herz, H.-M. *et al.* Histone H3 lysine-to-methionine mutants as a paradigm to study chromatin signaling. **345,** 1065–70 (2014).

201. Wu, G. *et al.* Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat. Genet.* **44,** 251–3 (2012).

202. Chan, K.-M. *et al.* The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Genes Dev.* **27,** 985–90 (2013).

203. Kim, S.-M. *et al.* Regulation of mouse steroidogenesis by WHISTLE and JMJD1C through histone methylation balance. *Nucleic Acids Res.* **38,** 6389–403 (2010).

204. Kuroki, S. *et al.* JMJD1C, a JmjC domain-containing protein, is required for long-term maintenance of male germ cells in mice. *Biol. Reprod.* **89,** 93 (2013).

205. Sroczynska, P. *et al.* shRNA screening identifies JMJD1C as being required for leukemia maintenance. *Blood* **123,** (2014).

206. Watanabe, S. *et al.* JMJD1C demethylates MDC1 to regulate the RNF8 and BRCA1-mediated chromatin response to DNA breaks. *Nat. Struct. Mol. Biol.* **20,** 1425–33 (2013).

207. Lu, J. & Matunis, M. J. A mediator methylation mystery: JMJD1C demethylates MDC1 to regulate DNA repair. *Nat. Struct. Mol. Biol.* **20,** 1346–8 (2013).

208. Sáez, M. A. *et al.* Mutations in JMJD1C are involved in Rett syndrome and intellectual disability. *Genet. Med.* **18,** 378–85 (2016).

209. Schneider Leonid. Manel Esteller, the Schrödinger cat of Barcelona. *For Better Science* (2016). doi:10.1080/15476286.2016.1191208

210. Wang, J., Park, J. W., Drissi, H., Wang, X. & Xu, R.-H. Epigenetic regulation of miR-302 by JMJD1C inhibits neural differentiation of human embryonic stem cells. *J. Biol. Chem.* **289,** 2384–95 (2014).

211. Nakajima, R., Okano, H. & Noce, T. JMJD1C Exhibits Multiple Functions in Epigenetic Regulation during Spermatogenesis. *PLoS One* **11,** e0163466 (2016).

212. Wolf, S. S., Patchev, V. K. & Obendorf, M. A novel variant of the putative demethylase gene, s-JMJD1C, is a coactivator of the AR. *Arch. Biochem. Biophys.* **460,** 56–66 (2007).

213. Chen, M. *et al.* JMJD1C is required for the survival of acute myeloid leukemia by functioning as a coactivator for key transcription factors. *Genes Dev.* **29,** 2123–39 (2015).

214. Cachon-Gonzalez, M. B. *et al.* Structure and expression of the hairless gene of mice. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 7717–21 (1994).

215. Ahmad, W. *et al.* Alopecia universalis associated with a mutation in the human hairless gene. *Science* **279,** 720–4 (1998).

216. Liu, L. *et al.* Hairless is a histone H3K9 demethylase. *FASEB J.* **28,** 1534–42 (2014).

217. Hsieh, J.-C. *et al.* Physical and functional interaction between the vitamin D receptor and hairless corepressor, two proteins required for hair cycling. *J. Biol. Chem.* **278,** 38665–74 (2003).

218. Moraitis, A. N., Giguère, V. & Thompson, C. C. Novel mechanism of nuclear receptor corepressor interaction dictated by activation function 2 helix determinants. *Mol. Cell. Biol.* **22,** 6831–41 (2002).

219. Potter, G. B. *et al.* The hairless gene mutated in congenital hair loss disorders encodes a novel nuclear receptor corepressor. *Genes Dev.* **15,** 2687–701 (2001).

220. Thompson, C. C. & Bottcher, M. C. The product of a thyroid hormone-responsive gene interacts with thyroid hormone receptors. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 8527–32 (1997).

221. Thompson, C. C. Hairless is a nuclear receptor corepressor essential for skin function. *Nucl. Recept. Signal.* **7,** (2009).

222. Malloy, P. J., Wang, J., Jensen, K. & Feldman, D. Modulation of Vitamin D Receptor Activity by the Corepressor Hairless: Differential Effects of Hairless Isoforms. *Endocrinology* **150,** 4950–4957 (2009).

223. AZEEM, Z. *et al.* Congenital atrichia with papular lesions resulting from novel mutations in human hairless gene in four consanguineous families. *J. Dermatol.* **38,** 755–760 (2011).

224. Mi, Y., Zhang, Y. & Shen, Y. Mechanism of JmjC-containing protein Hairless in the regulation of vitamin D receptor function. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1812,** 1675–1680 (2011).

225. Thompson, C. C., Sisk, J. M. & Beaudoin, G. M. J. Hairless and Wnt Signaling: Allies in Epithelial Stem Cell Differentiation. *Cell Cycle* **5,** 1913–1917 (2006).

226. Beaudoin, G. M. J., Sisk, J. M., Coulombe, P. A. & Thompson, C. C. Hairless triggers reactivation of hair growth by promoting Wnt signaling. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 14653–8 (2005).

227. Kumpf, S. *et al.* Hairless promotes PPARγ expression and is required for white adipogenesis. *EMBO Rep.* **13,** (2012).

Chapter 2
Development of a cell-permeable ester derivative of the JmjC inhibitor IOX1

2.1 Introduction

Aberrant histone methylation states are associated with a number of diseases, predominantly in cancer but also in neurological disorders such as autism and X-linked mental retardation (XLMR).[1–5] Although histone methylation was once considered irreversible, it is now known to be reversible, increasing its importance for pharmaceutical studies.[6]

The Jumonji C domain (JmjC) family is the largest family of histone lysine demethylases.[7] All JmjC subfamilies contain the conserved JmjC catalytic domain, however, each subfamily differs in their Fe^{II} and 2OG binding sites, properties that can be potentially exploited for the development of selective inhibitors .[8–10] Understanding the enzymatic mechanism of the JmjC KDMs has stimulated work to develop small-molecule inhibitors, which can be used for therapeutic intervention as well as chemical probes that modulate JmjC catalytic activity in order to investigate their biological roles. [11]

Chemical probes can be useful tools in deducing the underlying mechanism of biological processes.[12] Small-molecule inhibitors can be administered in a reversible, dose-dependent manner, whereas the use of genetic methods is often less controllable.[12]

The vast majority of the reported small-molecule inhibitors for JmjC KDMs are 2OG competitors.[13] Commonly used broad-spectrum inhibitors include *N*-oxalylglycine (NOG), which is an amide analogue of 2OG, as well as 2,4-pyridinedicarboxylic acid (2,4-PDCA) and 5-carboxy-8-hydroxyquinoline (IOX1) (**Figure 2.1**).[11] All of these broad-spectrum inhibitors occupy part of the 2OG binding site and are acids with iron-chelating pharmacophores. The polarity due to the acid often reduces the cell-permeability necessary for biological work.[10] Nevertheless, these molecules have been used as broad-spectrum

inhibitors as well as a template for derivatisation of more cell-permeable and selective inhibitors.[14–16]

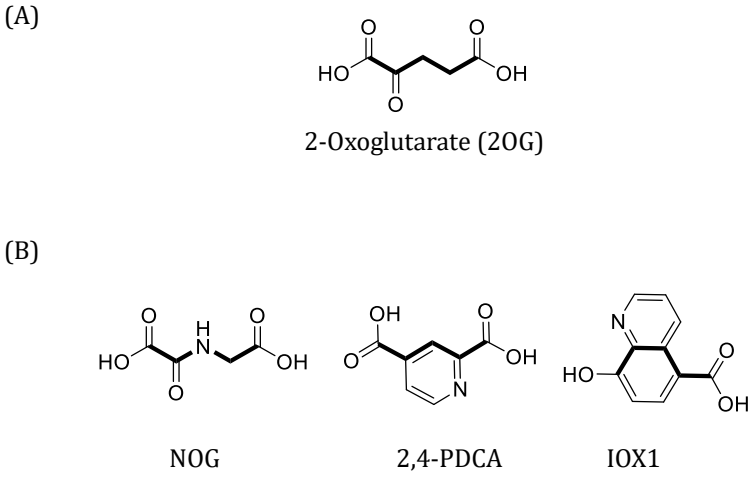


Figure 2.1 2-Oxoglutarate (2OG) analogues reported as broad-spectrum JmjC histone lysine demethylase (KDM) inhibitors. (A) 2-Oxoglutarate (2OG). (B) *N*-Oxalylglycine (NOG), 2, 4-pyridinedicarboxylic acid (2, 4-PDCA) and 8-hydroxyquinoline-5-carboxylic acid (IOX1).

Among these broad-spectrum inhibitors, IOX1 is reported to be the most potent against a representative panel of JmjC KDMs.[17] IOX1 was discovered by a quantitative high-throughput screening where it demonstrated clear structure activity relationships and was shown to inhibit KDM4E (IC_{50} value of 0.3 μM *in vitro*).[18] Since its discovery, IOX1 has been found to inhibit not only KDM4E, but also a broad panel of 2OG dependent oxygenases (KDM4A/C/D/E ,KDM3A, KDM6A, KDM2A, PHF8 and KDM5C).[17] Therefore, IOX1 has been declared a tool compound for 2OG oxygenases.[19] IOX1 was found to chelate with the active-site iron via its quinoline nitrogen and phenolic hydroxyl in a bidentate manner, causing iron translocation.[17] The carboxylic acid at the C-5 position of IOX1 was observed to interact with Tyr132 and Lys206 of KDM4A, similar to binding of the 2OG C-5 carboxylate (**Figure 2.2**).[18]

Although IOX1 exhibits an IC_{50} in the submicromolar to micromolar range against the tested 2OG oxygenases, its efficiency in cells is about two orders of magnitude lower (for KDM4A IC_{50} value of 0.2 μM *in vitro* vs. EC_{50} value of 86 μM in HeLa cells), possibly due to low cell permeability resulting from its polar C-5 carboxyl group.[17,18,20]

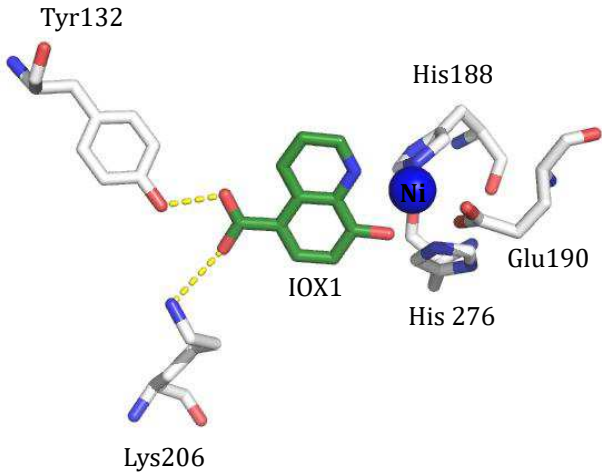


Figure 2.2 View from a crystal structure of IOX1 bound to KDM4A. Active site residues are shown as grey sticks, IOX1 is shown in green sticks. The Ni^{II} ion (replacing Fe^{II}) is shown as a blue sphere. Hydrogen bonds are shown in yellow (PDB ID - 3NJY).[18]

For cellular studies, a small-molecule tool compound must be sufficiently polar to be soluble in aqueous conditions, yet it should also be sufficiently lipophilic to cross cell membranes. Therefore, the compound needs to have both hydrophilic and lipophilic characteristics. A prodrug is a compound that on administration undergoes chemical conversion by metabolic processes before becoming an active agent.[21]

When a carboxylic acid is involved in target binding, as was assumed to be the case of IOX1, it might interfere with the membrane permeability of the compound as it is polar and ionisable.[22] Ester prodrug derivatives can be used to mask polar groups and therefore they decrease polarity and increase the hydrophobic character. Since esters are often readily hydrolysed by esterases to acids, the formation of an ester prodrug should aid in increasing transmembrane permeability without affecting the activity of the parent acid compound.[23] Providing that the alcohol leaving group is non-toxic, the use of a prodrug ester derivative can therefore be used as a general technique to increase membrane permeability in carboxylic acid chemical probes and drugs.[24]

In oral drug delivery, ester prodrugs are commonly used to enhance transepithelial transport of hydrophilic drugs by increasing the lipophilicity of the parent compound, resulting in enhanced transmembrane transport by passive diffusion.[25,26] The parent compound is expected to be released into the plasma through cleavage of the prodrug. An example for such a prodrug is the anti-viral agent Oseltamivir (Tamiflu) where the ethyl ester masks the carboxylic acid, resulting in improved oral bioavailability (**Figure 2.3 A**).[27]

Another example is the monoethyl ester Enalapril, which is orally bioavailable as opposed to its parent compound which requires administration by injection (**Figure 2.3 B**).[28]

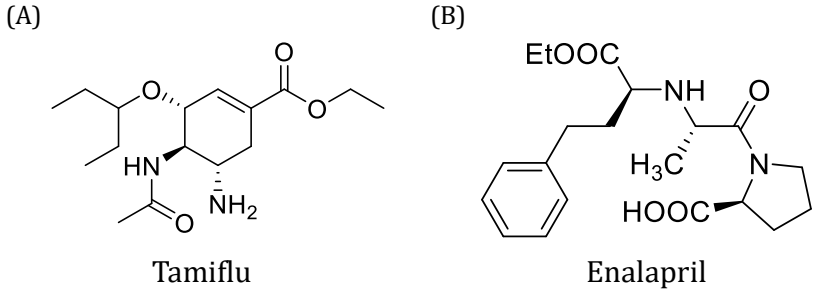


Figure 2.3 Structures of two commercially available ester prodrugs. (A) The anti-viral agent Oseltamavir (Tamiflu) ethyl ester. (B) Enalapril monoethyl ester for the treatment of hypertension.[27,28]

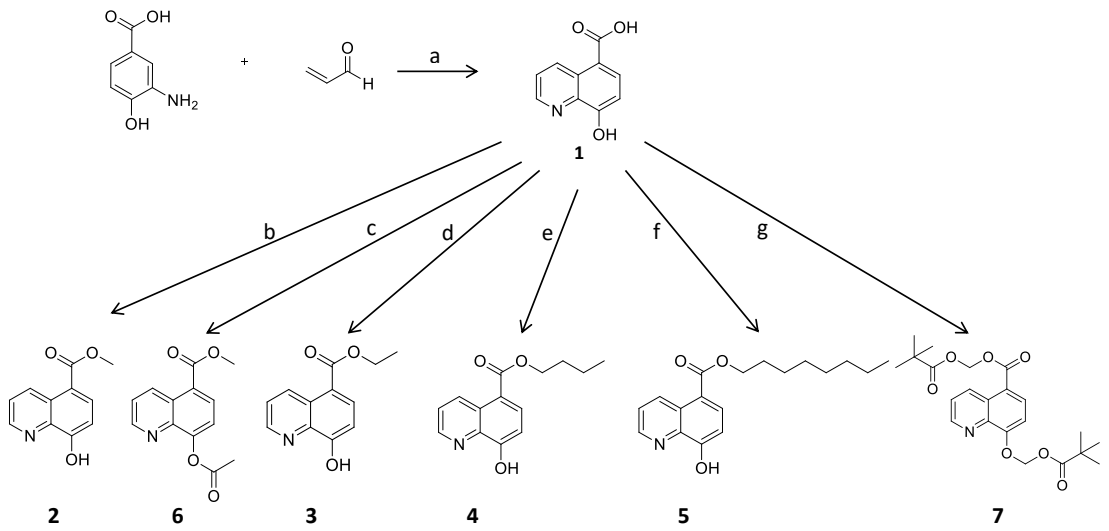
2.2 Objectives

The work described in this chapter aimed to improve the transmembrane permeability of IOX1. Ester derivatives with different lengths of alkoxy groups were synthesised as potential prodrugs. Characterisation of the IOX1 derivatives was carried out via *in vitro* inhibition studies as well as activity vs. cytotoxicity analysis in cell cultures. Further work investigated a 'lead' compound for intracellular uptake and stability using mass spectrometric studies. Finally, structure activity relationship (SAR) studies were carried out to investigate the *in vitro* selectivity against representatives of JmjC subfamilies with increased ester chain length.

*AlphaScreen measurements were carried out by Dr. Anthony Tumber (Structural Genomics Consortium (SGC), Oxford) and Dr. Tzu-Lan Yeh. Compounds **2** and **6** were provided by Prof. David Maloney (National Center for Advancing Translational Sciences (NCATS), NIH); Compound **3** was synthesised by Dr. Clarisse Lejeune. Dr. James Wickens provided guidance on mass spectrometric studies. This work was published by Schiller et al.*[29]

2.3 Synthesis of IOX1 and its ester derivatives

With the aim of improving the transmembrane permeability of IOX1 **1**, ester derivatives with different lengths of alkoxy groups were synthesised. To study the effect of increasing the alkoxy chain length, methyl- **2**, ethyl- **3**, *n*-butyl- **4** and *n*-octyl- **5** esters were synthesised (**Scheme 2.1**). To test whether improved permeability could be obtained, both the carboxylic acid and the phenol group were protected (methyl acetate diester **6**), and the activity of this compound was to be compared with the methyl ester derivative **2**. In addition, di-*tert*-butyl diacetate **7** was synthesised as it provides several hydrolysis pathways. See Chapter 11, Section 11.1, for detailed experimental procedures and compound characterization.



Scheme 2.1 **Ester derivatives of IOX1**. Reagents and conditions: (a) 6N HCl, 100°C, reflux, 2 h; (b) and (c) - *provided by Prof. David Maloney (NCGC, NIH)*; (d) *synthesised by Dr. Clarisse Lejeune;* (e) cat. H_2SO_4, butanol, reflux, 120°C, 21 h; (f) cat. H_2SO_4, octanol, reflux, 120°C for 20 h; (g) choloromethyl pivalate, Et3N, 1:1 acetone / DMF, reflux, 50°C for 3 h. See Chapter 11, Section 11.1, for more information.

2.3.1 Synthesis of IOX1

The Skraup reaction was employed to synthesise the quinolone ring of IOX1 **1** from commercial 3-amino-4-hydroxybenzoic acid and acrolein (**Scheme 2.1 a; Figure 2.4**). In this reaction, acrolein underwent an 1,4-addition and was added slowly over 30 minutes to

reduce polymerisation side reactions. The product was purified based on the difference in solubility in water depending on the pH. In a basic pH the molecule is negatively charged and therefore soluble in water. Acidification of the molecule in pH 4-5 neutralises the charge and makes the compound insoluble. The desired product, IOX1 **1**, was obtained with 95% purity as was determined by liquid chromatography–mass spectrometry (LC-MS) and therefore was used without further purification.

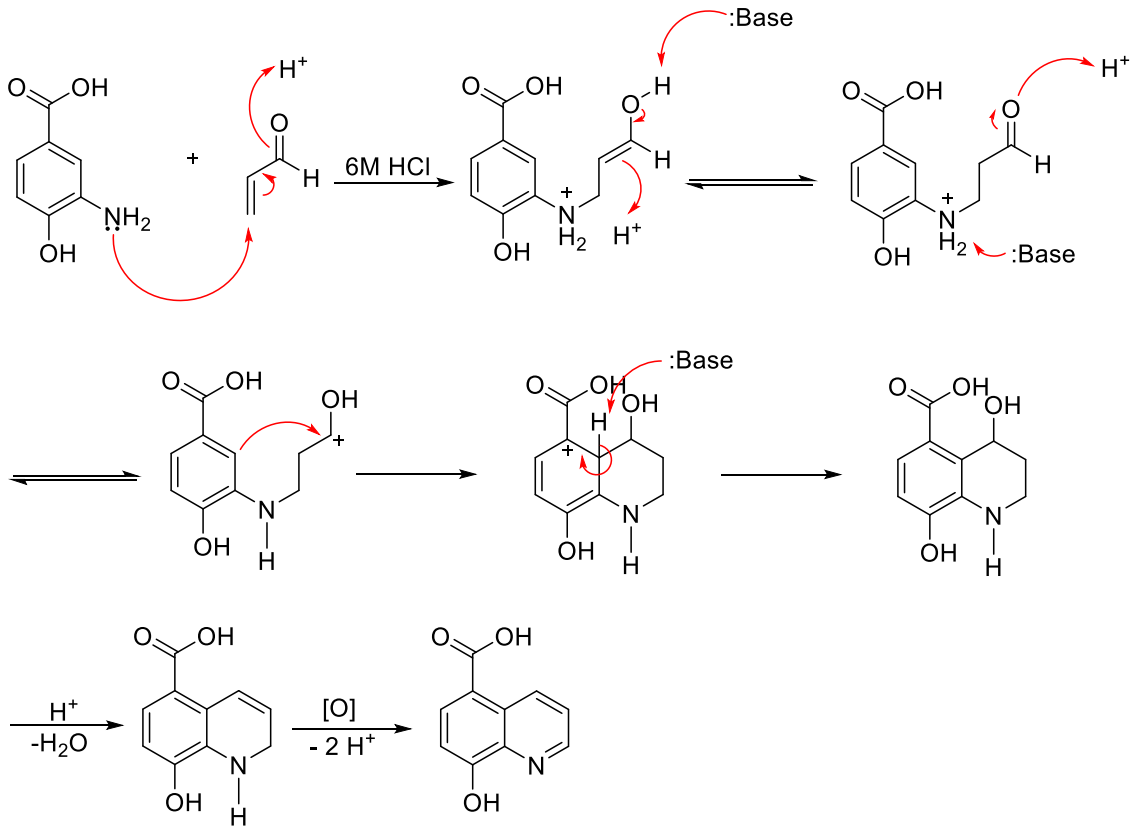


Figure 2.4 Skraup reaction mechanism to form IOX1 1. Acrolein undergoes 1,4-addition, intramolecular electrophilic aromatic substitution, then dehydration and oxidation.[30]

2.3.2 Synthesis of *n*-butyl- **4** and *n*-octyl- **5** ester derivatives

Synthesis of *n*-butyl- **4** and *n*-octyl- **5** ester derivatives was carried out by Fischer esterification of IOX1 **1**, wherein the carboxylic acid was treated with the appropriate alcohol in the presence of an inorganic acid catalyst (**Scheme 2.1 e-f; Figure 2.5**). The acid (H_2SO_4) reversibly protonates a small percentage of the carboxylic acid molecules, and the

protonated carboxylic acids are then susceptible to attack by the alcohol. The appropriate alcohol (e.g. *n*-Butanol for *n*-butyl ester **4** derivative) was used as the reaction solvent to drive equilibrium towards ester formation, following "Le Châtelier's principle". Similarly, the *n*-octyl ester derivative **5** was prepared by Fischer esterification, by using *n*-octanol as the solvent. After successful identification of reaction conditions to yield **5**, optimisations of purification conditions were carried out. Since *n*-octanol has a high boiling point (195°C), it was evaporated by rotary evaporation at 70°C and 15 mbar. The residual traces were washed with cyclohexane and purified by column chromatography affording product **5**.

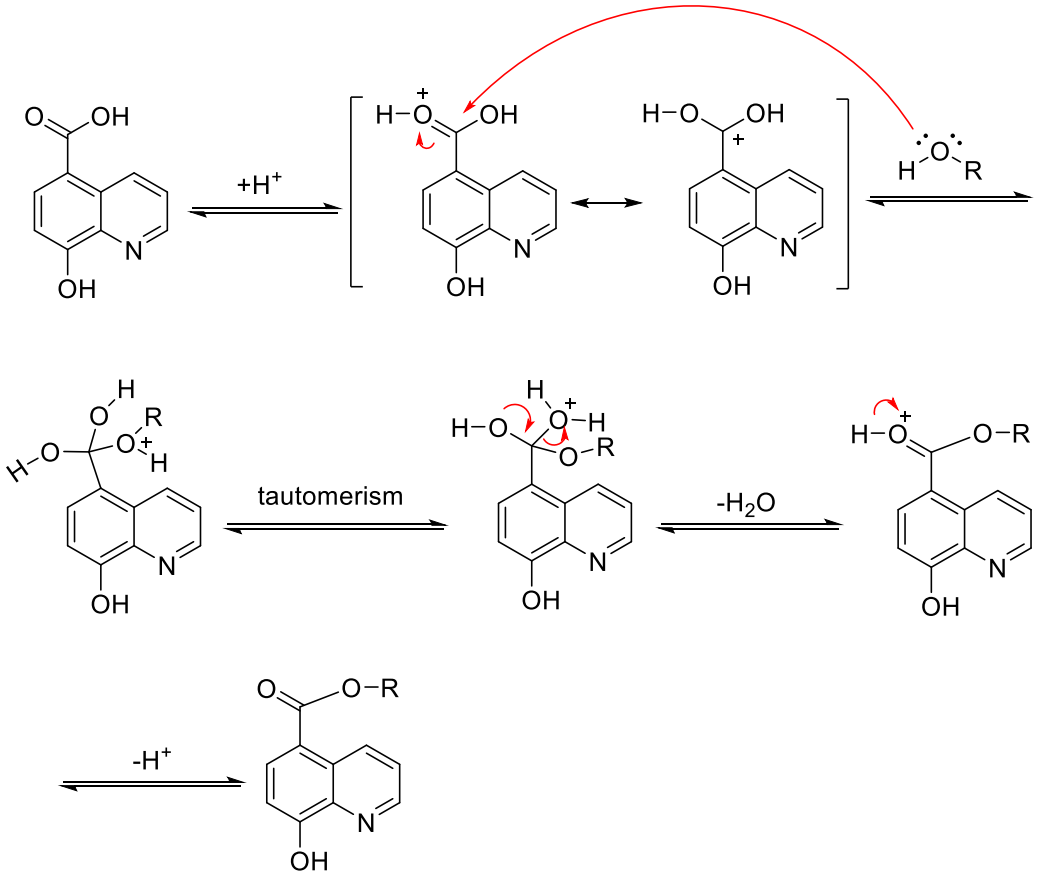


Figure 2.5 Fischer esterification mechanism to form *n*-butyl- 4 and *n*-octyl- 5 ester derivatives of IOX1.[31]

2.3.3 Synthesis of di-*tert*-butyl diacetate derivative **7**

Di-*tert*-butyl diacetate **7** was synthesised from IOX1 **1** using chloromethyl pivalate based on an optimised reaction conditions reported by Nudelman and co-workers, using triethylamine as a base in a solution of acetone / dimethylformamide (1:1) under reflux conditions (**Scheme 2.1 g**; **Figure 2.6**).[32,33] Although equimolar quantities of IOX1 **1** and

chloromethyl pivalate were used, a mixture of three products was detected. The main product appeared to be the diacetate **7**, which is formed via the double addition of chloromethyl pivalate. In addition, the mixture also contained the mono-addition product on either the phenol or on the acid. The propensity of **1** to undergo substitution on both positions might be explained by its poor solubility, making the actual concentration of IOX1 **1** in solution low in comparison with the concentration chloromethyl pivalate. Substitution on one position made the molecule less polar and therefore more soluble in the solvent, resulting in a preference towards double substitution. Esterification on both positions might provide improved cell permeability through increased lipophilicity.

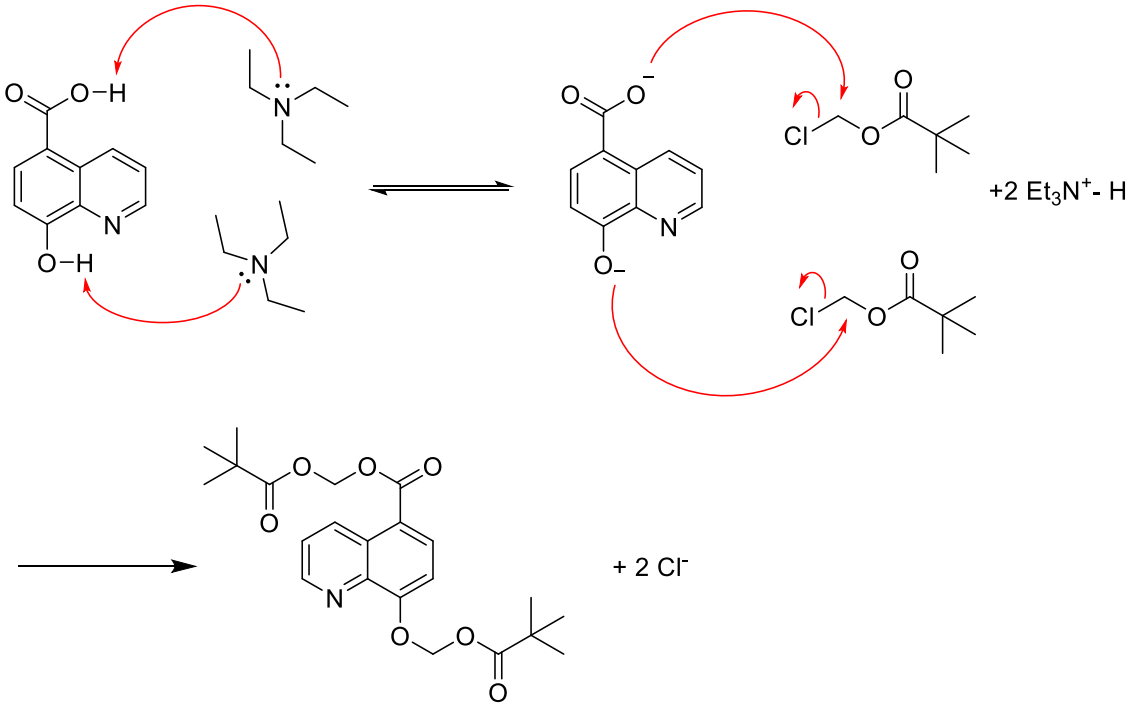


Figure 2.6 Mechanism of synthesis of di-*tert*-butyl diacetate 7 from IOX1 1. Deprotonation of the phenol and the carboxylic acid by trimethylamine enables nucleophilic substitution and the formation of di-*tert*-butyl diacetate derivative of IOX1.[32,33]

2.4 Investigating the cellular toxicity of IOX1 ester derivatives

IOX1 has been reported to be non-toxic to HeLa cells over a range of concentrations where it is active as a KDM inhibitor (CC_{50} value of 291.6 μM, EC_{50} value of 86.5 μM for KDM4A).[18] In order to test the toxicity of the ester derivatives of IOX1, and to compare it to that of IOX1, cytotoxicity tests were conducted using the CellTiter 96® AQueous One Solution cytotoxicity assay (Promega). This is a colorimetric method which enables determination of the number of viable cells using a tetrazolium compound (3-(4,5-dimethylthiazol-2-yl)-5-(3- carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium, MTS).[34,35] The MTS compound is bioreduced by metabolically active cells into a coloured Formazan product which is soluble in tissue culture medium.[36] Thus, measuring the absorbance at 490 nm is directly proportional to the number of living cells in culture (**Figure 2.7**).

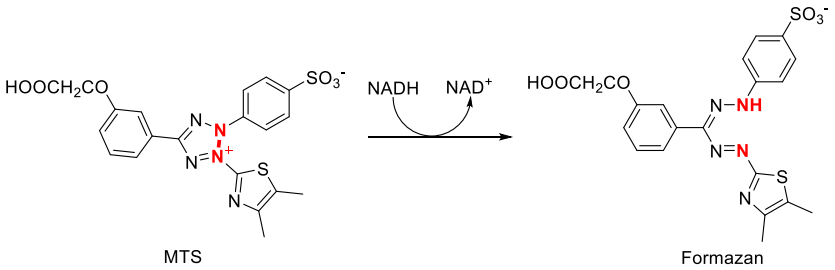


Figure 2.7 Structures of MTS and coloured Formazan product.[36]

Prior to use of the MTS tetrazolium assay, calibration of the cell seeding number was performed in order to achieve absorbance signal in the linear range after 24 hours of incubation. The value of 2000 cells/well was chosen to be the seeding number for cytotoxicity assays (**Figure 2.8 A**). Staurosporine, a natural product that inhibits protein kinases in cells and induces apoptosis, was used as a control for cytotoxicity and showed dose-response dependency (**Figure 2.8 B**).[37]

(A) (B)

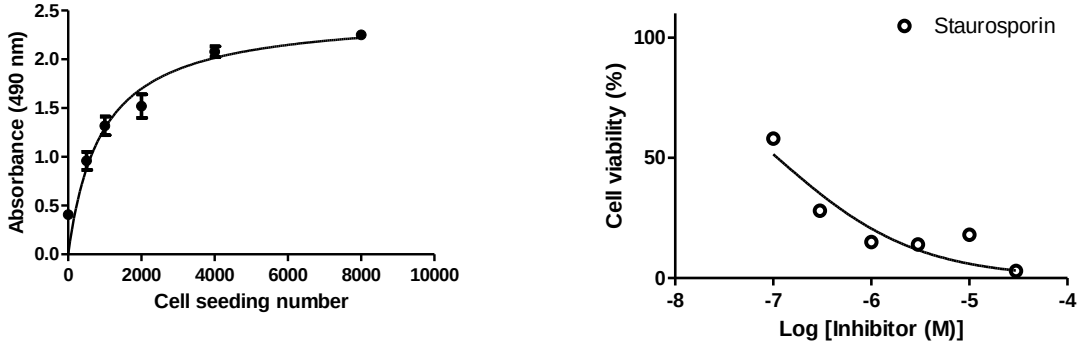


Figure 2.8 Cell viability control measurements. (A) Calibration of the cell seeding number needed for the absorption signal to be in the linear range after 24 h incubation. All measurements were conducted in technical triplicate and values represent mean and standard deviation. (B) Staurosporine at concentrations of 0.03 - 10 μM treatments as a control for cytotoxicity. See Chapter 11, Section 11.2.4.2, for more information.

The viability of HeLa cells was analysed after 24 hours following incubation with different concentrations (1–300 μM) of IOX1 **1** or its ester derivatives **2–7**. A vehicle DMSO was used to determine 100% cell viability. Media only wells were used as a baseline to normalise the signal intensity (see Chapter 11, Section 11.2.4.2, for more information).

Methyl ester derivative **2** was the most cytotoxic compound, with a CC_{50} value of 10 μM (**Figure 2.9**; see **Table 2.1** in Section 2.7.2). Di-substituted compounds **6** and **7** had CC_{50} values of 29 and 17 μM, respectively. Ethyl **3** and *n*-butyl **4** esters had similar CC_{50} values of 50 and 66 μM, respectively. Out of the tested compounds, only **7** resulted in complete toxicity at the highest concentration tested, while treatment with the other compounds led to between 25% and 60% viable cells. *n*-Octyl ester **5** was not cytotoxic within the tested concentration range, with a CC_{50} value greater than 300 μM and with over 50% viable cells at the highest concentration tested. This CC_{50} value is similar to the CC_{50} value obtained here for IOX1 **1** (> 300 μM), which is in agreement with the reported value of IOX1 (291.6 μM).[18]

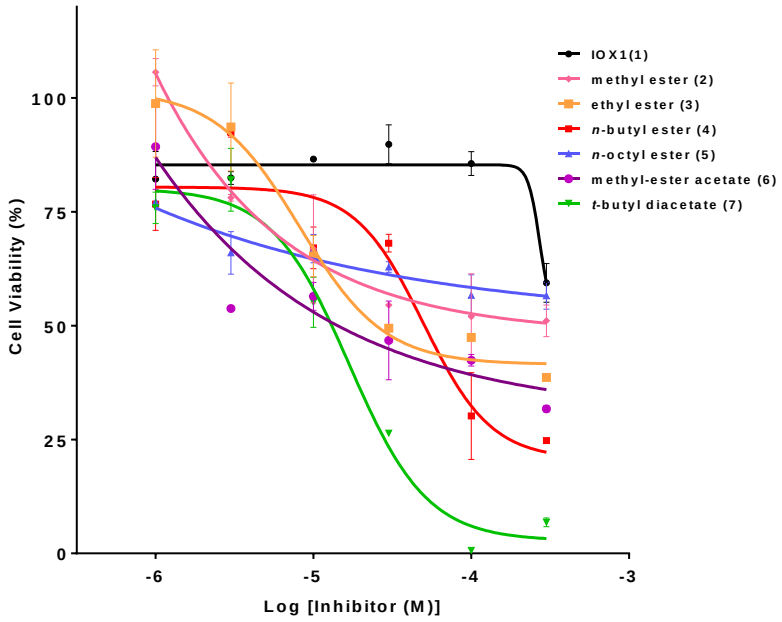


Figure 2.9 HeLa cell viability analysed following dosing with IOX1 ester derivatives. Cytotoxicity measurements of IOX1 **1** and its ester derivatives at varied concentrations (1 - 300 μM). All measurements were conducted in triplicate and values represent mean and standard deviation (see Chapter 11, Section 11.2.4.2, for more information). Figure reused with permission from Schiller *et al.*[29]

2.5 Investigating the cellular potency of IOX1 ester derivatives

Ester derivatisation of IOX1 was carried out in order to improve the transmembrane permeability of IOX1, and consequently to improve its cellular potency.

To quantitatively assess the effect of the IOX1 esters on the demethylation activity in cells, the reported immunofluorescence assay was adapted using KDM4A as a representative JmjC KDM.[18] FLAG-tagged KDM4A was transiently overexpressed in HeLa cells, and these were then treated with either a vehicle control (DMSO) or varying concentrations (1–300 μM) of IOX1 **1** or IOX1 ester derivatives **2–7**. After 24 hours of compound dosing, the cells were analysed by indirect immunofluorescence using an anti-FLAG tag antibody to identify cells overexpressing KDM4A, and an antibody for endogenous H3 K9me3 to quantify the level of this histone modification, reported to be regulated by KDM4A.[38] As a control, cells overexpressing the H188A catalytically deficient KDM4A variant were also analysed (see Chapter 11, Section 11.4.3, for more information).

Treatment with increasing concentrations of IOX1 **1** or the ester derivatives caused a dose-response-dependent increase in H3 K9me3 fluorescence intensity, implying KDM4A inhibition in cells. The cellular EC_{50} value of **1** was determined to be 100 μM (**Figure 2.10**; see **Table 2.1** in Section 2.7.2), correlating with the reported value (EC_{50} = 86.5 μM).[18]

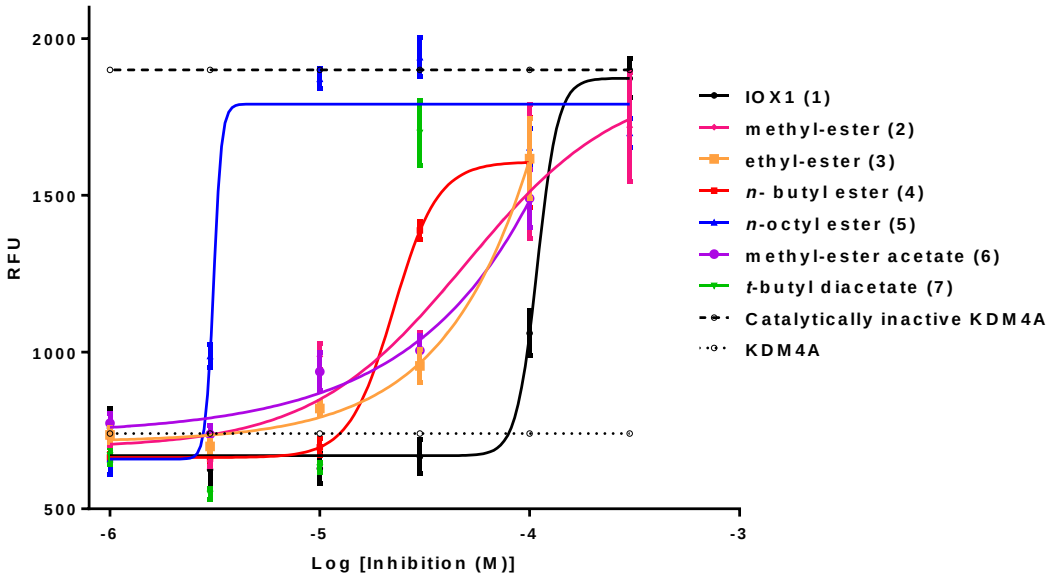


Figure 2.10 Analysis of H3 K9me3 inhibition of demethylation using immunofluorescence assays. Indirect immunofluorescence with anti-FLAG (Sigma F1804), anti-H3 K9me3 (Abcam Ab8898), and DAPI (Sigma D9564) staining in HeLa cells overexpressing FLAG-tagged KDM4A or the H188A catalytic inactive variant of KDM4A. Quantitation of H3 K9me3 levels are reflected in Relative Fluorescence Units (RFU). Data outside the quantification range were removed. EC_{50} values are listed in **Table 2.1** in Section 2.7.2. All measurements were conducted in triplicate and values represent mean and standard deviation (see Chapter 11, Section 11.4.3, for more information). Figure reused with permission from Schiller *et al.*[29]

The apparent cellular EC_{50} values of derivatives **2, 4** and **5** were substantially lower than that of IOX1 **1**, indicating better inhibition of KDM4A enzymatic activity. The smallest ester - methyl ester **2** - was the least active with an EC_{50} value of 88 μM, then the second chain in size- the ethyl-ester **3** - with 54 μM. The methyl-ester acetate **6** in which both the carboxylic acid and the phenolic position are substituted showed an EC_{50} value of 31 μM, and the *n*-butyl ester **4** had the same EC_{50} value as the di-*tert*-butyl diacetate **7** of 25 μM, both have an alkoxy chain length of four carbons. The most potent derivative was *n*-octyl ester **5** which was approximately 30-fold more active than IOX1 **1**, with an EC_{50} value of 3.8 μM (**Figure 2.11**).

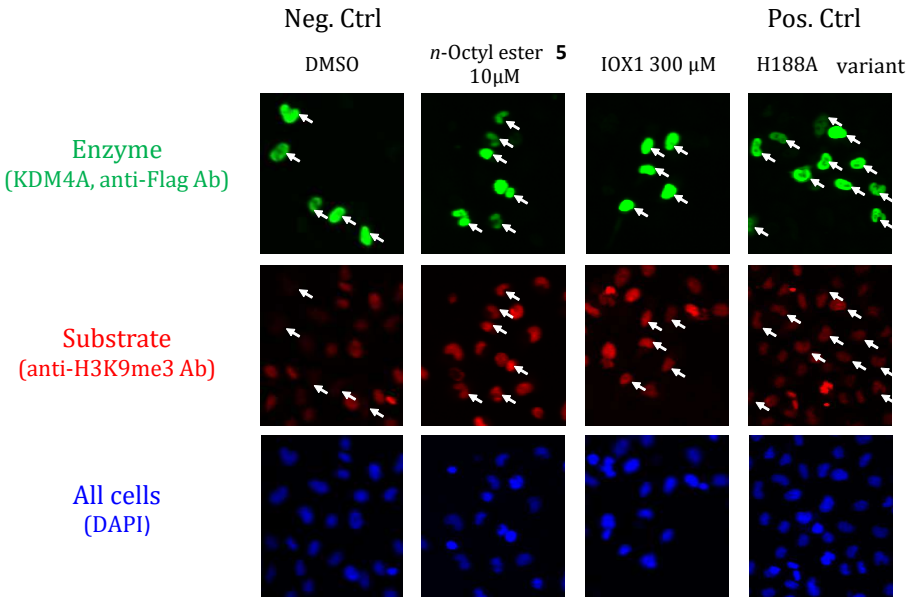


Figure 2.11 *n*-Octyl ester 5 increases H3 K9me3 levels in HeLa cells via KDM4A inhibition. Indirect immunofluorescence assays with anti-FLAG (green) and anti-H3 K9me3 (red) antibodies and with DAPI nuclear staining (blue) in HeLa cells overexpressing FLAG-tagged KDM4A. DMSO treatment has no effect on KDM4A demethylase activity, while IOX1 **1** (300 μM) and *n*-octyl ester **5** (10 μM) treatment resulted in increased H3 K9me3 substrate levels (white arrows indicate transfected cells, higher intensity of red signal indicates more substrate is present in the nucleus, likely due to decreased KDM4A enzymatic catalysis). The H188A catalytically inactive KDM4A variant does not affect H3 K9me3 levels (see Chapter 11, Section 11.4.3, for more information). Figure reused with permission from Schiller *et al*.[29]

2.6 Intracellular delivery measurements

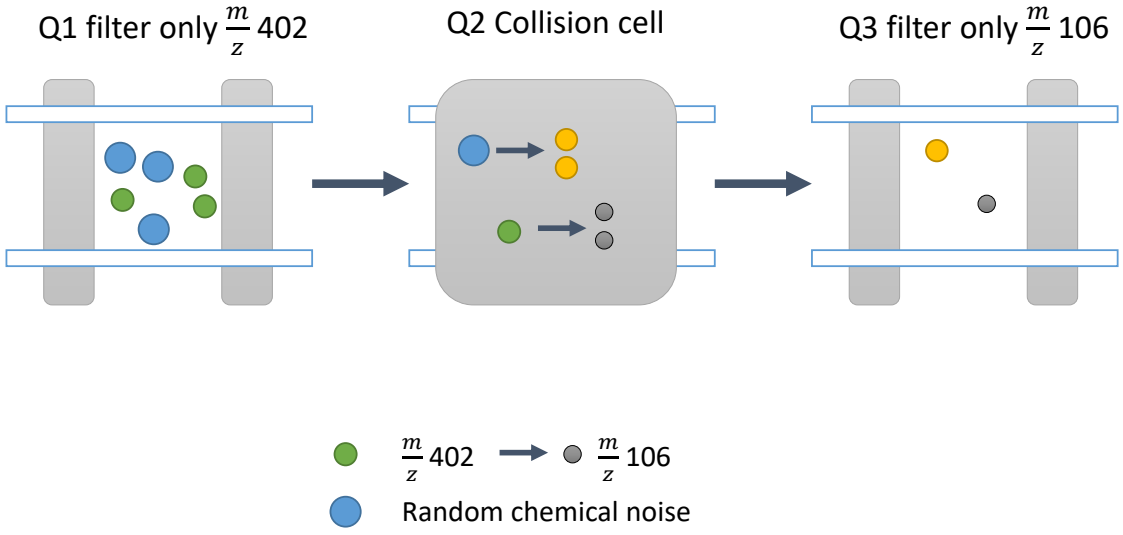
n-Octyl ester **5** exhibited the highest apparent inhibitory potency in cells, and was 30-fold more active than IOX1 **1** (EC_{50} value of 3.8 μM, see Section 2.5). To test whether the improved cellular potency of **5** is due to **5** being a prodrug form of 1 with improved cellular permeability, intracellular delivery assays were then performed. In this experiment the cellular permeabilities of compounds **1** and **5** were compared, and the level of hydrolysis of **5** to the parent compound (IOX1 **1**) in cells was investigated by Liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS).

The principles of the LC-MS/MS analysis method used here, and the development of the intracellular delivery assay are described in Sections 2.6.1 -2.6.3.

2.6.1 Principles of LC-MS/MS

Liquid chromatography (LC) is a robust separation technique which is based on competing interactions between analytes in solution (the mobile phase) and a bed of packed particles in a column (the stationary phase).[39] Mass spectrometry (MS) is a sensitive technique used to detect, identify and quantitate molecules based on their mass and charge (m/z) ratio.[40] Coupling of liquid chromatography (LC) with mass spectrometry (MS) began to be commonly used in the early 1990s.[39] Nowadays, LC-MS is being routinely used in research, providing a simple and robust technique which can be applied to a wide range of biological molecules.[41]

To improve the selectivity and sensitivity of the detection, triple quadrupoles, or tandem mass spectrometry (MS/MS) is often applied. In 'Multiple Reaction Monitoring' (MRM) mode, two stages of mass filtering are employed[42]. At the first stage of MS/MS molecules are ionised in the ion source and separated based on their m/z (Q1, **Scheme 2.2**). The selected analytes (parent ions) are then filtered and transferred to a collision cell (Q2). In Q2 the parent ions collide with a neutral gas at high pressure and fragment via collision-induced dissociation (CID). The resultant daughter ions are transferred into another scanning mass filter (Q3) and then detected on the basis of their *m/z*.

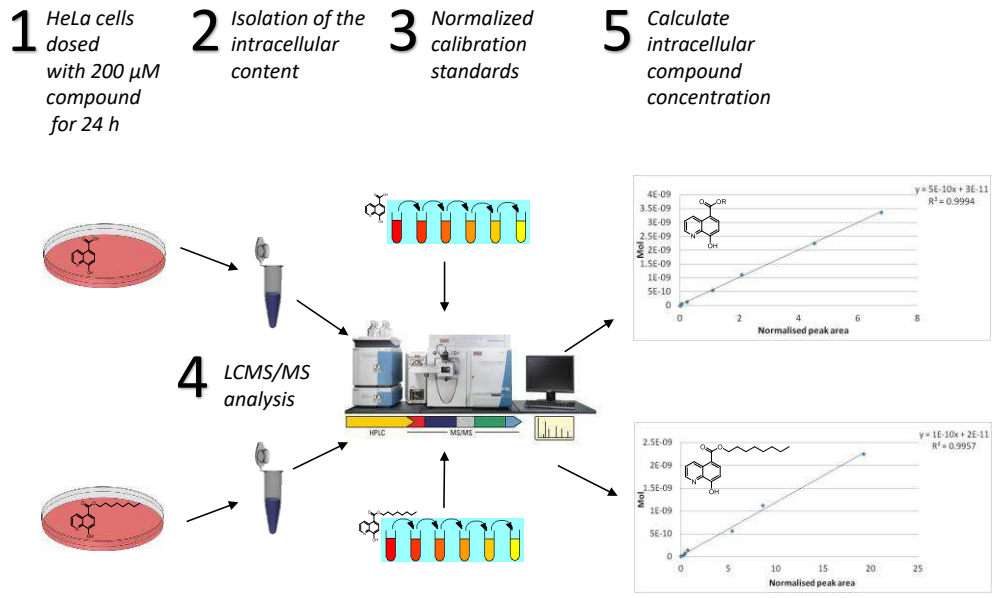


Scheme 2.2 Triple quadrupole mass spectrometer workflow. For high specificity, multi reaction monitoring (MRM) employing two mass analysers (Q1, Q3) and a collision cell (Q2) for collision-induced dissociation (CID) to fragment the analyte (parent ion) into multiple daughter ions. m/z 402 and 106 are examples for parent and daughter ions respectively.

LC-MS/MS was chosen as the preferred analytical technique to measure the intracellular quantities of IOX1 **1** and its *n*-octyl ester derivative **5** as it allows detection of specific molecules out of the cellular content in a highly sensitive and accurate manner. However, a thorough method development was required as described in Sections 2.6.2 and 2.6.3.

2.6.2 Development of an intracellular delivery assay

An intracellular delivery assay was developed based on a protocol from that of Kruidenier *et al.*[43] HeLa cells were dosed with IOX1 **1** or *n*-octyl ester **5** at a concentration of 200 µM and incubated for 24 hours (**Scheme 2.3**). The same cell line and cell culturing technique were used as described for IF experiments to allow comparison between the intracellular levels and the cellular inhibition potencies of the compounds (see Section 2.5).



Scheme 2.3 Outline for intracellular delivery assay procedure. See Chapter 11, Section 11.2.4.4, for more information.

The cells were then counted to enable measurements of moles per cell. Media dosed with compounds was washed to remove excess of compounds that did not enter the cells (last wash was analysed by LC-MS/MS as a control). The cells were lysed using 80% aqueous

methanol for a compatible lysis method for LC-MS analysis.[44,45] Samples were then vacuum concentrated and dissolved with water in proportion to the cell count. Each sample was mixed with 1000 ng/mL caffeine as an internal standard and relative quantitative data were collected by LC-MS/MS. To eliminate carry-over of compounds from one sample to another during LC-MS/MS, a wash with 10% formic acid (aq) was implemented at the end of each run. In addition, blank runs were performed in between measurements to test for any present residues from a previous run. To calculate the intracellular levels of the two forms, IOX1 **1** and *n*-octyl ester of IOX1 **5**, mass and retention times were compared with values for calibration standards (see Section 2.6.3 for calibration standard curves and Chapter 11, Section 11.2.4.4, for more information).

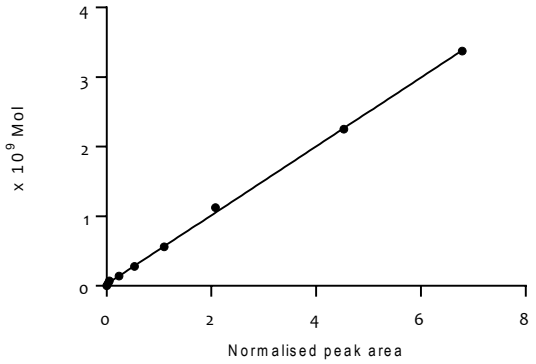
2.6.3 Calibration standards

To correct for variability between the samples, calibration standards were used in the intracellular delivery assay to identify and measure the intracellular levels of IOX1 **1** and *n*-octyl ester of IOX1 **5**.

Calibration standards were prepared by mixing **1** and **5** at a concentration range of 0-200 μM with caffeine as an internal standard (1000 ng/mL), which was chosen as it is a stable and easily detectible compound. To eliminate any ion suppression effects, dilutions were carried out using cell lysate of cells dosed with DMSO (**Figure 2.12**, see Chapter 11, Section 11.2.4.4, for more information)

Quantitative analysis was then carried out by measuring the peak integral in the mass spectra of each compound in the biological samples, then normalising the results using the peak integral of caffeine and performing a linear regression analysis to calculate the number of moles of compound per cell.

(A)



(B)

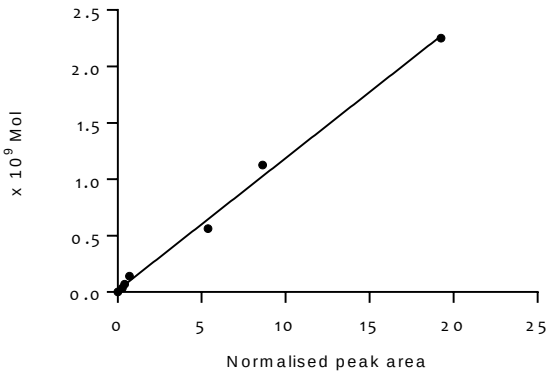
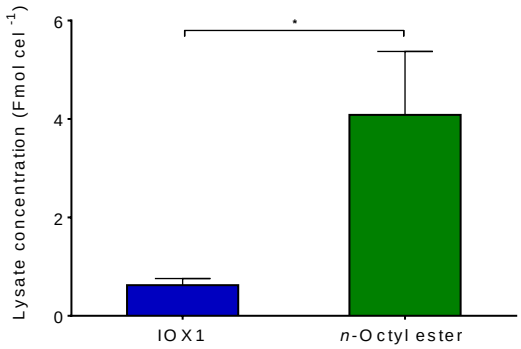


Figure 2.12 Normalised calibration curves used to determine predicted intracellular levels of 1 and 5 by linear regression. Data points represent number of moles compound in analyte vs. normalised peak area. Normalisation of peak area using caffeine internal standard. (A) Calibration curve for IOX1 **1**. (B) Calibration curve for *n*-octyl ester **5**. See Chapter 11, Section 11.2.4.4, for more information. Figure reused with permission from Schiller *et al*.[29]

2.6.4 Intracellular levels and stabilities of IOX1 **1** and its *n*-octyl derivative **5**

Following assay development as described in Sections 2.6.2 - 2.6.3, tests to compare the intracellular levels of **1** and **5**, as well as to measure hydrolysis in cells were carried out. The *n*-octyl ester **5** was found to be significantly more abundant in cells than IOX1 **1** (~six-fold, $P \leq 0.05$) indicating better cell permeability for **5** than **1** (**Figure 2.13 A**). Interestingly, only a small fraction (~2%) of **5** was observed to be hydrolysed to the parent compound IOX1 **1** **(Figure 2.13 B)**. These results imply that the unhydrolysed form of **5** might account for at least some of the KDM inhibitory activity observed in cells treated with ester **5**.

(A)



(B)

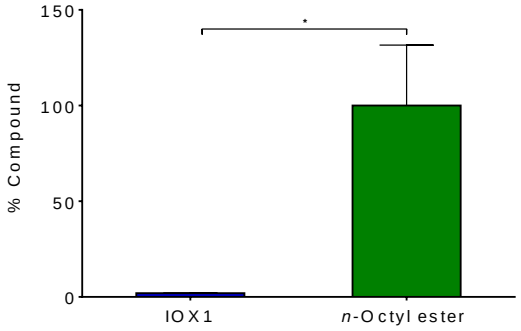


Figure 2.13 Intracellular delivery of IOX1 1 and *n*-octyl ester 5. (A) To measure the intracellular delivery of IOX1 **1** and the *n*-octyl ester **5** the compound were detected in the lysates of HeLa cells 24 h after the administration of either IOX1 **1** or *n*-octyl ester **5** at a concentration of 200 μM. (B) To measure the *n*-octyl ester **5** hydrolysis to IOX1 **1** the compounds were detected in the lysates of HeLa cells 24 h after the administration of the *n*-octyl ester **5** at a concentration of 200 μM. All measurements were conducted in triplicate and values represent mean and standard deviation. P-values were obtained from a two-tailed t-test using GraphPad Prism 6. * denotes P ≤ 0.05 (see Chapter 11, Section 11.2.4.4, for more information). *Dr. James Wickens provided guidance on the mass spectrometric studies.*

2.7 Comparison of *in vitro* inhibition between IOX1 **1** and its ester derivatives

To test whether the unhydrolysed IOX1 ester derivatives can inhibit KDMs, an *in vitro* inhibition assays were carried out. On the basis of crystallographic analysis, the C-5 carboxylic acid of IOX1 was hypothesised to be important for active site binding, therefore it might be expected that the ester derivatives would be substantially less potent than IOX1.[22]

To test this proposal, the ability of the compounds to inhibit the H3 K9me3 demethylation activity of isolated KDM4C was tested using an amplified luminescent proximity homogeneous assay (ALPHA) screen optimised for KDMs study.[46]

2.7.1 AlphaScreen technology

The 'Amplified Luminescent Proximity Homogeneous Assay' (ALPHA) screen is a bead-based technology which uses laser irradiation of donor beads at 680 nm to convert oxygen to an excited O_2. When an "acceptor bead" is nearby a "donor bead" (~200 nm) energy transfer occurs, which results in a luminescent signal at 520 to 620 nm.[47]

AlphaScreen protocols can be set up such that a biotinylated analyte is bound to streptavidin donor beads, and an antibody is conjugated to acceptor beads.[48] To measure KDM catalysis, the AlphaScreen assay was adapted to measure demethylation of the biotinylated histone peptide fragments when these are bound to donor beads, and a methylation state specific antibody bound to acceptor beads (**Figure 2.14**).[46]

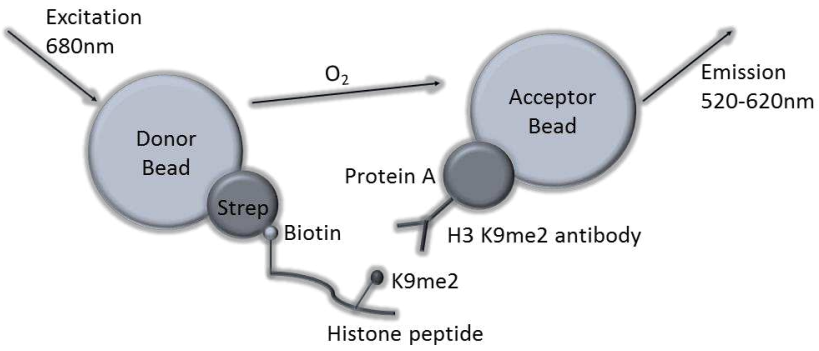


Figure 2.14 Principle of the AlphaScreen assay as adapted to measure KDM4 demethylation activity. Biotinylated-H3 K9me3 peptide is demethylated by KDM4 catalysis to give a biotinylated-H3 K9me2 product. The product peptide is then recognised by a methyl-state specific antibody (anti-H3 K9me2), allowing energy transfer due to proximity between the donor and acceptor beads. Figure adapted with permission from Kawamura *et al.*[46]

2.7.2 Determination of the *in vitro* inhibition of IOX1 **1** and its ester derivatives by AlphaScreen

The inhibitory potency of the unhydrolysed compounds was tested *in vitro* by measuring the catalytic activity of KDM4C histone demethylase using the AlphaScreen method.

KDM4C (5 nM) was pre-incubated with or without inhibitors, followed by incubation with the biotinylated substrate peptide (30 nM), co-factors and co-substrates at room temperature. The reaction was then quenched (EDTA). The AlphaScreen beads were incubated with peptide product and an antibody selective for the product (H3 K9me2, mAb1220, Abcam).

The resultant fluorescence signal was read using an 680 excitation / 570 emission filter, and IC_{50} values of the resultant inhibition curves were calculated using GraphPad Prism 6 after normalisation with DMSO controls.

Increasing the IOX1 concentration showed a dose-response decrease in demethylated H3 peptide, indicating that H3 K9me3 demethylation by KDM4C is inhibited by IOX1 **1**. The IC_{50} value was 0.6 μM, identical to that reported in the literature (**Figure 2.15**; **Table 2.1**).[17] Apart from the bulky di-*tert*-butyl diacetate derivative, **7**, the esters displayed similar activities in the micromolar range, with **5** being the most potent (IC_{50}=3.9 μM).

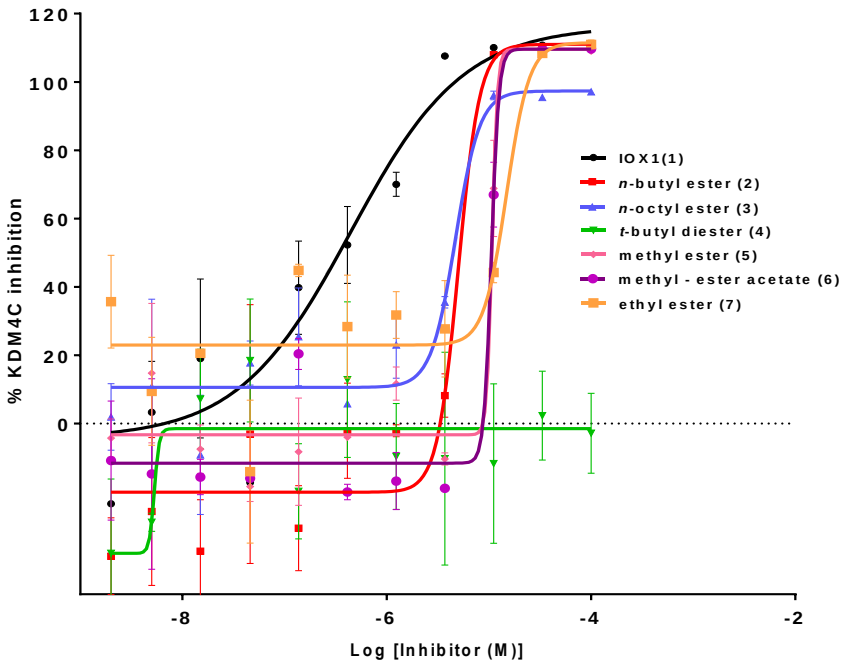


Figure 2.15 KDM4C peptide turnover assayed by AlphaScreen. Evaluation of the *in vitro* inhibition of KDM4C catalytic activity by IOX1 **1** and its ester derivatives. A counter screening with vehicle (DMSO) was used for normalisation. IC_{50} values are listed in **Table 2.1**. All measurements were conducted in quadruplicate and values represent mean and standard deviation. *In vitro activity measurements were carried out by Dr. Anthony Tumber (SGC, Oxford).* Figure reused with permission from Schiller *et al.*[29]

The *n*-octyl ester **5** was shown to be stable to hydrolysis in the AlphaScreen buffer according to LC-MS analysis results (**Figure 2.16**; see Chapter 11, Section 11.3.1, for more information).

The activity of derivative **5** and the other esters, as determined by the AlphaScreen assay, indicates that the C-5 ester derivatisation can be tolerated, while preserving some KDM inhibitory activity. Together with the results from the intracellular delivery assay which show that **5** is mainly present in cells in its unhydrolysed form (see Section 2.6.4), it seems likely that the detected inhibition in cells is due to **5** serving as an inhibitor as wells as a prodrug of IOX1.

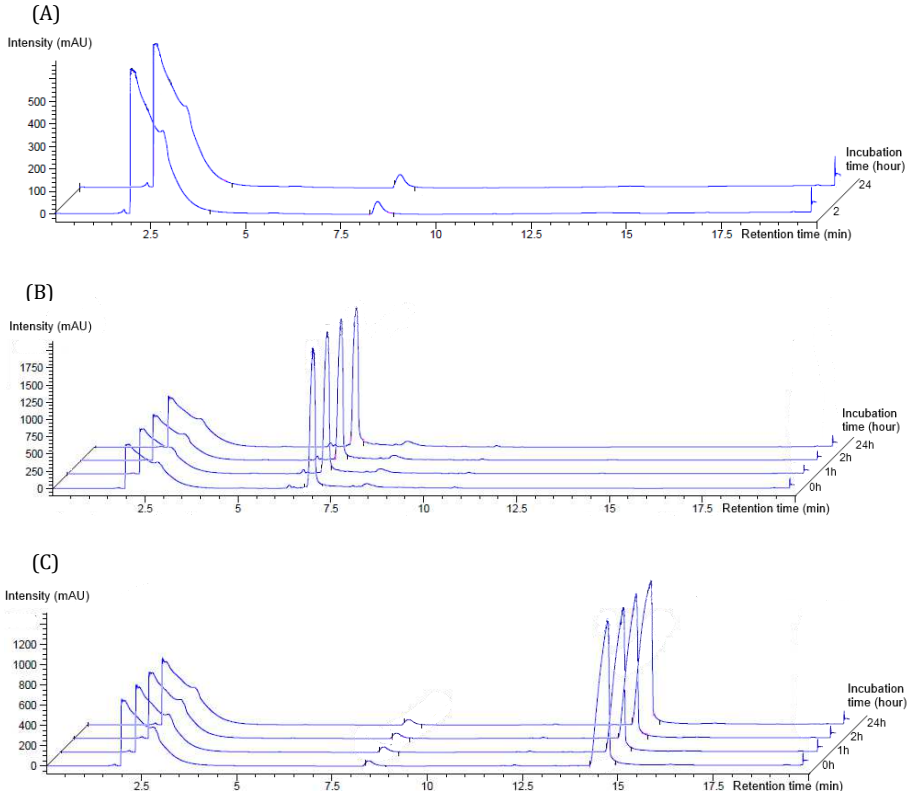
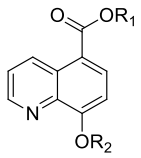
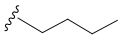
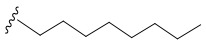
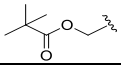
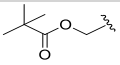


Figure 2.16 Measurements of the stability of IOX1 1 and *n*-octyl ester 5 in the AlphaScreen buffer by LC-MS. mAU: milliabsorbance units. Average retention times for observed species are: (A) AlphaScreen buffer (50 mM HEPES pH 7.5 supplemented with 0.1% BSA, 0.01% Tween20, Fe^{II} (10 μM), ascorbate (100μM) and 2OG (10 μM)), R_t = 2.0 and 8.5 minutes; (B) IOX1 **1** in AlphaScreen buffer, R_t =7.0 minutes; (C) Compound **5** in AlphaScreen buffer, R_t = 14.8 minutes. See Chapter 1, Section 11.3.1, for more information.

Table 2.1 Structure–activity relationships for IOX1 1 and its ester derivatives 2–7. Cell toxicity and activity measurements were conducted in triplicate, *in vitro* activity measurements were conducted in quadruplicate; values represent the mean (see **Figure 2.9**, **Figure 2.10** and **Figure 2.15**). *In vitro activity measurements were carried out by Dr. Anthony Tumber (SGC, Oxford).*

					
Compd	**R^1**	**R^2**	**Cell toxicity** HeLa CC_{50}[a] [μM]	**Activity in cells** HeLa KDM4A EC_{50}[b] [μM]	**Enzyme inhibition** KDM4C IC_{50}[c] [μM]
1	H	H	>300	100.0	0.6
2		H	10	50.0	10.7
3		H	66	>100	14.9
4		H	50	22.0	5.0
5		H	>300	3.8	3.9
6			29	>100	10.5
7			17	ND	>100

[a] CC_{50} values derived from HeLa cell viability assays. Green highlight represents low cytotoxicity, red highlight represents high cytotoxicity. [b] EC_{50} values derived from immunofluorescence assays of KDM4A enzymatic activity in HeLa cells. Green highlight represents high inhibition potency, red highlight represents no effect. [c] IC_{50} values derived from AlphaScreen assays of isolated KDM4C (5 nM). Green highlight represents high inhibition potency, red highlight represents no effect.

2.8 *In vitro* selectivity profiling

n-Octyl ester **5** was found to be the most potent ester derivative of IOX1 in its unhydrolysed form against isolated KDM4C (**Table 2.1**). A more extended set of AlphaScreen assays were then used to compare the selectivity of *n*-octyl ester **5** with that of IOX1 **1** and the shorter ester derivatives (methyl ester **2** and *n*-butyl ester **4**) against additional 2OG oxygenases.

AlphaScreen assays were performed using representatives of different JmjC KDM subfamilies (KDM4C, KDM4E, KDM2A, KDM3A, KDM5C and KDM6B) and the catalytic domain of a HIF prolyl hydroxylase (PHD2). The results support the classification of IOX1 **1**

as a broad-spectrum 2OG oxygenase inhibitor, with IC_{50} values in the micromolar range against all of the tested oxygenases (**Table 2.2**).[17]

Modification of IOX1 to methyl ester **2** gave an apparent relatively non-selective increase in IC_{50} values, all the IC_{50} values are higher compared to IOX1 though the extent of the increase is varied from < 2-fold increase for PHD2 to > 100-fold increase for KDM3A. Increasing the length of the ester alkoxy group to four carbons (as in **4**) created apparent selectivity towards a subset of the JmjC KDMs, and in particular the KDM4 subfamily. Further increasing the length of the alkoxy-group to eight carbons (as in **5**) narrowed the observed inhibitory activity to the KDM4 subfamily. Specifically, KDM4C was the most potently inhibited enzyme by the *n*-octyl ester derivative of IOX1 (**5**).

Table 2.2 *In vitro* selectivity for JmjC subfamilies depending on IOX1 1 ester chain length. IC_{50} values derived from *in vitro* AlphaScreen assays using representative JmjC KDMs (5 nM). Green highlight represents high inhibition potency, red highlight represents no effect. All measurements were conducted in quarduplicate and values represent the mean. *Measurements were carried out by Dr. Anthony Tumber (SGC, Oxford) and Dr. Tzu-Lan Yeh.*

JmjC protein	IOX1 1 IC_{50} [μM]	Methyl ester 2 IC_{50} [μM]	*n*-Butyl ester 4 IC_{50} [μM]	*n*-Octyl ester 5 IC_{50} [μM]
KDM4C	0.6	10.7	5.0	3.9
KDM4E	2.3	12.8	6.3	45.0
KDM2A	1.8	30.1	16.3	>100
KDM3A	0.1	14.5	29.4	>100
KDM5C	19.0	34.9	>100	>100
KDM6B	1.4	10.8	>100	>100
PHD2	33.0	41.1	>100	>100

2.9 Discussion

The 2-oxoglutarate (2OG)-dependent Jumonji C domain (JmjC) family is the largest family of histone lysine demethylases.[7] There is interest in developing small-molecule probes that modulate JmjC enzymatic activity to investigate their biological roles. 5-Carboxy-8-hydroxyquinoline (IOX1) is the most potent broad-spectrum inhibitor of 2OG oxygenases, including the JmjC demethylases, reported to date.[17] The inhibition efficiency of IOX1 in cells is about two orders of magnitude lower in compared to its *in vitro* potency (for KDM4A IC_{50} value of 0.2 μM *in vitro* vs. EC_{50} value of 86 μM in HeLa cells), possibly in part due to low cell permeability resulting from its polar C-5 carboxyl group.[17,18,20] It was hypothesised that the difference between the *in vitro* potency and the cellular potency is due to low transmembrane permeability, likely due to its C-5 carboxylic acid group. Designing ester derivatives as prodrugs is a common method to increase the cell permeability of an inhibitor while keeping the original structure once the prodrug is inside the cell.[24]

The work described in this chapter aimed to improve the cellular permeability of IOX1 using ester derivatisation as potential prodrugs. These studies led to the discovery of an *n*-octyl ester form of IOX1 **5** with enhanced cellular potency compared to IOX1 **1** (~30-fold), substantially improving its usability as a JmjC inhibitor for studies in cells. These findings are supported by a comparison of several IOX1 derivatives using *in vitro* inhibition studies, activity vs. toxicity analysis in cell cultures and mass spectrometric studies measuring intracellular uptake.

In vitro inhibition studies have shown that C-5 ester derivatives of IOX1 can retain JmjC KDM inhibitory activity. Of the tested esters, the *n*-octyl derivative **5** was the most potent *in vitro* against KDM4C. In cells, ester **5** was the least cytotoxic of the tested compounds and the most potent inhibitor of H3 K9me3 demethylation (EC_{50} = 3.8 μM). This is likely to be, at least in part, due to improved cell permeability of **5** compared with that of **1**, as detected in an intracellular delivery assay.

Interestingly, it seems that **5** is not, at least efficiently, hydrolysed in HeLa cells, though esterases are reported to be present and there are reported examples of short-chain ester hydrolysis.[49,50] It was noted however that enzymatic hydrolysis of esters decreases when the chain length is longer than six carbons.[23] Thus, it seems likely that at least some of the cellular activity of *n*-octyl ester **5** results from inhibition by the intact ester form.

An extended AlphaScreen with JmjC KDMs and PHD2 as a prolyl hydroxylase representative indicates that increasing the ester chain length to four carbons improves the selectivity towards JmjC KDMs. Most importantly, a chain length of eight carbons - as in derivative **5** - creates selectivity towards the KDM4 subfamily. The apparent relative selectivity of **5** for the KDM4 subfamily, at least compared with the parent IOX1 **1**, might be due to differences in the active sites of the JmjC proteins. Crystallographic evidence implies that the active site opening of the KDM4 demethylases is larger than that of other JmjC subfamilies, and in particular compared with the narrow binding pocket of the PHD family of hydroxylases.[22,32] Notably, docking predictions based on crystallographic analysis of KDM4A bound to IOX1 support the viability of *n*-octyl ester **5** binding KDM4, with the alkyl group occupying part of a region leading to the active site (*simulations by Dr. Hwanho Choi)*.[29] Trials to verify these predictions by crystallography met with difficulties in obtaining suitable crystallisation conditions (*crystallisation trails by Sarah Madden).*[51] Nonetheless, this initial characterisation suggests that an appropriate substitution of the IOX1 C-5 position could enable the generation of potent and selective JmjC KDM inhibitors that are active in cells.

In contrary to initial expectations, C-5 derivatisation of IOX1 in this work was shown to be tolerated, while preserving some KDM inhibitory activity.[22] Another C-5 derivatisation of IOX1, which led to a potent and selective KDM3A inhibitor, is described in Chapter 3, Section 3.6.

n-Octyl ester **5** could be used as a starting point for the development of new JmjC inhibitors as well as a cell-permeable tool compound in studies investigating the role of JmjC histone demethylases as therapeutic targets.

Notably, some other histone demethylase inhibitors reported in the literature contain an aliphatic chain possibly reflecting a general binding of aliphatic groups in this region.[52–54] The activity of **5** raises the question of whether other available ester prodrugs of JmjC inhibitors, such as NOG, 2,4-PCDA, GSK-J4, Methylstat and 2-hydroxyglutarate for which the cell permeable form commonly used is an octyl ester, could also be active in their ester prodrug forms, and whether systematic ester derivatisation could lead to increased cellular potencies.[43,52,55–57] It is important to note, however, that the results with different 2OG oxygenases reveal that ester derivatisation of IOX1, and possibly of other broad-spectrum KDM inhibitors including the aforementioned compounds, may confer selectivity not apparent in the parent inhibitor.

On a broader aspect, this work emphasises the importance of cell-based assays in search for inhibitors. Cell-based assays inform not only on inhibition potency, but also provide information on transmembrane permeability and toxicity.[58]

This work also highlights the need to take into consideration the biological activity that the prodrug itself or its pro-moieties might have when designing prodrugs for enhanced delivery of a compound of interest.[59] If the prodrug itself is also an inhibitor, as in the case of **5**, the inhibitory profile may change over time depending on the hydrolysis state of the prodrug.

2.10 References

1. Liu, G. *et al.* Genomic amplification and oncogenic properties of the GASC1 histone demethylase gene in breast cancer. *Oncogene* **28,** 4491–4500 (2009).

2. Wei, Y. *et al.* Global H3K4me3 genome mapping reveals alterations of innate immunity signaling and overexpression of JMJD3 in human myelodysplastic syndrome CD34+ cells. *Leukemia* **27,** 2177–86 (2013).

3. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466,** 368–72 (2010).

4. Qi, H. H. *et al.* Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. *Nature* **466,** 503–7 (2010).

5. Tzschach, A. *et al.* Novel JARID1C/SMCX mutations in patients with X-linked mental retardation. *Hum. Mutat.* **27,** 389 (2006).

6. Natoli, G., Testa, G. & De Santa, F. The future therapeutic potential of histone demethylases: A critical analysis. *Curr. Opin. Drug Discov. Devel.* **12,** 607–15 (2009).

7. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* **7,** 715–27 (2006).

8. McDonough, M. A., Loenarz, C., Chowdhury, R., Clifton, I. J. & Schofield, C. J. Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr. Opin. Struct. Biol.* **20,** 659–672 (2010).

9. Aik, W., McDonough, M. A., Thalhammer, A., Chowdhury, R. & Schofield, C. J. Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr. Opin. Struct. Biol.* **22,** 691–700 (2012).

10. Nowak, R. P. *et al.* Advances and challenges in understanding histone demethylase biology. *Curr. Opin. Chem. Biol.* **33,** 151–159 (2016).

11. Thinnes, C. C. *et al.* Targeting histone lysine demethylases — Progress, challenges, and the future. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839,** 1416–1432 (2014).

12. Frye, S. V. The art of the chemical probe. *Nat. Chem. Biol.* **6,** 159–161 (2010).

13. McAllister, T. E. *et al.* Recent Progress in Histone Demethylase Inhibitors. (2016).

14. Rose, N. R., McDonough, M. A., King, O. N. F., Kawamura, A. & Schofield, C. J. Inhibition of 2-oxoglutarate dependent oxygenases. *Chem. Soc. Rev.* **40,** 4364–97 (2011).

15. Mackeen, M. M. *et al.* Small-molecule-based inhibition of histone demethylation in cells assessed by quantitative mass spectrometry. *J. Proteome Res.* **9,** 4082–92 (2010).

16. Rotili, D. *et al.* Pan-histone demethylase inhibitors simultaneously targeting Jumonji C and lysine-specific demethylases display high anticancer activities. *J. Med. Chem.* **57,** 42–55 (2014).

17. Hopkinson, R. J. *et al.* 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem. Sci.* **4,** 3110

(2013).

18. King, O. N. F. *et al.* Quantitative high-throughput screening identifies 8-hydroxyquinolines as cell-active histone demethylase inhibitors. *PLoS One* **5,** e15535 (2010).

19. IOX1: A Cell-Penetrating Broad-Spectrum 2-Oxoglutarate Oxygenase Inhibitor. *SGC* Available at: http://www.thesgc.org/scientists/chemical_probes/IOX1.

20. Rose, N. R. *et al.* Plant growth regulator daminozide is a selective inhibitor of human KDM2/7 histone demethylases. *J. Med. Chem.* **55,** 6639–43 (2012).

21. Albert, A. Chemical Aspects of Selective Toxicity. *Nature* **182,** 421–423 (1958).

22. Mecinović, J., Loenarz, C., Chowdhury, R. & Schofield, C. J. *2-Oxoglutarate analogue inhibitors of prolyl hydroxylase domain 2*. *Bioorganic & Medicinal Chemistry Letters* **19,** (2009).

23. Hu, L. Prodrugs: effective solutions for solubility, permeability and targeting challenges. *IDrugs* **7,** 736–42 (2004).

24. Rautio, J. *et al.* Prodrugs: design and clinical applications. *Nat. Rev. Drug Discov.* **7,** 255–70 (2008).

25. Balant, L. P., Doelker, E. & Buri, P. Prodrugs for the improvement of drug absorption via different routes of administration. *Eur. J. Drug Metab. Pharmacokinet.* **15,** 143–53

26. Taylor, M. D. Improved passive oral drug delivery via prodrugs. *Adv. Drug Deliv. Rev.* **19,** 131–148 (1996).

27. Lew, W., Chen, X. & Kim, C. U. Discovery and development of GS 4104 (oseltamivir): an orally active influenza neuraminidase inhibitor. *Curr. Med. Chem.* **7,** 663–72 (2000).

28. Tabacova, S. A. & Kimmel, C. A. Enalapril: pharmacokinetic/dynamic inferences for comparative developmental toxicity. A review. *Reprod. Toxicol.* **15,** 467–78

29. Schiller, R. *et al.* A cell-permeable ester derivative of the JmjC histone demethylase inhibitor IOX1. *ChemMedChem* **9,** 566–71 (2014).

30. Yale, H. L. & Bernstein, J. The Skraup Reaction with Acrolein and its Derivatives. *J. Am. Chem. Soc.* **70,** 254–254 (1948).

31. Carey, F. A. *Organic chemistry*. (McGraw-Hill, 2006).

32. Rephaeli, A., Zhuk, R. & Nudelman, A. Prodrugs of butyric acid from bench to bedside: Synthetic design, mechanisms of action, and clinical applications. *Drug Dev. Res.* **50,** 379–391 (2000).

33. Nudelman, A., Levovich, I., Cutts, S. M., Phillips, D. R. & Rephaeli, A. The role of intracellularly released formaldehyde and butyric acid in the anticancer activity of acyloxyalkyl esters. *J. Med. Chem.* **48,** 1042–54 (2005).

34. Barltrop, J. A., Owen, T. C., Cory, A. H. & Cory, J. G. 5-(3-carboxymethoxyphenyl)-2-(4,5-dimethylthiazolyl)-3-(4-sulfophenyl)te trazolium, inner salt (MTS) and related analogs of 3-(4,5-dimethylthiazolyl)-2,5-diphenyltetrazolium bromide (MTT)

reducing to purple water-soluble formazans As cell-viability indica. *Bioorganic Med. Chem. Lett.* **1,** 611–614 (1991).

35. Promega Corporation. CellTiter 96 ® AQueous One Solution Cell Proliferation Assay. *Www.Promega.Com/Protocols/* 2014-12–15 (2012).

36. Riss, T. L. *et al. Cell Viability Assays*. *Assay Guidance Manual* (Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2004).

37. Chae, H.-J. *et al.* Molecular mechanism of staurosporine-induced apoptosis in osteoblasts. *Pharmacol. Res.* **42,** 373–381 (2000).

38. Kooistra, S. M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. *Nat. Rev. Mol. Cell Biol.* **13,** 297–311 (2012).

39. Michael P. Balogh. *The Mass Spectrometry Primer (Waters Series)*. (Waters Corporation, 2014).

40. Finehout, E. J. & Lee, K. H. An introduction to mass spectrometry applications in biological research. *Biochem. Mol. Biol. Educ.* **32,** 93–100 (2004).

41. Pitt, J. J. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin. Biochem. Rev.* **30,** 19–34 (2009).

42. Gross, J. H. K.M. Downard—Mass spectrometry: a foundation course. *Anal. Bioanal. Chem.* **383,** 8–9 (2005).

43. Kruidenier, L. *et al.* A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* **488,** 404–8 (2012).

44. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462,** 739–744 (2009).

45. Park, C.-H. *et al.* Determination of the Intracellular Concentrations of Metabolites in Escherichia coli Collected during the Exponential and Stationary Growth Phases using Liquid Chromatography-Mass Spectrometry. *Bull. Korean Chem. Soc.* **32,** 524–530 (2011).

46. Kawamura, A. *et al.* Development of homogeneous luminescence assays for histone demethylase catalysis and binding. *Anal. Biochem.* **404,** 86–93 (2010).

47. Ullman, E. F. *et al.* Luminescent oxygen channeling immunoassay: measurement of particle binding kinetics by chemiluminescence. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 5426–30 (1994).

48. Beaudet, L. *et al.* AlphaLISA immunoassays: the no-wash alternative to ELISAs for research and drug discovery. *Nat. Methods, Publ. online 01 December 2008; | doi10.1038/nmeth.f.230* **5,** (2008).

49. Paul, J. & Fottrell, P. Tissue-specific and species-specific esterases. *Biochem. J.* **78,** 418–24 (1961).

50. Panzarini, E., Tenuzzo, B. & Dini, L. Photodynamic therapy-induced apoptosis of HeLa cells. *Ann. N. Y. Acad. Sci.* **1171,** 617–26 (2009).

51. Madden, S. K. Studies on Histone Demethylases. A Thesis submitted for the Honour

School of Chemistry Part II. (Oxford, 2014).

52. Luo, X. *et al.* A selective inhibitor and probe of the cellular functions of Jumonji C domain-containing histone demethylases. *J. Am. Chem. Soc.* **133,** 9451–6 (2011).

53. Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19,** 17–30 (2011).

54. Suzuki, T. *et al.* Identification of the KDM2/7 Histone Lysine Demethylase Subfamily Inhibitor and its Antiproliferative Activity. *J. Med. Chem.* **56,** 7222–31 (2013).

55. Hamada, S. *et al.* Synthesis and activity of N-oxalylglycine and its derivatives as Jumonji C-domain-containing histone lysine demethylase inhibitors. *Bioorg. Med. Chem. Lett.* **19,** 2852–5 (2009).

56. Thalhammer, A. *et al.* Inhibition of the histone demethylase JMJD2E by 3-substituted pyridine 2,4-dicarboxylates. *Org. Biomol. Chem.* **9,** 127–35 (2011).

57. Chowdhury, R. *et al.* The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* **12,** 463–9 (2011).

58. Martinez, N. J. & Simeonov, A. Cell-based assays to support the profiling of small molecules with histone methyltransferase and demethylase modulatory activity. *Drug Discov. Today Technol.* **18,** 9–17 (2015).

59. Zengeya, T. T., Kulkarni, R. A. & Meier, J. L. Modular Synthesis of Cell-Permeating 2-Ketoglutarate Esters. (2015).

Chapter 3
Biochemical characterisation of KDM3A

3.1 Introduction

KDM3A was first identified in 2006 by Yamane and colleagues as an H3 K9me1/me2 KDM.[1] H3 K9me1/me2 are the only two reported KDM3A substrates, although KDM3A arginine demethylation of an artificial peptide sequence (H3 R9) was recently reported as a reaction catalysed by purified KDM3A (see Chapter 6, Section 6.3).[2,3]

KDM3A is a relatively large JmjC KDM, with a mass of 150 kDa. Although there is no crystal structure for KDM3A, a crystal structure for the JmjC domain of KDM3B in complex with NOG has been determined (PDB ID 4C8D). KDM3A domain architecture comprises a CH2C4 zinc finger domain followed by the Leu-Xaa-Xaa-Leu-Leu (LXXLL) motif, which is reported to mediate associations with nuclear receptors.[2,4] As in all members of the KDM3 subfamily, the JmjC catalytic domain is located in the *C*-terminal region of the overall protein (**Figure 3.1 A**).[5] The construct used in this study, KDM3A$_{515–1317,}$ contains all the assigned domains and motifs in KDM3A (**Figure 3.1 B,** for more information see Materials and methods, Chapter 11, Section 11.2.2).

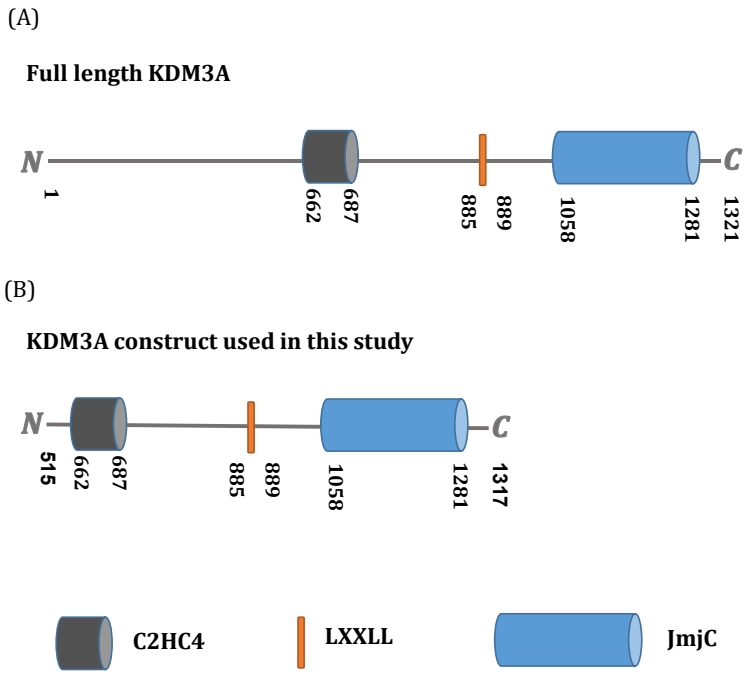


Figure 3.1 Domains and motifs in human KDM3A. (A) Full length KDM3A. (B) The KDM3A construct used in this study. CH2C4 zinc finger domain (grey) followed by the Leu-Xaa-Xaa-Leu-Leu (LXXLL) motif (orange) and the enzymatic JmjC domain at the *C*-terminal region (blue). Both CH2C4 and JmjC are essential for KDM3A demethylation activity.[2] LXXLL motifs are reported to mediate the association of KDM3A with nuclear receptors.[4] Amino acid residue positions refer to that of human KDM3A (UniProtKB - Q9Y4C1).

The zinc finger domain of KDM3A is reported to be essential for its catalytic activity, since truncations or deletions in this region abolish substrate demethylation.[1,2] Kinetic studies by Goda and colleagues, using recombinant enzymes, have shown that this zinc finger domain is involved in homodimerization of KDM3A *in vitro*.[6] In these studies, homodimerization of KDM3A was proposed to take part in a substrate channelling mechanism, in which two active sites cooperatively demethylate H3 K9me2 into the unmodified state by a two-step mechanism.[6] In this model, each domain in each monomer interacts with both the zinc finger and the JmjC domain of the other monomer, hence the requirement for zinc finger for efficient catalysis (**Figure 3.2**).

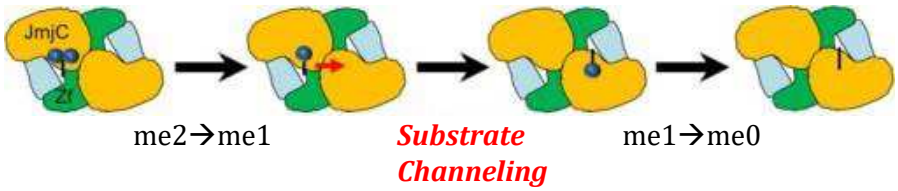


Figure 3.2 Proposed model of two-step demethylation. JmjC catalytic domain (yellow) and zinc finger (Zf) domain (green) are shown for each monomer of the KDM3A homodimer. Methylation of the H3 K9me2 substrate is represented by blue balls. Each site in the KDM3A homodimer has an equal chance of initial binding of the H3 K9me2 substrate. A conformational change occurs as a result of substrate binding. Me2 is converted to me1 by one monomer, and then me1 is converted to me0 by the other monomer. Figure was adapted with permission from Goda *et al.* [6]

Since its discovery in 2006, the physiological functions of KDM3A continue to unravel. To date KDM3A has been reported to be involved in hypoxic response, cancer, metabolism, spermatogenesis and sex determination (see Introduction, Chapter 1, Section 1.14.1, for more information, the potential roles of KDM3A in spermatogenesis is described in Chapter 5).[7–15]

Considering its varied physiological functions, development of selective inhibitors for KDM3A is of interest from both research and clinical perspectives. The small-molecule JIB-04 is a broad-spectrum JmjC KDM inhibitor, and *in vitro* and *in vivo* studies using this inhibitor have enabled the identification of a novel non-catalytic function for KDM3A (**Figure 3.3**).[16,17]

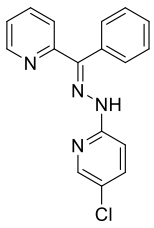


Figure 3.3 JIB-04 is a broad-spectrum JmjC KDM inhibitor.[16,17]

KDM3A is also inhibited by broad-spectrum 2OG analogue inhibitors such as *N*-oxalylglycine (NOG), which is an amide analogue of 2OG, its mimetic 2,4-pyridinedicarboxylic acid (2,4-PDCA) and 5-carboxy-8-hydroxyquinoline (IOX1).[18] Among these inhibitors IOX1 was shown to be the most potent against KDM3A *in vitro* (IC_{50} value

of 0.17 μM by AlphaScreen assays, see Chapter 2, Section 2.7.2).[19] To date, there is no selective inhibitor for the KDM3 subfamily, an appropriate derivatisation of IOX1 might lead to increased selectivity towards these enzymes, similar to the findings described in Chapter 2.

3.2 Objectives

The work in this Chapter describes the expression, purification and characterisation of KDM3A, a JmjC histone demethylase with varied physiological functions. MALDI-TOF MS and FDH coupled assays were utilised for the kinetic characterisation of KDM3A. Kinetic measurements for KDM3A substrates (H3 K9me2 and H3 K9me1) and co-substrate (2OG) are described. Studies investigating the effect of H3 K4me3 modification on the kinetics of KDM3A H3 K9 demethylation are also included. Preliminary characterisation of a KDM3A selective inhibitor was carried out.

*KDM3A recombinant baculovirus expression and purification were carried out under the guidance of Dr. James Dunford and Carina Giladi, respectively (Oppermann group, Oxford). pFB-CTHF10-LIC transfer vector was provided by Dr. Nicola Burgess-Brown, SGC. Formaldehyde standard curve for FDH was generated by Dr. Akane Kawamura. Compound **8** was provided by Prof. David Maloney (NCGC, NIH), compounds **9** and **10** were synthesised and tested by MALDI-TOF MS by Pauline Lang. AlphaScreen selectivity measurements were carried out by Dr. Anthony Tumber (SGC, Oxford).*

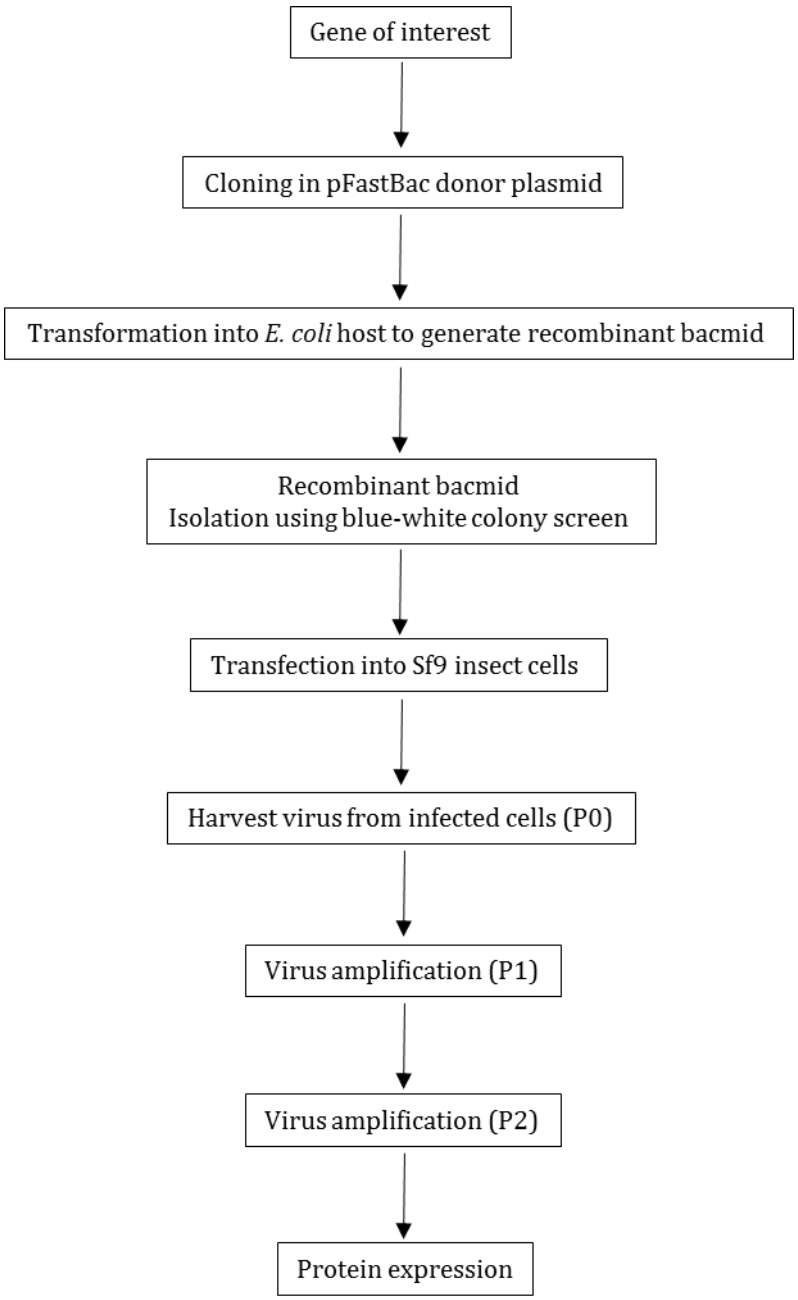
3.3 Expression of KDM3A

KDM3A requires both the catalytic JmjC and the CH2C4 zinc finger domains for its enzymatic activity.[2] For recombinant expression of a construct which accommodates both domains, the baculovirus system was chosen due to its suitability for the expression of large human genes.[20] The baculovirus is an insect-specific virus, which is used as an expression vector. In this expression system, a target gene is inserted into a nonessential genomic region of the virus via homologous recombination with a transfer vector containing the target gene.[21] The recombinant baculovirus is then used to infect cultured insect cells, resulting in the production of the target recombinant protein.

Baculovirus expression offers a number of advantages over other systems such as *E. coli* and mammalian expression. First and foremost, it is suited for the expression of large genes, which is essential for active recombinant KDM3A.[20] Second, since expression is performed in eukaryotic cells baculovirus expression ensures proper protein folding and post-translational modifications.[22] In terms of practicality, baculovirus is essentially non-pathogenic to mammals and therefore very safe to use.[23] Also, it is easy to scale up protein production using baculovirus and get high levels of recombinant gene expression.[24] The main drawback for producing recombinant protein in baculovirus is that the procedure is laborious and technically demanding. In fact, it takes a minimum of 19 days in order to get initial expression.[25] Homologous recombination of the target gene into the virus might require a few trials and the infected insect cells must be kept under optimal conditions to allow recombinant protein production.

$KDM3A_{515\text{-}1317}$ was produced as an *N*-terminally His_6-tagged protein in Sf9 insect cells using the Bac-to-Bac® baculovirus expression system (Invitrogen) following the procedure described by Rose *et al*.[26] The first step included cloning of DNA encoding for the catalytic domain of KDM3A, comprising residues T515 through S1317, into the pFastBac donor plasmid, which is the pFB-CTHF10-LIC transfer vector (**Scheme 3.1**, *provided by Dr. Nicola Burgess-Brown, SGC*).[23] The resultant plasmid was then transformed into DH10bac *E.* coli host strain. This strain contains a baculovirus shuttle vector (bacmid) into which the target gene is transformed. The resultant recombinant bacmid was isolated using blue – white colony screening.[23] The bacmid was then used to transfect healthy Sf9 insect cells. Infected cells were identified by light microscope, loss of the round cell shape compared to the null controls indicated virus infection (**Figure 3.4**). Recombinant virus was harvested from infected cells and was used to generate a high titer baculovirus stock in three

amplification cycles of cells infection (P0, P1, P2). P2 virus was used to generate protein at a large scale expression over 72 hours (see Materials and methods, Chapter 11, Section 11.2.3, for more information).



Scheme 3.1 Experimental outline for gene expression using the Bac-to-Bac® baculovirus expression system (Invitrogen).[23] See Materials and methods, Chapter 11, Section 11.2.3, for more information.

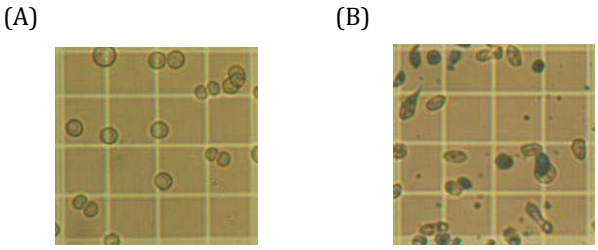


Figure 3.4 Identification of Sf9 insect cells infected with baculovirus using a light microscope. (A) Healthy cells. (B) Infected cells. 300X magnification. Figure reused with permission.[25]

3.4 Purification of KDM3A

Recombinant KDM3A$_{515\text{-}1317}$ protein is unstable and tends to degrade according to previous results by the SGC, Oxford (*Dr. Catrine Johansson, manuscript in preparation*). For optimal purity and catalytic activity, the protein was kept at 4°C during purification and the procedure be completed within a day.

The purification was adapted from the procedure described by Rose *et al*.[26] Sf9 insect cells were harvested and lysed by freeze thawing (50 mM HEPES-KOH pH 7.5, 300 mM KCl, 10 mM imidazole and 5% glycerol, cOmplete™ EDTA - free Protease Inhibitor Cocktail Sigma-Aldrich, 0.2 mg/mL DNAse I; Roche) and the homogenate (**H**) sample was collected for analysis. The lysate was clarified by high speed centrifugation (56000xg for 45 minutes), and the pellet (**P**) and supernatant (**S**) samples were collected. The protein was then purified by Ni-Sepharose (GE Healthcare) chromatography using batch binding. For protein binding, beads were incubated with lysate with slow rotation for 60 min. The flow-through (**FT**) was removed and beads were loaded into an empty column and washed using 20CV of Wash 1 (**W1**) buffer (50 mM HEPES-KOH pH 7.5, 300 mM KCl, 5% glycerol and 10 mM imidazole), followed by 4CV of Wash 2 (**W2**) buffer (30 mM imidazole). Protein was eluted with 5CV of elution buffer (250mM imidazole). Fractions containing protein were pooled and purified by gel filtration on a HiLoad 16/60 Superdex 200 column (GE Healthcare) equilibrated with 20 mM HEPES-KOH pH 7.5, 300 mM KCl and 5% glycerol (**Figure 3.5**). Fractions B5 and C1-C2 from gel filtration were chosen for KDM3A assays. They were pooled and concentrated to 63 μM, aliquoted, then rapidly frozen in liquid nitrogen and then stored for long term at -80°C (see Materials and methods, Chapter 11, Section 11.2.3).

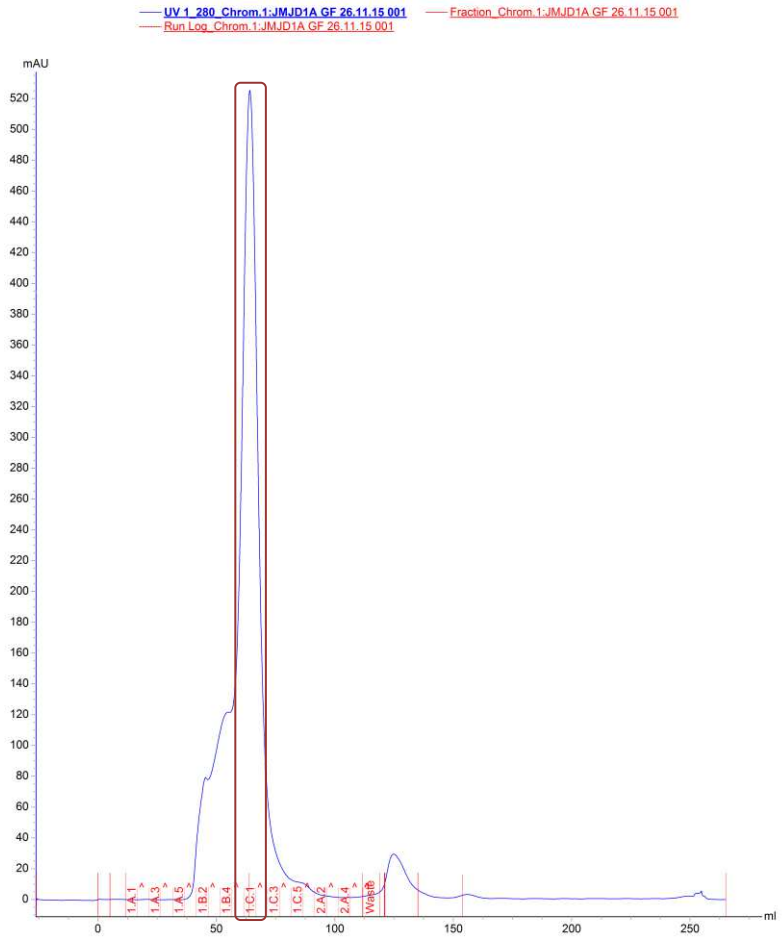


Figure 3.5 Gel filtration chromatogram of KDM3A purification. Gel filtration on a HiLoad 16/60 Superdex 200 column (GE Healthcare) equilibrated with 20 mM HEPES-KOH pH 7.5, 300 mM KCl and 5% glycerol. See Materials and methods, Chapter 11, Section 11.2.3, for more information. mAU signal (280 nM) vs. elution volume (mL). Letter and a number in red indicates fraction name. Eluted fractions are in frame (B5, C1, C2).

Samples from the different purification steps and the eluted protein were analysed using mass spectrometry, Coomassie staining and western blot analyses (**Figure 3.6** and **Figure 3.7**).

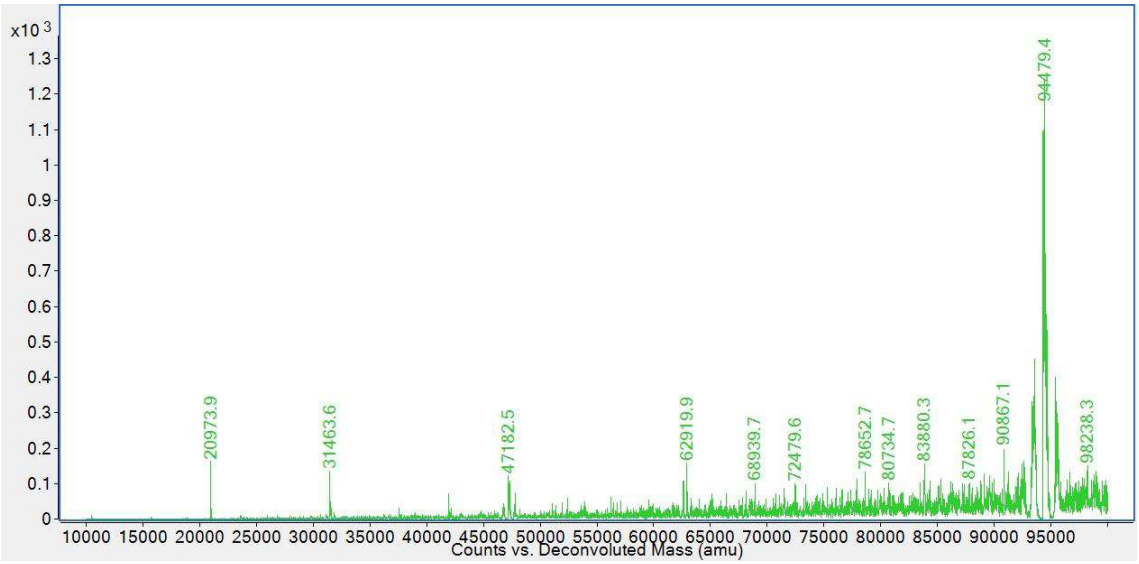


Figure 3.6 Mass spectrometric analysis of the eluted KDM3A. The calculated mass of KDM3A$_{515-1317}$ is 94.4 KDa. The main peak observed in this spectrum is 94.5 kDa, confirming that the eluted protein is KDM3A. *Mass spectrometric analysis was carried out by Carina Giladi (SGC, Oxford).*

The expected band size for the KDM3A$_{515-1317}$ construct by SDS-PAGE analysisis ~95 kDa. However, the observed band size was higher than 100 kDa. The observed band size was also higher than the predicted size for a full length KDM3A.[27] KDM3A protein degradation was also observed in this purification, in agreement with previous work, as can be seen from Coomassie staining and western blot analyses (**Figure 3.7**). Bands from the SDS-PAGE gel corresponding to KDM3A and the lower molecular bands (55 – 70 kDa) were analysed by Bruker HCT ion trap. MASCOT search results confirmed that these bands contain KDM3A, consistent with the western blot data (see **Appendix 1**, *analysis by Carina Giladi and Dr. Rod Chalk, Burgess-Brown group, SGC, Oxford)*. These results validate that the low molecular weight bands are indeed degradation products of KDM3A.

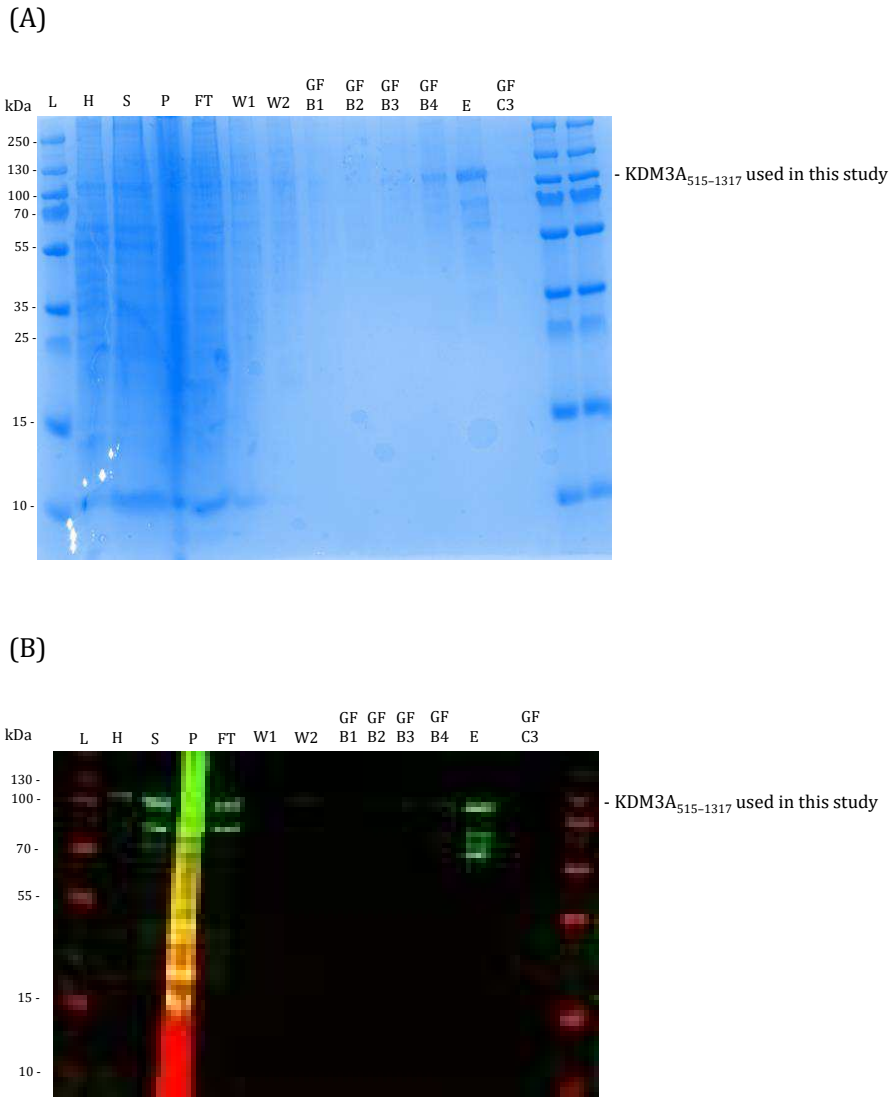


Figure 3.7 Samples collected during KDM3A purification. (A) Coomassie staining of SDS-PAGE protein gel. (B) Fluorescent western blot of SDS-PAGE protein gel using 6x-His epitope tag antibody as a primary antibody (Life Technologies, 710286, 1:1000, shown in green) and a Rabbit IgG as a secondary antibody (Life Technologies, 35568, 1:20,000). L denotes the ladder, H the homogenate, S the supernatant, P the pellet, FT the flow-through, W1 the wash 1, W2 the wash 2, GF the gel filtration fractions with a letter + number indicate the fraction name loaded in each lane and E the eluted protein (pooled from fractions B5, C1, C2). Numbers on the left of the gels present the molecular weight of each standard in KDa. Homogenate sample was 1:100 diluted, all other samples were 1:10 diluted. See Materials and methods, Chapter 11, Section 11.2.3, for more information.

3.5 Kinetic characterisation of KDM3A

In order to design assays to measure IC_{50} values of inhibitors, and to explore the catalytic efficiency for new substrates, knowledge of the steady-state kinetic parameters (K_M, k_{cat}) of KDM3A is required. KDM3A demethylation reactions were analysed using the Michaelis-Menten model to determine the apparent K_M and k_{cat} for the 2OG co-substrate and for histone fragment peptides. Apparent kinetic parameters for 2OG and K9-methylated H3 (H3(1-15)K9me2) peptide were determined by both matrix-assisted laser desorption/ionisation-time of flight mass spectrometry (MALDI-TOF MS) assay and a formaldehyde dehydrogenase (FDH) coupled assay. Kinetic data were compared between the two methods and with data from previous studies. Kinetic characterisation of additional substrate peptides was carried out using the FDH coupled assay only.

3.5.1 Kinetic characterisation by MALDI-TOF MS assay

Matrix-assisted laser desorption/ionisation-time of flight mass spectrometry (MALDI-TOF MS) assay can be used to analyse changes in the mass of peptide substrates.[28] In Matrix Assisted Laser Desorption/Ionization (MALDI), the sample is mixed with a saturating amount of matrix that absorbs UV (337 nm) and transfers it to heat. The surface of the analyte heats rapidly and evaporates. In Time of Flight Mass Spectrometry (TOF MS), charged ions are generated on the sample slide and accelerated by an electric field of a fixed potential difference. Due to the law of conservation of energy, the time of flight for the ions differs according to their mass-to-charge ratio (m/z), and is utilised for mass spectrometric characterisation.[29]

In MALDI-TOF MS activity assay, the enzyme is incubated with varying concentrations of the substrate or the co-substrate together with appropriate co-factors. The demethylation reaction is then stopped at different time points for mass spectrometric analyses. Time courses are then used to determine the initial rates of the reaction. This assay provides direct substrate to product ratio with high sensitivity and accuracy. However, it is a discontinuous assay which makes the results more prone to errors.

Conditions for KDM3A MALDI-TOF MS activity assay were adapted based on procedures for KDM3A activity assays reported by Rose *et al.* and Goda *et al.*[6,26] All reaction components were dissolved in 50 mM HEPES-KOH, pH 7.5 and standard co-factors concentrations are described in **Table 3.1**. Enzyme and substrate mixes were prepared

separately and combined to initiate the demethylation reaction, which was carried out at 37°C. Reactions were quenched with 1% formic acid, prior to spotting onto a MALDI-TOF MS plate and co-crystallisation with the MALDI-TOF MS matrix, alpha-cyano-4-hydroxycinnamic acid, CHCA. MALDI-TOF MS analysis in a positive ionisation mode was used to determine the extent of demethylation of substrate peptides by KDM3A. The percentage of enzymatic activity was calculated using **Equation 3.1,** and then normalised based on the no enzyme control reactions (see Materials and methods, Chapter 11, Section 11.3.3, for more information).

Table 3.1 Standard concentrations of components in KDM3A MALDI-TOF MS demethylation assay. The following conditions were used to measure KDM3A catalytic activity using MALDI-TOF MS unless otherwise stated. Assays were carried out in 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information).

	Component	Concentration in reaction
Enzyme mix	KDM3A	varied (single-digit μM)
	Fe^{II}	50 μM
	Ascorbate	100 μM
Substrate mix	Peptide	varied (5-500 μM)
	2OG	100 μM

$$\%\ Catalytic\ activity = \frac{\sum_{i=0}^{n} P}{\sum_{i=0}^{n} P + S}\ x\ 100$$

Equation 3.1 Calculation of the percentage of enzymatic activity based on MALDI-TOF MS analyses. P is the relative intensity of the peak of the product peptide and S is the relative intensity of the peak of the remaining substrate peptide in the sample. In cases where several products were observed, their sum of intensities is used to calculate the percentage of catalytic activity.

3.5.1.1 Time courses for KDM3A demethylation activity

MALDI-TOF MS activity assays were utilised to measure the extent of demethylation of the dimethylated substrate (H3(1-15)K9me2) as a function of time. Time courses were then used to determine the initial rates of the reaction for Michaelis-Menten kinetics.

Assays were carried out under the standard reaction conditions (see Section 3.5.1), with 1 μM KDM3A and varying H3(1-15)K9me2 peptide concentrations (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Percentage demethylation was calculated according to **Equation 3.1**. **Figure 3.8** presents example time courses of 10 and 100 μM peptide.

Demethylation by KDM3A reached saturation after five minutes, however the reaction was not complete and just over 50% of the substrate was catalysed into product. Interestingly, throughout the reaction the level of the intermediate substrate – me1- stayed constant and close to zero. These results are consistent with findings by Goda *et al.*[6] and supports their model for a processive mechanism, in which consecutive demethylation reactions occur without releasing the substrate.

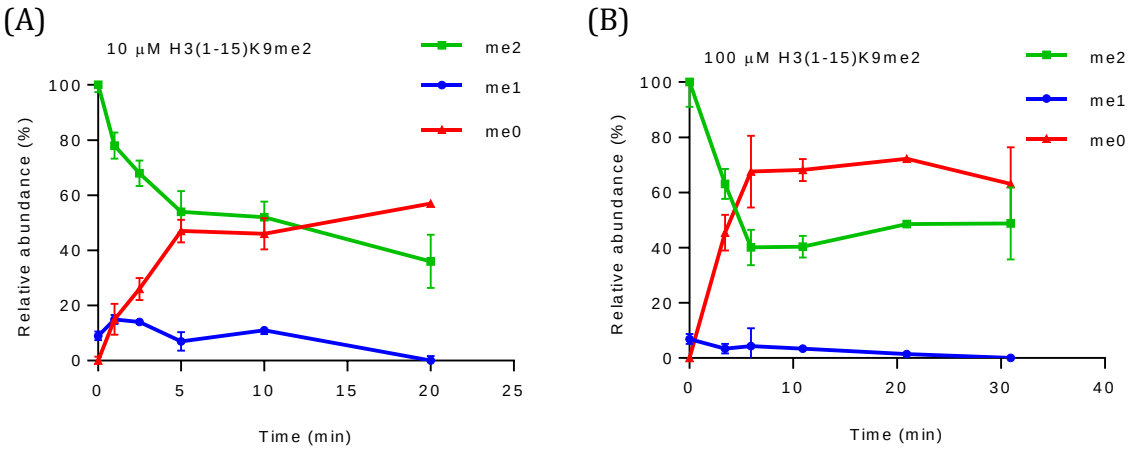


Figure 3.8 MALDI-TOF MS analyses of KDM3A demethylation activity using H3(1-15)K9me2 peptide as a substrate. Assays were carried out under the standard conditions described in Section 3.5.1 with 1 μM KDM3A. Percentage demethylation was calculated according to **Equation 3.1** (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) 10 μM substrate. (B) 100 μM substrate. All measurements were conducted in triplicate and values represent mean and standard deviation.

3.5.1.2 Determination of 2OG kinetic parameters for KDM3A

MALDI-TOF MS assays were used to determine the Michaelis-Menten constants for the co-substrate 2OG. To reduce the variability of the data, all measurements were performed simultaneously on a 96-well plate (see Materials and methods, Chapter 11, Section 11.3.3, for more information). 2OG concentrations were varied from 0 to 1000 μM, while the H3(1-15)K9me2 peptide concentration was kept in excess (100 μM, see Section 3.5.2) to ensure it is not a limiting factor for the initial reaction rate. Initial rates were plotted against 2OG concentration and fitted using the Michaelis-Menten equation using GraphPad Prism 6 to give an apparent K_M value of 22 ± 10 μM and k_{cat} value of 701 ± 39 10^{-3} s^{-1} (**Equation 3.2 i** in Section 3.5.2.3, **Figure 3.9** and **Table 3.2**).[31] The 2OG K_M value observed in this experiment is slightly higher than the reported value of 2OG K_M (5 ± 2 μM), where kinetic determination was carried out using the FDH coupled assay (same enzyme construct and peptide substrate were used in both experiments).[26] Therefore it was envisaged that additional measurements are required for the kinetic characterisation of KDM3A.

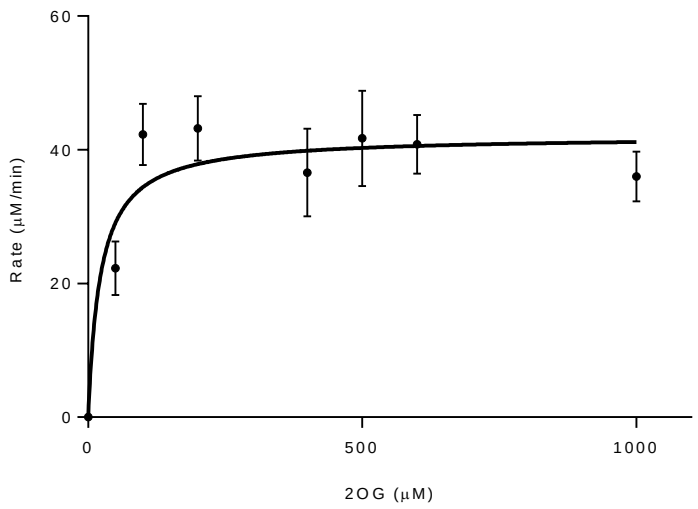


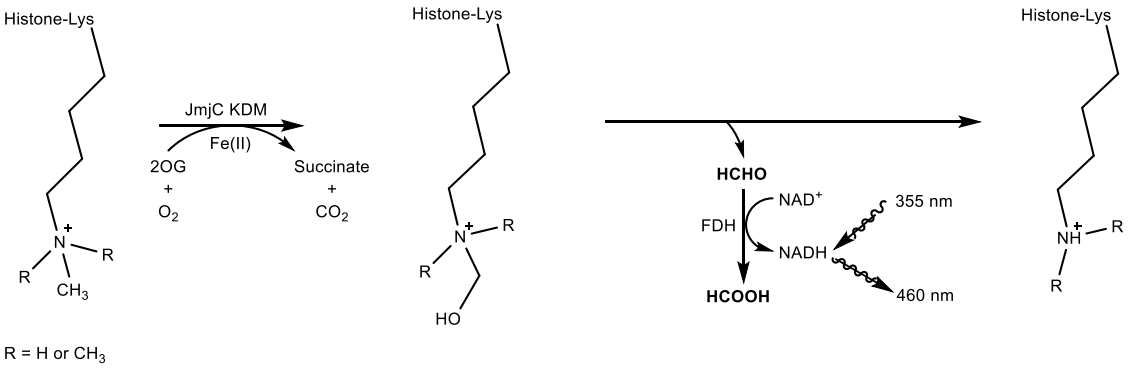
Figure 3.9 Kinetic characterisation of 2OG KDM3A using MALDI-TOF MS assay. Assays were carried out under the standard conditions described in Section 3.5.1**.** Initial velocity vs. 2OG concentration data were fitted with Michaelis–Menten curve using GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.3, for more information). All measurements were conducted in triplicate and values represent mean and standard deviation.

Table 3.2 Apparent kinetic parameters for KDM3A determined by MALDI-TOF MS assays. Assays were carried out under the standard conditions described in Section 3.5.1. Apparent kinetic parameters were calculated by the substrate inhibition model using GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.3, for more information). All measurements were conducted in triplicate and values represent mean and s.e.m. (n=3).

Substrate	Enzyme (μM)	Model used	K_M (μM)	V_{max} (μM/min)	k_{cat} ($\times 10^{-3}$ s^{-1})	k_{cat}/K_M (10^{-6} / s μM)
2OG	1.00	Michaelis-Menten	22 ± 10	42 ± 2	701 ± 39	31450

3.5.2 Kinetic characterisation by the FDH coupled assay

As an orthogonal screen to retest the apparent kinetic parameters determined by MALDI-TOF MS assay, kinetic measurements were performed using the formaldehyde dehydrogenase (FDH) coupled assay. The FDH coupled assay was first used by Lizcano *et al*. for methylamine oxidase (MAO) kinetic measurements.[33] This assay was later adapted to quantify formaldehyde release during KDM demethylation reaction by using formaldehyde as a substrate for a coupled enzyme reaction.[34] The coupled reaction is catalysed by the FDH enzyme, which oxidises formaldehyde to formic acid, with a concomitant reduction of NAD^+ to NADH (**Scheme 3.2**). NADH production is then detected by fluorescence spectroscopy using excitation at 355 nm and emission at 460 nm.[35]



Scheme 3.2 Principle of formaldehyde dehydrogenase (FDH)-coupled KDM assay. Formaldehyde produced by demethylase activity is coupled to oxidation to formic acid by formaldehyde dehydrogenase with the concomitant reduction of nicotinamide adenine dinucleotide (NAD^+) to NADH. Demethylase activity is measured by utilising NADH fluorescence (excitation at 355 nm and emission at 460 nm).

The main advantage of the FDH coupled assay over the MALDI-TOF MS assay is that it allows for real-time measurements of the reaction's progress by continuously monitoring it over time, making this method especially suitable for kinetic measurements. Therefore, there is no need for manual quenching of the reaction and the method is less prone to errors. Another advantage is the low enzyme concentrations required for measurements by the FDH coupled assay compared with the MALDI-TOF MS assay. The drawback of this method is that monitoring KDM catalysis is dependent on the performance of another enzyme, FDH. Therefore, saturating amounts of FDH and NAD^+ are used so that the second reaction is not rate limiting for the demethylation reaction. In addition, when screening for KDM inhibitors a preliminary screen to test for FDH inhibition is required.

The assay conditions were chosen based on procedures for KDM3A activity assays reported by Rose *et al.* and Goda *et al.*[6,26] Reaction components were dissolved in 50 mM HEPES-KOH, pH 7.5 and standard concentrations are described in **Table 3.3**. Enzyme and substrate mixes were prepared separately and combined to initiate the demethylation reaction, which was carried out at 37°C. Production of NADH was fluorescently monitored and readings were taken over 60 cycles with a cycle time of 30 seconds. A formaldehyde standard curve was used to convert fluorescence to demethylated product concentration (see Section 3.5.2.1). The linear range for the KDM reaction was used to calculate initial rates, which were plotted against substrate concentrations using GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.4, for more information).

Table 3.3 Standard concentration of components in the KDM3A FDH coupled assay. The following conditions were used to measure KDM3A catalytic activity using the FDH coupled assay throughout the study unless otherwise stated. See Materials and methods, Chapter 11, Section 11.3.4, for more information.

	Component	Concentration in reaction
Enzyme mix	KDM3A	varied (hundredths nM)*
	Fe^{II}	50 µM
	Ascorbate	100 µM
	FDH	0.5 U/50 µl reaction
Substrate mix	Peptide	varied (5-500 µM)
	2OG	100 µM
	NAD^+	500 µM

** See Section 3.5.2.1*

3.5.2.1 Assay calibration for measurements of KDM3A catalysis

To convert fluorescence to product concentration, a formaldehyde standard curve was made. Reaction mixture was prepared under standard conditions, as described in Section 3.5.2, omitting KDM3A while introducing formaldehyde at varying concentrations (0 – 160 μM) to the substrate mixture. Fluorescence levels due to FDH catalysis were plotted against formaldehyde concentrations, and the slope of the linear range was used as a rate of conversion between the relative fluorescence units RFU and product concentration (**Figure 3.10**).

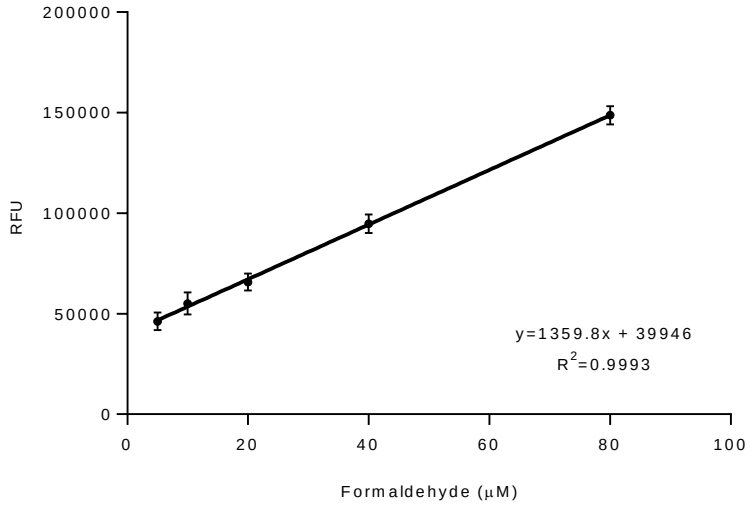


Figure 3.10 The linear range in formaldehyde standard curve for FDH enzyme. All measurements were conducted in triplicate and values represent mean and standard deviation. *Measurements were carried out by Dr. Akane Kawamura.*

A preliminary assay to select a suitable KDM3A concentration was carried out for each new substrate or KDM3A purification batch. KDM3A concentration was calibrated such that the fluorescence signal in the FDH coupled assay is significantly higher than the background noise while the linear range of the reaction is not too short.

The reaction mixture was prepared under standard conditions, as described in Section 3.5.2, with 25 μM H3(1-15)K9me2 peptide (apparent K_M, see Section 3.5.1.2) and varying concentrations of KDM3A (0 – 500 nM); the fluorescence was plotted against time.

Figure 3.11 presents an example for an assay to determine a suitable KDM3A concentration prior to FDH coupled activity measurements. The reaction mixture without KDM3A had a flat fluorescence signal over time, at the range of 60000 relative fluorescence units (RFU). This pilot reaction with 100 nM KDM3A did not reach saturation

over 30 minutes, while reactions with 250 and 500 nM KDM3A both reached saturation with a linear range of 9 and 2.5 minutes respectively. A concentration of 250 nM was chosen for following experiments as the linear range is long enough and the difference between the reaction slope and the no enzyme control measurement is clear (although lower KDM3A concentrations between 100-250 nM could have also been used). The described calibration procedure to find a suitable KDM3A concentration was performed prior to any measurement of new KDM3A substrates, as well as after each production of a new batch of recombinant KDM3A protein.

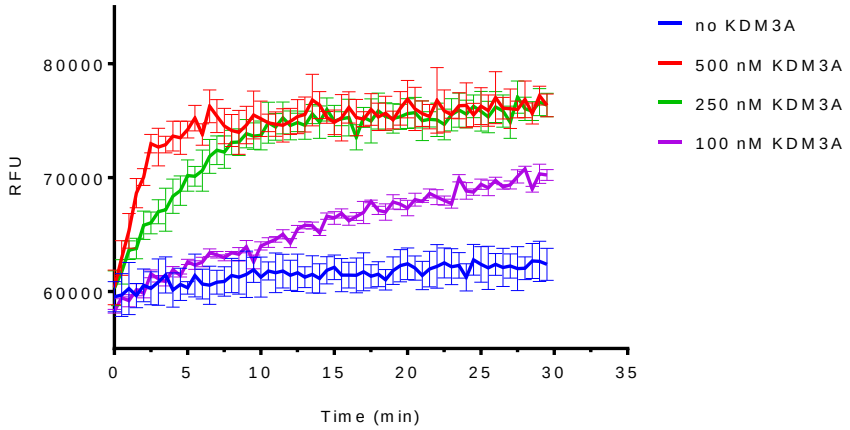


Figure 3.11 Calibration of KDM3A concentration for the FDH coupled assays. Reaction mixture was prepared under standard conditions, with varying concentrations of KDM3A (0 – 500 nM). Fluorescence (RFU) was plotted against time (see Materials and methods, Chapter 11, Section 11.3.4, for more information). All measurements were conducted in triplicate and values represent mean and s.e.m.

3.5.2.2 Determination of 2OG kinetic parameters for KDM3A

The FDH coupled assay was used as an orthogonal screen to re-determine the K_M and k_{cat} values for KDM3A with the co-substrate 2OG. FDH coupled assays were carried out under the standard conditions as described in Section 3.5.2. The H3(1-15)K9me2 substrate peptide was used in excess to ensure it was not limiting the reaction rate, yet below inhibiting concentrations (100 µM, see Section 3.5.1.2). Initial rates were plotted against 2OG concentration and the data were fitted to the Michaelis-Menten equation (**Equation 3.2 i**) using GraphPad Prism 6 to give an apparent 2OG K_M value of 30 ± 10 µM and an apparent k_{cat} value of 161 ± 8 10^{-3} s^{-1} (**Figure 3.12, Table 3.4**).

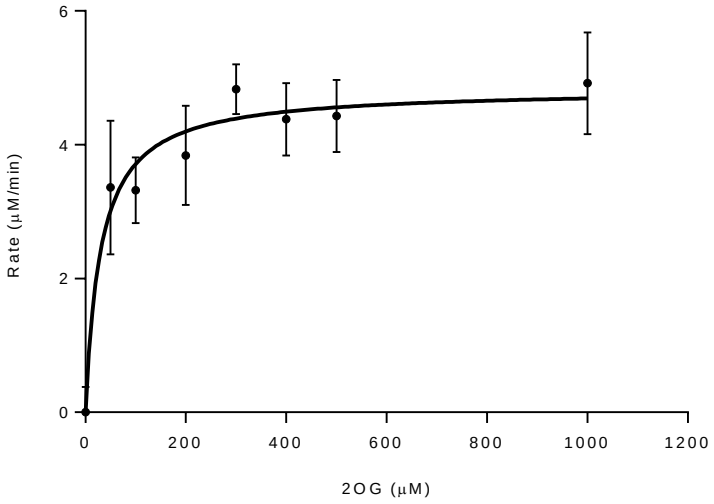


Figure 3.12 KDM3A catalysis of 2OG as determined by the FDH coupled assay. Assays were carried out under the standard conditions as described in Section 3.5.2.1. Michaelis-Menten plot of initial velocity vs. concentration data were fitted using GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.4, for more information). All measurements were conducted in triplicate and values represent mean and standard deviation.

Table 3.4 Apparent kinetic parameters for 2OG determined by FDH coupled assays. Assays were carried out under the standard conditions described in Section 3.5.2. Apparent kinetic parameters were calculated by the substrate inhibition model using GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.4, for more information). All measurements were conducted in triplicate and values represent mean and s.e.m. (n=3).

Substrate	**Enzyme (μM)**	**Model used**	**K_M (μM)**	**V_{max} (μM/min)**	**k_{cat} ($\times 10^{-3}$ s^{-1})**	**k_{cat} / K_M (10^{-6} / s μM)**
2OG	0.50	Michaelis-Menten	30 ± 10	4.8 ± 0.2	161 ± 8	5325

The apparent K_M value of 2OG closely resembles kinetic data from the MALDI-TOF MS assay (30 ± 10 μM by the FDH coupled assay vs. 22 ± 10 μM by the MALDI-TOF MS assay). Notably, both of these values are higher than the reported 2OG K_M value by Rose *et al.* (5 ± 2 μM), which was also determined by the FDH coupled assay using the same enzyme construct, expressed and purified using the same procedure, and peptide substrate.[26] On the other hand, the apparent k_{cat} determined by the FDH coupled assay is over four-fold lower than the value determined by MALDI-TOF MS (161 ± 8 10^{-3} s^{-1} vs. 701 ± 39 10^{-3} s^{-1} respectively). Since the FDH coupled assay is less prone to technical error due to

automated real-time measurements, the apparent k_{cat} determined by this assay is likely to be closer to the actual k_{cat} value than the value determined by MALDI-TOF MS.

3.5.2.3 Determination of H3(1-15)K9me2 kinetic parameters

FDH coupled assays were used to determine the kinetic parameters (K_M and k_{cat}) for KDM3A with the substrate peptide H3(1-15)K9me2. The assays were carried out under the standard conditions as described in Section 3.5.2. Increasing concentrations of the H3(1-15)K9me2 peptide (0 - 300 μM) were used, while 2OG was kept in excess to ensure it was not limiting the reaction rate (> three-fold K_M, 100 μM). Initial rates were plotted against peptide concentration and the data were fitted using GraphPad Prism 6.

Substrate inhibition by the H3(1-15)K9me2 peptide was observed at elevated peptide concentrations (>160 μM, **Figure 3.13**). Substrate inhibition usually occurs at high substrate concentrations were the substrate starts to bind a second site on the enzyme and thus blocks its catalytic activity.[30] In the case of KDM3A, the substrate can bind the second unit of the dimer and block substrate channelling (see Introduction Section 3.1). Due to the observed substrate inhibition at concentrations above 160 μM, the K_M and k_{cat} values were calculated using a substrate inhibition model (**Table 3.5** in Section 3.5.2.4). In this model, the Michaelis-Menten equation (**Equation 3.2 i**) is adjusted to include K_i, the dissociation constant for inhibitor binding (**Equation 3.2 ii**).[31,32]

(i)

$$v = \frac{v_{max}\,[S]}{K_{\mathrm{M}} + [S]}$$

(ii)

$$v = \frac{v_{max}\,[S]}{K_{\mathrm{M}} + [S] * (1 + \frac{[S]}{K_i})}$$

Equation 3.2 The Michaelis-Menten and the substrate inhibition equations. (i) The Michaelis-Menten equation, showing the relation between the substrate concentration and the initial reaction rate under steady-state conditions. (ii) The substrate inhibition equation. K_M is the Michaelis-Menten constant (the substrate concentration at which the reaction velocity is 50% of the V_{max}), [S] denotes substrate concentration and V_{max} the maximal enzyme velocity. K_i is the dissociation constant for the substrate (the concentration required to produce half maximum inhibition).[31,32]

The apparent K_M value was calculated to be 54 ± 13 μM and the apparent k_{cat} value was 790 ± 112 10^{-3} s^{-1}. Notably, previous studies by Goda *et al.* did not identify substrate inhibition by H3(1-15)K9me2 for KDM3A: the apparent K_M value was higher (106 ± 12 μM) and the apparent k_{cat} value was lower (275 ± 12 10^{-3} s^{-1}).[6] Data were obtained by the FDH coupled assay using a full length KDM3A (1-1321). The measurements by Goda *et al.* were taken every 3 minutes, while the linear range according to their report lasted 9 minutes.[6] Therefore, calculation of the initial rates used only three data-points which is likely to introduce errors. In the FDH coupled assays described herein, measurements were taken every 30 seconds to have a significant number of points for initial rate calculation. Another difference between the current study and that of Goda *et al.* is the expression and purification methods: Goda *et al.* expressed full length KDM3A in 293T cells and then used immunoprecipitation. Differences in construct and enzyme purification may also account, at least in part, for the observed differences.

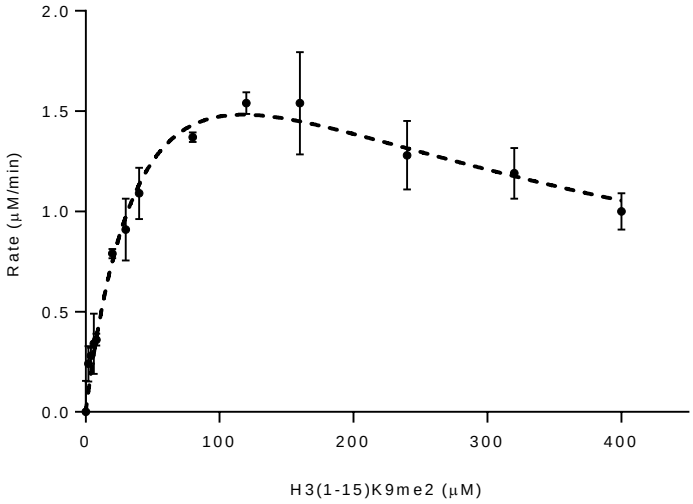


Figure 3.13 KDM3A catalysis of H3(1-15)K9me2. Assays were carried out under the standard conditions described in Section 3.5.2.1 for FDH coupled assay. Initial velocity vs. concentration data were fitted with substrate inhibition curve using GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.4, for more information). All measurements were conducted in triplicate and values represent mean and standard deviation.

3.5.2.4 Investigating the effect of H3 K4me3 modification on H3 K9me2 catalysis by KDM3A

Like KDM3A, KDM7B (PHF8) and KIAA1718 are KDMs that catalyse demethylation of H3 K9me2/me1 substrates.[36] Both have been reported to have modified enzymatic activity in the presence of H3 K4me3.[37] Both KDM7B and KIAA1718 bind H3 K4me3 via their plant homeodomain (PHD). However, the presence of the PTM on the substrate peptides promotes the catalytic activity of KDM7B, but diminishes it for KIAA1718.[37] This difference is proposed to be due to different effects on the enzyme conformation.[37] The H3 K4me3 modification has also been observed to be important in KDM3A catalysis with potentially novel substrates (oxidation of acetyllysines, as described in Chapter 7, oxidation of formyllysines, as described in Chapter 9). Although KDM3A does not have a PHD finger like KDM7B and KIAA1718, it might harbour a different cryptic methyl-reader like motif which interacts with H3 K4me3. Interestingly, Kuroki *et al*. have reported that ChIP analyses have reproducibly found correlation between H3 K4 methylation / H3 K9 demethylation in KDM3A target genes.[15] To test whether H3 K4me3 changes the catalytic efficiency of KDM3A for its H3 K9me2 substrate, kinetic analysis using H3(1-15)K4me3K9me2 peptide was carried out for comparison with the kinetic parameters of H3(1-15)K9me2.

The FDH coupled assays were used for determination of the apparent kinetic parameters of H3(1-15)K4me3K9me2 peptide. The assays were performed under the standard conditions described in Section 3.5.2 (see Materials and methods, Chapter 11, Section 11.3.4, for more information).

Comparison between the H3(1-15)K9me2 and H3(1-15)K4me3K9me2 substrates shows that, while there is an increase in rate with increasing substrate concentrations at lower substrate concentrations, at high substrate concentrations, the rate is decreased in with both peptides (**Figure 3.13 and Figure 3.14**). This observation implies that the reaction does not follow Michaelis Menten kinetics but rather has a more complex kinetic profile. In this analysis, data were analysed using a substrate inhibition model (**Table 3.5**); however one cannot rule out product inhibition – or any other complex kinetic models - and further detailed kinetic analysis (e.g. kinetics in the presence of differing concentrations of product in the case of product inhibition model), will be needed to identify the optimal kinetic model. Inhibition was more pronounced for H3(1-15)K4me3K9me2, as apparent by a lower K_i value (59 ± 17 vs. 256 ± 72 for H3(1-15)K9me2) with inhibition starting at a lower concentration (> 80 μM vs. > 160 μM for H3(1-15)K9me2).

Using the substrate inhibition model, the K_i of the H3(1-15)K4me3K9me2 peptide is much lower than that of the H3(1-15)K9me2 peptide, indicating more potent inhibition of KDM3A in the presence of K4 methylation. The K_M values for the H3(1-15)K9me2 and H3(1-15)K4me3K9me2 peptides are similar (54 ± 13 μM and 68 ± 20 μM, respectively). However, the k_{cat} value of the H3(1-15)K4me3K9me2 is more than double than the k_{cat} for H3(1-15)K9me2; this may reflect a faster turnover rate for KDM3A H3 K9me2 demethylation in the presence of H3 K4me3 (790 ± 112 10^{-3} s^{-1} vs. 2591 ± 570 10^{-3} s^{-1}, respectively). As a result, KDM3A catalytic efficiency, i.e. the k_{cat}/K_M ratio, is more than doubled in the presence of H3 K4me3 (38330 ×10^{-6} s^{-1} μM^{-1} for H3(1-15)K4me3K9me2 vs. 14649 ×10^{-6} s^{-1} μM^{-1} for H3(1-15)K9me2).

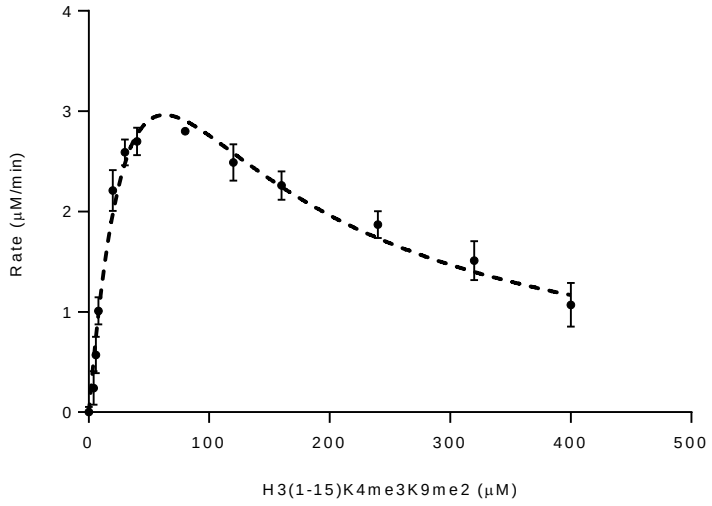


Figure 3.14 KDM3A catalysis of H3(1-15)K4me3K9me2. Measurements were carried out under the standard conditions described in Section 3.5.2.1 for FDH coupled assay. Initial velocity vs. concentration data were fitted with substrate inhibition curve using GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.4, for more information). All measurements were conducted in triplicate and values represent mean and standard deviation.

Table 3.5 Comparison of apparent kinetic parameters between H3(1-15)K9me2 peptide substrates with and without H3 K4me3 modification. Kinetic analyses using the FDH coupled assay. Measurements were carried out under the standard conditions described in Section 3.5.2.1. Apparent kinetic parameters were calculated by fitting the substrate inhibition model using GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.4, for more information). All measurements were conducted in triplicate and values represent mean and s.e.m. (n=3).

Substrate	Enzyme (µM)	Model used	K_i (µM)	K_M (µM)	V_{max} (µM/min)	k_{cat} ($\times 10^{-3}$ s^{-1})	k_{cat}/K_M (10^{-6} / s µM)
H3(1-15)K9me2	0.06	Substrate inhibition	256 ± 72	54 ± 13	2.8 ± 0.4	790 ± 112	14649
H3(1-15)K4me3K9me2	0.06	Substrate inhibition	59 ± 17	68 ± 20	9 ± 2	2591 ± 570	38330

3.5.2.5 Preliminary analysis of H3(1-15)K9me1 kinetics

Kinetic characterisation of an additional reported KDM3A substrate, H3 K9 monomethyllysine, was carried out using the H3(1-15)K9me1 fragment peptide.[1] It was envisaged that determination of the kinetic parameters for this substrate would enable the comparison of the affinity, turnover rate and substrate inhibition to the corresponding dimethylated substrate peptide. The FDH coupled assay was chosen as the preferred methods for measurement as it allows for continuous real-time measurements over the course of the reaction. The assays were carried out under the standard conditions described in Section 3.5.2 (see Materials and methods, Chapter 11, Section 11.3.4, for more information). 2OG was used in excess to ensure it did not limit the reaction rate (> three-fold K_M, 100 μM).

Similarly to the H3(1-15)K9me2 kinetic characterisation, substrate inhibition of KDM3A by the H3(1-15)K9me1 substrate peptide was observed (**Figure 3.15 A**). Inhibition by H3(1-15)K9me1 was more pronounced than inhibition by H3(1-15)K9me2, starting at a concentration of 50 μM compared to H3(1-15)K9me2 where inhibition was detected only at concentrations over 160 μM (**Figure 3.13**). Previous kinetic studies of this substrate did not identify substrate inhibition although the same peptide sequence was used.[6]

Data from the H3(1-15)K9me1 assays were fitted using the substrate inhibition model; however, the error was large (**Table 3.6**), likely due to insufficient number of data-points in this experiment. In order to get preliminary prediction for apparent K_M and k_{cat} values, the Lineweaver–Burk plot was used to calculate the kinetic parameters based on data-points below the substrate inhibition range (> 50 μM, **Figure 3.15 B**); the ratio of k_{cat}/K_M was then calculated as a measure of the catalytic efficiency for the substrate.

The preliminary apparent K_M value of H3(1-15)K9me1 was 9 ± 5 μM, which is lower than the apparent K_M value of H3(1-15)K9me2 (54 ± 13 μM**, Table 3.5**). A previous study reported a much higher K_M value of H3(1-15)K9me1 (95 ± 11 μM), this discrepancy could be due to differences in the number of measurements taken for determination of the initial rates as discussed in Section 3.5.2.3, and it could also be due to an insufficient number of data-points used for analysis.[6] The k_{cat} value of H3(1-15)K9me1 was 306 ± 92 μM. Interestingly, the apparent catalytic efficiency for H3(1-15)K9me1 KDM3A as indicated by the k_{cat}/K_M value was double than that of the catalytic efficiency for H3(1-15)K9me2 (29428 10^{-6} s^{-1} μM^{-1} vs. 14649 10^{-6} s^{-1} μM^{-1}).

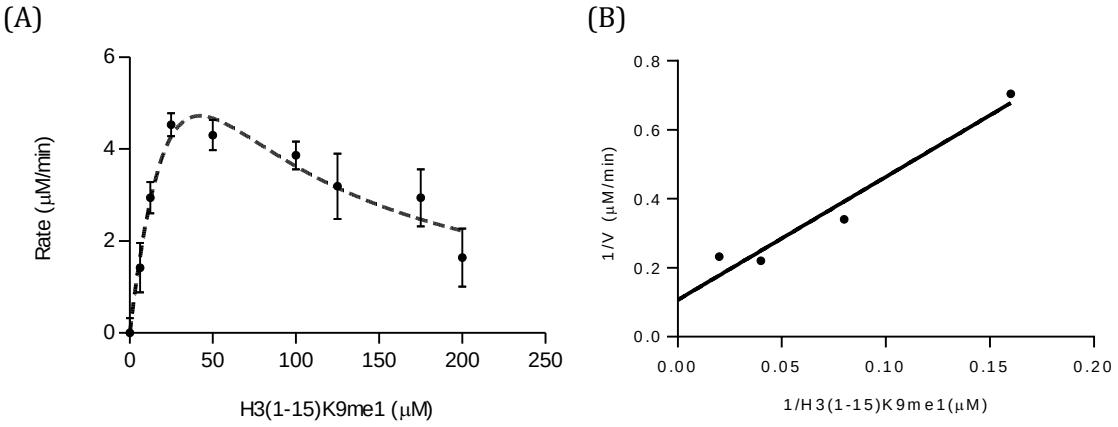


Figure 3.15 Michaelis-Menten plots for KDM3A measuring initial rates of reactions with H3(1-15)K9me1 using the FDH coupled assay. Assays were carried out under the standard conditions described in Section 3.5.2.1 for FDH coupled assay. Initial velocity vs. concentration data were fitted using GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.4, for more information). (A) Dotted line, fitted using a substrate inhibition curve. (B) Lineweaver–Burk plot analysing only the substrate concentration ranges with the assumption that the kinetics obeys the standard Michaelis-Menten in this concentration range (< 50 μM). All measurements were conducted in triplicate and values represent mean and standard deviation.

Table 3.6 Preliminary prediction of apparent kinetic parameters for H3(1-15)K9me1. Assays were carried out by FDH coupled assays under the standard conditions described in Section 3.5.2.1. Apparent kinetic parameters were calculated by fitting either the substrate inhibition model or by a Lineweaver–Burk plot for concentrations below the substrate inhibition range (< 50 μM) using GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.4, for more information). Measurements were conducted in technical triplicate.

Enzyme (μM)	Model used	K_M (μM)	V_{max} (μM/min)	k_{cat} ($\times 10^{-3}$ s^{-1})	k_{cat}/K_M (10^{-6} /s μM)
0.25	Lineweaver–Burk (Michaelis-Menten)	14 ± 4	6.0 ± 0.7	398 ± 45	29428
	Substrate inhibition	67 ± 51	20 ± 12	1312 ± 803	19644

3.6 Preliminary characterisation of a KDM3A inhibitor

KDM3A is inhibited by broad-spectrum 2OG analogue inhibitors (see Section 3.1).[19] IOX1, a broad-spectrum inhibitor, was shown to be potent against KDM3A (IC_{50} value of 0.17 μM using AlphaScreen assays), however, its polar C-5 carboxyl group might reduce transmembrane permeability.[19,38] To date, there is no selective inhibitor for KDM3A. An appropriate derivatisation of IOX1 might lead to increased selectivity towards this enzyme, similar to findings described in Chapter 2, where increased selectivity towards the KDM4 subfamily was observed with increasing the ester chain length at the C-5 position.

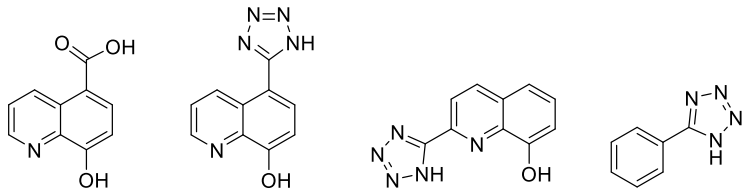
In previous work by Dr. Anthony Tumber (SGC, Oxford), IOX1 derivatives (*provided by Prof. David Malony (NCATS, NIH))* were screened as potential KDM3A inhibitors using AlphaScreen assays. In these screens, 5-(1*H*-tetrazol-5-yl)quinoline-8-ol (**8**) was identified as the lead compound for KDM3A with an IC_{50} value of 0.29 μM (**Figure 3.16**, *Dr. Anthony Tumber* (SGC, Oxford), *manuscript in preparation*).

8 is an IOX1 derivative in which the C-5 carboxylic acid is replaced by the commonly used bioisosteric tetrazole ring. A tetrazole ring is a good substitute for carboxylic acid as the acidity between the structures is comparable (pKa = ~4.5–4.9).[39] One advantage to using a tetrazole over a carboxylic acid is that the anion is more lipophilic than the corresponding carboxylate, which may aid in increasing transmembrane permeability.[40] However a tetrazole is slightly larger than a carboxylic acid, therefore the active site might need to expand to accommodate it. A cross screen for the tetrazole bioisostere **8** over 2OG dependent oxygenase representatives using AlphaScreen revealed an apparent selectivity for the tetrazole-bioisostere towards KDM3A over the tested 2OG dependent oxygenases (**Table 3.7**). This selectivity was not apparent in AlphaScreen cross screen for IOX1 (**1**, see Chapter 2, Section 2.7.2).

Table 3.7 Cross screen for compound 8 over 2OG dependent oxygenases. Enzyme concentration was 5 nM for all tested enzymes. Selectivity measurements were carried out using AlphaScreen. *Experiments were carried out by Dr. Anthony Tumber (SGC, Oxford).*

KDM3A IC_{50} [μM]	KDM4C IC_{50} [μM]	KDM4E IC_{50} [μM]	PHD2 IC_{50} [μM]	KDM2A IC_{50} [μM]	KDM5C IC_{50} [μM]	FIH IC_{50} [μM]
0.29	4.3	3	23.6	34% inhibition at 20 μM	43% inhibition at 20 μM	No inhibition

Following this work, an additional IOX1 **1** derivative was synthesised, in which the tetrazole group is located on the C-2 position instead on of the C-5 position (**compound 9**, **Figure 3.16**). In addition, a compound which lacks the 8HQ structure, and thereby should not chelate Fe^{II}, was used as a negative control in KDM3A inhibition assays (**compound 10**).



1 **8** **9** **10**

Figure 3.16 IOX1 and IOX1 derivatives to be tested for inhibition of KDM3A. 1 – IOX1, **8 -** tetrazole bioisostere of IOX1, **9** – an isomer of **8** and **10** – a compound lacking the 8HQ structure. ***8*** *was provided by Professor David Maloney (NCATS, NIH),* ***9*** *and* ***10*** *were synthesised by Pauline Lang.*

3.6.1 Measuring the efficiency of the FDH coupled assay in the presence of IOX1 derivatives

Prior to the IC_{50} analyses of IOX1 **1** and its derivatives **8-10**, the efficiency of the FDH coupled assay in the presence of these compounds was tested. Since the FDH assay is an enzyme-coupled assay, which is dependent on the catalytic activity of FDH, there is a possibility that inhibition by these small-molecule inhibitors may be the result of inhibition of FDH itself.

To ensure that the inhibitory effect reflects inhibition against KDM3A, IOX1 **1** and its derivatives **8-10** were pre-screened for FDH inhibition at 300 μM. The FDH concentration was chosen to be 50 μM as this lies within the linear range of the enzymatic activity for FDH (see Section 3.5.2.1). Co-factor and co-substrate concentrations were kept at the standard conditions (see Section 3.5.2). Catalytic activity in all complement assays was compared with the no inhibitor control and no FDH control experiments.

Measuring the initial rate of FDH activity indicated that most of the tested compounds substantially inhibited FDH, at least at high concentrations, with IOX1 **1** and the tetrazole bioisostere **8** being the most potent (over 80% inhibition, **Figure 3.17**). Therefore, the FDH coupled assay was deemed as invalid for IC_{50} determination of these compounds.

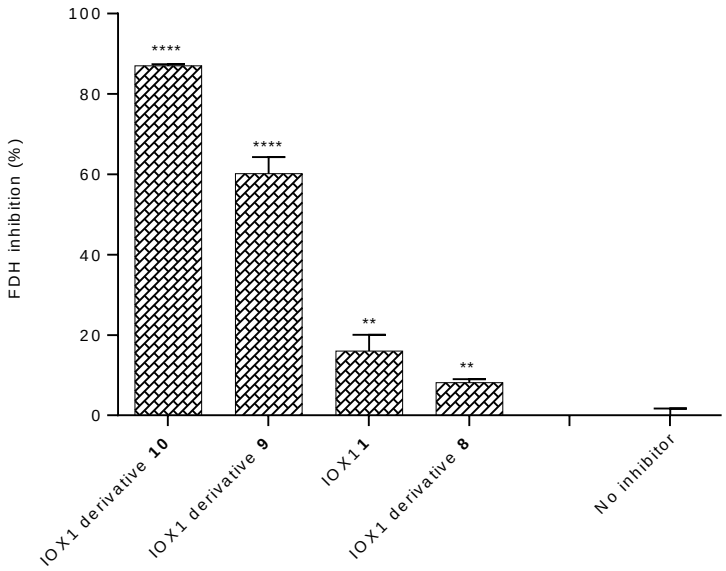


Figure 3.17 FDH catalytic activity in the presence of compounds to be evaluated as KDM3A inhibitors. Compounds were screened at 300 μM. Formaldehyde concentration was 50 μM. FDH catalytic activity was normalised according no inhibitor and no enzyme controls. All measurements were conducted in triplicate and values represent mean and standard deviation. P-values were obtained from a two-tailed t-test compared with no inhibitor control using GraphPad Prism 6. ** denotes $P \leq 0.01$ and **** denotes $P \leq 0.0001$.

3.6.2 IC_{50} values determination by MALDI-TOF MS assay

Since the FDH coupled assay was found to be unsuitable for IC_{50} measurements, MALDI-TOF MS assay was used as an alternative direct method to test for inhibition.

MALDI-TOF MS assays were carried out under the standard conditions described in Section 3.5.1. A preliminary assay to determine the linear range of the enzymatic activity was conducted, where the linear range was observed to be 5 minutes. To test whether the compounds inhibit KDM3A-catalysed demethylation, inhibition assays were carried out with IOX1 **1** and its derivatives **8-10** with varying concentrations (0 – 1000 μM). 2OG concentration was kept at the apparent K_M value (30 μM), while the peptide was used in excess to ensure it was not limiting the initial reaction rate. The reaction was quenched at different time points over the linear range and analysed by MALDI-TOF MS. Initial rates and IC_{50} values were calculated using GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.3, or more information).

As expected, compound **10**, which lacks the 8HQ structure**,** did not inhibit KDM3A demethylation (**Figure 3.18** and **Table 3.8**). IOX1 **1** was the most potent inhibitor with an IC_{50} value of 14 μM, while the tetrazole bioisostere **8** displays a slightly weaker inhibition at the micromolar range (IC_{50} value of 26 μM). Compound **9** seems to inhibit KDM3A catalysis only at 1000 μM.

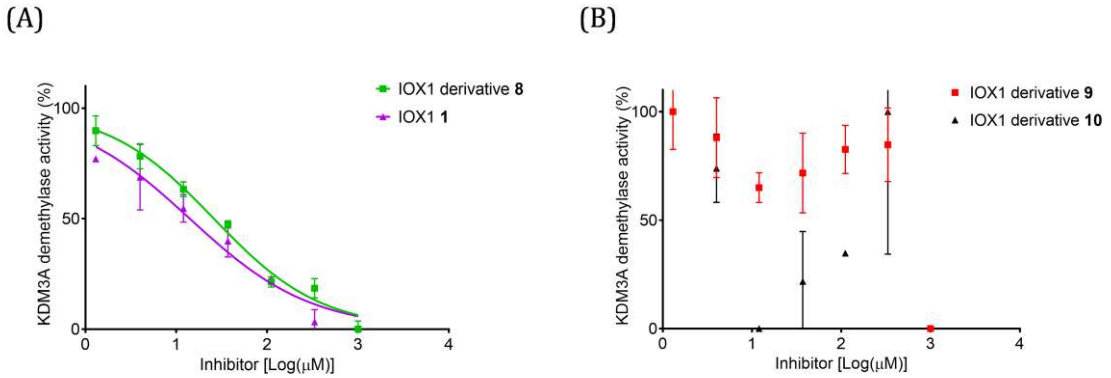


Figure 3.18 IC_{50} determination assayed by MALDI-TOF MS assays for IOX1 and derivative compounds tested against KDM3A. Evaluation of the *in vitro* inhibition of KDM3A catalytic activity by IOX1 and analogues. Compound **10** was used as a negative control. All measurements were conducted in triplicate and values represent mean and standard deviation (*measurements of derivatives **9** and **10** were carried out by Pauline Lang under my supervision*).

Table 3.8 IC_{50} values for IOX1 and derivative compounds determined by KDM3A inhibition curves. Enzyme, peptide and 2OG concentrations were 1 μM, 100 μM and 30 μM respectively.

Compound	IC_{50} (μM)	95% Confidence Interval (μM)
IOX1 1	14	9 to 23
8	26	18 to 37
9	> 1000 μM	-
10	> 1000 μM	-

3.7 Discussion

The work described in this chapter describes the initial biochemical characterisation of KDM3A. The work started with the production of an active KDM3A recombinant enzyme using a baculovirus expression system. This system was chosen due to its compatibility with expression of large genes, thus enabling expression of both the catalytic JmjC and the CH2C4 zinc finger domains of KDM3A which are necessary for its enzymatic activity (**Figure 3.1**).[2] Purification was carried out at 4°C, using affinity chromatography immediately followed by gel filtration to reduce protein degradation.

Initial kinetic characterisation of the isolated KDM3A enzyme was then carried out, with a focus on the steady-state kinetic parameters (K_M, k_{cat}) for the co-substrate 2OG and the methylated lysine substrates (using H3 fragment peptides). This characterisation was then used as the basis for a comparison between the catalytic efficiency of KDM3A for new substrates, and also for the determination of the saturating concentration to be used in inhibition studies.

The initial approach for kinetic characterisation was to use a MALDI-TOF MS activity assay. Time course measurements of the lysine demethylation reaction have identified low and constant levels of the mono-methylated intermediate. This data are consistent with the processive mechanism suggested by Goda *et al.*, in which KDM3A catalyses consecutive demethylation reactions without releasing its substrate.[6]

Kinetic measurements by MALDI-TOF MS activity assays is labour intense, with manual quenching of reactions needed at each time-point and can introduce error. Therefore, an

orthogonal assay was used to determine the steady-state kinetic parameters using the FDH coupled activity assay, in which measurements are taken continuously in real-time.

Notably, substrate inhibition at elevated substrate concentrations was observed, although it had not been reported in a previous study.[6] Therefore, the kinetic parameters were calculated using the substrate inhibition model described in **Equation 3.2 ii**. Inhibition by the monomethylated substrate (H3(1-15)K9me1) was observed at lower concentrations than by the dimethylated substrate (50 μM vs. 160 μM respectively). Preliminary prediction of the kinetic parameters for this substrate was carried out using the Lineweaver–Burk plot by calculating the kinetic parameters based on data-points below the substrate inhibition range (> 50 μM); Interestingly, the apparent catalytic efficiency for H3(1-15)K9me1 KDM3A as indicated by the k_{cat} /K_M value was double than that of the catalytic efficiency for H3(1-15)K9me2 (29428 10^{-6} s^{-1} μM^{-1} vs. 14649 10^{-6} s^{-1} μM^{-1}). It is important to note that product inhibition and/or substrate inhibition needs to be further investigated because both types of inhibition are not distinguishable with the current data. Moreover, it should be noted that experiments were carried out with fragment peptides of the natural substrates and therefore the kinetic characterisations is likely to vary from that of the full proteins.

The presence of H3 K4me3 was observed to be crucial in the KDM3A catalysed hydroxylation catalysis of novel substrates (see Chapter 7, Section 7.4.2 and Chapter 9, Section 9.4), although the enzyme lacks a PHD finger present in other H3 K9me2/1 KDMs (KDM7B and KIAA1718).[37] Kinetic analyses were carried out to test the effect of H3 K4me3 on lysine demethylation by KDM3A. The results revealed that an additional H3 K4me3 mark does not change the H3 K9me2 K_M value for KDM3A, but substrate inhibition appeared to be enhanced (**Figure 3.13, Figure 3.14** and **Table 3.5**). However, the turnover rate for catalysis is more than doubled for H3 K9me2 peptide with an additional H3 K4me3 mark relative to the corresponding peptide without this mark. Therefore, it might be possible that the turnover rate is increased in the presence of the H3 K4me3 mark due to faster product formation and release.

Following kinetic characterisations, studies then focused on preliminary characterisation of a selective inhibitor for KDM3A, a tetrazole bioisoster (**8**) of the broad-spectrum inhibitor IOX1 (**1**). Notably, IOX1 and its tested derivatives were observed to inhibit the FDH enzyme and therefore IC_{50} measurements could not be determined using the FDH coupled assay. These results emphasise the need of performing control FDH inhibition tests prior to any inhibition studies by the FDH coupled assay, which is a commonly used

method for measuring inhibitor potencies for KDMs. MALDI-TOF MS activity assays suggest that **8** is an inhibitor for KDM3A with potency in the micromolar range (**Figure 3.18**). AlphaScreen measurements indicate that **8** might exhibit selectivity for KDM3A over other 2OG dependent oxygenases (**Table 3.7**). **8** was observed to be technically challenging to synthesise (*data not shown*), however this preliminary data require further exploration. Testing the selectivity of the compound with additional 2OG enzymes is desirable. The potency of compound **8** in cells should also be tested, as well as its potency for the inhibition of novel reactions catalysed by KDM3A.

3.8 References

1. Yamane, K. *et al.* JHDM2A, a JmjC-Containing H3K9 Demethylase, Facilitates Transcription Activation by Androgen Receptor. *Cell* **125,** 483–495 (2006).

2. Brauchle, M. *et al.* Protein complex interactor analysis and differential activity of KDM3 subfamily members towards H3K9 methylation. *PLoS One* **8,** e60549 (2013).

3. Walport, L. J. *et al.* Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat. Commun.* **7,** 11974 (2016).

4. Plevin, M. J., Mills, M. M. & Ikura, M. The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends Biochem. Sci.* **30,** 66–9 (2005).

5. Klose, R. J. & Zhang, Y. Regulation of histone methylation by demethylimination and demethylation. *Nat. Rev. Mol. Cell Biol.* **8,** 307–318 (2007).

6. Goda, S., Isagawa, T., Chikaoka, Y., Kawamura, T. & Aburatani, H. Control of histone H3 lysine 9 (H3K9) methylation state via cooperative two-step demethylation by Jumonji domain containing 1A (JMJD1A) homodimer. *J. Biol. Chem.* **288,** 36948–56 (2013).

7. Schofield, Christopher JHausinger, R. P. *2-Oxoglutarate-Dependent Oxygenases*. (The Royal Society of Chemistry , 2015). doi:10.1039/9781782621959

8. Krieg, A. J. *et al.* Regulation of the histone demethylase JMJD1A by hypoxia-inducible factor 1 alpha enhances hypoxic gene expression and tumor growth. *Mol. Cell. Biol.* **30,** 344–53 (2010).

9. Mimura, I. *et al.* Dynamic change of chromatin conformation in response to hypoxia enhances the expression of GLUT3 (SLC2A3) by cooperative interaction of hypoxia-inducible factor 1 and KDM3A. *Mol. Cell. Biol.* **32,** 3018–32 (2012).

10. Tateishi, K., Okada, Y., Kallin, E. M. & Zhang, Y. Role of Jhdm2a in regulating metabolic gene expression and obesity resistance. *Nature* **458,** 757–761 (2009).

11. Inagaki, T. *et al.* Obesity and metabolic syndrome in histone demethylase JHDM2a-deficient mice. *Genes Cells* **14,** 991–1001 (2009).

12. Cho, H.-S. *et al.* The JmjC domain-containing histone demethylase KDM3A is a positive regulator of the G1/S transition in cancer cells via transcriptional regulation of the HOXA1 gene. *Int. J. Cancer* **131,** E179–E189 (2012).

13. Okada, Y., Tateishi, K. & Zhang, Y. Histone demethylase JHDM2A is involved in male infertility and obesity. *J. Androl.* **31,** 75–8

14. Okada, Y., Scott, G., Ray, M. K., Mishina, Y. & Zhang, Y. Histone demethylase JHDM2A is critical for Tnp1 and Prm1 transcription and spermatogenesis. *Nature* **450,** 119–23 (2007).

15. Kuroki, S. *et al.* Epigenetic regulation of mouse sex determination by the histone demethylase Jmjd1a. *Science* **341,** 1106–9 (2013).

16. Schneider, P. *et al.* Identification of a novel actin-dependent signal transducing module allows for the targeted degradation of GLI1. *Nat. Commun.* **6,** 8023 (2015).

17. Wang, L. *et al.* A small molecule modulates Jumonji histone demethylase activity and selectively inhibits cancer growth. *Nat. Commun.* **4,** 1157–1176 (2013).

18. Thinnes, C. C. *et al.* Targeting histone lysine demethylases — Progress, challenges, and the future. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839,** 1416–1432 (2014).

19. Hopkinson, R. J. *et al.* 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem. Sci.* **4,** 3110 (2013).

20. Kanai, Y., Athmaram, T. N., Stewart, M. & Roy, P. Multiple large foreign protein expression by a single recombinant baculovirus: A system for production of multivalent vaccines. *Protein Expr. Purif.* **91,** 77–84 (2013).

21. Sigma-Aldrich. Protein Expression Systems | Sigma-Aldrich. Available at: http://www.sigmaaldrich.com/technical-documents/articles/biology/protein-expression-systems.html.

22. EMBL Heidelberg. Protein Expression and Purification Core Facility - Cloning - Choice of Expression Systems - EMBL. Available at: https://www.embl.de/pepcore/pepcore_services/cloning/choice_expression_systems/.

23. *Instruction Manual Guide to Baculovirus Expression Vector Systems (BEVS) and Insect Cell Culture Techniques*. (Invitrogen life technologies).

24. Kost, T. A., Condreay, J. P. & Jarvis, D. L. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat. Biotechnol.* **23,** 567–575 (2005).

25. Shrestha, B., Smee, C. & Gileadi, O. *Methods in molecular biology (Clifton, N.J.)* **439**, 269–289 (2008).

26. Rose, N. R. *et al.* Plant growth regulator daminozide is a selective inhibitor of human KDM2/7 histone demethylases. *J. Med. Chem.* **55,** 6639–43 (2012).

27. Anti-KDM3A / JHDM2A antibody (ab80598) | Abcam. Available at: http://www.abcam.com/kdm3a-jhdm2a-antibody-ab80598.html.

28. Lewis, J. K., Wei, J. & Siuzdak, G. Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis. *Encycl. Anal. Chem. R.A. Meyers* 5880–5894

29. Shimadzu analytical and measuring instruments. Principles of MALDI-TOF Mass Spectrometry. Available at: http://www.shimadzu.com/an/lifescience/maldi/princpl1.html.

30. Reed, M. C., Lieb, A. & Nijhout, H. F. The biological significance of substrate inhibition: A mechanism with diverse functions. *BioEssays* **32,** 422–429 (2010).

31. Copeland, R. A. *Enzymes : a practical introduction to structure, mechanism, and data analysis*. (Wiley, 2000).

32. Haldane, J. B. S. *Enzymes*. (London: Longmans, Green and Co., 1930).

33. Lizcano, J. M., Unzeta, M. & Tipton, K. F. A spectrophotometric method for determining the oxidative deamination of methylamine by the amine oxidases. *Anal.*

Biochem. **286,** 75–9 (2000).

34. Jean-François *et al.* Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat. Struct. Mol. Biol.* **14,** 689–95 (2007).

35. Kerr & Kelly. An enzyme-based formaldehyde assay and its utility in a sponge sterol biosynthetic pathway. *J. Nat. Prod.* **62,** 201–2 (1999).

36. Suganuma, T. & Workman, J. L. Features of the PHF8/KIAA1718 histone demethylase. *Cell Res.* **20,** 861–862 (2010).

37. Horton, J. R. *et al.* Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat. Struct. Mol. Biol.* **17,** 38–43 (2010).

38. Schiller, R. *et al.* A cell-permeable ester derivative of the JmjC histone demethylase inhibitor IOX1. *ChemMedChem* **9,** 566–71 (2014).

39. Ballatore, C., Huryn, D. M. & Smith, A. B. Carboxylic acid (bio)isosteres in drug design. *ChemMedChem* **8,** 385–95 (2013).

40. Herr, R. J. 5-Substituted-1H-tetrazoles as carboxylic acid isosteres: medicinal chemistry and synthetic methods. *Bioorg. Med. Chem.* **10,** 3379–3393 (2002).

Chapter 4

Investigations into KDM3A inhibition by TCA cycle intermediates and 2OG analogue metabolites

4.1 Introduction

The tricarboxylic acid (TCA), or citric acid, cycle is a central metabolic pathway in all aerobic organisms. This cycle uses a series of enzymes to convert acetyl-CoA into carbon dioxide (**Figure 4.1**). The energy released in these series of reactions is eventually used to synthesise ATP.

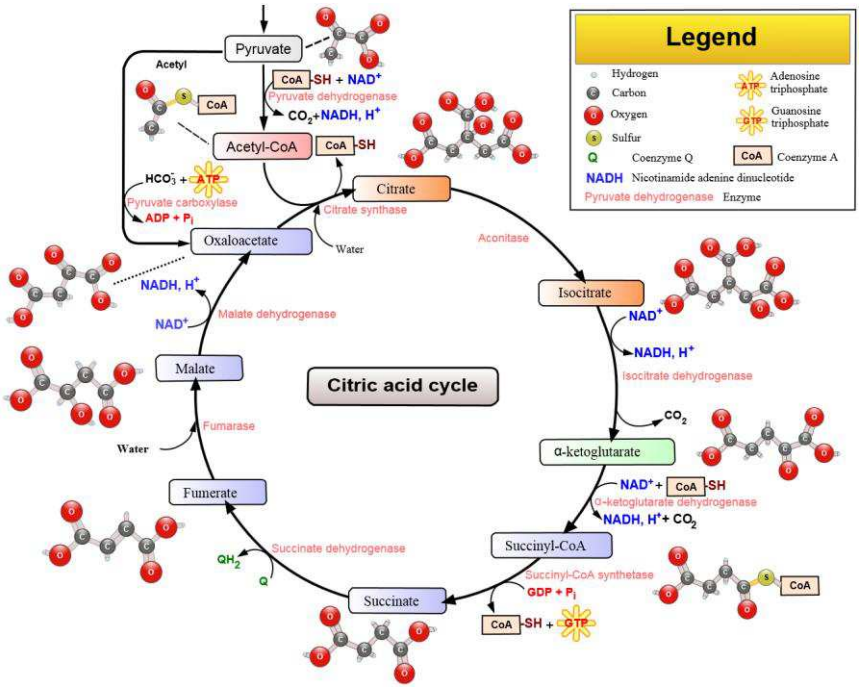


Figure 4.1 A scheme showing enzymes and metabolites in the tricarboxylic acid (TCA) cycle.[1] Reprinted under copyright licence.[2]

The intermediates in the pathway are also used for biosynthetic purposes. Chromatin modifying enzymes utilise TCA cycle intermediates (TCAI) and common metabolites (**Table 4.1**). The TCA cycle intermediates, and associated enzymes, thus have the potential to function as metabolic sensors to directly link metabolism and gene expression.[3]

Table 4.1 Epigenetic histone modifications catalysed by enzymes that utilise common metabolites as co-substrates.[4]

Histone modification	Enzymes	Metabolite co-factors
Methylation	Histone methyltransferases (HMT)	SAM
Demethylation	Histone lysine demethylase (KDMs)	FAD, 2OG
Acetylation	Histone acetyltransferases (HAT)	Acetyl-CoA
Deacetylation	Histone deacetylases (HDACs)	NAD^+

The abundance of TCAI and the rates of the reactions linking them indicates the energetic state of the cell.[4] Modulation of the activity of 'epigenetic enzymes', mediating transcriptional activation or repression, can thus be harnessed to maintain homeostasis.[4] For example, changes in acetylation levels depend on glucose levels; under high glucose concentrations the NAD^+/NADH ratio is low, which decreases the sirtuins HDACs activity that require NAD^+. In turn, this leads to higher histone acetylation levels that promotes gene transcription.[5]

The link between metabolism and epigenetics is also manifested in disease states where the normal homeostatic balance is broken. Altered metabolism is a hallmark of cancer, contributing to rapid proliferation and tumour growth.[6] Some tumours have abnormally elevated levels of fumarate and succinate resulting from mutations to the genes encoding the TCA cycle enzymes fumarate hydratase (FH) and succinate dehydrogenase (SDH), respectively.[7] These elevated concentrations of fumarate and succinate are proposed to be sufficient for the inhibition of the 2OG dependent oxygenases PHDs (hypoxia-inducible factor prolyl hydroxylases) by competitively binding into the active site instead of the co-substrate 2OG.[8–10] Since the PHDs are negative regulators of the hypoxia-inducible factor (HIF) and HIF target genes encode for proteins needed for tumour growth, accumulation of fumarate and succinate is proposed to promote tumour growth.[8,10,11] Similarly,

mutations in either of the two isoforms of the TCA cycle enzyme isocitrate dehydrogenase (IDH1/2) are common in glioma and acute myeloid leukemia.[12] These mutations cause a gain-of-function in IDH1/2 such that they additionally catalyse production of 2-hydroxyglutarate (D-2HG).[13] In some brain tumours, increased levels of D-2HG and L-2HG are also caused by mutations in D or L-2HG dehydrogenases.[14,15] Accumulation of these proposed 'oncometabolites' (to more than 10 mM in some cases) can inhibit 2OG dependent oxygenases with varying potencies suggesting that 2HG elevation might modulate chromatin modifications.[16,17]

4.2 Objectives

The work described in this chapter aimed to investigate the extent by which KDM3A is sensitive to inhibition by TCA intermediates and related metabolites. The work initially involved screening of a panel of 2OG dependent oxygenases against TCA cycle intermediates and other 2OG analogue metabolites by measuring T_m shifts using Differential Scanning Fluorimetry (DSF). Work was then focused on KDM3A inhibition, with single concentration inhibitor screen by the FDH coupled assay to parallel results from the DSF measurements. Control experiments were carried out to ensure that inhibition detected by this assay not due to FDH inhibition. IC_{50} determinations were then conducted by FDH coupled assays with selected potent metabolites.

The initial DSF screen with 2OG dependent oxygenases, except for measurements with KDM3A, were carried out by Dr. Inga Pfeffer.

4.3 Screening for TCAI thermal stabilisation with a panel of 2OG dependent oxygenases

Given the reported ability of fumarate, succinate and 2HG to inhibit a number of 2OG dependent oxygenases, it was considered possible that other 2OG dependent oxygenases might be able to bind TCAI and 2OG analogue metabolites.[16–18] A panel of fifteen recombinant 2OG dependent oxygenases was screened for interaction with twenty metabolites using T_m shift measurements by Differential Scanning Fluorimetry (DSF).[19]

DSF measures the stabilisation of a protein by a ligand during thermal denaturation, it is fast and requires low concentrations of protein compared to other screening methods.[19] The assay uses a fluorescence dye, SYPRO Orange, which has affinity for hydrophobic areas which in turn affect its fluorescence. During a denaturation process hydrophobic parts in the protein become exposed. The assay monitors the thermal unfolding of proteins by measuring the increase in the fluorescence of the dye (**Figure 4.2**). A potential inhibitor would be expected to interact with the protein, hence stabilising it, resulting in an increased melting temperature (T_m). An increase in T_m in most cases therefore correlates with ligand interactions with the protein that promotes protein stability.

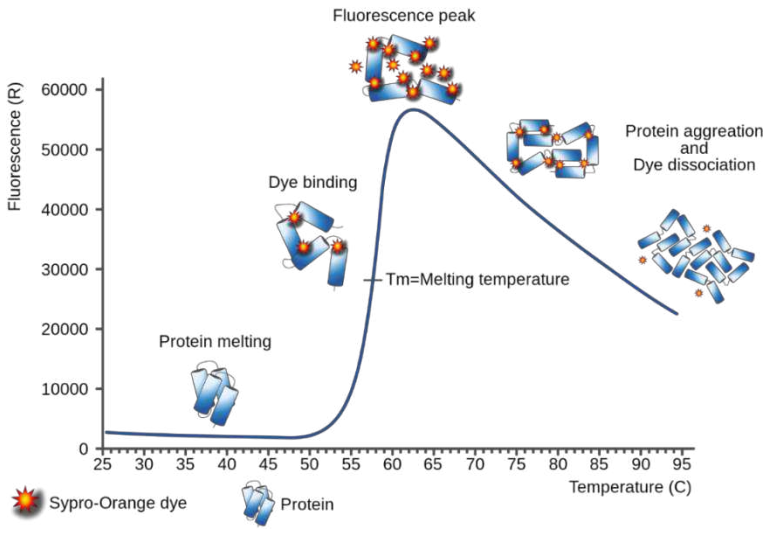


Figure 4.2 A typical protein melting curve obtained from DSF. The Sypro-Orange dye is symbolised as orange stars. Melting of the protein (in blue) exposes hydrophobic patches allowing dye molecules to bind and emits fluorescent light at 610 nm. Following the fluorescent peak a gradual decrease in fluorescence is observed due to protein aggregation. Reprinted under copyright licence.[20]

Thermal shift values are calculated as the difference between the T_m in the presence of a ligand and the T_m of the protein alone. Thermal shift values of different ligands are not always a good indication of their relative binding affinities. The thermal shift is determined by the contribution of enthalpy and entropy during the interaction.[19] In a hydrophobic interaction the entropy component is dominant and results in a large T_m shift compared to a more polar interaction. On the other hand, since the dye used is hydrophobic it can compete with a hydrophobic ligand and cause a decrease in the apparent interaction between the ligand and the protein.[21] All in all, comparison between hydrophobic and polar compounds might not reflect the actual binding affinity. In addition, since the method involves protein denaturation, the nature of the binding might be altered with the change in protein structure. A further caveat is that the fluorescent compounds being screened may interfere with the fluorescence readout. Taking into account the potential drawbacks, DSF allows ranking of compounds with similar physiochemical properties based on their relative stabilisation.[22]

All compounds were tested for thermal stabilisation of 2OG dependent oxygenases at 10 mM concentration, and the T_m shift is calculated relative to the protein alone (see Materials and methods, Chapter 11, Section 11.3.2, for more information). A 200 mM Hepes buffer was used at pH 7.5, pH was measured to verify that the metabolites in the concentration used for the assay do not reduce the pH value below 7. Conditions for the screening procedure were optimised for 2OG dependent oxygenases. Results for screening of the tested metabolites with KDM3A are presented in **Figure 4.3**; results from screening with a panel of 2OG dependent oxygenases are summarised in **Table 4.2**.

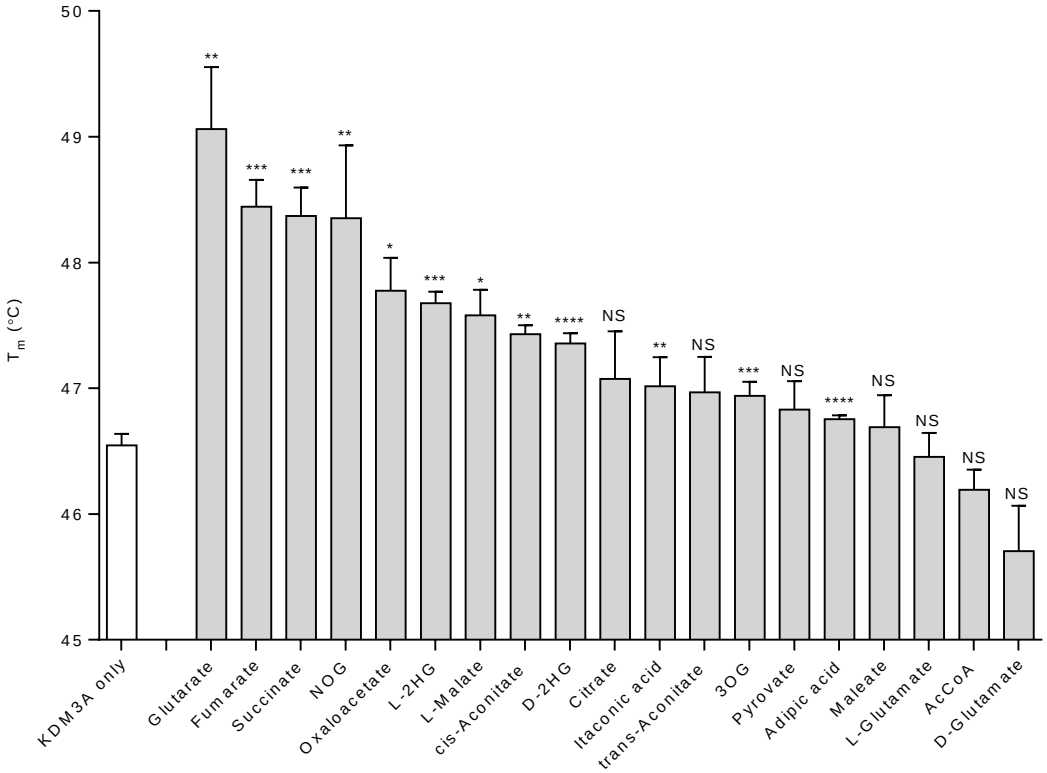
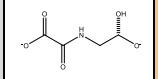
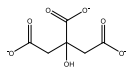
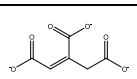
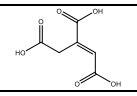
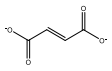
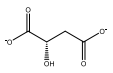
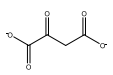
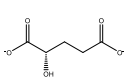
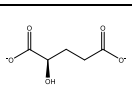
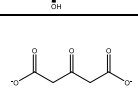
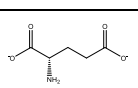
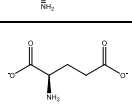
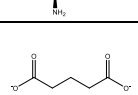
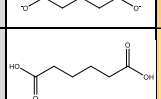
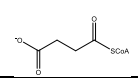
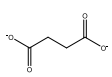
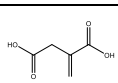


Figure 4.3 Tm measurements by DSF for TCAI and 2OG derivative metabolites against recombinant KDM3A. All measurements were conducted in triplicate and reported as averages with standard deviations. Statistical analyses were carried out using a two-tailed t-test compared with KDM3A only control using GraphPad Prism 6. NS denotes $P > 0.05$, * denotes $P \leq 0.05$, ** denotes $P \leq 0.01$, *** denotes $P \leq 0.001$ and **** denotes $P \leq 0.0001$.

Table 4.2 T_m shift measurements by DSF for TCAI and 2OG derivative metabolites against a recombinant 2OG dependent oxygenase panel. T_m shift data coloured according to a heat map, whereby red indicates a positive T_m shift and green indicates negative T_m shift. Average T_m shift for each enzyme is highlighted in blue at the end of each row and column. All measurements were conducted in triplicate and reported as averages. *All measurements, except for KDM3A, were carried out by Dr. Inga Pfeffer.*[23]

Compound	Structure	T_m Shift (°C)														
		FIH	PHD2	KDM4A	KDM4C	KDM4E	JMJD6	FBXL11	ogfoD1	BBOX	PHF8	Mina53	ycfD	NO66	AlkB ΔN11	KDM3A
		no tag	no tag	His tag	His tag	His tag	His tag	His tag	His tag	no tag	no tag	no tag	His tag	no tag	no tag	His tag
N -Oxalyl glycine (NOG)		6.03	11.07	4.49	3.77	3.62	2.77	1.92	14.23	1.88	-2.13	5.28	2.16	7.70	10.71	1.81
Pyruvate		0.09	5.02	3.87	6.43	7.06	1.24	0.11	6.04	-0.14	0.25	1.13	-0.13	6.61	1.39	0.28
Citrate		-1.29	-5.65	-0.91	-1.39	0.99	-1.50	-3.68	-0.11	-2.35	-0.89	1.63	-1.47	-0.19	-0.71	0.53
cis -Aconitate		1.35	-1.61	2.03	0.49	2.19	2.41	-2.29	8.94	0.39	-7.35	-2.39	3.05	0.36	6.31	0.88
trans -Aconitate		-0.15	-0.34	1.08	0.18	1.35	2.92	-2.35	7.09	0.57	-3.97	-3.63	3.02	0.66	6.60	0.42
Fumarate		0.87	7.17	1.54	0.57	3.12	4.34	2.77	11.52	0.78	-4.61	4.90	3.47	6.31	6.67	1.90
Maleate		-0.11	-1.64	-2.17	-1.43	-0.02	3.68	-2.87	5.26	0.35	-1.51	-1.07	1.19	1.70	-0.29	0.14
L-Malate		-0.78	0.42	0.98	0.58	1.16	2.58	0.43	8.84	0.19	-2.81	0.56	1.98	4.56	4.40	1.03
Oxaloacetate		-0.68	0.89	3.28	4.22	4.61	0.60	-0.29	4.17	-1.62	-3.25	0.68	-1.30	1.86	0.45	1.23
L-2-Hydroxyglutarate (L-2HG)		0.08	1.12	2.51	3.25	8.75	5.94	-	16.62	-1.59	1.02	10.16	3.04	10.01	6.55	1.13
D-2-Hydroxyglutarate (D-2HG)		0.01	6.89	4.17	6.22	6.86	2.69	1.43	10.71	0.99	-0.81	4.45	3.65	6.31	4.40	1.09
3-Oxoglutarate (3OG)		3.32	9.51	3.08	3.45	5.26	4.05	1.37	10.21	-	-	4.23	1.08	4.74	6.51	0.68
L-Glutamate		0.55	1.33	-0.36	0.05	1.35	1.03	-0.47	1.29	-	-	0.58	-0.11	1.36	0.21	0.19
D-Glutamate		0.78	3.23	0.23	0.13	1.03	1.28	0.55	1.47	-	-	0.77	0.21	0.87	0.81	-0.56
Glutarate		5.01	9.53	3.66	3.58	4.03	4.21	3.14	8.98	-	-	4.46	3.58	5.89	13.75	2.80
Adipic acid		3.25	5.65	2.32	2.78	3.14	3.03	1.89	7.54	-	-	2.55	2.01	3.23	7.83	0.49
Acetyl-Coenzyme A (AcCoA)		-1.04	0.25	-3.85	-4.33	-0.36	2.35	-0.64	0.62	-1.43	0.45	-1.64	-0.30	0.01	7.07	-0.35
Succinyl-Coenzyme A (SuccCoA)		-1.85	-0.07	1.20	1.52	1.58	2.27	0.16	5.17	-2.86	4.42	3.63	-3.69	0.85	6.83	-
Succinate		0.20	4.90	2.61	1.60	5.23	4.83	2.41	11.06	1.22	0.32	2.58	4.10	6.29	8.67	1.82
Itaconic acid		-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.75
Average for each enzyme		0.82	3.04	1.57	1.67	3.21	2.67	0.20	7.35	-0.26	-1.49	2.05	1.34	3.64	5.17	0.86

Overall, the results indicate that a number of compounds interact with 2OG dependent oxygenases to varying degrees, suggesting a possible relationship between metabolism in the cell and this superfamily of enzymes.

N-Oxalylglycine (NOG) was used as a positive control compound in this screen. NOG is a close structured analogue of 2OG and is commonly utilised as a broad spectrum inhibitor of 2OG dependent oxygenases.[24] NOG caused a relatively high T_m shift of 1.81°C for KDM3A compared to the other tested metabolites ($P \leq 0.01$) and its stabilisation effect was also apparent in some other 2OG dependent oxygenases such as PHD2 and OgfoD1. The most 'stabilising compound' for KDM3A was glutarate, with a value of 2.80°C ($P \leq 0.01$). 2OG is a derivative of glutarate, with an additional carbonyl group at the C-2 (see **Table 4.2** for structure). Like NOG, glutarate probably mimics 2OG quite well and therefore is likely to fit into the 2OG binding site. The relatively high stabilisation by glutarate was also observed in a number of 2OG dependent oxygenases, including AlkB ΔN11 and KDM4A.

Interestingly, the 2HG enantiomers were relatively good stabilisers for KDM3A (1.13°C with $P \leq 0.001$ for the L-form of 2HG, and 1.09°C with $P \leq 0.0001$ for its D-form). Additional KDMs, such as KDM4E and Mina53 were also stabilised by 2HG. These results support those of the results described in Chowdhury *et al.* which indicate that a number of 2OG dependent oxygenases might interact with these oncometabolites which are reported to accumulate in glioma, malignant glioblastoma and acute myeloid leukaemia.[17]

Similar to the general pattern, the addition of maleate had little effect on the stabilisation of KDM3A (0.14°C, no statistically significant stabilisation) possibly due to its relatively rigid *cis* structure that prevents it from binding to the 2OG site. Therefore, the general trends detected for the panel of oxygenases were also apparent in the KDM3A measurements. Itaconic acid, an analogue of 2OG produced in cells of macrophage lineage, was also tested.[25] Itaconic acid was a weak stabiliser for KDM3A (0.75°C with $P \leq 0.01$).

The DSF screen presented in **Table 4.2** highlighted two enzymes: OgfoD1(2-Oxoglutarate and Iron-Dependent Oxygenase Domain Containing 1) was found to be stabilised by most of the TCAI and 2OG analogue metabolites, as indicated by an average T_m shift value of ~7°C. OgfoD1, is a prolyl 3-hydroxylase of the small ribosomal subunit RPS23 that regulates protein translation termination efficiency.[26,27] It would be interesting to further investigate inhibition of this enzyme by metabolites as it may uncover a new link between metabolism and protein translation. The second enzyme is BBOX (Butyrobetaine (Gamma), 2-Oxoglutarate Dioxygenase). BBOX is an 2-oxoglutarate dependent oxygenase

that catalyses the final hydroxylation step in the biosynthesis of carnitine.[28] BBOX manifested an average T_m shift value which was close to zero (-0.26°C), indicating no stabilisation of the protein by 2OG analogues. Consistent with these observations, BBOX is reported to have a high K_m^{app} value of 2OG which suggests low affinity of 2OG to this enzyme, and consequently that it could act as a 2OG sensor.[29]

4.4 Single concentration inhibitor screen for KDM3A

Following the screening of a broad panel of 2OG dependent oxygenases using DSF, work focused on KDM3A aiming to build on the thermal stability results and examine the inhibitory effect of the tested metabolites. The panel of TCAI and 2OG analogues was then screened for inhibition of H3 K9me2 demethylation activity of isolated KDM3A using the FDH coupled assay as described in Chapter 3, Section 3.5.2. Compounds were screened for inhibition at a single fixed concentration of 1 mM, and their inhibition was normalised relative to a control reaction containing no inhibitor as the baseline, and NOG as a positive control for inhibition. 50 mM Hepes buffer was used at pH 7.5; the pH was measured to check that the concentration of the acidic metabolites used in the assay do not reduce the pH below 7.

KDM3A (250 nM) was incubated with the tested metabolites for 15 minutes prior to the addition of co-factors and H3(1-15)K9me2 peptide. The substrate peptide concentration was kept in excess of > 2x K_m^{app} (see Chapter 3, Section 3.5.2.3). The 2OG concentration in measurements was kept at the K_m^{app} concentration (2OG K_m^{app} value is 30 ± 10 μM, see Chapter 3, Section 3.5.2.2). The fluorescence signal was analysed every 30 seconds over 30 minutes at 37°C. Initial rates were calculated over the linear range of the obtained curves by linear regression (see Materials and methods, Chapter 11, Section 11.3.4, for more information).

Figure 4.4 summarises the results of the screen. D-2HG exhibited similar inhibition potency to that of NOG. Similar to the affinity screen data, the L-2HG enantiomer was also a strong inhibitor (76% inhibition at 1 mM); however, it was slightly less potent than the D-form of 2HG (88% inhibition at 1 mM). The DSF data indicated a stabilisation of KDM3A by the 2HG enantiomers, glutarate and oxaloacetate, which correlates with results of the single concentration inhibitor screen. Interestingly, pyruvate was a weak stabiliser for KDM3A with a T_m shift value that was a third than that of the average (0.28°C versus

0.86°C, no statistically significant stabilisation of KDM3A). However, pyruvate exhibited high inhibition potency (81% at 1 mM, $P \leq 0.001$) compared to NOG. Another difference between the thermal stability and the inhibition data was in the case of fumarate, which exhibited thermal stabilisation (1.9°C versus the average of 0.86°C, $P \leq 0.001$), but showed only 37% inhibition ($P \leq 0.01$).

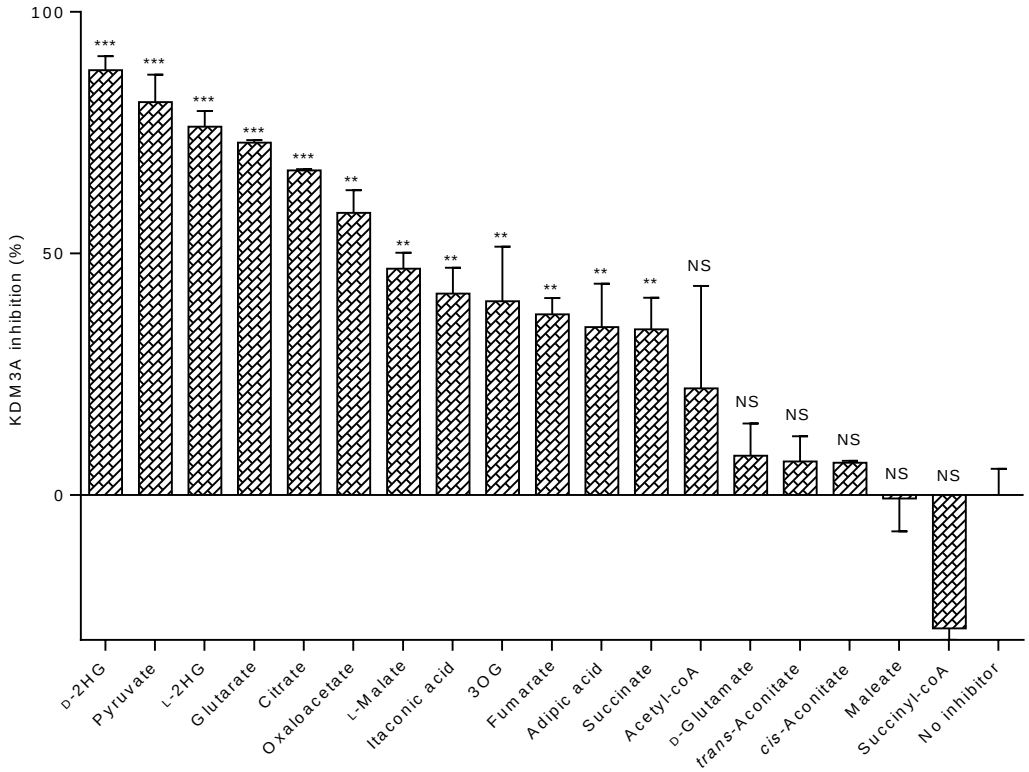


Figure 4.4 Single concentration inhibitor screen of KDM3A by TCAI and additional 2OG derivatives using the FDH coupled assay. All compounds were tested as KDM3A inhibitors at 1 mM concentration, and percentage inhibition normalised to a control reaction without inhibitor addition and to inhibition by NOG. 2OG concentration in measurements was kept at Kmapp concentration (see Chapter 3, Section 3.5.2.2). H3(1-15)K9me2 peptide substrate concentration was kept at > 2x Kmapp. All measurements were conducted in triplicate and values represent mean and standard deviation. Statistical analyses were carried out using a two-tailed t-test compared with no inhibitor control using GraphPad Prism 6. NS denotes $P > 0.05$, ** denotes $P \leq 0.01$, *** denotes $P \leq 0.001$.

4.5 Efficiency of the FDH coupled assay in the presence of 2OG analogues

Prior to IC_{50} analyses of selected 2OG analogues, the efficiency of the FDH coupled assay in the presence of these compounds was tested. Since the FDH assay is a coupled assay, which depends on the activity of FDH, there is a possibility that inhibition of 2OG analogues may be a result of inhibition of FDH itself. To ensure that the inhibitory effect observed in the previous section reflects inhibition against KDM3A, compounds that demonstrated over 50% inhibition were first screened for FDH inhibition. All compounds were screened at 1mM, which is the same concentration used for the inhibitor screen and also the highest concentration used in subsequent IC_{50} evaluations. The FDH concentration was chosen to be 50 μM, as it lies within the linear range of activity for FDH (see Chapter 3, Section 3.5.2.1). Co-factor and co-substrate concentrations were kept at the same concentrations as used for the previous KDM3A inhibition measurements (Section 4.4). FDH enzymatic activity was determined by measuring the initial rate of NADH production, as described in Chapter 3, Section 3.5.2. Activity was normalised according to no inhibitor and no enzyme control experiments. Measuring the initial rate of FDH activity indicated that none of the tested compounds substantially inhibited FDH (**Figure 4.5**), allowing the use of this method for IC_{50} determination. These results correlate with those reported by Chowdhury *et al.* stating that no inhibition is observed for FDH itself by D- or L-2HG.[17]

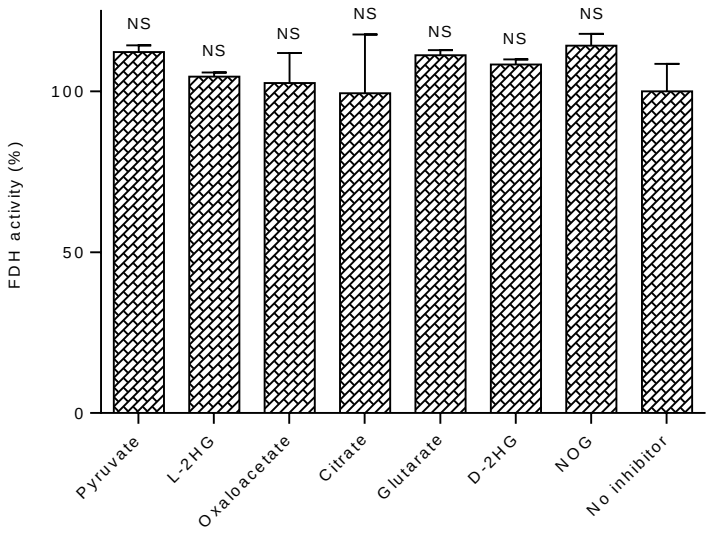


Figure 4.5 FDH activity in the presence of compounds to be evaluated as KDM3A inhibitors (1 mM). Formaldehyde concentration was 50 µM. FDH activity was normalised according no inhibitor and no enzyme controls. All measurements were conducted in triplicate and values represent mean and standard deviation. Statistical analyses were carried out using a two-tailed t-test compared with no inhibitor control using GraphPad Prism 6. NS denotes $P > 0.05$.

4.6 IC_{50} values determination of selected metabolites with KDM3A

After confirming that the inhibitory effect of the six most potent metabolites was due to KDM3A inhibition rather than FDH, these compounds were further evaluated by determining their IC_{50} values (**Figure 4.6, Table 4.3**). Assay conditions were similar to that of the previous assay. The 2OG concentration was kept at the K_m^{app} concentration (2OG K_m^{app} value is 30 ± 10 µM, see Chapter 3, Section 3.5.2.2), while inhibitors concentrations were varied between 3.7 µM and 1 mM. Initial rates were plotted against log (inhibitor concentration (M)); GraphPad Prism 6 was used to fit the resultant dose-response curves.

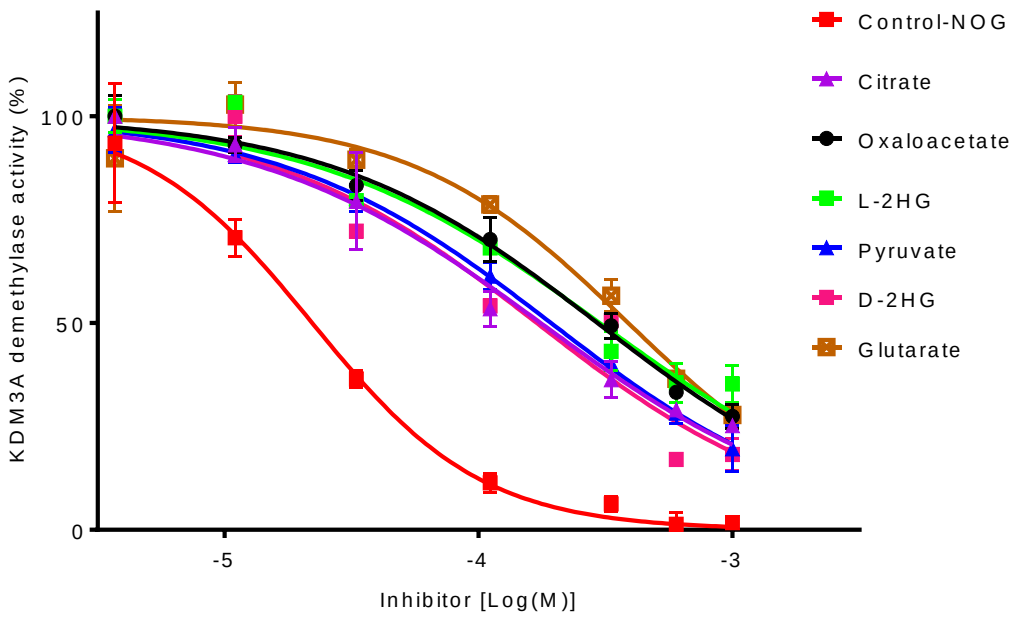


Figure 4.6 IC_{50} determination assayed by the FDH coupled assay for selected compounds tested against KDM3A. Evaluation of the *in vitro* inhibition of KDM3A (250 nM) catalytic activity by selected TCAI and 2OG analogues. NOG was used as a positive control. All measurements were conducted in triplicate and values represent mean and standard deviation.

Table 4.3 IC_{50} values of selected compounds determined by inhibition curves. KDM3A concentration was 250 nM.

Compound	IC_{50} (µM)	95% Confidence Intervals (µM)
NOG	22	20-25
D-2HG	169	77-373
Citrate	176	130-237
Pyruvate	193	182-205
Oxaloacetate	294	253-342
L-2HG	300	188-478
Glutarate	388	287-526

The IC_{50} of NOG was measured as a control. NOG had an IC_{50} value of 22 µM for KDM3A as determined by the FDH coupled assay. In this assay glutarate was the weakest inhibitor and D-2HG the most potent (388 and 169 µM respectively). However, the D- and L- forms

of 2-HG were less potent inhibitors of KDM3A compared to the reported values of the KDM4 representatives (169 / 300 μM for KDM3A compared with 24 / 26 μM for KDM4A and 79 / 97 μM for KDM4C, for D- and the L- 2-HG, respectively).[17] Focusing on KDM3A inhibition, the tested metabolites displayed similar potencies with IC_{50} values within the range of hundreds of micromolar and an average IC_{50} value of 253 μM, an order of magnitude higher than NOG.

4.7 Discussion

The work in this chapter describes the initial characterisation of 2OG dependent oxygenase inhibition by TCAI and related 2OG analogue metabolites. The work aimed to investigate the link between this family of proteins and metabolism, and to test whether 2OG dependent oxygenases have the potential to be metabolic sensors.

The initial approach to screening these compounds was to use thermal shift measurements by DSF to assess compounds stabilisation using a panel of fifteen recombinant 2OG dependent oxygenases. This work demonstrates that metabolites, other than the ones used as co-substrates, can also interact with 2OG-dependent oxygenases in a potentially much broader scope than reported in literature. Thermal shift analyses revealed a relatively strong stabilisation of some 2OG dependent oxygenases to the oncometabolite 2HG in its two enantiomeric forms compared with the other tested metabolites. 2HG is of particular interest in brain cancer research due to its common accumulation in glioblastoma.[14,15] DSF data show that OgfoD1, which regulates protein translation, was stabilised by most of the tested metabolites. Further investigations should be carried out to determine whether OgfoD1 is regulated by these metabolites.

KDM3A thermal stability was moderately sensitive to the tested metabolites. Further analyses were carried out for KDM3A measuring inhibition at a fixed concentration. Potent compounds were tested against FDH alone at their highest concentration to be used for subsequent IC_{50} measurements to confirm that they do not inhibit the reporter enzyme (FDH). IC_{50} values were then determined for selected compounds discovered in the inhibitor screen. Values were all in a similar range of hundreds micromolar, and within this range there were some differences in the potency ranking in comparison to the previous assay.

The physiological relevance of these *in vitro* results is unknown. If metabolites indeed regulate 2OG dependent oxygenases in cells it is probably context dependant, with respect to their subcellular concentrations and the intracellular concentration of 2OG. Notably, the apparent K_M value for 2OG (30 ± 10 μM, Chapter 3, **Table 3.4**) of KDM3A is an order of magnitude lower than the average IC_{50}s (measured with 2OG at the K_M value, see Chapter 3, Section 3.5.2.2), suggesting that significant *in vivo* inhibition of KDM3A by the tested metabolites under normal conditions is unlikely. However, physiologically relevant inhibition is probable in disease states where the normal metabolite balance is disturbed and specific metabolites accumulate, as is proposed in the case of the PHDs.[8] Further studies on the connection between 2OG dependent oxygenases and metabolism are needed. Knowledge gained by such studies might be useful in the early phase detection of disease states and for the development of 2OG dependent oxygenases inhibitors.

Although systematic inhibition studies as well as quantitative cellular assays have not yet been reported, the available inhibition data reveal that TCA cycle intermediates inhibit 2OG dependent oxygenases with varying degrees of potency with IC_{50} values ranging from the micromolar to the millimolar range.[18] This work is part of an ongoing study on TCAI and 2OG dependent oxygenases. Inhibition data for 2OG demethylases other than KDM3A are currently being collected. Kinetic characterisation of these enzymes would be useful in understanding the biological relevance of the results. Enzymes that showed relatively high stabilisation by certain metabolites, such as OgfoD1 and AlkB ΔN11, might be prioritised for cellular work. Metabolomic studies that provide concentration profiles for metabolites in healthy and diseased tissues are also emerging, and will help in identifying the extent to which specific 2OG dependent oxygenases are inhibited by TCAI.[30]

4.8 References

1. Lehninger, A. L., Nelson, D. L. (David L. & Cox, M. M. *Lehninger principles of biochemistry*. (W.H. Freeman, 2005).

2. Citric acid cycle with aconitate. Available at: https://commons.wikimedia.org/wiki/File:Citric_acid_cycle_with_aconitate_2.svg. (Accessed: 14th March 2016)

3. Phang, J. M., Liu, W. & Hancock, C. Bridging epigenetics and metabolism: role of non-essential amino acids. *Epigenetics* **8,** 231–6 (2013).

4. Donohoe, D. R. & Bultman, S. J. Metaboloepigenetics: interrelationships between energy metabolism and epigenetic control of gene expression. *J. Cell. Physiol.* **227,** 3169–77 (2012).

5. Guarente, L. Sirtuins, aging, and metabolism. *Cold Spring Harb. Symp. Quant. Biol.* **76,** 81–90 (2011).

6. DeBerardinis, R. J. & Thompson, C. B. Cellular metabolism and disease: what do metabolic outliers teach us? *Cell* **148,** 1132–44 (2012).

7. Kaelin, W. G. SDH5 mutations and familial paraganglioma: somewhere Warburg is smiling. *Cancer Cell* **16,** 180–2 (2009).

8. Hewitson, K. S. *et al.* Structural and mechanistic studies on the inhibition of the hypoxia-inducible transcription factor hydroxylases by tricarboxylic acid cycle intermediates. *J. Biol. Chem.* **282,** 3293–301 (2007).

9. Isaacs, J. S. *et al.* HIF overexpression correlates with biallelic loss of fumarate hydratase in renal cancer: Novel role of fumarate in regulation of HIF stability. *Cancer Cell* **8,** 143–153 (2005).

10. Selak, M. A. *et al.* Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF-α prolyl hydroxylase. *Cancer Cell* **7,** 77–85 (2005).

11. Semenza, G. L. Defining the role of hypoxia-inducible factor 1 in cancer biology and therapeutics. *Oncogene* **29,** 625–634 (2009).

12. Reitman, Z. J. & Yan, H. Isocitrate dehydrogenase 1 and 2 mutations in cancer: alterations at a crossroads of cellular metabolism. *J. Natl. Cancer Inst.* **102,** 932–41 (2010).

13. Leonardi, R., Subramanian, C., Jackowski, S. & Rock, C. O. Cancer-associated Isocitrate Dehydrogenase Mutations Inactivate NADPH-dependent Reductive Carboxylation. *J. Biol. Chem.* **287,** 14615–14620 (2012).

14. Dang, L., Jin, S. & Su, S. M. IDH mutations in glioma and acute myeloid leukemia. *Trends Mol. Med.* **16,** 387–97 (2010).

15. Struys, E. A. D-2-Hydroxyglutaric aciduria: unravelling the biochemical pathway and the genetic defect. *J. Inherit. Metab. Dis.* **29,** 21–9 (2006).

16. Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19,** 17–30 (2011).

17. Chowdhury, R. *et al.* The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* **12,** 463–9 (2011).

18. Rose, N. R., McDonough, M. A., King, O. N. F., Kawamura, A. & Schofield, C. J. Inhibition of 2-oxoglutarate dependent oxygenases. *Chem. Soc. Rev.* **40,** 4364–97 (2011).

19. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2,** 2212–21 (2007).

20. Florescence TM shift . Available at: http://www.bio.anl.gov/molecular_and_systems_biology/Sensor/sensor_images/assay_theory_figure.png. (Accessed: 14th March 2016)

21. Vivoli, M., Novak, H. R., Littlechild, J. A. & Harmer, N. J. Determination of protein-ligand interactions using differential scanning fluorimetry. *J. Vis. Exp.* 51809 (2014). doi:10.3791/51809

22. Holdgate, G. A. & Ward, W. H. J. Measurements of binding thermodynamics in drug discovery. *Drug Discov. Today* **10,** 1543–1550 (2005).

23. Pfeffer, I., Schofield, C. J., Davis, B. G. & Vi, A. Towards New Methods for Cancer Imaging by Investigation of Cancer Cell Metabolism PRS Transfer Report Written under supervision of. (2011).

24. Hamada, S. *et al.* Synthesis and activity of N-oxalylglycine and its derivatives as Jumonji C-domain-containing histone lysine demethylase inhibitors. *Bioorg. Med. Chem. Lett.* **19,** 2852–5 (2009).

25. Michelucci, A. *et al.* Immune-responsive gene 1 protein links metabolism to immunity by catalyzing itaconic acid production. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 7820–5 (2013).

26. Singleton, R. S. *et al.* OGFOD1 catalyzes prolyl hydroxylation of RPS23 and is involved in translation control and stress granule formation. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 4031–6 (2014).

27. Loenarz, C. *et al.* Hydroxylation of the eukaryotic ribosomal decoding center affects translational accuracy. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 4019–24 (2014).

28. Rydzik, A. M. *et al.* Oxygenase-catalyzed desymmetrization of N,N-dialkyl-piperidine-4-carboxylic acids. *Angew. Chem. Int. Ed. Engl.* **53,** 10925–7 (2014).

29. Rydzik, A. M. *et al.* Development and application of a fluoride-detection-based fluorescence assay for γ-butyrobetaine hydroxylase. *Chembiochem* **13,** 1559–63 (2012).

30. Emir, U. E. *et al.* Noninvasive Quantification of 2-Hydroxyglutarate in Human Gliomas with IDH1 and IDH2 Mutations. *Cancer Res.* **76,** 43–9 (2016).

Chapter 5

Investigations into KDM3A mediated regulation of intraflagellar transport proteins

5.1 Introduction

An emerging body of evidence suggests that KDM3A is required for male fertility and sex determination through its role in chromatin remodelling. Mice lacking KDM3A can undergo male-to-female sex reversal in non-Mendelian frequencies, a property which was shown to be due to the KDM3A epigenetic regulation of the mammalian Y chromosome sex-determining gene *Sry*.[1]

Round-headed spermatozoa are characteristic of globozoospermia, a severe disorder causing male infertility (**Figure 5.1**).[2] *Kdm3a* knockdown in mice results in these round-headed spermatozoa, which lack the acrosome cup required for fusion with the plasma membrane of the oocyte.[3] In addition, the abnormal sperm head in *Kdm3a* knockdown mice could be due to misregulation of DNA condensation by KDM3A, as a result of poor histone replacement by protamines, which gives rise to loose packaging of DNA during spermatogenesis.[4,5]

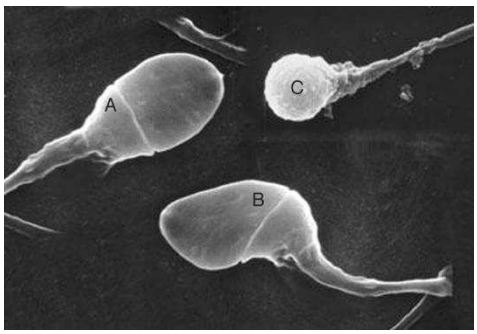


Figure 5.1 Sperm morphology affects male fertility. (A) Normal sperm head. (B) Abnormal sperm head. (C) Globozoospermia, a severe and rare infertility syndrome in which sperm heads lack acrosome caps. Figure used with permission.[2]

High levels of KDM3A were found in mice testes and are observed to be localised not only in the nucleus but also in the sub-compartments of sperm cells, which change during spermatid maturation.[3,4] These observations suggest a possible developmentally regulated cytoplasmic role for KDM3A that may contribute to the round headed phenotype of *Kdm3a* mutant spermatids. Interestingly, although KDM3A has been reported to regulate non-histone proteins (for additional information see Introduction, Chapter 1, Section 1.14.1) and was also observed to be localised in the cytoplasm during spermatogenesis, its potential non-histone related cytoplasmic roles in spermatogenesis, and in particular in flagella formation, are still largely unexplored.

Ciliogenesis is the process by which cilia are built and regulated.[6] Cilia, or flagella (the terms are interchangeable), are microtubule-based structures that extend from the cell surface. There are a number of functional roles for cilia. The best known role is to propel a cell, as is the function of the sperm flagellum.[7] Another role for cilia is in sensory signalling, including the nutrient sensors found in neurons that respond to leptin levels.[8] Although there is a large variety of cilia, their structure is highly conserved and is composed of a cytoskeletal scaffold with microtubules arrayed in specific configurations; these structures are termed axoneme.[9]

Ciliogenesis is highly coordinated with the cell cycle, and emerging data indicate a mutual regulation between the two processes.[7] The cilium resorbs by disassembly and gradually becomes shorter before cell division begins. After cell division, the cilium grows again by gradual assembly. Ciliogenesis requires a bidirectional transport along the axonemal microtubules. This transport is mediated by intraflagellar transport proteins (IFT) which form multi-subunit complexes.[10] IFTs are divided into two complexes; IFT complex B is required for anterograde transport contributing to flagellar assembly, while IFT complex A is required for retrograde transport during disassembly.

Intraflagellar transport is proposed to occur in several steps.[11] IFT-B proteins, the anterograde IFT, are assembled in the cell body around the neck of the cilium.[12] Powered by the motor protein kinesin-2, they then transport cargo and inactive retrograde IFT-dynein 2 motor protein from the cell body to the cilium tip.[13] The next steps involve events termed ‘turnaround’ where the cargo is unloaded and the dynein 2 is activated. IFT-A, the retrograde IFT, returns dissembled ciliary turnover proteins and inactive kinesin-2 motors back to the cell body for degradation or recycling, and the complex is disassembled (**Figure 5.2**).[9]

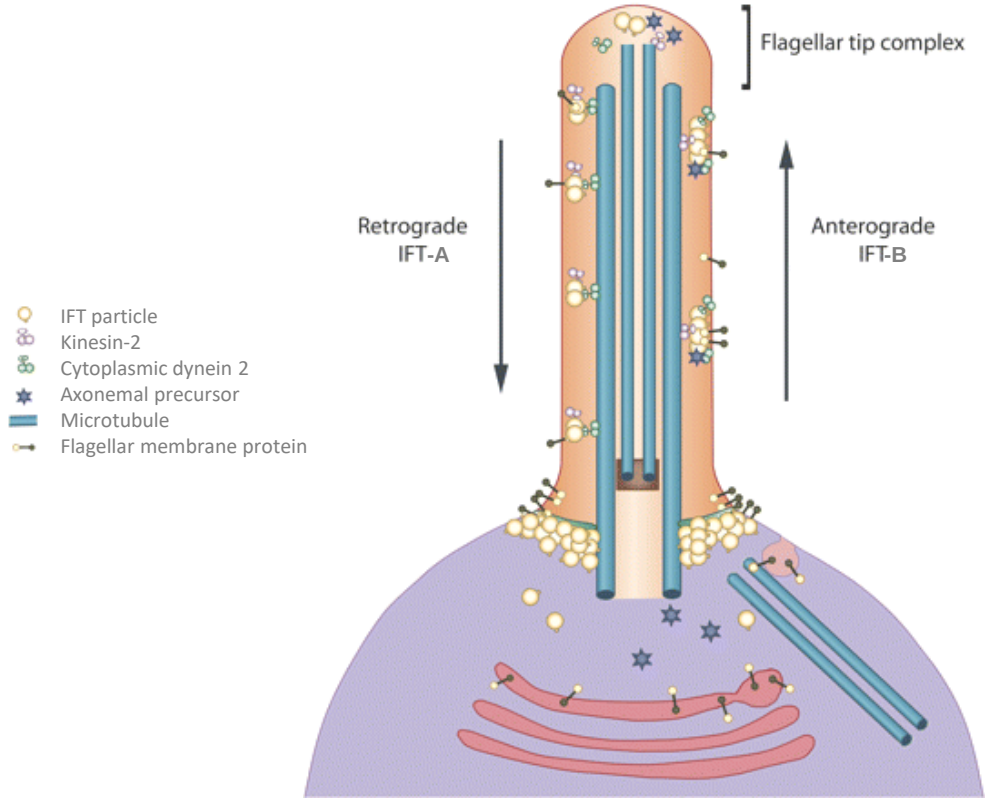


Figure 5.2 Diagram describing the mechanism of intraflagellar transport proteins (IFT). Intraflagellar transport is proposed to occur in several steps. IFT-B, the anterograde IFT, gather around the basal body region. IFT-B transport is then powered by motor protein kinesin-2. They transport cargo, including axonemal and membrane precursors and inactive retrograde IFT-dynein motor protein, from the cell body to the cilium tip. When IFT-B reaches the cilium tip a 'turnaround' process begins where the cargo is unloaded and dynein 2 is activated. Powered by dynein 2, IFT-A, the retrograde IFT, returns turned over proteins and inactive kinesin-2 motors back to the cytoplasm for degradation or recycling, and the complex is disassembled. Figure adapted with permission.[9]

IFT proteins are highly conserved among eukaryotes. Mutations in IFT disrupt ciliogenesis and are linked to various diseases such as Bardet-Biedl syndrome (BBS) and polycystic kidney disease (PDK).[14,15] IFT proteins were first identified in the single-cell green alga *Chlamydomonas reinhardtii.*[16] In this alga, six IFT proteins were found to comprise the IFT-A complex: IFT144, IFT140, IFT139, IFT122A, IFT122B and IFT43 (**Table 5.1**). IFT-B contains IFT88, IFT81, IFT74/72 variants, IFT52, IFT46 and IFT27 in its core, and dissociated components IFT172, IFT80, IFT57 and IFT20. Within the core, IFT74/72 variants interact with a dimer of IFT81 to form a tetramer.[17] IFT88 was shown to be essential in ciliogenesis.[18,19] Another essential interaction in the IFT-B core is the interaction between IFT74 and IFT81, which together form a tubulin-binding module through the actin binding domain Calponin in IFT81 and a highly basic domain in IFT74.[20]

Table 5.1 Components of the IFT complexes found in *Chlamydomonas reinhardtii*.[16]

Retrograde IFT-A	Anterograde IFT-B
IFT144	IFT172
IFT140	IFT88
IFT139	IFT81
IFT122A	IFT80
IFT122B	IFT74/72
IFT43	IFT57
	IFT52
	IFT46
	IFT27
	IFT20

Although components of these complexes have been identified, important questions about the intraflagellar transport process remain.[16] In particular, the process by which these components assemble to form a functional complex is largely unknown. One hypothesis is that IFT are regulated by post-translational modifications (PTMs). PTMs change the chemical properties of the modified residues and therefore are likely to affect protein-protein interaction.[21] In addition, PTMs offer a rapid, controllable and dynamic response mechanism. Such response is needed in bidirectional transport during ciliogenesis, where assembly and disassembly of the IFT complex is constantly turned over. This chapter describes work on lysine methylation of IFT and its effect on interactions with KDM3A. Sloboda's group have also identified methylated arginine residues in IFT, as will be further described in Chapter 6, Section 6.12.[22,23]

Previous work by Dr. Patricia Yeyati and Dr. Holger Kramer has revealed a possible relationship between IFT and KDM3A proteins during ciliogenesis. Studies with mouse models have revealed that *Kdm3a* mutant mice have abnormal length sperm flagella (**Figure 5.3**).

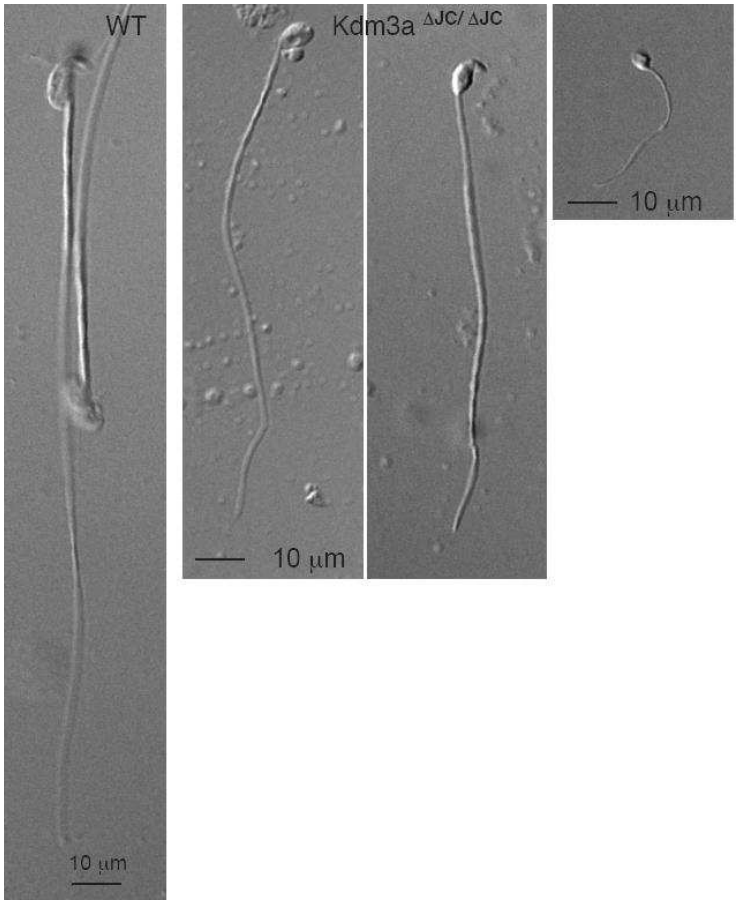


Figure 5.3 ***Kdm3*a mutant mouse models present short flagella**. Spermatozoa from wildtype (WT) and three *Kdm3a* mutants lacking the catalytic domain. *Kdm3a* mutant spermatozoa have rounded heads and short flagella (*study by Dr. Patricia Yeyati, manuscript accepted*).

Analysis of the IFT-B complex from mouse testes extracts by immunopurification and liquid chromatography-tandem MS (LC-MS/MS) has indicated the presence of methyllysine residues in a number of IFT proteins. The highest score predictions for methyllysines were detected for IFT74, IFT81 and IFT88 proteins; these methyllysine positions were then validated using synthetic peptides (*Dr. Patricia Yeyati and Dr. Holger Kramer, manuscript accepted*). Differential centrifugation of cytoplasmic extracts from adult mouse testes showed that KDM3A co-purifies with IFT88 and IFT81 (*Dr. Patricia Yeyati*, *manuscript in preparation*).

Work with the human retinal epithelium cell line hTert-RPE1 has also suggested a possible interaction between KDM3A and IFT. Reciprocal immunopurifications provide evidence of an interaction between IFT81 with KDM3A within hTert-RPE1 cells. IFT81 was pulled-down with KDM3A, and in the reciprocal experiment KDM3A was pulled-down with IFT81

(*Dr. Patricia Yeyati*, *manuscript accepted*). In addition, over-expressing IFT81 promotes ciliogenesis in KDM3A null cells (CRISPR/Cas9), but not in WT *KDM3A* cells (*Dr. Patricia Yeyati*, *manuscript accepted*). This increase, which could not be explained by differences in cell density, suggests that WT cells have an intrinsic constraint on ciliogenesis independent of the increased availability of IFT81 protein, while *KDM3A* null cells do not have such a constraint.

Together with previous observations of cytoskeletal defects in KDM3A mutant spermatids, it was hypothesised that KDM3A might regulate ciliogenesis by associating with methyllysine containing IFT proteins.[3]

5.2 Objectives

The work described in this chapter involved the characterisation of a potential relationship between KDM3A and IFT proteins from complex B. This work investigated the potential regulatory role of this interaction at several levels. First, direct protein-protein interactions were studied by measuring the impact of lysine methylated IFT fragment peptides on the activity, inhibition and thermal stability of isolated KDM3A using a panel of biochemical assays. Then, the potential regulation of KDM3A on IFT at the protein level was tested in cells. Finally, the possibility that IFT regulation occur at the transcriptional level via an 'epigenetic' role of KDM3A was investigated. This study required the development of a robust qRT-PCR assay. This work was extended to test the ability of KDM3A to recognise other potential IFT PTMs. Further work investigated the ability of other KDMs to demethylate methylated IFT fragment peptides.

Identification of putative lysyl-methylated residues in IFT proteins from mice was carried out by Dr. Patricia Yeyati and Dr. Holger Kramer. hTERT RPE-1 cell culturing, quantitative protein level determination and cloning of FLAG-tagged full length WT and H1120Y KDM3A constructs were carried out by Dr. Patricia Yeyati. JMJD1C was produced by Dr. Catrine Johansson and KDM4C was produced by Dr. Akane Kawamura. Expression of full length KDM3A, and production of KDM4A and KDM7B, were performed by Dr. Louise Walport.

5.3 Screening for IFT thermal stabilisation of KDM3A

Following results by collaborators indicating that a cytoplasmic pool of KDM3A might be associated with dimethyllysine-containing IFT proteins, the potential thermal stabilisation of KDM3A by dimethylated IFT fragments was tested. Differential scanning fluorimetry (DSF) was used to test for thermal stabilisation between recombinant KDM3A and IFT synthetic peptide fragments containing dimethylated lysine residues. The assay focused on IFT74(71-90)K81me2, IFT81(536-555)K545me2 and IFT88(61-80)K71me2 peptides which were considered most likely to be methylated according to LC-MS/MS analyses (*Dr. Patricia Yeyati and Dr. Holger Kramer, manuscript accepted*).

For DSF, 2 μM KDM3A was treated with 200 μM of the different peptides in the presence of SYPRO orange, and heated from 25°C to 80°C. *N*-Oxalylglycine (NOG) was used as a 2-oxoglutarate mimic to prevent demethylation of the ligand peptides by KDM3A.[24] A sample of KDM3A with no ligand was used as a baseline for thermal stabilisation determination, and a sample KDM3A in the presence of H3(1-15)K9me2 substrate peptide was used as a positive control.

The observed melting temperature (T_m) of KDM3A in the absence of peptides was 42.1 ± 0.2°C (**Figure 5.4**). There was a thermal stabilisation of 1.7°C in the presence of the H3(1-15)K9me2 substrate peptide. IFT74(71-90)K81me2 and IFT88(61-80)K71me2 peptides both showed an increase in T_m, similar in intensity to the thermal shift value of the positive control (ΔT_m values of 1.6 and 1.5°C, respectively). The IFT81(536-555)K545me2 peptide caused a smaller increase in T_m, suggesting an apparent weaker affinity (ΔT_m value of 1.0°C).

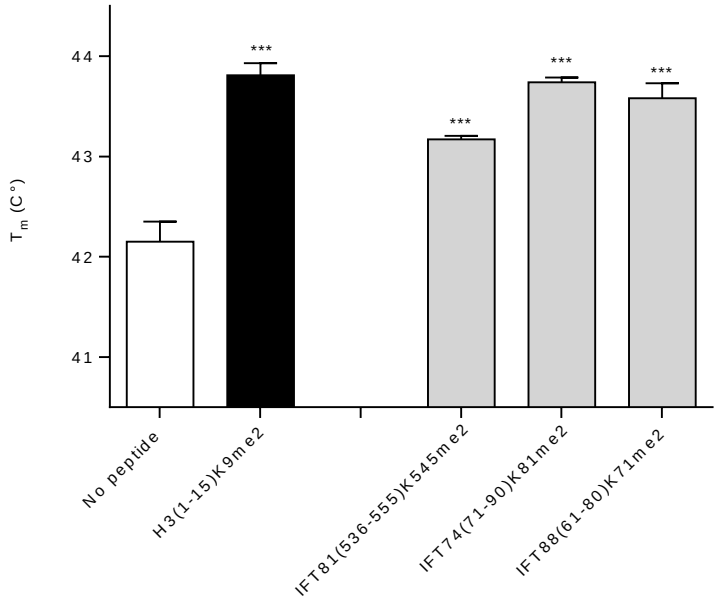


Figure 5.4 Thermal stabilisation of KDM3A by IFT peptides. T_m shift measurements by DSF indicating thermal stabilisation of KDM3A by synthetic peptide fragments of H3(1-15)K9me2 (control) and the dimethylated IFT proteins IFT74(71-90)K81me2, IFT81(536-555)K545me2 and IFT88(61-80)K71me2. Bars are plotted as the mean of triplicate with standard deviation. Statistical analyses were carried out using a two-tailed t-test compared with no peptide control using GraphPad Prism 6. *** denotes $P \leq 0.001$.

These observations indicate that the IFT peptides likely interact with KDM3A. However, DSF alone does not provide information regarding the site and type of interaction and does not always reflect the actual binding affinities and therefore additional characterisation was required to rank between the tested IFT peptides.

5.4 Screening for lysine demethylation of IFT by KDM3A

Following immunopurification results by collaborators which indicated that a cytoplasmic pool of KDM3A might be associated with dimethyllysine-containing IFT proteins, it was proposed to test whether KDM3A can post-translationally regulate IFT proteins by lysine demethylation.

To test whether IFT proteins might be non-histone substrates of KDM3A, an *in vitro* MALDI-TOF MS-based peptide demethylation assays were performed using recombinant KDM3A (**Table 5.2**). Dimethylated IFT fragment peptides were designed based on predictions of lysine methylation sites in IFT-B proteins according to LC-MS/MS analyses *(prediction by Dr. Patricia Yeyati and Dr. Holger Kramer, manuscript accepted*).

Table 5.2 Methylated IFT peptides screened against recombinant KDM3A. Target lysyl residues are shown in bold.

Peptide name	Protein	Target lysyl residue	Peptide sequence
IFT74(71-90)K81me2	IFT74	K81	PVTQQGLSGM-**Kme2-**TGMKGPQRQ**
IFT81(536-555)K545me2	IFT81	K545	AAGLESNRS-**Kme2-**LEQEVRRLRE**
IFT81(556-570)K563me2	IFT81	K563	EALQEES-**Kme2-**YHYTNAM*
IFT81(642-656)K649	IFT81	K649	LEQLMEA-**Kme2-**KQAFLKQ*
IFT81(642-656)K650	IFT81	K650	LEQLMEAK-**Kme2-**QAFLKQ*
IFT88(61-80)K71me2	IFT88	K71	SRPIATGYGS-**Kme2-**TSLTSSMGR**
IFT88(131-145)K138me2	IFT88	K138	PASPLEA-**Kme2-**KKDSPEE
IFT88(325-349)K332me2	IFT88	K332	FAIGDRE-**Kme2-**MKKAFQK

**Cysteine residues in the original protein sequence were replaced with alanine residues to block the formation of a disulfide bond. **Activity was tested with a 20 mer peptide instead of the standard length of 15 mer.*

The demethylation assays were carried out with IFT peptides, while the H3(1-15)K9me2 substrate peptide was used as a positive control. Peptides (10 μM) were incubated with recombinant KDM3A (2 μM) and co-factors under standard demethylation conditions for one hour at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Analyses were carried out by MALDI-TOF MS. A shift in peptide mass by -14 Da and -28 Da, compared to the negative controls with no enzyme, indicated mono- and di-demethylation, respectively.

Although KDM3A was shown to be active according to the positive histone fragment control which underwent > 80% demethylation, no observed demethylation or other enzymatic activity was detected with the dimethyllysine IFT peptides under the tested assay conditions (**Figure 5.5**). Note that small variations (≤ 0.5 Da) in the absolute experimentally observed values compared with the calculated values of the peptides was observed (see Chapter 11, Section 11.3.3).

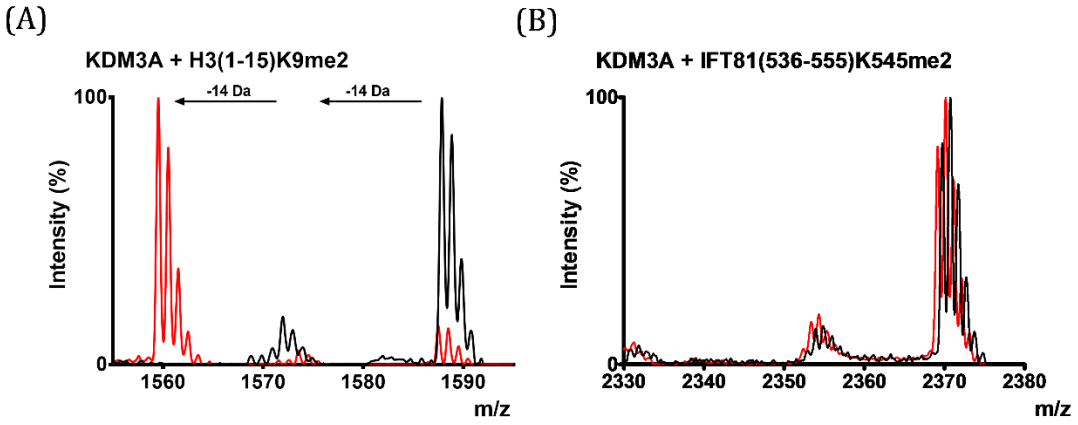


Figure 5.5 Representative MALDI-TOF MS spectra from IFT screen. (A) Demethylation of positive control peptide H3(1-15)K9me2. (B) Absence of demethylation of IFT81(536-555)K545me2 peptide. Peptides (10 μM) were incubated with KDM3A (2 μM) for one hour at 37°C, followed by analyses by MALDI-TOF MS (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Reactions with enzyme are shown in red, control reactions with no enzyme are shown in black.

5.5 Screening for lysine demethylation of IFT by full length KDM3A

Section 5.4 describes activity tests using recombinant KDM3A in which the *N*-terminal domain is truncated ($KDM3A_{515-1317}$, see Materials and methods, Chapter 11, Section 11.2.2, for more information). To test whether lysine demethylation activity can be observed by the full length KDM3A ($KDM3A_{1-1321}$), constructs encoding for *N*-terminally FLAG-tagged full length WT and H1120Y catalytically inactive variant proteins were overexpressed in HEK293T cells.[25] The recombinant proteins were then immunoprecipitated using their FLAG-tags. Both the WT and mutant forms of KDM3A showed degradation similar to that previously observed for recombinant KDM3A protein

(see Chapter 3, **Figure 3.7** and Chapter 6, **Figure 6.5**). Assays focused on the IFT74(71-90)K81me2, IFT81(536-555)K545me2 and IFT88(61-80)K71me2 peptides which were considered more likely to be methylated according to LC-MS/MS characterisation (*Dr. Patricia Yeyati and Dr. Holger Kramer, manuscript accepted*). Lysine demethylation activity was analysed by MALDI-TOF MS (see Materials and methods, Chapter 11, Section 11.3.3, for reaction conditions).

As a positive control, full length WT KDM3A was used and found to be catalytically active, catalysing demethylation of the H3(1-15)K9me2 peptide substrate, while the full length H1120Y variant KDM3A was found to be inactive (**Figure 5.6**). No activity was detected with any of the screened IFT peptides under the tested conditions, consistent with the absence of activity observed in assays with truncated recombinant enzyme.

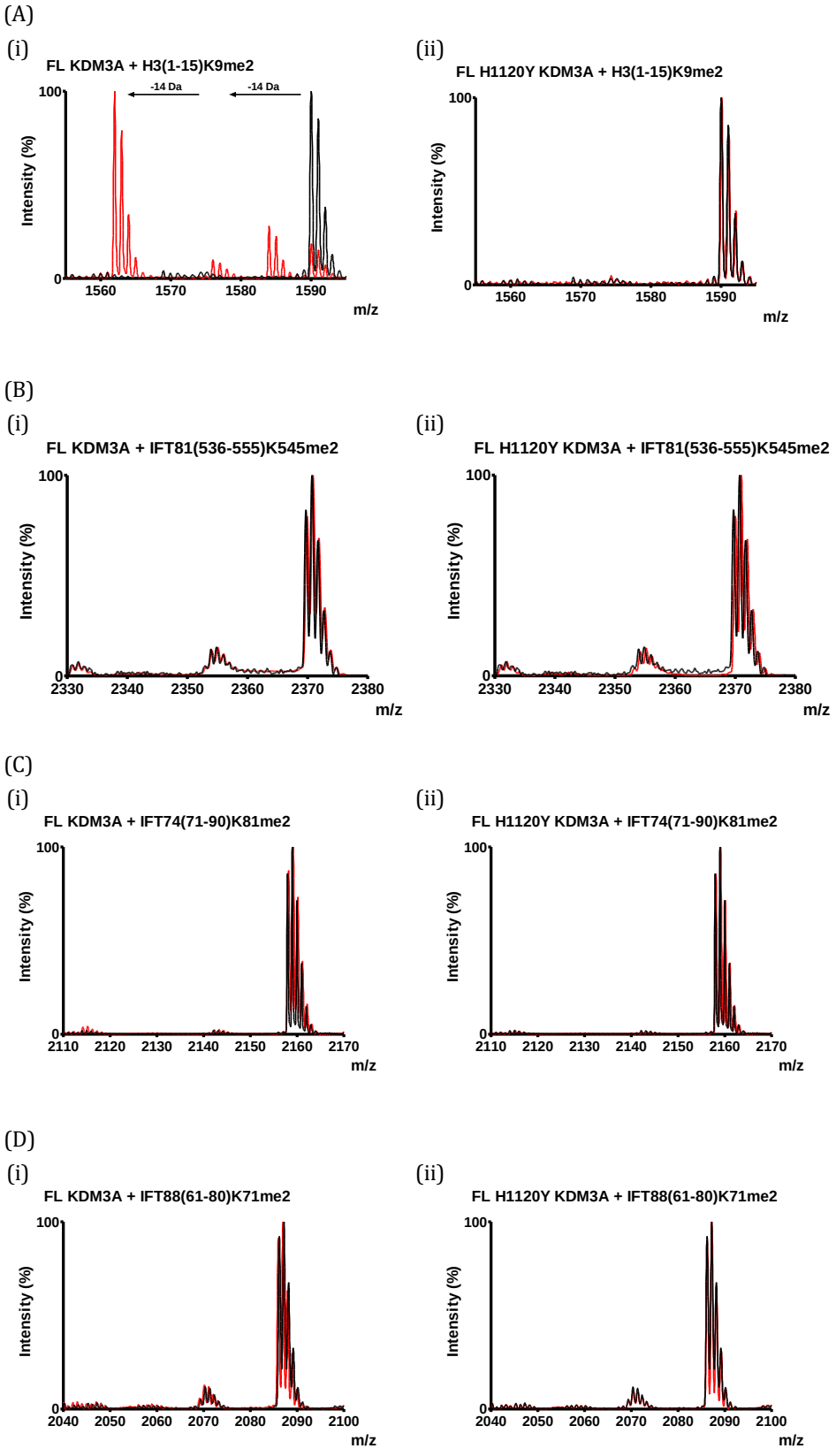

(A)
(i)
FL KDM3A + H3(1-15)K9me2
-14 Da
-14 Da
Intensity (%)
m/z
(ii)
FL H1120Y KDM3A + H3(1-15)K9me2
Intensity (%)
m/z
(B)
(i)
FL KDM3A + IFT81(536-555)K545me2
Intensity (%)
m/z
(ii)
FL H1120Y KDM3A + IFT81(536-555)K545me2
Intensity (%)
m/z
(C)
(i)
FL KDM3A + IFT74(71-90)K81me2
Intensity (%)
m/z
(ii)
FL H1120Y KDM3A + IFT74(71-90)K81me2
Intensity (%)
m/z
(D)
(i)
FL KDM3A + IFT88(61-80)K71me2
Intensity (%)
m/z
(ii)
FL H1120Y KDM3A + IFT88(61-80)K71me2
Intensity (%)
m/z

Figure 5.6 MALDI-TOF MS spectra showing a lack of demethylation of IFT peptides by full length KDM3A. FLAG-tagged full length (FL) WT and the catalytically inactive H1120Y variant KDM3A were immunoprecipitated from HEK293T cells and incubated with 10 μM of peptide for one hour under standard conditions (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) An H3(1-15)K9me2 histone peptide used as a positive control, (B) IFT81(536-555)K545me2, (C) IFT74(71-90)K81me2 and (D) IFT88(61-80)K71me2 peptides. For each peptide, (i) MALDI-TOF MS analysis following incubation with FLAG-tagged FL WT KDM3A and (ii) MALDI-TOF MS analysis following incubation with FLAG-tagged FL H1120Y catalytically inactive KDM3A. Reactions containing enzyme are shown in red, control reactions with no enzyme are shown in black. *Cloning of FLAG-tagged full length WT and H1120Y KDM3A constructs were carried out by Dr. Patricia Yeyati. Expression of recombinant full length KDM3A was performed by Dr. Louise Walport.*

5.6 KDM3A inhibition by IFT

The results described in Sections 5.3 - 5.5 suggest that the tested methyllysine IFT peptides are not substrates of KDM3A, but might still interact with it. The co-purification of IFT proteins with KDM3A and thermal stability data might therefore may reflect relationships that are independent of KDM3A catalysis (see Sections 5.1 and 5.3). There is precedent for non-histone, demethylase-independent regulation by KDM3A. Recent findings suggest that KDM3A stabilises Gli1 protein levels in a demethylase-independent manner.[26] In light of these findings, it was envisaged to determine whether KDM3A binds and regulates IFT in a demethylase-independent manner similar to Gli1 regulation. Binding of IFT might occur outside of KDM3A active site (whether via specific or unspecific binding), or IFT may bind to the KDM3A active site without undergoing demethylation, in a similar manner to the binding of a competitive inhibitor.

To test whether IFT peptides can bind to the KDM3A active site, MALDI-TOF-MS-based KDM3A inhibition studies were carried out. Inhibition was measured at the linear range of activity for lysine demethylation, which was determined to be five minutes for the reaction of 1 μM KDM3A with 10 μM H3(1-15)K9me2 peptide (see Chapter 3, Section 3.5.1.1). Inhibition assays were carried with 100 μM of the dimethyllysine IFT peptides together with 10 μM H3(1-15)K9me2 peptide and the reactions were incubated for 8 minutes before quenching. KDM3A demethylation activity was normalised according to control experiments with no enzyme or no IFT peptide. Assays were carried out in triplicate and analysed using MALDI-TOF MS and GraphPad Prism 6.0 (see Materials and methods, Chapter 11, Section 11.3.3, for more information).

The IFT74(71-90)K81me2 peptide had the weakest apparent inhibition (8% decrease in KDM3A activity compared to positive control), while the IFT81(536-555)K545me2 peptide had the strongest apparent inhibitory effect (48%, **Figure 5.7**). Although the DSF data suggested different stabilizing effects, the inhibition assay may better reflect the actual interaction between KDM3A and IFT peptides as it demonstrates that the peptide interactions with KDM3A affects demethylation activity. Interference with KDM3A enzymatic activity suggests that the tested IFT peptides, and in particular the IFT81(536-555)K545me2 peptide, bind to the KDM3A.

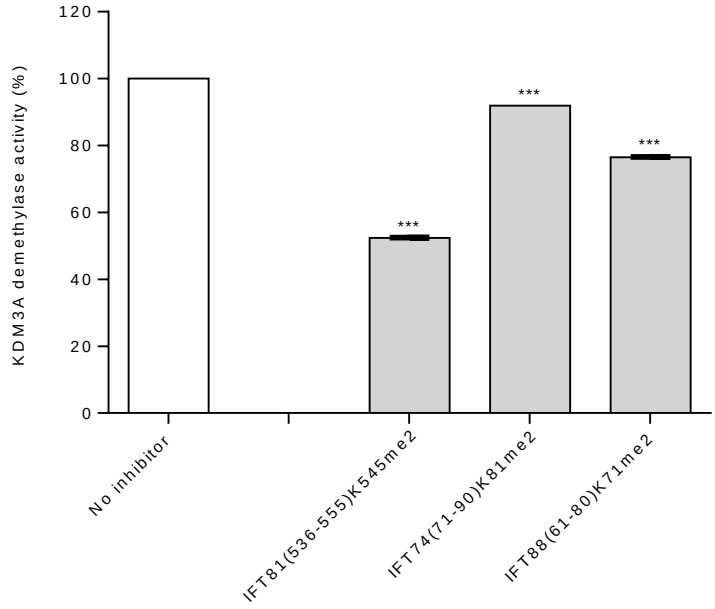


Figure 5.7 ***In vitro*** **inhibition assay of IFT peptides against KDM3A.** MALDI-TOF MS based demethylation assays on the histone substrate peptide H3(1-15)K9me2 in the presence or absence of synthetic methylated IFT peptides. Strongest inhibition was observed with the IFT81 peptide, harbouring methylated lysine 545. Inhibition assays were carried with 100 µM methylated IFT peptides together with 10 µM H3(1-15)K9me2 peptide and the reactions were incubated for 8 minutes before 1:1 quenching with methanol. See Materials and methods, Chapter 11, Section 11.3.3, for more information. Points are plotted as the mean of triplicate with standard deviation. Statistical analyses were carried out using a two-tailed t-test compared with no inhibitor control using GraphPad Prism 6, *** denotes $P \leq 0.001$.

5.7 Investigating the importance of IFT81 lysine methylation on inhibition of KDM3A

The results described in Section 5.6 indicate that the IFT81(536-555)K545me2 peptide inhibits KDM3A. One option for demethylase-independent regulation of IFT by KDM3A is that KDM3A binds IFT81 in a manner promoted by its methyllysine.

To further investigate whether dimethylation on K545 of IFT81 affects the affinity of the peptide to KDM3A, IC_{50} values of non-methylated or dimethylated IFT81(536-555) peptides against KDM3A were determined.

The formaldehyde dehydrogenase (FDH) coupled assay was used for inhibition measurements (for assay information see Chapter 3, Section 3.5.2), with the H3(1-15)K9me2 substrate peptide concentration kept at the K_M value (see Chapter 3, Section 3.5.2.3). Initial rates were plotted against inhibitor concentration (log[M]); GraphPad Prism 6.0 was used to fit the resultant dose-response curves (see Materials and methods, Chapter 11, Section 11.3.4, for more information).

The methylated IFT81 peptide displayed a greater inhibitory effect (IC_{50} = 24 μM) of the demethylase activity of KDM3A on histone substrates than the corresponding non-methylated peptide (IC_{50} = 270 μM) (**Figure 5.8**, **Table 5.3**). This result indicates that KDM3A preferentially binds the methylated form of IFT81, as might be anticipated given its known role as a lysine demethylase.

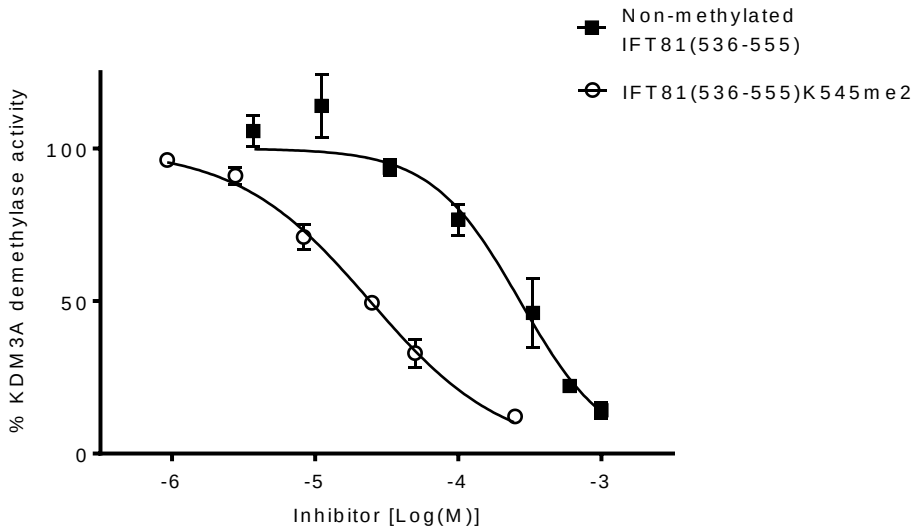


Figure 5.8 IC_{50} determination of non-methylated and dimethyllysine IFT81(536-555)K545 peptides with KDM3A. *In vitro* enzymatic reaction of KDM3A (100 nM) was performed in the absence (100% activity) or increasing concentrations of non-methylated or dimethylated IFT81. Demethylation activity was determined by the FDH coupled assay (see Materials and methods, Chapter 11, Section 11.3.4, for more information). Points are plotted as the mean of triplicate with standard deviation.

Table 5.3 IC_{50} values determined by the FDH coupled assay. KDM3A concentration was 100 nM.

Inhibitor	IC_{50} (μM)	95% Confidence Intervals (μM)
Non-methylated IFT81(536-555	270	182-401
IFT81(536-555)K545me2	24	21-28

5.8 Preliminary investigation of the mechanism of KDM3A inhibition by methylated IFT

The results described in Section 5.7 suggest that the presence of the dimethyllysine residue in the IFT81 fragment peptide strengthens the interaction between KDM3A and IFT81, as the modified peptide had a greater inhibitory potency than the non-methylated peptide. It was propsed that KDM3A is likely to bind the lysine methylated form of IFT81 through its JmjC domain. Thus, experiments to investigate the mechanism of inhibition (i.e. competitive, non-competitive or uncompetitive modes) were carried out.

Kinetic characterisation was carried out using the FDH coupled assay. In these experiments the concentrations of the substrate peptide, H3(1-15)K9me2, and the concentrations of the inhibitor, IFT81(536-555)K545me2, were varied (0 - 300 μM and 0 – 100 μM for the substrate and for the inhibitor, respectively). The apparent K_M and $V_{max\ values}$ were calculated and values were compared between no inhibitor and increasing inhibitor concentrations. Due to the observed substrate inhibition, kinetic parameters were calculated using both the Michaelis-Menten model by using data points below the substrate inhibition range and the substrate inhibition model.

The substrate inhibition model resulted in large errors when calculating K_M and $V_{max.}$ This is likely to be an indication of an insufficient number of data-points. Data from this assay was used for preliminary characterisation, though it is appreciated that further investigation with a larger data-set is required for former conclusions to be made. Notably, analysis by both models imply that increasing the inhibitor concentration decreases both K_M and V_{max} values; an exception was the assay with 100 μM inhibitor where using the substrate inhibition model resulted in a large error. A decrease in both the apparent K_M and V_{max} values is an indication of uncompetitive inhibition, in which the inhibitor binds the enzyme only after an enzyme-substrate (ES) complex has been formed.[27] The αK_i for uncompetitive inhibition was 170 ± 45 μM.

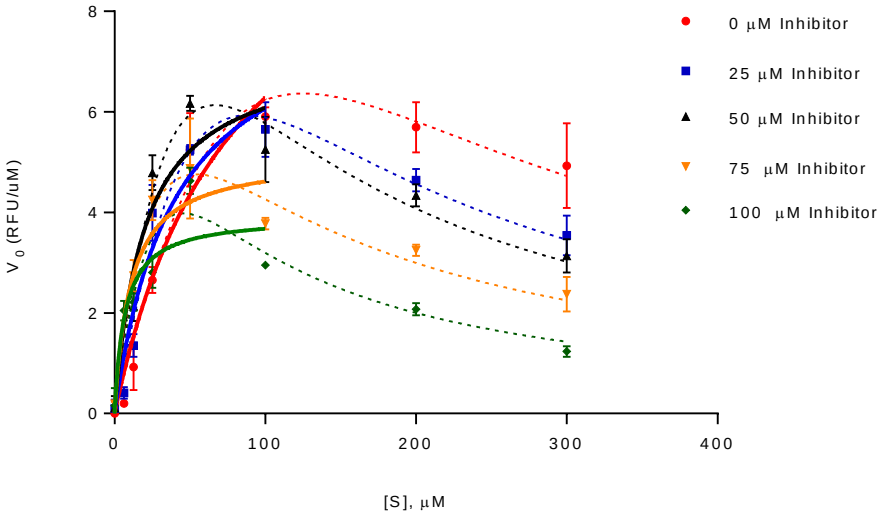
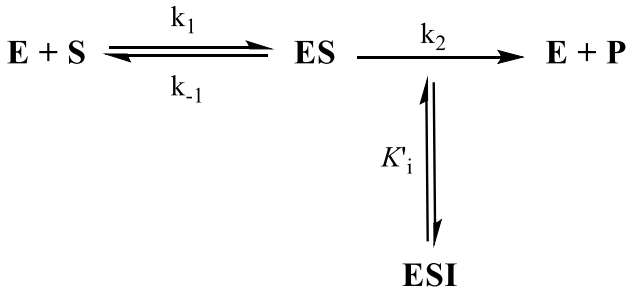


Figure 5.9 Michaelis-Menten plot to investigate the mechanism of inhibition of IFT81(536-555)K545me2 of KDM3A. The FDH coupled assay was used for KDM3A inhibition kinetic studies. Concentrations of the substrate peptide, H3(1-15)K9me2, and the concentrations of the inhibitor, IFT81(536-555)K545me2, were varied (0 - 300 µM and 0 – 100 µM for the substrate and for the inhibitor, respectively). The apparent K_M and V_{max} values were calculated using GraphPad Prism 6.0 (see Materials and methods, Chapter 11, Section 11.3.4, for more information). Dotted line, fitted with substrate inhibition curve. Solid line, fitted with Michaelis–Menten curve, analysing only the substrate concentration range where kinetics obeys the standard Michaelis-Menten (< 100 µM).All measurements were conducted in triplicate and values represent mean and standard deviation.

Table 5.4 Apparent K_M and V_{max} values of H3(1-15)K9me2 catalysis by KDM3A in the presence of increasing concentrations of IFT81(536-555)K545me2. Apparent kinetic parameters were calculated by fitting either the Michaelis-Menten model below substrate inhibition range or the substrate inhibition model using GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.4, for more information). All measurements were conducted in triplicate and values represent mean and standard deviation.

Model used	**Michaelis-Menten**		**Substrate inhibition**	
IFT81(536-555)K545me2 inhibitor (µM)	**Apparent K_M (µM)**	**Apparent V_{max} (µM/min)**	**Apparent K_M (µM)**	**Apparent V_{max} (µM/min)**
0	80 ± 28	11 ± 2	359 ± 448	43 ± 48
25	40 ± 11	8 ± 1	231 ± 207	37 ± 29
50	20 ± 6	7.3 ± 0.7	105 ± 52	25 ± 10
75	11 ± 4	5.1 ± 0.5	40 ± 16	12 ± 3
100	7 ± 3	3.9 ± 0.4	69 ± 58	16 ± 11

The preference of the inhibitor for binding to the ES complex over the non-substrate bound enzyme causes a relative loss of the ES complexes to form enzyme-substrate-inhibitor (ESI) complexes according to the Le Châtelier's principle (**Scheme 5.1**). As a result, an apparent greater affinity to the enzyme for the substrate is observed, hence the reduction in K_M. Reduction in V_{max} is a result of the enzyme inhibition which cannot be countered by adding more substrate. Given the behaviour of IFT81(536-555)K545me2 as an apparent uncompetitive inhibitor, it is likely that binding occurs outside the active site of KDM3A, arguing against the hypothesis for binding of IFT81 to the active site via its dimethyllysine.



Scheme 5.1 Michaelis-Menten reaction scheme for reversible uncompetitive inhibition. E represents the enzyme, S represents the substrate and I represents the reversible uncompetitive inhibitor which binds the enzyme only after an enzyme-substrate (ES) complex was formed. By decreasing the concentration of ES complex in the system and increasing the concentration of the ESI complex, the inhibitor shifts the equilibrium towards production of more ES complexes, resulting in apparent higher affinity to the enzyme for the substrate.[27]

5.9 Investigating whether KDM3A regulates IFT81 protein levels

Sections 5.6-5.7 describe investigations into the possibility that KDM3A non-catalytically regulates IFT via binding, similar to how KDM3A binds Gli1 and protects it from degradation, therefore stabilising its protein levels.[26] This hypothesis could explain the co-purification of KDM3A and IFT proteins and the lack of KDM3A catalysis on IFT peptides (see Sections 5.4 - 5.5). To further examine this hypothesis, it was envisaged to determine whether KDM3A regulates IFT protein levels in a demethylase-independent manner.

IFT81 protein levels were measured using quantitative LC-MS/MS and western blotting. hTert-RPE1 cells with the *IFT81*$^{cr3.7}$ genotype, which lack endogenous IFT81 through CRISPR/Cas9 knockout, were used as a negative control. IFT81 protein levels were compared between cells with WT and null *KDM3A* genotypes (*KDM3A*$^{Cr22.2}$; CRISPR/Cas9), which were either treated to induce ciliary growth or resorption (*cell lines generated by Dr. Patricia Yeyati, manuscript accepted*).

Data from both LC-MS/MS and western blotting experiments demonstrate that no change in IFT81 protein levels occurred between WT and null *KDM3A* cells (**Figure 5.10**). These results indicate that during ciliogenesis, IFT81 protein turnover is independent of KDM3A.

Only a small apparent decrease in IFT81 protein levels was detected by both quantitative mass spectrometry and western blotting experiments during resorption. Engel *et al.* have found that regulation of intra-ciliary transport is controlled by the IFT cargo size rather than by the quantity of IFT proteins per cilium.[28] The results of the new experiments correlate with the report by Engel *et al.* showing similar levels of IFT81 protein during growth and resorption.

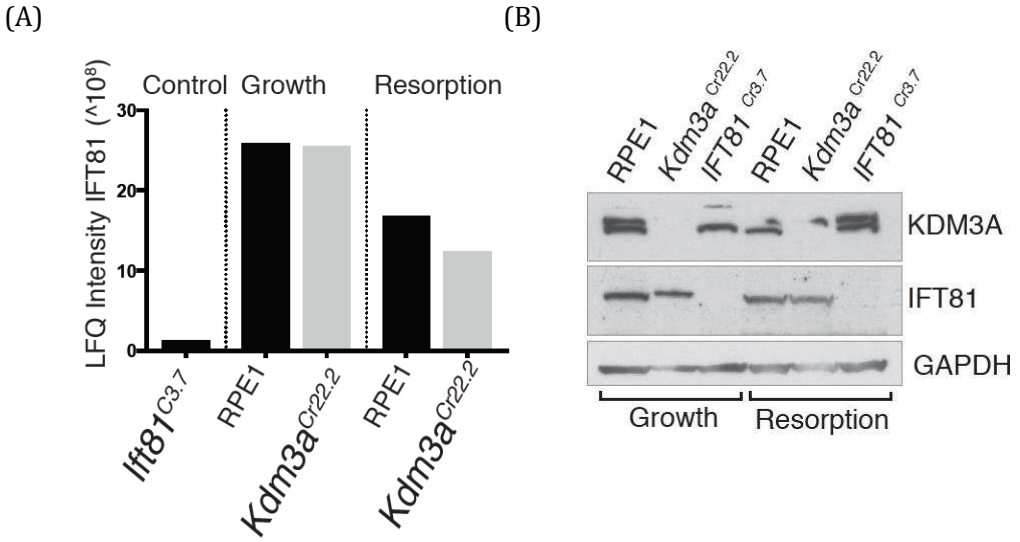


Figure 5.10 Dependence of IFT81 protein levels on KDM3A expression during ciliary growth and resorption. *IFT81*$^{cr3.7}$ cells lack endogenous IFT81 through CRISPR/Cas9 knockout. *KDM3A*$^{Cr22.2}$ cells (*KDM3A* null) do not express KDM3A protein through CRISPR/Cas9 knockout. (A) Quantitative mass-spectrometric analyses by LC-MS/MS of immunopurified IFT81 extracted from hTert-RPE1 and *KDM3A*$^{Cr22.2}$ mutants during ciliary growth and resorption. *IFT81*$^{Cr3.7}$ pools (combined extracts of growth and resorption) were used as a negative control. (B) Total cell extracts from growing and resorbing cultures blotted with the indicated antibodies. *Measurements were carried out by Dr. Patricia Yeyati.*

5.10 Investigating the epigenetic regulation on IFT expression by KDM3A

The presented data regarding KDM3A regulation of IFT imply that IFT fragment peptides are not substrates for KDM3A, at least under the tested conditions. In addition, KDM3A does not seem to regulate IFT81 protein levels during ciliogenesis. However, the DSF and inhibition data indicated that an interaction between KDM3A and IFT peptides might be possible although not necessarily specific, and that methylation of Lys545 in IFT81 improved the inhibition potency. One hypothesis that could explain these results is that KDM3A regulates IFT in a demethylase-independent manner by binding to methylated Lys545 in IFT81.

Independent of whether this data can be explained by direct protein-protein interactions, KDM3A might also epigenetically regulate flagella formation via transcriptional regulation of IFT. Previous work has demonstrated that KDM3A is involved in the regulation of flagella formation during spermatogenesis through its epigenetic role in chromatin remodeling.[4,5] To gain a comprehensive picture of the possible regulation of IFT by KDM3A, measurements of *IFT* transcription levels, in the presence or absence of KDM3A expression, were performed using Real Time quantitative Reverse Transcription PCR (qRT-PCR).

qRT-PCR is a widely used method to determine the transcript level of a gene in cells or tissue samples. The 'Real-Time' term indicates that measurements are performed during amplification instead of at the end of the reaction. The results achieved by this method can be quantitative with respect to the transcription levels of the target gene. This technique measures messenger RNA (mRNA) levels by using reverse transcriptase to convert the mRNA into the complementary DNA (cDNA) sequence which is then amplified by PCR.[29]

The qRT-PCR assay described herein uses the SYBR® Green fluorescence dye (Life technologies). This dye intercalates with double-stranded DNA (dsDNA), resulting in increased fluorescence an increase in the fluorescence signal is proportional to the dsDNA produced during the reaction.[30] The threshold cycle (C_t) is the cycle at which fluorescence due to amplification is greater than a predefined threshold. A lower C_t correlates with higher transcript expression levels. The difference in the C_t values between one treatment and a selected calibrator treatment, which is a reference treatment used for calibration, of the same gene is known as the 'delta C_t' (**Figure 5.11**).[29]

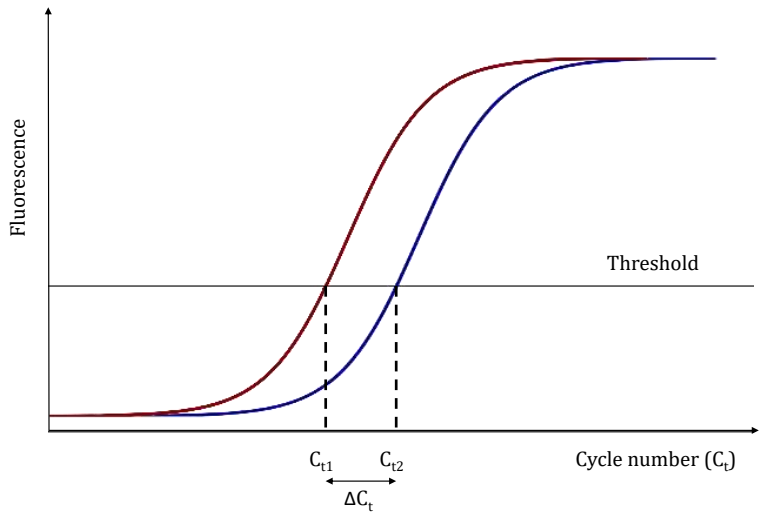


Figure 5.11 Illustration of qRT-PCR plot. Red plot represents transcript amplification during a selected calibrator treatment, blue plot represents transcript amplification during a tested treatment for the same gene. The threshold is an arbitrary level of fluorescence chosen to be above baseline variability. C_t is the cycle number at which the fluorescence signal exceeds the threshold. ΔC_t is the difference in the C_t values between the two treatments.

If primer efficiency was perfect, each amplification cycle would double the level of cDNA. Since in many cases primer efficiency is imperfect, the fold increase is a function of $(\text{efficiency})^n$, where n is the number of cycles. M.W. Pfaffl has developed a mathematical model for relative quantification in RT-PCR which takes into account the efficiency of the amplification for each primer (**Equation 5.1**).[31]

$$\text{Ratio} = (E_{\text{target}})^{\Delta C_t \text{ target (calibrator-treatment)}} / (E_{\text{ref}})^{\Delta C_t \text{ ref (calibrator-treatment)}}$$

Equation 5.1 The Pfaffl formula. E is primer efficiency, target is the target gene, ref is the reference control and ΔC_t is difference in the C_t values between one treatment and a selected calibrator treatment of the same gene.

The qRT-PCR technique is one of the most accurate and sensitive methods for measuring RNA levels.[32] However, since it is so sensitive it is essential to calibrate the conditions to avoid misleading results. Any genomic DNA contamination could bind the SYBR® Green dye and generate a false increase in signal. A false signal could also be a result of primer dimerization or suboptimal primer efficiency. In addition, a difference in the extraction of

mRNA between samples and performance of reverse transcription and PCR itself can also generate a false signal, and therefore normalisation using an appropriate reference gene is required.[33,34] All of these factors should be considered when designing qRT-PCR experiments.[35]

The next sections describe the design of a qRT-PCR experiment and the resultant data.

5.10.1 Primer design for maximising selectivity and efficiency

To get reliable mRNA transcription data, it is essential to design efficient and specific primers. For this assays, primers were either designed according to existing validated primers published in PrimerBank, or by using Primer-BLAST, a tool developed by the NCBI.[36,37] The Primer-BLAST tool checks primer pair selectivity against a selected database. To avoid binding of the primers to genomic DNA, primers were designed such that they must span an exon-exon junction in addition to the default design parameters (e.g. PCR product size and primer melting temperatures). A list of the selected primers is shown in **Table 5.5** of Section 5.10.2.

5.10.2 Determination of primer efficiency using standard curves

Standard curves are used as a quality control test for primers intended to be used in qRT-PCR assays to determine their efficiency. The ideal primer efficiency is 2 (100%), meaning that the amount of DNA is doubled at each amplification cycle. For accurate and reliable results it is highly recommended to use primers with efficiency in the range of 90% - 110%.[35]

In the assay, samples without cDNA template were used as controls to detect primer dimers. In addition, a control experiment in which RNA samples from each treatment were incubated without the reverse transcriptase enzyme was performed to test for the presence of genomic DNA contamination. To test for primer efficiency, cDNA from all samples to be tested in the experiment was mixed in equal amounts. Then, a two-fold serial dilution of this cDNA mixture was performed. For each primer set the threshold cycle (C_t) value was plotted for each of the dilutions against the log cDNA concentration (**Figure 5.12**).

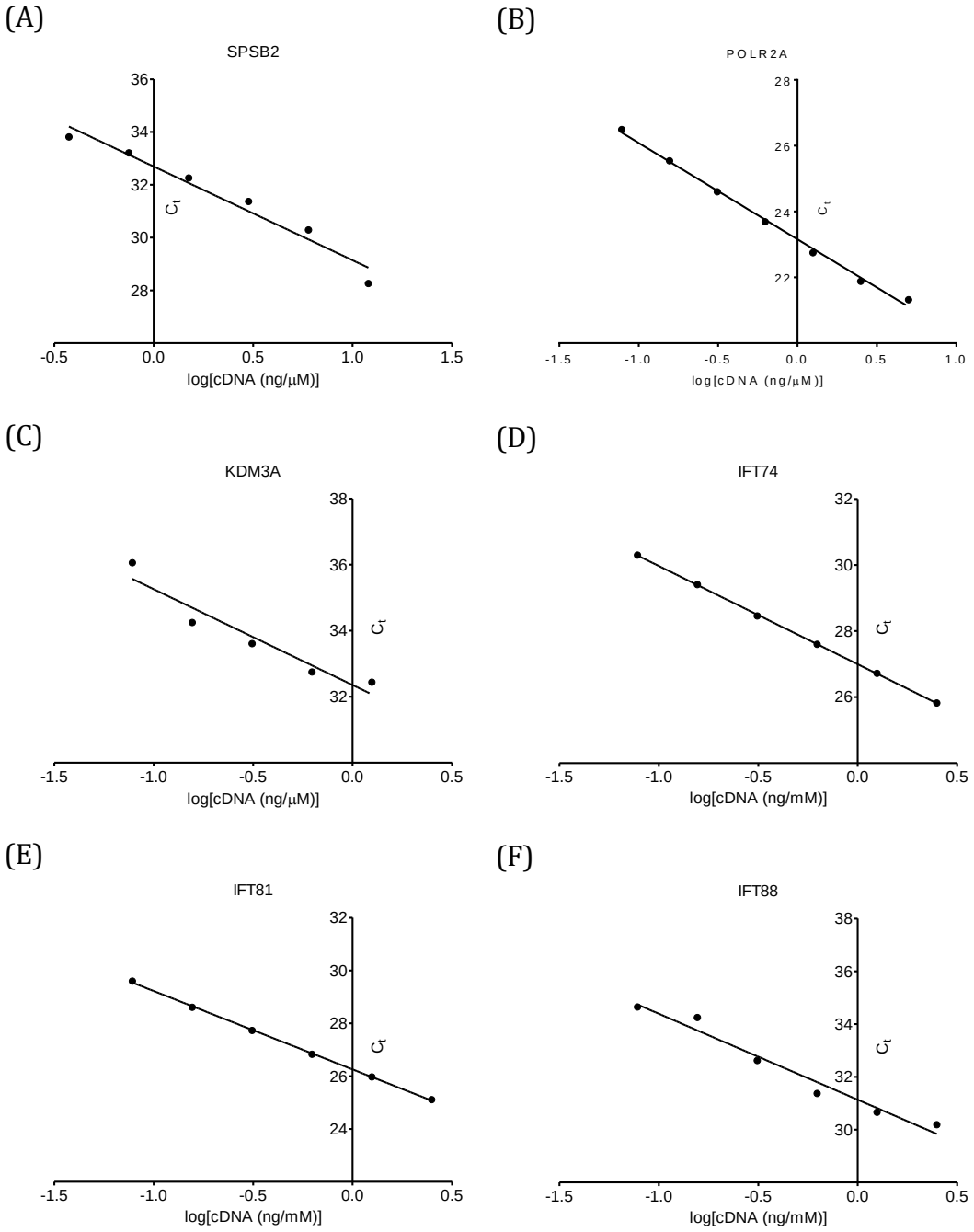


Figure 5.12 Standard curves for primer efficiency assay. For each primer set C_t values are plotted against a two-fold serial dilution of cDNA. cDNA in these assays was a 1:1 pool of cDNA from all the samples to be used in the assay (three biological repeats for each of the two genotypes and the two culture conditions). Primer concentrations were 900 nM (see Materials and methods, Chapter 11, Section 11.2.4.5, for more information). Primer sets were designed to test the transcription levels of the following genes: (A) *SPSB2*, (B) *POLR2A*, (C) *KDM3A*, (D) *IFT74*, (E) *IFT81* and (F) *IFT88*.

A linear fit was then used to calculate primer efficiency according to **Equation 5.2**.

$$Efficiency = 10^{-1/slope}$$

Equation 5.2 Calculation of efficiency using standard curve slope. Slopes between -3.1 and -3.6 gives reaction efficiencies between 90 and 110% which are typically acceptable.[35]

The C_t range for quantification is the dynamic range for each primer set in which the primer efficiency is optimal. During the first test some of the primers gave good correlation, but low efficiency, i.e. they were linear with a shallow gradient. The efficiency was optimised by increasing the primer concentration from 600 nM to 900 nM (see Materials and methods, Chapter 11, Section 11.2.4.5, for more information). In addition to primer efficiency measurements, melt curve analyses were carried out to determine whether a single product was produced in each amplification as indicated by a single peak in the curve (*data not shown*). **Table 5.5** summarises the sequence, C_t range, efficiency and the linear fit R^2 correlation for the chosen primer sets.

Table 5.5 Summary of standard curves data showing the sequence, C_t range, efficiency and R^2 correlation for each primer pair. Efficiencies were calculated based on the slopes of the standard curves (see Materials and methods, Chapter 11, Section 11.2.4.5, for more information). F- refers to forward sequence, R- refers to reverse sequence.

Gene	Sequence	C_t range	Efficiency (%)	R^2
SPSB2	**F-** CCGGGGTAAGAGGGGCTATT	28.3-33.8	96	0.96
	R- CCGCGTAGTGGTCAGTCTG			
POLR2A	**F-**GGGTGGCATCAAATACCCAGA	21.3-26.5	110	1.00
	R- AGACACAGCGCAAAACTTTCA			
KDM3A	**F-** AGGGGAGAGAAAGGGAGGAG	32.4-36.1	110	0.92
	R- TCCCAAACCAGCTTTGTCCA			
IFT74	**F-** TCAAGAGGTGGAGTTGGGTTA	25.8-30.3	108	1.00
	R- CATTGCAGTTGCCACTCGAAT			
IFT81	**F-**TCTCAATAAGGAGCCCTTTAGGA	25.1-29.6	109	1.00
	R- CACCAAACCCTGACGAAAAGT			
IFT88	**F-** TCCTGAAACTTCACGCAATCC	30.2-34.7	101	0.96
	R- GACCACCTGCATTAGCCATTC			

5.10.3 Reference gene selection

For accurate qRT-PCR, which allows comparison of mRNA transcription between different samples, it is necessary to choose an appropriate reference gene to be used for data normalisation. A good reference gene is one in which the transcription level stays constant under the different experimental procedures.[33] Differences in the total amount of cDNA in an assay can be formed by different amounts of total RNA, the quality of the RNA and variability during the synthesis of cDNA from RNA. Since a reference gene is exposed to the same procedure as the tested genes, its cDNA amplification allows the differences in total cDNA levels to be normalised between samples.[38] Primers for DNA-dependent RNA polymerase II (*POLR2A*) and for *Homo sapiens* splA/ryanodine receptor domain and SOCS box containing 2 (*SPSB2*) were tested as candidates for reference genes; they are both involved in housekeeping mechanisms (*POLR2A,* NCBI gene ID 5430, is involved in

transcription and *SPSB2,* NCBI gene ID 84727 is involved in protein degradation), and are therefore expected to have stable expression between the tested treatments.

To select a reference gene, the C_t levels of *POLR2A* and *SPSB2* cDNA were measured across two genotypes and two culture conditions for each reference gene candidate in hTert-RPE1 cells. Cells either expressed WT KDM3A or were *KDM3A* null cells (*KDM3A^{cr}22.2; CRISPR/Cas9 22 clone 2, cloned by Dr. Patricia Yeyati, manuscript accepted*), and were treated to either induce cilia resorption or growth. C_t values were calculated as the average of three biological repeats for each genotype and culture conditions, each with three technical measurements. Primers used for these measurements were validated in Section 5.10.2 (see Materials and methods, Chapter 11, Section 11.2.4.5, for more information).

POLR2A had a slightly lower C_t variability across the different samples compared to *SPSB2* (standard deviation of 0.4 vs. 0.5 respectively, **Figure 5.13**). In addition, *POLR2A* had a stronger mRNA expression than *SPSB2* (average C_t of 22.2 vs. 28.4 respectively). Therefore, *POLR2A* was chosen as a reference gene for normalisation.

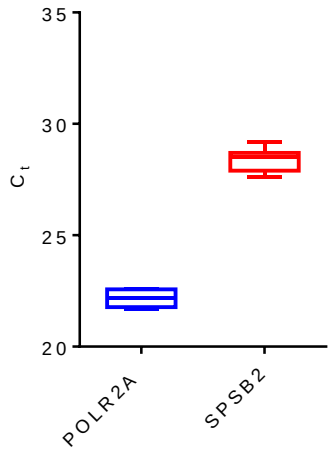


Figure 5.13 Reference gene selection. The box and whisker plot shows the median, lower/upper percentiles and the variability of the C_t values across two genotypes and two culture conditions for each reference gene candidate in hTert-RPE1 cells. Cells were either WT or null *KDM3A* cells, and were treated to either induce cilia resorption or growth. C_t values were calculated as the average of three technical measurements. Box and whiskers range represent the C_t variability over three biological repeats for each genotype and culture condition for expression of each gene (blue – *POLR2A* and red - *SPSB2*). *Dr. Patricia Yeyati carried out the cell cultures.*

5.10.4 Determination of the dependence of IFT expression levels on KDM3A expression

Sections 5.10.1-5.10.3 describe the development of a qRT-PCR method to determine whether KDM3A regulates *IFT* mRNA levels.

Human hTert-RPE1 cells expressing WT KDM3A or *KDM3A* null cells (*KDM3A*$^{cr22.2}$; CRISPR/Cas9 exon 22 clone 2, *cloned by Dr. Patricia Yeyati, manuscript accepted*) were treated to either induce cilia resorption or growth, each genotype and condition had three biological repeats. cDNA was used in qRT-PCR experiments to measure the levels of *IFT74, IFT81, IFT88* and *KDM3A* transcripts (see Materials and methods, Chapter 11, Section 11.2.4.5, for more information). Gene expression was measured with three technical replicates by qRT-PCR for each biological repeat.

Primer amplification efficiency was corrected by relative quantification based on the Pfaffl method, in which the measured efficiency for each primer set is incorporated into the amplification calculation.[31] For each gene, the average expression values were calculated, normalised with a reference gene (*POLR2A*) and compared to mRNA levels of an arbitrary selected calibrator treatment (resorption, expressing WT *KDM3A*) which was set to 100% (see Materials and methods, Chapter 11, Section 11.2.4.5, for more information).

As expected, the *KDM3A* null cells had much lower expression levels of *KDM3A* mRNA than WT KDM3A cells, regardless of the culture conditions (over four-fold, $P > 0.05$ and $P \leq 0.01$ for resorption and growth respectively, **Figure 5.14** A). *KDM3A* expression in the null cells still exists since the CRISPR/Cas9 editing of exon 22 in KDM3A null cells rendered the transcript out of frame, therefore some basal low levels of KDM3A transcript remains depending on the efficiency of nonsense mediated decay.[39] The mRNA levels of *IFT74*, *IFT81* and *IFT88* were unchanged between the different genotypes (WT or null *KDM3A*), suggesting that KDM3A is not an epigenetic regulator for these genes (no statistically significant changes, **Figure 5.14** B-D).

Interestingly, similar to measurement of IFT81 protein levels (see Section 5.9), the mRNA levels of the tested *IFT* genes between the different culture conditions was relatively constant (no statistically significant changes). These results suggest that transcription of the tested *IFT* genes during ciliogenesis may be constant, and correlate with reports by Engel *et al.* stating that a constant quantity of IFT proteins per cilium is maintained during ciliogenesis.[28]

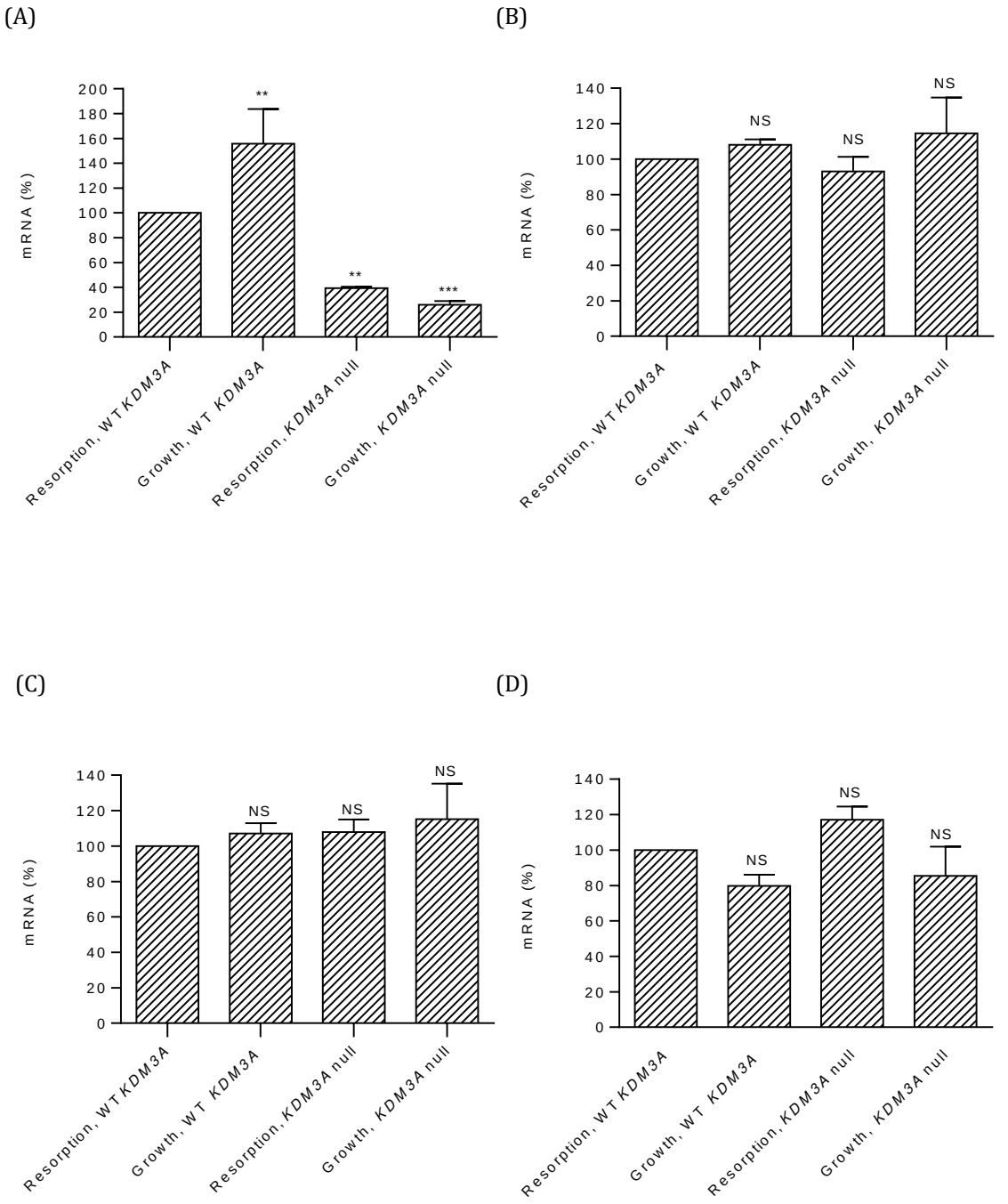


Figure 5.14 Relative mRNA expression of *KDM3A*, *IFT74*, *IFT81* and *IFT88* across two genotypes and two culture conditions in hTert-RPE1 cells. Cells were either WT or null *KDM3A* cells, and were treated to either induce cilia resorption or growth. Gene expression was measured by qRT-PCR. Efficiency was corrected by relative quantification based on the Pfaffl method.[31] Average expression values were calculated, normalised with a reference gene (*POLR2A*) and compared to mRNA levels of a selected calibrator treatment (resorption, WT *KDM3A*) which was arbitrarily set to 100%. Statistical analyses were carried out using two way ANOVA followed up by Dunnett's multiple comparisons tests using GraphPad Prism 6. NS denotes $P > 0.05$, ** denotes $P \leq 0.01$ and *** denotes $P \leq 0.001$. See Materials and methods, Chapter 11, Section 11.2.4.5, for more information. (A) - (D) The relative mRNA expression of *KDM3A, IFT74, IFT81* and *IFT88*, respectively. Bars are mean values ± S.D., representing three biological repeats for each genotype and culture condition, with three technical replicates for each biological repeat. *Dr. Patricia Yeyati carried out the preparations of cell pellets.*

5.11 Investigating KDM3A enzymatic activity on additional IFT PTMs

The results described in Chapters 6 and 7 describe the apparent activity of KDM3A on methylated arginine and acetylated lysine residues on histone fragment peptides.

In light of these novel data, the potential of IFTs to be substrates of KDM3A was retested. Of particular interest was the case of methylated arginine residues as these were identified on IFT proteins during resorption.[22] IFT fragment peptides containing either methylated arginines or acetylated lysines were designed and synthesised. These peptides were screened as substrates for KDM3A using MALDI-TOF MS based activity assays (see Materials and methods, Chapter 11, Section 11.3.3, for more information).

The description of the assays and the results are detailed in Chapter 6, Section 6.12 and Chapter 7, Section 7.11. No activity was detected for any of the screened methylated arginine or acetylated lysine IFT fragment peptides.

5.12 Screening for IFT lysine demethylation by other KDMs

The work by Dr. Patricia Yeyati and Dr. Holger Kramer has identified dimethyllysine residues on IFT-B proteins using LC-MS/MS. In this work, IFT fragment peptides were designed and tested as substrates for KDM3A. However, no catalysis was detected (see Section 5.4). To investigate whether lysine methylated IFTs are regulated by another KDM, activity tests were expanded to include recombinant forms of KDM4A, KDM4C, KDM7B (PHF8), and JMJD1C.

These KDMs were tested for activity against IFT74(71-90)K81me2, IFT81(536-555)K545me2 and IFT88(61-80)K71me2 peptides. For each KDM, a lysine methylated histone peptide known to be recognised by this demethylase was used as a positive control. The JMJD1C substrate is still unknown and so no positive control was available for this enzyme.[40]

Reactions were carried out under the standard conditions with incubation of 2 μM enzyme and 10 μM peptide for 1 h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3,

for more information). The only difference for KDM7B assays was the addition of salt to the reaction buffer (500 mM NaCl). Demethylated products were observed by MALDI-TOF MS for the positive control experiments (i.e., -14 Da shifts in the mass of the methylated peptide) compared to the no-enzyme control experiment, indicated demethylation activity. No reactions were observed for any of the tested IFT peptides following incubation with the tested KDMs (**Table 5.6** and **Figure 5.15**).

Table 5.6 Summary of recombinant KDM enzymes, their control histone substrate fragment peptide and the IFT fragment peptides tested in this study. - denotes lack of control peptide. X denotes absence of activity as detected by MALDI-TOF MS. *KDM4A and KDM7B were provided by Dr. Louise Walport, KDM4C was provided by Dr. Akane Kawamura and* JMJD1C was provided by *Dr. Catrine Johansson.*

Enzyme	Control peptide	IFT74(71-90)K81me2	IFT81(536-555)K545me2	IFT88(61-80)K71me2
KDM4A	H3(1-15)K9me3	X	X	X
KDM4C	H3(6-21)K9me3	X	X	X
KDM7B (PHF8)	H3(1-14)K4me3K9me2	X	X	X
JMJD1C	-	X	X	X

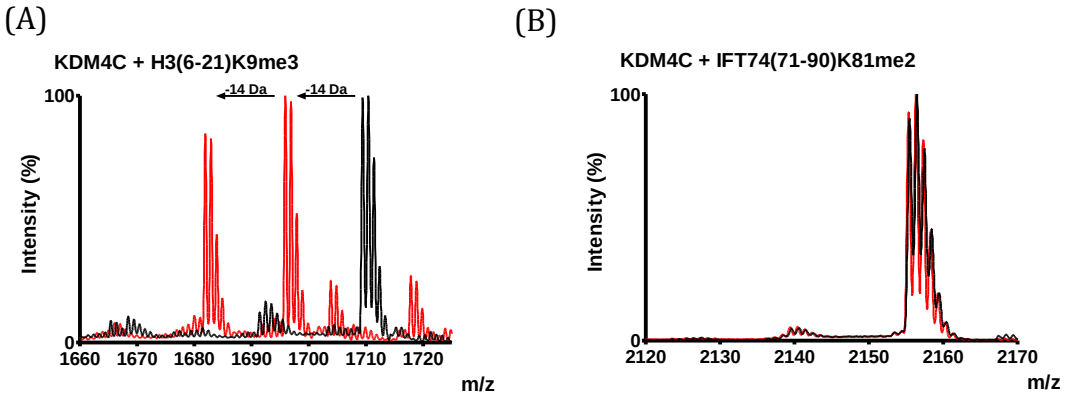


Figure 5.15 Sample MALDI-TOF MS spectra from the screen of IFT peptides showing the absence of KDM4C catalytic activity for IFT74(71-90)K81me2. (A) KDM4C demethylates the positive control peptide H3(6-21)K9me3. (B) Absence of KDM3A catalytic activity for IFT74(71-90)K81me2 peptide. Peptides (10 μM) were incubated with KDM4C (2 μM) under standard conditions, followed by analyses by MALDI-TOF MS (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Reactions with enzyme are shown in red, control reactions with no enzyme are shown in black. *KDM4C was provided by Dr. Akane Kawamura.*

5.13 Discussion

KDM3A is known to have an important role in male fertility, and is involved in sperm formation.[1,3–5] KDM3A is also known to interact with non-histone proteins in a catalytic and non-catalytic manner (see Introduction, Chapter 1), and was observed outside the nucleus at several cytoplasmic sites.[3,4]

Work done by collaborators has identified lysine methylation sites on a number of IFT proteins. The work has also shown that KDM3A co-purified with IFT81 and IFT88 during differential centrifugation and with IFT81 during immunopurification. In addition, work with hTert-RPE1 cells showed that over-expressing IFT81 promotes ciliogenesis in *KDM3A* null cells but not WT cells, suggesting that KDM3A negatively regulates ciliogenesis in response to increased availability of IFT81.

The work described in this chapter aimed to further investigate whether KDM3A regulates ciliogenesis thorough the regulation of IFT, and focuses on IFT74, IFT81 and IFT88, as these are the IFT-B proteins judged most likely to have methylated lysines. The work involved investigation at several levels of possible regulation from gene expression to post-translational regulation of KDM3A on IFT.

Gene expression analyses using qRT-PCR have indicated that *IFT* transcript levels are not dependent on KDM3A expression. These results suggest that KDM3A is unlikely to be a transcriptional regulator of *IFT74, IFT81 and IFT88*. Measurements of IFT81 protein levels, in the presence or absence of KDM3A, have indicated that KDM3A also does not regulate IFT81 protein levels.

No KDM3A demethylation activity was observed for the tested methylated IFT fragment peptides, which cover eight different potential methyl lysine positions on IFT74, IFT81 and IFT88. The results of experiments with recombinant KDM3A were in agreement with assays using full-length KDM3A. Furthermore, no KDM3A catalytic activity was observed in tests with IFT fragment peptides containing methylated arginines and acetyllysines, although KDM3A was observed to modify these PTMs when they are incorporated into histone peptides (see Chapter 6, Section 6.12 and Chapter 7, Section 7.11). Although KDM3A exhibits promiscuity in terms of the variety of modifications it catalyses, to date all of these novel reactions occurred on the H3 K9 position (see Chapter 6, Section 6.3, Chapter 7, Section 7.4.2 and Chapter 9, Section 9.3). Therefore, the available evidence is

that KDM3A seems to be highly sequence specific which might explain the lack of activity on IFT.

It is important to note that the observed lack of activity under the tested *in vitro* assay conditions could be due to the fact that the peptide fragments lack the conformation of the intact IFT proteins around the modified lysine sites. In contrast, using peptide substrates to detect activity on histone tail modifications may be easier as the tail is largely unstructured. Another drawback of these *in vitro* assays is that recognition of IFT by KDM3A might be dependent on other areas of the IFT proteins which are not represented in the tested peptides and longer peptides are required, similar to observations on length requirement for substrate peptides of hypoxia-inducible factor prolyl 4-hydroxylases.[41] In addition, a combination of post-translational modifications on IFT might be required to induce KDM3A catalytic activity, similar to results described in Chapter 7. All in all, *in vitro* experiments with isolated full length IFT proteins would have been preferable to test if IFT are substrates of KDM3. However, such work was beyond the scope of this thesis.

Thermal stability assays imply that KDM3A interacts with the tested IFT peptides, although interaction may not be specific. Inhibition assays using MALDI-TOF MS further show inhibition of KDM3A by IFT. However, these assays do not imply whether this interaction is specific or not, nor do they indicated the site on KDM3A in which this interaction occur i.e. whether it binds in the active site.

Fluorescence polarisation experiments were designed in order to validate the binding between IFT peptides and KDM3A. Unfortunately, although these assays gave robust data for a control protein (KDM4E), they were not reliable when KDM3A was tested due to KDM3A precipitation at high protein concentrations (*data not shown*). Follow up binding experiments can be developed using other methods that require lower protein concentration such as Microscale Thermophoresis (MST) and BioLayer Interferometry (BLI).[42,43]

Interestingly, the presence of the dimethyllysine residue in the IFT81 fragment peptide was responsible for a greater inhibitory effect on KDM3A than the corresponding non-methylated peptide (IC_{50} of 24 μM vs. 270 μM respectively), indicating that methylation strengthens the interaction between KDM3A and IFT81. The methylated form of IFT81 was therefore hypothesised to bind KDM3A by mimicking the dimethylated histone substrate and function as a competitive inhibitor. However, the results of kinetic characterisation of KDM3A inhibition implies that an uncompetitive inhibition mechanism

is most likely to occur, where dimethylated IFT81 only bind after the enzyme – substrate complex is formed. However, uncompetitive inhibition is rare and confirmation and further investigation should be carried out. Another hypothesis may be that IFT81 binds KDM3A through a methyllysine residue but instead of binding to the active site it binds in another site. H3 K4me3 was found to be important in KDM3A catalysis of novel substrates (Chapter 7, Section 7.4.2 and Chapter 9, Section 9.3).

KDM3A could potentially have a secondary methyl binding site in addition to the methylated substrate binding site, through which methylated IFT81might bind.

It is also possible that the KDM3A *in-vitro* inhibition results and the immunopurification data reflect non-specific binding (i.e. the results are not biologically relevant). Thus, KDM3A might not interact with IFT in cells. Instead, the regulation of KDM3A on ciliogenesis could be due to an indirect effect whereby KDM3A regulates other proteins related to ciliogenesis and these in turn modulate the IFT response. Indirect regulation might also explain the negative effect of KDM3A on ciliogenesis when over-expressing IFT81 in cells (see Section 15.1). More work to decipher the role of KDM3A in ciliogenesis is required.

Assuming that the *in vitro* results reflect an interaction between IFT proteins and KDM3A in cells, they imply that non-catalytic interactions take place. Similar to the non-catalytic regulation of KDM3A on Gli1, KDM3A could regulate IFT in a demethylase-independent manner. For instance, KDM3A could bind K545me2 on IFT81 and thereby protect this modification from removal by another demethylase. In this work, no activity by KDM4A, KDM4C, KDM7B or JMJD1C was detected; however, other KDMs might catalyse IFT demethylation.

SPOT peptide arrays if IFTs stained with recombinant KDM3A can be useful for identifying and mapping the exact positions of the interaction between KDM3A and IFT proteins *in vitro*.[44] One way to determine whether KDM3A regulates the levels of dimethyllysine on IFT would be to perform quantitative mass spectrometry experiments where IFT methylation between cells with WT and null *KDM3A* genotypes could be compared. Such an assay can be carried out by using Stable Isotope Labelling with Amino acids in Cell culture (SILAC) experiments.[45]

5.14 References

1. Kuroki, S. *et al.* Epigenetic regulation of mouse sex determination by the histone demethylase Jmjd1a. *Science* **341,** 1106–9 (2013).

2. Hirsh, A. Male subfertility. *BMJ* **327,** 669–72 (2003).

3. Kasioulis, I. *et al.* Kdm3a lysine demethylase is an Hsp90 client required for cytoskeletal rearrangements during spermatogenesis. *Mol. Biol. Cell* **25,** 1216–1233 (2014).

4. Okada, Y., Scott, G., Ray, M. K., Mishina, Y. & Zhang, Y. Histone demethylase JHDM2A is critical for Tnp1 and Prm1 transcription and spermatogenesis. *Nature* **450,** 119–23 (2007).

5. Liu, Z. *et al.* Jmjd1a demethylase-regulated histone modification is essential for cAMP-response element modulator-regulated gene expression and spermatogenesis. *J. Biol. Chem.* **285,** 2758–70 (2010).

6. Ishikawa, H. & Marshall, W. F. Ciliogenesis: building the cell's antenna. *Nat. Rev. Mol. Cell Biol.* **12,** 222–234 (2011).

7. Quarmby, L. M. & Parker, J. D. K. K. Cilia and the cell cycle? *J. Cell Biol.* **169,** 707–10 (2005).

8. Oh, E. C. *et al.* Metabolic regulation and energy homeostasis through the primary cilium. *Cell Metab.* **21,** 21–31 (2015).

9. Plotnikova, O. V., Pugacheva, E. N. & Golemis, E. A. in *Methods in Cell Biology* **94,** 137–160 (2009).

10. Rosenbaum, J. L. & Witman, G. B. Intraflagellar transport. *Nat. Rev. Mol. Cell Biol.* **3,** 813–825 (2002).

11. Pedersen, L. B., Veland, I. R., Schrøder, J. M. & Christensen, S. T. Assembly of primary cilia. *Dev. Dyn.* **237,** 1993–2006 (2008).

12. Deane, J. A., Cole, D. G., Seeley, E. S., Diener, D. R. & Rosenbaum, J. L. Localization of intraflagellar transport protein IFT52 identifies basal body transitional fibers as the docking site for IFT particles. *Curr. Biol.* **11,** 1586–1590 (2001).

13. Qin, H. *et al. Intraflagellar Transport Is Required for the Vectorial Movement of TRPV Channels in the Ciliary Membrane*. *Current Biology* **15,** (2005).

14. Qin, H., Rosenbaum, J. L. & Barr, M. M. *An autosomal recessive polycystic kidney disease gene homolog is involved in intraflagellar transport in C. elegans ciliated sensory neurons*. *Current Biology* **11,** (2001).

15. Blacque, O. E. & Leroux, M. R. Bardet-Biedl syndrome: an emerging pathomechanism of intracellular transport. *Cell. Mol. Life Sci.* **63,** 2145–2161 (2006).

16. Hao, L. & Scholey, J. M. Intraflagellar transport at a glance. *J. Cell Sci.* **122,** 889–92 (2009).

17. Lucker, B. F. *et al.* Characterization of the intraflagellar transport complex B core: direct interaction of the IFT81 and IFT74/72 subunits. *J. Biol. Chem.* **280,** 27688–96 (2005).

18. Pazour, G. J. *et al.* Chlamydomonas IFT88 and its mouse homologue, polycystic kidney disease gene tg737, are required for assembly of cilia and flagella. *J. Cell Biol.* **151,** 709–18 (2000).

19. Absalon, S. *et al.* Intraflagellar Transport and Functional Analysis of Genes Required for Flagellum Formation in Trypanosomes. *Mol. Biol. Cell* **19,** 929–944 (2007).

20. Bhogaraju, S. *et al.* Molecular basis of tubulin transport within the cilium by IFT74 and IFT81. *Science* **341,** 1009–12 (2013).

21. Khoury, G. A., Baliban, R. C. & Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **1,** (2011).

22. Schneider, M. J., Ulland, M. & Sloboda, R. D. A protein methylation pathway in Chlamydomonas flagella is active during flagellar resorption. *Mol. Biol. Cell* **19,** 4319–27 (2008).

23. Werner-Peterson, R. & Sloboda, R. D. Methylation of Structural Components of the Axoneme Occurs During Flagellar Disassembly. (2013).

24. Hamada, S. *et al.* Synthesis and activity of N-oxalylglycine and its derivatives as Jumonji C-domain-containing histone lysine demethylase inhibitors. *Bioorg. Med. Chem. Lett.* **19,** 2852–5 (2009).

25. Yamane, K. *et al.* JHDM2A, a JmjC-Containing H3K9 Demethylase, Facilitates Transcription Activation by Androgen Receptor. *Cell* **125,** 483–495 (2006).

26. Schneider, P. *et al.* Identification of a novel actin-dependent signal transducing module allows for the targeted degradation of GLI1. *Nat. Commun.* **6,** 8023 (2015).

27. Berg, J. M., Tymoczko, J. L. & Stryer, L. *Biochemistry*. (W H Freeman, 2002).

28. Engel, B. D., Ludington, W. B. & Marshall, W. F. Intraflagellar transport particle size scales inversely with flagellar length: revisiting the balance-point length control model. *J. Cell Biol.* **187,** 81–89 (2009).

29. Martin Whitman, T. S. S. at I. qPCR terminology—what does it mean? (2012).

30. Technologies, L. SYBR ® Select Master Mix User Guide. (2013).

31. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29,** 45e–45 (2001).

32. Gachon, C., Mingam, A. & Charrier, B. Real-time PCR: what relevance to plant studies? *J. Exp. Bot.* **55,** 1445–54 (2004).

33. Kozera, B. & Rapacz, M. Reference genes in real-time PCR. *J. Appl. Genet.* **54,** 391–406 (2013).

34. Huggett, J., Dheda, K., Bustin, S. & Zumla, A. Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun.* **6,** 279–284 (2005).

35. Quellhorst, G. (SABiosciences C. *A Systematic Guideline for Developing the Best Real-Time PCR Primers*. (2008).

36. Wang, X., Spandidos, A., Wang, H. & Seed, B. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res.* **40,** D1144–D1149 (2012).

37. Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13,** 134 (2012).

38. Radonić, A. *et al.* Guideline to reference gene selection for quantitative real-time PCR. *Biochem. Biophys. Res. Commun.* **313,** 856–862 (2004).

39. Baker, K. E. & Parker, R. Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Curr. Opin. Cell Biol.* **16,** 293–299 (2004).

40. Brauchle, M. *et al.* Protein complex interactor analysis and differential activity of KDM3 subfamily members towards H3K9 methylation. *PLoS One* **8,** e60549 (2013).

41. Koivunen, P., Hirsilä, M., Kivirikko, K. I. & Myllyharju, J. The length of peptide substrates has a marked effect on hydroxylation by the hypoxia-inducible factor prolyl 4-hydroxylases. *J. Biol. Chem.* **281,** 28712–20 (2006).

42. Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D. & Duhr, S. Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Commun.* **1,** 100 (2010).

43. Sultana, A., Lee, J. E., Sultana, A. & Lee, J. E. in *Current Protocols in Protein Science* 19.25.1-19.25.26 (John Wiley & Sons, Inc., 2015). doi:10.1002/0471140864.ps1925s79

44. Amartely, H., Iosub-Amir, A. & Friedler, A. Identifying protein-protein interaction sites using peptide arrays. *J. Vis. Exp.* e52097 (2014). doi:10.3791/52097

45. Ong, S.-E., Mittler, G. & Mann, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods* **1,** 119–26 (2004).

Chapter 6
Investigations into arginine demethylation by KDM3A

6.1 Introduction

Arginine methylation is a common post-translational modification (PTM), that is present on both nuclear and cytoplasmic proteins.[1,2] In mammals, arginine methylation is reported on the *N*ω-nitrogens, resulting in mono-, symmetric and asymmetric di- methyl forms (**Figure 6.1**).[3] The most abundant form of methylated arginine is an asymmetric dimethylarginine (ADMA). The levels of the symmetric dimethylated (SMDA) and the monomethylated (MMA) derivatives were found to be less than half of that of ADMA.[3]

The non-protonated form of the guanidino group of arginine contains four potential hydrogen bond donors. Methylation on arginine reduces the H-bonding capacity, increases the hydrophobicity and affects the structure of the guanidino group.[4] The methylated arginine structures may alter the affinity for binding partners, and may therefore regulate the biological consequences of these interactions such as protein stability, cellular localisation and transcription.[5–7]

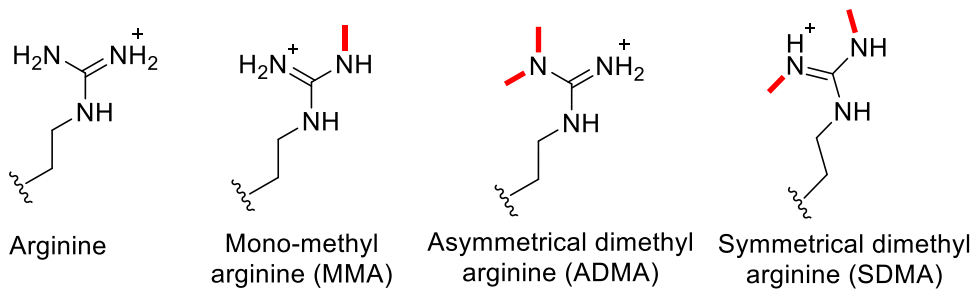


Figure 6.1 Identified forms of methylated arginines in mammals. Methylation is shown with red bonds.

Arginine methylation is especially common in shuttle proteins that regulate eukaryotic flagella and cilia dynamics.[8] As described in Chapter 5, Section 5.1, flagella and cilia can assemble and disassemble at the distal tip, a process that requires intraflagellar transport (IFT) proteins. During resorption, a number of flagellar proteins, including some IFT proteins, become methylated on arginine residues.[9,10] This observation suggests that arginine methylation may have a role in the flagellar disassembly process. Arginine methylation can disrupt H-bonding, and increase hydrophobic interactions, both of which could alter key protein-protein interactions which are necessary for disassembly. Since most cell types must resorb their flagella or cilia before mitosis, arginine methylation of IFT might regulate this process.[11,12]

In chromatin, arginine methylation occurs at a number of positions on all of the core histones, leading to either transcriptional gene activation or repression depending on the methylation position, the methylation state and the histone affinity of the binding partners.[8]

Protein Arginine Methyltransferases (PRMT) catalyse the addition of a methyl group to arginine using S-adenosyl methionine (SAM) as a co-factor and a methyl donor group.[13] PRMTs are further classified according to the methylation state they catalyse.[14] Although these PRMT enzymes are well described in the literature, definitive products of arginine demethylases have yet to be identified. There are reports of fluctuating arginine methylation levels while the total protein levels remain unchanged, suggesting removal of this PTM by demethylases rather than protein degradation.[15,16] Lysine methylation was also once assumed to be an irreversible PTM. However, lysine demethylases were discovered in 2004, a few years after the identification of histone methyltransferases (HMTs).[17,18] Contradicting reports have identified JMJD6 as an arginine demethylase, however the initial study has not been widely reproduced biochemically and JMJD6 has since also been reported as a lysine hydroxylase.[19]

In spite of conflicting evidence for JMJD6 enzymatic activity, 2OG dependent oxygenases are still good candidates for arginine demethylases due to the variety of oxidative reactions they can perform (see Introduction, Chapter 1, **Scheme 1.7**). A putative mechanism for arginine demethylation can be implied from observations on lysine analogue peptides. A recent study showed that KDM4E can hydroxylate an N^{ε}-methylisopropyllysine containing peptide, which has a similar structure to that of asymmetric N^{ω}-dimethylarginine, indicating that there is a possibility for KDM4E to demethylate arginine in a similar manner (**Figure 6.2 A**).[20] Arginine demethylation can

occur by hydroxylation of the arginine N^{ω}-methyl group followed by a loss of formaldehyde (**Figure 6.2 B and C**), similar to the catalysis of lysine demethylation (see Introduction, Chapter 1, Section 1.13, for more information).

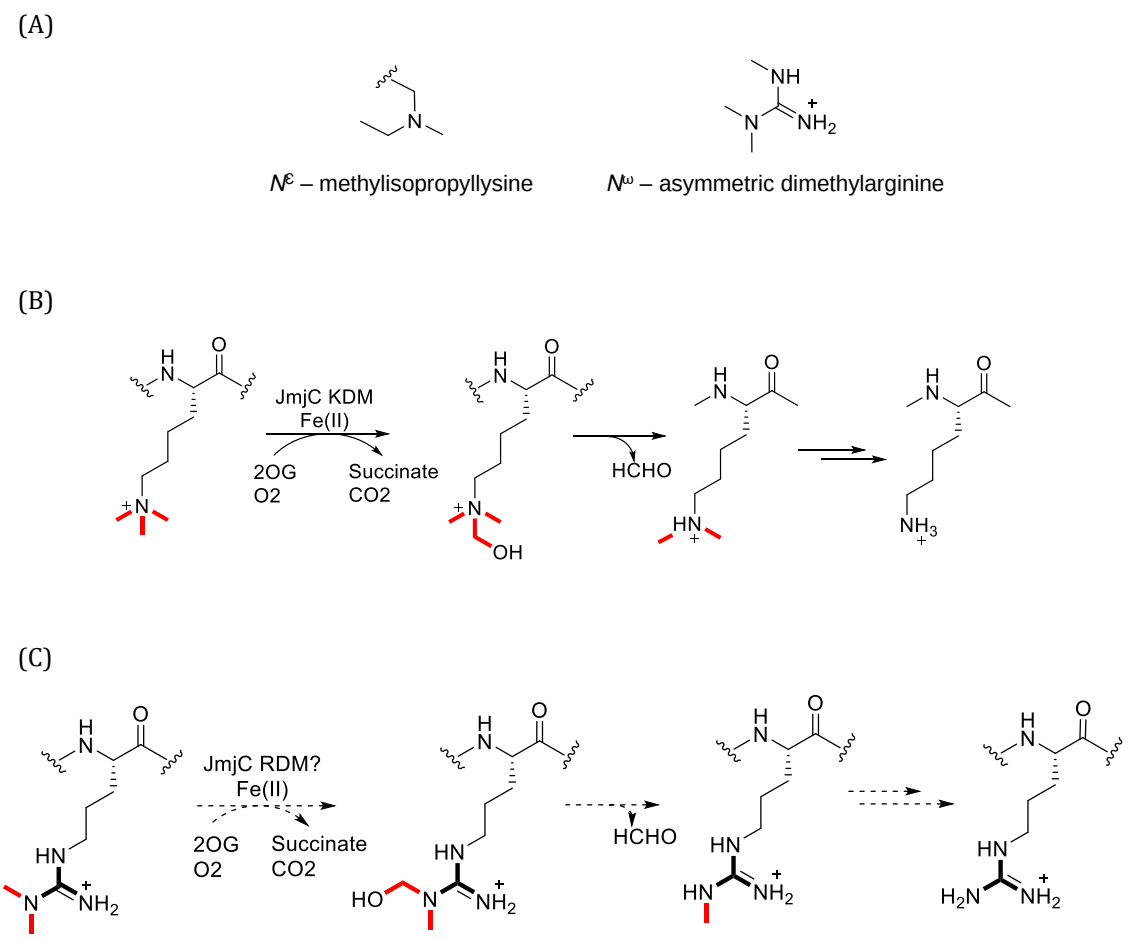


Figure 6.2 Putative outline of the catalytic mechanism for arginine demethylation by JmjC proteins. (A) Comparison between N^{ε}-methylisopropyllysine and asymmetric N^{ω}-dimethylarginine structures. (B) and (C) Comparison between the reported JmjC lysine demethylation (KDM) mechanism and the proposed arginine demethylation (RDM) catalysis respectively. Each JmjC oxidation reaction is coupled to the conversion of 2OG and oxygen to carbon dioxide, succinate and a reactive iron(IV)-oxo intermediate. This intermediate then promotes hydroxylation of the N^{ε}-methyl group to form an unstable hemiaminal which fragments to release formaldehyde and an non-methylated residue. Methyl groups are bold red lines, dashed arrows represent proposed reactions, HCHO is formaldehyde. Figure adapted with permission.[21]

Studies with lysine analogue peptides have also shown that KDM3A is enzymatically active not only on methylated lysine but on a variety of other substrates, revealing its potential

to have additional catalytic functions (see Chapter 7, Section 7.3, for more details on reactions with lysine analogues).

6.2 Objectives

The work described in this chapter aimed to investigate whether KDM3A catalyses N^{ω}-methylarginine demethylation. The work was initially focused on identifying arginine demethylation (RDM) activity with histone derived peptides using recombinant and full length overexpressed KDM3A isolated from human cells, as well as carrying out tests to determine co-factor and co-substrate dependence. Having identified KDM3A-catalysed RDM activity, more detailed kinetics were achieved to compare the catalytic efficiency of KDM3A for methylated lysine and arginine substrates. Collaborative work allowed further comparison of these kinetic parameters to that of other KDMs. A screen for arginine demethylase activity on natural arginine positions in histone tail peptides was performed. Further work included retesting KDM3A enzymatic activity in the presence of H3 K4me3. Finally, the potential for KDM3A to catalyse RDM of IFT fragment peptides was assessed.

Dr. Louise Walport produced all the recombinant KDM proteins, apart from KDM5C which was produced by Aleksandra Szykowska (SGC) and KDM3A. Dr. Louise Walport also synthesised most of the arginine methylated histone peptides, performed many of the MALDI screening experiments, and collaborated on tests with full length-IPs. Dr. Holger Kramer and Dr. Patricia Yeyati identified putative arginine methylated IFT proteins. Cloning of FLAG-tagged full length wildtype and H1120Y KDM3A constructs was carried out by Dr. Patricia Yeyati. This work was published by Walport et al.[21]

6.3 Identification of arginine demethylation by KDM3A

KDM3A and other KDMs were observed to react with lysine analogue substrates other than methylated lysine (see Chapter 7, Section 7.3).[20] Therefore, it was hypothesised that these KDMs may also be capable of reacting with other PTMs, and potentially to catalyse demethylation of methylated arginines.

Initial studies involved using modified H3 histone peptides that contained a methylated arginine residue in place of H3 K9. It was assumed that the sequence similarity between the natural and the mutated KDM3A substrate would help to direct the methylated arginine into the active site. Three forms of the corresponding methylated arginine peptides were synthesised, containing monomethylated arginine, asymmetric and symmetric dimethylarginine (MMA, ADMA and SMDA respectively). As these forms are chemically different, it was considered likely that only some may be accepted as substrates for a putative arginine demethylase.

To test whether KDM3A can catalyse demethylation of arginine residues, the arginine-methylated peptides (10 μM) and the corresponding H3(1-15)K9me2 peptide were incubated with recombinant KDM3A (2 μM) and co-factors under standard demethylation conditions for one hour at 37°C (for more information see Materials and methods, Chapter 11, Section 11.3.3 and Chapter 3, Section 3.5.1). Analyses were carried out by MALDI-TOF MS. A shift in peptide mass by -14 Da and -28 Da, compared to the non-enzyme negative controls, indicated mono and di- demethylation.

The positive control reaction with H3(1-15)K9me2 peptide indicated that KDM3A was active and can recognise the substrate sequence (**Figure 6.3 A**). Further, KDM3A was observed to demethylate all three forms of the methylated arginine residues, with almost a complete reaction for monomethyl arginine, partial reaction for the asymmetric form and limited activity for the symmetric form (**Figure 6.3 B-D** respectively).

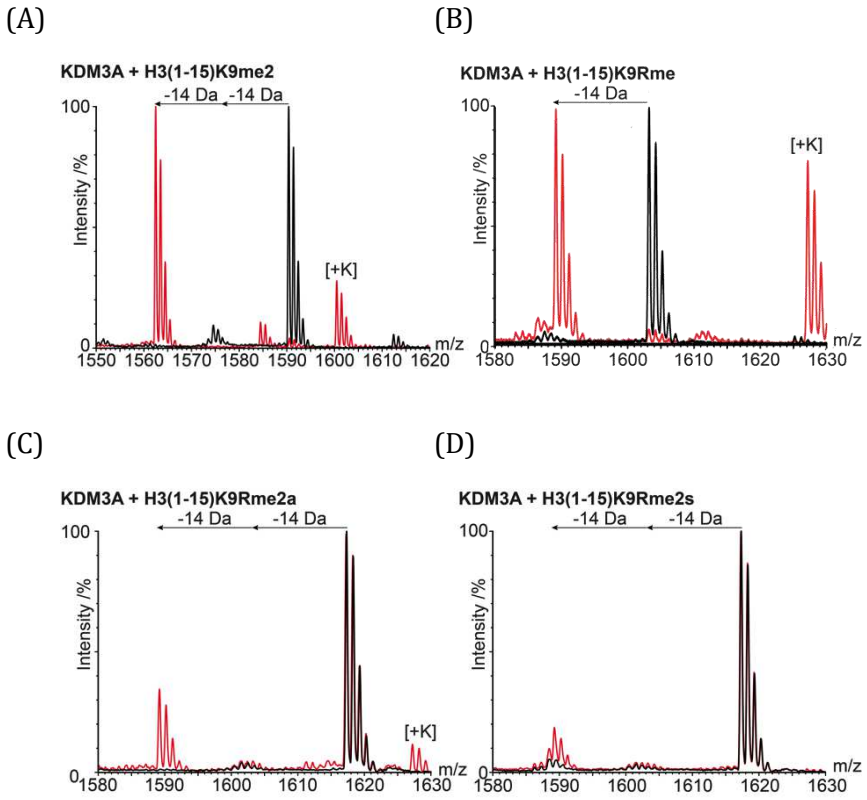


Figure 6.3 MALDI-TOF MS analyses of arginine demethylation of variant histone peptides (10 µM) as catalysed by recombinant KDM3A (2 µM) after 1 h at 37°C. (A) an H3(1-15)K9me2 histone peptide used as a positive control (B) an H3(1-15)K9Rme peptide (C) an H3(1-15)K9Rme2a peptide (D) an H3(1-15)K9Rme2s peptide. Red spectra show reactions including enzyme, control reactions with no enzyme are shown in black. Figure adapted with permission.[21] *Measurements were carried out by Dr. Louise Walport.*

6.4 Investigations into KDM3A co-substrate and co-factor dependence

Arginine demethylation by KDM3A was hypothesised to be catalysed by the JmjC domain, thus it was expected that a JmjC co-factor (Fe^{II}) and co-substrate (2OG) dependency would be observed. JmjC demethylases require Fe^{II} as a co-factor to mediate electron transfer during substrate oxidation. For substrate oxidation to occur JmjC KDMs also require 2OG as a co-substrate which is converted to succinate in a coupled reaction (**Figure 6.2**; see Introduction, Chapter 1, Section 1.13).

Experiments measuring enzymatic activity with and without co-factor (Fe^{II}) and co-substrate (2OG) were therefore performed, under identical conditions to those described

in Section 6.3. No reaction was observed in the absence of 2OG or Fe^{II}, supporting the proposed dependency on 2OG and Fe^{II} for arginine demethylation of all of the three arginine methylated forms (**Figure 6.4**). These results indicate that arginine demethylation is likely to occur via the JmjC domain in a similar catalytic mechanism to that of lysine demethylation, and correlate with the observed co-factor dependence for other KDM3A-catalysed modifications (see Chapter 7, Section 7.6.3 and Chapter 9, Section 9.5).

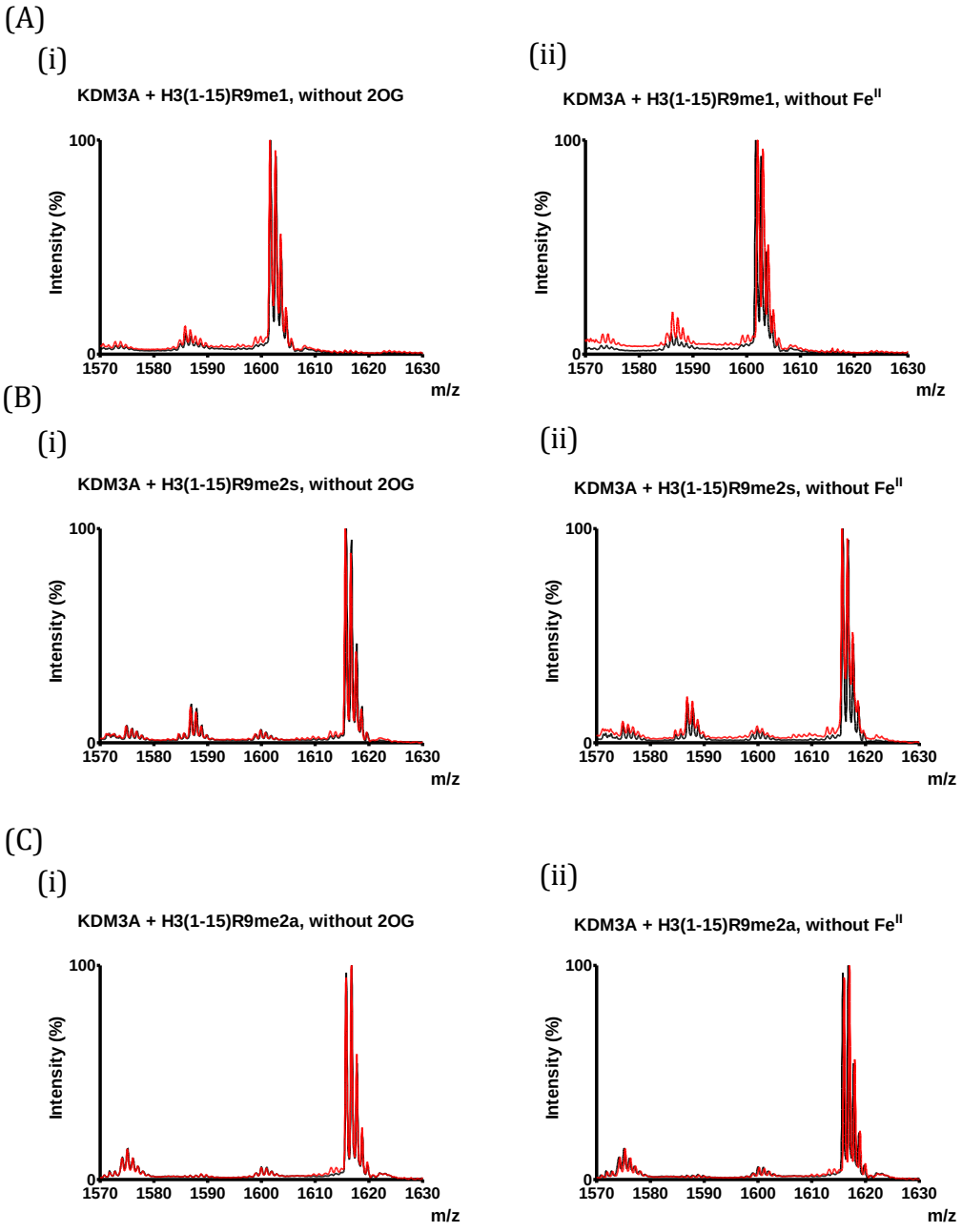


Figure 6.4 MALDI-TOF MS spectra showing lack of arginine demethylation activity by KDM3A when co-factors are absent from reaction mixture. Recombinant KDM3A (2 μM) was incubated with methylated arginine peptides (10 μM) under standard demethylation conditions, except for the lack of specified co-factors, for 1 h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) H3(1-15)K9Rme peptide, (B) an H3(1-15)K9Rme2s peptide and (C) an H3(1-15)K9Rme2a peptide. For each peptide (i) and (ii) shows MALDI-TOF MS spectra of reactions carried out without 2OG and Fe^{II} respectively. Reactions containing enzyme are shown in red, no enzyme control reactions are shown in black.

6.5 Arginine demethylation by full length KDM3A

The assays described above were performed with recombinant KDM3A, which is a truncated construct lacking the N' terminal domain (residues 515-1317, see Materials and methods, Chapter 11, Section 11.2.2 for more information).

To test whether arginine demethylation activity can be observed by full length KDM3A (residues 1-1321 see Chapter 11, Section 11.2.2 for more information), constructs encoding an *N*-terminal FLAG-tagged full length wildtype (WT) and H1120Y catalytically inactive variant proteins were overexpressed in HEK293T.[22] Expressed proteins were then immunoprecipitated using their FLAG-tags (**Figure 6.5**). Both the WT and the inactive variant proteins showed degradation similar to that previously observed for recombinant KDM3A (see Chapter 3, Section 3.4).

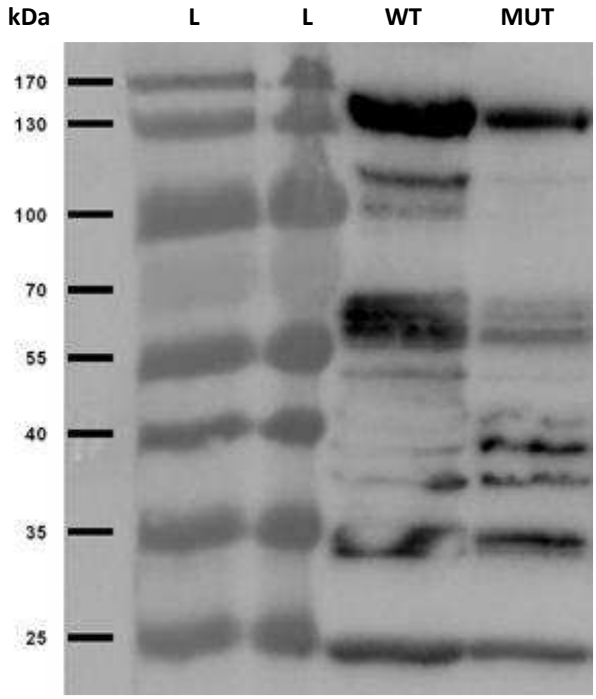


Figure 6.5 Western blot analysis of exogenous overexpression of full length FLAG-tagged KDM3A in HEK293T cells. Proteins were detected using mouse anti-FLAG antibody (F3165 Sigma, 1:1000) and goat anti-mouse-HRP (W4021 Promega, 1:10,000), see Materials and methods, Chapter 11, Section 11.2.4.6, for more information. Anti-FLAG beads were used to immunoprecipitated KDM3A for demethylation assays. L = molecular weight ladder, WT = wildtype KDM3A and MUT = H1120Y catalytically inactive KDM3A variant. The molecular mass of FLAG-tagged full length of KDM3A is 148 kDa. *Cloning of FLAG-tagged full length WT and H1120Y KDM3A constructs was carried out by Dr. Patricia Yeyati.*

Arginine demethylation activity was analysed by MALDI-TOF MS (see Materials and methods, Chapter 11, Section 11.3.3, for more information). KDM3A enzymatic activity with an H3(1-15)K9me2 peptide substrate was used as a positive control (**Figure 6.6 A**). Full length WT KDM3A was observed to catalyse demethylation of the monomethylarginine peptide, but not of the tested dimethylarginine peptides (**Figure 6.6 B-D**). This was consistent with the catalytic activity of the truncated recombinant KDM3A, which had higher levels of monomethylarginine demethylation activity compared with demethylation of dimethylated substrates (**Figure 6.3**). Importantly, no arginine demethylation activity was observed with the H1120Y catalytically inactive variant, indicating that the detected activity was KDM3A-catalysed rather than due to co-purified proteins in the KDM3A purification.

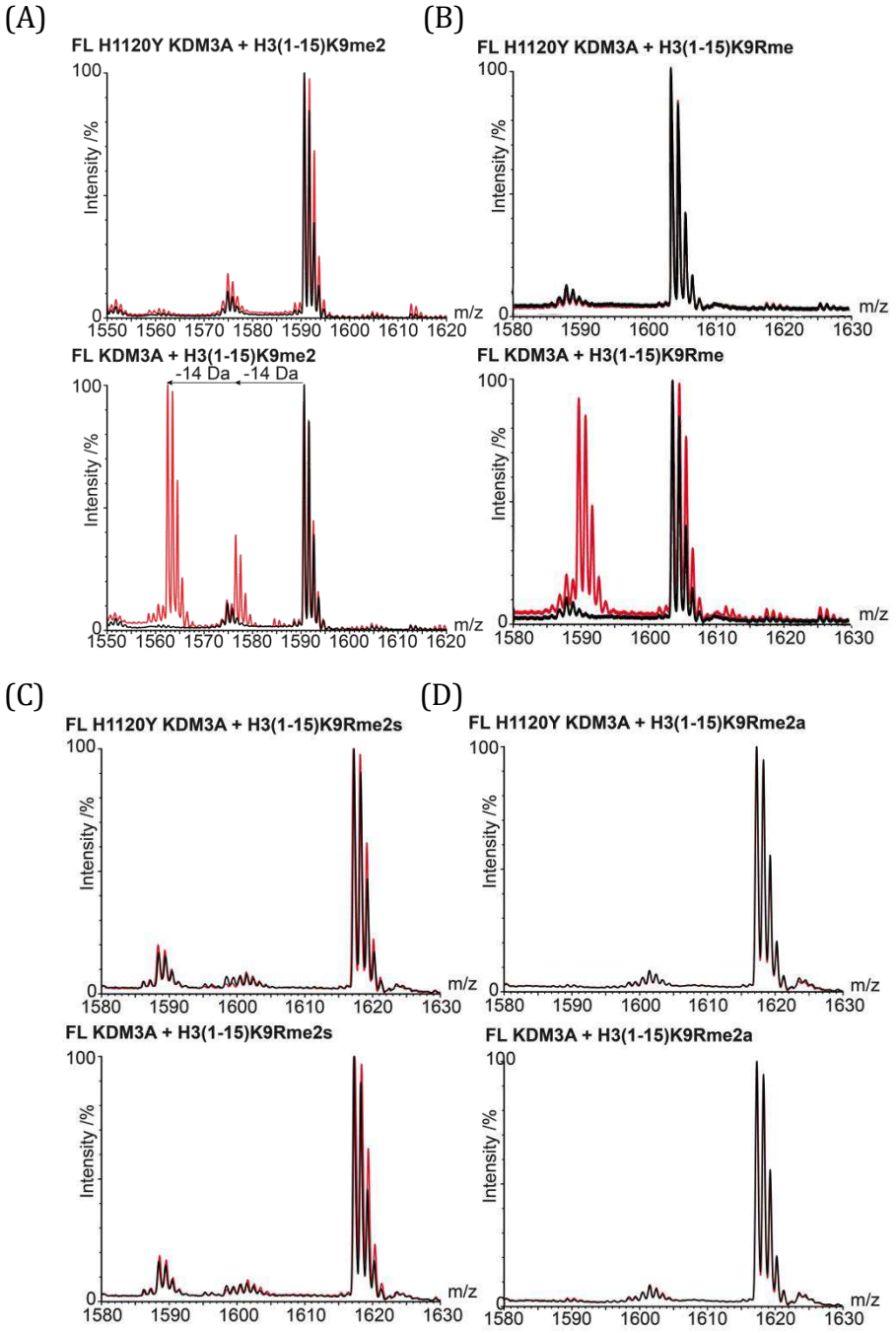


Figure 6.6 MALDI-TOF MS spectra showing demethylation of methylated lysine- and arginine-containing peptides by full length KDM3A. FLAG-tagged full length (FL) WT and H1120Y catalytically inactive variant KDM3A immunoprecipitated from HEK293T cells (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) An H3(1-15)K9me2 histone peptide used as a positive control (10 μM), (B) an H3(1-15)K9Rme peptide(10 μM), (C) an H3(1-15)K9Rme2a peptide(10 μM) and (D) an H3(1-15)K9Rme2s peptide (10 μM). Reactions containing enzyme are shown in red, no enzyme control reactions are shown in black. Figure adapted with permission.[21] *Measurements were carried out by Dr. Louise Walport.*

6.6 Screening for arginine demethylation by a panel of KDMs

Several KDMs were tested for demethylation of N^{ω}-methylated arginine residues as part of this study. These included KDM2A, KDM7B (PHF8), KDM4E, KDM5C and KDM6B in addition to KDM3A. Similar to KDM3A, peptides were designed such that the reported methylated lysine substrate was substituted with different forms of methylated arginine residues. The sequences of the peptides and the corresponding recombinant KDM for which they were tested are given in **Table 6.1**.

Reactions were performed under the same standard demethylation conditions used with KDM3A, with the only difference being the addition of salt to reactions with KDM6B and KDM7B (150 mM and 500 mM NaCl respectively).[21] Products were observed by MALDI-TOF MS. A -14 Da and -28 Da shift in peptide mass compared to the no enzyme controls indicated demethylation activity. Reactions with analogous methylated lysine sequences were tested as positive controls to confirm enzymatic activity.

Although no arginine demethylation activity was observed for KDM2A and KDM7B, activity was detected for KDM4E, KDM5C and KDM6B in addition to that of KDM3A (**Table 6.1**). Interestingly, activity for KDM6B was only observed for the asymmetric dimethylarginine modification when it was incorporated to a longer peptide (21 mer) rather than the standard length (15 mer).

Table 6.1 Recombinant KDM enzymes tested in this study, their lysine methylated peptide target sequence and the position where arginine modifications substituted the lysine. For each of the arginine methylation forms ticks denote methylation observed, crosses denote no reaction detected. *Measurements were carried out by Dr. Louise Walport.*

Enzyme	Lysine methylated peptide sequence	Position on histone	MMA (Rme1)	ADMA (Rme_{2a})	SMDA (Rme_{2s})
KDM3A	ARTKQTAR-**Kme2**-STGGKA	H3 K9	✓	✓	✓
KDM4E	ARTKQTAR-**Kme3**-STGGKA	H3 K9	✓	✓	✓
KDM5C	ART-**Kme3**-QTARKSTGGKA	H3 K4	✓	✓	✓
KDM6B*	KAPRKQLATKAAR-**Kme3**-SAPATGG	H3 K27	X	✓	X
KDM2A	APATGGV-**Kme2**-KPHRYRP	H3 K36	X	X	X
KDM7B (PHF8)	ART-Kme3-QTAR-**Kme2**-STGGK	H3 K4me3K9	X	X	X

**Reaction with 21 mer peptide instead of the standard 15 mer length.*

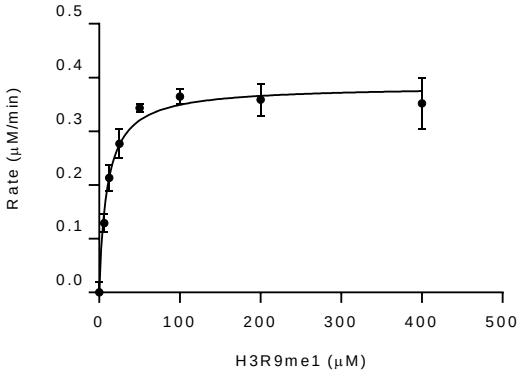
6.7 Kinetic characterisation of KDM3A arginine demethylation

Initial kinetic characterisation was carried out to compare the efficiency of KDM3A catalysis of the reported lysine methylated substrate (H3 K9me2) to the arginine methylated analogue substrates, as well as to compare the activity for the different arginine forms (Rme1, Rme2a, Rme2s).

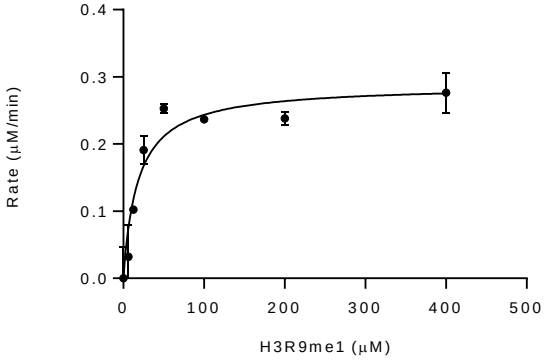
Demethylation was detected using a formaldehyde dehydrogenase (FDH) coupled demethylation assay, which is a convenient method for measuring enzyme kinetics. All reagents were kept in excess to ensure that the substrate peptide is the only rate limiting factor (for more information see Chapter 3, Section 3.5.2, and Materials and methods, Chapter 11, Section 11.3.4). For each of the tested peptides initial rates of KDM3A demethylation were measured in three biological replicates, each with three technical replicates. Michaelis-Menten plots for methylated-arginine substrates are presented in **Figure 6.7**. Apparent K_M and k_{cat} values were determined based on the Michaelis-Menten equation, and the ratio of k_{cat} / K_M was then calculated as a measure of the catalytic efficiency for the particular substrate (**Table 6.2**).

(A)

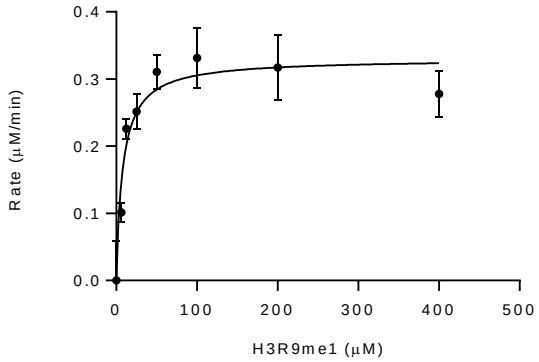
(i)



(ii)

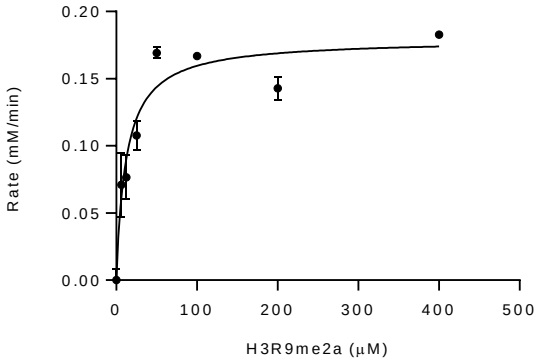


(iii)

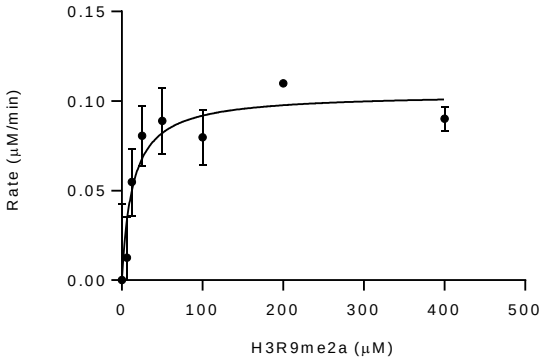


(B)

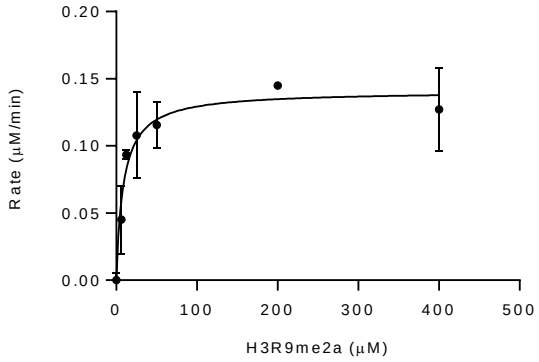
(i)



(ii)



(iii)



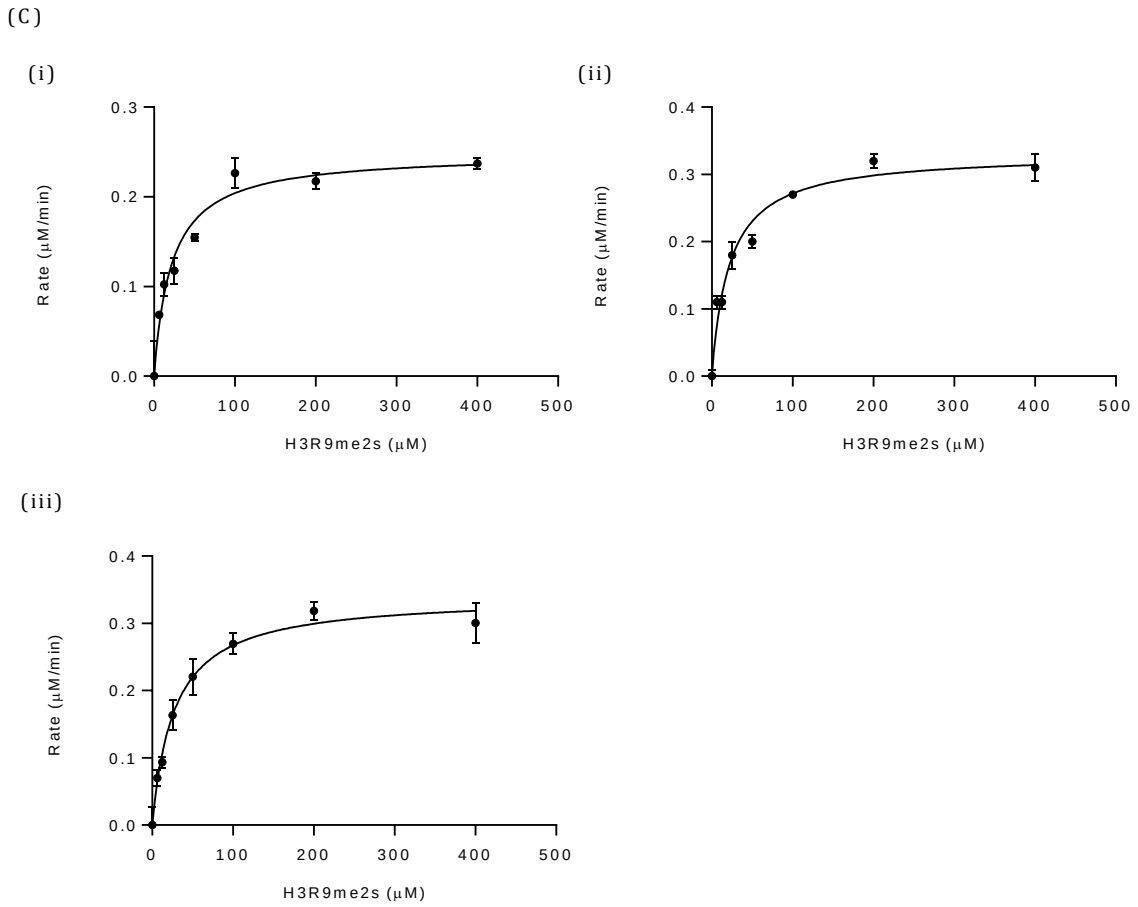


Figure 6.7 Michaelis-Menten plots for KDM3A measuring initial rates of reactions with the following peptides (A) H3(1-15)K9Rme1, (B) H3(1-15)K9Rme2a and (C) H3(1-15)K9Rme2s. (i), (ii) and (iii) are the biological replicates for each peptide. Data were determined by the FDH coupled assay, and were fitted to the Michaelis-Menten model using GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.4, for more information). All measurements were conducted in triplicate and represent mean and standard deviation.

Table 6.2 Kinetic parameters for KDM3A arginine demethylation of histone methylated-lysine and arginine peptide substrates. Apparent kinetic parameters (K_M and k_{cat} values) were determined by the FDH coupled assay. Michaelis-Menten plots for methylated-arginine substrates are presented in **Figure 6.7**. Data show mean (n = 3) for each biological replicate as well as the average of the biological replicates ± s.e.m.

Substrate	Biological replicate	Enzyme (μM)	K_M (μM)	k_{cat} ($\times 10^{-3}$ s^{-1})	k_{cat} / K_M ($\times 10^{-6}$ / s μM)
H3(1-15)K9Rme1	i	0.5	10	13	1277
	ii	0.5	18	10	526
	iii	0.5	8	11	1363
	Average		**12 ± 3**	**11 ± 1**	**908**
H3(1-15)K9Rme2a	i	0.5	12	6.0	483
	ii	0.5	14	3.5	257
	iii	0.5	9	4.7	527
	Average		**12 ± 1**	**5.0± 0.6**	**431**
H3(1-15)K9Rme2s	i	2.0	22	2.1	95
	ii	2.0	22	2.8	125
	iii	2.0	27	2.8	103
	Average		**24 ± 2**	**3.0 ± 0.3**	**127**

Substrate inhibition was observed for the lysine methylated peptide H3(1-15)K9me2 at concentrations above 160 μM. Substrate inhibition was also apparent for H3(1-15)K9me1 (see Chapter 3, Section 3.5.2.4, for primary data).

Interestingly, no substrate inhibition was detected for the arginine methylated peptides (**Figure 6.7 A-C**). K_M values of all arginine methylated peptides were lower than that of the lysine dimethylated substrate, indicating a better binding affinity to the active site (**Table 6.1 and Table 3.5** in Chapter 3). On the other hand, k_{cat} values of methylate arginine substrates were an order of magnitude lower than that of the dimethylated lysine

substrate, suggesting a much slower turnover of the methylated arginine residues compared to the lysine residue.

Among the methylated arginine peptides, the monomethylated form had the highest k_{cat}/K_M ratio and therefore seems to be catalysed most efficiently (908 $\times 10^{-6}$ / s μM). This is consistent with the MALDI-TOF MS data showing efficient conversion of the monomethyl peptide to the non-methylated product (**Figure 6.3** and **Figure 6.6**). The k_{cat}/K_M ratio for H3(1-15)K9Rme2a was two-fold lower than that of H3(1-15)9Rme1, and the symmetric peptide k_{cat}/K_M ratio was a further three-fold lower than that of the asymmetric peptide (908 vs. 431 and 127 $\times 10^{-6}$ / s μM). The combined kinetic results indicate different demethylation efficiencies for each of the methylated arginine forms, with preference for the monomethyl arginine residue, probably due to its low steric hindrance compared to the dimethylated forms.

6.8 Kinetic characterisation of arginine demethylation by additional KDMs

Preliminary kinetic characterisation was also carried out with additional KDMs which displayed arginine demethylation activity in the MALDI-TOF MS screen: KDM4E, KDM5C and KDM6B (see Section 6.6). Kinetic studies were performed using the FDH coupled assays in a similar manner to that of KDM3A studies, analogous lysine methylated peptide substrates were used as controls.

The rate constants for demethylation of the arginine methylated peptides by KDM4E and KDM5C, but not KDM6B, imply a much slower catalysis than those of the analogous lysine methylated peptide, in agreement with KDM3A kinetic parameters (**Table 6.3**). By contrast KDM3A, however, the apparent K_M values of arginine methylated substrates are lower than that of the natural substrate, likely to indicate on a weaker binding of these peptides to KDM4E/5C/6B. The catalytic efficiency, as expressed by the k_{cat}/K_M ratio clearly indicates that all of the tested KDM enzymes, including KDM3A, were more capable of demethylating lysine residues than arginines. As for arginine demethylation, KDM6B and KDM4E were the least efficient enzymes, while KDM3A was found to demethylate all arginine forms at a catalytic efficiency in the hundreds $\times 10^{-6}$ / s μM range.

Table 6.3 Kinetic parameters for arginine demethylation by KDMs for histone methylated-lysine/arginine peptide substrates. Apparent kinetic parameters were determined by the FDH coupled assay. Data show mean (n = 3) ± s.e.m. A positive control of histone methylated-lysine peptide substrate for each KDM as a basis for reference is marked in bold. *Measurements, except for KDM3A, were carried out by Dr. Louise Walport.*

Enzyme	Substrate	K_M (μM)	k_{cat} ($\times 10^{-3}$ s^{-1})	k_{cat}/K_M ($\times 10^{-6}$ / s μM)
KDM3A	**H3(1-15)K9me2**	**54 ± 13**	**790 ± 112**	**14649**
	H3(1-15)K9Rme2a	12 ± 1	5.0 ± 0.6	431
	H3(1-15)K9Rme2s	24 ± 2	3.0 ± 0.3	127
	H3(1-15)K9Rme	12 ± 3	11 ±1	908
KDM4E	**H3(1-15)K9me3**	**18 ± 4**	**171 ± 15**	**9575**
	H3(1-15)K9Rme2a	66 ± 5	4.6 ± 0.7	70
	H3(1-15)K9Rme2s	101 ± 11	8.7 ± 0.3	86
	H3(1-15)K9Rme	169 ± 36	2.7 ± 0.2	16
KDM5C	**H3(1-15)K4me3**	**8 ± 3**	**67 ± 10**	**8800**
	H3(1-15)K4Rme2a	29.0 ± 0.5	41 ± 4	1420
	H3(1-15)K4Rme2s	21 ± 6	4.6 ± 0.7	219
KDM6B	**H3(14-34)K27me3**	**6.7 ± 0.6**	**11.7 ± 0.1**	**1750**
	H3(14-34)K27Rme2a	676 ± 118	29 ± 4	43

6.9 Screening for arginine demethylation on wildtype histone derived peptides

In light of the observation that some KDMs can catalyse demethylation of arginine residues when they substitute for the natural methylated lysine, it was hypothesised that these KDMs might also catalyse demethylation of natural methylated arginine positions on histone tails.

Peptides for five wildtype methylated arginine positions on histones were tested against four KDMs that showed arginine demethylation activity with mutated peptide sequences (**Table 6.1**). Screening conditions were the same as those described in Section 6.6.

Interestingly, while KDM4E and KDM5C displayed activity against some of the tested peptides, KDM3A and KDM6B showed no activity for these peptides (**Table 6.4**) Therefore, these results demonstrate that whether or not arginine demethylation activity occurs is highly dependent on the sequence context, with KDM3A and KDM6B presenting particular preference for the sequence of the methylated lysine substrates.

Table 6.4 Arginine demethylation activity of the tested KDMs on peptides derived from WT histone sequences. Ticks indicate demethylation activity, crosses indicate no activity at the tested conditions. *Measurements were carried out by Dr. Louise Walport.*

KDM	H3(1-15)R2			H3(1-15)R8			H3(10-24)R17			H3(18-32)R26			H4(1-15)R3		
	me	me2a	me2s	me	me2a	me2s	me	me2a	me2s	me	me2a	me2s	me	me2a	me2s
KDM3A	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
KDM4E	✓	✓	✓	✓	✓	✓	X	X	X	✓	✓	✓	X	✓	✓
KDM5C	✓	✓	✓	X	✓	✓	X	X	X	X	X	X	X	✓	X
KDM6B	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

6.10 Testing the dependence of KDM3A arginine demethylation on the H3 K4me3 mark

KDM3A hydroxylation on acetyllysines was found to be promoted by the presence of trimethylation on the H3 K4 position (see Chapter 7, Section 7.4.2). In addition, KDM3A deformylation activity was also promoted by the presence of the H3 K4me3 mark (see Chapter 9, Section 9.4).

No demethylation activity was detected when KDM3A was incubated with H3(1-15)R8me peptide (see Section 6.9). To test whether arginine demethylation on natural arginine positions might also be dependent on adjacent lysine trimethylation, H3(1-15)K4me3R8me peptide was incubated with KDM3A under the same conditions as described in Section 6.3. Reaction with H3(1-15)R9me peptide was used as a positive control for KDM3A activity.

Although the KDM3A was active as apparent by catalysis of the H3(1-15)R9me peptide, MALDI-TOF MS analyses indicate lack of KDM3A activity against H3(1-15)K4me3R8me peptide (**Figure 6.8**). Therefore, H3 R8 demethylation does not seem to be dependent on trimethylation at the H3 K4 position under the tested conditions.

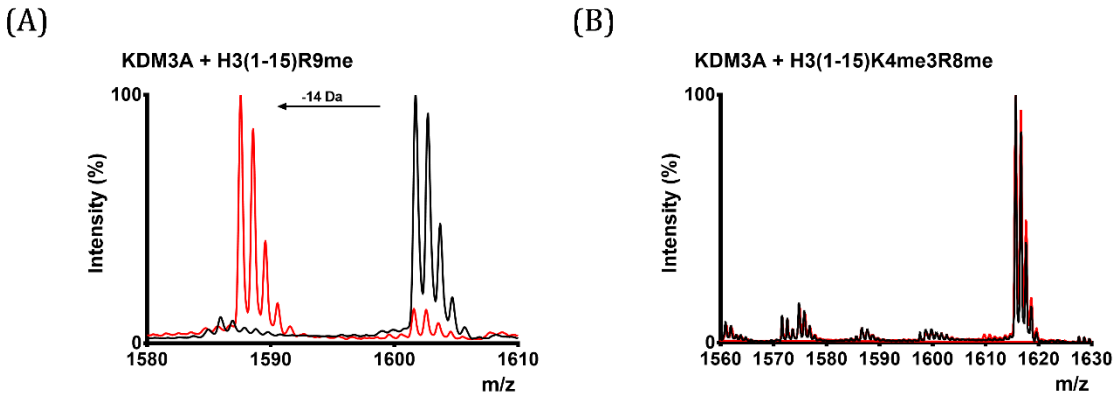


Figure 6.8 The presence of an H3 K4me3 mark does not improve demethylation of the H3 R8me residue. KDM3A (2 μM) was incubated with each peptide (10 μM) for 1 h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) MALDI-TOF MS spectra showing catalysis of H3(1-15)K9Rme peptide as a positive control for assay conditions. (B) MALDI-TOF MS spectra showing absence of activity for of H3(1-15)K4me3R8me peptide. Reactions containing enzyme are shown in red, no enzyme control reactions are shown in black.

6.11 Screening for arginine demethylation of non-histone substrates by KDMs

Given the detected arginine demethylation activity of some KDMs on peptides derived from WT histone tail sequences, it was hypothesised that these KDMs might also demethylate arginines on non-histone substrates.

A selection of non-histone peptides derived from nuclear proteins reported to be abundantly methylated on arginine residues were then synthesised and tested (**Table 6.5**).[23–28] These peptides were tested as potential arginine demethylase substrates by recombinant KDM5C, KDM4E and KDM6B under the assay conditions described in Section 6.6.

Table 6.5 Non-histone peptide sequences and the protein mark around which they were designed. Peptides were used to screen for potential methylarginine non-histone substrate of KDM5C, KDM4E and KDM6B.

Protein mark	Peptide sequence
53BP1(1394-1415) R1401me2a	GKAPVTP-**Rme2a**-GRGRRGRPPSRTTG
53BP1(1394-1415) R1403me2a	GKAPVTPRG-**Rme2a**-GRRGRPPSRTTG
53BP1(1394-1415) R1405me2a	GKAPVTPRGRG-**Rme2a**-RGRPPSRTTG
BRCA1(605-619)R610me2a	APKKN-**Rme2a**-LRRKSSTRH
hnRNPK(295-309)R296me2a	G-**Rme2a**-GGRGGSRARNLPL
hnRNPK(295-309)R299me2a	GRGG-**Rme2a**-GGSRARNLPL
FMR1(537-551)R544me2a	GGRGQGG-**Rme2a**-GRGGGFK
FMR1(537-551)R546me2a	GGRGQGGRG-**Rme2a**-GGGFK
p53(332-343)R333me2s	IRGRE-**Rme2s**-FEMFRE
p53(328-339)R335me2s	FTLQI-**Rme2s**-GRERFE
p53(330-341)R336me2s	LQIRG-**Rme2s**-ERFEMF

Most methylated arginine peptides were not substrates as indicated by MALDI-TOF MS analyses. However, in some cases clear demethylation was detected (**Figure 6.9**). Interestingly, almost complete demethylation was observed for KDM4E and KDM5C catalysis of peptides derived from heterogeneous nuclear ribonucleoprotein K (hnRNP K) and tumour suppressor p53-binding protein 1 (53BP1) respectively.

These results provide preliminary evidence for the potential of arginine demethylation beyond the context of transcriptional regulation.

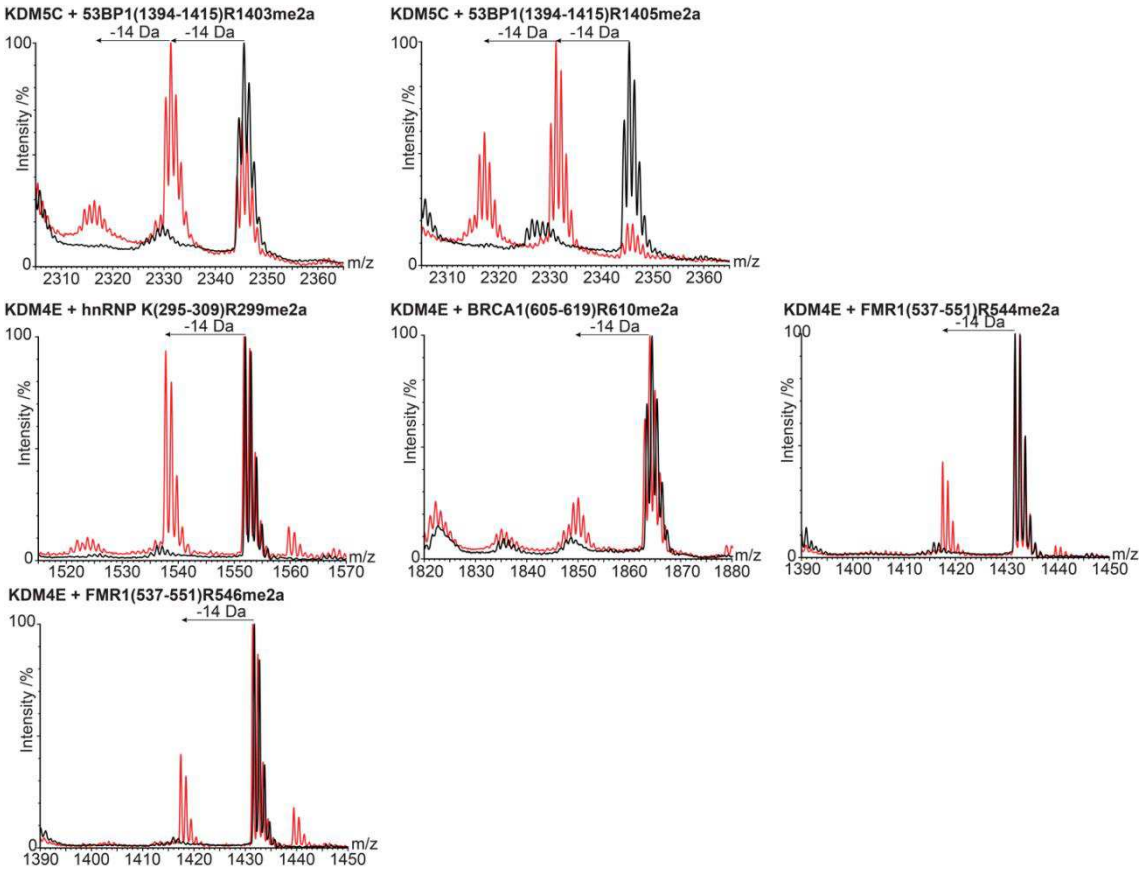


Figure 6.9 Arginine demethylation catalysis of non-histone peptides. MALDI-TOF MS spectra of observed demethylation reactions catalysed by recombinant KDM5C and KDM4E. Peptides (10 μM) were incubated with KDM5C or KDM4E (2 μM) for 1 h at 37°C. Reactions containing enzyme are shown in red, no enzyme control reactions are shown in black. Figure adapted with permission.[21] *Measurements were carried out by Dr. Louise Walport.*

6.12 Screening for arginine demethylation of IFT by KDM3A

The results in Section 6.11 demonstrate that KDMs can catalyse demethylation of non-histone substrate peptides in addition to histone peptides. In light of these findings, it was of interest to test whether a potential interaction between KDM3A and IFT proteins (see Chapter 5) is based on arginine demethylation.

Intraflagellar transport (IFT) proteins were found to be methylated on arginine residues during flagellar resorption (see Section 16.1). Schneider *et al.* have proposed that methylation on arginines is irreversible and therefore the methylated proteins generated are most likely not reused when resorption conditions are reversed. Instead, they assumed that the cytoplasm must contain a pool of nonmethylated forms of proteins to meet the requirement of these proteins for flagellar regeneration.[9]

Given the newly discovered ability of some KDMs to demethylate methylated arginine, it was considered possible that KDM3A might be able to recycle IFT proteins after resorption by reversing the arginine methylation state.

Eleven IFT peptides were synthesised with different forms of arginine methylation based on predictions by the PhosphoSitePlus® (PSP) database, and by PTM detection analyses by MS of IFT proteins extracted from mouse testes (see Chapter 5, Section 5.1 for more information, **Table 6.6** for peptide sequences).[29]

Table 6.6 Arginine-methylated IFT peptides screened against recombinant KDM3A. Target arginyl-residues are shown in bold. *Dr. Holger Kramer and Dr. Patricia Yeyati identified putative arginine methylated IFT proteins.*

Protein	Target arginyl-residue	Peptide sequence
IFT74	R51	GTARPGS-**Rme**-GCPIGTG
IFT74	R305	IGSPMEE-**Rme2**-EKLLKQI
IFT81	R345	EGKLSLY-**Rme**-QQASIIS
IFT81	R188	EKDQLIK-**Rme2a**-VEHLKKR
IFT81	R188	EKDQLIK-**Rme2s**-VEHLKKR
IFT88	R72	ISSTAVT-**Rme2a**-PIATGYG
IFT88	R72	ISSTAVT-**Rme2s**-PIATGYG
IFT88	R129	FDPLSQS-**Rme2s**-GPASPLE
IFT88	R129	FDPLSQS-**Rme2a**-GPASPLE
IFT70	R114	DNPAYHS-**Rme2a**-VLRLQAA
IFT70	R114	DNPAYHS-**Rme2s**-VLRLQAA

The arginine methylated peptides were screened under the same conditions as described in Section 06.3. H3(1-15)K9me2 peptide was tested as a control for enzyme activity. Samples were analysed for arginine demethylation activity by MALDI-TOF MS.

Although KDM3A was active when incubated with the histone H3(1-15)K9me2 substrate peptide, no arginine demethylation activity was observed for the tested IFT arginine methylated peptides. Increasing the enzyme concentration from 2 to 4 μM, as well as increasing the incubation time by another hour did not change the results (**Figure 6.10**).

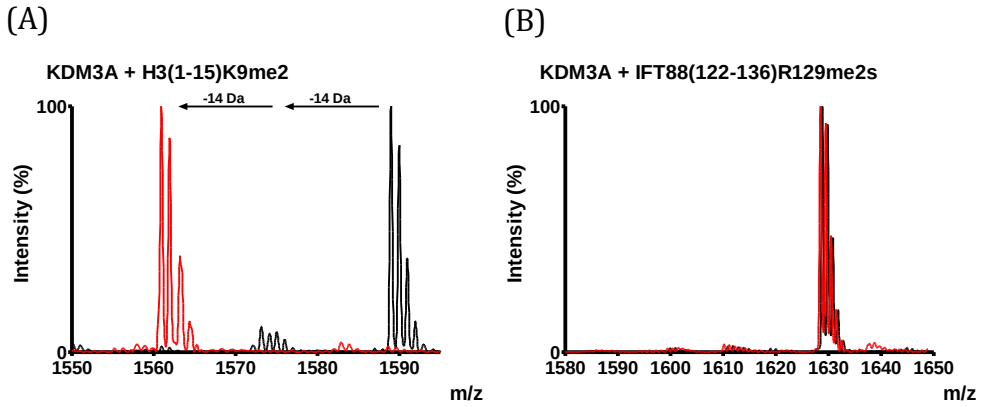


Figure 6.10 Example of MALDI-TOF MS spectra from IFT screen. (A) Demethylation of positive control peptide H3(1-15)K9me2. (B) Absence of demethylation of IFT88(122-136)R129me2s peptide. 15 mer peptides (10 µM) were incubated with KDM3A (4 µM) for 2 h at 37°C, followed by MALDI-TOF MS analyses (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Reactions with enzyme are shown in red, control reactions with no enzyme are shown in black.

6.13 Discussion

The work reported in this chapter describes the identification of arginine demethylation activity for KDM3A and additional KDMs from different JmjC subfamilies (KDM4E, KDM5C and KDM6B). Although Protein Arginine Methyltransferases (PRMTs) are well described, identification of a bone-fide arginine demethylases is still being actively studied. The current observations indicate that arginine methylation might be dynamic and not an irreversible PTM as previously assumed. This work also represents the first conclusive evidence for the existence of arginine demethylases *in vitro*. Moreover, it suggests that arginine and lysine demethylation can be performed by a similar catalysis mechanism and might be dependent on the same enzymes. Preliminary evidence indicates that arginine demethylation might occur on histone and non-histone substrates, in both an epigenetic and non-epigenetic context.

The results of MALDI-TOF MS screening assays reveal that some of the tested KDMs can catalyse demethylation of arginines *in vitro*. The peptide sequence as well as the arginine methylation forms (MMA, ADMA and SMDA) widely influence their catalytic activity. Kinetic studies on KDM3A have found that although the WT lysine methylated histone sequences (H3 K9me2 and H3 K9me1) showed substantial substrate inhibition for KDM3A, the analogous arginine methylated substrates do not cause the same effect.

The apparent low k_{cat} values KDM3A displayed for arginine methylated peptides may indeed represent slow catalysis, alternatively they might be a result of a combination of fast arginine demethylation and high uncoupled 2OG turnover. To test which assumption is the correct one, ^{1}H NMR experiments in which the 2OG turnover rate was measured were conducted. However, lack of an available appropriate buffer to maintain enzymatic activity as well as the need for high enzyme concentration in such assays hampered this work (*data not shown*). Testing these assumptions regarding KDM6B indicate that partial decoupling of 2OG oxidation and demethylation can occur.[21] This can be a result of a non-optimal substrate sequence or due to the use of a truncated substrate (i.e. peptide), as is the case with other 2OG dependent hydroxylases.[30]

Unlike some of the other tested KDMs, KDM3A does not appear to act on methylated arginine residues in WT histone positions or on non-histone peptides. This could be a consequence of a number of effects. Firstly, KDM3A *in vitro* activity might be sequence selective. This is an especially likely option given its ability to react with different lysine analogues and different lysine PTMs when these are incorporated at the H3 K9 position (see Chapter 7, Sections 7.3 and 7.4.2 and Chapter 9, Section 9.3). Secondly, the screen might not have been sufficiently comprehensive and there might be methylated arginine substrates yet to be discovered, which could be on histones or other proteins in the cell. Thirdly, peptides may not be a good enough mimic of the substrate, either because a full length protein is required or because of a lack in combinatorial modifications in the peptides used for the screen that might be required to increase catalytic efficiency. The catalytic efficiency of KDM3A lysine demethylation activity seems to be improved when H3 K4me3 is present in the peptide sequence (see Chapter 3, Section 3.5.2.5), and the same position seems to also induce reaction with H3 K9Ac and H3 K9for (see Chapter 7, Section 7.4.2 and Chapter 9, Section 9.4). However, no arginine demethylation was observed with H3(1-15)K4me3R8me peptide. Therefore, KDM3A arginine demethylation activity might be dependent on another combination of PTMs. It would be intriguing to follow up the biochemical work to decipher the combinatorial PTMs effect on arginine demethylation kinetics.

Arginine methylated IFT proteins, which accumulate during flagellar resorption, might be demethylated by other arginine demethylases. Therefore continuing tests with other identified arginine demethylases should be carried out.

To identify and confirm biologically relevant substrates, arginine demethylation by KDMs needs to be ultimately demonstrated in cells. Development of selective antibodies will be pivotal for such studies, which was beyond the scope of the work described in this thesis.

6.14 References

1. Ong, S.-E., Mittler, G. & Mann, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods* **1,** 119–26 (2004).

2. Uhlmann, T. *et al.* A method for large-scale identification of protein arginine methylation. *Mol. Cell. Proteomics* **11,** 1489–99 (2012).

3. Paik, W. K. & Kim, S. Natural occurrence of various methylated amino acid derivatives. *Protein methylation* 8–25 (1980).

4. Bedford, M. T. & Clarke, S. G. Protein arginine methylation in mammals: who, what, and why. *Mol. Cell* **33,** 1–13 (2009).

5. Bedford, M. T. *et al.* Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. *J. Biol. Chem.* **275,** 16030–6 (2000).

6. Hughes, R. M. & Waters, M. L. Arginine methylation in a beta-hairpin peptide: implications for Arg-pi interactions, DeltaCp(o), and the cold denatured state. *J. Am. Chem. Soc.* **128,** 12735–42 (2006).

7. Sprangers, R., Groves, M. R., Sinning, I. & Sattler, M. High-resolution X-ray and NMR Structures of the SMN Tudor Domain: Conformational Variation in the Binding Site for Symmetrically Dimethylated Arginine Residues. *J. Mol. Biol.* **327,** 507–520 (2003).

8. Di Lorenzo, A. & Bedford, M. T. Histone arginine methylation. *FEBS Lett.* **585,** 2024–31 (2011).

9. Schneider, M. J., Ulland, M. & Sloboda, R. D. A protein methylation pathway in Chlamydomonas flagella is active during flagellar resorption. *Mol. Biol. Cell* **19,** 4319–27 (2008).

10. Werner-Peterson, R. & Sloboda, R. D. Methylation of Structural Components of the Axoneme Occurs During Flagellar Disassembly. (2013).

11. Ehler, L. L., Holmes, J. A. & Dutcher, S. K. Loss of spatial control of the mitotic spindle apparatus in a Chlamydomonas reinhardtii mutant strain lacking basal bodies. *Genetics* **141,** 945–960 (1995).

12. Quarmby, L. M. & Parker, J. D. K. K. Cilia and the cell cycle? *J. Cell Biol.* **169,** 707–10 (2005).

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

14. Bedford, M. T. & Richard, S. Arginine methylation an emerging regulator of protein function. *Mol. Cell* **18,** 263–72 (2005).

15. Le Romancer, M. *et al.* Regulation of estrogen rapid signaling through arginine methylation by PRMT1. *Mol. Cell* **31,** 212–21 (2008).

16. Sakabe, K. & Hart, G. W. O-GlcNAc transferase regulates mitotic chromatin dynamics. *J. Biol. Chem.* **285,** 34460–8 (2010).

17. Shi, Y. Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119,** 941–53 (2004).

18. Bannister, A. J., Schneider, R. & Kouzarides, T. Histone Methylation. *Cell* **109,** 801–806 (2002).

19. Böttger, A., Islam, M. S., Chowdhury, R., Schofield, C. J. & Wolf, A. The oxygenase Jmjd6--a case study in conflicting assignments. *Biochem. J.* **468,** 191–202 (2015).

20. Hopkinson, R. J. *et al.* Is JmjC oxygenase catalysis limited to demethylation? *Angew. Chem. Int. Ed. Engl.* **52,** 7709–13 (2013).

21. Walport, L. J. *et al.* Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat. Commun.* **7,** 11974 (2016).

22. Yamane, K. *et al.* JHDM2A, a JmjC-Containing H3K9 Demethylase, Facilitates Transcription Activation by Androgen Receptor. *Cell* **125,** 483–495 (2006).

23. Jansson, M. *et al.* Arginine methylation regulates the p53 response. *Nat. Cell Biol.* **10,** 1431–1439 (2008).

24. Yang, J.-H. *et al.* Arginine methylation of hnRNPK negatively modulates apoptosis upon DNA damage through local regulation of phosphorylation. *Nucleic Acids Res.* **42,** 9908–24 (2014).

25. Liu, Q. & Dreyfuss, G. In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol. Cell. Biol.* **15,** 2800–2808 (1995).

26. Guendel, I. *et al.* Methylation of the Tumor Suppressor Protein, BRCA1, Influences Its Transcriptional Cofactor Function. *PLoS One* **5,** e11379 (2010).

27. Lee, Y.-H. & Stallcup, M. R. Minireview: Protein Arginine Methylation of Nonhistone Proteins in Transcriptional Regulation. doi:10.1210/me.2008-0380

28. Wei, H., Mundade, R., Lange, K. & Lu, T. Protein arginine methylation of non-histone proteins and its role in diseases. *http://dx.doi.org/10.4161/cc.27353* (2013).

29. PhosphoSitePlus: a resource for protein phosphorylation and other post-translational modifications. Available at: http://www.phosphosite.org/homeAction.action. (Accessed: 24th May 2016)

30. Ehrismann, D. *et al.* Studies on the activity of the hypoxia-inducible-factor hydroxylases using an oxygen consumption assay. *Biochem. J.* **401,** 227–34 (2007).

Chapter 7
Investigations into a putative new PTM catalysed by KDM3A

7.1 Introduction

Some of the most highly post-translationally modified proteins in eukaryotes are histones. PTMs on the *N*-terminus tail region of histones, in particular on H3, regulate chromatin conformation and gene transcription.[1] More than 20 types of histone PTMs have been reported and novel PTMs are being discovered due to advances in the development of specific antibodies and mass spectrometric methods.[2,3] Lysine is proposed to be the most modified amino acid, and lysine acetylation and methylation are amongst the most studied PTMs.[4] Whilst histone acetylation removes the positive charge on a lysine, and is in general transcriptionally activating, lysine methylation does not substantially alter the side-chain charge and can be either transcriptionally activating or repressing (see Introductory Chapter 1, Section 1.8).

Multiple marks can compete on the same position on a histone tail and result in a different outcome. Moreover, different combinations of PTMs on different positions along the histone tail can either counteract or synergise to co-regulate the same site. For instance, acetylation on H3 K9 competes with methylation on the same site and reverses the transcriptional repression caused by methylation, whereas the co-occurrence of H3 K9Ac and H3 K4me3 is often a hallmark of active genes.[5]

2OG dependent oxygenases can catalyse a versatile repertoire of oxidative reactions, including ring closure and epoxidation. However, only hydroxylation, and consequently demethylation, are reported to occur in humans.[6] The mechanism for the hydroxylation reaction is apparently conserved and includes sequential binding of iron, 2-oxoglutarate (2OG), substrate and molecular oxygen, which leads to oxidation of the substrate and is coupled to the conversion of 2OG to succinate.[7] JmjC demethylation occurs spontaneously following hydroxylation, since the oxidation product, the hemiaminal intermediate, is

unstable and fragments to form the demethylated product and formaldehyde (see Introductory Chapter 1, Section 1.13).[8]

Previous studies have revealed the potential of human JmjC demethylases to catalyse hydroxylation of chemical groups other than *N*ε-methyllysine, indicating that they may have more complex reactivity.[9–11] Although there is evidence for catalysis of uncharged substrates by KDMs in plants, the human KDMs tested under these studies (KDM4A, KDM4B, KDM4E, KDM7B and KDM2A) required a positively charged *N*ε-amino group and could not catalyse the uncharged acetyl-, formyl-, crotonyl-L-lysine residues or any of the tested uncharged lysine analogues.[9,10]

7.2 Objectives

The work described in this chapter aimed to investigate the possibility of KDM3A to act on *N*ε-acetyllysine residues. Work was initially focused on testing the need for charged substrates for KDM3A using lysine analogues. Studies then progressed to test *in vitro* activity on acetylated lysine. Detailed validation of a new reaction was then conducted and optimisation of reaction conditions served as the basis for kinetic characterisation. The biological relevance of the newly identified substrate sequence was predicted based on existing ChIP-seq data. Work was then expanded to address the possibility of other histone and non-histone lysine-acetylated sequences to be recognised as substrates for KDM3A and of other KDMs to react with acetylated lysines and involved designing peptides and testing reactions with recombinant KDMs. A probe into the binding affinity of the identified product for the bromodomain BRD4 was carried out to gain insights into the putative biological implications of the new modification.

*Unnatural lysine analogue peptides were synthesised by Gareth W. Langley. H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson. H3(1-15)K4me3 peptide and Fmoc-protected deuterated N*ε-acetyllysine *were synthesised by Roman Belle. KDM3B, JMJD1C and catalytically inactive KDM3A were provided by Dr. Catrine Johansson. KDM6B was provided by Kate Dunne. KDM4A was provided by Rebecca Hancock. ChIP-seq datasets were analysed by Tunc Morova and Prof. Nathan Lack. AlphaScreen measurements were carried out by David Zollman. Dr. Holger Kramer and Dr. Patricia Yeyati worked to identify acetylation positions on intraflagellar transport (IFT) proteins. Dr. Richard Hopkinson provided assistance with data analyses.*

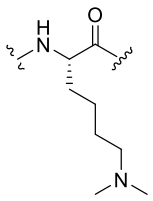
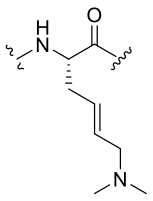
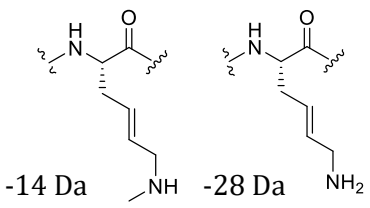
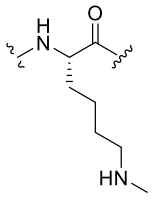
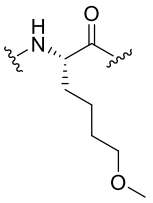
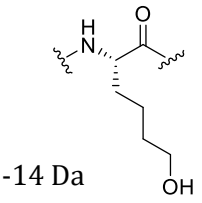
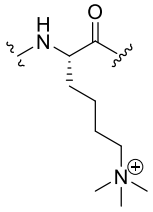
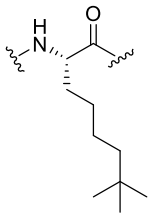
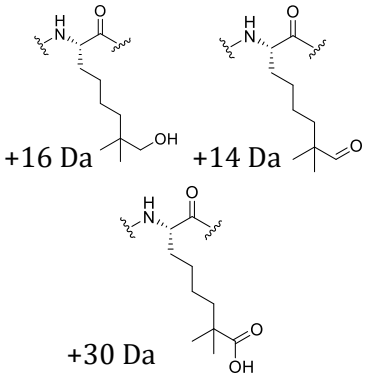
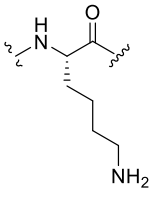
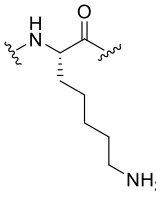
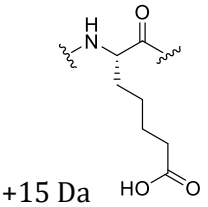
7.3 Probing KDM3A substrate repertoire using lysine analogues

A recent study reported on the use of lysine analogue containing peptides to test the dependence of KDM4 (KDM4A, KDM4B, KDM4E) and KDM5B on the preference of charge state and position of the ε-amino group of their substrate.[10] This study indicated that the tested KDMs may have some flexibility in substrate positioning, but are unable to catalyse oxidation of uncharged groups.

To test whether KDM3A has similar requirements for efficient catalysis, selected lysine analogue peptides were tested as potential substrates. To increase the chances of substrate recognition, H3 histone peptides were synthesised where the natural substrate for KDM3A, the target methylated lysyl-residue (K9), was replaced by a lysine analogue, on the basis of the hypothesis that the sequence similarity between the natural and the mutated substrate will help to direct the analogue into the active site.[12]

Four forms of lysine analogues were incorporated into peptides for experiments with recombinant KDM3A (**Table 7.1**). The **GL1** peptide contains the *N*ε-dimethyl-*trans*-4,5-dehydrolysine analogue which is geometrically constrained compared to the natural dimethyllysine, this analogue was aimed at testing for KDM3A flexibility in substrate binding. Demethylation of this analogue was observed with other KDMs, although crystallographic evidence predicts different binding conformations for the C4-C5 bond of the lysine residue.[10] The **GL2** peptide has an *O*-methyllysine analogue which mimics the natural monomethyllysine but is uncharged. This *O*-methyllysine analogue is similar to the ether substrate of plant 2OG dependent oxygenases, and was used to indicate whether KDM3A can demethylate ether substrates.[13] The **GL3** peptide is a carbon analogue of *N*ε-trimethylated lysine. In both **GL3** and **GL2** peptides the *N*ε-amino group is uncharged, thus allowing investigation of the need for charged substrates for KDM3A catalysis. The **GL4** peptide has an additional carbon on the *N*ε-amino group compared to a natural lysine residue which changes the position of the charge.

Table 7.1 Lysine analogue used in the KDM3A screen. Each analogue is shown alongside the natural modification it mimics as well as the structure of the putative KDM3A-catalysed products as apparent by MALDI-TOF MS mass shifts. All peptides contain the same sequence corresponding to the first 15 residues of histone H3, with different lysine analogues at position 9. *Lysine analogue peptides were synthesised by Gareth Langley.*

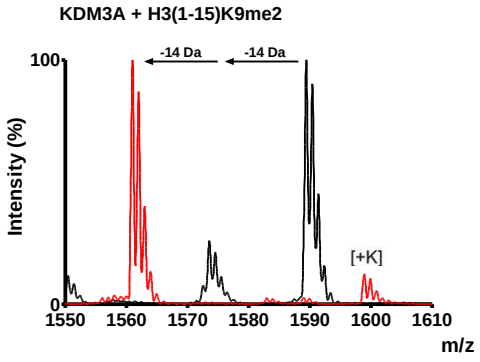
ARTKQTAR-**K**-STGGKA			
Peptide name	**Natural lysine modification**	**Lysine analogue substrate**	**Possible lysine analogue products**
GL1			
GL2			
GL3			
GL4			

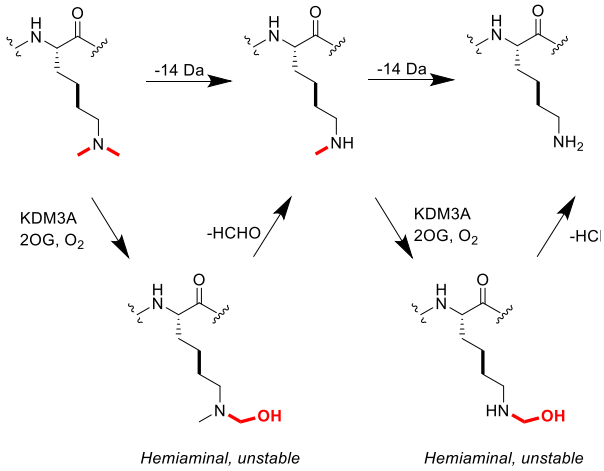
The lysine analogue peptides were tested as potential substrates for KDM3A. The corresponding "wildtype" H3(1-15)K9me2 peptide substrate was used as a positive control. Assays were carried out under standard conditions and were analysed by MALDI-TOF MS (see Chapter 3, Section 3.5.1 and Materials and methods, Chapter 11, Section

11.3.3, for more information). Spectra were compared to negative controls with no enzyme.

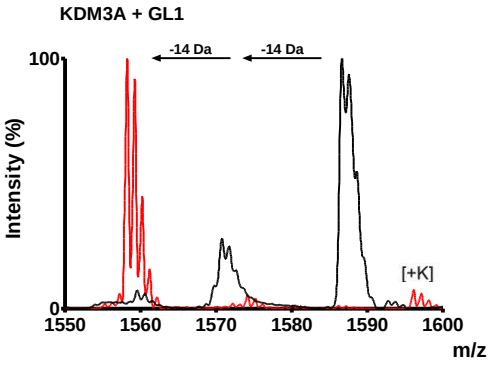
Remarkably, the MALDI-TOF MS analyses reveal that KDM3A catalysed modifications both with the "wildtype" peptide substrate and with all of the tested lysine analogue peptides (**Figure 7.1 A** and **Table 7.1**). Demethylation of the **GL1** unsaturated peptide, which is a geometrically constrained dimethyllysine analogue, was observed by mass shifts of -14 Da and -28 Da relative to the substrate peptide, indicating two consecutive demethylation reactions (**Figure 7.1 B**). Complete demethylation, as indicated by a -14 Da mass shift, was observed after incubation with the **GL2** *O*-methyllysine peptide, which is an uncharged monomethyllysine analogue. This may occur by the same JmjC lysine demethylation mechanism, as the hemiacetal is unstable and could fragment to the demethylated form (**Figure 7.1 C**). When KDM3A was incubated with the trimethyllysine carbon analogue, the **GL3** peptide, evidence for three oxidations was apparent as indicated by +14 Da, +16 Da and +30 Da shifts relative to controls with no enzyme (**Figure 7.1 D**). A single hydroxylation gives a hydroxylated product (+16 Da), two hydroxylations on the same carbon can result in an aldehyde after a H_2O loss (+14 Da) and three oxidations on the same carbon can form a carboxylic acid (+30 Da). Similar to the **GL3** peptide, two consecutive hydroxylations to carboxylic acid might also occur at the **GL4** peptide with an elongated lysine analogue, as indicated by an overall +15 Da mass shift (**Figure 7.1 E**). These results reveal that elongation of the N^{ε}-amino group by one carbon in peptide **GL4** does not affect KDM3A reactivity, as demonstrated by the almost complete conversion to the carboxylic acid product. Interestingly, an elongated allysine aldehyde intermediate was not detected by MALDI-TOF MS. This could be due to the complete carboxylation reaction that decreased the intermediate levels below the detection limit.

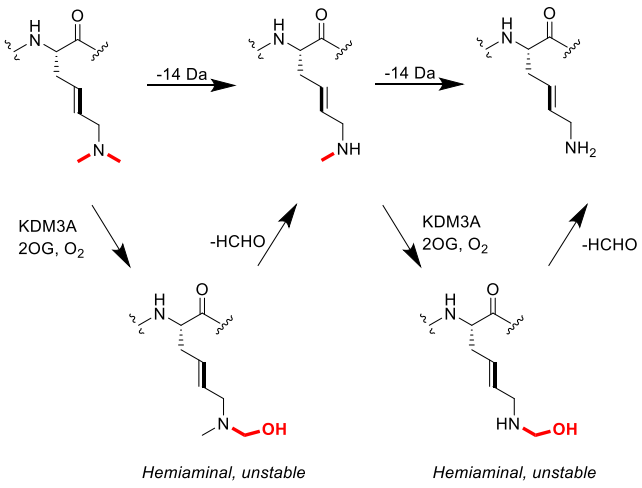
(A)


KDM3A + H3(1-15)K9me2
-14 Da
-14 Da
100
0
Intensity (%)
[+K]
1550 1560 1570 1580 1590 1600 1610
m/z

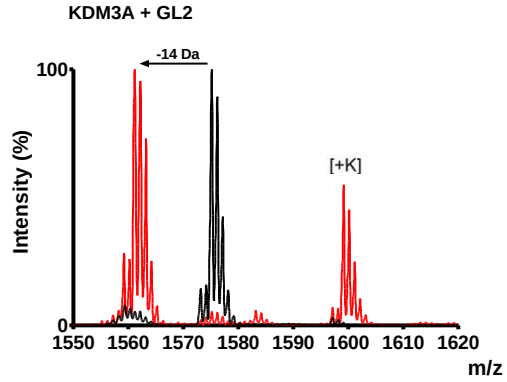

-14 Da
-14 Da
KDM3A
2OG, O2
-HCHO
KDM3A
2OG, O2
-HCHO
NH
NH2
OH
OH
Hemiaminal, unstable
Hemiaminal, unstable

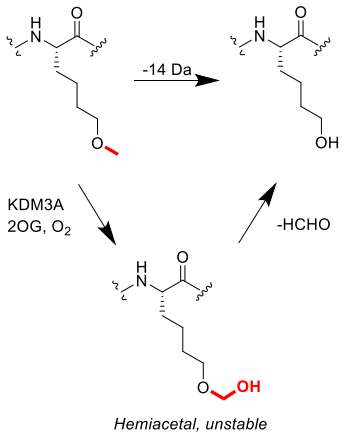
(B)


KDM3A + GL1
-14 Da
-14 Da
100
0
Intensity (%)
[+K]
1550 1560 1570 1580 1590 1600
m/z

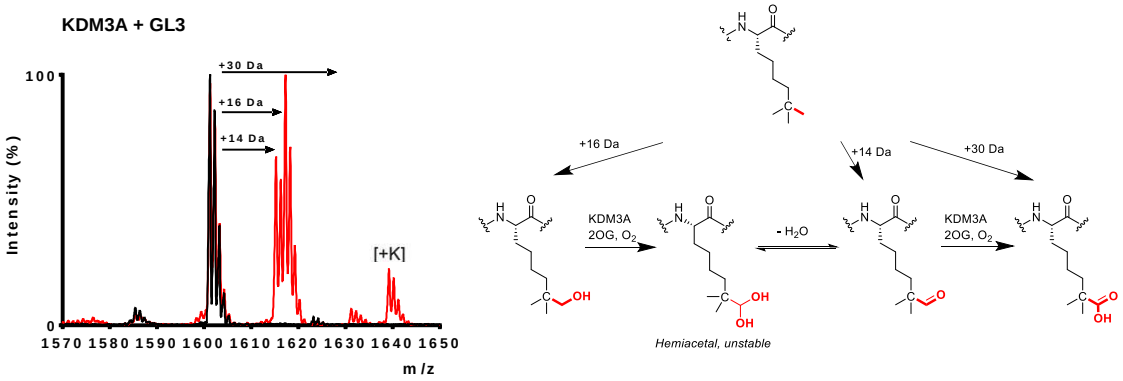

-14 Da
-14 Da
KDM3A
2OG, O2
-HCHO
KDM3A
2OG, O2
-HCHO
NH
NH2
OH
OH
Hemiaminal, unstable
Hemiaminal, unstable

(C)


KDM3A + GL2
-14 Da
100
0
Intensity (%)
[+K]
1550 1560 1570 1580 1590 1600 1610 1620
m/z


-14 Da
KDM3A
2OG, O2
-HCHO
OH
OH
Hemiacetal, unstable

(D)



(E)

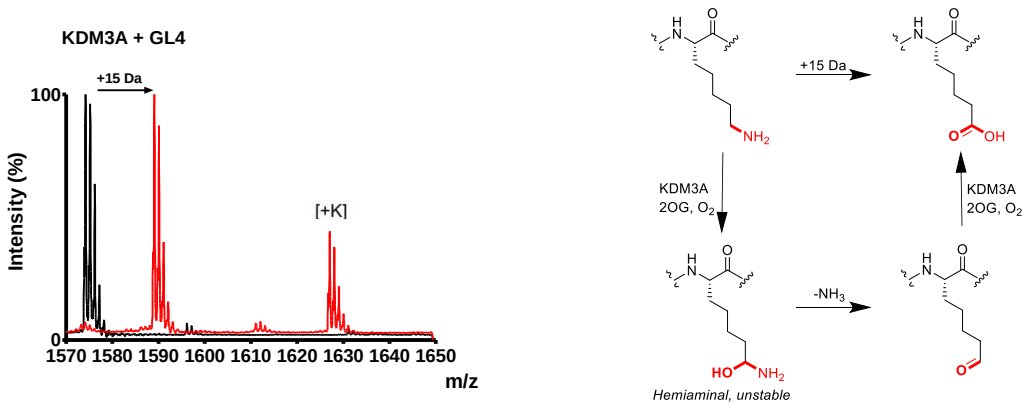


Figure 7.1 KDM3A-catalysed reactions of H3(1-15)K9 peptides with lysine analogues incorporated at position 9 (ARTKQTAR-K-STGGKA). On the left, MALDI-TOF MS spectra of KDM3A-catalysed reactions. Recombinant KDM3A (2 μM) was incubated with the appropriate H3 fragment peptide (10 μM) for 1h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). On the right, proposed mechanism for KDM3A catalysed reactions. (A) "wildtype" H3(1-15)K9me2 peptide substrate, demethylation was observed (mass shift corresponding to -14 Da). (B) Peptide **GL1** was observed to undergo demethylation (-14 Da). (C) Peptide **GL2** was observed to undergo demethylation (-14 Da). (D) For peptide **GL3** one hydroxylation gives hydroxyl-carbon product (+16 Da), two oxidations on the same carbon lead to aldehyde formation after loss of HCO (+14 Da) and three hydroxylations on the same carbon lead to carboxyl formation (+30 Da). (E) Peptide **GL4,** carboxylation was observed (+15 Da). Red spectra show reactions with enzyme and black spectra show the controls with no enzyme.

These results clearly indicate that KDM3A is flexible in terms of substrate conformation, and more importantly, that having a charged ε-amino group is not a requirement for efficient catalysis for all human JmjC KDMs. This information led the design of the experiments described in the following sections.

7.4 Identification of hydroxylation of acetyllysines by KDM3A

7.4.1 Tests with H3(1-15)K9Ac peptide

In light of the new data revealed by the lysine analogue peptide study, it was considered that KDM3A might interact with uncharged PTMs and thus might catalyse additional biologically relevant reactions other than demethylation.

Previous studies have shown that KDM4E, KDM7B and KDM2A are unable to react with *N*ε-acetyllysine peptides under the tested conditions.[9] To find out whether KDM3A could react with acetyllysine, activity assays were conducted with an H3 histone peptide where the natural dimethyllysine (K9me2) substrate for KDM3A was replaced by acetylated lysine (K9Ac). Since acetylation is reported to occur at the H3 K9 position, this substitution results in a biologically relevant sequence.[5,14]

H3(1-15)K9Ac peptide (10 μM) and the corresponding H3(1-15)K9me2 (10 μM) peptide were incubated with isolated recombinant KDM3A (2 μM) under standard demethylation conditions for 1 hour at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Analyses were carried out by MALDI-TOF MS, and enzymatic reactions were compared to the negative controls with no enzyme.

Although the enzyme was active according to the positive control with H3(1-15)K9me2 substrate peptide, no reaction was observed by MS with the H3(1-15)K9Ac peptide (**Figure 7.2**).

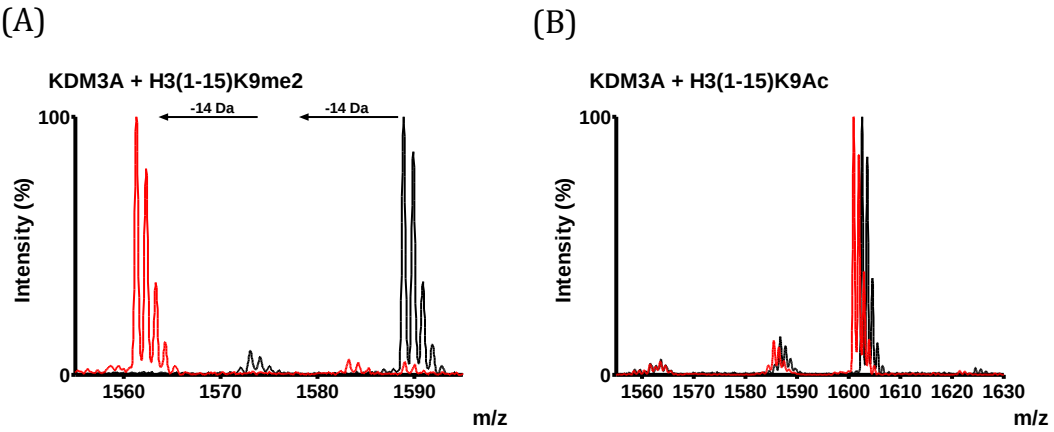


Figure 7.2 MALDI-TOF MS spectra showing no catalysis by KDM3A for H3(1-15)K9Ac peptide. KDM3A (2 μM) was incubated with peptides (10 μM) for 1 h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) Incubation with H3(1-15)K9me peptide was used as a positive control for enzyme activity. (B) Incubation with an H3(1-15)K9Ac peptide. Reactions containing enzyme are shown in red, no enzyme control reactions are shown in black.

7.4.2 Tests with H3(1-15)K4me3K9Ac peptide

KDM3A demethylation efficiency seems to improve in the presence of tri-methyllysine at position H3 K4 which is in proximity to the H3 K9me2 substrate (see Chapter 3, Section 3.5.2.5). It was therefore hypothesised that trimethylation at the H3 K4 position might also facilitate reaction on *N*ε-acetyllysine at position K9. Experiments were conducted with H3(1-15)K4me3K9Ac histone peptide and KDM3A under the same conditions as described in Section 7.4.1. MALDI-TOF MS spectra were compared with the non-enzyme containing negative controls, and activity for the H3(1-15)K9me2 peptide was used as a positive control. A mass shift of +16 Da, corresponding to a possible addition of a hydroxyl group was observed, providing a preliminary evidence for hydroxylation of acetylated lysine peptides by KDM3A. No indication for deacetylation by KDM3A was detected as suggested by the absence of a -42 Da mass shift in the peptide peak (**Figure 7.3**).

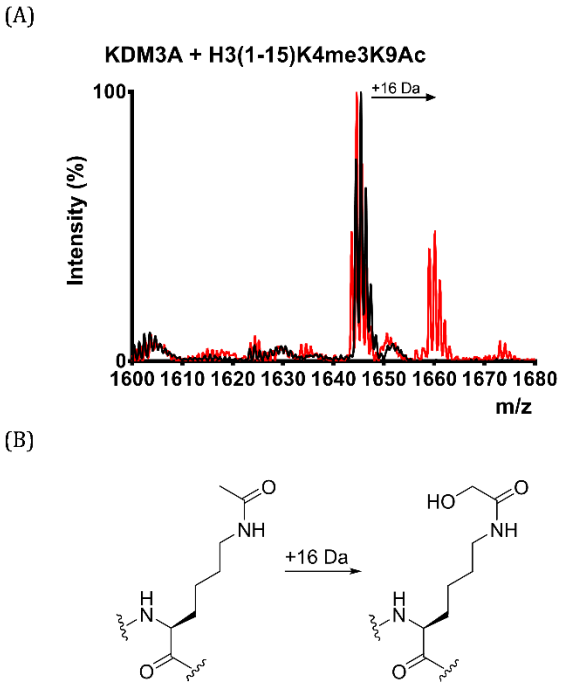


Figure 7.3 Preliminary predicted structure of the hydroxylated product from KDM3A - catalysis with KAc. (A) MALDI-TOF MS spectrum showing catalysis of H3(1-15)K4me3K9Ac peptide(10 μM) by KDM3A (2 μM) following incubation for 1 h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Peptide sequence was ART-Kme3-QTAR-KAc-STGGKA-NH_2. Reaction containing enzyme is shown in red, no enzyme control reactions are shown in black. 33% conversion to the hydroxylated product under the standard assay conditions was observed with this batch of enzyme. (B) Proposed hydroxylation product structure. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson.*

This result provides the first experimental evidence that JmjC KDMs might react with acetyllysines. Moreover, these data suggest further complexity in the histone code and the combinatorial dependence between different PTMs.

7.5 Investigations into the frequency of overlap between H3 K4me3 and H3 K9me2 marks in human cells

H3 K4me3 and H3 K9Ac modifications are reported to commonly coincide in mouse embryonic stem cells, and their co-occurrence suggests a coordinated regulation of active histone marks.[5]

To investigate whether H3 K4me and H3 K9Ac coexistence is also prevalent in human cells, analyses of ChIP-seq datasets from fourteen human derived cell lines were carried out, using ChIP-seq data available from the Encyclopedia of DNA Elements (ENCODE) project (ChIP-*seq datasets were analysed by Tunc Morova and Prof. Nathan Lack)*.[15] All ChIP-seq experiments have used the same antibodies and were analysed by the same bioinformatic methodologies.

Figure 7.4 summarises the percentage overlap of H3 K9Ac and H3 K4me3 in each of the tested cell lines following ChIP-seq analyses. Out of the fourteen cell lines tested, Human Umbilical Vein Endothelial Cells (HUVEC) had the highest percentage overlap between the two PTMs (49%), while HeLa S3 had the lowest overlap (15%). The average overlap was 35%. Overall, these data suggest that co-occurrence of H3 K9Ac and H3 K4me3 is relatively common in human cells.

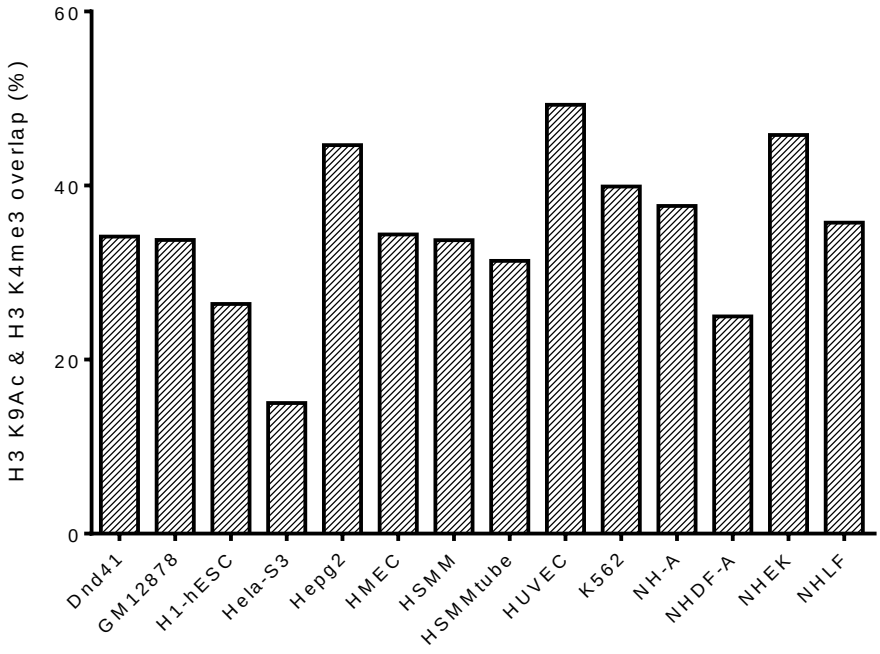


Figure 7.4 Frequency of overlap between H3 K9Ac and H3 K4me3 marks in various cell lines. Analyses of fourteen human cell lines with ENCODE ChIP-seq datasets for both marks.[15] ChIP-seq experiments have used the same antibodies and were analysed by the same bioinformatics methodologies, and are therefore comparable. *ChIP-seq datasets were analysed by Tunc Morova and Prof. Nathan Lack.*

7.6 Validation of *in vitro* hydroxylation by KDM3A

Experiments described in Section 7.4.2 have revealed previously unidentified putative hydroxylation reaction on an acetylated lysine residue catalysed by recombinant KDM3A *in vitro.* Experiments then aimed to validate the reaction product of KDM3A – catalysed modification of KAc, as well as to confirm that the reaction is catalysed by KDM3A itself, and is dependent on the reported JmjC 2OG co-substrate and Fe^{II} co-factor.

All experiments were carried out under the standard conditions as described in Section 7.4.2, unless otherwise stated. In each test, positive control reactions with H3(1-15)K4me3K9Ac peptide indicated that the enzyme was active. Analyses were conducted by MALDI-TOF MS. A shift in peptide mass was compared to the negative controls with no enzyme.

7.6.1 Identification of the modified residue using a reference control peptide

The work described in Section 7.4.2 indicated the possibility of hydroxylation on H3 K9Ac as catalysed by KDM3A, in the presence of H3 K4me3 PTM. In order to ascertain whether hydroxylation occurs on KAc rather than on another residue, reactions with a reference control peptide were carried out.

The control peptide used in this experiment was H3(1-21)K4me3, which contains only the tri-methylation mark, without acetylation at lysine 9. If the detected + 16 Da mass shift described in Section 7.4.2 was indeed due to a chemical change on the acetyl at the H3 K9 position, it is expected that KDM3A would not interact with the control peptide which lacks this acetyl.

No reaction was detected with the H3(1-21)K4me3 peptide (**Figure 7.5**). This result supports the assumption that the reaction is likely to occur on the H3 K9Ac mark rather than on another residue in the peptide.

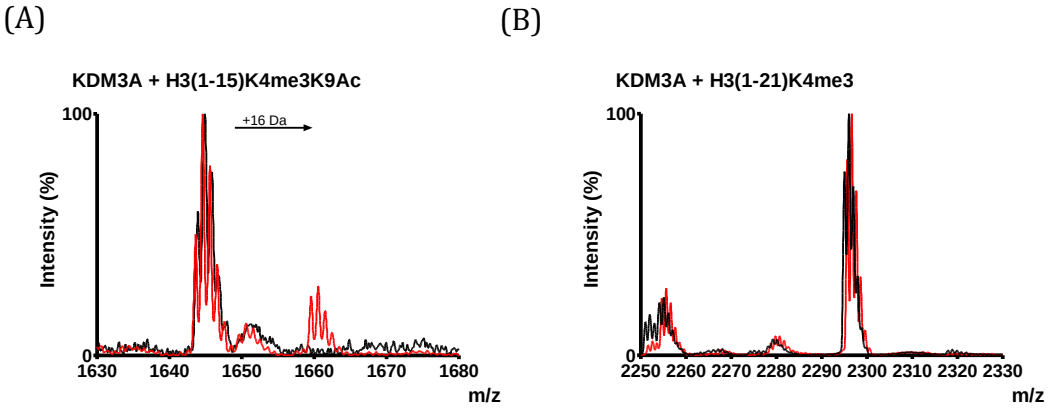


Figure 7.5 Absence of KDM3A enzymatic activity with H3(1-21)K4me3 reference control peptide. KDM3A (2 μM) was incubated with each peptide (10 μM) for 1 h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) MALDI-TOF MS spectrum showing catalysis of H3(1-15)K4me3K9Ac peptide as a positive control. 22% conversion to hydroxylated product was observed under the standard assay conditions with this batch of enzyme. (B) MALDI-TOF MS spectrum showing absence of KDM3A catalytic activity for H3(1-21)K4me3 peptide. Reactions containing enzyme are shown in red, no enzyme control reactions are shown in black. *H3(1-15)K4me3 peptide was synthesised by Roman Belle. H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson.*

7.6.2 Deciphering hydroxylation position using deuterated *N*ε-acetyllysine substrate

Section 7.6.1 suggests that the mass shift observed when incubating KDM3A with the H3(1-15)K4me3K9Ac peptide is likely due to reaction occurring on the H3 K9Ac residue. To find out the position within this lysine residue in which hydroxylation takes place, the reaction was monitored using deuterated *N*ε-acetyllysine substrate.

The deuterated peptide used in this experiment is identical to that tested in Section 7.4.2 (H3(1-15)K4me3K9Ac) with the exception of having a *d3*-acetyl at the K9 position instead of a protiated acetyl. It was expected that hydroxylation of the *d3*-acetyl group would result in a mass shift of +15 Da, rather than a +16 Da shift if it took place elsewhere in the peptide.

Incubation of KDM3A with the deuterated H3(1-15)K4me3K9*d3*-Ac peptide resulted in a +15 Da shift according to MALDI-TOF MS analyses, while incubation with the protiated H3(1-15)K4me3K9Ac peptide in the same assay caused a +16 Da shift in peptide mass (**Figure 7.6**). These results strongly support the proposal that the hydroxylation takes place on the acetyl CH_3 group of H3 K9.

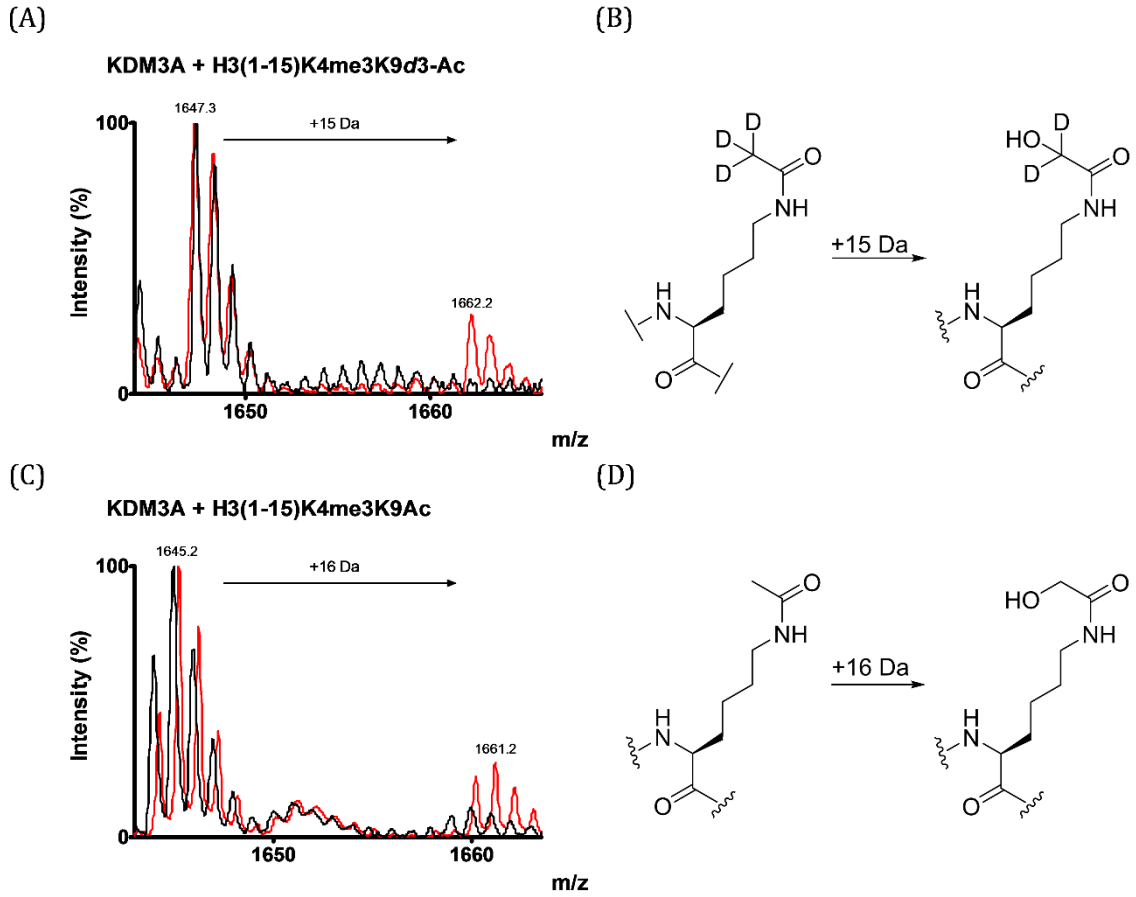


Figure 7.6 Deuterated substrate indicates that KDM3A hydroxylates at the acetyl. KDM3A (2 μM) was incubated with each peptide (10 μM) for 1 h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) MALDI-TOF MS spectrum showing catalysis of H3(1-15)K4me3K9*d*3-Ac peptide. (B) Acetyl hydroxylation on of the deuterated acetyl would result in a mass shift of +15 Da. (C) MALDI-TOF MS spectrum showing catalysis of H3(1-15)K4me3K9Ac peptide as a positive control. (D) Acetyl hydroxylation of the protiated acetyl would result in a mass shift of +16 Da. Reactions containing enzyme are shown in red, no enzyme control reactions are shown in black. *Fmoc-protected deuterated* N^{ε}-acetyllysine *was synthesised by Roman Belle. H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson.*

7.6.3 Investigations into JmjC co-substrate and co-factor dependence

Acetyllysine hydroxylation was hypothesised to be catalysed by the JmjC domain of KDM3A, thus it was expected that the same co-factor and co-substrate needed for demethylation would likely be required for hydroxylation. JmjC demethylases require Fe^{II} as a co-factor to mediate electron transfer during substrate oxidation. For substrate oxidation to occur JmjC KDMs also require 2OG as a co-substrate which converts to succinate (see Introduction, Chapter 1, Section 1.13).[8]

To investigate whether hydroxylation is dependent on 2OG and Fe^{II}, experiments measuring activity with and without co-factor and co-substrate and with other metals (Zn^{II}, Mn^{II}) were carried out. Other assay conditions were kept as described in Section 7.4.1. Data were analysed by MALDI-TOF MS.

Table 7.2 summarises the assays results. The data suggest that hydroxylation is dependent on 2OG and Fe^{II}, and that the replacement of Fe^{II} with other metal ions tested here could not recover the catalytic activity. These results indicate that hydroxylation of acetyllysines is likely to occur via the JmjC domain in a similar catalytic mechanism to that of demethylation, and correlates with data about co-factor dependence in additional novel PTM reactions catalysed by KDM3A (see Chapters 6, Section 6.4 and Chapter 9, Section 9.5).

Table 7.2 Hydroxylation of acetyllysines by KDM3A is dependent on 2OG and Fe^{II}. Summary of MALDI-TOF MS data of activity assays with KDM3A (2 μM) on H3(1-15)K4me3K9Ac peptide (10 μM) performed over 1h at 37°C. Assays were carried out under standard conditions, with 2OG (100 μM) and Fe^{II} (50 μM) except for the specified change described in the table (see Materials and methods, Chapter 11, Section 11.3.3, for more information). X denotes no reaction was detected. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson.*

Co-factor change from standard conditions	KDM3A hydroxylation catalysis
No 2OG	X
No Fe^{II}	X
Zn^{II} instead of Fe^{II} (50 μM)	X
Mn^{II} instead of Fe^{II} (50 μM)	X

7.6.4 Confirmation of KDM3A hydroxylation of acetyllysines by mutation study

To examine whether the hydroxylation reaction is catalysed by KDM3A rather than by another enzyme that had co-purified with it, experiments with a catalytically inactive variant were conducted. It was assumed that a construct that differs from the wildtype sequence by only one-point mutation would be expressed and purified in a similar manner, thus any impurity found with the wildtype protein would also be present there.

KDM3A and KDM3B have a conserved tyrosine residue in their active site (**Figure 7.7 A**). JMJD1C has a cysteine instead of this conserved tyrosine and there is no evidence for JMJD1C - catalysed demethylation (see Introduction, Chapter 1, Section 1.14, for sequence and structural comparison between the KDM3 subfamily).[17] X-ray crystallographic studies suggest that this tyrosine might facilitate interaction with the co-substrate 2OG, which can explain its importance in catalysis (**Figure 7.7 B**, PDB: 2YPD and 4C8D).

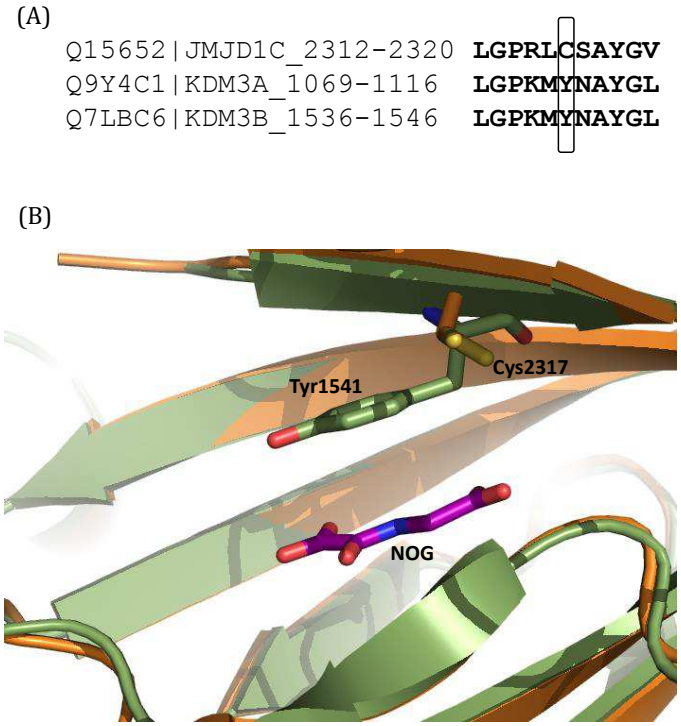


Figure 7.7 Conserved tyrosine residue at the active site of KDM3A and KDM3B is replaced by cysteine in JMJD1C. (A) Sequence alignment between JMJD1C (top), KDM3A (middle) and KDM3B (bottom). Partial sequences for JMJD1C, KDM3A and KDM3B (Uniprot numbers Q15652, Q9Y4C1 and Q7LBC6 respectively) were aligned using Clustal Omega EMBL-EBI.[16] Conserved tyrosine and its cysteine replacement in JMJD1C are highlighted. (B) Structural comparison between the JmjC domain of JMJD1C (orange, PDB: 2YPD) and KDM3B (green, PDB: 4C8D) in respect to 2OG (replaced here by NOG, pink) binding. The conserved Tyr1541 in KDM3B, the corresponding Cys2317 in JMJD1C and NOG are shown as sticks with labels.

Given the importance of this residue, it was expected that a Tyr1101Cys substitution would abolish KDM3A demethylation and hydroxylation of acetyllysines because both of these reactions are very likely 2OG - dependent.

Activity assays were carried out in order to ascertain the relative activities of the wildtype KDM3A and the available Tyr1101Cys variant with respect to demethylation and hydroxylation, followed by MALDI-TOF MS analyses as described in Section 7.4.1. The resulting data reveal that mutating Tyr1101 to cysteine in KDM3A caused loss of demethylation and hydroxylation activity (**Figure 7.8**). These results strengthen the assumption that the detected acetyllysine hydroxylation is catalysed by KDM3A rather than by another enzyme that had co-purified with it.

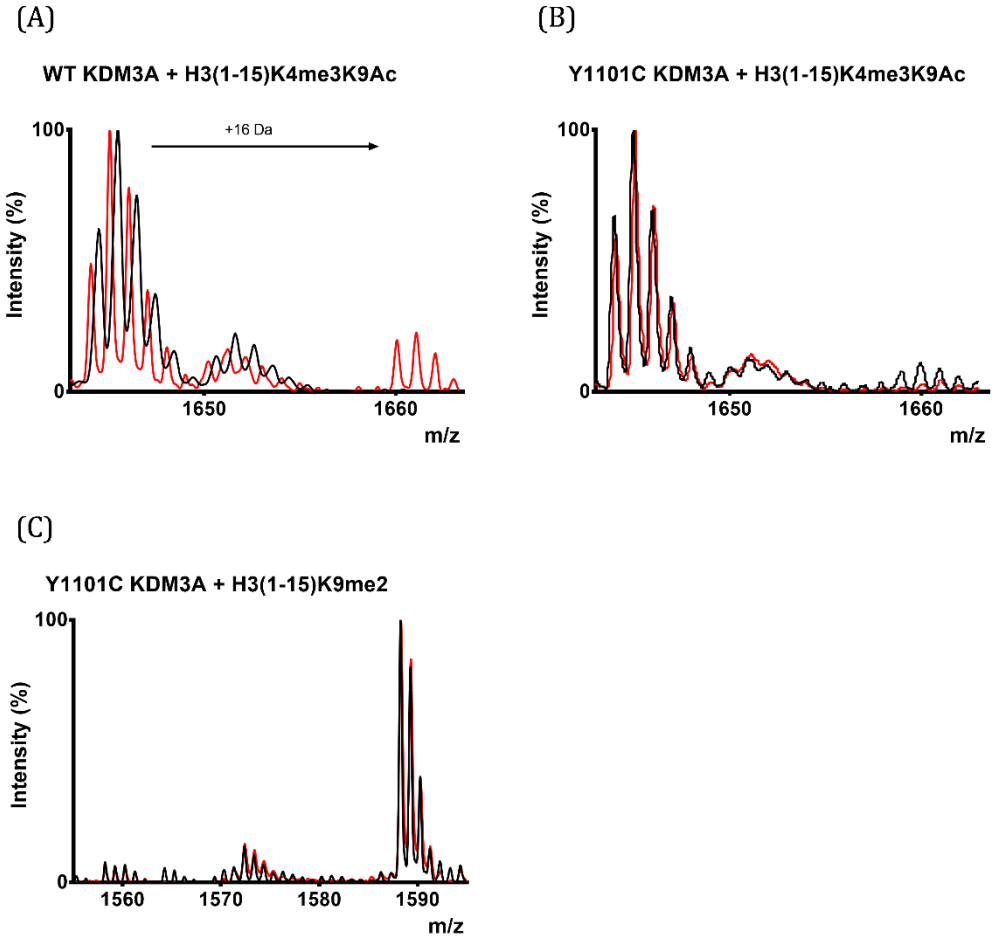


Figure 7.8 Y1101C KDM3A mutation is catalytically inactive. WT or variant KDM3A (2 µM) were incubated with each peptide (10 µM) for 1 h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) MALDI-TOF MS spectrum analysis showing catalysis of H3(1-15)K4me3K9Ac peptide by WT KDM3A as a positive control for assay conditions. (B) MALDI-TOF MS spectrum showing absence of catalytic activity of Y1101C KDM3A variant for H3(1-15)K4me3K9Ac peptide. (C) MALDI-TOF MS spectrum showing absence of catalytic activity of Y1101C KDM3A variant for of H3(1-15)K9me2 peptide. Reactions containing enzyme are shown in red, no enzyme control reactions are shown in black. *KDM3A catalytically inactive variant was provided by Dr. Catrine Johansson. H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson.*

7.7 Optimisation of *in vitro* assay conditions for hydroxylation of acetyllysines

The purpose of the following experiments described in this section was to optimise the hydroxylation reaction conditions. Maximising product formation should help in reaction characterisation and ensure that the best conditions are used for screening of additional substrates. Finding the optimal 2OG concentration will ensure that the co-substrate is not rate-limiting, which is important for kinetic characterisation of substrate efficiency.

7.7.1 Time course for hydroxylation of acetyllysines

Product formation for the hydroxylation reaction was measured over time under the conditions described in Section 7.4.1. The percentage of product formation was determined using MALDI-TOF MS by calculating the intensity of hydroxy-acetylated peptide product peak intensity and dividing this value by the sum of the intensities of substrate and product peaks (see Materials and methods, Chapter 11, Section 11.3.3, for more information). The hydroxylation product from peptide H3(1-15)K4me3K9Ac continued to be formed between 15 to 60 minutes of incubation (**Figure 7.9**). After 60 minutes of incubation no more hydroxylation product was formed. However, product levels only reached 15% of the total peptide concentration.

Since hydroxylation of acetyllysines product levels reached less than half of the level of demethylation product, and the reaction rate was much slower (see Chapter 3, Section 3.5.1.1), it seemed possible that substrate was turned over less efficiently in the reaction of hydroxylation than in demethylation. As all tests were carried out with peptide substrates, this comparison might not necessarily reflect the efficiency of the reaction in cells with protein substrates. Nevertheless, it was envisaged that validation of the use of saturating 2OG co-substrate concentration and determination of the steady-state kinetic parameters (K_M, k_{cat}) for hydroxylation of acetyllysines is essential for efficiency comparison between the reactions.

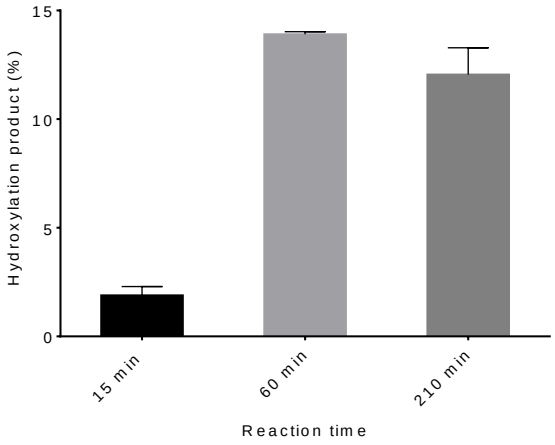


Figure 7.9 Percentage of hydroxylation product formation over time. Measuring percentage of product formation over time under standard assay conditions, in which KDM3A (2 μM) and H3(1-15)K4me3K9Ac peptide (10 μM) were incubated at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Percentage of product formation was calculated as the ratio between product intensity peak and the sum of the product and substrate intensity peaks as observed by MALDI-TOF MS spectra. Data plotted as the mean of triplicate with standard deviation. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson.*

7.7.2 Determination of saturating concentrations of 2OG for KDM3A hydroxylation of acetyllysines

The experiments described in Section 7.6.3 indicated that hydroxylation of acetyllysines by KDM3A requires 2OG. To maximise product formation and also to determine steady-state kinetic parameters (K_M, k_{cat}) the use of saturated 2OG co-substrate concentration must be validated to ensure that 2OG is not a rate limiting factor.

To investigate the optimal level of 2OG, activity assays were carried out with increasing concentrations of 2OG (0 to 3 mM), while keeping other factors under the standard conditions described in Section 7.4.1.

Figure 7.10 shows that the product ratio has increased in line with increasing concentrations of 2OG up to 250 μM of 2OG. Product formation at 2OG concentrations of 750 μM and above was slightly decreased, which might indicate on a weak co-substrate inhibition. These observations suggest that a concentration of 250 μM 2OG is optimal and will not limit the hydroxylation reaction rate.

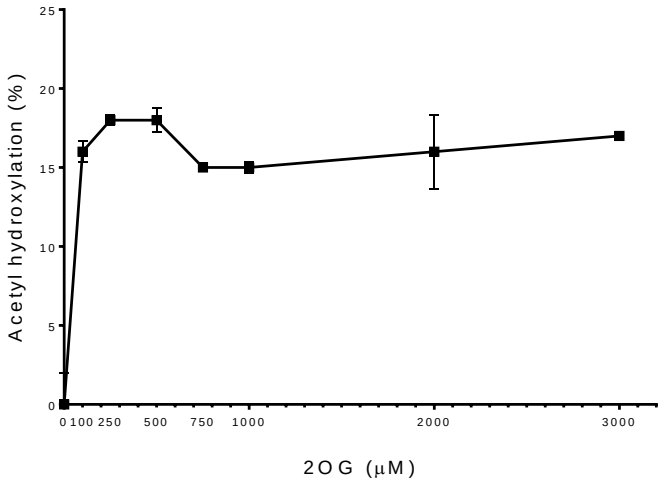


Figure 7.10 KDM3A hydroxylation catalysis under varying 2OG concentrations. Evaluation of KDM3A (2 μM) hydroxylation of acetyllysines activity on H3(1-15)K4me3K9Ac peptide (10 μM) for incubation with various concentrations of 2OG (0-3000 μM). Percentage product formation was determined using MALDI-TOF MS by calculating the intensity of hydroxy-acetylated peptide product peak area and dividing this value by the sum of the intensities of substrate and product peaks (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Points are plotted as the mean of triplicate with standard deviation. The *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson.*

7.8 Kinetic characterisation of KDM3A hydroxylation of acetyllysines

Preliminary kinetic characterisation was carried out in order to compare the catalytic efficiency of KDM3A hydroxylation of acetyllysines to that of lysine and arginine demethylation (see Chapters 3, Section 3.5.2, and Chapter 6, Section 6.7 for more information).

Since the FDH coupled assay can only detect reactions in which formaldehyde is released, kinetic characterisation of hydroxylation could not be carried out using this method.[17] Therefore, kinetic studies were performed using MALDI-TOF MS. The method employs manual quenching of the reaction at different time-points along the linear range of the reaction, and uses micromolar concentrations of enzyme as opposed to real-time monitoring of the reaction and use of nano-molar concentrations in FDH experiments.

All reagents were kept in excess to ensure that the substrate peptide was the only rate limiting factor for the reaction (see Section 7.7.2 and Materials and methods, Chapter 11,

Section 11.3.3, for more information). For kinetic characterisation, initial rates for the reaction of KDM3A with varying concentrations of H3(1-15)K4m3K9Ac peptide (0-400 μM) were measured. Kinetic parameters were calculated based on two biological repeats, each with three technical replicates (**Figure 7.11**). Apparent K_M and k_{cat} values were determined by fitting the Michaelis-Menten model using GraphPad Prism 6, and the ratio of k_{cat}/K_M was then calculated as a measure of the catalytic efficiency for the substrate (**Table 7.3**).

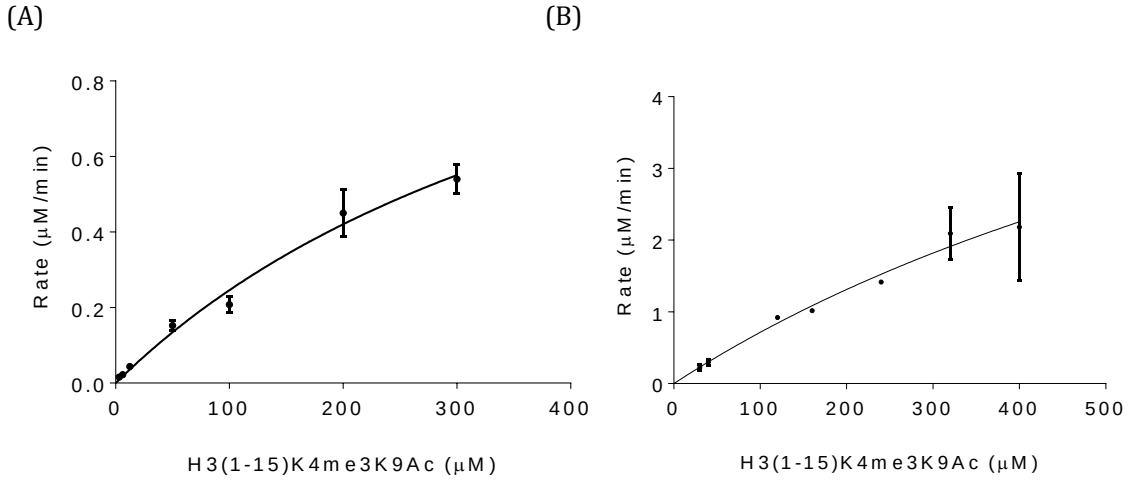


Figure 7.11 Michaelis-Menten graph measuring initial rates of hydroxylation reaction catalysed by KDM3A with the H3(1-15)K4me3K9Ac peptide Data were determined by MALDI-TOF MS assay, and was fitted using Michaelis-Menten model by GraphPad Prism 6 (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) and (B) are the biological replicates. Points are plotted as the mean of tree technical replicates with error bars showing ± s.d. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson.*

Table 7.3 Kinetic parameters for KDM3A hydroxylation with H3(1-15)K4me3K9Ac peptide substrate. Apparent kinetic parameters (K_M and k_{cat} values) were determined by MALDI-TOF MS assay (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Data show the mean (n = 3) of technical repeats for each biological replicate, as well as the average of the biological repeats ± s.e.m.

Biological replicate	Enzyme (μM)	K_M (μM)	k_{cat} ($\times 10^{-3}$ s^{-1})	k_{cat}/K_M (10^{-6} /s μM)
i	2	489 ± 192	12 ± 3	25
ii	3	1027 ± 542	45 ± 18	44
Average		**758 ± 269**	**28 ± 16**	**38**

The apparent K_M values for the two biological repeats varies with a large s.e.m. (489 and 1027 μM). This could be due to the low amounts of product formation when quenching the reaction at the linear range, which might lead to a low signal to noise ratio in MALDI-TOF MS spectra (maximal product formation under standard conditions was 20%, see Section 7.7.1., 7% of product was present when quenching the reaction at the linear range- *data not shown*). Nevertheless, the apparent average K_M value for the hydroxylation catalysis was relatively high (758 ± 269 μM), which is likely to indicate a low binding affinity of the tested substrate peptide to KDM3A. This K_M value was apparently an order of magnitude higher than that for the lysine and arginine demethylation reactions, suggesting that demethylation substrates are better recognised by KDM3A than hydroxylation substrates (**Table 7.4**). The apparent average k_{cat} for hydroxylation reaction was 28 ± 16 ×10-3 s^{-1}, which higher than that of the tested arginine methylated substrate but an order of magnitude lower than that of the lysine methylated peptide. These results might reflect on a slower turnover of the newly identified arginine demethylation and hydroxylation of acetyllysines compared to lysine demethylation by KDM3A.

Overall, the apparent catalytic efficiency for hydroxylation of acetyllysines by KDM3A as derived from the k_{cat} /K_M ratio was very low (38 ×10^{-6} /s μM). This value was an order of magnitude lower than the ratio for catalysis of the methylated arginine substrates, and two orders of magnitude lower than that of the lysine methylated substrate (**Table 7.4**). Notably, reaction did not reach saturation at both biological repeats at substrate concentrations of up to 400 μM.

Table 7.4 Kinetic parameters for KDM3A histone peptide substrates. Apparent kinetic parameters for H3(1-15)K4me3K9Ac were determined by MALDI-TOF MS analyses. Data show mean of biological repeats (n = 2) ± s.e.m. Apparent kinetic parameters for all other substrates were determined by formaldehyde dehydrogenase-coupled assay. Data show mean of biological repeats (n = 3) ± s.e.m. Histone N^{ε}-dimethylated-lysine peptide substrate is marked in bold. See Chapters 3, Section 3.5.2, and Chapter 6, Section 6.7 for more information.

Substrate	K_M (μM)	k_{cat} ($\times 10^{-3}$ s^{-1})	k_{cat}/K_M ($\times 10^{-6}$ /s μM)
H3(1-15)K9me2	**54 ± 13**	**790 ± 112**	**14649**
H3(1-15)K9Rme2a	12 ± 1	5.0 ± 0.6	431
H3(1-15)K9Rme2s	24 ± 2	3.0 ± 0.3	127
H3(1-15)K9Rme	12 ± 3	11 ±1	908
H3(1-15)K4me3K9Ac	758 ± 269	28 ± 16	38

The resulting data imply that the H3(1-15)K4me3K9Ac peptide is likely to be a poor substrate for KDM3A. However, the actual k_{cat}/K_M value for the hydroxylation catalysis might be different with full length histones and possibly other partner proteins that contain the full repertoire of PTMs, as opposed to the short peptide tested here.

7.9 Screening for N^{ε}-acetyllysine hydroxylation on other WT histone derived peptides

Considering the hydroxylation activity of KDM3A on the histone H3 fragment peptide H3(1-15)K4me3K9Ac, it was reasonable to propose that KDM3A might also react with additional natural acetylated positions detected on histone tails.

To investigate whether KDM3A can hydroxylate acetyllysines on other histone positions, peptides for eight reported acetylated lysine positions on histones H3 and H4 were tested (**Table 7.5**). Histone H3 fragment peptides were designed based on acetylated peptide sequences in the AltaBioscience peptide array.[18] Histone H4 derived peptides are reported binders for the BRD4 bromodomains, they were selected for this screen to test whether KDM3A can hydroxylate them.[19] Determination of the binding affinity of BRD4 to hydroxy-

acetyl lysine vs. *N*ε-acetyllysine is described in Section 7.10. Screening assays were using optimised conditions with 2OG concentration of 250 μM (see Section 7.7.2).

Interestingly, although KDM3A was shown to be active according to the reaction with H3(1-15)K4me3K9Ac, no reaction was observed for any of the tested H3 and H4 fragment peptides (**Figure 7.12**). Therefore, these results appear to suggest that hydroxylation of acetyllysines by KDM3A is sequence selective, i.e. it currently can only be detected for the H3(1-15)K4me3K9Ac peptide. No reaction was observed with the H3(1-15)K9Ac peptide (see Section 7.4.1), suggesting that activity is dependent on the K4me3 position for K9Ac hydroxylation. Therefore, a specific combination of modifications might also be necessary for detection of this reaction at other histone positions.

Table 7.5 Acetylated histone peptide sequences and the mark around which they were designed. Peptides were used as potential substrates for KDM3A. Target acetyl-residues are shown in bold.

Histone mark	Peptide sequence
H3 K14Ac	ARKSTGG**-KAc-**APRKQLA
H3 K18Ac	TGGKAPR**-KAc-**QLATKAA
H3 K23Ac	PRKQLAT**-KAc-**AARKSAP
H3 K27Ac	LATKAAR**-KAc-**SAPATGG
H3 K115Ac	NLCAIHA**-KAc-**RVTIMPK
H4 K5Ac	SGRG**-KAc-**GGKGLGKGGA
H4 K5,8Ac$_2$	SGRG**-KAc-**GG**-KAc-**GLGKGGA
H4 K12Ac	KGGKGLG**-KAc-**GGAKRHR
H4 K16Ac	GLGKGGA**-KAc-**RHRKVLR
H4 K12,16Ac$_2$	GKGLG**-KAc-**GGA**-KAc-**RHRKV

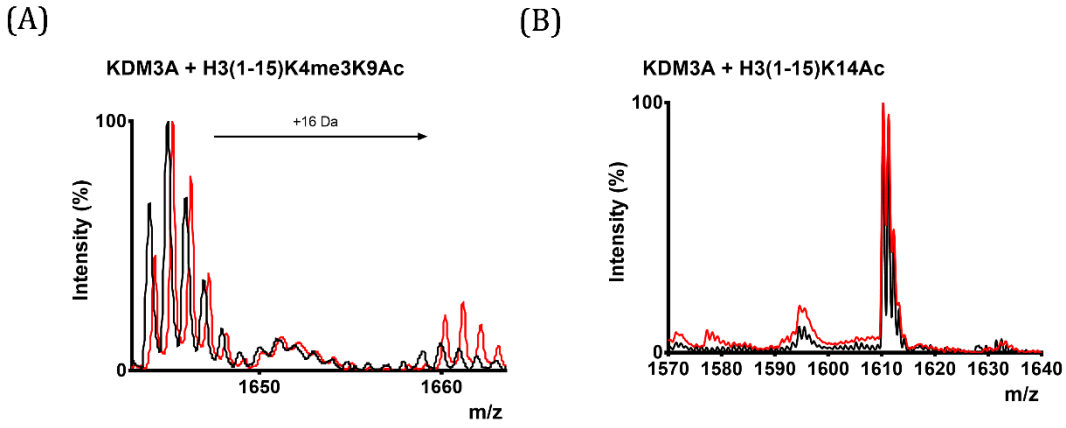


Figure 7.12 Example of MALDI-TOF MS spectra from histone peptides screen. (A) KDM3A hydroxylates the positive control peptide H3(1-15)K4me3K9Ac. (B) Absence of KDM3A catalytic activity for H3(1-15)K14Ac peptide. Peptides (10 μM) were incubated with KDM3A under optimised conditions, followed by analysis by MALDI-TOF MS (see Section 7.7.2). Reactions with enzyme are shown in red, control reactions with no enzyme are shown in black. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson.*

7.10 Probing for affinities for BRD4 and hydroxy-acetyllysine modification

The bromodomain (BRD) is a conserved structural module in chromatin-associated proteins and histone acetyltransferases reported to recognise acetyllysine residues on histone tails.[20] As hydroxylation might be a modification of lysine acetylation, hydroxylation of acetyllysines is likely to modulate the binding affinity of bromodomains to their histone tails ligands.

BRD4 binds to acetylated H4 histone acetylated lysine residues via its tandem bromodomains BD1 and BD2.[19] Although there was no prior evidence for KDM3A acetyl hydroxylation at BRD4 recognised histone H4 positions (see Section 7.9), it was chosen as a proof of principle for initial investigation to test the effect of hydroxylation on acetyllysines on binding to bromodomains.

Comparison of the binding affinity of H4(1-20)K5,8,12,16Ac4 (acetylation only), H4(1-20)K5AcOH K8,12,16Ac3 (hydroxy-acetylated at Lys5, acetylated on the other lysines) and H4(1-20)K5,8AcOH K12,K16Ac2 (hydroxy-acetylated at Lys5 and Lys8, acetylated on the other lysines) peptides to BRD4 was tested using the AlphaScreen displacement assay.[21] In

this assay the modified peptides were titrated for displacement of a biotinylated H4(1-20)K5,8,12,16Ac4 (acetylation only) peptide which was bound to BRD4. IC_{50} values were determined in GraphPad Prism 6 after normalisation of the AlphaScreen signal against the corresponding no peptide controls.

The acetylated H4(1-20)K5,8,12,16Ac4 peptide displaced its biotinylated peptide with an IC_{50} value of 3.9 μM, demonstrating that the acetylated peptide competes with biotinylated peptide for BRD4 binding (**Figure 7.13 and Table 7.6**). This IC_{50} value closely resembles the value reported by Philpott *et al*. for a similar peptide Y-H4(1-20)K5,8,12,16Ac4, in the same AlphaScreen displacement assay with BRD4 (IC_{50} value of 5.4 μM).[21] Competition assays with other acetylated and hydroxy-acetylated peptides revealed that H4 fragment peptides containing hydroxy-acetyllysine can bind BRD4. However, the more hydroxy-acetyllysine residues are present in the peptide, the lower is the affinity for BRD4 (IC_{50} values of 19.5 and 27.3 μM for mono- and di- hydroxy-acetyllysine peptides respectively). These results demonstrate that hydroxylation of acetyllysines could modulate the binding affinity of BRD4 to its histone tails ligand positions.

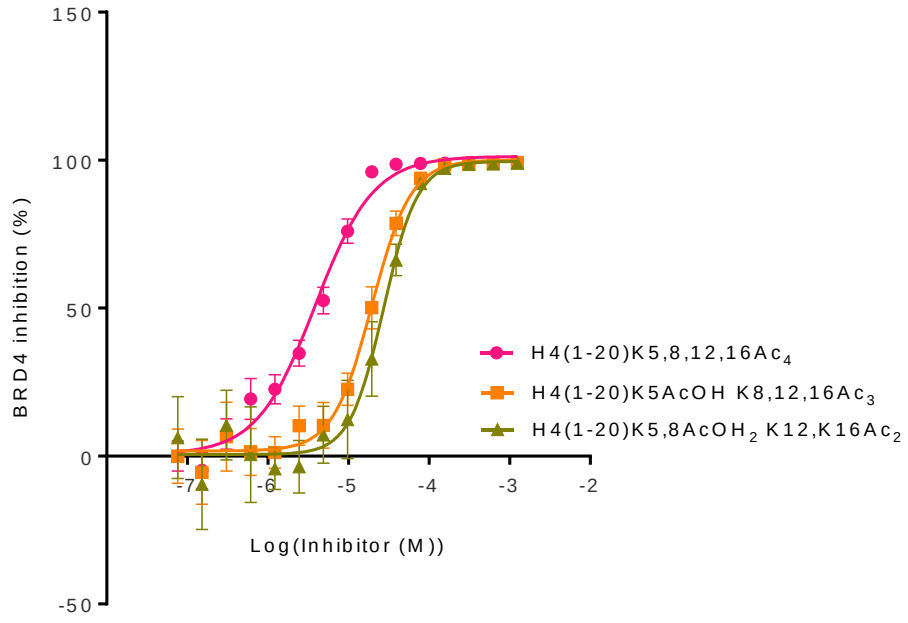


Figure 7.13 AlphaScreen displacement assay of BRD4 with lysine acetylated and hydroxy-acetylated peptides. Peptides affinities against BRD4 were measured by the AlphaScreen displacement assay. Titrations of each of the non-biotinylated versions of the histone H4 peptide was used to displace the biotinylated H4(1-20)K5,8,12,16Ac4 peptide for IC_{50} determination. All measurements were conducted in triplicate and values represent mean and standard deviation. *AlphaScreen measurements were carried out by David Zollman.*

Table 7.6 IC_{50} values of modified peptides displacement of biotinylated peptide from BRD4. Investigating the binding affinity of lysine acetylated and hydroxy-acetylated peptides to BRD4 using AlphaScreen displacement assay. IC_{50} values were calculated in GraphPad Prism 6 after normalisation of the AlphaScreen signal against the corresponding controls with no peptides. *AlphaScreen measurements were carried out by David Zollman.*

Peptide	IC_{50} (μM)	95% Confidence Intervals (μM)
H4(1-20)K5,8,12,16Ac4	3.9	3.1 - 4.8
H4(1-20)K5AcOH K8,12,16Ac3	19.5	15.6 – 24.4
H4(1-20)K5,8AcOH2 K12,K16Ac2	27.3	20.2-37.0

7.11 Screening for *N*ε-acetyllysine hydroxylation on IFT

In light of the novel observation of the *N*ε-acetyllysine hydroxylation reaction and the continued search for intraflagellar transport (IFT) substrates for KDM3A (Chapter 5), it was envisaged to screen for activity against acetylated IFT derived peptides.

MS analyses on IFT proteins extracted from mouse testes did not detect any acetylated lysine residues (*Dr. Patricia Yeyati and Dr. Holger Kramer, manuscript in preparation*). With lack of *N*ε-acetyllysine detection, eight peptides were designed based on sequences that were observed to interact with KDM3A, by choosing lysine residues instead of the detected methylated lysines on IFT and also by substitution of the methyllysine positions by acetyl lysines (**Table 7.7**). Additional peptide, IFT81(1-15)K596Ac was designed based on *N*ε-acetyllysine prediction by PhosphoSitePlus® (PSP) database.[22]

Activity assays were carried out, with optimised conditions of 250 μM 2OG (Section 7.7.2). H3(1-15)K4me3K9Ac peptide was tested as a control for KDM3A enzymatic activity. MALDI-TOF MS analyses implied that although KDM3A was active, no catalytic activity was observed for the tested IFT lysine acetylated peptides (**Figure 7.14**).

Table 7.7 Lysine-acetylated IFT peptides screened as potential substrates for KDM3A. Target acetylated-residues are shown in bold.

Protein	Target acetylated residue	Peptide sequence
IFT74	K81	QQGLTGM-**KAc**-TGTKGPQ
IFT81	K545	GLESNRS-**KAc**-LEQEVRR
IFT81	K563	EALQEES-**KAc**-YHYTNAM*
IFT81	K596	ISSDQQE-**KAc**-RKAIREQ
IFT81	K649	LEQLMEA-**KAc**-KQAFLKQ*
IFT81	K650	LEQLMEAK-**KAc**-QAFLKQ*
IFT88	K71	IATGYGS-**KAc**-TSLASSI
IFT88	K138	PASPLEA-**KAc**-KKDSPEE
IFT88	K332	FAIGDRE-**KAc**-MKKAFQK

**All alanine residues in the peptide sequence are replacing cysteine residues in the original protein sequence to block the formation of a disulfide bond.*

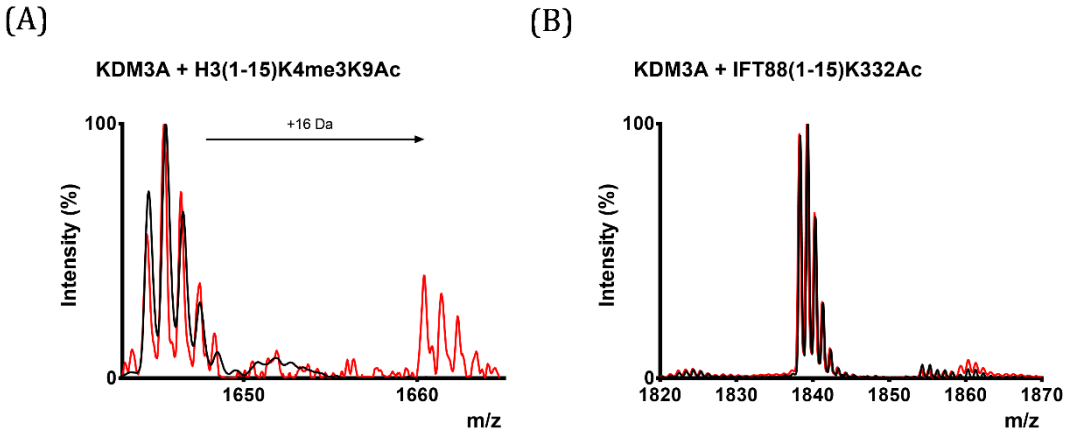


Figure 7.14 Example of MALDI-TOF MS spectra of IFT peptides screen. (A) KDM3A hydroxylates the positive control peptide H3(1-15)K4me3K9Ac. (B) Absence of KDM3A catalytic activity for IFT88(1-15)K332Ac peptide. Peptides (10 µM) were incubated for 2 h at 37°C with KDM3A (2 µM) under optimised conditions (see Section7.7.2). Reactions with enzyme are shown in red, control reactions with no enzyme are shown in black. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson.*

7.12 Screening for *N*ε-acetyllysine hydroxylation by a panel of KDMs

Previous work by Hopkinson *et al.* has demonstrated that KDM4E, KDM7B (PHF8) and KDM2A (FBXL11) are unable to react with histone peptides containing *N*ε-acetyllysine.[9] The work described in this section aimed to expand this screen for additional KDMs. The panel in these experiments included KDM6B, KDM4A and two additional KDM3 proteins – KDM3B and JMJD1C.

For each KDM test, a histone peptide containing the methylated mark recognised by this demethylase was used as a positive control. Acetylated substrate peptides were designed such that the methylated lysine target was substituted by *N*ε-acetyllysine. Since the JMJD1C substrate is unknown or might not exist, the acetylated peptide used for tests was identical to that of KDM3A. The sequences of the peptides and the corresponding recombinant KDM for which they were tested are given in **Table 7.8**.

Reactions were carried out under the standard conditions with incubation of 2 μM enzyme and 10 μM peptide for two hours at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). The only difference for KDM6B assays was the addition of salt to the reaction buffer (150 mM NaCl).

Analyses were carried out by MALDI-TOF MS. For KDM6B, KDM4A and KDM3B -14 Da shifts in the mass of the control peptide containing methyllysine, compared to the no-enzyme controls, indicated demethylation activity. No reactions were observed for the analogue acetylated peptides when these were incubated with KDM6B, KDM4A and KDM3B (**Figure 7.15 A-C**).

Interestingly, a -42 Da shift was detected for H3(1-15)K4me3K9Ac peptide after incubation with JMJD1C (**Figure 7.15 D**). This mass shift is likely to indicate that a deacetylation reaction has taken place as the mass of an acetyl group is 42 Da. If indeed this reaction would prove to be catalysed by JMJD1C, it would be the first identified substrate for this enzyme, and would be an evidence for a novel HDAC activity catalysed by the KDM family. The apparent JMJD1C deacetylation activity is described in Chapter 8.

Table 7.8 Recombinant KDM enzymes tested in this study, their lysine methylated peptide target sequence and the position where acetyl modification substituted the lysine. *KDM6B was provided by Kate Dunne. KDM4A was provided by Rebecca Hancock. KDM3B and JMJD1C were provided by Dr. Catrine Johansson.*

Enzyme	Histone mark	Peptide sequence
KDM6B	H3 K27me3	KAPRKQLATKAAR-**Kme3**-SAPATGG*
	H3 K27Ac	LATKAAR-**KAc**-SAPATGG
KDM4A	H3 K9me3	ARTKQTAR-**Kme3**-STGGKA
	H3 K4me3K9Ac	ART-Kme3-QTAR-**KAc**-STGGKA
KDM3B	H3 K9me2	ARTKQTAR-**Kme2**-STGGKA
	H3 K4me3K9Ac	ART-Kme3-QTAR-**KAc**-STGGKA
JMJD1C	H3 K4me3K9Ac	ART-Kme3-QTAR-**KAc**-STGGKA

**Reaction with 21 mer peptide instead of the standard 15 mer length.*

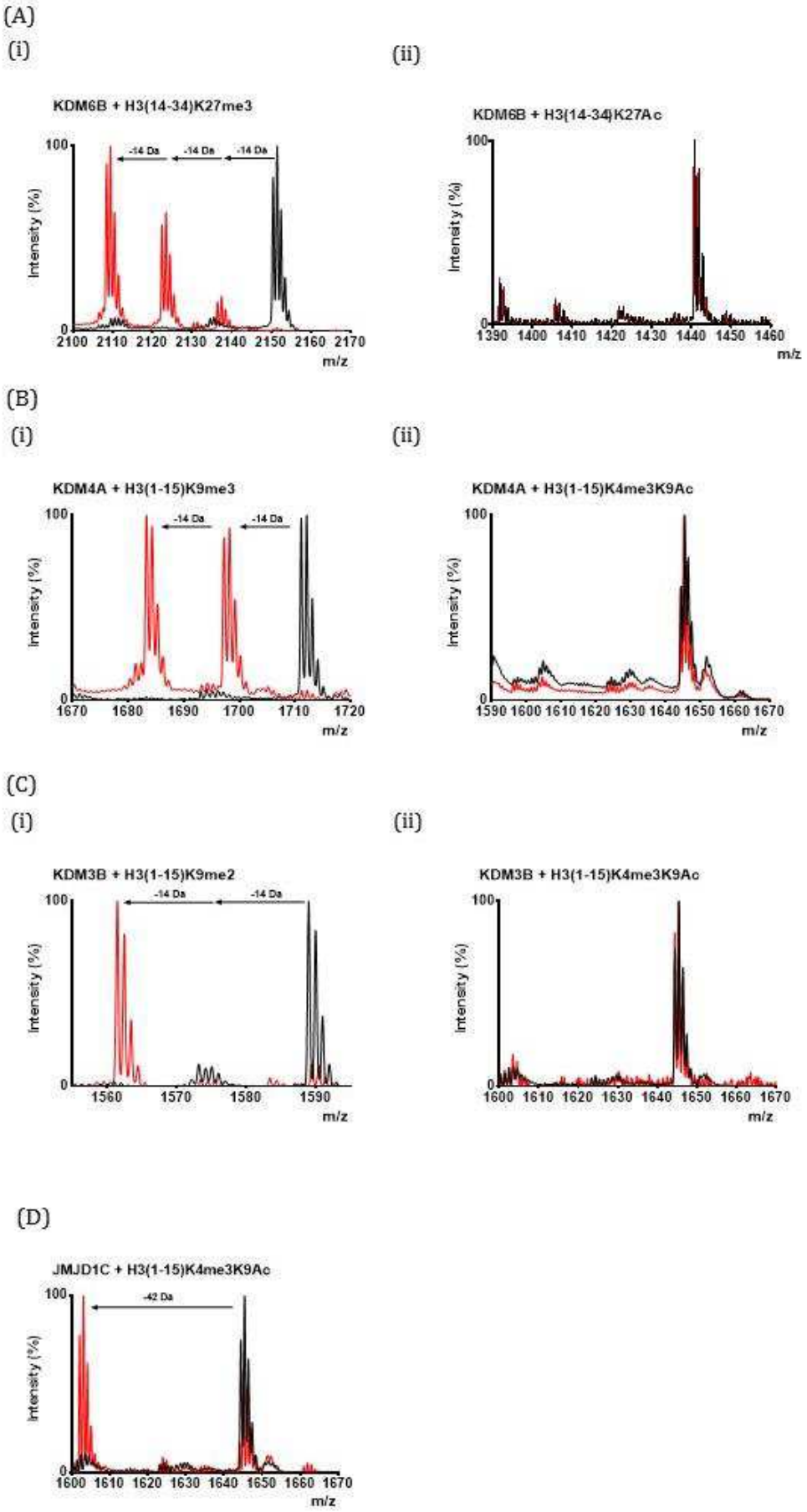

(A)
(i)
KDM6B + H3(14-34)K27me3
-14 Da
-14 Da
-14 Da
Intensity (%)
100
0
2100 2110 2120 2130 2140 2150 2160 2170
m/z
(ii)
KDM6B + H3(14-34)K27Ac
Intensity (%)
100
0
1390 1400 1410 1420 1430 1440 1450 1460
m/z
(B)
(i)
KDM4A + H3(1-15)K9me3
-14 Da
-14 Da
Intensity (%)
100
0
1670 1680 1690 1700 1710 1720
m/z
(ii)
KDM4A + H3(1-15)K4me3K9Ac
Intensity (%)
100
0
1590 1600 1610 1620 1630 1640 1650 1660 1670
m/z
(C)
(i)
KDM3B + H3(1-15)K9me2
-14 Da
-14 Da
Intensity (%)
100
0
1560 1570 1580 1590
m/z
(ii)
KDM3B + H3(1-15)K4me3K9Ac
Intensity (%)
100
0
1600 1610 1620 1630 1640 1650 1660 1670
m/z
(D)
JMJD1C + H3(1-15)K4me3K9Ac
-42 Da
Intensity (%)
100
0
1600 1610 1620 1630 1640 1650 1660 1670
m/z

Figure 7.15 MADLI-TOF MS spectra showing screening data for *N*ε-acetyllysine reactivity with a panel of recombinant KDMs. (A) KDM6B, (B) KDM4A, (C) KDM3B and (D) JMJD1C. For each protein (i) shows a positive control reaction with methylated substrate peptide and (ii) shows the activity of the protein for a peptide where the lysine methylated substrate residue was substituted by acetyllysine. Peptides (10 μM) were incubated with KDMs (2 μM) for 2 h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Reactions with enzyme are shown in red, control reactions with no enzyme are shown in black. *KDM6B was provided by Kate Dunne. KDM4A was provided by Rebecca Hancock. KDM3B and JMJD1C were provided by Dr. Catrine Johansson. H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson.*

7.13 Discussion

The work described in this chapter provides the first evidence of JmjC KDMs catalysed hydroxylation of an acetyllysine residue, specifically on the acetyl methyl group. The discovery was made possible by a *bone fide* approach using knowledge gained by studies with unnatural lysine analogues, which showed various JmjC KDM catalysed reactions with uncharged residues.

The lysine analogue studies revealed that KDM3A can also react with an *O*-methyllysine analogue (peptide **GL2**), and therefore is the first human KDM to be found to react with an ether substrate. To date *O*-demethylation catalysis by 2OG dependant oxygenases has only been detected by the opium poppy enzymes thebaine 6-*O*-demethylase (T6ODM) and codeine *O*-demethylase (CODM) which are involved in morphine biosynthesis.[13] As part of morphine biosynthesis, these enzyme catalyse demethylation of the intermediate thebaine: demethylation on one site by T6ODM produces codeinone whereas demethylation at a second site by CODM forms oripavine (**Figure 7.16**).

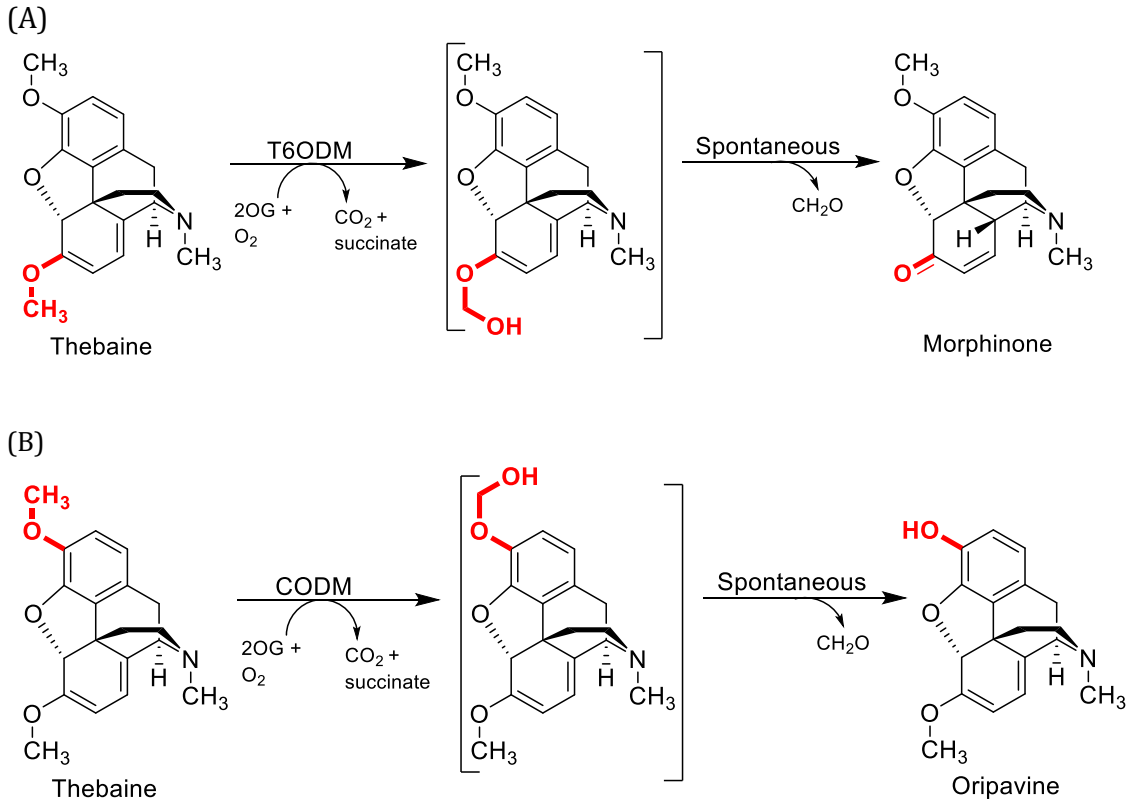


Figure 7.16 Steps in the morphine biosynthesis in opium poppy, showing two routes for *O*-demethylation of theabaine. (A) Thebaine 6-*O*-demethylase (T6ODM). (B) Codeine *O*-demethylase (CODM).[13]

KDM3A also catalysed hydroxylation reaction with a tri-alkyl lysine analogue (peptide **GL3**), in a similar manner to that of the 5-methylcytosine hydroxylations catalysed by the TET enzymes (**Figure 7.17**).[23] Both the tri-alkyl lysine analogue and the *O*-methyllysine analogue are uncharged, revealing the ability of KDM3A to recognise neutral substrates. Reaction with a tri-alkyl lysine analogue is of particular interest as KDM3A can only accept di- and mono- methylated lysine residues as demethylation substrates and is unable to catalyse reaction with tri-methyllysine. Since tri-methyllysine and the tri-alkyl lysine analogue have a similar size, it is likely that there is another reason for the lack of reactivity of KDM3A towards tri-methyllysines. It is possible that a positive charge on a KDM3A substrate interferes with KDM3A catalysis rather than being required for activity as detected with other KDMs.[10] A tri-methyllysine substrate might, for example, be blocked from access to the KDM3A substrate binding site due to the presence of a positively charged residue at the entrance of the KDM3A substrate binding site. Since dimethyllysines and mono-methyllysines can be less protonated or deprotonated, they

could still access the site in this scenario. If KDM3A catalytic activity is indeed dependent on substrate charge it might be regulated by pH changes in cells with physiological implications. When acid-base homeostasis in cells is disrupted, KDM3A enzymatic activity might be modified, and as a result the expression of KDM3A target genes might alter. At a low pH, such as in an acidosis state, more substrates would be protonated and KDM3A catalytic activity might be reduced. At a higher pH state, such as in alkalosis conditions, KDM3A might even demethylate a tri-methyllysine.[24–26] Therefore, KDM3A could in effect act as a pH sensor in cells. The potential of pH regulation of KDM3A catalytic activity requires further investigation and experimental evidence. At a chemical level, the difference between the requirements for substrate protonation between KDM3A and other KDMs, such as the KDM4 subfamily, might be utilised in the development of selective KDM3A inhibitors.[10]

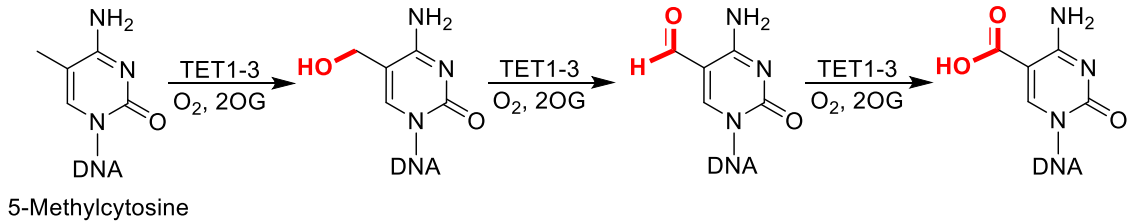
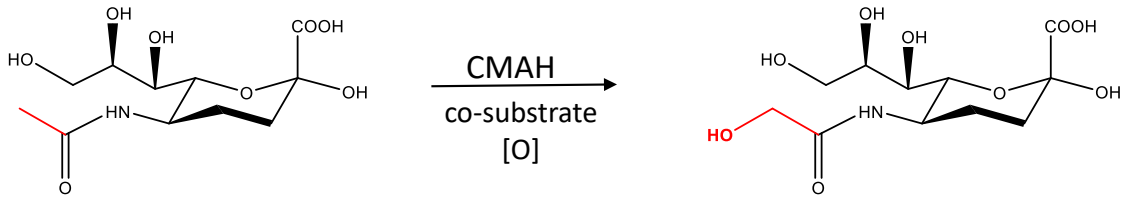


Figure 7.17 Catalytic pathway for oxidation of 5-Methylcytosine (5mC) as catalysed by the TET1-3 enzymes.[23]

Overall, the studies with lysine analogue peptides demonstrate the versatility of reactions that can be catalysed by KDM3A and its ability to catalyse reactions with uncharged substrates. This understanding led to reactivity tests of KDM3A with acetyllysines, which are uncharged, that revealed a hydroxylation reaction.

The only enzyme reported to date to hydroxylate an acetyl group is cytidine monophospho-*N*-acetylneuraminic acid hydroxylase (CMAH), which is a metallo-β-lactamase fold enzyme.[27] CMAH is present in most, if not all, mammals except humans, and catalyses the hydroxylation of sialic acid found at the termini of sugar chains on the surface of plasma membrane proteins (**Figure 7.18**).[28] The mechanism by which CMAH catalyses hydroxylation of an acetyl methyl group is currently unknown as there is no structure available for this enzyme.



N-acetylneuramic acid
(Neu5Ac)

N-glycolylneuramic acid
(Neu5Gc)

Figure 7.18 Cytidine monophospho-*N*-acetylneuraminic acid hydroxylase (CMAH) catalysed reaction. Substrate atoms directly involved in enzyme catalysis are shown in red. [27]

Like KDM3A, CMAH is likely to be an oxygenase. CMAH is reported to be stimulated by NADH, cytochrome b5, and NADH-cytochrome b5 reductase.[29,30] KDM3A hydroxylation on the other hand, appears to be dependent on 2OG and Fe^{II}, suggesting that the mechanism is probably similar to other JmjC catalysed reactions (see Introduction Section 7.1).[7]

The hydroxylation reaction appears to be highly dependent on the combinatorial effect of the PTMs of the H3 K4me3K9Ac sequence. Hydroxylation of K9Ac was only observed in the presence of H3 K4me3. No reaction was detected on the K4me3 control peptide, and experiments with deuterated acetyl confirm that the hydroxylation occur on the K9 Ac. Therefore, the H3 K4me3 position seems to be required for KAc substrate recognition demonstrating the need in further understanding the combinatorial readout of multiple PTMs. Interestingly, H3 K4me3 appear to also induce KDM3A demethylation efficiency although methylation at this site was not a requirement for activity (see Chapter 3, Section 3.5.2.5).

Other JmjC KDMs are reported to have modified catalytic activity in the presence of H3 K4me3, for example KDM7B and KIAA1718 (JHDM1D). KDM7B and KIAA1718 bind H3 K4me3 via their plant homeodomain (PHD). However, the presence of this PTM on the substrate peptides induces the catalytic activity for KDM7B but diminishes the activity for KIAA1718 due to a different effect on enzyme conformation.[31] Another example is the KDM4 subfamily members, which are reported to have enhanced demethylation activity *in vitro* on the H3 K9 position in the presence of trimethylated H3 K4.[32]

Although KDM3A does not have a PHD finger, the protein might harbour a different methyl-reader domain that can interact with H3 K4me3, which could have similar characteristics to that of the Cys4-His-Cys3 zinc binding motif found in PHD fingers.[33]

Domains prediction using NCBI Conserved Domain Database (CDD) and ExPASy Prosite yielded the JmjC and zinc finger domains, however no similarity to other domains was identified.[12,34–36]

Kinetic characterisation indicates that the hydroxylation of *N*ε-acetyllysine peptides by KDM3A is catalysed with poor efficiency compared to demethylation of lysine and arginine methylated peptides in the tested conditions. However, this newly discovered reaction *in vitro* may still have biological implications in cells. The combination of the two marks, K4me3 and K9Ac on histone 3 is enriched in promoter regions.[5] This sequence was found to be common in human, with an average co-occurrence of 35% for fourteen analysed cell lines. Therefore, it is plausible that hydroxylation of K9Ac would lead to biological outcomes influencing gene activation.

Since hydroxylation is a modification of the well-studied lysine acetylation, hydroxylation of acetyllysines could modulate the binding affinity of bromodomains (BRDs), histone acetyltransferases (HATs) and histone deacetylases (HDACs). Although KDM3A catalysis was not detected on BRD4 ligand positions, displacement experiments have demonstrated that hydroxylation of acetyllysines could interfere with the binding affinity of BRD4 to its histone tail ligand positions.

There are a number of enzymes reported to have preference for the H3 K9Ac position, such as the Spt7 and CECR2 BRDs, Gcn5, CBP and p300 which are HATs with a bromodomain, and also several HDACs, such as SIRT6.[37–41] HDACs are classified in four classes depending on domain organisation and sequence homology. These classes can be divided to two groups according to their catalytic mechanism. The majority of the identified HDACs belong to classes I/II/IV which use Zn^{II} and water to directly hydrolyse an *N*ε-acetyllysine residue.[42] Class III HDACs, however, uses NAD^+ as a co-factor, which is converted to 2'-O-Acetyl-ADP Ribose and nicotineamide during deacetylation of the *N*ε-acetyllysine substrate (see Introduction, Chapter 1).[43]

Both groups of HDACs are likely to interact with hydroxy-acetyllysine. Sir5, a human HDAC from class III was crystallised with a peptide of histone 3 succinyl lysine 9 and act as desuccinylase (PDB: 3RIY).[44] Since hydroxy-acetyllysine is a much smaller group than succinyl lysine it seems probable that Sir5 would bind hydroxy-acetyllysine. Studies with *N*ε-acetyllysine analogues showed that Sir2 and Hst2 class III HDACs and also HDAC8, a class I HDAC, can deacetylate hydroxy-acetyllysine peptides.[45,46] These studies have observed opposite substrate preference for the two groups of HDACs: in these tests, class III HDACs, hydroxy-acetyllysine peptide had weaker binding (four-fold) and slower reaction rate (three-fold) compared to *N*ε-acetyllysine peptide. On the other hand, HDAC8

from class I seemed to have higher affinity and faster turnover for the hydroxylated *N*ε-acetyllysine vs. the *N*ε-acetyllysine substrate. Whether one group of HDACs can catalyse hydroxy-acetyllysine residue more efficiently than the *N*ε-acetyllysine substrate still requires further investigation, however overall it seems reasonable that hydroxylation on acetyllysines could be recognised by both groups of HDACs.

KDM3A was catalytically active only with the H3 K4me3K9Ac mark. No catalysis was detected with any other tested H3 and H4 peptides bearing reported *N*ε-acetyllysine positions. No reaction was also detected for non-histone acetyllysines on peptides derived from IFT when these were incubated with KDM3A. It should be noted however that the tested peptides had only the acetylation mark, it is possible that an additional mark including a methyl group could improve substrate recognition by KDM3A, as was the case for H3 K9Ac hydroxylation catalysis. For example, a peptide fragment of H3 K9me3K14Ac might be better recognised by KDM3A than a peptide fragment of H3 K14Ac. Similarly, a peptide fragment of H3 K18me3K23Ac might be a substrate for KDM3A although the corresponding H3 K23Ac is not. An extended screen for potential acetyllysine substrates is currently being carried out.

Recent studies discovered a range of lysine acylation PTMs.[2] Since KDM3A recognises acetyllysines it is likely for KDM3A to react with additional acyl-PTMs as described in Chapter 9. To date, KDM3A is the only KDM that was observed to hydroxylate acetyllysines. However, tests with JMJD1C indicate on a possible deacetylation reaction which is elaborated in Chapter 8.

One important objective of a future study would be to test for the presence of this novel PTM on histones and link it with KDM3A function in a cellular context. Mass spectrometric analyses on histones extracted from KDM3A modulated cells (knock-out, wildtype, catalytically inactive variant, overexpressed) are currently conducted in order to confirm the presence of hydroxy-acetyllysine in cells and to investigate whether this PTM is KDM3A regulated. Antibodies against hydroxy-acetyllysine should be generated and validated. These antibodies can then be used to enrich the histone samples if the new PTM levels are below the detection limit. Cellular immunofluorescence assays can also be developed to further investigate the hydroxylation of acetyllysines. In addition, specific antibodies against hydroxy-acetyllysine could be utilised in ChIP-seq experiments to decipher the localisation of this PTM and thus get an insight into its biological function.

7.14 References

1. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403,** 41–5 (2000).

2. Huang, H., Lin, S., Garcia, B. A. & Zhao, Y. Quantitative proteomic analysis of histone modifications. *Chem. Rev.* **115,** 2376–418 (2015).

3. Dai, L. *et al.* Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nat. Chem. Biol.* **10,** 365–70 (2014).

4. Azevedo, C. & Saiardi, A. Why always lysine? The ongoing tale of one of the most modified amino acids. *Adv. Biol. Regul.* **60,** 144–50 (2016).

5. Karmodiya, K., Krebs, A. R., Oulad-Abdelghani, M., Kimura, H. & Tora, L. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* **13,** 424 (2012).

6. Clifton, I. J. *et al.* Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J. Inorg. Biochem.* **100,** 644–69 (2006).

7. Hewitson, K. S., Granatino, N., Welford, R. W. D., McDonough, M. A. & Schofield, C. J. Oxidation by 2-oxoglutarate oxygenases: non-haem iron systems in catalysis and signalling. *Philos. Trans. A. Math. Phys. Eng. Sci.* **363,** 807-28-40 (2005).

8. Walport, L. J., Hopkinson, R. J. & Schofield, C. J. Mechanisms of human histone and nucleic acid demethylases. *Curr. Opin. Chem. Biol.* **16,** 525–34 (2012).

9. Hopkinson, R. J. *et al.* Is JmjC oxygenase catalysis limited to demethylation? *Angew. Chem. Int. Ed. Engl.* **52,** 7709–13 (2013).

10. Langley, G. W. *et al.* Analysis of JmjC Demethylase-Catalyzed Demethylation Using Geometrically-Constrained Lysine Analogues. *ACS Chem. Biol.* **11,** 755–62 (2016).

11. Walport, L. J. *et al.* Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat. Commun.* **7,** 11974 (2016).

12. Yamane, K. *et al.* JHDM2A, a JmjC-Containing H3K9 Demethylase, Facilitates Transcription Activation by Androgen Receptor. *Cell* **125,** 483–495 (2006).

13. Hagel, J. M. & Facchini, P. J. Dioxygenases catalyze the O-demethylation steps of morphine biosynthesis in opium poppy. *Nat. Chem. Biol.* **6,** 273–5 (2010).

14. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40,** 897–903 (2008).

15. ENCODE: Encyclopedia of DNA Elements – ENCODE. Available at: https://www.encodeproject.org/. (Accessed: 1st June 2016)

16. EMBL-EBI. Clustal Omega.

17. Jean-François *et al.* Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat. Struct. Mol. Biol.* **14,** 689–95 (2007).

18. AltaBioscience histone peptides. Available at: https://altabioscience.com/custom-synthesis/peptide-synthesis/histone-peptides-and-microarrays/.

19. Jung, M. *et al.* Affinity map of bromodomain protein 4 (BRD4) interactions with the histone H4 tail and the small molecule inhibitor JQ1. *J. Biol. Chem.* **289,** 9304–19 (2014).

20. Sanchez, R. & Zhou, M.-M. The role of human bromodomains in chromatin biology and gene transcription. *Curr. Opin. Drug Discov. Devel.* **12,** 659–65 (2009).

21. Philpott, M. *et al.* Bromodomain-peptide displacement assays for interactome mapping and inhibitor discoveryw. *Mol. BioSyst. Mol. BioSyst* **7,** 2899–2908 (2011).

22. PhosphoSitePlus: a resource for protein phosphorylation and other post-translational modifications. Available at: http://www.phosphosite.org/homeAction.action. (Accessed: 24th May 2016)

23. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502,** 472–9 (2013).

24. Simon, D. B. *et al.* Gitelman's variant of Barter's syndrome, inherited hypokalaemic alkalosis, is caused by mutations in the thiazide-sensitive Na–Cl cotransporter. *Nat. Genet.* **12,** 24–30 (1996).

25. Rickman, T., Garmany, R., Doherty, T., Benson, D. & Okusa, M. D. Hypokalemia, metabolic alkalosis, and hypertension: Cushing's syndrome in a patient with metastatic prostate adenocarcinoma. *Am. J. Kidney Dis.* **37,** 838–846 (2001).

26. Smeets, E. E. J., Pelc, K. & Dan, B. Rett Syndrome. *Mol. Syndromol.* **2,** 113–127 (2012).

27. Pettinati, I., Brem, J., Lee, S. Y., McHugh, P. J. & Schofield, C. J. The Chemical Biology of Human Metallo-β-Lactamase Fold Proteins. *Trends Biochem. Sci.* **41,** 338–355 (2016).

28. Varki, A. Multiple changes in sialic acid biology during human evolution. *Glycoconj. J.* **26,** 231–45 (2009).

29. Suzuki, A. Genetic basis for the lack of N-glycolylneuraminic acid expression in human tissues and its implication to human evolution. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* **82,** 93–103 (2006).

30. Kawano, T. *et al.* Molecular Cloning of Cytidine Monophospho-N-acetylneuraminic Acid Hydroxylase. Regulation of species- and tissue-specific expression of N-glycolylneuraminic acid. *J. Biol. Chem.* **270,** 16458–16463 (1995).

31. Horton, J. R. *et al.* Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat. Struct. Mol. Biol.* **17,** 38–43 (2010).

32. Lohse, B. *et al.* Posttranslational Modifications of the Histone 3 Tail and Their Impact on the Activity of Histone Lysine Demethylases In Vitro. *PLoS One* **8,** e67653 (2013).

33. Schindler, U., Beckmann, H. & Cashmore, A. R. HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. *Plant J.* **4,** 137–150 (1993).

34. Brauchle, M. *et al.* Protein complex interactor analysis and differential activity of KDM3 subfamily members towards H3K9 methylation. *PLoS One* **8,** e60549 (2013).

35. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43,** D222–D226 (2015).

36. Sigrist, C. J. A. *et al.* New and continuing developments at PROSITE. *Nucleic Acids Res.* **41,** D344–D347 (2013).

37. Hassan, A. H. *et al.* Selective recognition of acetylated histones by bromodomains in transcriptional co-activators. *Biochem. J.* **402,** 125–33 (2007).

38. Filippakopoulos, P. & Knapp, S. The bromodomain interaction module. *FEBS Lett.* **586,** 2692–704 (2012).

39. Cieniewicz, A. M. *et al.* The bromodomain of Gcn5 regulates site specificity of lysine acetylation on histone H3. *Mol. Cell. Proteomics* **13,** 2896–910 (2014).

40. Michishita, E. *et al.* SIRT6 is a histone H3 lysine 9 deacetylase that modulates telomeric chromatin. *Nature* **452,** 492–6 (2008).

41. Henry, R. A., Kuo, Y.-M. & Andrews, A. J. Differences in specificity and selectivity between CBP and p300 acetylation of histone H3 and H3/H4. *Biochemistry* **52,** 5746–59 (2013).

42. Haberland, M., Montgomery, R. L. & Olson, E. N. The many roles of histone deacetylases in development and physiology: implications for disease and therapy : Nature Reviews Genetics. *Nature reviews. Genetics* **10,** 32–42 (2009).

43. Barneda-Zahonero, B. & Parra, M. Histone deacetylases and cancer. *Mol. Oncol.* **6,** 579–89 (2012).

44. Du, J. *et al.* Sirt5 is a NAD-dependent protein lysine demalonylase and desuccinylase. *Science* **334,** 806–9 (2011).

45. Smith, B. C. & Denu, J. M. Acetyl-lysine analog peptides as mechanistic probes of protein deacetylases. *J. Biol. Chem.* **282,** 37256–65 (2007).

46. Smith, B. C. & Denu, J. M. Sir2 deacetylases exhibit nucleophilic participation of acetyl-lysine in NAD+ cleavage. *J. Am. Chem. Soc.* **129,** 5802–3 (2007).

Chapter 8

Investigations into a putative histone deacetylation catalysis by JMJD1C

8.1 Introduction

The demethylation activity of JMJD1C (also known as KDM3C) is controversial.[1–3] Similarly to other KDM3 members, JMJD1C has a CH2C4 zinc finger domain followed by a Leu-Xaa-Xaa-Leu-Leu (LXXLL) motif, and a JmjC domain in its sequentially *C*-terminal region. However, JMJD1C is almost double in length compared to KDM3A, and is also larger than KDM3B and Hairless (2540 compared to 1321, 1761 and 1189 aa, respectively in humans, see Introduction, Chapter 1). The JmjC domain of JMJD1C shares only about 50% sequence similarity with the other KDM3 members, this compares with about 80% similarity between the validated histone JMJD1 demethylases KDM3A and KDM3B (see Introduction, Chapter 1).

Although the carboxylate facial triad residues are conserved, a number of key residues found in KDM3A and KDM3B are different in JMJD1C. These include a threonine in its zinc finger domain (Thr667), which has been observed to be important for H3 K9me1 demethylation by KDM3A.[3] In JMJD1C, this threonine is replaced by an alanine.[3] An alanine to threonine mutation, however, did not restore JMJD1C H3 K9me1/me2 demethylase activity *in vitro*.[3] In its JmjC domain, JMJD1C is missing a conserved lysine residue, involved in 2OG binding in KDM3A and KDM3B (Lys1259 in KDM3A and Lys1699 in KDM3B), which is replaced by a glutamine (Gln2476). The change in charge, from a positive to a neutral residue, may result in a weaker interaction of JMJD1C with 2OG. Another conserved residue important for 2OG binding in KDM3A/B is a tyrosine (Tyr1101 in KDM3A and Tyr1541 in KDM3B), which in JMJD1C is replaced by cysteine (Cys2317) (see Chapter 7, Section 7.6). These sequence differences likely indicate different co-substrate interactions for JMJD1C.

Interestingly, recent structural studies of the JmjC domain in JMJD1C reveal oxidation of Cys2317 to a sulfinic acid following incubation with certain TCA cycle intermediates (*Dr.*

Catrine Johansson, manuscript in preparation). Thus, it is possible that JMJD1C utilises other co-factors and/or substrate then KDM3A/B. In addition, a putative deacetylase activity was detected *in vitro* using recombinant JMJD1C and an acetylated histone peptide (see Chapter 7, Section 7.12).

8.2 Objectives

The work described in this chapter aimed to gain preliminary data as to whether JMJD1C is responsible for catalysing apparent deacetylation of N^{ε}-acetyllysine of a histone H3 fragment, an activity which was observed while screening KDMs with acetylated histone peptides. The work aimed to probe whether the apparent JMJD1C-catalysed deacetylation reaction is indeed catalysed by JMJD1C itself, or by a co-purified catalyst. The work included experiments that tested the dependence of the putative deacetylation reaction on H3 K4 methylation and the known JmjC 2OG co-substrate and Fe^{II} co-factor. In addition, tests with different batches and different constructs of JMJD1C, including available JMJD1C JmjC domain variants, were carried out. The efficiency of the catalysis was tested in the presence of various TCA cycle intermediates as putative co-substrates. The study was then expanded to investigate the sequence selectivity of the reaction, the possibility of other histone sequences to be recognised as substrates for deacetylation.

All JMJD1C proteins, including JMJD1C variants, were produced by Dr. Catrine Johansson. Non-histone artificial N-terminal acetyl and H3(1-15)K4me3K9Ac peptides were synthesised by Dr. Richard Hopkinson.

8.3 Testing for histone demethylation catalysis by JMJD1C

Considering the conflicting data reported in the literature regarding the histone demethylation activity by JMJD1C, its catalytic activity was tested *in vitro* using the H3(1-15)K9me2 peptide.[1–3]

Recombinant JMJD1C$_{1760\text{-}2540}$ protein, which includes the JmjC domain and the zinc finger, was expressed in baculovirus, and was purified using Ni-Affinity chromatography followed by gel filtration (See **Appendix 2** for analysis of JMJD1C protein purification, *protein was produced by Dr. Catrine Johansson*).

JMJD1C (2 μM) was incubated with H3(1-15)K9me2 peptide (10 μM) under standard demethylation conditions for 2 hours at 37°C. Analyses were carried out by MALDI-TOF MS, and enzymatic reactions were compared to the non-enzyme containing negative controls (see Materials and methods, Chapter 11, Section 11.3.3).

No reaction was observed by MS with the dimethylated H3 fragment peptides (**Figure 8.1**), demonstrating that, under the conditions tested, JMJD1C could not demethylate this histone H3 peptide.

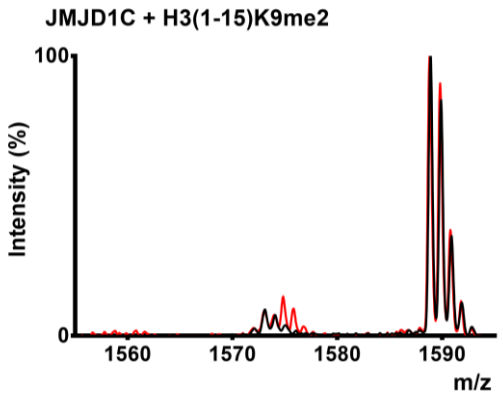


Figure 8.1 MALDI-TOF MS spectra showing a lack of demethylation catalysis by JMJD1C with H3(1-15)K9me2 peptide. JMJD1C (2 μM) was incubated with peptide (10 μM) for 2 h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Reactions containing enzyme are shown in red, no enzyme control reactions are shown in black. *JMJD1C protein was produced by Dr. Catrine Johansson.*

8.4 Detection of a putative deacetylation activity by JMJD1C

KDM3A preparations were observed to hydroxylate the *N*ε-acetyllysine at position K9 of the H3 histone fragment peptide H3(1-15)K4me3K9Ac (see Chapter 7, Section 7.4.2). Therefore, JMJD1C was tested with this peptide to explore if other JmjC proteins had similar activity. Assay conditions were kept as described in Section 8.3. Data were analysed by MALDI-TOF MS.

Interestingly, a two hour incubation at 37°C with JMJD1C resulted in a mass shift of -42 Da, with a 65% peptide conversion to this product (**Figure 8.2**). This mass shift corresponds to a putative deacetylation, resulting in the H3(1-15)K4me3 peptide. A very small peak corresponding to a +16 Da shift might also indicate on a hydroxylated product, however the intensity of this peak is similar to the noise level. The assay conditions and additional results are described in detail in Chapter 7, Section 7.12.

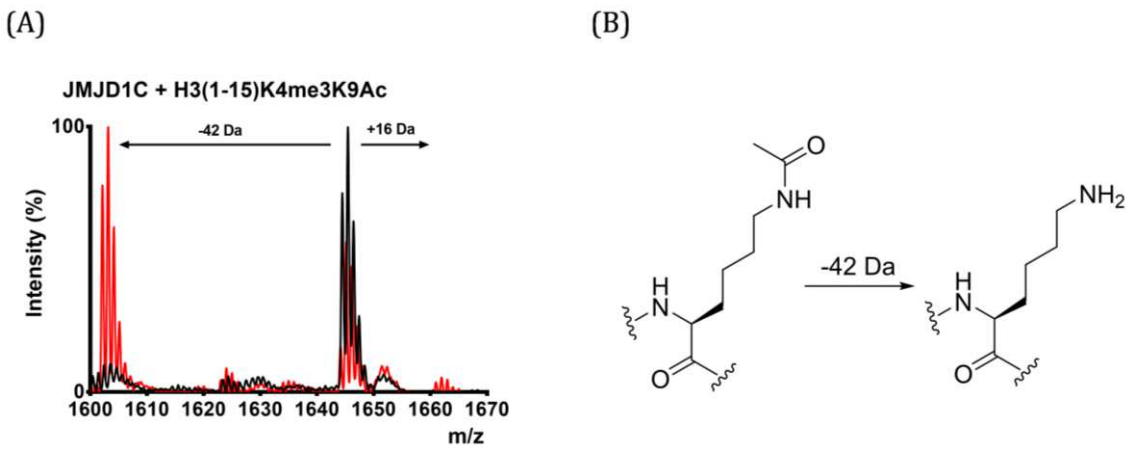


Figure 8.2 Detection of a putative deacetylation activity by JMJD1C. Activity assays with JMJD1C (2 µM) on H3(1-15)K4me3K9Ac peptide (10 µM) performed over 2 h at 37°C. Assays were carried out under standard conditions (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) MALDI-TOF MS spectrum. Reaction containing enzyme is shown in red, no enzyme control reaction is shown in black. (B) Proposed deacetylation reaction. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson. JMJD1C protein was produced by Dr. Catrine Johansson.*

8.5 Time course measurements

Time course measurements were carried out to further characterise the supposed deacetylation reaction, and were carried out under the standard conditions described in Section 8.3.

These results show that the amount of the observed deacetylated H3(1-15)K4me3 product peptide in the presence of a JMJD1C (2 μM) enzyme mixture increases in a time dependent manner (**Figure 8.3**). The levels of observed deacetylated product are doubled during the second hour (from 18% after 1 h to 38% by 2h), and then doubled again by eight hours (76%). These results suggest that under the tested conditions the reaction rate is relatively slow as it does not reach saturation after two hours of incubation. For comparison, KDM3A lysine demethylation catalysis under these conditions reached saturation within minutes (see Chapter 3, Section 3.5.1.1). KDM3A hydroxylation of acetyllysines under these conditions was slower than demethylation, with saturation by one hour (see Chapter 7, Section 7.7.1). Therefore, it is likely that the conditions used in these tests are not optimal for the putative deacetylation catalysis. In addition, the observed reactivity after two hours implies that the catalytic source for this reaction is stable under these conditions.

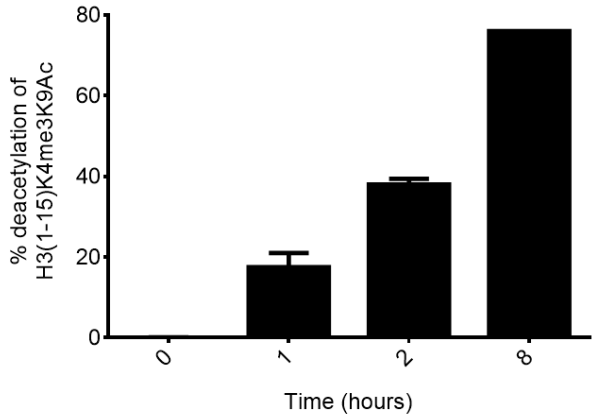


Figure 8.3 Time course measurements of the putative deacetylation reaction. Measurement of the percentage observed deacetylation as apparent by a -42 Da shift in the mass of the H3(1-15)K4me3K9Ac peptide (10 μM) following incubation with JMJD1C (2 μM) under standard conditions at 37°C. Percentage of product formation was calculated as the ratio between product intensity peak and the sum of the product and substrate intensity peaks as observed by MALDI-TOF MS spectra. Data plotted as the mean of duplicate with standard deviation. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson. JMJD1C protein was produced by Dr. Catrine Johansson.*

8.6 Investigating the dependency of the putative deacetylation reaction on the H3 K4me3 mark

The previous chapter described work showing that the KDM3A catalysed hydroxylation of the *N*ε-acetyllysine at position K9 of the histone fragment peptide H3(1-15)K4me3K9Ac appears to be promoted by H3 K4 methylation (see Chapter 7, Section 7.4.2). It was therefore of interest to ascertain whether the putative JMJD1C deacetylation activity is also affected by H3 K4 methylation.

JMJD1C (2 μM) was incubated with either H3(1-15)K4me3K9Ac (10 μM), or the corresponding H3(1-15)K9Ac (10 μM) peptide, under the standard reaction conditions for one hour at 37°C. Analyses were carried out by MALDI-TOF MS, and enzymatic reactions were compared to the non-enzyme containing negative controls. Comparison of the percentage deacetylation observed for both peptides suggests that the apparent catalytic efficiency is similar whether the H3 K4 position is methylated or not (**Figure 8.4**).

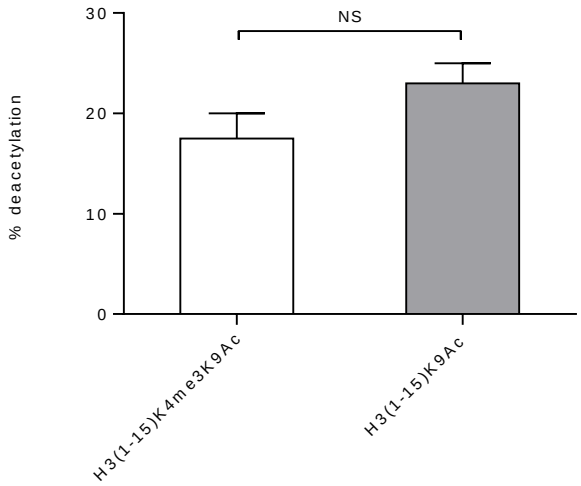


Figure 8.4 Investigating the dependency of the putative deacetylation catalysis on the H3 K4me3 mark. JMJD1C (2 μM) was incubated with H3(1-15)K4me3K9Ac and H3(1-15)K9Ac peptides (10 μM) for 1 h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Comparison of percentage observed deacetylation was calculated as the ratio between product intensity peak and the sum of the product and substrate intensity peaks, as observed by MALDI-TOF MS spectra. Data plotted as the mean of duplicate with standard deviation. Statistical analysis was carried out using a two-tailed t-test using GraphPad Prism 6, NS denotes $P > 0.05$. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson. JMJD1C protein was produced by Dr. Catrine Johansson.*

8.7 Validation of *in vitro* putative deacetylation by JMJD1C

Experiments described in Section 8.4 identified a novel putative deacetylation reaction catalysed by recombinant JMJD1C *in vitro.* Subsequent experiments then aimed to probe whether the reaction is catalysed by JMJD1C itself, or by a co-purifying catalyst.

All experiments were carried out under the standard conditions as described in Section 8.3, unless otherwise stated. Analyses were conducted by MALDI-TOF MS, the observed shift in peptide mass was compared to non-enzyme containing negative controls.

8.7.1 Testing for co-factors dependence

As JMJD1C is a JmjC protein, its catalytic activity might be expected to be dependent on Fe^{II} as a co-factor and 2OG as a co-substrate (for additional information see Introduction, Chapter 1). To investigate whether the observed putative deacetylation reaction is dependent on 2OG and Fe^{II}, experiments measuring activity with, without, and with a different co-factor and co-substrate were carried out. Other assay conditions were kept as described in Section 8.3.

35% of the H3(1-15)K4m3K9ac peptide was observed to deacetylate after two hours of incubation under the standard conditions (**Figure 8.5 A**). Similar amount of peptide (31%) was converted to the "deacetylated" form following incubation with a reaction mixture lacking 2OG, indicating that 2OG is not essential for the putative deacetylation catalysis (**Figure 8.5 B**). 46% of peptide was "deacetylated" when incubated with a reaction mixture lacking Fe^{II}, indicating that additional Fe^{II} is not essential for the catalysis (**Figure 8.5 C**). Fe^{II} might even cause a mild inhibitory effect, as more peptide was converted in its absence, however, this observation should be further investigated. Pre-incubation of the enzyme with EDTA was carried out to chelate metal ions that could have been co-purified with the enzyme, such as Fe^{II}, Mn^{II} and Zn^{II}. A small decrease in the amount of the converted peptide from 35% to 27% was observed following pre-incubation with EDTA (**Figure 8.5 D**). More experiments are required to define whether the responsible enzyme is metal-dependent. Most of the apparent deacetylation was detected following incubation with Mn^{II}, indicating that catalysis might be manganese dependent, although more experiments are required to confirm this observation (48%, **Figure 8.5 E**). Interestingly, replacement of Fe^{II} with Zn^{II}

completely abolished reaction (**Figure 8.5 F**). Zinc might act as an 'allosteric' inhibitor through coordination to the zinc binding site (see Discussion Section 8.10).

The data suggest that the observed apparent deacetylation reaction is not dependent on 2OG, or Fe^{II}, and is inhibited by Zn^{II}. The decrease in activity followed by pre-incubation with EDTA indicates that the responsible enzyme might be metal-dependent, and the increase in activity in the presence of Mn^{II} suggests that this metal is a candidate co-factor.

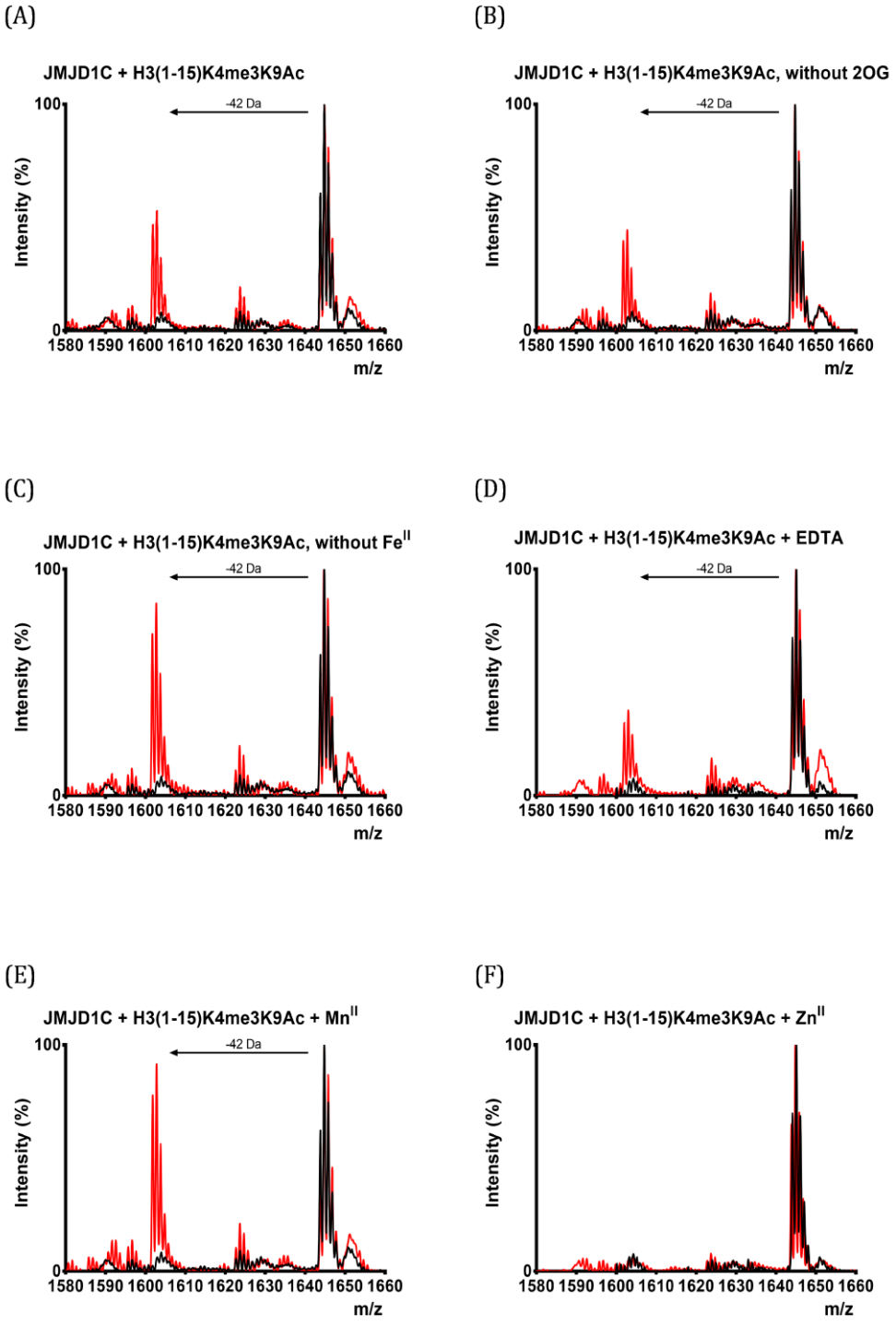


Figure 8.5 Testing for co-factors dependence. MALDI-TOF MS data of activity assays with JMJD1C (2 μM) on H3(1-15)K4me3K9Ac peptide (10 μM) performed over 2 h at 37°C. Assays were carried out under standard conditions, with 2OG (100 μM) and Fe^{II} (50 μM) except for the specified changes (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) Reaction under the standard conditions. (B) Reaction mixture where 2OG is absent. (C) Reaction mixture where Fe^{II} is absent. (D) Pre-incubation of JMJD1C with 3mM EDTA, followed by addition of reaction mixture where Fe^{II} is absent. (E) Replacement of Fe^{II} with Mn^{II} (50 μM). (F) Replacement of Fe^{II} with Zn^{II} (50 μM). Reactions containing enzyme are shown in red, no enzyme control reactions are shown in black. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson. JMJD1C protein was produced by Dr. Catrine Johansson.*

8.7.2 Probing putative deacetylation with different protein productions and constructs of JMJD1C

Experiments with different batches of recombinant JMJD1C were carried out to investigate whether co-purifying catalyst or JMJD1C are responsible for the observed "deacetylation" reaction.

First, reaction was conducted with a different batch of JMJD1C, expressed using the same construct ($JMJD1C_{1760-2540}$) and purification conditions to test whether the putative deacetylation activity is not specific for the first batch used. Both protein batches exhibited "deacetylation" activity as apparent by a -42 Da shift in peptide mass, albeit a slightly reduced activity for the second batch compared to the first one (**Figure 8.6 A – B**). It is known that the activities of many JmjC demethylases vary between various purifications and therefore this observation may not be surprising (*Dr. Catrine Johansson, manuscript in preparation*). The reaction was also conducted with a truncated version of $JMJD1C_{2157-2497,}$ encompassing only the JmjC domain (see Materials and methods, Chapter 11, Section 11.2.2, for construct information). Using this construct, no mass change was observed, indicating on a lack of activity (**Figure 8.6 C**).

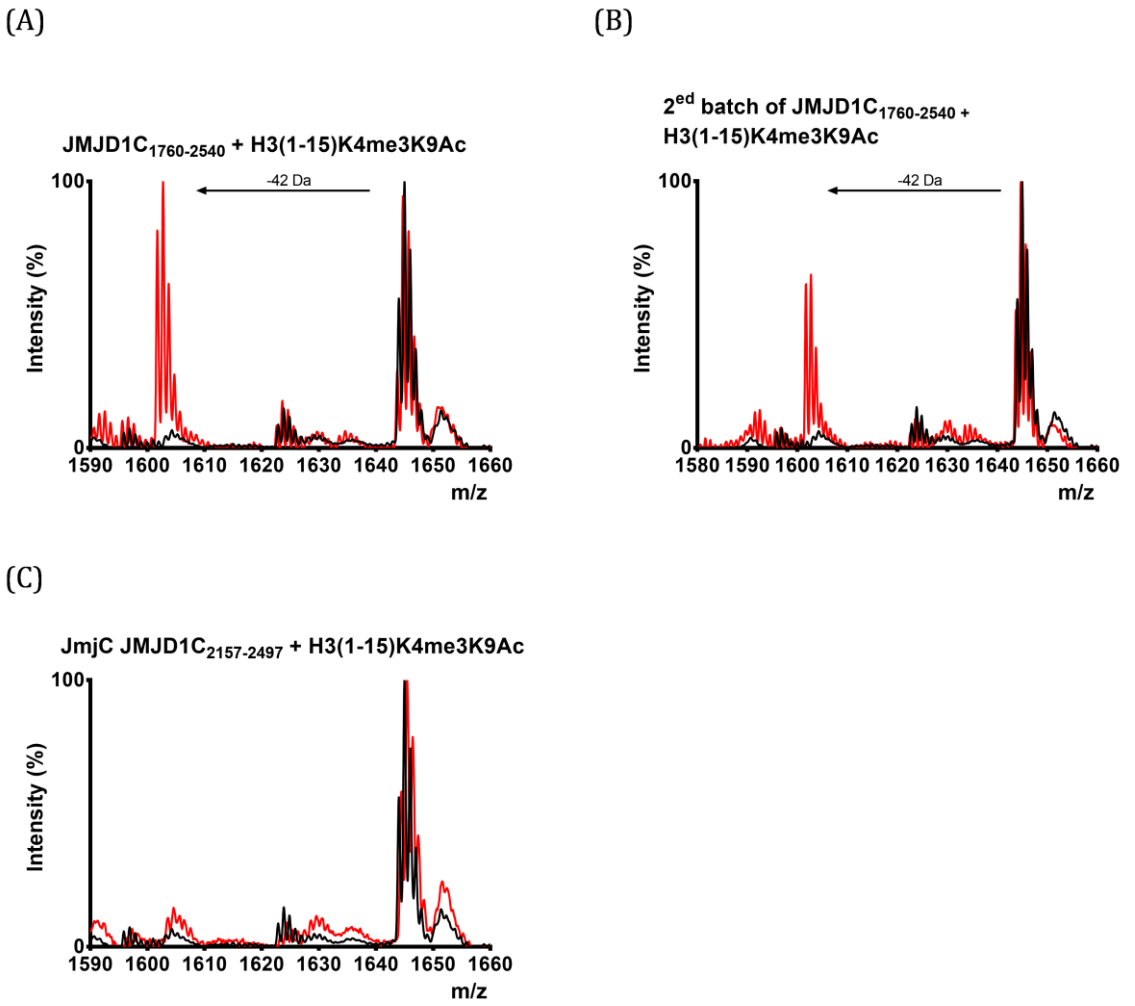


Figure 8.6 Probing putative deacetylation with different protein productions and constructs of JMJD1C. MALDI-TOF MS data of activity assays with JMJD1C (2 μM) on H3(1-15)K4me3K9Ac peptide (10 μM) performed over 2h at 37°C. Assays were carried out under standard conditions (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) Reaction using baculovirus produced His-GST-JMJD1C$_{1760-2540}$ (B) Reaction using a second batch of baculovirus produced His-GST-JMJD1C$_{1760-2540}$ than that used for previous tests. (C) Reaction using the *E. coli* produced JmjC-domain of a His-JMJD1C$_{2157-2497}$ construct. Reactions containing enzyme are shown in red, no enzyme control reactions are shown in black. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson. JMJD1C proteins were produced by Dr. Catrine Johansson.*

To conclude, these experiments showed that the putative deacetylation activity is observed with different batches of recombinant JMJD1C, albeit activity seems to change between batches which might be due to different purities. It is tempting to propose that the absence of a deacetylation activity when incubating the peptide with the JmjC domain construct is due to dependency of JMJD1C catalysis on sequences beyond the JmjC domain such as the zinc finger. However, another possible explanation for the lack of activity could be the

presence of different co-eluting proteins in the different expression systems (*E. coli* for JmjC JMJD1C$_{2157-2497}$ vs. insect cells for JMJD1C$_{1760-2540}$). It should be noted that KDM3A and KDM3B were also produced via expression in insect cells, and no deacetylase activity is detected for these enzymes (see Chapter 7, Sections 7.4.2 and 7.12). Another difference between the proteins is their purification methods due to their different tags, which has the potential to affect impurities (His-tag for JmjC JMJD1C$_{2157-2497}$ vs. His-GST-tag JMJD1C$_{1760-2540}$).

8.7.3 Examination of the putative deacetylation reaction using mutation study

The JmjC domain of JMJD1C has a cysteine at position 2317, instead of a conserved tyrosine.[4] X-ray crystallographic studies on KDM3B suggested that this conserved tyrosine might facilitate interaction with the co-substrate 2OG. Thus, a cysteine replacement might explain the lack of 2OG requirement for JMJD1C (see Chapter 7, Section 7.6.4 for sequence alignment and structural comparison). Interestingly, the thiol of this cysteine was observed to oxidise to a sulfinic acid under certain conditions (*Catrine Johansson, manuscript in preparation*, see Section 8.8 for related experiments).

To investigate whether the putative deacetylation reaction is catalysed by JMJD1C, rather than by a catalyst that had co-purified with it, experiments with a C2317Y variant of JMJD1C were performed. The C2317Y JMJD1C variant was purified in a similar manner as the wildtype JMJD1C, thus it is likely that it was co-purified with similar impurities to the wildtype.

Activity assays were carried out as described previously in Section 8.3. The resultant data revealed that mutating Cys2317 to tyrosine in JMJD1C reduced the putative deacetylation activity by about half (from 38% conversion to 23%), but that it was still present (**Figure 8.7**). These results strengthen the hypothesis that the detected deacetylation is catalysed by JMJD1C rather than by another co-purifying enzyme. These results also suggest that C2317 may be important for a putative deacetylation but that is not the only residue required for catalysis, because substituting it did not abolish activity. It should be noted that C2317Y variant does not confer demethylation activity for JMJD1C, demonstrating that the conserved tyrosine in KDM3A/B is also not the only requirement for demethylation catalysis *(data not shown, Catrine Johansson, manuscript in preparation*).

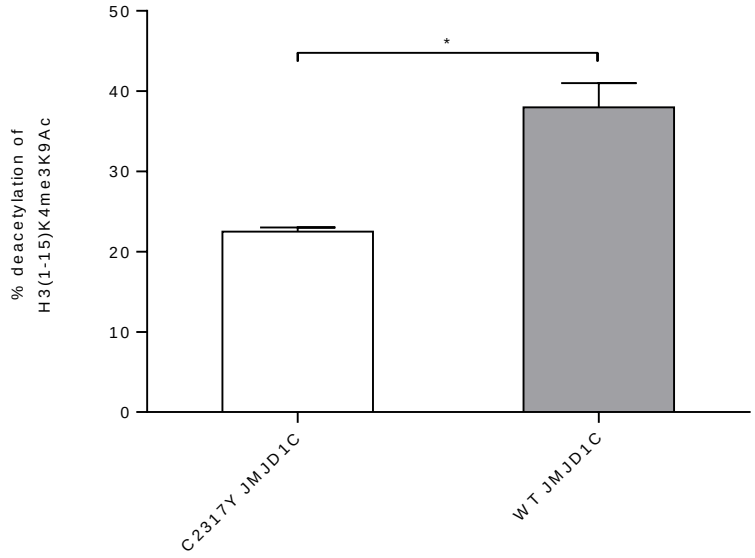


Figure 8.7 Examination of the putative deacetylation reaction using C2317Y variant. Comparison of percentage "deacetylation" of an H3(1-15)K4me3K9Ac peptide after incubation of either WT JMJD1C (2 μM) or the corresponding C2317Y JMJD1C variant (2 μM) under standard conditions, with 2 hours incubation at 37°C. Percentage of product formation was calculated as the ratio between product intensity peak and the sum of the product and substrate intensity peaks as observed by MALDI-TOF MS spectra. Data plotted as the mean of duplicate with standard deviation. Statistical analysis was carried out using a two-tailed t-test using GraphPad Prism 6, * denotes $P \leq 0.05$. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson. JMJD1C proteins were produced by Dr. Catrine Johansson.*

8.8 Investigation of other putative co-substrates

The work described in Section 8.7.3 suggests the importance of Cys2317 in the active site of JMJD1C for catalysis of its putative deacetylation reaction. In addition to these findings, recent studies have revealed that D-α-hydroxyglutaric acid (D-2HG) and citrate, but not 2OG, promotes oxidation of Cys2317 to sulfinic acid (*Dr. Catrine Johansson, manuscript in preparation*). Having confirmed that 2OG is not required for the apparent deacetylation reaction, it was envisaged to investigate whether either of the mentioned TCA cycle intermediates can stimulate catalysis of the putative deacetylation.

JMJD1C (2 μM) was incubated with either of 2OG (100 μM) or D-2HG or citrate (100 and 500 μM) and an H3(1-15)K4me3K9Ac (10 μM) peptide for 2 hours at 37°C. D-2HG did not cause an increase in the apparent deacetylation rate compared to 2OG (**Figure 8.8**). However, incubation of the reaction mixture with either 100 or 500 μM citrate seemed to stimulate

reaction by about 25% compared to incubation with 100 μM 2OG (from 38% to 47% "deacetylation"), although the increase was not statistically significant and more repeats are required to determine that effect .

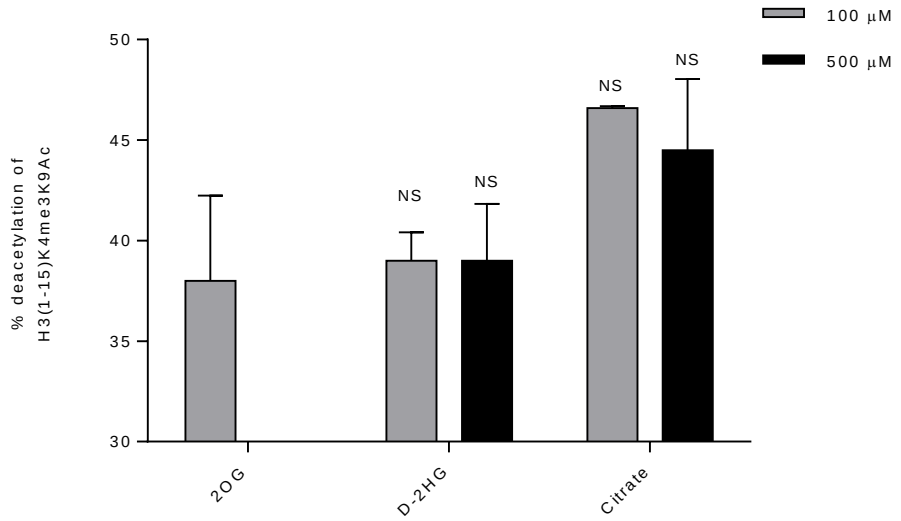


Figure 8.8 Investigation of other potential co-substrate dependence. Comparison of percentage "deacetylation" of an H3(1-15)K4me3K9Ac peptide after incubation of JMJD1C (2 μM) with either 2OG (100 μM) or D-2HG or citrate (100 & 500 μM) for 2 hours at 37°C. Percentage of product formation was calculated as the ratio between product intensity peak and the sum of the product and substrate intensity peaks as observed by MALDI-TOF MS spectra. Data plotted as the mean of duplicate with standard deviation. Statistical analyses were carried out using a two-tailed t-test compared with 2OG control using GraphPad Prism 6. NS denotes $P > 0.05$. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson. JMJD1C protein was produced by Dr. Catrine Johansson.*

8.9 Screening for possible deacetylation on other histone fragment peptides

Having confirmed that the H3 K9Ac mark is an *in vitro* substrate for a putative deacetylation using the JMJD1C preparations, as apparent by a -42 Da shift in peptide mass, it was reasonable to propose that deacetylation might also be detected with additionally reported acetylated positions on histone tails.

To investigate deacetylation on other histone positions, peptides for ten peptides with reported acetylated lysine positions on histones H3 and H4 were tested (**Table 8.1**). Histone H3 fragment peptides were designed based on acetylated peptide sequences in the

AltaBioscience peptide array.[5] Histone H4 derived peptides are reported binders for the BRD4 bromodomains.[6] As a control for selectivity, an available non-histone artificial *N*-terminal acetyl peptide was added to the screen. Screening assays were conducted under standard conditions (see Section 8.3).

The *N*-terminal acetyl peptide was not deacetylated, although this could be due to position effect and a better negative control would be a peptide with KAc in the centre (**Figure 8.9 A**). However, all the tested acetylated histone fragment peptides were observed to deacetylate to varying degrees when incubated with JMJD1C, the sequences containing two acetyllysines were "deacetylated" twice (**Figure 8.9 B**). Therefore, these results appear to suggest that although deacetylation would not occur on any *N*-terminal acetyl peptide, the putative JMJD1C catalysed histone lysine deacetylation is not sequence selective.

Table 8.1 Acetylated peptide sequences and the mark around which they were designed. Peptides were used as potential substrates for JMJD1C. Target acetyl-residues are shown in bold. *Non-histone artificial N-terminal acetyl peptide was synthesised by Dr. Richard Hopkinson.*

Histone mark	Peptide sequence
H3 K14Ac	ARKSTGG**-KAc-**APRKQLA
H3 K18Ac	TGGKAPR**-KAc-**QLATKAA
H3 K23Ac	PRKQLAT**-KAc-**AARKSAP
H3 K27Ac	LATKAAR**-KAc-**SAPATGG
H3 K115Ac	NLCAIHA**-KAc-**RVTIMPK
H4 K5Ac	SGRG**-KAc-**GGKGLGKGGA
H4 K5,8Ac$_2$	SGRG**-KAc-**GG**-KAc-**GLGKGGA
H4 K12Ac	KGGKGLG**-KAc-**GGAKRHR
H4 K16Ac	GLGKGGA**-KAc-**RHRKVLR
H4 K12,16Ac$_2$	GKGLG**-KAc-**GGA**-KAc-**RHRKV
Artificial	**Ac**-AAAARAAAA

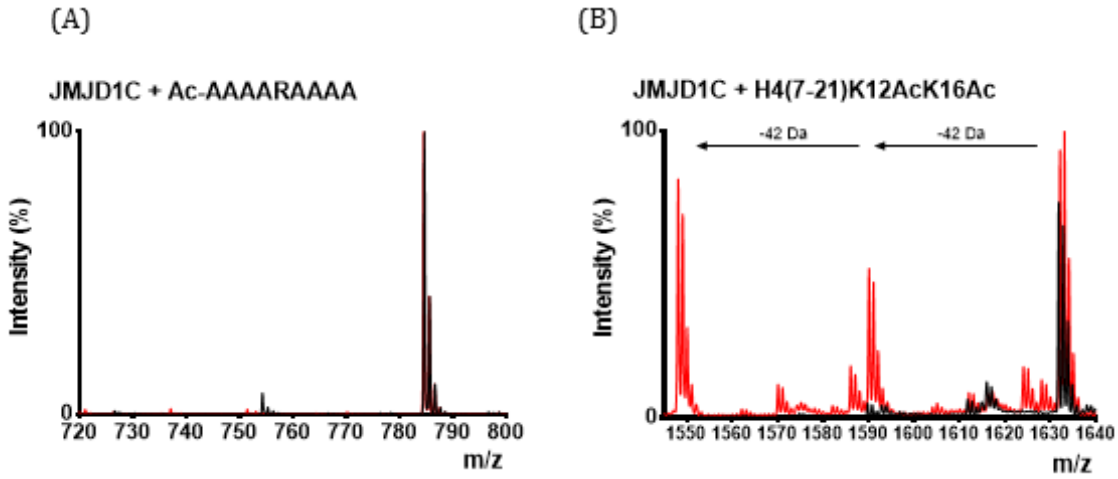


Figure 8.9 Example of MALDI-TOF MS spectra of acetylated peptides in the screen. JMJD1C (2 μM) was incubated with peptides for 2 hours at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) Incubation with a non-histone artificial *N*-terminal acetyl peptide (10 μM). (B) Incubation with an H4(1-15)K12AcK16Ac peptide (10 μM). Reactions containing enzyme are shown in red, no enzyme control reactions are shown in black. *Non-histone artificial N-terminal acetyl peptide was synthesised by Dr. Richard Hopkinson. JMJD1C protein was produced by Dr. Catrine Johansson.*

8.10 Discussion

The work in this chapter has documented the preliminary initial characterisation of histone peptide putative deacetylation reactions in samples containing JMJD1C. Initial work focused on testing for JMJD1C histone demethylation activity and revealed no demethylation activity under the *in vitro* conditions used. MALDI-TOF MS based studies were then carried out in order to gain preliminary data to identify whether JMJD1C is responsible for the observed apparent deacetylation catalysis.

The apparent deacetylation reaction does not seem to be dependent on 2OG and Fe^{II}, raising the possibility that the reaction might be catalysed by another catalyst that co-elutes with JMJD1C, rather than JMJD1C itself. Another possibility is that JMJD1C catalysis is not dependent on the same co-factors required for demethylation, as can also be inferred by its low sequence similarity to other KDM3 subfamily members. It would be interesting to test the activity in the presence of other metals with different oxidation states. Interestingly, addition of zinc to the reaction mixture completely abolished deacetylation. Zinc inhibition has been previously reported for PHD3, a 2OG dependent prolyl-hydroxylase.[7] This study showed that Zn^{II} does not compete with Fe^{II} for the active site of PHD3, but instead that Zn^{II}

binds to cysteines in an 'allosteric site' that leads to inhibition. For further investigation, it would be beneficial to determine the IC_{50} for JMJD1C inhibition, as well as carrying out kinetic characterisation to identify the type of inhibition by zinc (i.e. whether the inhibition is allosteric or not). The apparent deacetylation activity appeared to increase following incubation with Mn^{II}.

A different protein batch of JMJD1C also exhibited a putative deacetylation activity, showing that the reaction is not an artefact related to a specific batch. On the other hand, an impurity produced in insect cells might have been co-purified with JMJD1C in both batches. Co-purification is possible, albeit unlikely, as purification involves two chromatography columns (Immobilized Metal ion Affinity Chromatography (IMAC) followed by gel filtration). If indeed an enzyme, for instance a HDAC, co-eluted with JMJD1C through both columns, it suggests a strong interaction between the two proteins which would also be intriguing to study.

Interestingly, a construct preparation containing just the JmjC domain of JMJD1C was not active as a deacetylase. This could imply that other regions of JMJD1C, such as the zinc finger domain, are important for deacetylation catalysis, in agreement with the requirement for the Zn-binding domains for the demethylase activity of several KDMs containing these domains including KDM3A, KDM5 and KDM6 enzymes.[3,8,9] Another hypothesis for the absence of activity is the different production procedures between the two constructs, which could result in different co-eluting impurities. It would be worthwhile to express and purify the JmjC domain of JMJD1C under the same conditions as for the longer construct (i.e. expression with a His-GST-tag in a baculovirus system). To retest for Fe^{II} dependence it would also be desirable to produce a JMJD1C construct with mutations on the conserved His-X-Asp/Glu-X_N-His motif, however this is likely to affect the coordination of another metal that JMJD1C might be dependent on. Another useful JMJD1C variant to produce would be Gln2476Lys, involving in 2OG binding (see Introduction Section 1).

Studies using Cys2317Y variant of JMJD1C showed reduced catalytic activity, this appears to suggest that this residue is important for catalysis. On the other hand, this difference could also stem from batch to batch variation of a potentially co-eluting enzyme. Interestingly, Cys2317 of JMJD1C was recently observed to undergo oxidation into a sulfinic acid in the present of D-2HG and citrate (*Dr. Catrine Johansson, manuscript in preparation)*. This could either be a result of a crystallographic artefact, or it could reflect a factor facilitating the catalysis. Under the tested conditions described here, citrate was shown to slightly stimulate the apparent deacetylation catalysis although the increase was not

statistically significant and more repeats are required to determine that effect. Since the time course measurements in the presence of 2OG showed a slow reaction, it seems plausible that another co-substrate is required for efficient catalysis.

The work was expanded to investigate the sequence selectivity of the reaction. Unlike the acetyllysine hydroxylation catalysed by KDM3A, the apparent deacetylation activity was not dependent on H3 K4 methylation. Moreover, putative deacetylation was observed following incubation of JMJD1C with the entire tested set of H3 and H4 fragment peptides. However, no deacetylation of an *N*-terminal acetyl peptide was detected. The apparent lack of deacetylation selectivity is not uncommon for HDAC catalysis. Compared to histone demethylases, little is known about the selectivity of histone deacetylases. Reports show that different HDAC subtypes differ in their substrate selectivity, and that some HDACs exhibit a very low specificity.[10–13]

Assuming that the observed putative deacetylation described in this work is indeed due to JMJD1C mediated catalysis, there could be several theoretical mechanisms for this catalysis. A small peak in the MALDI-TOF MS spectra corresponding a +16 Da shift in peptide mass might imply that the classical hydroxylation mechanism occurs prior to deacetylation. However, data from co-factors control experiments show that an addition of 2OG is not required for deacetylation. Therefore, the conserved 2OG dependent hydroxylation mechanism might not be employed here by JMJD1C. In this scenario, JMJD1C may catalyse deacetylation by simple hydrolysis via binding and activation of the carbonyl to make it more susceptible to attack by a water molecule. Such mechanism does not require a co-substrate such as 2OG or co-factor such as Fe^{II}. On the other hand, this putative mechanism might imply that an insect-cell hydrolase co-eluted with JMJD1C could also catalyse this deacetylation reaction, as no unique JmjC related properties are involved. Evidence of oxidation of the active site cysteine to sulfinic acid facilitated by citrate, and a possible increase in "deacetylation" activity in the presence of citrate and manganese should also be taken into account when trying to propose a putative JMJD1C-catalysed deacetylation mechanism.

All in all, this work has revealed a possible histone deacetylation catalysis by JMJD1C. Together with KDM3A catalysis of acetyl hydroxylations, and catalysis of another acyl modifications by both enzymes (see Chapters 7 and 9), it seems that possible members of the KDM3 subfamily might regulate both methylation and acylation. However, the available preliminary *in vitro* results gained in this study are not sufficient to either eliminate or validate the hypothesis that JMJD1C is the catalyst responsible for the observed

deacetylation of histone fragment peptides. Current efforts are focused on testing for deacetylation in cells by immunofluorescence assay, using overexpressed JMJD1C and an anti-acetyllysine antibody. Results from these tests are expected to be more conclusive compared to tests with a purified protein, where impurity issues might hamper analysis. Depending on these results, a comprehensive biochemical and kinetic characterisation will be followed to complete the data from the presented preliminary observations.

8.11 References

1. Kim, S.-M. *et al.* Regulation of mouse steroidogenesis by WHISTLE and JMJD1C through histone methylation balance. *Nucleic Acids Res.* **38,** 6389–403 (2010).

2. Watanabe, S. *et al.* JMJD1C demethylates MDC1 to regulate the RNF8 and BRCA1-mediated chromatin response to DNA breaks. *Nat. Struct. Mol. Biol.* **20,** 1425–33 (2013).

3. Brauchle, M. *et al.* Protein complex interactor analysis and differential activity of KDM3 subfamily members towards H3K9 methylation. *PLoS One* **8,** e60549 (2013).

4. Clifton, I. J. *et al.* Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J. Inorg. Biochem.* **100,** 644–69 (2006).

5. AltaBioscience histone peptides. Available at: https://altabioscience.com/custom-synthesis/peptide-synthesis/histone-peptides-and-microarrays/.

6. Jung, M. *et al.* Affinity map of bromodomain protein 4 (BRD4) interactions with the histone H4 tail and the small molecule inhibitor JQ1. *J. Biol. Chem.* **289,** 9304–19 (2014).

7. Na, Y.-R. *et al.* Selective inhibition of the hypoxia-inducible factor prolyl hydroxylase PHD3 by Zn(ii). *Chem. Commun.* **51,** 10730–10733 (2015).

8. Horton, J. R. *et al.* Characterization of a Linked Jumonji Domain of the KDM5/JARID1 Family of Histone H3 Lysine 4 Demethylases. *J. Biol. Chem.* **291,** 2631–46 (2016).

9. Sengoku, T. & Yokoyama, S. Structural basis for histone H3 Lys 27 demethylation by UTX/KDM6A. *Genes Dev.* **25,** 2266–77 (2011).

10. Riester, D., Hildmann, C., Grünewald, S., Beckers, T. & Schwienhorst, A. Factors affecting the substrate specificity of histone deacetylases. *Biochem. Biophys. Res. Commun.* **357,** 439–45 (2007).

11. Blander, G. *et al.* SIRT1 shows no substrate specificity in vitro. *J. Biol. Chem.* **280,** 9780–5 (2005).

12. Garske, A. L. & Denu, J. M. SIRT1 top 40 hits: use of one-bead, one-compound acetyl-peptide libraries and quantum dots to probe deacetylase specificity. *Biochemistry* **45,** 94–101 (2006).

13. Cosgrove, M. S. *et al.* The structural basis of sirtuin substrate affinity. *Biochemistry* **45,** 7511–21 (2006).

Chapter 9
Detection of a putative lysine deformylation by KDM3A and JMJD1C

9.1 Introduction

Lysine acetylation is the most studied histone modification of the acyl-type PTMs. Over the past decade, a substantial number of new histone lysine acylation PTMs that have been discovered, including lysine formylation (**Figure 9.1**; for more information see Introduction, Chapter 1, Section 1.4).[1–4]

(A) (B)

Figure 9.1 Lysine acylation PTMs. (A) The structure of a formyllysine residue. (B) The structure of an acetyllysine residue.

Apart from formylation of the methionine α-amino group for translation initiation, formylation of lysine is a PTM uniquely associated with histones and other nuclear proteins.[4–6] Formyllysines were first identified on histone H1 by mass spectrometry.[6] Formylation is proposed to occur as a non-enzymatic secondary modification in cells during oxidative DNA damage, which may explain why formylation is only detected in nuclear proteins.[4] An alternative proposal is that lysine formylation occurs enzymatically by a transformylase which uses formyl-tetrahydrofolate (THF), the formyl source of formyl

methionyl-tRNA, to catalyse formylation. However, such an enzyme has not yet been identified.[7] It is also important to note that silver staining protocols that use formaldehyde can result in ε-formylation of lysine residues, and therefore use of this technique should be avoided when identifying formylation sites.[8]

Whether lysine formylation on histones is catalysed enzymatically or is a secondary side effect reaction, formylation competes with other lysine modifications such as acetylation and methylation that could have been present at that position.[9] Thus, formylation is likely to have epigenetic implications. Enzymes that catalyse the removal of a formyl group from a lysine residue were only reported recently; the bacterial SrtN and the mammalian SIRT1, both sirtuin-type HDACs, were identified as deformylases (see Introduction, Chapter 1, **Scheme 1.4**).[10]

9.2 Objectives

The work described in this chapter investigated the potential of a catalytic promiscuity of KDM3A and JMJD1C with respect to deformylation of *N*ε-formyllysines. Following observations of a putative deformylation reaction on histone peptides, the dependence of catalysis on H3 K4 methylation status for each enzyme was investigated. The study then focused on KDM3A-mediated putative deformylation, testing for dependence on the JmjC co-factors and for the possibility of other histone formyllysine sequences to be recognised as substrates. Further work explored whether other KDMs react with formyllysines *in vitro*.

JMJD1C and KDM3B were produced by Dr. Catrine Johansson. KDM4A was provided by Rebecca Hancock. H3(1-15)K4me3K9for and H3(1-15)K4me3K9Ac peptides were synthesised by Dr. Richard Hopkinson.

9.3 Identification of a possible lysine deformylation by KDM3A and JMJD1C

Previous chapters have described the possible *in vitro* hydroxylation and deacetylation of the histone fragment peptide H3(1-15)K4me3K9Ac by KDM3A and JMJD1C, respectively (see Chapter 7, Section 7.4.2 and Chapter 8, Section 8.4). JMJD1C was also observed to catalyse deacetylation of a number of additional acetylated histone fragment peptides (Chapter 8). Although previous work by Hopkinson *et al.* has demonstrated that the JmjC KDMs KDM4E, KDM7B (PHF8) and KDM2A (FBXL11) are unable to deformylate histone peptides containing N^{ε}-formyllysine, it was proposed to test whether KDM3A and JMJD1C might remove this modification.[11]

Experiments were conducted with an H3(1-15)K4me3K9for histone peptide (10 μM) and either KDM3A or JMJD1C (2 μM) under the standard conditions (for more information see Materials and methods, Chapter 11, Section 11.3.3). MALDI-TOF mass spectra were compared with the negative controls in the absence of the enzyme.

Interestingly, an almost complete conversion to a -28 Da product peak in the peptide mass was observed following two hours of incubation of H3(1-15)K4me3K9for with KDM3A, corresponding to a possible deformylation (84% conversion, **Figure 9.2 A, C**). The same product was also detected following incubation with JMJD1C, although the relative conversion was lower when compared to the KDM3A reaction (24%, **Figure 9.2 B**). These findings provide preliminary evidence that members of the KDM3 subfamily, and JmjC histone demethylases in general, show a putative deformylation activity.

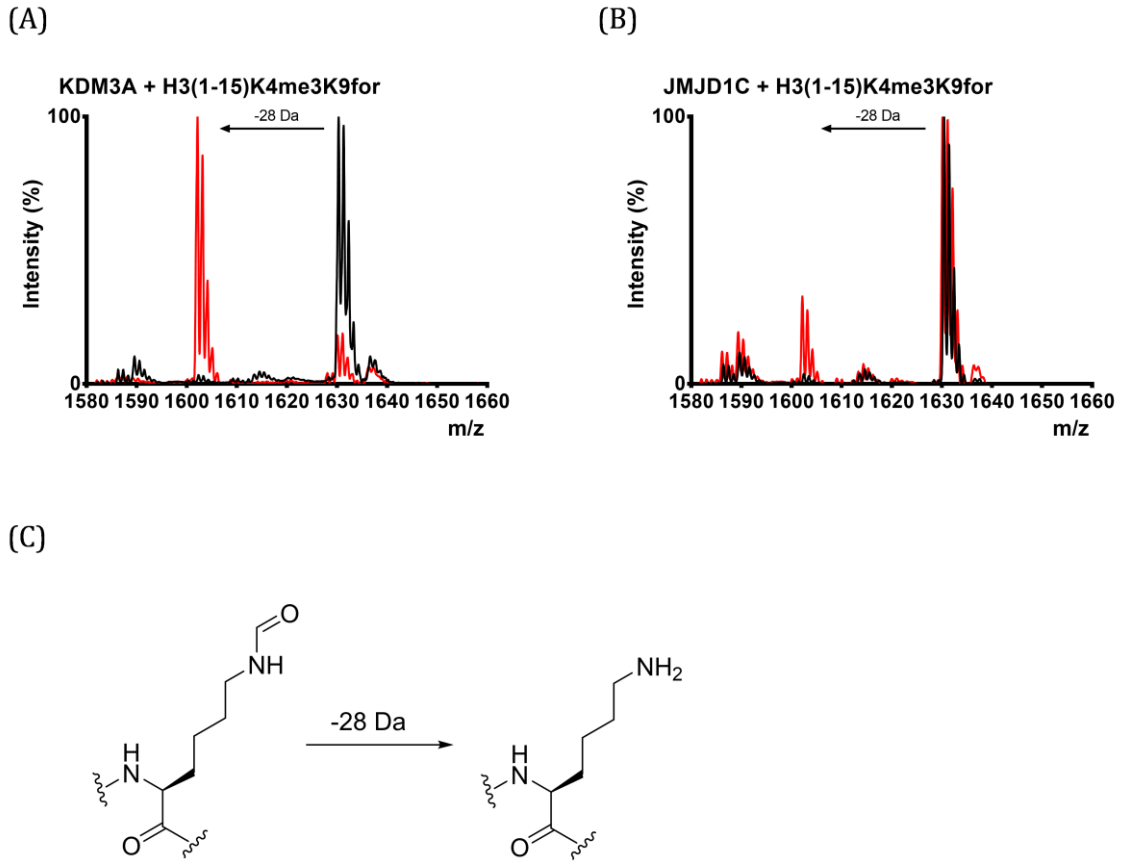


Figure 9.2 Detection of putative deformylation activity for KDM3A and JMJD1C. Activity assays with KDM3A (2 μM) and JMJD1C (2 μM) with an H3(1-15)K4me3K9for peptide (10 μM) were performed over 2 h at 37°C. Assays were carried out under standard conditions (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) MALDI-TOF mass spectrum of H3(1-15)K4me3K9for peptide following incubation with KDM3A. (B) MALDI-TOF mass spectrum of H3(1-15)K4me3K9for peptide following incubation with JMJD1C. Reactions containing enzyme are shown in red, control reactions with no enzyme are shown in black. (C) Proposed deformylation reaction. *H3(1-15)K4me3K9for peptide was synthesised by Dr. Richard Hopkinson. JMJD1C protein was produced by Dr. Catrine Johansson.*

9.4 Investigation of the dependency of the apparent deformylation activity on H3 K4me3

Previous chapters described a dependency of KDM3A hydroxylation of acetyllysines on the methylation status of H3 K4, and a lack of such dependency for JMJD1C apparent deacetylation (see Chapters 7, Section 7.4 and Chapter 8, Section 8.6). Considering these findings, apparent deformylation activities by both enzymes was tested for their dependence on H3 K4 methylation.

KDM3A and JMJD1C were incubated with either H3(1-15)K4me3K9for or H3(1-15)K9for peptide under standard reaction conditions at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Analyses were carried out by MALDI-TOF MS, and enzymatic reactions were compared to negative controls not containing enzyme.

Comparison of the extent of "deformylation" by KDM3A between both substrates suggests that the apparent catalytic efficiency of KDM3A is stimulated when the H3 K4 position is methylated (**Figure 9.3**). These data correlate with the KDM3A dependency on H3 K4 methylation for hydroxylation of H3 K9Ac (see Chapter 7, Section 7.4). The apparent deformylation by JMJD1C was completely abolished in the absence of H3 K4 methylation (**Figure 9.4**). Interestingly, the putative deacetylation activity of JMJD1C was not dependent on the methylation state of H3 K4 (see Chapter 8, Section 8.6). Note that the spectrum presented in **Figure 9.4 B** for reaction containing enzyme is saturated, as the second isotopic peak of 1589 Da which supposed to be lower than the first one appears at 100 % intensity. Therefore all peaks, including nose, have higher intensity compared to the spectrum of the control reaction in this figure

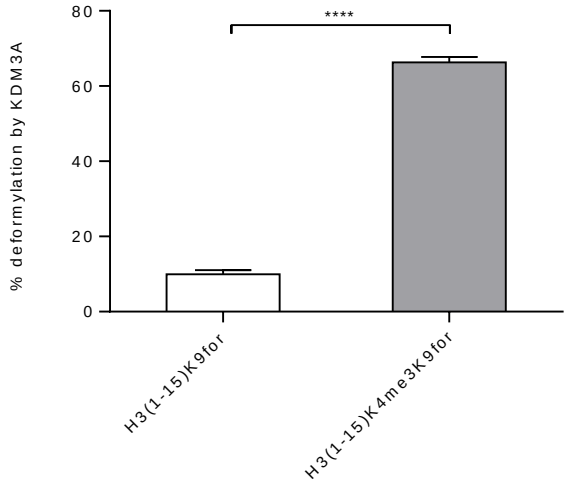


Figure 9.3 Comparison between the putative deformylation activities of KDM3A with H3(1-15)K4me3K9for and H3(1-15)K9for. KDM3A (1 µM) was incubated with the H3 peptides (10 µM) for 1 h at 37°C. Percentage "deformylation" was calculated as the ratio between product intensity peak and the sum of the product and substrate intensity peaks, as determined by MALDI-TOF MS. Data plotted as the mean of triplicate with standard deviation. Statistical analysis was carried out using a two-tailed t-test using GraphPad Prism 6, **** denotes P ≤ 0.0001. *H3(1-15)K4me3K9for peptide was synthesised by Dr. Richard Hopkinson.*

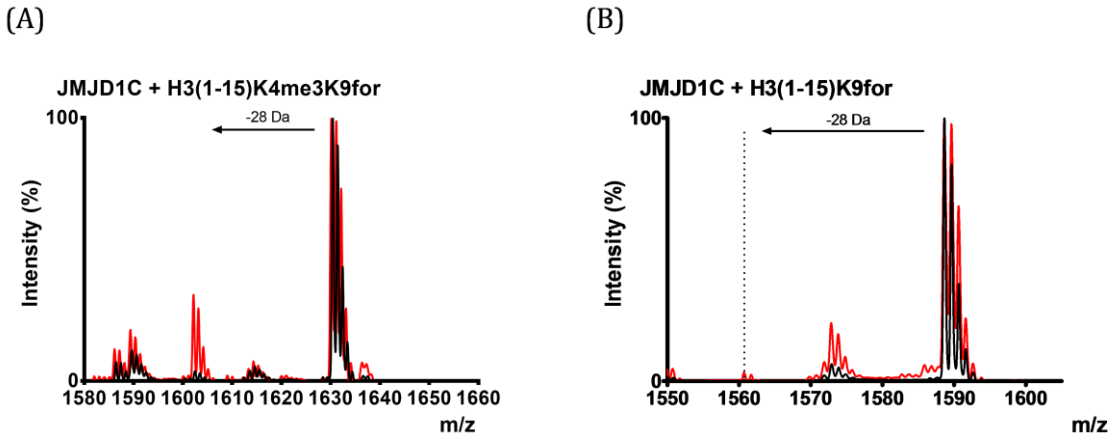
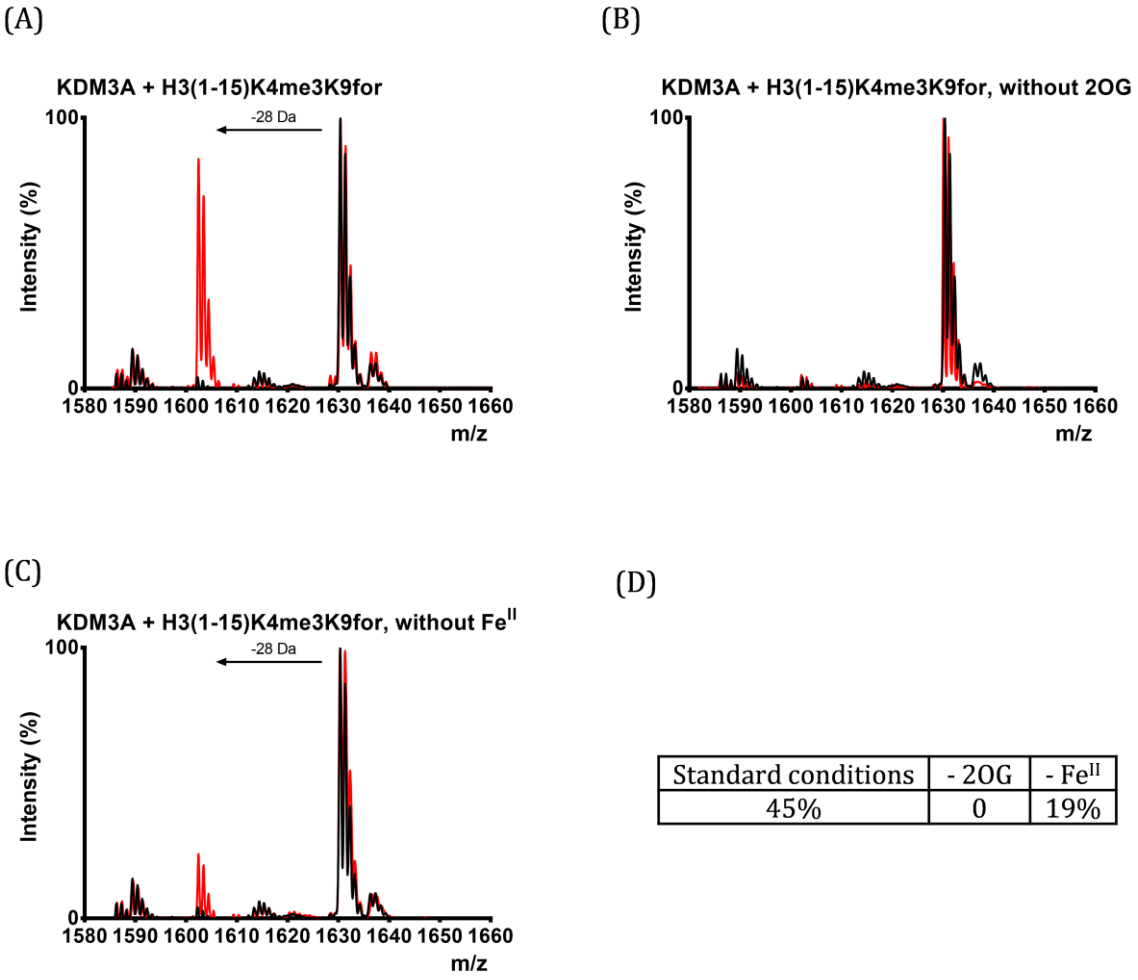


Figure 9.4 JMJD1C dependency of a putative deformylation activity on H3 K4me3 methylation. JMJD1C (2 µM) was incubated with the H3 peptides (10 µM) for 2 h at 37°C. Assays were carried out under standard conditions (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) MALDI-TOF mass spectrum monitoring the incubation of H3(1-15)K4me3K9for peptide with JMJD1C. (B) MALDI-TOF mass spectrum monitoring the incubation of H3(1-15)K9for peptide with JMJD1C. Reactions containing enzyme are shown in red, control reactions with no enzyme are shown in black. *H3(1-15)K4me3K9for peptide was synthesised by Dr. Richard Hopkinson. JMJD1C protein was produced by Dr. Catrine Johansson.*

9.5 Testing the co-factor dependence of KDM3A putative deformylation catalysis

To investigate whether the observed apparent deformylation reaction by KDM3A is dependent on JmjC co-substrate 2OG and co-factor Fe^{II}, experiments measuring activity with and without co-factor and co-substrate were carried out.

KDM3A (1 μM) was incubated with H3(1-15)K4me3K9for (10 μM) under standard reaction conditions for one hour at 37°C, the reaction was analysed by MALDI-TOF MS (see Materials and methods, Chapter 11, Section 11.3.3, for more information). About 45% "deformylation" was observed under standard conditions with 2OG and Fe^{II} (**Figure 9.5 A**). No reaction was observed in the absence of 2OG (**Figure 9.5 B**), strengthening the assumption that the apparent deformylation reaction is JmjC mediated and is based on 2OG-dependent catalysis. Low levels of the "deformylated" product were observed in reaction mixtures that did not contain Fe^{II} (19%, **Figure 9.5 C**). These low levels might suggest that KDM3A co-purified with some iron. This co-purified iron was apparently sufficient for low levels of "deformylation". In contrast, hydroxylation of acetyllysine by KDM3A was completely abolished without Fe^{II}. However, the efficiency of hydroxylation is likely to be much lower than that of the putative deformylation and therefore low levels of activity might not have been detectable (see Chapter 7, Section 7.8).



Standard conditions	- 2OG	- Fe^{II}
45%	0	19%

Figure 9.5 Testing co-factor dependence of KDM3A apparent deformylation catalysis. MALDI-TOF MS data of activity assays of KDM3A (1 μM) with the H3(1-15)K4me3K9for peptide (10 μM) performed over 1 h at 37°C. Assays were carried out under standard conditions, with 2OG (100 μM) and Fe^{II} (50 μM) except for the specified changes (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) Reaction under the standard conditions. (B) Reaction in the absence of 2OG. (C) Reaction in the absence of Fe^{II}. Reactions containing enzyme are shown in red, control reactions with no enzyme are shown in black. (D) A table summarising the resultant data. *H3(1-15)K4me3K9for peptide was synthesised by Dr. Richard Hopkinson.*

9.6 Screening for KDM3A putative deformylation of other formylated histone fragment peptides

The *N*-terminal tails of a number of variant histones are reported to contain formyllysine sites.[12] To investigate whether KDM3A can deformylate these reported positions, histone fragment peptides containing these marks were tested (**Table 9.1**).

Screening assays were carried out under standard reaction conditions and were analysed by MALDI-TOF MS (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Although KDM3A was shown to be enzymatically active (with respect to the positive control experiment with H3(1-15)K4me3K9Ac), no deformylation activity was observed with any of the tested formylated histone fragment peptides (**Figure 9.6**). These results indicate selectivity of KDM3A catalysed reactions for the H3 K9. However, since KDM3A putative deformylation was observed to be induced in the presence of H3 K4 methylation (see Section 9.4), a specific combination of methylation and formylation modifications might be necessary for the deformylation of other histone positions.

Table 9.1 Formylated histone peptide sequences. Target formyllysine residues are shown in bold. Peptides were designed based on reported formylated histone sites found in human and mouse tissues, and were tested as substrates for KDM3A.[12]

Histone mark (*N*-tail)	Peptide sequence
H1.2 K17for	AAAPPAE-**Kfor**-APVKKKA
H2B K5for	MPEPA-**Kfor**-SAPAPKKGS
H3 K23for	APRKQLAT-**Kfor**-AARKSA
H3 K18for	TGGKAPR-**Kfor**-QLATKAA
H4 K12for	GKGGKGLG-**Kfor**-GGAKRH
H4 K31for	DNIQGIT-**Kfor**-PAIRRLA

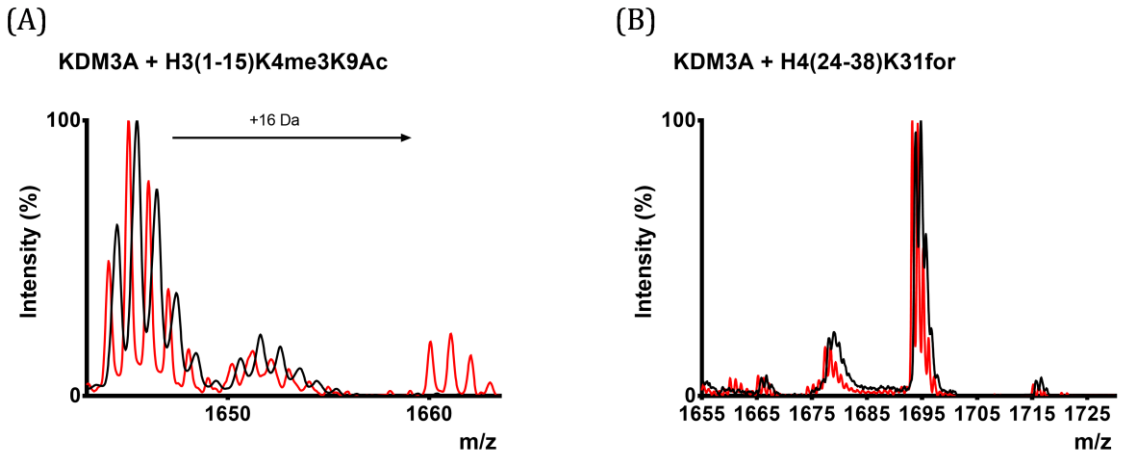


Figure 9.6 Example of MALDI-TOF MS spectra from histone peptide screen. MALDI-TOF MS data of activity assays with KDM3A (2 μM) on H3(1-15)K4me3K9for peptide (10 μM) performed over 2 h at 37°C. Assays were carried out under standard conditions (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) KDM3A hydroxylates the positive control peptide H3(1-15)K4me3K9Ac under the experiment conditions. (B) Absence of KDM3A activity for the H4(24-38)K31for peptide. Reactions with enzyme are shown in red, control reactions with no enzyme are shown in black. *H3(1-15)K4me3K9Ac peptide was synthesised by Dr. Richard Hopkinson.*

9.7 Testing for *N*ε-formyllysine demethylation by other KDMs

Previous work by Hopkinson *et al.* has reported that the JmjC KDMs KDM4E, KDM7B (PHF8) and KDM2A (FBXL11) are unable to catalyse deformylation of histone peptides containing *N*ε-formyllysine.[11]

The work described in this section was designed to screen other KDMs that have selectivity for the H3 K9 position: KDM3B and KDM4A. For each KDM tested, a histone peptide containing the H3 K9me2 mark (known to be recognised by these demethylases) was used as a positive control. The substrate peptide used in these tests was formylated at the H3 K9 site: H3(1-15)K4me3K9for.

Reactions were carried out under the standard conditions with incubation of 2 μM enzyme and 10 μM peptide for two hours at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). Analyses were carried out by MALDI-TOF MS. For KDM3B and KDM4A, -14 Da shifts in the mass of the lysine methylated control peptide were observed, indicating demethylation activity (**Figure 9.7**). However, no reactions were

observed for the analogous formylated peptides when they were incubated with either KDM3B or KDM4A.

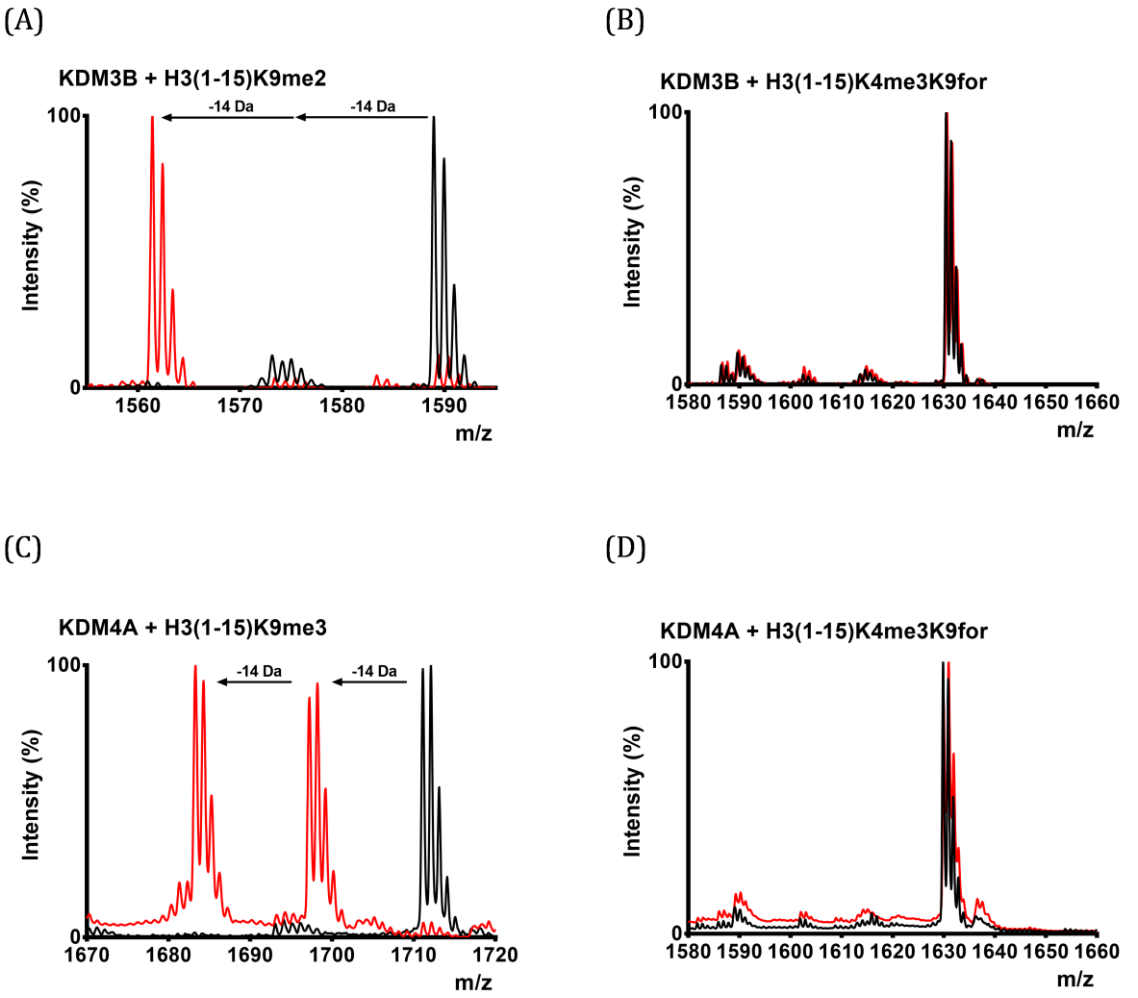


Figure 9.7 MADLI-TOF MS spectra showing data for the impact of recombinant KDMs on *N*ε-formyllysine. Peptides (10 μM) were incubated with KDMs (2 μM) for 2 h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, for more information). (A) A positive control reaction in which KDM3B was incubated with H3(1-15)K9me2. (B) Incubation of KDM3B with an H3(1-15)K4me3K9for peptide. (C) A positive control reaction in which KDM4A was incubated with H3(1-15)K9me2. (D) Incubation of KDM4A with an H3(1-15)K4me3K9for peptide. Reactions with enzyme are shown in red, control reactions with no enzyme are shown in black. *KDM4A was provided by Rebecca Hancock. KDM3B was provided by Dr. Catrine Johansson. H3(1-15)K4me3K9for peptide was synthesised by Dr. Richard Hopkinson.*

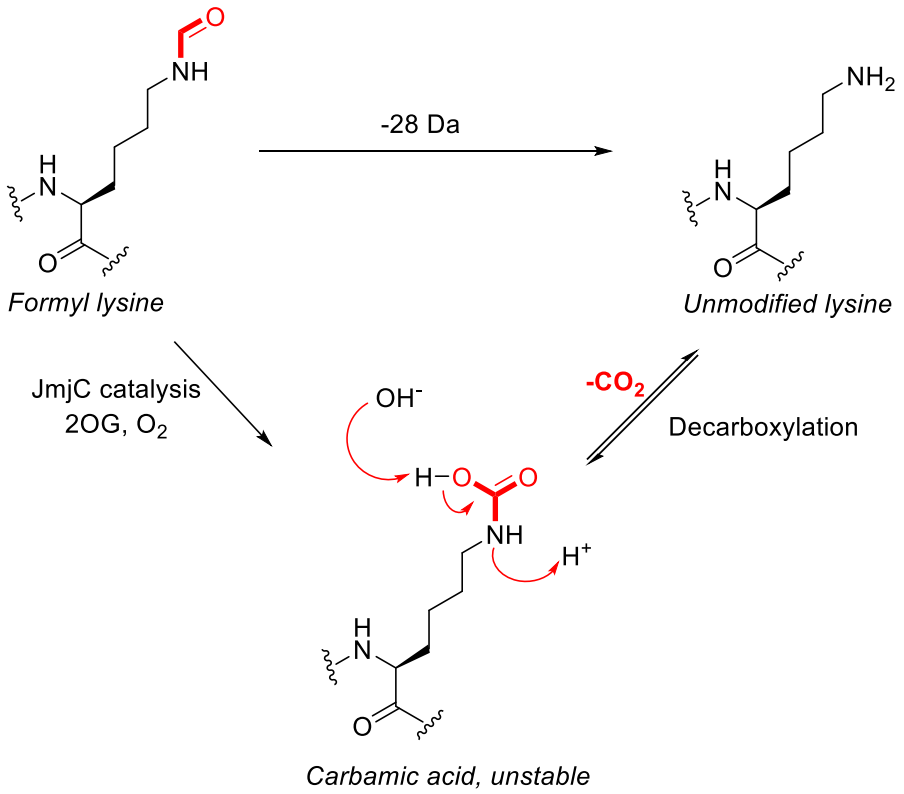
9.8 Discussion

Formylation is a recently discovered modification that adds to the repertoire of lysine acylation PTMs.[6] Compared to the well-studied lysine acetylation, little is known about the biological roles of lysine formylation and the factors that regulate this PTM. It is unclear whether lysine formylation on histones is catalysed enzymatically or is a secondary side effect reaction due to DNA damage. Nonetheless, the presence of formylation may block other lysine modifications, and therefore has epigenetic implications.[9]

A recent study identified two sirtuin type HDACs that can catalyse the removal of a formyl group on lysine residues (SrtN and SIRT1) *in vitro*.[10] The results presented in this chapter suggest that lysine deformylation may be carried out by enzymes other than the reported HDACs. The work described in this chapter provides the first evidence of a putative deformylation of N^{ε}-formyllysine residues as mediated by JmjC KDMs.

These findings strengthen the assumption that both KDM3A and JMJD1C can react with acyl type histone modifications, and correlate with their putative catalysis of acetyllysine hydroxylation and deacetylation (Chapters 7 and 8). However, there is also the possibility that the apparent deformylation by recombinant JMJD1C may result from the presence of a co-purified impurity which could catalyse both the removal of acetyl- and formyl- groups. In this case, deformylation could occur by simple hydrolysis. Such mechanism would not be dependent on the JmjC co-factors. The putative deacetylation reaction that was observed during incubation with JMJD1C was not dependent on 2OG and Fe^{II}. Further investigation into the co-factors dependence of the putative deformylation catalysis by JMJD1C may shed light on the catalytic source of the reaction. Interestingly, reaction by both enzymes seemed to be induced by H3 K4me3. Future screening for formylated substrates should test for deformylation in the presence of Kme3 at the -5 position.

Putative deformylation mediated by KDM3A was observed to be dependent on the JmjC co-factors (**Figure 9.5**). Although there is no precedence for JmjC-mediated deformylation, a mechanism for a putative JmjC-mediated deformylation is by hydroxylation to form a carbamic acid (**Scheme 9.1**). The unstable carbamic acid can then undergo spontaneous decarboxylation to release carbon dioxide and the unmodified lysine. A +16 Da shift in peptide mass corresponding to the formation of carbamic acid was not observed in MALDI-TOF mass spectra. However, the half-life of such an intermediate would likely to be very short and thus it may not be detected by mass spectrometric analyses (see Section 9.3).



Scheme 9.1 Proposed mechanism for KDM3A deformylation. JmjC-mediated hydroxylation forms a carbamic acid. The unstable carbamic acid then undergo spontaneous decarboxylation to release carbon dioxide and the free unmodified lysine. The formyl group is labelled in bold red lines.

If this initial data on KDM3 reactivity with acyl type modifications presented herein is relevant *in vivo*, members of the KDM3 might regulate both methylation and acylation marks. The observed dependency of H3 K9 putative deformylation on H3 K4 methylation for both KDM3A and JMJD1C emphasises the combinatorial regulation between different histone PTMs, a phenomenon that would be beneficial to further investigate for a better understanding of the epigenetic code. Future study should also screen for additional acyl-type modifications as potential KDM3 substrates. Another important objective for future study is the kinetic characterisation of the new deformylation reaction. In light of the complete apparent deformylation reaction that was observed for KDM3A mediated catalysis under the tested conditions, it is possible that the catalytic efficiency for deformylation might be even greater than that of the demethylation reaction.

9.9 References

1. Huang, H., Lin, S., Garcia, B. A. & Zhao, Y. Quantitative proteomic analysis of histone modifications. *Chem. Rev.* **115,** 2376–418 (2015).

2. Tan, M. *et al.* Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* **146,** 1016–1028 (2011).

3. Dai, L. *et al.* Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nat. Chem. Biol.* **10,** 365–70 (2014).

4. Jiang, T., Zhou, X., Taghizadeh, K., Dong, M. & Dedon, P. C. N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 60–5 (2007).

5. Sherman, F., Stewart, J. W. & Tsunasawa, S. Methionine or not methionine at the beginning of a protein. *BioEssays* **3,** 27–31 (1985).

6. Wisniewski, J. R., Zougman, A., Krüger, S. & Mann, M. Mass spectrmetric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. *Mol. Cell. Proteomics* **6,** 72–87 (2007).

7. Tibbetts, A. S. & Appling, D. R. Compartmentalization of Mammalian Folate-Mediated One-Carbon Metabolism. *Annu. Rev. Nutr.* **30,** 57–81 (2010).

8. Osés-Prieto, J. A., Zhang, X. & Burlingame, A. L. Formation of epsilon-formyllysine on silver-stained proteins: implications for assignment of isobaric dimethylation sites by tandem mass spectrometry. *Mol. Cell. Proteomics* **6,** 181–92 (2007).

9. Lin, H., Su, X. & He, B. Protein Lysine Acylation and Cysteine Succination by Intermediates of Energy Metabolism. *ACS Chem. Biol.* **7,** (2012).

10. Seidel, J., Klockenbusch, C. & Schwarzer, D. Investigating Deformylase and Deacylase Activity of Mammalian and Bacterial Sirtuins. *ChemBioChem* **17,** 398–402 (2016).

11. Hopkinson, R. J. *et al.* Is JmjC oxygenase catalysis limited to demethylation? *Angew. Chem. Int. Ed. Engl.* **52,** 7709–13 (2013).

12. Wisniewski, J. R., Zougman, A. & Mann, M. Nepsilon-formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. *Nucleic Acids Res.* **36,** 570–7 (2008).

Chapter 10

Summary and future prospects

10.1 Summary

"Every global traveller has experienced the disorientation of being unable to speak the local language. At present, those in the field of chromatin regulation and epigenetics are in a similar position. Like the traveller who hears words he or she does not understand, we are faced with observations that cannot be neatly categorized within previous models." Berger *et al.*[1]

Histones, the basic packaging units of eukaryotic DNA, are heavily decorated with a variety of post-translational modifications (PTMs).[2] In 2000, the "histone code hypothesis" was proposed, which posited that PTMs can act to form a language read by specialised proteins to regulate the structure of chromatin, and ultimately to mediate gene expression.[3,4] Although we have made an enormous progress in some aspects of epigenetics, elucidation of the full complexity of the control of gene expression by histone PTMs and the epigenetic enzymes that regulate them is only now beginning to emerge.

The histone lysine demethylase (KDM) family of epigenetic enzymes remove methyl marks from lysines on histone tails; by doing so they change the epigenetic code and regulate gene expression; The Jumonji C domain (JmjC) family is the largest family of KDMs.[5] The work described in this thesis aimed to biochemically investigate the inhibitor and substrate diversity of the JmjC KDMs, which has led to the identification of new inhibitors and substrates and revealed a dependence between adjacent PTMs.

There is an interest in developing small-molecule probes that modulate JmjC catalytic activity to investigate their biological roles.[6] 5-Carboxy-8-hydroxyquinoline (IOX1) is the most potent broad-spectrum inhibitor of JmjC demethylases reported.[7] However, IOX1 suffers from low cell permeability.[8] **Chapter 2** described structure–activity relationship studies leading to the discovery of an *n*-octyl ester form of IOX1 **5** with an improved cellular potency (EC_{50} value of 100 to 4 μM) and selectivity towards the KDM4 subfamily (see Chapter 2**, Table 2.1 and 2.2**). These findings were supported by *in vitro* inhibition and selectivity studies, activity versus toxicity analyses in cell cultures and intracellular uptake

measurements. The *n*-octyl ester of IOX1 should find utility as a starting point for the development of JmjC inhibitors such as branched and non-ester analogues, and as a cell-permeable tool compound for studies investigating the roles of JmjC KDMs in epigenetic regulation.

In **Chapter 3**, the production and kinetic characterisation of KDM3A, a JmjC histone demethylase with varied physiological functions, was carried out.[9] MALDI-TOF MS and FDH coupled assays were developed for the kinetic characterisation of KDM3A. Kinetic measurements for reported KDM3A substrates (H3 K9me2 and H3 K9me1) revealed substrate inhibition at elevated substrate concentrations, observations which were not described in previous reports (see Chapter 3, **Figure 3.13 and Figure 3.15**).[10] Studies investigating the effect of H3 K4me3 on the kinetics of KDM3A H3 K9 demethylation discovered that the H3 K4me3 mark appears to enhance the substrate inhibition, but it also improves the catalytic efficiency of the demethylation reactions (see Chapter 3, **Figure 3.14**; **Table 10.1**). Results from these kinetic characterisations were then used as the basis for subsequent comparisons between the catalytic efficiency of KDM3A for reported and novel substrates.

In **Chapter 4**, TCA cycle intermediates were evaluated as co-substrate mimic inhibitors for KDM3A. KDM3A was screened against TCA cycle intermediates by T_m shift measurements using DSF as well as FDH coupled–enzyme assays. IC_{50} determinations were then conducted with selected potent metabolites. The reported inhibition data reveal that TCA cycle intermediates inhibit 2OG dependent oxygenases with varying degrees of potency, with IC_{50} values ranging from the micromolar to the millimolar range.[11] The observed data suggest that KDM3A thermal stability was moderately sensitive to the tested metabolites in comparison to other 2OG oxygenases (see Chapter 4, **Table 4.2**). IC_{50} determinations validated the inhibition at the 100 μM range by a subset of the tested metabolites (see Chapter 4**, Table 4.3**). Metabolomic studies which provide concentration profiles for metabolites in healthy and diseased tissues are now emerging, and will help in identifying whether specific 2OG oxygenases can act as metabolic sensors.[12]

The work presented in **Chapter 5** involved analyses of KDM3A potential for non-epigenetic regulation during the formation of sperm flagellum. KDM3A is reported to be epigenetically involved in sperm development.[13–15] KDM3A is also proposed to interact with non-histone proteins in a catalytic and non-catalytic manner.[16,17] Intraflagellar transport proteins (IFTs) are proteins participating in flagella formation.[18] Work by collaborators has identified lysine methylation sites on a number of IFT proteins, and suggested a potential interaction

between KDM3A and IFTs (*Dr. Patricia Yeyati and Dr. Holger Kramer, manuscript in preparation*). The work described in this chapter used this data as the basis for further investigation, aiming to study this potential interaction at several levels of regulation. Under the tested *in vitro* assay conditions methylated IFT peptides were not found to be substrates for KDM3A. The FDH coupled assays revealed that the presence of the dimethyllysine residue in the IFT81 fragment peptide causes a greater inhibitory effect on KDM3A than the corresponding non-methylated peptide (IC_{50} of 24.2 μM vs. 270.4 μM respectively, see Chapter 5, **Figure 5.8**). This may indicate that methylation strengthens the interaction between KDM3A and IFT81. The methylated form of IFT81 was therefore hypothesised to function as a competitive inhibitor. However kinetic characterisation did not find evidence to support this proposal. Moreover, measurements of IFT81 protein levels in cells, in the presence or absence of KDM3A, indicated that KDM3A does not regulate IFT81 protein levels. qRT-PCR assay was developed for gene expression analyses, the results indicated that KDM3A does not affect the mRNA expression levels of *IFT74, IFT81* and *IFT88*. The work was then extended to test the ability of KDM3A to recognise other potential IFT PTMs. No activity was detected for any of the screened methylated arginine or acetylated lysine IFT fragment peptides. Further experiments will seek to determine whether KDM3A regulates the levels of methylated lysines on IFT in cells using a SILAC-based MS.[19] SPOT peptide arrays of IFTs stained with recombinant KDM3A could be useful in identifying and mapping the exact positions of the interaction between KDM3A and IFT proteins *in vitro*.[20,21]

Chapters 6 to 9 addressed the exploration of novel *in vitro* substrates for KDM3 (KDM3A and JMJD1C) mediated catalysis, including: methylated arginines, lysine analogues, acetylated and formylated lysines (**Table 10.1**). KDM3A was discovered to catalyse arginine demethylation *in vitro,* although the catalytic efficiency was lower than that of lysine demethylation (see Chapter 6**, Table 6.3**). Studies described in **Chapter 7** with unnatural lysine analogues, showed various KDM3A-catalysed reactions with uncharged residues (see Chapter 7, **Table 7.1**). This knowledge was then utilised to search for novel PTM substrates for KDM3A. The work described in this chapter provides the first evidence of JmjC KDM catalysed hydroxylation of an N^{ε}-acetyllysine residue (see Chapter 7, **Figure 7.3**). The H3 K4me3 position seems to be required for acetyllysine substrate recognition *in vitro*, demonstrating the need for further understanding of the combinatorial readout of multiple PTMs. Kinetic characterisation indicated that hydroxylation of acetyllysine-containing peptides by KDM3A is probably catalysed with poor efficiency in comparison to demethylation of lysine and arginine methylated peptides (**Table 10.1**). One important objective of future studies would be to test for the presence of this novel PTM on histones

and link it with KDM3A in a cellular context. To date, KDM3A is the only KDM that was observed to hydroxylate acetyllysines. However, activity tests of histone fragment peptides containing acetyllysines with JMJD1C indicated a putative deacetylation catalysis (see Chapter 8, **Figure 8.2 and Table 8.1**). **Chapter 8** describes preliminary data as to whether JMJD1C is responsible for this apparent deacetylation. The available initial *in vitro* results gained in this study were not sufficient to either eliminate or validate the hypothesis that JMJD1C is the catalyst. Current efforts are focused on testing for deacetylation by JMJD1C in cells. **Chapter 9** presents preliminary evidence for an additional novel acyl-PTM substrate for KDM3A and JMJD1C; both enzymes were observed to deformylate N^{ε}-formyllysine residues on histone H3 fragment peptides (see Chapter 9, **Figure 9.2**). Based on the preliminary data presented in this chapter, a mechanism for JmjC-mediated deformylation was proposed (see Chapter 9, **Figure 9.8**). Interestingly, H3 K4 methylation was also observed to enhance putative deformylation for both KDM3A and JMJD1C catalysed reactions, as in the case for H3 K9 demethylation/hydroxylation of acetyls.

Taken together, the work described in this thesis suggests that the catalytic activity of JmjC KDMs is not limited to lysine demethylation. If the data on KDM3 reactivity proves to be biologically relevant, members of the KDM3 subfamily might provide a regulatory link between methylation and acylation marks. Such a link will further highlight the complex relationships between histone PTMs and the epigenetic enzymes that regulate them. The observed dependency of H3 K9 catalysis on H3 K4 methylation adds another layer of complexity to the histone code.

Table 10.1 Novel putative *in vitro* histone PTM substrates for KDM3A identified in this work. Recognised histone N^{ε}-methylated-lysine peptide substrates are marked in bold. Apparent kinetic parameters were determined by the FDH coupled assay except where otherwise stated. Green highlight represents high efficiency, yellow highlight represents medium efficiency and red highlight represents low efficiency. *Methylated arginine peptides were synthesised by Dr. Louise Walport. Acetylated and formylated peptides were synthesised by Dr. Richard Hopkinson.*

Substrate peptide	Putative reaction	Efficiency k_{cat}/K_M (10^{-6} /s μM)	Relevant Chapter
H3(1-15)K9me2	**Lysine demethylation**	**14649** [a]	**Chapter 3**
H3(1-15)K9me1	**Lysine demethylation**	**29428** [b]	**Chapter 3**
H3(1-15)K4me3K9me2	Lysine demethylation	38330 [c]	Chapter 3
H3(1-15)K9Rme2a	Arginine demethylation	431	Chapter 6
H3(1-15)K9Rme2s	Arginine demethylation	127	Chapter 6
H3(1-15)K9Rme	Arginine demethylation	908	Chapter 6
H3(1-15)K4me3K9Ac	Hydroxylation of acetyllysine	38[d]	Chapter 7
H3(1-15)K4me3K9for	Lysine deformylation	ND, 84%[e]	Chapter 9

[a] Substrate inhibition at H3(1-15)K9me2 concentrations of > 200 μM was observed. Kinetic parameters were calculated based on the substrate inhibition model. [b] Substrate inhibition at H3(1-15)K9me1 concentrations of > 50 μM was observed. Preliminary kinetic parameters predications were calculated by Lineweaver–Burk plotting for data below substrate inhibition range. *[c] Substrate inhibition at H3(1-15)K4me3K9me2 concentrations of > 80 μM was observed. Kinetic parameters were calculated based on the substrate inhibition model. [d] Apparent kinetic parameters were determined by MALDI-TOF MS based assays [e] Percentage conversion of substrate peptide to the product form under the standard conditions as determined by MALDI-TOF MS based assays: KDM3A (2 μM) with an H3(1-15)K4me3K9for peptide(10 μM), 2OG (100 μM) and Fe^{II} (50 μM) were incubated over 2 h at 37°C (see Materials and methods, Chapter 11, Section 11.3.3, and Chapter 9, Section 9.3, for more information).*

10.2 Future recombinant KDM3A production and assay development

Although recombinant KDM3A is active at relatively low enzyme concentrations compared to other KDMs, recombinant KDM3A tends to degrade over time (see Chapter 3, ~60-70 kDa bands in **Figure 3.7**).[22] This degradation introduces complications with protein production procedures and assay development. The following section summarises conclusions and suggestions for the development of future KDM3A-related biochemical research.

As described in Chapter 3, KDM3A catalysis requires both the JmjC and the zinc finger domains (see Chapter 3, **Figure 3.1**). Therefore, the KDM3A construct is relatively large (94.4 KDa) and requires production in insect cells using the baculovirus system. Expression using baculovirus requires the use of fresh insect cells with low passage and specially designated incubators and hoods.[23] Use of the Sf-900™ II SFM medium (ThermoFisher scientific) resulted in good levels of production of KDM3A in insect cells (see Chapter 11, Section 11.2.3).

Since recombinant KDM3A degrades over time, purification should be carried out at 4°C and be completed as quickly as possible. The first report of recombinant KDM3A purification described a five chromatography column procedure, which resulted in a large amount of degraded protein.[24] In this thesis, a two column procedure, using nickel affinity and gel filtration chromatographies, based on the procedure described by Rose *et al* is used.[25] A step gradient of imidazole is recommended to improve purification; this might also help enable the removal of the protease likely to be responsible for KDM3A degradation. To minimise degradation by reducing purification time, it may be possible to carry out purification using nickel affinity chromatography alone. If the gel filtration step is omitted, the activity of the enzyme should be tested in the presence of imidazole. The enzyme was concentrated up to 63 μM without any observed precipitation, concentration to ~150 μM resulted in protein precipitation (see Chapter 11, Section 11.2.3). Storage conditions may further be optimised, for example, by using high concentrations of salt (> 500mM) and glycerol (20%). After thawing, gel filtration can be run in order to remove any aggregated protein and exchange to the assay buffer.

Several main difficulties should be taken into account when developing assays to study KDM3A function. Due to the degradation of the enzyme over time, any assay development should be designed such that all the samples are measured simultaneously, e.g. using an

appropriate plate system and a multi-channel pipette, similarly to the procedure used for kinetic measurements (See Chapter 3, Section 3.5). KDM3A should be defrosted as late as possible and kept at 4°C until used. Assay development efforts faced challenges due to the stability of recombinant KDM3A. Kinetic measurements introducing variations in O_2 (aq) concentrations are problematic as the current facilities are designed for gas equilibration of one sample at a time and therefore do not allow simultaneous measurements. Assays which required the use of a high concentration stock of KDM3A, such as fluorescence polarisation measurements, were hampered due to protein precipitation (*results not described in thesis*).

Overall, future work should focus on developing a more stable KDM3A construct, or identifying an additive, which could be a natural partner of KDM3A, to stabilise it.

10.3 Future prospects

The work described in this thesis identified new inhibitors and substrates for the JmjC KDMs and revealed a dependence between adjacent PTMs (for new inhibitors see Chapter 2, Chapter 3, Section 3.6 and Chapter 4; for new substrates see Chapters 6-9; for data on dependence between adjacent PTMs see Chapter 7, Section 7.4 and Chapter 9, Section 9.4). Clearly, there are many questions to explore and possible directions for future work have been suggested in the discussion sections at the end of each chapter. Of particular interest is the discovery of novel reactions catalysed by KDM3. Further research should extend this work and explore the relevance of the findings to a biologically relevant context.

Specifically, focusing on KDM3A hydroxylation of acetyllysines, possible aims for future investigations include:

i) <u>To confirm the hydroxylation reaction *in vitro*</u>

KDM3A mediated hydroxylation of the acetyl group on an acetyllysine histone fragment peptide was demonstrated using deuterated *N*ε-acetyllysine substrate (see Chapter 7, **Figure 7.6**). Additional experiments can be carried out to validate the reaction *in vitro:*

a. Use of two-dimensional nuclear magnetic resonance spectroscopy (2D NMR) to confirm the hydroxylation position by comparing the product spectra of an enzymatic reaction with H3(1-15)K4me3K9Ac with that of a synthetic reference peptide containing hydroxylated acetyllysine.

b. Conduct $^{18}O_2$ experiments to unequivocally prove that the –OH group is derived from atmospheric oxygen as catalysed by KDM3A.

ii) <u>To decipher the biochemical basis of substrate recognition by KDM3A</u>

KDM3A is the first human KDM to be found to react with uncharged substrates (see Chapter 7, Sections 7.3 and 7.4.2; Chapter 9, Section 9.3). It was hypothesised that a positive charge on a KDM3A substrate interferes with KDM3A catalysis since a tri-alkyl lysine analogue that have a similar size to tri-methyllysine was catalysed by KDM3A (see Chapter 7, Section 7.3). KDM3B catalyses demethylation of H3 K9me1/2 marks but not H3 K9me3, which is similar to KDM3A.[26] KDM3B has 83% protein sequence identity to KDM3A in its JmjC domain (see Chapter 1, **Figure 1.11**). The results presented in this thesis suggest that KDM3B does not catalyse acetyllysine hydroxylation under the tested *in vitro* conditions (see and Chapter 7, **Figure 7.15**).

a. As there is no reported crystal structure for KDM3A, homology modelling can be used to build a model of KDM3A based on the structures of other KDMs. Comparison between a model of the KDM3A active site and the reported crystal structure of the JmjC domain of KDM3B (PDB ID - 4C8D) could be carried out to investigate the reasons why KDM3A can accept uncharged substrates but KDM3B apparently cannot. For instance, positively charged residues unique to the KDM3A substrate binding site might regulate substrate recognition by repulsion of positively charged substrates. *In silico* simulations can then be utilised to identify key residues candidates which may be involved in binding to H3 K9me and H3 K9Ac, these residues could then be further investigated using site directed mutagenesis.

b. It is of interest to investigate KDM3B *in vitro* substrate scoping and compare it with those of KDM3A. Although KDM3B did not react with an acetyllysine histone 3 fragment peptide (see Chapter 7, **Figure 7.15**), it could be useful to thoroughly test its substrate variability experimentally by using lysine analogue peptides, similar to the work described in Chapter 7, Section 7.3. Data from this assay might strengthen or weaken hypotheses arising from the structural comparisons regarding the differences between KDM3A and KDM3B substrate recognition.

c. Testing the *in vitro* catalytic activity of KDM3A and KDM3B at different pH values is of interest. If KDM3A catalytic activity is indeed dependent on substrate charge, the catalytic activity might be regulated by pH changes in cells which affect the substrates protonation state. A MALDI-TOF MS activity assay could be carried out in varying pH using appropriate buffers; KDM3B catalytic activity could be tested and compared with that of KDM3A at different pH values.

iii) <u>To investigate the dependence of KDM3A-catalysed hydroxylation of acetyllysines on the H3 K4me3 mark</u>

Hydroxylation of acetyllysine by KDM3A was observed with a H3(1-15) K4me3K9Ac peptide. Hydroxylation of H3 K9Ac was only observed in the presence of H3 K4 trimethylation.

a. Further work is needed to understand the requirement of H3 K4me3 for the KAc reaction and understand whether it is sequence specific or general to Kme3 at positions -4/-5 to an acetyllysine. Testing whether trimethylation of a lysine at positions -4/-5 to that of the acetyllysine on histones might result in KDM3A substrate recognition is of interest, since the distance between Kme3 and KAc is the same as H3 K4me3K9Ac. For instance, although a peptide fragment of H3 K14Ac is not hydroxylated by KDM3A, H3 K9me3K14Ac may be a substrate (see Chapter 7, Section 7.9). Similarly, a peptide fragment of H3 K18me3K23Ac may be a substrate for KDM3A although the corresponding H3 K23Ac was not (see Chapter 7, Section 7.9). To investigate the sequence specificity of KDM3A alanine scanning could be used, systematically replacing each residue in the H3(1-15)K4me3K9Ac peptide.[27] A peptide containing only alanine residues before and after the K4 and K9 positions can be utilised to rule out sequence specificity. An extended screen for potential acetyllysine substrates across the proteome is suggested below (Section 10.3, Subsection iv, e).

b. Measurement of the binding affinity of KDM3A to H3 K9Ac and H3 K4me3K9Ac peptides to test whether methylation on the H3 K4 position increases the binding affinity to KDM3A. Such measurements can be carried out, using Microscale Thermophoresis (MST) in which fluorescence is used to monitor the motion of molecules along a microscopic temperature gradient formed in a glass capillary, with the advantage of using a low

sample volume.[28] Another possible method for binding affinity determination is BioLayer Interferometry (BLI), which is an optical analytical technique that analyses the interference pattern of reflected white light due to binding between a ligand immobilized on the biosensor tip surface and an analyte in solution.[29] BLI has the advantage of being a label-free technology. Note that a methyl-reader domain may only be created at the dimer interface. Although specific residues required for dimerisation are not reported, it was demonstrated that both the JmjC and the zinc finger domains are essential for dimer formation.[10] To test the hypothesis that a methyl-reader domain is created at the dimer interface, use of Isothermal Titration Calorimetry (ITC) for binding and stoichiometry measurements can be carried out.[30] Prior to this test, analytical gel filtration or analytical ultracentrifugation should be carried out to conform the dimerisation of the $KDM3A_{515-1317}$ construct.

c. Although KDM3A does not have a PHD finger, it may harbour a different methyl-reader domain that can interact with H3 K4me3. Modelling of KDM3A may shed light on putative residues that are similar to that of the zinc binding motif found in PHD domains.[31]

iv) <u>To investigate whether hydroxylation of acetyllysines is a novel PTM present on histones and study its biological function and KDM3A regulation in a cellular context</u>

a. Mass spectrometric analysis on histones prepared from KDM3A modulated cells (knock-out, wildtype, catalytically inactive mutant, overexpressed) can be conducted in order to confirm the presence of hydroxyl-acetyl lysine as a novel PTM on histones and to investigate whether this PTM is KDM3A regulated. Use of a synthetic hydroxylated acetyllysine reference peptide identical to the expected trypsin digest product may be used to facilitate the identification of this PTM. Deuterated acetic anhydride can be used to block free lysine positions in a procedure similar to that described by Mackeen *et al*.[32] This procedure will ensure that trypsin can only digest arginines and will result in longer, and therefore more hydrophobic, peptides which can be better separated in the C18 LCMS column.

b. Quantification of the abundancy of this novel PTM. If hydroxyl-acetyl lysines can indeed be detected as a novel PTM on histones, then quantification of

the stoichiometry of this modification should be performed. This can be carried out using stable isotope labelling by amino acids in cell culture (SILAC)-based methods and PITC (Edman's Reagent) similar to the approaches used for lysine 2-hydroxyisobutyrylation and lysine formylation abundancy quantification.[33,34] Quantitative proteomics procedures of histone PTMs has been thoroughly reviewed by Huang *et al*.[35]

c. Generation of selective antibodies against hydroxy-acetyllysine could be very useful. Antibodies can be used to enrich the hydroxy-acetyllysine containing peptides in digested histone samples used for proteomics analysis if the PTM levels are below the detection limit in a similar procedure used for histidine phosphorylation detection.[36] Cellular immunofluorescence assays can also be developed using the new antibody to further verify KDM3A activity. In addition, these antibodies can be utilised in ChIP-seq experiments to decipher the localisation of this PTM and thus gain more insight into its biological function.

d. Expansion of the *in vitro* tests to study the effects of hydroxylation on an acetyllysine on bromodomains and histone deacetylases. Since hydroxylation is a modification of the well-studied lysine acetylation, hydroxylation of acetyllysines is likely to modulate the binding affinity of bromodomains (BRDs), histone acetyltransferases (HATs) and histone deacetylases (HDACs). Displacement studies described in Chapter 7, Section 7.10 have demonstrated that hydroxylation of acetyllysines could decrease the binding affinity of BRD4 to its histone tails ligands. Tests with additional bromodomains should be carried out, in particular with those reported to have a preference for the H3 K9Ac position such as Spt7 and CECR2 BRDs.[37,38] A collaboration with Eli Lilly is currently under way to test for the effect of acetyllysine hydroxylation on HDAC activity. A number of zinc-dependent HDACs have been found to recognise this modification (*Eli Lilly, unpublished data)*. Further kinetic characterisation as well as tests with NAD^+-dependent HDACs should be explored.

e. A high throughput approach could be used to identify candidate novel substrates across the proteome for KDM3A by designing peptides based on post-translational modification databases such as PhosphoSitePlus and dbPTM.[39,40] The tested substrate peptides should contain an acetyllysine modification and a tri-methyllysine modification at positions -4/-5 of the

acetyllysine to resemble the H3 K4me3K9Ac substrate. Potential substrates could also have the 'reverse' arrangement, where the acetyllysine modification is at position -4/-5 of the tri-methyllysine. A preliminary search across the proteome based on the UniProt database identified 317,445 proteins with the pattern K-X-X-X-X-K, i.e. - two lysines which are five residues apart, out of 553,656 proteins; more than half of the reported proteins across the proteome.[41] A more narrowed search, based on the reported PTM sites deposited on PhosphoSitePlus, found 5432 methyllysine sites on 2978 proteins, and 46203 acetyllysine sites on 14665 proteins (some proteins contain multiple hits).[39] The combination of a methyllysine modification at position -5 of an acetyllysine appeared 76 times on the search results. These hits include variants and orthologs, as well as multiple hits on the same protein (see **Appendix 3**).[39] Therefore, only about 0.2% of the reported acetyllysine sites have a reported methyllysine at position -5. Notably, GO annotations have indicated that two thirds of the hits are DNA/RNA binding proteins (22 out of 36 distinct proteins). While this information may shed some light on the biological distribution of the Kme-X-X-X-X-KAc pattern across the proteome, it may also reflect a bias in the database towards this type of proteins. Future work could involve the synthesis of these potential substrate peptide fragments and *in vitro* KDM3A activity screen to determine whether any of these are substrates / binders. SPOT peptide arrays could be utilised to generate a peptide library at the scale of thousands of peptides.[20,21,42] Therefore, SPOT arrays can facilitate the screening of all possible substrates, including those with the 'reverse' arrengment of an acetyllysine preceding a methyllysine and those with a -4 distance between the two PTMs to provide a comprehensive peptide library for screening as potential substrates / binders. This library would then be incubated with KDM3A under *in vitro* reaction conditions, and the detection of potential novel substrates for KDM3A can be achieved using a reaction product selective antibody (ie. hydroxy-acetyllysine antibody). If binding assays to KDM3A are employed, anti- His_6-tag antibody can be utilised. It should be noted that the reported dissociation constant (K_D) values for H3 K9me0/1/2 peptides are in the single-digit micromolar range, however affinities may vary between different substrates.[10] Another potential method to detect enzymatic reactions could be using peptide arrays is by

'Self-Assembled monolayers with Matrix-assisted laser Desorption-Ionization' (SAMDI) mass spectrometry, similarly to the procedure used by Gurard-Levin *et al.* to study the specificities of histone deacetylases.[43]

In addition to studying KDM3A hydroxylation of acetyllysines, a number of additional investigations could be carried out following the findings of novel reactions for KDM3 described in this thesis. These include:

I. The promiscuity of KDM3A could be further explored by testing its reactivity with other residues and acyl-type modifications. It is of interest is to test whether KDM3A can hydroxylate isobutyryl-lysine to form the lysine 2-hydroxyisobutyrylation (Khib), which has been identified by MS[34] Another interesting substrate to test would be an H3 K9 peptide where the lysine is replaced with methionine. The H3 K27M mutation is prevalent in a subtype of aggressive paediatric brain cancers, and the H3 K9M variant has been reported to be recognised by KDM3B.[44–46]
II. Chapter 8 describes a putative histone deacetylase activity for KDM3C, a KDM with no apparent demethylase activity, Deacetylation is an unprecedented activity for 2OG oxygenases.[26,47] Current efforts are focused on testing for deacetylation in cells by an immunofluorescence assay, using overexpressed JMJD1C and an anti-acetyllysine antibody.
III. Chapter 9 describes KDM3A/3C (but not KDM3B) catalysed deformylation. While the biological functions of formyllysines are still unclear, no human deformylase have been identified to date; hence this may be of great importance if the activity was validated in cellular context. Future screening for formylated substrates should test for deformylation in the presence of Kme3 at the -4/-5 position.

Overall, the work described in this thesis has identified the potential of JmjC KDMs to catalyse reactions beyond lysine demethylation and opened many possibilities for further discoveries on the function of JmjC oxygenases.

10.4 References

1. Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447,** 407–412 (2007).

2. Gardner, K. E., Allis, C. D. & Strahl, B. D. OPERating ON Chromatin, a Colorful Language where Context Matters. *J. Mol. Biol.* **409,** 36–46 (2011).

3. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293,** 1074–80 (2001).

4. Turner, B. M. Cellular Memory and the Histone Code. *Cell* **111,** 285–291 (2002).

5. Walport, L. J., Hopkinson, R. J. & Schofield, C. J. Mechanisms of human histone and nucleic acid demethylases. *Curr. Opin. Chem. Biol.* **16,** 525–34 (2012).

6. Frye, S. V. The art of the chemical probe. *Nat. Chem. Biol.* **6,** 159–161 (2010).

7. Hopkinson, R. J. *et al.* 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem. Sci.* **4,** 3110 (2013).

8. IOX1: A Cell-Penetrating Broad-Spectrum 2-Oxoglutarate Oxygenase Inhibitor. *SGC* Available at: http://www.thesgc.org/scientists/chemical_probes/IOX1.

9. Schofield, Christopher JHausinger, R. P. 2-Oxoglutarate-Dependent Oxygenases. *The Royal Society of Chemistry* (2015).

10. Goda, S., Isagawa, T., Chikaoka, Y., Kawamura, T. & Aburatani, H. Control of histone H3 lysine 9 (H3K9) methylation state via cooperative two-step demethylation by Jumonji domain containing 1A (JMJD1A) homodimer. *J. Biol. Chem.* **288,** 36948–56 (2013).

11. Rose, N. R., McDonough, M. A., King, O. N. F., Kawamura, A. & Schofield, C. J. Inhibition of 2-oxoglutarate dependent oxygenases. *Chem. Soc. Rev.* **40,** 4364–97 (2011).

12. Emir, U. E. *et al.* Noninvasive Quantification of 2-Hydroxyglutarate in Human Gliomas with IDH1 and IDH2 Mutations. *Cancer Res.* **76,** 43–9 (2016).

13. Kasioulis, I. *et al.* Kdm3a lysine demethylase is an Hsp90 client required for cytoskeletal rearrangements during spermatogenesis. *Mol. Biol. Cell* **25,** 1216–1233 (2014).

14. Okada, Y., Scott, G., Ray, M. K., Mishina, Y. & Zhang, Y. Histone demethylase JHDM2A is critical for Tnp1 and Prm1 transcription and spermatogenesis. *Nature* **450,** 119–23 (2007).

15. Liu, Z. *et al.* Jmjd1a demethylase-regulated histone modification is essential for cAMP-response element modulator-regulated gene expression and spermatogenesis. *J. Biol. Chem.* **285,** 2758–70 (2010).

16. Schneider, P. *et al.* Identification of a novel actin-dependent signal transducing module allows for the targeted degradation of GLI1. *Nat. Commun.* **6,** 8023 (2015).

17. Ramadoss, S., Guo, G. & Wang, C.-Y. Lysine demethylase KDM3A regulates breast cancer cell invasion and apoptosis by targeting histone and the non-histone protein

p53. *Oncogene* **36**, 47-59 (2017).

18. Rosenbaum, J. L. & Witman, G. B. Intraflagellar transport. *Nat. Rev. Mol. Cell Biol.* **3,** 813–825 (2002).

19. Ong, S.-E., Mittler, G. & Mann, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods* **1,** 119–26 (2004).

20. Frank, R. Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron* **48,** 9217–9232 (1992).

21. Amartely, H., Iosub-Amir, A. & Friedler, A. Identifying protein-protein interaction sites using peptide arrays. *J. Vis. Exp.* **93** (2014).

22. Walport, L. J. *et al.* Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat. Commun.* **7,** 11974 (2016).

23. *Instruction Manual Guide to Baculovirus Expression Vector Systems (BEVS) and Insect Cell Culture Techniques*. (Invitrogen life technologies).

24. Yamane, K. *et al.* JHDM2A, a JmjC-Containing H3K9 Demethylase, Facilitates Transcription Activation by Androgen Receptor. *Cell* **125,** 483–495 (2006).

25. Rose, N. R. *et al.* Plant growth regulator daminozide is a selective inhibitor of human KDM2/7 histone demethylases. *J. Med. Chem.* **55,** 6639–43 (2012).

26. Brauchle, M. *et al.* Protein complex interactor analysis and differential activity of KDM3 subfamily members towards H3K9 methylation. *PLoS One* **8,** e60549 (2013).

27. Morrison, K. L. & Weiss, G. A. Combinatorial alanine-scanning. *Curr. Opin. Chem. Biol.* **5,** 302–307 (2001).

28. Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D. & Duhr, S. Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Commun.* **1,** 100 (2010).

29. Rich, R. L. & Myszka, D. G. Higher-throughput, label-free, real-time molecular interaction analysis. *Anal. Biochem.* **361,** 1–6 (2007).

30. Keller, S. *et al.* High-Precision Isothermal Titration Calorimetry with Automated Peak-Shape Analysis. *Anal. Chem.* **84,** 5066–5073 (2012).

31. Sanchez, R. & Zhou, M.-M. The PHD finger: a versatile epigenome reader. *Trends Biochem. Sci.* **36,** 364–372 (2011).

32. Mackeen, M. M. *et al.* Small-molecule-based inhibition of histone demethylation in cells assessed by quantitative mass spectrometry. *J. Proteome Res.* **9,** 4082–92 (2010).

33. Jiang, T., Zhou, X., Taghizadeh, K., Dong, M. & Dedon, P. C. N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 60–5 (2007).

34. Dai, L. *et al.* Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nat. Chem. Biol.* **10,** 365–70 (2014).

35. Huang, H., Lin, S., Garcia, B. A. & Zhao, Y. Quantitative proteomic analysis of histone

modifications. *Chem. Rev.* **115,** 2376–418 (2015).

36. Fuhs, S. R. & Hunter, T. pHisphorylation: the emergence of histidine phosphorylation as a reversible regulatory modification. *Curr. Opin. Cell Biol.* **45,** 8–16 (2017).

37. Hassan, A. H. *et al.* Selective recognition of acetylated histones by bromodomains in transcriptional co-activators. *Biochem. J.* **402,** 125–33 (2007).

38. Filippakopoulos, P. & Knapp, S. The bromodomain interaction module. *FEBS Lett.* **586,** 2692–704 (2012).

39. Hornbeck, P. V. et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2015).

40. dbPTM. Available at: http://dbptm.mbc.nctu.edu.tw/. (Accessed: 22nd February 2017).

41. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45,** D158–D169 (2017).

42. Filippakopoulos, P. *et al.* Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **149,** 214–231 (2012).

43. Gurard-Levin, Z. A., Kim, J. & Mrksich, M. Combining Mass Spectrometry and Peptide Arrays to Profile the Specificities of Histone Deacetylases. *ChemBioChem* **10,** 2159–2161 (2009).

44. Schwartzentruber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482,** 226–231 (2012).

45. Herz, H.-M. *et al.* Histone H3 lysine-to-methionine mutants as a paradigm to study chromatin signaling. **345,** 1065–70 (2014).

46. Wu, G. *et al.* Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat. Genet.* **44,** 251–3 (2012).

47. Hopkinson, R. J. *et al.* Is JmjC oxygenase catalysis limited to demethylation? *Angew. Chem. Int. Ed. Engl.* **52,** 7709–13 (2013).

Chapter 11
Materials and methods

11.1 Chemical procedures

11.1.1 General experimental

Reagents solvents and commercial peptides Reagents and solvents were, unless otherwise stated, from Sigma Aldrich, Alfa Aesar or Acros. Solvents were removed under reduced pressure using a Buchi™ rotary evaporator. Water was purified using an Elix® UV-10 system. H3(1-15)K4me3K9me2, H3(1-15)K9me2 and H3(1-15)K9me1 peptides were commercially synthesised at >95% purity and purchased from GLBiochem.

Analytical thin layer chromatography (TLC) was carried out on Merck silica gel 60 F254 aluminium supported thin layer chromatography sheets. Visualisation was by absorption of UV light (λ_{max} 254 nm).

Flash Column chromatography was carried out using a Biotage SP1 automated flash column chromatography platform, eluting with indicated solvents under a positive pressure of compressed air.

Melting points were determined using a Leica Galen III hot stage melting point apparatus and microscope.

Infrared spectra were obtained as a thin film on sodium chloride discs. The spectra were recorded on a Bruker Tensor 27 spectrometer and a representative number of absorption maxima are reported in wavenumbers (cm^{-1}).

^{1}H NMR spectra were recorded using a Bruker DPX400 (400 MHz) using deuterochloroform or DMSO-d_6 as a reference for internal deuterium lock. The chemical shifts data are given as δ in units of parts per million (ppm) relative to tetramethylsilane (TMS) where δ (TMS) = 0.00 ppm. The multiplicity of each signal is indicated by: s (singlet); app. br s (apparent broad singlet); d (doublet); t (triplet); q (quartet); dd (doublet of doublets); ddd (doublet of doublet of doublets); td (triplet of doublets) or m (multiplet). The number of protons (n) for

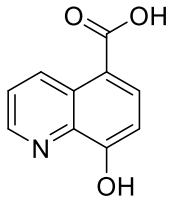
a given resonance signal is indicated by nH. Coupling constants (*J*) are expressed in Hz and are recorded to the nearest 0.5 Hz. Coupled proton coupling constants (*J*) are averaged and reported to the nearest 0.5 Hz.

^{13}C NMR spectra were recorded using a Bruker AV400 (101 MHz) spectrometer using the PENDANT or DEPT Q pulse sequences with broadband proton decoupling and internal deuterium lock. The chemical shift values of each signal are given as δ in units of parts per million (ppm) relative to tetramethylsilane (TMS) where δC (TMS) = 0.00 ppm. ^{1}H and ^{13}C spectra were assigned using 2D NMR experiments including COSY and HSQC.

Mass spectra were acquired on Agilent technologies 6120 quadruple LC/MS spectrometer using electrospray ionisation, operating in positive or negative mode, from sample solutions in MeOH. *m/z* values are reported in Daltons and followed by their percentage abundance in parentheses. High resolution mass spectra (HRMS) were recorded using a Bruker MicroTOF machine internally calibrated with polyalanine.

11.1.2 Synthetic procedures and characterisation for compounds

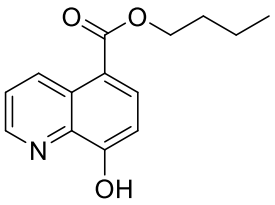
8-Hydroxyquinoline-5-carboxylic acid (1)[1]



Acrolein (550 mg, 0.65 mL, 9.81 mmol, 1.5 eq.) was added dropwise over 30 min to a solution of 3-amino-4-hydroxybenzoic acid (1.0 g, 6.54 mmol, 1.0 eq.) in 6 N HCl. The reaction mixture was refluxed at 100 °C for 2 h in a 50 mL round bottom flask fitted with a jacketed water condenser. Upon completion of the reaction based on TLC analysis, the reaction mixture was allowed to cool to room temperature and the pH was adjusted to pH 9 with aqueous ammonia. The mixture was then filtered and the filtrate was acidified to pH 4-5 with 10% aqueous acetic acid. The resulting precipitate was obtained by filtration, washed with water (10 mL) and dried in vacuo to yield the desired product **1** as a brown powder (642 mg, 3.40 mmol, 52%). R_f = 0.35 (CH_2Cl_2 / MeOH (3:1)); m.p. 272-273 °C (decomposition) (Literature value: 278-280 °C)[1]; ^{1}H NMR (400 MHz, DMSO-d_6) δ = 7.13 (1 H, d, *J* = 8.0 Hz, C(Ar)*H*), 7.70 (1 H, dd, *J* = 8.5, 4.0 Hz, C(Ar)*H*), 8.25 (1 H, d, *J* = 8.0 Hz, C(Ar)*H*),

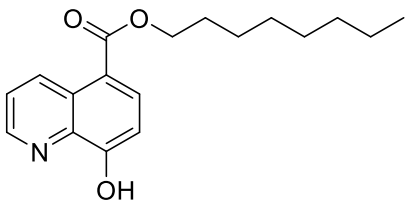
8.92 (1 H, dd, *J* = 4.0, 1.5 Hz, C(Ar)*H*), 9.47 (1 H, dd, *J* = 8.5, 1.5 Hz, C(Ar)*H*) ppm; ^{13}C-NMR (101 MHz, DMSO-d_6) δ = 111.0 (*C*(Ar)), 117.2 (*C*(Ar)), 124.1 (*C*(Ar)), 128.9 (*C*(Ar)), 134.4 (*C*(Ar)), 135.3 (*C*(Ar)), 139.1 (*C*(Ar)), 149.0 (*C*(Ar)), 158.7 (*C*(Ar)), 168.5 (*C*OO) ppm; FT-IR ν_{max}: 3210 (OH), 1684 (C=O) cm^{-1}; LRMS *m*/z (ESI$^+$) 191 ([M+2H] $^{2+}$, 100%); HRMS (ESI$^-$) calculated for $C_{10}H_6NO_3$ ([M-H]$^-$): 188.0353; found: 188.0351. These data are consistent with those previously reported.[1]

_n_-Butyl 8-hydroxyquinoline-5-carboxylate (4)



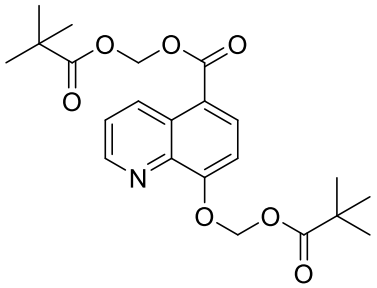
A solution of **1** (100 mg, 0.53 mmol) in *n*-butanol (10 mL) and one drop of concentrated H_2SO_4 was refluxed at 120 °C for 21 h in a 50 mL round bottom flask fitted with a jacketed water condenser. Upon completion of the reaction, based on TLC analysis, the reaction mixture was neutralised with 20% aqueous NaOH and excess *n*-butanol was removed in vacuo. The residue was dissolved in ethylacetate (EtOAc, 10 mL), washed with sat. $NaHCO_3$ (aq) (10 mL) and water (2 x 10 mL)), dried over Na_2SO_4 and concentrated in vacuo. The resulting brown gum was purified by flash silica gel chromatography (gradient elution 12-50% CH_2Cl_2 in cyclohexane (*c*Hex)) to give **4** as a sand-yellow solid (33 mg, 0.14 mmol, 25%). R_f = 0.30 (EtOAc / *c*Hex (1:1)); m.p. 76.5-77.7 °C; ^{1}H NMR (400 MHz, $CDCl_3$) δ = 1.01 (3 H, t, *J* = 7.5 Hz, C*H*$_3$), 1.39 - 1.63 (2 H, m, C*H*$_2$CH$_3$), 1.68 - 2.02 (2 H, m, C*H*$_2$CH$_2$CH$_3$), 4.38 (2 H, t, *J* = 7.0 Hz, C*H*$_2$CH$_2$CH$_2$CH$_3$), 7.17 (1 H, d, *J* = 8.0 Hz, C(Ar)*H*), 7.59 (1 H, dd, *J* = 9.0 , 4.0 Hz, C(Ar)*H*), 8.36 (1 H, d, *J* = 8.0 Hz, C(Ar)*H*), 8.81 (1 H, dd, *J* = 4.0, 1.5 Hz, C(Ar)*H*), 9.52 (1 H, dd, *J* = 9.0, 1.5 Hz, C(Ar)*H*) ppm; ^{13}C NMR (101 MHz, CDCl3) δ = 13.8 (*C*H$_3$), 19.4 (*C*H$_2$CH$_3$), 30.9(*C*H$_2$CH$_2$CH$_3$), 64.7 (O*C*H$_2$), 108.6 (*C*(Ar)), 116.9 (*C*(Ar)), 123.4 (*C*(Ar)), 128.0 (*C*(Ar)), 133.7 (*C*(Ar)), 135.4 (*C*(Ar)), 137.9 (*C*(Ar)), 147.8 (*C*(Ar)), 156.5 (*C*(Ar)), 166.4 (*C*OO) ppm; FT-IR ν_{max}: 3200 (OH), 2959-2873 (C-H), 1693 (C=O) cm^{-1}; LRMS *m*/z (ESI$^+$) 246 ([M+H]$^+$, 100%); HRMS *m*/z (ESI$^+$) calculated for $C_{14}H_{16}NO_3^+$, [M+H]$^+$ = 246.1125; found [M+H]$^+$ = 246.1121.

***n*-Octyl 8-hydroxyquinoline-5-carboxylate (5)**



A solution of **1** (100 mg, 0.53 mmol) in *n*-octanol (10 mL) and one drop of concentrated H_2SO_4 was refluxed at 120 °C for 20 h in a 50 mL round bottom flask fitted with a jacketed water condenser. Upon completion of the reaction based on TLC analysis, the reaction mixture was neutralised with 20% aqueous NaOH. Excess *n*-octanol was removed by rotary evaporation at 70 °C and 15 mbar. The residue was further purified by flash silica gel chromatography (eluting with 6 column volumes of *c*Hex followed by gradient elution with 0 to 100% CH_2Cl_2 in *c*Hex) to give **5** as a light-yellow solid (39 mg, 0.13 mmol, 25%). R_f = 0.40 (CH_2Cl_2 / MeOH (19:1)); m.p. 82.0-84.1 °C; ^{1}H NMR (400 MHz, $CDCl_3$) δ = 0.89 (3 H, t, *J* = 6.5 Hz, C*H*$_3$), 1.24 - 1.56 (10 H, m, $CH_3(CH_2)_5$), 1.76 - 1.90 (2 H, m, $COOCH_2CH_2$), 4.37 (2 H, t, *J* = 6.5 Hz, $COOCH_2$), 7.18 (1 H, d, *J* = 8.5 Hz, C(Ar)*H*), 7.59 (1 H, dd, *J* = 8.5, 4.0 Hz, C(Ar)*H*), 8.36 (1 H, d, *J* = 8.5 Hz, C(Ar)*H*), 8.81 (1 H, dd, *J* = 4.0, 1.5 Hz, C(Ar)*H*), 9.52 (1 H, dd, *J* = 8.5, 1.5 Hz, C(Ar)*H*) ppm; ^{13}C NMR (101 MHz, CDCl3) δ = 14.1(*C*H3), 22.7 (CH_3CH_2), 26.2 ($COO(CH_2)_2CH_2$), 28.8 ($COOCH_2CH_2$), 29.2 ($COO(CH_2)_4CH_2$) , 29.3 ($COO(CH_2)_3CH_2$), 31.8 ($COO(CH_2)_5CH_2$), 65.0 ($COOCH_2$), 108.6 (*C*(Ar)), 116.9 (*C*(Ar)), 123.4 (*C*(Ar)), 128.0 (*C*(Ar)), 133.7 (*C*(Ar)), 135.4 (*C*(Ar)), 137.9 (*C*(Ar)), 147.8 (*C*(Ar)), 156.5 (*C*(Ar)), 166.4 (*C*OO) ppm; FT-IR ν_{max}: 3297 (OH), 2857-2955 (C-H), 1702 (C=O) cm^{-1}; LRMS *m*/z (ESI$^+$) 302([M+H]$^+$, 100%); HRMS *m*/z (ESI$^+$) calculated for $C_{18}H_{24}NO_3^+$, [M+H]$^+$ = 302.1651; found [M+H]$^+$ = 302.1744.

(Pivaloyloxy)methyl-8-((pivaloyloxy)methoxy)quinoline-5-carboxylate (7)



To a mixture of **1** (100 mg, 0.53 mmol, 1.0 eq.) and chloromethyl pivalate (77 µL, 0.53 mmol, 1.0 eq.) in a mixture of acetone / dimethylformamide (1:1, 10 mL) was added Et_3N (54 mg, 0.53 mmol, 1.0 eq.). The mixture was refluxed at 50 °C for 3 h, whereupon a small amount of precipitate was formed. The precipitate was filtered and washed with acetone (20 mL). The filtrate was evaporated and the residue was partitioned between 5% w/w aqueous $NaHCO_3$ (30 mL) and EtOAc (20 mL). The organic phase was collected, dried over $MgSO_4$ and concentrated in vacuo. The residue was further purified by flash silica gel chromatography (gradient elution of 0 to 40% EtOAc in *c*Hex (15 column volumes) followed by 5 column volumes of 40% EtOAc in *c*Hex) to give **7** as a yellow oil (52 mg, 0.13 mmol, 24%). R_f = 0.45 (*c*Hex / EtOAc (3:2)); ^{1}H NMR (400 MHz, $CDCl_3$) δ = 1.22 (9 H, s, $(CH_3)_3$), 1.25 (9 H, s, $(CH_3)_3$), 6.07 (2 H, s, OCH_2O), 6.16 (2 H, s, OCH_2O), 7.32 (1 H, d, *J* = 8.0 Hz, C(Ar)*H*), 7.60 (1 H, dd, *J* = 9.0, 4.0 Hz, C(Ar)*H*), 8.38 (1 H, d, *J* = 8.0 Hz, C(Ar)*H*), 9.01 (1 H, dd, *J* = 4.0, 1.5 Hz, C(Ar)*H*), 9.47 (1 H, dd, *J* = 9.0, 1.5 Hz, C(Ar)*H*) ppm; ^{13}C NMR (101 MHz, DMSO-d_6) δ = 26.5 (2x$(CH_3)_3$), 38.3 ($C(CH_3)_3$), 38.4 ($C(CH_3)_3$), 80.0 (OCH_2O), 85.4 (OCH_2O), 111.3 (*C*(Ar)), 118.4 (*C*(Ar)), 123.7 (*C*(Ar)), 127.9 (*C*(Ar)), 132.7 (*C*(Ar)), 133.4 (*C*(Ar)), 139.5 (*C*(Ar)), 150.0 (*C*(Ar)), 156.6 (*C*(Ar)), 164.0 (C(5)*C*OO), 176.3 (*C*OC$(CH_3)_3$), 176.5 (*C*OC$(CH_3)_3$) ppm; FT-IR ν_{max}: 2975 (C-H), 1725 (C=O) cm^{-1}; LRMS *m*/z (ESI$^+$) 418 ([M+H]$^+$, 100%); HRMS *m*/z (ESI$^+$) calculated for $C_{22}H_{27}NO_7$$^{+,}$ [M+H]$^+$ = 417.1787; found [M+H]$^+$ = 417.1801.

11.1.3 Peptide synthesis

Synthesis was carried out by standard solid phase peptide synthesis (SPPS) using fluorenylmethyloxycarbonyl (Fmoc) - protected amino acids, 1-hydroxybenzotriazole (HOBt) and diisopropylcarbodiimide (DIC) as activating reagents at a 1:1 ratio, and piperidine as Fmoc deprotection reagent. Peptides were either synthesised using a Multipep peptide synthesis machine (Intavis AG Bioanalytical Instruments, Germany) for screening proposes or the Liberty Blue automated microwave peptide synthesiser (CEM Corporation, USA) for kinetic studies. Peptides used for screening assays were not purified. All syntheses were carried out in dimethylformamide (DMF). Standard N^{α}-Fmoc-protected amino acids amino acids were purchased from Novabiochem or CS Bio, and modified amino acids (such as methylated residues) were purchased from GL Biochem, except for deuterated acetyllysine which was synthesised in the lab (*Roman Belle*). All peptides were synthesised as *C*-terminal amides. Fmoc-protected amino acids were loaded either on a Tentagel S-RAM resin (Multipep) or on a Rink Amide MBHA resin LL as described.[2,3] After synthesis was completed, the resin was removed from the synthesiser, washed 4 x with dichloromethane (DCM) and dried overnight in a vacuum desiccator. Peptide sequences are given in the individual relevant chapters.

11.1.3.1 Peptide cleavage and HPLC purification

Peptide cleavage and HPLC purification were carried out as described by Walport *el al*.[3] Cleavage of the peptides from the resin was achieved by incubation for three hours with trifluoroacetic acid (TFA):triethylsilane (97.5:2.5), which also removed the side-chain protecting groups. The peptide was precipitated with diethyl ether from the cleavage mixture and the resultant solid dissolved in MilliQ water and lyophilized overnight.

Preparative HPLC for peptide purification was carried out using a Jasco LC Net II / ADC HPLC using a Grace/Vydac 218GCC1520 column (C18, 10-15 µM, Prep Guard Cartridge 10x20mm). Peptides were purified using water (0.1% TFA) and acetonitrile (0.1% TFA) as solvents, and the typical purification gradient for peptide purification was 5-20% acetonitrile in water. Peaks were detected at 220 nm and collected automatically, and the mass and purity of the fractions were confirmed by MALDI-TOF MS. Fractions containing purified peptide (> 95% purity) were pooled and lyophilised in 50 mL Falcon tubes. Peptides were stored either as lyophilised solids or as solutions in water at -80°C.

11.2 Biological procedures

11.2.1 General experimental

11.2.1.1 pH measurements

pH measurements were made using a Hanna Instruments HI 9321 microprocessor pH meter fitted with a 5 mm diameter electrode. Before use the probe was calibrated in either the range of pH 4-7 or pH 7-10, as appropriate, using standard calibration solutions (Fisher Scientific) and stored in 4 M KCl solution between uses. pH of solutions was adjusted with stock solutions of HCl or NaOH as necessary.

11.2.1.2 SDS-PAGE gel electrophoresis

NuPAGE® Novex® Bis-Tris Precast gels (4-12%, 1 mm, 15 well, Life Technologies) were used for SDS-PAGE gel electrophoresis. Samples were prepared for loading in NuPAGE® Lithium Dodecyl Sulfate (LDS) Sample Buffer (4 x, Life Technologies), supplemented with 100 µM dithiothreitol (DTT) and run in NuPAGE® MES SDS Running Buffer (20 x, Life Technologies). Coomassie staining was carried out using Pierce Coomassie Brilliant Blue R-250 (Thermo Scientific) according to the manufacturer's protocol.

11.2.1.3 Western blot

Proteins were transferred onto a nitrocellulose membrane (0.2 µM pore size, Invitrogen) using an iBlot Gel Transfer Device (Invitrogen), program P2 for 6 minutes. The membrane was then blocked for 16 h with 50% SeaBlock Reagent (Merck Millipore) in PBS (4°C). After washing (3 x 5 min, PBS-T) primary antibodies (specified in the individual relevant chapters) were added to the membrane and incubated for 1 h (room temperature, shaking). The membrane was washed again (3 x 5 min, PBS-T) and secondary antibodies (specified in the individual relevant chapters) were added before incubating (1 h, room temperature, dark, shaking). Following 3 further washes, the membrane was imaged on an Odyssey® CLx Infrared Imaging System (LI-COR) and analysed using Image Studio™software (LI-COR).

11.2.1.4 Plasmid purification

The plasmid was transformed into *E. coli* XL10 Gold cells (Stratagene). The transformed cells were grown on a LB Agar plate supplemented with the appropriate antibiotics. A single colony from the plate was used to inoculate 10 mL 2TY medium containing appropriate antibiotic and was incubated overnight (220 rpm shaking, 37˚C, 20 h). The plasmid DNA was purified using a Fermentas GeneJET™ Miniprep or Maxiprep kit, according to the manufacturer's protocol.

11.2.2 Plasmids and protein constructs

Plasmids used in this study are described in **Table 11.1**.

Table 11.1 Plasmids used in this study and their origins.

Encoded human protein	Vector	Tag (*N*-terminal)	Antibiotic resistance	Use	Source
KDM4A$_{1–1,064}$	pcDNA3-*N*-FLAG-LIC	FLAG	Ampicillin	Mammalian cell expression	Prof. Yi Zhang, University of North Carolina at Chapel Hill.[4]
H118A catalytic inactive variant of KDM4A$_{1–1,064}$	pcDNA3-*N*-FLAG-LIC	FLAG	Ampicillin	Mammalian cell expression	Prof. Yi Zhang, University of North Carolina at Chapel Hill.[4]
Full length KDM3A$_{1–1321}$	pcDNA3-*N*-FLAG	FLAG	Ampicillin	Mammalian cell expression	Prof. Yi Zhang, University of North Carolina at Chapel Hill.[5]
H1120Y catalytic inactive variant of KDM3A$_{1–1321}$	pcDNA3-*N*-FLAG	FLAG	Ampicillin	Mammalian cell expression	Prof. Yi Zhang, University of North Carolina at Chapel Hill.[5]
KDM3A$_{515–1317}$	pFB-CT10HF-LIC	His$_6$	Ampicillin	Production in baculovirus	Structural Genomics Consortium (SGC), Oxford University.[6]

Human recombinant isolated proteins used in this study are described in **Table 11.2**.

Table 11.2 Recombinant proteins used in this study and their sources.

Protein	Tag (*N*-terminal)	Source
$KDM3A_{515–1317}$	His_6	Professor Christopher J. Schofield (CJS) group, Oxford University.[6]
$KDM3B_{882-1761}$	His_6	SGC, Oxford University.[6]
$JMJD1C_{1760-2540}$	His_6-Glutathione S-transferase (GST)	SGC, Oxford University.[6]
JmjC $JMJD1C_{2157-2497}$	His_6	SGC, Oxford University.[6]
$KDM4A_{1-359}$	His_6	CJS group, Oxford University.[7]
KDM4C	His_6	CJS group, Oxford University.[8]
$KDM6B_{1141–1,590}$ (JMJD3)	His_6	CJS group, Oxford University.[6]
$KDM7B_{1–447}$ (PHF8)	His_6	CJS group, Oxford University.[9]

11.2.3 KDM3A expression and purification

$KDM3A_{515-1317}$ was produced as an *N*-terminally His_6-tagged protein in Sf9 insect cells using the Bac-to-Bac® baculovirus expression system (Invitrogen) based on the procedure described by Rose *et al*.[6] DNA encoding for the catalytic domain of KDM3A, comprising residues T515 through S1317, was cloned into the pFB-CTHF10-LIC vector. The resultant plasmid was transformed into the DH10bac vector. The resulting 'bacmid' was transfected into Sf9 cells grown in Sf-900™ II SFM medium (ThermoFisher scientific) to produce the S11 baculovirus. P2 virus was used to generate protein at a large scale, expression occurring

over 72 hours. Sf9 insect cells were harvested and lysed (50 mM HEPES-KOH pH 7.5, 300 mM KCl, 10 mM imidazole and 5% glycerol, cOmplete™ EDTA - free Protease Inhibitor Cocktail; Sigma-Aldrich, 0.2 mg/mL DNAse I; Roche); homogenate sample was collected. For optimal purity and catalytic activity, the protein was kept at 4°C during purification and the procedure be completed within a day. The lysate clarified by high speed centrifugation (56000xg for 45 minutes), pellet and supernatant samples were collected. The protein was then purified by Ni-Sepharose (GE Healthcare) by batch binding. For protein binding, beads were incubated with lysate at slow rotation for 60 min. Supernatant was removed and beads were loaded into column and washed using 20CV of Wash 1 buffer (50 mM HEPES-KOH pH 7.5, 300 mM KCl, 5% glycerol and 10 mM imidazole), followed by 4CV of Wash 2 buffer (30 mM imidazole). Protein was eluted with 5CV of elution buffer (250mM imidazole). Fractions containing protein were pooled and purified by gel filtration on a HiLoad 16/60 Superdex 200 column (GE Healthcare) equilibrated with 20 mM HEPES-KOH pH 7.5, 300 mM KCl and 5% glycerol. Fractions containing protein from gel filtration were pooled and concentrated to 63 μM (concentration to ~150 μM resulted in protein precipitation), aliquoted, then rapidly frozen in liquid nitrogen and then stored for long term at -80°C.

11.2.4 Mammalian cell culture procedures

11.2.4.1 HeLa cell culture

All procedures were carried out in a Class II Biological Safety Cabiner (Nuaire Nu 437-400E). The human cervical carcinoma HeLa cell line was from the American Type Cultures Collection (ATCC, Manassas, VA) and cultured in Dulbecco's modified Eagle medium (DMEM; Invitrogen Gibco Cell Culture Products, Carlsbad, CA) supplemented with 10% fetal calf serum (FCS, Invitrogen), 1% Glutamax (Invitrogen) 6 and 1% penicillin-streptomycin (Lonza). Cultures were kept at 37°C in a humidified atmosphere of 95% air and 5% CO_2. Cells were detached from the culture flask with a solution of 0.05% trypsin and 0.02% Ethylenediaminetetraacetic acid (EDTA; Sigma) and then washed and suspended in complete culture medium with 10% FCS.

Passaging HeLa cells: The medium was removed from the cells via aspiration and cells were washed with phosphate-buffered saline (PBS). 0.25% Trypsin-ethylenediaminetetraacetic acid (EDTA) was added and the cells incubated (5% CO_2, 37°C)

until cells were released from the surface of the plate (approx. 3 min). Cells were then gently resuspended in fresh medium, before transferred to new plates containing fresh pre-warmed medium.

Cell Storage: Cells were frozen in DMEM containing 20% FBS and 10% DMSO and stored in a -80°C freezer for short term storage (less than 1 year) and in a liquid nitrogen dewar for longer term storage.

Transfection: Transfection was carried out with cells at ~50% confluency as judged using a Motic AE20 (Ted Pella) microscope. Cells were transiently transfected with plasmid DNA using DNA stock solution and Fugene® HD (Promega) diluted in OptiMEM minimal medium (Life Technologies) according to company's manual.[10] DNA for transfection and transfection reagent were separately diluted in OptiMEM® and incubated at room temperature for 5 min. The two reagents were then mixed and left at room temperature for 10 min to give a final DNA concentration of 0.02 μg/μL and a Fugene® HD: DNA ratio of 1:3. The medium on cells in the 96 well plates was replaced and the DNA-Fugene® HD solution (5 μL) added to each appropriate well and mixed thoroughly.

11.2.4.2 Viability analyses in HeLa cells

Antiproliferative activities of compounds were determined using 5-(3-carboxymethoxyphenyl)-2-(4,5-dimethylthiazolyl)-3-(4-sulfophenyl)te trazolium, inner salt (MTS) assay.[11] HeLa cells were seeded into 96-well plates (2000 cells/well) and cultured at 37°C for 24 h to achieve 70% confluency. Subsequently, the medium was replaced with DMEM medium containing the tested compounds in different concentrations of 1-300 μM in 1% DMSO. Staurosporine in 0.03-10 μM final concentration was used as a control for cytotoxicity. After 24 h of treatment, the medium was replaced with CellTiter 96® Aqueous One Solution Reagent (Promega) and incubated for 4 h. CC_{50} values were calculated in Prism 6 software after normalisation against corresponding 1% DMSO treated cells and 1% DMSO in media (no cells) controls.

11.2.4.3 Cellular immunofluorescence demethylase assay

A protocol adapted from that of King *et al.*, 2010 was used for investigation of cellular demethylation inhibition.[12] HeLa cells were maintained in DMEM media supplemented with 10% fetal bovine serum (FBS), penicillin and streptomycin. Cells were transiently transfected with either FLAG-tagged KDM4A or the H188A catalytic inactive variant of KDM4A using Fugene® HD. Inhibition studies were initiated 24 h after cellular transfection and compounds were added to a final concentration of 1-300 μM in 1% DMSO. Cells were washed with PBS and fixed using formaldehyde (4%). H3 K9me3 levels were measured in cells following 24 h incubation with compound. All cells were stained with an anti-FLAG mouse monoclonal antibody (Sigma F1804), rabbit anti-H3 K9me3 (Abcam Ab8898) and DAPI (Sigma D9564) for DNA. Alexa Fluor® 488 antibody (Life technologies A-21121) and Alexa Fluor® 568 antibody (Life technologies A-11011) were used to fluorescently label the FLAG and H3 K9me3 primary antibodies. Image acquisition was conducted using an Operetta High Content Imaging System (PerkinElmer) and image analysis was performed with Harmony High Content Imaging and Analysis Software (PerkinElmer). The nuclei were automatically identified by DAPI staining using the standard parameters. Cells expressing low amounts of the exogenous demethylase, where FLAG-tag staining was dim, were omitted from analyses by setting a minimum threshold. EC_{50} values were calculated in Prism 6 after normalisation by setting DMSO treated KDM4A variant transfected cells to 0% inhibition.

11.2.4.4 Intracellular delivery assay

The protocol used was adapted from that of Kruidenier *et al.*, 2010.[13] HeLa cells were dosed with a 200 μM solution of the compound in DMSO or DMSO alone. After 24 h of incubation the medium was aspirated and the cells were washed with PBS. The number of cells in each sample was determined using a Motic AE20 (Ted Pella) microscope. Extraction of the compounds from the cells was performed by lysis using 80% aqueous methanol. All samples were shaken for 10 min on a vortex mixer and then centrifuged for 10 min at 15000 rpm. The supernatant was removed and the methanol solution was evaporated using a SpeedVac machine. Samples were dissolved with water in proportion to the cell count and an aliquot of the resulting supernatant was mixed with caffeine at 1000 ng/mL as the internal standard. Samples were analysed by reverse phase LC MS/MS using a heat assisted electrospray interface in positive ion mode. The instrument used was a Waters Quattro

Micro triple quadrupole mass spectrometer coupled to liquid chromatography CTC PAL and Waters Acquity UPLC. The column was Gemini C18 HPLC with 3.0 mm internal diameter (Phenomenex). Nominal MRM (Multiple Reaction Monitoring) transitions for analytes were 190.0 to 146.0 for IOX1 **1**, 302.2 to 190.0 for ester **5** and 195.0 to 138.1 for Caffeine. MRM methods were ran over a 5 minutes gradient running from 1% ACN + 0.5% acetic acid (aq) to 90% ACN + 10% formic acid (aq), held for 2 minutes and returned to the starting conditions over 0.1 minutes and remained at the starting conditions for 3 minutes. Samples were assayed against calibration standard curve over the range of 0 to 200 μM prepared with diluted cell lysate of cells dosed with DMSO. Quantitative analyses were made by measuring the peak area of the compounds, normalised using caffeine, followed by linear regression analyses to calculate the number of moles per cell.

11.2.4.5 Real-time qRT-PCR

RNA extraction and reverse transcription: total cellular RNA was isolated from the cells using the RNeasy Mini, RNA isolation kit (Qiagen). The RNA was eluted from the RNeasy Mini columns with 30 μl of RNase-free water. RNA samples were treated with deoxyribonuclease (TurboDNAse, Ambion). The amount of total RNA isolated from the cells was quantified using a NanoDrop spectrophotometer (2100 Bioanalyzer). Samples were then divided into 1500 ng RNA per 100 μl of cDNA synthesis reaction using the TaqMan Archival Reverse Transcription Kit (Applied Biosystems). To minimise potential variability, the cDNA for all RNA samples has been generated in the same reverse transcriptase reaction mixture. Genomic DNA contamination was assessed by real-time qRT-PCR, using mixes without the reverse transcriptase enzyme.

Primers design and efficiency measurements: primers were either designed according to existing validated primers published in PrimerBank, or by using Primer-BLAST (NCBI).[14,15]

Oligonucleotides were purchased from Sigma. cDNA from all samples to be tested in the experiment was mixed in equal amounts to be used as a template for primer efficiency measurements. PCR efficiency and dynamic range for the primers were determined by running real-time PCRs on serial dilutions of the cDNA template. Primers sequences and efficiency data are described in the individual relevant chapters.

Quantitative RT-PCR: Real-time quantitative RT-PCR was performed using SYBR Green chemistry.[16] For each reaction, 2x SYBR Green Quantitect mastermix (Qiagen) was added to 900 nM forward and reverse primer. cDNA template was added to 8 ng per PCR reaction. Reaction mixes were aliquoted in a volume of 40 µl/well onto a 96-well optical reaction plate (Applied Biosystems) in technical triplicates. The reaction plate was then transferred to the CFX96 Touch Real-Time PCR detection system (BIO-RAD). The following cycling conditions were used: 50°C for 2 min (stage1); 95°C for 10 min (stage 2); 95°C for 15 s; 60°C for 1 min (stage 3) X 40. The specificity of the PCR primers was checked using dissociation curve analysis at the end of each run. Fold change in mRNA expression was determined by the Pfaffl method.[17] Results were normalised to a reference gene and a selected calibrator sample which was arbitrarily set to 100%. All experiments were repeated three independent times (three biological repeats). Statistical analyses were carried out using one way ANOVA followed up by Dunnett's multiple comparisons tests using GraphPad Prism 6.

11.2.4.6 On-bead demethylation assays

A protocol adapted from that of Walport *et al.* was used.[3] HEK293T cells were transiently transfected with KDM3A. Cells were washed in PBS buffer and lysed in 50mM HEPES, pH 7.5, 150mM NaCl, 0.5% NP40, 1% EDTA-free protease inhibitor cocktail and DNAse (10 mg ml^{-1}). Lysate was diluted five-fold with buffer containing 50mM HEPES, pH 7.5, 150mM NaCl before being incubated with 20 ml anti-FLAG M2 Magnetic Beads (Sigma-Aldrich) for 3 h at 4°C. Beads were washed four times with lysis buffer (500 ml) and two times with 50mM HEPES (500 ml; or 50mM HEPES) before MS analysis. Beads were resuspended in a final volume of 30 ml for use in activity assays. Demethylation reactions were conducted with 2 ml of the immobilized anti-FLAG bead suspension with peptide (10 mM) and cofactors/substrates. The demethylation products were detected by MALDI MS (conditions given in Section 11.3.3). Western blotting of lysate samples was carried out to confirm exogenous overexpression of all enzymes (conditions given in Section 11.2.1.3).

11.3 Biochemical assays

11.3.1 Test of compound stability in AlphaScreen assay buffer

3 mM of the tested compounds were incubated in a solution composed of the tested buffer for AlphaScreen assay (50 mM HEPES pH 7.5 supplemented with 0.1% BSA, 0.01% Tween20, Fe^{II} (10 µM), ascorbate (100µM) and 2OG (10 µM)) in room temperature. Samples were taken for LCMS analysis after 0, 1, 2 and 24 h of incubation. LCMS retention times (t_r) are quoted to the nearest 0.1 min. LCMS was performed on a WATERS Sunfire equipped with a C18 column (150 mm x 4.6 mm, 5 µm) using a linear gradient of solvent A (water + 0.1% CF_3CO_2H) and solvent B (acetonitrile + 0.1% CF_3CO_2H), eluting at a flow rate of 1 mL/min and monitoring at 254 nm: 2% B over 2 min, 2% B to 100% B over 16 min and 100% B over 2 min.

11.3.2 Differential scanning fluorimetry (DSF) assay

A protocol adapted from that of Niesen *et al.* was used.[18] Samples were loaded in triplicate into a white 96-well PCR plate (Bio-Rad) with each well containing a final volume of 20 µl. The PCR plates were sealed with optical quality sealing tape (Bio-Rad). The concentration of KDM3A in each well was 2 µM with 50 µM $MnCl_2{\cdot}4H_2O$ and 2X SYPRO Orange (Invitrogen) in 200 mM Hepes buffer, pH 7.5. TCAI metabolites were used at a concentration of 10 mM. Peptides were used at a concentration of 200 µM. DSF experiments were carried out using MiniOpticon Real-Time PCR System (Bio-Rad). The samples were heated from 25 to 80°C at the rate of 1°C/minute. A single fluorescence measurement was taken each minute. Melting temperatures were determined by performing a curve fit to the Boltzmann equation (**Equation 11.1**).

$$y = LL + \frac{(UL - LL)}{1 + exp\left(\frac{T_m - x}{a}\right)}$$

Equation 11.1 The Boltzmann equation. LL and UL are the values of minimum and maximum intensities, respectively, and a denotes the slope of the curve within T_m.[18]

The degree of thermal shift was calculated by comparing the melting temperature of KDM3A in the presence of compounds against negative controls with no enzyme. 200 μM H3(1-15)K9me2 peptide fragment was used as a positive control for binding.

11.3.3 MALDI-TOF MS based activity assays

Conditions for MALDI-TOF MS activity assay were carried out as reported by Walport *et al.*[3] All reagents were prepared in 50 mM HEPES buffer pH 7.5 unless otherwise stated, except the ferrous iron, which was prepared as a stock solution (100 mM) in 20 mM aqueous HCl. This stock was then diluted in water. Standard co-factors concentrations used in assays are described in **Table 11.3**, unless otherwise stated. Enzyme and substrate mixes were prepared separately and combined into sample-wells of a 96-well plate to a final volume of 30 μl to initiate the enzymatic catalysis, which was carried out at 37°C.

Table 11.3 Standard concentrations of components in MALDI-TOF MS activity assay. The following conditions were used to measure catalytic activity using MALDI-TOF MS, unless otherwise stated.

	Component	Concentration in reaction
Enzyme mix	Enzyme	varied (single-digit μM)
	Fe^{II}	50 μM
	Ascorbate	100 μM
Substrate mix	Peptide	varied (5-500 μM)
	2OG	100 μM

For potential novel substrate screening reactions, 10 μM of the relevant peptides was used. The reactions were incubated for 1 h, unless otherwise stated. To test for inhibition, the enzyme was pre-incubated with an inhibitor for 15 min at 37°C before initiation of the reaction with all other reagents. Where necessary for inhibitor solubilisation, 1% DMSO was included in the assay buffer. Inhibition assays were carried with varied concentrations of the tested inhibitors together with methylated H3 fragment substrate peptide, and IC_{50} values were calculated from the reaction rate during the linear phase of catalysis. For kinetic studies, all samples containing the various substrate concentrations were kept under the

same conditions and measured simultaneously for each time-point during the time-course experiment.

Reactions were quenched with 1% formic acid. A 96-spot MALDI plate was spotted with 1 µL of α-cyano-4-hydroxycinnamic acid (CHCA MALDI matrix) and 1 µL of sample per spot. Samples were left to air dry before being analysed by MALDI-TOF MS.

Samples were analysed using a MALDI-TOF mass spectrometer (Waters MALDI Micro MX) in the positive ion reflectron mode. A laser energy of 200-250, flight tube voltage 12000 V, reflectron voltage 5200 V. The mass window collected was determined based upon the peptides in the sample. Spectra were processed (background subtraction: polynomial order 3, below curve (%) 10, tolerance 0.010; smoothing: smooth window (channels) ±3, number of smooths 5, Savitzky Golay method) using MASSLYNX 4.0 (Waters). Note that in some cases small variations (≤ 2 Da) in the absolute experimentally observed values compared with the calculated values of the peptides were observed. However, in all cases the mass shift between the product and the substrate masses was unchanged.

The percentage of enzymatic activity was calculated using **Equation 11.2** and was normalised based on control reactions with no enzyme.

$$\%\ Catalytic\ activity = \frac{\sum_{i=0}^{n} P}{\sum_{i=0}^{n} P + S}\ x\ 100$$

Equation 11.2 Calculation of the percentage of enzymatic activity based on MALDI-TOF MS analyses. P is the relative intensity of the peak of the product peptide and S is the relative intensity of the peak of the remaining substrate peptide in the sample. In cases where several products were observed, their sum of intensities is used to calculate the percentage of catalytic activity.

All data were processed with GraphPad Prism 6, fitting either dose-response curves with variable Hill slope (for IC_{50} determination) or Michaelis-Menten curves (for kinetics). In all assays, non-enzyme containing negative controls were used for data normalisation. A methylated H3 fragment peptide was used as a positive control for KDM catalysis. All measurements were carried out in triplicate unless otherwise stated.

11.3.4 Formaldehyde dehydrogenase (FDH) coupled assay

KDM3A assays were carried out as reported at 37°C.[3] All reagents were prepared in 50 mM HEPES buffer pH 7.5 unless otherwise stated, except the ferrous iron, which was prepared as a stock solution (100 mM) in 20 mM aqueous HCl. This stock was then diluted in water. Standard co-factors concentrations used in assays are described in **Table 11.4**, unless otherwise stated. To test for inhibition, enzyme was pre-incubated with inhibitor for 15 min before initiation of the reaction with all other reagents. Where necessary for inhibitor solubilisation, 1% DMSO was included in the assay buffer. Enzyme and substrate mixes were prepared separately and combined to initiate catalysis in clear-bottom black 384-well plates (Greiner Bio-One).

Table 11.4 Standard concentration of components in KDM3A the FDH coupled assay. The following conditions were used to measure KDM3A catalytic activity using the FDH coupled assay throughout the study unless otherwise stated.

	Component	Concentration in reaction
Enzyme mix	KDM3A	varied (hundrents nM)
	Fe^{II}	50 µM
	Ascorbate	100 µM
	FDH	0.5 U/50 µl reaction
Substrate mix	Peptide	varied (5-500 µM)
	2OG	100 µM
	NAD^+	500 µM

Production of NADH was fluorescently monitored over 60 cycles with a cycle time of 30 seconds using a PHERAstar FS (BMG Labtech) plate reader with 355 nm excitation and 460nm emission. Formaldehyde standard curve was used to convert fluorescence to demethylated product concentration. The linear range for the KDM reaction was used to calculate initial rates, as determined by the slope of the fluorescence signal recorded. All data were processed with GraphPad Prism 6, fitting either dose-response curves with variable Hill slope (for IC_{50} determination) or Michaelis-Menten curves (for kinetics). In all assays, non-enzyme containing negative controls were used for data normalisation. A

methylated H3 fragment peptide was used as a positive control for KDM catalysis. All measurements were carried out in triplicate unless otherwise stated.

11.4 References

1. Rotili, D. *et al.* A photoreactive small-molecule probe for 2-oxoglutarate oxygenases. *Chem. Biol.* **18,** 642–54 (2011).

2. Ge, W. *et al.* Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat. Chem. Biol.* **8,** 960–2 (2012).

3. Walport, L. J. *et al.* Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat. Commun.* **7,** 11974 (2016).

4. Klose, R. J. *et al.* The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* **442,** 312–316 (2006).

5. Yamane, K. *et al.* JHDM2A, a JmjC-Containing H3K9 Demethylase, Facilitates Transcription Activation by Androgen Receptor. *Cell* **125,** 483–495 (2006).

6. Rose, N. R. *et al.* Plant growth regulator daminozide is a selective inhibitor of human KDM2/7 histone demethylases. *J. Med. Chem.* **55,** 6639–43 (2012).

7. Ng, S. S. *et al.* Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **448,** 87–91 (2007).

8. Chowdhury, R. *et al.* The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* **12,** 463–9 (2011).

9. Loenarz, C. *et al.* PHF8, a gene associated with cleft lip/palate and mental retardation, encodes for an Nepsilon-dimethyl lysine demethylase. *Hum. Mol. Genet.* **19,** 217–22 (2010).

10. Promega. FuGENE® HD Transfection Reagent Protocol. (2013). Available at: https://www.promega.co.uk/resources/protocols/technical-manuals/101/fugene-hd-transfection-reagent-protocol/.

11. Barltrop, J. A., Owen, T. C., Cory, A. H. & Cory, J. G. 5-(3-carboxymethoxyphenyl)-2-(4,5-dimethylthiazolyl)-3-(4-sulfophenyl)te trazolium, inner salt (MTS) and related analogs of 3-(4,5-dimethylthiazolyl)-2,5-diphenyltetrazolium bromide (MTT) reducing to purple water-soluble formazans As cell-viability indica. *Bioorganic Med. Chem. Lett.* **1,** 611–614 (1991).

12. King, O. N. F. *et al.* Quantitative high-throughput screening identifies 8-hydroxyquinolines as cell-active histone demethylase inhibitors. *PLoS One* **5,** e15535 (2010).

13. Kruidenier, L. *et al.* A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* **488,** 404–8 (2012).

14. Pedersen, L. B., Veland, I. R., Schrøder, J. M. & Christensen, S. T. Assembly of primary cilia. *Dev. Dyn.* **237,** 1993–2006 (2008).

15. Deane, J. A., Cole, D. G., Seeley, E. S., Diener, D. R. & Rosenbaum, J. L. Localization of intraflagellar transport protein IFT52 identifies basal body transitional fibers as the docking site for IFT particles. *Curr. Biol.* **11,** 1586–1590 (2001).

16. Technologies, L. SYBR ® Select Master Mix User Guide. (2013).

17. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29,** 45e–45 (2001).

18. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2,** 2212–21 (2007).

Appendix 1

Mascot Search Results for SDS-PAGE gel bands corresponding to KDM3A (see Chapter 3, Section 3.4).

Analysis by Carina Giladi, Oppermann group and Dr. Rod Chalk, Burgess-Brown group, Structural Genomics Consortium, Oxford

Mascot Search Results for the ~115 kDa band

Search title : Auto submitted by BioTools
Database : SwissProt 2012_09 (538010 sequences; 190998508 residues)
Timestamp : 26 Jun 2014 at 16:52:32 GMT
Protein hits : KDM3A_HUMAN Lysine-specific demethylase 3A OS=Homo sapiens GN=KDM3A PE=1 SV=4
KDM3A_MOUSE Lysine-specific demethylase 3A OS=Mus musculus GN=Kdm3a PE=1 SV=1
DNAA_STREM Chromosomal replication initiator protein DnaA OS=Streptococcus equi subsp. zooepidemicus (strain MGCS10565) GN=dnaA PE=3 SV=1
RPOD_BORBU RNA polymerase sigma factor rpoD OS=Borrelia burgdorferi (strain ATCC 35210 / B31 / CIP 102532 / DSM 4680) GN=rpoD PE=3 SV=2
FARS_ARTAN (E)-beta-farnesene synthase OS=Artemisia annua GN=CASC125 PE=1 SV=1
ATP8_APILI ATP synthase protein 8 OS=Apis mellifera ligustica GN=mt:ATPase8 PE=3 SV=1
BIOB_RHOPB Biotin synthase OS=Rhodopseudomonas palustris (strain BisB18) GN=bioB PE=3 SV=1
MED30_AEDAE Mediator of RNA polymerase II transcription subunit 30 OS=Aedes aegypti GN=MED30 PE=3 SV=1
TBP_METKA TATA-box-binding protein OS=Methanopyrus kandleri (strain AV19 / DSM 6324 / JCM 9639 / NBRC 100938) GN=tbp PE=3 SV=1
NFI_DICT6 Endonuclease V OS=Dictyoglomus thermophilum (strain ATCC 35947 / DSM 3960 / H-6-12) GN=nfi PE=3 SV=1
KIF4B_HUMAN Chromosome-associated kinesin KIF4B OS=Homo sapiens GN=KIF4B PE=1 SV=2
DADA1_RHILO D-amino acid dehydrogenase 1 small subunit OS=Rhizobium loti (strain MAFF303099) GN=dadA1 PE=3 SV=1
SYCP1_MESAU Synaptonemal complex protein 1 (Fragment) OS=Mesocricetus auratus GN=SCP1 PE=2 SV=1
BCCIP_CAEBR Protein BCCIP homolog OS=Caenorhabditis briggsae GN=CBG11215 PE=3 SV=1
L_MMVR RNA-directed RNA polymerase L OS=Maize mosaic virus (isolate Maize/United States/Reed/2005) GN=L PE=3 SV=1
COI1_ARATH Coronatine-insensitive protein 1 OS=Arabidopsis thaliana GN=COI1 PE=1 SV=1
TBD2B_XENTR TBC1 domain family member 2B OS=Xenopus tropicalis GN=tbc1d2b PE=2 SV=1
DDX49_DROME Probable ATP-dependent RNA helicase Dbp45A OS=Drosophila melanogaster GN=Dbp45A PE=2 SV=2
PRSA2_BACCR Foldase protein PrsA 2 OS=Bacillus cereus (strain ATCC 14579 / DSM 31) GN=prsA2 PE=3 SV=1
PUR7_NAUPA Phosphoribosylaminoimidazole-succinocarboxamide synthase OS=Nautilia profundicola (strain ATCC BAA-1463 /

Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 52 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a nonprobabilistic basis for ranking protein hits.

Peptide Summary Report
Significance threshold p< 0.05 Max. number of hits 20
1. **KDM3A_HUMAN** Mass: 149472 Score: 466 Matches: 17(10) Sequences: 8(5) emPAI: 0.10
Lysine-specific demethylase 3A OS=Homo sapiens GN=KDM3A PE=1 SV=4

Mascot Search Results for the ~70 kDa band

Search title : Auto submitted by BioTools
Database : SwissProt 2012_09 (538,010 sequences; 190,998,508 residues)
Timestamp : 26 Jun 2014 at 17:27:26 GMT
Type of search : MS/MS Ion Search
Enzyme : Trypsin
Fixed modifications : Carbamidomethyl (C)
Variable modifications : Oxidation (M)
Mass values : Monoisotopic
Protein mass : Unrestricted
Peptide mass tolerance : ± 1.5 Da
Fragment mass tolerance : ± 1.3 Da
Max missed cleavages : 4
Instrument type : ESITRAP
Number of queries : 306
Score distribution
Peptide score distribution. Ions score is –10log(P), where P is the probability that the observed match is a random event. Individual ions scores > 53 indicate identity or extensive homology (p<0.05).
[Deprecated] Score distribution for family members in the first 50 proteins. Protein scores are derived from ions scores as a nonprobabilistic basis for ranking protein families.
Protein families 1–2 (out of 2)

1
Accession Score Description
1 **KDM3A_HUMAN** 133 Lysinespecific
demethylase 3A OS=Homo sapiens GN=KDM3A PE=1 SV=4
2
Accession Score Description
1 ATP8_APILI 74 ATP synthase protein 8 OS=Apis mellifera ligustica GN=mt:ATPase8 PE=3 SV=1

Mascot Search Results for the ~60 kDa band

Search title : Auto submitted by BioTools
Database : SwissProt 2012_09 (538010 sequences; 190998508 residues)
Timestamp : 3 Jul 2014 at 05:51:07 GMT
Protein hits : **KDM3A_HUMAN** Lysine-specific demethylase 3A OS=Homo sapiens GN=KDM3A PE=1 SV=4
BIOB_RHOPB Biotin synthase OS=Rhodopseudomonas palustris (strain BisB18) GN=bioB PE=3 SV=1
RH27_ARATH DEAD-box ATP-dependent RNA helicase 27 OS=Arabidopsis thaliana GN=RH27 PE=2 SV=2
MUTS_RICAE DNA mismatch repair protein MutS OS=Rickettsia africae (strain ESF-5) GN=mutS PE=3 SV=1
COI1_ARATH Coronatine-insensitive protein 1 OS=Arabidopsis thaliana GN=COI1 PE=1 SV=1
RTC5_CANTT Restriction of telomere capping protein 5 OS=Candida tropicalis (strain ATCC MYA-3404 / T1) GN=RTC5 PE=3 SV=1
RS10_MYCGA 30S ribosomal protein S10 OS=Mycoplasma gallisepticum (strain R(low / passage 15 / clone 2)) GN=rpsJ PE=3 SV=2
ATPB_GEOLS ATP synthase subunit beta OS=Geobacter lovleyi (strain ATCC BAA-1151 / DSM 17278 / SZ) GN=atpD PE=3 SV=1
DNLI_PYRKO DNA ligase OS=Pyrococcus kodakaraensis (strain ATCC BAA-918 / JCM 12380 / KOD1) GN=lig PE=1 SV=2
ATPB_DESHD ATP synthase subunit beta OS=Desulfitobacterium hafniense (strain DCB-2 / DSM 10664) GN=atpD PE=3 SV=1
BYST_NEMVE Bystin OS=Nematostella vectensis GN=bysl PE=3 SV=1
POLY_DROME Retrovirus-related Pol polyprotein from transposon gypsy OS=Drosophila melanogaster GN=pol PE=4 SV=1
PLD5_MOUSE Inactive phospholipase D5 OS=Mus musculus GN=Pld5 PE=2 SV=1
RNH2_VIBC3 Ribonuclease HII OS=Vibrio cholerae serotype O1 (strain ATCC 39541 / Ogawa 395 / O395) GN=rnhB PE=3 SV=1
PHCB_SPIPL C-phycocyanin beta chain OS=Spirulina platensis GN=cpcB PE=1 SV=2
GCSPB_PROA2 Probable glycine dehydrogenase [decarboxylating] subunit 2 OS=Prosthecochloris aestuarii (strain DSM 271 / SK 413) GN=gcvPB PE=3 Y4102_ARATH Putative uncharacterized protein At4g01020, chloroplastic OS=Arabidopsis thaliana GN=At4g01020 PE=4 SV=1
ATPB_SPHAL ATP synthase subunit beta OS=Sphingopyxis alaskensis (strain DSM 13593 / LMG 18877 / RB2256) GN=atpD PE=3 SV=1
SYT_LACLA Threonine--tRNA ligase OS=Lactococcus lactis subsp. lactis (strain IL1403) GN=thrS PE=3 SV=1
NOG1_MOUSE Nucleolar GTP-binding protein 1 OS=Mus musculus GN=Gtpbp4 PE=2 SV=3
Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 53 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a nonprobabilistic basis for ranking protein hits.
Peptide Summary Report
Significance threshold p< 0.05 Max. number of hits 20
Standard scoring MudPIT scoring Ions score or expect cutoff

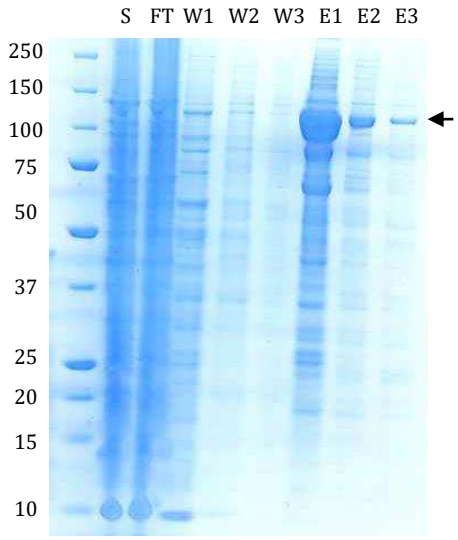
1. **KDM3A_HUMAN** Mass: 149472 Score: 292 Matches: 13(7) Sequences: 5(3) emPAI: 0.06
Lysine-specific demethylase 3A OS=Homo sapiens GN=KDM3A PE=1 SV=4
Check to include this hit in error tolerant search or archive report

Appendix 2

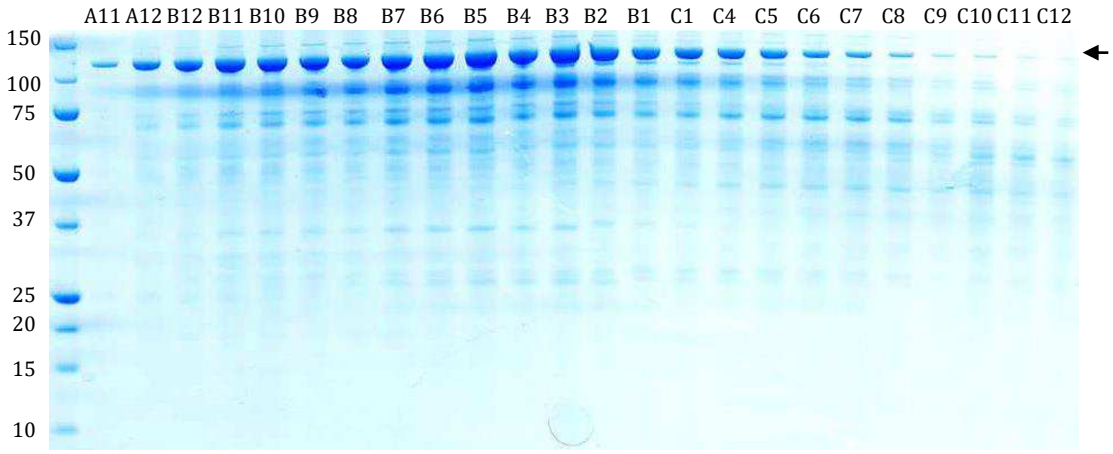
Analysis of JMJD1C protein purification (see Chapter 8, Section 8.3)

Protein was produced by Dr. Catrine Johansson

(A)



(B)



(C)

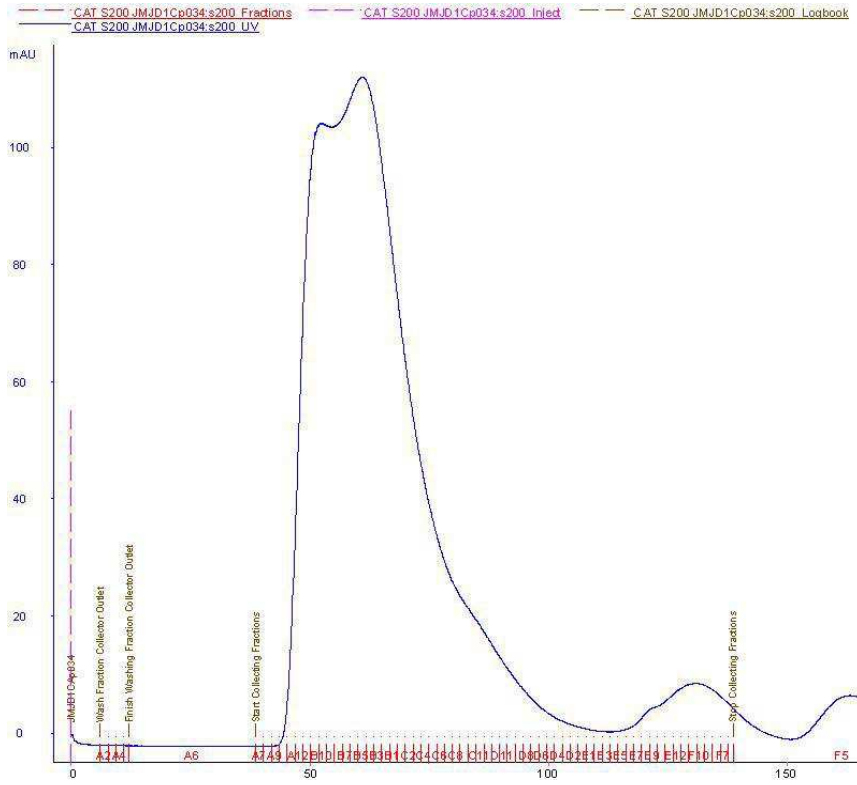


Figure A2.1 Analysis of JMJD1C protein samples during purification. (A) Coomassie staining of SDS-PAGE protein gel following fractions from Ni-Sepharose purification. S denotes the supernatant, FT the flow-through, W1- wash with 10 mM Imidazole, W2 - wash with 20 mM Imidazole, W3 - wash with 30 mM Imidazole, E1 – elution fraction 1 with 250 mM Imidazole, E2 – elution fraction 2 with 250 mM Imidazole, E3 – elution fraction 3 with 500 mM Imidazole. (B) Coomassie staining of SDS-PAGE protein gel of fractions from gel filtration purification on a HiLoad 16/60 Superdex 200 column (GE Healthcare). A letter and a number indicate the fraction name loaded in each lane. Numbers on the left of the gels present the molecular weight of each standard in KDa. Arrows on the right indicate the bands representing the JMJD1C protein (117 KDa). (C) Gel filtration chromatogram of JMJD1C purification. *Protein was produced by Dr. Catrine Johansson.*

Appendix 3

Reported sites of a methyllysine modification at position -5 relative to an acetyllysine (see Chapter 10, Section 10.3).

Table A3.1 PhosphoSitePlus data of reported sites of a methyllysine modification at position -5 relative to an acetyllysine across the proteome.[1] The following reported acetyllysine marks can be screened as candidate novel substrates of KDM3A.

Main GO-molecular function	Protein name	Protein	Organism	Uniprot	Methyl position	Acetyl position
CCR7 chemokine receptor binding	C-C motif chemokine 21	CCL21	human	O00585	115	120
protein binding involved in protein folding	T-complex protein 1 subunit zeta	CCT6A	human	P40227	5	10
ATP-dependent helicase activity	ATP-dependent RNA helicase DDX1	DDX1	human	Q92499	234	239
pyruvate dehydrogenase (NAD+) activity	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial	DLAT	mouse	Q8BMF4	358	363
tRNA binding	Elongation factor 1-alpha 1	EEF1A1	human	P68104	36	41
tRNA binding	Elongation factor 1-alpha 1	EEF1A1	human	P68104	79	84
translation factor activity, RNA binding	Elongation factor 1-alpha 2	EEF1A2	human	Q05639	36	41
structural constituent of cytoskeleton	Protein 4.1	EPB41	human	P11171	361	366
rRNA methyltransferase activity	pre-rRNA processing protein FTSJ3	FTSJ3	human	Q8IY81	788	793
misfolded protein binding	78 kDa glucose-regulated protein	GRP78	human	P11021	596	601
chromatin DNA binding	Histone H1.4	H1E	mouse	P43274	21	26
DNA binding	Histone H2B type 1-B	H2B	human	P33778	15	20

Main GO-molecular function	Protein name	Protein	Organism	Uniprot	Methyl position	Acetyl position
DNA binding	Histone H2B type 1-B	H2B	human	P33778	23	28
DNA binding	Histone H2B type 1-B	H2B	human	P33778	120	125
DNA binding	Histone H2B type 1-A	H2B1A	human	Q96A08	13	18
DNA binding	Histone H3.1	H3	mouse	P68433	4	9
nucleosomal DNA binding	Histone H3.3	H3	cow	Q5E9F8	4	9
DNA binding	Histone H3.2	H3	chicken	P84229	4	9
DNA binding	Histone H3.1	H3	mouse	P68433	9	14
DNA binding	Histone H3.1	H3	mouse	P68433	18	23
DNA binding	Histone H3.1	H3.1	human	P68431	4	9
DNA binding	Histone H3.1	H3.1	human	P68431	9	14
DNA binding	Histone H3.1	H3.1	human	P68431	18	23
nucleosomal DNA binding	Histone H3.3	H3F3A	human	P84243	4	9
nucleosomal DNA binding	Histone H3.3	H3F3A	human	P84243	9	14
nucleosomal DNA binding	Histone H3.3	H3F3A	human	P84243	18	23
nucleosomal DNA binding	Histone H3.3C	H3F3C	human	Q6NXT2	5	10
nucleosomal DNA binding	Histone H3.3C	H3F3C	human	Q6NXT2	10	15
nucleosomal DNA binding	Histone H3.3C	H3F3C	human	Q6NXT2	19	24
chromatin binding	Histone H3.2	HIST2H3 A/C/D	human	Q71DI3	5	10
chromatin binding	Histone H3.2	HIST2H3 A/C/D	human	Q71DI3	10	15
chromatin binding	Histone H3.2	HIST2H3 A/C/D	human	Q71DI3	19	24
DNA binding	Histone H3.1t	HIST3H3	human	Q16695	5	10
DNA binding	Histone H3.1t	HIST3H3	human	Q16695	10	15
DNA binding	High mobility group protein HMG-I/HMG-Y	HMGA1	human	P17096	62	67
RNA binding	Heterogeneous nuclear ribonucleoprotein A1	hnRNP A1	human	P09651	161	166
microtubule motor activity	Kinesin-like protein KIF3C	KIF3C	human	O14782	163	168
DNA binding	DNA/RNA-binding protein KIN17	KIN	human	O60870	306	311
ATP-dependent DNA	X-ray repair cross-	Ku80	human	P13010	660	665

Main GO-molecular function	Protein name	Protein	Organism	Uniprot	Methyl position	Acetyl position
helicase activity	complementing protein 5					
chromatin DNA binding	Mediator of RNA polymerase II transcription subunit 1	MED1	human	Q15648	1006	1011
actin binding	Myosin-9	MYH9	human	P35579	651	656
chromatin DNA binding	Transcription factor p65	NFkB-p65	human	Q04206	310	315
ADP-ribose diphosphatase activity	ADP-sugar pyrophosphatase	NUDT5	human	Q9UKK9	205	210
DNA binding	POU domain, class 5, transcription factor 1	Oct4	human	Q01860	151	156
RNA binding	RNA-binding protein 7	RBM7	human	Q9Y580	45	50
calcium ion binding	Reticulocalbin-1	RCN1	human	Q15293	81	86
RNA binding	60S ribosomal protein L7	RPL7	human	P18124	156	161
RNA binding	Ubiquitin-40S ribosomal protein S27a	RPS27A	human	P62979	6	11
GTPase activator activity	SH3 domain-binding protein 1	SH3BP1	human	Q9Y3L3	142	147
chromatin binding	Serine hydroxymethyltransferase, mitochondrial	SHMT2	human	P34897	469	474
myosin light chain kinase activity	Myosin light chain kinase, smooth muscle	smMLCK	human	Q15746	603	608
cadherin binding	Sorting nexin-2	SNX2	human	O60749	442	447
methylated histone binding	Spindlin-1	SPIN1	human	Q9Y657	2	7
DNA binding	Nuclear transition protein 2	TP2	rat	P11101	83	88
DNA binding	Nuclear transition protein 2	TP2	rat	P11101	88	93
structural constituent of ribosome	Ubiquitin-60S ribosomal protein L40	UBA52	human	P62987	6	11
	Polyubiquitin-B	UBB	human	P0CG47	6	11
	Polyubiquitin-B	UBB	human	P0CG47	82	87
	Polyubiquitin-B	UBB	human	P0CG47	158	163
RNA binding	Polyubiquitin-C	UBC	human	P0CG48	6	11
RNA binding	Polyubiquitin-C	UBC	human	P0CG48	82	87
RNA binding	Polyubiquitin-C	UBC	human	P0CG48	158	163
RNA binding	Polyubiquitin-C	UBC	human	P0CG48	234	239
RNA binding	Polyubiquitin-C	UBC	human	P0CG48	310	315
RNA binding	Polyubiquitin-C	UBC	human	P0CG48	386	391
RNA binding	Polyubiquitin-C	UBC	human	P0CG48	462	467

Main GO-molecular function	Protein name	Protein	Organism	Uniprot	Methyl position	Acetyl position
RNA binding	Polyubiquitin-C	UBC	human	P0CG48	538	543
RNA binding	Polyubiquitin-C	UBC	human	P0CG48	614	619
DNA binding	Zinc finger protein 484	ZNF484	human	Q5JVG2	315	320

1. Hornbeck, P. V. et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. **43**, D512–D520 (2015).