Histone Lysine and DNA Methylation: Dynamic Marks in the Chromatin

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



Roman Belle

St Cross College

Trinity term 2019



Abstract

Epigenetics concerns the regulation of genes through heritable dynamic modifications. Part of the epigenetic regulation is the methylation of DNA and histones. DNA methylation is introduced by DNA methyltransferases to give 5-methylcytosine (^{5m}C) which is involved in both transcriptional activation and silencing depending on the location in the genome. ^{5m}C can be oxidized by 2OG-dependent Ten-Eleven Translocation (TETs) in sequential order to give 5-hydroxymethylcytosine (^{5hm}C), 5-formylcytosine (^{5f}C) and 5-carboxylcytosine (^{5ca}C).[1] The 5-oxidizedcytosines (^{5ox}C) are linked to an active readout by chromatin binding modules participating in signalling pathways, nucleosome stabilisation and active demethylation. Abnormal epigenetic patterns caused by TET proteins are linked to multiple diseases such as myelodysplastic syndrome (MDS), acute myeloid leukaemia (AML) and solid tumours making them of scientific interest.[2–4]

In Chapter 2, biochemical activity assays based on AlphaScreen™ luminescence, isothermal titration calorimetry (ITC), MALDI-TOF and RapidFire mass spectrometry (RF-MS) were developed to characterise TET enzymes. Steady state kinetics on 2OG co-factor using AlphaScreen™ provides evidence for changes in hTET2 K_M (2OG) dependent on the low complexity insert (LCI). ITC enabled enzyme kinetics on single stranded DNA substrate providing a K_M (23.8 ± 1.7 µM) while MALDI-TOF gave a double stranded K_M (1.02 ± 0.23 µM) revealing a likely difference in hTET substrate preference. Investigations using RF-MS with hTET2$_{CD\Delta LCI}$ and ^{5m}C DNA substrate enabled continuous monitoring of ^{5ox}C production. The biochemical parameters of TET were then applied to an inhibitor screening using AlphaScreen™ and RF-MS. Clinically relevant 2OG-dependent oxygenase inhibitors IOX4 and Vadadustat were identified to inhibit TET proteins. Furthermore, oncometabolite D-2HG was observed to inhibit hTET2 ($IC_{50:}$ ~16 µM) but substantially higher for TET1 (IC_{50}: 417 µM), demonstrating the different sensitivities of the hTET enzymes to oncometabolite concentrations. In summary, the work presented here describes the biochemical characterisation of hTET enzymes as well as robust and sensitive assay methods for high throughput inhibitor screening. Starting inhibitor scaffolds for the TET proteins were identified, which can aid to the development of new potent inhibitors and chemical probes.

Beside DNA methylation, also posttranslational modifications (PTMs) on histones contribute to epigenetic changes. *N*$^{\varepsilon}$-lysine methylation is among the most frequently observed histone PTMs and is introduced by histone lysine methyltransferases (KMTs) and removed by histone lysine demethylases (KDMs).[5] Chromatin binding modules are associated to lysine methylation and provide the active readout enabling epigenetic signalling.[6] The transcriptional implication of *N*$^{\varepsilon}$-lysine methylation is dependent on the site of histone methylation and can function both transcriptional silencing as well as activating. Aberrant

methylation patterns by silencing or overexpression of KMTs, KDMs or chromatin binding modules can result in disease.[7] KDM5A, a histone H3K4me$_{3-1}$ demethylase, is of scientific interests as its linked to Acute Megakaryoblastic Leukaemia (AMKL), gastric cancer and breastcancer.[8–10]

In Chapter 3, the H3K4 methylation states with respect to histone demethylase KDM5A is explored in detail. KDM5s have multiple methyl-lysine binding PHD-fingers, in addition to the catalytic JmjC-domain. To understand the function of the PHD3 in relation to the catalytic activity of KDM5A, histone H3 peptide-based molecules consisting of *N*$^{\varepsilon}$-triethyl-lysine (H3K4et$_3$) and *N*$^{\varepsilon}$-piperidine-lysine (H3K4pip) were produced. ITC measurements demonstrate that H3K4et$_3$ and H3K4pip bind as strongly as H3K4me$_3$ to PHD3 but much weaker to PHD1. MALDI-TOF activity assays of KDM5A-D using H3K4et$_3$ reveal it is a poor substrate for the JmjC-domain and H3K4et$_3$ has little effect on catalytic activity when considered as inhibitor. H3K4pip is a substrate for KDMs and is likely oxidized on the piperidine ring by KDM5s. KDM5 family members show different levels of catalytic activity towards H3K4pip suggesting different JmjC-domain architectures.

In Chapter 4, the recognition pockets of *N*$^{\varepsilon}$-trimethyl-lysine chromatin binding modules are explored using sterically enlarged *N*$^{\varepsilon}$-trimethyl-lysines on histone peptides. Histone peptide variants were produced and subjected to ITC analysis against a panel of chromatin binding modules. Increased alkylation revealed relatively similar binding to H3K4et$_3$ for SPIN2B$_{TTD}$, KDM4A$_{TTD}$, SGF29$_{TTD}$, TAF3$_{PHD3}$, KDM5A$_{PHD3}$, but was typically decreased when acquired to H3K4^{n}Bu$_3$. BPTF$_{PHD}$ and KDM7B$_{PHD}$ showed a 5–10-fold increase in the binding affinity for H3K4et$_3$ relative to H3K4me$_3$. The increased binding affinity of H3K4et$_3$ to K9 demethylation by KDM7B-JmjC reveals a similar K_M but a ~13-fold higher V_{max}. This demonstrates how the affinity of the chromatin binding module drives the catalytic efficiency of the linked JmjC-domain, highlighting the importance of inter-domain cross talk in epigenetic regulation.

In Chapter 5, the effect of *N*$^{\alpha}$-lysine stereochemistry was investigated against histone *N*$^{\varepsilon}$-methyl-lysine associated proteins using ITC, MALDI-TOF, molecular dynamic (MD) simulations and RF-MS experiments. The most striking result emerged from chromatin binding modules that were able to bind and recognise unnatural D-*N*$^{\varepsilon}$-trimethyl-lysine, albeit at 10–35-fold decreased binding affinity relative to the natural L-*N*$^{\varepsilon}$-trimethyl-lysine, and was further supported by MD simulations. KMTs and KDMs show good activity against natural L-*N*$^{\varepsilon}$-(trimethyl)-lysine; however, only traces or no level of turnover for D-*N*$^{\varepsilon}$-(trimethyl)-lysine substrate. The obtained data provide new insight into the plasticity of chromatin binding modules over KMTs and KDMs which can form the basis of target specific inhibitor and chemical probe development.

Acknowledgement

First, I would like to thank Assoc. Prof. Akane Kawamura and Prof. Tom Brown for accepting me in their group, guiding me and giving me this fantastic opportunity to develop myself. Their supervision throughout my DPhil was truly excellent and it was equally inspiring to work for them. I would also like to thank them for their support during my medical leave.

The work carried out in this thesis would not have been possible without the fruitful collaboration with Prof. Jasmin Mecinović, Abbas Al temimi, Dr. Kiran Kumar, Prof. Robert Paton and Dr. Bhaskar Bhusan. It was wonderful experience to work together and giving me insights into other fields of science. I would also give my sincere thanks to the EPSRC funding body for supporting the work carried out in this thesis.

I am also indebted by Dr. James Wickens and Elizabeth Pires for teaching and advising me in using the mass-spectrometry facilities; Dr. Louise Walport and Dr. Chris Lohans for advice in protein production and purification; Georgina Berridge and Dr. Anthony (Tony) Tumber for their assistance and help using the RapidFire instruments; Dr. Catrine Johansson and Eidarus (Eddy) Salah for their help in production of the most challenging KDM and TET proteins; Prof. Christopher J. Schofield and Dr. Richard J. Hopkinson for their advice and guidance in synthetic organic chemistry and chemical biology and finally, Adam Hardy for maintaining the laboratories.

My thanks go to my friendships with Bojana Selinšek, Jos Kamps, Marie Bernkopf, Cyrille Thinnes, Robert Lesniak, David Zollman, Dong Zhang, James Holt-Martin (housekettle), Alexander Erdman, Ludovic Halby, Joanna Bonicci, Mark White, Yijia Zhang, Oliver Coleman, Klemensas Simelis,

William Figg, Tobias John, Raphael Reinbold, Timothy Suits, Cheryn Tan and Darryl McClymont for making Oxford such a wonderful time for me.

Ik zou ook in het bijzonder Casper Geritssen, Joost Teunissen, Daniël Geritssen, Rick Schut, Maarten Roetman en Bastiaan Lammers willen bedanken hun vriendschap. Het is niet altijd makkelijk om dit te onderhouden als je ~500 km uit elkaar woont, maar samen bijkletsen blijft als van ouds gezellig.

Als laatste, zou ik mijn diepste dank willen betuigen aan mijn ouders, zus en oma voor hun steun, hun raad en luisterend oor tijdens mijn DPhil. Het ware leuke, inspiratie volle maar ook zware tijden met wie ik het zonder niet gered zou hebben.

Declaration

I declare that the work performed in this thesis is entirely my own unless a reference to a person or published source has been made. Artwork in this thesis is painted by Hanna Belle-Helan and is gratefully used as inside front-cover and as introduction cover to the Chapters 1–5,7[†].

Date

Signed

[†]The cover of Chapter 3 is based upon a KDM5A-PHD3 crystal structure (PDB: 3GL6).

Contents

Abstract ..... **i**

Acknowledgement ..... **iii**

Declaration ..... **v**

Contents ..... **vi**

Abbreviations ..... **xiii**

Chapter 1. Introduction to the epigenetic landscape **1**

1.1 Introduction to the chromatin **2**

1.1.1 DNA structure and packaging in the nucleus 2

1.2 TET facilitate the iterative oxidation of ^{5m}C to ^{5hm}C, ^{5f}C and ^{5ca}C **6**

1.2.1 ^{5m}C is a common DNA modification 6

1.2.2 DNMTs methylate ^{5}C using SAM 6

1.2.3 TETs enzymes are 2OG-dependent oxygenases that facilitate ^{5m}C oxidation 8

1.2.4 ^{5ox}C are part of ^{5m}C demethylation pathway to C 10

1.2.5 Reaction mechanism of TET enzymes 11

1.2.6 Substrate specificity of TET enzymes 12

1.2.7 The abundance of ^{5ox}C are different among tissues 14

1.2.8 Biological relevance of ^{5ox}C and TET enzymes 15

1.2.9 Chemical probes are vital to the biochemical characterisation of TET enzymes 17

1.3 Post translational modifications on histones **18**

1.3.1 Histone N^{ε}-lysine methylation is a dynamic mark and important in gene regulation 21

1.3.2 Histone lysine methyltransferases methylate histone N^{ε}-lysine residues 23

1.3.3 Amine oxidase KDM1 can demethylate N^{ε}-methyl-lysine 25

1.3.4 The JumonjiC domain of KDM2–7 can demethylate N^{ε}-methyl-lysine 26
1.3.5 The KDMs can demethylate various N^{ε}-methyl-lysine residues 29
1.3.6 Biological relevance of KDMs 31
1.3.7 Inhibitor development for KDMs 33
1.3.8 Nuclear chromatin N^{ε}-methyl-lysine binding modules 35
1.3.9 Creating highly selective KDM and chromatin binding module inhibitors remains challenging 36
1.4 Thesis objectives 37

Chapter 2. Biochemical characterisation, assay development and inhibitor screening for human TETs 40

2.1 Contributors to this Chapter 41
2.2 Introduction to TET proteins 42
2.2.1 TETs and ^{5ox}C play multiple functional roles in the cell 42
2.2.2 *In vitro* enzymatic assay to study the catalytic activity of TETs 43
2.3 Chapter objectives 45
2.4 Results 46
2.4.1 Production of human TET proteins 46
2.4.2 AlphaScreen™ technology as *in vitro* ^{5hm}C-DNA detection method 47
2.4.3 ITC analysis of substrate enzyme kinetics of $^{5m}C_{ss}$DNA 2.5 against hTET2$_{CD}$ 52
2.4.4 MALDI-TOF steady state enzyme kinetics of hTET2$_{CD}$ 54
2.4.5 Activity assay for hTET2$_{CD\Delta LCI}$ using RapidFire Mass spectrometry 56
2.4.6 Inhibition of hTET1$_{CD}$, hTET2$_{CD}$ and hTET2$_{CD\Delta LCI}$ measured using AlphaScreen™ 60
2.4.7 Inhibition of hTET2$_{CD\Delta LCI}$ measured using RapidFire Mass spectrometry 65
2.4.8 Cyclic peptide development as inhibitor for hTET1$_{CD}$ 68
2.5 Discussion 72
2.6 Conclusion 77
2.7 Future perspective 79

2.8 Materials and Methods 80

2.8.1 General conditions 80

2.8.2 Protein production for *in vitro* analysis 80

2.8.3 Inhibition of hTET proteins using AlphaScreen™ 82

2.8.4 Enzyme kinetics of $hTET2_{CD}$ of $^{5m}C_{ds}$DNA using MALDI-TOF 85

2.8.5 Enzyme kinetics of $^{5m}C_{ss}$DNA for $hTET2_{CD}$ of ITC analysis 85

2.8.6 RapidFire mass spectrometry (RF-MS) for $hTET2_{CD\Delta LCI}$ inhibition 86

Chapter 3. Investigating the PHD3-finger of KDM5A 90

3.1 Contributors to this Chapter 91

3.2 Introduction to histone H3 *N*ε-lysine methylation 92

3.2.1 Histone *N*ε-lysine methylation in the chromatin 92

3.2.2 Introduction to demethylase KDM5A 94

3.2.3 Biological relevance of KDM5s as nuclear proteins 97

3.3 Approaches to study the function KDM5A-PHD3 98

3.3.1 Small molecule probes for PHDs 98

3.3.2 Histone peptide probes 98

3.4 Chapter objectives 101

3.5 Results 103

3.5.1 Synthesis of functionalized *N*ε-trimethyl-lysine and preparation of H3 mimic peptides 103

3.5.2 Protein production of $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$ 105

3.5.3 Binding studies of $H3_{1-21}K4et_3$ 3.1 and $H3_{1-21}K4pip$ 3.8 against $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$ using ITC 108

3.5.4 AlphaScreen™ displacement of $H3_{1-21}K4et_3$ 3.1 and $H3_{1-21}K4pip$ 3.8 against $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$ 112

3.5.5 Production of KDM5 proteins 116

3.5.6 FDH measurements to investigate the demethylation activity of KDM5 proteins 118

3.5.7 MALDI-TOF measurements to determine if $H3_{1-21}K4et_3$ 3.1 and $H3_{1-21}K4pip$ 3.8 are alternative substrates KDM5 proteins 121

3.5.8 $H3_{1-21}K4et_3$ 3.1 and $H3_{1-21}K4pip$ 3.8 as inhibitors for KDM5s JmjC-domain 130

3.5.9 Inhibition of $KDM1A_{FL}$ by $H3_{1-21}K4et_3$ 3.1, $H3_{1-21}K4me_3$ 3.2 and $H3_{1-21}K4pip$ 3.8 132

3.5.10 MALDI-TOF measurements to determine the inhibition of $KDM5A_{C2}$ demethylation by $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$ 133

3.5.11 Exploring the KDM5 domain boundaries 134

3.5.12 Putative role of KDM5A-PHD2 135

3.6 Discussion 138

3.7 Conclusion 145

3.8 Future perspective 147

3.9 Materials and Methods 148

3.9.1 General Experimental conditions 148

3.9.2 ITC analysis 148

3.9.3 AlphaScreen™ displacement 148

3.9.4 MALDI-TOF assay 150

3.9.5 ^{1}H NMR spectroscopy to determine $H3_{1-21}K4pip$ 3.8 hydroxylation site by KDM5 proteins 153

3.9.6 FDH assay 153

3.9.7 Fmoc-Amino acid and histone H3 mimic peptide synthesis 154

3.9.8 Recombinant protein production 154

Chapter 4. The effect of sterically enlarged H3 N^{ε}-trimethyl-lysine 4 on the binding affinity to chromatin binding modules 160

4.1 Contributors to this Chapter 161

4.2 Introduction to reader proteins 162

4.2.1 Chromatin binding domains in different shapes and forms 163

4.3 Chapter objectives 165

4.4 Results 166

4.4.1 Panel of selected chromatin reader modules 166

4.4.2 Synthesis of di-alkylated Fmoc-protected amino acids and Solid Phase Peptide Synthesis 167

4.4.3 ITC analysis of $KDM4A_{TTD}$ and $KDM5A_{PHD3}$ against di-alkylated histone H3 lysine 4 mimic peptides ($H3_{1-21}$) 4.1–4.7. 169

4.4.4 Synthesis of N^{ε}-tri-alkylated N^{α}-Fmoc-protected lysine residues and incorporation in histone mimic peptides ($H3_{1-21}$) 172

4.4.5 ITC binding studies of histone H3 mimic peptides containing N^{ε}-tri-alkylated lysine 4 residues against Tandem Tudor Domain reader proteins 174

4.4.6 ITC binding studies of histone H3 mimic peptides containing N^{ε}-tri-alkylated lysine 4 residue against Plant Homeodomain reader proteins 177

4.4.7 ITC binding studies of histone H3 mimic peptides containing N^{ε}-tri-alkylated lysine 4 residues against chromo domain reader protein $CHD1_{DCD}$ 181

4.4.8 ITC binding studies of histone H3 mimic peptides containing N^{ε}-tri-alkylated lysine 4 against KDM7B; surface groove *versus* cavity insertion 183

4.4.9 KDM7B JmjC catalytic activity on tighter binding of the PHD; $H3_{1-21}K4me_3K9me_2$ 4.21 vs $H3_{1-21}K4et_3K9me_2$ 4.22 185

4.5 Discussion 188

4.6 Conclusion 192

4.7 Future perspective 194

4.8 Materials and methods 196

4.8.1 General conditions 196

4.8.2 ITC analysis 196

4.8.3 Fluorescence polarization assay 197

4.8.4 Activity assay of $KDM7B_{PHDJmjC}$ with $H3_{1-14}K4me_3K9me_2$ substrate using MALDI-TOF 197

4.8.5 Incorporation of Fmoc-trimethyl-lysine derivatives in histone mimic peptides using microwave assisted solid phase peptide synthesis 198

4.8.6 Synthesis of Fmoc-Lys($me_{2,3}$)-OH derivatives 199

4.8.7 Protein expression 214

Chapter 5. Investigating the effect on Histone *C*$^{\alpha}$-Lysine stereochemistry on catalysis and recognition by KMTs, KDMs and chromatin binding modules 221

5.1 Contributors to this Chapter 222

5.2 Introduction 223

5.2.1 In the context of epigenetic regulation 224

5.3 Chapter objectives 225

5.4 Results 226

5.4.1 Synthesis of Fmoc-protected lysine monomers and production of peptides 226

5.4.2 ITC analysis of histone H3 D- *versus* L-*N*$^{\varepsilon}$-trimethyl-lysine 4 mimic peptides against chromatin binding modules 228

5.4.3 MD simulating interactions of reader domains to H3 D- and L-*N*$^{\varepsilon}$-trimethyl-lysine 231

5.4.4 Histone methyltransferase (KMTs) can methylate D-*N*$^{\varepsilon}$-lysine residues in histone H3 H4 substrates 235

5.4.5 KDM activity on D-*N*$^{\varepsilon}$-methylate-lysine residues in histone H3 peptide mimic substrates 238

5.5 Discussion 248

5.6 Conclusion 250

5.7 Future perspective 251

5.8 Materials and Methods 252

5.8.1 ITC analysis of chromatin binding modules against histone H3 mimic peptides (carried out by Dr. Jasmin Mecinović) 252

5.8.2 Computational studies applying molecular dynamics (carried out by Dr. Kiran Kumar and Dr. Robert Paton) 252

5.8.3 *In vitro* assays of Histone methyltransferases testing histone mimic peptides 254

5.8.4 *In vitro* assays of lysine specific demethylases examining histone mimic peptide using RF-MS and ^{1}H NMR 255

5.8.5 RF-MS instrument configuration 256

5.8.6 Solid Phase Peptide Synthesis 258

5.8.7 Protein production of chromatin binding modules, KMTs and KDMs 261

5.8.8 Synthetis of N^{α}-Fmoc-N^{ε}-trimethyl-lysine analogues 264

Chapter 6. Summary and Conclusions 271

Chapter 7. Materials and methods 278

7.1 Synthetic methods 278

7.1.1 General experimental conditions 278

7.1.2 Synthesis of Fmoc-protected lysine derivatives 279

7.1.3 Solid Phase Peptide Synthesis 286

7.2 Microbiological methods and techniques 288

7.2.1 General materials and methods for microbiology 288

7.2.2 Production of competent cells 288

7.2.3 Preparation of antibiotics 289

7.2.4 Preparation of 2TY media or Terrific broth (TB) 289

7.2.5 Preparation of agar plate 289

7.2.6 Transformation of plasmid in *E. coli* 290

7.2.7 Preparation of HisTrap™ HP 290

7.2.8 SDS-PAGE preparation and analysis 291

Chapter 8. References 294

Chapter 9. Appendix 314

Abbreviations

2,4-PDCA	2,4-pyridinedicarboxylic acid
2HG	2-hydroxyglutarate
2OG	2-oxoglutarate
2TY	2 Tryptone/Yeast extract
6×His-tag	6×histidine-tag
^{1m}A	1-*N*-methyladenine
^{3m}C	3-*N*-methylcytosine
^{3m}T	3-*N*-methylthymine
^{6m}A	6-*N*-methyladenine
^{5ca}C	5-carboxylcytosine
^{5f}C	5-formylcytosine
^{5hm}C	5-hydroxymethylcytosine
^{5m}C	5-methylcytosine
^{5ox}C	5-oxidizedmethylcytosine
αKG	α-keto-glutarate
Asc	sodium-L-ascorbate
µW-SPPS	Microwave assisted solid phase peptide synthesis
AIRE	Autoimmune regulator
Alpha	Amplified luminescence proximity homogeneous assay
AML	Acute myeloid leukaemia
AMKL	Acute Megakaryoblastic Leukemia
ARID	AT-rich interaction domain
ADP	Adenosine 5'-diphosphate
AM	Active modification
ATP	Adenosine 5'-triphosphate
BER	Base excision repair
Boc	tert-butyloxycarbonyl
Boc_2O	di-tert-butyl dicarbonate
BPTF	Bromodomain PHD Transcription Factor
BSA	Bovin Serum Albumin

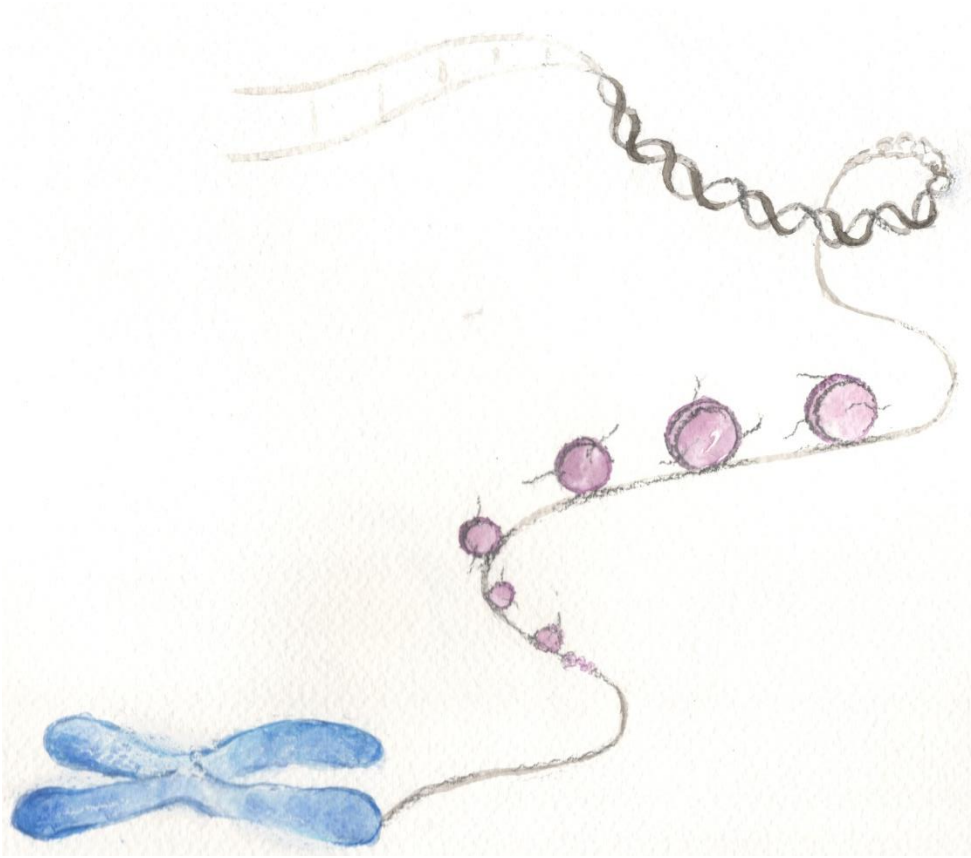
C	Celsius
Cbz	Carboxybenzyl zervas
CLL	Chronic lymphocytic leukaemia
CD	Catalytic domain
ChD	Chromo domain
CHD1	Chromodomain Helicase/ATPase Domain
CMML	Chronic myelomonocytic leukaemia
COSY	correlation spectroscopy
CpA	cytidine phosphate adenosine
CpC	cytidine phosphate cytidine
CpG	cytidine phosphate guanosine
CpT	cytidine phosphate thymidine
CV	Column volume
Da	Dalton
DCD	Double chromo domain
DIC	*N,N'*-diisopropylcarbodiimide
DIEA	*N*,*N*-diisopropylethylamine
DMAP	*4*-(dimethylamino)pyridine
DMF	*N*,*N*-dimethylformamide
DMSO	Dimethylsulfoxide
DNA	2'-deoxyribonucleic acid
DNMT	DNA methyltransferase
DOT	Disruptor of telomeric silencing
DSBH	Double stranded β-helix
DTT	1,4-dithiothreitol
DTTD	Double tandem tudor domain
DUH	2'-deoxyuridine-hydroxylase
E. coli	*Escherichia coli*
EDT	1,2-ethanedithiol
EDTA	Ethylenediaminetetraacetic acid
et	ethyl
EPR	Electron paramagnetic resonance
equiv	equivalents

ESC	embryonic stem cell
ESI	Electron spray ionisation
FA	Formic acid
FAD	Flavin adenine dinucleotide
FDH	Formaldehyde dehydrogenase
FIH	Factor inhibiting HIF
FIT	flexible *in vitro* translation
Fmoc	Fluorenylmethyloxycarbonyl
FP	Fluorescence polarisation
FPLC	Fast protein (/performance) liquid chromatography
FTO	Fat mass and obesity associated protein
g	gram
H1	Histone H1
H2A	Histone H2A
H2B	Histone H2B
H3	Histone H3
H4	Histone H4
HCCA	α-Cyano-4-hydroxy cinnamic acid
HDAc	Histone deacetylase
HEK	Human embryonic kidney
HEPES	4-[2-hydroxyethyl]-piperazin-1-ethanesulphonic acid
HOBt	1-hydroxybenzotriazole
HRP	HorseRadishPeroxidase
hrs	hours
HPLC	High performance liquid chromatography
IDH	Isocitrate dehydrogenase
ipr	isopropyl
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Infrared
ITC	Isothermal titration calorimetry
JARID1	JumonjiC ARID 1
JmjC	JumonjiC
JmjN	JumonjiN

K_a	association constant
Kac	N^{ε}-acetyl-lysine
KAT	Lysine acetyltransferase
Kbhb	N^{ε}-[C^{β}-hydroxybutyryl]-lysine
Kcr	N^{ε}-crotonyl-lysine
K_{cat}	Enzyme substrate turnover rate
K_d	dissociation constant
KDM	Lysine demethylase
Khib	N^{ε}-[C^{β}-hydroxyisobutyryl]-lysine
K_M	Michaelis constant
Kmal	N^{ε}-malonyl-lysine
Kme_1	N^{ε}-monomethyl-lysine
Kme_2	N^{ε}-dimethyl-lysine
Kme_3	N^{ε}-trimethyl-lysine
KMT	Lysine methyltransferase
Kpr	N^{ε}-propionyl-lysine
Ksuc	N^{ε}-succinyl-lysine
L	Liter
LC-MS	Liquid chromatography – mass spectrometry
LID	Little imaginal discs
LSD1	Lysine specific demethylase 1
M	$mol \cdot L^{-1}$
MALDI	Matrix assisted laser desorption ionization
MBT	Malignant brain tumor
MD	Molecular dynamics
MDS	Myelodysplastic syndrome
me	methyl
MES	*N*-morpholine ethanesulphonic acid
mHz	megahertz
min	minutes
MLL1	Mixed lineage leukaemia 1
M.P.	Melting point
MRM	multiple reaction monitoring

Mw	Molecular weight
m/z	mass over charge ratio
NAD	β-nicotinamide adenine
NMP	*N*-methyl-2-pyrrolidone
NMR	Nuclear magnetic resonance
NOG	*N*-oxalylglycine
NUP98	Nuclear Pore complex
OAA	Octylammonium acetate
OSu	*N*-oxysuccinimide
PD	Passive dilution
PDB	Protein data bank
PHD	Plant homeo domain
PHF	PHD-finger containing protein
ppm	parts per million
PRMT	protein arginine methyltransferase
ProHD	Prolyl hydroxylase domain
PTM	Post translation modification
R_f-C_{18}	Reverse phase-C_{18}
RF-MS	RapidFire Mass Spectrometer
Rme_1	N^{ω}-monomethyl-arginine
Rme_{2a}	N^{ω},N^{ω}'-dimethyl-arginine (asymmetric)
Rme_{2s}	N^{ω},N^{ω}-dimethyl-arginine (symmetric)
RNA	Ribonucleic acid
RPM	rounds per minute
RT	room temperature
SAH	*S*-5'-adenosyl-L-homocysteine
SAM	*S*-5'-adenosyl-L-methionine
SDS-PAGE	sodium dodecyl sulphate – polyacrylamide gel electrophoresis
sec	seconds
SET	Su(var) enhancer of zeste Trithorax
SGF29	SAGA associated Factor 29
SPE	Solid phase extraction
Sph	O^{β}-phosphoryl-serine

SPPS	Solid phase peptide synthesis
SPIN	Spindlin
t	time
TAF3	TATA Box binding protein Associated Factor
TCA	tricarboxylic acid
TCEP	2-tris(carboxyethyl)phosphine hydrochloride
TD	Tudor domain
TDG	Thymine DNA glycosylase
TET	Ten-Eleven Translocation
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
TIPS	Triisopropylsilane
TLC	Thin layer chromatography
TOCSY	Total correlation spectroscopy
TOF	Time of Flight
Tph	O^{β}-phosphoryl-threonine
Tris	tris(hydroxymethyl)aminomethane
TSS	transcription starting site
TTD	tandem tudor domain
TTS	transcription termination site
UV	ultraviolet
VIS	Visible
V_{max}	maximum reaction velocity
Yph	*O*-phosphoryl-tyrosine



Chapter 1. Introduction to the epigenetic landscape

1.1 Introduction to the chromatin

Life evolved for millions of years by adequately storing cellular information to function and to transfer this information during replication. In the human body, DNA holds the blueprint of every cell and despite each cell contains virtually identical DNA, the individual cellular function can be very different. An important aspect to cellular diversity is the spatio-temporal and context dependent regulation of DNA which, through the transcription-translational machinery, influences downstream protein expression. The study that embodies the regulatory control of DNA is epigenetics and is defined as: "An Epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alteration in the DNA sequence".[11]

A natural part of evolution is the changes that occur to the genetic code in the form of silent, missense, nonsense mutations or translocations. As is the case for every gene, the epigenetic regulation ultimately determines the genotypic or phenotypic outcome. Altered genotype and phenotype can result in biological advantages such as resistance to disease or enhanced physical endurance; however it could also result in uncontrolled proliferation, cellular reprogrammed development, disease(s), apoptosis or acute cell death. In order to be able to address an anomaly, it is vital to understand the impact of the alteration and what kind of down-stream consequences this may have to gene regulation. This thesis concerns the recent advances I made to the field of epigenetics.

1.1.1 DNA structure and packaging in the nucleus

Cellular DNA has the approximate length of two meters in its linear form in humans, but is situated into a confined eukaryotic cell nucleus; hence, careful organization is required. Packing of DNA occurs via its assembly into chromatin, where DNA is wound around histone complexes to form nucleosomes providing efficient DNA folding of up to 10 000 times[12]. Nucleosomes modulate active or silenced

gene transcription by creating an open chromatin, known as euchromatin, or a densely packed chromatin, known as heterochromatin. Each nucleosome consists of DNA (approximately 146–148 bases) wound around an octameric histone core complex containing two copies of histone H2A, H2B, H3 and H4. To form the histone octamer, a histone H3 and H4 form a heterodimer [H3-H4] and two dimers result in a tetramer [$H3_2$-$H4_2$]. Histone H2A and H2B form a heterodimer [H2A-H2B] which can bind two times to the [$H3_2$-$H4_2$] tetramer forming the histone octameric complex (Fig. 1.1.1).

The physiologically most common structure of DNA, known as B-DNA, is wrapped around the octameric histone. By over-winding by 1.65 base-pair (bp)[13,14] on top of the 10.5 bp per helical turn, left handed super helical turn to strengthen the formation of the nucleosome complex. Additional nucleosome stabilization is created by the binding of histone H1 to the nucleosome allowing a typical inter-nucleosome DNA linker lengths of ~36–40 bp in humans.[15,16]

Interestingly, histone H3 and histone H4 were discovered to be of archaeal origin; however, unlike eukaryotes, archaeal do not possess histone H2A or H2B. The leading theory is that archaea and eukarya shared a common ancestor.[17] Protein sequence examination of histones found in eukarya demonstrate that the four canonical histones are distinctive in amino acid sequence but manifest structurally similar architectures suggesting that H2A and H2B may have evolved from H3 and H4 after divergence. In humans, histones are between 103 (H4) to 130 amino acids (H2A) in length and contain each three α-helixes at the C-terminus with N-terminal protruding tails from the centre core. Interestingly, each histone is subjected to an extensive array of post translational modifications (PTMs). These frequently small dynamic modifications provide control of the DNA accessibility ultimately leading to activate or silence transcriptional gene control. In the silenced form, the DNA is believed to be arranged in 30 nm fibres.[18]

The importance of histones to the cell is reflected in the fact that the canonical histones are encoded by intron-less repetitions of >10 gene copies and found in three clusters on chromosomes 1q21, 1q42 and 6p22.[19] Histone H3 and H2A, unlike H4 and H2B, are observed in many variants. Eight variants have been found for H3: H3.1, H3.2, H3.3, CenH3, H3.t (H3.4), H3.5, H3.X and H3.Y. Replication-dependent H3.1 is only found in mammals while H3.2 is found in all metazoans. These two histone H3s are considered to be the canonical variants and differs only by one amino acid (H3.1: C, H3.2: S). H3.3 has, in contrast to H3.1 and H3.2, lower sequence similarity and is not found in the three histone clusters but observed as single gene copies. The gene encoding for H3.3 typically contains introns and is uniquely replication independent.[20] H3.3 is linked to transcriptionally active chromatin, enriched near the promotor site and correlates to locally elevated levels of N^{ε}-dimethyl-lysine 4 ($H3K4me_2$).[20,21] Histone H3.t is solely observed by mRNA analysis in testis and plays an active role in the early stages of spermatogenesis.[22,23] Like H3.t, H3.5 is found in high abundance in testis tissue situated in high frequency in euchromatic environments; however, it does not retain in mature sperm.[24,25] The centromere-specific histone H3 variant (CenH3) resembles lower sequence similarity relative to the other histone H3 variants and is involved in kinetochore assembly.[26,27] H3.X and H3.Y are found in primates and have considerably lower abundance.[28] H3.X and H3.Y are described to be located in primary human tissues (bone, breast, lung, testis and brain)[28] and H3.Y is linked to both stress and transcriptional activation.[29] Like H3, many variants exist beside canonical H2A, including H2A.X and H2A.Z that are found to be linked to DNA repair and chromatin remodelling.[30]

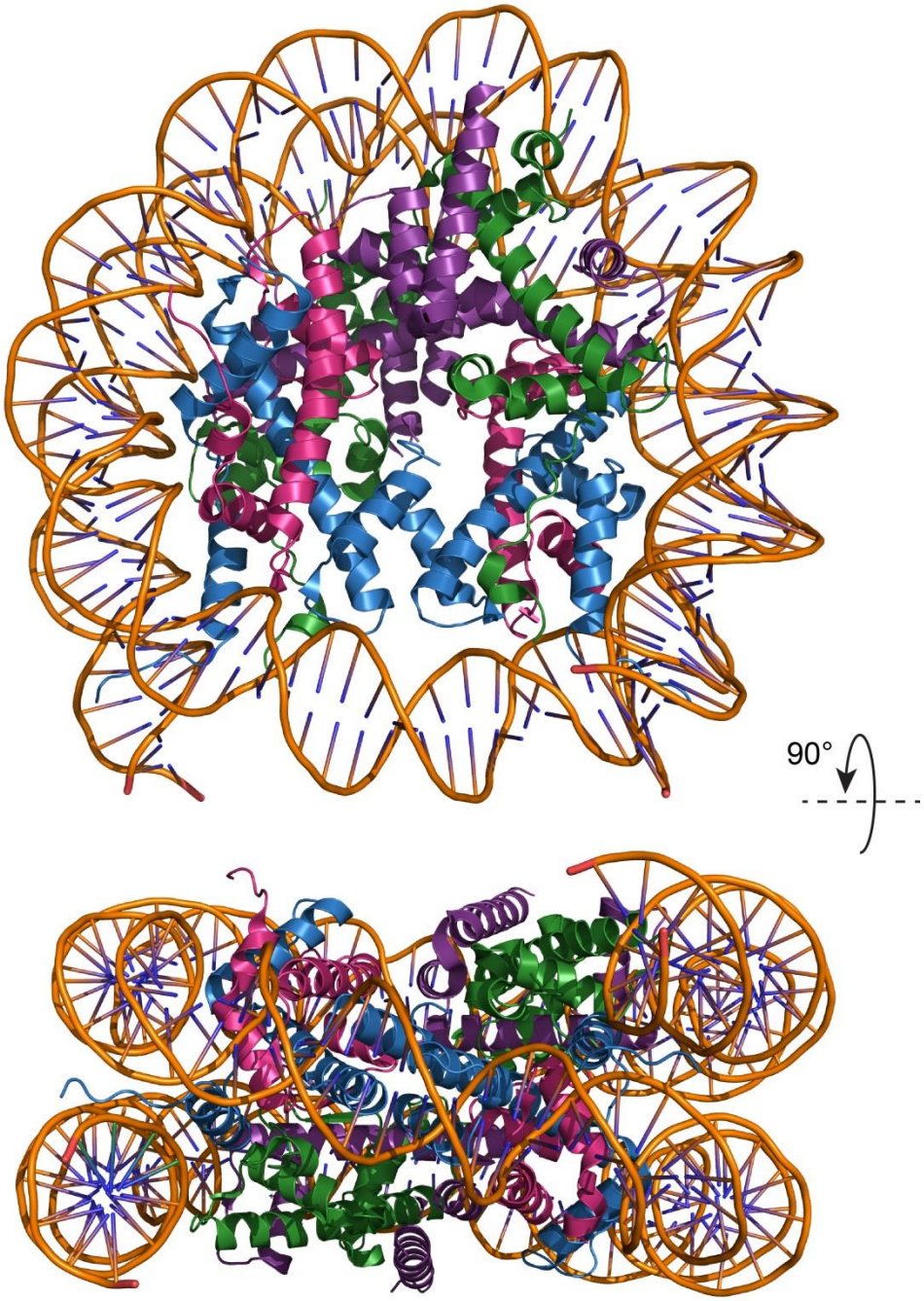


Figure 1.1.1: Crystal structure of a nucleosome (top and side view) consisting of two of each histone H2A (green), H2B (purple), H3 (blue), H4 (pink) double stranded DNA (orange) (PDB: 3AFA).

1.2 TET facilitate the iterative oxidation of ^{5m}C to ^{5hm}C, ^{5f}C and ^{5ca}C

Accurate gene regulation is of critical importance for the optimal functioning of the cell. Part of this regulation involves epigenetic modifications made directly to the base or the sugar of the DNA without interfering with the underlying genetic code. These modifications are perhaps the simplest form of epigenetic gene regulation directly mounted on the DNA, most commonly on the cytosine base.

1.2.1 ^{5m}C is a common DNA modification

A widely studied epigenetic DNA modification is 5-methylcytosine (^{5m}C) which is observed throughout the human body with relatively stable levels found among tissues (3.5–4.5% of all cytosine are methylated).[31] There is a high abundance of ^{5m}C found at CpG island near the promotor region of a silenced gene were levels as high as 60–80%[32] are observed. ^{5m}C is generally associated with gene silencing; however, the function is strongly dependent on the position within the genome. DNA methylation, for example at enhancer regions in the DNA, has been linked to gene activation. Furthermore ^{5m}C is associated to ageing[33], and is correlated to inheritance and major life events.[34]

1.2.2 DNMTs methylate ^{5}C using SAM

DNA methyltransferases (DNMTs) can install a methyl group on the 5 position of C by using *S*-adenosyl-L-methionine (SAM) as a co-factor to yield ^{5m}C and *S*-adenosyl-L-homocysteine (SAH). There are to date three DNMTs identified that can perform this reaction, namely DNMT1 and DNMT3A/B. DNMT1 prefers the catalysis of hemi-methylated DNA[35] and is considered to maintain methylation levels while DNMT3A/B are found to introduce *de novo* methyl groups.[36] In the proposed catalytic mechanism for DNMTs, the DNA binds to the protein flipping out a base that is directed

towards the active site (Fig. 1.2.1). A cysteine located in the active site of DNMT attacks ^{6}C leaving a stabilized lone electron pair on ^{5}C that reacts with the methyl group of the SAM creating ^{5m}C and SAH. This event causes a direct steric clash resulting in β-elimination followed by the release of the DNA substrate.[37] The DNMTs are vital to the cellular function and aberrant methylations patterns are frequently linked to diseases.[38] DNMTs are also associated to cancers such as acute myeloid leukaemia (AML), breast, lung, liver and prostate cancer.[39]

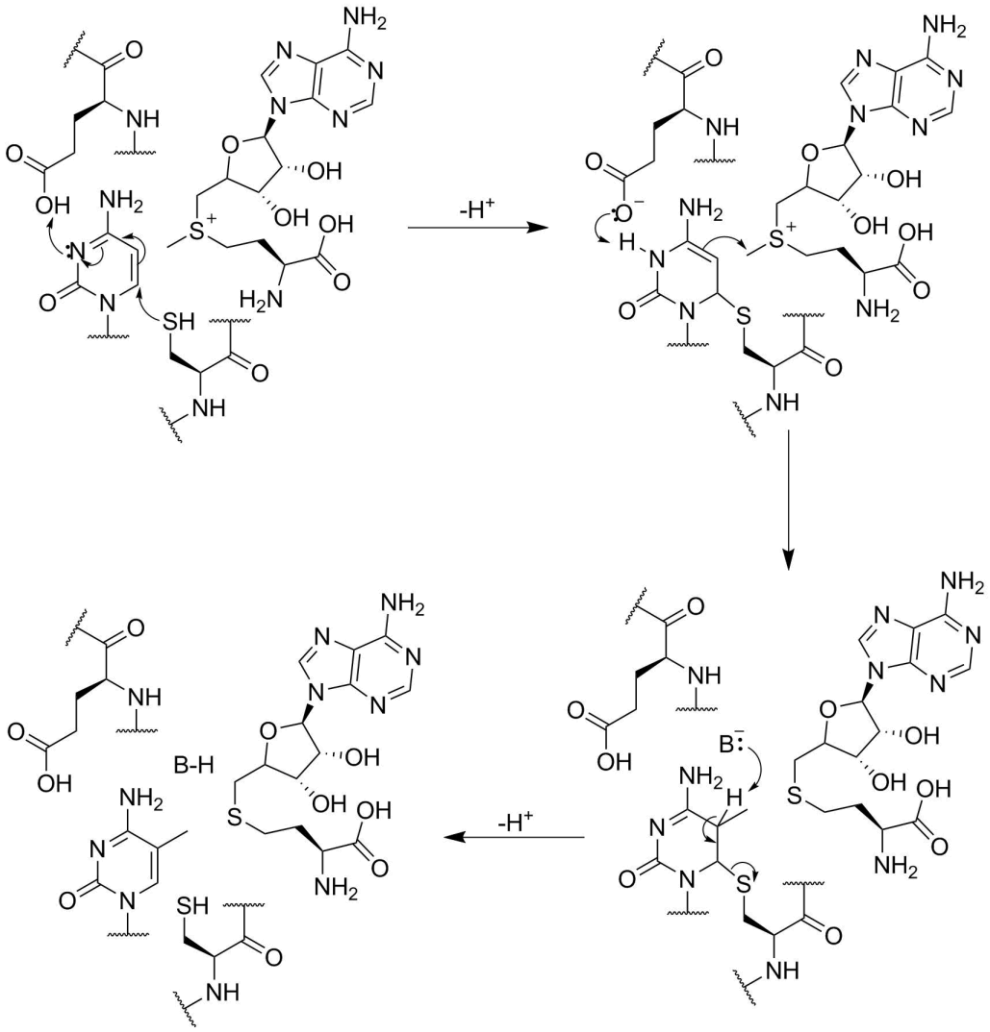


Figure 1.2.1: Proposed catalytic mechanism[37] of C methylation by DNMT1 using the co-factor SAM to produce ^{5m}C and SAH mediated by C1226 and E1266. The base facilitating the β-elimination in this mechanism is likely to be water.[40]

1.2.3 TETs enzymes are 2OG-dependent oxygenases that facilitate ^{5m}C oxidation

In 2009, Kriaucionis *et al.* convincingly reported the discovery of 5-hydroxymethylcytosine (^{5hm}C) on DNA as epigenetic modification that is abundant in human brain cells[41]. TET proteins were assigned to introduce this modification by ^{5m}C oxidation using the catalytic 2OG-dependent oxygenase domain (Fig. 1.2.2).[42] Shortly after, TETs were assigned as a family of enzymes that catalyse the oxidation of ^{5m}C in iterative steps to ^{5hm}C, 5-formylcytosine (^{5f}C) and 5-carboxylcytosine (^{5ca}C) (Fig. 1.2.2). TET enzymes, like other 2OG-dependent dioxygenase superfamily members (see also section 1.3.4), use Fe^{II} catalysis with 2OG and O_2 as co-factors for catalysing oxidations.[42,43] Characteristic of this protein superfamily is the jelly-roll fold of the double stranded β-helix (DSBH) with the active site placed within the core. To date, three TET proteins have been discovered, all containing a split DSBH catalytic core, where the two halves are covalently connect through a lower complexity insert (LCI) which is predicted to be unstructured.[44] A cysteine rich region (CRR) flanks the DSBH on both the N- and C-termini and stabilizes the DSBH. Removal of the CRR abolishes catalytic activity.[45] In the DSBH of crystallized hTET2, the Fe^{II} is coordinated by H1382, D1384 and H1881 residues, and 2OG coordinates to Fe^{II} in a bidentate manner (Fig. 1.2.3). The C5-carboxylate of 2OG interacts with R1261, H1416, R1896 and S1898. Although the DSBH is highly conserved among the hTETs, the LCI is not and is believed to serve different roles. The LCI can be truncated for all three TET proteins but does impact the catalytic efficiency of the catalytic domain (CD).[45–47] Additionally, hTET1 and hTET3 have an N-terminal CXXC zinc finger that is believed to guide TET proteins to CpG (enriched) regions. This domain is absent in hTET2 but instead, hTET2 interacts with IDAX (CXXC4) that facilitates a similar function.[48]

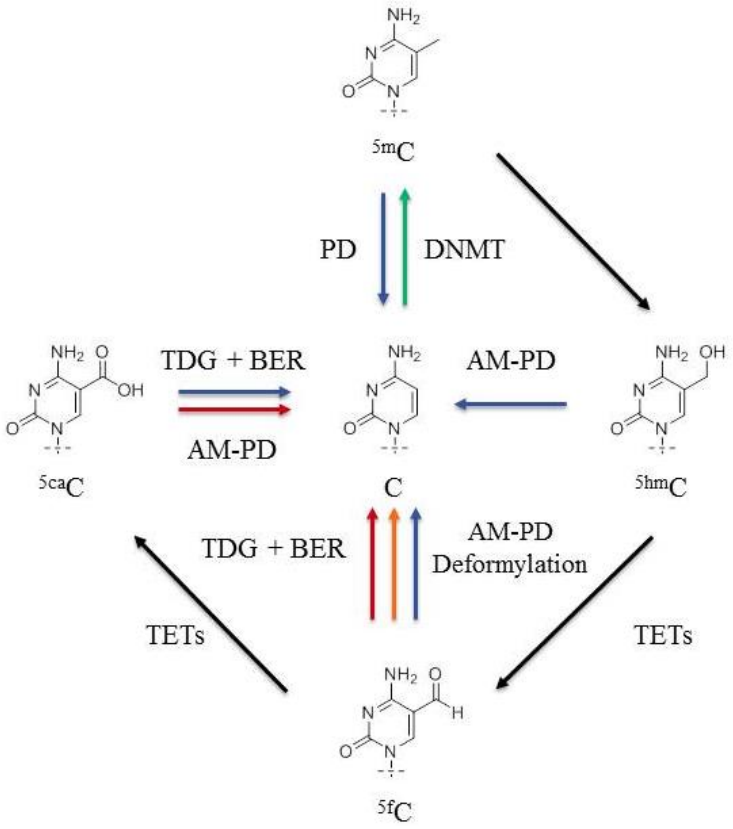


Figure 1.2.2: Cytosine methylation, oxidation and demethylation pathway. (AM)-PD: (active modification) - passive dilution; TET: Ten-Eleven Translocation; DNMT: DNA methyltransferase; TDG: Thymine-DNA glycosylase; BER: Base excision repair.[49]

Binding of double stranded DNA by TET is achieved by a strong interaction between the lysine and arginine residues near the active site and the DNA phosphate backbone. The concave orientation of the catalytic domain forces the disruption of the hydrogen bonding between ^{5m}C-G enabling ^{5m}C to enter the active site placing the 5-methyl of cytosine adjacent to Fe^{II}.

Interestingly, Alkylation repair proteins (AlkB), a subfamily of the 2OG-dependent dioxygenase are homolog and structurally related to TETs. Known members of the family are ABHs and Fat mass and Obesity associated proteins (FTO). The AlkB family also act on nucleic acid substrates providing demethylation for 1-*N*'-methyladenine (^{1m}A), 6-*N*'-methyladenine (^{6m}A), 3-*N*'-methylcytosine (^{3m}C) and 3-*N*'-methylthymine (^{3m}T).[50–52]

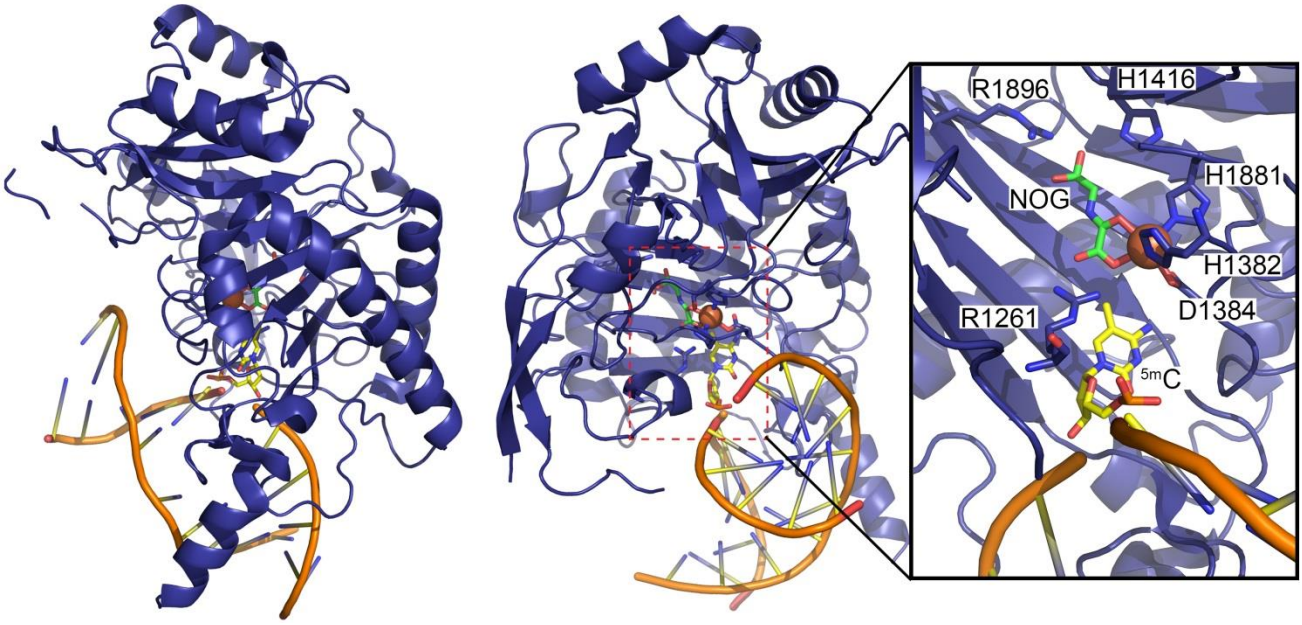


Figure 1.2.3: Crystal structure of hTET2 (blue), DNA (orange helix) and a close-up of the active site. Atoms highlighted are nitrogen: blue; oxygen: red; phosphate: orange. (PDB: 4NM6)[49]

1.2.4 ^{5ox}C are part of ^{5m}C demethylation pathway to C

In mammals, ^{5m}C methylation is a reversible dynamic mark; however, no evidence to date has been found that this can happen through a single step demethylation process where ^{5m}C is converted back to C. There are currently five proposed pathways described in the literature enabling the removal of ^{5m}C or ^{5ox}C to provide unmodified C (Fig. 1.2.2). First, a passive dilution (PD) pathway is proposed were modified ^{5m}C is being depleted by subsequent cell divisions creating an overall reduced concentration of ^{5m}C but always with one remaining sister cell containing the original ^{5m}C installed strand. A report by Mayer *et al.* demonstrated that a decrease in ^{5m}C could not entirely be accounted for via a passive dilution mechanism indicating that DNA remodelling takes place at a faster pace than cell division and therefore, alternative pathways are likely.[53] The second proposed pathway is a combination of active modification and passive dilution (AM-PD); ^{5m}C is oxidized mediated by TETs to ^{5ox}C followed by depletion in consecutive rounds of cell proliferation. The third method involves a combination of proteins using a multi-step active pathway. Herein is ^{5m}C sequentially oxidized to give ^{5f}C or ^{5ca}C that is then recognized by Thymidine DNA Glycosylase (TDG) reading ^{5f}C / ^{5ca}C-G as a T-

G base pair mismatch. TDG hydrolyses the nucleotide between the base and sugar creating an a-basic ribose.[54,55] By means of Base Excision Repair (BER) a C is installed generating an unmodified C-G pair. This active demethylation process has to date only been identified on DNA C-G pairs, although ^{5ox}C modifications have been observed on RNA.[56,57] Finally, there is a pathway recently described by Carell *et al.* that shows labelled ^{5f}C to be directly converted to give C in a cellular context.[58] Labelling experiments using ^{5m}C, ^{5hm}C or ^{5ca}C that would result in demethylation, dehydroxymethylation or decarboxylation for direct C-C bond cleavage were not observed. Although ^{5f}C may undergo a deformylation pathway, no corresponding enzyme(s) has / have been identified. Indeed, also ^{5ox}C are dynamic marks in the chromatin and are believed to be removed by similar pathways as described for ^{5m}C. A fifth pathway has been proposed where ^{5hm}C is deaminated by AID to result in 5-hydroxyUracil (^{5hm}U) and like, T:G mismatch, also ^{5hm}U:G can be removed by TDG *in vitro*.[59] The deamination pathway is however, controversial as the AID protein mainly functions on single-stranded DNA and has a decreased affinity towards ^{5m}C but no activity towards ^{5hm}C relative to C.[60]

1.2.5 Reaction mechanism of TET enzymes

The proposed catalytic mechanism of TET is believed to be similar to other 2OG-dependent dioxygenases described in section 1.3.4. In brief, in the first step, water is replaced by coordination of 2OG to the Fe^{II}. Double stranded B-DNA binds to the TET protein disrupting the hydrogen bonding between the ^{5m}C-G pair protruding the ^{5m}C to the inside of the active site releasing another water molecule. In the next step, molecular oxygen binds to the Fe^{II} forming a Fe^{III} complex. In two sequential steps the oxygen reacts with 2OG giving Fe^{IV} that then rearranges to form a highly reactive $O{=}Fe^{IV}$, CO_2 and succinate. The iron reacts with hydrogen of 5-methyl of the ^{5m}C nucleotide to create a radical ^{5m}C and Fe^{III} followed by the transfer of the hydroxyl group forming ^{5hm}C and Fe^{II}. DNA, CO_2 and succinate leave the active site and water coordinate to the Fe^{II} completing the catalytic cycle. Similar pathways are assumed to occur for ^{5hm}C to ^{5f}C and for ^{5f}C to ^{5ca}C but have, due to mechanistic

properties, a decreased efficiency in catalytic turnover by TET enzymes.[61] It is noteworthy that although ^{5hm}C is observed, TET enzymes might not necessarily release the DNA prior to the next oxidation step.

1.2.6 Substrate specificity of TET enzymes

TET enzymes are known to react on a plethora of substrates but with significantly different catalytic efficiencies. The substrate at which TET proteins are at its peak efficiency, in other words: reaching the highest level of turnover, is still a matter of scientific debate. Different types of known TET substrates are shown in Figure 1.2.4.

TET protein studies showed that ^{5m}C (K_M: 0.48 µM, K_{cat}: $2.12 \cdot 10^{-3}$ min^{-1}) on double stranded DNA is favoured over higher oxidation states, with an increase in K_M for ^{5hm}C (K_M: 0.90 µM, K_{cat}: $0.63 \cdot 10^{-3}$ s^{-1}) and ^{5f}C (K_M: 1.30 µM, K_{cat}: $0.46 \cdot 10^{-3}$ min^{-1}) accompanied with a significant slower turnover (Fig. 1.2.4 Green).[61] It is hypothesized that proton extraction as well as proton intra-inter stabilization may contribute to different rates observed. Furthermore, efforts were devoted to study the activity of TET in DNA CpGs that are of hemi-methylation and symmetric methylation revealing that the cytosine on the opposing strand does not govern the catalytic turnover.[62]

Interestingly, *Naegleria gruberi* Tet1 (NgTet1) and murine Tet (mTet) proteins also mediate thymine oxidation to give 5-hydroxymethyluracil (^{5hm}U), and 5-formyluridine (^{5f}U) *in vitro* and *in vivo* (Fig. 1.2.4 Purple).[62,63] While 5-carboxyuridine (^{5ca}U) has not been detected for mTet, it was observed for NgTet1.[62] Thymine as alternative substrate is substantially worse relative to ^{5m}C and as a result quantification remains challenging due to the relative low conversion levels by both NgTet and mTet.

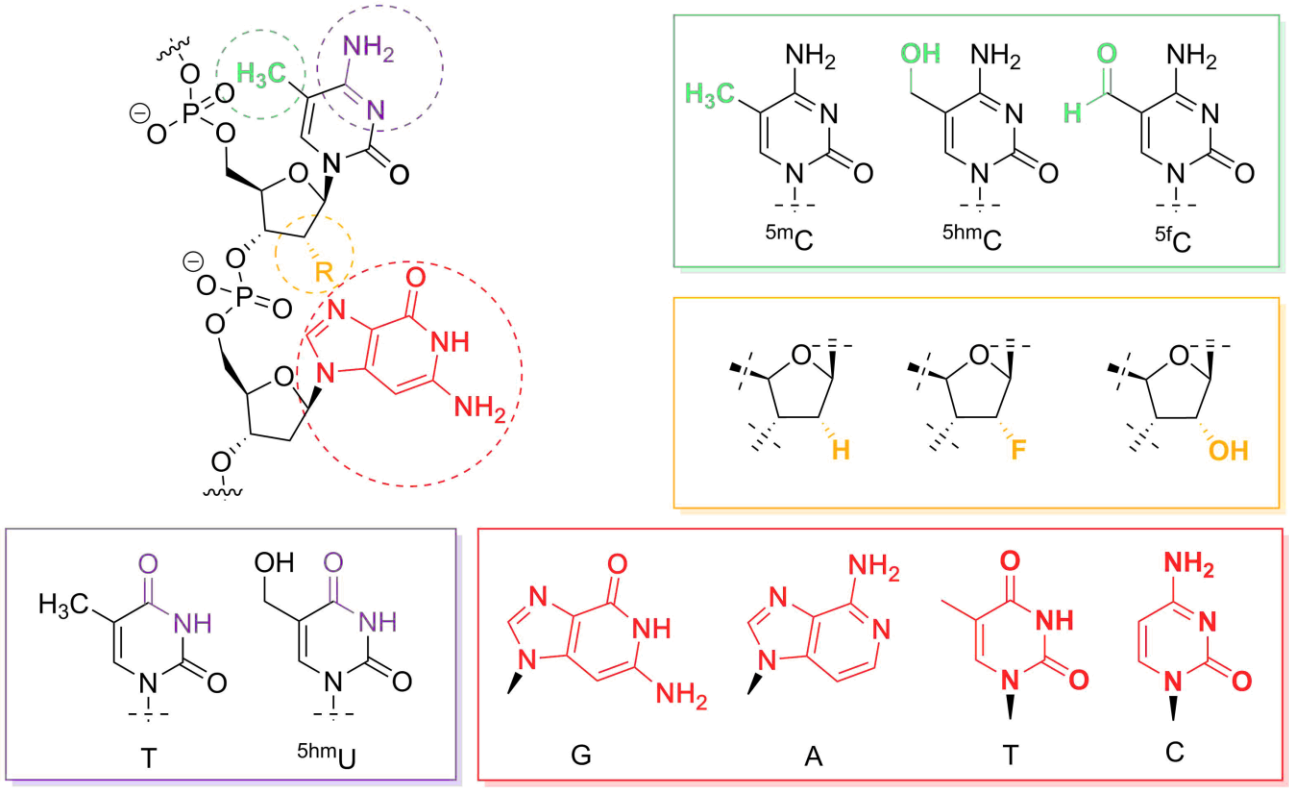


Figure 1.2.4: Schematic representation of DNA modifications that influence the substrate turnover-rate by TET proteins. Green box: 5-cytosine modifications; Yellow box: 2'-modifications; Purple box: thymine modifications; and Red box: bases pairing to cytosine. Furthermore, TETs can react on both single and double stranded substrates.

Modifications on the 2'- of DNA such as 2'-hydroxy (RNA) and 2'-(*R*)-fluorine show a decrease in oxidative performance (Fig. 1.2.4 Yellow).[57,64,65] hTET2 crystal structure reveals an interaction between base sugar of the ^{5m}C nucleotide and S1286 of hTET2 suggesting that 2'- modifications may influence the binding affinity and thus the turnover rate.

It is generally believed that the TET catalytic domain is not actively involved in DNA sequence recognition. This is partially reflected in the non-sequence specificity and oxidation rate dependence on the DNA length.[61] A longer oligonucleotide typically decreases the rate and the overall DNA sequence typically plays a little role in affecting the rate observed.[61] Interestingly however, the 3'-pairing nucleotide interacts with the TET protein altering the catalytic activity (Fig. 1.2.4 Red). Hashimoto *et al.* observed that in NgTet1, CpG is optimal for activity with decreasing efficiency for

CpA (~1.8-fold), CpT (~3.8-fold) and CpC (~5.8-fold).[66] Similar observations were made for hTET2 were DNA in ^{5m}CpG context demonstrated +85% conversion but ^{5m}CpC and ^{5m}CpA showed conversion rates of <10% and <2%, respectively.[45] The TET2-DNA crystal structure complex reveals that G7:^{5m}C7' interacts with Y1294 when the ^{5m}C6 is in the active site, and it is likely that A:T, C:G and T:A do not have this favourable interaction to Y1294 and consequently reducing the catalytic activity.

Finally, TET proteins can oxidize single stranded and double stranded DNA; however, the efficiency of the double stranded or single stranded DNA as substrates are currently a matter of debate.[65,67]

1.2.7 The abundance of ^{5ox}C are different among tissues

Approximately 3.5% – 4.5% of all cytosine in the human genomic DNA is methylated and this remains relatively evenly distributed among different cell lines.[31] The abundance of ^{5hm}C, ^{5f}C and ^{5ca}C is substantially lower with (typically) each iterative oxidation and is in contrast to ^{5m}C highly cell line specific. A relatively high level abundance of ^{5hm}C, a stable DNA modification in cells,[68] is found in the central nervous system (0.35–0.65% % ^{5hm}C relative to G) while low levels are observed in for example testis and spleen (0.03–0.05% ^{5hm}C relative to G). Purkinje neurons are observed to contain 40% ^{5hm}C relative to ^{5m}C which is approximately ~1.8% of total ^{5hm}C relative to G.[41]

Oxidized products of ^{5hm}C, ^{5f}C and ^{5ca}C, are believed to be part of an active indirect demethylation pathway. Despite this, ^{5hm}C and ^{5f}C are found to be stable modifications.[69] The abundance of ^{5f}C is substantially lower than ^{5hm}C, ranging between 12 parts per million (ppm) found in the brain to 0.5 ppm in lungs as detected in mice[69]; however, some variation in numbers have been observed.[43] The levels of ^{5ca}C in mice embryonic stem cells are ~5 ppm ^{5ca}C/ C. The abundance is lower in 12-week

old female mice were the highest recorded abundance were found to be the liver at ~0.8 ppm ^{5ca}C / C.[69]

It is noteworthy that the genomic abundance of ^{5ox}Cs are mostly independent of each other; a lower oxidation state abundance does not necessarily reflect a higher oxidation state(s) abundance. However, in some case an correlation between ^{5hm}C-^{5f}C and ^{5f}C-^{5ca}C has been detected.[43,70]

1.2.8 Biological relevance of ^{5ox}C and TET enzymes

The functional role of TET proteins is progressively better understood by the scientific community but, the full implication on a cellular level, which is accompanied with role of ^{5ox}C, remains only partially understood. Here, an overview of the functional role of the modified cytosine and TET proteins are described.

An increased abundance of ^{5hm}C is found near active genes region, particular at the transcription starting site and in exons[71,72] and there is a direct correlation between cellular proliferation and the ^{5hm}C abundance. In contrast, ^{5m}C is considered an epigenetic mark linked primarily to gene silencing when found in the promotor region.[68] While a study by Raiber *et al.* proposed that ^{5f}C influences the structure of DNA,[73] Hardwick *et al.* did not confirm this and proposed that it maybe context specific.[74] Furthermore, ^{5f}C has been proposed to serve as anchoring points *in vitro* and *in vivo* for histone lysine to form a stable Schiff base nucleosomal complex.[75] This enables stabilization of the DNA-histone complex preventing disassembly or rotary movement influencing nucleosomal organization. Furthermore, the marks ^{5f}C and ^{5ca}C are actively probed by chromatin reader modules or other enzymes with a recognition site.[76] Examples include the CXXC of hTET1, that promotes gene transcription by positioning the oxygenase active site near other ^{5m}C / ^{5ox}C.[77] A reader modules is also found in RNA

polymerase II that actively reads the DNA and slows down significant in the presence of ^{5f}C and ^{5ca}C while the presence of C, ^{5m}C of ^{5hm}C only shows little effect.[78]

TET1 was discovered when the gene encoding for the CXXC domain of TET1 was observed to fuse with myeloid mixed lineage leukaemia 1 (MLL1, KMT2A) gene resulting in elevated levels of ^{5hm}C leading to AML.[3,79,80] TET2 and TET3 were identified shortly after and TET functional assignment was made in 2009.[42] TET1 is also found to function in the HMGA2–TET1–HOXA9 pathway as a tumour suppressor in breast and prostate cancer.[81,82] Additionally, TET1 is down regulated in some solid tumours like colorectal and gastric cancer.[83]

TET2 is linked to haematopoietic malignancies[84] and is the most mutated gene in myelodysplastic syndrome (MDS).[2] Furthermore, *TET2* gene is mutated in AML resulting in either a frameshift, silent mutation, preliminary termination (nonsense) or change in amino acid (missense) that may lead to a decrease of the catalytic efficiency or protein truncation.[85] TET1, TET2 and TET3 have also been associated with human lymphoid malignancies including chronic lymphocytic leukaemia (CLL).[86,87] A lower abundance of ^{5hm}C has been detected in pancreatic, liver, prostate and lung solid tumours which is directly associated to a decrease in *TET* gene expression.[4]

TET2 mutations are mutually exclusive with mutations in Isocitrate dehydrogenase (IDH) in AML.[88] Interestingly, IDH1/2 wild-type converts isocitrate in 2OG; however, when the mutation IDH1: R132H or IDH2: R140Q/R140K occurs, isocitrate is converted to 2-hydroxyglutarate (2HG) which can accumulate in cells up to 30 mM.[89,90] A recent study demonstrated that the 2HGs are inhibitory to TET2 protein (IC_{50}: D-2HG: 5.3 mM, IC_{50}: L-2HG: 12.4 mM).[46]

1.2.9 Chemical probes are vital to the biochemical characterisation of TET enzymes

The combined observations reveal important cellular functions of TETs in development and disease, including both suppressor and oncogenic roles. TETs may, in some cases, be a therapeutic target and to further understand the biological role of TETs, chemical inhibition of the TET function could assist in elucidating their cellular functions and their role in development. No potent or selective TET inhibitors have been published to date, yet there is a significant need for such chemical probes.

1.3 Post translational modifications on histones

DNA cytosine methylation and methyl-cytosine oxidation are epigenetic modifications that are directly made to the DNA. Epigenetic marks are also found in another layer where the modifications are introduced after translation of the protein, known as PTMs. One major set of proteins that are extensively post-translational modified are amino acids found in humans histones. The most common studied PTMs on human histones are lysine / arginine methylation[91], histone lysine acetylation[92] and phosphorylation;[93] however, also other PTMs have been discovered including: β-N-acetylglucosamine[94], citrullination[95], glutathionylation[96], serotonylation.[97] formylation[98], propylation[99], crotonylation[100], butyrylation[101], β-hydroxybutyrylation[102], hydroxyisobutyrylation[103], malonylation[104], succinalation[104], biotinylation[105], ubiquitinylation[106], sumoylation[107] and mono-ADP-ribosylation[108]. The amino acids that are observed to contain PTM(s) on human histone H3.1 and H3.2 are displayed in Figure 1.3.1. The corresponding schematic amino acids with structures are displayed in Figure 1.3.2. In this thesis the focus of study of Chapter 3, 4 and 5 concerns histone lysine methylation and is therefore described in further detail below.

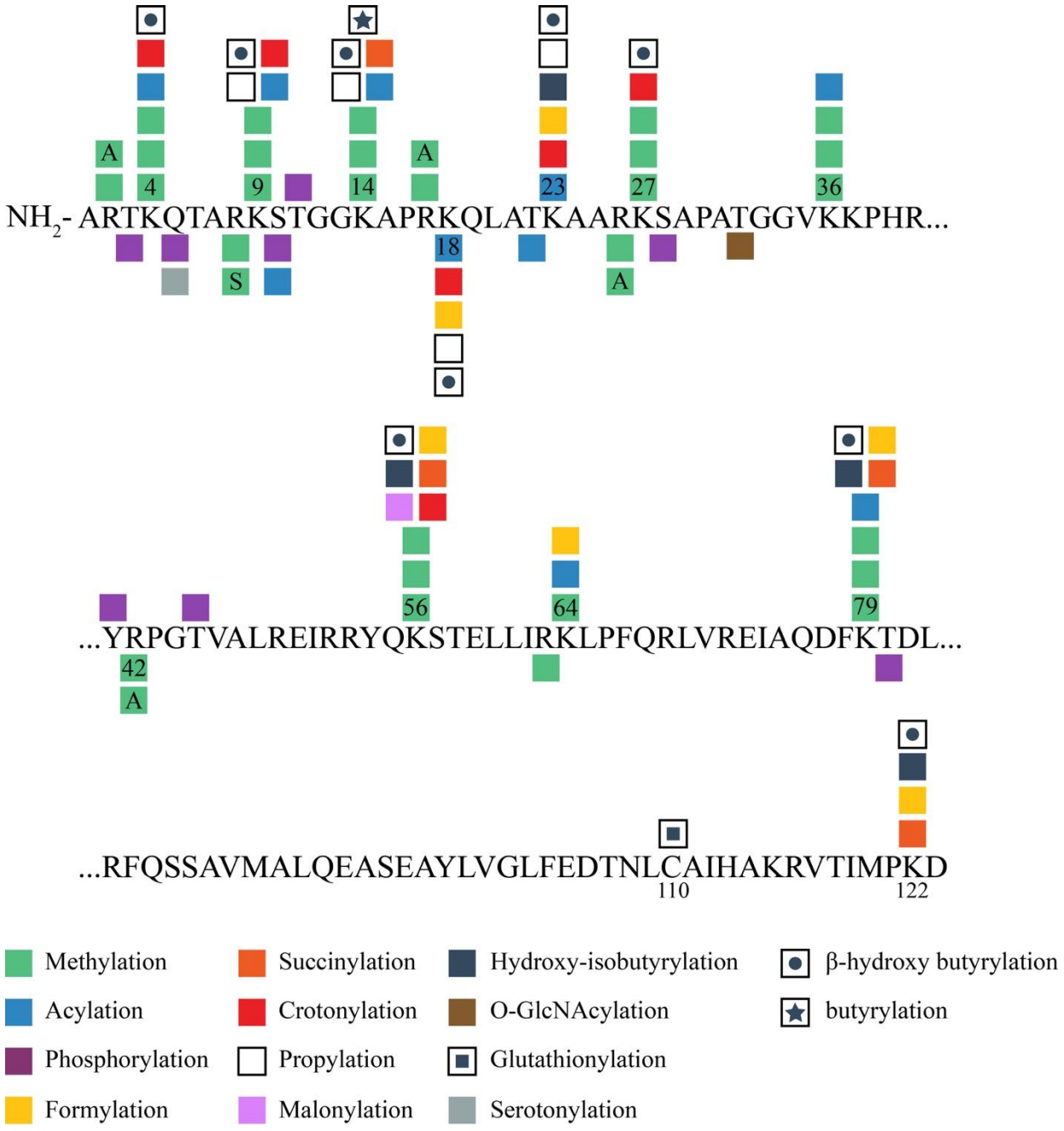


Figure 1.3.1: Overview of PTMs identified in human histone H3.1 and H3.2 highlighted as different squares above and below the sequence. Sequence displayed as H3.2. N^{ω},N^{ω}-dimethyl-arginine is annotated as S and N^{ω},N^{ω}'-dimethyl-arginine is annotated with A.

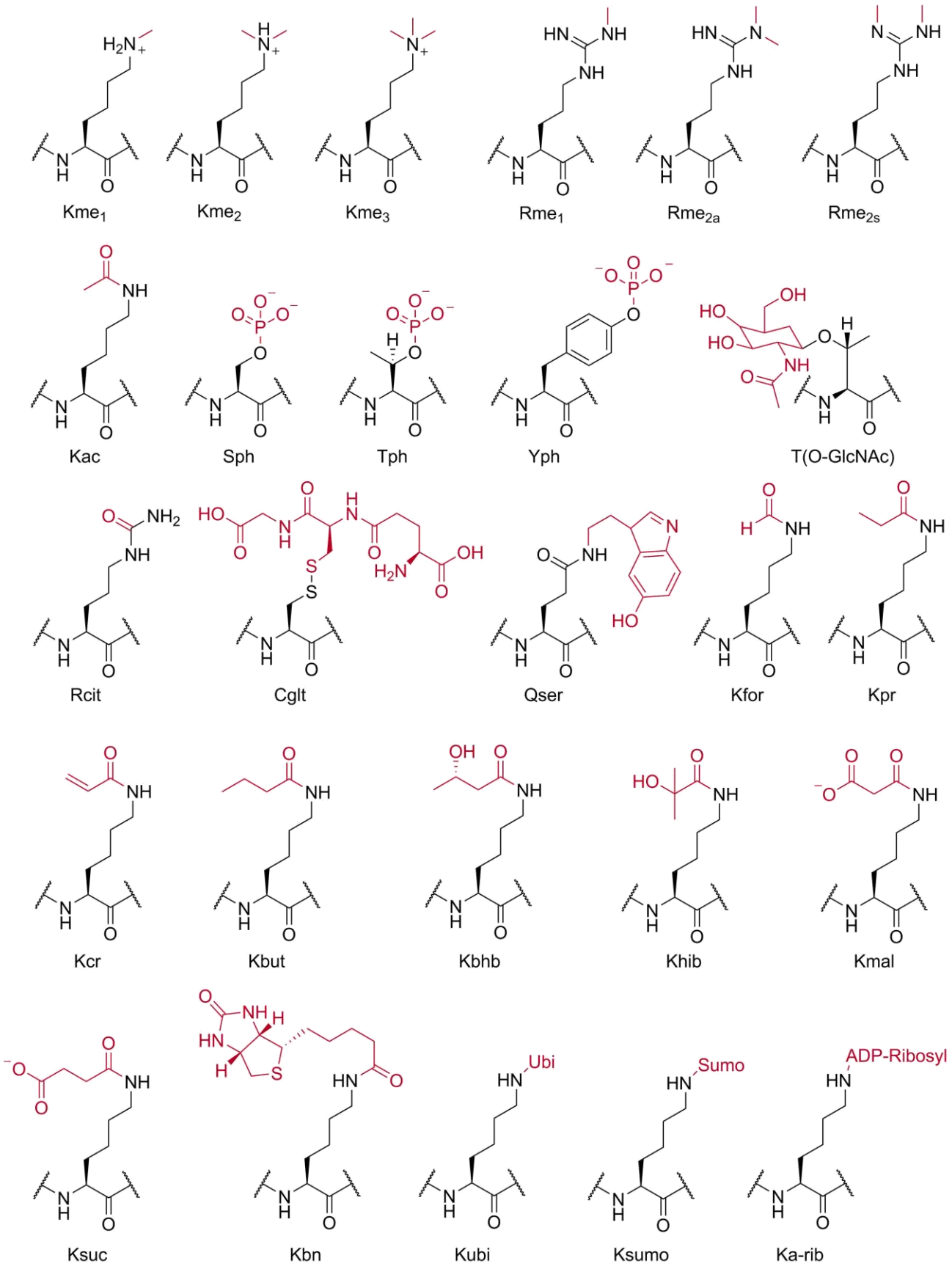


Figure 1.3.2: Schematic representation of PTMs found in human H3.1 and H3.2, with modifications (highlighted in red) on their designated amino acids. Kme_1: N^{ε}-monomethyl-lysine; Kme_2: N^{ε}-dimethyl-lysine; Kme_3: N^{ε}-trimethyl-lysine; Kac: N^{ε}-acetyl-lysine; Rme_1: N^{ω}-monomethyl-arginine; Rme_{2a}: N^{ω},N^{ω}'-

dimethyl-arginine (asymmetric); Rme_{2s}: N^{ω},N^{ω}-dimethyl-arginine (symmetric); Sph: O^{β}-phosphoryl-serine; Tph: O^{β}-phosphoryl-threonine; Yph: *O*-phosphoryl-tyrosine; T(O-GlcNAc): O^{β}-*N*-acetylglucosamino-threonine; Rcit: Citrulline; Cglt: S^{β}-glutathionyl-cysteine; Qser: N^{δ}-serotonyl-glutamine; Kfor: N^{ε}-formyl-lysine; Kpr: N^{ε}-propyonyl-lysine; Kcr: N^{ε}-crotonyl-lysine; Kbut: N^{ε}-butyryl-lysine; Kbhb: N^{ε}-[C^{β}-hydroxybutyryl]-lysine; Khib: N^{ε}-[C^{β}-hydroxyisobutyryl]-lysine; Kmal: N^{ε}-malonyl-lysine; Ksuc: N^{ε}-succinyl-lysine; Kbn: N^{ε}-biotinyl-lysine; Kubi: N^{ε}-ubiquitinyl-lysine; Ksumo: N^{ε}-sumoyl-lysine and Ka-rib: N^{ε}-mono-ADP-ribosyl-lysine.

1.3.1 Histone N^{ε}-lysine methylation is a dynamic mark and important in gene regulation

Methylation of histone residues has been observed on basic amino acids: lysine, arginine and histidine. Arginine can be methylated on the guanidine group to give N^{ω}-monomethyl-arginine (Rme_1) followed by secondary methylation that can result in symmetric asymmetric N^{ω},N'^{ω}-dimethyl-arginine (Rme_{2a}) or N^{ω},N^{ω}-dimethyl-arginine (Rme_{2s}).[109] Histidine has been observed to be only single methylated,[110] but histone lysine methylation can occur up to three stages by iterative methylation to give N^{ε}-monomethyl- (Kme_1), N^{ε}-dimethyl- (Kme_2) and N^{ε}-trimethyl-lysine (Kme_3).[111–113]

Histone lysine methylation is observed in humans on all four histones but most abundantly found on histone H3. The most studied methylation sites on H3 are K4, K9, K27, K36, and K79.[114] Three types of chromatin modifiers are associated to histone lysine methylation: 1) histone lysine methyltransferases (KMTs, writers) that add a methyl group; 2) chromatin binding modules (readers), proteins that recognize the methylation states of N^{ε}-lysine's and 3) histone lysine demethylases (KDMs, erasers) which reverse N^{ε}-lysine methylation. In this thesis, KMTs and KDMs are annotated following the latest nomenclature[115]; however, previously used names are mentioned.

N^{ε}-lysine methylation was originally considered an irreversible modification with relatively slow turnover rate in comparison to lysine acetylation.[116] This theory was disproved in 2004 with the discovery of lysine specific demethylase KDM1A (LSD1), an amine oxidase that uses flavin adenine

dinucleotide (FAD) and molecular oxygen as co-factors.[117] Shortly after, many more demethylases (>20) were discovered.

Histone N^{ε}-lysine methylation is strictly controlled in the cell and is linked to transcriptional activation and silencing.[118] H3K4me$_3$, H3K36me$_3$ and H3K79me$_3$ are typically associated to transcriptional active genes while H3K9me$_3$, H3K27me$_3$ and H4K20me$_3$ are enriched at repressive genes. Lee *et al.* illustrates an example of different histone lysine methylation levels that may be found at a gene *loci* (Fig. 1.3.3).[119] When a gene is repressed, a baseline level of H4K20me$_3$, H3K9me$_{2,3}$ and H3K27me$_3$ is observed, with increased levels of H3K27me$_3$ at the promotor site. When the gene is active, specific histone methylation signatures are observed across the gene.

Near an active gene, the levels of H3K9me$_{2,3}$, H3K27me$_{2,3}$ and H4K20me$_3$ remains the same until the promotor where repressing methylation marks are depleted and an increase in H3K4me$_{1\text{-}3}$ is observed. H3K4 methylation is only present at the promotor as well as near the transcription start site (TSS) while downstream the gene mainly H3K79me$_{1\text{-}3}$ is found as well as H2BK5me$_1$, H3K9me$_1$, H3K27me$_1$ and H4K20me$_1$. H3K36me$_3$ is observed near the end of the active gene, and at the transcription termination site. An increase of the repressing methylation marks H3K9me$_{2,3}$, H3K27me$_{2,3}$ and H4K20me$_3$ are observed.

A recent report suggests that H3K14me$_3$ is correlated to transcriptional activation in HEK293T cells but this was not confirmed in human embryonic stem cells suggesting a cell line dependent function.[120] H3K56me$_3$ is found in low levels and observed to coincide with H3K9 methylation.[121] Additionally, histone *N*$^{\varepsilon}$-lysine methylation is known to modulate gene regulation based on bivalent intra-histone or intra-nucleosome crosstalk overruling the individual effect.[122] Interestingly, Zee *et al.* demonstrated that the half-life of lysine methylation is dependent on the position of lysine methylation in the histone complex as well as the methylation state.[123]

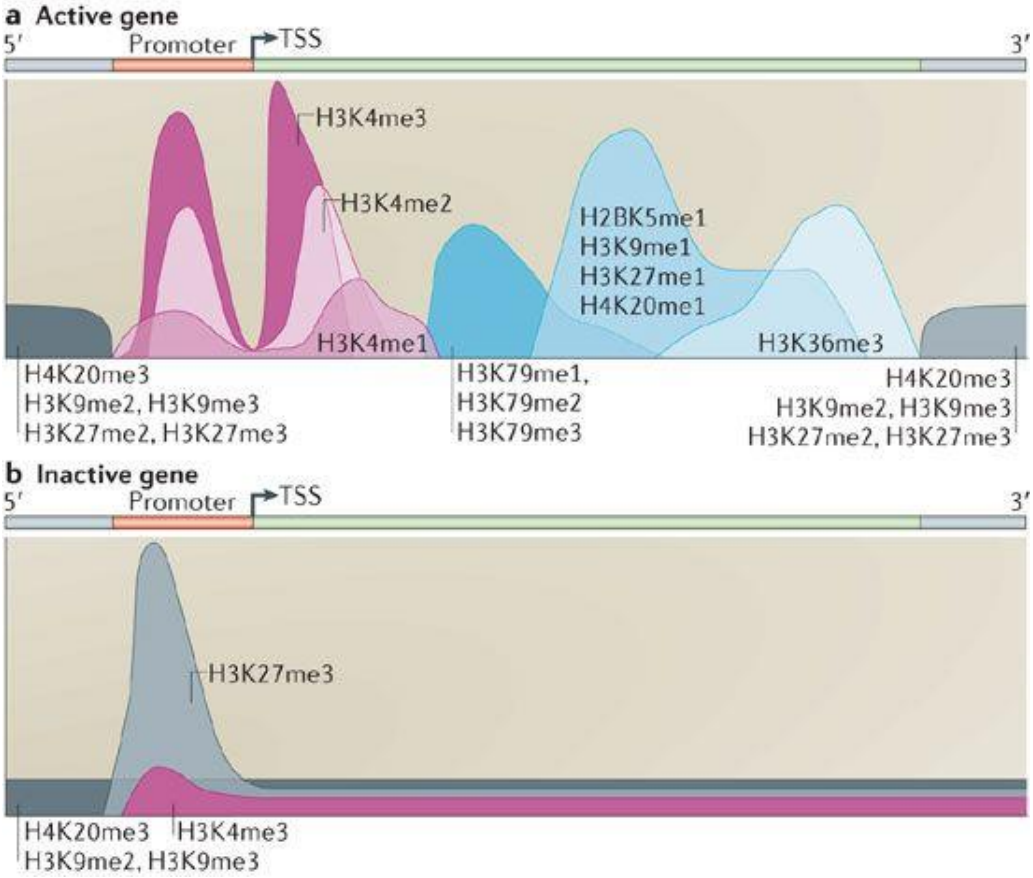


Figure 1.3.3: Schematic representation of histone lysine methylation at the designated position in the nucleosomes of active genes (a) and silenced genes (b). Used with permission from Springer Nature.[5]

1.3.2 Histone lysine methyltransferases methylate histone *N*$^{\varepsilon}$-lysine residues

Histone *N*$^{\varepsilon}$-lysine methyl groups are installed by KMTs that uses *S*-Adenosyl-L-Methionine (SAM) as a co-factor and as methyl source. In the proposed catalytic mechanism, neutral *N*$^{\varepsilon}$-lysine attacks the methyl of SAM generating the *N*$^{\varepsilon}$-methyl-lysine product and *S*-Adenosyl-L-homocysteine (SAH) (Fig.

1.3.4). This process can occur in sequential order in a bi-bi reaction to ultimately produce *N*ε-trimethyl-lysine.[124] It is however, the underlying histone sequence and KMT domain architecture that determines which lysine is methylated and the number of added methyl groups.[125]

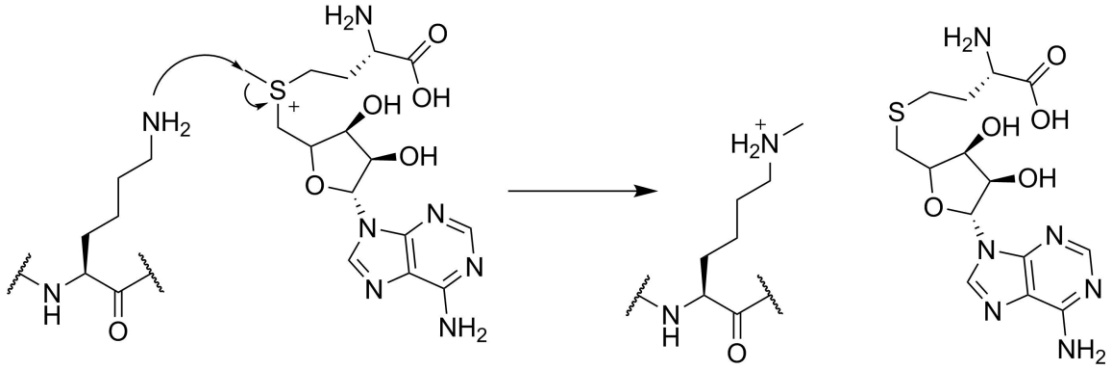


Figure 1.3.4: Proposed catalytic mechanism of histone lysine methyltransferases (KMTs) catalysing *N*ε-lysine residue methylation in the presence of SAM leading to *N*ε-methyl-lysine and SAH.

Many KMTs to date have been discovered containing either a Su(var) enhancer of zeste Trithorax (SET) domain or a disruptor of telomeric silencing-1 (DOT) domain.[126] The SET domain is typically ~130 amino acids in length and is characterised by the many short α/β-helixes with three highly conserved segments at both termini. Separation of the conserved segments is observed with various insertions. The DOT domain is only found in DOT1L methyltransferase which performs H3K79 methylation at all three stages using SAM as co-factor and methyl source.[127]

Many KMTs are linked to diseases,[91] as exemplified in mixed lineage leukaemia 1 (MLL1, KMT2A) and DOT1L (KMT4). MLL1, a H3K4 *N*ε-mono- and *N*ε-dimethyl-lysine transferase, is known to translocate with an array of proteins with the most common being AF4, AF9, ENL and AF10. This results in the production of the corresponding onco-fusion proteins leading to abolished KMT4A methyltransferase activity. Interestingly, AF4, AF9, ENL and AF10 interact with DOT1L[128], a protein known to play an important role in cellular development and is vital to gene transcription by increasing the levels of H3K79 methylation. The MLL1-fusion proteins cause the recruitment of DOT1L resulting in an upregulation of K79 methylation mark and induces the altered expression of MLL1 associated genes that leads, depending on the MLL-fusion protein, to different leukaemia's.[129–132] DOT1L has

also been linked to initiation and progression of breast cancer by upregulating H3K79 methylation and histone H3 acetylation, which interferes with HDAC1 and DNMT1.[133]

1.3.3 Amine oxidase KDM1 can demethylate *N*ε-methyl-lysine

Histone lysine demethylation is facilitated by KDMs that contain either an amine oxidase domain or a JmjC-domain. KDM1A (LSD1, AOF2) has an amine oxidase domain and is involved in H3K4me$_{2,1}$ demethylation using FAD and molecular oxygen as co-factors.[117] In the proposed catalytic mechanism, *N*ε-dimethyl- or *N*ε-monomethyl-lysine 4 is bound in the active site adjacent to the FAD. H3K4me$_{2,1}$ residues are deprotonated by KDM1A K661 followed by a hydride transfer from the neutral *N*ε-methyl-lysine to the FAD[134] oxidizing lysine 4 and reducing the FAD (Fig. 1.3.5). The newly formed iminium ion can, in the presence of water, hydrolyse to form an unstable hemiaminal that readily decomposes to release formaldehyde and the demethylated H3K4me$_1$ product. FADH is oxidized back to FAD upon exposure to molecular oxygen generating hydrogen peroxide in the process.

Recent literature proposes that the *N*ε-lysine 4 lone pair is essential to the hydride transfer mechanism[134,135] as such, a protonated *N*ε-methyl-lysine 4 or *N*ε-trimethyl-lysine 4 is not demethylated. KDM1A is a homologue of KDM1B (LSD2, AOF1) and a more distant homologue of mono amine oxidase A/B (MAO A/B) that is also found to react via a hydride transfer mechanism.[136] Interestingly, not all amine oxidase flavoproteins perform the reaction via a hydride transfer mechanism, some react through a radical transfer mechanism instead.[137]

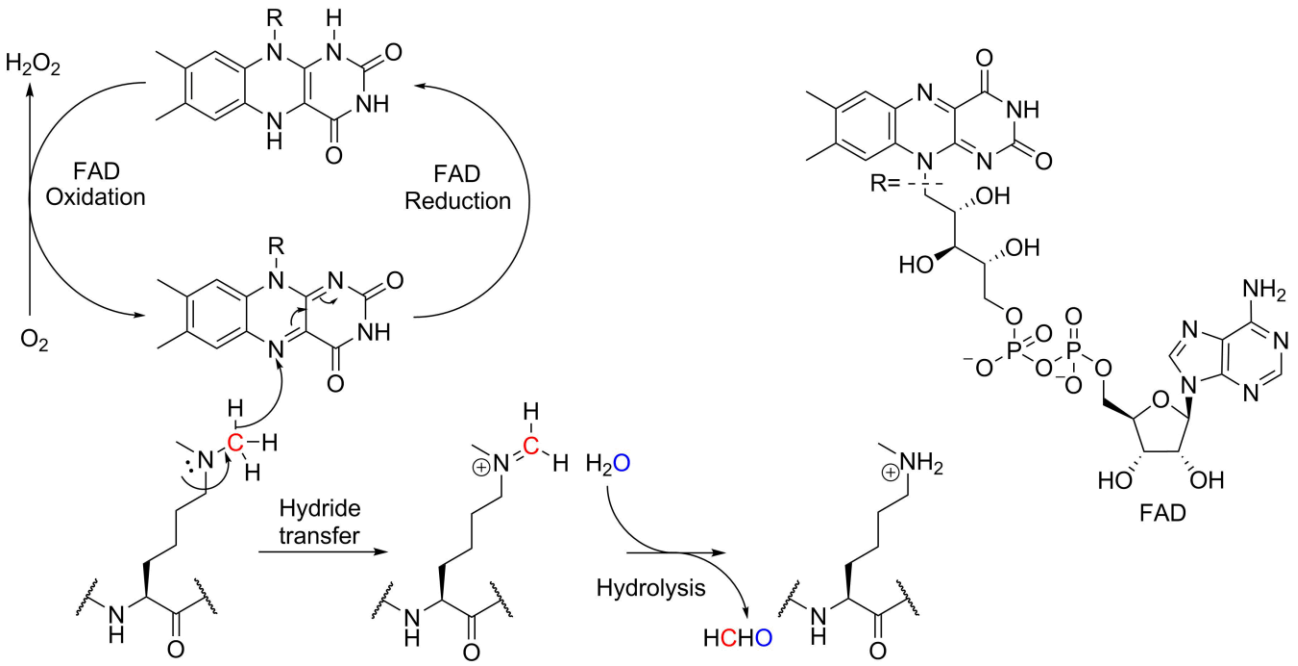


Figure 1.3.5: Proposed catalytic mechanism of KDM1A using FAD as co-factor to demethylate H3K4me$_{2,1}$ substrate. FADH regeneration is mediated by molecular oxygen to oxidize FADH giving FAD and hydrogen peroxide.

A vital element of KDM1A is the ability to 'read' the histone code. This is reflected in abolished H3K4me$_{2,1}$ demethylation activity in the presence of H3S10ph or decrease activity in the presence of H3K9ac.[138] Furthermore, the binding of transcriptional factor core-repressor CoREST to the tower domain of KDM1A is believed to modulate the active site by and improves the demethylation rate of N^{ε}-methyl-lysine 4 substrate.[139]

KDM1A is important for cell proliferation and is interestingly linked to DNMT1 regulation.[140] KDM1A is also involved in many cancers including: prostate cancer[141], bladder cancer[142,143], breast cancer[144,145], lung and colorectal carcinoma[142] making KDM1A an interesting pharmacological target.

1.3.4 The JumonjiC domain of KDM2–7 can demethylate N^{ε}-methyl-lysine

Shortly after the first report of KDM1A, histone lysine demethylases KDM2–7, a JmjC- subclass of the 2OG-dependent oxygenase superfamily, were discovered to catalyse histone N^{ε}-lysine demethylation. With ~60 proteins identified as 2OG-dependent dioxygenases[146] in humans, KDMs are

the largest sub-group (>20) to contain a JmjC-domain. Interestingly, the large 2OG-dependent oxygenase superfamily has a broad scope of reactions that they can catalyse, which includes: desaturation, hydroxylation, ring closure, ring opening, epoxidation and chlorination (Fig. 1.3.6).[147]

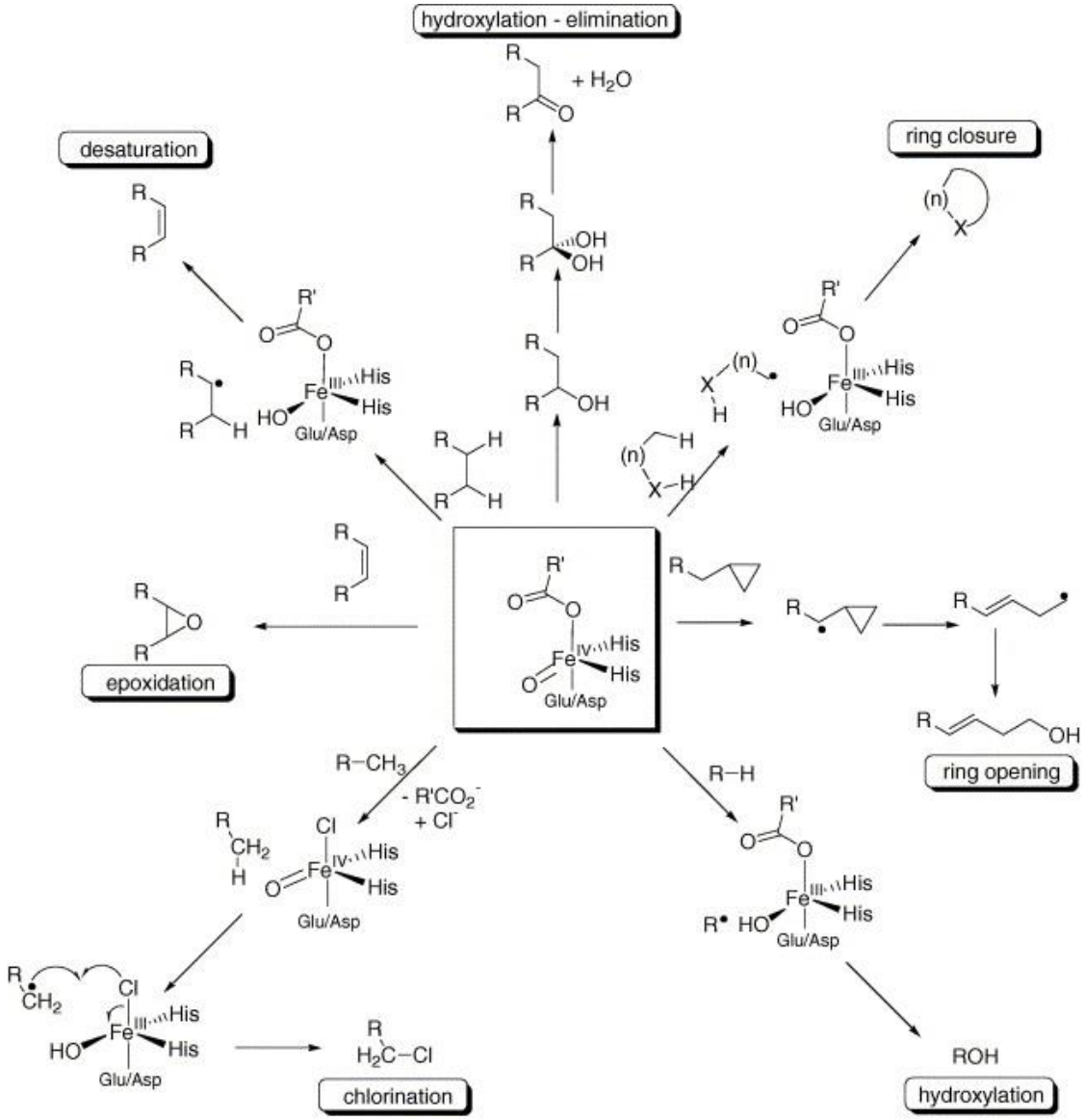


Figure 1.3.6: Schematic representation of reactions that 2OG-dependent oxygenase's are able to perform. Used with permission from Elsevier.[147]

The 2OG-dependent dioxygenases use non-haem Fe^{II}, 2-oxoglutarate (2OG) and molecular oxygen to perform the reaction. The domain is evolutionary and structurally conserved, with coordination of a catalytic Fe^{II}, bound by a triad of residues, HXD/E…H.[148] This active site is positioned within the asymmetric double stranded β-helix (DSBH) fold, also known as the 'jelly-roll' fold. The HXD/E coordination site is typically situated at the βII/βIII loop while a singular coordinating H is found on

the βVII strand but the location can vary. Although, two H and one D/E is the default triad configuration, protein KDM7C (PHF2) coordinates Fe^{II} using H249, D251 and Y321, substituting a tyrosine for a histidine.[149]

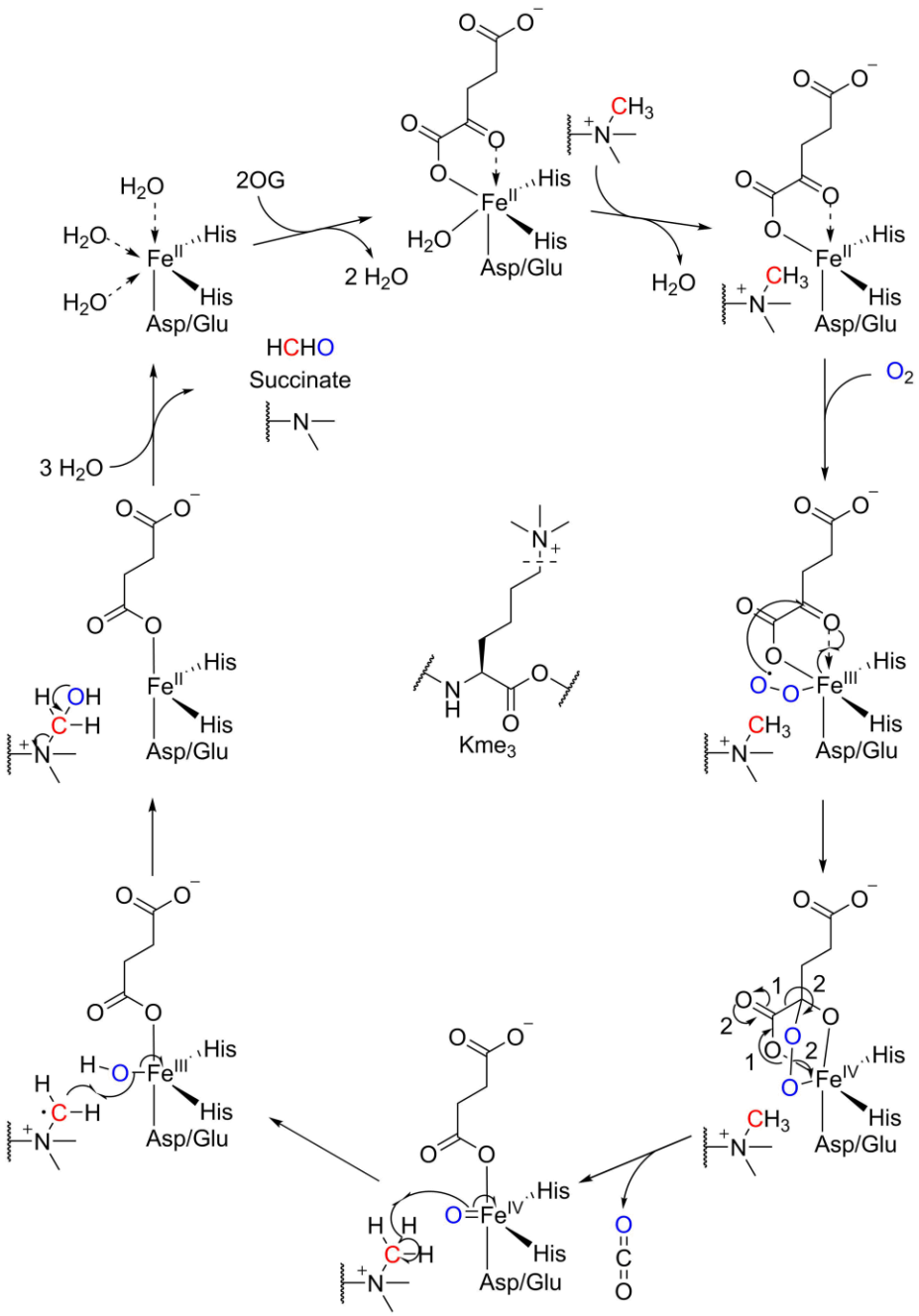


Figure 1.3.7: Proposed catalytic mechanism of 2OG-dependent dioxygenase enzymes here depicted with *N*ε-trimethyl-lysine substrate.

It is generally believed that KDMs follow a very similar mechanistic pathway to other oxygenase proteins (Fig. 1.3.7). The catalytic cycle is initiated by the displacement of two water molecules coordinating to Fe^{II} by 2OG.[150] When the lysine substrate binds (here displayed as *N*ε-trimethyl-

lysine,) the N^{ε}-lysine is positioned adjacent to the Fe^{II} displacing a coordinating water molecule. Molecular oxygen enters and reacts with the Fe^{II} generating Fe^{III} followed by a radical oxygen atom reacting with 2-position of 2OG and producing a Fe^{IV} complex. The complex rearranges to produce CO_2, succinate, and $O{=}Fe^{IV}$. The highly reactive ferrous metal reacts with the substrate by a radical hydrogen transfer forming a $HO\text{-}Fe^{III}$. The hydroxyl specie is then transferred to the radical substrate to give Fe^{II} and the hydroxylated substrate. In the catalytic cycle described in figure 1.3.7, the produced hemiaminal N^{ε}-hydroxymethyl-dimethyl-lysine is unstable and decomposes releasing formaldehyde and the corresponding demethylated N^{ε}-dimethyl-lysine product. The KDM catalytic cycle is completed by the release of the substrate, succinate, CO_2, and the entry of water.

Evidence in the proposed mechanism is found in the observations that substrate, O_2 and 2OG mediated by Fe^{II} catalysis yields the oxidized product succinate and CO_2 with improved activity in the presence of reducing agent (L-ascorbic acid)[151], as demonstrated in detailed kinetic studies using model 2OG oxygenase enzymes, such as AlkB, 2'-deoxyuridine-hydroxylase (DUH), ProHD and Factor inhibiting HIF (FIH).[152–155] The coordinated water to Fe^{II} is believed to be displace upon binding to the substrate.[148] Labelling experiments reveal that one of the oxygen atoms in succinate and CO_2 is sourced from molecular oxygen. Fe^{IV} intermediate was identified using electron paramagnetic resonance (EPR) and Mössbauer spectroscopy while confirmation of the Fe^{IV}-oxo specie was achieved using Raman spectroscopy and x-ray absorption spectroscopy using Taurine:α-KG dioxygenase (TauD)).[156–160]

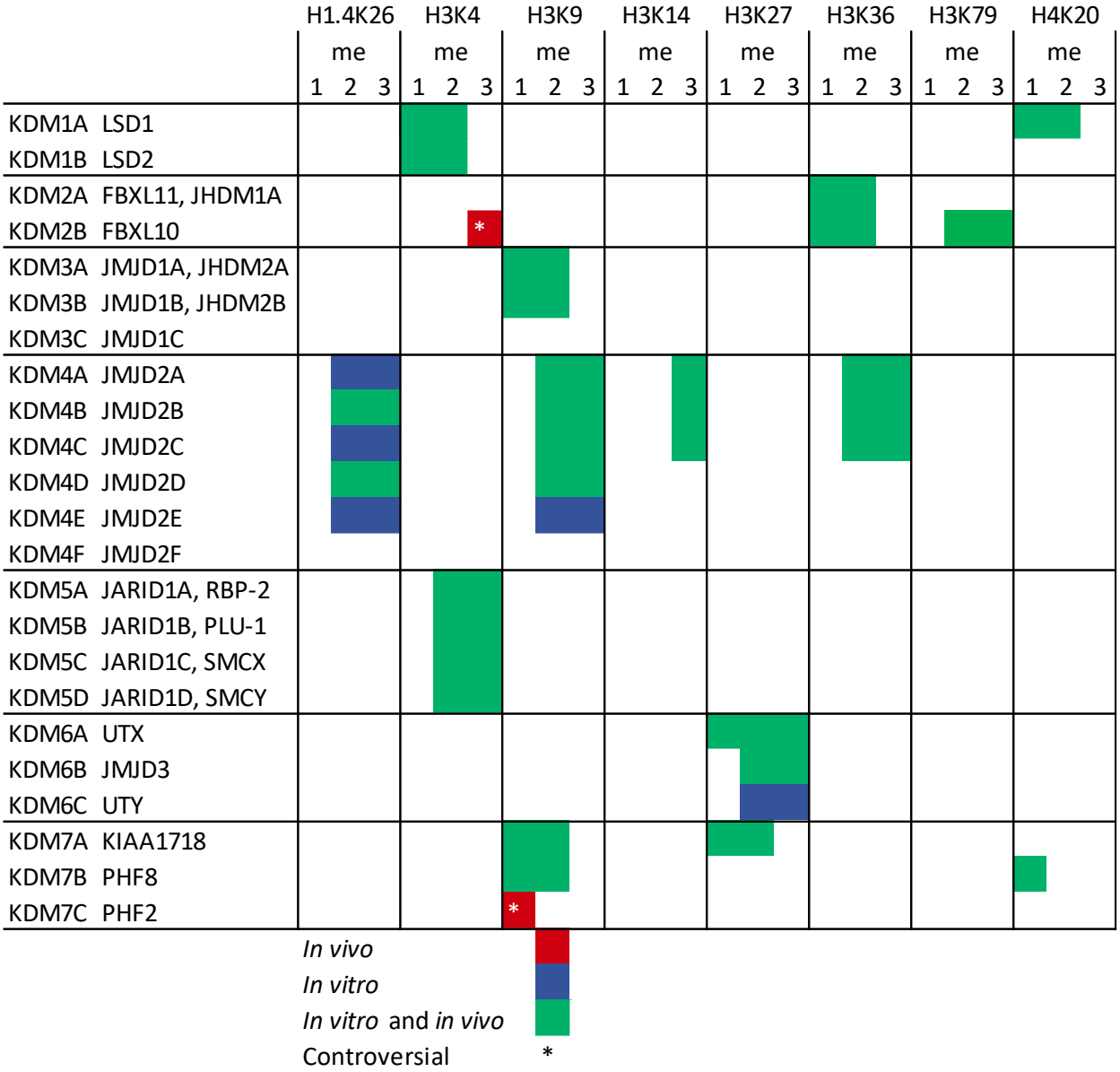
1.3.5 The KDMs can demethylate various N^{ε}-methyl-lysine residues

Lysine demethylases 2–7 (KDMs) are a major sub-group of the 2OG-dependent dioxygenases.[161] In order to demethylate specific histone lysine residues, the KDM is guided by histone peptide sequence recognition and consequently, KDMs are only able to demethylate specific sites on histones. The KDM active site architecture modulates the degree of demethylation for example: KDM4A is described to

favour higher methylation levels ($K9me_{3/2}$) with strongly reduced catalytic activity over lower methylation states ($K9me_1$). In contrast, KDM7B cannot demethylate $K9me_3$ but demonstrates activity for $K9me_2$ and $K9me_1$ demethylation. The currently known histone substrates for each KDM in humans is summarised in Table 1.3.1. Note that in some cases the substrates have only been discovered in an *in vitro* or *in vivo* setting and some assignments are controversial. JMJD5 (KDM8) for example was initially assigned as a $H3K36me_2$ demethylase[162] but this has not been independently confirmed.[163] Additionally, some KDMs also demonstrate activity on non-histone substrates suggesting a diverse biological role.[164,165]

Most identified KDMs contain ancillary domains with distinct properties. Chromatin binding modules such as Plant homeo-domains (PHDs) are observed in KDM2, KDM4A-C, KDM5 and KDM7 while tandem tudor domains (TTDs) are found in all KDM4 proteins. Both PHD and TTD are known to interact with histone N^{ε}-methyl-lysine residues and in some cases, such as KDM7B and KDM5A, the PHD finger modulates the demethylation activity.[166,167] Additionally, a DNA interaction domain, A-T Rich interaction domain (ARID) and CXXC zinc finger are found in KDM5s and KDM2s, respectively.[168,169] These ancillary domains are not exclusive to KDMs but are also described in other epigenetic regulating enzymes, as exemplified by the presence of CXXC in hTET proteins and ARID in JARID2 and ARID 1-5.[1,170] The PLU interaction domain is an exception to this and is exclusively to the KDM5 proteins.[171] Other domains that are found within the KDMs are F-BOX domain (KDM2), leucine rich repeat (LRR, KDM2) and C5H2C2 zinc binding finger (KDM5).[172]

Table 1.3.1: Known histone substrates for KDMs. KDM1A,[117,173] KDM1B,[174] KDM2A,[172] KDM2B,[172,175–177] KDM3A,[178] KDM3B,[179] KDM4A[120,180,181], KDM4B,[120,181–184] KDM4C,[120,181,182,185] KDM4D,[181,182,184,186] KDM4E,[181,187,188] KDM4F,[188] KDM5A,[189] KDM5B,[190] KDM5C,[191] KDM5D,[192] KDM6A,[193] KDM6B,[193] KDM6C,[194] KDM7A,[167,195] KDM7B,[167,196,197] and KDM7C.[149,198]



1.3.6 Biological relevance of KDMs

KDMs are of critical importance to the epigenetic landscape and are in part responsible for the histone lysine methylation levels. Disruption caused by, for example, overexpression, mutation or translocation may result in aberrant methylation patterns that are frequently associated with (a cascade of) cellular malfunction(s) and potentially disease(s). Studies show that KDMs can modulate the

activation or expression of oncogenes and tumour suppressor genes.[5,91] Interestingly, KDMs have been linked to multiple diseases and in some cases, have been proposed as attractive therapeutic targets. KDM2A is found to promote lung tumorigenesis[199] and KDM2B is highly expressed in bladder carcinoma[200] and leukaemias.[201] Overexpression of KDM3A is found in ovarian cancer[202], metastasized prostate adenocarcinoma[203] and colorectal cancer.[204] KDM3B levels are upregulated in bloods cells in patients that have acute lymphoblastic leukaemia[179] while mutations in JMJD1C, a protein with high sequence similarity to KDM3A/B, has been linked to intellectual disability and were found to contribute to the development of Rett syndrome.[205] KDM4 proteins have been linked to lung sarcomatoid carcinoma[206], breast cancer[207,208], bladder cancer[143], medulloblastoma[209] and oesophageal cancer.[210] KDM5A is found to be silenced or downregulated in retinablastoma[211] and fused with NUP98 caused by a translocation, observed in acute leukaemia.[8] KDM5B is overexpressed in breast cancer[212], prostate cancer[190] and bladder cancer.[213] KDM5C is linked to mutations with intellectual disability[214,215] and renal carcinoma[216] while a deletion of KDM5D was found in prostate cancer.[217] KDM6A is linked to several mutations in chronic myelomonocytic leukaemia (CMML)[218] and renal clear cell carcinoma.[219] KDM6A overexpression was found in breast cancer[207] while KDM6B is known to be overexpressed in haematological malignancies, liver and lung carcinomas.[220] Furthermore, a deletion of KDM6B enhances pancreatic proliferation.[221] Downregulation of KDM6C is correlated to cardiovascular risk.[222] KDM7A is linked to prostate cancer where KDM7A is involved in the regulation of the androgen receptor activity and therefore hormone dependent proliferation.[223] A mutation in KDM7B (truncation of the PHD-finger) is found to result in mental X-linked retardation accompanied with cleft in lip and palate.[196,224] PHF2 is linked to breast cancer[225] and head and neck squamous cell carcinoma.[226] The many diseases and developmental disorders KDMs are involved in indicate their vital importance to human life and have the potential to be therapeutic targets.

1.3.7 Inhibitor development for KDMs

KDMs are important epigenetic factors in the human cells that are only partially understood. In order to biochemical characterise the KDMs, the use of selective chemical inhibitors has proven to be useful. Rose *et al.* reported the first small molecule inhibitors for KDMs demonstrating micro-molar inhibition using the 2OG mimicking iron chelating inhibitors *N*-Oxalylglycine (NOG) and 2,4-pyridinedicarboxylic acid (2,4-PDCA) for KDM4E.[187] Both NOG and 2,4-PDCA were later found to be pan-2OG-dependent oxygenase inhibitor. Generating selective inhibitors for the ~20 KDMs have been met with great challenges. A common approach is to target the active site by generating 2OG co-factor mimics, or Fe(II) metal chelators. While this might provide selective over other non-2OG-dependent dioxygenases proteins, the 2OG active site is structurally similar among the JmjC-KDMs making it difficult to produce selective inhibitors or chemical probes. This is exemplified in 5-carboxy-8-hydroxyquinoline (IOX1), where a high-throughput screening against KDM4E identified a pan-2OG oxygenase inhibitor.[227] Nevertheless, improvements in selectivity have been made over the years, as evident for pyridine based inhibitor KDOAM25 against KDM5B. Tumber *et al.* demonstrated that KDOAM25 is a potent inhibitor for KDM5B (IC_{50} of 19 nM) while a 100-fold decrease was observed (IC_{50}: 4800 nM) against a structurally similar KDM4C.[228] Only limited potency (>50 µM) was observed in cellular assays. The structurally different pyrazolo-pyrimidine derivative CPI-455, developed by Genentech Inc. and Constellation Pharmaceutical Inc., shows an IC_{50} of 10 nM for KDM5A as determined using time resolved fluorescence resonance energy transfer assay. The compound inhibited in excess of 200-fold selectivity over KDM2B, KDM3B, KDM4C, KDM6A and KDM7B while also demonstrating cellular activity (EC_{50}: 5.2 µM). Little variation was found in the KDM5 family when tested against KDM5B (IC_{50}: 3 nM) and KDM5C (IC_{50}: 14 nM) suggesting that it could potentially be a cellular KDM5-specific probe.

Selectivity remains a critical element in the development of KDM inhibitors. After a screening campaign and lead optimization, GSK-J1 (*N*-[2-(2-Pyridinyl)-6-(1,2,4,5-tetrahydro-3*H*-3-benzazepin-3-yl)-4-pyrimidinyl]-β-alanine) was developed by GSK and the Structural Genomics Consortium. GSK-J1 is a potent inhibitor of KDM6B (IC_{50}: 60 nM) which binds to the active site Fe(II) in a bidentate manner displacing 2OG; however, it is selective over KDM2/3/4/7[229] but mildly inhibits KDM5B/C (< 5–10 fold IC_{50}).[230] Prodrug GSK-J4, a GSK-J1 ethyl ester, is cell-penetrating and leads to an increase in H3K27 methylation levels with a cellular dose response of ~9 μM in HeLa cells, suggesting on target protein effects.

To avoid potential off-target inhibition linked to metal chelation or cellular changes in 2OG, alternative inhibitors were considered that aim to be non-2OG competitive inhibitors. KDMs are observed to bind the histone substrate typically with K_M values ranging between 10^{-4}–10^{-6} M with high sequence specificity suggesting that small (5–21 amino acids) peptide fragments may function as inhibitors.[167] However, while improving KDM selectivity, there are concerns over the affinity for histone substrates for other reader proteins (K_d: 10^{-6}–10^{-8} M) [231], and therefore histone peptide fragments may show significant off-target effects. To address this, Woon *et al.* conjugated a NOG derivative to a histone H3 peptide fragment with H3K9 methylation.[232] High potency was achieved with KDM4A (IC_{50}: 0.27 μM) and KDM4E (IC_{50}: 0.09 μM) with an excess of 100-fold selectivity in potency over other KDMs (KDM2A, KDM3A, KDM6B, KDM7B).

Kawamura *et al.* reported a different approach to inhibitor development of KDM4A by applying mRNA-display to generate potent cyclic peptides inhibitors.[233] Inhibition potency for cyclic peptide 2 (CP2) was found to be IC_{50}: 42 nM while binding affinity was measured at K_d: 30 nM. KDM4B,C showed similar values to KDM4A but against KDM4D,E showed significant higher IC_{50} values of ~150-fold and higher. Attempts were made to introduce small modifications to the cyclic peptides in

order to improve cell permeability and enhanced cell stability. CP2.1–3 yielded a similar binding strength and good cell permeability; however, this was accompanied with potential cell toxicity.

1.3.8 Nuclear chromatin *N*ε-methyl-lysine binding modules

Chromatin binding modules manifest the ability to selectively recognise histone *N*ε-methylation-lysine resulting in either an increased or decreased binding affinity depending on the protein residues architecture interaction with *N*ε-lysine. Examples of reader proteins include: inhibitor of growth 2 (ING2) which has the highest affinity with histone H3 *N*ε-trimethyl-lysine (H3K4me$_3$), interacts with Histone lysine deacetylase (HDAC1) and can be linked to active gene repression.[234] Furthermore, KDM5A-PHD1 is a *N*ε-lysine 4 sensor, that contains the highest binding strength in the absence of methyl group and has been described to improve the demethylation rate of the KDM5A JmjC-domain.[166] TAF3 is a chromatin binding module with a preference for H3K4me$_3$ and is an integral part of the TAF complex that is involved in transcription.[235]

To date, many histone *N*ε-lysine methylation readers domains were discovered containing a Plant homeo domains (PHDs)[236] or a domain of the Royal superfamily. The Royal superfamily has five sub-domains namely: chromo domain (ChD)[237], double chromo domain (DCD)[238], tandem tudor domain (TTD)[239], double tandem tudor domain (DTTD)[180] and malignant brain tumour (MBT).[240]

Chromatin binding modules that sense *N*ε-methyl-lysine levels use a combination of structural and physical properties to differentiate between methylation levels. Proteins that prefer *N*ε-dimethyl- and *N*ε-trimethyl-lysine typically contain an ‘aromatic cage’ consisting of tryptophan, tyrosine and phenylalanine frequently accompanied by hydrophobic residues or acidic residues. A major contributing factor to be strong cation-π interactions[241,242] between the *N*ε-trimethylammonium warhead and aromatic residues while hydrophobic desolvation effects play a smaller role.[243] Recent

reports also suggest that the release of 'high energetic' water may be a contributing factor.[244] Chromatin binding modules that possess a higher binding strength for N^{ε}-monomethyl lysine or lysine are typically found to contain fewer aromatic residues and on the contrary, amino acids that enable hydrogen bonding to N^{ε}-lysine are frequently binding partners. Furthermore, these proteins are commonly observed to also introduce amino acids that can create sterical hindrance for certain methylation states.

There are found to be two types of domain architectures: surface groove and cavity insertion. Readers that are categorised to have a high affinity for higher N^{ε}-lysine methylation states typically have a surface groove binding motif while readers with a high affinity for lower N^{ε}-lysine methylation states possess typically a cavity insertion mode of binding. [6]

To date, many chromatin binding modules have been identified to be an integral part in reading the chromatin a disruption to this can lead to diseases. Examples of cellular disruptions are the translocation of KDM5A-PHD3 with NUP98 creating a NUP98-KDM5A-PHD3 fusion protein found in acute leukaemia.[8] BPTF, a N^{ε}-trimethyl-lysine binder is linked to deregulation of c-MYC oncogene, a gene that is known to be dysregulated in many human tumours.[245]

1.3.9 Creating highly selective KDM and chromatin binding module inhibitors remains challenging

KDMs belong to a larger family of 2OG oxygenases, which are structurally very similar and it is therefore challenging to develop target selective inhibitors. Part of the complexity lies in the combination of structural similarities in KDMs substrate recognition (*e.g.* N^{ε}-trimethyl-lysine) and a similar 2OG binding pocket prohibiting the exploitation of any individual binding mode. Secondly, targeting histone substrate binding sites of KDMs may mean that the inhibitor could cross-react with

chromatin binding modules / KMTs due to similar modes of N^{ε}-methylated lysine or histone sequence recognition. Finally, KDMs have evolved to exhibit a 2OG cavity with an adjacent N^{ε}-methyl-lysine binding pocket and central Fe^{II}. It is therefore vital to study and understand the boundaries and limitations of these proteins in for example substrate recognition or oxidative activity so that these could be potentially exploited to ultimately provide selective probes and inhibitors.

1.4 Thesis objectives

This thesis aims to unravel some of the issues described in section 1.2.9 and 1.3.9 with the key objectives outlined here:

- Biochemical characterisation of human TET 1 and TET2. (Chapter 2)
- Study the level of inhibition against human TET1 and TET2 by generic, (clinically relevant) target-specific 2OG-dependent oxygenase inhibitors, TCA cycle intermediates and oncometabolites. (Chapter 2)
- Biophysical and biochemical characterisation of KDM5A JmjC-domain and KDM5A-PHD3. (Chapter 3)
- Develop probe-like chemical tools to study N^{ε}-trimethyl-lysine specific chromatin reader binding modules. (Chapter 3, 4 and 5)
- Investigate the properties of the chromatin reader modules aromatic cage using sterically enlarged N^{ε}-trimethyl-lysine derivatives. (Chapter 3, 4 and 5)
- Study the plasticity of KDMs JmjC-domain acceptance towards unnatural N^{ε}-tri/dimethyl-lysine derivatives. (Chapter 3 and 5)
- Determine if the JmjC-domain catalytic activity can by modulated by the PHD of KDM7B or PHD3 of KDM5A. (Chapter 3 and 4)

Parts of the work presented here are also described, an extension of, or related to work outlined in the theses written by Dr. Yijia Zhang[†], Dr. Richard J. Hopkinson[†], Dr. Louise Walport[†], Dr. Rebecca Louise Hancock[†], Dr. Rachel Shamo Schiller[†], Dr. Kiran Kumar[†], Dr. Hanna Tarhonskaya[†] and Abbas Al temimi[‡*].

† Bodleian library, University of Oxford, Oxford, United Kingdom
‡ Radboud University Nijmegen Thesis Repository, Nijmegen, the Netherlands
* Manuscript in preparation



Chapter 2. Biochemical characterisation, assay development and inhibitor screening for human TETs

2.1 Contributors to this Chapter

Recombinant hTET2$_{CD}$ and hTET2$_{CD\Delta LCI}$ were produced by Aleksandra Szykowska (Structural Genomics Consortium) and Eidarus Salah (Kawamura / Schofield group), respectively. Hiroaki Suga's laboratory at the University of Tokyo carried out the mRNA-display selection, cyclic peptide synthesis, SPR and DOT-blot assay, while Shigeyuki Yokoyama's laboratory (RIKEN) carried out the cloning and production of recombinant hTET1$_{CCD}$.

Publication arising from work discussed in this Chapter:

Kosuke Nishio, Roman Belle, Takayuki Katoh, Akane Kawamura, Toru Sengoku, Kazuharu Hanada, Noboru Ohsawa, Mikako Shirouzu, Shigeeyuki Yokoyama and Hiroaki Suga. Thioether Macrocylcic Peptides Selected Against TET1 Compact Catalytic Domain Inhibit TET1 Catalytic Activity, *ChemBioChem*, **19**, 979-985 (2018).

2.2 Introduction to TET proteins

Epigenetic modifications on DNA impact gene regulation, transcription, and protein expression. The most common studied modifications are 5-methylcytosine (^{5m}C), and the oxidized derivatives: 5-hydroxylmethylcytosine (^{5hm}C), 5-formylcytosine (^{5f}C) and 5-carboxylcytosine (^{5ca}C). ^{5m}C is produced by methylating cytosine (C) facilitated by DNA methyltransferases (DNMTs) using *S*-adenosyl-L-methionine (SAM) as methylation source. ^{5hm}C, ^{5f}C and ^{5ca}C are produced in sequential oxidation by Ten-Eleven Translocation (TET) proteins using the 2OG-dependent oxygenase domain performing Fe^{II} catalysis in the presence of co-factors 2OG and molecular oxygen.

2.2.1 TETs and ^{5ox}C play multiple functional roles in the cell

The function of the ^{5m}C and ^{5ox}C in the cell are only partially understood. At the promotor-region of a gene is ^{5m}C associated to a gene silencing whereas, ^{5hm}C is linked to transcriptional activation.[71,72] The function of ^{5m}C; however, is positional dependent as observed for ^{5m}C at the enhancer region[34] were it is believe to be transcriptional activating. Interestingly, Reik *et al.* reported a correlation between age and ^{5m}C patterns allowing for age determination by DNA methylation sequencing.[33] ^{5f}C is linked to nucleosome stabilization by forming a Schiff base with the histone lysine's.[75] Combined, the ^{5m}C and ^{5ox}C are part of an active signalling pathway reflected in the many chromatin binding modules found interacting with these epigenetic DNA modifications.[246] The 5-oxidized-methylcytosine (^{5ox}C) by TETs have multiple implications to the chromatin. The ^{5ox}C are involved in passive and active ^{5m}C demethylation pathways. While ^{5ox}C are considered to be part of passive dilution through active modification followed by subsequent cell division; active ^{5m}C TET mediated oxidation to ^{5f}C and ^{5ca}C can follow the removal by thymine DNA glycosylase (TDG) Base Excision Repair (BER) linked pathway to generate a 'demethylated' C:G pair.[55] Beside this, other demethylation pathways have been proposed linked to TET proteins.[58]

The collective evidence demonstrates the importance of hTETs in the chromatin and concomitant aberrant epigenetic patterns as a result of TET malfunction could lead to disease. The functional role of TET proteins is currently only partially understood and the use of chemical probes could aid to a better understanding.

2.2.2 *In vitro* enzymatic assay to study the catalytic activity of TETs

To date, various methodologies have been developed to detect and quantify the oxidation of ^{5m}C by TET proteins. The analytical techniques used are based on one or more of the four categories: mass spectrometry, radiolabelling, antibodies, and luminescence.

A number of researchers have reported LC-MS/MS and LC- tandem triplequadrupole MS based methods where nucleosides ^{5m}C, ^{5hm}C, ^{5f}C and ^{5ca}C are quantified. Oligonucleotides can be purified (*e.g.* from TET reaction) and enzymatically digested to obtain nucleosides, and quantitated on LC-MS/MS using nucleoside standards to yield high levels of sensitivity (femtomole range).[43,61] Alternatively, to avoid the digestion step and to be able to directly analyse the oligonucleotide, MALDI-TOF MS can be used for the detection of intact oligonucleotides. While this techniques works, it is unable to differentiate between ^{5hm}C and ^{5f}C while also the sensitivity is reduced proportionally to the length of oligonucleotide.[46,247] As alternative methods, radiolabelling techniques have been developed. DNA can be digested with MspI and calf alkaline phosphatase to give the nucleosides, and phosphorylated using [γ-^{32}P]ATP by T4 polynucleotide kinase. The mixture is then treated with DNase I and Phosphodiesterase, separated on a TLC plate, exposed to X-ray film and analysed to provide relative abundance of ^{5m}C and ^{5ox}C.[248] The second radiolabel method uses 2-oxo[1-^{14}C]-glutarate as cofactor and quantifies the level of formed $^{14}CO_2$ as a result of the consumption by TETs facilitated reactions.[249]

For antibody assays, Szulwach *et al.* demonstrated a Dot-blot method that utilizes antibodies for the quantitation[250] of ^{5m}C and as alternative, Epigentek developed a commercially available ELISA assay. The two methods rely on the specificity of the ^{5hm}C-specific antibodies for the detection of ^{5hm}C in the presence of ^{5m}C and ^{5ox}C. Finally, succinate, a by-product of TET catalysis, can be coupled to the production of ATP generating bioluminescence, and this forms the basis of Succinate-Glo™ assay (Promega).

The current literature lacks the use of screening technologies that are sensitive, robust, rapid, low volume and applicable for high throughput assays. While the mass spectrometry-based LC-MS/MS show good accuracy, the oligonucleotide digestion and the relative sample rate makes this method low throughput. The MALDI-TOF attempts to address these issues by directly analysing the oligonucleotide without any purification; however, this requires the use of desalting step by the addition of cation exchange resin to aid crystallisation with the MALDI matrix, which also prohibits sample storage. Additionally, the MALDI-TOF requires manual data collection and analysis to gain acceptable results for quantitation. The Succinate-Glo™ assay relies heavily on the readout of succinate which may also be formed during uncoupled turnover and relies on a multi-step enzyme coupled assay. In order to develop new methods applicable for high-throughput assays, the assay has to be sensitive (<1.0 µM quantification of product), robust, reproducible, and able to be miniaturised (*e.g* assay volumes <50 µL). It is also important to develop multiple assays to independently verify the results to avoid assay-dependent interference or false positives / negatives, and if the technique permits detailed kinetic analysis.

2.3 Chapter objectives

The aim of this Chapter is to:

- Develop robust, sensitive and low volume enzyme assays for assessing the catalytic activity of TETs for kinetic analysis and inhibitor screening.
- Explore the catalytic activity on hTETs dioxygenases.
- Assess (clinically) relevant generic 2OG-dependent oxygenase, KDM, HIF hydroxylase and HDAC inhibitors against hTET proteins.
- Test the effect of TCA-cycle intermediates and TCA-cycle related oncometabolites on the hTET activity.

Throughout this Chapter, references made to the catalytic domain are described in capitals (e.g. CD) while references made to the catalytic domain construct are described in subscript (e.g. hTET1$_{CD}$).

2.4 Results

2.4.1 Production of human TET proteins

One of the key aspects of developing robust assays is a reliable source of active target enzyme. hTETs have historically been very difficult to produce recombinantly although some methods were reported at the start of this work. A literature survey was carried out to determine the most suitable hTET protein constructs for our biochemical and biophysical studies and the constructs were selected based on the following criteria: A) it is a hTET catalytic domain CD containing construct, which has been reported to be catalytically active (*i.e.* hydroxylate ^{5m}C); B) hTET construct enables the recombinant expression of sufficient quantity and quality of protein for activity assays and C) the recombinant proteins are stable (*e.g.* during assay period, as well as for crystallography). With this criteria, the plasmids for the recently reported protein constructs hTET1 and hTET2 were requested and kindly gifted by Hu *et al.*[61] The hTET1, construct consists of the CD with split DSBH and LCI (hTET1$_{CD}$, E1418–V2136). The construct encoding for hTET2, like hTET1 contains the CD but the LCI is removed and replaced by a GGGGS linker (hTET2$_{CD\Delta LCI}$, S1129–P1936 with L1482–D1842 replaced by 3×GGGGS-linker, Fig. 2.4.1).

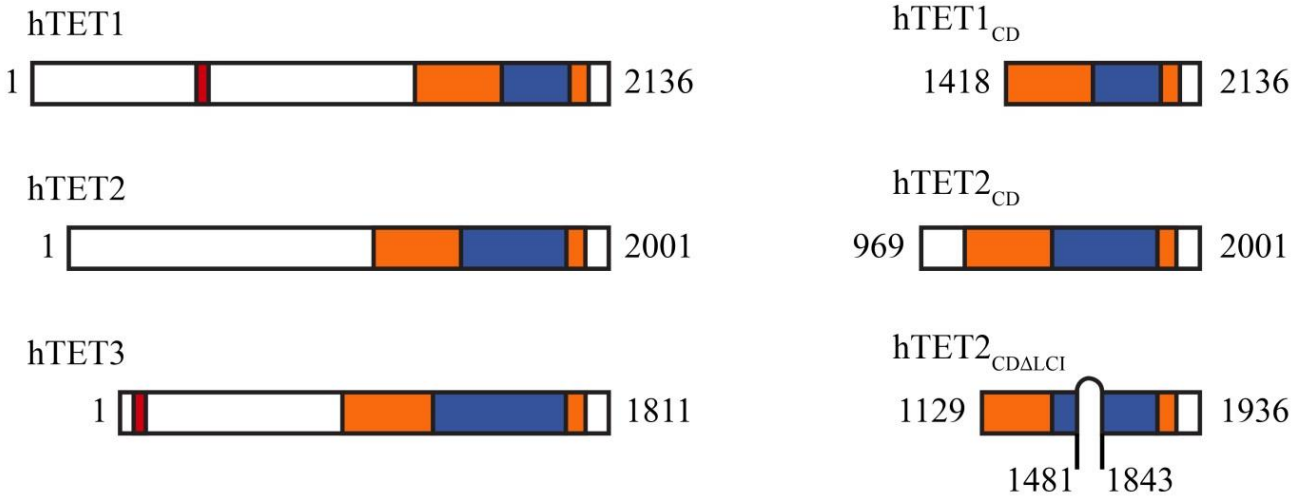


Figure 2.4.1: Three known hTET proteins (l) and hTET constructs used (r) with domains highlighted red: CXXC-domain, orange: split double stranded beta helix and blue: low complexity insert (LCI). CD: catalytic domain.

Following literature procedures, expression trials of hTET1$_{CD}$ (E1418-V2136, Fig. 2.4.1) in phage resistant and pRARE2 carrying *E. coli* yielded insoluble protein *(by Eidarus Salah)* and experiments were not further pursued. Thus, to perform kinetic studies on hTET1$_{CD}$ (E1418–V2136, Fig. 2.4.1), hTET1 protein containing N-terminal 3×flag-tag produced in Baculo virus / insect cells (Sf9) was sourced from Active Motif. Expression trial of N-terminal His-tagged hTET2$_{CD\Delta LCI}$ were carried out and yielded low-levels of active protein (see section 2.8.2). These conditions were used for protein expression at 6 litre scale to give 9.6 mg (15 µM, 1.0 mg·mL^{-1}) of hTET2$_{CD\Delta LCI}$ of pure protein *(by Eidarus Salah).*

To understand how the LCI affects the inhibition of hTET2, and to produce comparative data to hTET1$_{CD}$, the hTET2 CD with LCI (hTET2$_{CD}$, Q969–Y2001, Fig. 2.4.1) was produced in baculo virus / insect cells (Sf9) expression systems (*by Aleksandra Szykowska*) to give hTET2$_{CD}$ (~10 mg·mL^{-1}).

2.4.2 AlphaScreen™ technology as *in vitro* ^{5hm}C-DNA detection method

To investigate the catalytic activity of hTETs, AlphaScreen™ based assay was developed. AlphaScreen™ methodology is a luminescence-based assay which relies on the transfer of energy between donor and acceptor beads. For the TET assay setup, the streptavidin coated donor and Protein A acceptor bead pair was used. The streptavidin beads can capture the biotin-labelled single strand 32 base DNA (^{5m}C$_{ss}$DNA-[Biotin]) **2.1** (Fig. 2.4.2)[66] which contains a single modified ^{5m}C that can be oxidized by hTET to give ^{5hm}C. If ^{5hm}C is produced, it can be recognized by the ^{5hm}C-specific polyclonal antibody that is bound to the Protein A coated acceptor bead, thus establishing a donor-acceptor bead link. Chemiluminescence induced by a light pulse (680 nm) follows the release of singlet molecular oxygen from the donor beads that can react with nearby acceptor beads resulting in radiative emission (520–620 nm). However, if ^{5m}C is not converted to ^{5hm}C, the ^{5hm}C specific-antibody provides

poor binding to ^{5m}C and does not establish a (strong) donor-acceptor bead link resulting in a no or a low signal.

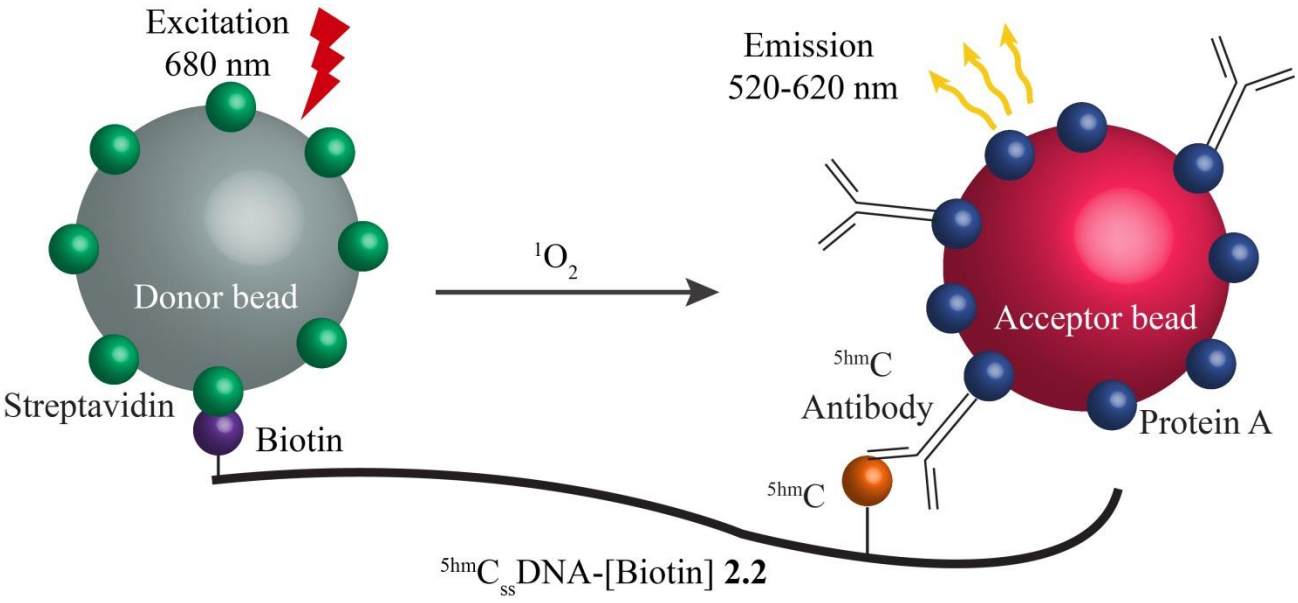


Figure 2.4.2: Schematic representation of AlphaScreen™. The donor (grey) bead is coated with streptavidin (green) that binds to ^{5hm}C (orange) containing single strand DNA with biotin tag (purple). The acceptor bead (pink) binds to ^{5hm}C nucleotide through the ^{5hm}C-specific antibody bound to protein A (blue) which is attached to the acceptor bead. Excitation at 680 nm will release singlet molecular oxygen ($^1O_2$) and nearby acceptor beads will convert this to chemical luminescence (520–620 nm).

To investigate the potential use of AlphaScreen™ as a method to determine TET oxidative activity, a DNA sequence, previously reported as a substrate for *Naegleria gruberi*[66], was chosen. This sequence contains ^{5m}CpG, provides an equal amount of each base in duplex and contains appropriate DNA length (>20 bp and <50 bp). Crude synthetic DNA, prepared by ATDBio Ltd (Southampton), was purified using HPLC. Selected fractions from the major peak were lyophilized and the obtained product was mass and purity confirmed by LC-MS (>95% purity).

AlphaScreen™ assays were conducted initially with double stranded DNA [Biotin]-5′-TCG GAT GTT GTG GGT CAG ^{5m}CGC ATG ATA GTG TA-3′ ($^{5m}C_{ds}$DNA-[Biotin]) **2.1**; while it was possible to obtain standard curves, the linear range was variable between experiments. This is likely caused by subtle differences in DNA concentration of each strands, as well as the potentially restricted access of

binding of the ^{5hm}C-specific antibody to the ^{5hm}C site duplex in the double stranded DNA. Since hTET proteins also efficiently catalyse the oxidation of ^{5m}C on single stranded DNA[65,67], all AlphaScreen™ assays were further carried out using single stranded DNA to gain maximal signal / noise and reproducibility.

To understand the selectivity of the ^{5hm}C-specific antibody and to determine the linear region of the AlphaScreen™ assay, standard curves were prepared containing: A) single ^{5m}C in 32 base-pair single stranded DNA with 5'-[Biotin] ($^{5m}C_{ss}$DNA-[Biotin]) **2.1**, B) single ^{5hm}C in 32 base-pair single stranded DNA with 5'-[Biotin] ($^{5hm}C_{ss}$DNA-[Biotin] **2.2**) and C) $^{5hm}C_{ss}$DNA-[Biotin] **2.2** with compensating levels of $^{5m}C_{ss}$DNA-[Biotin] **2.1** to give 10 nM DNA total (Fig. 2.4.3). DNA levels beyond 10 nM resulted in rapid saturation of the signal and detected a hook effect. The ^{5hm}C-specific antibody showed good selectivity between $^{5hm}C_{ss}$DNA-[Biotin] **2.2** (Fig. 2.4.3, red squares) in comparison to $^{5m}C_{ss}$DNA-[Biotin] **2.1** (Fig. 2.4.3, blue diamonds) with a difference in signal of >10-fold beyond 0.1 nM.

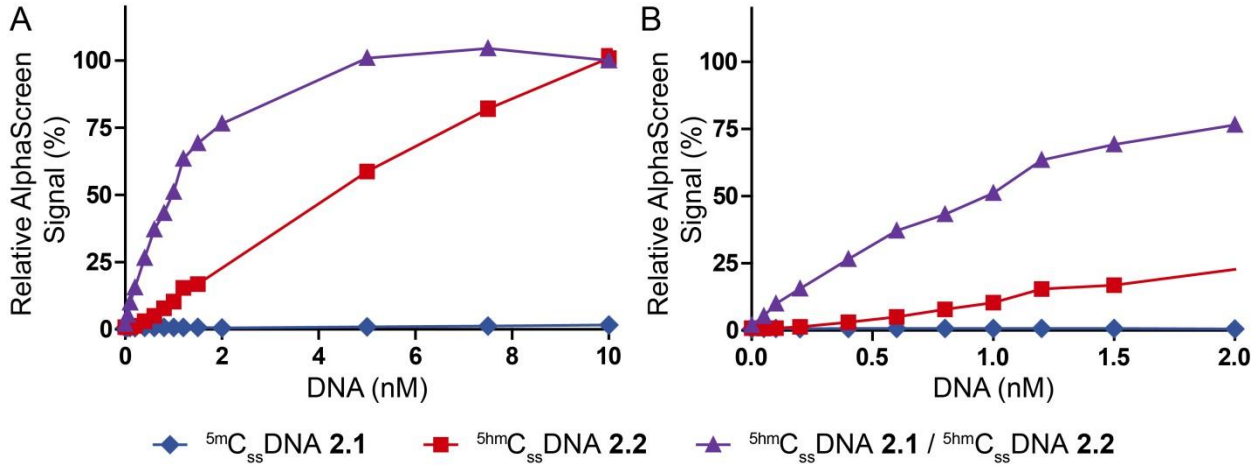


Figure 2.4.3: AlphaScreen™ assay $^{5hm}C_{ss}$DNA standard curve. (A) $^{5m}C_{ss}$DNA **2.1** (diamond, blue), $^{5hm}C_{ss}$DNA **2.2** and [$^{5hm}C_{ss}$DNA **2.2** / $^{5m}C_{ss}$DNA **2.1** (10 nM)] (triangle, purple) were tested showing low affinity of the antibody by ^{5m}C but high for ^{5hm}C. n= 4, error given as s.d. The AlphaScreen™ signal was normalised to the highest signal (saturation, 100%). (B) The linear region of the standard curve (0-1.5 nM). Optimization of AlphaScreen™ is required for each new batch and novel antibody/substrate combination as results may vary.

Furthermore, relatively small standard deviations were observed for the concentrations tested in a single assay and between assays, demonstrating good reproducibility. To understand the effect of a

constant level of DNA (10 nM) during an enzymatic reaction, $^{5m}C_{ss}$DNA-[Biotin] **2.1** + $^{5hm}C_{ss}$DNA-[Biotin] **2.2** (Fig. 2.4.3, purple triangle) was tested and the linear region of detection is found between 0–1.5 nM of $^{5hm}C_{ss}$DNA-[Biotin] **2.2** in the presence of $^{5m}C_{ss}$DNA-[Biotin] **2.1** (10–8.5 nM) with good correlation (R^2: 0.98).

To optimise the AlphaScreen™ for TETs, the most readily available TET protein, hTET2$_{CD\Delta LCI}$, was used. A time-course was carried out in the presence of different hTET2 $_{CD\Delta LCI}$ concentrations (0.25–100 nM), $^{5m}C_{ss}$DNA-[Biotin] **2.1** (10 nM), sodium ascorbate (Asc) (100 μM), $(NH_4)_2Fe(II)(SO_4)_2$ (10 μM) and disodium 2-oxoglutarate (2OG) (10 μM) in buffer [HEPES (50 mM), NaCl (150 mM), BSA (0.1%), Tween20® (0.01%) in MQ (pH 7.0)]. The reaction was quenched with EDTA solution [EDTA (30 mM) in MQ (pH 4.2)] and bead mixture [Donor (1:62.5), Acceptor beads (1:62.5), polyclonal ^{5hm}C-specific antibody (1:4000) in buffer] was added after all time-point were completed. The plate was incubated (35–60 min) at RT and scanned using Envision Perkin Elmer plate reader. Analysis of the obtained data reveal that ^{5hm}C levels increase over time in an enzyme dependent manner indicating catalytic activity of hTET2$_{CD\Delta LCI}$ (Fig. 2.4.4 A,B). The concentration of ^{5hm}C produced from hTET catalytic activity was estimated based on the standard curve acquired on the same assay plate.

Interestingly, 1 nM and 2 nM hTET2$_{CD\Delta LCI}$ showed good signal for ^{5hm}C between 0 and 20 min but limited signal increase beyond 20 min suggesting saturation of the beads. Both 10 nM and 100 nM of hTET2$_{CD\Delta LCI}$ showed rapid formation of ^{5hm}C followed by a rapid decline in signal. This is likely caused by subsequent oxidation of ^{5hm}C by hTET2$_{CD\Delta LCI}$ to give ^{5f}C and / or ^{5ca}C which may not be (well) recognized by ^{5hm}C-specific antibody. hTET2$_{CD\Delta LCI}$ at 0.5 nM and 0.25 nM reveal good activity, demonstrating that as little as 2.5 fmol of protein is required per assay data point. hTET2$_{CD}$ showed similar activity to hTET2$_{CD\Delta LCI}$ while hTET1$_{CD}$ demonstrate significantly lower activity compared to hTET2$_{CD\Delta LCI}$ (Fig. 2.4.4 C,D). To perform biochemical characterisation on TET enzymes, the optimal signal / noise window and linear regions were identified. Based on the observations made displayed in

figure 2.4.4, the final concentrations were set to 10 nM for hTET1$_{CD}$ (0–30 min, R^2: 0.97), 1 nM for hTET2$_{CD}$ (0–10 min, R^2: 0.93) and 1 nM for hTET2$_{CD\Delta LCI}$ (0–10 min, R^2: 0.97).

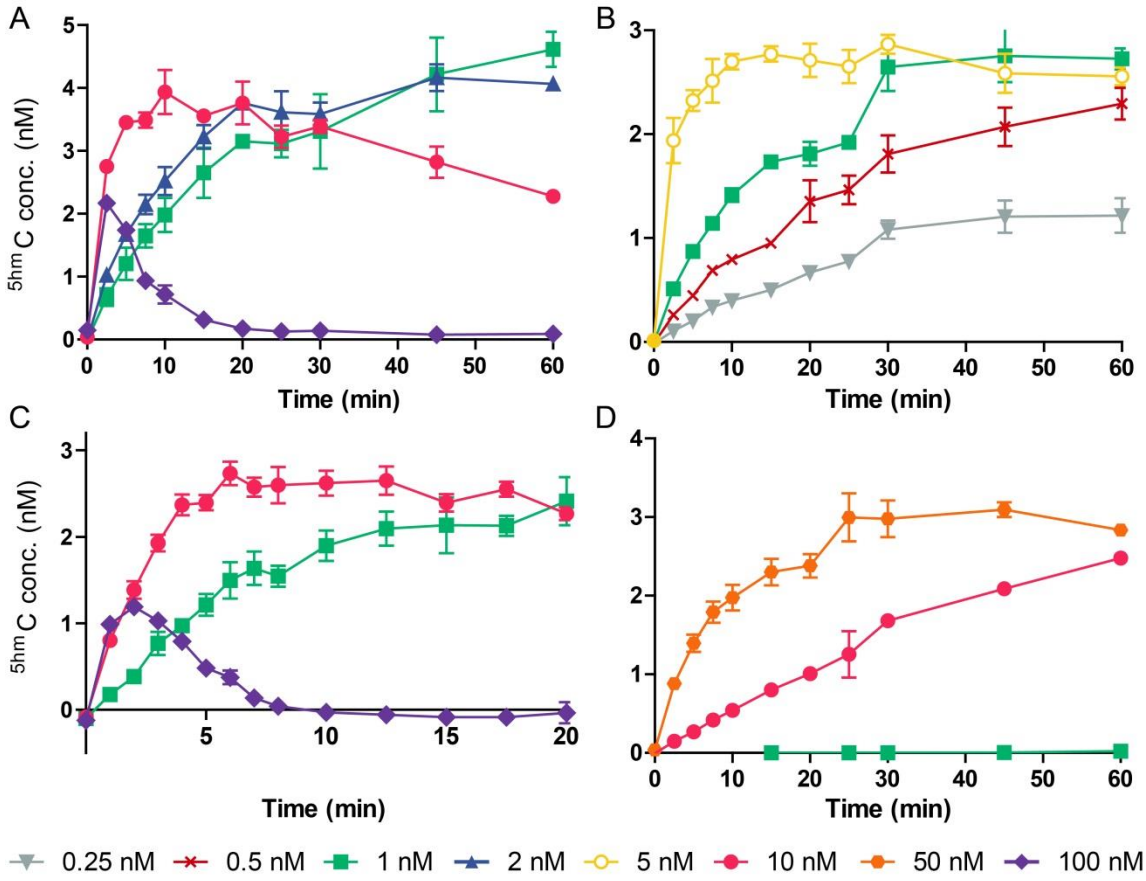


Figure 2.4.4: AlphaScreen™ activity assay of different concentrations of hTET2$_{CD\Delta LCI}$ (A, B), hTET2$_{CD}$ (C) and hTET1$_{CD}$ (D). Experimental conditions: $^{5m}C_{ss}$DNA-[Biotin] **2.1** (10 nM), Asc (100 μM), $(NH_4)_2Fe(II)(SO_4)_2$ (10 μM), 2OG (10 μM) in buffer. Concentration of product estimate based on side-by-side standard curve. n=3–4, error bars depicted as s.d.

Having established the optimal conditions for AlphaScreen™ for TET enzyme activity assay using single stranded DNA, attempts were made to determine the apparent kinetic parameters of these enzymes. Increase in the $^{5m}C_{ss}$DNA-[Biotin] **2.1** substrate concentration (>10 nM), led to the rapid formation of saturating levels of signal prohibiting steady state Michaelis-Menten enzyme kinetics on DNA. However, apparent kinetic parameters for the 2OG co-factor against TET proteins were successfully measured. With variable 2OG concentrations and limiting amounts of DNA (10 nM $^{5m}C_{ss}$DNA-[Biotin] **2.1**) in the presence of Asc (100 μM) and $(NH_4)_2Fe(II)(SO_4)_2$ (10 μM), apparent 2OG K_M (K_M^{App}) was obtained. Proteins hTET1$_{CD}$ (10 nM), hTET2$_{CD}$ (1 nM) and hTET2$_{CD\Delta LCI}$ (1 nM)

were tested at single time-point determination providing under these conditions a K_M^{App} of 11.1 ± 3.5 µM for $hTET2_{CD}$ and K_m^{App}: 4.8 ± 1.6 µM for $hTET2_{CD\Delta LCI}$ (Fig. 2.4.5). Interestingly, the difference observed between $hTET2_{CD}$ and $hTET2_{CD\Delta LCI}$ implies that the LCI may influence access to the 2OG binding site. $hTET1_{CD}$ showed relatively low K_M^{App} of 2.4 ± 0.3 µM.

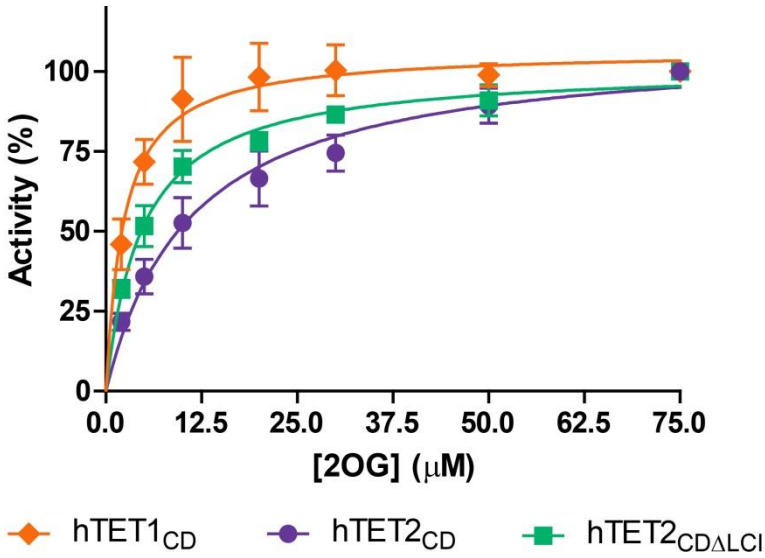


Figure 2.4.5: Apparent K_M determination using AlphaScreen™ of 2OG under limiting DNA (10 nM) for $hTET1_{CD}$ (10 nM), $hTET2_{CD}$ (1 nM) and $hTET2_{CD\Delta LCI}$ (1 nM). n= 5, error given as s.d.

2.4.3 ITC analysis of substrate enzyme kinetics of $^{5m}C_{ss}$DNA 2.5 against $hTET2_{CD}$

With the current AlphaScreen™ conditions, DNA substrate enzyme kinetics could not be analysed. In order to obtain steady state Michaelis-Menten enzyme kinetics on $^{5m}C_{ss}$DNA **2.5** substrate for $hTET2_{CD}$, Isothermal Titration Calorimetry (ITC) was employed. A detailed description of the ITC instrument can be found in section 3.5.3 but briefly discussed here. In this assay the ITC instrument is used to monitor the progression of the TET enzymatic reaction as function of the DNA concentration. By closely monitoring the temperature in the sample and reference cell, the energy difference between the enzymatic reaction and negative control can be measured. The difference in energy observed is directly proportional to the reaction velocity.

Steady state kinetics studies were carried out for hTET2$_{CD}$ and $^{5m}C_{ss}$DNA **2.5** substrate using ITC. A protein solution was prepared consisting of hTET2$_{CD}$ (1.0 μM) in co-factor solution: {Asc (2.0 mM), $(NH_4)_2Fe(II)(SO_4)_2$ (200 μM), 2OG (200 μM) in buffer [HEPES (50 mM), NaCl (150 mM) in MQ (pH 7.5)]} (280 μL) and placed in the sample cell (25.0 °C). DNA solution consisting of $^{5m}C_{ss}$DNA (340 μM) in co-factor solution was titrated (2.0 μL) in intervals of 150 sec; each injection increasing the substrate concentration in the cell. The reaction velocity is unknown and it is therefore assumed that <30% ^{5hm}C is produced prior to the next injection.

The negative control was made by performing the same reaction in the absence of $^{5m}C_{ss}$DNA **2.5** (Fig. 2.4.6 A). The recorded raw data was baseline corrected against the negative control and the average reaction velocity was calculated from the data acquired 84–144 sec after each injection (Fig. 2.4.6 B). Under these conditions, the enzymatic kinetic parameters for hTET2$_{CD}$ using $^{5m}C_{ss}$DNA **2.5** were found to be: K_M 23.8 ± 1.7 μM, a V_{max}: 0.802 ± 0.0278 μCals·sec^{-1} min^{-1} and K_{cat}: 0.802 ± 0.0278 μCals·sec^{-1} min^{-1} μM·Enz^{-1}. The V_{max} and K_{cat} are difficult to interpret and to understand these values better, secondary experiments need to be conducted. The details of these experiments are discussed in section 2.5.

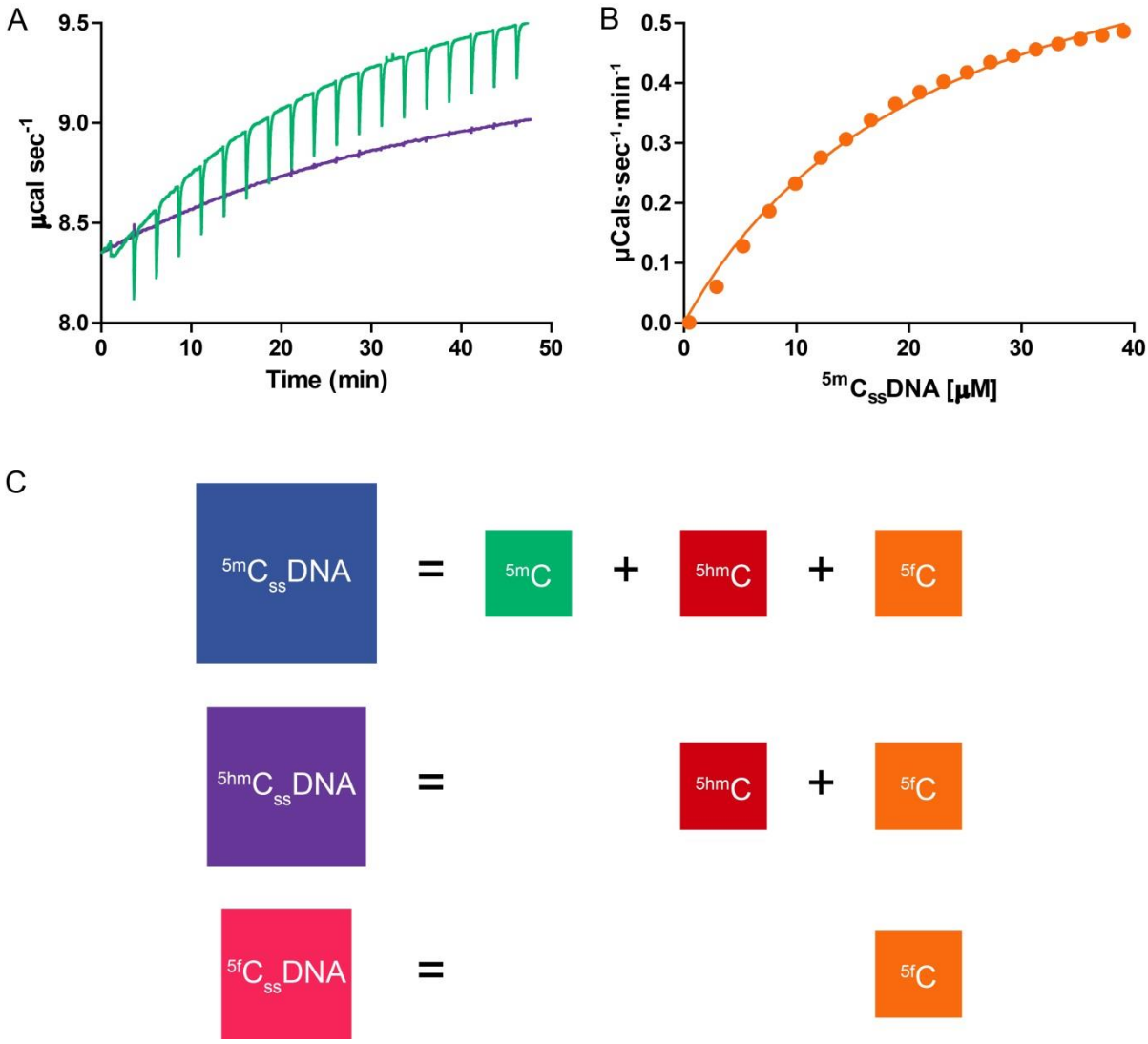


Figure 2.4.6: Enzyme kinetics on $^{5m}C_{ss}$DNA 2.5 against hTET2$_{CD}$ using ITC. A) raw data of hTET2$_{CD}$ (1.0 µM) enzyme kinetics in the presence of $^{5m}C_{ss}$DNA substrate (340 µM) (green) and absence of substrate (purple) containing Asc (2.0 mM), $(NH_4)_2Fe(II)(SO_4)_2$ (200 µM) and 2OG (200 µM) in buffer. B) Δ Energy between reaction and control experiment based on 1 min average. K_M: 23.8 ± 1.7 µM, V_{max}: 0.802 ± 0.028 µCals·sec^{-1}, K_{cat}: 0.802 ± 0.028 µCals·sec^{-1}·min^{-1}·µM Enz^{-1}. C) Schematic representation of heat derived from total conversion of substrates.

2.4.4 MALDI-TOF steady state enzyme kinetics of hTET2$_{CD}$

To study the relative levels of ^{5hm}C produced from ^{5m}C by TET proteins using MS and to perform enzyme kinetics on the substrates, a MALDI-TOF assay was developed based on a previous report procedure with minor alterations.[46] Firstly, optimization experiments were conducted to establish the sensitivity of the MALDI instrument. DNA was tested for detection using 3-hydroxypicolinic acid (3-

HPA) supplemented with diammonium citrate as the MALDI matrix on a ground steel target plate with the MALDI-TOF in linear reflectron mode. Although, initial examination of 32 bases single and double stranded DNA were detected, a higher sensitivity was found with 12 base pair palindromic DNA (5'-AGG AGG ^{5m}CCT CCT-3' **2.33**). Note that due to the broad peaks and the relatively small difference in mass between ^{5hm}C and ^{5f}C (2 Da), the individual modifications could not be quantified directly. Thus assays were carried out in the linear range of ^{5hm}C product formation and before further oxidation was observed. Next, the activity of hTET2$_{CD}$ (0.4 μM) was tested under the following assay conditions: ^{5m}C$_{ds}$DNA **2.33** (2.0 μM), Asc (500 μM), $(NH_4)_2Fe(II)(SO_4)_2$ (50 μM), 2OG (100 μM) in buffer [HEPES (50 mM), NaCl (50 mM) in MilliQ (pH 7.0)] and the reaction was quenched at time-points using cation exchange resin hydrogel. The samples in the supernatant were immediately spotted, air dried on a 3-HPA premade target plate and analysed on the MALDI-TOF (Fig. 2.4.7 A,B). Progression of the reaction was calculated based on the relative peak surface area and normalized against the known DNA concentration. Michaelis-Menten enzyme kinetics were performed on hTET2$_{CD\Delta LCI}$ with ^{5m}C$_{ds}$DNA **2.33** as substrate (Fig. 2.4.7 C). Under these conditions, a substrate K_M of 1.02 ± 0.23 μM was obtained, which is similar to literature values.[61] However, V_{max}: 0.119 ± 0.011 μM·min^{-1} and K_{cat}: 0.296 ± 0.028 min^{-1} are significantly higher than the reported values. This is potentially due to higher recombinant TET activity / purity, the use of different substrates, co-factor composition, buffer conditions and the presence of the LCI. MALDI-TOF provides an assay for determining kinetic properties for hTET2$_{CD}$; however, the complex and laborious sample preparation and analysis prohibits its use for high throughput screening of inhibitors.

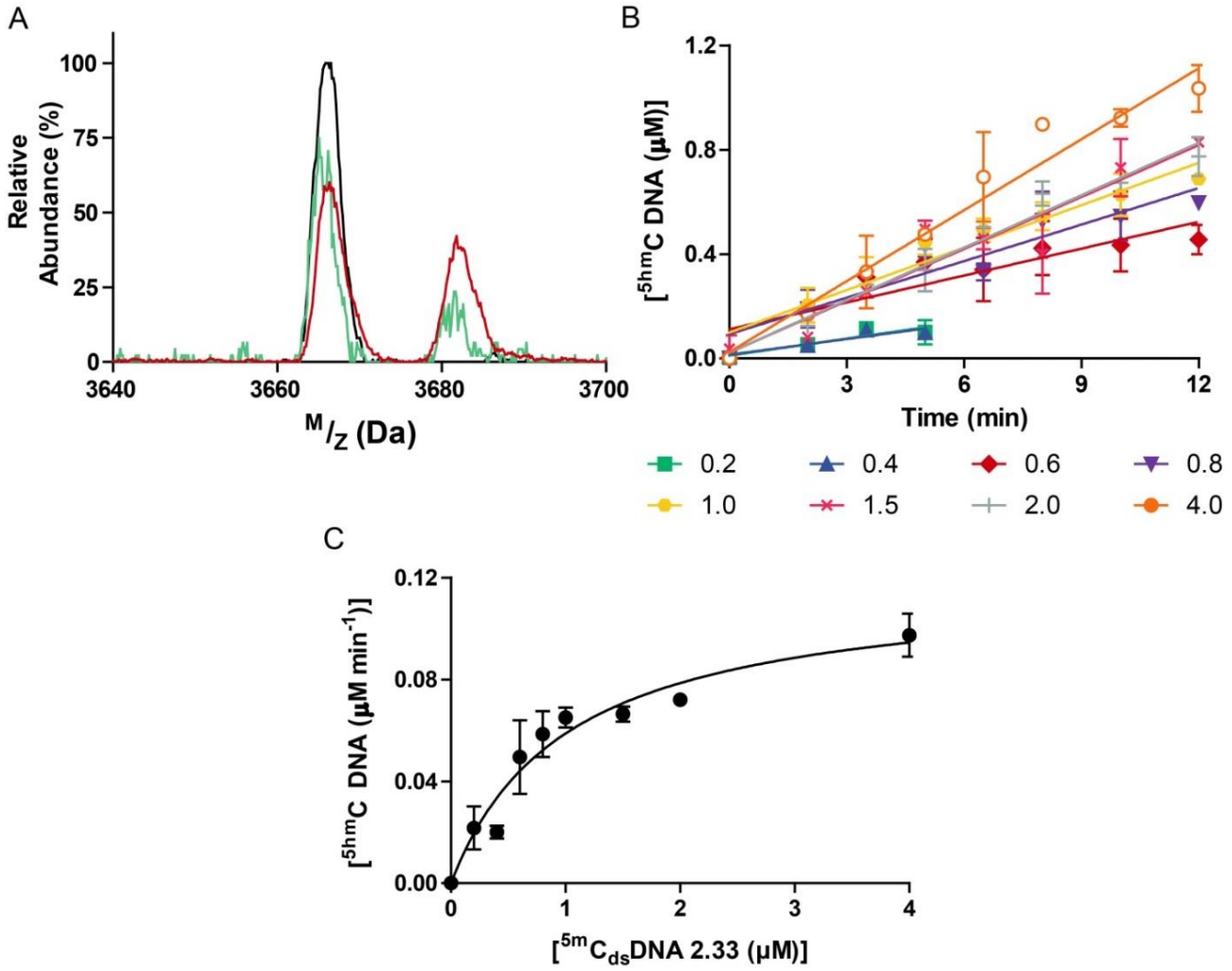


Figure 2.4.7: Time course and enzyme kinetics with $^{5m}C_{ds}$DNA 2.33 against hTET2$_{CD}$ (0.4 µM) in the presence of Asc (500 µM), Fe^{II} (50 µM) and 2OG (100 µM) measured by MALDI-TOF. A) Representative time-course at $^{5m}C_{ds}$DNA 2.33 (2.0 µM) at t= 0 min (black), 6.5 min (green) and 12 min (red); B,C) enzyme kinetics with TET activity curves at different concentrations DNA (B) and corresponding velocities plotted (C). n= 2, error given as s.d.

2.4.5 Activity assay for hTET2$_{CD\Delta LCI}$ using RapidFire Mass spectrometry

Whilst the data obtained from MALDI-TOF is informative, providing a quantitative readout of relative substrate / product abundance in a reaction is relatively slow. Therefore, an alternative mass spectrometry-based assay was developed. Agilent Rapid-Fire mass spectrometry (RF-MS) was chosen as it advantageous over 'classical' LC-MS due to the fast SPE cartridge-based LC purification which aids rapid sample acquisitions and enables real-time multi reaction monitoring (MRM). In the RF-MS

instrument setup, the sample is vacuum aspirated, mixed with ion-pairing agent Octyl Ammonium Acetate (OAA) and DNA retained on the DNA on an R_f-C_{18} Solid Phase Extraction (SPE) cartridge. The sample is washed with OAA buffer, eluted by a mixture of acetonitrile/water (80%) and directed towards the QTOF mass spectrometer for real-time measurements.

In a similar fashion to MALDI-TOF, 32 bases $^{5m}C_{ss}$DNA **2.5** gave a lower signal in comparison to the shorter 12 bases oligonucleotides while at the same time enhance catalytic turnover by hTETs. 12-bp double stranded palindromic DNA ($^{5m}C_{ds}$DNA **2.33**) was used. An activity assay was conducted with $^{5m}C_{ds}$DNA **2.33** (4.0 μM), Asc (200 μM), $(NH_4)_2Fe(II)(SO_4)_2$ (50 μM), 2OG (10 μM) in buffer [HEPES (50 mM) in MQ (pH 7.0)] and the reaction was initiated with the introduction of $hTET2_{CD\Delta LCI}$ (0.4 μM). RF-MS was set in MRM mode with t=0 min measurement in the absence of protein followed by first injection after 56 sec and subsequent sample acquisitions each 69 sec, detecting the DNA charge states [M-2H^{2-}], [M-3H^{3-}] and [M-4H^{4-}]. Catalytic turnover was observed, with the formation of ^{5hm}C and minor levels of ^{5f}C formation (Fig. 2.4.8 A) in a time-dependent manner. To promote the observation of all four-potential species (^{5m}C and ^{5ox}C), the experimental conditions were changed providing higher levels of 2OG (50 μM) and $hTET2_{CD\Delta LCI}$ (0.8 μM) while decreasing the concentration of $^{5m}C_{ds}$DNA **2.33** (1.0 μM) (Fig. 2.4.8 B,C).

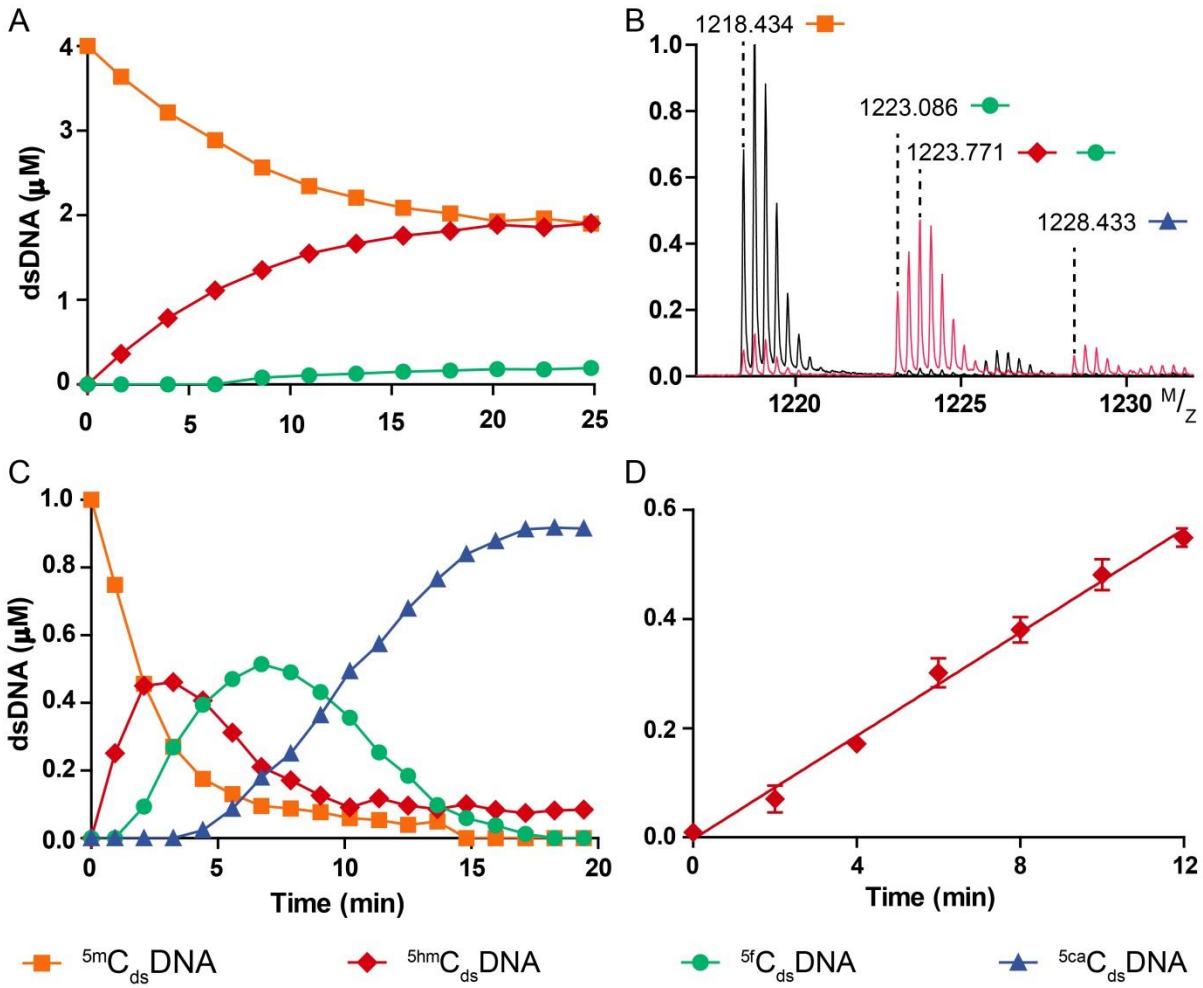


Figure 2.4.8: RF-MS analysis of $^{5m}C_{ds}$DNA 2.33 in the presence of hTET2$_{CD\Delta LI}$ with substrate and products $^{5m}C_{ds}$DNA 2.33 (orange square), ^{5hm}C (red diamond), ^{5f}C (green circle) and $^{5ca}C_{ds}$DNA 2.36 (blue triangle). A) Activity assay of $^{5m}C_{ds}$DNA **2.33** (4.0 µM) in the presence of hTET2$_{CD\Delta LCI}$ demonstrating the formation of ^{5hm}C and ^{5f}C. B,C) Activity assay of ^{5m}C (1.0 µM) showing with decreased substrate and increased hTET2$_{CD\Delta LCI}$ (0.8 µM) concentration the formation of ^{5m}C, ^{5hm}C, ^{5f}C and ^{5ca}C. B) black: t= 0 min and pink: t= 5.57 min, product mass depicted as $[M-3H]^{3-}$. D) Determination of the linear region by RF-MS using ^{5m}C (2.0 µM) and hTET2$_{CD\Delta LCI}$ (0.4 µM) as quenched assay. Observed correlation R^2= 0.99.

Due to the small difference in masses between ^{5f}C and ^{5hm}C (2 Da), the overlap in their isotopic patterns prohibit direct analysis. To quantitate the levels of oxidative products formed, the relative ratio of isotopes was calculated for ^{5m}C and compared to acquired ^{5m}C data, finding relative similar values for the isotope peak surface area. Next, the composition of ^{5f}C third isotopic peak was determined relative to the mono isotopic peak of ^{5f}C and excess area calculated was assigned to ^{5hm}C. The substrate and product peaks were visualized as shown in figure 2.4.8 C. Analysis of the raw data reveals a clear time dependent decrease of ^{5m}C and simultaneous increase in ^{5hm}C, with peaks at 3.23 min followed by rapid conversion to ^{5f}C. After 4.40 min ^{5ca}C is detected and all four species were found in solution.

Over time, nearly all substrates were consumed and at 20 min ^{5ca}C was the predominant peak with little ^{5m}C, ^{5hm}C and ^{5f}C left.

Recently, it was reported that TET proteins are also able to oxidize, with lower efficiency, thymidine in similar fashion to ^{5m}C to give 5-hydroxyuracil (^{5hm}U) nucleotide followed by iterative oxidations to give 5-formyluracil (^{5f}U) and 5-carboxyuracil (^{5ca}U).[62,63] Examination of the collected data here, cannot rule out that some thymidine present in ^{5m}C$_{ds}$DNA **2.33** are oxidized. The oxidative pattern for T would provide the same mass shift as for ^{5m}C; however, a combination of 5-oxidized methyl cytosine (^{5ox}C) and oxidized-thymine (^{5ox}U$_{ds}$DNA) can only be found in certain cases as shown in figure 2.4.9. For example, while a combination of masses of ^{5f}C-^{5hm}U$_{ds}$DNA is possible, a mass of ^{5ca}C-^{5hm}U$_{ds}$DNA is not observed. Further MS/MS analysis could provide evidence for the potential formation of oxidized thymine (^{5ox}U).

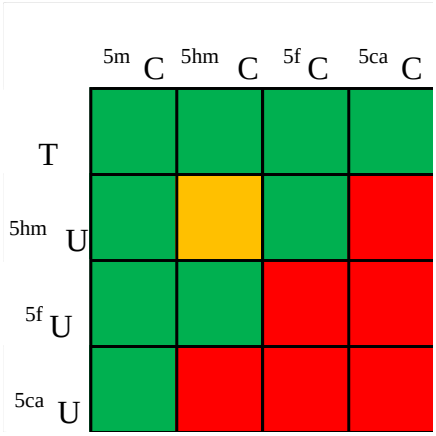


Figure 2.4.9: Schematic representation of possible combination of oxidation of ^{5m}C$_{ds}$DNA 2.34 and T by hTET2$_{CDALCI}$ protein. Green represents masses observed and red are not found. The mass displayed in orange is observed but only occur in small quantities.

2.4.6 Inhibition of $hTET1_{CD}$, $hTET2_{CD}$ and $hTET2_{CD\Delta LCI}$ measured using AlphaScreen™

AlphaScreen™ and RF-MS methods can be efficiently used to determine the conversion levels of ^{5m}C to ^{5hm}C and in the case for RF-MS also ^{5f}C and ^{5ca}C. These methods are also suitable for examining the potential decrease in catalytic activity, in other words, these methods could be used for the assessment of inhibitor activity. Chemical inhibitors are useful to study the 2OG-dependent dioxygenase domain for: A) functional biological role; B) study the effect on disease and C) evolutionary relation to other 2OG-dependent domain containing protein within the chromatin. In this study, potential inhibitors of hTETs were selected based on: A) known 2OG-dependent oxygenase inhibitor; B) structurally mimicking 2OG or DNA; and if applicable C) importance to the intended target (*e.g.* inhibitor is clinically relevant or most potent inhibitor to target) or disease related to TET. With these criteria in mind, a panel of inhibitors was selected consisting of: generic 2OG-dependent domain inhibitors, Prolyl hydroxylase (ProHD) inhibitors, Lysine specific demethylases (KDM) inhibitors and HDAC inhibitor (Fig. 2.4.10). Furthermore, nucleosides, oncometabolites and TCA cycle intermediates were screened for their level of inhibition. Dose response screens were prepared using Labcyte Echo Dispenser from 10 mM DMSO-stock (TCA cycle intermediate from 100 mM MilliQ solutions) source-stock providing 100 µM inhibitor top concentrations.

AlphaScreen™ provided a 2OG K_M^{App} for $hTET1_{CD}$ and $hTET2_{CD/CD\Delta LCI}$ ranging from 11.1 µM to 2.4 µM and therefore the 2OG concentration was set to 10 µM for all AlphaScreen™ inhibition assays. Observations made using ITC revealed a substrate K_M of 23.8 ± 1.7 µM $^{5m}C_{ss}$DNA **2.5**, but due to AlphaScreen™ properties, inhibition screens are limited to 10 nM $^{5m}C_{ss}$DNA-[Biotin] **2.1**. The inhibitors were incubated (10 min) with $hTET1_{CD}$ (10 nM), $hTET2_{CD}$ (1.0 nM) or $hTET2_{CD\Delta LCI}$ (1.0 nM) followed by the addition of DNA mix [$^{5m}C_{ss}$DNA-[Biotin] **2.1** (10 nM), Asc (100 µM), $(NH_4)_2Fe(II)(SO_4)_2$ (10 µM), 2OG (10 µM) in buffer]. Reactions were quenched with EDTA after 10

min in the case of hTET2$_{CD/CD\Delta LCI}$ or 30 min for hTET1$_{CD}$. AlphaScreen™ beads were added, the mixture was incubated (30–60 min) and the wells were examined using Envision plate reader. Each screening plate carried IOX1 as positive-control and reference molecule. The values obtained are displayed in Table 2.4.1 with corresponding structures and representative IC_{50} curves in figure 2.4.10 and figure 2.4.11, respectively.

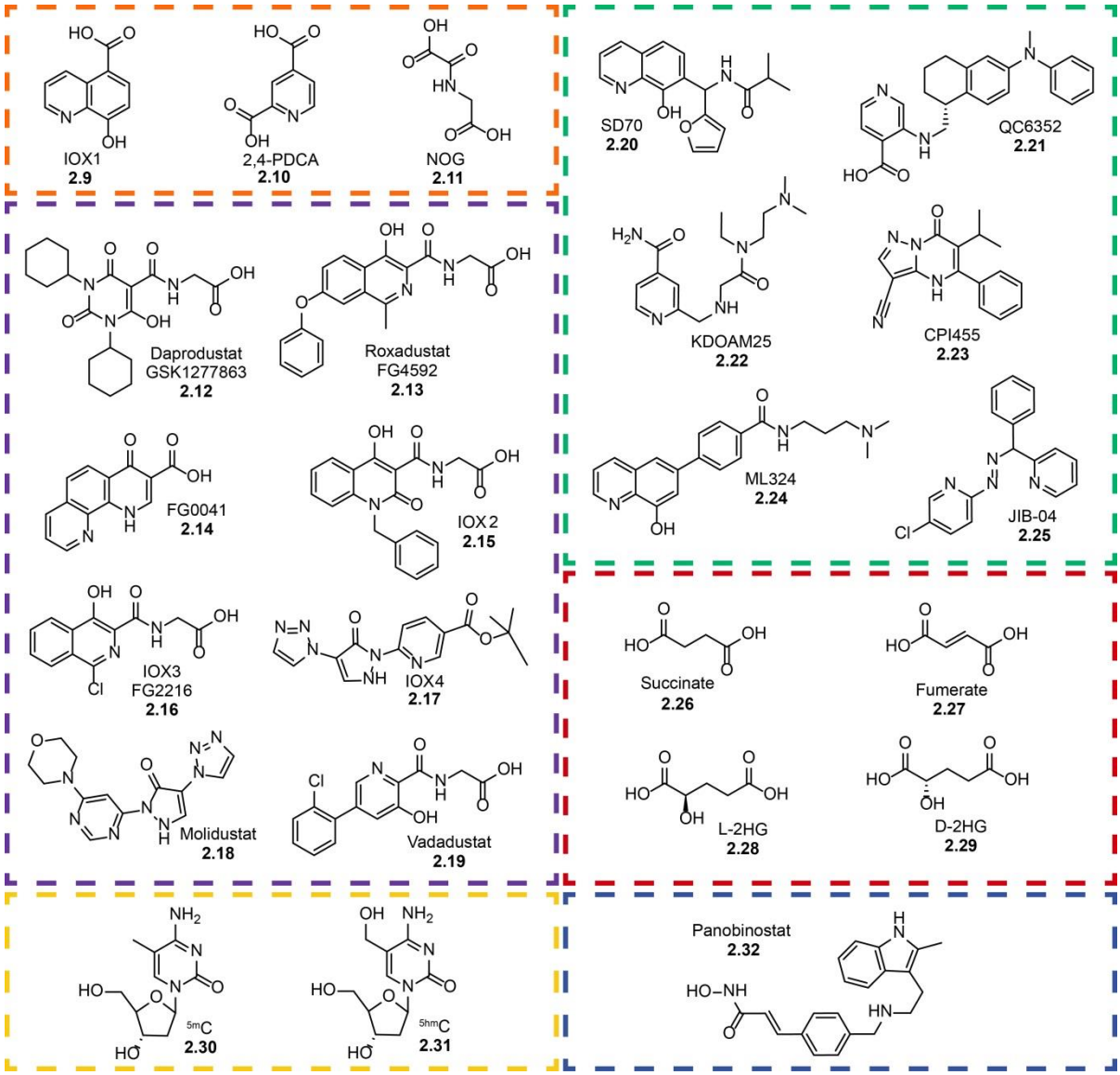


Figure 2.4.10: Schematic representation of inhibitors for generic 2OG-dependent dioxygenases (orange), Prolyl hydroxylase (purple), nucleosides (yellow), KDMs (green), TCA cycle intermediates, oncometabolites (red) and histone lysine deacetylase inhibitor (blue).

The 2OG-dependent domain containing proteins were found to be catalytically inhibited by pan-specific 2OG competitive inhibitors. To study this for TETs, IOX1 **2.9**, 2,4-Pyridinedicaboxylic acid (2,4-PDCA), **2.10** and NOG **2.11** were tested. hTET1$_{CD}$ showed considerate inhibition at 0.87 µM

(pIC_{50}: 6.06 ± 0.60) for IOX1 **2.9**, while 2,4-PDCA **2.10** (pIC_{50}: 5.62 ± 0.70) and NOG **2.11** (pIC_{50}: 4.91 ± 0.18) were found to inhibit significantly weaker for $hTET1_{CD}$.

Table 2.4.1: pIC_{50} of values of generic-, ProHD-, KDM-, HDAC-inhibitors, TCA cyclic intermediate, oncometabolites and nucleosides against $hTET1_{CD}$, $hTET2_{CD}$ and $hTET2_{CD\Delta LCI}$ tested using AlphaScreen™. Values shown as percentage revealed an incomplete IC_{50} curve and therefore, percentage activity is given at 100 µM inhibitor concentration. N.I.: not inhibitory (>90% activity at 100 µM), A.I.: assay inhibitory and N.D.: not determined. n=2–4, error given as s.d.

			AlphaScreen [pIC_{50}(M)]		
		Target	$hTET1_{CD}$	$hTET2_{CD}$	$hTET2_{CD\Delta LCI}$
IOX1	**2.9**	2OG Oxygenases	6.06 ± 0.60	5.30 ± 0.24	5.62 ± 0.25
2,4-PDCA	**2.10**	2OG Oxygenases	5.62 ± 0.70	5.29 ± 0.06	5.08 ± 0.07
NOG	**2.11**	2OG Oxygenases	4.91 ± 0.18	5.05 ± 0.17	4.90 ± 0.02
GSK1278863	**2.12**	ProHD	62.7% ± 8.5	69.5% ± 5.0	73.4% ± 8.45
Roxadustat	**2.13**	ProHD	68.6% ± 2.6	N.I.	N.I.
FG0041	**2.14**	ProHD	5.65 ± 0.22	5.61 ± 0.13	5.39 ± 0.05
IOX2	**2.15**	ProHD	42.6% ± 16.4	68.4% ± 7.9	67.6% ± 1.8
IOX3 (FG2216)	**2.16**	ProHD	58.4% ± 17.4	55.2% ± 3.2	54.6% ± 11.5
IOX4	**2.17**	ProHD	4.57 ± 0.16	4.83 ± 0.12	4.80 ± 0.13
Molidustat	**2.18**	ProHD	47.6% ± 17.4	N.I.	82.5% ± 7.3
Vadadustat	**2.19**	ProHD	5.22 ± 0.22	5.13 ± 0.19	5.31 ± 0.11
SD70	**2.20**	KDM4	64.3% ± 21	53.0% ± 2.8	44.8% ± 1.8
QC6352	**2.21**	KDM4	71.4% ± 0.4	70.2% ± 9.7	38.0% ± 1.5
KDOAM25	**2.22**	KDM5	N.I.	N.I.	70.6% ± 4.3
CPI455	**2.23**	KDM5	N.I.	N.I.	N.I.
ML324	**2.24**	JmjC-KDMs	5.56 ± 0.39	5.81 ± 0.08	6.12 ± 0.01
JIB-04	**2.25**	JmjC-KDMs	4.54 ± 0.44	5.52 ± 0.01	5.44 ± 0.01
Succinate	**2.26**	TCA cycle intermediate	3.96 ± 0.02	3.86 ± 0.13	3.82 ± 0.27
Fumarate	**2.27**	TCA cycle intermediate	4.08 ± 0.02	3.69 ± 0.06	3.99 ± 0.23
L-2HG	**2.28**	Oncometabolite	3.00 ± 0.05	4.89 ± 0.04	5.30 ± 0.14
D-2HG	**2.29**	Oncometabolite	3.38 ± 0.02	4.81 ± 0.02	5.21 ± 0.31
^{5m}C	**2.30**	Nucleoside	A.I.	A.I.	A.I.
^{5hm}C	**2.31**	Nucleoside	A.I.	A.I.	A.I.
Panobinostat	**2.32**	HDAC	5.38 ± 0.27	5.41 ± 0.24	5.67 ± 0.01

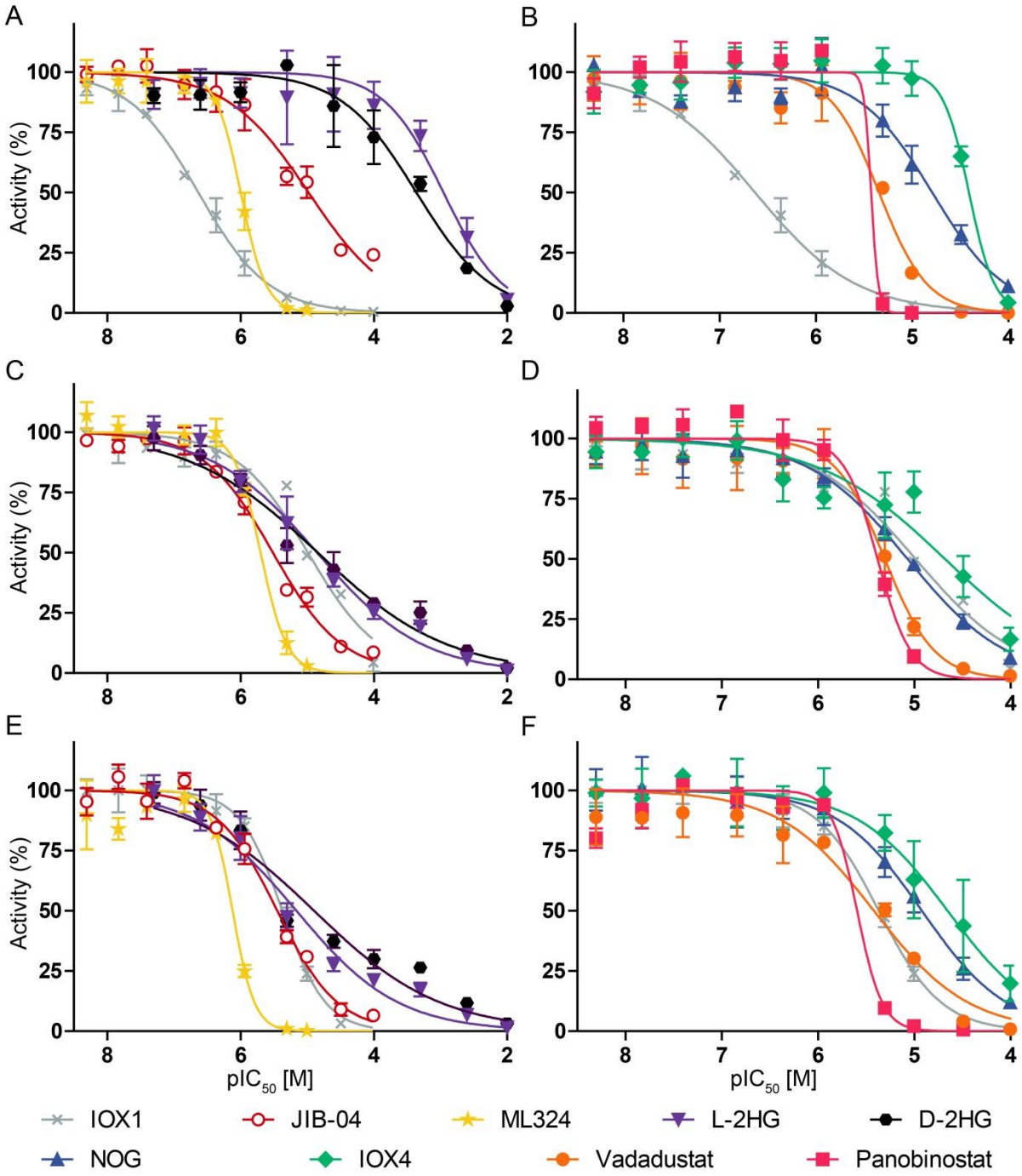


Figure 2.4.11: Dose response curves of various chromatin modifier inhibitors against hTET1$_{CD}$ (A,B), hTET2$_{CD}$ (C,D) and hTET2$_{CD\Delta LCI}$ (E,F). n=2–3, error given as s.d.

Amongst ProHD inhibitors tested against hTET1$_{CD}$, IOX4 **2.17** (pIC$_{50}$: 4.57 ± 0.16), FG0041 **2.14** (pIC$_{50}$: 5.65 ± 0.22) and Vadadustat **2.19** (pIC$_{50}$: 5.22 ± 0.22) showed moderate inhibition of TETs with values of IC$_{50}$: 26.9 μM, IC$_{50}$: 2.24 μM and IC$_{50}$: 6.02 μM, respectively. GSK1278863 **2.12**, Roxadustat **2.13**, IOX2 **2.15**, IOX3 **2.16** and Molidustat **2.18** were found to contain low levels of inhibition of hTET1$_{CD}$ at 100 μM. KDM4 specific inhibitors SD70 **2.20** (64.3 ± 21% at 100 μM) and

QC6352 **2.21** (71.4 ± 0.4% at 100 μM) were observed to be poor inhibitors for hTET1$_{CD}$ similar to KDM5 specific inhibitors KDMOAM25 **2.22** (N.I.) and CPI455 **2.23** (N.I.). ML324 **2.24**, a KDM pan-specific inhibitor was found to inhibit hTET1$_{CD}$ at IC$_{50}$: 2.75 μM (pIC$_{50}$: 5.56 ± 0.39) while JIB-04 **2.25** was found at IC$_{50}$: 28.8 μM (pIC$_{50}$: 4.54 ± 0.44). TCA cycle intermediate succinate **2.26**, was found to inhibit at IC$_{50}$: 110 μM (pIC$_{50}$: 3.96 ± 0.02) while TCA cycle intermediate fumarate **2.27** inhibits in a similar range of 83.2 μM (pIC$_{50}$: 4.08 ± 0.02). Interestingly, oncometabolites L-2HG **2.28** and D-2HG **2.29** were determined to be 50% inhibitory at 1.0 mM (pIC$_{50}$: 3.00 ± 0.05) and 417 μM (pIC$_{50}$: 3.38 ± 0.02), respectively. Attempts were made to assess the inhibition capabilities of ^{5m}C **2.30** and ^{5hm}C **2.31** nucleosides but were found to likely inhibit the ^{5hm}C-specific antibody depleting the signal. Pan-specific Histone Deacetalase (HDAC) inhibitor Panobinostat **2.32** was found to inhibit hTET1$_{CD}$ at IC$_{50}$: 4.17 μM (pIC$_{50}$: 5.38 ± 0.27) with a steep hill slope. Although the cause is unknown, it is tempting to speculate that this is caused by protein precipitation.[251]

In a complementary fashion, hTET1$_{CD}$, hTET2$_{CD}$ and hTET2$_{CD\Delta LCI}$ were tested against the panel of set 2OG oxygenase inhibitors (Fig. 2.4.11 C-F, Table 2.4.1). The majority of 2OG oxygenase inhibitors were found to inhibit with relatively similar inhibition profile to hTET2$_{CD}$ and hTET2$_{CD\Delta LCI}$ as hTET1$_{CD}$. IOX1 **2.9** gave an IC$_{50}$ at 5.01 μM (pIC$_{50}$: 5.30 ± 0.24) for hTET2$_{CD}$ and IC$_{50}$ of 2.39 μM for hTET2$_{CD\Delta LCI}$ (pIC$_{50}$: 5.62 ± 0.25). NOG **2.11**, a 2OG mimic and observed in the hTET2$_{CD\Delta LCI}$ crystal structures[45] inhibits at IC$_{50}$: 12.6 μM (pIC$_{50}$: 4.90 ± 0.02). 2,4-PDCA **2.10** was determined to inhibit hTET2$_{CD}$ (pIC$_{50}$: 5.29 ± 0.06) and hTET2$_{CD\Delta LCI}$ (pIC$_{50}$: 5.08 ± 0.07) weaker than hTET1$_{CD}$ showing a minor level of distinction between hTET1 and hTET2 proteins. The ProHD inhibitors demonstrated to have a similar pattern of inhibition for hTET1 in comparison to hTET2 proteins tested and little variation was found between constructs hTET2$_{CD}$ and hTET2$_{CD\Delta LCI}$. GSK1278863 **2.12**, Roxadustat **2.13**, IOX2 **2.15**, IOX3 **2.16** and Molidustat **2.18** are weak inhibitors (IC$_{50}$ >100 μM) while FG0041 **2.14** was observed to have an IC$_{50}$ of 2.45 μM (pIC$_{50}$: 5.61 ± 0.13) and 4.07 μM (pIC$_{50}$: 5.39 ± 0.05) for hTET2$_{CD}$ and hTET2$_{CD\Delta LCI}$, respectively. Vadadustat **2.19** shows single digit micro-molar

inhibition at IC_{50}: 7.41 μM (pIC_{50}: 5.13 ± 0.19) and IC_{50}: 4.90 μM (pIC_{50}: 5.31 ± 0.11) for $hTET2_{CD}$ and $hTET2_{CD\Delta LCI}$. IOX4 **2.17** reveals comparable levels of inhibition for $hTET2_{CD}$ (pIC_{50}: 4.83 ± 0.12) and $hTET2_{CD\Delta LCI}$ (pIC_{50}: 4.80 ± 0.13) relative to $hTET1_{CD}$.

All KDM inhibitors tested with hTET2 proteins showed slightly higher levels of inhibition over hTET1, but the inhibition levels were detected to be very similar for ProHD inhibitors among hTET1 and hTET2 proteins. KDM specific QC6352 **2.21**, SD70 **2.20** and KDOAM25 **2.22** were found to inhibit hTET2 weakly at 100 μM, while ML324 **2.24** showed IC_{50} at 1.54 μM (pIC_{50}: 5.81 ± 0.08) for $hTET2_{CD}$ and sub-micro molar IC_{50}: 0.76 μM (pIC_{50}: 6.12 ± 0.01) for $hTET2_{CD\Delta LCI}$. Furthermore, TCA cycle intermediates inhibit $hTET2_{CD}$ at high micro-molar range with succinate **2.26** at IC_{50}: 138 μM (pIC_{50}: 3.86 ± 0.13) and fumarate **2.27** IC_{50}: 204 μM (pIC_{50}: 3.69 ± 0.06). The most striking result emerged from oncometabolites L/D-2HG **2.28** / **2.29** which inhibited $hTET2_{CD}$ at IC_{50}: 12.9 μM (pIC_{50}: 4.89 ± 0.04) and IC_{50}: 15.5 μM (pIC_{50}: 4.81 ± 0.02). Even stronger inhibition was found with $hTET2_{CD\Delta LCI}$ showing IC_{50}: 5.01 μM (pIC_{50}: 5.30 ± 0.14) for L-2HG **2.28** and IC_{50}: 6.16 μM for D-2HG **2.29** (pIC_{50}: 5.21 ± 0.31). Comparing this to $hTET1_{CD}$, $hTET2_{CD/CD\Delta LCI}$ showed approximately two orders of magnitude higher levels of inhibition for the L/D-2HG. Additionally, HDAC inhibitor Panobinostat **2.32** was found to inhibit $hTET2_{CD}$ and $hTET2_{CD\Delta LCI}$ at IC_{50}: 3.89 μM (5.41 ± 0.24) and IC_{50}: 2.13 μM (pIC_{50}: 5.67 ± 0.01), respectively.

2.4.7 Inhibition of $hTET2_{CD\Delta LCI}$ measured using RapidFire Mass spectrometry

To confirm the observations made using AlphaScreen™ a secondary orthogonal screening assay is required. RF-MS was used as validation assay and is mass-spectrometry based. To investigate the level of inhibition of $hTET2_{CD\Delta LCI}$ using mass spectrometry, RF-MS was employed. To prepare RF-MS for inhibition studies using TET, a discontinuous assay setup was developed. RF-MS requires a larger reaction volume (>50 μL) and is significantly less sensitive than AlphaScreen™. Thus more enzyme

for is required analysis restricting to only the use of hTET2 $_{CD\Delta LCI}$ for this purpose. In the optimization experiments carried out, efforts were made to efficiently quench the enzymatic reactions. EDTA was tested but was found to suppress the DNA mass signal, which suggested that EDTA retains on the SPE cartridge and therefore other common methods to inactivate 2OG-dependent dioxygenases were considered, including acidic quench, basic quench, rapid heating or methanol precipitation. However, none of these are compatible with the assay or may interfere with DNA stability. Interestingly, small molecule inhibitor NOG **2.11** (1.0 mM) showed good inhibition of hTET2$_{CD\Delta LCI}$ without suppressing the DNA signal during mass acquisitions, thus NOG **2.11** was used to quench the reaction.

Next, an optimal assay window was determined. The targeted conversion of ^{5m}C to ^{5hm}C was set to 20–30% allowing for a maximum signal / noise window but minimizing the formation of ^{5f}C and ^{5ca}C. Time-course pre-quenched activity assay showed that $^{5m}C_{ds}$DNA **2.33** (2.0 μM), Asc (200 μM), $(NH_4)_2Fe(II)(SO_4)_2$ (50 μM), 2OG (10 μM) in buffer [HEPES (50 mM) in MQ (pH 7.0)] and hTET2$_{CD\Delta LCI}$ (0.4 μM) gave excellent correlation (R^2: 0.99) (Fig. 2.4.8 D).

The optimised experimental conditions were then applied to inhibition studies of hTET2$_{CD\Delta LCI}$. Enzyme mix was incubated for 5 min with the inhibitor (pre-dispensed with Cyte Echo) and the reaction initiated by the addition of [co-factors, substrate] mixture. The mixture was reacted for 10 min and quenched with an excess of NOG (1.0 mM final). Robust and reproducible IC_{50} curves were obtained using this method (Fig. 2.4.10, Fig. 2.4.12, Table 2.4.2).

IOX1 **2.9** and 2,4-PDCA **2.10** IC_{50} values are in agreement with obtained values for AlphaScreen™, however, NOG **2.11** was found to inhibit with an order of magnitude more potent in RF-MS. ProHD inhibitor Molidustat **2.18** was found to be a weak inhibitor while IOX3 **2.16** (pIC_{50}: 5.29 ± 0.04), GSK1277863 **2.12** (pIC_{50}: 4.38 ± 0.02), IOX2 **2.15** (pIC_{50}: 4.78 ± 0.02) showed a higher potency in the RF-MS relative to AlphaScreen™. The level of inhibition by RF-MS for IOX4 **2.17** at 30.2 μM

(pIC_{50}: 4.52 ± 0.14), FG0041 **2.14** at 5.37 μM (pIC_{50}: 5.27 ± 0.01) and Vadadustat **2.19** at 1.94 μM (pIC_{50}: 5.71 ± 0.02) were observed to be similar to AlphaScreen™. Poor inhibition was found with KDM4 specific inhibitors SD70 **2.20** / QC6352 **2.21** and KDM5 specific inhibitors KDMOAM25 **2.22** / CPI455 **2.23**.

Table 2.4.2: pIC_{50} of values of generic-, ProHD-, KDM-, HDAC-inhibitors, TCA cyclic intermediate, oncometabolites and nucleosides against $hTET1_{CD}$, $hTET2_{CD}$ and $hTET2_{CD\Delta LCI}$ tested using AlphaScreen™ and RF-MS. Values shown as percentage revealed an incomplete IC_{50} curve therefore, percentage activity is given at 100 μM inhibitor concentration. N.I.: not inhibitory (>90% activity at 100 μM), A.I.: assay inhibitory and N.D.: not determined. n=2–4, error given as s.d.

			AlphaScreen [pIC_{50}(M)]			RF-MS [pIC_{50} (M)]
		Target	$hTET1_{CD}$	$hTET2_{CD}$	$hTET2_{CD\Delta LCI}$	$hTET2_{CD\Delta LCI}$
IOX1	**2.9**	2OG Oxygenases	6.06 ± 0.60	5.30 ± 0.24	5.62 ± 0.25	5.78 ± 0.03
2,4-PDCA	**2.10**	2OG Oxygenases	5.62 ± 0.70	5.29 ± 0.06	5.08 ± 0.07	5.83 ± 0.05
NOG	**2.11**	2OG Oxygenases	4.91 ± 0.18	5.05 ± 0.17	4.90 ± 0.02	6.02 ± 0.04
GSK1278863	**2.12**	ProHD	62.7% ± 8.5	69.5% ± 5.0	73.4% ± 8.45	4.38 ± 0.02
Roxadustat	**2.13**	ProHD	68.6% ± 2.6	N.I.	N.I.	N.D.
FG0041	**2.14**	ProHD	5.65 ± 0.22	5.61 ± 0.13	5.39 ± 0.05	5.27 ± 0.01
IOX2	**2.15**	ProHD	42.6% ± 16.4	68.4% ± 7.9	67.6% ± 1.8	4.78 ± 0.02
IOX3 (FG2216)	**2.16**	ProHD	58.4% ± 17.4	55.2% ± 3.2	54.6% ± 11.5	5.29 ± 0.04
IOX4	**2.17**	ProHD	4.57 ± 0.16	4.83 ± 0.12	4.80 ± 0.13	4.52 ± 0.14
Molidustat	**2.18**	ProHD	47.6% ± 17.4	N.I.	82.5% ± 7.3	N.I.
Vadadustat	**2.19**	ProHD	5.22 ± 0.22	5.13 ± 0.19	5.31 ± 0.11	5.71 ± 0.02
SD70	**2.20**	KDM4	64.3% ± 21	53.0% ± 2.8	44.8% ± 1.8	41.7% ± 9.3
QC6352	**2.21**	KDM4	71.4% ± 0.4	70.2% ± 9.7	38.0% ± 1.5	66.5% ± 4.3
KDOAM25	**2.22**	KDM5	N.I.	N.I.	70.6% ± 4.3	N.I.
CPI455	**2.23**	KDM5	N.I.	N.I.	N.I.	86.7% ± 0.5
ML324	**2.24**	JmjC-KDMs	5.56 ± 0.39	5.81 ± 0.08	6.12 ± 0.01	4.76 ± 0.14
JIB-04	**2.25**	JmjC-KDMs	4.54 ± 0.44	5.52 ± 0.01	5.44 ± 0.01	47.8% ± 1.1
Succinate	**2.26**	TCA cycle intermediate	3.96 ± 0.02	3.86 ± 0.13	3.82 ± 0.27	4.44 ± 0.04
Fumarate	**2.27**	TCA cycle intermediate	4.08 ± 0.02	3.69 ± 0.06	3.99 ± 0.23	4.77 ± 0.06
L-2HG	**2.28**	Oncometabolite	3.00 ± 0.05	4.89 ± 0.04	5.30 ± 0.14	5.44 ± 0.04
D-2HG	**2.29**	Oncometabolite	3.38 ± 0.02	4.81 ± 0.02	5.21 ± 0.31	5.46 ± 0.06
^{5m}C	**2.30**	Nucleoside	A.I.	A.I.	A.I.	N.I.
^{5hm}C	**2.31**	Nucleoside	A.I.	A.I.	A.I.	N.I.
Panobinostat	**2.32**	HDAC	5.38 ± 0.27	5.41 ± 0.24	5.67 ± 0.01	4.62 ± 0.08

Pan specific inhibitor JIB-04 **2.25** inhibitor was not confirmed as a hit using the RF-MS assay and showed only 52% ± 1.1% activity at 100 μM inhibitor concentration. ML324 **2.24** revealed to be the most potent KDM inhibitor in the panel at 17.4 μM (pIC_{50}: 4.76 ± 0.14) but over an order of magnitude weaker in inhibition as observed for AlphaScreen™.

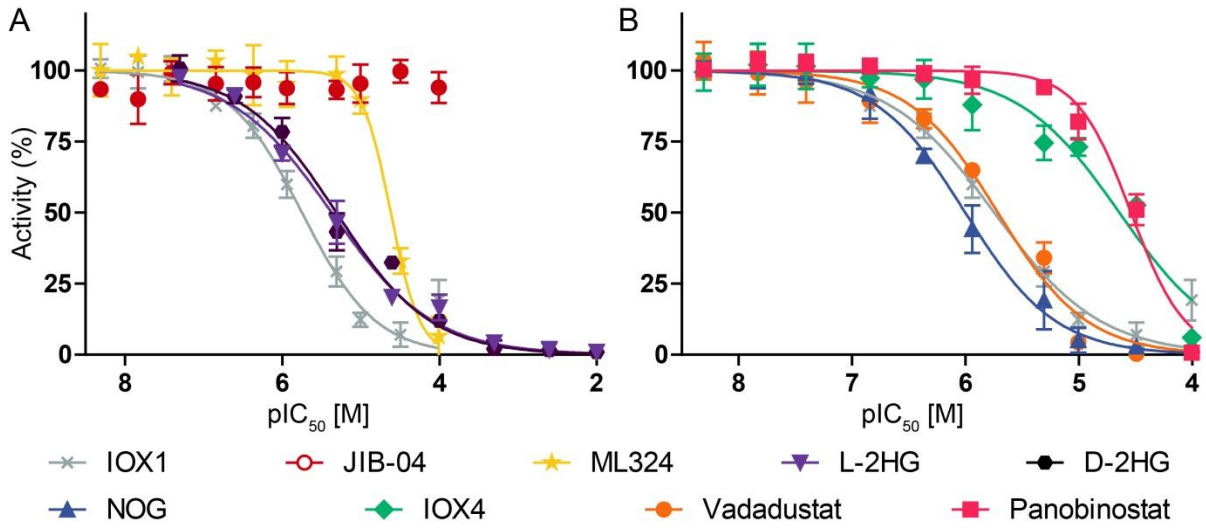


Figure 2.4.12: Dose response curves of chromatin modifier inhibitors against hTET2$_{CD\Delta LCI}$ using RF-MS. n=2–3, error given as s.d.

TCA cycle intermediates showed some level of increased inhibition in the RF-MS assay with succinate **2.26** at 36.2 μM (pIC$_{50}$: 4.44 ± 0.04) and fumarate **2.27** at 17.0 μM (pIC$_{50}$: 4.77 ± 0.06). Oncometabolites L/D-2HG **2.28** / **2.29** were also observed with RF-MS to inhibit substantially at 3.63 μM (pIC$_{50}$: 5.44 ± 0.04) and 3.47 μM (pIC$_{50}$: 5.46 ± 0.06), respectively. Nucleosides ^{5m}C **2.30** and ^{5hm}C **2.31** were tested using RF-MS and showed no significant level of inhibition at 100 μM suggesting that the DNA strand is vital for binding and potentially for the enzymatic reaction. HDAC inhibitor Panobinostat **2.32** demonstrated an IC$_{50}$ of 24.0 μM (pIC$_{50}$: 4.62 ± 0.08) resembling an approximate order of magnitude increase relative to AlphaScreen™.

RF-MS demonstrates to be a method for assessing the potency of inhibitors against hTET2$_{CD\Delta LCI}$ and validate some of the findings using AlphaScreen™.

2.4.8 Cyclic peptide development as inhibitor for hTET1$_{CD}$

To date, only a few inhibitors have been reported to inhibit hTETs. To produce potentially potent and selective inhibitors for hTET1, Random nonstandard Peptide Integrated Discovery (RaPID) mRNA display cyclic peptide selection was carried out. The results of this research are published in

ChemBioChem.[47] The production of hTET1$_{CD/CCD}$ mRNA RaPID display, production hTET1$_{CD/CCD}$, DOTBlot and SPR were *conducted by Kosuke Nishio, Takayuki Katoh, Toru Sengoku, Kazuharu Hanada, Noboru Ohsawa, Mikako Shirouzu, Shigeeyuki Yokoyama and Hiroaki Suga*.

In brief, in the RaPID selection, the mRNA display is combined with flexible *in vitro* translation (FIT) whereby the start codon AUG is reprogrammed using flexizyme[252] to start with the initiator amino acid *N*-chloroacetyl-D-Tyrosine. The following sequence encoding for the peptide library sequence consists of NNK$_{4-15}$ were N represents the four bases found in RNA (A, U, G, C) and K represents U or G. This enables a generation of random peptide sequences with each position representing one of 19 amino acids (methionine is excluded). The sequence follows by encoding for UGC (cysteine) and a sequence encoding for a triple repeat of Glycine-Serine finalized by the stop codon. The cysteine installed can react spontaneously with *N*-chloroacetyl-D-Tyrosine to form thioether linked cyclic peptide. Using this broad mRNA sequence library, a panel of macrocycles consisting of 6–17 amino acids can be screened. It is worth to note that a puromycin linker attaches the mRNA at the 3'- sequence with a Poly Ethylene Glycol (PEG) linker. This enables covalent linking of the mRNA to the *in vitro* expressed peptide. Furthermore, the target protein in this method is hTET1$_{CD}$, tagged with an N-terminal Avitag™ encoding for N-terminal biotinylation to immobilize the protein using magnetic streptavidin coated beads (Dynabeads M280).

With the library screen and hTET1$_{CD}$ protein produced, seven successive rounds of RaPID selection were performed refining only the most potent cyclic peptides and the corresponding sequences of the peptides were transformed into *E. coli*. Enriched sequences were sequenced and analysed (*by Kosuke Nishio)*.

Cyclic peptide hits identified from the library generated, T$_C$P1 **2.42** were produced synthetically using SPPS and purified using HPLC (*by Kosuke Nishio)*. The binding affinity of T$_C$P1 **2.42** was tested

against immobilized hTET1$_{CD}$ on a sensor chip using surface plasmon resonance (SPR) and using DOT blot analysis. The obtained data was proven ambiguous suggesting potentially weak protein-peptide interactions that could have resulted as a side-effect from protein immobilization or interference of the LCI.

To avoid cyclic peptides being generated to an LCI-CD interface, a hTET1 compact catalytic domain (hTET1$_{CCD}$) was produced containing a truncated N-terminus, C-terminus and LCI relative to hTET1$_{CD}$ but hTET1$_{CCD}$ remained catalytically active. The RaPID selection procedure was repeated with the same mRNA library. The three cyclic peptides with the highest sequence occurrence based on next generation response were made synthetically using SPPS and purified using HPLC. The cyclic peptides TiP1–3 **2.39–2.41** were measured against hTET1$_{CD}$ and hTET1$_{CCD}$ using SPR (Table 2.4.3). The binding affinity (K_d) revealed double digit nano-molar binding affinities against hTET1$_{CCD}$ for TiP1-2 **2.39–2.40** but triple digit nano-molar against hTET1$_{CD}$.

Because the TiP1–3 **2.39**–**2.41** showed excellent binding affinities, the effect on the activity was measured against hTET1$_{CD}$ and hTET2$_{CD}$ using AlphaScreen™ together with reference molecule IOX1 **2.9** (Fig. 2.4.13). The IOX1 **2.9** IC_{50} data showed comparable values as observed in section 2.4.6 (Table 2.4.1), in part validating the screening experiments and confirming the reproducibility of the assay.

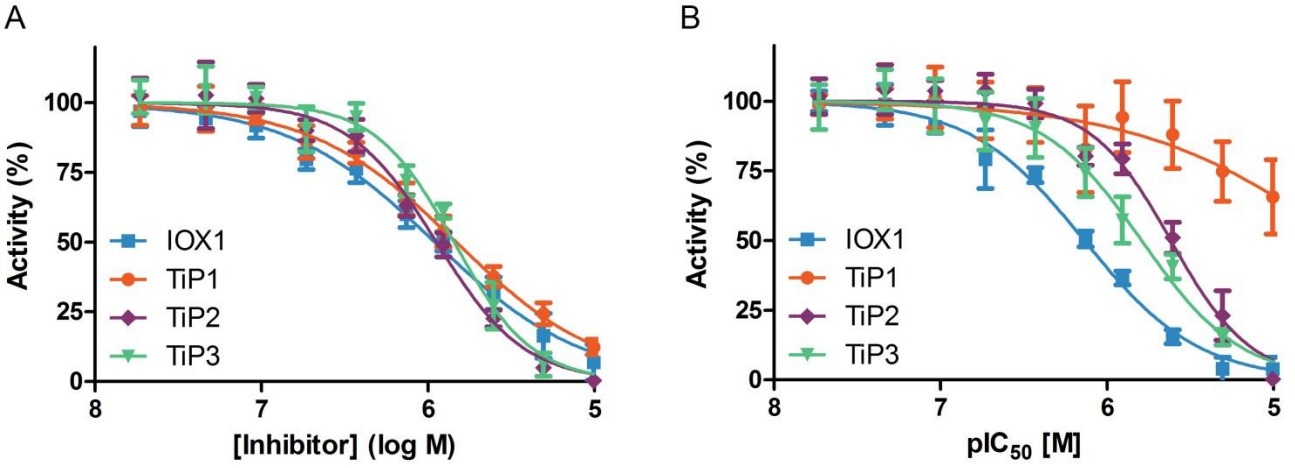


Figure 2.4.13: dose response curves of cyclic peptides TiP 1–3 2.39–2.41 against hTET1$_{CD}$ (A) or hTET2$_{CD}$ (B) as determined using AlphaScreen™. n=4–8, error given as s.d.

The cyclic peptides TiP 1–3 **2.39**–**2.41** demonstrate low single digit micro-molar affinities for hTET1$_{CD}$ (Table 2.4.3) and most striking results emerged from TiP 1 **2.39** against hTET2$_{CD}$ that showed some level of selectivity relative to hTET1$_{CD}$. TiP 2–3 **2.40–2.41** showed similar results between hTET1$_{CD}$ and hTET2$_{CD}$. The data obtained here identifies the first TET-specific generation of cyclic peptides revealing single digit micro-molar affinities.

Table 2.4.3: Binding affinity (SPR) and displacement of cyclic peptides (AlphaScreen™) against reference IOX1. n.d.: not determined.

		AlphaScreen		SPR					
		IC_{50} (μM)		K_D (μM)		k_a (10^5 $m^{-1}{\cdot}s^{-1}$)		k_d (10^{-2} s^{-1})	
		hTET1$_{CD}$	hTET2$_{CD}$	hTET1$_{CD}$	hTET1$_{CCD}$	hTET1$_{CD}$	hTET1$_{CCD}$	hTET1$_{CD}$	hTET1$_{CCD}$
IOX1	**2.9**	1.10	0.72	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
TiP1	**2.39**	1.48	>10	0.113	0.023	0.34	1.6	0.37	0.37
TiP2	**2.40**	1.13	2.42	0.116	0.049	0.20	0.74	0.36	0.36
TiP3M15L	**2.41**	1.44	1.62	0.216	0.217	0.23	1.6	3.5	3.5

2.5 Discussion

The conversion of $^{5m}C_{ss}$DNA-[Biotin] **2.1** to $^{5hm}C_{ss}$DNA-[Biotin] **2.2** mediated by hTET1$_{CD}$ and hTET2$_{CD/CD\Delta LCI}$ can be monitored using AlphaScreen™. The relatively low amount of DNA and enzyme required makes this an attractive method for the determination of TET activity and for inhibition studies. AlphaScreen™ enabled the detection of steady state kinetics of 2OG K_M^{App}: 4.8 ± 1.6 µM for hTET2$_{CD\Delta LCI}$ and K_M^{App}: 11.1 ± 3.5 µM for hTET2$_{CD}$ which is relatively similar to the previous reported K_M: 15.7 ± 2.8 µM using the same hTET2$_{CD\Delta LCI}$ construct.[46] The difference observed between hTET2$_{CD}$ and hTET2$_{CD\Delta LCI}$ implies that the LCI may influence the 2OG binding site, while hTET1$_{CD}$ showed a lower K_M^{App}: 2.4 ± 0.3 µM relative to the hTET2s.

AlphaScreen™ assay set up excludes its use for substrate DNA steady state kinetic analysis, due to the need for low substrate concentration (10 nM); however, ITC enabled the determination K_M of $^{5m}C_{ss}$DNA **2.5** substrate (23.8 ± 1.7 µM) which was found to be ~40-fold higher than observed for double stranded DNA.[61] Although a V_{max} and K_{cat} were determined, the current unit prevent direct comparison to reported data. To make the V_{max} and K_{cat} values more meaningful, a secondary experiment has to be acquired. A known amount of DNA substrate has to be converted to completion to afford the enthalpy of the reaction (energy per mol substrate). However, since TETs can convert ^{5m}C in iterative fashion to ^{5ca}C as well as ^{5hm}C and ^{5f}C, all substrates have to converted to ^{5ca}C to provide the enthalpy for each substrate (Fig. 2.4.6 C). With the enthalpy value, the corresponding reaction velocities can be calculated giving a more meaningful K_{cat} and V_{max}. The downside of the ITC method are the relatively long acquisition times and 'black-box' method (*i.e.* there is no indication of the ^{5m}C or ^{5ox}C composition).

MALDI-TOF experiments revealed that TET activity can be monitored as reported and enables Michaelis-Menten enzyme kinetics on double stranded DNA to give a K_M of 1.02 ± 0.23 µM, similar

to reported values of K_M: 0.48 ± 0.19 μM.[61] Although the effect of inhibitors on TET activity can be tested, the use of DMSO increased the difficulty of manual data point analysis (data not shown). In addition to this, under the experimental conditions, the MALDI-TOF cannot accurately distinguish between ^{5hm}C and ^{5f}C without additional derivatization.[253]

RF-MS proved to be a suitable alternative MS method to MALDI-TOF and permits the acquisition of double stranded DNA. To our knowledge the RF-MS method showed for the first continuous monitoring of the production of ^{5ox}C from ^{5m}C using TETs.

AlphaScreen™ and RF-MS were found to be suitable methods for high-throughput screening. While AlphaScreen™ is a useful screening method due to the low consumption of DNA (10 nM) and TET protein (1–10 nM), it also manifests fast data point acquisition (<1 sec). Downsides of AlphaScreen™ are the low concentration of DNA (10 nM, 10 μL) as this is not near the DNA substrate K_M, the limited use of Fe^{II}, potential interference of compounds with ^{5hm}C-antibody, streptavidin or beads, and potential absorbance of light signal (~520–620, ~680 nm) by inhibitors. In contrast, the RF-MS aided in the study of inhibitor development by enabling the monitoring of near the K_M of double stranded DNA (2.0 μM). Furthermore, the RF-MS is, relative to AlphaScreen™, slow with acquisition times of ~1 min and high consumption of DNA (2.0 μM, 25 μL) and protein (0.4 μM) but can monitor also simultaneously the production ^{5f}C or ^{5ca}C should that be necessary.

The inhibition of NOG **2.11** and 2,4-PDCA **2.10** against $hTET1_{CD}$ (pIC_{50}: 4.91 / 5.62) and $hTET2_{CD}$ (pIC_{50}: 5.05 / 5.29) are in range with reported values for other 2OG-dependent oxygenases with NOG **2.11** or 2,4-PDCA **2.10** typically found in a range of pIC_{50}: 6.52–3.52 (KDM6B, KDM7B) or pIC_{50}: 6.40–3.84 (KDM4A, KDM7B), respectively.[227] Of the ProHD2 inhibitors tested, IOX4 **2.17** and Vadadustat **2.19** inhibited while others such as GSK1278863 **2.12**, Roxadustat **2.13** and Molidustat **2.18**, showed weak inhibition (IC_{50}: ~100 or >100 μM) suggesting no direct structural link between

TET1/2 or ProHDs. IOX4 **2.17** showed reasonable inhibition ($hTET1_{CD}$ IC_{50}: ~15 μM) but was found to inhibit ProHD2 stronger at IC_{50}: 0.54 μM.[254] FG0041 **2.14**, a 1,10-phenanthroline derivative and 2OG mimic, is found to have relatively good inhibition to $hTET1_{CD}$ and $hTET2_{CD/CD\Delta LCI}$ while ProHD2 demonstrated a similar IC_{50} value of 10 μM.[255] ProHD3/2 inhibitor Vadadustat **2.19**, a phase 3 clinical candidate for chronic kidney disease[256], demonstrated single digit micro-molar inhibition levels to $hTET1_{CD}$ and $hTET2_{CD/CD\Delta LCI}$ (IC_{50}: 7.4–4.9 μM) but was found to inhibit ProHD2 at 1.33 μM (RF-MS).[254] Due to the relatively similar IC_{50} values, Vadadustat **2.19** may reveals off-target effects.

While some ProHD inhibitors inhibit hTET1/2, none of the tested KDM specific inhibitors did. KDM4 inhibitors SD70 **2.20**[257] / QC6352 **2.21**[258] and KDM5 inhibitors KDMOAM25[228] **2.22** / CPI455[259] **2.23**, showed little inhibition for the hTET1 and hTET2 constructs tested; however, the more generic KDM inhibitor ML324 **2.24** showed single digit micro-molar to sub-micro-molar inhibition using AlphaScreen™ (IC_{50}: 2.75–0.76 μM). Substantially higher values were found using RF-MS at IC_{50}: 17.4 μM for $hTET2_{CD\Delta LCI}$. In similar fashion, JIB-04 **2.25**[260] demonstrated a relatively low level of inhibition for hTETs'; particular $hTET2_{CD\Delta CLI}$ IC_{50}: 3.63 μM showed a good level of inhibition using AlphaScreen™. This observation was not confirmed for $hTET2_{CD\Delta LCI}$ with RF-MS demonstrating 47.8% activity at 100 μM inhibitor concentration. A potential reason for this observation could be that JIB-04 **2.25** is in part responsible for chelating to Fe^{II} limiting the accessibility to TET proteins. In the AlphaScreen™ assay, the level of Fe^{II} used is 10 μM Fe^{II} but in the RF-MS assay, 50 μM of Fe^{II} is used instead. There is possibility that as a result, a higher concentration of JIB-04 **2.25** is required for adequate Fe^{II} chelation, in essence shifting the level of inhibition observed while not actually inhibiting the $hTET2_{CD\Delta LCI}$ protein.

TET2 mutations have been linked to mutation in IDH1/2 where certain mutations are observed to be mutually exclusive in AML.[88] IDH1/2 would normally, as part of the Krebs cycle, convert isocitrate

into 2OG but, upon mutation, oncometabolite D-2HG **2.29** is produced instead. Sudhamalla *et al.* found that D-2HG **2.29** inhibits hTET2$_{CD\Delta LCI}$ at IC_{50}: 5.3 mM.[46] Here, inhibition studies using AlphaScreen™ against D-2HG **2.29** showed that hTET1$_{CD}$ has an IC_{50} of 416 µM and hTET2$_{CD\Delta LCI}$ an IC_{50} of 6.17 µM. RF-MS confirmed hTET2$_{CD\Delta LCI}$ values with IC_{50} of 3.47 µM suggesting a much higher impact than previously considered which may explain the mutual exclusivity between hTET2-IDH1/2. With cellular levels up to 30 mM[89] it is possible that D-2HG **2.29** can act inhibitory towards hTET2 but also towards hTET1. TCA-cycle intermediates succinate **2.26** and fumarate **2.27** were detected to have inhibitory levels of IC_{50}: ~100 µM for the TETs using AlphaScreen™.

The nucleosides ^{5m}C **2.30** and ^{5hm}C **2.31** were examined using AlphaScreen™ but were found to interfere with the assay. When tested in the presence of hTET2$_{CD\Delta LCI}$ using RF-MS, no level of inhibition was found at 100 µM. This data suggests that the DNA strand back bone that provides interaction to the TET proteins[45] are of vital importance to binding and catalytic turnover. Similar observations were made for small oligonucleotides, where ^{5m}CGC was not found to be a substrate but longer 6-meric oligo CG^{5m}CGGC was.[67]

With the three TET protein constructs measured under the same AlphaScreen™ conditions, a direct comparison can be made and potential trends can be observed. In general, little variation is found in IC_{50} values for the inhibitors screened between hTET1$_{CD}$ and hTET2$_{CD}$ with the exception for JIB-04 **2.25** and the D/L-2HG **2.28**, **2.29**. Between hTET2$_{CD}$ and hTET2$_{CD\Delta LCI}$ only a small difference (ΔpIC_{50} <0.41) between pIC_{50} values was found for using AlphaScreen™ of the inhibitors tested. Some differences are found between AlphaScreen™ and RF-MS in pIC_{50} values depending on the inhibitor tested. For NOG **2.11**, RF-MS showed a higher level of inhibition (ΔpIC_{50}: 1.12) than AlphaScreen™ while a lower inhibition was found for JIB-04 **2.25** (ΔpIC_{50}: >1.0) and ML324 **2.24** (ΔpIC_{50}: 1.36). Overall, these combined findings suggest that hTET1 and hTET2 are structurally very similar and as a result, it may be challenging to generate selective inhibitors for the individual proteins.

In collaboration with Nishio *et al.* the first panel of (cyclic peptide) inhibitors were produced for hTET1 showing single digit micro-molar inhibition. Interestingly, TiP1 **2.39** was found to possess some selectivity (>5-fold) for hTET1$_{CD}$ and hTET2$_{CD}$ tested. It is; however, not yet assessed as to what extend this is selective over hTET3-CD or other 2OG-dependent oxygenases.

2.6 Conclusion

The use of AlphaScreen™ empowered a robust, repeatable and fast method to determined hTET activity and revealed a K_M^{App} of 2OG of $hTET1_{CD}$ (K_M^{App}: 2.4 ± 0.3 µM), $hTET2_{CD}$ (K_M^{App}: 11.1 ± 3.5 µM) and $hTET2_{CD\Delta LCI}$ (K_M^{App}: 4.8 ± 1.6 µM) in line with reported values. ITC provided the biophysical characterisation for single stranded DNA (K_M: 23.8 ± 1.7 µM) showing a substantially higher K_M than previously observed for double stranded DNA. The MALDI-TOF obtained data validates the DNA substrate K_M of 1.02 ±0.23 to reported values[61] but was found unsuitable as high throughput screening method. The use of RF-MS facilitated the real-time detection of ^{5m}C oxidation by $hTET2_{CD\Delta LCI}$ in sequential steps to ultimately yield ^{5ca}C.

AlphaScreen™ and RF-MS were successfully applied as two high throughput screening methods that can determine the inhibitory effects of compounds against $hTET1_{CD}$ and $hTET2_{CD}$ proteins. The screened inhibitors showed that generic 2OG-dependent inhibitors with TET proteins tested and are in a similar range to previously described levels of 2OG-dependent oxygenases. The ProHD2/3 inhibitors, in particular, IOX4 **2.17**, FG0041 **2.14** and Vadadustat **2.19**, inhibit $hTET1_{CD}$ and $hTET2_{CD/CD\Delta LCI}$ as demonstrated by AlphaScreen™ and confirmed using RF-MS. The target-specific KDM4 and KDM5 inhibitors showed poor inhibition against hTET1/2 proteins; however, the generic KDM inhibitor ML324 **2.24** revealed single digit micro-molar levels against $hTET2_{CD\Delta LCI}$. TCA cycle intermediates succinate **2.26** and fumarate **2.27** were found to inhibit at IC_{50}: ~100 µM. Oncometabolite D-2HG **2.29** was found to inhibit $hTET2_{CD\Delta LCI}$ with an IC_{50}: 6.17 µM but $hTET1_{CD}$ with an IC_{50}: 416 µM which may explain the mutual exclusivity observed between $hTET2_{CD}$ and IDH1/2 upon mutation in AML.[88] Furthermore, L-2HG **2.28** and D-2HG **2.29** are the only small molecules tested as inhibitors that show two orders of magnitude difference in inhibition levels between $hTET1_{CD}$ and $hTET2_{CD}$. The nucleosides ^{5m}C **2.30** and ^{5hm}C **2.31** revealed no inhibition at

100 μM concentration against $hTET2_{CD\Delta LCI}$ as determined with RF-MS. This suggests that other nucleotides in a DNA strand are of vital importance to binding and potential catalytic turnover.

Using mRNA display, the first cyclic peptides target $hTET1_{CD}$ specific were produced with AlphaScreen™ revealing single digit micro-molar IC_{50} values. In, particular TiP1 **2.39** demonstrated some selectivity for $hTET1_{CD}$ in relation to $hTET2_{CD}$ tested.

2.7 Future perspective

Having established two high throughput screening methods to determine potential inhibition of TET proteins, it is now possible to perform inhibitor library screening against hTETs. Select small molecule and cyclic peptide inhibitors were identified and could provide new starting points for the development of selective TET inhibitors. For example, inhibitors such as IOX4 **2.17**, Vadadustat **2.19** or ML324 **3.24** revealed to have some level inhibition for hTET1/2$_{CD}$ proteins and can be used as suitable scaffold for the development of inhibitors. In addition, the development of a cellular based assay and the screening of the tested inhibitors here could aid to understand the effect of these inhibitors on cellular level.

The screening efforts made here, validated the selectivity of certain inhibitors over hTET1/2 exemplified by GSK1278863 **2.12**, Roxadustat **2.13** and CPI455 **2.23**. On the contrary, some inhibitors such as IOX4 **2.17**, Vadadustat **2.19** or ML324 **3.24** revealed to have some level of inhibition for hTET1/2$_{CD}$ proteins. Therefore, interpretation of particular cellular data or *in vivo* screening should be taken with caution as these compounds may show off-target effects.

The cyclic peptides developed show promising results for the inhibition of hTET1$_{CD}$ and in part hTET2$_{CD}$. However, the selectivity in relation to hTET3 and other 2OG-dependent domain containing proteins such as FTO, ABH-family and KDMs is yet to be determined. Furthermore, the mRNA selection also provides other cyclic peptides not discussed here of which the level of inhibition is unknown and may provide an alternative to TiP1–3.

Overall, the work in this Chapter describes the first set of μM inhibitors for the human TET enzymes which can function as the starting point for further medicinal chemistry refinement.

2.8 Materials and Methods

2.8.1 General conditions

General synthetic and microbiological conditions are described in Chapter 7.1 and 7.2.

2.8.2 Protein production for *in vitro* analysis

$hTET1_{CD}$ for AlphaScreen™

$hTET1_{CD}$ (E1418–V2136) was obtained from Epigentek (cat#: E12002-1), based on NP_085128.1 sequence. The proteins was produced with a N-terminal Flag-tag (>80% purity).

$hTET2_{CD\Delta LCI}$ for AlphaScreen™ and RF-MS (provided by Eidarus Salah)

The plasmid encoding for $hTET2_{CD\Delta LCI}$ (D1129–S1481_linker(GGGGSGGGGSGGGGS)_N1843–G1936) in pET-28 vector was gratefully gifted by Yanhui Xu. The construct was transformed into phage resistant *E. coli* BL21 (DE3) containing pRARE2, a single colony was picked and a starter culture was initiated. Large scale growth containing kanamycin (100 $mg \cdot L^{-1}$) was inoculated with starter culture and grown at 37 °C with shaking. At OD_{600}: ~1.0, cells were cooled to 18 °C and induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) (0.5 mM) and grown with shaking for 18 hrs. The cells were harvested by centrifugation, suspended in lysis buffer [Imidazole (20 mM), TCEP (0.5 mM), Glycerol (5%), NaCl (500 mM), HEPES (50 mM) in water (pH 7.4)] supplemented with protease inhibitor (Complete, EDTA-Free protease inhibitor cocktail (Roche Diagnotics LTD) (1:2000) and sonicated (4.0 °C). The lysed cells were centrifuged and the supernatant containing soluble proteins was collected and His-tagged hTET2 was purified using Ni NTA column. The column was washed, protein was eluted [Imidazole (300 mM), TCEP (0.5 mM), Glycerol (5%), NaCl (500 mM), HEPES (50 mM) in water (pH 7.4)] and concentrated using centrifugation and spin column. The concentrate

was loaded onto size exclusion gel filtration column (Superdex S200, GE Healthcare) in gel filtration buffer [TCEP (0.5 mM), HEPES (20 mM), NaCl (150 mM), glycerol (5%) in MilliQ (pH 7.4)]. Fractions were pooled based on LC-MS and SDS-PAGE (Fig. 2.8.1 A) observations yielding, 9.6 mg of $hTET2_{CD\Delta LCI}$ (1.0 $mg \cdot mL^{-1}$).

$hTET2_{CD}$ for AlphaScreen™, ITC and MALDI-TOF (*provided by Aleksandra Szykowska*)

Human TET2 construct was prepared from hTET2 (Q969–Y2001) gene by Ligase independent cloning (LIC) in Baculo vector pFB-CT10HF-LIC (Addgene plasmid: 39191). The corresponding construct was amplified with Tn7 in a DH10Bac cell line followed by transfection into Sf9 insect cells using jetPrime® transfection reagent (Polyplus transfection), and grown in Insect-Xpress medium (Lonza Bioscience, cat#: BE12-730Q) at 27 °C for 72 hrs while shaking (100 rpm). The cells were pelleted by centrifugation and suspended in lysis buffer [Imidazole (10 mM), TCEP (0.5 mM), Glycerol (5%), HEPES (100 mM), NaCl (500 mM) in water (pH 7.4)] in the presence of protease inhibitor cocktail (Calbiochem setIII, EDTA free, MERCK) followed by lysing in a Dounce homogeniser. The suspension was centrifuged and TET protein was isolated using Nickel-Sepharose 6 FF (GE Healthcare) for 1 hrs at 4 °C while mixing. The supernatant was removed by gravity filtration and the beads were washed with Imidazole (50 mM) in buffer, Imidazole (70 mM) in buffer and Imidazole (100 mM) in buffer. Elution was achieved with Imidazole (250 mM) in buffer and the eluted filtrate was concentrated. The concentrate was purified using size exclusion gel chromatography (Superdex 200, 16/60, GE Healthcare) in the buffer [TCEP (0.5 mM), Glycerol (5.0%), HEPES (50 mM), NaCl (500 mM) in MilliQ (pH 7.4)]. Selected fractions based on SDS-PAGE (Fig. 2.8.1 B) and MS/MS tryptic digest analysis were pooled, concentrated and stored at −80 °C.

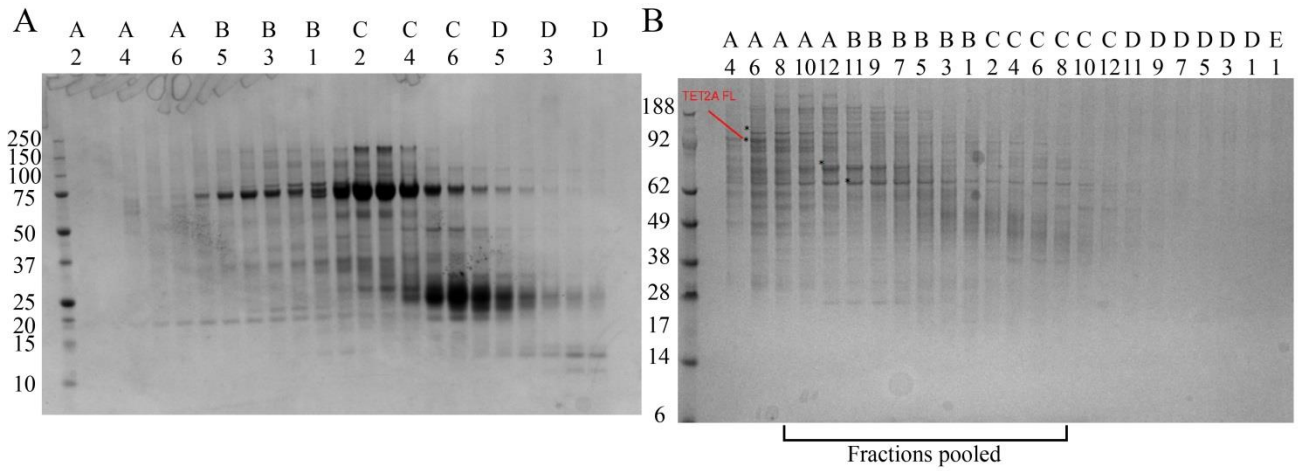


Figure 2.8.1: Visualization of S200 size exclusion fractions on SDS-PAGE of hTET2$_{CD}$ (A) and hTET2$_{CD\Delta LCI}$ (B) displaying the fraction (x-axis) and kDa (y-axis).

2.8.3 Inhibition of hTET proteins using AlphaScreen™

DNA purification

[Biotin]-5'-TCG GAT GTT GTG GGT CAG ^{5m}CGC ATG ATA GTG TA-3' ($^{5m}C_{ss}$DNA-[Biotin] **2.1**) and [Biotin]-5'-TCG GAT GTT GTG GGT CAG ^{5hm}CGC ATG ATA GTG TA-3' ($^{5m}C_{ss}$DNA-[Biotin] **2.2**) 32 base pair single stranded DNA were produced by ATDBio Ltd (Southampton, UK) at 30 µmol scale. The crude was purified in 6 runs (approximate 5 µmol per injection) using Gilson 811C mixer and Gilson 118 UV/VIS detector programmed at 298 nm monitoring wavelength and 1.00 intensity. Recordings were made using Kipp & zonon recorder 2 mm·min^{-1} at 20 mV sensitivity and suppression off. Luna 10µ C8 721971_4, 100 Å 250 x 10 mm column was used in combination with a solvent composition: 0.0–3.0 min 0.0% B, 3%–10% B over 3.0–3.5 min, 10%–50% B over 3.5–15.0 min, 15.0–15.5 min 50%–100% B, 15.5 to 17.0 min 100% B, 17.0 to 17.5 min 100%–0% B, 17.5–20 min 0.0%. Buffer A: ammonium bicarbonate (100 mM, pH <8.0 typically ~7.5) in MilliQ, Buffer B: ammonium bicarbonate (50 mm, pH <8.0 typically ~7.5) in acetonitrile (50 v/v%) and MilliQ (50 v/v%). The mass and purity of selected fractions were confirmed using Bruker microTOF II (ESI-TOF, negative mode), >95% purity. Two predominant peaks were observed with the first eluting be the unlabelled DNA and the second the biotin labelled DNA.

Standard activity assay

Buffer [HEPES (50 mM) (Apollo Scientific Limited, Code: BI8181), NaCl (150 mM) (Fisher Bioreagents®, Cat#: BP358-10), Tween20® (0.01%) (Promega, Ref#: H5151), BSA (0.1%) (PerkinElmer, 7.5%, Cat#: CR84-100) in MilliQ (pH 7.0)] was prepared and ^{5hm}C- antibody (Active Motif, 5-hydroxymethylcytosine cat#: 39769), was 200× diluted (1 antibody µL + 199 µL buffer). Beads mixture was prepared [Acceptor beads (62.5× dilution), Donor Beads (62.5× dilution) (AlphaScreen™ PerkinElmer® IgG detection Kit (Protein A) cat#: 6760617M, Lot#: 1571768), antibody solution (20× dilution) in buffer] in cold room (~4 °C) and dark vicinity followed by incubation for at least 30 min in a dark environment.

A solution DNA mixture solution was prepared consisting of: $^{5m}C_{ss}$DNA **2.1** (20 nM), Sodium L-ascorbate (Asc) (200 µM) (Sigma cat#: 11140), Ammonium iron(II) sulphate hexahydrate ($(NH_4)_2Fe(II)(SO_4)_2$) (20 µM) (Sigma Aldrich cat#: 215406-100G), α-Ketoglutaric acid disodium salt (2OG) (20 µM) (Sigma cat#: K3752-100G) in buffer (5.0 µL). The DNA mixture was pipetted using E1-ClipTip™ (Thermo Scientific, 16-ch 2–125 µL, part#: 4671060) and ClipTip™ tips (ThermoFisher Scientific, cat#: 94410150) in ProxiPlate™ -384 (PerkinElmer, White Shallow 384-well, part#: 6008280). At time point 0 min, the DNA mixture was pre-quenched with EDTA solution (30 mM in MilliQ, pH 4.2) (Sigma cat#: E5134-500G) (5.0 µL). Enzyme mixture in buffer (5.0 µL) was added and wells were quenched with EDTA (5.0 µL) at the appropriate time points. To the plate was added, in a dark vicinity, beads mixture (5.0 µL) and incubated in the dark for 35–60 min, typically for 45 min. The plate was scanned using PerkinElmer Envision 2104 Multilabel Plate Reader and data acquired processed using a combination of MicroSoft™ Excel™ 2010 and GraphPad Prism ® 5 (v. 5.04). Each time-points was prepared in triplicate. Standard curve was prepared from $^{5m}C_{ss}$DNA-[Biotin] **2.1** and $^{5hm}C_{ss}$DNA-[Biotin] **2.2** source stocks in buffer (10.0 µL), EDTA and Beads mixture was added as described for the activity assay.

IC_{50} inhibition assay

IC_{50} AlphaScreen™ inhibition assay is derived from the standard activity assay and adjustments are described.

ProxiPlates™ were prepared using Echo Acoustic dispenser (Labcyte, Echo 550) from DMSO (Fischer Scientific, cat#: D/4120/PB08) stock solution containing inhibitor (10 mM) using source and intermediate plates 384 LDV plates (LabCyte, cat#: LP-0200). Echo dispensing was performed in 2.5 nL iterative steps up to 100 nL DMSO and wells were backfilled to 100 nL DMSO in necessary. Dilution series were made robotically.

To ProxiPlate™ was added hTET in buffer (Table 2.8.1) (5.0 µL) using ClipTip-E1 and incubated for 10 min. To the mixture was added DNA mixture containing $^{5m}C_{ss}$DNA **2.1** (20 nM), Asc (200 µM), $(NH_4)_2Fe(II)(SO_4)_2 \cdot 6H_2O$ (20 µM), 2OG (20 µM) in buffer (5.0 µL) and reacted for the appropriate time (Table 2.8.1) followed by quenching with EDTA (30 mM, pH 4.2 in MilliQ) (5.0 µL). The beads mixture was added and incubated in the dark 40–60 min and samples were measured using the PerkinElmer plate reader. Acquired data was analysed with Microsoft Excel™ and GraphPad Prism®. Plates were prepared with n= ≥3 and each plate was performed as technical replicate.

Table 2.8.1: TET time and concentration in inhibition assay.

	Reaction time	TET concentration
$hTET1_{CD}$	30 min	10 nM
$hTET2_{CD}$	10 min	1 nM
$hTET2_{CD\Delta LCI}$	10 min	1 nM

2.8.4 Enzyme kinetics of hTET2$_{CD}$ of $^{5m}C_{ds}$DNA using MALDI-TOF

Activity assay

To a solution of palindromic $^{5m}C_{ds}$DNA **2.33** 5'-ACC AC^{5m}C GGT GGT-3' (10.0 µM), Sodium L-ascorbate (1.0 mM) (Sigma cat#: 11140), Ammonium iron(II) sulphate hexahydrate (100 µM) (Sigma Aldrich cat#: 215406-100G), α-Ketoglutaric acid disodium salt (200 µM) (Sigma cat#: K3752-100G) in buffer [HEPES (50 mM), NaCl (150 mM) in MilliQ (pH 7.1)] (5.0 µL) in ProxiPlate™ (PerkinElmer®, cat#: 6008280) was added hTET2$_{CD}$ (0.8 µM) in buffer (5.0 µL). All wells were quenched with hydrogel [Bio-Rad cation exchange resin (AG 50W-X8 biotech grade, 100-200 mesh, cat#: 1435441) in MilliQ] (10 µL). On target plate (Bruker MTP ground steel Bruker) was spotted [3-HPA (50 mg·mL^{-1}) (Sigma-Aldrich, 3-hydroxypicolinic acid, 56197-250MG), Ammonium citrate dibasic (10 mg·mL^{-1}) (Fluka®, cat#: 09831-100G) in (acetonitrile: [0.1% trifluoroacetic acid (Acros Organics, cat#: 139721000) in MilliQ] (1:1))] (1.0 µL), dried at 37 °C followed by sample (1.0 µL) and dried at 37 °C. Spots were measured on Bruker Ultraflex II MALDI-TOF/TOF in linear positive mode. Acquired data was analysed using Microsoft Excel™ (2010) and GraphPad Prism® (v5.04).

2.8.5 Enzyme kinetics of $^{5m}C_{ss}$DNA for hTET2$_{CD}$ of ITC analysis

Activity assay

A co-factor-buffer solution was prepared containing Asc (2 mM), $(NH_4)_2Fe(II)(SO_4)_2$•$6H_2O$ (200 µM), 2OG (200 µM) in a solution of HEPES (20 mM) and NaCl (100 mM) in MilliQ (pH 7.5). hTET2$_{CD}$ (5 µL, 80 µM) was diluted in co-factor buffer (395 µL) to hTET2$_{CD}$ (400 µL, 1.0 µM). DNA was prepared by diluting 5'-TCG GAT GTT GTG GGT CAG ^{5m}CGC ATG ATA GTG TA-3' ($^{5m}C_{ss}$DNA, obtained side product from unlabelled $^{5m}C_{ss}$DNA, see section 2.8.2 DNA purification) (40 µL, 1189 µM) in co-factor buffer (100 µL) giving solution of $^{5m}C_{ss}$DNA (140 µL, 339.7 µM). Solution were prepared deep well 96-well plate (Fisher Brand™ 96-Well DeepWell™, Polypropylene Microplates, Non-Sterile, 12-566-611) with pre-rinse solution co-factor buffer. hTET2$_{CD}$ was placed in the sample cell and the DNA

solution was titrated to the reference cell. The negative control was used as co-factor buffer in the absence of DNA. Operation of the ITC was done using Microsoft Windows 7 and MicroCal PEAQ-ITC Automated Operation software and MicroCal PEAQ-ITC Automated control software (v 1.1.3.0). Obtained data were analysed using MicroCal PEAQ-ITC Analysis Software (v 1.1.0.1262), MicroSoft Excel, and GraphPad Prism (v. 5.04). Calculations were made by adjusting the negative control to the same entering power, and the sample was subtracted from the corrected control. Points were taken based on 1 minute average between 84 sec and 144 sec.

ITC settings:

For ITC acquisitions, Automated ITC PEAQ was used by Malvern using default settings with the program: plate prerinse quick. Reference power: 10 µCal·sec^{-1}, cell volume fill: 280 µL, gain feedback: high, cell temperature: 25.0 °C, tray temperature: 18.0 °C stirring speed: 750 rpm, Injections: 19, 1st injection: 0.4 µL, injection 2–18: 2.0 µL, injection speed: 0.5 µL·sec^{-1}, initial delays: 60 sec, Injection spacing: 150 sec.

2.8.6 RapidFire mass spectrometry (RF-MS) for hTET2$_{CD\Delta LCI}$ inhibition

Default RF-MS settings

Acquisitions were performed using Agilent 360 RapidFire in couple to a QTOF6530.

Samples were vacuum aspirated (50 µL, ms), pass through SPS cartridge (Type A: C_{18} Agilent) with buffer A [Octyl ammonium acetate (6 mM) in water (LiChroSolv® for LC-MS, VWR cat#: (1.15333.2500)], washed with buffer A (4000 ms) and eluted with buffer B [80% Acetonitrile (LiChroSolv® for LC, VWR, 1.00030.2500) in Water] (4000 ms). Pumps were set as pump 1: 1.5 mL·min^{-1}, pump 2: 1.25 mL·min^{-1} and pump 3: 0.8 mL·min^{-1}. Wash station were equipped with inorganic and organic wash of water and acetonitrile, respectively. Between each sample the two rounds of inorganic wash and organic wash was performed. Each sample taken was time stamped.

QTOF settings was set to Gas Temperature: 350 °C, Flow rate: 9 L·min^{-1}, Nebulizer: 40 PSIG, Sheath Gas Temperature: 400 °C, Flow rate: 11 L·min^{-1}, Fragmentor: 135 V, Skimmer: 65 V, OCT1 RF Vpp: 750 V. Acquisition pace: 5 spectra·sec^{-1}. Collected data was processed using Agilent Qualitative analysis (v. B.06.00), MicroSoft Excel™ and GraphPad Prism® (v. 5.04). Data were visualized with Adobe® Illustrator® (v. 15.0)

RF-MS in multi reaction monitoring mode

A solution of co-factors sodium ascorbate (240 µM), $(NH_4)_2Fe(II)(SO_4)_2 \cdot 6H_2O$ (60 µM), disodium 2-oxoglutarate (60 µM) and $^{5m}C_{ds}$DNA substrate (1.2 µM) in buffer [HEPES (50 mM) in MilliQ (pH 7.3)] (550 µL) was prepared in plate (Greiner, Masterblock® (PP, 2.0 mL, V-bottom, 96-well, Ref#: 780271). LC-MS t= 0 min was acquired (~50 µL aspiration) followed by the addition of hTET2$_{CD\Delta LCI}$ (2.4 µM, 100 µL) solution, mixed and samples acquisition was continued with 69.4 sec interval. First mixture time point was manually observed, subsequent time points were automatically time stamped.

RF-MS in pre-quenched mode

Plates were prepared using LabCyte Echo 550 with source plate and intermediate plates (Labcyte, Echo® (384PP) cat#: P-105525) from DMSO (Fischer Scientific, cat#: D/4120/PB08) stock solution containing inhibitor (10 mM). Iterative volumes of 2.5 nL were dispensed up to 250 nL and back filled with DMSO to 250 nL, if necessary. Intermediate plates were prepared robotically and collected in destination plate (Greiner, Microplate (PP, 384 well), Ref#: 781280). A solution of hTET2$_{CD\Delta LCI}$ (0.8 µM) was added to the destination plate using liquid handling robot (Thermo scientific, Multidrop™ Combi, equipped with Small tube metal tip dispensing cassette, cat#: 24073296) in buffer [HEPES (50 mM) in MilliQ (pH 7.3)] (12.5 µL). The solution was incubated for 5.0 min and DNA mixture was added containing $^{5m}C_{ds}$DNA (4.0 µM), sodium ascorbate (400 µM), $(NH_4)_2Fe(II)(SO_4)_2$ (100 µM), disodium 2-oxo-glutarate (10 µM) in buffer (12.5 µL). Solution were quenched at the required time

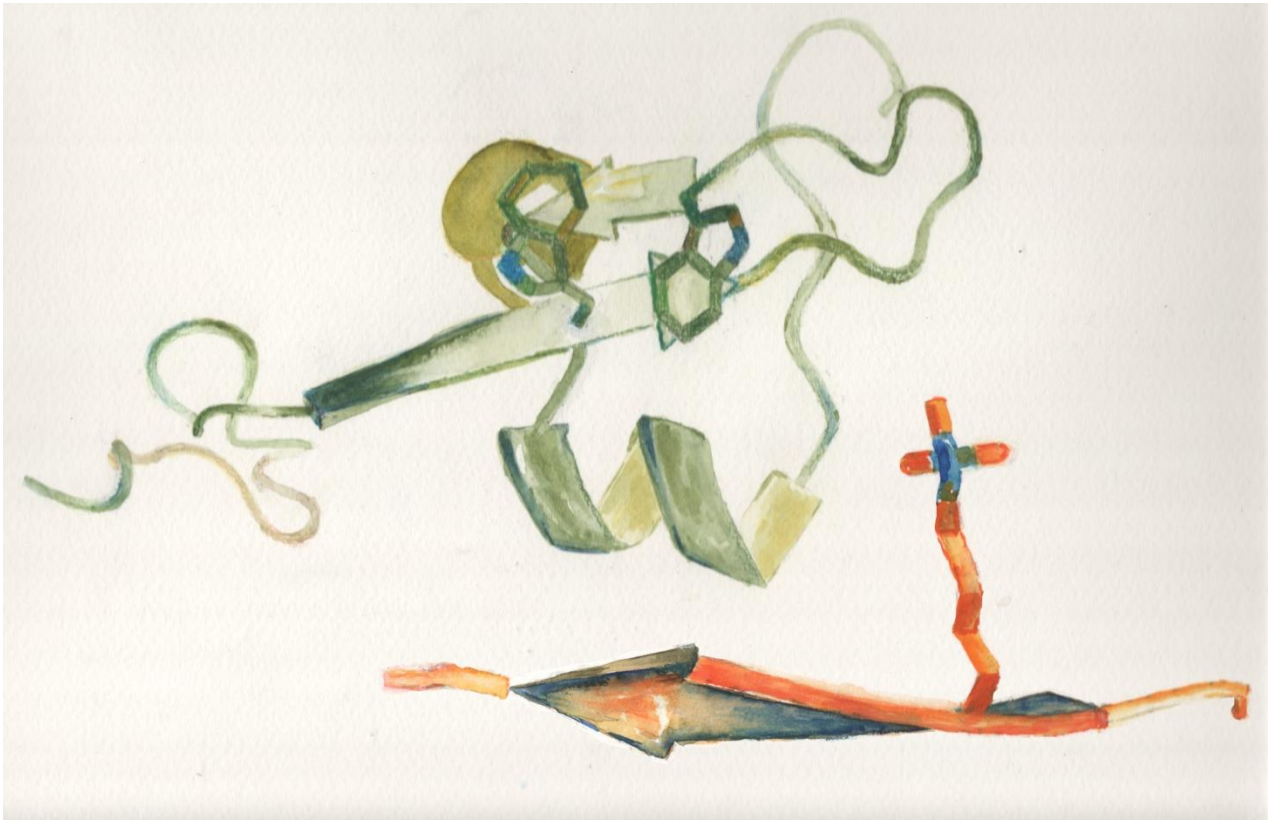
with N-Oxalylglycine (NOG) (2.0 mM in MilliQ) (25.0 µL) and plate were measured using RF-MS in default mode.

Adjustments were made for oncometabolites and TCA cycle intermediates

L/D-2HG, succinate and fumarate solutions were made in MilliQ and echo and intermediate stocks in MilliQ were made manually. Plates were not back filled with MilliQ and plates were dried upon air prior to addition of $hTET2_{CD\Delta LCI}$ mixture. The assay performed without further alterations.

Experiments to determine linear range in pre-quenched mode

In order to determine the linear region of $hTET2_{CD\Delta LCI}$ minor modifications were made. DNA mixture was added using 16 multi-channel pipette E1-ClipTip™ (Thermo Scientific, 16-ch 2–125 µL, part#: 4671060) and ClipTip ™ tips (ThermoFisher Scientific, cat#: 94410150) in the Greiner 384 well plate followed by addition of enzyme mixture. The DNA mixture contained an additional 2% DMSO and the rest as proceeded previously described.



Chapter 3. Investigating the PHD3-finger of KDM5A

3.1 Contributors to this Chapter

Dr. Catrine Johansson produced $KDM5A_{C1}$, $KDM5A_{C2}$, $KDM5B_{C1}$ and Aleksandra Szykowska produced $KDM5C_{C1}$. $KDM1A_{FL}$ was expressed by Raphael Reinbold and $KDM5D_{C1}$ was made by Eidarus Salah. 2TY media was prepared by H. Manso Jubier for $KDM1A_{FL}$.

3.2 Introduction to histone H3 *N*$^{\varepsilon}$-lysine methylation

3.2.1 Histone *N*$^{\varepsilon}$-lysine methylation in the chromatin

Transcriptional regulation is partially mediated by the read out of histone *N*$^{\varepsilon}$-lysine methylation positioned at various lysines throughout histones in nucleosomal complexes. This post translational modification (PTM) has been linked to different transcriptional outcomes, depending on the level of methylation or the lysine position within the nucleosomal complex. For example, histone H3 *N*$^{\varepsilon}$-trimethyl-lysine 4 (H3K4me$_3$) is considered to be transcriptionally activating, leading to downstream protein production, while H3K4me$_0$ and H3K9me$_3$ have been associated to transcriptional silencing.[119,261] The dynamics of lysine methylation are regulated by three types of proteins (Fig. 3.2.1). Installation of methyl groups is achieved by histone lysine *N*$^{\varepsilon}$-methyltransferases (KMTs) (Fig. 3.2.1 A), such as the KMT2 (Mixed-Lineage Leukaemia (MLL)) family[262] or KMT7 (SET domain 7 (SETD7)),[263] resulting in iterative methylation of lysine to *N*$^{\varepsilon}$-monomethyl- (K4me$_1$), *N*$^{\varepsilon}$-dimethyl- (K4me$_2$) and to *N*$^{\varepsilon}$-trimethyl-lysine (K4me$_3$) (See also Chapter 1.2.2). Active readout is provided by chromatin binding modules which facilitate downstream effects. Examples include: binding of TATA box-binding protein-associated factor (TAF3$_{PHD}$, Fig. 3.2.1 B),[235] a Plant Homeodomain (PHD) finger that is part of a transcriptional factor TFIID, or Chromo ATPase/Helicase DNA binding domain (CHD1$_{DCD}$)[264] involved in the regulation of ATP dependent nucleosome assembly and movement through the chromatin landscape. Both TAF3 and CHD1 have been demonstrated to have increased binding affinity to higher methylation states on H3K4.

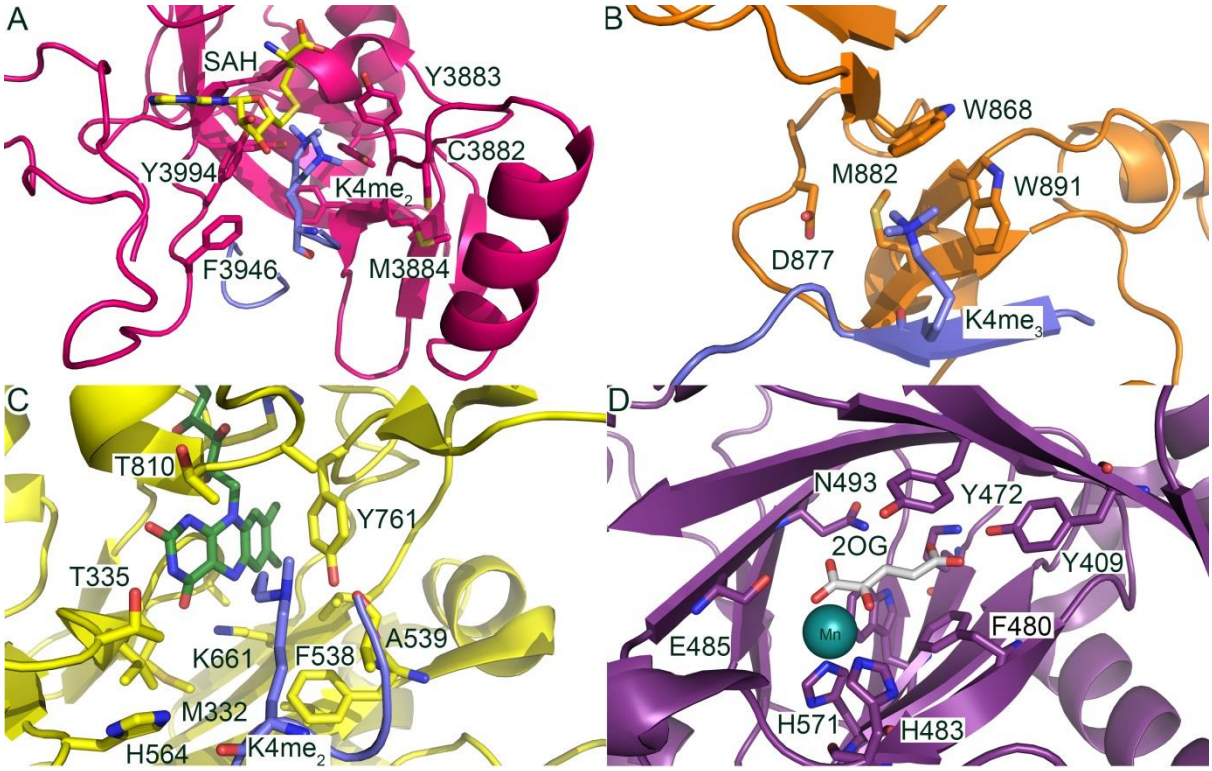


Figure 3.2.1: Crystal or solution structures of H3K4 associated chromatin effector proteins. A) Active site of KMT2A (PDB: 2W5Z) displaying co-substrate SAM (yellow), H3K4me$_2$ (blue) and KMT2A (pink); B) Solution structure of chromatin reader module TAF3 (orange) with H3K4me$_3$ (blue) bound (PDB: 2K17); C) KDM1A (yellow) active site, FAD (green) and peptide containing H3K4me$_2$ derivative (blue) (PDB: 2UXN) and D) Active site of KDM5A (purple) with manganese (cyan) and 2OG (grey) (PDB: 5E6H). Heteroatoms oxygen (red), nitrogen (blue) and sulphur (yellow) displayed are coloured.

Lysine methylation is reversed by histone lysine specific demethylases (KDMs) containing either an amine oxidase domain or a JumonjiC (JmjC) domain. KDM1A (Lysine specific histone demethylase 1 (LSD1)) is the first discovered histone lysine demethylase and possesses the capacity to demethylate H3K4me$_{2,1}$ using an amine oxidase domain (Fig. 3.2.1 C).[117] KDM5s, also known as JumonjiC (JmjC) AT-Rich Interaction Domain 1 (JARID1) proteins, are JmjC-domain containing protein that is able to demethylate histone H3 *N*$^{\varepsilon}$-methyl-lysine 4 (H3K4me$_{3\text{-}1}$) and are vital to the chromatin landscape.[191,192,265,266] KDM5s are the only known proteins capable of demethylating the H3K4me$_3$ mark.

3.2.2 Introduction to demethylase KDM5A

KDM5s belong to the 2OG oxygenase superfamily and form a second class of histone lysine demethylases. Four KDM5 proteins have been identified: KDM5A (RBP2, JARID1A), KDM5B (PLU-1, JARID1B), KDM5C (SMCX, JARID1C) and KDM5D (SMCY, JARID1D). Genes encoding for *KDM5A* and *KDM5B* are found on chromosome 1, whereas *KDM5C* is found on the X chromosome and *KDM5D* on the Y chromosome. With the wild-type KDM5s being between 1544 (KDM5B) and 1690 (KDM5A) amino acids in length, the molecular weight of KDM5 enzymes exceed that of an unmodified single nucleosome indicating the size of the protein. KDM5s contain multiple domains: the JumonjiN (JmjN) domain, two or three Plant Homeodomains (PHDs), an AT-Rich interaction (ARID) domain, a JmjC-domain (Fig. 3.2.1 D), a C5HC2 zinc-binding domain and a PLU-1 domain. The multi-domain architecture of KDM5s adds to the complexity in understanding their molecular mechanism and complicates their biochemical characterisation. Consequently, it is vital to understand the individual domains prior to understanding the cellular implication of their biochemistry / biophysical properties.

A schematic representation of the KDM5 sequence and domain alignment is displayed in figure 3.2.2. The JmjN-domain (Fig. 3.2.2, light-blue (KDM5A: E18-R76)) is found in the KDM4 and KDM5 subfamilies and is proposed to be a stabilization domain that supports JmjC activity. A unique feature of the KDM5 proteins is that the JmjN- and JmjC-domains are split by an insertion of the ARID and PHD1 domains. Deletion constructs have demonstrated that in the absence of the JmjN-domain, the JmjC catalytic activity is diminished[267] but deletion constructs in which ARID, PHD1/2/3 are removed and JmjN-JmjC are connected by a small linker remain active.[268] Initial studies suggested that JmjN- and JmjC-domains were always found together[269] but was later disproved when KDM6 or KDM7 were found to lack the JmjN-domain but demonstrate demethylation activity.[167,229] The ARID (KDM5A: R85-Q173) domain (Fig. 3.2.2, green) is a DNA binding domain with preference for CG rich

sequences, in particular, a CCGCCC motif.[168] Studies suggest that this is due to the K152 directing binding towards to G^{270} which is considered in other ARID classes to be mutated to serine or threonine for non-specific or T DNA binding properties, respectively. *In vitro* studies demonstrated that KDM5A cannot operate in the absence of the ARID domain. It is thought that the binding of ARID domain guides KDM5A within chromatin to a specific DNA sequence containing the nucleosome complex and projects the JmjC-domain in proximity of the nearby H3 tails for transcriptional regulation.[168]

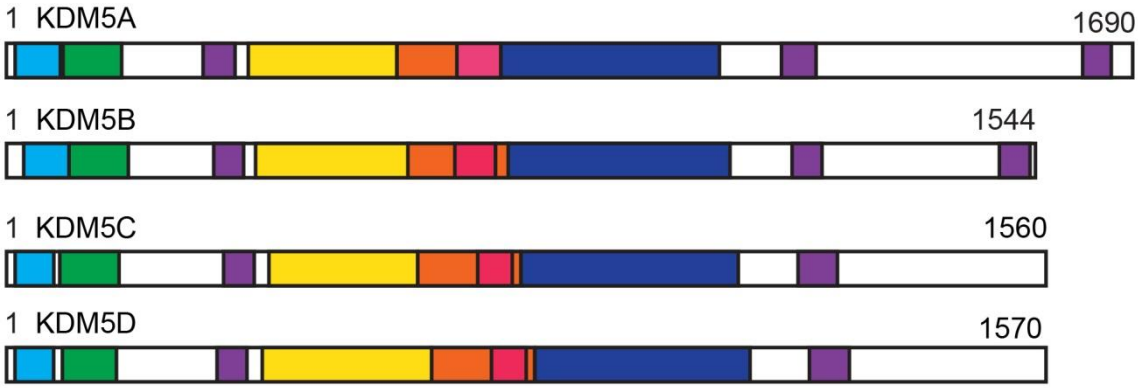


Figure 3.2.2: Schematic representation of domains and structural elements found in KDM5s. The domain boundaries are determined using BLAST sequence alignment and Phyre 2 structural prediction, described in section 3.5.11. Light blue: JmjN-domain, green: ARID-domain, purple: PHD, yellow: JmjC-domain, orange: alpha-helix, pink: C5HC2 zinc-finger and blue: PLU-domain. Residue numbers at the start and end of the proteins are denoted.

The exact sequence preference for the ARID-domain is ambiguous; Scibetta *et al.* reported a different sequence for ARID binding namely: GCACA/C but equally concluded a GC-rich specificity.[271] The JmjC-domain of KDM5A (KDM5A: L363-F586, Fig. 3.2.2, yellow) facilitates catalytic demethylation of substrates H3K4$me_{3\text{-}1}$ to yield lower methylation states in iterative steps. Note that H3K4me_1 demethylation is only observed with recombinant protein *in vitro*. A detailed description of the reaction mechanism is given in Chapter 1.3.4 but briefly considered here. In the presence of the co-factor 2OG, molecular oxygen, and substrate H3K4$me_{3\text{-}1}$, the catalytic Fe^{II} in the JmjC-active site is converted to the highly reactive Fe^{IV}=O species that hydroxylates N^{ε}-methyl lysine to form an unstable hemiaminal, which ultimately releases the demethylated substrate product, formaldehyde, succinate and carbon dioxide. The core of the KDM5A JmjC-domain of KDM5A consists of a double stranded β-helix (DSBH), bound to Fe^{II} metal, which is stabilized by two twisted β-sheets and a third antiparallel β-

sheet between JmjN-JmjC.[268] Coordination of iron is achieved by H483, E485 and H571. The co-factor 2OG binds in a bidentate manner to Fe^{II}. In addition to this, the JmjC-domain also contains an α-helix (KDM5A: F586-W742 Fig. 3.2.2, orange) and zinc binding domain C5HC2, of KDM5A (KDM5A: C676-W742, Fig. 3.2.2, pink) that are vital to the JmjC catalytic activity. The function of this PLU domain of KDM5A (KDM5A: V743-L1070), found to be a WYFC-rich region remains to date unknown.

KDM5s have three (KDM5A, KDM5B) or two (KDM5C, KDM5D) PHD-finger domains, with each PHD-finger numbered from N- to C-terminus in order (Fig, 3.2.2). PHD are characterised by their typical length of 50–80 amino acids sequence and binding preference to two zinc atoms using cysteine and histidine motifs. The structural confirmation is defined by two stranded anti-parallel β-sheets and frequently an α-helix position on the C-terminus. PHD fingers are a family of domains that have high affinity for various histone PTMs, including H3K4$me_{3,2}$, H3K4me_0 or H3K14Ac and may have sensing capabilities for H3R2 methylation.[272]

The PHD1 of KDM5A (V295-E343) (Fig. 3.2.2, purple) is believed to bind histone H3 tails, sensing the level of methylation on lysine 4, preferring lower methylation states.[166,273] PHD1 is considered to be part of an integral feedback loop to the JmjC-domain, promoting demethylase activity.

Limited information is currently available for the PHD2 domain of KDM5A (KDM5A: F1163-C1215). PHD3 of KDM5A (KDM5A: C1615-C1658) binds methylated lysine on H3K4 with greater affinity towards higher methylation states, while H3K9me_{3-0}, H3K27_{3-0}, H3K36_{3-0} and H3K$79_{3,2}$ were found to bind weakly. Crystal (PDB: 3GL6) and solution (PDB: 2KGI) structures revealed the typical anti-parallel β-sheet and α-helix motifs. The histone H3 N^{ε}-trimethyl-lysine substrate binds in an "aromatic cage" consisting of W1625 and W1635, and the histone H3 sequence forms an anti-parallel β-sheet extension with the PHD finger.[8] Frequently, the aromatic cage consists of three or four residues

that are responsible for recognising higher methylation levels; however, the aromatic cage architecture varies among different reader proteins. The aromatic cage typically consists out of hydrophobic and aromatic residues amino acids as for example: tryptophan, tyrosine, phenylalanine, methionine.[6] In KDM5A-PHD3, the H3K4me$_3$ is only bound to two tryptophan side chains providing cation-π interactions, hydrophobic effect and displacement of high energy water.[244] Furthermore, KDM5A-PHD3 has a secondary binding pocket where Q1627, D1629 and interacts with H3R2.[8]

3.2.3 Biological relevance of KDM5s as nuclear proteins

KDM5s play an important role in regulation but are also strongly linked to the proliferation and development of many cancers.[274] Recent publications report that translocation of the genes *NUP98* to *KDM5A* cause the formation of a fusion protein that is linked to Acute Megakaryoblastic Leukemia (AMKL), a form of blood cancer. In this translocation, only a part of the *KDM5A* gene (t11;12 p15;13) is translocated resulting in the fusion of the KDM5A PHD3 finger with NUP98, and thereby leading to the cellular expression of oncogenic NUP98-KDM5A$_{PHD3}$ fusion protein.[8] KDM5A is also linked to various cancers including gastric cancers, breast cancer and retinoblastoma.[9,10,275,276]

Kdm5a knockout in mice demonstrated no significant phenotypic changes although haematological abnormalities were observed.[189] On the contrary, KDM5A is found to be vital at several stages in development. The Kdm5a analogue of *Drosophila melanogaster* is known as little imaginal discs (LID). Homozygous mutations within LID cause high larvae lethality, with few surviving beyond the early pupal stage.[277] In humans, KDM5B is found to be strongly upregulated in breast cancer,[212] while KDM5C mutations have shown to be linked to mental-X retardation.[278,279]

3.3 Approaches to study the function KDM5A-PHD3

The involvement of the KDM5 family members in gene silencing as well as their link to various cancers make KDM5A an interesting target to study. To date however, little is known about the detailed molecular mechanism of their function, including the functional crosstalk of the KDM5A PHD3 finger to the JmjC-domain.

3.3.1 Small molecule probes for PHDs

To investigate the functional role of the KDM5A-PHD3 finger, a novel selective probe-like molecule is required that binds exclusively to the PHD3 and does not cross-react/bind to the KDM5A PHD1-finger or JmjC-domain. While small molecule probe inhibitors would be the reagent of choice, to date, limited inhibitors have been identified that bind N^{ε}-trimethyl-lysine PHD-fingers in general. The first KDM5A-PHD3 targeting inhibitors, amiodarone and derivatives, were reported as PHD-finger inhibitors.[280] However, we have confirmed that the weak binding affinity (high micro molar range of displacement) and non-specific mode of action (demonstrate pan-specificity between the PHD3 and JmjC-domains) of these small molecules make them ineffective as functional chemical probes (Bhushan, Erdmann, Zhang, Belle *et al. Bioorganic Med. Chem*. **26**, 2984–2991 (2018)).[281]

3.3.2 Histone peptide probes

Alternative probe-like molecules generated from histone H3 peptide fragments of different lengths have been considered. In KDM5s, it has been shown that the KDM5A JmjC-domain requires an 18-mer peptide ($H3_{1-18}$) for catalytic activity with decreasing activity for smaller peptide lengths.[166] On the other hand, chromatin readers such as PHD3 of KDM5A demonstrate a high binding affinity of 10-mer peptides or less,[231] suggesting short peptide fragments may be suitable as specific inhibitors of KDM5A-PHD3. A potential limitation to this approach is that short peptide fragments may still inhibit

the KDM5A JmjC-domain, but this was not observed previously. One other concern is the selectivity across other H3K4me$_3$ reader domains; however, variations in the "aromatic cage" of binding modules may provide some degree of selectivity.

Inspection of the crystal structure of KDM5A-PHD3 revealed that the PHD finger has a unique surface groove binding configuration between the *N*$^{\varepsilon}$-trimethyl-lysine peptide and KDM5A$_{PHD3}$ protein (PDB: 3GL6) (Fig. 3.3.1 cyan).[8] This could potentially allow for binding of an enlarged analogues of the *N*$^{\varepsilon}$ ammonium group of trimethyl-lysine forming a similarly strong complex to endogenous *N*$^{\varepsilon}$-trimethyl-lysine. Provided that the enlarged *N*$^{\varepsilon}$-lysine is selective for KDM5A-PHD3 and does not inhibit the JmjC-domain or bind to the PHD1 finger, it could serve as valuable tool for dissecting the functions of the PHD3.

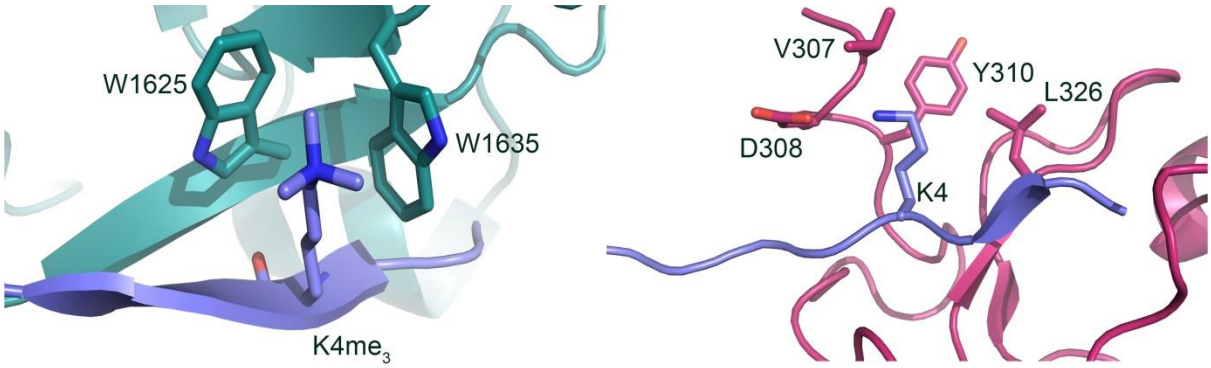


Figure 3.3.1: H3K4 (blue) binding site of crystal structure KDM5A-PHD3 (cyan) (PDB: 3GL6) and solution structure KDM5B-PHD1 (pink) (PDB: 2MNZ). Heteroatoms oxygen (red), nitrogen (blue) and sulphur (yellow) displayed are coloured.

To gain insight into whether targeting of the *N*$^{\varepsilon}$-trimethyl-lysine recognition pocket for PHD3 over the catalytic JmjC-domain is possible, it is important to assess the recognition / interactions that the histone peptides make with JmjC-domain. To date, there are no crystal structures of KDM5 JmjC-domain in complex with a peptide available, making interpretation of the active site difficult. The crystal structure of closely related histone lysine demethylase KDM4A in complex to H3K9me$_3$ peptide (PDB: 2OQ6) shows that an enlarged *N*$^{\varepsilon}$-lysine head could experience strong steric hindrance (Fig. 3.3.2). However, the lysine side chain may exhibit some freedom of movement potentially providing space for enlarged

N^{ε}-trimethyl-lysine.[282] Furthermore, JmjC-domains of KDM2A, KDM5E and KDM7B have been observed to have decreased or abolished demethylase activity with sterically enlarged N^{ε}-dimethyl-lysine analogues depending on their size, relative to canonical N^{ε}-dimethyl-lysine substrate.[283] N^{ε}-diethyl-lysine, an N^{ε}-dimethyl-lysine derivative, was not enzymatically modified by the KDMs tested; however, the KDM6B JmjC-domain has subsequently been shown to possess a higher level of flexibility and is able to de-ethylate and oxidize N^{ε}-diethyl-lysine.[284]

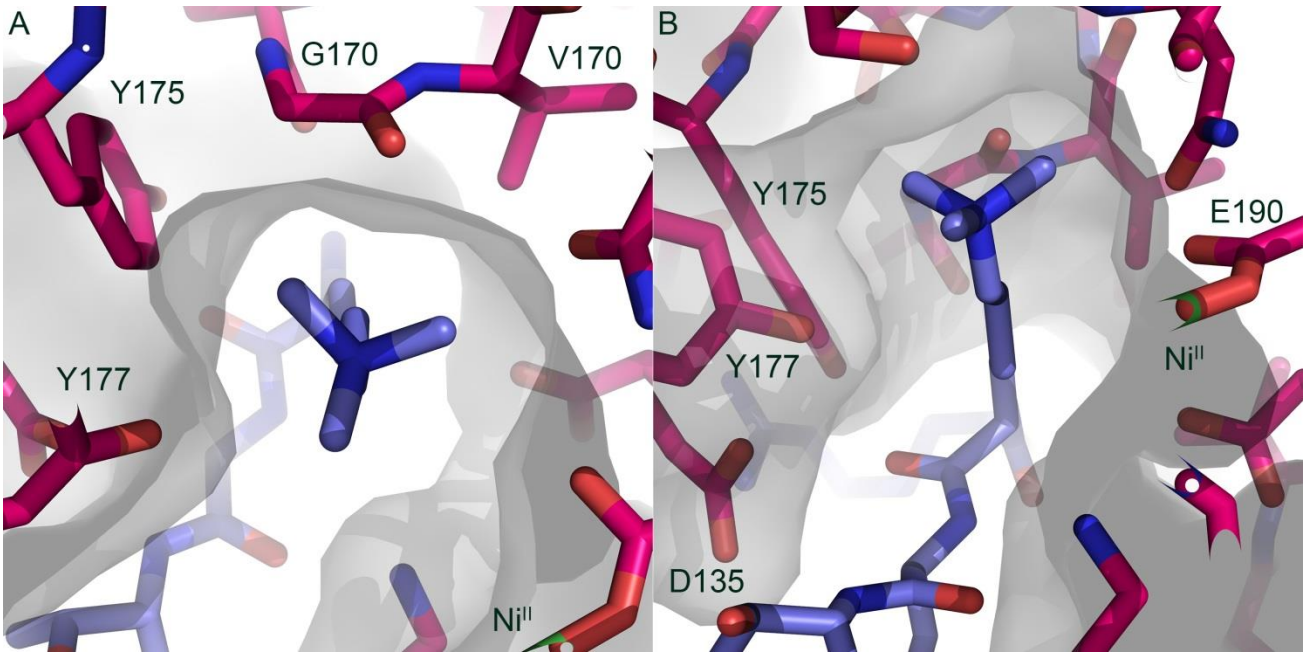


Figure 3.3.2: Crystal structure of KDM4A (pink) in complex to histone H3 peptide (light blue) displaying the N^{ε}-trimethyl-lysine in the active site from (A) top view and (B) side view (PDB: 2OQ6). Heteroatoms oxygen (red) and nitrogen (blue) displayed are coloured.

KDM5A-PHD3 and KDM5A-PHD1 recognize the same histone H3 sequence. The affinity for different methylation levels is the opposite between the two domains demonstrating different recognition architectures (Fig. 3.3.1). Studies of KDM5A-PHD1 show that un-methylated H3K4 binds the strongest whereas H3K4me$_3$ binds the weakest.[166,285] Larger N^{ε}-substituted-lysine derivatives have not been tested on KDM5A-PHD3.

3.4 Chapter objectives

The role of the PHD3-finger within KDM5A remains only partially understood. Key questions to be answered are: Why does KDM5A have three domains (PHD1, JmjC and PHD3) with different functions acting on the same histone H3 mark? What is the role of the PHD3 finger in relation to other domains? Does the PHD3-finger function as 'anchoring' site to nucleosomes or does the PHD3-finger work synergistically with the JmjC to enhance demethylation activity? In this Chapter, various histone H3 peptide variants were synthesized and systematically analysed against individual domains and multi-domain constructs of KDM5. Particular focus is on N^{ε}-triethyl-lysine ($H3_{1\text{-}21}K4et_3$ **3.1**) and N^{ε}-piperidine lysine ($H3_{1\text{-}21}K4pip$ **3.8**), which were designed and characterised as potential peptide probe-like molecules for KDM5A-PHD3 (Fig. 3.4.1).

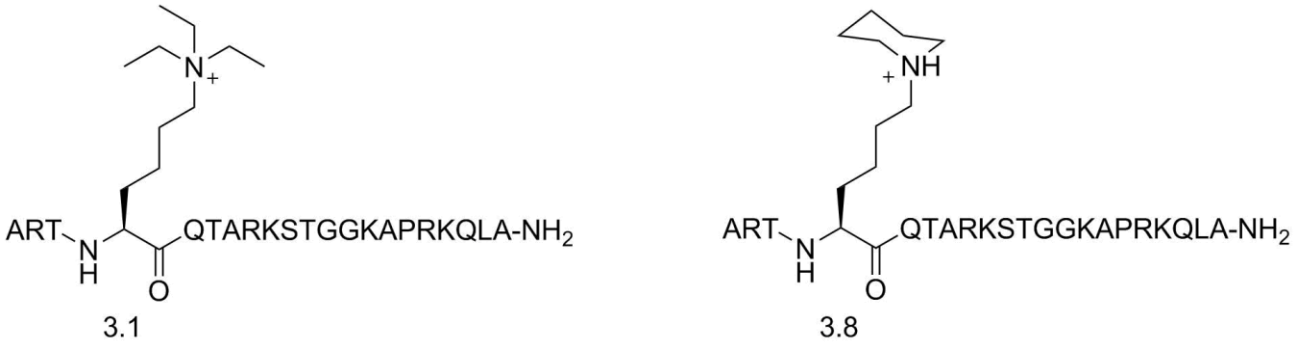


Figure 3.4.1: Schematic representation of probe-like histone mimic peptides targeting KDM5A-PHD3.

The aims of this Chapter are:

- To synthesize and characterise $H3_{1\text{-}21}K4et_3$ **3.1** and $H3_{1\text{-}21}K4pip$ **3.8** peptides as potential mechanistic inhibitors / chemical tools to study KDM5A-PHD3.
- To determine the binding selectivity, if any, of $H3_{1\text{-}21}K4et_3$ **3.1** and $H3_{1\text{-}21}K4pip$ **3.8** towards KDM5A-PHD3 over KDM5A-PHD1.
- To determine whether $H3_{1\text{-}21}K4et_3$ **3.1** and $H3_{1\text{-}21}K4pip$ **3.8** are substrates for the JmjC-domain and if the JmjC-domain natural substrate demethylation is inhibited by these peptides.

- To investigate whether the PHD3-finger alters the demethylation activity of KDM5A-JmjC by potential crosstalk between the PHD3-finger and the JmjC-domain.
- To explore the level of inhibition the demethylation activity of KDM5A-JmjC by KDM5A-PHD3.
- To carry out bioinformatics analysis of the KDM5 domain architecture and to investigate the potential function of the KDM5A-PHD2 finger.

References made to the domain are described in capitals (e.g. PHD3) while a reference made to the protein construct are made in subscript (e.g. KDM5A$_{PHD3}$).

3.5 Results

3.5.1 Synthesis of functionalized N^{ε}-trimethyl-lysine and preparation of H3 mimic peptides

To prepare histone H3 N^{ε}-lysine modified peptides $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** using Fmoc chemistry in solid phase peptide synthesis (SPPS), precursor Fmoc-Lys(pip·HCl)-OH **3.10** and Fmoc-Lys(et_3·Cl)-OH **3.16** were synthesised. The target molecule synthesis was commenced from the commercially available Fmoc-Lys(Boc)-OH **3.9**. Boc deprotection of Fmoc-Lys(Boc)-OH **3.9** was carried out using neat formic acid, followed by the removal of acid *in vacuo* (Fig. 3.5.1).

The solid product (838 mg, 2.02 mmol) was further used to perform reductive alkylation using glutaraldehyde and sodium cyanoborohydride ($NaBH_3CN$). After R_f-C_{18} purification and lyophilisation, the product was obtained in decent yield (52%) and protonated using hydrochloride in diethyl ether ($HCl{\cdot}Et_2O$) in methylene chloride. This yielded Fmoc-Lys(pip·HCl)-OH **3.10**, which was subsequently used for SPPS. The synthesis of Fmoc-Lys(et_3·Cl)-OH **3.16** is based on previous reported procedures to generate Boc-Lys(H)-O^{t}Bu **3.13** from Boc-Lys(Cbz)-OH **3.11**.[286] Tertbutyl ester protection of Boc-Lys(Cbz)-OH (13.6 mmol) **3.11** was carried out using tert-butanol and di-tertbutyl dicarbonate (Boc_2O) and 4-(dimethylamino)-pyridine (DMAP) to obtain Boc-Lys(Cbz)-O^{t}Bu **3.12** in good yield (11.6 mmol, 85%). The Boc-Lys(Cbz)O^{t}Bu **3.12** (5.43 mmol) group was efficiently deprotected by hydrogenation using palladium on activated charcoal (Pd/C) and H_2 (g) to give Boc-Lys(H)-O^{t}Bu **3.13** in excellent yield (4.86 mmol, 90%). In the next step, Boc-Lys(H)-O^{t}Bu **3.13** (3.0 mmol) was reductively alkylated using $NaBH_3CN$ and excess acetaldehyde. After an extractive workup step and concentration *in vacuo*, the resultant N^{ε}-diethyl-lysine **3.14** residue was ethylated using ethyl iodide under reflux conditions in acetonitrile. The crude solid was purified using [SiO_2] furnishing Boc-Lys(et_3·I)-O^{t}Bu **3.15** (1.13 mmol, 46% over two steps). In two subsequent steps, the

Boc and terbutyl ester protecting groups were removed using trifluoroacetic acid (TFA) followed by N^{α}-lysine protection in the presence of Fmoc-OSu and sodium bicarbonate in dioxane / water. The lysine derivative was purified using R_f-C_{18} column purification and isolated by freeze drying to provide the lyophilized product as white solid (1.13 mmol, 81%). This was then protonated using $HCl{\cdot}Et_2O$, resulting in Fmoc-Lys($et_3{\cdot}Cl$)-OH **3.16** suitable for use in SPPS.

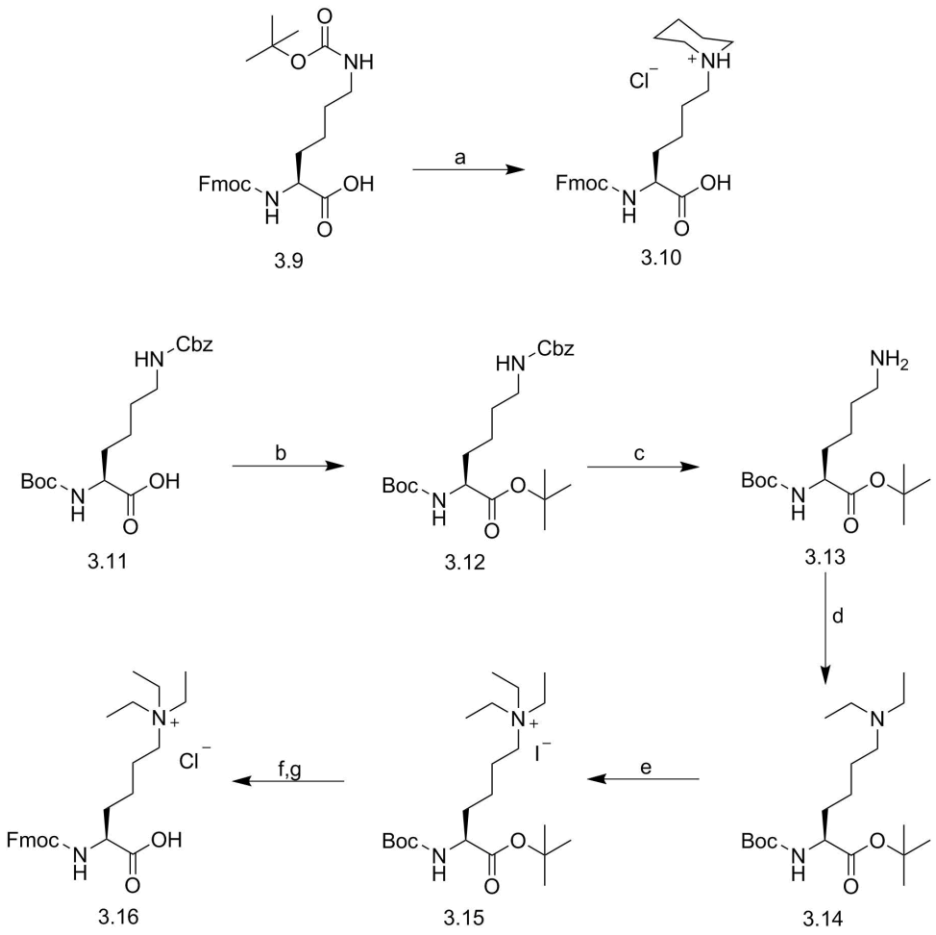


Figure 3.5.1: Synthesis of Fmoc-Lys(pip·HCl)-OH 3.10 and Fmoc-Lys(et_3·Cl)-OH 3.16. a) formic acid, 16 hrs then glutaraldehyde (8.0 equiv.), $NaBH_3CN$ (2.1 equiv.), 3 hrs at 0 °C, 16 hrs at RT, R_f-C_{18} purification, $HCl{\cdot}Et_2O$, 52%; b) Boc_2O (1.5 equiv.), DMAP (0.2 equiv.), ^{t}BuOH, purification [SiO_2], 2.5 hrs, 85%; c) Pd / C (2.1 mol%), $H_{2\,(g)}$ in EtOH, 1.5 hrs, purification [SiO_2], 90%; d) acetaldehyde (5.9 equiv.), $NaBH_3CN$ (2.2 equiv.), EtOH, 0.67 hrs; e) iodoethane (2.0 equiv.), acetonitrile, 80 °C, 1 hrs, purification [SiO_2], 46%; f) TFA / CH_2Cl_2, 32 hrs; g) Fmoc-OSu (1.2 equiv.), $NaHCO_3$ (8.4 equiv.), milliQ / dioxane, 2 hrs, R_f-C_{18} purification, 81%.

The unnatural amino acid was incorporated into the peptide using microwave assisted (μW) SPPS, a method that enables fast coupling of amino acids and deprotection of the terminal Fmoc-group. Manufacturers recommended conditions were used with Fmoc-protected amino acids with, when

applicable, acid labile side chain protecting groups. Solid phase Rink amide resin was used as a solid support to provide peptides containing C-terminal amides. Diisopropyl carbodiimide (DIC), OxymaPure® and diisopropylethylamine (DIEA) in DMF were used as coupling partners, while Fmoc deprotection was achieved using piperidine (20 v/v%) in DMF. Synthesised peptides were cleaved from the resin and deprotected using a solution of TFA containing scavengers [H_2O, triisoproyl silane (TIPS) and ethane-1,2-dithiol (EDT)] and HPLC purified using an R_f-C_{18} column. Of the lyophilized peptide was the purity confirmed using HPLC and peptides for assays were standardized using ^{1}H NMR with sodium 3-(trimethylsilyl)-2,2,3,3-d_4 propionate (TSP) in D_2O as internal standard. $H3_{1\text{-}21}K4me_{3\text{-}0}$ **3.2–3.5**, $H3_{1\text{-}21}K4Ac$ **3.6**, $H3_{1\text{-}21}K4A$ **3.7** peptides were synthesised using the same procedure as control peptides. Standardized peptide solutions (concentration determined by NMR) were used for all assays except isothermal titration calorimetry (ITC) analysis as this required buffer specific solutions to minimize heat of dilution.

3.5.2 Protein production of $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$

To carry out biophysical and biochemical characterisation of binding affinities of $H3_{1\text{-}21}K4et_3$ **3.1** and $H3_{1\text{-}21}K4pip$ **3.8,** the $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$ reader proteins were produced through recombinant expression. KDM5 constructs were kindly gifted by the Structural Genomics Consortium (SGC) (Oxford, UK). Construct information of proteins used in this Chapter is given below (Fig. 3.5.2).

In brief, recombinant $KDM5A_{PHD3}$ (A1542–K1660, Fig. 3.5.2) and $KDM5A_{PHD1}$ (K283–Q365, Fig. 3.5.2) were produced as a single domain and expressed in pGTVL2 vector in *E.coli*. This produced N-terminal 6×His-tag and GST-tag labelled proteins. $KDM5A_{PHD3}$ was produced according to previous reported procedure with minor alterations.[281] Phage resistant *E. coli* cells (pRARE) transformed with BL21 (DE3) *KDM5A-PHD1/3* in pGTVL2 were grown in 4 litres of broth (supplemented with

kanamycin), and expression induced to yield 89.3 g ($KDM5A_{PHD3}$) or 74.5 g ($KDM5A_{PHD1}$) of cell pellet. The pellet was split and 20.9 g ($KDM5A_{PHD3}$) or 22.0 g ($KDM5A_{PHD1}$) of pellet was used for protein purification. The cells were lysed by sonication, and His-tagged proteins were isolated using nickel affinity purification followed by gel filtration (Superdex75, 300 mL). Selected fractions were pooled based on their purity ($KDM5A_{PHD3}$: B1–B3 C1–C4, $KDM5A_{PHD1}$: C2–C7) and concentrated to yield from these purifications 123.8 mg (22.3 $mg \cdot mL^{-1}$) and 94.7 mg (18.43 $mg \cdot mL^{-1}$) of protein for $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$, respectively. SDS-PAGE (Fig. 3.5.3 A/B) indicated the correct mass and high purity of overexpressed proteins (>95%). Protein mass spectrometry analysis (LC-MS) revealed the correct masses for $KDM5A_{PHD3}$ (Fig. 3.5.3 C,E) and $KDM5A_{PHD1}$ (Fig. 3.5.3 D,F).

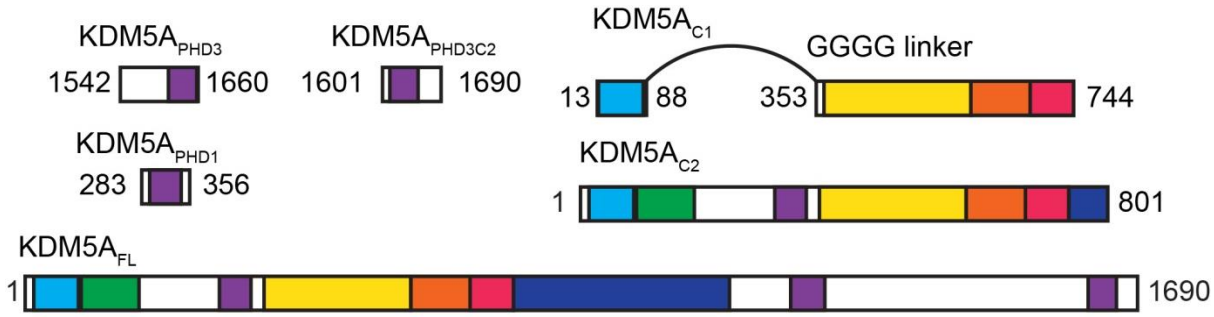


Figure 3.5.2: Schematic representation of KDM5A protein length and domains used in this Chapter. Purple: PHD-domain, light blue: JmjN-domain, green: ARID-domain, yellow: JmjC-domain, orange: α-helix, pink: C5HC2-zinc finger, Blue: PLU-domain.

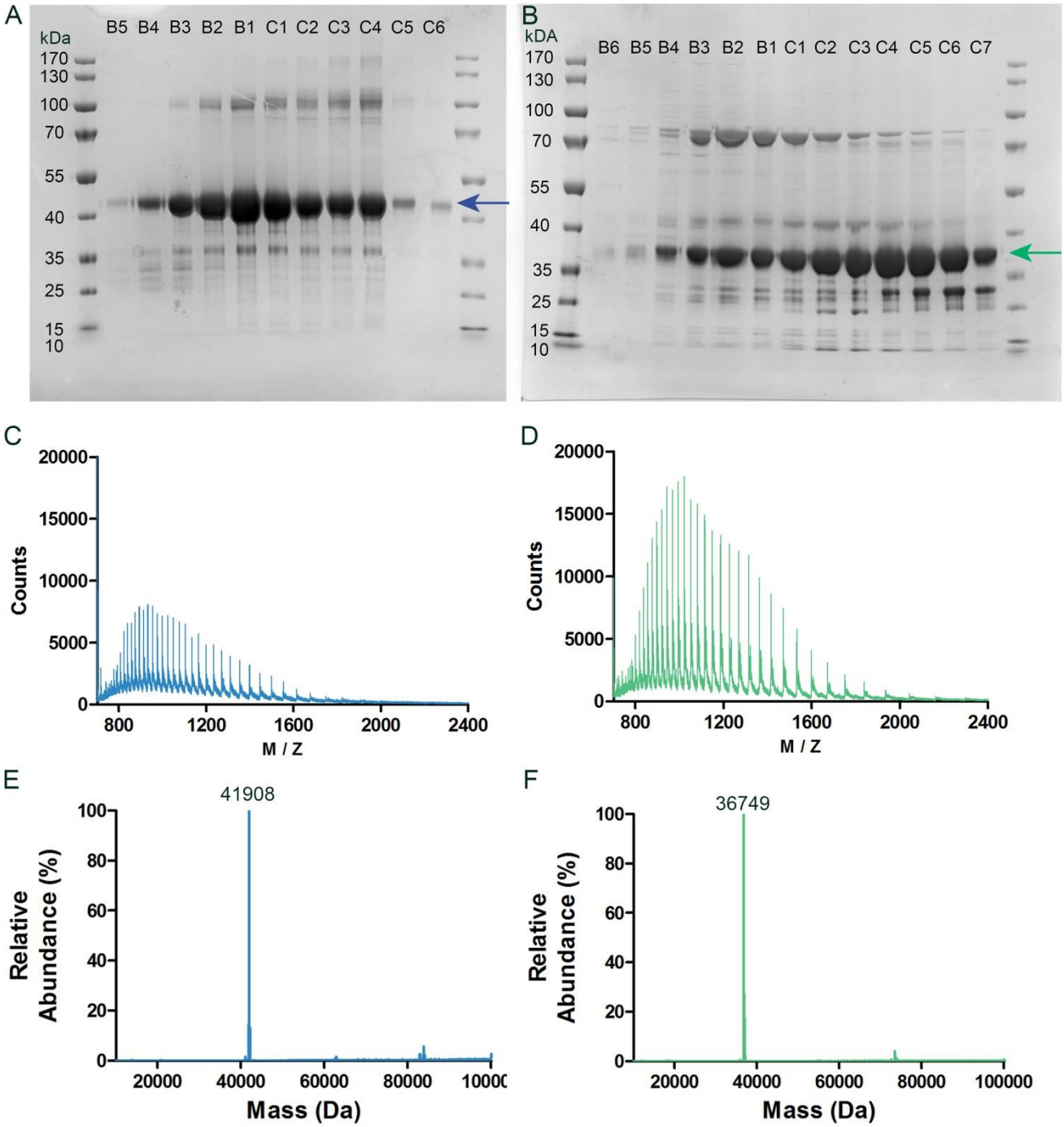


Figure 3.5.3: SDS-PAGE (A,B), raw LC-MS data (C,D) and deconvoluted LC-MS data (E,F) of recombinant KDM5A$_{PHD3}$ (blue, A,C,E) and KDM5A$_{PHD1}$ (green, B,D,F). A) SDS-PAGE gel (gradient 4–12%) of KDM5A$_{PHD3}$ (Calc. mass: 42 kDa), pooled: B3-B1 and C1-C4, concentrated: 123.8 mg (22.26 mg·mL^{-1}); B) SDS-PAGE gel (gradient 4–12%) KDM5A$_{PHD1}$ (Calc. mass: 37 kDa), pooled: C2-C7, concentrated: 94.7 mg (18.43 mg·mL^{-1}); C) Raw LC-MS (RF-MS) data of KDM5A$_{PHD3}$; D) Raw LC-MS (RF-MS) data of KDM5A$_{PHD1}$; E) Deconvoluted LC-MS (RF-MS) data of KDM5A$_{PHD3}$ (Calc. mass: 41908, Observed: 41908); F) Deconvoluted LC-MS (RF-MS) data of KDM5A$_{PHD1}$ (Calc. mass: 36748, Observed: 36749).

3.5.3 Binding studies of $H3_{1-21}K4et_3$ 3.1 and $H3_{1-21}K4pip$ 3.8 against $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$ using ITC

To investigate the binding affinity of $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** against $KDM5A_{PHD3}$, ITC measurements were conducted. ITC is a used to study biomolecular interactions and is considered the 'gold-standard' in determining binding affinity.[287] It provides high precision measurement of the strength of binding (K_d, μM), as well as the changes in Gibbs free energy (ΔG, kcal·mol^{-1}), enthalpy (ΔH, kcal·mol^{-1}) and entropy (ΔS, kcal·mol^{-1}·K^{-1}) of binding, allowing for the detailed thermodynamic characterization of the molecular interactions observed. ITC is based on the detection of the release or absorption of energy in the form of heat upon the interaction of molecules or in the case of a (catalytic) reaction. The ITC instrument contains two cells: a reference cell and a sample cell, that are attached but separated by individual heating elements and temperature sensors. The cells are situated in a slightly colder environment than the pre-set temperature, thereby cooling the cells. However, by directing power towards the heating elements, a constant (pre-set) temperature is established. By titrating titrant to the sample containing sample cell, a favourable interaction or reaction can occur that releases energy in the form of heat, thus altering the temperature of the sample cell. As the sample cell temperature is altered, the ITC instrument changes the energy input to the sample cell relative to the reference cell to maintain an equal temperature. This difference in power input is recorded (μcal·sec^{-1}) and used to determine binding strength or reaction kinetics. The binding affinity (K_a, μM^{-1}) is determined from the steepest slope of the titration curve and the initial heat relates to the enthalpy (ΔH, kcal·mol^{-1}).

$$K_d = \frac{1}{K_a} \qquad \text{Equation 3.1}$$

$$\Delta G = RT \ln(K_d) \qquad \text{Equation 3.2}$$

$$\Delta G = \Delta H - T\Delta S \qquad \text{Equation 3.3}$$

$$-T\Delta S = \Delta G - \Delta H \qquad \text{Equation 3.4}$$

Equation 3.1 gives the K_d value that is also used to determine the ΔG in equation 3.2 at a known cell temperature. With known ΔG and ΔH values, rearranging of equation 3.3 to equation 3.4 provides $-T\Delta S$.

The key advantages of ITC binding assays over other biophysical methods include label free analysis, a broad range of compatible aqueous buffers, real-time experiment monitoring and the option to recover the samples after analysis. The disadvantages of ITC are the relatively slow measurements in comparison to other fluorescence or luminescence-based assays (*e.g.* AlphaScreen™ displacement or BioLayer Interferometry (BLI)), interference by buffer mismatch and the requirement for relatively high reagent volumes and concentrations. Additionally, values detected for potentially observed binding affinities between protein-protein or protein-ligand using ITC do not reveal the confirmation of binding. In other words, the instrument provides is informative with respect to the binding affinity but does not elucidate which interactions contribute to this. Other techniques could be used to supplement the ITC results, such as NMR or crystallography.

In this Chapter, binding data for KDM5A$_{PHD1}$ or KDM5A$_{PHD3}$ were obtained using automated MicroCal PEAQ-ITC (Malvern), with reader proteins in the sample cell and histone H3 tail mimic peptides titrated in. The protein and peptide buffer [HEPES (50 mM), NaCl (50 mM) in MQ (pH 7.5, filtrated and degassed)] solutions were matched by dissolving solid peptides in buffers used for the KDM5A$_{PHD1}$ and KDM5A$_{PHD3}$ gel-filtration purification step. ITC experiments were conducted using manufacturer's program 'Plates-prerinse syringe clean' providing the most extensive cleaning steps between each experiment. The cell temperature was set to 25.0 °C with tray temperatures at 15.0 °C to enhance protein and peptide stability. As validation of the data obtained here and to compare it to reported values, peptides H3$_{1-21}$K4me$_3$ **3.2**, H3$_{1-21}$K4me$_2$ **3.3**, H3$_{1-21}$K4me$_1$ **3.4**, H3$_{1-21}$K4me$_0$ **3.5** were prepared for analysis.[8,166,231] Furthermore, to verify selectivity of the reader domain, H3$_{1-21}$K4Ac **3.6** and H3$_{1-21}$K4A **3.7** were made. ITC experiments were carried out for H3$_{1-21}$K4 **3.1**–**3.8** using

KDM5A$_{PHD3}$ and KDM5A$_{PHD1}$ proteins. ITC recordings (n≥2, displayed in Table 3.5.1, Fig. 3.5.4), revealed that the binding affinity of KDM5A$_{PHD3}$ is the strongest for *N*$^{\varepsilon}$-trimethyl-lysine H3$_{1-21}$K4me$_3$ **3.2** (K_d: 0.117 ± 0.022 μM) and lower methylation states showed decreasing binding affinities for H3$_{1-21}$K4me$_2$ **3.3** (K_d: 0.415 ± 0.035 μM), H3$_{1-21}$K4me$_1$ **3.4** (K_d: 1.58 ± 0.04 μM) and H3$_{1-21}$K4me$_0$ **3.5** (K_d: 6.21 ± 0.96 μM).

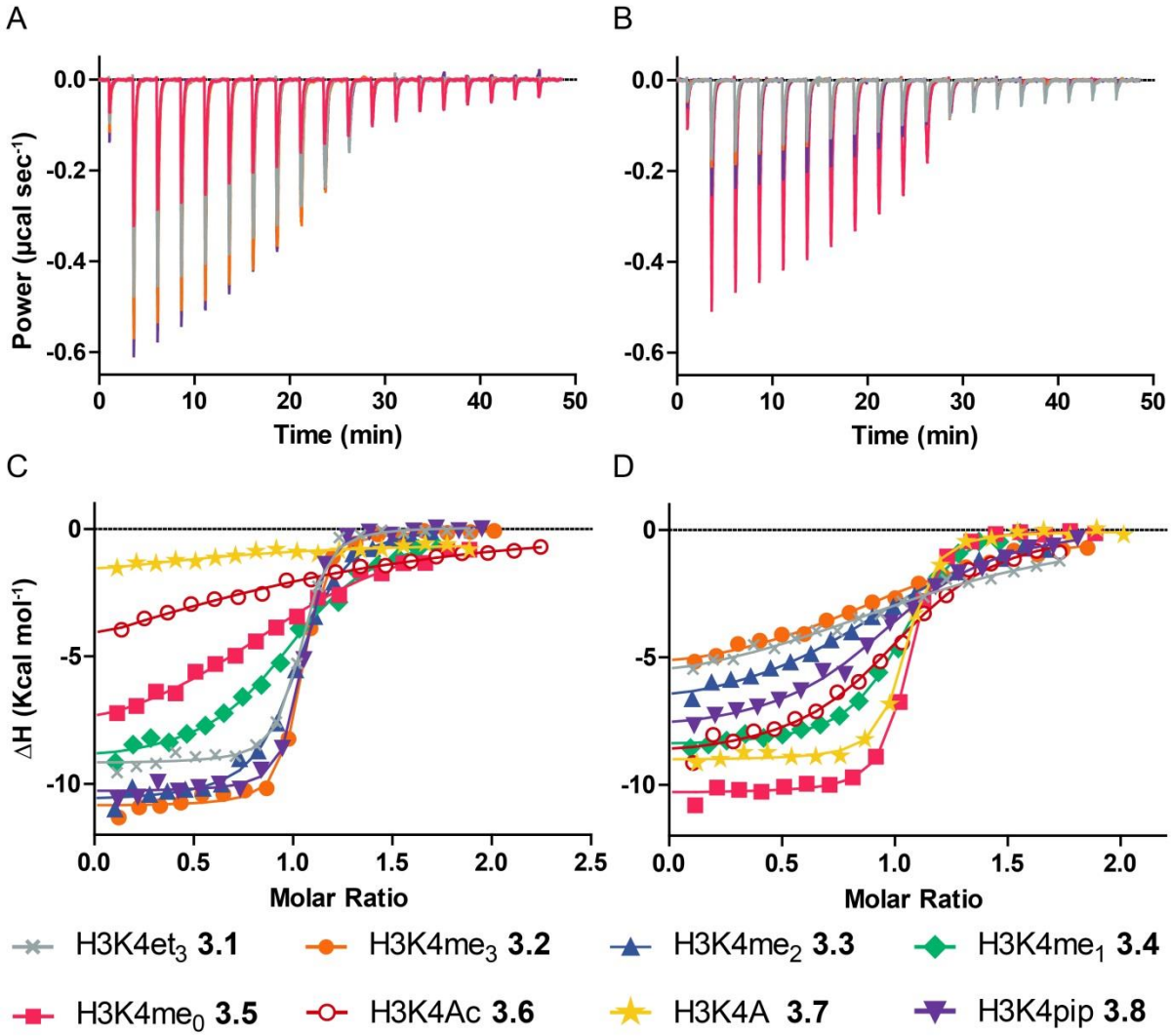


Figure 3.5.4: ITC binding of H3$_{1-21}$K4 peptides 3.1–3.8 binding to KDM5A$_{PHD3}$ (A, C) or KDM5A$_{PHD1}$ (B, D) shown as raw data (peptide H3$_{1-21}$K4 3.1, 3.2, 3.5, 3.8) (A, B) and peak integration (peptide H3$_{1-21}$K4 3.1–3.8) (C, D).

This demonstrates that the KDM5A$_{PHD3}$ is correctly folded, and the binding affinity are in agreement with reported values (K_d: 0.75, 0.93, 2.79 and 20.7 μM for H3K4me$_3$, H3K4me$_2$, H3K4me$_1$ and H3K4me$_0$, respectively).[8] While a similar trend is observed, subtle differences in the K_d values are found compared to Wang *et al.* This could be due to factors such as buffer conditions (*i.e.* buffer, salt

concentration and pH), differences in the H3 peptide length and protein construct, protein / peptide purity or data analysis.

Interestingly, peptides $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** bind $KDM5A_{PHD3}$ with K_d: 0.101 ± 0.012 μM and K_d: 0.086 ± 0.011 μM, respectively. This is stronger but of relatively similar strength to $H3_{1-21}K4me_3$ **3.2**. The enthalpy for $H3_{1-21}K4et_3$ **3.1** decreased (ΔΔH: 1.62 kcal·mol^{-1}) however, this is (over) compensated by entropy (Δ-TΔS: −1.69 kcal·mol^{-1}). $H3_{1-21}K4pip$ **3.8** depicts a similar trend with decreased entropy (ΔΔH: 0.50 kcal·mol^{-1}) and compensation by entropy (Δ-TΔS: −0.70 kcal·mol^{-1}). To validate that charge on N^{ε}-lysine and the lysine 4 side chain are of importance to strong binding, $H3_{1-21}K4Ac$ **3.6** and $H3_{1-21}K4A$ **3.7** were examined. Weak binding (K_d: >10 μM) was observed for $H3_{1-21}K4Ac$ **3.6** and $H3_{1-21}K4A$ **3.7** under the conditions tested. $H3_{1-21}K4A$ **3.7** is at least 100-fold weaker in binding than $H3_{1-21}K4me_3$ **3.2**. This suggests that the lysine 4 side chain maybe of importance to a strong histone H3 – reader complex. The ITC measurements revealed that $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** contain comparable binding affinities to $H3_{1-21}K4me_3$ **3.2** for $KDM5A_{PHD3}$ and significantly different than $H3_{1-21}K4me_0$ **3.5**.

Table 3.5.1: ITC data of $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$ against $H3_{1-21}K4$ peptides 3.1–3.8. n≥2, error given as s.d.

		$KDM5A_{PHD3}$				$KDM5A_{PHD1}$			
		K_d μM	ΔG kcal mol^{-1}	ΔH kcal mol^{-1}	-TΔS kcal mol^{-1}	K_d μM	ΔG kcal mol^{-1}	ΔH kcal mol^{-1}	-TΔS kcal mol^{-1}
$H3_{1-21}K4et_3$	**3.1**	0.101 ± 0.012	-9.55 ± 0.07	-9.18 ± 0.09	-0.37 ± 0.16	6.86 ± 1.00	-7.06 ± 0.08	-6.94 ± 0.32	-0.12 ± 0.39
$H3_{1-21}K4me_3$	**3.2**	0.117 ± 0.022	-9.47 ± 0.11	-10.80 ± 0.39	1.32 ± 0.47	8.37 ± 0.70	-6.93 ± 0.05	-7.31 ± 0.047	0.37 ± 0.51
$H3_{1-21}K4me_2$	**3.3**	0.415 ± 0.035	-8.71 ± 0.05	-10.90 ± 0.10	2.20 ± 0.15	4.03 ± 0.04	-7.36 ± 0.01	-7.66 ± 0.08	0.29 ± 0.08
$H3_{1-21}K4me_1$	**3.4**	1.58 ± 0.04	-7.92 ± 0.01	-9.49 ± 0.20	1.57 ± 0.21	0.469 ± 0.047	-8.64 ± 0.06	-8.63 ± 0.25	-0.01 ± 0.31
$H3_{1-21}K4me_0$	**3.5**	6.21 ± 0.96	-7.12 ± 0.10	-9.37 ± 0.57	2.27 ± 0.65	0.086 ± 0.012	-9.64 ± 0.09	-10.04 ± 0.53	0.40 ± 0.59
$H3_{1-21}K4Ac$	**3.6**	31.4 ± 20.3	-6.31 ± 0.46	-9.16 ± 4.44	2.86 ± 4.90	1.07 ± 0.24	-8.16 ± 0.14	-9.18 ± 0.02	1.01 ± 0.16
$H3_{1-21}K4A$	**3.7**	Not Determined				0.098 ± 0.004	-9.56 ± 0.021	-8.93 ± 0.04	-0.63 ± 0.05
$H3_{1-21}K4_{pip}$	**3.8**	0.086 ± 0.011	-9.65 ± 0.08	-10.30 ± 0.10	0.62 ± 0.04	2.33 ± 0.06	-7.69 ± 0.02	-5.86 ± 0.09	-1.83 ± 0.10

Next, to understand the binding affinities of $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** to $KDM5A_{PHD1}$, similar ITC experiments with peptides $H3_{1-21}K4$ **3.1–3.8** were carried out. The collected data (Table 3.5.1) confirm previous reports that $KDM5A_{PHD1}$ binds with higher preference to lower methylation states.[8,166] $H3_{1-21}K4me_3$ **3.2** was found to bind with a K_d of 8.37 ± 0.70 μM while $H3_{1-21}K4me_0$ **3.5**

was found to have a K_d of 0.086 ± 0.012 μM; a difference of approximately 100-fold in binding strength. Peptides $H3_{1-21}K4$ **3.2–3.5** showed that a decreasing methylation state correlates to an increase in enthalpy (maximum ΔΔH of 2.73 kcal·mol^{-1}) while the entropy (maximum Δ−TΔS of 0.41 kcal·mol^{-1}) remained relatively similar. Modified N^{ε}-trimethyl-lysine derivatives $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** were tested capturing a similar trend as compared to $H3_{1-21}K4me_3$ **3.2** obtaining low binding values of K_d: 6.86 ± 1.00 μM and K_d: 2.33 ± 0.06 μM, respectively. $H3_{1-21}K4pip$ **3.8** was demonstrated to have relatively low enthalpy (ΔΔH: 1.45 kcal·mol^{-1}) but strong entropic compensation (Δ-TΔS: −2.20 kcal·mol^{-1}) in comparison to $H3_{1-21}K4me_3$ **3.2**. Furthermore, $H3_{1-21}K4Ac$ **3.6** (K_d: 1.07 ± 0.24 μM) was seen to bind approximately 12-fold weaker than $H3_{1-21}K4me_0$ **3.5** but surprisingly, $H3_{1-21}K4A$ **3.7** (K_d: 0.098 ± 0.007 μM) bound of analogues strength as to $H3_{1-21}K4me_0$ **3.5**. Although the $H3_{1-21}K4me_0$ **3.5** and $H3_{1-21}K4A$ **3.7** contain comparative binding affinities, the enthalpy (ΔΔH: 1.07 kcal·mol^{-1}) and entropy (Δ-TΔS: −1.03 kcal·mol^{-1}) values are different.

In summary, the ITC data on the reader domains $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$ reveal that $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** bind with similar affinities to cognate histone $H3_{1-21}K4me_3$ **3.2**.

3.5.4 AlphaScreen™ displacement of $H3_{1-21}K4et_3$ 3.1 and $H3_{1-21}K4pip$ 3.8 against $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$

Next, AlphaScreen™ (Perkin Elmer) functional peptide displacement assays were optimised for KDM5A-PHD-fingers. This universal proximity-based luminescence assay method can be used for measuring binding events, and can also be adapted to measure catalytic turnover. A detailed description of AlphaScreen™ is given in Chapter 2.4. Briefly, in functional peptide displacement assay configuration, Streptavidin-coated donor and Ni-chelate-coated acceptor beads are used. C-terminal biotinylated peptide ($H3_{1-21}K4me_3$-[Biotin]) binds to the streptavidin coated donor bead. When His-

tagged KDM5A$_{PHD3}$, which associates with Ni-chelate acceptor beads, binds to the peptide H3$_{1-21}$K4me$_3$-[Biotin], which binds to streptavidin-functionalised donor beads, the donor and acceptor beads come into close proximity, and energy transfer can occur upon irradiation at 680 nm. By introducing competitive peptides to the H3$_{1-21}$K4me$_3$-[Biotin] peptide that disrupt its interaction with the KDM5A$_{PHD3}$ finger, the donor-acceptor beads pair are dissociated and diminishing the transfer of singlet molecular oxygen ultimately leading to signal reduction. Note that each bead may contain multiple linking sites and could therefore potentially link to multiple beads and consequently, the bead interaction measured is not necessary 1:1 stoichiometry between H3$_{1-21}$K4me$_3$-[Biotin] and the displacement peptide H3$_{1-21}$K4me$_3$ **3.2**. For this reason, the IC$_{50}$ values obtained are relative and functions as measure of 'rank order' analysis. AlphaScreen™ assay for KDM5A$_{PHD1}$ was carried out as described for KDM5A$_{PHD3}$ using H3$_{1-21}$K4me$_0$-[Biotin] as displacement peptide. To prevent non-specific binding, the buffer used for AlphaScreen™ measurements was supplemented with Tween20® (0.01%) and Bovine Serum Albumin (BSA, 0.1%).

To determine the linear region and best signal / noise, an AlphaScreen™ peptide displacement optimisation assay was carried out by preparing a dilution series of H3$_{1-21}$K4me$_3$-[Biotin] (25 nM) and KDM5A$_{PHD3}$ (25 nM) generating an 8×8 2-dimensional matrix. Signal increase was observed with higher concentrations of both peptide and protein indicating binding (Fig. 3.5.5 A). The optimal concentrations (highest signal / the linear range) were determined as follows: H3$_{1-21}$K4me$_3$-[Biotin] peptide (3.125 nM, R^2: 0.996) and KDM5A$_{PHD3}$ (1.563 nM, R^2: 0.999) (Fig. 3.5.5 B). KDM5A$_{PHD1}$ displacement protocol was adapted from KDM5A$_{PHD3}$ finding optimal conditions for KDM5A$_{PHD1}$ (1.563 nM, R^2: 0.999) and H3$_{1-21}$K4me$_0$-[biotin] peptide (3.125 nM, R^2: 0.989) (Fig. 3.5.5 B).

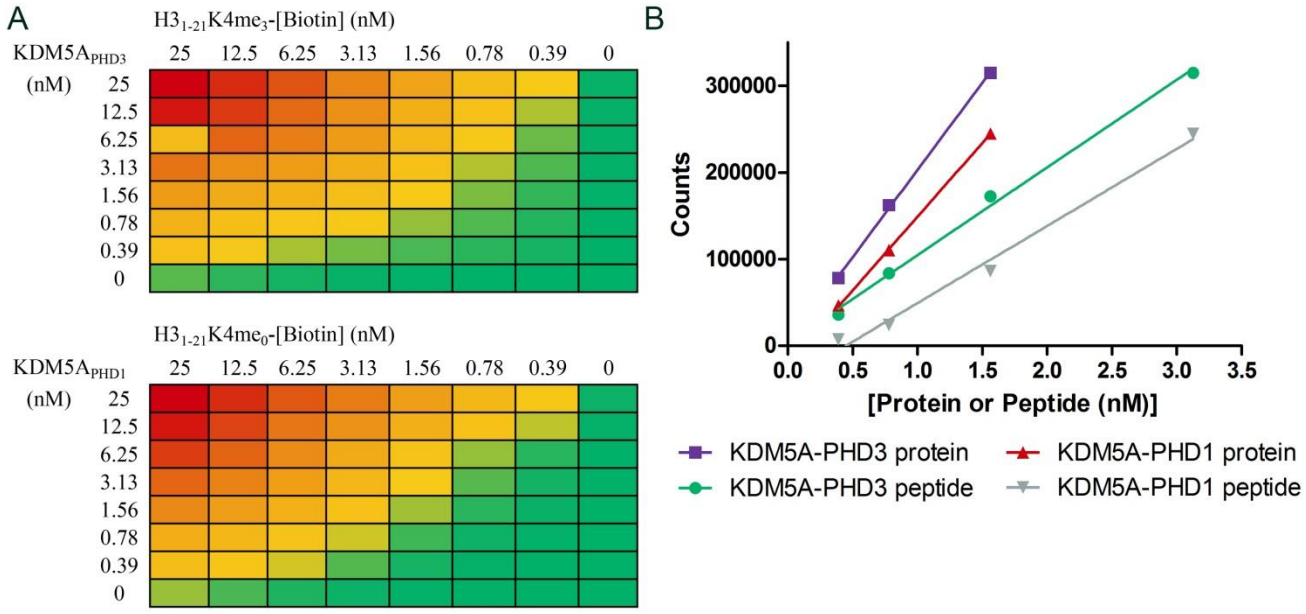


Figure 3.5.5: Linearity determination of AlphaScreen™ with KDM5A$_{PHD3}$ and KDM5A$_{PHD1}$ against peptide H3$_{1-21}$K4me$_3$-[Biotin] and H3$_{1-21}$K4-[Biotin], respectively. A) Plate setup and signal detected. Green: <10^4, red: >10^6. B) Linear regions for KDM5A$_{PHD3}$ (1.563 nM, R^2: 0.999), KDM5A$_{PHD1}$ (1.563 nM, R^2: 0.999) and the corresponding peptides H3$_{1-21}$K4me$_3$-[Biotin] (3.125 nM, R^2: 0.996), H3$_{1-21}$K4me$_0$-[Biotin] (1.563 nM, R^2: 0.989).

With established experimental conditions, AlphaPlates™ were prepared using a Cyte Echo dispenser from source peptide stocks H3$_{1-21}$K4 **3.1**–**3.8** (10 mM in MilliQ), creating a top concentration of 100 µM and a two-fold dilution series, and peptide displacement assays were carried out. AlphaScreen™ results are summarized in figure 3.5.6 and Table 3.5.2.

For KDM5A$_{PHD3}$, H3$_{1-21}$K4me$_3$ **3.2** displaces the H3$_{1-21}$K4me$_3$-[Biotin] peptide at IC$_{50}$: 0.554 ± 0.105 µM in agreement with previous reported observations (0.22 µM).[281] The IC$_{50}$ values obtained with AlphaScreen™ demonstrate a similar trend observed using ITC; the higher the *N*$^{\varepsilon}$-lysine 4 methylation state, the greater the level of displacement of the biotinylated peptide. H3$_{1-21}$K4et$_3$ **3.1** and H3$_{1-21}$K4pip **3.8** were observed to have IC$_{50}$ values of 0.515 ± 0.003 µM and 0.325 ± 0.030 µM; similar to H3$_{1-21}$K4me$_3$ **3.2** (0.554 ± 0.105 µM). Results for KDM5A$_{PHD1}$ demonstrate that H3$_{1-21}$K4me$_0$ **3.5** (IC$_{50}$: 3.89 ± 0.80 µM) is the preferred lysine methylation state, with decreasing preference for H3$_{1-21}$K4me$_1$ **3.4** (IC$_{50}$: 18.7 ± 3.5 µM), H3$_{1-21}$K4me$_2$ **3.3** (25.0 ± 4.93% displacement at 100 µM) and H3$_{1-21}$K4me$_3$

3.2 (22.3 ± 1.5% displacement at 100 μM). In general, $H3_{1-21}K4et_3$ **3.1** (19.3 ± 4.7% displacement at 100 μM) and $H3_{1-21}K4pip$ **3.8** (39.6 ± 4.6% displacement at 100 μM) had comparable IC_{50} values to $H3_{1-21}K4me_3$ **3.2** and $H3_{1-21}K4me_2$ **3.3** for both PHD-fingers.

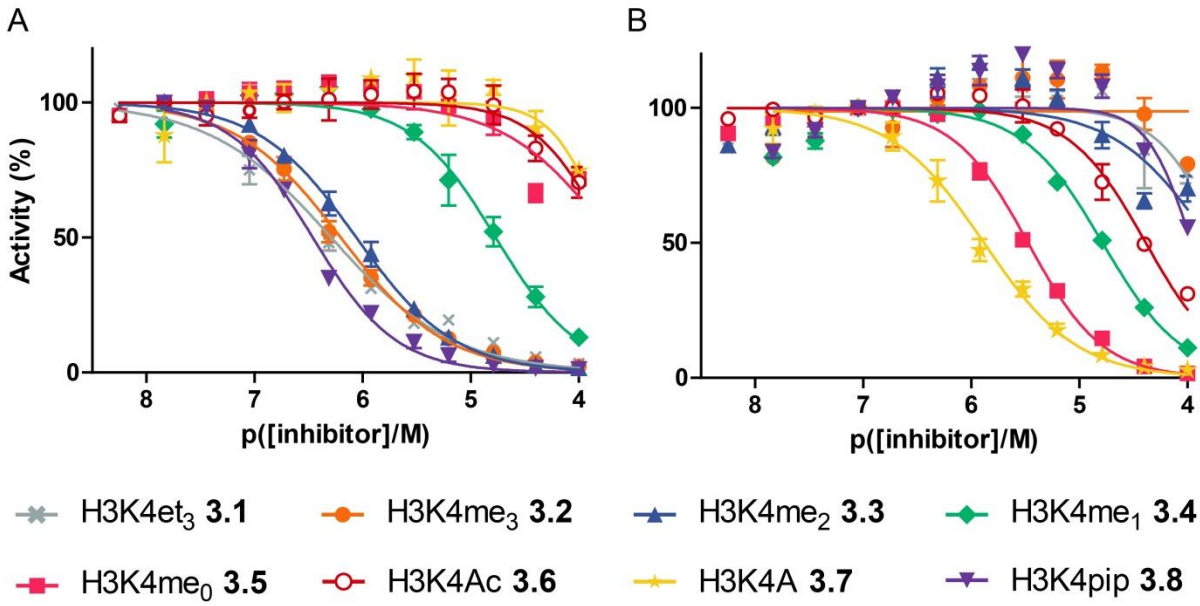


Figure 3.5.6: IC_{50} curves of AlphaScreen™ displacement assay using proteins $KDM5A_{PHD3}$ (A) or $KDM5A_{PHD1}$ (B) against $H3_{1-21}K4$ 3.1–3.8 peptides. n= 2, error given as s.d.

To understand the correlation between the data obtained using ITC and AlphaScreen™, the results were plotted revealing a good correlation for $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$ (R^2: 0.91, R^2:0.91) (Fig. 3.5.7). Note that peptides (*e.g.* $H3_{1-21}K4Ac$ **3.6**) with incomplete IC_{50} curve do not provide an accurate displacement value and are therefore not included in the graph.

Table 3.5.2: IC_{50} values of AlphaScreen™ of $KDM5A_{PHD3}$ or $KDM5A_{PHD1}$ in complex to – $H3_{1-21}K4me_{3/0}$-[Biotin] displaced by $H3_{1-21}K4$ 3.1–3.8. Values given in percentages refer to percentage displacement at top peptide concentration of 100 µM. n= 2–4 samples, error given as s.d.

		$KDM5A_{PHD3}$	$KDM5A_{PHD1}$
		IC_{50} µM	
$H3_{1-21}K4et_3$	**3.1**	0.515 ± 0.003	19.3 ± 4.7% at 100 µM
$H3_{1-21}K4me_3$	**3.2**	0.554 ± 0.105	22.3 ± 1.5% at 100 µM
$H3_{1-21}K4me_2$	**3.3**	0.954 ± 0.108	25.0 ± 4.9% at 100 µM
$H3_{1-21}K4me_1$	**3.4**	15.9 ± 1.53	18.7 ± 3.5
$H3_{1-21}K4me_0$	**3.5**	40.7 ± 12.4% at 100 µM	3.93 ± 0.80
$H3_{1-21}K4Ac$	**3.6**	21.9 ± 3.4% at 100 µM	65.1 ± 3.1 at 100 µM
$H3_{1-21}A4$	**3.7**	31.0 ± 1.4% at 100 µM	1.40 ± 0.11
$H3_{1-21}K4pip$	**3.8**	0.325 ± 0.030	39.6 ± 4.6% at 100 µM

AlphaScreen™ functional displacement assay corroborates the ITC data and confirms that $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** can bind to $KDM5A_{PHD3}$ with similar potency and in a similar manner to $H3_{1-21}K4me_3$ **3.2**; all three are good binders to $KDM5A_{PHD3}$ but weak binders to $KDM5A_{PHD1}$ with at least >20-fold selectivity for PHD3 over PHD1 of KDM5A.

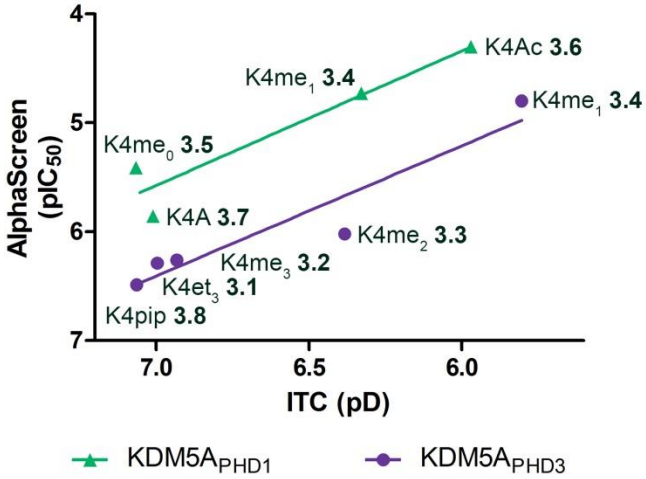


Figure 3.5.7: Correlation between obtained ITC data and AlphaScreen™ displacement of $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$. Good correlation was found for $KDM5A_{PHD3}$ (R^2: 0.91) and $KDM5A_{PHD1}$ (R^2: of 0.91).

3.5.5 Production of KDM5 proteins

Having established that $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** are good binders to $KDM5A_{PHD3}$ but weak binders to $KDM5A_{PHD1}$, studies were conducted to determine if these peptides can function as

substrates or inhibitors of the KDM5A JmjC-domain. To start, different constructs of KDM5 proteins were designed and expressed using Baculo virus in combination with insect Sf9 cell line, a parental clonal isolate from IPLB-Sf21-AE, as shown in figure 3.5.2 and figure 3.5.8.[288]

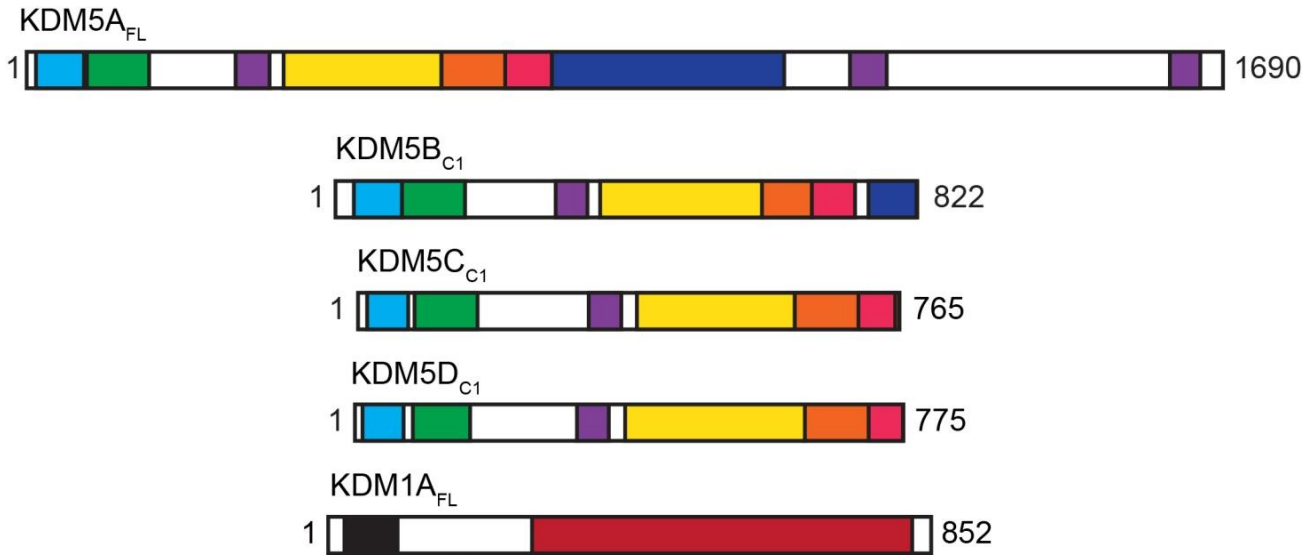


Figure 3.5.8: Schematic representation of KDM5A-D and KDM1A$_{FL}$ protein length and domains used in this Chapter. Purple: PHD-domain, light blue: JmjN-domain, green: ARID-domain, yellow: JmjC-domain, orange: α-helix, pink: C5HC2-zinc finger, Blue: PLU-domain, black: swirm-domain and red: amine-oxidase.

KDM5A$_{C1}$ was produced (*by Dr. Catrine Johansson*) in the absence of PHD1, PHD2, PHD3 and the ARID domain while KDM5A$_{C2}$ was produced by without PHD2 and PHD3 (*by Dr. Catrine Johansson*). KDM5A-Full length (KDM5A$_{FL}$) was purchased from Active Motif. These three KDM5A constructs enable the study of the different domains and the importance of domain-cross talk, as well as assigning functions to certain domains. Furthermore, histone lysine demethylases KDM5B$_{C1}$ (*Dr. Catrine Johansson*), KDM5C$_{C1}$ (*Aleksandra Szykowska*), KDM5D$_{C1}$ (*Eidarus Salah*) and amine oxidase KDM1A$_{FL}$ (*Raphael Reinbold*) were produced to determine the potential inhibition, activation or selectivity across the other H3K4 demethylases.

3.5.6 FDH measurements to investigate the demethylation activity of KDM5 proteins

To test the produced KDM5 proteins for catalytic activity, MALDI-TOF assays were carried out. The relative ratios of substrate and product(s) were measured by quenching assays at different time-points, and initial rates were determined. Attempts were made to use MALDI-TOF for characterising KDM5s using steady state Michaelis-Menten enzyme kinetics. However, the initial results revealed that the Michaelis constant (K_M) of the KDM5 are below the level of quantification for MALDI-TOF ($K_M \leq 1.0$ µM for $H3_{1\text{-}21}K4me_3$ **3.2**). As an alternative approach to determine the enzymatic properties of KDM5s, the enzyme-coupled fluorescence based formaldehyde dehydrogenase (FDH) assay was used.[187] In this method, formaldehyde released as a result of KDM5-catalysed demethylation of $H3_{1\text{-}21}K4me_{3,2,1}$ **3.2–3.4** substrate, is oxidized by FDH enzyme (Fig. 3.5.9 A) into formic acid with the concomitant reduction of β-Nicotinamide adenine dinucleotide (NAD^+) into disodium β-Nicotinamide adenine dinucleotide (NADH). NADH production can be monitored by excitation at 366 ± 3 nm leading to the generation of a fluorescence signal at 460 ± 10 nm. Thus, the formation of NADH can be quantitated by measuring the fluorescence readout in a continuous assay setup.[289]

The FDH method used was, based on previous reported procedures[290–292], Pherastar FX (BMGLabtech) plate reader equipped with the FI 350 460-detection module was used, providing excitation of NADH at 350 nm and fluorescence detection at 460 nm. Note that for this assay, the buffer is supplemented with Tween20® (0.01%) to avoid non-specific binding of protein to the plate. Standard assay conditions for the FDH assay consisted of $H3_{1\text{-}21}K4me_3$ **3.2** at various concentrations in the presence sodium ascorbate (1.0 mM) (Asc), $(NH_4)_2Fe(II)(SO_4)_2$ (50 µM), disodium 2-oxoglutarate (200 µM) (2OG), NAD^+ (250 µM) and FDH (1.0 µM) in buffer [HEPES (50 mM), NaCl (50 mM), Tween20® (0.01%) in MQ (pH 7.5)] with variable enzyme ranging between 0.15 µM to 0.50 µM. For enzyme kinetics with variable 2OG concentration, the $H3_{1\text{-}21}K4me_3$ **3.2** concentration

was set to 5.0 μM. The reactions were monitored after initiation of the reaction (*e.g.* with enzyme mixture) measuring the fluorescence in 30 sec intervals. The collected data points were standardized and baseline corrected against no peptide or 2OG controls. Quantification was carried out using a formaldehyde standard curve (Fig. 3.5.9 B).

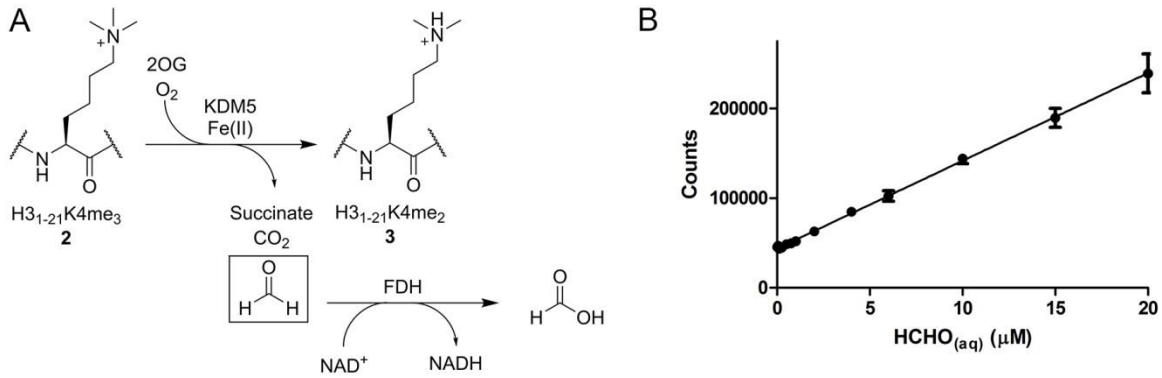


Figure 3.5.9: Schematic representation of FDH assay and standard curve using this method of formaldehyde. A) KDM5A uses O_2 and 2OG to oxidize a methyl group on $H3_{1\text{-}21}K4me_3$ **3.2** generating succinate, CO_2, formaldehyde and the demethylated peptide $H3_{1\text{-}21}K4me_2$ **3.3**. The formaldehyde is oxidized by FDH to give formic acid and reducing NAD^+ to NADH which can be measured by fluorescence. B) Standard curve of formaldehyde converted by FDH to formic acid converting NAD^+ to NADH. R^2: 0.99, n= 3 and error bars as s.d.

Determining an accurate concentration of formaldehyde for the standard curve is not a trivial task. Formaldehyde is a gas under standard conditions and unstable in aqueous solutions. Commercial solutions of formaldehyde are typically stabilized with methanol (15%) which may influence measurements. Alternative sources were considered such as paraformaldehyde being heated ('cracked') in water to produced formaldehyde. However, this may also introduce a degree of uncertainty due to the volatile nature of formaldehyde. The standard curve used was made from a commercially obtained source in the buffer with corresponding co-factors Asc, 2OG and Fe^{II}. The linear region was found to be between 0.05 to 20 μM with a correlation of R^2: 0.99 and applied across all measurements.

With assay conditions optimized, steady state enzyme kinetics were performed on $KDM5A_{C1}$ and $KDM5A_{C2}$ against $H3_{1\text{-}21}K4me_3$ **3.2** obtaining K_M: 0.887 ± 0.053 μM and K_M: 1.87 ± 0.15 μM,

respectively (Fig. 3.5.10 A, Table 3.5.3). This difference of ~2.1-fold shows that the ARID and PHD1 domains decrease the K_M value, which is indicative of increase in binding strength. Furthermore, the catalytic efficiency of $KDM5A_{C2}$ (K_{cat}: 0.264 ± 0.008 min^{-1}) was decreased by ~2.2 fold in comparison to $KDM5A_{C1}$ (K_{cat}: 0.581 ± 0.009 min^{-1}).

Attempts were made to perform steady state kinetics on $KDM5A_{FL}$ but the protein was provided in a tris(hydroxymethyl)aminomethane (Tris) buffer which interferes with the FDH assays. Tris is potentially a substrate for FDH enzyme resulting in elevated levels of coupled NAD^+ turnover giving a high NADH background.

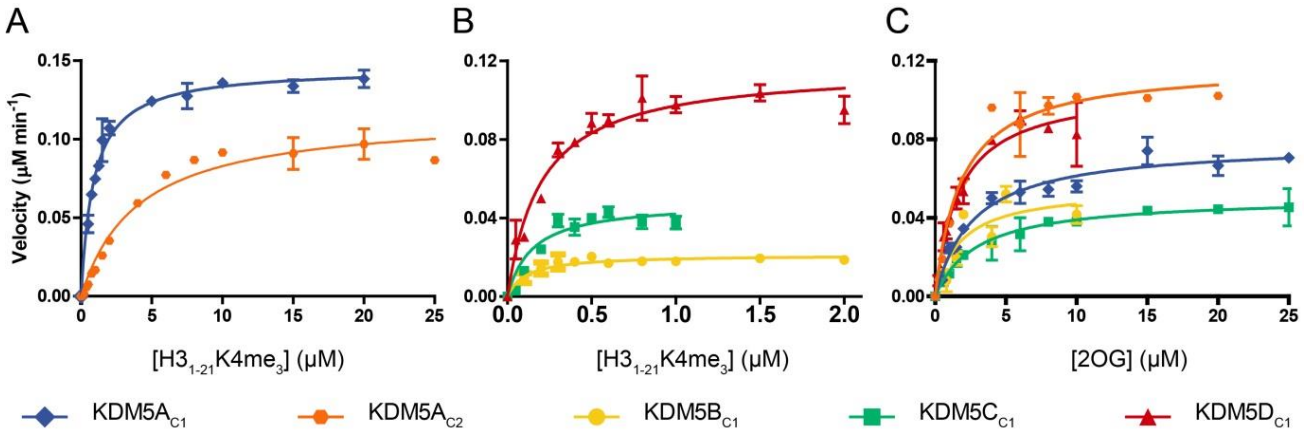


Figure 3.5.10: Steady state Michaelis-Menten kinetics curves of $KDM5A_{C1}$, $KDM5A_{C2}$, $KDM5B_{C1}$, $KDM5C_{C1}$ and $KDM5D_{C1}$ of $H3_{1-21}K4me_3$ 3.2 (A, B) substrate or 2OG (C) as determined using FDH assay. n= 2–3, error given as s.d.

$KDM5B_{C1}$ (0.2 μM) was found to have a K_M^{App} of 0.109 ± 0.021 μM and a K_{cat} 0.106 ± 0.004 min^{-1}, which is substantially lower than comparative construct for $KDM5A_{C2}$. The relatively low K_M^{App} coupled to the limited demethylation turnover, prohibited the use of further reduction in enzyme concentration. This limits the steady state kinetics assumption that [$H3_{1-21}K4me_3$] >> [$KDM5B_{C1}$] and indicates that the K_M value can only be lower than currently observed.

Table 3.5.3: KDM5 enzyme kinetics of $H3_{1-21}K4me_3$ 3.2 or 2OG substrate. n= 2–3, error given as st. error.

		$KDM5A_{C1}$	$KDM5A_{C2}$	$KDM5B_{C1}$	$KDM5C_{C1}$	$KDM5D_{C1}$
$H3_{1-21}K4me_3$ **3.2**	K_{cat} (min^{-1})	0.581 ± 0.009	0.264 ± 0.008	0.106 ± 0.004	0.482 ± 0.040	0.778 ± 0.025
	K_M (μM)	0.887 ± 0.053	1.87 ± 0.15	0.109 ± 0.021	0.147 ± 0.050	0.201 ± 0.025
	V_{max} (μM min^{-1})	0.145 ± 0.002	0.132 ± 0.004	0.021 ± 0.001	0.048 ± 0.004	0.117 ± 0.004
2OG	K_{cat} (min^{-1})	0.312 ± 0.010	0.237 ± 0.008	0.150 ± 0.006	0.506 ± 0.022	1.07 ± 0.04
	K_M (μM)	2.71 ± 0.31	1.83 ± 0.26	1.11 ± 0.22	2.98 ± 0.46	1.69 ± 0.19
	V_{max} (μM min^{-1})	0.078 ± 0.003	0.116 ± 0.004	0.030 ± 0.001	0.051 ± 0.002	0.107 ± 0.004

$KDM5C_{C1}$ (K_{cat}: 0.482 ± 0.040 μM·min^{-1}) and $KDM5D_{C1}$ (K_{cat}: 0.778 ± 0.025 μM·min^{-1}) were found to have a similar K_{cat} to $KDM5A_{C1}$ (K_{cat}: 0.581 ± 0.009) for $H3_{1-21}K4me_3$ **3.2**. However, the K_M of $H3_{1-21}K4me_3$ **3.2** against $KDM5C_{C1}$ (K_M^{App}: 0.147 ± 0.050 μM) and $KDM5D_{C1}$ (K_M^{App}: 0.201 ± 0.025 μM) were closer to the value determined for $KDM5B_{C1}$ (K_M^{App}: 0.109 ± 0.021 μM) (Fig. 3.5.10 B). Steady state kinetics using 2OG (Fig. 3.5.10 C, Table 3.5.3) as variable substrate revealed a relatively small range of K_M observed with $KDM5C_{C1}$ (K_M: 2.98 ± 0.46 μM) the highest and $KDM5B_{C1}$ the lowest (K_M: 1.11 ± 0.22 μM). $KDM5A_{C1}$ (K_M: 2.71 ± 0.31 μM) showed to have relatively similar properties as to $KDM5A_{C2}$ (K_M: 1.83 ± 0.26 ± μM) indicating that the 2OG active site is less affected by the deletion of the ARID and PHD1 domains than the peptide binding site. Observed V_{max} and corresponding K_{cat} were found to be of similar range between $H3_{1-21}K4me_3$ **3.2** and 2OG kinetics across the KDM5s. Data here are interestingly observed with lower K_M for $H3K4me_3$ substrate and 2OG than found by Horton *et al.* using the FDH assay.[268]

3.5.7 MALDI-TOF measurements to determine if $H3_{1-21}K4et_3$ 3.1 and $H3_{1-21}K4pip$ 3.8 are alternative substrates KDM5 proteins

Results obtained in the previous section (3.5.6) demonstrated that KDM5A-D enzymes are catalytically active and can demethylate $H3_{1-21}K4me_3$ **3.2** substrate. The FDH assay is limited to indirect formaldehyde detection and provides no information about other potential enzymatic activities on alternative substrates such as hydroxylation or dealkylation. To provide supporting evidence,

MALDI-TOF analysis was used to identify conversions made to the peptides. Using this method, the default buffer conditions were used with additive TCEP·HCl to improve demethylation activity. This time-point based assay was quenched using an excess of formic acid and samples were co-spotted with α-cyano 4-hydroxy-cinnamic acid (HCCA) MALDI-matrix on a polished steel target plate for MALDI-TOF analysis. Time course demethylation assays of $H3_{1\text{-}21}K4me_3$ **3.2**, $H3_{1\text{-}21}K4et_3$ **3.1** and $H3_{1\text{-}21}K4pip$ **3.8** were carried out at standard and elevated enzyme concentrations of KDM5s: $KDM5A_{C1}$ (0.5 μM, 2.5 μM), $KDM5A_{C2}$ (0.5 μM, 2.5 μM), $KDM5B_{C1}$ (0.25 μM, 1.25 μM), $KDM5C_{C1}$ (0.25 μM, 1.25 μM) and $KDM5D_{C1}$ (0.25 μM, 1.25 μM). Reaction mixtures were performed in the presence of co-factors (Asc: 1.0 mM, Fe^{II}: 50 μM, 2OG: 200 μM, TCEP: 5.0 mM).

Representative spectra for $H3_{1\text{-}21}K4me_3$ **3.2** (5.0 μM) and $H3_{1\text{-}21}K4et_3$ **3.1** (5.0 μM) with proteins $KDM5A_{C2}$ (0.5 μM, 2.5 μM) and $KDM5C_{C1}$ (0.25 μM, 1.25 μM) are displayed in figure 3.5.11. Analysis of the data reveal that $H3_{1\text{-}21}K4me_3$ **3.2** is rapidly demethylated with peaks observable for each iterative demethylation state showing a −14 Da mass shift resulting in $H3_{1\text{-}21}K4me_2$ **3.3** and $H3_{1\text{-}21}K4me_1$ **3.4**.

Under standard assay conditions, only traces of $H3_{1\text{-}21}K4me_0$ **3.5** were observed for all five tested constructs (Fig. 3.5.11 A,E). An increased concentration of protein led, as expected ([Enz] $\propto$ v), to faster demethylation for all proteins examined and produced higher levels of $H3_{1\text{-}21}K4me_0$ **3.5** (Fig. 3.5.11 B,F). $H3_{1\text{-}21}K4et_3$ **3.1**, tested as a substrate, did not show a mass change for all five proteins under normal conditions (Fig. 3.5.11 C,G). However, under higher enzyme conditions and prolonged incubation, $KDM5C_{C1}$ and $KDM5C_{D1}$ showed low levels of +16 Da peak after 60 min (Fig. 3.5.14 H).

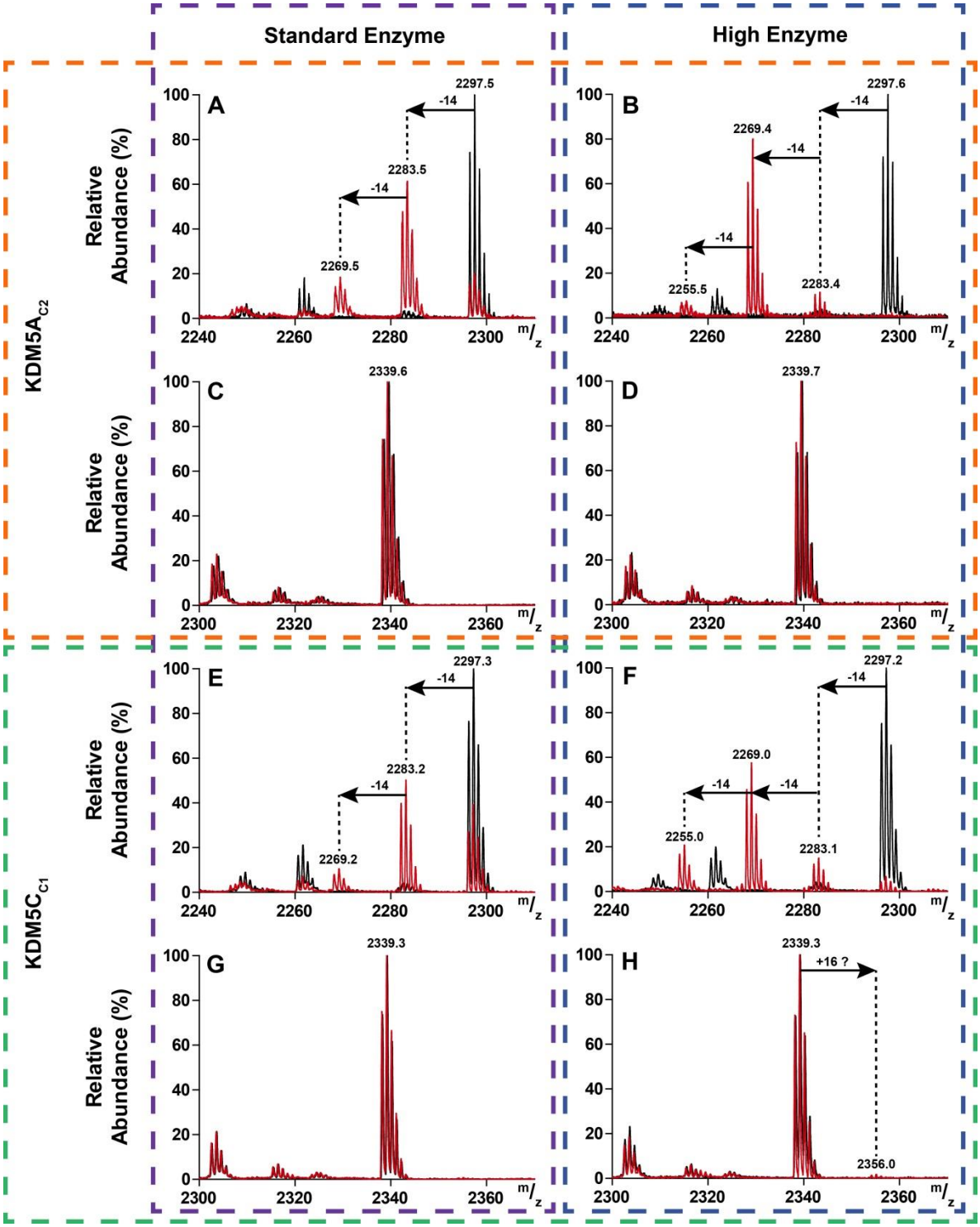


Figure 3.5.11: MALDI-TOF spectra of KDM5A$_{C2}$ (orange dashed box, A-D) or KDM5C$_{C1}$ (green dashed box, E-H) demethylation of H3$_{1-21}$K4me$_3$ 3.2 (A-B, E-F) or H3$_{1-21}$K4et$_3$ 3.1 (C-D, G-H). The purple dashed box contains standard condition with KDM5A$_{C2}$ (0.5 µM) and KDM5C$_{C1}$ (0.25 µM) (black line: 0 min, red line: 30 min). The blue dashed box contains higher enzyme conditions with elevated levels of KDM5A$_{C2}$ (2.5 µM) and KDM5C$_{C1}$ (1.25 µM) and prolong reaction times (black line: 0 min, red line: 60 min). KDM5A$_{C2}$ and KDM5C$_{C1}$ show that H3$_{1-21}$K4me$_3$ **3.2** can be demethylated under standard and high enzyme conditions. KDM5C$_{C1}$ demonstrates that under high enzyme conditions H3$_{1-21}$Ket$_3$ **3.1** may potentially be hydroxylated (H). Peptide (5.0 µM), Asc (1.0 mM), Fe^{II} (50 µM), 2OG (200 µM), TCEP (5.0 mM).

Considering the nature of the KDM JmjC-domain containing proteins, it is likely that an oxygen is introduced forming a hydroxylated C^{β}-ethyl group as part of the N^{ε}-Lysine head group. The low levels of product formed prohibited further MALDI-TOF/TOF analysis.

Interestingly, +16 Da peaks were observed for KDM5A-D when incubated with $H3_{1-21}K4pip$ **3.8**, at both standard and high enzyme concentrations ($KDM5B_{C1}$ and $KDM5D_{C1}$ displayed in Fig. 3.5.12). The natural histone mimic substrate $H3_{1-21}K4me_3$ **3.2** showed demethylation under standard enzyme as well as high enzyme conditions. Interestingly, a +16 Da mass shift was seen with $H3_{1-21}K4pip$ **3.8** as the substrate when treated with co-factors and enzymes KDM5A-D_{C1} and $KDM5A_{C2}$. The tested conditions suggest potential hydroxylation of the piperidine ring and initial analysis shows a clear difference between KDM5 proteins. $KDM5A_{C1-2}$ and $KDM5B_{C1}$ (Fig, 3.5.12 C, D) indicated a lower hydroxylation proficiency in generating $H3_{1-21}K4pipOH$ in comparison to $KDM5C_{C1}$ and $KDM5C_{D1}$ (Fig. 3.5.12 G,H). Steady state Michaelis-Menten enzyme kinetics were then used to define the kinetic properties using MALDI-TOF (Table 3.5.4, Fig. 3.5.13). $KDM5C_{C1}$ and $KDM5D_{C1}$ was found to exhibit K_M values of 2.53 ± 0.43 μM and 1.64 ± 0.31 μM for $H3_{1-21}K4pip$ **3.8**, respectively. But $KDM5A_{C1}$ (K_M: 24.6 ± 6.8 μM) and $KDM5B_{C1}$ (K_M: 10.2 ± 2.3 μM) were found to contain substantially higher K_M values with nearly ~15-fold difference between $KDM5A_{C2}$ and $KDM5D_{C1}$.

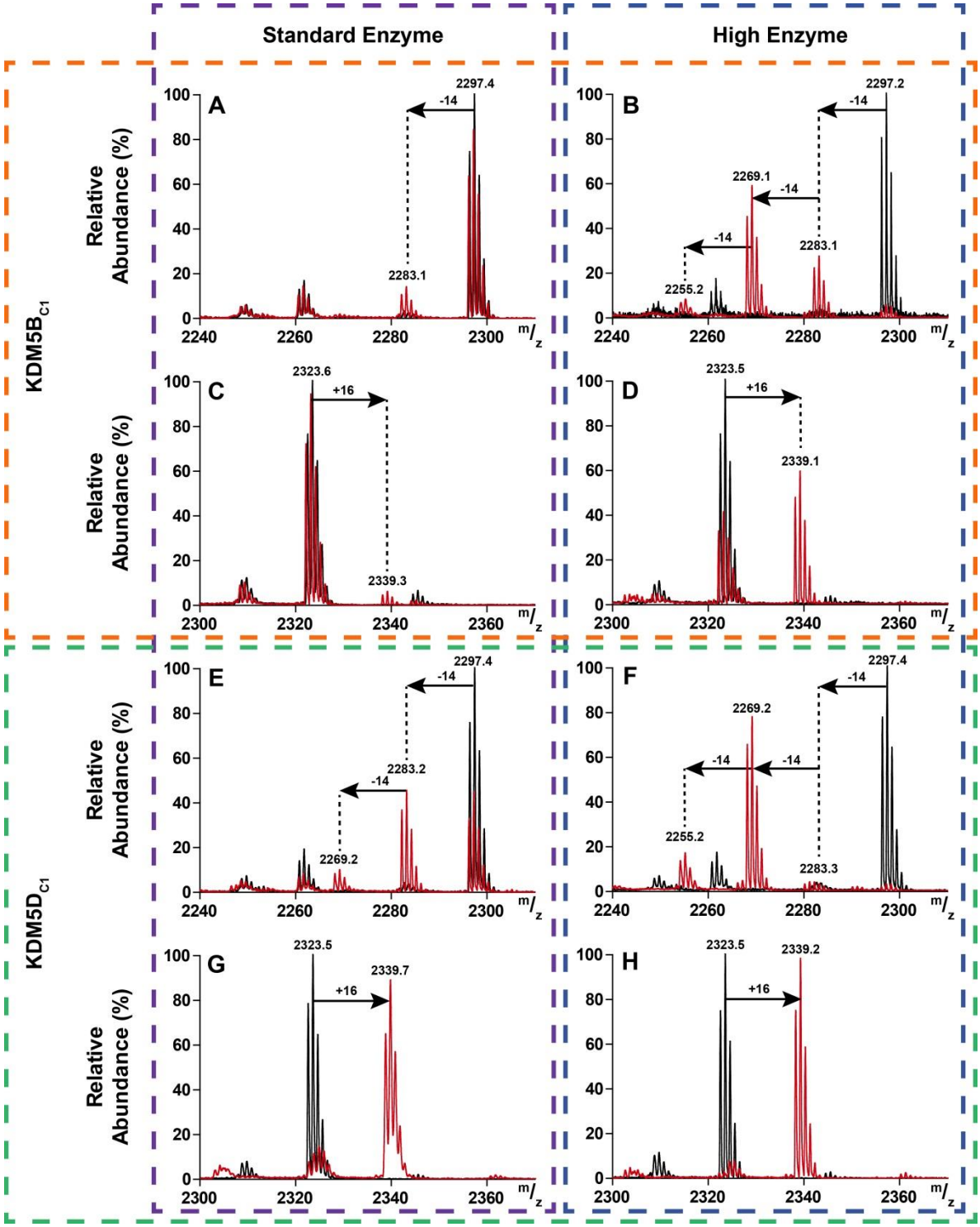


Figure 3.5.12: MALDI-TOF spectra of KDM5B$_{C1}$ (orange dashed box, A-D) or KDM5D$_{C1}$ (green dashed box, E-H) demethylation of H3$_{1-21}$K4me$_3$ 3.2 (A-B, E-F) or H3$_{1-21}$K4pip 3.8 (C-D, G-H). The purple dashed box contains standard condition with KDM5B$_{C1}$ (0.25 µM) and KDM5D$_{C1}$ (0.25 µM) (black line: 0 min, red line: 30 min). The blue dashed box represents high enzyme conditions with elevated levels of KDM5B$_{C1}$ (1.25 µM) and KDM5D$_{C1}$ (1.25 µM) and prolonged reaction times (black line: 0 min, red line: 60 min). KDM5B$_{C1}$ and KDM5D$_{C1}$ show that H3$_{1-21}$K4me$_3$ **3.2** can be demethylated under standard and high enzyme conditions. KDM5B$_{C1}$ showed low levels of hydroxylation under standard conditions whereas KDM5D$_{C1}$ also hydroxylated H3$_{1-21}$K4pip **3.8** efficiently under normal conditions. Peptide (5.0 µM), Asc (1.0 mM), Fe^{II} (50 µM), 2OG (200 µM), TCEP (5.0 mM).

The obtained K_M values are inverse proportional correlate to the K_{cat} values; a higher K_M showed a significantly lower K_{cat}. Whereas the K_{cat} of $KDM5C_{C1}$ (K_{cat}: 1.47 ± 0.07 μM·min^{-1}·μM Enz^{-1}) and $KDM5D_{C1}$ (K_{cat}: 1.26 ± 0.06 μM·min^{-1}·μM Enz^{-1}) revealed relatively efficient catalysis, $KDM5A_{C2}$ (K_{cat}: 0.128 ± 0.015 μM·min^{-1}·μM Enz^{-1}) and $KDM5B_{C1}$ (K_{cat}: 0.308 ± 0.034 μM·min^{-1}·μM Enz^{-1}) have substantially lower level of turnover, with ~11.5-fold difference in catalytic turnover between $KDM5A_{C2}$ and $KDM5C_{C1}$. Although observations for $KDM5A_{C1}$ were made showing limited formation of the +16 Da peak, relatively slow turnover prohibited accurate kinetic assessment using MALDI-TOF analysis without altering the protocol.

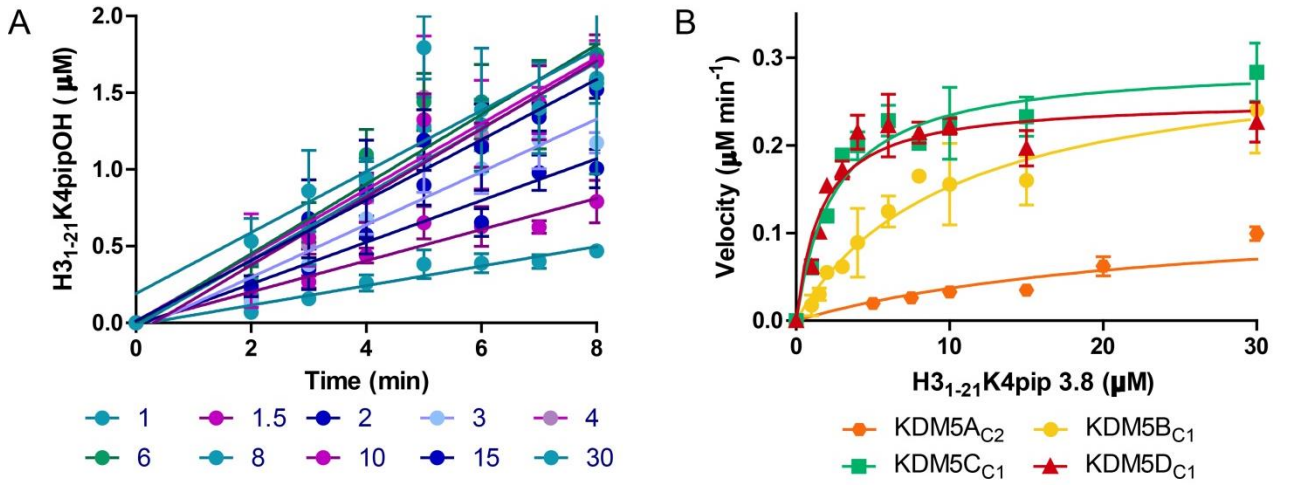


Figure 3.5.13: Steady state Michaelis-Menten enzyme kinetics of $H3_{1-21}$K4pip 3.8 substrate to form $H3_{1-21}$K4pipOH by KDM5 enzymes measured using MALDI-TOF. A: Time-course of $KDM5D_{C1}$ with at different $H3_{1-21}$K4pip **3.8** (1–30 μM) substrate concentrations showing product formation of $H3_{1-21}$K4pipOH over time (min), B: Best fit of steady state kinetics of $KDM5A_{C2}$ (1.2 μM), $KDM5B_{C1}$ (1.0 μM), $KDM5C_{C1}$ (0.2 μM) and $KDM5D_{C1}$ (0.2 μM) using $H3_{1-21}$K4pip **3.8** substrate showing the reaction velocity (μM·min^{-1}) as function of substrate concentration ranging between 0–30 μM ($KDM5A_{C2}$: 0–80 μM). Asc (1.0 mM), Fe^{II} (50 μM), 2OG (200 μM), TCEP (5.0 mM). n= 2, error given as s.d.

To define the +16 Da modification site on $H3_{1-21}$K4pip **3.8**, MALDI-TOF/TOF was employed to fragment peptides using laser induced dissociation (LID) giving rise to B and Y ions. $KDM5C_{C1}$ and $KDM5D_{C1}$ were selected for MALDI-TOF/TOF analysis as these proteins demonstrated to be able to hydroxylate the $H3_{1-21}$K4pip **3.8** to near completion without stoichiometric peptide and enzyme conditions. To enhance the chances of complete sequence coverage in MALDI-TOF/TOF experiments, the $H3_{1-21}$K4pip **3.8** (0.5 mM) peptide concentration was increased with corresponding increased levels of co-factors (Asc: 2.0 mM, Fe^{II}: 100 μM, 2OG: 2.0 mM, TCEP: 2.5 mM) and

$KDM5C_{C1}/D_{C1}$ (3.6 / 3.2 µM). After reacting for 16 hrs a (>90%) conversion level was obtained based on peak area and the sample was subjected to MALDI-TOF/TOF measurements.

Table 3.5.4: Steady state enzyme kinetics of $H3_{1-21}$K4pip 3.8 reacted by $KDM5A_{C2}$, $KDM5B_{C1}$ $KDM5C_{C1}$ and $KDM5C_{D1}$ to produce $H3_{1-21}$K4pipOH measured using MALDI-TOF. Error given as std. error.

	K_m (µM)	K_{cat} (min^{-1})	V_{max} (µM·min^{-1})
$KDM5A_{C2}$	24.6 ± 6.8	0.128 ± 0.015	0.128 ± 0.015
$KDM5B_{C1}$	10.2 ± 2.3	0.308 ± 0.034	0.308 ± 0.034
$KDM5C_{C1}$	2.53 ± 0.43	1.46 ± 0.07	0.294 ± 0.014
$KDM5D_{C1}$	1.64 ± 0.31	1.26 ± 0.06	0.252 ± 0.012

Analysis of the MALDI-TOF/TOF data demonstrates sequence coverage of 71% peptide for $KDM5C_{C1}$ and $KDM5D_{C1}$ based on B and Y ions detected (Figure 3.5.14 C). With the modification site down to amino acid resolution, evidence of the exact position on lysine 4 remained to be determined. Potential sites of modification are shown in figure 3.5.14 A, B. If hydroxylation takes place adjacent to the N^{ε}-Lysine nitrogen (Fig. 3.5.14 A, red and orange circles) a hemiaminal would be formed that could result in the loss of water to give a −2 Da shift relative to $H3_{1-21}$K4pip **3.8**, which is not observed. The hemiaminal could undergo potential hydrolysis of either the C^{ε}-Lysine-piperidine (Fig. 3.5.14 A, orange) to give a homo glutamic acid derivative that would show a −69 Da shift (Fig. 3.5.14 B, orange box) or hydrolysis of C^{α}-piperidine (Fig. 3.5.14 A, red) resulting in ring opening (Fig. 3.5.14 B, red box). A mass change of −69 Da relative to $H3_{1-21}$K4pip **3.8** is not observed however, hydrolysis of C^{α}-piperidine cannot be detected directly as the hemiaminal and the aldehyde derivatives have the same mass. To derivatize the potentially formed aldehyde, acetyl hydrazine was added in 10- and 100-fold excess to $H3_{1-21}$K4pipOH and reacted for 24 hrs. There was no change in mass observed suggesting that formation of the aldehyde (Fig. 3.5.14 B, orange or red box) is unlikely. Evidence obtained using mass spectrometry concludes that the hydroxylation product is either the lysine alkyl side chain or piperidine C^{β} or C^{γ} with corresponding hydroxylation products displayed in figure 3.5.14 B in the green or purple box, respectively.

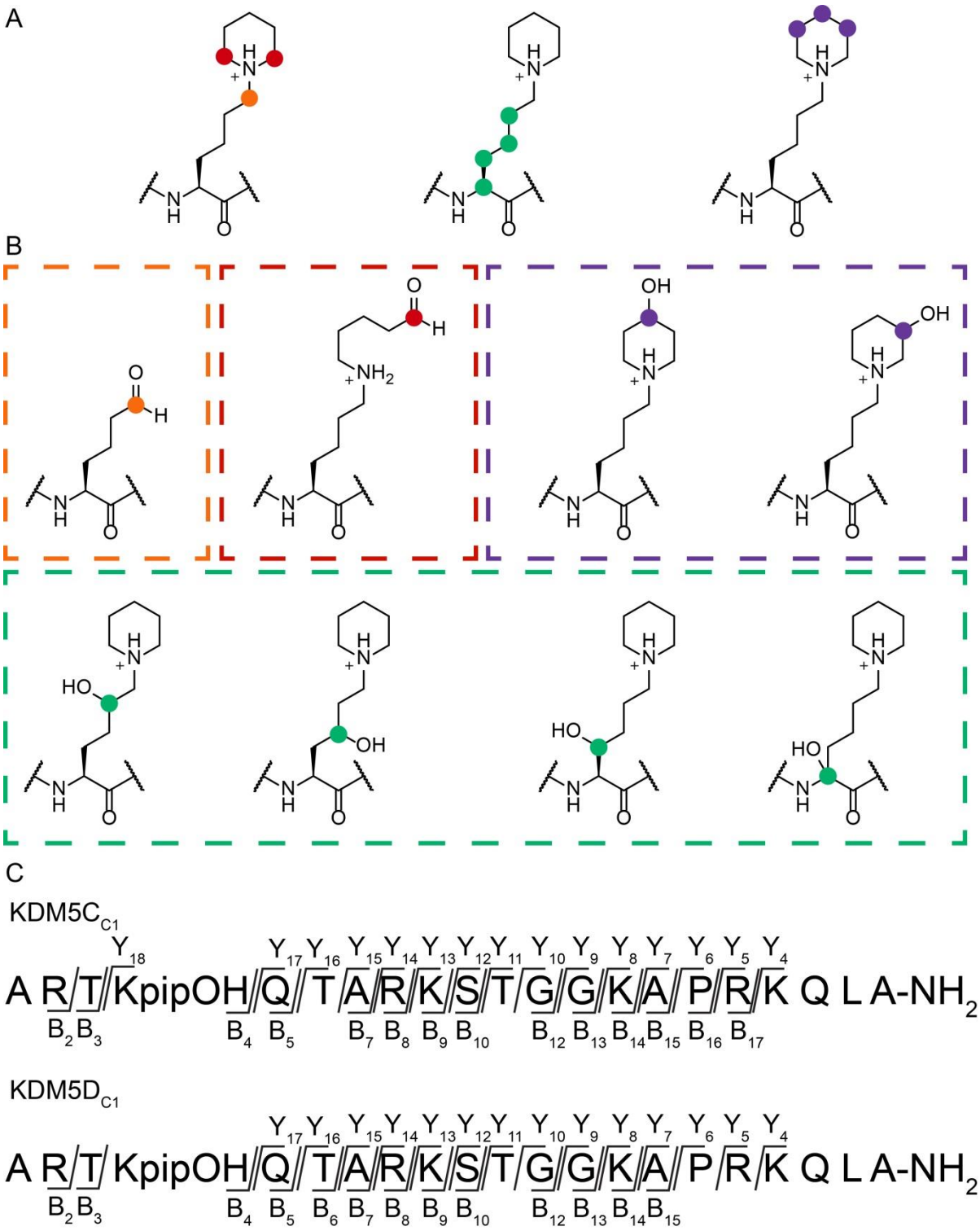


Figure 3.5.14: Schematic representation of H3$_{1-21}$K4pipOH possible sites of hydroxylation by KDM5D$_{C1}$ enzyme. A) Potential H3$_{1-21}$K4pipOH hydroxylation sites; B) Potentially observed mass spectra in MALDI-TOF analysis with expected mass shift of +16 Da for red, purple, green or −69 Da for orange. C) MALDI-TOF/TOF sequence coverage of H3$_{1-21}$K4pipOH. Hydroxylation site that could react by derivatization acetyl hydrazine (red, orange), hydroxylation site at lysine backbone (green) and piperidine ring of lysine 4 (purple). Potential products of hydroxylation colour coded and boxed.

To determine the exact location of hydroxylation ^{1}H and ^{13}C NMR techniques were used. Material for analysis was produced using $H3_{1-21}K4pip$ **3.8** (1.0 mM) co-factors (Asc: 2.0 mM, Fe^{II}: 100 µM, 2OG: 2.0 mM, TCEP: 2.5 mM) and $KDM5D_{C1}$ (6.35 µM) at 250 µL scale showing >91% conversion after 27 hrs as determined using MALDI-TOF. The peptide was purified using HPLC, lyophilized and dissolved in phosphate buffer (pH 7.1) in D_2O and subjected to NMR measurements.

Using acquired ^{1}H NMR, COSY and TOCSY spectra and assigned the ^{1}H NMR spectrum of $H3_{1-21}K4me_0$ allowed for the identification of the K4pip epsilon protons of $H3_{1-21}K4pip$ **3.8**.

^{1}H NMR, COSY and TOCSY experiments in combination with peptide overlay show protons of $H3_{1-21}K4pip$ **3.8** appearance and shift of the epsilon protons (Fig. 3.5.15 A,B). Interestingly, the spectrum $H3_{1-21}K4pipOH$ shows a shift of the K4pip epsilon protons and a shift of the piperidine protons (Fig. 3.5.15 B,C). It is likely that the hydroxylation at the delta carbon of lysine would have more pronounced effect on the epsilon protons while it is reasonable to assume that this would have little effect on the piperidine protons. The shift of the K4pipOH epsilon protons is relatively small ($\Delta\delta$: ~0.033 ppm) whereas piperidine protons signals were not observed. Combining this data with obtained MALDI-TOF derivatization results, the most likely places of hydroxylation are C^{β} or C^{γ} of the piperidine ring however, there is no conclusive analytical evidence that support the regiochemistry.

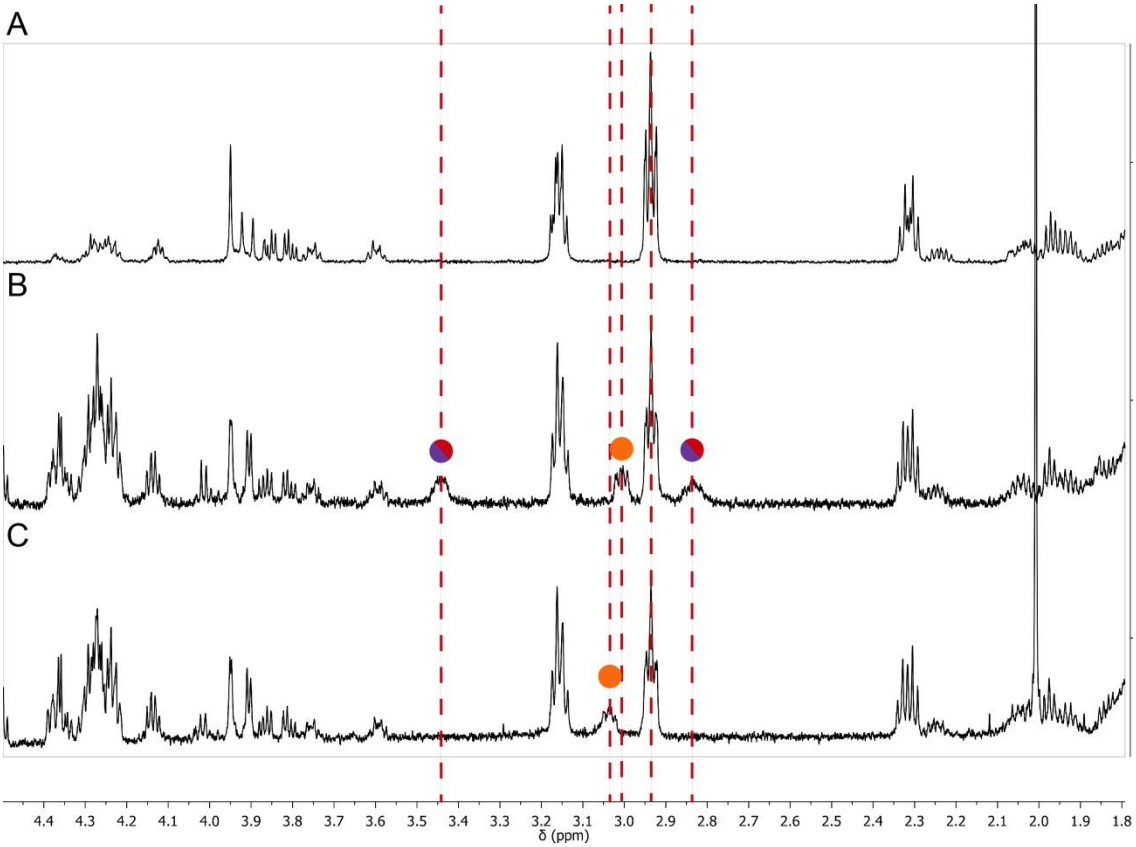


Figure 3.5.15: ^{1}H NMR spectra of $H3_{1-21}K4me_0$ 3.5 (A), $H3_{1-21}K4pip$ 3.8 (B) and $H3_{1-21}K4pipOH$ (C) aligned. Protons of piperidine ring (red/purple circles) epsilon protons of $H3_{1-21}K4pipOH$ (orange). After hydroxylation, piperidine protons (red/purple circles) and epsilon protons (orange) of lysine shift.

3.5.8 $H3_{1-21}K4et_3$ 3.1 and $H3_{1-21}K4pip$ 3.8 as inhibitors for KDM5s JmjC-domain

Next, $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** were tested as inhibitors (at 10 µM) of the JmjC-domain, by screening against all available KDM5 JmjC-domain containing constructs using $H3_{1-21}K4me_3$ **3.2** (5.0 µM) as substrate. The results, shown in figure 3.5.16, demonstrate no significant inhibition of $KDM5A_{C1-C2}$, $KDM5C_{C1}$ or $KDM5D_{C1}$.

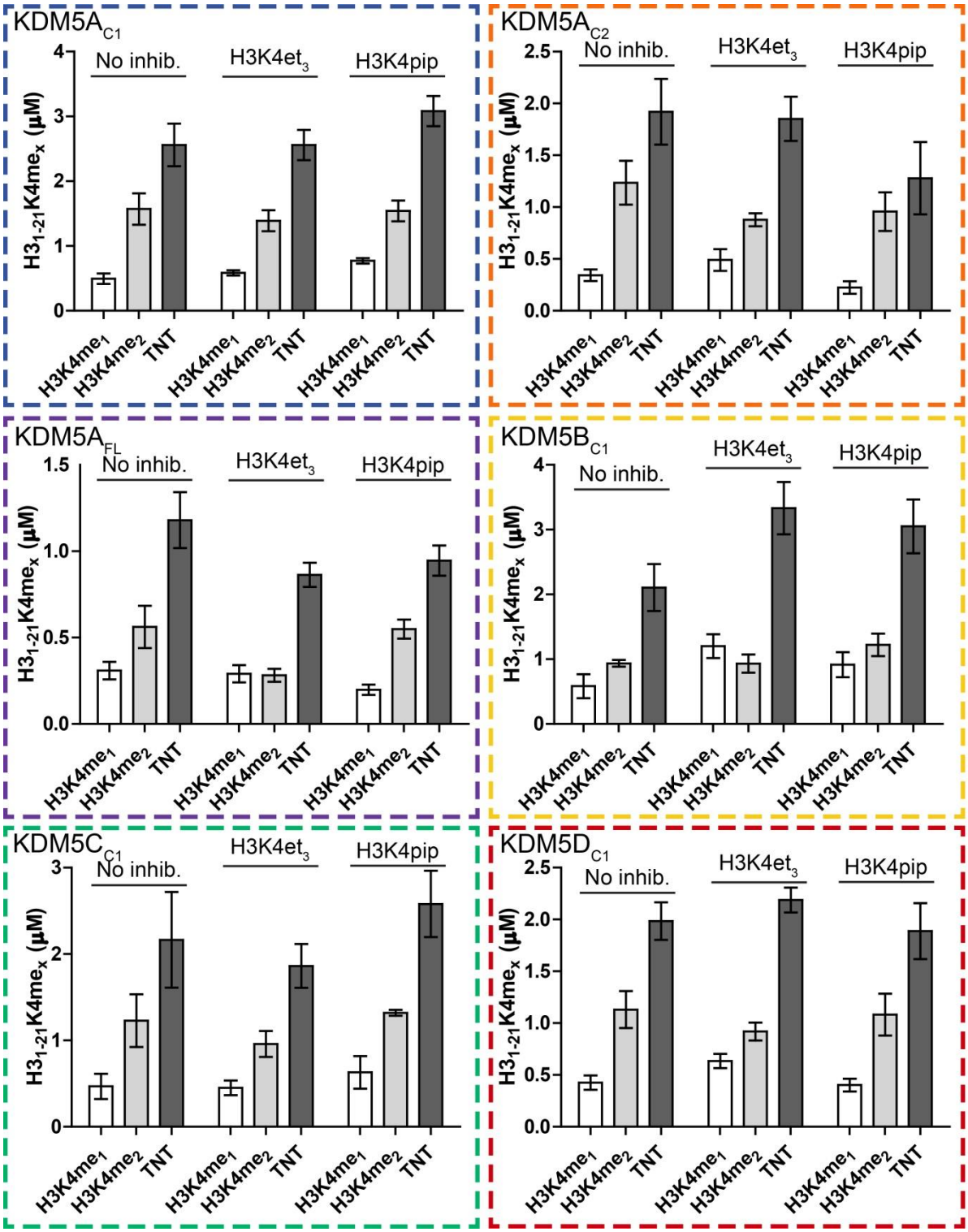


Figure 3.5.16: Single point inhibition assay of KDM5A$_{C1}$ (300 nM, blue box), KDM5A$_{C2}$ (300 nM, orange box), KDM5A$_{FL}$ (250 nM, purple box), KDM5B$_{C1}$ (500 nM, yellow box), KDM5C$_{C1}$ (200 nM, green box) and KDM5D$_{C1}$ (200 nM, red box) by H3$_{1-21}$K4et$_3$ 3.1 (10.0 µM), H3$_{1-21}$K4pip 3.8 (10.0 µM) or no inhibitor (no inhib.) analysed using MALDI-TOF. H3$_{1-21}$K4me$_3$ 3.2 (5.0 µM), Asc (1.0 mM), Fe^{II} (50 µM), 2OG (200 µM), TCEP (5.0 mM). n≥ 3, error bars given as s.d. TNT: total number of turnover.

Although potential weak inhibition was observed for H3$_{1-21}$K4et$_3$ **3.1** against KDM5A$_{FL}$, while it showed mild enhancement of demethylation activity for KDM5B$_{C1}$. However, technical replicates of

this experiment did not confirm these results and showed no sign of significant demethylase inhibition or acceleration. Furthermore, $H3_{1-21}K4pip$ **3.8** was tested as a competitive substrate against the KDM5 panel of demethylases (Fig. 3.5.16). Inspecting the mass spectra, $KDM5A_{FL}$ showed 1.51 ± 0.59% conversion to give $H3_{1-21}K4pipOH$ while $KDM5A_{C1}$ and $KDM5B_{C1}$ showed level of <1%. $KDM5A_{C2}$, $KDM5C_{C1}$ and $KDM5D_{C1}$ did not show any formation of $H3_{1-21}K4pipOH$. Since $H3_{1-21}K4pip$ **3.8** (10 µM) is a poor substrate in the presence of $H3_{1-21}K4me_3$ **3.2** (5.0 µM), it was also tested as an inhibitor. While demethylation levels appeared to be mildly affected for $KDM5A_{FL}$ and $KDM5B_{C1}$ (Fig. 3.5.16), this was not significant.

3.5.9 Inhibition of $KDM1A_{FL}$ by $H3_{1-21}K4et_3$ 3.1, $H3_{1-21}K4me_3$ 3.2 and $H3_{1-21}K4pip$ 3.8

To study the potential inhibition of KDM1A by $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8**, $KDM1A_{FL}$ was recombinantly expressed, purified *(by Raphael Reinbold)* and found to be active as determined using MALDI-TOF analysis. To test potential inhibition of $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8**, $KDM1A_{FL}$ was subjected to a time-course and samples were analysed using MALDI-TOF instrument. The linear region was determined using $H3_{1-21}K4me_1$ **3.4** substrate (5.0 µM) obtaining a linear region between 0 and 15 min with good correlation of (R^2: 0.97) (Fig. 3.5.17 A). Inhibition of $KDM1A_{FL}$ (0.15 µM) was tested with $H3_{1-21}K4me_1$ **3.4** (5.0 µM) substrate and $H3_{1-21}K4et_3$ **3.1** (10 µM), $H3_{1-21}K4pip$ **3.8** (10 µM) or $H3_{1-21}K4me_3$ **3.2** inhibitor (10 µM) (Fig. 3.5.17 B). Interestingly, $H3_{1-21}K4et_3$ **3.1** showed limited inhibition (~29%) at 10 µM while $H3_{1-21}K4pip$ **3.8** resulted in ~41% inhibition. As control experiment $H3_{1-21}K4me_3$ **3.2** showed ~59% inhibition suggesting that N^{ε}-trimethyl-4-lysine might occupy the amine oxidase active site.

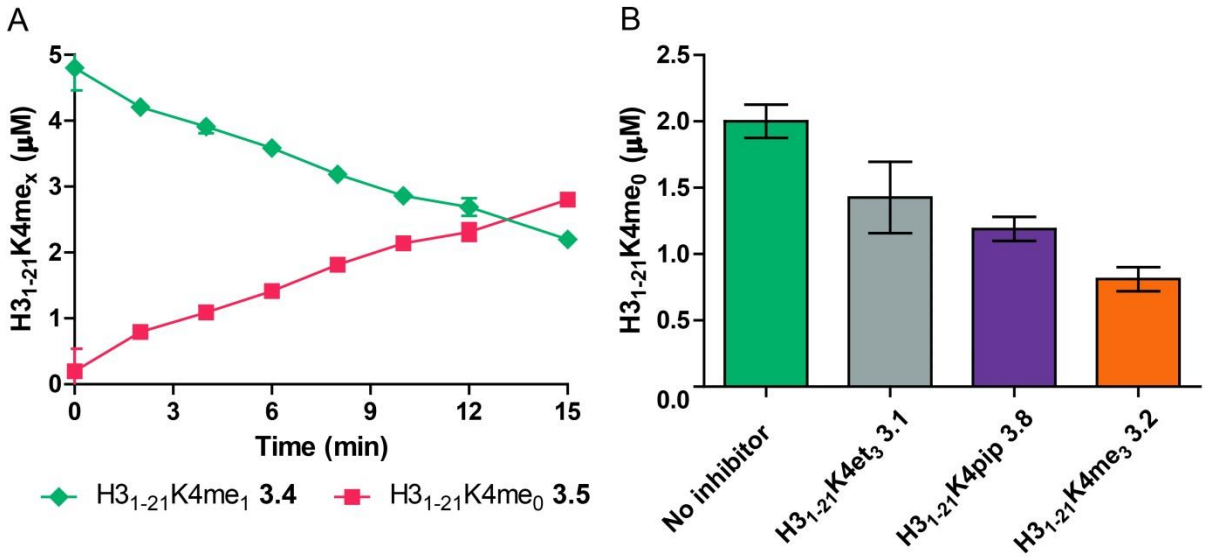


Figure 3.5.17: Linear region of $H3_{1-21}K4me_1$ 3.4 (5.0 µM) demethylation by $KDM1A_{FL}$ (A, R^2: 0.97) and single point inhibition of $KDM1A_{FL}$ (0.15 µM) with substrate $H3_{1-21}K4me_1$ 3.4 (5.0 µM) by $H3_{1-21}K4et_3$ 3.1 (grey), $H3_{1-21}K4pip$ 3.8 (purple) or $H3_{1-21}K4me_3$ 3.2 (10 µM) or in the absence of inhibitor (green). Measured using MALDI-TOF. n≥3 error given as s.d.

3.5.10 MALDI-TOF measurements to determine the inhibition of $KDM5A_{C2}$ demethylation by $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$

Next, the effect of single domain constructs of the PHD-fingers on KDM5A catalytic activity was explored. This is disease relevant, as PHD-finger translocations and fusion oncoproteins have been reported. For example, NUP98 is known to form a fusion partner with $KDM5A_{PHD3}$ creating oncoprotein NUP98-$KDM5A_{PHD3}$ that is linked to Acute Megakaryoblastic Leukaemia (AMKL) (described in section 3.2.2). To simulate *in vitro* conditions of (highly) expressed NUP98-$KDM5A_{PHD3}$, $KDM5A_{PHD3}$ was tested for its effect on $KDM5A_{C2}$ substrate demethylation. Providing an estimated potency of inhibition. $H3_{1-21}K4me_3$ **3.2** (5.0 µM) was pre-incubated with a 2-fold serial dilution series of $KDM5A_{PHD3}$ (100 µM top concentration) in the presence of co-factors, followed by the addition of $KDM5A_{C2}$ (0.5 µM) to start the reaction. The reaction was quenched (12 min) and the demethylation products were measured using MALDI-TOF. Results of demethylation were represented as the total number of turnover (TNT), multiple demethylation products were obtained ($H3_{1-21}K4me_2$ **3.3** and $H3_{1-21}K4me_1$ **3.4)** (Fig. 3.5.18). $KDM5A_{PHD3}$ was found to reduce the level of

demethylation, at pIC_{50} 5.42 ± 0.03 (IC_{50}: 3.80 μM), which is close to the theoretical IC_{50} limit of 2.5 μM (½ [$H3_{1-21}K4me_3$ **3.2**]). To determine how the inhibition of $KDM5A_{PHD3}$ compares relative to untagged PHD3-finger and the PHD1-finger, $KDM5A_{PHD3C2}$ and $KDM5A_{PHD1}$ were also tested. $KDM5A_{PHD3C2}$ (E1601-S1690) does not have a tag, unlike $KDM5A_{PHD3}$ (A1542-K1660) and was found to have a pIC_{50} of 5.63 ± 0.02 (IC_{50}: 2.34 μM) which is comparable to findings for $KDM5A_{PHD3}$. Furthermore, $KDM5A_{PHD1}$ was found to inhibit significantly weaker with a pIC_{50} of 4.51 ± 0.06 (IC_{50}: 30.9 μM).

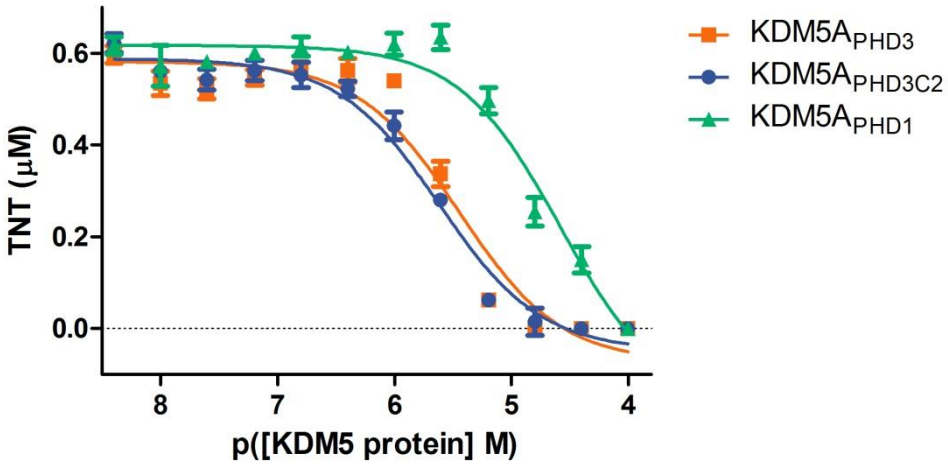


Figure 3.5.18: Inhibition of demethylase of $KDM5A_{C2}$ substrate $H3_{1-21}K4me_3$ 3.2 by $KDM5A_{PHD3}$, $KDM5A_{PHD3b}$ and $KDM5A_{PHD1}$ reader domains as determined using MALDI-TOF. TNT: total number of turnover. n=3, error given as s.d.

3.5.11 Exploring the KDM5 domain boundaries

Current literature remains ambiguous about the domain boundaries for the KDM5s and no universal assessment for all four proteins have been proposed.[268,292] To give an accurate estimation for all four KDM5 proteins, the domain boundaries were identified using a combination of Basic Local Alignment Search Tool (BLAST) and Protein Homology / analogY Recognition Engine (Phyre) 2. BLAST is a software algorithm (blastp (protein-protein BLAST)) provided by the National Centre of Biotechnology Information (NCBI) that operates by aligning a given sequence to previously identified protein sequences in the database.[293] Sequence matching allows for the identification of domains and reveals similar primary amino acids sequences. This could ultimately show potentially key conserved

residues (*e.g.* catalytic / recognition motifs), and elucidate possible related functions and evolutionary origin.[294] Potential sequence alignments identified are scored by a level of identity and similarity; frequently only partial sequence matches to other domains or proteins are found. KDM5 sequences were submitted to the Phyre 2 program for 3D structural predictions. Phyre 2 is a bioinformatics tool that renders 3D structures of a given sequence based on sequence homology to previously established protein structures.[295] With combined data sets in hand, domains lengths were defined based on their sequence homology to other proteins (BLAST) and observations made from secondary structural motifs (Phyre2 3D model). The result is displayed in figure 3.2.2 and Table 9.1.

3.5.12 Putative role of KDM5A-PHD2

Literature reports have identified the binding motif for KDM5A-PHD3, KDM5A-PHD1 and the KDM5A-JmjC. What is not yet clear is the potential binding sequence and function of KDM5A-PHD2. If a probe is designed for KDM5A-PHD3, could it interact with KDM5A-PHD2 as well? To gain insight into the potential function of PHD2, sequence alignment of PHD2 was performed using a combination of BLAST followed by Clustal Omega (Fig. 3.6.4).[293,296] Initial sequence comparisons identified alignment to other KDM5 proteins of *Homo sapiens* and related species however, strong correlation was also found with KDM7 PHD fingers. The KDM7A (KIAA1718), KDM7B (PHF8) and PHF2 zinc fingers are known to be potent (KDM7A K_d: 0.3 µM, KDM7B K_d: 0.95 µM[167]) binders of histone H3 N^{ε}-trimethyl-lysine (H3K4me$_3$), and contain an aromatic binding 'cage' formed by two conserved tyrosines, a methionine and a tryptophan.[167,198] Sequence alignment of KDM5A-PHD2 to KDM7A-C-PHD show that tyrosine (Y7 of KDM7B) is a phenylalanine (F1163 of KDM5A), a tyrosine (Y14 of KDM7B) is deleted but methionine (M20 of KDM7B and M1174 of KDM5A) and tryptophan (W29 of KDM7B and W1183 of KDM5A) are conserved (Fig. 3.5.19). Interestingly, other chromatin binding modules that have the high binding affinities for N^{ε}-trimethyl-lysine namely: ING4,

YNG1 or RAG2, contain a tryptophan, a methionine and a tyrosine.[297–299] This suggests that KDM5A-PHD2 might be a *N*ε-trimethyl-lysine binder with currently unknown sequence recognition.

```
KDM5A-PHD2  ----------------------------------------VKFCICRK--TASGFMLQCE    1178
KDM7A-PHD   MAGAAAAVAAGAAAGAAAAAVSVA----APGRASAPPPPPPVYCVCRQPYDVNRFMIECD    56
KDM7B-PHD   MNRSRAIVQRGRVLPPPAPLDTTNLAGRRTLQGRAKMASVPVYCLCRLPYDVTRFMIECD    60
KDM7C-PHD   -----------------------------------MATVPVYCVCRLPYDVTRFMIECD    24
                                                  :*:**    .. **::*:

KDM5A-PHD2  LCKDWFHNSCVPLPKSSSQKKGSSWQAKEVKFLCPLCMRSRR----------    1220
KDM7A-PHD   ICKDWFHGSCVGVEEHHAVD--------IDLYHCPNCAVLHGSSLMKKRRNW    100
KDM7B-PHD   MCQDWFHGSCVGVEEEKAAD--------IDLYHCPNCEVLHGPSIMKK----    100
KDM7C-PHD   ACKDWFHGSCVGVEEEEAPD--------IDIYHCPNCEKTHGKSTL------    62
             *:****.*** : :  : .          : ** *   :
```

Figure 3.5.19: Sequence alignment of KDM5A-PHD2 with KDM7A, KDM7B and KDM7C PHD fingers using Clustal Omega. Columns annotated with: '*' are identical amino acids, ':' are very similar amino acids, '.'are moderately comparable amino acids, unannotated amino acids which do not have comparable properties and '-' are gaps in the sequence alignment. Highlighted amino acids are residues in the aromatic cage binding the *N*ε-trimethyl-lysine of the substrate of KDM7 (based on structure KDM7B (PDB: 3KV4)). Highlighted in yellow: minor changes in amino acid, red: deletion and green: identical amino acids between KDM5A-PHD2 and KDM7A/B/C-PHD.

To generate a model structure of KDM5A-PHD2, Phyre2 was employed. KDM5A-PHD2 (F1163-C1215) was modelled with 99.5% confidence covering 92% of the sequence (Fig. 3.5.20). In the program PyMOL, structure alignment was achieved using SUPER[300] against KDM7B (PDB: 3KV4) demonstrating a very similar structure with β-strand and α-helix overlay.

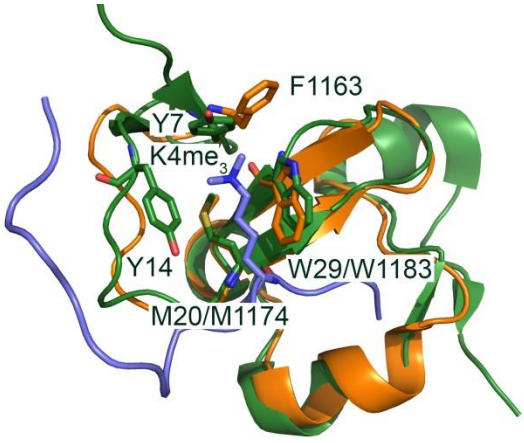


Figure 3.5.20: Phyre2 generated structure of KDM5A-PHD2 (orange) aligned to {KDM7B (green) (PDB: 3KV4) with H3K4me$_3$ bound (Blue)}. Structure alignment made using PyMOL applying the function Super. Atoms coloured in red: oxygen, blue: nitrogen, golden/yellow: sulphur and grey: zinc. Amino acid labels 7–29 refer to KDM7B while amino acids 1163-1183 refer to KDM5A-PHD2.

Additionally, the aromatic residues are observed on the same protein as determined with the KDM7B crystal structure. The results of Phyre2 model support the hypothesis that KDM5A-PHD2 is a putative N^{ε}-trimethyl-lysine binder. A potential shortcoming of this method is that the PHD2 domain Phyre 2 structure might be (partially) built based on KDM7-PHD fingers as these resemble close sequence alignment to KDM5A-PHD2. A more accurate structure may be generated if the build is established from the global pool of PHD-finger structures providing an orthogonal method for sequence alignment in combination with PyMOL Super alignment. To gain a higher degree of confidence in the KDM5A-PHD2 structure, some experimentally obtained structural data is required such as ^{1}H, ^{13}C, ^{15}N NMR or crystallography.

3.6 Discussion

In this Chapter, the PHD finger binding / recognition of KDM5A were explored. A series of lysine analogues at H3K4 sites were synthesized and characterised to understand the binding affinity and catalytic turnover of KDM5 proteins. ITC and AlphaScreen™ assays demonstrated that $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** follow a similar binding pattern as for $H3_{1-21}K4me_3$ **3.2** relative to $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$. Namely, tighter binding to $KDM5A_{PHD3}$ (K_d [ITC] 89–117 nM) than $KDM5A_{PHD1}$ (K_d [ITC]: 2.3–8.3 μM).

Given the correlations found between using ITC and AlphaScreen™, it is likely that $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** may have a similar binding configuration to the PHD3-finger relative to $H3_{1-21}K4me_3$ **3.2**. A decrease in binding strength was found for $H3_{1-21}K4me_{3-0}$ **3.2–3.5** for lower methylation states to $KDM5A_{PHD3}$ while an increase in binding strength is observed for fewer methylation states against $KDM5A_{PHD1,}$ consistent with literature observations.[8,166] $H3_{1-21}K4Ac$ **3.6** was observed to be poor binder for both readers tested. The N^{ε}-lysine 4 side chain was essential in the binding interactions in the hydrophobic aromatic cage in PHD3, as $H3_{1-21}K4A$ **3.7** showed a >100-fold decrease in K_d. A surprising observation for $H3_{1-21}K4A$ **3.7** was that it forms a stronger binding complex with $KDM5A_{PHD1}$ than the natural substrate $H3_{1-21}K4me_0$ **3.5**. These findings suggest that steric hindrance and the absence of hydrogen bonding of enlarged N^{ε}-Lysine $H3_{1-21}K4me_3$ **3.2**, $H3_{1-21}K4et_3$ **3.1** may play an important role in differentiation between lower and higher methylation levels.

The ITC observations indicate enthalpy-entropy compensation but the exact reason for this phenomenon remains speculative. One possibility is that $H3_{1-21}K4et_3$ **3.1** can, to some extent, accommodate the electron cloud towards the di-tryptophan aromatic and hydrophobic cage of $KDM5A_{PHD3}$ measured against $H3_{1-21}K4me_3$ providing weaker cation-π interactions. On the contrary,

a weaker interaction with the aromatic cage allows for a higher degree of freedom of the lysine 4 side chain which is entropically favourable.[244]

$H3_{1-21}K4et_3$ **3.1** was found to be a poor substrate for the JmjC-domain of KDM5 proteins tested. In the presence of high concentrations of $KDM5C_{C1}$ (1.25 μM) or $KDM5D_{C1}$ (1.25 μM) and prolonged incubation, $H3_{1-21}K4et_3$ **3.1** (5.0 μM) showed traces (+16 Da peak) of potential C^{β}-ethyl hydroxylation. When tested as an inhibitor, $H3_{1-21}Ket_3$ **3.1** (10 μM) did not inhibit the demethylation activity of KDM5 proteins toward $H3_{1-21}K4me_3$ under the tested conditions. This suggests that $H3_{1-21}K4et_3$ **3.1** peptide is not an effective substrate-competitive inhibitor of the JmjC-domain.

Interestingly, $H3_{1-21}K4pip$ **3.8** can function as a substrate of KDM5 proteins. The most likely site of hydroxylation is the C^{β} or C^{γ} of the piperidine ring; however, the efficiency of this reaction is KDM5 protein dependent. $KDM5C_{C1}$ (K_{cat}: 1.26 min^{-1}) and $KDM5D_{C1}$ (K_{cat}: 1.46 min^{-1}) show high rates of hydroxylation but KDM5B (K_{cat}: 0.31 min^{-1}) and $KDM5A_{C2}$ (K_{cat}: 0.13 min^{-1}) in particular show a lower catalytic efficiency (K_{cat}) which inversely proportionally correlates to the K_M observed. Given that the peptides possess only a single site of modification and the KDM5s tested are of similar size it is likely that the active site, which contains the ferrous metal, is of critical importance. To shed light on this observation, KDM5 JmjC-domain alignment was performed using Clustal Omega (Fig. 3.6.1).

```
5B      ARDYTLRTFGEMADAFKSDYFNMPVHMVPTELVEKEFWRLVSTIEEDVTVEYGADIASKE    469
5A      -----LQSFGEMADNFKSDYFNMPVHMVPTELVEKEFWRLVSSIEEDVIVEYGADISSKD    417
5C      -----LQSFGEMADSFKADYFNMPVHMVPTELVEKEFWRLVNSIEEDVTVEYGADIHSKE    448
5D      -----LQSFGEMADSFKSDYFNMPVHMVPTELVEKEFWRLVSSIEEDVTVEYGADIHSKE    438
             *::****** **:**********************.:***** ******* **:

5B      FGSGFPVRDGKIKLSPEE------------------------------EEYLDSGWNLN    498
5A      FGSGFPVKDGRRKILPEE------------------------------EEYALSGWNLN    446
5C      FGSGFPVSDSKRHLTPEE------------------------------EEYATSGWNLN    476
5D      FGSGFPVSNSKQNLSPEEKRQSLTVLTRLISSFWAQAVLPPWPPKVLGLQEYATSGWNLN    498
        ******* :.: :: ***                                 :**  ******

5B      NMPVMEQSVLAHITADICGMKLPWLYVGMCFSSFCWHIEDHWSYSINYLHWGEPKTWYGV    559
5A      NMPVLEQSVLAHINVDISGMKVPWLYVGMCFSSFCWHIEDHWSYSINYLHWGEPKTWYGV    506
5C      VMPVLEQSVLCHINADISGMKVPWLYVGMVFSAFCWHIEDHWSYSINYLHWGEPKTWYGV    536
5D      VMPVLDQSVLCHINADISGMKVPWLYVGMVFSAFCWHIEDHWSYSINYLHWGEPKTWYGV    558
         ***::****.**..**.***:******* **:**************************

5B      PGYAAEQLENVMKKLAPELFVSQPDLLHQLVTIMNPNTLMTHEVPVYRTNQCAGEFVITF    620
5A      PSHAAEQLEEVMRELAPELFESQPDLLHQLVTIMNPNVLMEHGVPVYRTNQCAGEFVVTF    566
5C      PSLAAEHLEEVMKKLTPELFDSQPDLLHQLVTLMNPNTLMSHGVPVVRTNQCAGEFVITF    596
5D      PSLAAEHLEEVMKMLTPELFDSQPDLLHQLVTLMNPNTLMSHGVPVVRTNQCAGEFVITF    628
        *. ***:**:**: *:**** ***********:****.** * *** **********:**

5B      PRAYHSGFNQGFNFAEAVNFCT---------------------------------------    640
5A      PRAYHSGFNQGYNFAEAVNF-----------------------------------------    596
5C      PRAYHSGFNQGYNFAEAVNF-----------------------------------------    617
5D      PRAYHSGFNQGYNFAEAVNFCTADWLPAGRQCIEHYRRLRRYCVFSHEELICKMAAFPET    678
        ***********:********
```

Figure 3.6.1: Sequence alignment of KDM5A-D JmjC-domain using Clustal omega. Highlighted are the Fe^{II} chelating residues (pink), conserved residues in KDM4A- KDM5 N^{ε}-trimethyl-lysine binding pocket (green), KDM4A vs. KDM5 trimethyl-lysine pocket mismatch (yellow), a KDM5A-B vs KDM5C-D mismatch (blue) and KDM5A/B and KDM5C-D mismatch (grey).

Alignment of these sequences revealed high sequence similarity among the KDM5 proteins with a sequence identity of ~70% and sequence similarity of ~82%. However, the available crystal structures of KDM5 JmjC-domains do not contain a bound histone H3 substrate, making it difficult to assign of amino acids that direct the cause of catalytic differentiation among the KDM5s. KDM5s are known to be similar in sequence to the KDM4s[301] and therefore KDM5B and KDM4A JmjC-domains were aligned (Fig. 3.6.2). KDM4A crystal structure was used to assign residues that interact with the N^{ε}-trimethyl-lysine 9 (Fig. 3.6.3, PDB: 2OQ6). Using both data set allowed for the prediction of residues that KDM5 proteins may use to interact with N^{ε}-trimethyl-lysine 4.

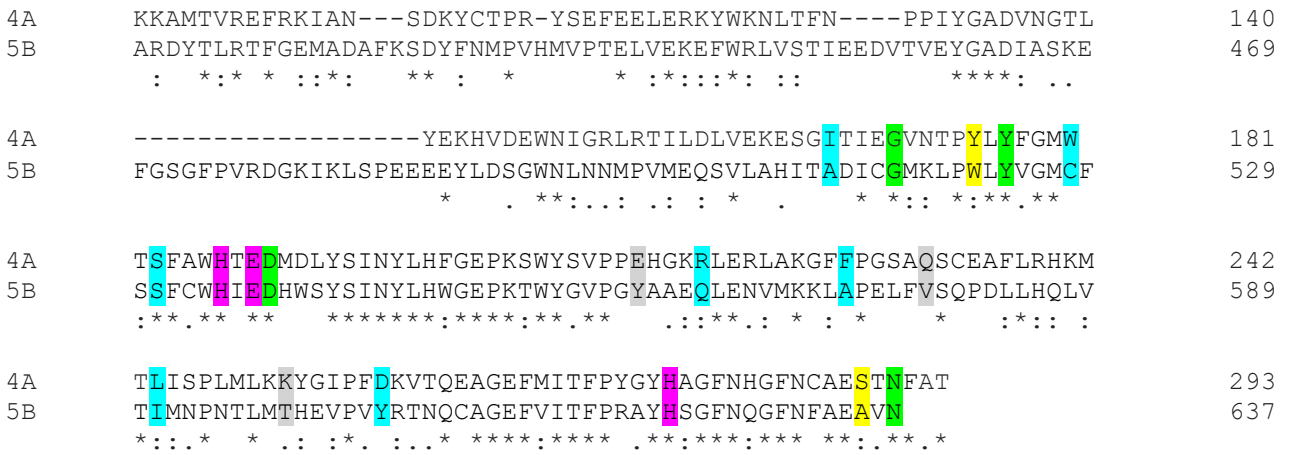


Figure 3.6.2: Alignment of KDM4A and KDM5B JmjC-domain sequences provided using clustal omega. Highlighted are the Fe^{II} chelating residues (pink), KDM4A- KDM5 N^{ε}-trimethyl-lysine pocket match (green), KDM4A vs. KDM5 trimethyl-lysine pocket mismatch (yellow), a KDM5A-B vs KDM5C-D mismatch (cyan) and KDM5A/B and KDM5C-D mismatch (grey).

All five proteins have identical Fe^{II} chelating residues, via two conserved histidine residues and a single glutamate residue. Furthermore, N^{ε}-trimethyl-lysine 9 in the KDM4A pocket is surrounded by six amino acids (G170, Y175, Y177, D191, S288, N290), four of which are identical and two of similar sterical or aromatic properties to KDM5B sequence. These six amino acid residues within the KDM4 active site (assigned using the KDM4 crystal structure), are identical across the KDM5s. It is therefore likely that the nearby amino acid residues surrounding the N^{ε}-trimethyl-lysine 9 are not directly responsible for the observed difference in binding and catalytic efficiency among KDM5s for $H3_{1-21}$K4pip **3.8** substrate.

Considering the ability of KDM5s to hydroxylate $H3_{1-21}$K4pip **3.8**, it is surprising that in a competitive assay of $H3_{1-21}K4me_3$ **3.2** (5.0 μM) and $H3_{1-21}$K4pip **3.8** (10 μM) (at maximum), trace levels of hydroxylation are observed. This suggests $H3_{1-21}$K4pip **3.8** to be significantly poor substrate relative to $H3_{1-21}K4me_3$ **3.2**. Due to the relatively low K_M of $H3_{1-21}K4me_3$ **3.2** towards the KDM5 proteins a direct comparison was not possible. In a separate $KDM1A_{FL}$ competition assay, $H3_{1-21}$K4pip **3.8** and $H3_{1-21}K4me_3$ **3.2** were found to have some inhibitor capabilities but it was not a substrate.

Piperidine and its derivatives are commonly found scaffolds in many pharmaceutical drugs.[302] In the field of epigenetic drug discovery, nitrogen-containing saturated heterocycles (*e.g.* piperidine, piperazine, pyrolydine) are often used as a lysine mimic, with notable examples such as methyltransferase inhibitors such as GSK126, GSK343 and EPZ005687.[303] Given the observations made here, it is important to consider that these piperidine-containing inhibitors that mimic histones may also be modified by KDM5 proteins. This was originally considered to be unlikely as previous studies showed that KDM2A, KDM4E and KDM7B were unable to catalytically hydroxylate N^{ε}-diethyl-lysine substrate, a derivative of N^{ε}-piperidine-lysine.[283] However, literature reports later indicated that KDM6B is capable of doing this.[284] While this study revealed that KDM5s cannot hydroxylate N^{ε}-triethyl-lysine, +16 Da peaks were observed for N^{ε}-piperidine-lysine which may suggest conformational constrain may aid in optimal substrate positioning and enhance catalytic turnover. However, further work is needed to identify the site of oxidation.

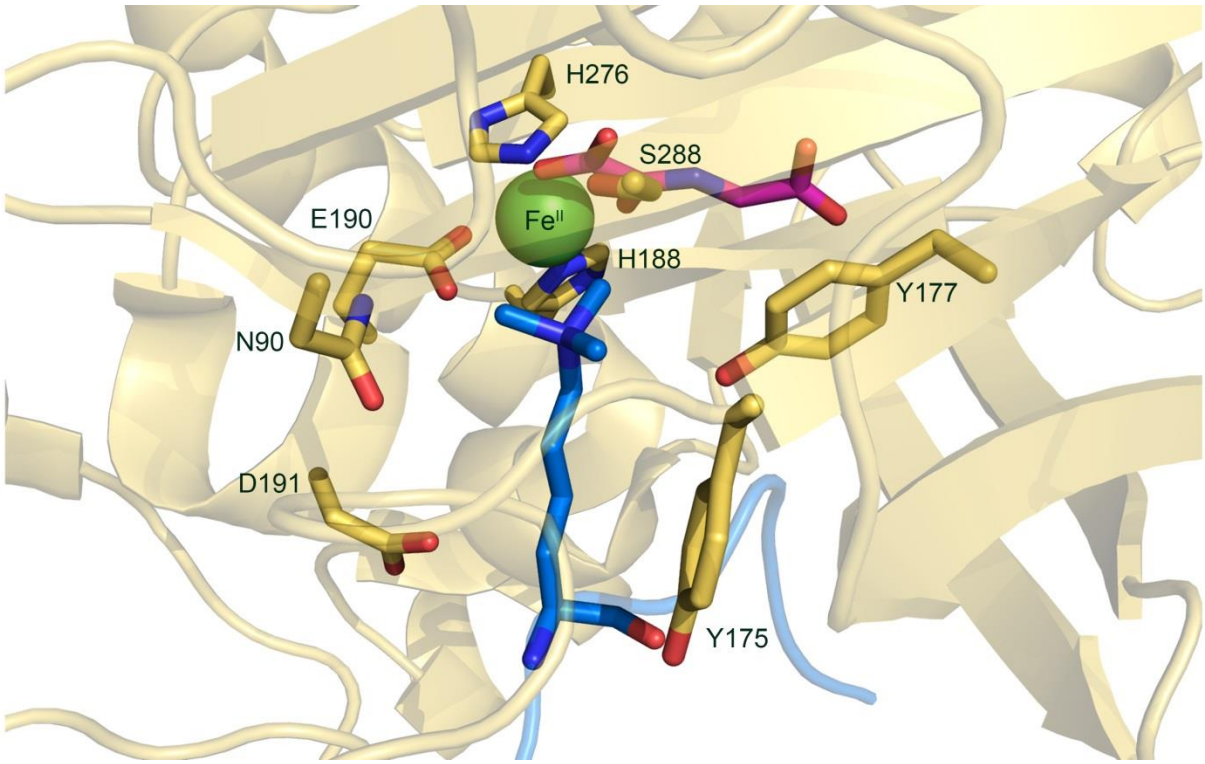


Figure 3.6.3: Crystal structure of KDM4A (yellow) visualizing the catalytic active site containing Fe^{II} (green) and 2OG (pink) highlighting residues interacting with N^{ε}-trimethyl-lysine 9 (light blue). Atoms oxygen (red) and nitrogen (blue) are colour coded.

The KDM5A domain complexity feeds the debate on the mechanism of action of KDM5A on the chromatin. The protein could, for example, bind nucleosomal DNA via the ARID-domain while binding one or more histone H3 tails. This could either be with the JmjC-domain and / or PHD3 in the case of higher methylation states histone H3 tail or with KDM5A$_{PHD1}$ when methylation is absent. Determining whether KDM5A PHD3 finger has a role in regulating JmjC activity or exclusively functions as a passive read out of nucleosome remains challenging to study in an *in vitro* setting. A complicating factor is that both the JmjC-catalysed demethylation substrate (H3$_{1\text{-}21}$K4me$_3$ **3.2**) as well as the potential PHD3 activator (H3$_{1\text{-}21}$K4et$_3$ **3.1**) can compete for binding to the KDM5A-PHD3 finger. In other words, if acceleration or deceleration takes place, this can be initiated by the JmjC substrate H3$_{1\text{-}21}$K4me$_3$ **3.2**. A potential method of investigating this is to study the demethylation rate of histone H3 in intact intra-nucleosome containing H3^{A}K4me$_3$ and H3^{B}K4et$_3$ or to H3^{A}K4me$_3$ and H3^{B}K4me$_0$.

The inhibition of H3K4me$_3$ **3.2** substrate for KDM5A$_{C2}$ protein by the KDM5A$_{PHD3}$ finger is of significant importance. Elevated levels of KDM5A$_{PHD3}$ due to the expression of fusion protein NUP98-KDM5A$_{PHD3}$ could orchestrate the rise of higher methylation states on H3K4. The direct effect of this however, remains to be elucidated. It is possible that since H3K4me$_3$ is associated with gene activation and could potentially enhance levels of transcription of genes, binding of increased levels of KDM5A$_{PHD3}$ might prevent the binding of (apart from KDM5 demethylases), other proteins including transcription factors such as sub unit TAF3. This would block potential transcription from taking place. KDM5A is not the only protein to form a fusion partner with NUP98, and other examples with PHF23 and BPTF have been reported.[304,305] This observation may function as a general proof of concept that overexpression of chromatin binding modules could lead to KDM5 inhibition through substrate sequestration.

Note that as $KDM5A_{PHD3}$ is a strong $H3K4me_3$ binder that likely functions as a substrate inhibitor, preventing $KDM5A_{C2}$ demethylation and it is therefore unlikely that $KDM5A_{PHD3}$ classifies among the classical inhibitor types as competitive, non-competitive, or uncompetitive protein inhibitor.

3.7 Conclusion

Work in this Chapter describes the use of $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** to study the KDM5A-PHD3 finger. ITC and AlphaScreen™ observations revealed that $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** have a similar binding affinity to $KDM5A_{PHD3}$ as $H3_{1-21}K4me_3$ **3.2**. On the contrary, $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** were determined to be weaker binders (~80-fold) relative to $H3_{1-21}K4me_0$ **3.5** for $KDM5A_{PHD1}$.

Observations made using the FDH assay revealed that KDM5 constructs produced were catalytically active. Steady state enzyme kinetics assays for KDM5A show that the PHD1 and the ARID domains influence JmjC catalytic activity, with KDM5A-JmjC demonstrating a higher K_M (~2.1-fold) in their absence but a faster K_{cat} (~2.2-fold) with $H3_{1-21}K4me_3$ **3.2** as a substrate. Parameters for constructs KDM5A-D reveal a generally lower rate of demethylation and a lower K_M relative to previous observations.[268] MALDI-TOF mass spectrometry studies demonstrated that $H3_{1-21}K4et_3$ **3.1** is a poor substrate for KDM5 constructs. Under elevated enzyme concentrations (*e.g.* 1.25 μM) of KDM5C and KDM5D, only a trace (<5%) peak of +16 Da mass shift was observed that may indicate C^{β}-ethyl-N^{ε}-diethyl-lysine hydroxylation. $H3_{1-21}K4pip$ **3.8** was found to be a substrate for KDM5A-D protein indicating hydroxylation (+16 Da) at the C^{β} or C^{γ} of the N^{ε}-piperidine ring determined using MALDI-TOF and 1-D ^{1}H NMR spectroscopy. Interestingly, KDM5C and KDM5D showed a significant higher K_{cat} (~10 fold greater) than was found for KDM5A. The increase in K_{cat} was observed to be inversely proportional correlated to the K_M with substrate K_M values for KDM5C and KDM5D being ~10-fold lower than for KDM5A. KDM5B protein was found to behave in between $KDM5A_{C2}$ and KDM5C/D containing a K_M that is ~2.4-fold lower but a K_{cat} that is ~2.4-fold higher than $KDM5A_{C2}$. It remains unknown if this is caused by individual residues or secondary protein structure formation. It is important to note that kinetic parameters of KDM5A-D using $H3_{1-21}K4pip$ **3.8** do suggest that the JmjC catalytic site is likely to be structurally different across these constructs.

The peptides $H3_{1-21}K4et_3$ **3.1** and $H3_{1-21}K4pip$ **3.8** were examined for their inhibitory properties against KDM5A-D constructs containing the JmjC-domain. Collected observations do not indicate that the PHD3 finger of KDM5A acts as an allosteric inhibitor or activator of the JmjC-domain. The data show that little inhibition (<20%) was found at 10 μM inhibitor peptide and 5.0 μM $H3_{1-21}K4me_3$ **3.2** substrate.

MALDI-TOF activity screening for $KDM5A_{C2}$ catalytic inhibition by $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$ demonstrate considerate amounts of inhibition, suggesting that aberrantly high levels of chromatin binding modules may disrupt cellular functions.

3.8 Future perspective

The *in vitro* experiments remain inconclusive about potential allosteric activation or inhibition of the JmjC-domain by binding the PHD3 finger using $H3_{1-21}K4me_3$ **3.2**, $H3_{1-21}K4et_3$ **3.1** or $H3_{1-21}$K4pip **3.8**. A potential method to investigate this is to produce two nucleosomal complexes with heteromeric histone H3. One nucleosome would contain [$H3^A$-$K4me_3$, $H3^B$-$K4et_3$] while the other would be [$H3^A$-$K4me_3$, $H3^B$-$K4me_0$]. This would allow KDM5A to bind to the nucleosome and interact with $H3^B$ through the PHD3 finger and demethylate $H3^A$ the JmjC-domain. This system could be used to investigate the effect of modifications on $H3^B$K4 on the rate of demethylation of $H3^A$K4.

N$^{\varepsilon}$-triethyl-Lysine proved to be a useful tool in discriminating between the PHD3 and the JmjC-domain suggesting that *N*$^{\varepsilon}$-triethyl-Lysine can be implemented in inhibitors targeting chromatin binding modules. Furthermore, the use of $H3_{1-21}$K4pip **3.8** demonstrated that the KDM5 proteins may contain differences in their structures that could be exploited for the development of intra-KDM5 specific JmjC-domain inhibitors.

3.9 Materials and Methods

3.9.1 General Experimental conditions

General experimental conditions are described in Chapter 7.1.1.

3.9.2 ITC analysis

ITC measurements were taken using automated Malvern GE ITC PEAQ with histone mimic peptide (135 µL / sample, 300 µM) titrated to the sample cell contain the protein KDM5A (385 µL / sample, 30 µM). Samples were prepared in buffer [HEPES (50 mM), NaCl (50 mM), TCEP (0.5 mM) in MQ (pH 7.5)] and placed in a deep well 96-well plate (Fisher Brand™ 96-Well DeepWell™, Polypropylene Microplates, Non-Sterile, 12-566-611) prior to analysis. Automated ITC acquisitions were according to manufacturer's set program 'plates pre-rinse syringe clean' using solution Decon90 (14% in MilliQ) and Methanol (HPLC grade). The ITC instrument was controlled computer (MS windows 7) using MicroCal PEAQ-ITC Automated control software (v 1.21). Analysis of the raw acquired data was performed using MicroCal PEAQ-ITC Analysis Software (v 1.22).

Instrument settings used are as follows:

Reference power: 10 µcal·sec^{-1}; 19 injections, 1: 0.4 µL, 2–19: 2.0 µL at 2.0 sec·µL^{-1}; stirring speed: 750 rpm; initial delay: 60 sec; delay between injection: 150 sec; cell feedback: high; cell temperature: 25 °C; tray temperature: 15 °C; protein cell fill: 280 µL.

3.9.3 AlphaScreen™ displacement

AlphaScreen™ experiments were carried to assess the displacement of histone mimic peptides from KDM5A$_{PHD3}$ and KDM5A$_{PHD1}$. The peptide used for KDM5A$_{PHD3}$ was H3$_{1-21}$K4me$_3$–[Biotin] (ART

Kme_3QT ARK STG GKA PRK QLA GGK(Biotin)-NH_2) (ANASPEC Peptide) and for $KDM5A_{PHD1}$ was used: $H3_{1-21}$–[Biotin] (ART KQT ARK STG GKA PRK QLA GGK(Biotin)-NH_2) (ANASPEC Peptide, Purity: 97%). Peptides, proteins and AlphaScreen™ beads (AlphaScreen™ Histidine Detection Kit (Nickel Chelate) (6760619, lot #: 2457886)) were prepared in buffer [HEPES (50 mM), NaCl (150 mM), BSA (0.1% (Perkin Elmer, BSA (DTPA purified), 7.5% (Cat: CR84-100)), Tween20 (0.01%, Tween®20, Promega (Ref#: H5151)) in MQ (pH 7.5)]. Analysis was performed on AlphaScreen AlphaPlate™ -384 SW (part#: 6008350) plates and multichannel liquid dispensing was carried out using ThermoScientific E1-ClipTip. All measurements were taken with Perkin Elmer EnVision 2104 multilabel reader controlled using Wallac EnVision Manager (v. 1.12). Peptide displacement plates were prepared from 10 mM Peptide stocks and dispensed using Echo Acoustic dispenser (Labcyte, Echo 550). Obtained data were processed using Microsoft Excel (2010) and GraphPad Prism (v. 5.04).

Optimization of AlphaScreen™ conditions

To test the linearity of the AlphaScreen™ beads, KDM5A reader protein and peptide were prepared at 4× final concentration and a 6-step 2-fold serial diluted for 6× to provide an 8×8 matrix with the final concentrations (nM) as shown in figure 3.9.1. KDM5A reader protein at the appropriate concentration (5.0 µL) was added to buffer (5.0 µL) and biotinylated peptide (5.0 µL) and incubated for 30 min. To the solution was added to AlphaScreen™ beads in buffer (62.5× dilution, 5.0 µL, final assay volume: 20 µL), incubated (30–45 min) and the plate was read by Perkin Elmer EnVision. Obtained data were analysed and revealed that $KDM5A_{PHD3}$ and $KDM5A_{PHD1}$ concentrations are within the linear region at 1.563 nM protein and 3.125 nM biotinylated peptide.

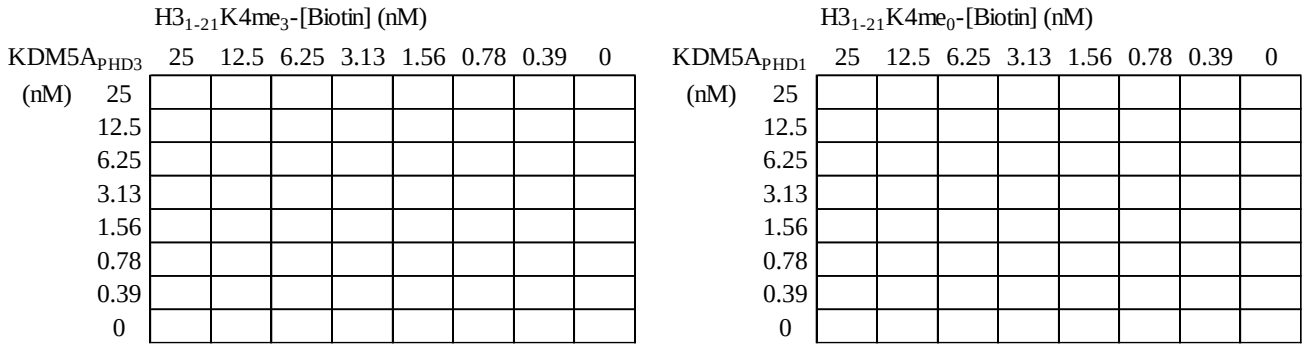


Figure 3.9.1: schematic representation of 8×8 matrix for KDM5APHD3/1 AlphaScreen™ displacement assays.

AlphaScreen™ Peptide displacement

AlphaPlates™ containing increasing concentration of histone-mimic peptides were prepared using Echo dispenser from source stock (10 mM in MQ, 2.5 nL increments) in 384- low dead volume (LDV) plates (LabCyte, cat#: LP-0200 source plate. Dilutions of the source stock were made in MQ where necessary. The plate was left to evaporate to dryness upon air (~15 min) prior to commencing the experiment. To the plate was added buffer (5.0 µL) and KDM5A$_{PHD3}$ or KDM5A$_{PHD1}$ protein (1.563 nM in buffer, 5.0 µL) and incubated (15 min) followed by the addition of biotinylated peptide (3.125 nM in MQ, 5.0 µL) and incubation (30 min). The AlphaScreen™ beads were added (62.5× dilution, 5.0 µL) and the mixture was incubated (30 min), scanned and the acquired data was fitted model 'log(inhibitor) vs. normalized response –variable slope' using GraphPad Prism (v. 5.04)

3.9.4 MALDI-TOF assay

MALDI-TOF assay conditions were adjusted and optimized depending on the protein and the type of assays conducted. Experiments were performed on AlphaScreen AlphaPlate™ -384 SW (part#: 6008350) and reagents were dispensed using multichannel liquid using ThermoScientific E1-ClipTip. Plates for dose response experiments were prepared using Echo Acoustic dispenser (Labcyte, Echo 550) with peptides that were synthesized, purified and standardized against (sodium 3-(trimethylsilyl)-

2,2,3,3-d_4 propionate (TSP) in D_2O using ^1H NMR on Bruker AVIII 700 (5 mm $^1H(^{13}C/^{15}N)$ helium cryoprobe). For peptide preparation see Chapter 7.1.3. Reactions were quenched (unless stated otherwise) with formic acid (FA) (5.0 µL, 1% in MQ) and samples (1.0 µL) were co-spotted with MALDI matrix solution [α-Cyano-4-hydroxy cinnamic acid (HCCA, Sigma-Aldrich (product #: 70990-1G-F)) (saturated sol. HCCA in 50:50:0.1 MQ ACN Trifluoro acetic acid (TFA))] (1.0 µL) on Bruker MTP 384 target plate polished steel BC (Part #: 8280781). Samples were measured using MALDI-TOF/TOF Bruker Daltonics AutoflexSpeed controlled by Compass for flex series (v 1.4) and Flex control (v. 3.4, build 135.10) in positive Reflectron mode. The mass analysis was performed on the MALDI-TOF and processed using Flex analysis (v. 3.4, build 78), Microsoft Excel (2010) and Graphpad Prism (v. 5.04).

Standard MALDI-TOF experiment

To a plate containing peptide mix [sodium ascorbate (2 mM), $(NH_4)_2Fe(II)(SO_4)_2$ (100 µM), 2-oxoglutarate (400 µM), $H3_{1\text{-}21}K4me_3$ **3.2** (10 µM), tris(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl) (10 or 5 mM) in buffer [HEPES (50 mM), NaCl (50 mM) MQ (pH 7.5)] (5.0 µL) was added [$KDM5A_{C1}$ protein (0.5 µM) in buffer] (5.0 µL), incubated at room temperature and quenched using formic acid in MQ (1.0 v/v% in MQ, 5.0 µL).

Peptides and proteins used with their corresponding concentrations are displayed in Table 3.9.1.

Table 3.9.1: typical protein and peptide concentrations in MALDI-TOF assay

	Protein		Peptide	
	Standard (µM)	High (µM)	Peptide substrate	Standard (µM)
$KDM5A_{C1}$	0.50	2.50	$H3_{1-21}K4me_3$ **2**	5.0
$KDM5A_{C2}$	0.50	2.50	$H3_{1-21}K4me_3$ **2**	5.0
$KDM5A_{FL}$	0.15	0.25	$H3_{1-21}K4me_3$ **2**	5.0
$KDM5B_{C1}$	0.25	1.25	$H3_{1-21}K4me_3$ **2**	5.0
$KDM5C_{C1}$	0.25	1.25	$H3_{1-21}K4me_3$ **2**	5.0
$KDM5D_{C1}$	0.25	1.25	$H3_{1-21}K4me_3$ **2**	5.0
$KDM1A_{FL}$	0.15	0.25	$H3_{1-21}K4me_1$ **4**	5.0

MALDI-TOF/TOF

A prepared peptide mix [sodium ascorbate (4.0 mM), $(NH_4)_2Fe(II)(SO_4)_2$ (200 µM), 2-oxoglutarate (4.0 mM), $H3_{1-21}K4pip$ **3.8** (1.0 mM), tris(2-carboxyethyl)phosphine hydrochloride (TCEP HCl) (5 mM) in buffer [HEPES (50 mM), NaCl (50 mM) MQ (pH 7.5)] (5.0 µL) was added to $KDM5C_{C1}$ protein (7.15 µM) in buffer (5.0 µL) or $KDM5D_{C1}$ protein (6.35 µM) in buffer (5.0 µL). The reaction mixture was incubated (16 hrs) and samples were diluted (100-fold in MQ, 5.0 µM final peptide concentration) and analysed using MALDI-TOF. Under these conditions, >95% of the peptide was converted to product (+16 Da, $H3_{1-21}K4pipOH$). To execute MALDI-TOF/TOF analysis, the reaction mixture was diluted (10-fold in MQ, 50 µM) and measured in positive reflectron mode MALDI-TOF/TOF mode.

MALDI-TOF $KDM5A_{PHD3}$ or $KDM5A_{PHD1}$ inhibition

A solution of $KDM5A_{PHD3}$ (100 µM), sodium ascorbate (2.0 mM), $(NH_4)_2Fe(II)(SO_4)_2$ (100 µM), disodium 2-oxo-glutarate (400 µM), $H3_{1-21}K4me_3$ **3.2** (10 µM), TCEP (5.0 mM) was prepared in buffer [HEPES (50 mM), NaCl (50 mM) in MilliQ (pH 7.5)] (5.0 µL) and 2.5-fold serial diluted to give a total of 12 dose response points. The reaction was initiated with $KDM5A_{C2}$ (1.0 µM, 5.0 µL) in buffer and quenched using formic acid solution (1 v/v% in MilliQ, 5.0 µL) after 12 min. Samples were measured as previously stated.

3.9.5 ^{1}H NMR spectroscopy to determine $H3_{1-21}$K4pip 3.8 hydroxylation site by KDM5 proteins

To a solution {$H3_{1-21}$K4pip **3.8** (2.0 mM), sodium ascorbate (4.0 mM), $(NH_4)_2Fe(II)(SO_4)_2$ (200 µM), 2OG (4.0 mM), TCEP·HCl (5.0 mM) in Buffer [HEPES (50 mM), NaCl (50 mM) in MQ (pH 7.5)} (125 µL) was added {$KDM5D_{C1}$ (12.7 µM) in buffer (HEPES (20 mM), NaCl(150 mM), glycerol (5 v/v%) in MQ (pH 7.5)} (125 µL), mixed using pipette and after 27 hrs the mixture was assed using MALDI-TOF. Peak integration revealed that >91% product was formed. The mixture was diluted with MilliQ (250 µL) and spin filtrated using (Amicon Ultra-0.5 30K, 14000 × g, 30 min). The filtrate was HPLC purified and selected fractions, as determined using MALDI-TOF, were pooled, lyophilized and dissolved in [(sodium phosphate (50 mM) in D_2O (pH 7.1) (128 µL), (TSP (10 mg·mL^{-1}) in D_2O, 4 µL) and D_2O (28 µL)]. The sample was measured on a Bruker ^{1}H NMR.

3.9.6 FDH assay

The FDH experimental conditions varied depending on the KDM examined. Peptides used were synthesized, purified and standardized (sodium 3-(trimethylsilyl)-2,2,3,3-d_4 propionate (TSP) in MQ, ^{1}H NMR on Bruker AVIII 700 (5 mm ^{1}H(^{13}C/^{15}N) helium cryoprobe). Samples were dispensed using ThermoScientific E1-ClipTip for multichannel pipetting in a black clear bottom plate [Microplate 384 well, Ps, µClear®, black, non-binding (Ref #:781906)] and measured on BMG Pherastar FX using optic module FI 280 350 (FI 1205A1) at a rate of 2 scans per minute. The instrument was processed using MARS (v 2.41) and data was analysed using a combination of MARS data analysis software, Microsoft Excel (2010) and Graphpad Prism (v. 5.04). Enzyme kinetics fitted using the function 'Michaelis-Menten' in GraphPad.

Standard FDH assay

A prepared solution containing [sodium ascorbate (2 mM), $(NH_4)_2Fe(II)(SO_4)_2$ (100 µM), 2-oxoglutarate (400 µM), $H3_{1-21}K4me_3$ **3.2** (10 µM), β-Nicotinamide adenine dinucleotide hydrate (β-NAD^+) (500 µM) in buffer (HEPES (50 mM), NaCl (50 mM), Tween20 (0.01%) MQ (pH 7.5))] (25.0 µL) was added to the plate followed by enzyme solution [$KDM5A_{C1}$ (0.50 µM), Formaldehyde dehydrogenase (2.0 µM) in buffer] (25.0 µL). The reaction was monitored using the BMG PHERAstar (25 min, scanning every 30 sec) and raw data was analysed as stated earlier.

Various concentrations of $KDM5A_{C1}$ (0.25 µM), $KDM5A_{C2}$ (0.50 µM), $KDM5A_{FL}$ (0.15 µM), $KDM5B_{C1}$ (0.25 µM), $KDM5C_{C1}$ (0.10 µM) and $KDM5D_{C1}$ (0.15 µM) were used dependent upon the activity. Steady state enzyme kinetics of the peptide was performed with variable $H3_{1-21}K4me_3$ **3.2** concentrations and 200 µM 2OG while kinetics of 2OG were determined using variable concentrations of 2OG and $H3_{1-21}K4me_3$ **3.2** at 5.0 µM ($KDM5A_{C1/C2}$) or 2.0 µM ($KDM5B_{C1}$ C_{C1} D_{C1}).

3.9.7 Fmoc-Amino acid and histone H3 mimic peptide synthesis

Amino acids and histone H3 mimic peptides were produced as described in Chapter 7.1.2 and Chapter 7.1.3, respectively.

3.9.8 Recombinant protein production

Protein production of $KDM5A_{PHD3}$

Plasmid encoding for A1542–K1660 of KDM5A in vector pGTVL2 was transformed into *E. coli* phage resistant BL21 DE3 strain carrying pRARE and the cells were grown on kanamycin (30 $\mu g \cdot mL^{-1}$) and chloroamphenicol (34 $mg \cdot L^{-1}$) plates. A single colony was used to initiate a starter culture (100 mL, 2TY media) and incubated overnight (200 rpm, 37 °C). Media (Terrific broth, Fisher Bioreagents™ (product #: 12891660)) was inoculated with starter of culture (1 mL L^{-1} TB) with

shaking (175 rpm) to an OD_{600} of ~0.8 followed by cooling (17 °C). The cells were induced with IPTG (0.2 mM) and grown overnight with shaking (175 rpm, 17 °C). The cells were harvested by centrifugation (11899 × g, 10 min) to obtain 89.3 g of pellet, which was stored (−80 °C).

Purification of the protein was performed at 4.0 °C (including suspending, column purification and centrifugation). To 20.9 gram of pellet was added Lysis buffer [100 mL, HEPES (50 mM), NaCl (500 mM), imidazole (10 mM) in MQ (pH 7.5), filtrated through 0.2 µm filter], DNAseI and protease inhibitor (Roche, EDTA free). Cells were thawed by stirring to create a smooth mixture (~ 45 min, 4 °C) followed by sonication (12 min, alternating between 2 sec on, 3 sec off). The mixture was centrifuged (48384 × g, 20 min), the supernatant decanted, filtered and loaded on to a Histrap™ HP column (5 mL (1 CV = 5 mL), 2 $mL \cdot min^{-1}$) attached to a FPLC. The Histrap was washed with washing buffer [10 CV, HEPES (50 mM), NaCl (500 mM), imidazole (40 mM) in MQ (pH 7.5), filtrated through 0.2 µm filter] and eluted in a gradient elution with elution buffer [10 CV, 0 to 100%, HEPES (50 mM), NaCl (500 mM), imidazole (250 mM) in MilliQ (pH 7.5), filtrated through 0.2 µm filter]. The fractions containing $KDM5A_{PHD3}$ (indicated by SDS-PAGE) were pooled, concentrated (Amicon® Ultra-15 30K, 4255 × g) and purified in 2 runs over Gel filtration (Superdex 75, 300 mL) with gel filtration buffer [HEPES (50 mM), NaCl (50 mM), in MQ (pH 7.5), filtrated (0.2 µm) and degassed] (20 mL Loading, 300 mL elution, 1.0 $mL \cdot min^{-1}$). The eluted fractions were visualised on SDS-PAGE (NuPAGE™ 4–12% Bis-tris gel, cat#: NP0335BOX) (Fig. 3.6.5.) stained with Instant*Blue*™ (expedeon, cat#: ISB1L) and $KDM5A_{PHD3}$ containing fractions were pooled and concentrated (Amicon® Ultra-15 30K, 4255 × g) obtaining 123.8 mg of protein as determined using ThermoScientific NanoDrop one (ext. coefficient: 64580, ProtParam by ExPASy).

Protein production of KDM5A$_{PHD1}$

KDM5A$_{PHD1}$ gene (K283–Q356) in vector pGTVL2, was purified as described in KDM5A$_{PHD3}$ with minor alterations. A total of 74.5 g of pellet was produced of which 22 grams was used for purification furnishing 94.7 mg of protein (ext. coefficient: 52830, ProtParam by ExPASy). Furthermore, to the Lysis buffer was added TCEP·HCl (0.5 mM).

Protein production of KDM5A$_{PHD3C2}$ *(produced by Dr. Catrine Johansson)*

Protein KDM5A$_{PHD3C2}$ was produced from the plasmid containing vector pGTVL2 and theKDM5A gene encoding for amino acid E1601–S1690 generously gifted by the SGC, Oxford. The plasmid was transformed into competent *E. coli* BL23 (DE3) phage resistant equipped with pRARE2 plasmid cell line. Cells were grown in Terrific Broth media containing chloramphenicol (34 mg·L^{-1}) and kanamycin (50 mg·L^{-1}) (37 °C) until an OD_{600}: 2.0 was reached. The cells were cooled (18 °C), induced with IPTG (0.1 mM) and grown over night. The cells were harvested by centrifugation (5000 × g, 15 min) and stored in the freezer (−80 °C). The cells were suspended in lysis buffer [Imidazole (10 mM), TCEP·HCl (0.5 mM), glycerol (5 v/v%), NaCl (500 mM), HEPES (50 mM) in MilliQ (pH 7.5)] with added protease inhibitor cocktail (Sigma) followed by sonication. The mixture was centrifuged and the lysate was purified using nickel affinity purification. The protein tag was cleaved using TEV protease and incubated overnight (4 °C). The protein was purified using gel filtration chromatography (Superdex 200, GE healthcare) in the buffer [TCEP·HCl (0.5 mM), glycerol (5 v/v%), NaCl (500 mM), HEPES (10 mM) in MilliQ (pH 7.5)]. Selected fractions were pooled and concentrated using spin concentrator (Amicon). The protein mass was confirmed using ESI-TOF-TOF Agilent LC/MSD. Mass calc.: 10159.4 Da, mass observed: 10159.6 Da.

Protein production of KDM1A$_{FL}$ ***(produced by Raphael Reinbold)***

KDM1A$_{FL}$ construct (2.0 µL, 30 ng·µL^{-1}) encoding for the gene (M1–M852) in vector pET-15b was transformed into phage resistant *E. coli* with pRARE2 plasmid by heat shocking (42 °C, 40 sec) in competent cells (20 µL) and incubation on ice (2 min). To the cells was added 2TY media (900 µL) and incubated (1 hrs, 37 °C). The cells gently mixed and transferred (100 µL) to an agar plate containing chloramphenicol (30 mg·L^{-1}) and ampicillin (100 mg·L^{-1}) antibiotics. The plate was incubated overnight (37 °C) and a single colony was used to induce the starter culture of 2TY media (100 mL) supplemented with ampicillin (100 mg·L^{-1}). Large scale growth of 2TY (2 L) media (containing ampicillin (100 mg·L^{-1})) was induced with the starter culture (10 mL·L^{-1}) and grown with shaking (175 rpm, 37 °C) until an OD_{600}: 0.6–0.8 was reached (5.5 hrs). Cells were then induced with IPTG (0.1 mM), cooled (19 °C) and 16 hrs. Cells harvested by centrifugation (11899 × g, 8000 rpm, 10 min) and stored (−80 °C). The cell pellet was suspended in lysis buffer and lysed using sonication (3.0 sec on, 3.0 sec, total time: 5 min) on ice. The lysate was isolate by centrifugation (48384 × g, 35 min), and the supernatant was decanted off and filtered (0.45 µm) prior to purification. The lysate was loaded on a HisTrap™ (5 mL), washed with wash buffer and eluted using elution buffer. The fractions were combined based on SDS-PAGE, concentrated (4255 × g) and subjected to size exclusion gel chromatography (Superdex 200, 330 mL) using gel filtration buffer. Fractions were pooled and concentrated using centrifugation and stored (−80 °C). The protein concentration was determined by NanoDrop (Thermo Scientific).

KDM5A, B demethylases ***(produced by Dr. Catrine Johansson)***

KDM5A$_{C1}$ (P13–S744 (ΔL88–G353 replace by: 4×G)), KDM5A$_{C2}$ (M1–L801) and KDM5B$_{C1}$ (M1–R822) were produced according to previous reported protocols.[228,292]

KDM5C$_{C1}$ ***(produced by Aleksandra Szykowska)***

The KDM5C$_{C1}$ protein was expressed according to previous reported protocol.[306]

$KDM5D_{C1}$ *(produced by Eidarus Salah)*

The KDM5D (M1–D775) in the vector pFB-LIC-Bse were transformed into Baculo virus. Exponentially growing Sf9 cells (2×10^6 cells·mL^{-1}) were infected with high titre Baculo virus stock and incubated in shaker flasks. Cells were shaken at 90 rpm at 27 °C. After 60 hrs post-infection, the cell suspensions were harvested (15 min, 800 × g, 4 °C), and the cell pellets suspended in PBS. After another centrifugation step (15 min, 800 × g, 4 °C), the cell pellets were stored at −80 °C. The cell pellet was thawed, suspended in lysis buffer [Imidazole (20 mM), NaCl (200 mM), TCEP (0.5 mM), HEPES (50 mM) in water (pH: 7.4)] (100 mL) and sonicated (2 min, amplitude 35%, on ice). The cell lysate was cleared by centrifugation (60 min, 36000 × g, 4 °C). The supernatant was combined with Ni Sepharose (7.5 mL) and loaded onto a gravity flow column. The column was washed with binding buffer [Imidazole (20 mM), NaCl (500 mM), TCEP (0.5 mM), glycerol (5%), HEPES (50 mM) in water (pH: 7.4)] (100 mL), wash buffer 1 [Imidazole (40 mM), NaCl (1.0 M), TCEP (0.5 mM), HEPES (50 mM) in water (pH: 7.4)] (50 mL) and wash buffer 2 [Imidazole (60 mM), NaCl (500 mM), TCEP (0.5 mM), HEPES (50 mM) in water (pH: 7.4)]. Elute Buffer [Imidazole (300 mM), NaCl (500 mM), TCEP (0.5 mM), HEPES (50 mM) in water (pH: 7.4)] (25 mL) was passed through to elute the protein from the column and the Eluted protein was concentrated (~5 mL). Finally, the protein was polished by gel filtration using an ÄKTAxpress system with an S200 16/600 column and GF buffer [NaCl (500 mM), glycerol (5%), TCEP (0.5 mM), HEPES (20 mM) in water (pH 7.4)]. Fractions containing protein were pooled, concentrated and stored at −80 °C.



Chapter 4. The effect of sterically enlarged H3 N^{ε}-trimethyl-lysine 4 on the binding affinity to chromatin binding modules

4.1 Contributors to this Chapter

SPIN2B$_{TTD}$ was produced by Dr. Catrine Johansson and KDM7B$_{PHD}$ was produced by Dr. Yijia Zhang. TAF3$_{PHD}$, BPTF$_{PHD}$, SGF29$_{TTD}$, KDM5A$_{PHD3}$ and KDM4A$_{TTD}$ were expressed and purified by Dr. Bas J. G. E. Pieters. ITC experiments of H3$_{1-21}$K4^{n}bu$_3$ **4.15** against BPTF$_{PHD}$ and KDM4A$_{TTD}$ and H3$_{1-21}$K4 **4.1, 4.11–4.14** against KDM7B$_{PHDJmjC}$ were conducted by Dr. Jasmin Mecinović. FP assay and enzyme kinetics on KDM7B$_{PHDJmjC}$ were performed by Dr. Yijia Zhang.

4.2 Introduction to reader proteins

Histone N^{ε}-lysine methylation is a dynamic PTM that shapes the chromatin landscape. This PTM can be installed iteratively up to three stages on the N^{ε}-lysine to ultimately give N^{ε}-trimethyl-lysine. N^{ε}-methyl-lysine is associated to both transcriptional activation and silencing depending on the methylation state and the position of the lysine residue within the histone octomeric complex.

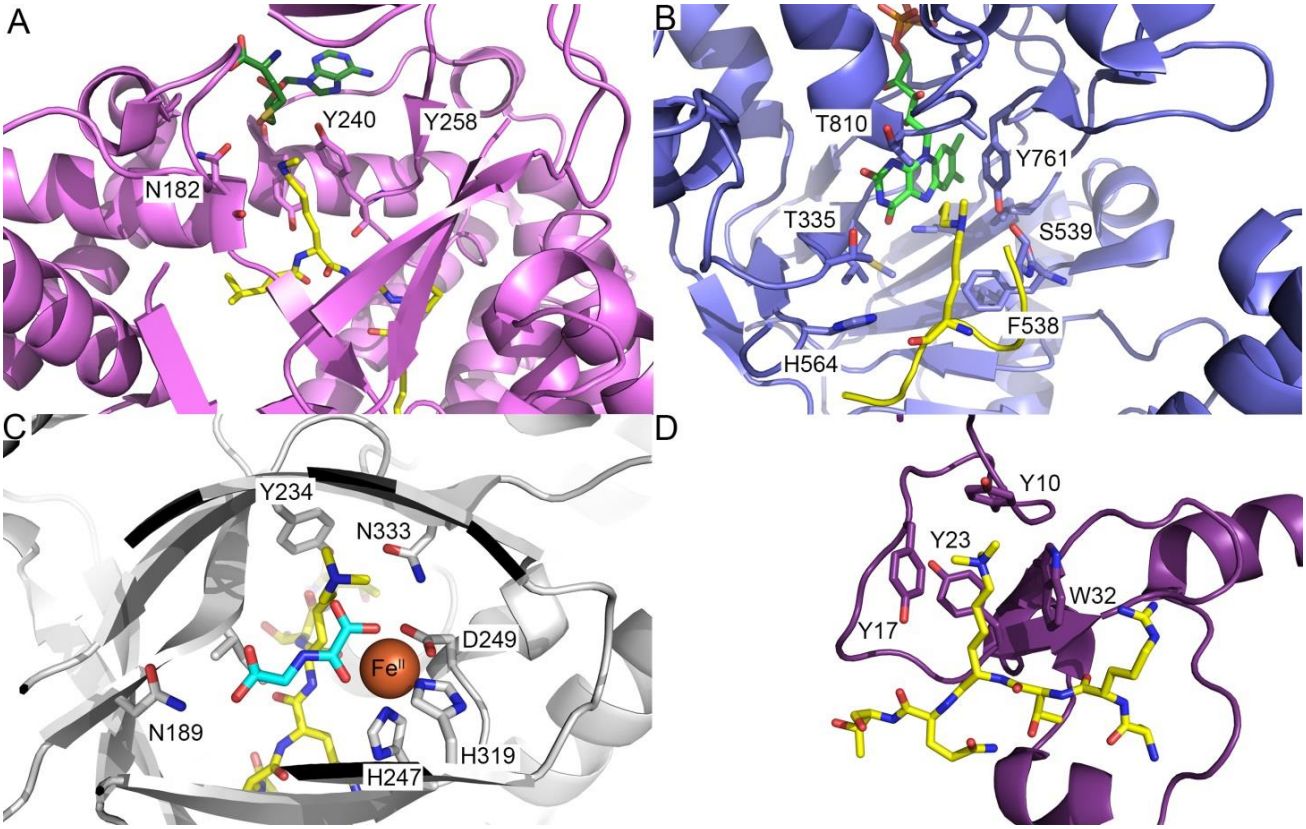


Figure 4.2.1: Crystal structure of chromatin lysine methylation associated proteins bound to histone H3 mimic peptides (yellow). A) Histone lysine methyltransferase SMYD2 (Pink) with substrate and SAM (dark green) (PDB: 3S7D), B) Co-factor FAD (bright green) and substrate bound to histone lysine demethylase KDM1A (light blue) (PDB: 2UXN), C) substrate, Iron (orange) and 2OG (cyan) bound to KDM7B containing (grey) (PDB: 3KV4) and D) substrate bound to reader protein BPTF-PHD (purple) (PDB: 2FJ6). Oxygen (red), nitrogen (blue) and sulphur (yellow) atoms are separately highlighted.

Associated to this histone lysine methylation are three types of proteins: A) histone methyltransferases (writers) that install a methyl at the designated histone lysine residues using co-factor *S*-adenosyl-L-methionine (SAM); B) histone lysine specific demethylases (erasers) containing either an amine oxidase domain or JumonjiC (JmjC) domain that remove the methyl groups; and C) chromatin binding

modules (readers) which sense a specific methylation level and facilitate a downstream effect (Fig. 4.2.1).

4.2.1 Chromatin binding domains in different shapes and forms

Chromatin binding modules are found throughout the chromatin, each domain playing roles in recognising histone lysine (and in some cases arginine) methylation states. Multiple domains to date have been identified that bind to histone *N*ε-trimethyl-lysine. These domains include the "Royal superfamily" of domains containing structurally related protein folds including: Tudor (TD), PWWP, malignant brain tumour (MBT), chromodomains (CD) and the Plant Homeodomains (PHD).[307] The focus in this Chapter is primarily on all class of methyl-lysine readers that bind H3K4, extending on the work carried out in Chapter 3 on the PHD-fingers associated with KDM5A. The chromatin binding modules used in this Chapter possess an increasing binding affinity with each subsequent higher methylation level ($K4me_0 << K4me_1 < K4me_2 < K4me_3$) as has been reported. However, it is worth noting that some reader domains recognize unmethylated lysines / unmodified histone tails; for example, AIRE-PHD1 preferentially binds to $K4me_0$.[235,308] In some cases, such as KDM5A, two reader domains are found within the same protein and recognise lysine methylation at the same H3K4 site, but with different methylation state selectivity (see also Chapter 3).[8,166]

There have been many reports on the significance of histone *N*ε-trimethyl-lysine recognition by chromatin binding proteins in transcription factor recruitment, epigenetic repression and nucleosome assembly among other operations.[235,238,309] Small molecule inhibitors could be a vital tool to explore the role of these proteins in the cell. However, few inhibitors of methyl-lysine reader domains have been reported to date.[280,310–312] While some inhibitors have been identified for KDM5A-PHD3, they showed double digit micro-molar IC_{50} inhibition and displayed strong off-target inhibition to the JmjC-domain of KDMs.[280,281] A potential reason for the limited reports could be that only few chemical tools

have been established to target *N*ε-trimethyl-lysine reader recognition improving both potency and selectivity.

What is not yet certain is to what extent chromatin binding modules, known to be histone *N*ε-trimethyl-lysine binders, can accommodate larger modified *N*ε-lysine substituents. Greater plasticity of the binding pocket of reader proteins may be advantageous if steric enlargement or modified *N*ε-lysine substituents can be accommodated by the reader proteins but not recognised by KMTs and KDMs. Work on histone H3 *N*ε-dialkylated-lysine has shown that KDM7B, KDM2B and KDM4E cannot catalyse the oxidation/dealkylation of *N*ε-diethyl-lysine, whilst KDM6B can.[283,284] Furthermore, Chapter 3 demonstrates that *N*ε-triethyl-lysine is not a substrate for JmjC-domain of KDM5A-D.

This Chapter investigates the effect of sterically enlarged *N*ε-trimethyl-lysine on binding to a panel of reader proteins. By expanding the survey across multiple reader domains / proteins, a better fundamental insight into reader protein binding and recognition can be gained. This allows exploration of the binding diversity among reader protein domains which can potentially be exploited for inhibitor development.

4.3 Chapter objectives

The aim of this Chapter is to:

- Synthesise a library of di-/ tri-alkylated N^{ε}-lysine containing histone mimic peptides.
- Investigate the effect of sterically enlarged H3 N^{ε}-dimethyl-lysine 4 and H3 N^{ε}-trimethyl-lysine 4 functional groups on the binding affinity to chromatin binding modules that preferential binding for H3 N^{ε}-trimethyl-lysine 4.
- Determine the potential trends in binding affinities between PHD, TTDs and DCD of the readers examined.
- Elucidate the size of histone H3 N^{ε}-lysine mimic peptides at which the binding affinity starts to significantly decrease for chromatin binding modules.
- Determine the potential differences in enzymatic activity of $KDM7B_{PHDJmjC}$ by modulating the H3 N^{ε}-trimethyl-lysine size.

In this Chapter reference made to the domain are described in capitals (e.g. PHD) while a reference made to construct domain is described in subscript (e.g. $TAF3_{PHD}$).

4.4 Results

4.4.1 Panel of selected chromatin reader modules

To examine the potential differences in the binding strength between chromatin binding modules and sterically enlarged N^{ε}-trimethyl-lysine, a representative panel of reader proteins was selected based on following criteria: A) the protein is a known binder to histone H3 lysine 4 (H3K4) with the preference for higher N^{ε}-lysine methylation states; B) the reader protein adds diversity to the panel by domain or aromatic 'cage' architecture; C) the protein can be expressed in good quantity and purity using *E. coli* expression systems for ITC analysis; D) the protein has a known functional importance within the chromatin landscape.

Based on these selection criteria, the following proteins were identified and produced:

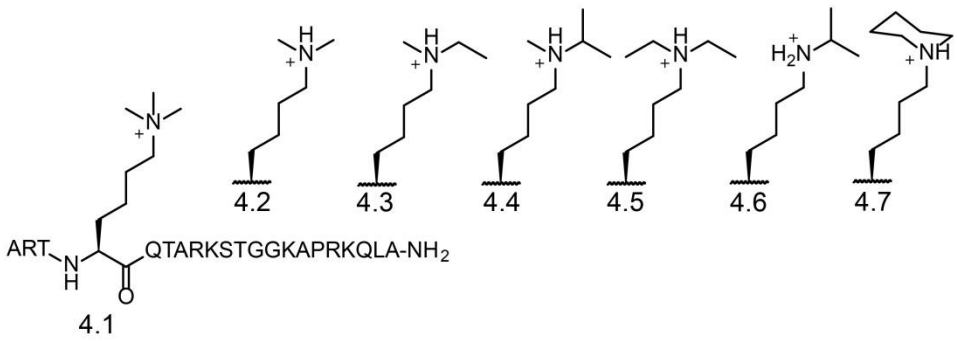
- TAF3$_{PHD}$ *(Dr. Bas J.G.E. Pieters)*, KDM7B$_{PHD}$ *(Dr. Yijia Zhang)*, KDM5A$_{PHD3}$ *(Dr. Bas J.G.E. Pieters)*, KDM7B$_{PHDJmjC}$ containing PHD.
- SGF29$_{TTD}$ *(Dr. Bas J.G.E. Pieters)*, KDM4A$_{TTD}$ *(Dr. Bas J.G.E. Pieters)* and SPIN2B$_{TTD}$ *(Dr. Catrine Johansson)* containing TTDs.
- CHD1$_{DCD}$ containing DCD.

The PHD3-finger of KDM5 used in this Chapter differs in length to the PHD3-finger in Chapter 3 and was chosen based on protein availability; however, both constructs contain the entirety of the PHD3-finger. KDM5A-PHD3 in this Chapter is annotated as KDM5A$_{PHD3}$ (M1489–V1641) and KDM5A-PHD3 of Chapter 3 as KDM5A$_{PHD3C2}$ (A1542–K1660).

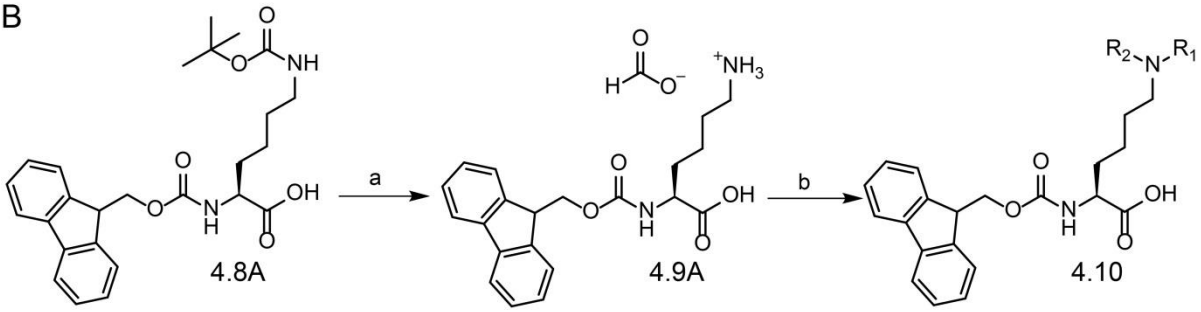
4.4.2 Synthesis of di-alkylated Fmoc-protected amino acids and Solid Phase Peptide Synthesis

To study larger N^{ε}-tri/dimethyl-lysine analogues, N^{ε}-lysine analogues were designed with different alkyl substituents (Fig. 4.4.1 A). The modifications introduced to N^{ε}-lysine were designed with the simplest form of sterically enlargement in mind, to not generate types of interactions that normally are not observed (*e.g.* π-π interactions or restriction on the rotational freedom of lysine). The larger methyl derivatives selected were: ethyl, isopropyl and piperidine derivative. Synthesis of the monomers was commenced from the commercially available starting material Fmoc-Lys(Boc)-OH **4.8A** (Fig. 4.4.1 B). The Boc-group of Fmoc-Lys(Boc)-OH **4.8A** was deprotected using neat formic acid or hydrochloric acid in diethyl ether ($HCl{\cdot}Et_2O$). The crude Fmoc-Lys-OH·HCOOH **4.9A** or Fmoc-Lys-OH·HCl **4.9B** was reductively alkylated using the corresponding aldehyde derivatives (Fig. 4.4.1 C) and sodium cyanoborohydride ($NaBH_3CN$) followed by acidic quench and reverse phase (R_f-C_{18}) purification. The collected fractions were pooled and lyophilized based on HPLC trace and mass confirmation. To yield Fmoc-Lys(me_2)-OH **4.10A** or Fmoc-Lys(et_2)-OH **4.10D**, reductive alkylation experiments were carried out in the presence of an excess formaldehyde or acetaldehyde, $NaBH_3CN$ and Fmoc-Lys-OH·HCl **4.9B** (for Lys(me_2)) or Fmoc-Lys-OH·HCOOH **4.9A** (for Lys(et_2)) obtaining the target molecules with yields of 82%, 69%, respectively. Attempts were made to produced Fmoc-Lys(me_1)-OH, a single alkylated N^{ε}-lysine molecule under limiting formaldehyde conditions and lower compound concentrations; however, this gave rise to a mixture of no-, single- and double-methylated lysine derivatives that were found to be difficult to isolate. To obtain, Fmoc-Lys(me, et)-OH **4.10B**, commercial available Fmoc-Lys(me, Boc)-OH **4.8B** was deprotected using neat formic acid followed by reductive alkylation with $NaBH_3CN$ and acetaldehyde resulting in Fmoc-Lys(me, et)-OH **4.10B** in 67% yield.

A



B



C

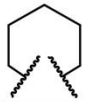
	R_1	R_2	Carbonyl 1	Carbonyl 2	Lyophilized yield
4.10A	methyl	methyl	HCHO(aq)	-	82%
4.10B	methyl	ethyl	CH_3CHO	-	67%
4.10C	methyl	isopropyl	Acetone	HCHO(aq)	82%
4.10D	ethyl	ethyl	CH_3CHO	-	69%
4.10E	isopropyl	H	Acetone	-	71%
4.10F			Gluteraldehyde	-	54%

Figure 4.4.1: Design and synthetic route to produced dialkylated N^{ε}-lysine analogues. A) Histone H3 mimic peptides containing sterically enlarged mono- and di-alkylated N^{ε}-lysine residues derived from the natural occurring H3K4$me_{2,1}$. B) Representative synthetic route to produce dialkylated N^{ε}-lysine derivatives suitable for SPPS. a: formic acid, O/N; b: carbonyls (5.4–9.2 equiv.), $NaBH_3CN$ (0.96–2.3 equiv.), 0.75–19 hrs then TFA quench and R_f-C_{18} purification. C) Table of produced dialkylated N^{ε}-lysine derivatives with corresponding aldehydes, yields and incorporated peptides displayed.

Experiments conducted using Fmoc-Lys(HCOOH)-OH **4.9A** and $NaBH_3CN$ in the presence of an excess amount of acetone showed only single isopropylation under the experimental conditions. This was likely due to steric hindrance of a secondary isopropyl group allowing the exclusive furnishing of Fmoc-Lys(ipr)-OH **4.10E** in good yield (71%). Following this route, reductive methylation of Fmoc-Lys(ipr)-OH **4.10E** using formaldehyde and $NaBH_3CN$ gave Fmoc-Lys(me, ipr)-OH **4.10C** in a single pot synthesis furnishing the product in a yield of 82%. Fmoc-Lys(pip)-OH **4.10F** was produced from

Fmoc-Lys(HCOOH)-OH **4.9A**, glutaraldehyde and $NaBH_3CN$ yielding a sufficient quantity for SPPS (54%). All obtained products were protonated with $HCl{\cdot}Et_2O$ to improve stability in DMF solutions and long-term storage. Protonation of the products produced hydroscopic solids and therefore yields were based on the lyophilized product. Identification of the products was achieved with high resolution mass spectrometry, ^1H NMR and ^{13}C NMR confirming the formation of the desired product (see methods).

The Fmoc-Lysine derivatives were incorporated in histone mimic peptide using microwave assisted solid phase peptide synthesis (μW-SPPS) applying Fmoc chemistry and synthesised following the same general procedure as described previously (Chapter 3). Minimal purity was >91%, with typical purity of >95% as determined using LC-MS of $H3_{1-21}K4$ **4.1**–**4.7**.

4.4.3 ITC analysis of $KDM4A_{TTD}$ and $KDM5A_{PHD3}$ against di-alkylated histone H3 lysine 4 mimic peptides ($H3_{1-21}$) 4.1–4.7.

To examine the binding affinity of histone H3 mimic di-alkylated N^{ε}-lysine 4 peptides **4.1–4.7** against reader proteins, ITC was employed. ITC experiments provide thermodynamic data on the binding strength (K_d, μM), Gibbs free energy (ΔG, kcal·mol^{-1}), enthalpy (ΔH, kcal·mol^{-1}) and entropy (−TΔS, kcal·mol^{-1}). A detailed description of this technique can be found in section 3.5.3. Representatives from TTD of $KDM4A_{TTD}$ and PHD-finger $KDM5_{PHD3}$ were chosen initially, since these are associated with KDMs and are well characterised.[8,313]

Proteins $KDM4A_{TTD}$ and $KDM5A_{PHD3}$ were measured using automated ITC200 (Malvern) with reader proteins in the cell (25.0 °C) and $H3_{1-21}$ peptides solution as titrant (19 injections). $KDM4A_{TTD}$ and $KDM5A_{PHD3}$ were examined against the panel of di-alkylated peptides $H3_{1-21}K4$ **4.1**–**4.7** and the results are shown in figure 4.4.2 and Table 4.4.1.

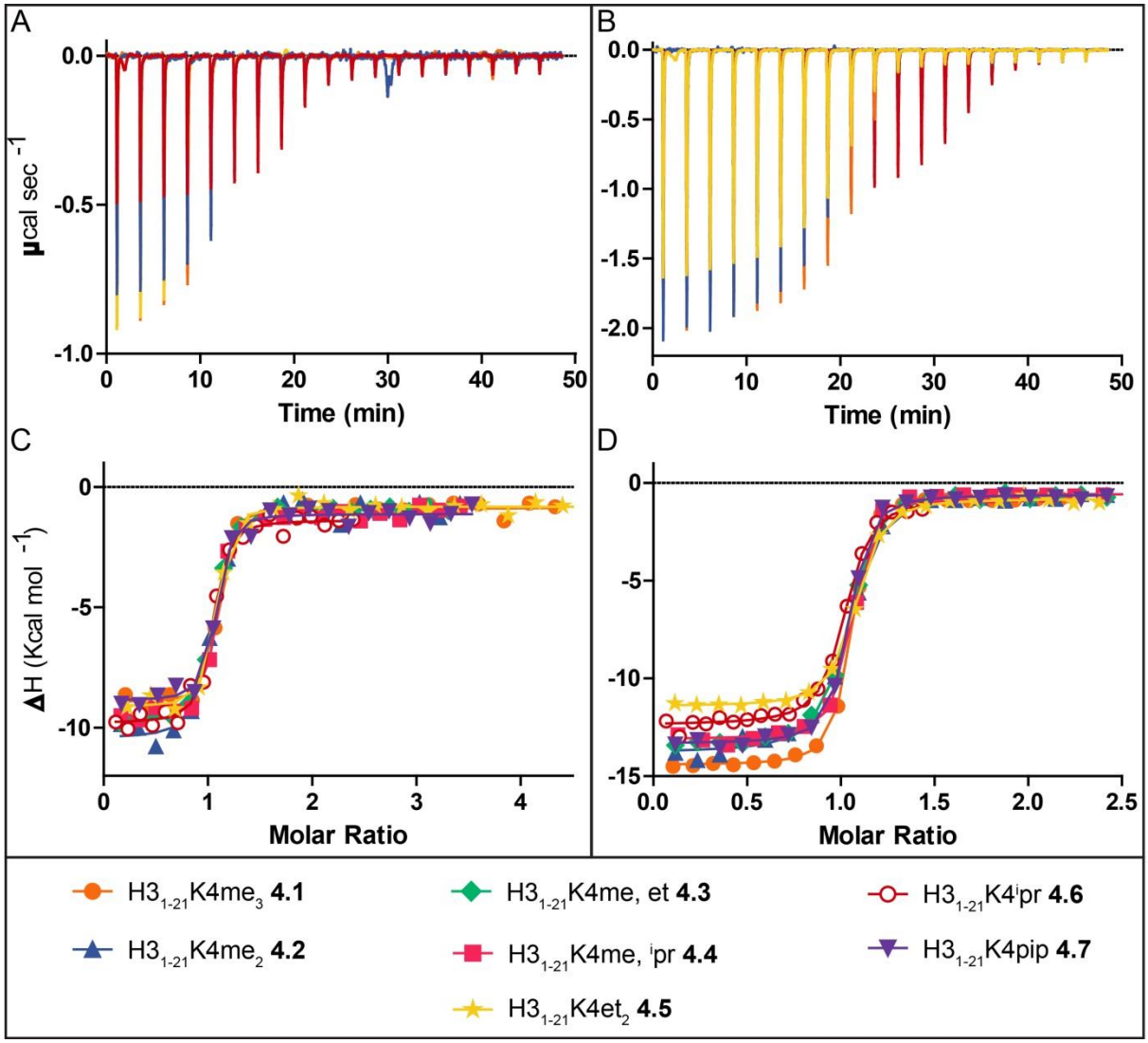


Figure 4.4.2: ITC measurements of KDM4A$_{TTD}$ (A,C) and KDM5A$_{PHD3}$ (B,D) against histone mimic peptides H3$_{1-21}$K4 4.1–4.7 (colour coded). A,B: representative raw ITC data of H3$_{1-21}$K4me$_3$ **4.1**, H3$_{1-21}$K4me$_2$ **4.2**, H3$_{1-21}$K4et$_2$ **4.5** and H3$_{1-21}$K4^{i}pr **4.6**. C,D: integrated ITC data of H3$_{1-21}$K4 **4.1–4.7** measured against KDM4A$_{TTD}$ (C) and KDM5A$_{PHD3}$(D). Corresponding ITC values are displayed in Table 4.4.1.

For KDM4A$_{TTD}$, H3$_{1-21}$K4me$_3$ **4.1** showed binding at sub micro-molar range with a K_d of 107 ± 16 nM and corresponding ΔH: −13.42 kcal·mol^{-1} and −TΔS: 3.92 kcal·mol^{-1}. This is in agreement with reported data.[244,313,314] As ITC data can vary depending on temperature, protein/peptide purity and buffer conditions, H3$_{1-21}$K4me$_3$ **4.1** was used as a point of reference. H3$_{1-21}$K4me$_2$ **4.2** was found to bind more weakly than H3$_{1-21}$K4me$_3$ **4.1** (K_d: 314 ± 18 nM) which is in agreement with reported data.[231,244,313] Interestingly, all unnatural alkylation modifications of H3$_{1-21}$K4 **4.3–4.7** tested against KDM4A$_{TTD}$ demonstrated an increase in binding affinity comparative to H3$_{1-21}$K4me$_2$ **4.2** but did not

exceed the binding affinity found for $H3_{1-21}K4me_3$ **4.1**. Use of $H3_{1-21}K4me,et$ **4.3**, $H3_{1-21}K4me,{}^{i}pr$ **4.4,** $H3_{1-21}K4et_2$ **4.5** and $H3_{1-21}K4pip$ **4.7** revealed that bulkier substituents resulted in a decrease in enthalpy, but, a more favourable entropy improving the overall binding strength. The strongest binder was found to be $H3_{1-21}K4et_2$ **4.5** at K_d: 123 ± 3 nM demonstrating a reduction in enthalpy of ΔΔH: 0.47 kcal·mol^{-1} but an increase in entropy of Δ−TΔS: −1.00 kcal·mol^{-1}. The largest enthalpy–entropy compensation was observed for $H3_{1-21}K4pip$ **4.7** of ΔΔH: 2.40 kcal·mol^{-1} and Δ−TΔS: 2.54 kcal·mol^{-1} relative to $H3_{1-21}K4me_2$ **4.2**. Strong binding was also observed for *N*$^{\varepsilon}$-lysine singly alkylated $H3_{1-21}K4{}^{i}pr$ **4.6** with an enthalpy of ΔΔH: 2.19 kcal·mol^{-1} and entropy of Δ−TΔS: 2.51 kcal·mol^{-1}.

Table 4.4.1: ITC data obtained of chromatin binding module $KDM4A_{TTD}$ and $KDM5A_{PHD3}$ against histone H3 mimic peptides ($H3_{1-21}K4$ 4.1–4.7). n ≥ 2, error given as s.d.

		K_d nM	ΔG kcal·mol^{-1}	ΔH kcal·mol^{-1}	-TΔS kcal·mol^{-1}	K_d nM	ΔG kcal·mol^{-1}	ΔH kcal·mol^{-1}	-TΔS kcal·mol^{-1}
		$KDM4A_{TTD}$				**$KDM5A_{PHD3}$**			
$H3_{1-21}K4me_3$	**4.1**	107 ± 16	-9.51 ± 0.08	-13.42 ± 0.19	3.92 ± 0.13	48 ± 20	-10.00 ± 0.06	-9.94 ± 0.01	-0.06 ± 0.04
$H3_{1-21}K4me_2$	**4.2**	314 ± 18	-8.91 ± 0.11	-13.09 ± 0.21	4.18 ± 0.28	272 ± 74	-8.97 ± 0.16	-9.39 ± 0.06	0.42 ± 0.10
$H3_{1-21}K4me,et$	**4.3**	227 ± 3	-9.06 ± 0.01	-12.71 ± 0.06	3.65 ± 0.06	109 ± 1	-9.46 ± 0.06	-8.81 ± 0.20	-0.66 ± 0.26
$H3_{1-21}K4me,{}^{i}pr$	**4.4**	149 ± 6	-9.31 ± 0.01	-12.71 ± 0.07	3.40 ± 0.08	63 ± 12	-9.83 ± 0.11	-8.48 ± 0.06	-1.35 ± 0.17
$H3_{1-21}K4et_2$	**4.5**	123 ± 3	-9.44 ± 0.02	-12.62 ± 0.04	3.18 ± 0.02	78 ± 10	-9.70 ± 0.08	-8.34 ± 0.09	-1.36 ± 0.01
$H3_{1-21}K4{}^{i}pr$	**4.6**	181 ± 3	-9.22 ± 0.02	-10.90 ± 0.04	1.67 ± 0.05	77 ± 22	-9.71 ± 0.17	-8.28 ± 0.06	-1.43 ± 0.11
$H3_{1-21}K4pip$	**4.7**	236 ± 49	-9.04 ± 0.12	-10.69 ± 0.26	1.64 ± 0.38	54 ± 8	-9.92 ± 0.09	-7.90 ± 0.04	-2.02 ± 0.13

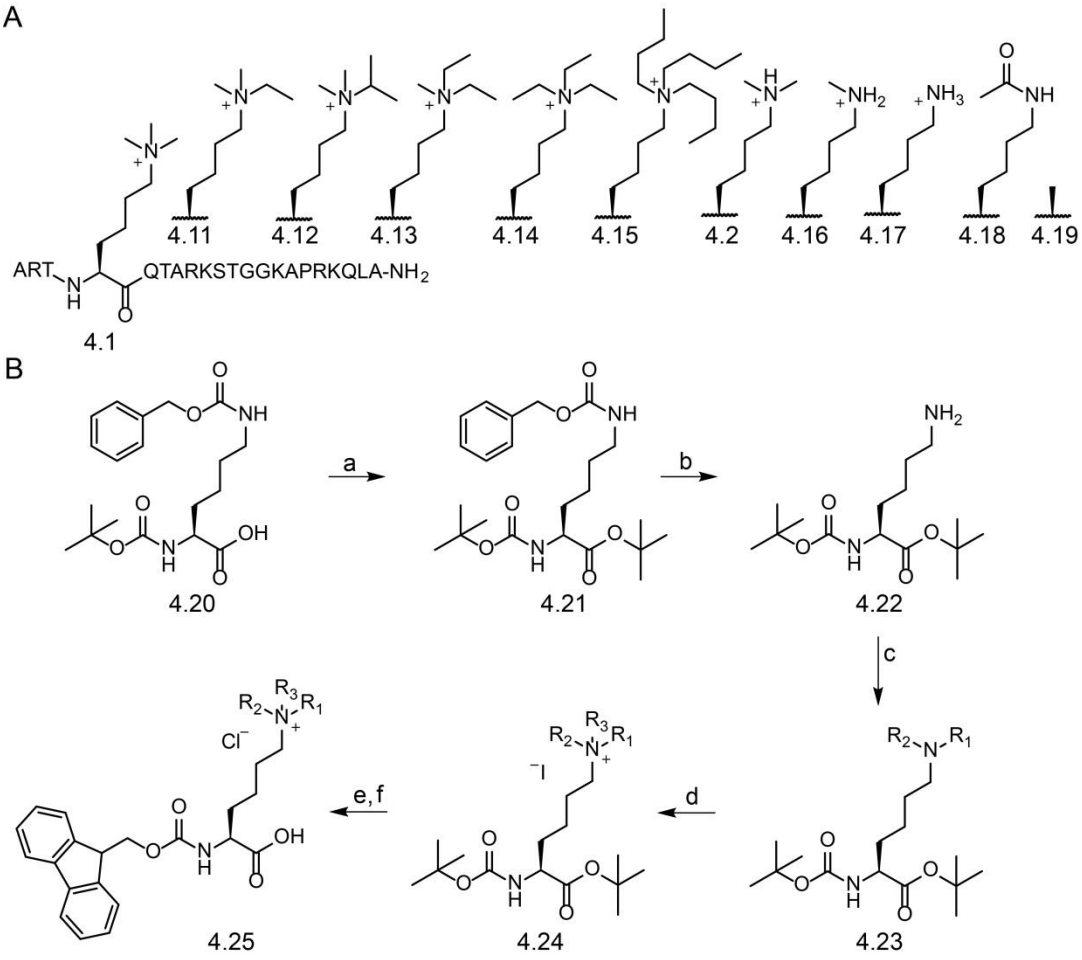
ITC analysis of $KDM5A_{PHD3}$ against $H3_{1-21}K4me_3$ **4.1** (K_d: 48 ± 20 nM) and $H3_{1-21}K4me_2$ **4.2** (K_d: 272 ± 52 nM) provided values in agreement with previously observed data.[8,231,244] Furthermore, ITC measurements conducted with $KDM5A_{PHD3C2}$ (see section 3.5.4) showed that $H3_{1-21}K4me_3$ **4.1** (K_d: 117 ± 25 nM) and $H3_{1-21}K4me_2$ **4.2** (K_d: 415 ± 49 nM) bound more weakly than was observed for $KDM5A_{PHD3}$. This is likely due to the different buffer conditions and protein construct length at which the binding events were measured. The binding affinity to $KDM5A_{PHD3}$ increased for all di-alkylated peptides $H3_{1-21}K4$ **4.3–4.7** tested relative to $H3_{1-21}K4me_2$ **4.2**. $H3_{1-21}K4pip$ **4.7** was found to bind the strongest at K_d: 54 ± 8 nM, representing a ~5-fold stronger binder relative to $H3_{1-21}K4me_2$ **4.2** and provides comparable binding to $H3_{1-21}K4me_3$ **4.1**. However, a loss in enthalpy was observed (ΔΔH: 1.49 ± 0.03 kcal·mol^{-1}) but strong entropy compensation (Δ−TΔS: −2.43 ± 0.09 kcal·mol^{-1}) caused an increase in binding strength relative to $H3_{1-21}K4me_2$ **4.2**. $H3_{1-21}K4{}^{i}pr$ **4.6** (K_d: 77.4 ± 15.2 nM) was

found to be an equally good binder as $H3_{1-21}K4me,iPr$ **4.4** (K_d: 62.9 ± 8.5 nM) despite $H3_{1-21}K4^{i}pr$ **4.6** lacking a methyl group and containing a single alkylation of N^{ε}-lysine 4.

The tested set of larger N^{ε}-mono- and N^{ε}-dialkylated-lysine 4 substituents showed that an increase in binding strength can be achieved relative to N^{ε}-dimethyl-lysine for both $KDM4A_{TTD}$ and $KDM5A_{PHD3}$**.** Although the N^{ε}-mono- and N^{ε}-dialkylated-lysine 4 substituents showed a stronger binding affinity against the reader proteins tested, the binding strength was found to be weaker or comparable to N^{ε}-trimethyl-lysine 4.

4.4.4 Synthesis of N^{ε}-tri-alkylated N^{α}-Fmoc-protected lysine residues and incorporation in histone mimic peptides ($H3_{1-21}$)

Studies of di-alkylated N^{ε}-lysine 4 residues demonstrated that stronger binding is achieved for the tested peptides $H3_{1-21}K4$ **4.3–4.7** relative to peptide $H3_{1-21}K4me_2$ **4.2** against chromatin binding modules $KDM4A_{TTD}$ and $KDM5A_{PHD3}$. To examine the boundaries of acceptance by reader proteins, a further set alkylated N^{ε}-lysine histone H3 mimic peptides were next prepared (**4.11–4.19**), using similar larger alkylation substituents (Fig. 4.4.3 A). Initial attempts were made to produce Fmoc-Lys(me_2, et)-OH from Fmoc-Lys(me_2)-OH **4.10A** by alkylation in the presence of ethyliodide and taking advantage of the potentially more nucleophilic N^{ε}-lysine. However, it was proven to be difficult to target the N^{ε}-lysine without affecting the C-terminal carboxylic acid. To produce exclusively N^{ε}-lysine alkylation, the synthetic route was changed and described in detail in section 3.5.3 and displayed in Figure 4.4.3 B. Briefly, following literature procedure, Boc-Lys(Cbz)-OH **4.20** was esterified to a tertbutyl ester using Boc_2O, DMAP in tert-butanol to give Boc-Lys(Cbz)-O^{t}Bu **4.21** (85%).[315] Following a literature procedure, deprotection of the benzyl group was achieved by catalytic hydrogenation using $H_{2(g)}$ and Pd / C providing Boc-Lys(H)-O^{t}Bu **4.22** in 90%.[315]



C

Peptide		Yield		Lyophilized yield	R_1	R_2	R_3
4.11	**4.24A**	35%	**4.25A**	60%	methyl	methyl	ethyl
4.12	**4.24B**	48%	**4.25B**	65%	isopropyl	methyl	methyl
4.13	**4.24C**	49%	**4.25C**	81%	ethyl	ethyl	methyl
4.14	**4.24D**	46%	**4.25D**	69%	ethyl	ethyl	ethyl
4.15	**4.24E**	65%	**4.25E**	20%	butyl	butyl	butyl

Figure 4.4.3: A) Histone $H3_{1-21}K4$ mimic peptides containing *N*ε-trimethyl-lysine 4 derived sterically enlarge modifications, B) Synthetic route for the formation of Fmoc-Lys(me_3)-OH related molecules suitable for peptide incorporation using SPPS, C) Table of Fmoc-monomers yields obtained in steps d and f and the corresponding peptide number. a) Boc_2O (1.46 equiv.), DMAP (0.2 equiv.), ^tBuOH, [SiO_2], 2.5 hrs, yield (85%); b) Pd / C (2.1 mol%), $H_{2\,(g)}$ in EtOH, 90%; c) corresponding carbonyls (5.0–6.2 equiv.), $NaBH_3CN$ (2.1-2.2 equiv.), 0.67–23 hrs, EtOH; d) iodoalkane (2.3–4.4 equiv.), acetonitrile, 80 °C, 0.67–48 hrs, 35–65%; e) TFA / CH_2Cl_2, 32 hrs; f) Fmoc-OSu (1.1–1.2 equiv.), $NaHCO_3$ (4.8–8.4 equiv.), milliQ / dioxane, 1-2 hrs, 20–81%.

The formation of the quaternary *N*ε-lysine intermediate molecule **4.24** was made in two consecutive steps from **4.22**. By altering the aldehyde and alkylation reagent composition, different *N*ε-lysine combinations were made (Fig. 4.4.3 C). As expected, more sterically hindered trialkylated *N*ε-Lysine products required longer reaction times for both reductive amination and alkylation furnishing **4.24** in

yields ranging between 35%–65%. The Boc-Lys(alkylated)-O^{t}Bu **4.24** were in a single pot synthesis deprotected from the acid labile tertbutyl-and Boc-protecting groups using trifluoroacetic acid (TFA) and N-terminal protection was installed using Fmoc-OSu providing compatibility with SPPS method used. The Fmoc-protected amino acids were turned into hydrochloride salt to enhance stability and incorporated into histone mimic peptide using μW-SPPS, purified using HPLC and lyophilized. Mass confirmation of the pure peptide was achieved with MALDI-TOF and LC-MS. The purity of the peptides was determined based on LC-MS UV trace finding purities exceeding >95%. Control peptides $H3_{1-21}K4me_1$ **4.16**, $H3_{1-21}K4me_0$ **4.17**, $H3_{1-21}K4Ac$ **4.18** and $H3_{1-21}K4A$ **4.19** were produced using this work-flow and commercially available Fmoc-monomers.

4.4.5 ITC binding studies of histone H3 mimic peptides containing *N*$^{\varepsilon}$-tri-alkylated lysine 4 residues against Tandem Tudor Domain reader proteins

To broaden the scope of chromatin binding modules in addition to $KDM4A_{TTD}$ (*by Dr. Bas J.G.E. Pieters*) and $KDM5A_{PHD3}$ (*by Dr. Bas J.G.E. Pieters*), a panel of the $H3K4me_3$ reader proteins representing TTD, PHD and CHD containing proteins were explored. $SGF29_{TTD}$ (*by Dr. Bas J.G.E. Pieters*), $TAF3_{PHD}$ (*by Dr. Bas J.G.E. Pieters*), $BPTF_{PHD}$ (*by Dr. Bas J.G.E. Pieters*), $KDM7B_{PHD}$ (*by Dr. Yijia Zhang*), $KDM7B_{PHDJmjC}$, $CHD1_{DCD}$ were produced in good quantity according to reported procedures.[231,306,316] A new method was developed for $SPIN2B_{TTD}$ (*by Dr. Catrine Johansson*). ITC binding studies were conducted using the produced histone H3 peptide variants against reader proteins $TAF3_{PHD}$, $KDM5A_{PHD}$, $KDM4A_{TTD}$, $KDM7B_{PHDJmjC}$, $SGF29_{TTD}$, $BPTF_{PHD}$, with the automated ITC200 (Malvern) and reader proteins $SPIN2B_{TTD}$, $KDM7B_{PHD}$, $CHD1_{DCD}$ using automated ITC-PEAQ (Malvern) (Fig. 4.4.4, Fig. 4.4.5, Table 4.4.2, Table 4.4.3). Measurements on both ITC instruments were carried out using the same robotic liquid-handling method, cell volume and cell temperature. The difference between the instrument setup is that the first injection of the PEAQ is 0.4

µL while the ITC200 injects 2.0 µL; however, the subsequent 18 injections are of equal quantity (2.0 µL).

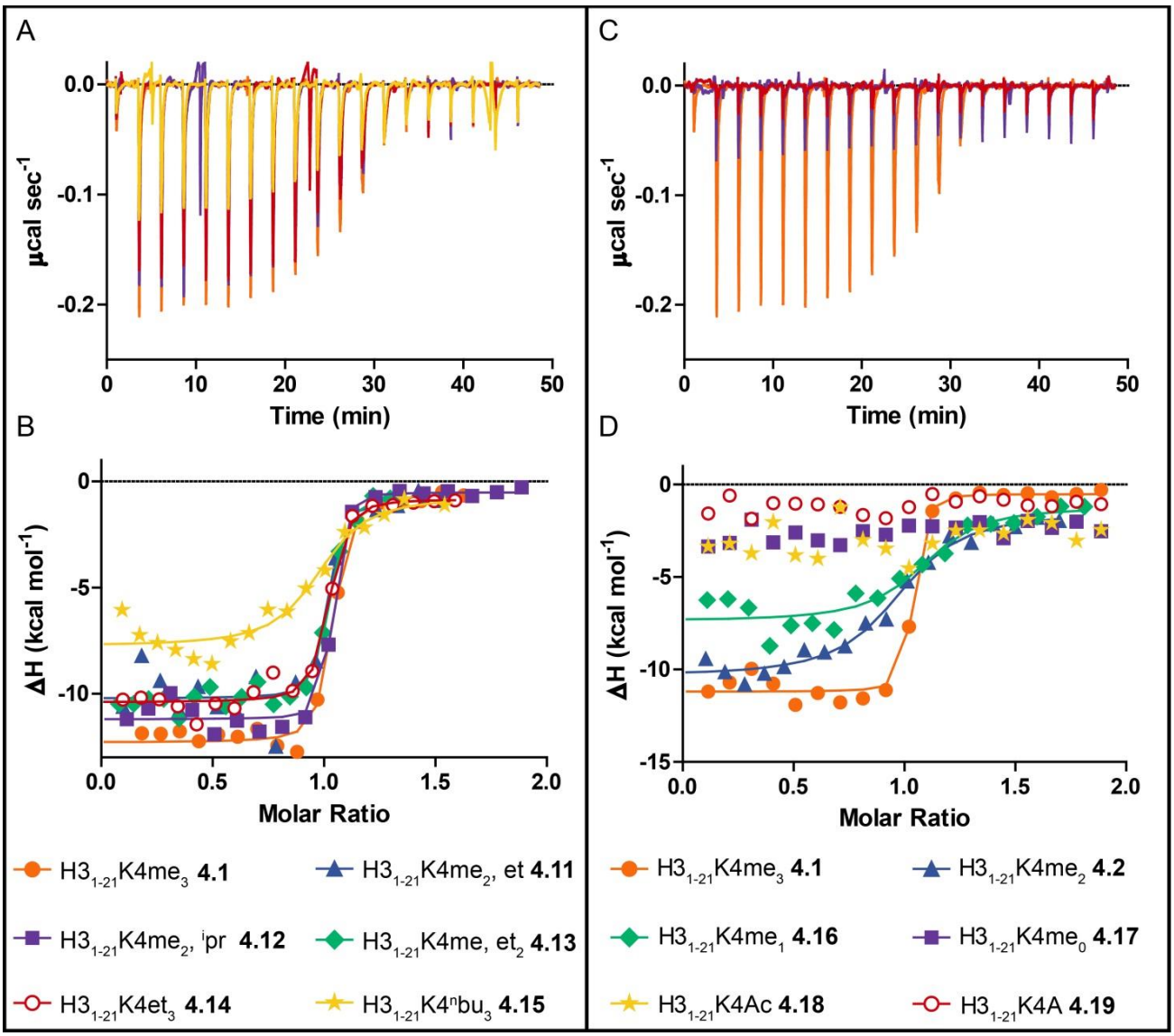


Figure 4.4.4: ITC binding data of SPIN2B$_{TTD}$ against H3$_{1-21}$K4 4.1, 4.11–4.15 (A,B) or H3$_{1-21}$K4 4.1, 4.2, 4.16–4.19 (colour coded). A,C) Raw ITC data of titrated and B,D) Integrated and molar ratio corrected ITC data of histone H3$_{1-21}$K4 peptides. Calculated affinity and thermodynamic values are displayed in Table 4.4.2.

ITC measurements of TTD proteins SPIN2B$_{TTD}$ (Fig. 4.4.4), KDM4A$_{TTD}$ and SGF29$_{TTD}$ against H3$_{1-21}$K4me$_3$ **4.1** (Table 4.4.2), demonstrate good binding affinity with K_d: 8.21 ± 5.02 nM, 107 ± 16 nM and 1050 ± 30 nM, respectively. This is in agreement with reported values.[231,244,313,317] Estimations of comparable binding strength and pattern for SPIN2B$_{TTD}$ were made based on reported SPIN1. SPIN2B and SPIN1 identical residues interacting with the *N*$^{\varepsilon}$-lysine and high sequence identity (79%) and

similarity (88%).[318] SPIN2B is likely to possess a similar pattern and sub-micro molar range of binding for $H3_{1-21}K4me_3$ **4.1**.

Examination of peptides $H3_{1-21}K4$ **4.11–4.14** against TTD containing readers showed for $SPIN2B_{TTD}$ (Fig. 4.4.4 A,B), $KDM4A_{TTD}$ and $SGF29_{TTD}$ only small variations in binding strength. The maximum decrease is observed for $SGF29_{TTD}$ by 2.3-fold between $H3_{1-21}K4me_3$ **4.1** and $H3_{1-21}K4et_3$ **4.14**. The peptide $H3_{1-21}K4^{n}bu_3$ **4.15** was revealed to drastically reduce the binding strength to $SPIN2B_{TTD}$ and $KDM4A_{TTD}$ *(Tested by Dr. Jasmin Mecinović)* by approximately ~23-fold and ~5-fold relative to $H3_{1-21}K4me_3$ **4.1** gaining similar levels of binding affinity as $H3_{1-21}K4me_2$ **4.2**. This data suggests that the bulky N^{ε}-tributyl-lysine may be a breaking point at which decrease in binding affinity is observed. Control experiments for $SPIN2B_{TTD}$ demonstrate that lower methylation states $H3_{1-21}K4me_2$ **4.2** and $H3_{1-21}K4me_1$ **4.16** bind increasingly weaker losing substantially in enthalpy with minor improvements in entropy (Fig. 4.4.4 C,D).

Table 4.4.2: ITC data of chromatin binding module $KDM4A_{TTD}$, $SGF29_{TTD}$, $SPIN2B_{TTD}$ against histone H3 mimic peptides ($H3_{1-21}K4$ 4.1, 4.2, 4.11–4.19). N.D.: not determined (SPIN2B K_d >5.0 µM). N.M.: not measured. $n \geq 2$, error given as s.d.

		K_d nM	ΔG kcal·mol^{-1}	ΔH kcal·mol^{-1}	-TΔS kcal·mol^{-1}	K_d nM	ΔG kcal·mol^{-1}	ΔH kcal·mol^{-1}	-TΔS kcal·mol^{-1}
		$KDM4A_{TTD}$				**$SGF29_{TTD}$**			
$H3_{1-21}K4me_3$	**4.1**	107 ± 16	-9.51 ± 0.08	-13.42 ± 0.19	3.92 ± 0.13	1050 ± 30	-8.15 ± 0.02	-7.58 ± 0.22	-0.57 ± 0.20
$H3_{1-21}K4me_2,et$	**4.11**	106 ± 8	-9.51 ± 0.04	-13.24 ± 0.01	3.73 ± 0.04	1760 ± 230	-7.87 ± 0.10	-6.91 ± 0.05	-0.96 ± 0.15
$H3_{1-21}K4me_2,^{i}pr$	**4.12**	127 ± 5	-9.39 ± 0.03	-13.18 ± 0.06	3.79 ± 0.08	1700 ± 90	-7.87 ± 0.03	-6.25 ± 0.04	-1.62 ± 0.07
$H3_{1-21}K4me,et_2$	**4.13**	104 ± 5	-9.52 ± 0.03	-12.98 ± 0.06	3.46 ± 0.08	1640 ± 60	-7.89 ± 0.02	-5.52 ± 0.13	-2.37 ± 0.11
$H3_{1-21}K4et_3$	**4.14**	105 ± 3	-9.52 ± 0.01	-12.31 ± 0.11	2.79 ± 0.13	2380 ± 190	-7.67 ± 0.05	-6.54 ± 0.10	-1.13 ± 0.14
$H3_{1-21}K4^{n}bu_3$	**4.15**	528 ± 57	-8.55 ± 0.08	-8.63 ± 0.22	0.07 ± 0.13	N.M.			
		$SPIN2B_{TTD}$							
$H3_{1-21}K4me_3$	**4.1**	8.21 ± 5.02	-11.13 ± 1.04	-10.86 ± 0.75	-0.27 ± 1.04				
$H3_{1-21}K4me_2,et$	**4.11**	16.6 ± 0.6	-10.60 ± 0.0	-12.65 ± 0.35	2.01 ± 0.33				
$H3_{1-21}K4me_2,^{i}pr$	**4.12**	18.1 ± 4.0	-10.62 ± 0.12	-10.40 ± 0.14	-0.22 ± 0.27				
$H3_{1-21}K4me,et_2$	**4.13**	14.1 ± 3.4	-10.75 ± 0.12	-9.28 ± 0.71	-1.46 ± 0.80				
$H3_{1-21}K4et_3$	**4.14**	12.1 ± 1.8	-10.81 ± 0.08	-9.58 ± 0.32	-1.24 ± 0.23				
$H3_{1-21}K4^{n}bu_3$	**4.15**	192 ± 66	-9.19 ± 0.23	-7.29 ± 0.61	-1.91 ± 0.84				
$H3_{1-21}K4me_2$	**4.2**	174 ± 24	-9.23 ± 0.08	-8.79 ± 0.20	-0.45 ± 0.28				
$H3_{1-21}K4me_1$	**4.16**	333 ± 61	-8.85 ± 0.11	-7.92 ± 0.79	-0.93 ± 0.90				

ITC analysis revealed that for $SPIN2B_{TTD}$ higher alkylation states mostly gave weaker enthalpy for $H3_{1-21}K4$ **4.11–4.15** with compensating levels of entropy in comparison to $H3_{1-21}K4me_3$ **4.1**. The

peptide $H3_{1-21}K4me_2$,et **4.11** is interesting due it being the only $SPIN2B_{TTD}$-peptide complex with a strong increase in favourable enthalpy (ΔΔH: −1.79 kcal·mol^{-1}) but less favourable entropy (Δ−TΔS: 2.28 kcal·mol^{-1}) relative to $H3_{1-21}K4me_3$ **4.1**. On the other hand, $H3_{1-21}K4^nbu_3$ **4.15** showed a decrease in enthalpy of ΔΔH: 3.57 kcal·mol^{-1} but an entropy compensation of Δ−TΔS: −1.64 kcal·mol^{-1} relative to $H3_{1-21}K4me_3$ **4.1**. The peptides $H3_{1-21}K4me_0$ **4.17**, $H3_{1-21}K4Ac$ **4.18** and $H3_{1-21}K4A$ **4.19** were determined to possess weak binding, the affinity of which could not be determined under the experiment conditions, estimated to be >5.0 μM. $KDM4A_{TTD}$ revealed little variation in binding strength between $H3_{1-21}K4me_3$ **4.1** and **4.11–4.13**. A large drop in enthalpy (ΔΔH: 4.79 kcal·mol^{-1}) was found for $H3_{1-21}K4^nbu_3$ **4.15** relative to $H3_{1-21}K4me_3$ **4.1** with a large entropic compensation (Δ−TΔS: −3.85 kcal·mol^{-1}). $SGF29_{TTD}$ was observed to have a similar binding strength for $H3_{1-21}K4$ **4.11–4.14** relative to $H3_{1-21}K4me_3$ **4.1**; however, levels of enthalpy-entropy compensation were detected. In particular, $H3_{1-21}K4me_2$,et **4.13** showed that the enthalpy weakened ΔΔH: 2.06 kcal·mol^{-1} while entropy improved by Δ−TΔS: −1.80 kcal·mol^{-1} when compared to $H3_{1-21}K4me_3$**4.1**. In general, the binding strength between $H3_{1-21}K4me_3$ **4.1** and **4.11–4.14** is relatively similar among tested TTD proteins. Both $KDM4A_{TTD}$ and $SPIN2B_{TTD}$ show a significant decrease in binding affinity for $H3_{1-21}K4^nbu_3$ **4.15** which suggests that a limit has been reached in accommodating bulkier N^{ε}-lysine analogues.

4.4.6 ITC binding studies of histone H3 mimic peptides containing N^{ε}-tri-alkylated lysine 4 residue against Plant Homeodomain reader proteins

To understand how PHD-readers compare to TTD-readers, binding affinity experiments of PHD-reader proteins with histone mimic peptides were carried out using ITC (Fig. 4.4.5, Table 4.4.3). All protein binding data obtained for $H3_{1-21}K4me_3$ **4.1** modified peptides were in agreement with literature publications.[8,167,231,244,317,319] The most striking result emerged from the analysis of the collated and comparative ITC data. Based on the obtained data, the tested PHD-proteins can be classed into two

sub-groups. TAF3$_{PHD}$ and KDM5A$_{PHD3}$ readers show, in comparison to the TTD-readers, similar binding strength across H3$_{1-21}$K4 **4.11–4.14** peptides examined with marginal decreases in affinity. Furthermore, the second group comprise of KDM7B$_{PHD}$ and BPTF$_{PHD}$ demonstrate a general increase in binding strength for H3$_{1-21}$K4 **4.11–4.14**.

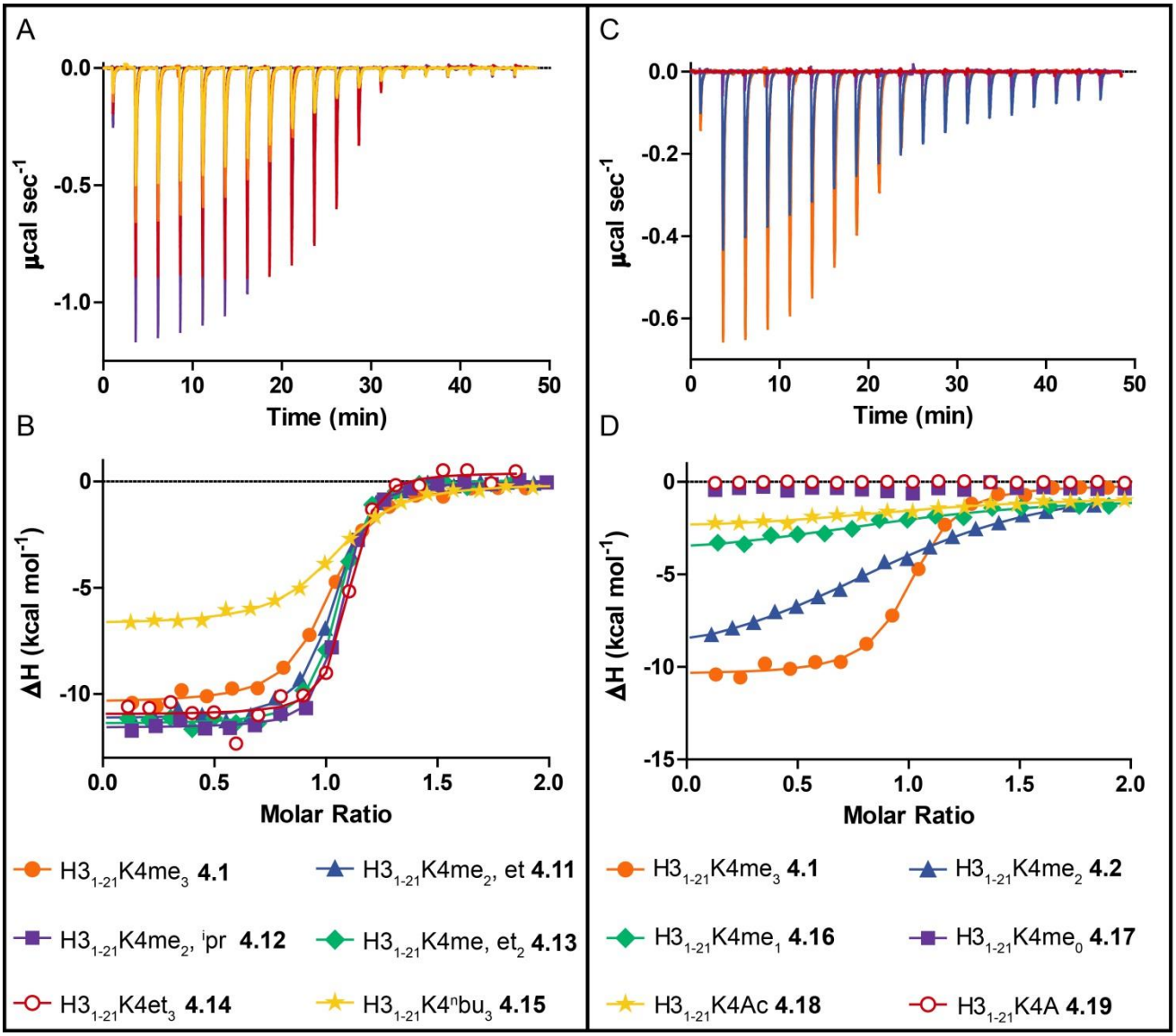


Figure 4.4.5: ITC binding data of KDM7B$_{PHD}$ against H3$_{1-21}$K4 4.1, 4.11–4.15 (A,B) or H3$_{1-21}$K4 4.1, 4.2, 4.16–4.19 (colour coded). A,C) Raw ITC data of titrated and B,D) Integrated and molar ratio corrected ITC data of histone H3$_{1-21}$K4 peptides. Calculated affinity and thermodynamic values are displayed in Table 4.4.3.

TAF3$_{PHD}$ revealed a similar binding strength to the tested H3$_{1-21}$K4 **4.1**, **4.11–4.14**. H3$_{1-21}$K4me$_2$, et **4.11** was determined to be the strongest binder at K_d: 10.0 ± 3.3 nM while H3$_{1-21}$K4et$_3$ **4.14** at K_d: 18.8 ± 0.6 nM was found to be the weakest binder. Small variations were observed in enthalpy H3$_{1-21}$K4 **4.1**, **4.11–4.14** peptides. H3$_{1-21}$K4me$_3$ **4.1** was determined to have the highest enthalpy (ΔH: −11.22 ±

0.11 kcal·mol^{-1}) but the least favourable entropy (−TΔS: 0.65 ± 0.21 kcal·mol^{-1}); H3$_{1-21}$K4me$_2^i$pr **4.12** was found to contain least favourable enthalpy (ΔH: −10.23 ± 0.06 kcal·mol^{-1}) and the most favourable entropy (−TΔS: −0.35 ± 0.13 kcal·mol^{-1}). ITC acquisitions of H3$_{1-21}$K4 **4.1**, **4.11–4.14** against KDM5A$_{PHD3}$ were determined to follow a similar trend as seen for TAF3$_{PHD}$ where minor levels of change in binding strength among peptides was observed. H3$_{1-21}$K4et$_3$ **4.14** (K_d: 74.1 ± 2.3 nM) was found to be similar to the natural substrate H3$_{1-21}$K4me$_3$ **4.1** (K_d: 47.5± 19.7 nM) but the enthalpy (ΔΔH: 1.98 kcal·mol^{-1}) was substantially weaker with some level of entropic compensation (Δ−TΔS: −1.67 kcal·mol^{-1}). KDM5A$_{PHD3}$ and TAF3$_{PHD}$ gave comparable trends; however, BPTF$_{PHD}$ presented a binding pattern also observed with KDM7B$_{PHD}$. H3$_{1-21}$K4me$_3$ **4.1** (K_d: 325 ± 50 nM) were found to contain similar binding strength to previous observations[319] but an increase in strength is found with higher alkylation state H3$_{1-21}$K4 **4.11–4.14**. Here, H3$_{1-21}$K4et$_3$ **4.14** was determined the strongest binder (K_d: 34 ± 1 nM) an ~10-fold improvement to H3$_{1-21}$K4me$_3$ **4.1**. This observation can mainly be attributed to a gain in enthalpy (ΔΔH: −1.98 kcal·mol^{-1}) but some increase in entropy (Δ−TΔS: −0.16 kcal·mol^{-1}) is observed in comparison to H3$_{1-21}$K4me$_3$ **4.1**. H3$_{1-21}$K4^{n}bu$_3$ **4.15** (K_d: 151 ± 57 nM) was found to be a weaker binder than H3$_{1-21}$K4et$_3$ **4.14** (~4-fold) but remained stronger than H3$_{1-21}$K4me$_3$ **4.1** (~2-fold). Relative to H3$_{1-21}$K4me$_3$ **4.1**, H3$_{1-21}$K4me,et$_2$ **4.13** was found to show the highest difference in enthalpy (ΔΔH: −1.98 kcal·mol^{-1}) with less favourable entropy (Δ−TΔS: 1.31 kcal·mol^{-1}).

Experiments with KDM7B$_{PHD}$ revealed that H3$_{1-21}$K4me$_3$ **4.1** binds with good binding strength of K_d: 391 ± 39 nM. Interestingly, peptides H3$_{1-21}$K4 **4.11–4.13** with higher alkylation states show consistent improvements of binding strength that can be attributed to an enthalpy increase with less favourable entropy. H3$_{1-21}$K4et$_3$ **4.14** is the strongest binder found at K_d: 67 ± 22 μM with both an increase in enthalpy (ΔΔH: −0.59 kcal·mol^{-1}) and entropy (Δ−TΔS: −0.48 kcal·mol^{-1}). Although H3$_{1-21}$K4et$_3$ **4.14** showed a ~5.8-fold increase in binding for KDM7B$_{PHD}$, a decrease in binding strength was found for H3$_{1-21}$K4^{n}bu$_3$ **4.15** by ~2.0 fold (K_d: 793 ± 34 nM) relative to H3$_{1-21}$K4me$_3$ **4.1**. H3$_{1-21}$K4^{n}bu$_3$ **4.15**

showed a large decrease in enthalpy (ΔΔH: 3.99 kcal·mol^{-1}) with major gains in entropy (Δ−TΔS: −3.57 kcal·mol^{-1}). Despite the reduction in binding strength for $H3_{1-21}K4^{n}bu_3$ **4.15** relative to $H3_{1-21}K4me_3$ **4.1**, this peptide is stronger than $H3_{1-21}K4me_2$ **4.2** (K_d: 6.81 ± 16 µM). As previously demonstrated, lower methylation states $H3_{1-21}K4me_2$ **4.2** > $H3_{1-21}K4me_1$ **4.16** > $H3_{1-21}K4me_0$ **4.17** bind with decreasing lower affinity to $KDM7B_{PHD}$.[167] Taken together, these results show that $H3_{1-21}K4$ **4.11–4.14** are similar or stronger binders than $H3_{1-21}K4me_3$ **4.1** for both TTD and PHD chromatin binding modules. However, unlike observed for TTDs, three of the five PHD containing proteins ($BPTF_{PHD}$, $KDM7_{PHD}$, $KDM7B_{PHDJmjC}$) show an increase in binding of $H3_{1-21}K4$ **4.11–4.14.** Furthermore, as seen for TTD, the binding affinity for $BPTF_{PHD}$ and $KDM7B_{PHD}$ decreases for $H3_{1-21}K4^{n}Bu_3$ **4.15** relative to $H3_{1-21}K4et_3$ **4.14**.

Table 4.4.3: ITC data of chromatin binding module $KDM5A_{PHD3}$, $BPTF_{PHD}$, $TAF3_{PHD}$, $KDM7B_{PHD}$ and $KDM7B_{PHDJmjC}$ against histone H3 mimic peptides ($H3_{1-21}K4$ 4.1, 4.2, 4.11–4.19). N.D.: not determined ($KDM7B_{PHD}$ K_d > 25 µM). N.M.: not measured. n ≥ 2, error given as s.d.

		K_d nM	ΔG kcal·mol^{-1}	ΔH kcal·mol^{-1}	-TΔS kcal·mol^{-1}	K_d nM	ΔG kcal·mol^{-1}	ΔH kcal·mol^{-1}	-TΔS kcal·mol^{-1}
		$KDM5A_{PHD3}$				**$BPTF_{PHD}$**			
$H3_{1-21}K4me_3$	**4.1**	47.5 ± 19.7	-10.00 ± 0.06	-9.94 ± 0.01	-0.06 ± 0.04	328 ± 50	-8.85 ± 0.08	-10.94 ± 0.13	2.09 ± 0.22
$H3_{1-21}K4me_2$,et	**4.11**	46.6 ± 8.6	-10.01 ± 0.11	-9.70 ± 0.01	-0.31 ± 0.10	116 ± 2	-9.46 ± 0.03	-12.48 ± 0.10	3.02 ± 0.07
$H3_{1-21}K4me_2$,^{i}pr	**4.12**	58.6 ± 19.4	-9.83 ± 0.09	-9.96 ± 0.08	0.16 ± 0.13	109 ± 22	-9.51 ± 0.13	-12.41± 0.04	2.90 ± 0.16
$H3_{1-21}K4me$,et_2	**4.13**	58.0 ± 26.8	-9.90 ± 0.28	-8.82 ± 0.06	-1.09 ± 0.34	104 ± 1	-9.52 ± 0.01	-12.92 ± 0.01	3.40 ± 0.01
$H3_{1-21}K4et_3$	**4.14**	74.1 ± 2.3	-9.69 ± 0.04	-7.96 ± 0.19	-1.73 ± 0.15	34 ± 1	-10.19 ± 0.01	-12.12 ± 0.06	1.93 ± 0.08
$H3_{1-21}K4^{n}bu_3$	**4.15**	N.M.				151 ± 57	-9.34 ± 0.26	-10.34 ± 0.11	1.00 ± 0.21
		$TAF3_{PHD}$				**$KDM7B_{PHD-JmjC}$**			
$H3_{1-21}K4me_3$	**4.1**	17.8 ± 2.9	-10.57 ± 0.10	-11.22 ± 0.11	0.65 ± 0.21	1990 ± 81	-7.78 ± 0.04	-11.42 ± 0.09	3.64 ± 0.13
$H3_{1-21}K4me_2$,et	**4.11**	10.0 ± 3.3	-10.93 ± 0.20	-11.13 ± 0.06	0.20 ± 0.14	897 ± 17	-8.25 ± 0.01	-10.58 ± 0.05	2.33 ± 0.04
$H3_{1-21}K4me_2$,^{i}pr	**4.12**	17.9 ± 2.8	-10.58 ± 0.08	-10.23 ± 0.06	-0.35 ± 0.13	644 ± 58	-8.45 ± 0.05	-11.10 ± 0.09	2.65 ± 0.14
$H3_{1-21}K4me$,et_2	**4.13**	15.2 ± 2.1	-10.67 ± 0.08	-10.53 ± 0.07	-0.14 ± 0.15	557 ± 11	-8.54 ± 0.01	-11.99 ± 0.12	3.45 ± 0.11
$H3_{1-21}K4et_3$	**4.14**	18.8 ± 0.6	-10.54 ± 0.02	-10.64 ± 0.04	0.10 ± 0.03	388 ± 38	-8.75 ± 0.04	-12.12 ± 0.04	3.37 ± 0.00
		$KDM7B_{PHD}$							
$H3_{1-21}K4me_3$	**4.1**	391 ± 39	-8.68 ± 0.13	-10.26 ± 0.34	1.51 ± 0.37				
$H3_{1-21}K4me_2$,et	**4.11**	176 ± 2	-9.22 ± 0.01	-10.95 ± 0.05	1.74 ± 0.04				
$H3_{1-21}K4me_2$,^{i}pr	**4.12**	110 ± 16	-9.50 ± 0.09	-11.48 ± 0.08	1.96 ± 0.08				
$H3_{1-21}K4me$,et_2	**4.13**	119 ± 8	-9.45 ± 0.04	-11.40 ± 0.10	1.97 ± 0.16				
$H3_{1-21}K4et_3$	**4.14**	67 ± 22	-9.83 ± 0.20	-10.85 ± 0.27	1.03 ± 0.46				
$H3_{1-21}K4^{n}bu_3$	**4.15**	793 ± 34	-8.33 ± 0.02	-6.27 ± 0.09	-2.06 ± 0.12				
$H3_{1-21}K4me_2$	**4.2**	6810 ± 160	-7.06 ± 0.02	-10.35 ± 0.14	3.31 ± 0.14				
$H3_{1-21}K4me_1$	**4.16**	13470 ± 3530	-6.83 ± 0.16	-3.92 ± 0.78	-2.76 ± 0.94				
$H3_{1-21}K4me_0$	**4.17**	N.D.							
$H3_{1-21}K4Ac$	**4.18**	13265 ± 6840	-6.75 ± 0.34	-2.03 ± 0.27	-4.73 ± 0.61				
$H3_{1-21}A4$	**4.19**	N.M.							

4.4.7 ITC binding studies of histone H3 mimic peptides containing N^{ε}-tri-alkylated lysine 4 residues against chromo domain reader protein $CHD1_{DCD}$

To study the impact of sterically enlarged N^{ε}-lysine on the double chromo domain containing reader, recombinant $CHD1_{DCD}$ (A268–R443) was produced in *E. coli.* From 4 L of terrific broth was obtained 26.0 g of pellet, approximately 6.5 g of pellet was purified by nickel affinity purification and size exclusion yielding 68.3 mg of protein (9.87 mg·mL^{-1}). Mass confirmation using LC-MS (RF-MS) and SDS-PAGE indicated the correct mass and high purity (Fig. 4.8.2 A,B). Experiments of $CHD1_{DCD}$ binding to histone H3 mimic peptides using ITC instrumentation were carried out (Table 4.4.4), with corresponding representative binding curves shown in figure 4.4.6. The binding of $CHD1_{DCD}$ to $H3_{1-21}K4me_3$ **4.1** was found to be 1.60 ± 0.51 µM, in agreement with previously reported data.[264]

Similar binding values were observed for $H3_{1-21}K4me_2$,et **4.11–4.14** relative to $H3_{1-21}K4me_3$ **4.1**. ITC analysis identified some variation for $H3_{1-21}K4me_2$,et **4.11** gaining enthalpy (ΔΔH: −0.61 kcal·mol^{-1}) but loss in entropy (Δ−TΔS: 0.59 kcal·mol^{-1}). In contrast, $H3_{1-21}K4me_2$, ipr **4.12** showed enthalpic loss (ΔΔH: 2.10 kcal·mol^{-1}) but entropic gain (Δ−TΔS: −2.21 kcal·mol^{-1}) suggesting that the ethyl or isopropyl might be differently accommodated by the $CHD1_{DCD}$ aromatic cage. The most striking result emerged from $H3_{1-21}K4^{n}bu_3$ **4.15** that showed no detectable level of binding to $CHD1_{DCD}$ (K_d: >50 µM) (Fig. 4.4.6 yellow) and was observed to be entropically driven. This is contrary to the observations made for $KDM4A_{TTD}$, $BPTF_{PHD}$ and $KDM7B_{PHD}$ where $H3_{1-21}K4^{n}bu_3$ **4.15** showed a smaller (~2–5 fold) decrease in binding relative to $H3_{1-21}K4me_3$ **4.1**. However, this might be similar to $SPIN2B_{TTD}$ where a relative large decrease of ~23-fold was observed between $H3_{1-21}K4^{n}bu_3$ **4.15** and $H3_{1-21}K4me_3$ **4.1**.

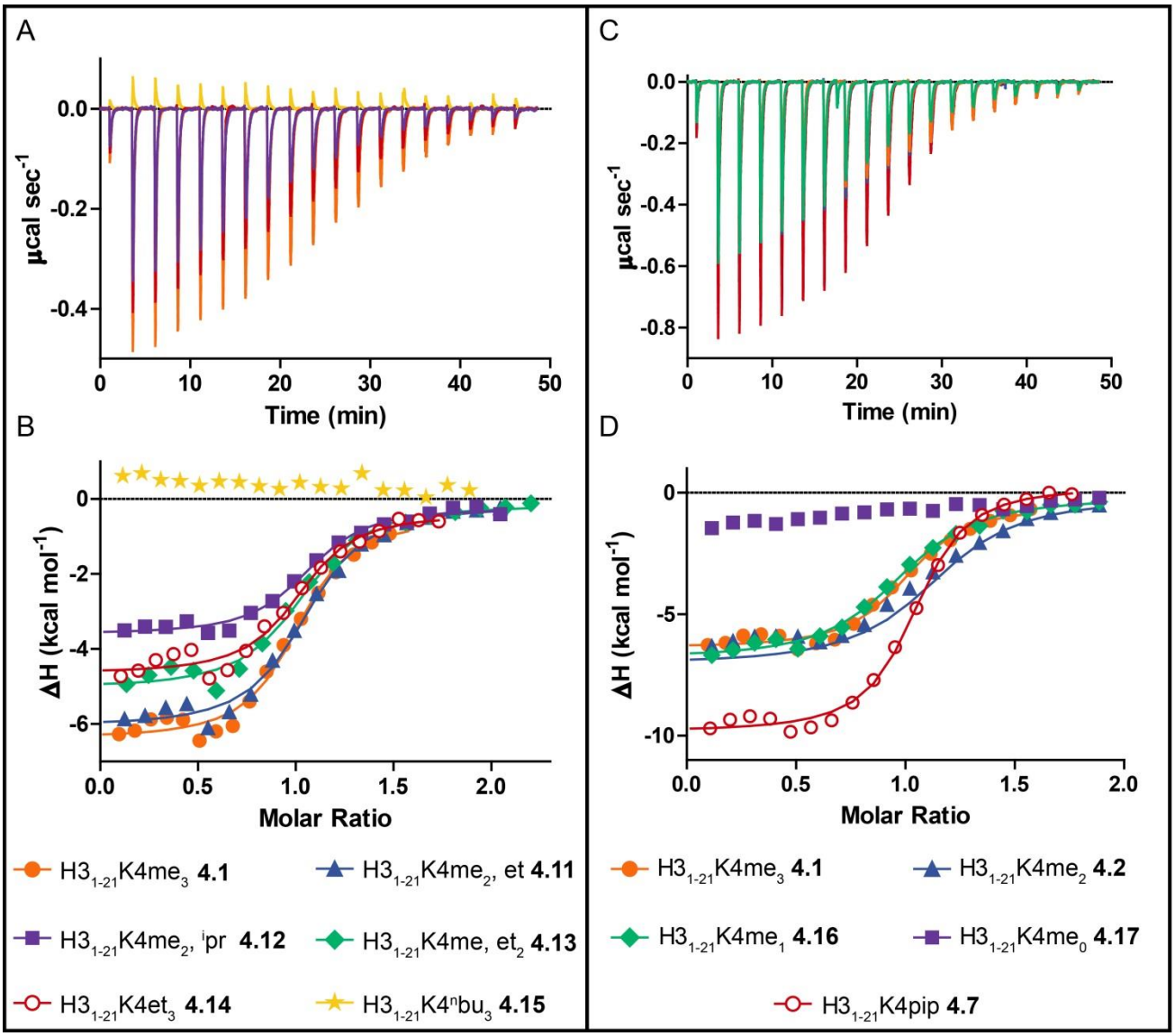


Figure 4.4.6: Representative ITC binding data of raw (A, C) and (B, D) of $H3_{1-21}K4$ 4.1, 4.2, 4.7, 4.11–4.17 peptides titrated to $CHD1_{DCD}$. Calculated affinity and thermodynamic values are displayed in Table 4.4.4.

Furthermore, control peptides $H3_{1-21}K4me_2$ **4.2** (K_d: 1.90 ± 0.16 µM), $H3_{1-21}K4me_1$ **4.16** (K_d: 1.70 ± 0.04 µM) showed relatively good binding while a sudden decrease in binding strength was found for $H3_{1-21}K4me_0$ **4.17** (K_d: >50 µM). This is unexpected as observations made for $SPIN2B_{TTD}$ and $KDM7B_{PHD}$ showed a more gradual decrease in binding strength for each subsequent lower methylation level. Despite none of higher trialkylated N^{ε}-lysine 4 **4.11–4.15** residues giving a stronger binding strength for $CHD1_{DCD}$, $H3_{1-21}K4pip$ **4.7** surprisingly revealed a small increase of ~2-fold over $H3_{1-21}K4me_3$ **4.1** enhancing enthalpy ($\Delta\Delta H$: −4.49 kcal·mol^{-1}) at the cost of entropy ($\Delta -T\Delta S$: 4.01

kcal·mol^{-1}). This data implies that alkylation of N^{ε}-lysine is vital to binding but that a wide range of alkylation levels are accepted.

Table 4.4.4: ITC binding values of histone H3 mimic peptides $H3_{1-21}K4$ 4.1, 4.2, 4.7, 4.11–4.18 measured against $CHD1_{DCD}$. N.D.: not determined (>50 µM). n≥2, errors given as ± s.d.

		$CHD1_{CHD}$			
		K_d	ΔG	ΔH	-TΔS
		µM	kcal·mol^{-1}	kcal·mol^{-1}	kcal·mol^{-1}
$H3_{1-21}K4me_3$	**4.1**	1.60 ± 0.51	-7.94 ± 0.19	-5.48 ± 0.41	-2.45 ± 0.24
$H3_{1-21}K4me_2$,et	**4.11**	1.51 ± 0.20	-7.95 ± 0.08	-6.09 ± 0.28	-1.86 ±0.35
$H3_{1-21}K4me_2$,^{i}pr	**4.12**	1.30 ± 0.03	-8.04 ± 0.01	-3.38 ± 0.01	-4.66 ± 0.01
$H3_{1-21}K4me$,et_2	**4.13**	1.42 ± 0.20	-7.98 ± 0.08	-4.84 ± 0.10	-3.15 ± 0.18
$H3_{1-21}K4et_3$	**4.14**	1.31 ± 0.11	-8.03 ± 0.04	-4.19 ± 0.08	-3.84 ± 0.04
$H3_{1-21}K4^{n}bu_3$	**4.15**	N.D.			
$H3_{1-21}K4me_2$	**4.2**	1.90 ± 0.16	-7.81 ± 0.05	-7.31 ± 0.16	-0.50 ± 0.21
$H3_{1-21}K4me_1$	**4.16**	1.70 ± 0.04	-7.88 ± 0.01	-6.25 ± 0.08	-1.63 ± 0.09
$H3_{1-21}K4me_0$	**4.17**	N.D.			
$H3_{1-21}K4Ac$	**4.18**	N.D.			
$H3_{1-21}K4pip$	**4.7**	0.696 ± 0.01	-8.40 ± 0.01	-9.97 ± 0.18	1.56 ± 0.16

4.4.8 ITC binding studies of histone H3 mimic peptides containing N^{ε}-tri-alkylated lysine 4 against KDM7B; surface groove *versus* cavity insertion

The N^{ε}-trimethyl-lysine binding site of $KDM7B_{PHD}$ can be categorized as surface groove binding configuration; however, the longer $KDM7B_{PHDJmjC}$ construct with PHD-domain adjacent to JmjC-domain reveals a more enclosed environment for N^{ε}-trimethyl-lysine tilting the mode of recognition to cavity insertion (Fig. 4.4.7).

To determine the effect of the peptide affinity on dual PHD-finger and JmjC- catalytic domain of KDM7B, $KDM7B_{PHDJmjC}$ (M37-N483) was produced from *E. coli*. The protein was isolated using nickel affinity purification followed by tag cleavage by His TEV protease. The target protein was obtained after gel filtration purification and concentrated to give 32 mg of 10.1 mg·mL^{-1} protein (Fig. 4.8.3). Protein purity and mass confirmation was achieved using SDS-PAGE and RF-MS, respectively.

As additional confirmation, KDM7B$_{PHDJmjC}$ was examined for enzymatic demethylation of histone H3 *N*$^{\varepsilon}$-lysine 9 using MALDI-TOF MS. In a time-course assay, H3$_{1-14}$K4me$_3$K9me$_2$ **4.20** (10 µM) was treated with co-factors: Asc (100 µM), $(NH_4)_2Fe(II)(SO_4)_2$ (10 µM), 2OG (100 µM) in the presence of KDM7B$_{PHDJmjC}$ (1.0 µM) and quenched at appropriate time points using methanol. After 20 min, more than 90% of the starting material was converted to give H3$_{1-14}$K4me$_3$K9me$_1$ or final product H3$_{1-14}$K4me$_3$K9me$_0$ demonstrating good activity (Fig. 4.4.8 A, Fig. 4.8.1).[167]

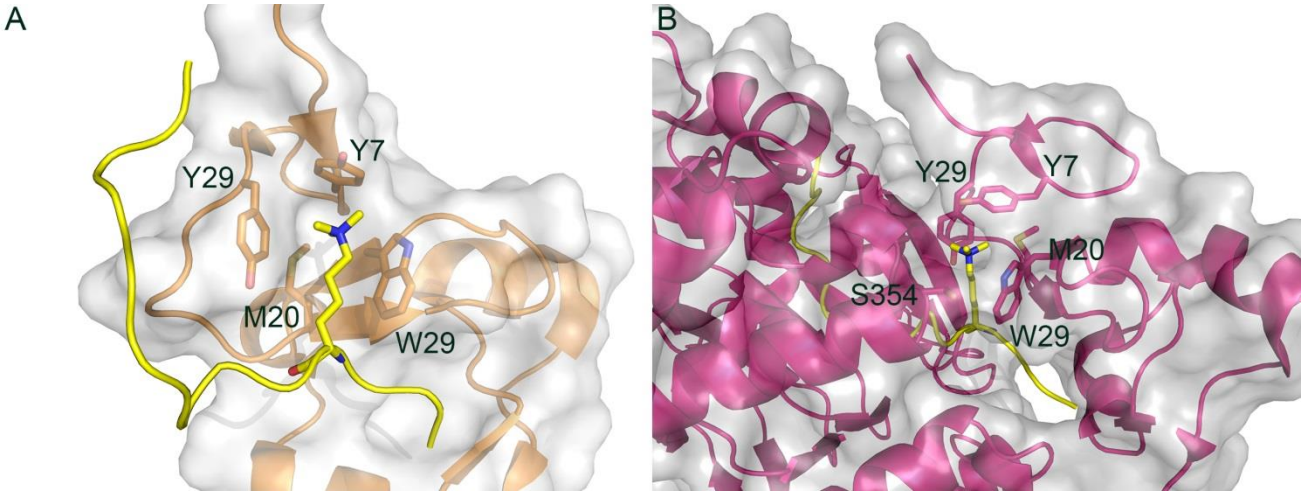


Figure 4.4.7: Crystal structures of KDM7B$_{PHD}$ (A, orange) and KDM7B$_{PHDJmjC}$ (B, pink) in complex with H3K4me$_3$ (yellow). Heteroatoms are highlights namely: oxygen (red), nitrogen (blue) and sulphur (yellow). (PDB: 3KV4)

KDM7B$_{PHDJmjC}$ was tested against H3$_{1-21}$K4 **4.1**, **4.11–4.14** peptides using ITC and the obtained data reveal that H3$_{1-21}$K4me$_3$ **4.1** binds well to KDM7B$_{PHDJmjC}$ (K_d: 1990 ± 81 nM) (Table 4.4.3). This is approximately 5-fold weaker than was determined with KDM7B$_{PHD}$ (K_d: 391 ± 39 nM). KDM7B$_{PHDJmjC}$ binding to H3$_{1-21}$K4 derivatives **4.11–4.15** was observed to generally bind weaker than for KDM7B$_{PHD}$ however, a similar trend was found between KDM7B$_{PHD}$ and KDM7B$_{PHDJmjC}$ where sterically enlarged *N*$^{\varepsilon}$-trimethyl-lysine derivatives bind more strongly. The peptide values of H3$_{1-21}$K4 **4.1**, **4.11–4.14** with both proteins reveals good correlation of R^2: 0.98 suggesting that secondary residues from the JmjC near the aromatic cage in KDM7B$_{PHDJmjC}$, but absent in KDM7B$_{PHD}$, might only weakly interact with *N*$^{\varepsilon}$-lysine 4 (Fig. 4.4.8 B).

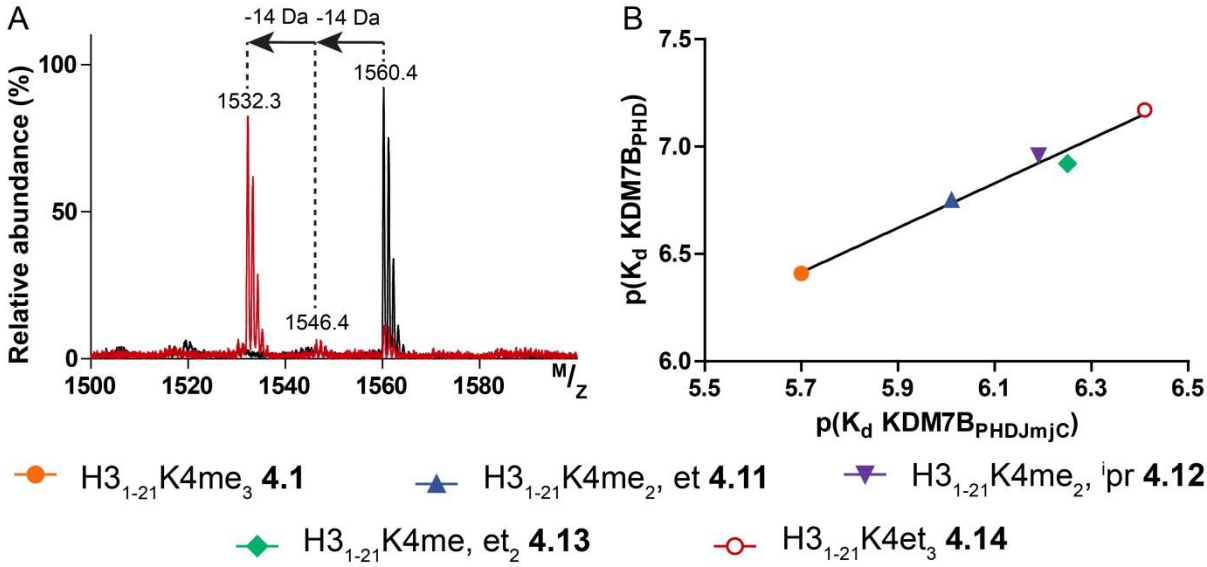


Figure 4.4.8: MALDI-TOF activity of KDM7B$_{PHDJmjC}$ and ITC binding data of H3$_{1-21}$K4 4.1, 4.11–4.14 measured against KDM7B$_{PHDJmjC}$ and KDM7B$_{PHD}$. A) Activity assay of KDM7B$_{PHDJmjC}$ (1.0 µM) with substrate H3$_{1-14}$K4me$_3$K9me$_2$ **4.20** (10 µM) and co-factors [Asc (100 µM), Fe(II)SO$_4$ (10 µM), 2OG (100 µM)] in buffer. Black is t= 0 min, red is t= 20 min. B) Correlation of ITC peptide H3$_{1-21}$K4 **4.1**, **4.11–4.14** binding data between KDM7B$_{PHDJmjC}$ and KDM7B$_{PHD}$ revealing an R^2 of 0.98.

4.4.9 KDM7B JmjC catalytic activity on tighter binding of the PHD; H3$_{1-21}$K4me$_3$K9me$_2$ 4.21 vs H3$_{1-21}$K4et$_3$K9me$_2$ 4.22

ITC analysis demonstrated that the singular PHD finger in KDM7B$_{PHD}$ as well as the combined KDM7B$_{PHDJmjC}$ manifests the capacity to bind H3$_{1-21}$K4me$_3$ **4.1** and higher alkylation states H3$_{1-21}$K4 **4.11–4.14**. The strongest binding protein-peptide complex was found with H3$_{1-21}$K4et$_3$ **4.14** providing at least 5-fold stronger binding relative to the natural interaction partner H3$_{1-21}$K4me$_3$ **4.1**. To investigate the effect of enhanced binding strength due to a stronger H3K4me$_3$–PHD interaction on the KDM7B JmjC-domain enzymatic activity, peptides H3$_{1-15}$K4me$_3$K9me$_2$ **4.21** and H3$_{1-15}$K4et$_3$K9me$_3$ **4.22** were produced using earlier described methods.

To understand the addition of methyl groups in the form of *N*$^{\varepsilon}$-dimethyl-lysine 9, the binding affinities of **4.21** and **4.22** against KDM7B$_{PHDJmjC}$ were evaluated. With the peptides in hand, experiments were conducted using Fluorescence Polarization (FP) to study the influence of the combined, K4 and K9 modifications to KDM7B$_{PHDJmjC}$ binding (Fig. 4.4.9 A) *(Performed by Dr. Yijia Zhang)*. In this FP

assay, the displacement of $H3_{1-15}K4me_3K9me_2$-FITC **4.23** in complex to $KDM7B_{PHDJmjC}$ revealed an IC_{50} of 2.6 µM (pIC_{50}: 5.58 ± 0.04) for $H3_{1-15}K4me_3K9me_2$ **4.21** but a ~4-fold increase in displacement is observed for $H3_{1-15}K4et_3K9me_2$ **4.22** with an IC_{50} of 0.70 µM (pIC_{50}: 6.16 ± 0.04). This demonstrates that the modified K4 peptide $H3_{1-15}K4et_3K9me_2$ **4.22** is also capable of binding to the $KDM7B_{PHDJmjC}$ protein and the increased binding strength relative to $H3_{1-15}K4me_3K9me_2$ **4.21** is in agreement to observations made using ITC for peptides that lack K9 methylation.

To understand the effect of *N*ε-triethyl-lysine 4 relative to *N*ε-trimethyl-lysine 4 on the JmjC-domain catalytic activity of $KDM7B_{PHDJmjC}$, Michaelis-Menten steady state enzyme kinetics were carried using MALDI-TOF *(performed by Dr. Yijia Zhang)* (Fig. 4.4.9 B). Both peptides revealed substrate inhibition and therefore the computational model was adjusted to take this in account. $H3_{1-15}K4me_3K9me_2$ **4.21** was found to have a K_M of 22.6 ± 4.6 µM and a V_{max} of 0.477 ± 0.048 µM·min^{-1}. Interestingly, the K_M^{App}, a constant associated to binding strength, was determined to be 19.9 ± 11.7 µM for $H3_{1-15}K4et_3K9me_2$ **4.22** which is similar to the K_M^{App} of $H3_{1-15}K4me_3K9me_2$ **4.21** within error. Due to limited sensitivity by the MALDI-TOF instrument, lower peptide concentrations of $H3_{1-15}K4et_3K9me_2$ **4.22** could not be acquired. As was observed for $H3_{1-15}K4me_3K9me_2$ **4.21**, $H3_{1-15}K4et_3K9me_2$ **4.22** demonstrated product inhibition was observed. The most striking result emerged from the V_{max}, obtaining 6.31 ± 2.68 µM·min^{-1} for $H3_{1-21}K4et_3K9me_2$. This is approximately 13-fold higher rate than observed for the natural substrate $H3_{1-15}K4me_3K9me_2$ **4.21**. The obtained data are an estimate based on the data found for $H3_{1-15}K4et_3K9me_2$ **4.22** and product $H3_{1-15}K4me_3K9me_0$ **4.24** as $H3_{1-15}K4et_3$ $K9me_1$ **4.23** intermediate state was demethylated at such high rate that only low levels were detected. This suggests strongly that $H3_{1-15}K4et_3K9me_1$ **4.28** is demethylated faster than $H3_{1-15}K4et_3K9me_2$ **4.21**. Furthermore, this work supports the notion that the reader domains can significantly modulate the activity of the associated catalytic domain, by regulating combinatorial PTM readout and act as 'anchoring' chromatin effector domain.

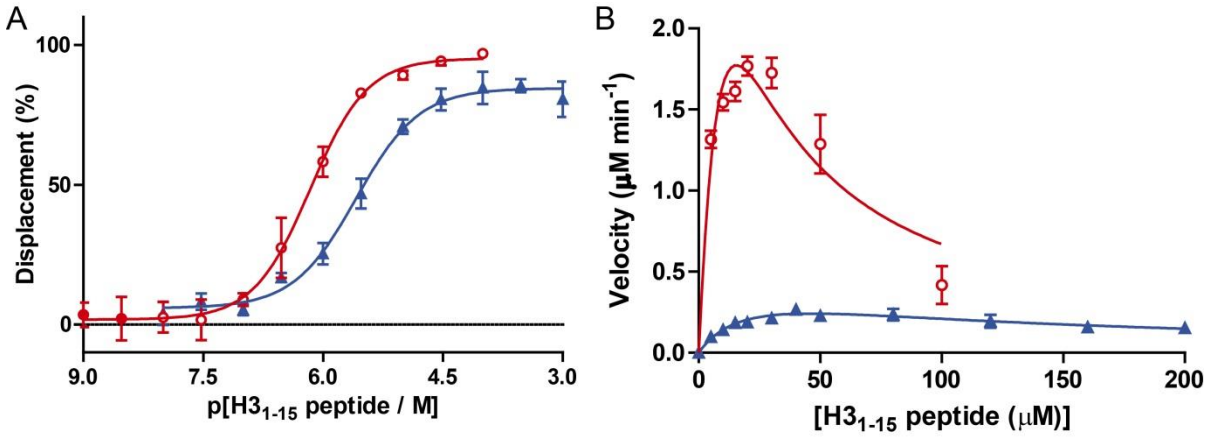


Figure 4.4.9: FP binding (A) and Michaelis-Menten steady state enzyme kinetics (B) of $H3_{1-15}K4me_3K9me_2$ 4.21 (blue triangles) against $H3_{1-15}K4et_3K9me_2$ 4.22 (red circles) $KDM7B_{PHDJmjC}$. Figure adapted from Dr. Yijia Zhang.[320] n= 3, error given as s.d.

4.5 Discussion

The examined panel of chromatin binding modules all demonstrated the capability to bind sterically enlarged trimethyl-N^{ε}-Lysine derivatives at least up to the triethyl level; only $BPTF_{PHD}$ showed comparable affinity for the largest alkyl group tested (n-butyl), in $H3_{1-21}K4^{n}bu_3$ **4.15**. The binding strength relative to $H3_{1-21}K4me_3$ **4.1** was, however, dependent on the protein. $CHD1_{DCD}$, $SPIN2B_{TTD}$, $SGF29_{TTD}$, $KDM4A_{TTD}$, $KDM5A_{PHD3}$ and $TAF3_{PHD}$ showed little variation (~ ±2.3-fold) in binding strength between the natural $H3_{1-21}K4me_3$ **4.1** mimic peptide and the bulkier analogues $H3_{1-21}K4$ **4.11–4.14** (annotated as Class A herein, Table 4.5.1). Whilst the binding strength is maintained for class A for bulkier N^{ε}-lysine analogues, strikingly, $KDM7B_{PHD}$, $KDM7B_{PHDJmjC}$ and $BPTF_{PHD}$ (annotated as Class B herein, Table 4.5.1) demonstrated an increase in binding strength of ~5–10 fold for $H3_{1-21}K4et_3$ **4.14** over $H3_{1-21}K4me_3$ **4.1**. This increase in binding strength can be partially attributed to an increase in enthalpy; however, there was no clear trend in entropy, showing improved or decreased data depending on the binding module-peptide complex.

Table 4.5.1: Chromatin binding modules are classified based on binding affinity observed between $H3_{1-21}K4et_3$ 4.14 and $H3_{1-21}K4me_3$ 4.1. Class A shows little variation between the modifications while Class B shows a clear increase in binding strength observed.

Class A			Class B
PHD	TTD	CHD1	PHD
$KDM5A_{PHD3}$	$KDM4A_{TTD}$	$CHD1_{DCD}$	$BPTF_{PHD}$
$TAF3_{PHD}$	$SPIN2B_{TTD}$		$KDM7B_{PHD}$
	$Sgf29_{TTD}$		$KDM7B_{PHDJmjC}$

Surprisingly, the obtained data reveal no clear distinction in N^{ε}-Lysine peptide recognition attributed to either reader domain TTD, PHD or DCD tested. This suggests that features allocated to a domain such as fold or metal binding do not exclusively contribute to the observations made between class A and class B. Additionally, this study confirms previous reports that the side chain of lysine 4 and multi-level methylation on N^{ε}-lysine 4 residue are critical to the formation of a strong binding complex.[244]

The absence of N^{ε}-methylated lysine 4 diminishes strong binding affinity, confirming that the strong complex formation found for peptides **4.1–4.7** and **4.11–4.14** is due to N^{ε}-lysine side chain.

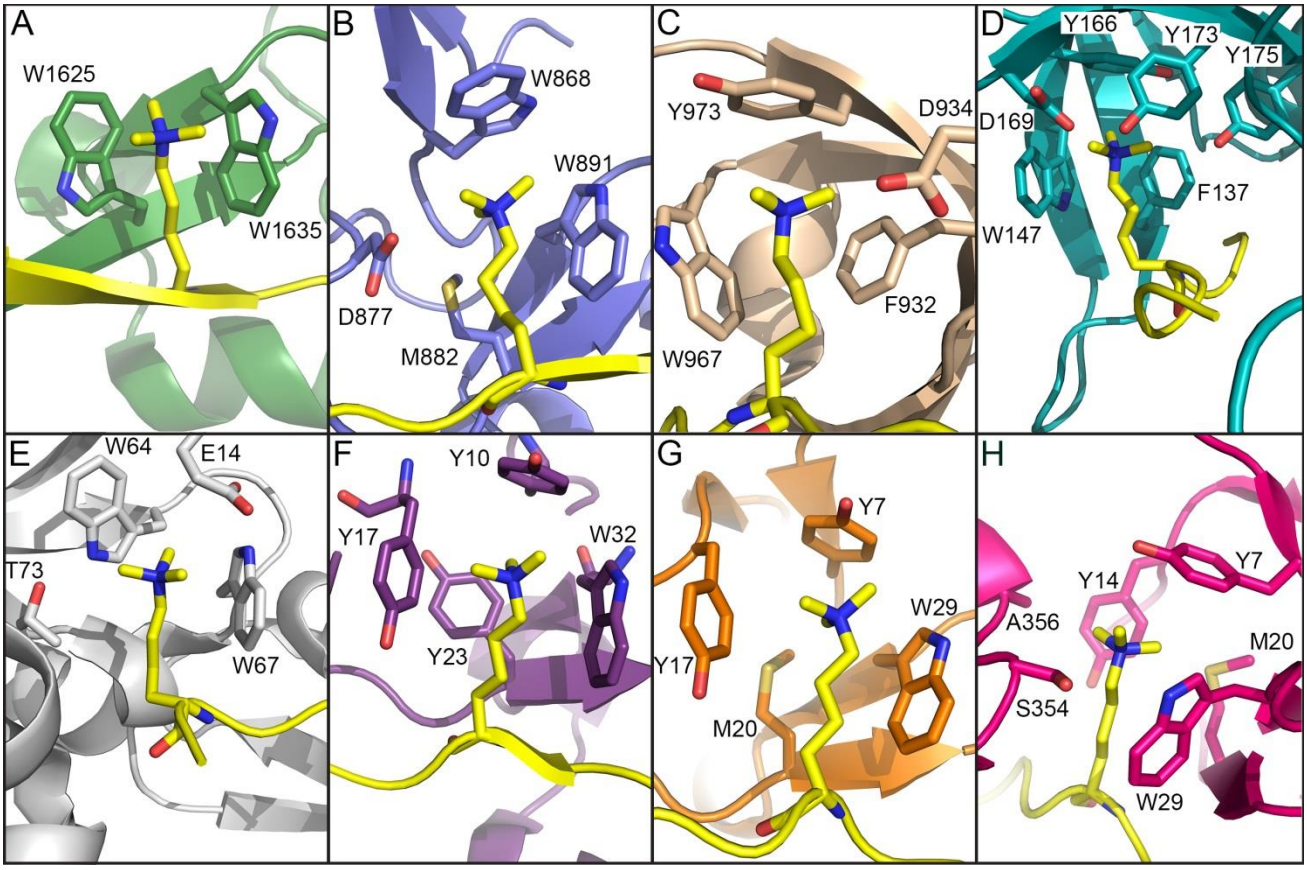


Figure 4.5.1: Chromatin binding modules class A (A-E) and class B (F-H) bound to H3K4me$_3$ peptide (yellow) with key 'aromatic cage' residues are highlighted important in N^{ε}-methyl-lysine binding. A) KDM5A$_{PHD3}$ (green) (PDB: 3GL6); B) TAF3$_{PHD}$ (light blue) (PDB: 5WXH); C) KDM4A$_{TTD}$ (beige) (PDB: 2GFA); D) SPIN2B$_{TTD}$ (cyan) (PDB: 5LUG); E) CHD1$_{DCD}$ (grey) (PDB: 2B2W); F) BPTF$_{PHD}$ (purple) (PDB: 2F6J); G) KDM7B$_{PHD}$ (orange); H) KDM7B$_{PHDJmjC}$ (pink). Oxygen, nitrogen and sulphur atoms are highlighted in red, blue and yellow, respectively.

Crystal structures of the protein-H3 peptide complexes reveal key residues in the histone H3 lysine 4 binding at the aromatic cages of each reader domain (Fig. 4.5.1). The residues interacting with the N^{ε}-Lysine are found to be typically aromatic amino acids: tryptophan, tyrosine and phenylalanine that are frequently accompanied with the hydrophobic residue methionine and acidic residues aspartate or glutamate. A major driving force to the formation of a strong complex is considered to be the cation-π interaction between the N^{ε}-Lysine and aromatic residues of the reader aromatic cage.[243] Furthermore, contributing factors are salt-bridging between the N^{ε}-trialkylated-lysine ammonium head group and

acidic residues in the aromatic cages. Recent literature also suggested that the release of 'high energetic water' from the reader aromatic cages maybe a strong contributors in binding strength gained.[244]

Two modes of aromatic cages positioning within the protein have been found: 1) surface groove and 2) cavity insertion.[6] Proteins in this panel tested, except for KDM7B$_{PHDJmjC}$, are found to be a surface groove motif providing potentially more space to bind to bulkier *N*$^{\varepsilon}$-trialkylated-lysine analogues. The construction of the aromatic cage may play an important role in the different observations found for the proteins in class A and B.

The aromatic cage architecture of KDM5A$_{PHD3}$ only contains W1625 and W1635 providing an open cage (Fig. 4.5.1 A) but most of the proteins tested contain three or four residues as found for TAF3$_{PHD}$, KDM4A$_{TTD}$, SGF29$_{TTD}$, CHD1$_{DCD}$, BPTF$_{PHD}$, KDM7B$_{PHD}$ and KDM7B$_{PHDJmjC}$ (Fig. 4.5.1 A-C, E-H). There is possibility that proteins SPIN2B$_{TTD}$, KDM7B$_{PHD}$, KDM4A$_{TTD}$ and BPTF$_{PHD}$ that possess sides covering *N*$^{\varepsilon}$-trimethyl-lysine, *e.g.* contain a more closed aromatic cage, would experience difficulty in accommodating more sterically demanding *N*$^{\varepsilon}$-lysine analogues. On the contrary, SPIN2B$_{TTD}$, KDM7B$_{PHD}$, KDM4A$_{TTD}$ and BPTF$_{PHD}$ were revealed to bind the bulkiest analogue tested H3$_{1-21}$K4^{n}bu$_3$ **4.15** with some loss in binding strength observed but maintaining a strong affinity. The proteins SPIN2B$_{TTD}$ and KDM7B$_{PHD}$ show that H3$_{1-21}$K4^{n}bu$_3$ **4.15** binds weaker than H3$_{1-21}$K4me$_3$ **4.1** but stronger than H3$_{1-21}$K4me$_2$ **4.2**. Surprisingly, CHD1$_{DCD}$ is the only protein that was observed to not bind H3$_{1-21}$K4^{n}bu$_3$ **4.15** with good affinity (>50 μM) despite binding of H3$_{1-21}$K4et$_3$ **4.14** with equivalent capacity as H3$_{1-21}$K4me$_3$ **4.1**.

Interestingly, the charge on the *N*$^{\varepsilon}$-lysine nitrogen is an effect of electron deficiency that is not solely allocated to the *N*$^{\varepsilon}$-lysine nitrogen atom but partially distributed over nearby alkylated members of the ammonium head group. By increasing the *N*$^{\varepsilon}$-alkyl group the charge is distributed through a larger space that potentially, in the right aromatic cage configuration, could provide better cation-π

interactions to all corresponding members. There is a possibility that this may be the case for class B proteins ($BPTF_{PHD}$, $KDM7B_{PHD}$, $KDM7B_{PHDJmjC}$) explaining the gain in binding strength observed.

It remains difficult to elucidate the changes observed in thermodynamic values acquired. The addition of a single methylene group (*i.e.* $H3_{1-21}K4me_3$ **4.1** vs. $H3_{1-21}K4me_{2,et}$ **4.11**) impacts multiple interactions such as: water solvation in the bound / unbound state, cation-π interactions and hydrogen bonding resulting in an sum of different factors that alter the values observed.

4.6 Conclusion

In conclusion, the nine chromatin binding modules tested show low micro-molar to low nano-molar binding affinity to their target histone H3 *N*ε-trimethyl-lysine 4 mimic peptides as demonstrated previously using ITC. Bulkier analogues $H3_{1-21}K4me_2$,et **4.11**, $H3_{1-21}K4me_2$,^{i}pr **4.12**, $H3_{1-21}K4me$,et_2 **4.13**, $H3_{1-21}K4et_3$ **4.14** bind strongly to $KDM4A_{TTD}$, $SPIN2B_{TTD}$, $SGF29_{TTD}$, $TAF3_{PHD}$, $KDM5A_{PHD3}$ and $CHD1_{DCD}$ showing similar binding affinity to the natural $H3_{1-21}K4me_3$ **4.1** substrate. The values reveal that most proteins tested decrease the enthalpy with bulkier *N*ε-trimethyl-lysine 4 head group but this is typically compensated by favourable entropy. Interestingly, $KDM7B_{PHD}$, $KDM7B_{PHDJmjC}$ and $BPTF_{PHD}$ show an increased in binding strength for $H3_{1-21}K4$ **4.11–4.14** relative to $H3_{1-21}K4me_3$ **4.1**. This observation is partially due to an enthalpic improvement, but it depends on the complex whether this entropic contribution is favourable or unfavourable. Obtained data suggests that there might be binding sub-group within the PHDs showing either maintained binding or increased binding with $H3_{1-21}K4$ **4.11–4.14**. The most sterically demanding $H3_{1-21}K4^{n}bu_3$ **4.15** bound with a similar magnitude to $H3_{1-21}K4me_3$ **4.1** ($KDM7B_{PHD}$, $BPTF_{PHD}$) or $H3_{1-21}K4me_2$ **4.2** ($SPIN2B_{TTD}$, $KDM5A_{PHD3}$). Interestingly, $CHD1_{DCD}$ revealed weak binding affinity (>50 µM) to $H3_{1-21}K4^{n}bu_3$ **4.15** while $H3_{1-21}K4et_3$ **4.14** was found to be a comparable binder to $H3_{1-21}K4me_3$ **4.1**. This suggests that $H3_{1-21}K4^{n}bu_3$ **4.15** is the level of *N*ε-lysine alkylation at which the overall binding affinity starts to decrease.

There was no difference found in the binding patterns between the surface groove binding protein $KDM7_{PHD}$ and the cavity insertion type binding of $KDM7B_{PHDJmjC}$ as determined using ITC. FP binding studies showed that *N*ε-lysine 9 methylation did not decrease the IC_{50} of $H3_{1-15}K4et_3K9me_2$ **4.22** relative to $H3_{1-15}K4me_3K9me_2$ **4.21** and was in agreement with ITC data recorded. Surprisingly, enhanced demethylation activity is observed for $KDM7B_{PHDJmjC}$ when the histone H3 *N*ε-lysine 4 steric bulk is increased. Steady state Michaelis-Menten enzyme kinetics demonstrated that the K_M is similar

for both peptides $H3_{1-15}K4et_3K9me_2$ **4.22** and $H3_{1-15}K4me_3K9me_2$ **4.21** but $H3_{1-15}K4et_3K9me_2$ **4.22** exhibited a 13-fold increase in V_{max}. It remains unclear to what extent this effect is due to enhanced binding or active site modulation / stabilization.

4.7 Future perspective

This investigation sheds light on the binding strength of N^{ε}-trimethyl-lysine chromatin binding modules by sterically enlarged N^{ε}-lysine residue and to what extent this is tolerated. This is interesting from a fundamental perspective but may have future applications. A study discussed in Chapter 3 reveals that using N^{ε}-triethyl-lysine over N^{ε}-trimethyl-lysine enables the targeting of KDM5A-PHD3 in the context of three domains acting on the same histone H3 peptide sequence. This result proposes that a larger N^{ε}-lysine could be a new tool for inhibitor development to target selectively chromatin N^{ε}-trimethyl-lysine binding modules over KDMs. Additionally, the relatively open chromatin binding modules observed from the crystal structures in combination with ITC values suggest that one of the groups of N^{ε}-trimethyl-lysine can be functionalized to exhibit a secondary function. This R-group could be modified to contain a linker motif connecting to a dye for biochemical assays or enable the incorporation of biotin tag for pull-down assays.

Studies of KDM7B$_{PHDJmjC}$ detailing the effect of the N^{ε}-triethyl-lysine 4 relative to N^{ε}-trimethyl-lysine 4 demonstrate an increase in enzymatic catalysis on the N^{ε}-dimethyl-lysine 9. Further studies may address the kinetic parameters for detailed N^{ε}-dimethyl-lysine 9 and N^{ε}-monomethyl-lysine 9 examination of the exact enzymatic parameters. This also highlights potential development strategy targeting the PHD finger to modulate the associate catalytic domain.

Evolution has allowed for the optimization of the binding modules to create a recognition complex that can distinguish between: N^{ε}-lysine, N^{ε}-acetylated-lysine and N^{ε}-methylated-lysine. As a result of this, it is most likely that N^{ε}-trimethyl-lysine is the preferred binding motif of the reader proteins tested. The fact that bulkier N^{ε}-Lysine analogues examined across all nine chromatin binding modules in this study demonstrated equal or higher binding affinity with sterically enlarger N^{ε}-trimethyl-lysine derivatives is, therefore, peculiar. All things considered, these data may imply the possibility that

ethylation, (iso)propylation or other forms of alkylation on *N*ε-lysine could be readable post translational modifications in the epigenetic chromatin environment. To date, none of these modifications have been reported on histone H3. The lack of observation of these PTMs in nature could be due to limited amounts of analyte obtained from biological samples diverting most experiments carried out to highly sensitive detection techniques such of mass-spectrometry. A limitation to this can be that isomers can be difficult to separate and modifications like one *N*ε-ethyl-lysine could be miss assigned as *N*ε-dimethyl-lysine residue within the same histone H3 protein (fragment). Alternatively global histone analysis by mass spectrometry may miss this due to the relatively low abundance of *N*ε-ethyl-lysine.

4.8 Materials and methods

4.8.1 General conditions

General methods and conditions can be found in Chapter 7.1 and the experimental conditions for compounds Synthesis of monomers Fmoc-Lys(me_2·HCl)-OH **4.10A**, Fmoc-Lys(pip·HCl)-OH **4.10F**, Boc-Lys(Cbz)-O^{t}Bu **4.21**, Boc-Lys(H)-O^{t}Bu **4.22**, Boc-Lys(et_3·I)-O^{t}Bu **4.24** and Fmoc-Lys(et_3·Cl)-OH **4.25** are described in Chapter 7.1.2. Peptide containing N^{ε}-acetyl-, N^{ε}-monomethyl- or trimethyl-lysine were produced using Fmoc-Lys(Ac)-OH, Fmoc-Lys(me, boc)-OH, (Novabiochem®, cat#: 8.52106.8500) or Fmoc-Lys(me_3·Cl)-OH (GL Biochem, cat#: 21152).

4.8.2 ITC analysis

Chromatin binding module acquisitions for $KDM5A_{PHD3}$, $TAF3_{PHD}$, $BPTF_{PHD}$, $SGF29_{TTD}$, $KDM7B_{PHDJMJC}$, $KDM4A_{TTD}$ against histone mimic peptides were measured on automated ITC200 (Malvern). $CHD1_{DCD}$, $SPIN2B_{TTD}$ and $KDM7B_{PHD}$ were measured using automated ITC-PEAQ (Malvern). The experimental buffer conditions for each protein is displayed in Table 4.8.1.

Table 4.8.1: ITC experimental buffer conditions

	Buffer
$BPTF_{PHD}$	TRIS (50 mM), NaCl (20 mM) in water (pH 7.5)
$CHD1_{CHD}$	HEPES (50 mM), NaCl (50 mM), TCEP (0.5 mM) in MilliQ (pH 7.5)
$KDM4A_{TTD}$	TRIS (50 mM) in water (pH 7.5)
$KDM5A_{PHD3}$	TRIS (50 mM), NaCl (20 mM) in water (pH 7.5)
$KDM7B_{PHD}$	HEPES (20 mM), NaCl (200 mM) in MilliQ (pH 7.5)
$KDM7B_{PHD-JMJC}$	HEPES (10 mM), NaCl (500 mM) in MilliQ (pH 7.5)
$SGF29_{TTD}$	TRIS (25 mM), NaCl (50 mM), DTT (1 mM) in water (pH 7.5)
$SPIN2B_{TTD}$	TRIS (10 mM), NaCl (100 mM), NaH_2PO_4 (1 mM) in water (pH 8.0)
$TAF3_{PHD}$	TRIS (50 mM) in water (pH 7.5)

SPIN2B$_{TTD}$ is stored in a glycerol buffer and is buffer exchanged prior to ITC analysis using Spin column (Bio-Spin™ 6 columns, BioRad, cat#: 732-6221) following manufactures recommended procedure. The spin column was primed 5× using SPIN2B$_{TTD}$ ITC buffer and kept cool on ice during operation. The concentration of the protein was determined using Nanodrop and estimated based on three reads.

Instrument settings

Reference power: ITC200: 5 µcal·sec^{-1}, ITC-PEAQ: 10 µcal·sec^{-1} with initial spacing of 60 sec and spacing between injections of 150 sec. ITC200 was 19 injections (2.0 µL, 0.5 µL·sec^{-1}) while ITC-PEAQ did one small injection (0.4 µL, 0.5 µL·sec^{-1}) followed by 18 normal injections (2.0 µL, 0.5 µL·sec^{-1}). The stirring speed was 750 rpm, the cell volume 280 µL and the cell temperature 25.0 °C. All plates were executed using manufactures default settings: Plate prerinse syringe clean.

4.8.3 Fluorescence polarization assay

Dose response curves containing either H3$_{1-15}$K4me$_3$K9me$_2$ **4.21** or H3$_{1-15}$K4et$_3$K9me$_2$ **4.22** were prepared in the presence of KDM7B$_{PHDJmjC}$ (600 nM), H3$_{1-21}$K4me$_3$K9me$_2$-FITC **4.23** (50 nM), Ni(II)SO$_4$ (100 µM) in buffer [HEPES (50 mM), NaCl (50 mM), Tween20® (0.1 v/v%) in MilliQ (pH 7.0)]. The solutions were plated (384- well, black, no-binding microplate (Greiner Bio-one)), incubated in the dark and FP was measured using PHERAstar FX plate reader (BMG).

4.8.4 Activity assay of KDM7B$_{PHDJmjC}$ with H3$_{1-14}$K4me$_3$K9me$_2$ substrate using MALDI-TOF

A solution of {sodium ascorbate (1.0 mM, 5.0 µL), Fe(II)SO$_4$ (100 µM, 5.0 µL), KDM7B$_{PHDJmjC}$ (10 µM, 5.0 µL) in buffer [HEPES (100 mM), NaCl (500 mM) in MilliQ (pH 7.5) (20 µL)]} was prepared followed by the addition of {disodium 2-oxo-glutarate (1.0 mM, 5.0 µL) and H3$_{1-21}$K4me$_3$K9me$_2$ **4.21**

(100 µM, 10.0 µL)}. The enzymatic reaction was performed in triplicate at 37 °C and time points (t= 0, 20, 40 and 60 min) were taken from the well (5.0 µL) and quenched in methanol (5.0 µL). The samples were spotted (1.0 µL) with [α-cyano-4-hydroxycinnamic acid (sat. 10 $mg \cdot mL^{-1}$) in acetonitrile/water (1:1, containing 0.1% trifluoroacetic acid)] (1.0 µL) on a Waters target plate and air dried. The samples were measured using Waters microTOF and acquired data were analysed using MassLynx (v. 4.00.00), Microsoft Excel (v. 14.0.7188.5002) and Graphpad Prism (v. 5.04).

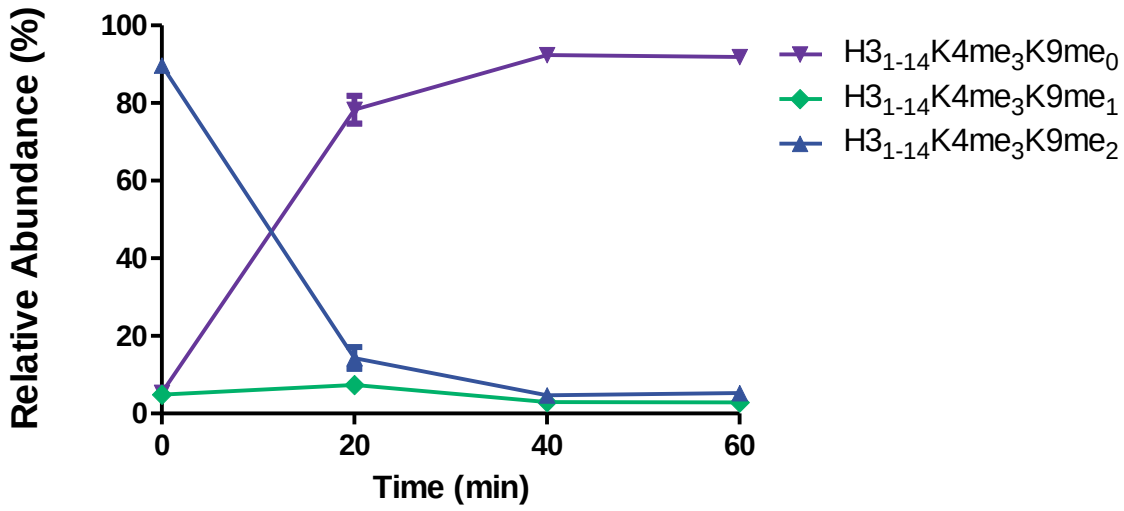
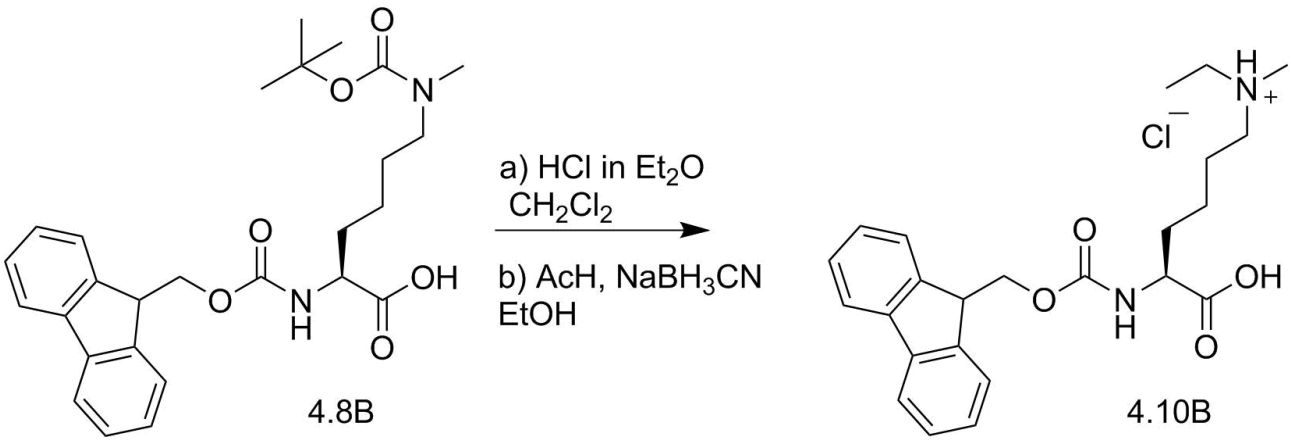


Figure 4.8.1: Time-course of KDM7B$_{PHDJmjC}$ demethylation of the substrate H3$_{1-14}$K4me$_3$K9me$_2$ 4.21.

4.8.5 Incorporation of Fmoc-trimethyl-lysine derivatives in histone mimic peptides using microwave assisted solid phase peptide synthesis

Histone mimic peptides are made according general method displayed in Chapter 7.1.3.

4.8.6 Synthesis of Fmoc-Lys(me$_{2,3}$)-OH derivatives

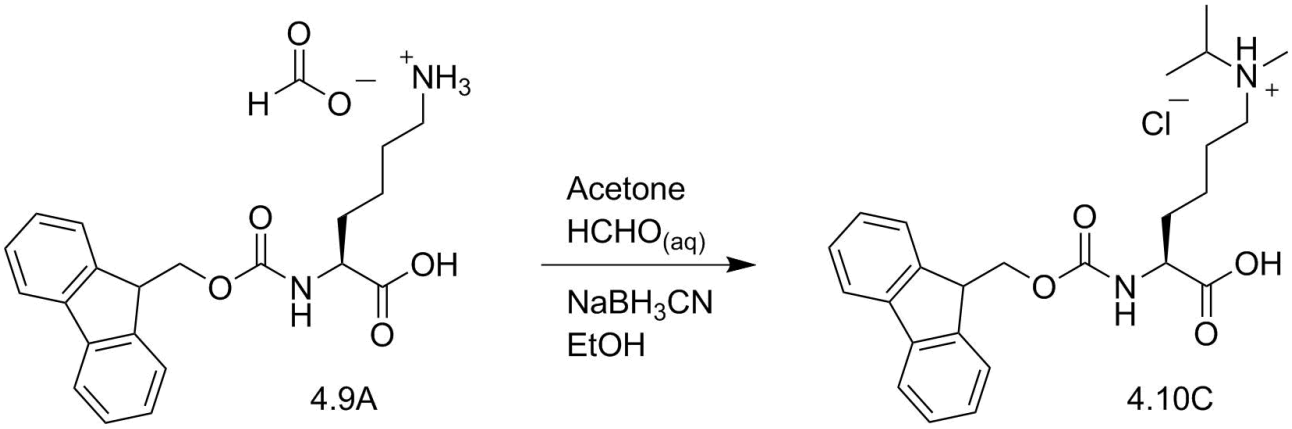


Fmoc-Lys(me, et·HCl)-OH[283]

(*S*)-2-[Fluorenylmethyloxycarbonyl]amino-6-[methyl, ethyl]amino-hexanoic acid hydrochloric acid 4.10B

The method to produced **4.10B** was adapted from a previous report procedure with minor alterations[283]. To a solution of Fmoc-Lys(me, Boc)-OH (573 mg, 1.19 mmol, 1.0 equiv.) in CH_2Cl_2 (5 mL, 0.24 M) was added HCl in Et_2O (2.0 mL, 2.0 M, 4.0 mmol, 3.3 equiv.) while stirring at RT. The mixture was continued stirring for 16 hrs followed by the removal of volatiles under reduced pressure. To a solution of Fmoc-Lys(me·HCl)-OH in EtOH (15 mL, 0.08 M), acetaldehyde (404 µL, 7.2 mmol, 6.0 equiv.) was added drop wise followed by the slow addition of $NaBH_3CN$ (169 mg, 2.69 mmol, 2.3 equiv.). The mixture was stirred for 1.5 hrs and progress of the reaction was monitored by ESI-MS. TFA (380 µL, 4.57 mmol, 3.84 equiv.) was added to quench the reaction and 1 mL of H_2O was added followed by stirring at RT for 15 min. Volatiles were removed *in vacuo* and the crude was loaded on a Grace 40 g R_f-C_{18} column for HPLC purification. Solution A: [TFA in MilliQ], solvent B: [TFA (0.1 v/v%) in acetonitrile]. Program: 3 CV (5% B) equilibration, loading ~15 mL (25% B), 2 CV (5% B), 20 CV (5–100% B). Selected fractions were lyophilized (yielding, 326 mg, 0.79 mmol, 67%), combined and dissolved in CH_2Cl_2 (5 mL, 0.16 M). To the solution was added slowly HCl in Et_2O (1 mL, 2 M, 2 mmol, 1.68 equiv.) while stirring and volatiles were removed *in vacuo*. Azeotropically Et_2O was removed with CH_2Cl_2 *in vacuo* (4× 10 mL). mp: 74 °C (decomp.); IR υ_{max}: 3392, 3061, 2944, 1708,

1533, 1201 cm^{-1}; $[\alpha]^{25}_D$ −1.3 (c = 0.011 g·mL^{-1} MeOH); ^{1}H NMR 400 MHz, Solvent: DMSO-d_6, δ (ppm): 12.50 (s, 1 H, $NH^{\alpha}CHCOO\underline{H}$), 10.10 (s, 1 H, $^{\varepsilon}N\underline{H}(CH_2CH_3)CH_3$), 7.90 (d, 2 H, *J*= 7.5 Hz, 2× Fmoc-$C_4\underline{H}$), 7.73 (q, 2 H, *J*= 4.0, 3.5, 4.0 Hz, 2× Fmoc-$C_1\underline{H}$), 7.68 (d, 1 H, *J*= 8.0 Hz, $N\underline{H}^{\alpha}CHCOOH$), 7.42 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-$C_3\underline{H}$), 7.33 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-$C_2\underline{H}$), 4.34-4.26 (m, 2 H, $OC\underline{H}_2CH$), 4.26-4.19 (m, 1 H, $OCH_2C\underline{H}$), 3.94 (dt, 1 H, *J*= 9.0 4.5 Hz, $NH^{\alpha}C\underline{H}COOH$), 3.18-3.07 (m, 1 H, $^{\delta}CH_2{}^{\varepsilon}CH\underline{H}NH$), 3.06-2.96 (m, 2 H, $NH(C\underline{H}_2CH_3)CH_3$), 2.96-2.84 (m, 1 H, $^{\delta}CH_2{}^{\varepsilon}C\underline{H}HNH$), 2.67 (d, 3 H, *J*= 5.0 Hz, $NH(CH_2CH_3)C\underline{H}_3$), 1.80-1.55 (m, 4 H, $^{\alpha}CH^{\beta}C\underline{H}_2{}^{\gamma}CH_2{}^{\delta}C\underline{H}_2{}^{\varepsilon}CH_2$) 1.43-1.27 (m, 2 H, $^{\beta}CH_2{}^{\gamma}C\underline{H}_2{}^{\delta}CH_2$), 1.20 (t, 3 H, *J*= 7.0 Hz, $NH(CH_2C\underline{H}_3)CH_3$); ^{13}C NMR 100 MHz (MeOD-d_4) δ (ppm): 175.5, 158.7, 145.1, 142.5, 128.8, 128.2, 126.2, 120.9, 67.9, 56.4, 54.9, 52.3, 48.4, 39.7, 32.0, 24.6, 24.0, 9.6; ESI-MS in MeOH $[M+H]^+$, calc: 411.2 found: 411.2; HR-MS (ESI+) in MeOH as $C_{24}H_{31}N_2O_4^+$, calc: 411.2278 found: 411.2261.

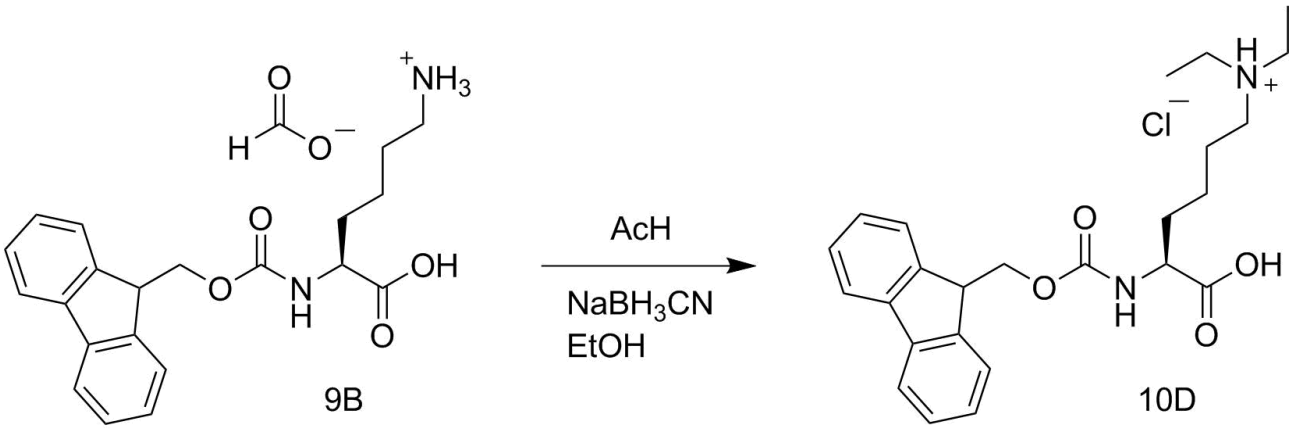


Fmoc-Lys(me, ipr·HCl)-OH[283]

(*S*)-2-[Fluorenylmethyloxycarbonyl]amino-6-[methyl, isopropyl]amino-hexanoic acid hydrochloric acid 4.10C

Adjustments were made according to previous reported protocol for the production of **4.10C**.[283] Fmoc-Lys(HCOOH)-OH (1.00 g, 2.5 mmol, 1.0 equiv.) was suspended in ethanol (20 mL, 0.125 M) and a portion of acetone (1.0 mL, 13.6 mmol, 5.4 equiv.) was added slowly. The reduction process was initiated using $NaBH_3CN$ (172 mg, 2.7 mmol, 1.08 equiv.) and the mixture was stirred for 0.75 hrs. After ESI-MS indicated reaction completion, formaldehyde solution (0.90 mL, 37% sol in H_2O, 15.0

mmol, 6.0 equiv.) was added drop wise followed by another portion of $NaBH_3CN$ (169 mg, 2.7 mmol, 1.08 equiv.). ESI-MS indicated completion of the methylation after 1.0 hrs and the reaction was quenched with TFA (0.30 mL). The mixture was passes through a filter, the solids were washed with ethanol (3× 10 mL) and the supernatant was evacuated under reduced pressure. The mixture was loaded was solubilized in a minimal acetonitrile water mixture and reverse phase purified (KPC18-HS 30 gram R_f-C_{18} by Biotage). Program: 3 CV (5% B) equilibration, 2 CV (5% B), 20 CV (5%–100% B). The fractions were pooled as indicated by UV trace and ESI-MS and lyophilized yielding a white solid (871 mg, 82%). The solid was dissolved in CH_2Cl_2 (~10 mL) and protonated using HCl in Et_2O solution (2.0 mL, 2.0 M, 4.0 mmol, 1.60 equiv.). Volatiles were removed *in vacuo* and 2× CH_2Cl_2 (~10 mL) was added and removed under reduced pressure to give a white solid. mp: 78 °C (decomp); IR υ_{max}: 3400, 2976, 2362, 1709, 1532, 1251, 1201 cm^{-1}; $[\alpha]^{25}_D$ −0.6 (c = 0.020 g·mL^{-1} MeOH); ^1H NMR 400 mHz Solvent: DMSO-d_6, δ (ppm): 12.67 (s, 1 H, $NH^{\alpha}CHCOO\underline{H}$), 9.84 (s, 1 H, $^{\varepsilon}CH_2N\underline{H}CH_3(CH_3CHCH_3)$), 7.92 (d, 2 H, *J*= 8.0 Hz, 2× Fmoc-$C_4\underline{H}$), 7.75 (dd, 2 H, *J*= 4.5 3.0 4.5 Hz, 2× Fmoc$C_1\underline{H}$), 7.69 (d, 1 H, *J*= 8.0 Hz, $N\underline{H}^{\alpha}CHCOOH$), 7.44 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-$C_3\underline{H}$), 7.35 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-$C_2\underline{H}$), 4.36-4.20 (m, 3 H, $OC\underline{H}_2C\underline{H}$), 3.95 (m, 1 H, $NH^{\alpha}C\underline{H}COOH$), 3.55-3.45 (m, 1 H, $^{\varepsilon}NHC(CH_3\underline{H}CH_3)$) 3.10-2.98 (m, 1 H, $^{\delta}CH_2{}^{\varepsilon}C\underline{H}HNH$), 2.97-2.84 (m, 1 H, $^{\delta}CH_2{}^{\varepsilon}CH\underline{H}NH$), 2.62 (d, 3 H, *J*= 5.0 Hz, $NHC\underline{H}_3(CH_3CHCH_3)$), 1.82-1.60 (m, 4 H, $^{\alpha}CH^{\beta}C\underline{H}_2{}^{\gamma}CH_2{}^{\delta}C\underline{H}_2{}^{\varepsilon}CH_2$), 1.45-1.31 (m, 2 H, $^{\beta}CH_2{}^{\gamma}C\underline{H}_2{}^{\delta}CH_2$), 1.25 (d, 3 H, *J*= 6.0 Hz, $NHCH_3(C\underline{H}_3CHCH_3)$), 1.20 (dd, 3 H, *J*= 7.0, 2.0 Hz, $NHC(CH_3HC\underline{H}_3)$); ^{13}C NMR 100 mHz Solvent: DMSO-d_6, δ (ppm): 174.3, 156.7, 144.3, 141.2, 128.1, 127.5, 125.7, 120.6, 66.1, 56.3, 54.1, 52.2, 47.1, 35.0, 30.7, 23.9, 23.4, 17.2, 15.6; ESI-MS in MeOH, $[M+H]^+$ calc.: 425.2 found: 425.3; HR-MS (ESI+) in MeOH as $C_{25}H_{33}N_2O_4^+$, $[M+H]^+$: calc.: 425.2435, found: 425.2422.

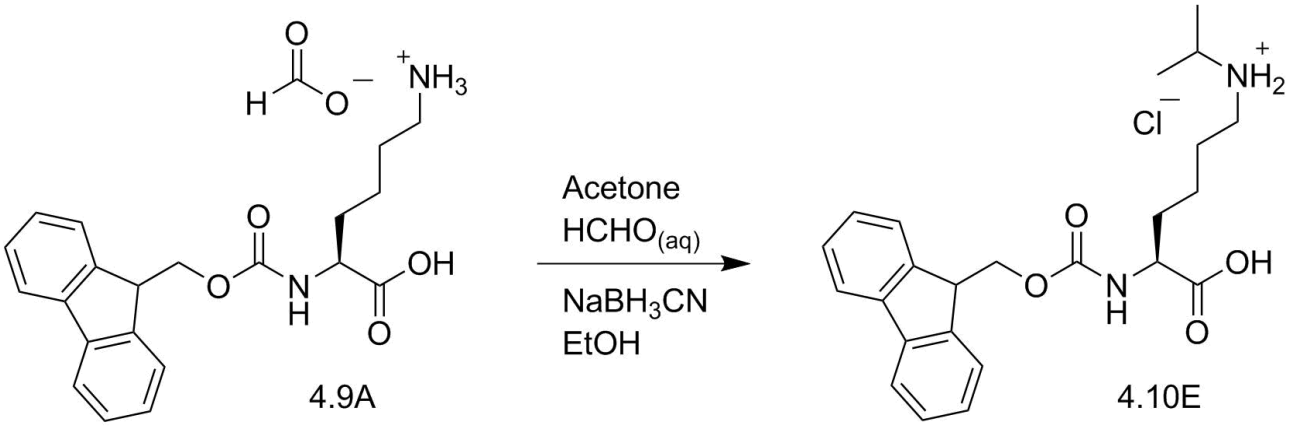


Fmoc-Lys(et$_2$·HCl)-OH[283]

(*S*)-2-[Fluorenylmethyloxycarbonyl]amino-6-[diethyl]amino-hexanoic acid hydrochloric acid 4.10D

Previous protocol was used as inspirational source for the synthesis of **4.10D**.[283] To a suspension of Fmoc-Lys(HCOOH)-OH (801 mg, 1.93 mmol, 1.0 equiv.) in EtOH (20 mL, 0.1 M), acetaldehyde (1.0 mL, 17.8 mmol, 9.23 equiv.) was added followed by slow addition of $NaBH_3CN$ (268 mg, 4.25 mmol, 2.2 equiv.). The mixture was stirred for 45 min creating over time a clear solution. TFA (300 µL) was added drop wise and the mixture was stirred for 3 min. *In vacuo* volatiles were removed and the crude was loaded on a Grace 40 g R_f-C_{18} column. Solution A: [TFA in MilliQ], solvent B: [TFA (0.1 v/v%) in acetonitrile]. Program: 3 CV (5%) equilibration, loading (~6 mL, 25%), 2 CV (5%), 20 CV (5–100%) MeCN in H_2O (containing 0.01% TFA). Selected fractions were lyophilised yielding 564 mg of white solid (1.33 mmol, 69%). The solid was dissolved in CH_2Cl_2 (5 mL, 0.27 M) and to the solution was added HCl in Et_2O (2.0 mL, 2.0 M, 4.0 mmol, 3.0 equiv.) and volatiles were removed *in vacuo*. Azeotropically Et_2O was removed with CH_2Cl_2 *in vacuo* (4× 10 mL). mp: 76 °C (decomp.); IR υ_{max}: 3407, 2978, 2362, 1695, 1200, 1130 cm^{-1}; $[\alpha]^{25}_D$ +6.5 (c = 0.010 g·mL^{-1} MeOH); ^{1}H NMR 400 MHz Solvent: DMSO-d_6, δ (ppm): 12.67 (s, 1 H, NH$^{\alpha}$CHCOO<u>*H*</u>), 7.90 (d, 2 H, *J*= 7.5 Hz, 2× Fmoc-C$_4$<u>*H*</u>), 7.72 (q, 2 H, *J*= 5.0, 2.5, 5.0 Hz, 2× Fmoc-C$_1$<u>*H*</u>), 7.67 (d, 1 H, *J*= 8.0 Hz, N<u>*H*</u>$^{\alpha}$CHCOOH), 7.43 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-C$_3$<u>*H*</u>), 7.33 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-C$_2$<u>*H*</u>), 4.36-4.18 (m, 3 H, OC<u>*H*</u>$_2$C<u>*H*</u>), 3.97 (m, 1 H, NH$^{\alpha}$C<u>*H*</u>COOH), 3.21 (q, 6 H, *J*= 7.0 Hz, $^{\varepsilon}$N(C<u>*H*</u>$_2$CH$_3$)$_2$), 3.16-3.00 (m, 2 H, $^{\delta}$CH$_2$$^{\varepsilon}$C<u>*H*</u>$_2$N),

1.83-1.71 (m, 1 H, $^{\alpha}CH^{\beta}CH\underline{H}^{\gamma}CH_2$), 1.71-1.48 (m, 3 H, $^{\alpha}CH^{\beta}C\underline{H}H^{\gamma}CH_2{}^{\delta}C\underline{H}_2{}^{\varepsilon}CH_2$), 1.42-1.28 (m, 2 H, $^{\beta}CH_2{}^{\gamma}C\underline{H}_2{}^{\delta}CH_2$), 1.15 (t, 6 H, *J*= 7.0 Hz $N(CH_2C\underline{H}_3)_2$); ^{13}C NMR 100 MHz Solvent: MeOD-d_4, δ (ppm): 177.8, 158.3, 145.2, 142.5, 128.8, 128.2, 126.2, 120.9, 67.8, 56.2, 52.7, 48.4, 48.1, 32.9, 24.3, 23.8, 9.0; ESI-MS in MeOH $[M+H]^+$, calc.: 425.2 found: 425.2; HRMS (ESI+) in MeOH $[M+H]^+$ as $C_{25}H_{33}N_2O_4^+$, calc: 425.2435 found: 425.2421.

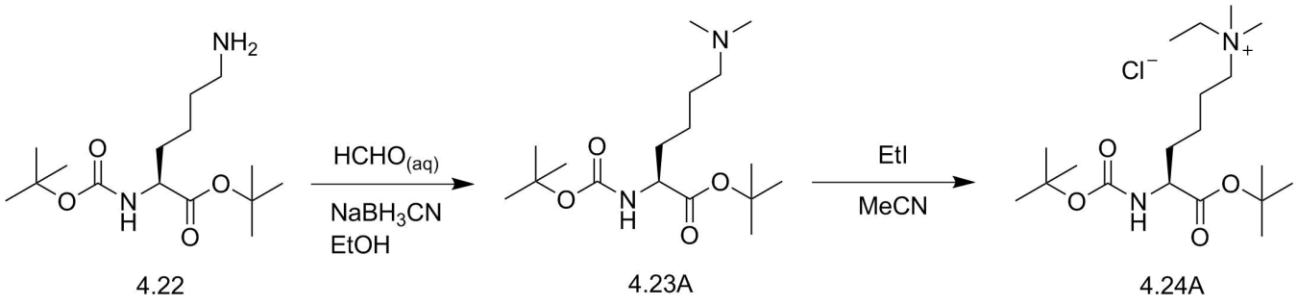


Fmoc-Lys(^{i}pr·HCl)-OH[283]

(*S*)-2-[Fluorenylmethyloxycarbonyl]amino-6-[isopropyl]amino-hexanoic acid hydrochloric acid 4.10E

The method for producing **4.10E** was adapted from a previous report procedure with minor alterations.[283] Fmoc-Lys(ipr HCl)-OH was prepared using Fmoc-Lys(HCOOH)-OH (1.012 g, 2.5 mmol, 1.0 equiv.) $NaBH_3CN$ (179 mg, 2.7 mmol, 1.08 equiv.) and acetone (1 mL, 13.6 mmol, 5.4 equiv.) following for the procedure of Fmoc-Lys(ipr HCl)-OH; however, the methylation step was not performed and formaldehyde and 2nd $NaBH_3CN$ was not added. The mixture was quenched after 0.75 hrs. mp: 98 °C (decomp.); IR υ_{max}: 3400, 2954, 2362, 1703, 1528, 1253, 1050 cm^{-1}; $[\alpha]^{25}_D$ −3.3 (c = 0.014 g·mL^{-1} MeOH); ^{1}H NMR 400 MHz Solvent: DMSO-d_6, δ (ppm): 8.34 (s, 1 H, $^{\varepsilon}CH_2N\underline{H}_2(CH_3CHCH_3)$), 7.90 (d, 2 H, *J*= 7.5 Hz, 2× Fmoc-$C_4\underline{H}$), 7.72 (dd, 2H, *J*= 4.5, 3, 4.5 Hz, 2× Fmoc-$C_1\underline{H}$), 7.55 (d, 1 H, *J*= 8.0 Hz, $N\underline{H}^{\alpha}CHCOOH$), 7.42 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-$C_3\underline{H}$), 7.33 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-$C_2\underline{H}$), 4.35-4.19 (m, 3 H, $OC\underline{H}_2CH$, $OCH_2C\underline{H}$), 3.91 (dt, 1 H, *J*= 9.0 4.5 Hz, $NH^{\alpha}C\underline{H}COOH$), 3.25 (hept, 1 H, *J*= 6.5 Hz, $^{\varepsilon}CH_2NH_2(CH_3C\underline{H}CH_3)$), 2.90-2.81 (m, 1 H,

$^{\delta}CH_2{}^{\varepsilon}C\underline{H}_2NH_2(CH_3CHCH_3)$), 1.80-1.68 (m, 1 H, $^{\alpha}CH^{\beta}C\underline{H}H^{\gamma}CH_2$), 1.68-1.47 (m, 3 H,$^{\alpha}CH^{\beta}CH\underline{H}^{\gamma}CH_2{}^{\delta}C\underline{H}_2{}^{\varepsilon}CH_2$), 1.42-1.30 (m, 1 H, $^{\beta}CH_2{}^{\gamma}C\underline{H}_2{}^{\delta}CH_2$), 1.19 (d, 6 H, *J*= 6.5 Hz, $^{\varepsilon}CH_2NH(C\underline{H}_3CHC\underline{H}_3)$); ^{13}C NMR 100 MHz Solvent: MeOD-d$_4$, δ (ppm): 175.6, 158.7, 145.1, 142.5, 128.8, 128.1, 126.2, 120.9, 67.9, 55.0, 51.7, 48.3, 40.5, 32.0, 28.0, 23.9, 19.2; ESI-MS in MeOH, $[M+H]^+$ calc.: 411.2, found: 411.3; HR-MS (ESI+) in MeOH as $C_{24}H_{31}N_2O_4^+$, $[M+H]^+$ calc.: 411.2278, found: 411.2265.

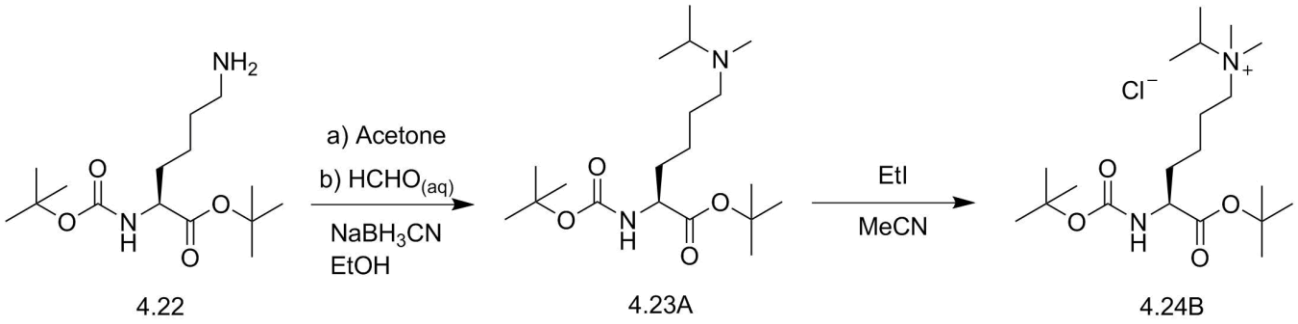


Boc-Lys(me$_2$, et·I)-O^{t}Bu

tertbutyl (*S*)-2-[tertbutoxycarbonyl]amino-6-[dimethyl ethyl]amino-hexanoate iodide 4.24A

To a solution of Boc-Lys(H)-O^{t}Bu **4.22** (567 mg, 1.91 mmol, 1.0 equiv.) in EtOH (9.6 mL, 0.1 M) was added drop wise HCHO (850 µL, 37% sol., 11.4 mmol, 6.0 equiv.) followed by the slow addition of $NaBH_3CN$ (265 mg, 4.0 mmol, 2.1 equiv.). The solution was stirred for 5 hrs followed by the addition HCHO (425 µL, 37% sol., 5.7 mmol, 3.0 equiv.) and $NaBH_3CN$ (115 mg, 1.83 mmol, 0.96 equiv.) and stirred for 1 hrs. sat. $NaHCO_3$ sol. (40 mL) was combined with the slightly milkish suspension and the mixture was extracted with CH_2Cl_2 (3× 50 mL). After drying the organic layer over $MgSO_4$, volatiles were removed under reduced pressure obtaining crude colourless oil. To the crude Boc-Lys(me$_2$)-O^{t}Bu was added anhydrous MeCN (4 mL, 0.47 M) and iodoethane (350 µL, 4.35 mmol, 2.3 equiv.) followed by heating to 80 °C for 40 min equipped with cooler while stirring. The solution was allowed to cooling down to RT and the crude was concentrated *in vacuo*. Purification was achieved with HPLC. Grace column 40g [SiO_2], Solvent A: CH_2Cl_2, solvent B: MeOH. Program: 3 CV equilibration, loading 4 mL (0% B), 16CV (0 to 15% B) and the correct fractions were combined and volatiles were removed under

reduced pressure obtaining a colourless oil (326 mg, 0.67 mmol, 35%). IR υ_{max}: 3448, 2976, 1707, 1366, 1156 cm^{-1}; $[\alpha]^{25}_D$ −11.0 (c = 0.011 g·mL^{-1} MeOH); ^{1}H NMR 400 MHz Solvent: $CDCl_3$-d_1, δ (ppm): 5.14 (bd, 1 H, *J*= 8.5 Hz, N*H*$^{\alpha}$CHCOO), 4.18-4.07 (m, 1 H, NH$^{\alpha}$C*H*COO), 3.71-3.57 (m, 3 H, $^{\delta}CH_2{}^{\varepsilon}C\underline{H}HN(CH_3)_2(C\underline{H}_2CH_3)$), 3.46-3.38 (m, 1 H, $^{\delta}CH_2{}^{\varepsilon}CH\underline{H}N(CH_3)_2(CH_2CH_3)$), 3.34 (s, 6 H, $^{\varepsilon}CH_2N(C\underline{H}_3)_2(CH_2CH_3)$), 1.95-1.75 (m, 3 H, $^{\alpha}CH^{\beta}C\underline{H}H^{\gamma}CH_2{}^{\delta}C\underline{H}_2{}^{\varepsilon}CH_2$), 1.75-1.60 (m, 3 H, $^{\alpha}CH^{\beta}CH\underline{H}^{\gamma}C\underline{H}_2{}^{\delta}CH_2$), 1.47 (s, 9 H, $^{\alpha}CHCOOC(C\underline{H}_3)_3$), 1.43 (s, 9 H, $^{\alpha}NHCOOC(C\underline{H}_3)_3$), 1.44-1.39 (m, 3 H, $NCH_3(CH_2C\underline{H}_3)_2$); ^{13}C NMR 100 mHz Solvent: $CDCl_3$-d_1, δ (ppm): 171.8, 155.7, 82.2, 79.9, 57.7, 49.2, 48.6, 40.0 32.3, 28.4, 28.1, 26.7, 22.7, 8.7; ESI-MS (+) in MeOH, $[M+H]^+$ calc: 359.3 found: 359.3; HR-MS ESI (+) in MeOH as $C_{19}H_{39}N_2O_4^+$, $[M+H]^+$, calc: 359.2904, found: 359.2902.

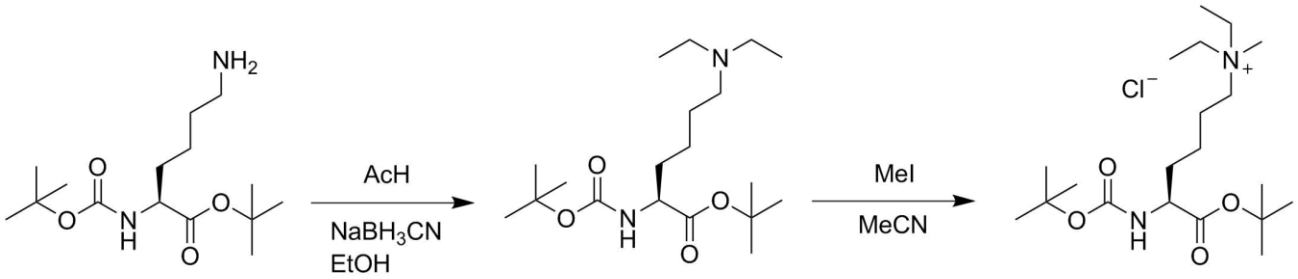


Boc-Lys(me$_2$, ^{i}pr·I)-OH

tertbutyl (S)-2-[tertbutoxycarbonyl]amino-6-[dimethyl, isopropyl]amino-hexanoate iodide 4.24B

To a solution of Boc-Lys(H)-O^{t}Bu (604 mg, 2.0 mmol, 1.0 equiv.) in MeOH (8.0 mL, 0.25 M) was added drop wise acetone (750 µL, 10.2 mmol, 5.0 equiv.) followed by the slow addition of $NaBH_3CN$ (137 mg, 2.2 mmol, 1.1 equiv.). The solution was stirred for 1 hrs monitoring with ESI-MS, followed by the addition HCHO (744 µL, 37% sol., 10.0 mmol, 5.0 equiv.) and slowly $NaBH_3CN$ (137 mg, 2.2 mmol, 1.1 equiv.) was added. The mixture was stirred for 1 hrs and sat. $NaHCO_3$ sol. (40 mL) was combined and was extracted with CH_2Cl_2 (3× 35 mL). After drying the organic layer over $MgSO_4$, volatiles were removed under reduced pressure obtaining a crude oil. To a solution of Boc-Lys(me, ^{i}pr)-O^{t}Bu in anhydrous MeCN (6 mL, 0.5 M) was added iodomethane (380 µL, 6.0 mmol, 3.0 equiv.)

followed by heating to 80 °C for 1 hrs equipped with cooler while stirring. The solution was allowed to cool to RT followed by concentrating *in vacuo*. The crude was purified on Grace column 24 g [SiO_2]. Solvent A: CH_2Cl_2, solvent B: MeOH. 3 CV (0% B) equilibration, loading 5 mL (0% B), 12 CV (0%–10% B) and the correct fractions were combined and volatiles were removed under reduced pressure obtaining a slightly yellowish hydroscopic solid (467 mg, 48%, 0.95 mmol). mp: 137.6–139.2 °C; IR υ_{max}: 3447, 2980, 1705, 1366, 1153 cm^{-1}; $[\alpha]^{25}_D$ −14.2 (c = 0.011 g·mL^{-1} $CHCl_3$); ^{1}H NMR 400 MHz Solvent: $CDCl_3$-d_1, δ (ppm): 5.13 (bd, 1 H, *J*= 7.5 Hz, N*H*$^{\alpha}$CHCOO), 4.17 (bdt, 1 H, *J*= 5.0 13 Hz, NH$^{\alpha}$C*H*COO), 3.98 (hept., 1 H, *J*= 6.4 Hz, N(CH_3C*H*CH_3)(CH_3)$_2$), 3.63 (dt, 1 H, *J*= 5.0 13.0 Hz, $^{\delta}CH_2{}^{\varepsilon}$CH*H*N($CH_3CHCH_3$)($CH_3$)$_2$), 3.50-3.37 (m, 1 H, $^{\delta}CH_2{}^{\varepsilon}$C*H*HN ($CH_3CHCH_3$)($CH_3$)$_2$), 3.27 (s, 6 H, $^{\varepsilon}$N(CH_3CHCH_3)(C*H*$_3$)$_2$), 2.00-1.76 (m, 3 H, $^{\alpha}$CH$^{\beta}$CH*H*$^{\gamma}CH_2{}^{\delta}$C*H*$_2{}^{\varepsilon}CH_2$), 1.66-1.50 (m, 3 H, $^{\beta}$C*H*H$^{\gamma}$C*H*$_2{}^{\delta}CH_2$), 1.47 (s, 9 H, $^{\alpha}$CHCOOC(C*H*$_3$)$_3$), 1.43 (s, 9 H, NHCOOC(C*H*$_3$)$_3$); ^{13}C NMR 125 mHz solvent: $CDCl_3$-d_1, δ (ppm): 171.6, 158.8, 82.6, 80.0, 65.3, 62.6, 53.1, 48.7, 32.6 28.5, 28.2, 22.3, 21.8, 17.1, 17.0; ESI-MS (+) in MeOH, $[M+H]^+$ calc: 373.3 found: 373.3; HR-MS ESI (+) in MeOH as $C_{20}H_{41}N_2O_4^+$, $[M+H]^+$, calc: 373.3061 found: 373.3069.

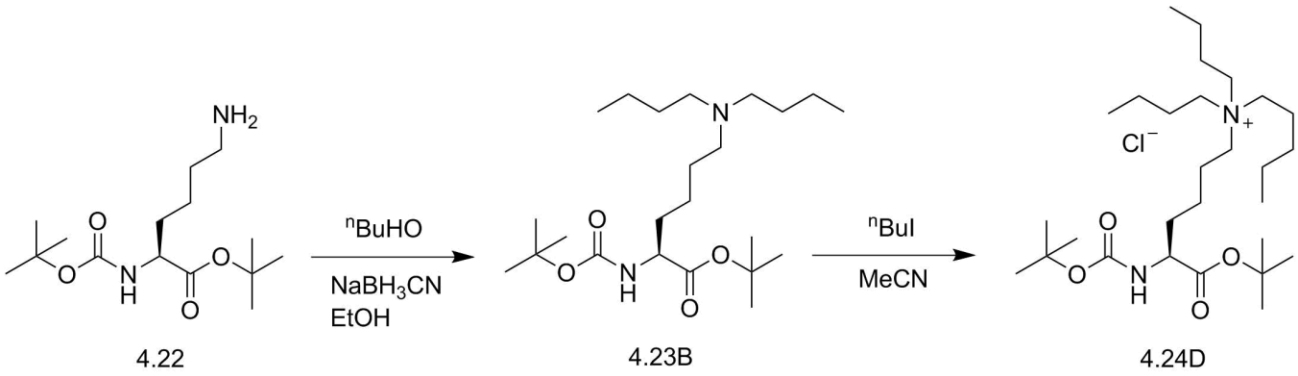


Boc-Lys(me, et_2·I)-O^{t}Bu

tertbutyl (S)-2-[tertbutoxycarbonyl]amino-6-[diethyl methyl]amino-hexanoate iodide 4.24C

To a solution of Boc-Lys(H)-O^{t}Bu (587 mg, 1.94 mmol, 1.0 equiv.) in MeOH (8 mL, 0.25 M) was added drop wise acetaldehyde (680 µL, 12.0 mmol, 6.2 equiv.) followed by the slow addition of $NaBH_3CN$ (283 mg, 4.4 mmol, 2.2 equiv.). The solution stirred for 40 min followed by the addition of sat. $NaHCO_3$ sol. (40 mL) and the mixture was extracted with CH_2Cl_2 (3× 35 mL). After drying the

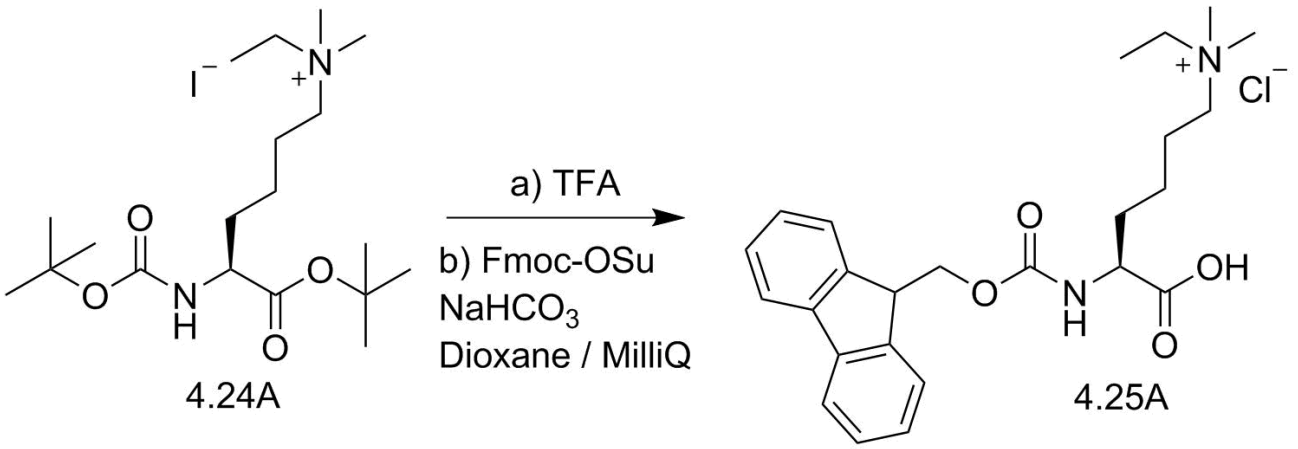
organic layer over $MgSO_4$, volatiles were removed under reduced pressure obtaining a crude red oil. To a solution of Boc-Lys(et_2)-O^{t}Bu in anhydrous MeCN (6 mL, 0.33 M) was added iodomethane (380 µL, 6.0 mmol, 3 equiv.) followed by heating to 80 °C for 40 min equipped with cooler while stirring. The solution was allowed to cooling down to RT followed by concentrating *in vacuo*. Grace column 24 g [SiO_2]. Solvent A: CH_2Cl_2, solvent B: MeOH. Program: 3 CV (0% B), loading 5 mL (0% B), 12 CV (0%–10% B) over 12 CV and the correct fractions were combined and volatiles were removed under reduced pressure obtaining a slightly yellowish oil (467 mg, 49%, 0.95 mmol). IR υ_{max}: 3440, 2977, 2361, 1707, 1366, 1156 cm^{-1}; $[\alpha]^{25}_D$ −12.5 (c = 0.010 g·mL^{-1} $CHCl_3$); ^{1}H NMR 400 MHz Solvent: $CDCl_3$-d_1, δ (ppm): 5.13 (bd, 1 H, *J*= 8.0 Hz, N*H*$^{\alpha}$CHCOO), 4.10-3.99 (m, 1 H, NH$^{\alpha}$C*H*COO), 3.48 (q, 4 H, *J*= 7.0 Hz, $^{\gamma}CH_2NCH_3(CH_2CH_3)_2$), 3.38 (dt, 1 H, *J*= 13.0 5.0 Hz, $^{\delta}CH_2{}^{\varepsilon}CHHNCH_3(CH_2CH_3)_2$), 3.28-3.16 (m, 1 H, $^{\delta}CH_2{}^{\varepsilon}CHHNCH_3(CH_2CH_3)_2$), 3.17 (s, 3 H, $^{\varepsilon}CH_2NCH_3(CH_2CH_3)_2$), 1.84-1.67 (m, 3 H, $^{\alpha}CH^{\beta}CHH^{\gamma}CH_2{}^{\delta}CH_2{}^{\varepsilon}CH_2$), 1.66-1.50 (m, 3 H, $^{\alpha}CH^{\beta}CHH^{\gamma}CH_2{}^{\delta}CH_2$), 1.40 (s, 9 H, $^{\alpha}$CHCOOC(CH_3)$_3$), 1.37 (s, 9 H, NHCOOC(CH_3)$_3$), 1.35 (t (partially visible), 6 H, *J*= 7.0 Hz, $^{\varepsilon}NCH_3(CH_2CH_3)_2$). ^{13}C NMR 100 mHz Solvent: $CDCl_3$-d_1, δ (ppm): 171.6, 155.7, 82.3, 79.9, 60.6, 57.08 53.2, 48.3, 32.4, 28.4, 28.1, 22.3, 21.6, 8.5; ESI-MS (+) in MeOH, $[M+H]^+$ calc: 373.3 found: 373.3; HR-MS ESI (+) in MeOH as $C_{20}H_{41}N_2O_4^+$, $[M+H]^+$, calc: 373.3061 found: 373.3062.



Boc-Lys(bu_3·I)-O^{t}Bu

tertbutyl (*S*)-2-[tertbutoxycarbonyl]amino-6-[tributyl]amino-hexanoate iodide 4.24E

To a solution of Boc-Lys(H)-O^{t}Bu **4.22** (599 mg, 2.0 mmol, 1.0 equiv.) in MeOH (10 mL, 0.2 M) was added drop wise n-buteraldehyde (540 µL, 37% sol., 6.0 mmol, 3.0 equiv.) followed by the slow addition of $NaBH_3CN$ (264 mg, 4.2 mmol, 2.1 equiv.). The solution was stirred for 23 hrs followed by the addition of sat. $NaHCO_3$ sol. (40 mL) and the mixture was extracted with CH_2Cl_2 (3× 35 mL). After drying the organic layer over $MgSO_4$, volatiles were removed under reduced pressure obtaining a crude colourless oil. To a solution of Boc-Lys(bu_2)-O^{t}Bu in anhydrous MeCN (10 mL, 0.2 M) was added butyl-iodide (1.0 mL, 8.8 mmol, 4.4 equiv.) followed by heating to 80 °C for 48 hrs equipped with cooler while stirring. The solution was allowed to cool to RT followed by concentrating *in vacuo*. Grace column 40 g [SiO_2]. Solvent A: CH_2Cl_2, solvent B: MeOH. 3 CV (0% B) equilibration, loading 5 mL (0% B), 12 CV (0%–10% B) over 12 CV and the correct fractions were combined and volatiles were removed under reduced pressure obtaining a slightly yellowish oil (770 mg, 65%, 1.3 mmol). IR υ_{max}: 2980, 1709, 1461, 1251, 1153 cm^{-1}; $[\alpha]^{25}_D$ −12.3 (c = 0.012 g·mL^{-1} $CHCl_3$); ^{1}H NMR 500 MHz Solvent: $CDCl_3$-d_1, δ (ppm): 5.14 (bd, 1 H, *J*= 8.0 Hz, N$\underline{H}^{\alpha}$CHCOO), 4.20-4.07 (m, 1 H, NH$^{\alpha}$C$\underline{H}$COO), 3.46-3.38 (m, 1 H, $^{\delta}CH_2{}^{\varepsilon}CH\underline{H}N(CH_2CH_2CH_2CH_3)_3$), 3.40 (t, 6 H, *J*= 9.0 Hz, $^{\gamma}CH_2N(C\underline{H}_2CH_2CH_2CH_3)_3$), 3.30-3.21 (m, 1 H, $^{\delta}CH_2{}^{\varepsilon}C\underline{H}HN(CH_2CH_2CH_2CH_3)_3$), 1.95-1.74 (m, 3 H, $^{\alpha}CH^{\beta}CH\underline{H}^{\gamma}CH_2{}^{\delta}C\underline{H}_2{}^{\varepsilon}CH_2$), 1.74-1.60 (quint, 6 H, *J*= 8.0 Hz, $^{\gamma}CH_2N(CH_2C\underline{H}_2CH_2CH_3)_3$) 1.64-1.50 (m, 3 H, $^{\beta}C\underline{H}H^{\gamma}C\underline{H}_2{}^{\delta}CH_2$),1.47 (s, 9 H, $^{\alpha}CHCOOC(C\underline{H}_3)_3$), 1.43 (s, 9 H, $NHCOOC(C\underline{H}_3)_3$), 1.02 (t, *J*= 7.5 Hz, 9 H, $^{\gamma}CH_2N(CH_2CH_2CH_2C\underline{H}_3)_3$), 0.99-0.92(m, 6 H, $^{\gamma}CH_2N(CH_2CH_2C\underline{H}_2CH_3)_3$); ^{13}C NMR 125 mHz ($CDCl_3$-$d_1$) δ (ppm): 171.6, 155.7, 82.0, 80.0, 59.5, 53.9, 53.2, 33.0, 28.5, 28.1, 24.4 20.0 22.4, 21.7, 13.9; ESI-MS (+) in MeOH, $[M+H]^+$ calc: 471.4 found: 471.4; HR-MS ESI (+) in MeOH as $C_{27}H_{55}N_2O_4^+$, $[M+H]^+$, calc: 471.4156 found: 471.4151.

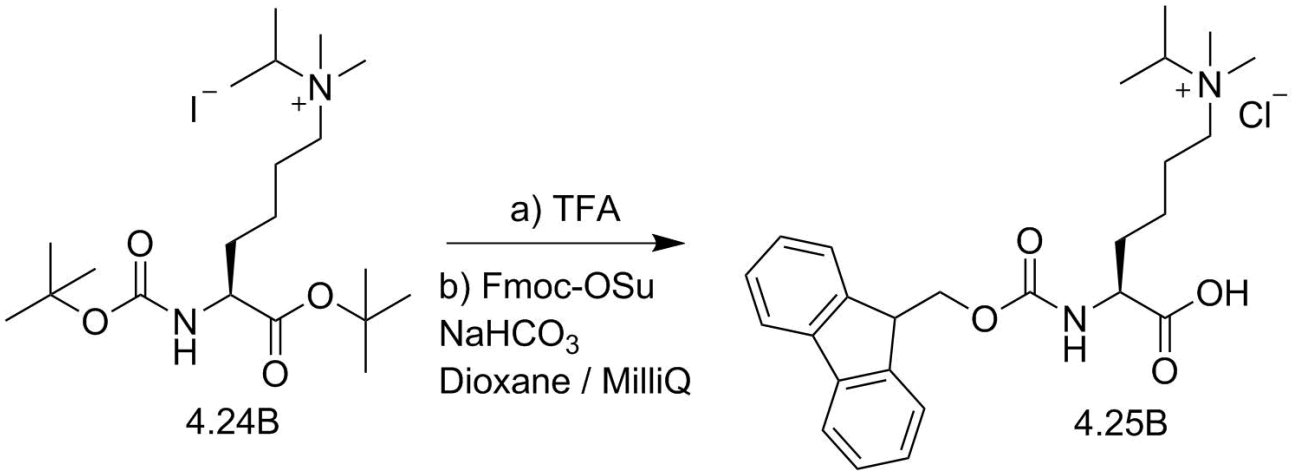


Fmoc-Lys(me_2, et·Cl)-OH

(*S*)-2-[Fluorenylmethyloxycarbonyl]amino-6-[dimethyl ethyl]amino-hexanoic acid hydrochloric acid 4.25A

Boc-Lys(me_2, et·I)-O^{t}Bu **4.24A** (975 mg, 2.0 mmol, 1.0 equiv.) was dissolved in TFA (15.0 mL, 0.13 M) and stirred for 32 hrs followed by the removal of volatiles under reduced pressure. The crude yellow/orange oil was dissolved in a solution of 1,4-dioxane and H_2O (1:1, 10 mL, 0.09 M). To the solution was added $NaHCO_3$ (814 mg, 9.7 mmol, 4.8 equiv.) observing gas release followed by introduction of Fmoc-OSu (745 mg, 2.2 mmol, 1.1 equiv.) to the mixture while stirring. After 2 hrs, the mixture was concentrated to ~4 mL and loaded on R_f-C_{18} column. Grace 40 g R_f-C_{18} column. Solvent A: [TFA (0.1 v/v%) in MilliQ], Solvent B: [TFA (0.1 v/v%) in acetonitrile]. Program: 3 CV (5% B) equilibration, loading 6 mL (25% B), 2 CV (5% B), 20 CV (5%–100% B). Selected fractions were pooled and lyophilised obtaining 506 mg (1.19 mmol) of a yellowish solid (60% yield). The white solid was dissolved in CH_2Cl_2 (5 mL) and HCl in Et_2O (1.0 mL, 2 M, 2 mmol, 3.45 equiv.) and volatiles were removed *in vacuo* obtaining a white solid mp: 77 °C (decomp.); $[\alpha]^{25}_D$ −0.9 (c = 0.005 g·mL^{-1} MeOH); IR υ_{max}: 3405, 2981, 1706, 1382, 1154 cm^{-1}; ^{1}H NMR 500 mHz Solvent: MeOD-d_4, δ (ppm): 7.81 (d, 1 H, *J*= 7.5 Hz, 2× Fmoc-C_4<u>*H*</u>), 7.68 (dd, 2 H, *J*= 7.5 4.5 Hz, 2× Fmoc-C_1<u>*H*</u>), 7.40 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-C_3<u>*H*</u>), 7.32 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-C_2<u>*H*</u>), 4.45-4.39 (m, 1 H, OC<u>*H*</u>HCH), 4.36-4.30 (m, 1 H, OCH<u>*H*</u>CH), 4.26-4.17 (m, 2 H, OCH_2C<u>*H*</u>, ND$^{\alpha}$C<u>*H*</u>COOH), 3.38 (q, 2 H, *J*= 7.0 Hz, $^{\varepsilon}CH_2N(CH_3)_2(C$<u>*H*</u>$_2CH_3)$), 3.29-3.20 (m, 2 H, $^{\delta}CH_2{}^{\varepsilon}C$<u>*H*</u>$_2NH$), 3.04 (s, 6 H,

$^{\varepsilon}CH_2N(C\underline{H}_3)_2(CH_2CH_3)$), 2.00-1.89 (m, 1 H, $^{\alpha}CH^{\beta}CH\underline{H}^{\gamma}CH_2$), 1.86-1.69 (m, 3 H, $^{\alpha}CH^{\beta}C\underline{H}H^{\gamma}CH_2{}^{\delta}C\underline{H}_2{}^{\varepsilon}CH_2$), 1.54-1.40 (m, 2 H, $^{\beta}CH_2{}^{\gamma}C\underline{H}_2{}^{\delta}CH_2$), 1.34 (t, 3 H, *J*= 7.5 Hz, $^{\varepsilon}CH_2N(CH_3)_2(CH_2C\underline{H}_3)$); ^{13}C NMR 125 mHz Solvent: MeOD-d_4, δ (ppm): 175.5, 158.7, 145.2, 142.5, 128.8, 128.2, 126.3, 121.0, 68.0, 64.5, 60.7, 54.8, 50.7, 48.4 31.2, 23.8, 22.9, 8.4; ESI-MS (+) in MeOH, $[M+H]^+$ calc: 425.2 found: 425.2; HR-MS ESI (+) in MeOH $[M+H]^+$ as $C_{25}H_{33}N_2O_4^+$, calc: 425.2435 found: 425.2428.

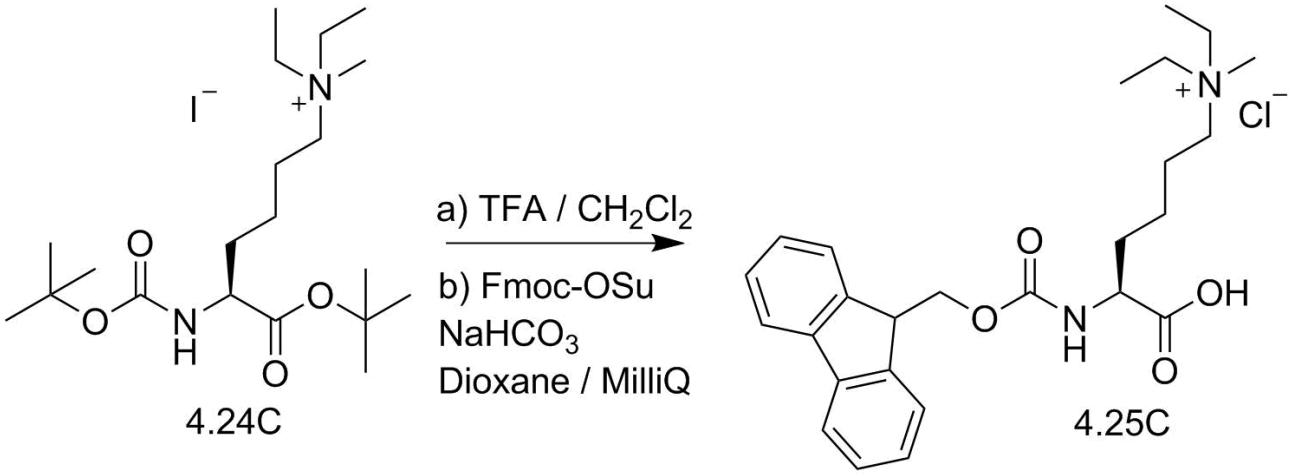


Fmoc-Lys(me$_2$, ^{i}Pr·Cl)-OH

(*S*)-2-[Fluorenylmethyloxycarbonyl]amino-6-[dimethyl isopropyl]amino-hexanoic acid hydrochloric acid 4.25B

Boc-Lys(me$_2$, ^{i}Pr·I)-O^{t}Bu **4.24B** (447 mg, 0.89 mmol, 1.0 equiv.) was dissolved in TFA (10 mL, 0.09 M) and stirred for 32 hrs followed by the removal of volatiles under reduced pressure. The crude yellow/orange oil was dissolved in a solution of 1,4-dioxane and H_2O (1:1, 10 mL, 0.09 M). To the solution was added $NaHCO_3$ (598 mg, 7.12 mmol, 8.0 equiv.) observing gas release followed by introduction of Fmoc-OSu (365 mg, 1.07 mmol, 1.20 equiv.) to the mixture while stirring. After 2 hrs, the mixture was concentrated to ~4 and loaded on R_f-C_{18} column. Grace 40 g R_f-C_{18} column. Solvent A: [TFA (0.1 v/v%) in MilliQ], solvent B: [TFA (0.1 v/v%) in acetonitrile]. Program: 3 CV (5% B) equilibration, loading 6 mL (25% B), 2 CV (5% B), 20 CV (5–100%). Selected fractions were combined and lyophilised obtaining 256 mg (0.58 mmol) of a yellowish solid (65% yield). The white

solid was dissolved in CH_2Cl_2 (5 mL) and $HCl{\cdot}Et_2O$ (1.0 mL, 2 M, 2 mmol, 3.45 equiv.) and volatiles were removed *in vacuo* obtaining a white solid. mp: 88 °C (decomp); IR υ_{max}: 3391, 2889, 2360, 1707, 1450, 1224, 1037 cm^{-1}; $[\alpha]^{25}_D$ −0.5 (c = 0.012 $g{\cdot}mL^{-1}$ MeOH); ^{1}H NMR 400 MHz (DMSO-d_6) δ (ppm): 12.71 (s, 1 H, $NH^{\alpha}CHCOO\underline{H}$), 7.90 (d, 2 H, *J*= 7.5 Hz, 2× Fmoc-$C_4\underline{H}$), 7.77-7.70 (m, 2 H, 2× Fmoc-$C_1\underline{H}$), 7.68 (d, 1 H, *J*= 8.0 Hz, $N\underline{H}^{\alpha}CHCOOH$), 7.42 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-$C_3\underline{H}$), 7.33 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-$C_2\underline{H}$), 4.36-4.18 (m, 3 H, $OC\underline{H}_2C\underline{H}$), 3.97 (dt, 1 H, *J*= 9.0 8.0 Hz, $NH^{\alpha}C\underline{H}COOH$), 3.65 (hept, 1 H, *J*= 6.5 Hz, $N(CH_3)_2(CH_3C\underline{H}CH_3)$), 3.30-3.17 (m, 2 H, $^{\delta}CH_2{}^{\varepsilon}C\underline{H}_2N$), 2.91 (s, 6 H, $N(C\underline{H}_3)_2(CH_3CHCH_3)$), 1.82-1.72 (m, 1 H, $^{\alpha}CH^{\beta}CH\underline{H}^{\gamma}CH_2$), 1.72-1.57 (m, 3 H, $^{\alpha}CH^{\beta}C\underline{H}H^{\gamma}CH_2{}^{\delta}C\underline{H}_2{}^{\varepsilon}CH_2$), 1.41-1.31 (m, 2 H, $^{\beta}CH_2{}^{\gamma}C\underline{H}_2{}^{\delta}CH_2$), 1.30-1.19 (m, 6 H, $N(CH_3)_2(C\underline{H}_3CHC\underline{H}_3)$; ^{13}C NMR 100 MHz Solvent: MeOD-d_4, δ (ppm): 175.5, 158.7, 145.2, 142.5, 128.8, 128.2, 126.2, 121.0, 67.9, 61.3, 57.5, 54.8, 48.4, 47.7, 32.0, 23.8, 22.5, 8.1; ESI-MS (+) in MeOH, $[M+H]^+$ calc: 439.3 found: 439.3; HR-MS (ESI+) in MeOH as $C_{26}H_{35}N_2O_4^+$, $[M+H]^+$: calc.: 439.2591, found: 439.2582.

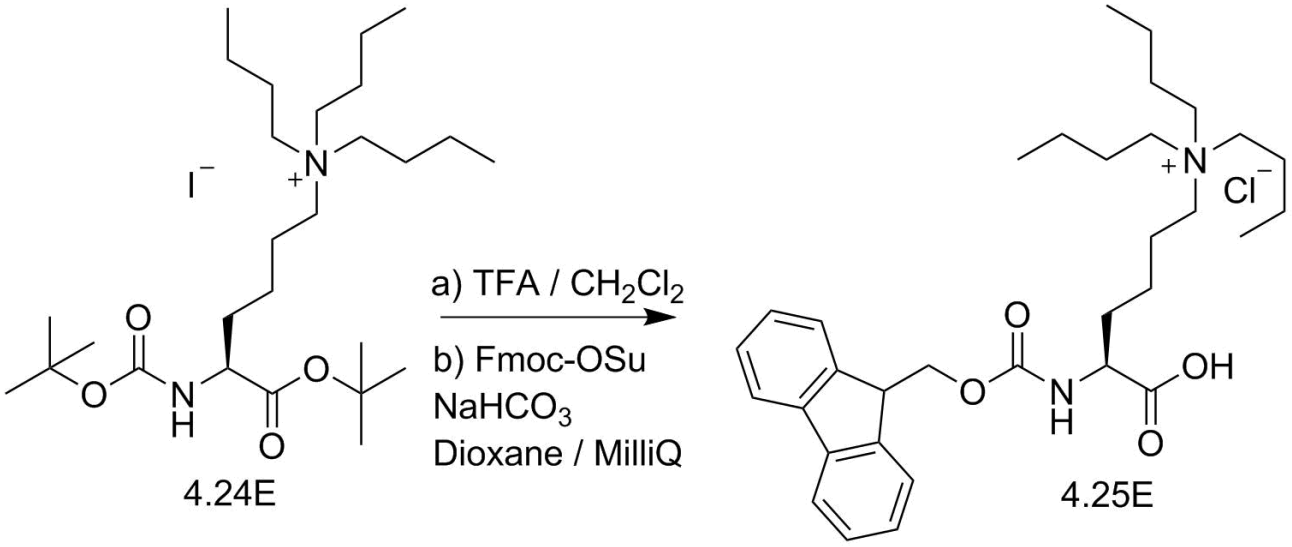


Fmoc-Lys(me, $et_2{\cdot}Cl$)-OH

(*S*)-2-[Fluorenylmethyloxycarbonyl]amino-6-[diethyl methyl]amino-hexanoic acid hydrochloric acid 4.25C

Boc-Lys(me, $et_2{\cdot}I$)-O^{t}Bu **4.24C** (702 mg, 1.4 mmol, 1.0 equiv.) was dissolved in [TFA / CH_2Cl_2 (5:1, 12 mL, 0.12 M)] and stirred for 32 hrs followed by the removal of volatiles under reduced pressure.

The crude yellow/orange oil was dissolved in a solution of 1,4-dioxane and H_2O (1:1, 14 mL, 0.1 M). To the solution was added $NaHCO_3$ (653 mg, 7.77 mmol, 5.6 equiv.) observing gas release followed by introduction of Fmoc-OSu (583 mg, 1.73 mmol, 1.24 equiv.) to the mixture while stirring. After 2 hrs, the mixture was concentrated to ~4 mL and loaded on R_f-C_{18} column. Grace 40 g R_f-C_{18} column. Solvent A: [TFA (0.1 v/v%) in MilliQ], Solvent B: [TFA (0.1 v/v%) in acetonitrile] Program: 3 CV (5% B) equilibration, loading added 6 mL (25% B), 2 CV (5% B), 20 CV (5–100% B). Selected fractions were combined and lyophilised obtaining 188 mg (0.43 mmol) of a light yellow solid (31% yield). The solid was dissolved in CH_2Cl_2 (5 mL) and HCl in Et_2O (1.0 mL, 2.0 mmol, 2 M 4.67 equiv.) and volatiles were removed *in vacuo* obtaining an off white solid. mp: 116 °C (decomp); $[\alpha]^{25}_D$ −0.4 (c = 0.010 g·mL^{-1} MeOH); IR υ_{max}: 3401, 2980, 1706, 1393, 1154 cm^{-1}; ^{1}H NMR 400 MHz Solvent: MeOD-d_4, δ (ppm): 7.79 (d, 2 H, *J*= 7.5 Hz, Fmoc-$C_4\underline{H}$), 7.67 (t, 2 H, *J*= 8.0 Hz, 2× Fmoc-$C_1\underline{H}$), 7.42 (t, 2 H, *J*= 7.0 Hz, 2× Fmoc-$C_3\underline{H}$), 7.33 (t, 2 H, *J*= 7.0 Hz, 2× Fmoc-$C_2\underline{H}$), 4.42-4.27 (m, 2 H, $OC\underline{H}_2CH$), 4.23-4.15 (m, 2 H, $OCH_2C\underline{H}$, $ND^{\alpha}C\underline{H}COOH$), 3.31 (q, 4 H, *J*= 7.0 Hz, $^{\varepsilon}CH_2NCH_3(C\underline{H}_2CH_3)_2$), 3.24-3.13 (m, 2 H, $^{\delta}CH_2{}^{\varepsilon}C\underline{H}_2NH$), 2.94 (s, 3 H, $^{\varepsilon}CH_2NC\underline{H}_3(CH_2CH_3)_2$), 2.00-1.88 (m, 1 H, $^{\alpha}CH^{\beta}CH\underline{H}^{\gamma}CH_2$), 1.84-1.63 (m, 3 H, $^{\alpha}CH^{\beta}C\underline{H}H^{\gamma}CH_2{}^{\delta}C\underline{H}_2{}^{\varepsilon}CH_2$), 1.54-1.39 (m, 2 H, $^{\beta}CH_2{}^{\gamma}C\underline{H}_2{}^{\delta}CH_2$), 1.27 (t, *J*= 7.5 Hz, 6 H, $^{\varepsilon}CH_2NCH_3(CH_2C\underline{H}_3)_2$); ^{13}C NMR 100 MHz Solvent: MeOD-d_4, δ (ppm): 175.5, 158.7, 145.2, 142.5, 128.8, 128.2, 126.2, 121.0, 67.9, 61.3, 57.5, 54.8, 48.4, 47.7, 32.0, 23.8, 22.5, 8.1; ESI-MS (+) in MeOH, $[M+H]^+$ calc: 439.3 found: 439.3; HR-MS ESI (+) in MeOH $[M+H]^+$, calc: 439.2591 found: 439.2604.



Fmoc-Lys(bu_3·Cl)-OH

(*S*)-2-[Fluorenylmethyloxycarbonyl]amino-6-[tributyl]amino-hexanoic acid hydrochloric acid 4.25E

Boc-Lys(Bu_3·I)-O^{t}Bu **4.24E** (758 mg, 1.27 mmol, 1.0 equiv.) was dissolved in TFA (12 mL, 0.11 M) and stirred for 32 hrs followed by the removal of volatiles under reduced pressure. The crude yellow/orange oil was dissolved in a solution of 1,4-dioxane and H_2O (1:1, 12 mL, 0.11 M). To the solution was added $NaHCO_3$ (477 mg, 5.68 mmol, 4.47 equiv.) observing gas release followed by introduction of Fmoc-OSu (532 mg, 1.55 mmol, 1.22 equiv.) to the mixture while stirring. After 1 hrs, the mixture was concentrated to ~4 mL and loaded on R_f-C_{18} column. Grace 40 g R_f-C_{18} column. Solvent A: [TFA (0.1 v/v%) in MilliQ], solvent B: [TFA (0.1 v/v%) in MilliQ]. 3 CV (5% B) equilibration, loading added 6 mL (25% B), 2 CV (5% B), 20 CV (5%–100% B). Selected fractions were combined and lyophilised obtaining 134 mg (0.25 mmol) of a slightly yellow solid (20%). The white solid was dissolved in CH_2Cl_2 (5 mL) and HCl in Et_2O (0.5 mL, 1.0 mmol, 2 M, 4.0 equiv.) and volatiles were removed *in vacuo* obtaining a slightly yellow solid. mp: 75 °C (decomp); IR υ_{max}: 2963, 2362, 1712, 1536, 1193, 1130 cm^{-1}; $[\alpha]^{25}_D$ −0.7 (c = 0.010 g·mL^{-1} MeOH); ^{1}H NMR 400 MHz Solvent: MeOD-d_4, δ (ppm): 7.81 (d, 2 H, *J*= 7.0 Hz, Fmoc-$C_4\underline{H}$), 7.72-7.62 (m, 2 H, *J*= 8.0 Hz, 2× Fmoc-$C_1\underline{H}$), 7.40 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-$C_3\underline{H}$), 7.31 (t, 2 H, *J*= 7.0 Hz, 2× Fmoc-$C_2\underline{H}$), 4.45-4.28 (m, 2 H, OC$\underline{H}_2$CH), 4.27-4.18 (m, 2 H, OCH$_2$C$\underline{H}$, ND$^\alpha$C$\underline{H}$COOH), 3.28-3.08 (m, 8 H, *J*= 7.0 Hz,

$^{\delta}CH_2{}^{\varepsilon}C\underline{H_2}N(C\underline{H_2}CH_2CH_2CH_3)_3$), 2.02-1.90 (m, 1 H, $^{\alpha}CH^{\beta}CH\underline{H}^{\gamma}CH_2$), 1.83-1.70 (m, 2 H, $^{\alpha}CH^{\beta}C\underline{H}H^{\gamma}CH_2{}^{\delta}CH\underline{H}^{\varepsilon}CH_2$), 1.70-1.56 (m, 7 H, $^{\gamma}CH_2{}^{\delta}C\underline{H}H^{\varepsilon}CH_2$, $^{\varepsilon}N(CH_2C\underline{H_2}CH_2CH_3)_3$), 1.53-1.31 (m, 8 H, $^{\gamma}C\underline{H_2}{}^{\delta}CH_2{}^{\varepsilon}CH_2N(CH_2CH_2C\underline{H_2}CH_3)_3$), 1.00 (t, J= 7.5 hz, 9 H, $^{\varepsilon}N(CH_2CH_2CH_2C\underline{H_3})_3$); ^{13}C NMR 100 MHz Solvent: MeOD-d_4, δ (ppm): 175.6, 158.7, 145.2, 142.5, 128.8, 128.2, 126.2, 121.0, 68.0, 59.49, 59.39, 54.6, 48.3, 32.1, 24.7, 23.6, 22.0, 20.7, 13.9; ESI-MS (+) in MeOH, $[M+H]^+$ calc: 537.4 found: 537.4; HR-MS (ESI+) in MeOH as $C_{33}H_{49}N_2O_4^+$, $[M+H]^+$: calc.: 537.3687, found: 537.3683.

4.8.7 Protein expression

$CHD1_{DCD}$

$CHD1_{DCD}$ was gifted by Cheryl Arrowsmith (Structural Genomic Consortium, Toronto) through Addgene (58755) in vector: pET28-MHL (E268-R443) and the expression was performed according to reported procedure.[316]

$CHD1_{DCD}$ was transformed into phage resistant BL21 DE3 *E. coli* cell line and grown on agar plate containing antibiotics (Kanamycin (30 µg·mL^{-1}) and Chloroamphenicol (34 µg·mL^{-1})). A starter culture was grown overnight in a 2TY media (100 mL) containing kanamycin (30 µg·mL^{-1}) and used to inoculate (10 mL·L^{-1} media) 4× 1 litre Terrific broth (Fisher Bioreagents™ (product #: 12891660)). The cultures in flasks were shaken at 175 rpm at 37 °C until an OD_{600}: ~0.6 was reached followed by cooling down to 17 °C and induced at OD_{600}: ~0.8 using IPTG (0.5 mM). After shaking overnight, the cells were collected using centrifugation (11899 × g, 8000 rpm, 10 min), weighted (26.03 g) and stored (−80 °C). Cell pellet (12.88 gram) was suspended in a mixture of binding buffer [Imidazole (10 mM), TCEP (0.5 mM), HEPES (50 mM), NaCl (500 mM), in MilliQ (pH 7.5) (filtered 0.2 µm)] (100 mL), DNAse (I) (Sigma-Aldrich, cat#: 10104159001) and protease inhibitor (cOmplete™, EDTA-free, Roche, cat#: 05056489001) at 4 °C. A smooth paste was made and the mixture was sonicated on ice for 12 min (2 sec on, 3 sec off). The mixture was centrifuged (48384 × g, 20 min), the lysate was

filtered (0.45 µm) and loaded on Histrap™ column (5 mL, (1 CV = 5 mL), 2 mL·min^{-1}) using FPLC (ÄKTA, GE Healthcare). The Histrap was treated with 10 CV of wash buffer [Imidazole (40 mM), TCEP (0.5 mM), HEPES (50 mM), NaCl (500 mM), in MilliQ (pH 7.5)] and the protein eluted was gradient eluted from 0 to 100% Elution buffer [Imidazole (250 mM), TCEP (0.5 mM), HEPES (50 mM), NaCl (500 mM), in MilliQ (pH 7.5)] over 10 CV. The obtained fractions were pooled based on FPLC trace chromatogram and concentrated (Amicon® Ultra-15 10K, 4255 × g). Half of this was purified using size exclusion column (S75, GE Healthcare) in combination with the FPLC using gel filtration buffer [HEPES (50 mM), NaCl (50 mM), in MQ (pH 7.5), filtrated through 0.2 µm filter and degassed] (300 mL). Selected fractions were pooled C2-C10 and concentrated (Amicon® Ultra-15 10K, 4255 × g) yielding 68.3 mg of protein (6.92 mL, 9.872 mg·mL^{-1}). The protein concentration was determined by Nanodrop One (Thermoscientific) using computed extinction coefficient: 50880 (ProtParam, ExPASy).

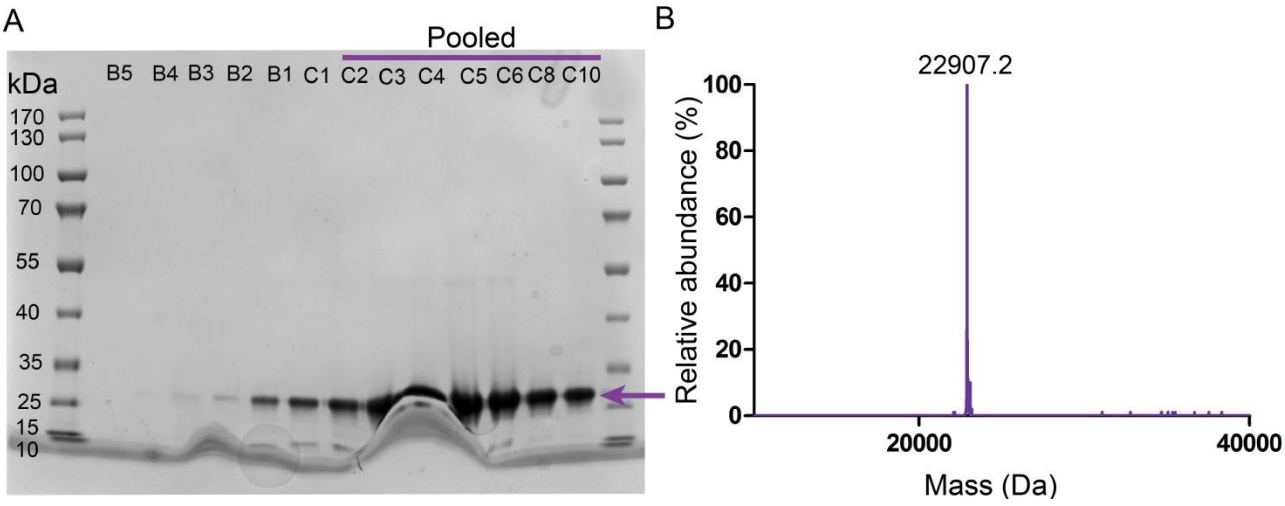


Figure 4.8.2: A) SDS-PAGE of $CHD1_{DCD}$ fractions after Superdex 75 gel filtration. Expected mass is 22.9 kDa. Fractions C2-C10 were pooled together. B) LC-MS (RF-MS) deconvoluted mass spectrum of $CHD1_{DCD}$, expected molecular weight: 22906.4 Da.

$SPIN2B_{TTD}$

The plasmid containing the vector pNIC-CTHF the gene SPIN2B encoding for (V50-S258) was used for protein production. The protein is made with c-terminal TEV cleavage site, 6×His-tag and Flag-

tag. The cloned gene in pNIC-CTHF was transformed into *E. coli* strain (BL21) DE3 that is phage resistant and is equipped with pRARE2 ^{t}RNA plasmid. From a single colony grown on an agar plate, was inoculated a starter culture (100 mL) containing kanamycin (50 μg·mL^{-1}) and chloroamphenicol (34 μg·mL^{-1}) was shaked overnight (37 °C). Terrific broth media containing kanamycin (50 μg·mL^{-1}) and chloroamphenicol (34 μg·mL^{-1}) was inoculated and cells were grown (37 °C) while shaking until an OD_{600}: ~2.0 was reached and the media was cooled to 18 °C and induced with IPTG (0.1 mM). The mixture was shaked overnight, cells were collected by centrifugation and stored in the freezer (−80 °C). The pellet was suspended in lysis buffer [HEPES (50 mM), NaCl (500 mM), Imidazole (10 mM), glycerol (5%), TCEP (0.5 mM) in MilliQ (pH 7.5)] and protease inhibitor cocktail (Sigma-Aldrich) until a smooth paste was formed and the cells were sonicated on ice. The solution was cleared using centrifugation, filtrated and loaded on Histrap™ (GE, Healtcare) and purified using stepwise increase of imidazole. The right fractions were pooled, treated with TEV protease overnight (4.0 °C) and concentrated. The proteins was purified using gel filtration size exclusion (Superdex 200, GE Healthcare) with gel filtration buffer [HEPES (10 mM), NaCl (500 mM), glycerol (5%), TCEP (0.5 mM) in MilliQ (pH 7.5)]. The pooled fractions were treated with Histrap™ column to remove any TEV cleavage protein and His-tagged protein. Selected fractions were combined and concentrated to give of protein and stored (−80 °C). The protein purity and mass was confirmed with SDS-PAGE and ESI-TOF-TOF.

TAF3$_{PHD}$

TAF3$_{PHD}$ received generously from M. Th. Marc Timmers encoding TAF3 gene (R857-K924) and the protein was produced according to previous report procedures.[321]

KDM5A$_{PHD3}$

KDM5A$_{PHD3}$ (M1489-V1641) was generously gifted by C. David Allis and expressed according to previously reported procedure.[8]

$KDM4A_{TTD}$

$KDM4A_{TTD}$ encoding for Q897-P1011 was generously gifted by George Mer and produced as stated earlier.[313]

$KDM7B_{PHD}$

KDM7B gene encoding for M37-P116 was cloned in pNIC28-BSA4 vector. The plasmid was transformed into *E. coli* and plated on an agar plate containing kanamycin (30 $\mu M \cdot mL^{-1}$). A single colony was picked and inoculated a starter culture (100 mL, 2TY media) containing kanamycin (30 $\mu M \cdot mL^{-1}$) and shaked overnight (250 rpm). 2TY (12× 600mL) media containing kanamycin (30 $\mu M \cdot mL^{-1}$) was inoculated with starter culture (10 $mL \cdot L^{-1}$ media) and grown while shaking at 37 °C until an OD_{600}: 0.8–1.2 of was reached. IPTG (0.1 mM) was added and grown shaking (220 rpm, 18.0 °C) overnight. Cells were harvested by centrifugation (8000 rpm, 4.0 °C, 10 min), and stored at −80 °C. Cell pellet was suspended in binding buffer [HEPES (50 mM), NaCl (500), Imidazole (20 mM), glycerol (5%) in MilliQ (pH 7.5)] (5.0 $mL \cdot g^{-1}$ pellet) and stirred (4.0 °C) to form a smooth mixture followed by sonication (3 sec on, 4 sec off, 4 min total), centrifugation (22000 rpm, 13486 × g, 4.0 °C, 30 min) and decantation of the supernatant. The supernatant was filtered (0.45 µM) and loaded on a Histrap™ (1 CV = 5 mL) (GE, Healthcare) column using ÄKTA FPLC system (GE, Healthcare). The column washed with binding buffer (15 CV), wash buffer [HEPES (50 mM), NaCl (500), Imidazole (40 mM), glycerol (5%) in MilliQ (pH 7.5)] (5 CV) and the protein was eluted using single step gradient elution buffer [Imidazole (250 mM), HEPES (50 mM), NaCl (500), glycerol (5%) in MilliQ (pH 7.5)] (2 CV). To the pooled selected fractions was added EDTA (1.0 mM) and TEV protease (TEV mass to $KDM7B_{PHD}$ mass, 1:50) to react overnight (12–16 hrs, 4.0 °C). The fractions were concentrated (~2.0 mL) and purified using Superdex 75 (GE, Healthcare) and gel filtration buffer [HEPES (50 mM), NaCl (200 mM), TCEP (1.0 mM) in MilliQ (pH 7.0), degassed] (1 $mL \cdot min^{-1}$).

Selected fractions based on SDS-PAGE were concentrated (Amicon® Ultra 3.5K MWCO) obtaining $KDM7_{PHD}$ (11.6 mg·mL^{-1}, 1.3 mM).

$KDM7B_{PHDJmjC}$

pNH-TrxT containing sequence encoding for $KDM7B_{PHDJmjC}$ M37 – N483

From transformed plasmid in competent *E. coli* BL21 (DE3), a single colony was grown in starter culture (100 mL) with kanamycin (30 µg·mL^{-1}). 8× 1 litre Terrific broth [Tryptone (12 g·L^{-1}), Yeast extract (24 g·L^{-1}), Glycerol (4 mL·L^{-1}), KH_2PO_4 (17 mM), K_2HPO_4 (72 mM) in MilliQ] were grown using 10 mL·L^{-1} media and kanamycin media (30 µg·mL^{-1}). The mixture was shaken at 170 rpm at 37 °C until an OD_{600} of ~1.4 is reached and media was induced with IPTG (300 µM) an cooled to 18 °C. Cells were harvested using centrifugation (10 min, 8000 rpm, 10967 × g, 4.0 °C), combined and frozen (−80 °C). Cells were suspended in lysis buffer [Imidazole (20 mM), HEPES (50 mM), NaCl (500 mM) in MilliQ (pH 7.5)] (4.0 °C), sonicated (3 sec on, 4 sec off, 7 min) and the lysate was centrifuged (20 min, 20 000 rpm, 75600 × g). The lysate was filtered (0.45 µm) and purified using nickel affinity purification to HisTrap™ (5 mL, 1 CV = 5 mL) (GE Healthcare) using ÄKTA FPLC (GE Healthcare) (1 mL·min^{-1}). The HisTrap™ was washed with wash buffer [Imidazole (40 mM), HEPES (50 mM), NaCl (500 mM) in MilliQ (pH 7.5)] over 6 CV and gradient eluted using elution buffer [Imidazole (250 mM), HEPES (50 mM), NaCl (500 mM) in MilliQ (pH 7.5)] (10 CV). Selected fractions were pooled, concentrated, diluted with gel filtration buffer [HEPES (10 mM), NaCl (500 mM) in MilliQ (pH 7.5)] and His-TEV protease was added followed by cleavage overnight. The protein was concentrated and loaded on size exclusion purified Superdex75 (GE Healthcare) (300 mL) using gel filtration buffer (300 mL, 1 mL·min^{-1}). Fractions (2.0 mL) were pooled based on SDS-page (Fraction 1–5) using (Amicon® Ultra-15 30K, 4255 × g) obtaining 32 mg of protein (194 µM, 10.09 mg·mL^{-1}). The protein concentration was determined by Nanodrop One (Nanodrop ND-1000 by Thermo Scientific) using computed extinction coefficient: 66350 (ProtParam, ExPASy). The protein was mass confirmed using RF-MS with calculated: 51620.16 and measured: 51621.88.

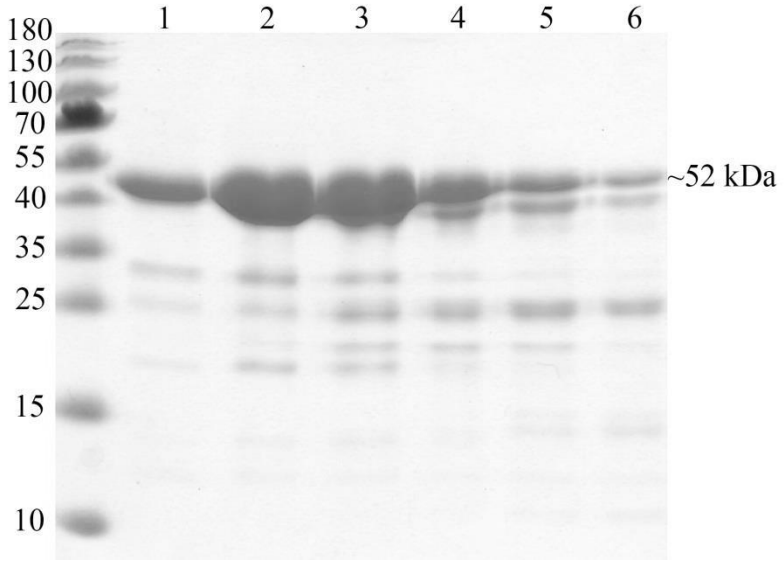


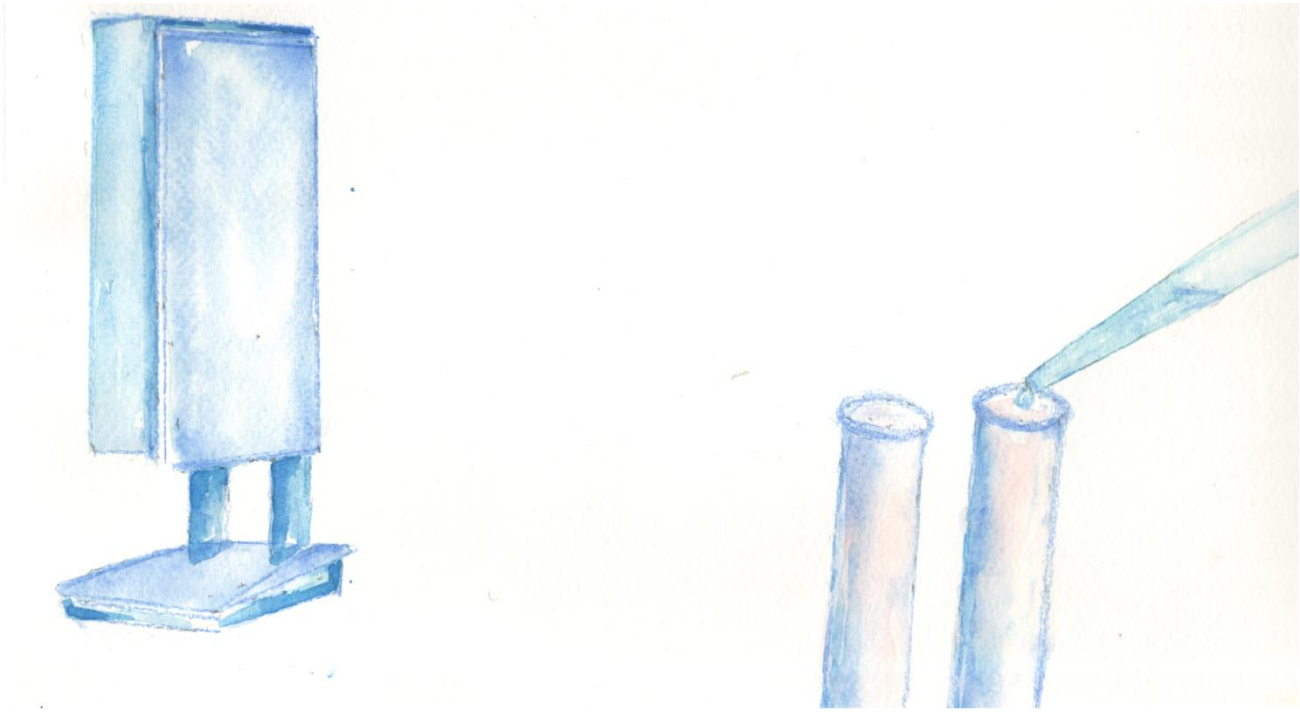
Figure 4.8.3: SDS-PAGE gel of KDM7B$_{PHDJmjC}$ after gel filtration were fractions 2-6 were pooled. Predicted mass of~52 kDA is observed visualized using Coomassie Brilliant Blue. Ladder displayed in kDA.

SGF29$_{TTD}$

SGF29$_{TTD}$ (R115-K293) was generously gifted by Jinrong Min and the protein was made according to reported literature procedure.[322]

BPTF$_{PHD}$

The construct encoding for BPTF$_{PHD}$ (L2583-N2751) was kindly provided by C. David Allis and the protein was expressed in *E. coli* as previously published.[319]



Chapter 5. Investigating the effect on Histone C^{α}-Lysine stereochemistry on catalysis and recognition by KMTs, KDMs and chromatin binding modules

5.1 Contributors to this Chapter

Production of histone methyltransferases (KMTs) KMT7 (SETD7), KMT5A (SEDT8), KMT1C (G9a), KMT1D (GLP), the corresponding KMTs substrates and KMT enzymatic assays were carried out by Abbas H. K. Al Temimi. Reader proteins: $TAF3_{PHD}$, $KDM5A_{PHD3}$, $KDM4A_{TTD}$, $BPTF_{PHD}$ and $SGF29_{TTD}$, were produced by Dr. Bas J.G.E. Pieters. ITC analysis was executed by Dr. Jasmin Mecinović. Molecular Dynamics was carried out by Kiran Kumar and Dr. Robert S. Paton. RapidFire analysis was assisted by Dr. Anthony Tumber and ^{1}H NMR was assisted by Dr. Richard J. Hopkinson. Dr. Catrine Johansson produced KDMs $KDM5B_{JmjC}$ and Aleksandra Szykowska gifted protein $KDM4A_{JmjC}$, $KDM4D_{JmjC}$ and $KDM5C_{JmjC}$. $KDM1A_{C1\text{-}corest}$ was produced by Dr. James E. Dunford and $KDM4E_{JmjC}$ was expressed by Dr. Richard J. Hopkinson. 2TY media for $KDM4E_{JmjC}$ was prepared by H. Manso Jubier.

Publication arising from this Chapter:

Roman Belle, *et al.* Investigating d-lysine stereochemistry for epigenetic methylation, demethylation and recognition. *Chemical Communications*, **55**, 13264-13267 (2018).

5.2 Introduction

The effects of N^{ε}-alkyl substitutions on lysine residues on the binding affinities to chromatin binding modules are explored in Chapter 3 and 4. In this Chapter, the importance of the stereochemistry of modified lysines for binding recognition and catalysis by associated chromatin effector proteins is explored.

Amino acids, the building blocks for proteins, are predominantly found with a Laevorotatory (L) stereochemistry in cellular context. The ribosomal synthesis of peptides and proteins uses highly selective translational machinery to exclusively incorporate L amino acids. However, numerous studies have demonstrated that dextrorotatory (D) amino acids are found in biological systems.

One way D-amino acids are formed is by external physical properties that cause racemization[323–325] as part of a thermodynamic equilibrium influenced by temperature and pH. However, D-residues are also actively produced in the cell. The pyridoxal 5'-phosphate (PLP)-dependent enzyme alanine racemase from *Bacillus stearothermophilus* converts L-alanine to D-alanine, which is a vital building block for the synthesis of cell walls.[326] Recent reports show that serine racemase in mammals can selectively catalyse conversion of L-serine to D-serine using PLP and Mg(II).[327] It is proposed that D-serine is an integral component of neurotransmitter glutamate were it binds the N-methyl-D-aspartate (MDMA) receptor in conjunction to glutamate.[328] Furthermore, D-aspartate has been observed in mammals indicating similar physiological properties to the L-stereoisomer where it can be absorbed, released, synthesized and degraded.[329,330] In rats, D-aspartate is detected in high concentration in the rat brain, pineal gland, pituitary gland, adrenal gland and is involved in prolactin production.[329,331–333] D-aspartate has also been detected in human tissues including in the aorta, lung, skin, brain and bone; it is proposed that D-aspartate may take part in aging and stress.[334–339] D-aspartate residue is also found at approximately 12% of histone H2B (D-isoAsp25) in mammalian brain, and highly enriched in active

chromatin.[340,341] While L-aspartate is incorporated during protein synthesis, it can form L-isoaspartate over time as a consequence of aging-induced trans-amidation. Protein L-isoaspartyl methyltransferase is a protein that catalyses methylation of the C^{α}-carboxylic acid in order to reverse this process. However, during the multi-step mechanism the imide intermediate can racemize to give the D-aspartate residue.[340]

5.2.1 In the context of epigenetic regulation

While D-lysine has not been yet observed in histones, it is important to study the molecular recognition and plasticity of lysine readers and modifiers. The change of stereo chemistry from L-lysine to D-lysine may have a profound impact on the binding affinity of chromatin binding modules and therefore transcriptional regulation.

5.3 Chapter objectives

The aim of this Chapter is to investigate the effect of changing the stereochemistry from L to D (S to R) of histone H3 N^{ε}-methyl-lysine residues on their binding to chromatin binding modules, KMTs and KDMs. The following questions are addressed in this Chapter:

- What is the change in binding affinity between histone H3 mimic peptides containing L- and D-N^{ε}-trimethyl-lysine with chromatin binding modules?
- Can molecular dynamics (MD) be used to model D-N^{ε}-trimethyl-lysine residues in protein binding site and calculate their binding affinities relative to L-N^{ε}-trimethyl-lysine?
- Can KMTs catalyse methylation of D-N^{ε}-lysine residues?
- Are KDMs able to catalyse demethylation of D-N^{ε}-trimethyl- or D-N^{ε}-dimethyl-lysine residues?

In this Chapter reference made to the domain are described in capitals (*e.g.* JmjC) while a reference made to construct domain is described in subscript (*e.g.* KDM4A$_{JmjC}$).

5.4 Results

The findings described in this Chapter are published in *Chemical Communications*.[342]

5.4.1 Synthesis of Fmoc-protected lysine monomers and production of peptides

Initially, Fmoc protected lysine monomers were synthesised as displayed in figure 5.4.1.

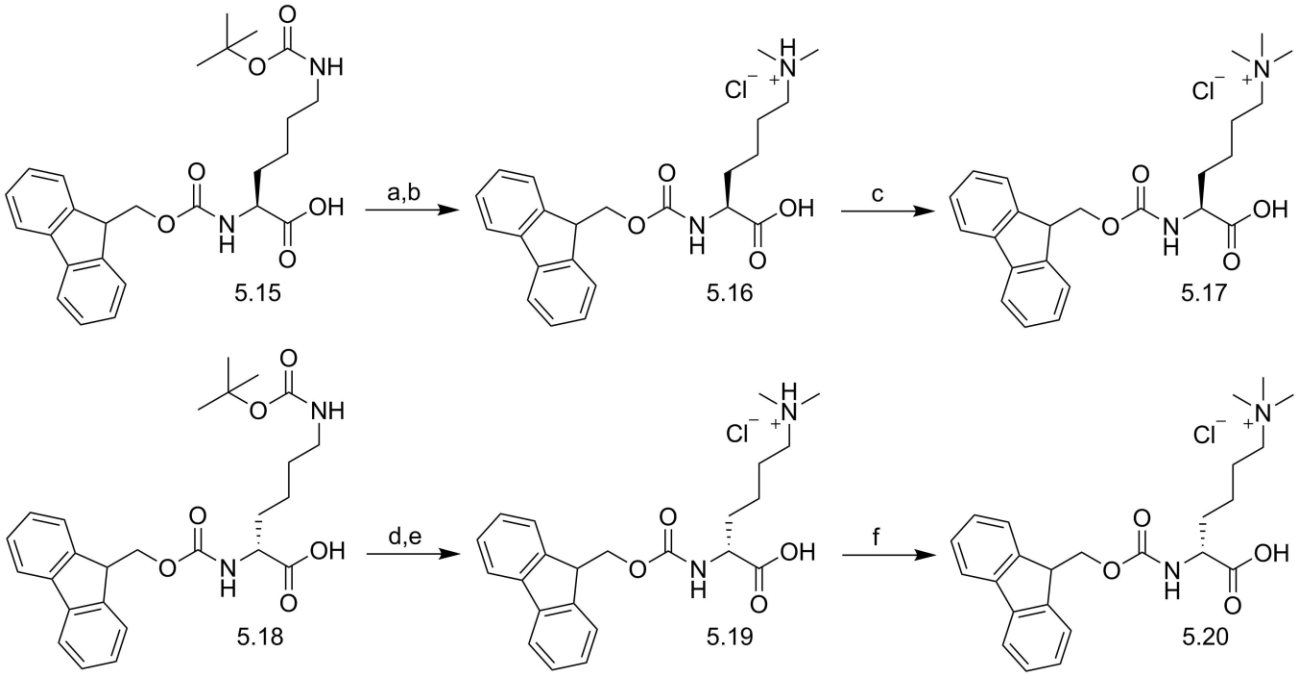


Figure 5.4.1: Synthetic pathway for the preparation of Fmoc-$^{D/L}$Lys(me$_2$·HCl)-OH 5.16/5.19 and Fmoc-$^{D/L}$Lys(me$_3$·HCl)-OH 5.17/5.20. a) HCl·Et$_2$O (5.0 equiv.), in CH_2Cl_2, 20 °C, 16 hrs; b) $HCHO_{(aq)}$ (6.0 equiv.), $NaBH_3CN$ (2.2 equiv.) in EtOH, 45 min then TFA quench, 15 min, R_f-C_{18} purification, (82%), HCl·Et$_2$O (2.0 equiv.), 20 °C; c) CH_3I (20 equiv.), $NaHCO_3$ (1.1 equiv.) for 48 hrs, then CH_3I (13.4 equiv.) for 24 hrs in EtOH, 20 °C, R_f-C_{18} purification, (62%), HCl·Et$_2$O (2.78 equiv.), 20 °C; d) HCl·Et$_2$O (5.0 equiv.), in CH_2Cl_2, 20 °C, 16 hrs; e) formalin (9.0 equiv.), $NaBH_3CN$ (2.9 equiv.) in EtOH, 45 min then TFA quench, 15 min, R_f-C_{18} purification, (71%), HCl·Et$_2$O (1.83 equiv.), 20 °C; f) CH_3I (10 equiv.), $NaHCO_3$ (1.1 equiv.) in EtOH, 48 hrs, R_f-C_{18} purification, (77%) HCl·Et$_2$O (3.8 equiv.), 20 °C.

Starting from commercially available Fmoc-^{L}Lys(Boc)-OH **5.15**, the acid labile Boc was deprotected using hydrochloric acid in diethyl ether (HCl·Et$_2$O) followed by reductive amination using formaldehyde and $NaBH_3CN$ in ethanol. The R_f-C_{18} purified product was lyophilized and protonated using HCl·Et$_2$O to give Fmoc-^{L}Lys(me$_2$·HCl)-OH **5.16**. The protonated methylated lysine derivatives (*i.e.* Fmoc-^{L}Lys(me$_2$·HCl)-OH **5.16**) were found to be hydroscopic and provided limited stability,

therefore, the yield of the product was estimated based on lyophilized solid obtained, after R_f-C_{18} purification. Fmoc-^{L}Lys(me_2·HCl)-OH **5.16** was treated with $NaHCO_3$ and iodomethane (CH_3I) in ethanol and after 2 days supplemented with ($NaHCO_3$ + CH_3I). The reaction progress was carefully monitored using ESI-MS. The Fmoc-^{L}Lys(me_3)-OH intermediate product was purified using R_f-C_{18} and lyophilized, obtaining a yield of 62%. The substance was protonated using HCl·Et_2O to promote stability in DMF and the product giving Fmoc-^{L}Lys(me_3·HCl)-OH **5.17**. The products Fmoc-^{D}Lys(me_2·HCl)-OH **5.19** and Fmoc-^{D}Lys(me_3·HCl)-OH **5.20** were synthesized in similar fashion to the L-Lysine stereoisomers but with minor alterations of the reagent composition yielding intermediate products at 71% and 77%, respectively.

The synthesized and commercially available monomers were incorporated in histone mimic peptides using either manual SPPS (**5.3–5.8**) or μW-SPPS (**5.1–5.2, 5.9–5.14**) applying Fmoc- chemistry as described in Chapter 3. A list of synthesised peptides is shown in Table 5.4.1. Peptide purity was determined using LC-MS UV-trace and MS-trace.

Table 5.4.1: Overview of peptides produced with modification and displayed with corresponding peptide number and target protein.

	Peptide	Modification	Number	Annotation in text
Chromatin binding modules	Histone 3 A1-S10	$^{L}K4me_3$	**5.1**	$H3_{1-10}{}^{L}K4me_3$ **5.1**
	Histone 3 A1-S10	$^{D}K4me_3$	**5.2**	$H3_{1-10}{}^{D}K4me_3$ **5.2**
KMT7	Histone 3 A1-A15	$^{L}K4me_0$	**5.3**	$H3_{1-15}{}^{L}K4me_0$ **5.3**
	Histone 3 A1-A15	$^{D}K4me_0$	**5.4**	$H3_{1-15}{}^{D}K4me_0$ **5.4**
KMT5A	Histone 4 G13-Q27	$^{L}K20me_0$	**5.5**	$H3_{13-27}{}^{L}K4me_0$ **5.5**
	Histone 4 G13-Q27	$^{D}K20me_0$	**5.6**	$H3_{13-27}{}^{D}K4me_0$ **5.6**
KMT1C, KMT1D	Histone 3 A1-A15	$^{L}K9me_0$	**5.7**	$H3_{1-15}{}^{L}K9me_0$ **5.7**
	Histone 3 A1-A15	$^{D}K9me_0$	**5.8**	$H3_{1-15}{}^{D}K9me_0$ **5.8**
KDM1A	Histone 3 A1-A21	$^{L}K4me_2$	**5.9**	$H3_{1-21}{}^{L}K4me_2$ **5.9**
	Histone 3 A1-A21	$^{D}K4me_2$	**5.10**	$H3_{1-21}{}^{D}K4me_2$ **5.10**
$KDM4A_{JmjC}$, $KDM4D_{JmjC}$, $KDM4E_{JmjC}$	Histone 3 A1-A15	$^{L}K9me_3$	**5.11**	$H3_{1-15}{}^{L}K9me_3$ **5.11**
	Histone 3 A1-A15	$^{D}K9me_3$	**5.12**	$H3_{1-15}{}^{D}K9me_3$ **5.12**
$KDM5B_{JmjC}$, $KDM5C_{JmjC}$	Histone 3 A1-A21	$^{L}K4me_3$	**5.13**	$H3_{1-21}{}^{L}K4me_3$ **5.13**
	Histone 3 A1-A21	$^{D}K4me_3$	**5.14**	$H3_{1-21}{}^{D}K4me_3$ **5.14**

5.4.2 ITC analysis of histone H3 D- *versus* L-N^{ε}-trimethyl-lysine 4 mimic peptides against chromatin binding modules

In order to study the binding affinity of D- relative to L-N^{ε}-trimethyl-lysine, a panel of chromatin binding modules was selected according to the following criteria: A) the reader is a known binder to histone H3K4 or H3K9 with increasing affinity for higher methylation states ($Kme_0 << Kme_1 < Kme_2$, $< Kme_3$); B) the binding affinity is known to be strong for the reader-histone H3 mimic peptide complex (K_d: <10 μM), C) recombinant proteins can be produced in sufficient quantity for Isothermal titration calorimetry (ITC) analysis from *E. coli* expression systems and D) the reader protein adds diversity to the reader proteins with a different domain or different N^{ε}-trimethyl-lysine 'aromatic binding cage' architecture. Based on these selection criteria, PHD containing proteins ($KDM5A_{PHD3}$, $BPTF_{PHD}$ and $TAF3_{PHD}$) and TTD-containing proteins ($KDM4A_{TTD}$ and $SGF29_{TTD}$) were selected. All proteins were produced *by Dr. Bas Pieters*. The binding strength of chromatin reader domains towards

the histone H3 mimic peptides containing L- and D-N^{ε}-trimethyl-lysine 4, were evaluated using ITC *by Dr. Jasmin Mecinović*. A detailed description of the function and operation of the ITC instrument can be found in section 3.5.3.

The PHD proteins showed good binding affinity to the histone H3 peptide $H3_{1-10}{}^{L}K4me_3$ **5.1**, which is in agreement to previously reported data (Appendix Fig. 9.1, Table 5.4.2).[8,317,319] The strongest binder was $TAF3_{PHD}$ at K_d: 0.082 μM and the weakest binding was determined for $BPTF_{PHD}$ at K_d: 0.44 μM. Inspection of the raw and processed data acquisitions of $H3_{1-10}{}^{D}K4me_3$ **5.2** against PHD readers show a decrease in enthalpy and a decrease in slope observed for all reader proteins. This observation is reflected in the decrease in binding affinity found for the PHD proteins tested with the lowest $TAF3_{PHD}$ (~9-fold) / $BPTF_{PHD}$ (~9-fold) and the highest $KDM5A_{PHD3}$ (~11-fold). While $TAF3_{PHD}$ showed a relatively small decrease in enthalpy of binding (ΔΔH: 1.7 kcal·mol^{-1}), $BPTF_{PHD}$ revealed a substantially larger decrease (ΔΔH: 5.8 kcal·mol^{-1} for $H3_{1-10}{}^{D}K4me_3$ **5.2** in comparison to $H3_{1-10}{}^{L}K4me_3$ **5.1**). An equally small increase in entropy was observed for $TAF3_{PHD}$ at Δ-TΔS: −0.8 kcal·mol^{-1}, but a steep increase was observed for $BPTF_{PHD}$ at Δ-TΔS: −4.2 kcal·mol^{-1}. $KDM5A_{PHD3}$ was found to contain intermediate levels of enthalpy and entropy compensation.

Measurements of TTD readers, $KDM4A_{TTD}$ and $SGF29_{TTD}$, show good binding towards the natural $H3_{1-10}{}^{L}K4me_3$ **5.1** histone H3 mimic substrate (Fig. 9.2, Table 5.4.2) at strengths of K_d: 1.1 μM and K_d: 1.7 μM, respectively. ITC analysis of $H3_{1-10}{}^{D}K4me_3$ **5.2** was performed showing substantially decreased association strengths of ~35-fold for $KDM4A_{TTD}$ and ~29-fold for $SGF29_{TTD}$. A high loss in enthalpy was observed for $KDM4A_{TTD}$ (ΔΔH: 10.6 kcal·mol^{-1}) with high entropic improvement of Δ-TΔS: −8.4 kcal·mol^{-1} showing a relatively high level of enthalpy-entropy compensation.

Table 5.4.2: Obtained ITC values of $H3_{1-10}{}^{L}K4me_3$ 5.1 and $H3_{1-10}{}^{D}K4me_3$ 5.2 binding against PHD-fingers ($KDM5A_{PHD3}$, $BTPF_{PHD}$, $TAF3_{PHD}$) and TTD-domains ($KDM4A_{TTD}$, $SGF29_{TTD}$).[342] D.I.B.: relative difference in binding between $H3_{1-10}{}^{D}K4me_3$ **5.2** and $H3_{1-10}{}^{L}K4me_3$ **5.1**. n= 3–5 measurements, error given as s.d.

	K_d (μM)	D.I.B.	ΔG (kcal·mol^{-1})	ΔH (kcal·mol^{-1})	-TΔS (kcal·mol^{-1})
			$H3_{1-10}{}^{L}K4me_3$ 5.1		
$KDM5A_{PHD3}$	0.11		-9.5 ± 0.1	-11.1 ± 0.1	1.6 ± 0.1
$BPTF_{PHD}$	0.44		-8.7 ± 0.1	-13.2 ± 0.1	4.5 ± 0.1
$TAF3_{PHD}$	0.082		-9.7 ± 0.1	-10.1 ± 0.1	0.4 ± 0.2
$KDM4A_{TTD}$	1.1		-8.1 ± 0.1	-13.1 ± 0.2	5.0 ± 0.2
$SGF29_{TTD}$	1.7		-7.9 ± 0.1	-7.7 ± 0.1	-0.2 ± 0.1
			$H3_{1-10}{}^{D}K4me_3$ 5.2		
$KDM5A_{PHD3}$	1.2	11	-8.1 ± 0.1	-7.8 ± 0.1	-0.3 ± 0.1
$BPTF_{PHD}$	3.8	9	-7.4 ± 0.1	-7.7 ± 0.1	0.3 ± 0.1
$TAF3_{PHD}$	0.73	9	-8.4 ± 0.1	-8.0 ± 0.1	-0.4 ± 0.2
$KDM4A_{TTD}$	39	35	-6.0 ± 0.2	-2.6 ± 0.1	-3.4 ± 0.2
$SGF29_{TTD}$	50	29	-5.9 ± 0.1	-2.3 ± 0.1	-3.6 ± 0.1

Earlier studies on the same chromatin binding modules in complex to $H3_{1-10}$K4G report a K_d of 88 μM for $KDM5A_{PHD3}$ and a K_d of 36 μM for $TAF3_{PHD}$ (>500-fold).[244] These values are substantially weaker than those found for the D/L-N^{ε}-trimethyl-lysine peptides which suggests that the side-chain of D-N^{ε}-trimethyl-lysine 4, like L-N^{ε}-trimethyl-lysine 4, is involved in the interaction forming the reader-peptide complex. However, the D-lysine side chain may be less accommodating towards binding in the aromatic 'cage' or may (partially) disrupt favourable interactions between neighbouring residues in the protein complex. Furthermore, the backbone could potentially be partially rotated locally to maximize interaction of D-N^{ε}-trimethyl-lysine to the 'aromatic cage' at the cost of favourable interactions to the amide-bonds.

5.4.3 MD simulating interactions of reader domains to H3 D- and L-*N*ε-trimethyl-lysine

ITC revealed that reader proteins can accommodate D-*N*ε-trimethyl-lysine 4 residue but at the cost of binding affinity. To understand the binding properties of the formed complex, temporal atomistic molecular dynamics were calculated (*performed by Dr. Kiran Kumar and Dr. Robert S. Paton*).

Previous reports demonstrated that MD can be a useful method to understand the binding of H2AK5ac, H3K14ac, H4K12ac to BRPF1 using Amberff99sb force field.[343] Additionally, studies were performed on KDM5C using Amberff14sb to compute the binding affinity for H3K9me$_3$ and H3K4me$_3$ for the PHD1 and JmjC-domain, respectively.[344] Here, computational models were carried out modelling deposited crystal structures of KDM5A$_{PHD3}$ (PDB: 2KGI), BPTF$_{PHD}$ (PDB: 2F6J), TAF3$_{PHD}$ (PDB: 2K17), KDM4A$_{TTD}$ (PDB: 2GFA) and SGF29$_{TTD}$ (PDB: 3ME9) containing H3^{L}K4me$_3$ substrate using AMBER12 (Fig. 5.4.2).[345] The structures were solvated in water (TIP3P)[346] in a 10.0 Å octahedral cage, while sodium and chloride ions were added to counter charged protein residues. D-*N*ε-trimethyl-lysine 4 was modelled based on the L-*N*ε-trimethyl-lysine 4, and the substrates were minimized and equilibrated (Fig. 5.4.2). Each simulation was computed for 10 nano seconds (ns) with iterative steps of 20 picoseconds (ps). Snapshots at 0, 5 and 10 ns binding configurations are displayed for the PHD fingers in Fig. 5.4.3 and TTD in Fig. 5.4.4. Analysis of the simulations reveals the distance between the aromatic cage of the reader relative to *N*ε-trimethyl-lysine (Fig. 9.3, Fig. 9.4). The most similar orientation was found for KDM5A$_{PHD3}$ W1625 and W1635 in both *N*ε-^{L}K4me$_3$ and *N*ε-^{D}K4me$_3$ at t= 0 ns but, SGF29$_{TTD}$ was the only protein in which the D-lysine residues did not have the same orientation as the L-lysine residue after minimization and equilibration. The D-lysine residue however, stabilized throughout the computational analysis.

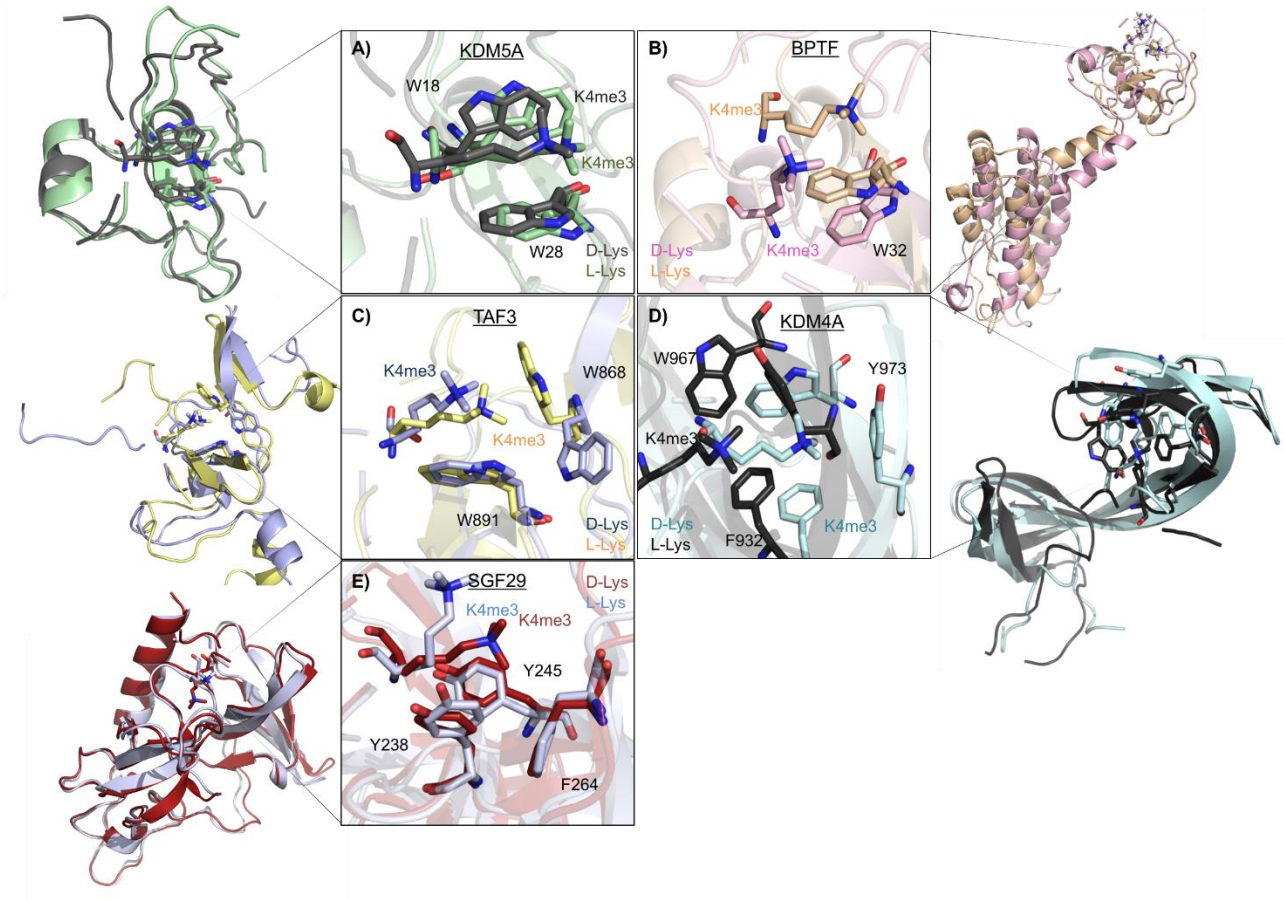


Figure 5.4.2: Crystal structures of reader proteins aromatic cage examined with overlay between ^{D}K4me$_3$ *versus* ^{L}K4me$_3$ peptide at the start (t=0 ns) of the computation after minimization and equilibration.[342] A) KDM5A$_{PHD3}$ in complex to ^{L}K4me$_3$ (light green) or ^{D}K4me$_3$ (dark grey) (PDB: 2KGI); B) BPTF$_{PHD}$ in complex to ^{L}K4me$_3$ (beige) or ^{D}K4me$_3$ (pink) (PDB: 2F6J); C) TAF3$_{PHD}$ in complex to ^{L}K4me$_3$ (yellow) or ^{D}K4me$_3$ (light blue) (PDB: 2K17); D) KDM4A$_{TTD}$ in complex to ^{L}K4me$_3$ (black) or ^{D}K4me$_3$ (cyan) (PDB: 2GFA); E) SGF29$_{TTD}$ in complex to ^{L}K4me$_3$ (light blue) or ^{D}K4me$_3$ (red) (PDB: 3ME9). Atom annotation, oxygen: red, nitrogen: blue. This figure is reproduced with permission from the Royal Society of Chemistry.[342]

Interestingly, based on the MD simulations, none of the bound peptides were dissociated from the reader proteins. KDM5A$_{PHD3}$ (W1625, W1635) and TAF3$_{PHD}$ (W868, W891) determined to have big variations in histone H3 C^{α} movement for the duration of the simulation (Fig. 9.3). Herein, KDM5A$_{PHD3}$ W1625 and W1635 showed little difference in distance between the aromatic cage and N^{ε}-trimethyl-lysine with similar observations made for KDM4A$_{TTD}$ (F932, W967 and Y238) and SGF29$_{TTD}$ (F164, Y238, Y245). TAF3$_{PHD}$, on the other hand revealed to have relatively high movement towards ^{D}K4me$_3$ with W868 and W891 but the relative changes in backbone orientation were similar to that of KDM4A$_{TTD}$, SGF29$_{TTD}$ or BPTF$_{PHD}$. These findings collected using MD

simulations propose that ^{L}K4me$_3$ is the preferred substrate but ^{D}K4me$_3$ retains some affinity towards the reader protein.

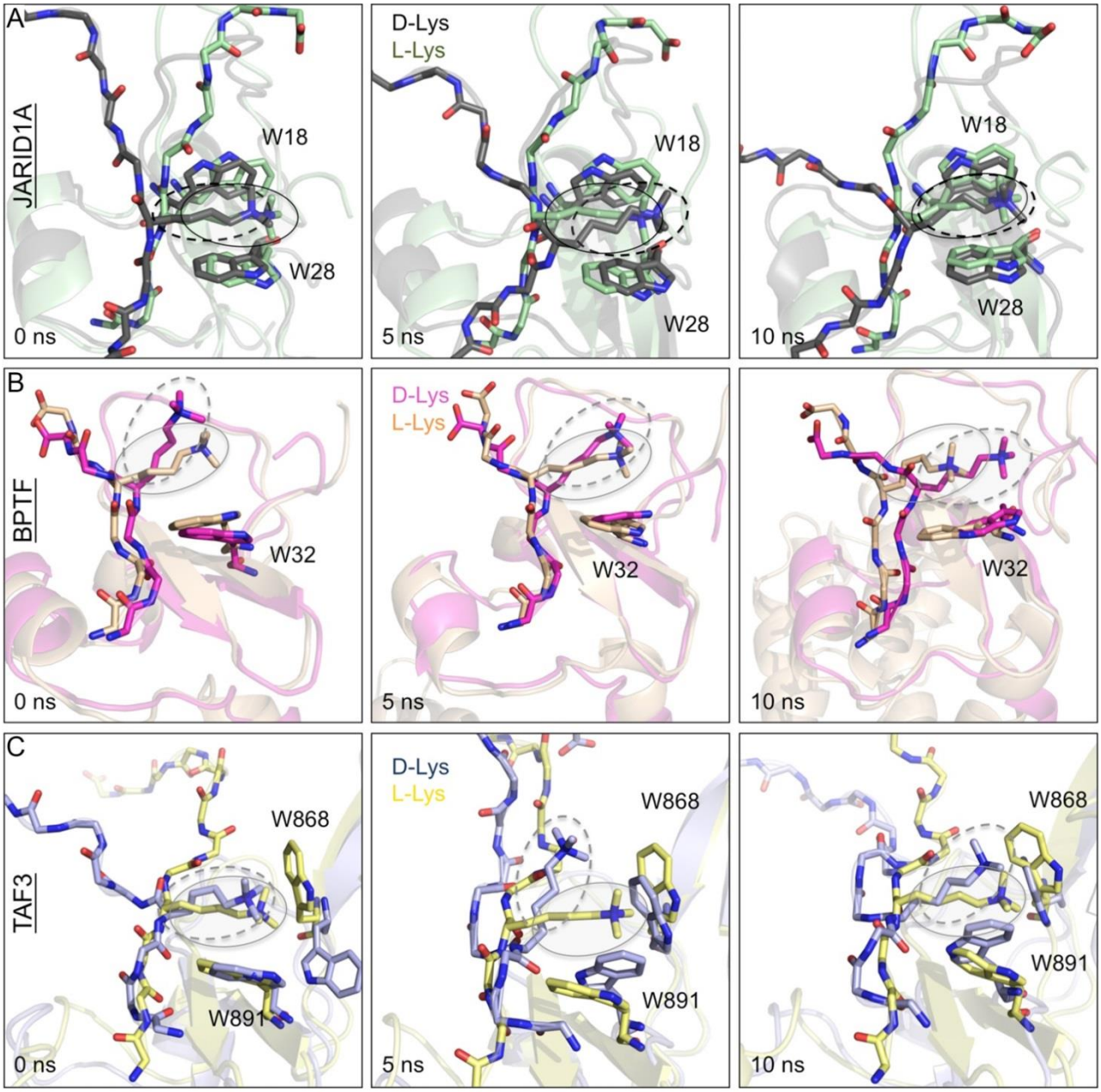


Figure 5.4.3: Molecular dynamic simulation snap shot at t= 0, 5 and 10 ns frame interval of: A) KDM5A$_{PHD3}$–H3 complex D-lysine (dark grey), L-lysine (green) (PDB: 2KGI); B) BPTF$_{PHD}$–H3 complex D-lysine (pink), L-lysine (beige) (PDB: 2F6J) and C) TAF3$_{PHD}$–H3 complex D-lysine (light blue), L-lysine (yellow) (PDB: 2K17). This figure is reproduced with permission from the Royal Society of Chemistry.[342]

During the simulations, the electrostatics (E_{ele}) were calculated between protein aromatic residues and the N^{ε}-^{L}K4me$_3$ or N^{ε}-^{D}K4me$_3$. In most cases, strong cation-π interactions were observed, with D-N^{ε}-trimethyl-lysine being less favourable relative to L-N^{ε}-trimethyl-lysine. SGF29$_{TTD}$, however, revealed

a favourable ΔE_{ele} for N^{ε}-^{D}K4me$_3$, although this could be caused by the different orientations of this residues at the beginning of the simulation.

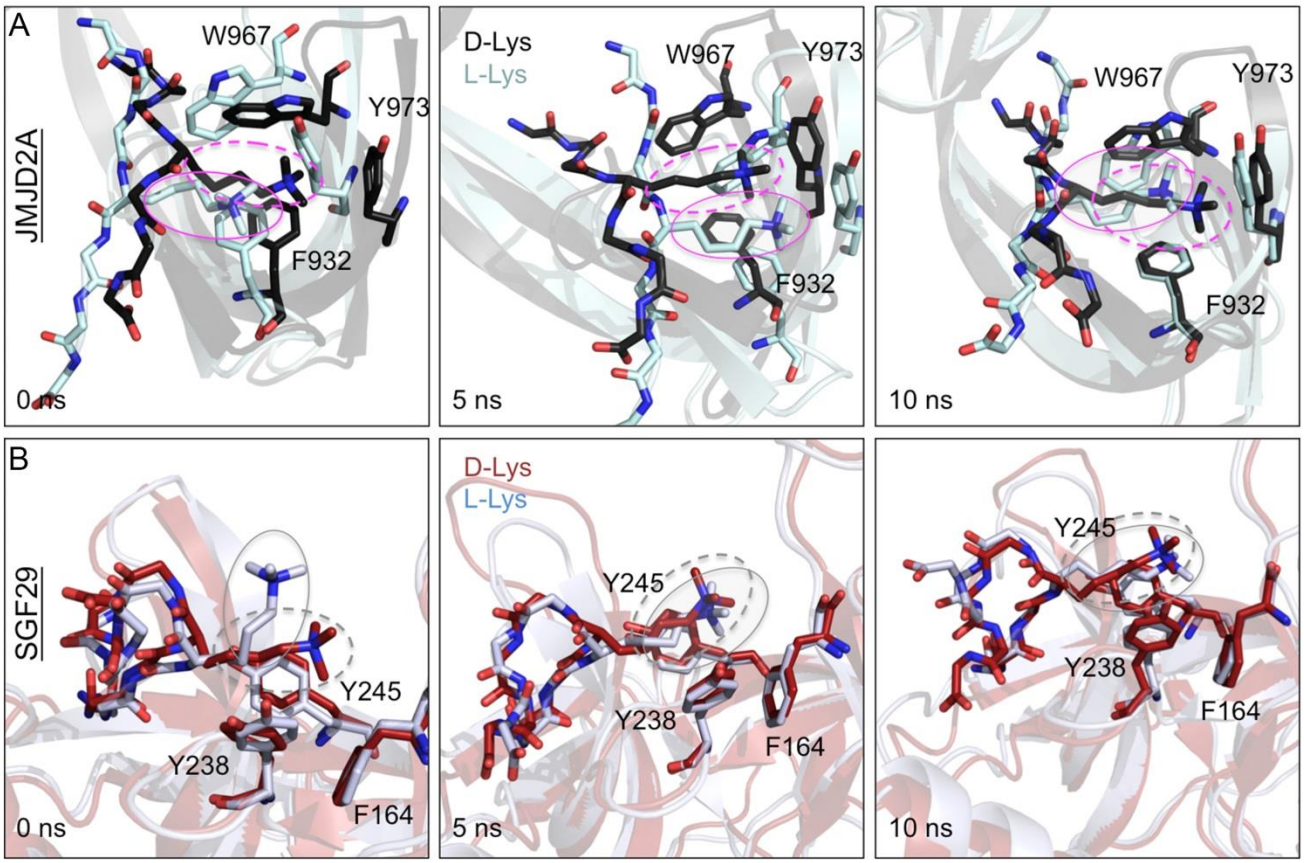


Figure 5.4.4: Simulation snapshot computed using MD based on TTD containing proteins at t= 0, 5, 10 ns. A) KDM4A$_{TTD}$ in complex to ^{D}K4me$_3$ (light cyan) or ^{L}K4me$_3$ (black) peptide (PDB: 2GFA) and B) SGF29$_{TTD}$ in complex to ^{D}K4me$_3$ (Bordeaux red) or ^{L}K4me$_3$ (light purple) peptide (PDB: 2ME9). This figure is reproduced with permission from the Royal Society of Chemistry.[342]

The evidence from ITC and MD suggest that L-N^{ε}-trimethyl-lysine is the prefered binder over D-N^{ε}-trimethyl-lysine however, the D-lysine stereo-chemistry provides micromolar affinity when complexed to chromatin binding modules.

5.4.4 Histone methyltransferase (KMTs) can methylate D-N^{ε}-lysine residues in histone H3 substrates

Work in this section (5.4.4) has been conducted by Abbas Al Temimi. The *in vitro* and *in silico* binding studies using ITC and MD illustrate that D-N^{ε}-trimethyl-lysine remains a strong binding companion to L-N^{ε}-trimethyl-lysine.

To investigate this effect on histone lysine methyltransferases (KMTs), four representative proteins were selected based on: A) histone lysine specificity; B) high construct methylation efficiency and C) ease of expression in *E. coli*. The tested proteins were KMT7 (SETD7), KMT5A (SETD8), KMT1C (G9a) and KMT1D (GLP). KMT7 catalyses N^{ε}-mono-methylation on H3K4 while KMT5A catalyses N^{ε}-mono-methylation on H4K20. KMT1C and KMT1D catalyse N^{ε}-mono-/di-/tri-methylation on H3K9. The complementary histone H3 or H4 mimic peptides were prepared containing either D- or L-lysine stereochemistry for the target methylation site.

The activity of KMT7 (2.0 μM) and KMT5A (2.0 μM) was determined with substrates $H3_{1\text{-}15}{}^{L}K4me_0$ **5.3** or $H4_{13\text{-}27}{}^{L}K20$ **5.5** (100 μM), respectively. The reaction reached completion after 1 hour at 37 °C in the presence of SAM (200 μM) (Fig. 5.4.5 A,B). Corresponding control experiments showed that, in the absence of enzyme and SAM or in the absence of SAM, only traces of methylated peptide were observed. On the contrary, $H3_{1\text{-}15}{}^{D}K4$ **5.4** (100 μM) peptide was not efficiently methylated under these conditions but revealed only traces masses that may be assigned to mono methylation (Fig. 5.4.5 A,B). Attempts were made to promote methylation of the D-lysine residue by incubating the $H3_{1\text{-}15}{}^{D}K4$ peptide **5.4** with elevated levels of KMT7 (10 μM) or KMT5A (10 μM), SAM (1.0 mM) and extended exposure time (6 hrs). However, this did not yield higher levels of observable methylated D-lysine peptide.

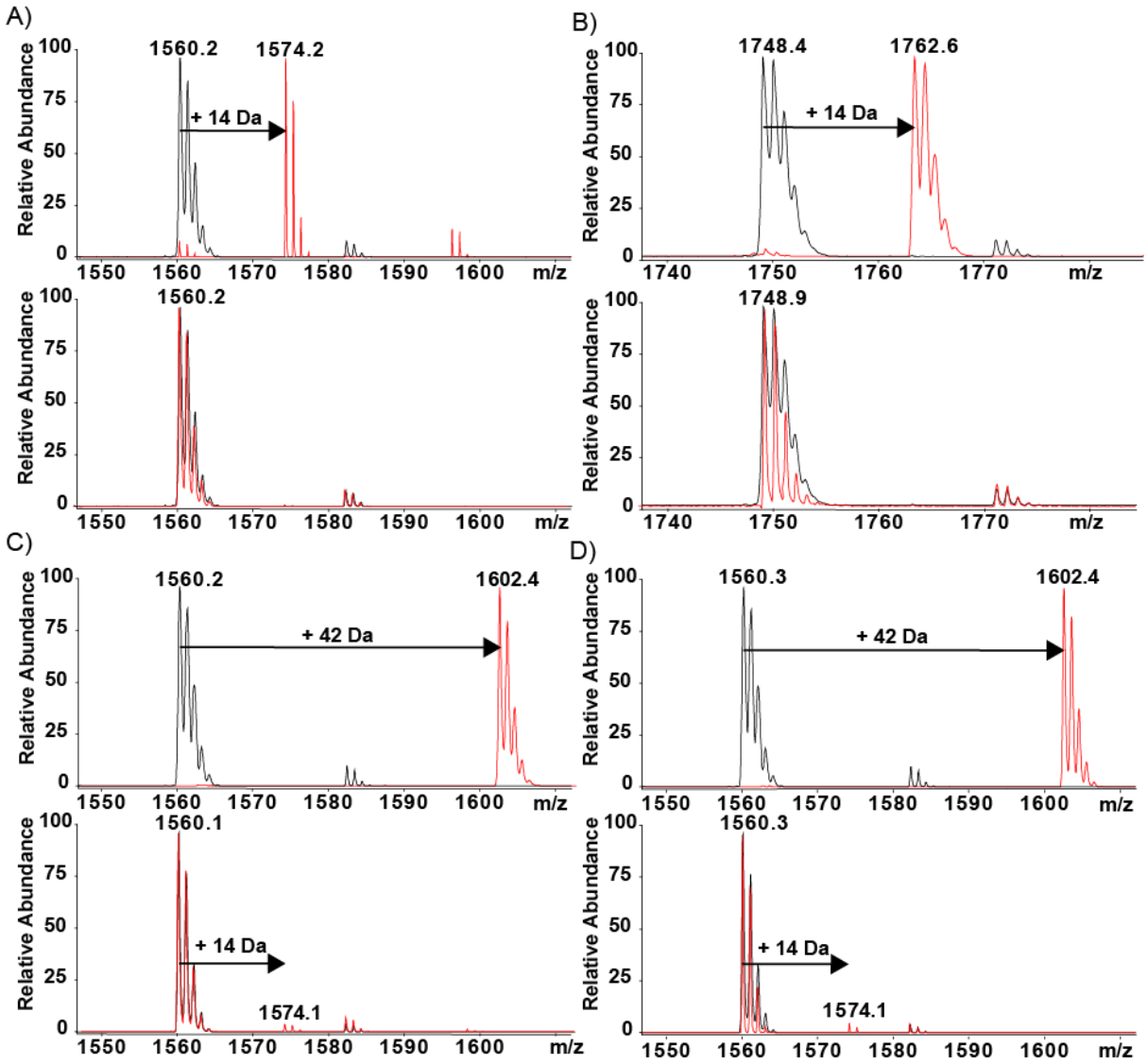


Figure 5.4.5: MALDI-TOF mass spectra of KMT methylation of L-lysine (top panel) and D-lysine residues (bottom panel) at t=0 min (black) and t= 60 min (red). A) KMT7 with $H3_{1\text{-}15}{}^{L}K4$ **5.3** (upper) and $H3_{1\text{-}15}{}^{D}K4$ **5.4** (lower); B) KMT5A with $H4_{13\text{-}27}{}^{L}K20$ **5.5** and $H4_{13\text{-}27}{}^{D}K20$ **5.6**; C) KMT1C with $H3_{1\text{-}15}{}^{L}K9$ **5.7** and $H3_{1\text{-}15}{}^{D}K9$ **5.8**; D) KMT1D with $H3_{1\text{-}15}{}^{L}K9$ **5.7** and $H3_{1\text{-}15}{}^{D}K9$ **5.8**. This figure is reproduced with permission from the Royal Society of Chemistry.[342]

While observations of KMT7 and KMT5A show traces of methylation on their designated D-Lysine substrate, KMT1D and KMT1C revealed some evidence for D-lysine methylation. Mass spectrometry indicated that KMT1D (2.0 μM) and KMT1C (2.0 μM), when incubated with $H3_{1\text{-}15}{}^{D}K9$ **5.8** (100 μM) and SAM (500 μM), produced traces of methylated product. Increasing the KMT1D or KMT1C concentrations to 10 μM resulted in approximately 30% conversion of the monomethylated $H3_{1\text{-}15}{}^{D}K9me_1$ **5.8B** product. Traces of the dimethylated $H3_{1\text{-}15}{}^{D}K9me_2$ **5.8C** (1 hrs, 37 °C) were also

observed. Incubation of the mixture for 6 hrs produced even higher levels of methylation, with predominantly the trimethylated $H3_{1-15}{}^{D}K9me_3$ **5.8D** product of KMT1D and KMT1C being observed.

To understand the impact of the D-lysine residue relative to the L-stereoisomer, MALDI-TOF based competition assays were conducted. Due to the identical mass shifts of D- *versus* L-peptides, a shorter $H3_{1-14}{}^{L}K9$ **5.21** histone H3 mimic peptide was produced and purified (Fig. 5.4.6 A). The $H3_{1-14}{}^{L}K9$ **5.21** peptide was tested against KMT1C (2.0 μM) protein, which revealed complete methylation of the short peptide (Fig. 5.4.6 B). In a competition assay with equal quantities of $H3_{1-15}{}^{L}K9$ **5.7** (100 μM) and $H3_{1-14}{}^{L}K9$ **5.21** (100 μM) peptides, lower levels of methylation were observed due to the increased quantity of total peptide (200 μM). However, very similar methylation levels were observed for each peptide (Fig. 5.4.6 C). Under the same conditions, $H3_{1-14}{}^{L}K9$ **5.21** (100 μM) and $H3_{1-15}{}^{D}K9$ **5.8** (100 μM) were tested, which revealed formation of $H3_{1-14}{}^{L}K9me_2$ **5.21C** and $H3_{1-14}{}^{L}K9me_3$ **5.21D** peptides (Fig. 5.4.6 D). In a competition experiment between $H3_{1-14}{}^{L}K9$ **5.21** (100 μM) and $H3_{1-15}{}^{D}K9$ **5.8** (100 μM), only trace amounts of $H3_{1-15}{}^{D}K9me_1$ were observed, indicating that $H3_{1-14}{}^{L}K9$ **5.21** is a far superior substrate over $H3_{1-15}{}^{D}K9$ **5.8**. Since $H3_{1-15}{}^{D}K9$ **5.8** was found to be a poor substrate for KMT1C, it might also be considered as an inhibitor. When figure 5.4.6 panels B and D are compared, the levels of methylated peptide of $H3_{1-14}{}^{L}K9me_2$ **5.21C** and $H3_{1-14}{}^{L}K9me_3$ **5.21D** presented show only a slight decrease of $H3_{1-14}{}^{L}K9me_3$ **5.21D** formation. This suggests that $H3_{1-15}{}^{D}K9$ **5.8** induces a relatively weak level of inhibition against KMT1C under the experimental conditions.

These recordings reveal that KMT1D and KMT1C histone methyltransferases can tolerate D-Lysine stereochemistry but at the cost of efficiency. KMTs KMT7 and KMT5A on the contrary are potentially only able to catalyse methylation of D-lysine peptide at trace levels

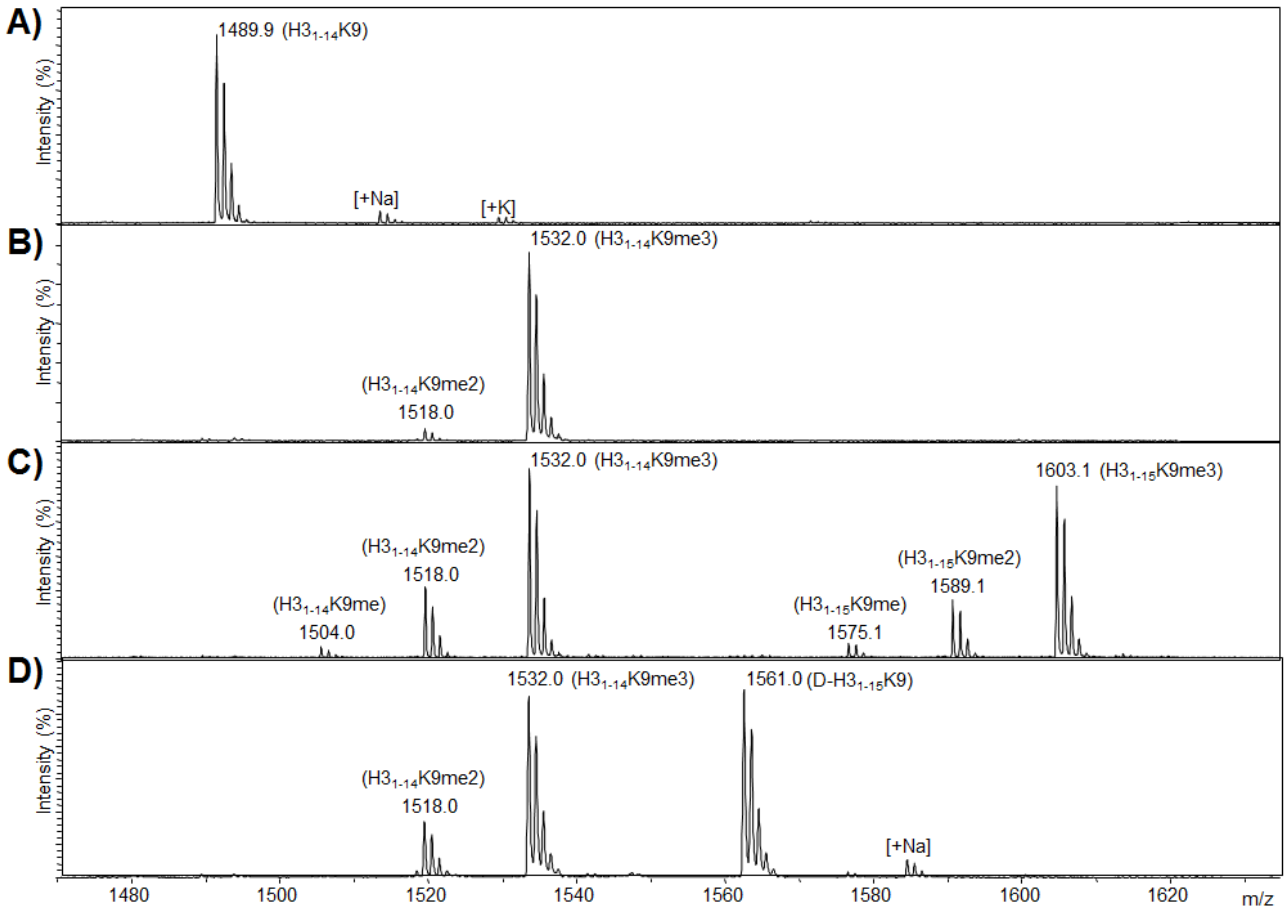


Figure 5.4.6: Competition examination of KMT1C against D- *versus* L-lysine determined using MALDI-TOF. A) $H3_{1-14}{}^{L}K9$ **5.21** peptide (100 μM); B) Activity assay of KMT1C (2.0 μM) against $H3_{1-14}{}^{L}K9$ **5.21** (100 μM) to provide full conversion to product $H3_{1-14}{}^{L}K9me_3$ **5.21D** (100 μM) after incubation (1 hrs, 37 °C); C) competition assay of $H3_{1-14}{}^{L}K9$ **5.21** (100 μM) and $H3_{1-15}{}^{L}K9$ **5.7** (100 μM) against KMT1C (2.0 μM) (1 hrs, 37 °C); D) $H3_{1-14}{}^{L}K9$ **5.21** (100 μM) in the presence of $H3_{1-15}{}^{D}K9$ **5.8** (100 μM) and KMT1C (2.0 μM) (1 hrs, 37 °C). This figure is reproduced with permission from the Royal Society of Chemistry.[342]

5.4.5 KDM activity on D-N^{ε}-methylate-lysine residues in histone H3 peptide mimic substrates

To determine the substrate specificity on L- and D-N^{ε}-trimethyl-lysine for lysine demethylases (KDMs), a panel of KDMs containing either the JmjC or amine oxidase domain were produced. Firstly, KDMs were selected based on: A) the known expression construct; B) ability to efficiently catalyse demethylation of N^{ε}-dimethyl- or N^{ε}-trimethyl-lysine and C) their type of domain (JmjC or Amine oxidase). Of the amine oxidases, $KDM1A_{C1\text{-}corest}$ was selected and produced in *E. coli (Produced by*

Dr. James E. Dunford). JmjC-domain containing proteins $KDM4A_{JmjC}$ *(Aleksandra Szykowska)*, $KDM4D_{JmjC}$ *(Aleksandra Szykowska)* and $KDM4E_{JmjC}$ (*Dr. Richard J. Hopkinson)* were recombinantly produced in *E. coli*. $KDM5B_{JmjC}$ and $KDM5C_{JmjC}$, were produced in insect cells (Sf9) / Baculo Virus expression systems *(by Dr. Catrine Johansson and Aleksandra Szykowska, respectively)*. The corresponding histone H3 mimic peptide substrates were synthesized using μW-SPPS and the crude peptide was HPLC purified (>91% purity, typical ~95%).

To examine the demethylation rates of N^{ε}-methylated-lysine substrates by KDMs, Agilent RapidFire mass spectrometry (RF-MS) was employed. RF-MS enables relatively fast and sensitive determination of substrate to product(s) conversion based on high resolution mass spectrometry. In these experiments, samples are processed within 12 to 14 sec using robotic controlled sipper and vacuum aspiration equipment. Due to the rapid assessment, progress of the enzymatic reactions was monitored real-time and in multiple reaction monitoring (MRM) mode. RF-MS, unlike MALDI-TOF, uses electron spray ionisation (ESI) to inject the molecules into the MS resulting in multiple charge states observed for the used peptides. The overall conversion is based on the individual peak integral divided by the sum of all integrals and normalized to 100 or the peptide concentration as determined using RF-MS. Spectra visualized are represented as deconvoluted peptide masses. Initially, the catalytic demethylation activity of enzymes $KDM1A_{FL}$, $KDM4A_{JmjC}$, $KDM4D_{JmjC}$, $KDM4E_{JmjC}$, $KDM5B_{JmjC}$, $KDM5C_{JmjC}$ was determined using the natural histone H3 substrate (Fig. 5.4.7, 5.4.8 and 5.4.9). All RF-MS experiments were performed in MRM mode at room temperature (22.0 °C).

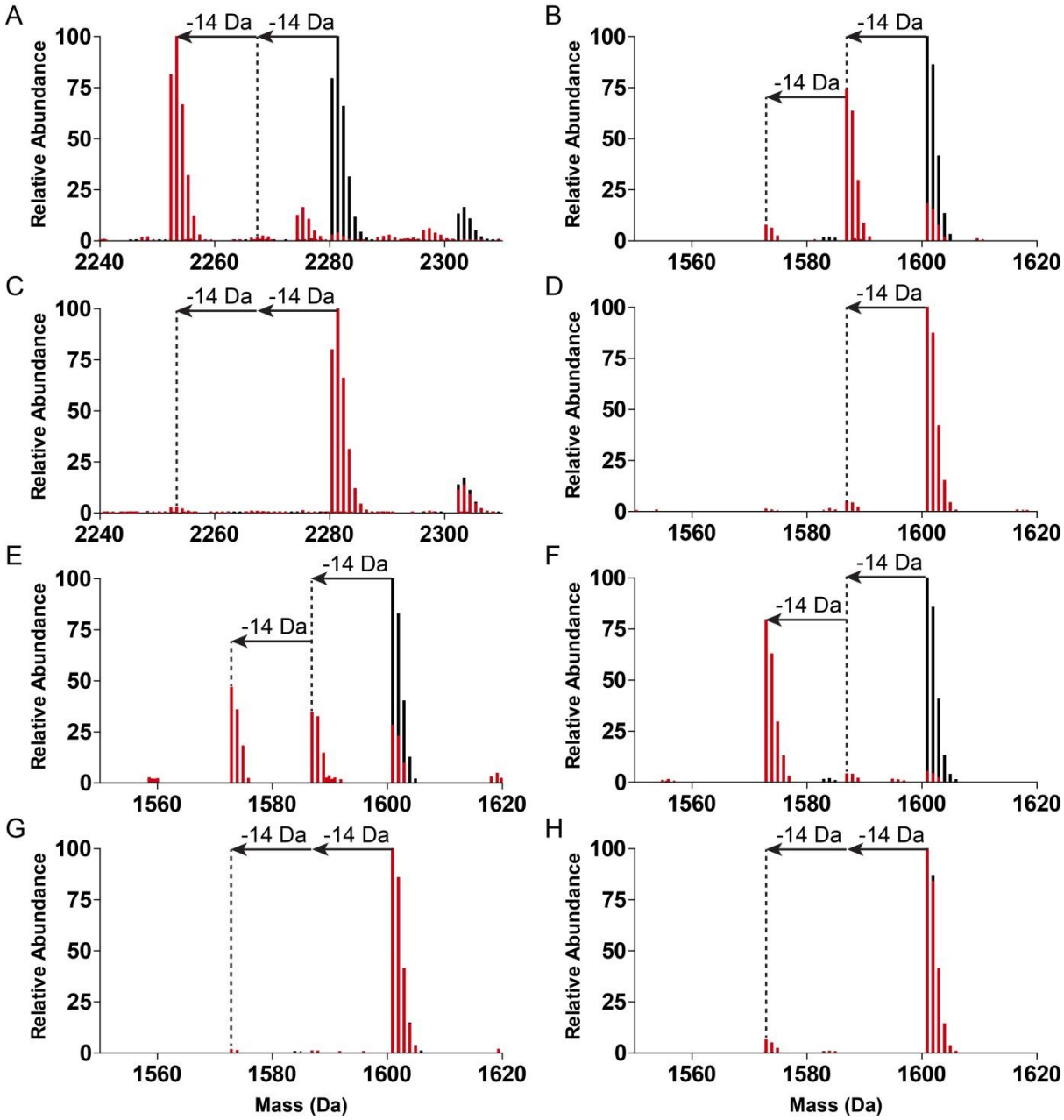


Figure 5.4.7: Deconvoluted enzymatic reactions on L- or D-*N*$^{\varepsilon}$-trimethyl-lysine containing histone H3 mimic peptides using LC-MS normalized based on observed peak surface area. KDM1A$_{C1\text{-}corest}$ (200 nM) (A,C) using H3$_{1\text{-}21}$^{L}K4me$_2$ **5.9** (10.0 µM) (A) and H3$_{1\text{-}21}$^{D}K4me$_2$ **5.10** (10.0 µM) (C). KDM4A$_{JmjC}$ (200 nM) (B,D), KDM4D$_{JmjC}$ (200 nM) (E,G) and KDM4E$_{JmjC}$ (200 nM) (F,H) against peptide H3$_{1\text{-}15}$^{L}K9me$_3$ **5.11** (10.0 µM) (B,E,F) and H3$_{1\text{-}15}$^{D}K9me$_3$ **5.12** (10.0 µM) (D,G,H). This figure is reproduced with permission from the Royale Society of Chemistry.[342]

KDM1A$_{C1\text{-}corest}$ (200 nM) catalysed demethylation of H3$_{1\text{-}21}$^{L}K4me$_2$ **5.9** (10 µM) efficiently to give product masses that could be assigned to H3$_{1\text{-}21}$^{L}K4me$_1$ **5.9B** followed by iterative demethylation resulting in near complete substrate conversion to H3$_{1\text{-}21}$^{L}K4me$_0$ **5.9C** (Fig. 5.4.7 A, Fig. 5.4.9 A (full

lines)). Interestingly, KDM1A$_{C1\text{-}corest}$ was unable to catalyse demethylation of the H3$_{1\text{-}21}$^{D}K4me$_2$ **5.10** stereoisomer, showing only traces of demethylated product H3$_{1\text{-}21}$^{D}K4me$_0$ **5.10C** under the same conditions.

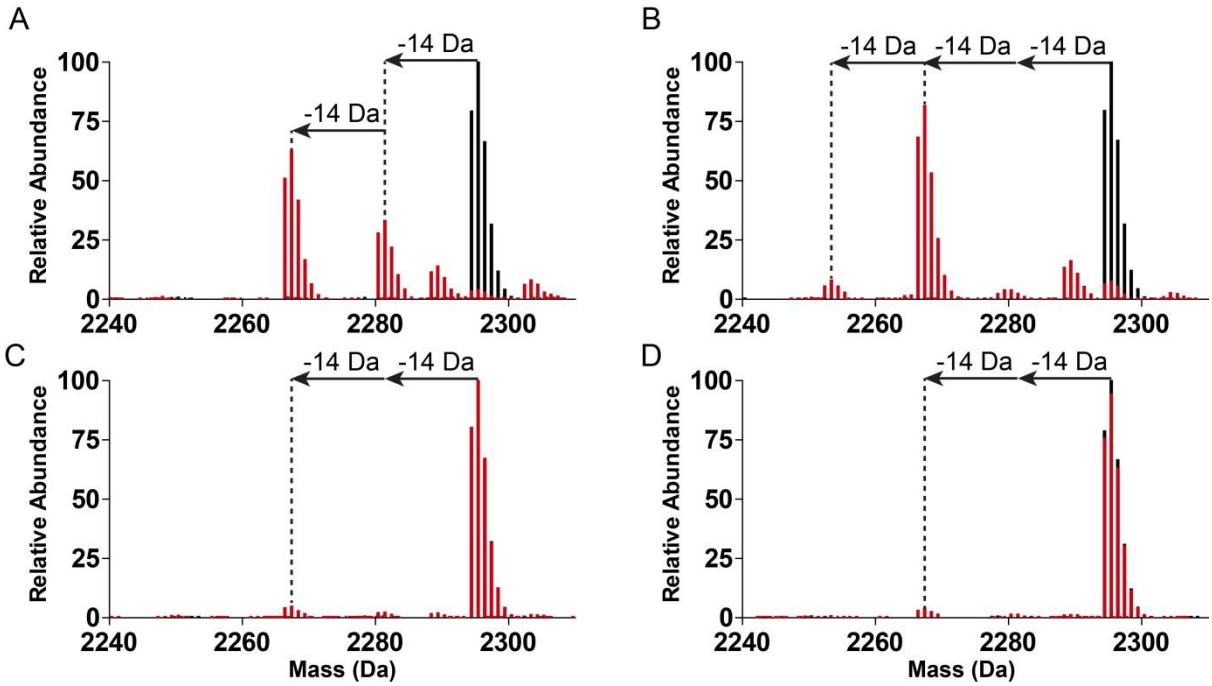


Figure 5.4.8: Deconvoluted LC-MS observations based from KDM5B$_{JmjC}$ (200 nM) (A, C) and KDM5C$_{JmjC}$ (200 nM) (B, D) with substrate H3$_{1\text{-}21}$^{L}K4me$_3$ 5.13 (5.0 μM) (A, B) and H3$_{1\text{-}21}$^{D}K4me$_3$ 5.14 (5.0 μM) (C, D). Clear demethylation is observed for KDM5B$_{JmjC}$ and KDM5C$_{JmjC}$ in the presence of ^{L}K4 peptide, but little is observed for ^{D}L4. Black t=0 in the absence of enzyme and red is t> 60 min. This figure is reproduced with permission from the Royal Society of Chemistry.[342]

KDM4A$_{JmjC}$, KDM4D$_{JmjC}$ and KDM4E$_{JmjC}$ were tested for their demethylation capability using RF-MS in the presence of sodium ascorbate (Asc) (100 μM), $(NH_4)_2Fe(II)(SO_4)_2$ (10 μM) and disodium 2-oxoglutarate (2OG) (30 μM) in buffer [2-*N*-morpholino ethane sulfonate (MES) (50 mM) in MilliQ (pH 7.0)] and H3$_{1\text{-}15}$K9me$_3$ **5.11** (10 μM). The formation of H3$_{1\text{-}15}$^{L}K9me$_2$ **5.11B** was predominantly observed for KDM4A$_{JmjC}$ reaction after ~1 hrs, with low levels of H3$_{1\text{-}15}$^{L}K9me$_1$ **5.11**. Demethylation experiments with KDM4E$_{JmjC}$ demonstrated higher demethylation than with KDM4A$_{JmjC}$ under the same conditions – the major peaks observed after 1 hrs correspond to H3$_{1\text{-}15}$^{L}K9me$_1$ **5.11C**, H3$_{1\text{-}15}$^{L}K9me$_2$ **5.11B** and H3$_{1\text{-}15}$^{L}K9me$_3$ **5.11** peptides (Fig. 5.4.7 B,F, Fig. 5.4.9 B,D). KDM4D$_{JmjC}$ was observed to manifest catalytic activity between that of KDM4A$_{JmjC}$ and KDM4E$_{JmjC}$ (Fig. 5.4.7 E, Fig. 5.4.9 C).

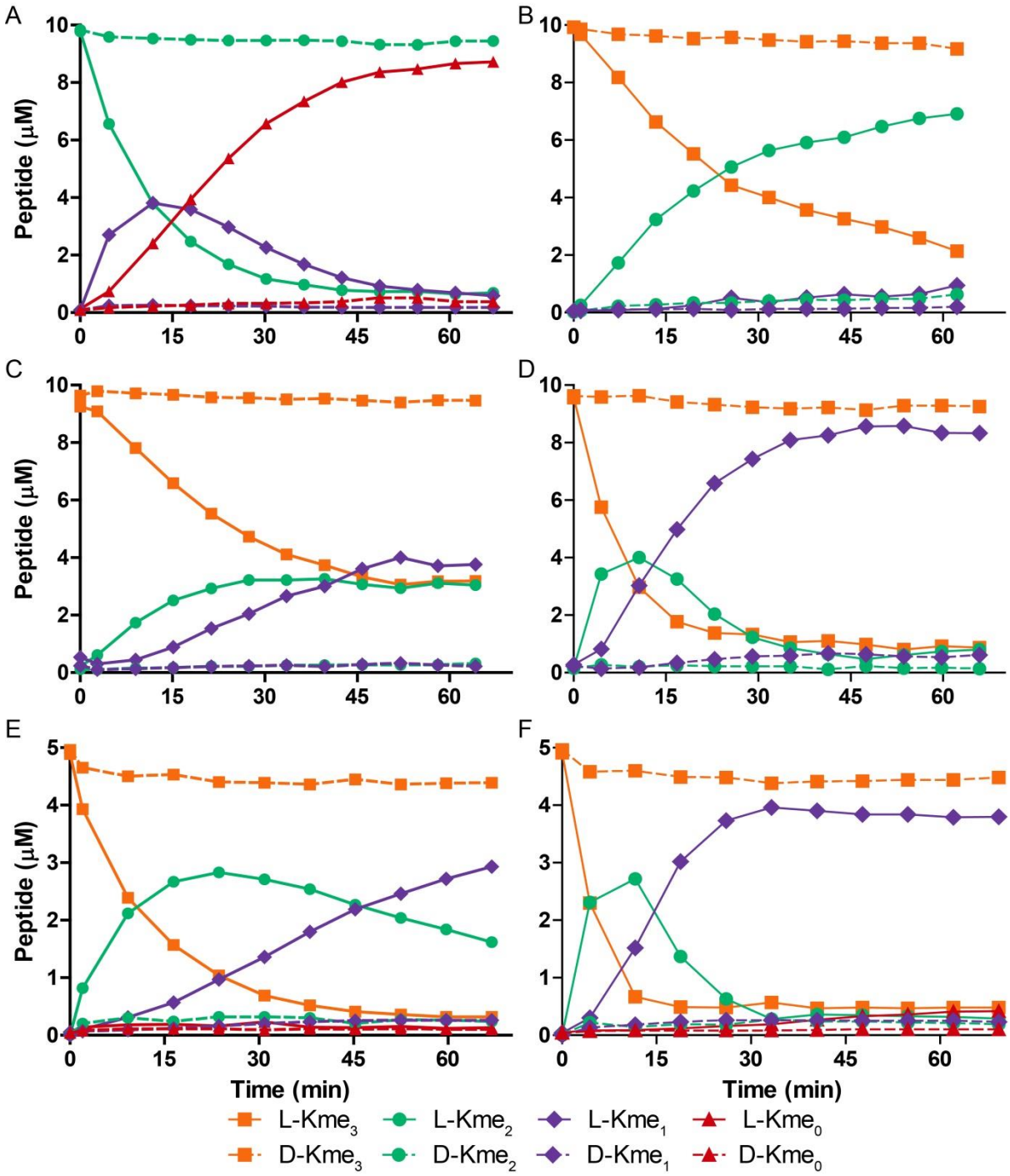


Figure 5.4.9: Time-course LC-MS activity assay of KDM enzymatic reaction against histone H3 substrate containing L-lysine (full lines) or D-Lysine substrates (dashed lines) displaying corresponding decrease in substrate concentration and increase in product formation. Intensity is normalized to the substrate concentration and peptide composition based on the relative peak area observed. A) $KDM1A_{FL}$ (200 nM) with $H3_{1\text{-}21}{}^{L}K4me_2$ **5.9** and $H3_{1\text{-}21}{}^{D}K4me_2$ **5.10** substrates (10.0 μM); B,C,D) $KDM4A_{JmjC}$ (200 nM) (B), $KDM4D_{JmjC}$ (200 nM) (C), $KDM4E_{JmjC}$ (200 nM) with $H3_{1\text{-}15}{}^{L}K9me_3$ **5.11** and $H3_{1\text{-}15}{}^{D}K9me_3$ **5.12** substrates (10.0 μM); E,F) $KDM5B_{JmjC}$ (200 nM) (E), $KDM5C_{JmjC}$ (200 nM) (F) with $H3_{1\text{-}21}{}^{L}K4me_3$ **5.13** and $H3_{1\text{-}21}{}^{D}K4me_3$ **5.14** substrates (5.0 μM). This figure is reproduced with permission from the Royal Society of Chemistry.[342]

The substrate stereoisomer $H3_{1-15}{}^{D}K9me_3$ **5.12** was then tested against $KDM4A_{JmjC}$ (Fig. 5.4.7 D, Fig. 5.4.9 B), $KDM4D_{JmjC}$ (Fig. 5.4.7 G, Fig. 5.4.9 C) and $KDM4E_{JmjC}$ (Fig. 5.4.7 H, Fig. 5.4.9 D), which revealed only traces of $H3_{1-15}{}^{D}K9me_2$ **5.12B** or $H3_{1-15}{}^{D}K9me_1$ **5.12C** after 1 hour.

The activity of $KDM5B_{JmjC}$ and $KDM5C_{JmjC}$ was then determined using LC-MS. A time-course, carried out with $H3_{1-21}{}^{L}K4me_3$ **5.13** substrate (5.0 μM), Asc (100 μM), (2OG) (10 μM), $(NH_4)_2Fe(II)(SO_4)_2$ (10 μM) in buffer [MES (50 mM), NaCl (50 mM), pH 7.0 in MilliQ]. $KDM5B_{JmjC}$ (Fig. 5.4.8 A, Fig. 5.4.9 E) and $KDM5C_{JmjC}$ (Fig. 5.4.8 B, Fig. 5.4.9 F), showed demethylation of $H3_{1-21}{}^{L}K4me_3$ **5.13** to give the $H3_{1-21}{}^{L}K4me_2$ **5.9** and $H3_{1-21}{}^{L}K4me_1$ **5.9B**. In addition, $KDM5C_{JmjC}$ also produced low levels of $H3_{1-21}{}^{L}K4me_0$ **5.9C**. Time-courses conducted with $KDM5B_{JmjC}$ and $KDM5C_{JmjC}$ with $H3_{1-21}{}^{D}K4me_3$ **5.14** produced only traces of $H3_{1-21}{}^{D}K4me_1$ **5.10B,** indicating that D-lysine is not a favourable substrate for the KDM5 protein constructs tested. To understand the effect of the lysine stereochemistry, the catalytic activity of $KDM4E_{JmjC}$, a representative JmjC-KDM, was investigated using ^1H NMR spectroscopy. A sample of $H3_{1-15}{}^{L}K9me_3$ **5.11** peptide (392 μM), Asc (1.0 mM), $(NH_4)_2Fe(II)(SO_4)_2$ (100 μM), 2OG (800 μM) in ammonium formate buffer (40 mM, pH 7.5) with ~10% D_2O composition was initiated with $KDM4E_{JmjC}$ (11.5 μM) and the reaction was monitored using AV-700 NMR equipped with inverse helium cryo-probe. Acquisition was conducted every 113 sec interval and the first and final acquisitions are displayed in figure 5.4.10.

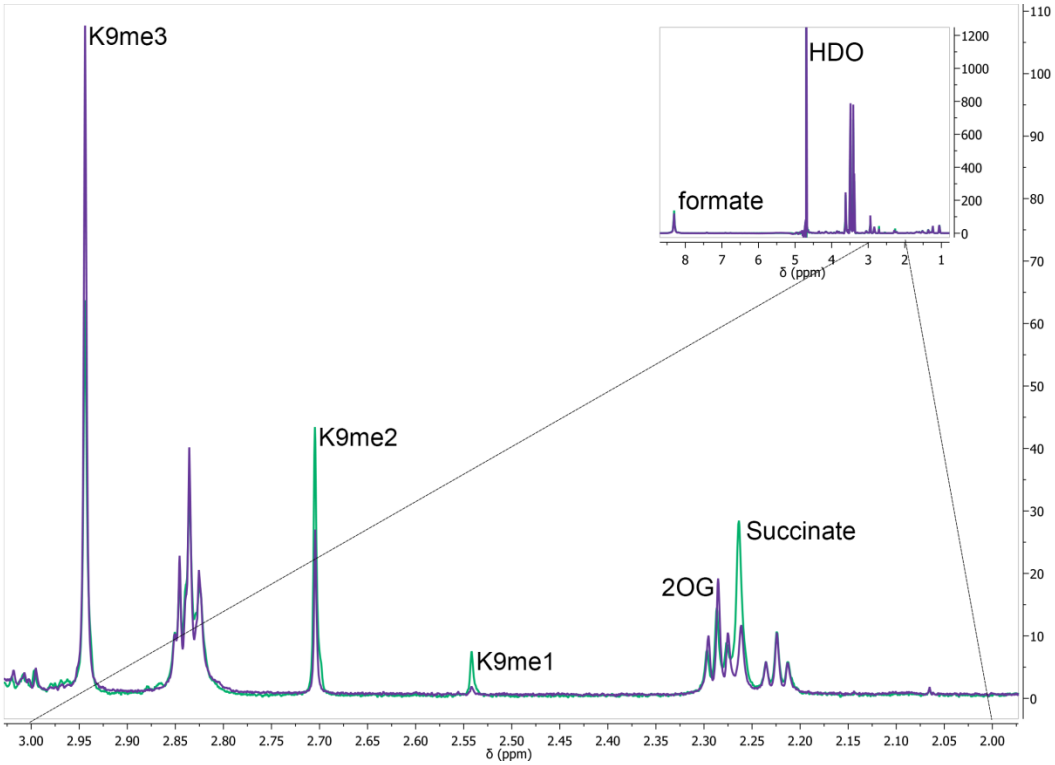


Figure 5.4.10: ^{1}H NMR of assay of H3$_{1-15}^{L}$K9me$_3$ 5.11 (392 µM) in the presence of KDM4E$_{JmjC}$ (11.5 µM) displaying the first acquisition (t= 6.1 min, purple) and the final acquisition (t= 38.2 min, green). The relevant peak signals from the mixture are annotate. Analysis of the data reveal that the first 6.1 min contained the highest initial rate but also showed activity during acquisition decreasing the K9me$_3$ and peak but showing an increase in K9me$_2$, K9me$_1$ and succinate peaks live ^{1}H NMR. This figure is reproduced with permission from the Royal Society of Chemistry.[342]

The corresponding time-course is visualized in figure 5.4.12 A. Consistent with RF-MS, the NMR analyses demonstrate that the concentration of H3$_{1-15}^{L}$K9me$_3$ **5.11** decreases over time while peaks correspond to H3$_{1-15}^{L}$K9me$_2$ **5.11B** and H3$_{1-15}^{L}$K9me$_1$ **5.11C** increase. Concomitant reduction in 2OG resonances and an increase in the succinate resonance was also observed. By calculating the required amount of succinate in relation to the amount of demethylated product, the amount of coupled and uncoupled 2OG can be estimated. While some level of uncoupled 2OG is present after the first acquisition, the level of uncoupled 2OG does not significantly change throughout the experiment (Fig. 5.4.12 A). Under the same conditions, H3$_{1-15}^{D}$K9me$_3$ **5.12** substrate was tested showing only an initial burst of demethylation with little formation of the demethylated product during the acquisition period (Fig. 5.4.11, Fig. 5.4.12 B). This confirms observations made using RF-MS that D-lysine is a

significantly worse substrate to L-lysine substrate stereoisomer. While coupled turnover was low during the time-course, uncoupled turnover increased continuously which suggests that the enzyme remains active during the acquisition time. The same experiment was performed visualizing the levels of uncoupled turnover in the absence of substrate showing some level of succinate formation over time (Fig. 5.4.12 D).

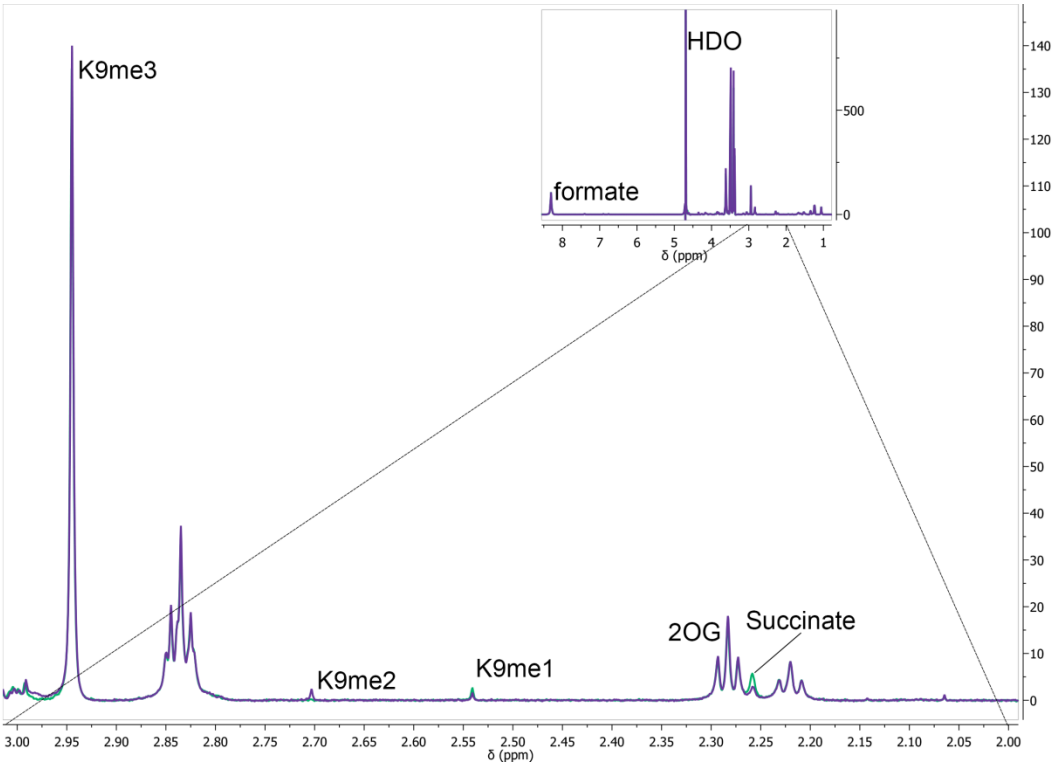


Figure 5.4.11: ^{1}H NMR analysis of H3$_{1\text{-}15}^{D}$K9me$_3$ 5.12 (392 µM) in the presence of KDM4E$_{JmjC}$ (11.5 µM) displaying the first acquisition (t= 5.3 min, purple) and the final acquisition (t= 37.2 min, green). The relevant peak signals from the mixture are annotate. Analysis of the mixture reveals only trace levels of peaks are formed that can be linked to *N*$^{\varepsilon}$-dimethyl- and *N*$^{\varepsilon}$-monomethyl-lysine. This figure is reproduced with permission from the Royal Society of Chemistry.[342]

The initial rate of succinate formation (first 5 acquisition points) is compared in figure 5.4.13, which shows uncoupled turnover in the absence of substrate, with an initial rate of 0.56 ± 0.09 µM·min^{-1}. In the presence of H3$_{1\text{-}15}^{D}$K9me$_3$ **5.12** a slight increase is turnover rate is observed 0.84 ± 0.24 µM·min^{-1} but not statistically significant suggesting that the peptide may only have a small effect (Fig. 5.4.13). Experiments using the natural H3$_{1\text{-}15}^{L}$K9me$_3$ **5.11** (392 µM) show good turnover and succinate formation (9.12 ± 0.14 µM·min^{-1}) with relatively low uncoupled turnover seen as total turnover (9.70 ± 0.42 µM·min^{-1}).

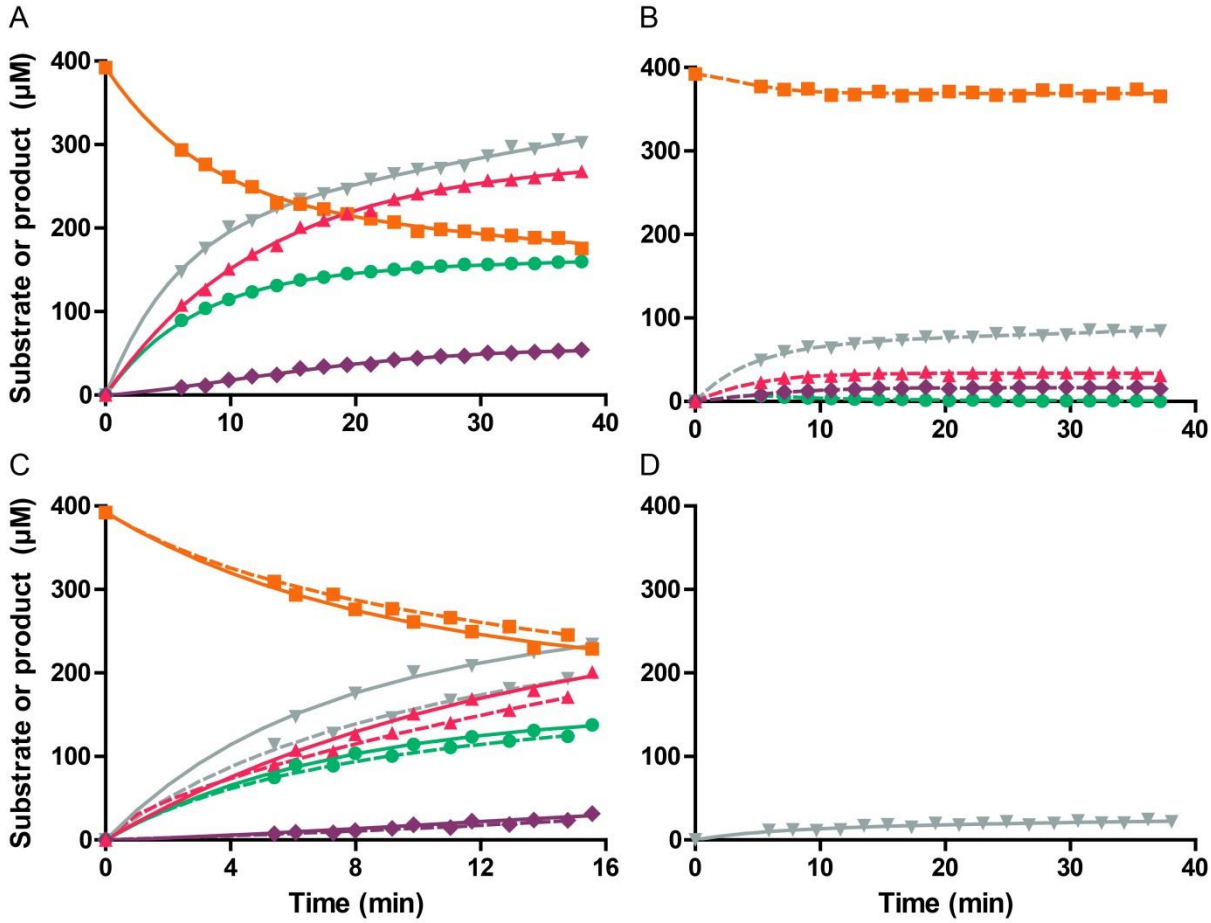


Figure 5.4.12: Time-course of KDM4E$_{JmjC}$ (11.5 μM) and substrate H3$_{1-15}^{L/D}$K4me$_3$ 5.11 / 5.12 (392 μM) monitoring the levels of succinate (800 μM) (grey), coupled succinate (pink), *N*$^{\varepsilon}$-trimethyl-lysine (orange), *N*$^{\varepsilon}$-dimethyl-lysine (green) and *N*$^{\varepsilon}$-monomethyl-lysine (purple) measured using ^{1}H NMR. A) catalytic reaction of KDM4E$_{JmjC}$ in the presence of H3$_{1-15}^{L}$K9me$_3$ **5.11** and 2OG; B) catalytic reaction of KDM4E$_{JmjC}$ in the presence of H3$_{1-15}^{D}$K9me$_3$ **5.12** and 2OG; C) Enzymatic reaction of KDM4E$_{JmjC}$ in a competition assay of H3$_{1-15}^{L}$K9me$_3$ **5.11** and H3$_{1-15}^{D}$K9me$_3$ **5.12** (dashed lines) by KDM4E$_{JmjC}$ in overlay with only H3$_{1-15}^{L}$K9me$_3$ **5.11** (also seen in A); D) uncoupled 2OG turn-over by KDM4E$_{JmjC}$. This figure is reproduced with permission from the Royal Society of Chemistry.[342]

In a competition assay with H3$_{1-15}^{D}$K9me$_3$ **5.12** (392 μM) and H3$_{1-15}^{L}$K9me$_3$ **5.11** (392 μM), the rate of H3$_{1-15}^{L}$K9me$_3$ **5.11** succinate formation was not significantly altered, suggesting relatively low levels of competition/inhibition by H3$_{1-15}^{D}$K9me$_3$ **5.12** (Fig. 5.4.12 C, Fig. 5.4.13) to KDM4E$_{JmjC}$.

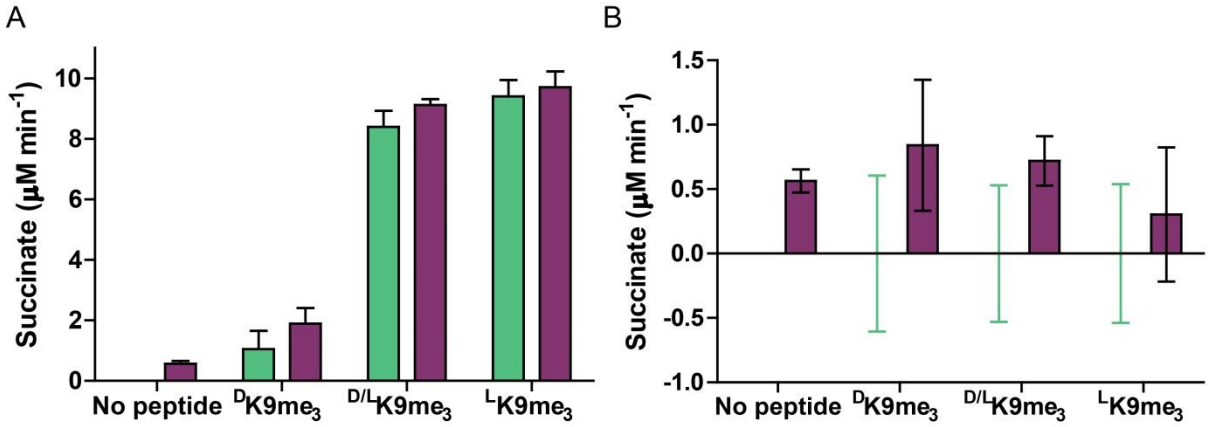


Figure 5.4.13: ^{1}H NMR assay estimation of the initial rate of succinate formation based on coupled (green) and uncoupled (purple) turnover by KDM4E$_{JmjC}$ (11.5 μM). No peptide refers to conditions in the absence of peptide, ^{D}K9me$_3$ to H3$_{1-15}$^{D}K9me$_3$ 5.12 (392 μM), $^{D/L}$K9me$_3$ to the competition assay of H3$_{1-15}$$^{L/D}$K9me$_3$ (392 μM) and ^{L}K9me$_3$ to the presence H3$_{1-15}$^{L}K9me$_3$ 5.11 (392 μM). A) Total initial rate, B) Relative contribution of uncoupled turn in total succinate formation. n=2–3, error given as s.d. This figure is reproduced with permission from the Royal Society of Chemistry.[342]

Collectively, the data show that no significant enhancement of uncoupled turnover detected using ^{1}H NMR depending on the presence or absence of H3$_{1-15}$^{L}K9me$_3$ **5.11** or H3$_{1-15}$^{D}K9me$_3$ **5.12**.

5.5 Discussion

The chromatin binding modules tested using ITC show that PHD-domain containing proteins ($KDM5A_{PHD}$, $BPTF_{PHD}$, $TAF3_{PHD}$) are able to bind both the natural L-N^{ε}-trimethyl-lysine and the D-N^{ε}-trimethyl-lysine stereoisomers. Although single L-lysine derivatives are observed to bind in the sub-micro-molar range, an approximate 10-fold decrease in binding strength is observed. Some level of selectivity within the reader domain for D-lysine binding was observed. While PHD containing proteins showed loss in binding strength of ~10-fold, TTD containing proteins ($KDM4A_{TTD}$, $SGF29_{TTD}$) were observed to decrease by ~30-fold suggesting that domain architecture may have some influence on the recognition level in L- or D-lysine 4.

In Chapter 3.5.3, we observed that replacement of K4 to alanine (K4A) in histone H3 mimic peptide reduces binding to $KDM5A_{PHD3}$ by at least 2 orders of magnitude. It is therefore reasonable to assume, the lysine residue is a major driving force to binding and thus complex formation. As the binding of the D-N^{ε}-trimethyl-lysine containing peptide is weaker, but not as weak as for $H3_{1-21}$K4A, the evidence suggests that the D-N^{ε}-trimethyl-lysine sidechain contributes to complex formation and is a valid binding replacement for L-N^{ε}-trimethyl-lysine. Similar observations are reported for $KDM5A_{PHD3}$, $BPTF_{PHD}$, $TAF3_{PHD}$, $KDM4A_{TTD}$ and $SGF29_{TTD}$ (same proteins used in this study), were K4 was replaced by G4 revealing in excess of 500-fold decrease in binding strength, supporting the idea that the D-N^{ε}-trimethyl-lysine side chain plays a major role in binding complex formation. However, other favourable interactions may be disturbed.

The molecular dynamics calculations show good alignment with the ITC data obtained. According to the computational results, L-lysine forms a stronger binding complex that is less perturbed during the 10 ns simulations than counterpart D-lysine, which showed greater variability in overall movement between the protein-peptide complexes, suggesting a weaker interaction. During the simulation time,

none of the peptide-protein complexes were found to dissociate, supporting the idea that D-lysine and the reader domains form a relatively strong peptide-protein complex.

Unlike chromatin binding modules, KMTs and KDMs were found to be strongly influenced by the change in C^{α}-lysine stereochemistry. KMT7 and KMT5A showed good activity on the L-lysine substrates; however, when D-lysine derivative peptides were used, only traces peaks that could be assigned to methyl lysine products were observed upon elevated enzyme concentration and higher concentration of SAM. More significant methylation was observed with, KMT1C and KMT1D when using prolonged incubation times and elevated protein conditions. This suggests that SET domains show some level of plasticity depending on the KMT investigated. Interestingly, the KDMs examined did not catalyse demethylation of the D-lysine substrate beyond trace levels (<5%). Increased levels of protein did not yield higher levels of demethylation.

The observation of trace levels of substrate conversion should be treated with caution. There is a possibility that low level (<5%) of L-lysine is introduced due to impure commercially obtained starting materials Fmoc-^{D}Lys(Boc)-H **5.18** reagents or introduced during synthetic chemistry or SPPS chemistry as a result of racemization. It is therefore reasonable to assume that D-N^{ε}-trimethyl-lysine is not a substrate for KDMs.

5.6 Conclusion

In conclusion, chromatin binding modules are found to be more flexible than KMTs or KDMs examined. While the binding strength for D-K4me$_3$ or L-K4me$_3$ peptides for PHD and TTD proteins are decreased by 10-fold or 32-fold, respectively, strong micro-molar affinity is maintained. This phenomenon is support by MD computational studies showing weaker affinity for the D-lysine relative to L-lysine but in all cases is the protein-peptide complex maintained throughout the simulation. KMTs KMT7 and KMT5A demonstrated that only trace of D-lysine substrate is methylated while L-lysine residues are fully converted to the corresponding methylated products. Interestingly, a higher proficiency of D-lysine residues methylation is observed with extended exposure times and higher enzyme conditions for KMT1D / KMT1C. However, combined D- and L-lysine substrates in a competition assay show the L-lysine derivative is superior to the D-lysine substrate. LC-MS experiments show that examined demethylases KDM1A$_{C1\text{-}corest}$, KDM4A$_{JmjC}$, KDM4D$_{JmjC}$, KDM4E$_{JmjC}$, KDM5B$_{JmjC}$ and KDM5C$_{JmjC}$ may be able to convert D-lysine substrate to traces of demethylated product. Additional ^{1}H NMR analysis confirms these observations for KDM4E$_{JmjC}$ and competition assays suggest that D-lysine has a poor binding affinity to the KDM4E$_{JmjC}$ protein suggesting minimal interaction at 100 μM. Overall, selectivity in plasticity is observed in PTM associated recognition of L/D-N^{ε}-methyl-lysine between chromatin binding modules relative to KDMs and KMTs.

5.7 Future perspective

The chromatin binding modules showed that for the PHDs, there was only ~10-fold decrease in the binding strength for the D-lysine residues relative to the L-lysine. Therefore, if D-N^{ε}-(methyl-) lysine residues are present on histone / non-histone substrates, it is likely that chromatin binding module can recognise this, whilst KTMs / KDMs that are less likely to react at these sites. Furthermore, these findings could have implications in inhibitor development where the change in stereochemistry could function as tool to create a selective small molecule or macromolecule (*e.g.* cyclic peptide) that exclusively targets the chromatin binding module while avoiding off-target effects to KMTs or KDMs.

5.8 Materials and Methods

5.8.1 ITC analysis of chromatin binding modules against histone H3 mimic peptides (carried out by Dr. Jasmin Mecinović)

Acquisitions were performed on ITC200 (GE Healthcare). Protein samples were prepared according to Table 5.8.1 and histone mimic peptide was titrated to the chromatin binding module containing cell. Samples were acquired 3–5 times. Curve fitting of the obtained data was achieved using Origin (Microcal Inc. v 6.0) in one mode binding configuration.

Table 5.8.1: Peptide, protein concentrations and corresponding buffer composition.

	Peptide	Protein	Buffer
$KDM5A_{PHD3}$	430 µM	33 µM	Tris (50 mM), NaCl (20 mM) in water (pH 7.5)
$BPTF_{PHD}$	630 µM	40 µM	Tris (50 mM), NaCl (20 mM) in water (pH 7.5)
$TAF3_{PHD}$	200 µM	20 µM	Tris (50 mM) in water (pH 7.5)
$KDM4A_{TTD}$	1400 µM	110 µM	Tris (50 mM) in water (pH 7.5)
$SGF29_{TTD}$	320 µM	30 µM	Tris (25 mM), NaCl (50 mM), DTT (1.0 mM) in water (pH 7.5)

5.8.2 Computational studies applying molecular dynamics (carried out by Dr. Kiran Kumar and Dr. Robert Paton)

All computational studies were carried out by Dr Kiran Kumar and Prof Robert Paton. Chromatin binding modules in this study were subjected to molecular dynamics based on PDB structures of $KDM5A_{PHD3}$ (PDB ID: 2KGI), $BPTF_{PHD}$ (PDB ID: 2F6J), $TAF3_{PHD}$ (PDB ID: 2K17), $KDM4A_{TTD}$ (PDB ID: 2GFA) and $SGF29_{TTD}$ (PDB ID: 3ME9) using the default $^{L}K4me_3$-peptide as binding partner. Substituent $^{D}K4me_3$ was generated from the $^{L}K4me_3$ template. Hydrogens were introduced using LEaP and protein force field partial charges was defined using Amberff12SB (AMBER12).[345]

$KDM4A_{TTD}$ coordination of non-covalent bond partner Zn(II) was adapted from previous study.[347,348] The proteins simulated were configured in a 10 Å octahedral cage and solvated in water (TIP3P)[346] and neutralization of the complex was achieved with counter-ions sodium and chloride. Minimization of steepest-descent and conjugated-gradient (10^3 cycles each) of the final models was achieved to avoid close van der Waals contact was done using option *sander* in AMBER12 software while $K4me_3$ partial charges were determined using Restrained Electrostatic Potential (RESP) charges.[349] The system was computed at 1 atm (isotropic molecule-based scaling) at times step of 2.0 fs and temperature of 310K was used applying PMEMD for heating the system. Furthermore, Particle Mesh Ewald summation[350] was applied as level of summation to compute the bond length constraints and long-distance electrostatic interactions for all hydrogen atoms added and a set dielectric constant of 1.0 calculated at intervals (20 ps) followed by visualization using VMD.[351] Cation pi systems were defined based on non-hydrogen containing aromatic residues, phenylalanine, tyrosine and tryptophan depending on the protein aromatic cage. N^{ε}-Lysine ammonium head group and aromatic cage electrostatic energies were determine using NAMDEnergy Plugin 1[351] for the duration of 10 ns with 20 ps intervals.

Table 5.8.2 MD simulations of modified K4me$_3$ residues of histone mimic peptide computed using RESP method (HF/6-31G*) displaying the Cartesian coordinates against RESP charge. This Table is reproduced with permission from the Royal Society of Chemistry.[342]

L-Kme3				RESP	D-Kme3				RESP
Atom	X	Y	Z	Charge	Atom	X	Y	Z	Charge
N	-3.252	1.483	0.199	-0.9943	N	-3.414	-0.900	1.164	-1.0164
C	-3.165	0.204	-0.472	0.4440	C	-3.321	-0.605	-0.250	0.4854
C	-1.866	-0.521	-0.080	-0.2969	C	-3.718	0.831	-0.528	0.3440
C	-0.621	0.257	-0.504	0.1166	O	-4.108	1.587	0.299	-0.4506
C	0.672	-0.381	0.025	0.0526	H	-4.013	-0.243	1.624	0.3815
C	1.868	0.461	-0.408	-0.2183	C	-1.898	-0.871	-0.778	-0.2915
N	3.234	-0.002	0.067	0.2357	H	-1.692	-1.930	-0.647	0.0611
C	-4.344	-0.715	-0.183	0.3660	H	-1.862	-0.677	-1.848	0.0611
O	-5.155	-0.486	0.652	-0.4634	C	-0.833	-0.049	-0.048	0.1206
C	3.545	-1.375	-0.436	-0.3779	H	-0.969	-0.187	1.018	0.0133
C	4.253	0.948	-0.477	-0.3779	H	-0.978	1.010	-0.251	0.0133
C	3.320	0.011	1.560	-0.3779	C	0.594	-0.447	-0.451	0.0194
H	-3.594	1.349	1.132	0.3760	H	0.736	-1.502	-0.244	0.0386
H	-3.923	2.072	-0.255	0.3760	H	0.724	-0.301	-1.518	0.0386
H	-3.147	0.375	-1.547	-0.0071	C	1.590	0.399	0.338	-0.2223
H	-1.850	-1.516	-0.521	0.0726	H	1.448	1.448	0.119	0.1366
H	-1.864	-0.657	1.000	0.0726	H	1.440	0.256	1.400	0.1366
H	-0.581	0.310	-1.590	0.0062	N	3.063	0.124	0.091	0.2428
H	-0.709	1.273	-0.139	0.0062	C	3.427	0.371	-1.337	-0.3832
H	0.756	-1.394	-0.353	0.0291	H	3.132	1.373	-1.607	0.1845
H	0.615	-0.439	1.107	0.0291	H	4.496	0.263	-1.446	0.1845
H	1.931	0.496	-1.487	0.1332	H	2.927	-0.344	-1.967	0.1845
H	1.761	1.474	-0.046	0.1332	C	3.852	1.059	0.954	-0.3832
H	-4.395	-1.630	-0.781	0.0133	H	4.905	0.881	0.797	0.1845
H	3.443	-1.388	-1.511	0.1835	H	3.609	2.076	0.687	0.1845
H	4.559	-1.623	-0.162	0.1835	H	3.601	0.882	1.988	0.1845
H	2.868	-2.086	0.008	0.1835	C	3.420	-1.280	0.460	-0.3832
H	4.208	0.936	-1.555	0.1835	H	3.118	-1.464	1.480	0.1845
H	5.235	0.641	-0.149	0.1835	H	2.921	-1.967	-0.202	0.1845
H	4.041	1.942	-0.112	0.1835	H	4.489	-1.402	0.367	0.1845
H	3.055	0.995	1.919	0.1835	H	-4.002	-1.209	-0.852	-0.0316
H	2.648	-0.724	1.969	0.1835	H	-3.625	1.142	-1.573	0.0272
H	4.333	-0.224	1.851	0.1835	H	-3.771	-1.820	1.325	0.3815

5.8.3 *In vitro* assays of Histone methyltransferases testing histone mimic peptides

Standard KMTs methylation assay (using KMT5A) (adapted from reported procedure).[352] **(carried out by Abbas Al Temimi)**

An assay mixture was prepared H4$_{13\text{-}27}$^{L}K20 (100 μM), *S*-(5'-Adenosyl)-L-methionine chloride dihydrochloride (200 μM) in buffer [Tris (50 mM) in MilliQ (pH 8.0)] and initiated with KMT5A (2.0 μM) in buffer (final volume: 100 μL). The mixture was incubated for 1 hrs at 37 °C followed by the quench of a sample (5.0 μL) with methanol (5.0 μL). The sample mixed with MALDI Matrix [HCCN

(10 mg·mL^{-1}) in MilliQ : MeCN (containing 0.1% TFA) (1:1)] mixture and spotted on MALDI-TOF ground steel target plate (MS 96) and air dried. Examination of the samples was measured in positive reflectron mode and spectra were analysed using FlexAnalysis software.

The conditions may change depending on the KMT tested and the enzyme concentration. The details are displayed in Table 5.8.3.

Table 5.8.3: Overview of MALDI-TOF assay conditions for KMTs.

	Protein conc	Buffer	SAM	Substrate
SETD7 SETD8	2.0 µM, 10.0 µM	Tris (50 mM) in MilliQ (pH 8.0)	0.2, 1.0 mM	$H3_{1-15}{}^{L}K4$ or $H3_{1-15}{}^{D}K4$ (100 µM) $H4_{13-27}{}^{L}K20$ or $H4_{13-27}{}^{D}K20$ (100 µM)
GLP G9a			0.5, 1.0 mM	$H3_{1-15}{}^{L}K9$ or $H3_{1-15}{}^{D}K9$ (100 µM)

5.8.4 *In vitro* assays of lysine specific demethylases examining histone mimic peptide using RF-MS and ^{1}H NMR

^{1}H NMR analysis of peptide concentration

Prior to executing the RF-MS and ^{1}H NMR assays, the concentration of peptides was determined using AVIII 700 mHz equipped with helium cryoprobe. Peptide samples were prepared in D_2O containing internal standard sodium 3-(trimethylsilane)[2,2,3,3-$^{2}H_4$]-propionate (TSP) (10 mg·mL^{-1}) (75 µL) and measured with extended t=0 delay (5 sec). The concentration was calculated based on the obtained peak integrals of $Kme_{3/2}$ relative to the trimethyl group of TSP.

5.8.5 RF-MS instrument configuration

Demethylation of histone mimic peptides by KDMs was assessed using Agilent RapidFire 360 connected to Agilent QTOF 6530 set in MRM mode. Retention of the peptide of interested was accomplished on SPE cartridges C_4 or C_{18}, equilibration and washing was achieved with {Formic acid in MilliQ (0.1%)} and elution with [Acetonitrile (85%) + {FA in MilliQ (0.1%)} (15%)]. Acquisition cycle consisted of:

1. Aspiration (400 ms, 1.5 $mL \cdot min^{-1}$) through a loop (10 µL)
2. Washing (4500 ms, 1.5 $mL \cdot min^{-1}$)
3. Elution (4500 ms, 1.25 $mL \cdot min^{-1}$)
4. Equilibration (500 ms, 1.25 $mL \cdot min^{-1}$)
5. Repeat of step 1–4, aspirating water
6. Repeat of step 1–4 aspirating acetronitrile
7. Repeat of step 5,6

After each time stamped measurement was the obtained data analysed using Agilent Masshunter Qualitative Analysis (B.06.00), following by data separated using Mathworks® Matlab R2015 and processed using MicroSoft Excel™ and Prism© GraphPad. Relative demethylation was determined based on peak area detected from Z = 2–7 and standardized to known peptide concentration.

Standard time-course assay of KDM demethylation by RF-MS in MRM mode

A peptide mixture was prepared: [$H3_{1\text{-}15}{}^{L}K9me_3$ **5.11** (12 µM), sodium ascorbate (120 µM), $(NH_4)_2Fe(II)(SO_4)_2$ (12 µM), disodium 2-oxoglutarate (36 µM) in buffer {MES (50 mM) in MilliQ (pH 7.0)}] (550 µL) in 96-well block. RF-MS acquired t= 0 sec and enzyme mixture [$KDM4E_{JmjC}$ (1.2 µM) in buffer] (100 µL) was added, mixed using pipette and reacted while RF-MS performed acquisitions in MRM mode time stamping each acquisition. The reaction mixture was measured every

6–7 min for the duration of 60 to 70 min. Details regarding the mixture composition for each KDM tested can be found in Table 5.8.4.

Table 5.8.4: mixture composition of KDMs tested. This Table is reproduced with permission from the Royal Society of Chemistry.[342]

KDMs		Substrate	Buffer	Ascorbate	$(NH_4)_2Fe(II)(SO_4)_2$	2OG	TCEP
$KDM1A_{FL\text{-}Corest}$ (0.2, 1.0 μM)	LSD1	$H3_{1\text{-}21}{}^{L/D}K4me_2$ **5.9, 5.10** (10 μM)	TRIS (50 mM) in MilliQ (pH 8.0)	N.A.	N.A.	N.A.	N.A.
$KDM4A_{JmjC}$ (0.2, 1.0 μM)	JMJD2A	$H3_{1\text{-}15}{}^{L/D}K9me_3$ **5.11, 5.12** (10 μM)	MES (50 mM) in MilliQ (pH 7.0)	100 μM	10 μM	30 μM	N.A.
$KDM4D_{JmjC}$ (0.2, 1.0 μM)	JMJD2D	$H3_{1\text{-}15}{}^{L/D}K9me_3$ **5.11, 5.12** (10 μM)	MES (50 mM) in MilliQ (pH 7.0)	100 μM	10 μM	30 μM	N.A.
$KDM4E_{JmjC}$ (0.2, 1.0 μM)	JMJD2E	$H3_{1\text{-}15}{}^{L/D}K9me_3$ **5.11, 5.12** (10 μM)	MES (50 mM) in MilliQ (pH 7.0)	100 μM	10 μM	30 μM	N.A.
$KDM5B_{JmjC}$ (0.2, 1.0 μM)	JARID1B	$H3_{1\text{-}15}{}^{L/D}K9me_3$ **5.13, 5.14** (5.0 μM)	MES (50 mM), NaCl (50 mM) in MilliQ (pH 7.0)	100 μM	10 μM	30 μM	5.0 mM
$KDM5C_{JmjC}$ (0.2, 1.0 μM)	JARID1C	$H3_{1\text{-}15}{}^{L/D}K9me_3$ **5.13, 5.14** (5.0 μM)	MES (50 mM), NaCl (50 mM) in MilliQ (pH 7.0)	100 μM	10 μM	30 μM	5.0 mM

^{1}H NMR time-course assay of $KDM4E_{JmjC}$ tested against histone mimic peptides

A mixture was prepared starting from buffer [ammonium formate-$^2H_5$ (AFN) (50 mM) in MilliQ (pH 7.5), 38.0 μL] followed by sodium ascorbate (Asc) (10 mM, 7.5 μL) in buffer, $(NH_4)_2Fe(II)(SO_4)_2$ (1.0 mM) in D_2O (7.5 μL), disodium 2-oxoglutarate (4 mM) in buffer (15 μL), $H3_{1\text{-}15}{}^{L}K9me_3$ **5.11** (4.9 μM) (6 μL) and the enzymatic reaction was started by the addition of [$KDM4E_{JmjC}$ (864 μM) in buffer {HEPES (50 mM), NaCl (500 mM) in MilliQ (pH 7.5)} (1.0 μL)]. The mixture was carefully mixed and transferred into an NMR tube (75 μL, 2 mm (Bruker, cat #:67539) and inserted into AVIII 700 mHz NMR. ^{1}H NMR measurements were taken in consecutive order every ~113 sec for the approximate duration of 30 min at 25 °C using method excitation sculpting suppression baseline optimisation.

Acquired data was analysed using MestReNova (v 11.0.2) and concentration of $K9me_3$, $K9me_2$ and $K9me_1$, 2OG and succinate was determined based on the peak integral taking in account potential changes in protons composition corresponding to the observed signal. Uncoupled turnover is computed from the relative peak integrals of 2OG and succinate. Coupled turnover is estimate based on the observed concentration of $K9me_2$ and $K9me_1$, registering $K9me_1$ as double turnover. In estimating the rate over turnover, the first 5 acquisition points are averaged and the error is derived from the standard

deviation. Peptides $H3_{1-15}{}^{L}K9me_3$ 5.11 (A) and $H3_{1-15}{}^{D}K9me_3$ 5.12 (B) in TSP and AFN are displayed in figure 5.8.1.

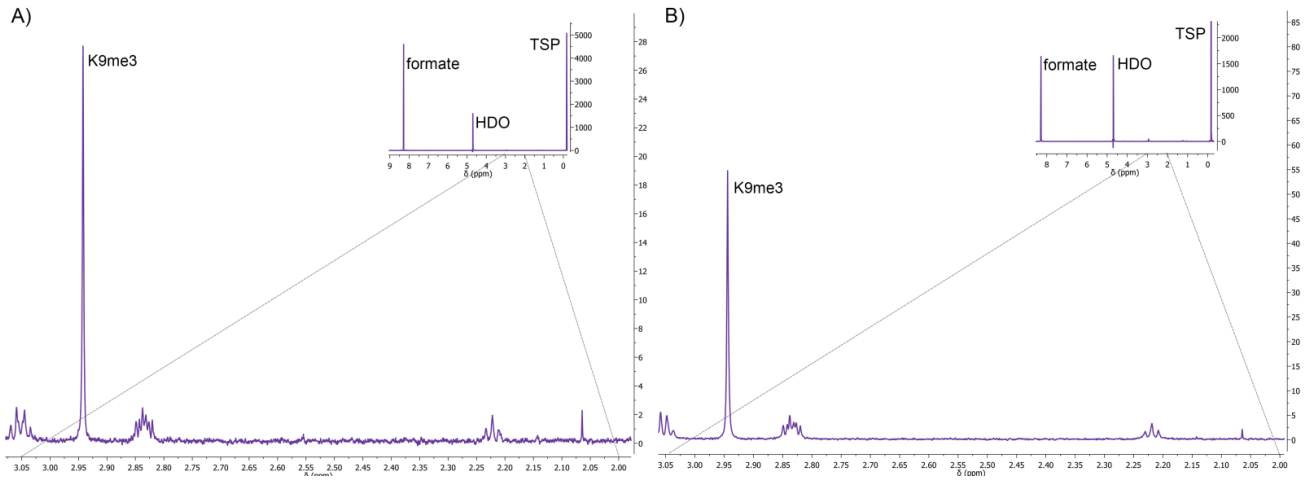


Figure 5.8.1: ^{1}H NMR spectra of $H3_{1-15}{}^{L}K9me_3$ 5.11 (A) and $H3_{1-15}{}^{D}K9me_3$ 5.12 (B) and TSP in AFN buffer. This figure is reproduced with permission from the Royal Society of Chemistry.[342]

5.8.6 Solid Phase Peptide Synthesis

SPPS for chromatin binding modules and KDMs

Peptides were synthesised from C- to N-terminus on Rink amide resin (Breiphol) on polystyrene support (0.1 mmol scale, 0.66 mmol·g^{-1}) applying Fmoc-chemistry, type chemistry using microwave assisted automated peptide synthesizer Liberty Blue (CEM). Commercially available amino acids building blocks were purchased as Fmoc-protected, and if required, provided with acid labile side chain protection groups. Couplings of Fmoc-amino acids (0.2 mmol, 0.2 M, 2.5 mL, 5 equiv.) were performed in the presence of 1-hydroxybenzotriazole (HObt) (0.5 mmol, 0.25 M, 2.0 mL, 5 equiv.) and *N,N'*-diisopropylcarbodiimide (DIC) (0.5 mmol, 1 M, 0.5 mL, 5.0 equiv.) in *N,N*-dimethylformamide (DMF) for 10 min at 75 °C. Arginine was found to be unstable under these conditions and was double coupled for 2 hrs at room temperature. Fmoc-Lys(me_3·Cl)-OH and Fmoc-Lys(me_2·HCl)-OH (0.3 mmol, 0.12 M, 2.5 mL, 3.0 equiv.) were double coupled using 1-[bis(dimethylamino)methylele]-1 H-1,2,3triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU) (0.4 mmol, 0.2 M, 2.0 mL, 4.0 equiv.), *N,N*-diisopropylethylamine (DIEA) in N-methyl

pyrrolidine (NMP) (0.5 mmol, 1 M, 0.5 mL, 5.0 equiv.) and coupled for 2 min at 95 °C. Washing of the resin was achieved with DMF and deprotection of the Fmoc was carried out with piperidine in DMF (20 v/v%) in two rounds with fresh piperidine in DMF every round for 30 sec at 75 °C, followed by 3.5 min at 89 °C. Deprotection is monitored by UV to estimate the level of deprotected Fmoc-groups. If the signal was high, another 3.5 min at 89 °C was carried out. After coupling completion, the resin was transferred to the HT loader and filtered. The resin was thoroughly washed with 3× DMF (~5 mL), 4× diethylether (~5 mL) and dried upon air. Acid labile protecting groups were removed using a deprotection mixture [1,2-ethandithiol (EDT) (2.5%), triisopropylsilane (TIPS) (2.5%), MilliQ (2.5%) in trifluoroacetic acid (TFA) (92.5%)] (~2.5 mL) for the duration of 4 hrs. The mixture was filtrated through a syringe filter and content of supernatant was precipitate in cold diethylether (-26 °C). The solid was pelleted using centrifugation (4255 × g, 3 min, 4.0 °C) and the liquid was decanted off and the solid was air dried (~1 hrs). The solid was suspended in MilliQ (~3 mL) and lyophilized to give the crude product. Crude peptide was suspended in MilliQ, filtrated (0.45 µm) and purified using HPLC Dowex Ultimate 3000 equipped with R_f-C_{18} column Grace Vydac 218 TP101522. Solvent A [TFA (0.1 v/v%) in MilliQ] and B [TFA (0.1 v/v%) in Acetonitrile]. Program: 2%–20% B over 35 min (15 mL·min^{-1}). Mass confirmation was achieved by sample spotting (1.0 µL) and MALDI matrix [α-cyano-4-hydroxycinnamic acid (sat. 10 mg·mL^{-1}) in acetonitrile/water (1:1, containing 0.1% trifluoroacetic acid)] (1.0 µL) on a Waters target plate. The spots were air dried, measured and acquired data were analysed using MassLynx (v. 4.00.00). Based on MALDI-TOF the fractions were selected, pooled and lyophilized. Purity of the peptide was determined using Agilent LC-MS (1200 series) coupled to an Agilent MS 6120 Quadrupole equipped with Grace Vydac 218TP C_{18} 5µ with solvent A [TFA (0.1 v/v%) in MilliQ] and B [TFA (0.1 v/v%) in acetonitrile]. Column program: 1%B for min 0–1, 1%–20% for min 1–8, 20%–100% for min 8.0–8.5, 100% for min 8.5–9.5, 100%-1% for min 9.5–10, 1% for min 10–12 at 1.5 mL·min^{-1} monitoring ions in positive ion mode and UV (210 nm). Purity was calculated based on the peak area of the UV spectrum detected.

SPPS for KMTs

Histone mimic peptides were produced in similar fashion as to chromatin binding module and KDM peptides. Histone H3 mimic peptides were made using Fmoc-chemistry on Rink resin (Breiphol) to provide C-terminal amide while histone H4 mimic peptides were synthesized from Fmoc-Ala-wang resin to give a C-terminal carboxylic acid. Amino acids with reactive functional groups were protected with acid labile groups namely: Fmoc-Arg(Pbf)-OH, Fmoc-Thr(^{t}Bu)-OH, Fmoc-$^{D/L}$Lys(Boc), Fmoc-Gln(Trt)-OH, Fmoc-Ser(^{t}Bu)-OH, Fmoc-His(Trt)-OH, Fmoc-Asp(^{t}Bu)-OH and Fmoc-Asn(Trt)-OH. Deprotecting of the Fmoc group was performed using Piperidine in DMF (20 v/v%) for 30 min at room temperature. The Fmoc protected amino acids (3.0 equiv.) were coupled with N,N'-diisopropylaminecarbodiimide (DIC) (1 M in DMF, 3.3 equiv.) and 1-hydroxybenzotriazole (HObt) (1 M in DMF, 3.6 equiv.) until the Kaiser test indicated reaction completion. After completion of peptide synthesis, the resin was washed with DMF and diethylether followed by deprotection of acid labile protecting group and cleavage of the solid support using a mixture [TIPS (2.5%), water (2.5%) and TFA (95%)] for the duration of 4 hrs. Solids were removed and the crude peptide was precipitated in cold diethylether. The precipitate was pelleted using centrifugation and liquid was removed be decantation. Solids were suspended in water (~3 mL) and freeze-dried overnight on ilShin Freeze Dryer (ilShin). The crude was purified using Shimadzu preparative HPLC equipped with Phenomenex® Gemini-NX 3U C18 110A (dimensions: 150 × 21.2 mm) column that is stabilized at 30 °C. Solvent A: [TFA (0.1 v/v%) in water], Solvent B: [TFA (0.1 v/v%) in acetonitrile], Flowrate: 10 mL·min^{-1}. Program: 3% B for 0–3 min, 3%–15% B for 3–12 min, 15%–30% B for 12.0–17.0 min, 30-100% B for 17–19 min, 100% B for 19–21 min, 100%–3% B for 21–30 min. Selected fraction were pooled and lyophilized to obtain the pure solid peptide and purity was determined using HPLC Shimadzu LC10A HPLC system containing the column: Phenomenex® Prodigy ODS3 5 µm partical size and 110Å pore size (dimensions: 150 × 4.60 mm). Solvent A: [TFA (0.1 v/v%) in water], Solvent B: [TFA (0.1 v/v%) in acetonitrile], Flowrate: 10 mL·min^{-1}. Program: 5% B for min 0–1, 5%–100% B for min 1–30, 100% B for min 30.0–35.0, 100%–5% B for 35–49 min monitoring 214 nm. Mass

confirmation was achieved with MALDI-TOF, ESI-MS Thermofisher Finnigan LCQ Advantage Max positive mode. LC-MS Thermofisher Thermo Finigan LCQ-Fleet positive mode with column connected Phenomenex® Gemini NX C18, particle size: 3.0 µm (dimensions: 50 × 2.0 mm). Solvent A: [formic acid (0.1 v/v%) in water], Solvent B: [formic acid (0.1 v/v%) in acetonitrile] flowrate: 0.2 $mL \cdot min^{-1}$. Program: 5% B for 0–1 min, 5%–100% B for 1–50 min.

5.8.7 Protein production of chromatin binding modules, KMTs and KDMs

Expression of Chromatin binding modules

The proteins $KDM5A_{PHD3}$, $BPTF_{PHD}$, $TAF3_{PHD}$, $KDM4A_{TTD}$ and $SGF29_{TTD}$ are produced according to previously reported procedures.[231]

Expression of KMT5A (KMT5A)

The plasmid encoding for the encoding for the KMT5A was transformed into *E. coli* Rosetta BL21 (DE3) PlysS cell line and the protocol was adapted from literature.[352] From a starter culture, the cells were grown (1 L Terrific broth media with antibiotic kanamycin) until after 2.5 hrs an OD_{600}: 0.6 is reached at 37 °C. The culture was induced with IPTG (1.0 mM) and cooled to 16 °C while shaking overnight. The cells were harvested using centrifugation and the cell pellet was suspended in lysis buffer [β-mercaptoethanol (5 mM), Imidazole (20 mM), Tris (20 mM), NaCl (500 mM), Glycerol (5%) in water (pH 7.8)]. The mixture was lysed by sonication and centrifuged to give the supernatant. The supernatant was incubated with Nickel-NTA beads for the duration of 2.5 hrs at 4 °C, washed with lysis buffer and then the protein was eluted with elution buffer [β-mercaptoethanol (5 mM), Imidazole (250 mM), Tris (20 mM), NaCl (500 mM), Glycerol (5%) in water (pH 7.8)]. The correct fractions were pooled and concentrated in spin concentrator (Amicon® Ultra 3.5K MWCO) using centrifugation and the protein was purified using gel filtration (Superdex 75 (GE Healthcare)) and gel filtration buffer [β-mercaptoethanol (10 mM), Tris (20 mM), NaCl (100 mM), in water (pH 7.5)] at flowrate: 0.5

$mL \cdot min^{-1}$. Fractions were pooled, concentrated, snap frozen (Liquid nitrogen) and stored (−80 °C) based on the mass of the fractions observed and purity visualized with SDS-page.

Expression of KMT7 (SETD7)

KMT7 was produced according to KMT5A protocol with minor alterations from previous reports.[353] IPTG: 0.4 mM, lysis buffer: PBS containing β-mercaptoethanol (2 mM) Imidazole (5 mM), NaCl (250 mM), glycerol (5 v/v%) in water; wash buffer: Imidazole (50 mM), Tris (20 mM), NaCl (250 mM), glycerol (5 v/v%) in water (pH: 8.0); elution buffer: Imidazole (250 mM), Tris (20 mM), NaCl (250 mM), glycerol (5 v/v%) in water (pH: 8.0); Gel filtration buffer: Tris (20 mM), NaCl (20 mM) in water (pH: 7.5).

Expression of KMT1C (G9a)

The plasmid encoding for KMT1C (913–1193) in vector pET28a-LIC was used to produce KMT1C protein following the protocol of KMT5A with small changes in buffer composition according to published study.[354] Induced: IPTG (1.0 mM) and Zinc(II)Chloride (0.1 mM); Lysis buffer: PBS containing β-mercaptoethanol (2 mM), NaCl (25 mM), glycerol (5 v/v%), Trixton X-100 (0.1%) in water (pH: 7.4) containing EDTA-free protease inhibitors; Wash buffer: Imidazole (50 mM), Tris (20 mM), NaCl (250 mM), glycerol (5 v/v%) in water (pH: 8.0); Elution buffer: Imidazole (250 mM), Tris (20 mM), NaCl (250 mM), glycerol (5 v/v%) in water (pH: 8.0); Gel filtration buffer: Tris (20 mM), NaCl (150 mM) in water (pH: 8.0).

Expression of KMT1D (GLP)

KMT1D protein was produced in similar fashion as to KMT5A with alterations in buffer conditions adapted from an earlier investigation.[354] The plasmid encoding for KMT1D (951–1235) in vector pET28a-LIC was used. Induced: IPTG (0.1 mM) and Zinc(II)Chloride (0.1 mM); Lysis buffer: PBS containing β-mercaptoethanol (2 mM), NaCl (25 mM), glycerol (5 v/v%), Trixton X-100 (0.1%) in

water (pH: 7.4) containing EDTA-free protease inhibitors; Wash buffer: Imidazole (50 mM), Tris (20 mM), NaCl (250 mM), glycerol (5 v/v%) in water (pH: 8.0); Elution buffer: Imidazole (250 mM), Tris (20 mM), NaCl (250 mM), glycerol (5 v/v%) in water (pH: 8.0); Gel filtration buffer: Tris (20 mM), NaCl (150 mM) in water (pH: 8.0).

Expression of KDMs

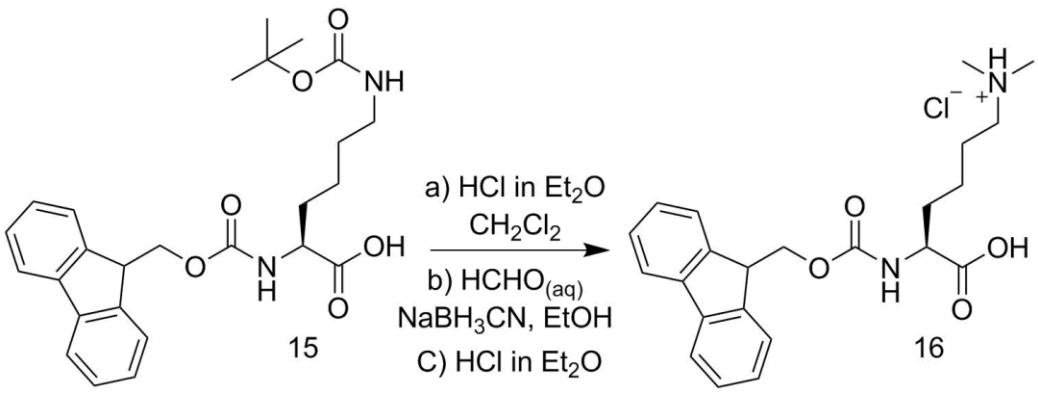
$KDM5B_{JmjC}^{292}$, $KDM5C_{JmjC}^{306}$, $KDM4D_{JmjC}$, $KDM4E_{JmjC}^{187}$ and $KDM4A_{JmjC}^{282}$ were produced from previously described literature reports.

Expression of $KDM1A_{C1\text{-}Corest}$

Plasmid containing vector pET11b with incorporated gene encoding for KDM1A (E169–A852) and Plasmid containing vector pET15b with incorporated gene encoding for CoREST (Gratefully received from H. Greschik and R. Schüle, University of Freiburg Medical centre, Germany) were transformed into *E. coli* Rosetta BL21 (DE3) and a starter cultures were made. Terrific broth media supplemented with FAD (2 $mg \cdot L^{-1}$), was independently inoculated with starter culture media and shaken at 37 °C until the right optical density was observed followed by the induction of IPTG (0.5 mM) and the cells were grown overnight at 18 °C. The cells were collected after centrifugation and the cell pellet was suspended in lysis buffer [Tris (20 mM), KCl (100 mM), $MgCl_2$ (25 mM), glycerol (10 v/v%) in water (pH 8.0)] and sonicated. The supernatant was obtained after centrifugation and KDM1A and CoREST were individually purified using Nickel chelation affinity-tag purification. KDM1A and CoREST were mixed in stoichiometric amounts to give the homodimer and purified as complex on Sephadex S200 (GE Healthcare) gel filtration column using lysis buffer. Obtained fractions indicating the content of KDM1A-CoREST complex were concentrated (10 $mg \cdot mL^{-1}$) in Spin concentrator 30 kDa, aliquoted to small samples (100 µL), snap frozen and stored (−80 °C).

5.8.8 Synthetis of N^{α}-Fmoc-N^{ε}-trimethyl-lysine analogues

Details about the instrumentation and standard conditions are described in materials methods, general experimental conditions.

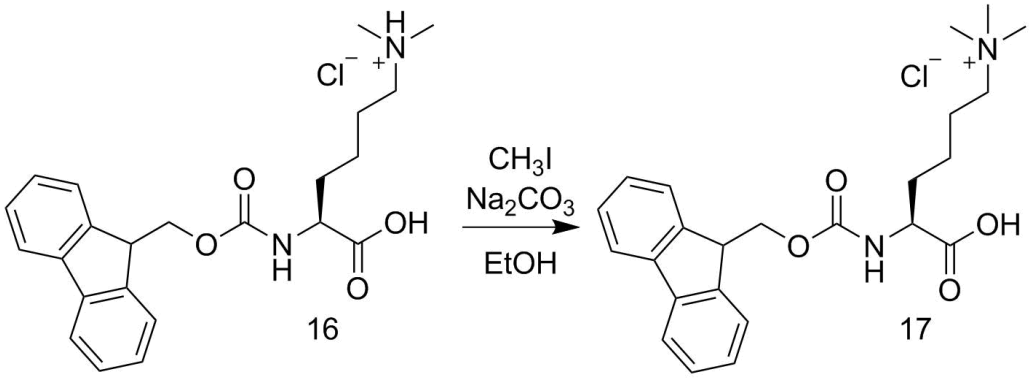


Fmoc-^{L}Lys(me$_2$·HCl)-OH

(S)-2-[9-Fluoronyloxycarbonyl]amino-6-[dimethyl]amino-hexanoic acid hydrochloric acid 5.16

Fmoc-^{L}Lys(Boc)-OH (469 mg, 1.0 mmol, 1.0 equiv.) was dissolved in dichloromethane (7.5 mL) and HCl in diethyl ether (2.5 mL, 5.0 mmol, 2 M, 5.0 equiv.) while mechanically stirred at room temperature. The mixture was reacted for the duration of 16 hrs followed by the removal of volatiles *in vacuo*. The crude was mass confirmed (ESI-MS) and dissolved in ethanol followed by the dropwise addition of formaldehyde (0.447 mL, 6.0 mmol, 37% sol., 6.0 equiv.). Portions sodium cyanoborohydride (138 mg, 2.2 mmol, 2.2 equiv.) was added after 0.75 hrs of stirring, TFA (0.320 mL) and methanol (10 mL) was added drop wise followed stirring for 0.25 hrs. The crude was concentrated under reduced pressure and purified using R_f-C_{18} (Grace 40g). Solvent A: [TFA (0.1 v/v%) in MilliQ], Solvent B: [TFA (0.1 v/v%) in acetonitrile]. Program: 5% B for 2 CV, 5–100% B over 20 CV. Fractions containing the product were identified based on monitoring wave length (254 nm) and low resolution mass, pooled and lyophilized furnishing 324 mg (0.82 mmol, 82%) white promiscuous product. The product was dissolved in dichloromethane (10 mL) and HCl in diethyl ether (1.0 mL, 2.0 mmol, 2 M, 1.68 equiv.) was added drop wise. Volatiles were evacuated *in vacuo*, diethyl ether was azeotropic removed with dichloromethane (4× ~10 mL) to receive a white hydroscopic solid.

mp: 67 °C (decomp.); IR υ_{max}: 3405, 2951, 1695, 1202, 1132 cm^{-1}; $[\alpha]^{25}_D$ +15 (c = 0.010 g·mL^{-1} MeOH); ^{1}H NMR 700 MHz, Solvent: DMSO-d_6, δ (ppm): 12.63 (s, 1 H, $^{\alpha}$CHCOO*H*), 10.31 (s, 1 H, N*H*$(CH_3)_2$), 7.90 (d, 2 H, *J*= 7.5 Hz, Fmoc-C_4*H*), 7.73 (t, 2 H, *J*= 7.0 Hz, 2× Fmoc-C_1*H*), 7.67 (d, 1 H, *J*= 8.0 Hz, N*H*$^{\alpha}$CHCOO), 7.42 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-C_3*H*), 7.33 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-C_2*H*), 4.32-4.26 (m, 2 H, OC*H*$_2$CH-Fmoc), 4.22 (t, 1 H, *J*= 7.0 Hz, OCH_2C*H*-Fmoc), 3.94 (dt, 1 H, *J*= 4.5, 9.0 Hz, NH$^{\alpha}$C*H*COO), 3.04-2.94 (m, 2 H, $^{\delta}CH_2{}^{\varepsilon}C$*H*$_2NH(CH_3)_2$), 2.71 (d, 6 H, *J*= 5.0 Hz, $^{\gamma}CH_2NH(C$*H*$_3)_2$), 1.77-1.70 (m, 1 H, $^{\alpha}CH^{\beta}CH$*H*$^{\gamma}CH_2$), 1.70-1.58 (m, 3 H, $^{\alpha}CH^{\beta}CH$*H*$^{\gamma}CH_2{}^{\delta}C$*H*$_2{}^{\varepsilon}CH_2$), 1.41-1.29 (m, 2 H, $^{\beta}CH_2{}^{\gamma}C$*H*$_2{}^{\delta}CH_2$). ^{13}C NMR 101 MHz, Solvent: MeOD-d_4, δ (ppm): 178.2, 158.3, 145.1, 142.5, 128.8, 128.2, 126.2, 120.9, 67.7, 58.6, 56.5, 48.4, 43.2, 32.9, 25.0, 23.7; ESI-MS (+) in MeOH as $[M+H]^+$ calc: 397.2 found: 397.2; HR-MS ESI (+) in MeOH as $[M+H]^+$ ($C_{23}H_{29}N_2O_4^+$) calc: 397.2122 found: 397.2114. The obtained spectroscopic values are in agreement with earlier literature studies.[355]

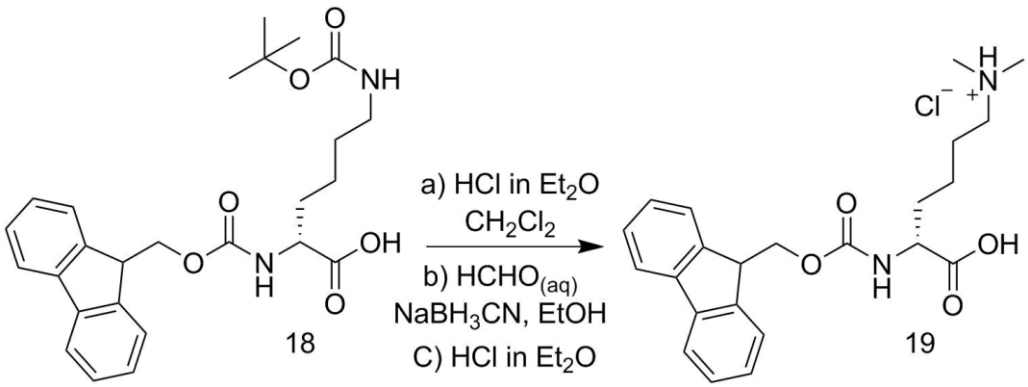


Fmoc-^{L}Lys(me$_3$·Cl)-OH

(S)-2-[9-Fluoronyloxycarbonyl]amino-6-[trimethyl]amino-hexanoate hydrochloride 5.17

To solution of Fmoc-^{L}Lys(me$_2$·HCl)-OH (514 mg, 1.19 mml, 1.0 equiv.) in ethanol (30 mL), iodomethane (1.50 mL, 24.0 mmol, 20.0 equiv.) was added dropwise followed by sodium carbonate (206 mg, 2.51 mmol, 2.1 equiv.). The mixture was mechanically stirred for 48 hrs monitoring reaction completion using ESI-MS. Additional iodomethane (1.00 mL, 16.0 mmol, 13.4 equiv.) was supplemented and after an additional 24 hrs of stirring at room temperature, volatiles were removed

under reduced pressure and the crude was purified using R_f-C_{18} chromatography. Grace 40 column, Solvent A: [TFA (0.1 v/v%) in MilliQ], Solvent B: [TFA (0.1%) in acetonitrile]. Program: 5% B for 2 CV, 5%–100% B over 24 CV. Fractions based on UV absorbance (254) and ESI-MS were combined and freeze-dried to give a white promiscuous solid producing 352 mg (0.72 mmol) of product (61%). The solid was dissolved in dichloromethane (2 mL) and HCl in diethyl ether (1.0 mL, 2.0 mmol, 2 M, 2.78 equiv.) was added, stirred for 5 min and *in vacuo* a hydroscopic white solid was obtained. Additional traces of diethyl ether were removed using azeotropic removal with dichloromethane (3× 10 mL). ^{1}H NMR 400 mHz, Solvent: DMSO-d_6, δ (ppm): 12.75 (bs, 1 H, NH$^{\alpha}$CHCOO*H*), 7.92 (d, 2 H, *J*= 7.5 Hz, 2× Fmoc-C_4*H*), 7.77-7.72 (m, 2 H, 2× Fmoc-C_1*H*), 7.71 (d, 1 H, *J*= 8.0 Hz, N*H*$^{\alpha}$CHCOO), 7.44 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-C_3*H*), 7.35 (apparent dt, 2 H, *J*= 1.0, 7.5 Hz, 2× Fmoc-C_2*H*), 4.38-4.21 (m, 3 H, OC*H*$_2$CH-Fmoc, OCH$_2$C*H*-Fmoc), 3.98 (dt, 1 H, *J*= 4.5 Hz, 9 Hz, NH$^{\alpha}$C*H*COOH), 3.35-3.20 (m, 2 H, $^{\delta}$CH$_2$$^{\varepsilon}$C*H*$_2$N(CH$_3$)$_3$), 3.04 (s, 9 H, $^{\varepsilon}$CH$_2$NH(C*H*$_3$)$_3$), 1.84-1.56 (m, 4 H, $^{\alpha}$CH$^{\beta}$C*H*$_2$$^{\gamma}CH_2$$^{\delta}$C*H*$_2$$^{\varepsilon}CH_2$), 1.42-1.28 (m, 2 H, $^{\beta}$CH$_2$$^{\gamma}$C*H*$_2$$^{\delta}CH_2$). ESI-MS (+) in MeOH as $[M+H]^+$ calc.: 411.2, found: 411.2. HR-MS ESI (+) in MeOH as $[M+H]^+$ ($C_{24}H_{31}N_2O_4^+$) calc.: 411.2278, found: 411.2280. Spectroscopic data are in agreement to earlier reported study.[355]

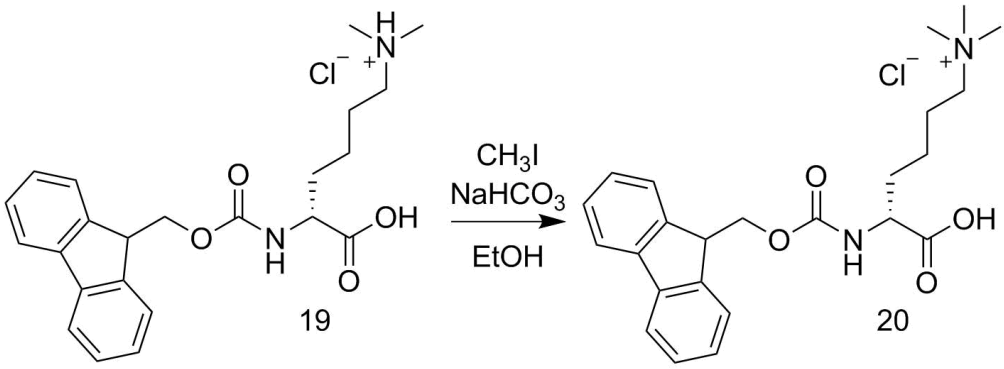


Fmoc-^{D}Lys(me$_2$·HCl)-OH

(R)-2-[9-Fluoronyloxycarbonyl]amino-6-[dimethyl]amino-hexanoic acid hydrochloric acid 5.19

The commercially available Fmoc-^{D}Lys(Boc)-OH **18** (718 mg, 1.53 mmol, 1.0 equiv.) was dissolved in dichloromethane (50 mL) and HCl in diethyl ether (3.8 mL, 7.6 mmol, 2.0 M, 5.0 equiv.) was added in dropwise fashion and after 16 hrs of mechanically stirring crude solid was obtained after evacuation

of volatiles. A solution of the crude in ethanol (15 mL) was prepared and formaldehyde (0.677 mL, 9.0 mmol, 37%$_{(aq)}$, 5.88 equiv.) was added dropwise followed by the portion addition of sodium cyanoborohydride (202 mg , 3.2 mmol, 2.9 equiv.). The mixture was allowed to react for 1 hrs followed by quenching with TFA (300 µL) and stirring for additional 0.25 hrs. The mixture was concentrated and purified using R_f-C_{18} purification. Column: Grace 40 gram, solvent A: [TFA (0.1 v/v%) in MilliQ], Solvent B: [TFA (0.1 v/v%) in acetonitrile]. Program: 4 CV equilibration (5% B), loading (~10 mL, 25% B), 2 CV (5% B), 24 CV (5–100% B). The purified compound was isolated by lyophilisation furnishing 423 mg of product (1.09, 71%). The solid was dissolved in dichloromethane and HCl in diethyl ether was added drop wise. The mixture was stirred for 5 min and volatiles were removed under reduced pressure followed by azeotropic removal of diethyl ether with dichloromethane (3× 10 mL) giving a hydroscopic white solid. mp: 64 °C (decomp.); IR υ_{max}: 3407, 2361, 1696, 1394, 1250, 1202, 1132, 1079 cm^{-1}; $[\alpha]^{25}_D$ −10 (c = 0.005 g·mL^{-1} MeOH); ^{1}H NMR 500 MHz, Solvent: MeOD-d_4, δ (ppm): 7.79 (d, 2 H, *J*= 7.5 Hz, Fmoc-C_4<u>*H*</u>), 7.69-7.61 (m, 2 H, 2× Fmoc-C_1<u>*H*</u>), 7.38 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-C_3<u>*H*</u>), 7.30 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-C_2<u>*H*</u>), 4.40-4.29 (m, 2 H, OC<u>*H*</u>$_2$CH-Fmoc), 4.21 (t, 1 H, *J*= 7.0 Hz, OCH$_2$C<u>*H*</u>-Fmoc), 4.05 (dd, 1 H, *J*= 3.0, 8.0 Hz, N^{2}H$^\alpha$C<u>*H*</u>COO^{2}H), 3.11-2.99 (m, 2 H, $^\delta$CH$_2$$^\varepsilon$C<u>*H*</u>$_2$NH), 2.83 (s, 6 H, $^\varepsilon$CH$_2$NH(C<u>*H*</u>$_3$)$_2$), 1.91-1.80 (m, 1 H, $^\alpha$CH$^\beta$C<u>*H*</u>H$^\gamma$CH$_2$), 1.79-1.65 (m, 3 H, $^\alpha$CH$^\beta$CH<u>*H*</u>$^\gamma$CH$_2$$^\delta$C<u>*H*</u>$_2$$^\varepsilon$CH$_2$), 1.47-1.37 (m, 2 H, $^\beta$CH$_2$$^\gamma$C<u>*H*</u>$_2$$^\delta$CH$_2$); ^{13}C NMR 125 MHz, Solvent: MeOD-d_4, δ (ppm): 178.4, 158.2, 145.3, 142.6, 128.8, 128.2, 126.2, 120.9, 67.7, 58.8 56.6, 48.4, 43.3, 33.1, 25.1, 23.5; ESI-MS (+) in MeOH as $[M+H]^+$ calc: 397.2 found: 397.2; HR-MS ESI (+) in MeOH as $[M+H]^+$ ($C_{23}H_{29}N_2O_4^+$) calc: 397.2122 found: 397.2123.

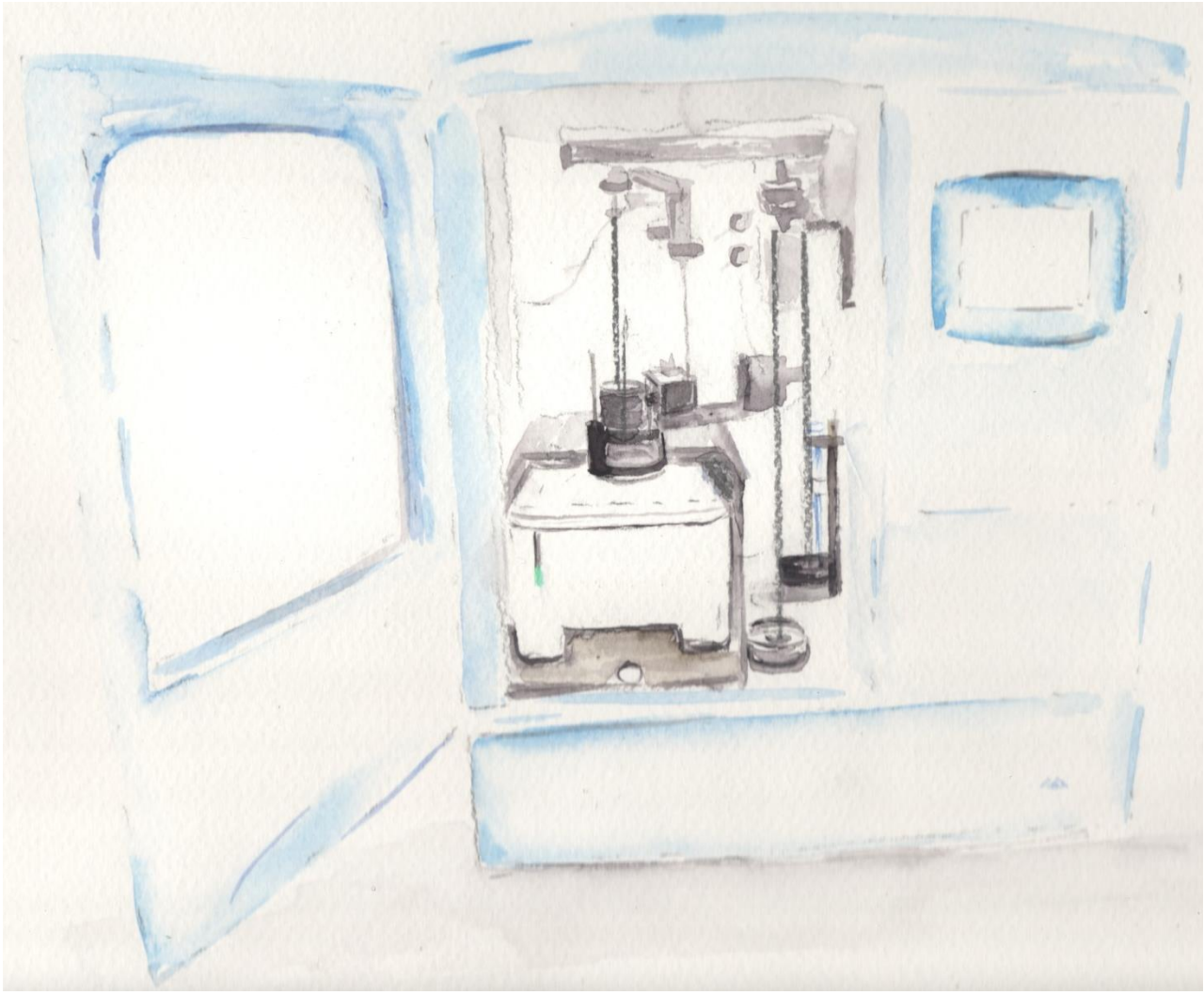


Fmoc-^{D}Lys(me$_3$·Cl)-OH

(R)-2-[9-Fluoronyloxycarbonyl]amino-6-[trimethyl]amino-hexanoate hydrochloride 5.20

A solution containing Fmoc-^{D}Lys(me$_2$·Cl)-OH **19** (1.04 mmol, 1.0 equiv.) in ethanol (10 mL) was prepared and iodomethane (0.650 mL, 10.4 mmol, 10 equiv.) was added dropwise followed by sodium bicarbonate (96 mg, 1.14 mmol, 1.1 equiv.). The mixture was stirred for 48 hrs and volatiles were evacuated under reduced pressure to give the crude. The crude solid was purified using reverse phase purification. Column: R$_f$-C$_{18}$ Grace 40 g, Solvent A: [TFA (0.1 v/v%) in MilliQ], Solvent B: [TFA (0.1 v/v%) in acetonitrile]. Equilibration: 5% B, Loading: 25% B, 10 mL, Program: 5% B for 4 CV, 5%–100% B over 20 CV. Selected fractions were pooled and lyophilized to furnish 328 mg (0.80 mmol, 77%) of product as white solid. The solid was dissolved in dichloromethane (5 mL), HCl in diethyl ether (2 mL, 4.0 mmol, 2.0 M, 3.8 equiv.) was added drop wise, stirred for 5 min and volatiles were removed *in vacuo*. Trace of diethyl ether facilitated by azeotropic distillation with dichloromethane (3× 10 mL) under reduced to give a white hydroscopic solid. mp: 72 °C (decomp.); IR υ_{max}: 3404, 2953, 2362, 1691, 1250, 1128 cm^{-1}; $[\alpha]^{25}_{D}$ −5 (c = 0.012 g·mL^{-1} MeOH); ^{1}H NMR 400 MHz, Solvent: MeOD-d$_4$, δ (ppm): 7.77 (d, 2 H, *J*= 7.5 Hz, 2× Fmoc-C$_4$*H*), 7.65 (t, 2 H, *J*= 8.0 Hz, 2× FmocC$_1$*H*), 7.37 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-C$_3$*H*), 7.29 (t, 2 H, *J*= 7.5 Hz, 2× Fmoc-C$_2$*H*), 4.42-4.26 (m, 2 H, OC*H*$_2$CH-Fmoc), 4.23-4.15 (m, 1 H, OCH$_2$C*H*-Fmoc), 4.09 (dd, 1 H, *J*= 3.5, 5.0 Hz, ND$^{\alpha}$C*H*COOD), 3.30-3.18 (m, 2 H, $^{\delta}$CH$_2$$^{\varepsilon}$C*H*$_2$N(CH$_3$)$_3$), 3.06 (s, 9 H, $^{\varepsilon}$CH$_2$N(C*H*$_3$)$_3$), 1.96-1.84 (m, 1 H, $^{\alpha}$CH$^{\beta}$CH*H*$^{\gamma}$CH$_2$), 1.84-1.60 (m, 3 H, $^{\alpha}$CH$^{\beta}$CH*H*$^{\gamma}$CH$_2$$^{\delta}$C*H*$_2$$^{\varepsilon}CH_2$N), 1.49-1.36 (m, 2 H, $^{\beta}$CH$_2$$^{\gamma}$C*H*$_2$$^{\delta}CH_2$); ^{13}C NMR 101 MHz, Solvent: MeOD-d$_4$, δ (ppm): 177.8, 158.2, 145.2, 142.5, 128.8, 128.2, 126.2, 121.0, 67.7, 56.3, 53.4, 48.4, 33.0, 31.9, 23.5, 23.3; ESI-MS (+) in MeOH as [M+H]$^+$

calc: 411.2 found: 411.3; HR-MS ESI (+) in MeOH as $[M+H]^+$ ($C_{24}H_{31}N_2O_4^+$) calc.: 411.2278, found: 411.2275.



Chapter 6. Summary and Conclusions

DNA regulation is an integral part of the cellular functions and abnormal regulation, due to changes in the DNA (*e.g.* mutations) or epigenetic modifications, can result in diseases, apoptosis or acute cell death. Elucidating the cellular implication of associated proteins is challenging and chemical probes can play an important role in this matter as exemplified in 5-azacytidine for DNMTs and JQ1 for BET-family bromodomains.[356,357] Epigenetic regulation is mediated through (frequently small) modifications on DNA and proteins. The widely studied DNA modification ^{5m}C, introduced by DNMTs, can be iteratively oxidized by TETs using the catalytic 2OG-dependent oxygenase domain to give ^{5hm}C, ^{5f}C and ^{5ca}C.[43] The ^{5ox}C modifications are involved in important cellular functions including epigenetic signalling pathways, transcriptional activation, transcriptional silencing and nucleosomal stabilization.[75] TETs are linked to diseases such as AML, CLL and solid tumors.[4,85,86] Despite understanding some of the functions of TET enzymes, little biophysical and biochemical characterisation has been conducted to date. Additionally, the use of chemical probes could aid in the understanding of TETs in its relation to cellular levels of ^{5m}C, ^{5ox}C and what impact TETs have in diseases; however, there are currently no potent or selective inhibitors available.

Chapter 2 outlines the work carried out to biophysically and biochemically characterise hTET1/2$_{CD}$ and hTET2$_{CD\Delta LCI}$ proteins. The chemiluminescence assay AlphaScreen™ was developed to study TET activity and although the experimental conditions prohibited DNA substrate Michaelis-Menten steady state kinetics, the K_M^{App} of 2OG was obtained successfully. The difference observed between hTET2$_{CD}$ and hTET2$_{CD\Delta LCI}$ (~2.3-fold) implies that the LCI may influence the 2OG binding site whilst hTET1$_{CD}$ shows a significant lower K_M^{App} (~4.5-fold) relative to hTET2$_{CD}$. To evaluate substrate kinetics, ITC and MALDI-TOF mass spectrometry assays were developed demonstrating a ~23-fold higher K_M for single stranded DNA (K_M: 23.8 ± 1.7 μM) than double stranded DNA (K_M of 1.02 ± 0.23 μM) measured using ITC and MALDI-TOF, respectively. The sample preparation and analysis for MALDI-

TOF were found to be complex and therefore a RF-MS activity assay was developed. This method enables continues monitoring of the sequential oxidation of ^{5m}C by hTET2$_{CD\Delta LCI}$. To study the changes of TET activity during incubation with generic and target-specific 2OG-dependent oxygenase inhibitors, oncometabolites, nucleosides and TCA cycle intermediates, AlphaScreen™ and RF-MS assays were used. Interestingly, hTET1/2$_{CD}$ were found to be inhibited by generic inhibitors IOX1 **2.9**, 2,4-PDCA **2.10** and NOG **2.11** (IC_{50}: 0.87–12.5 µM). The ProHD inhibitors show that some, such as GSK1278863 **2.12** and Molidustat **2.18,** inhibit weakly (>50 µM) but others, such as IOX4 **2.17** (IC_{50}: ~18 µM) and Vadadustat **2.19** (IC_{50}: ~6 µM), show good potency against hTET1/2$_{CD}$. KDM-specific inhibitors SD70 **2.20**, QC6352 **2.21**, KDOAM25 **2.22** and CPI455 **2.23** inhibited poorly (>50 µM) while ML324 **2.24** showed good levels of inhibition (IC_{50}: ~1.5 µM). Nucleosides ^{5m}C **2.30** and ^{5hm}C **2.31** were not inhibitory at 100 µM suggesting that DNA nucleotides contribute significantly to a strong binding. While TCA cycle intermediates succinate **2.26** and fumarate **2.27** show relatively weak inhibition for hTET1/2$_{CD}$ (IC_{50}: >100 µM), strong inhibition was found for D-2HG **2.29** for hTET2$_{CD}$ using AlphaScreen™ (IC_{50}: 15.5 µM) and RF-MS (hTET2$_{CD\Delta LCI}$ IC_{50}: 3.5 µM). In contrast, D-2HG (IC_{50}: 417 µM) inhibits hTET1$_{CD}$ considerably weaker. This result indicates the potential link to the mutual exclusivity observed between mutations in hTET2 and IDH1/2 in AML.[88] The advances made in screening methodologies aids in determining the potency of an inhibitor and enables high throughput library screens. This, together with the inhibitor screens here, provides a starting point for the development of chemical probes that can aid in elucidating the biological role of TETs.

Histones are part of the epigenetic landscape and subjected to an extensive array of epigenetic modifications in the form of PTMs. *N*$^{\varepsilon}$-lysine methylation is a frequently observed dynamic epigenetic modification that has a location dependent function where it is associated to be transcriptional activating as well as silencing.[91] Histone *N*$^{\varepsilon}$-lysine methylation is introduced by KMTs but removed by KDMs. An active read out of the mark is provided by chromatin binding modules that recognize the methylation state.

The KDM5A H3K4me$_{3-1}$ demethylase is of scientific interest as it is responsible for the down regulation of the transcriptional activating mark H3K4me$_{3,2}$. KDM5A contains a JmjC-domain, PHD1 and PHD3 which act on the same histone H3 sequence[8,166] and KDM5A is associated to AMKL[8], gastric and breast cancers among others making it an interesting target to study.[9,10,275,276] PHD3 is known to have an increased binding affinity for higher H3K4 methylation states; however, the full implication of PHD3 on KDM5A is only partially understood. To investigate PHD3, *N*$^{\varepsilon}$-triethyl-lysine, *N*$^{\varepsilon}$-piperidine-lysine and *N*$^{\varepsilon}$-trimethyl-lysine derivatives were synthesized and incorporated in histone H3-mimic peptides. ITC experiments were carried out on PHD3 revealing a similar strong binding affinity (K_d: ~0.10 µM) for H3$_{1-21}$Ket$_3$ **3.1** and H3$_{1-21}$K4pip **3.8** peptides relative to H3$_{1-21}$K4me$_3$ **3.2**. Similar to H3$_{1-21}$K4me$_3$ **3.2**, H3$_{1-21}$K4et$_3$ **3.1** and H3$_{1-21}$K4pip **3.8** are also poor binders (K_d: ~5 µM) of PHD1 as determined using ITC. To evaluate H3$_{1-21}$K4et$_3$ **3.1** and H3$_{1-21}$K4pip **3.8** as substrate of the KDM5A-D JmjC-domain, MALDI-TOF activity assays were conducted showing that H3$_{1-21}$K4et$_3$ **3.1** is not a substrate; however, H3$_{1-21}$K4pip **3.8** was hydroxylated by all KDM5 constructs tested. Steady state enzyme kinetics showed inverse proportional correlation in K_M *versus* V_{max}. KDM5A/B showing a high K_M (10–24 µM) but low K_{cat} (0.12–0.31 min^{-1}) while KDM5C/D show a low K_M (1.6–2.5 µM) but a high K_{cat} (1.3–1.5 min^{-1}). The data obtained here reveal that *N*$^{\varepsilon}$-triethyl-lysine is an exclusive binder to PHD3 with substantially weaker affinity for PHD1 and JmjC providing a potential interesting tool for chromatin binding module inhibition. Whilst H3$_{1-21}$K4et$_3$ **3.1** and H3$_{1-21}$K4pip **3.8** are examined as potential inhibitors or activators of the PHD3 domain, it remains unclear if the JmjC-catalytic activity alters due to inter-domain feedback-loop. The use of different nucleosome substrates can potentially facilitate in a deeper understanding of the KDM5A-PHD3 functional role. Additionally, the difference in acceptance among KDM5 proteins for H3$_{1-21}$K4pip hydroxylation suggests a slightly altered JmjC-domain architecture that can be exploited for the development of selective inhibitors.

While many chromatin binding modules that focus on binding of N^{ε}-trimethyl-lysine have been identified and linked to diseases[6] including the KDM5A-PHD3[8], few inhibitors to date have been developed, none of which are potent and selective.[281] In Chapter 4, the sterically enlarged N^{ε}-triethyl-lysine, a derivative of N^{ε}-trimethyl-lysine, is studied in further detail against different chromatin binding modules. N^{ε}-trimethyl-lysine derivatives were produced synthetically, incorporated in H3 mimic peptides and the chromatin binding modules, $KDM4A_{TTD}$, $SGF29_{TTD}$, $SPIN2B_{TTD}$, $KDM5A_{PHD3}$, $TAF3_{PHD}$, $KDM7B_{PHD}$, $KDM7B_{PHDJmjC}$, and $CHD1_{DCD}$ were produced in *E. coli*. The reader proteins and $H3_{1-21}K4me_3$ **4.1** derivatives were then subjected to binding studies using ITC. The values show that $KDM4A_{TTD}$, $SGF29_{TTD}$, $SPIN2B_{TTD}$, $KDM5A_{PHD3}$, $TAF3_{PHD}$ and $CHD1_{DCD}$ have similar levels of binding affinity for $H3_{1-21}K4me_2$,et **4.11**, $H3_{1-21}K4me,et_2$ **4.13**, $H3_{1-21}K4me_2,^{i}pr$ **4.12** and $H3_{1-21}K4et_3$ **4.14** relative to $H3_{1-21}K4me_3$ **4.1**. Most strikingly, $KDM7B_{PHD}$, $KDM7B_{PHDJmjC}$ and $BPTF_{PHD}$ reveal a 5–10-fold increase in binding strength for $H3_{1-21}K4et_3$ **4.14** relative to $H3_{1-21}K4me_3$ **4.1**. The most sterically demanding peptide, $H3_{1-21}K4^{n}Bu_3$ **4.15**, showed typically a weaker binding affinity than the strongest binder for all chromatin binding modules tested suggesting a decreasing point in binding affinity in relation to the N^{ε}-lysine size. In order to gain further insight into the relation of K9 demethylation by the JmjC-domain of $KDM7B_{PHDJmjC}$ to $K4me_3$ or $K4et_3$ PHD binding, MALDI-TOF experiments were carried out revealing a similar K_M but a ~13-fold increase in V_{max} for $H3_{1-21}K4et_3K9me_2$ **4.22** relative to $H3_{1-21}K4me_3K9me_2$ **4.21**. This suggests that KDM7B JmjC activity can be modulated through the PHD-finger. Furthermore, the observation that chromatin binding modules can accommodate larger N^{ε}-trimethyl-lysine substituent opens the possibility that these can accommodate larger PTMs that may not yet have been identified.

The cellular machinery is highly selective in incorporating only L-amino acid during protein synthesis; however, D-amino acids have been widely observed in nature but also in histones. In Chapter 5 D- and L-N^{ε}-methyl-lysine are evaluated against chromatin binding modules, KMTs, and KDMs to gain a better understanding in the method of recognition and protein plasticity. The binding affinity of

KDM5A$_{PHD3}$, TAF3$_{PHD}$, BPTF$_{PHD}$, SGF29$_{TTD}$, and KDM4A$_{TTD}$ against histone H3 mimic peptides H3$_{1-10}$^{L}K4me$_3$ and H3$_{1-10}$^{D}K4me$_3$ was determined using ITC and revealing that H3$_{1-10}$^{L}K4me$_3$ binds similarly to reported values but H3$_{1-10}$^{D}K4me$_3$ gave a decreased binding affinity of ~10-fold or ~32-fold for PHD and TTD readers, respectively. MD confirmed these observations demonstrating that H3^{L}K4me$_3$ is the favoured substrate but H3^{D}K4me$_3$ remains a strong binder throughout the simulations (20 ns) for all reader proteins tested. To understand the plasticity of KMTs, MALDI-TOF assays were conducted using KMT1C, KMT1D, KMT5A and KMT7 revealing good methylation activity against the natural H3^{L}K4 or H3^{L}K9 substrate; however, only trace of methylation was observed for H3^{D}K4 or H3^{D}K9. Interestingly, KMT1C and KMT1D showed higher levels of methylation under prolonged exposure, increased enzyme and SAM levels. Despite this, a competition assay using both the L- and D-substrate indicated that L-substrate is the preferred substrate. Similar observations were made using KDMs where a demethylation assay was carried out using RF-MS. The corresponding natural substrates for KDM1A$_{C1\text{-}corest}$, KDM4A$_{JmjC}$, KDM4D$_{JmjC}$, KDM4E$_{JmjC}$, KDM5B$_{JmjC}$ and KDM5C$_{JmjC}$ showed efficient demethylation levels with the H3$_{1-21}$^{L}K4me$_{3/2}$ **5.9/13** substrate yet H3$_{1-21}$^{D}K4me$_{3/2}$ **5.10/14** substrate was only demethylated to potentially traces. An increased amount of enzyme did not yield increased levels of demethylated product. The combined data provides evidence that chromatin binding modules have a higher plasticity in comparison to KMTs and KDMs. This could mean on a cellular level that upon a change in stereo chemistry, the mark is unlikely to be dynamic but can still be 'read'. Furthermore, this observation can be utilized to selectively investigate chromatin binding modules and may allow for the chemical functionalization of the *C*$^{\alpha}$-lysine.

Overall, the data presented here gives a deeper insight in demethylases and chromatin binding modules. Furthermore, the collected data suggests the use of new chemical tools in the form of *N*$^{\varepsilon}$-lysine enlargement and *N*$^{\alpha}$-lysine stereochemistry that can be utilized for the development of potent and selective inhibitors and chemical probes for chromatin binding modules. These chemical probes

could assist in understanding biological systems and have the potential to find their way into therapeutics.



Chapter 7. Materials and methods

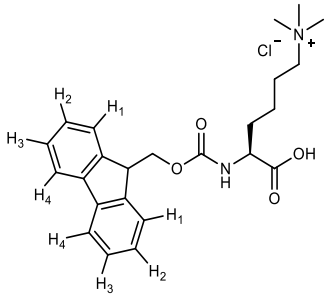
7.1 Synthetic methods

7.1.1 General experimental conditions

Solvents for monomer synthesis and anhydrous solvents were obtained from Sigma Aldrich. All reactions were executed under air atmosphere unless stated otherwise. Hydrogenation was introduced using a needle, adapter and balloon. Purifications were performed on Biotage flash purification system with Grace Solve Revelaris or Biotage [SiO_2] and R_f-C_{18} columns. GraceResolv standard setting were used for column: 4 g (12 mL·min^{-1}, 1 Column Volume (CV): 5mL), 12 g (24 mL·min^{-1}, 16 mL), 24 g (24 mL·min^{-1}, 1 CV: 32 mL), 40 g (40 mL·min^{-1}, 1 CV: 50 mL), 80 g (60 mL·min^{-1}, 1 CV: 100 mL), 120 g (80 mL·min^{-1}, 1 CV: 160 mL). Biotage standard settings: 10 g (36 mL·min^{-1}, 1 CV: 15 mL), 25 g (75 mL·min^{-1}, 1 CV: 33 mL), 50 g (100 mL·min^{-1}, 1 CV: 66 mL) and 100 g (100 mL·min^{-1}, CV: 132 mL). Fraction collection was at 254 nm and monitoring at 230 nm. Reagents were obtained from Alfa Aeser, ChemImpex, Sigma Aldrich, Bachem, or Acros and used as received. Yields noted are purified and isolated material, unless stated otherwise. Merck thin layer chromatography (TLC) was used to monitor reaction progress equipped with 254 nm coating and staining was achieved with $KMnO_4$ or Ninhydrin in combination with vigorous heating. ^{1}H NMR spectra were acquired using 400 mHz and ^{13}C NMR was performed at 101 mHz on Brucker electrospin 400 with samplXpress. The spectra X-axis are denoted in (δ) ppm and peak splitting is described in Hz. Abbreviations used referring to the peak observed are: s= singlet; d= doublet; t=triplet; q= quadruplet; *b*= broad and m= multiplet. Assigned protons are noted as *italic* and <u>underlined</u>. Deuterated solvents for NMR were $CDCl_3$-d_1, DMSO-d_6 and MeOD-d_4 obtained from Sigma Aldrich. Infrared spectra were acquired using IR Bruker tensor 27. Optical rotation data was furnished with Schmidt haensch polirameter with 100 mm cell at 25.0 ºC. Low resolution mass spectrometry was obtained employing Agilent 6120 quadrople LC-MS Agilent 1260 equipped with a Waters sample manager 2777. LC-MS analysis was used Agilent 6120

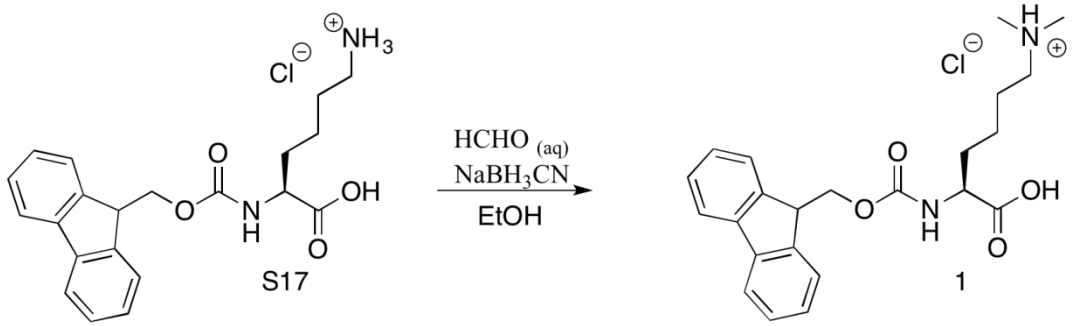
quadrupole and LC-MS Agilent 1200. All small molecules were determined in positive ion mode using electron spray ionisation (ESI). Lyophilisation for the removal of water was accomplished on Scanvac coolsafe.

Assignment of Fluorenyl methyloxycarbonyl (Fmoc) protons depicted with Fmoc-Lys(me_3·Cl)-OH as example:



7.1.2 Synthesis of Fmoc-protected lysine derivatives

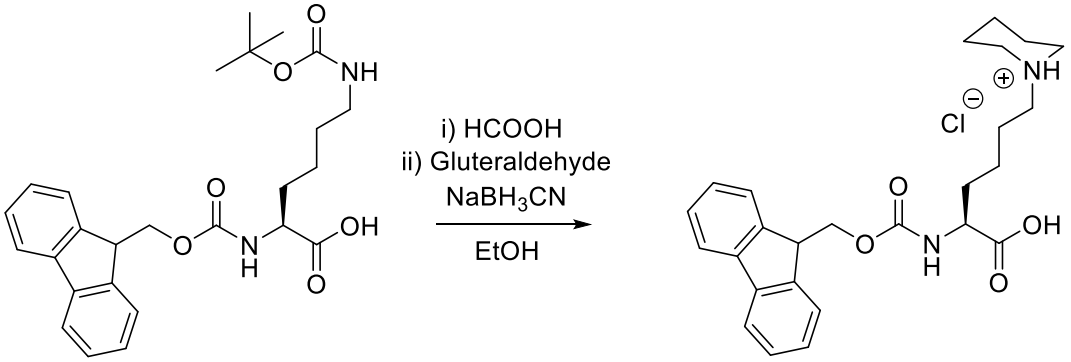
Fmoc-Lys(me_2·HCl)-OH[358]



(*S*)-2-[Fluorenylmethyloxycarbonyl]amino-6-di[methyl]amino-hexanoic acid hydrochloric acid 4.9a or 5.16

To a solution of Fmoc-Lys(Boc)-OH (469 mg, 1.0 mmol, 1.0 equiv.) in CH_2Cl_2 (7.5 mL, final: 0.1 M), was added HCl in Et_2O (2.5 ml, 2 M, 5.0 mmol, 5.0 equiv.) followed by stirring for 16 hrs. Volatiles were removed *in vacuo* giving a white solid. The solid was dissolved in EtOH; to the stirred solution was added HCHO (447 µL, 37% sol., 6.0 mmol, 6.0 equiv.) and $NaBH_3CN$ (138 mg, 2.2 mmol, 2.2

equiv.). The mixture was stirred for 45 min, TFA (320 µL) + MeOH (10 mL) was added to the suspension forming a solution that was stirred for 15 min. The mixture was concentrated and purified on HPLC column. Grace 40 g R_f-C_{18} column. 3 CV (5%) equilibration, loading (6 mL, 25%), 2 CV (5%), 20 CV (5–100%) MeCN in H_2O (containing 0.01% TFA). Selected fractions were combined and lyophilized obtaining 324 mg of a white solid (82%). The white solid was dissolved in CH_2Cl_2 was added HCl in Et_2O (1 mL, 2 M, 2 mmol, 1.68 equiv.) and volatiles were removed *in vacuo*. Azeotropically, Et_2O was removed with CH_2Cl_2 (4× 10 mL) *in vacuo*. mp: 67 °C (decomp.); IR υ_{max}: 3405, 2951, 1695, 1202, 1132 cm^{-1}; $[\alpha]^{25}_D$ +15.0 (c = 0.010 g·mL^{-1} MeOH); ^{1}H NMR 700 MHz (DMSO-d_6) δ (ppm): 12.64 (s, 1 H, $^{\alpha}$CHCOO*H*), 10.32 (s, 1 H, N*H*$(CH_3)_2$), 7.90 (d, *J*= 7.5 Hz, 2 H, Fmoc-C_4*H*), 7.73 (t, *J*= 7.0, 7.5 Hz, 2 H, 2× FmocC_1*H*), 7.68 (d, *J*= 7.5 Hz, 1 H, N*H*$^{\alpha}$CHCOOH), 7.42 (t, *J*= 7.5 7.0 Hz, 2 H, 2×FmocC_3*H*), 7.33 (t, *J*= 7.5 7.0 Hz, 2 H, 2×FmocC_2*H*), 4.33-4.26 (m, 2 H, OC*H*$_2$CH), 4.23 (t, *J*= 7.0 Hz, 1 H, OCH_2C*H*), 3.94 (dt, *J*= 9.0 5.0 Hz, 1 H, NH$^{\alpha}$C*H*COOH), 3.03-2.95 (m, 2 H, $^{\delta}CH_2{}^{\varepsilon}C$*H*$_2$NH), 2.71 (d, *J*= 5.0 Hz, 6 H, NH(C*H*$_3$)$_2$), 1.77-1.70 (m, 1 H, $^{\alpha}CH^{\beta}CH$*H*$^{\gamma}CH_2$), 1.70-1.58 (m, 3 H, $^{\alpha}CH^{\beta}CH$*H*$^{\gamma}CH_2{}^{\delta}C$*H*$_2{}^{\varepsilon}CH_2$), 1.41-1.29 (m, 2 H, $^{\beta}CH_2{}^{\gamma}C$*H*$_2{}^{\delta}CH_2$). ^{13}C NMR 100 MHz (MeOD-d_4) δ (ppm): 178.2, 158.3, 145.2, 142.5, 128.8, 128.2, 126.2, 120.9, 67.7, 58.6, 56.5, 48.4, 43.2, 32.9, 25.0, 23.7; ESI-MS in MeOH, $[M+H]^+$, calc: 397.2 found: 397.2. HR-MS (ESI+) in MeOH as $C_{23}H_{29}N_2O_4^+$, calc: 397.2113 found: 397.2122.

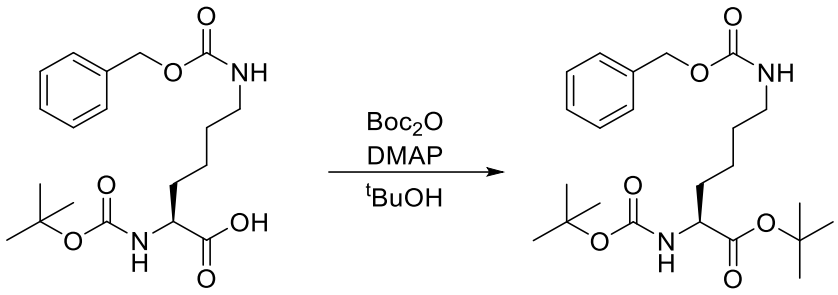


Fmoc-Lys(pip·HCl)-OH

(S)-2-[fluorenylmethyloxycarbonyl]amino-6-[1-piperidine]-hexanoic acid hydrochloric acid 3.10 or 4.10F

To a Fmoc-Lys(Boc)-OH (10.232 g, 21.8 mmol) in a round bottom flask (100 mL), was added neat formic acid (40 mL) followed by stirring for 16 hrs. After the reaction was completed indicated by TLC, volatiles were removed under reduced pressure to give a colourless oil. Three subsequent round of azeotropic removal of formic acid using $CHCl_3$ yield a white crystalline solid. Fmoc-Lys(HCOOH)-OH (838 mg, 2.02 mmol, 1.0 equiv.) was suspended in EtOH (100 mL, 0.02 M) and cooled down to 0 °C while mechanically stirred. After 15 min glutaraldehyde (1.51 mL, 16.0 mmol, 8.0 equiv.) was added followed by $NaBH_3CN$ (269 mg, 4.28. 2.1 equiv.). The mixture was stirred at 0 °C for 3 hrs while aloud to gradually warm up to room temperature (RT) and stirred for another 16 hrs. The mixture was quenched with trifluoroacetic acid (TFA) (300 µL, 3.91 mmol, 1.9 equiv.) and stirred for 15 min. The mixture was concentrated *in vacuo* and the crude was loaded on a Grace 40 g R_f-C_{18} column for purification. Program: 4 column volume (CV) (5%) equilibration; loading (~15 mL, 25%); 2 CV (5%); 24 CV (linear gradient of 5%–100%) MeCN in H_2O (containing 0.01% TFA). Lyophilized fractions yielded a white solid (457 mg, 1.05 mmol, 52%). The compound was dissolved in CH_2Cl_2 (5 mL) and HCl was added in Et_2O (2 mL, 2 M, 4 mmol, 3.8 equiv.). Azeotropically, Et_2O was removed with CH_2Cl_2 *in vacuo* (3× 15 mL) giving a hydroscopic white solid. mp: 93 °C (decomp); IR υ_{max}: 3680, 2922, 1707, 1033 cm^{-1}; $[\alpha]^{25}_D$ −0.1 (c = 0.014 g·mL^{-1} MeOH); ^{1}H NMR 400 MHz (DMSO-d_6) δ (ppm): 12.68 (s, 1 H, $^{\alpha}$CHCOO<u>*H*</u>), 9.97 (s, 1 H, N<u>*H*</u>(CH_2CH_3)CH_3), 7.90 (d, *J*= 7.0 Hz, 2 H, Fmoc-C_4<u>*H*</u>), 7.73 (q, *J*= 4.5, 3.0, 4.0 Hz, 2 H, 2× FmocC_1<u>*H*</u>), 7.69 (d, *J*= 8.0 Hz, 1 H, N<u>*H*</u>$^{\alpha}$CHCOOH), 7.42 (t, *J*= 8.0 Hz, 2 H, 2× FmocC_3<u>*H*</u>), 7.33 (t, *J*= 7.5 Hz, 2 H, 2×FmocC_2<u>*H*</u>), 4.34-4.26 (m, 2 H, OC<u>*H*</u>$_2$CH), 4.26-4.18 (m, 1 H, OCH_2C<u>*H*</u>), 3.94 (dt, *J*= 9.0 3.0 Hz, 1 H, NH$^{\alpha}$C<u>*H*</u>COOH), 3.37 (bd, *J*= 12 Hz, 2 H, NH(CH<u>*H*</u>)$_2$(CH_2)$_2CH_2$), 3.37 (bd, *J*= 12 Hz, 2 H, NH(CH<u>*H*</u>)$_2$(CH_2)$_2CH_2$), 2.95 (bq, *J*= 7.5 6 7.5 Hz, 2 H, $^{\delta}CH_2{}^{\varepsilon}C$<u>*H*</u>$_2$NH) 2.87-2.73 (m, 2 H, NH(C<u>*H*</u>H)$_2$(CH_2)$_2CH_2$), 1.81-1.54 (m, 8 H, $^{\alpha}CH^{\beta}C$<u>*H*</u>$_2{}^{\gamma}CH_2{}^{\delta}C$<u>*H*</u>$_2{}^{\varepsilon}CH_2$, NH($CH_2$)$_2$(C<u>*H*</u>$_2$)$_2CH_2$), 1.47-1.20 (m, 4 H, $^{\beta}CH_2{}^{\gamma}C$<u>*H*</u>$_2{}^{\delta}CH_2$, NH(CH_2)$_2$(CH_2)$_2$C<u>*H*</u>$_2$)). ^{13}C NMR 101 MHz (MeOD-d_4) δ (ppm): 175.5, 158.7, 145.1, 142.5, 128.8, 128.2, 126.2, 120.9, 67.9, 57.9, 54.9, 54.2, 48.3, 32.0, 24.4, 24.2, 24.1, 22.7; ESI-MS (+) in MeOH, $[M+H]^+$ calc: 437.2 found: 437.2; HR-MS ESI (+) in MeOH as $C_{26}H_{33}N_2O_4^+$, $[M+H]^+$, calc: 437.2435

found: 437.2427.



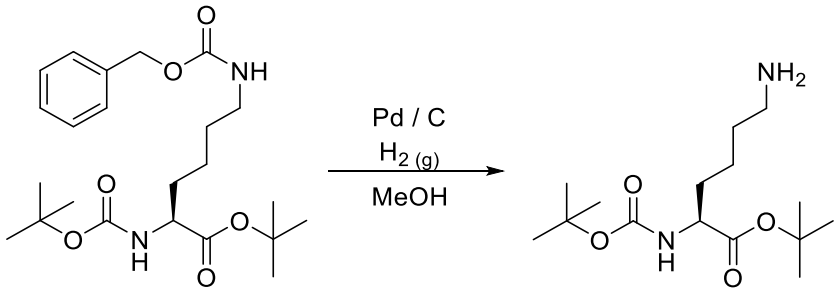
Boc-Lys(Cbz)-O^{t}Bu[315]

tertbutyl (S)-2-[tertbutoxycarbonyl]amino-6-[benzyloxycarbonyl]amino-hexanoate

3.12 or 4.21

Method adapted from previous report procedure with minor alterations.[315] To a solution of Boc-Lys(Cbz)-OH (5.186 g, 13.63 mmol, 1 equiv.) in anhydrous tert-Butanol (24 mL, 0.56 M), Boc_2O (4.360 g, 20.0 mmol, 1.47 equiv.) was added, followed by 4-(dimethylamino)pyridine (DMAP) (315 mg, 2.56 mmol, 0.19 equiv.) under inert (N_2) atmosphere while mechanically stirred at RT. The colour less solution turned over time yellow and after TLC indicated reaction completion (2.5 hrs), volatiles were removed *in vacuo*. The crude was purified on Grace 80 g [SiO_2] column using ethyl acetate (EA) in cyclohexane. Settings: equilibration 3CV; loading (~15 mL, 15%); 12 CV (linear gradient of 15%–40%) EA in cyclohexane. Selected fractions were pooled and concentrated giving colourless oil (5.044 g, 85%). IR υ_{max}: 2980, 1703, 1250 (m), 1153; $[\alpha]^{25}_D$ −13.2 (c = 0.005 g·mL^{-1} $CHCl_3$); ^{1}H NMR 400 MHz ($CDCl_3$-d_1) δ (ppm): 7.39-7.28 (m, 5H, $(Ar\text{-}H)_5$), 5.09 (s, 2 H, NHCOOC$\underline{H}_2$Ar), 5.05 (bd, *J*= 7.5 Hz, 1 H, N$\underline{H}^{\alpha}$CHCOO), 4.80 (bs, 1 H, $^{\varepsilon}CH_2N\underline{H}COO$), 4.20-4.12 (m, 1 H, $NH^{\alpha}C\underline{H}COO$), 3.19 (q, *J*= 6.5 Hz, 2 H, $^{\delta}CH_2{}^{\varepsilon}C\underline{H}_2NH$), 1.84-1.70 (m, 1 H, $^{\alpha}CH^{\beta}CH\underline{H}^{\gamma}CH_2$), 1.66-1.49 (m, 3 H, $^{\alpha}CH^{\beta}C\underline{H}H^{\gamma}CH_2{}^{\delta}C\underline{H}_2{}^{\varepsilon}CH_2$), 1.48-1.30 (m, 2 H, $^{\beta}CH_2{}^{\gamma}C\underline{H}_2{}^{\delta}CH_2$), 1.45 (s, 9 H, $^{\alpha}CHCOOC(C\underline{H}_3)_3$), 1.43 (s, 9 H, $NHCOOC(C\underline{H}_3)_3$); ^{13}C NMR 101 mHz ($CDCl_3$-$d_1$) δ (ppm): 172.0, 156.5, 155.5, 136.7, 128.5, 128.16, 128.10, 81.9, 79.7, 66.6, 53.8, 40.8, 32.7, 29.5, 28.4, 28.1, 22.4; ESI-MS (+) in MeOH, $[M+Na]^+$ calc: 459.2 found: 459.3; HR-MS ESI (+) in MeOH $[M+Na]^+$ as $C_{23}H_{36}N_2O_6Na^+$, calc:

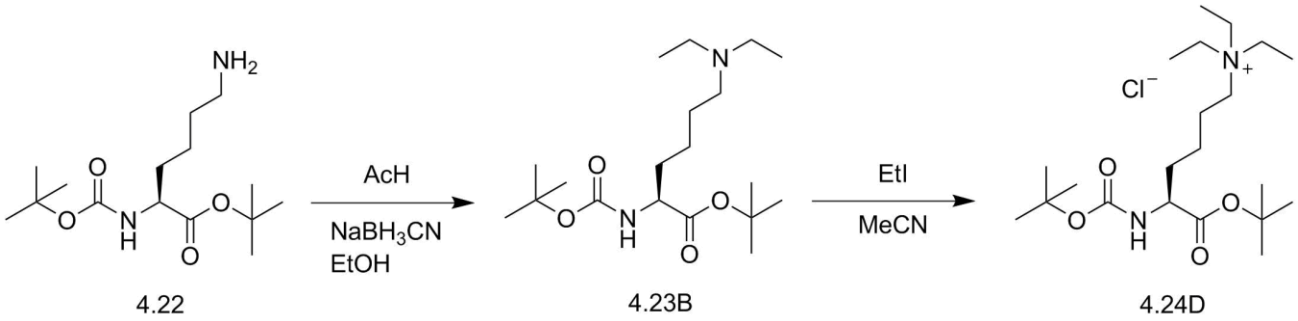
459.2466 found: 459.2464.



Boc-Lys(H)-O^{t}Bu[286]

tertbutyl (S)-2-[tertbutoxycarbonyl]amino-6-amino-hexanoate 3.13 or 4.22

Method adapted from previous report procedure with minor alterations.[286] To a solution of Boc-Lys(Cbz)-O^{t}Bu (2.370 g, 5.43, 1.0 equiv.) in MeOH (40 mL, 0.14 M), was added Palladium on carbon (118 mg, 10 w%, 2.1 mol%) introducing an inert atmosphere and then a hydrogen atmosphere (1 bar) while mechanically stirred. The mixture was vigorous stirred as a black suspension and after TLC indicated full deprotection (1.5 hrs) the solution was vacuum filtrated through a short path of Celite®. The Celite® was washed with MeOH (3× 10 mL) and the supernatant was concentrated *in vacuo*. The crude was purified by column chromatography Grace 24 g. [SiO_2]. Program: 3 CV equilibration (1:5:84 NH_4OH in H_2O, MeOH, CH_2Cl_2); loading (~5 mL, 1:5:84 NH_4OH in H_2O, MeOH, CH_2Cl_2); 12 CV (1:5:94 NH_4OH in H_2O, MeOH, CH_2Cl_2 to 1:15:84 NH_4OH in H_2O, MeOH, CH_2Cl_2, over a linear gradient). Selected fractions were combined and concentrated under reduced pressure yielding a colour less oil (1.470 g, 4.86 mmol, 90%). IR υ_{max}: 2977, 1710, 1366, 1153 cm^{-1}; $[\alpha]^{25}_D$ −23.8 (c = 0.007 g·mL^{-1} $CHCl_3$); ^{1}H NMR 400 mHz ($CDCl_3$-d_1) δ (ppm): 5.02 (bd, *J*= 7.5 Hz, 1 H, N*H*$^{\alpha}$CHCOO), 4.17 (b dt, *J*= 5.5 13 Hz , 1 H, NH$^{\alpha}$C*H*COO), 2.68 (t, *J*= 7.0 Hz, 2 H, $^{\delta}CH_2{}^{\varepsilon}CH_2NH_2$), 1.84-1.70 (m, 1 H, $^{\alpha}CH^{\beta}CHH^{\gamma}CH_2$), 1.84-1.70 (m, 1 H, $^{\alpha}CH^{\beta}CHH^{\gamma}CH_2$), 1.66-1.54 (m, 1 H, $^{\alpha}CH^{\beta}CHH^{\gamma}CH_2$), 1.52-1.39 (m, 2 H, $^{\gamma}CH_2{}^{\delta}CH_2{}^{\varepsilon}CH_2$), 1.46 (s, 9 H, $^{\alpha}$CHCOOC(CH_3)$_3$), 1.44 (s, 9 H, NHCOOC(CH_3)$_3$) 1.39-1.29 (m, 2 H, $^{\beta}CH_2{}^{\gamma}CH_2{}^{\delta}CH_2$); ^{13}C NMR 100 mHz ($CDCl_3$-$d_1$) δ (ppm): 172.0, 155.4, 81.6, 79.4, 53.9, 41.7, 33.0, 32.6, 28.3, 27.9, 22.5; ESI-MS (+) in MeOH, $[M+H]^+$ calc: 303.2 found: 303.2; HR-MS ESI (+) in MeOH as $C_{15}H_{31}N_2O_4^+$ $[M+H]^+$, calc: 303.2278 found: 303.2276.

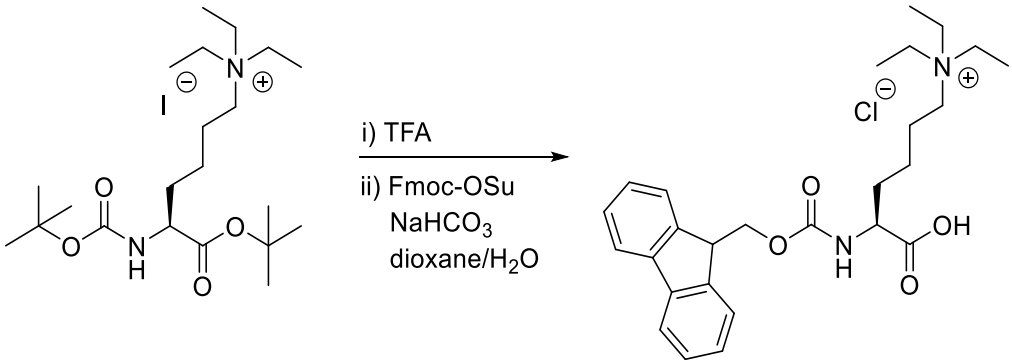


Boc-Lys(et$_3$ I)-O^{t}Bu

tertbutyl (S)-2-[tertbutoxycarbonyl]amino-6-[triethyl]amino-hexanoate iodide 3.15 or 4.24D

To a solution of Boc-Lys(H)-O^{t}Bu (905 mg, 3.0 mmol, 1.0 equiv.) in EtOH (12 mL, 0.25 M) was added drop wise acetaldehyde (1 mL, 17.6 mmol, 5.9 equiv.) followed by the slow addition of $NaBH_3CN$ (417 mg, 6.6 mmol, 2.2 equiv.). The solution stirred for 40 min until ESI-MS indicated starting material consumption followed by the addition of sat. $NaHCO_3$ sol. (40 mL). The aqueous phase was extracted with CH_2Cl_2 (3× 35 mL), the organic layer was dried over $MgSO_4$ and volatiles were removed under reduced pressure obtaining a crude red oil. To a solution of Boc-Lys(et$_2$)-O^{t}Bu in anhydrous MeCN (6 mL, 0.5 M) was added iodoethane (480 µL, 6.0 mmol, 2.0 equiv.) followed by heating the cooler equipped mixture to 80 °C for 1 hrs while stirring. ESI–MS indicated the absence of the diethyl-lysine intermediate followed by gradually cooling the sample to room temperature and pursued by concentrating *in vacuo*. The crude was purified on Grace column 40g [SiO_2]. Program: 3 CV equilibration (0%); loading (5 mL, 0%); 12 CV (linear gradient of 0%–10%) MeOH in CH_2Cl_2. The correct fractions were combined as indicated by the UV-VIS trace, volatiles were removed under reduced pressure obtaining a white hydroscopic solid (701 mg, 46%, 1.37 mmol). mp: 136 °C (decomp.); IR υ_{max}: 2980, 1707, 1392, 1153 cm^{-1}; $[\alpha]^{25}_D$ −13.6 (c = 0.017 g·mL^{-1} $CHCl_3$); ^{1}H NMR 400 MHz ($CDCl_3$-d$_1$) δ (ppm): 5.13 (bd, *J*= 8.0 Hz, 1 H, N*H*$^{\alpha}$CHCOO), 4.17 (bdt, *J*= 5.0 8 Hz , 1 H, NH$^{\alpha}$C*H*COO), 3.48 (q, *J*= 7.0 7.5 7.0 Hz, 6 H, N(C*H*$_2$CH$_3$)$_3$), 3.38 (dt, *J*=7.0 13.0 Hz, 1 H, $^{\delta}$CH$_2$$^{\varepsilon}$CH*H*Net$_3$), 3.28-3.16 (m, 1 H, $^{\delta}$CH$_2$$^{\varepsilon}$C*H*HNet$_3$), 1.84-1.70 (m, 1 H, $^{\alpha}$CH$^{\beta}$C*H*$_2$$^{\gamma}CH_2$), 1.84-1.70 (m, 1 H, $^{\alpha}$CH$^{\beta}$CH*H*$^{\gamma}$CH$_2$), 1.66-1.50 (m, 4 H, $^{\beta}$CH2$^{\gamma}$C*H*$_2$$^{\delta}$C*H*$_2$$^{\varepsilon}CH_2$), 1.47 (s, 9 H, $^{\alpha}$CHCOOC(C*H*$_3$)$_3$),

1.43 (s, 9 H, NHCOOC(C<u>*H*</u>$_3$)$_3$), 1.40 (t, *J*= 7.0 Hz, 9 H, N(CH$_2$C<u>*H*</u>$_3$)$_3$); ^{13}C NMR 125 mHz (CDCl$_3$-d$_1$) δ (ppm): 171.6, 155.9, 82.6, 80.0, 57.7, 53.9, 53.0, 32.7, 28.5, 28.2, 22.4, 21.4, 8.4; ESI-MS (+) in MeOH, [M+H]$^+$ calc: 387.3 found: 387.3; HR-MS ESI (+) in MeOH as $C_{21\ H43}N_2O_4^+$ [M+H]$^+$, calc: 387.3217 found: 387.3222.



Fmoc-Lys(et$_3$·Cl)-OH

(S)-2-[fluorenylmethyloxycarbonyl]amino-6-[triethyl]amino-hexanoic acid hydrochloric acid 3.16 or 4.25D

Boc-Lys(et$_3$·I)-O^{t}Bu (679 mg, 1.39 mmol, 1.0 equiv.) was dissolved in TFA (10 mL, 0.09 M) and stirred for 32 hrs followed by the removal of volatiles under reduced pressure. The crude orange oil was dissolved in a solution of 1,4-dioxane and H_2O (1:1, 14 mL, 0.1 M) and $NaHCO_3$ (977 mg, 11.6 mmol, 8.4 equiv.) was added observing gas release. Fmoc-OSu (545 mg, 1.61 mmol, 1.16 equiv.) was introduced to the mixture while mechanically stirred. After 2 hrs, the mixture was partially concentrated (~4 mL) and the crude was purified on Grace 40 g R_f-C_{18} column. Settings: 3 CV (5%) equilibration; loading (added 6 mL, 25%); 2 CV (5%); 20 CV (5–100%) MeCN in H_2O (containing 0.01% TFA). Selected fractions were combined and lyophilised giving 510 mg (1.13 mmol) of a yellowish solid (81% yield). The solid was dissolved in CH_2Cl_2 (5 mL), HCl in Et_2O (1.0 mL, 2 M, 2 mmol, 3.45 equiv.) was drop wise added. Azeotropically, Et_2O was removed with CH_2Cl_2 *in vacuo* (3× 15 mL) giving a hydroscopic off white solid. mp: 91 °C (decomp); IR υ_{max}: 3369, 2980, 2360, 1708, 1252 cm^{-1}; $[\alpha]^{25}_D$ −0.3 (c = 0.010 g·mL^{-1} MeOH); ^{1}H NMR 400 MHz (DMSO-d$_6$) δ (ppm):

12.67 (s, 1 H, NH$^{\alpha}$CHCOO<u>*H*</u>), 7.90 (d, *J*= 7.5 Hz, 2 H, 2×FmocC$_4$<u>*H*</u>), 7.72 (q, *J*= 5.0, 2.5, 5.0 Hz, 2 H, 2×FmocC$_1$<u>*H*</u>), 7.67 (d, *J*= 8.0 Hz, 1 H, N<u>*H*</u>$^{\alpha}$CHCOOH), 7.43 (t, *J*= 7.5 Hz, 2 H, 2×FmocC$_3$<u>*H*</u>), 7.33 (t, *J*= 7.5 Hz, 2 H, 2×FmocC$_2$<u>*H*</u>), 4.36-4.18 (m, 3 H, OC<u>*H*</u>$_2$C<u>*H*</u>), 3.97 (m, 1 H, NH$^{\alpha}$C<u>*H*</u>COOH), 3.21 (q, *J*= 7.0 Hz, 6 H, N(C<u>*H*</u>$_2$CH$_3$)$_3$), 3.16-3.00 (m, 2 H, $^{\delta}$CH$_2^{\varepsilon}$ C<u>*H*</u>$_2$N), 1.83-1.71 (m, 1 H, $^{\alpha}$CH$^{\beta}$CH<u>*H*</u>$^{\gamma}$CH$_2$), 1.71-1.48 (m, 3 H, $^{\alpha}$CH$^{\beta}$C<u>*H*</u>H$^{\gamma}$CH$_2^{\delta}$C<u>*H*</u>$_2^{\varepsilon}$CH$_2$), 1.42-1.28 (m, 2 H, $^{\beta}$CH$_2^{\gamma}$C<u>*H*</u>$_2^{\delta}$CH$_2$), 1.15 (t, *J*= 7.0 Hz, 9 H, N(CH$_2$C<u>*H*</u>$_3$)$_3$); ^{13}C NMR 126 MHz (DMSO-d$_6$) δ (ppm): 173.7, 156.2, 143.8, 140.7, 127.7, 127.1, 125.3, 120.2, 65.5, 55.8, 53.5, 52.0, 46.7, 30.2, 22.6, 20.5, 7.2; ESI-MS (+) in MeOH, [M+H]$^+$ calc: 453.3 found: 453.3; HR-MS ESI (+) in MeOH [M+H]$^+$ as $C_{27}H_{37}N_2O_4^+$, calc: 453.2748 found: 453.2740.

7.1.3 Solid Phase Peptide Synthesis

Peptides were synthesized and purified as described before.[342] Briefly, peptides were prepared using microwave assisted solid phase peptide synthesis using a Liberty Blue by CEM. Peptides were built from C- to N-terminus using Fmoc chemistry. Breipohl (rink resin) (0.1 mmol, ~0.6 mmol g^{-1}) was soaked in DMF (5 mL) and transferred in the reaction vessel using HT12 high throughput loader. Fmoc deprotection was achieved with piperidine in DMF (20 v/v%, 4 mL). Washing of the resin was achieved with DMF and coupling of the amino acid was done using Fmoc-protected amino acids (0.2 M, 2.5 mL), *N,N'*-diisopropylcarbodiimide (DIC) (0.5 M, 1.0 mL) and Oxyma Pure [OxymaPur (1.0 M, 0.5 mL), *N,N*-diisopropylethylamine (DIEA) (0.1 M). After the peptide was built, the resin was washed with 3× DMF (~10 mL DMF), 4–5× (~10 mL Et$_2$O) and dried upon air over night. The peptide as deprotected using a mixture of 1,2-ethanedithiol (EDT) / triisopropylsilane (TIPS) / MQ in trifluoro acetic acid (TFA) (2.5 : 2.5 : 2.5 : 92.5, 3.0 mL). After 3 hrs, the mixture was precipitated in cold (−25 °C) Et$_2$O (30 mL) while stirring and a pellet was formed after centrifugation (4255 × g, 3 min). Solvent was removed by decantation and the pellet was suspended in MQ (5 mL) and lyophilized. The crude peptide was suspended in MQ (1.8 mL), filtered (0.45 µm) and purified on JASCO PU-2086 plus

system with column: Gemini-NX 5u C18 110A. Program: 1–20% (45 min), 20–100% (2.5 min), 100% (10 min) acetonitrile in MQ (degassed) containing 0.1% TFA (20 mL·min^{-1}). Selected fractions were lyophilized and the mass and purity was determined using MALDI-TOF and LC-MS.

Table 7.1.1: peptides produced with their corresponding mass and purity observed.

				LC-MS							MALDI-TOF
Sequence	**Modification**	**Chemical formula**	**Exact mass**							**Purity**	
				$[M+3H]^{3+}$	$[M+4H]^{4+}$	$[M+5H]^{5+}$	$[M+3H]^{3+}$	$[M+4H]^{4+}$	$[M+5H]^{5+}$		$[M+1H]^{+}$
$H3_{1-21}$	$K4et_3$	$C_{100}H_{184}N_{36}O_{28}$	2337.41	780.1	585.4	468.5	780.1	585.4	468.5	>98.0%	2337.6
$H3_{1-21}$	$K4me_3$	$C_{97}H_{178}N_{36}O_{28}$	2295.36	766.1	574.8	460.1	766.1	574.8	460.0	97.0%	2295.5
$H3_{1-21}$	$K4me_2$	$C_{96}H_{176}N_{36}O_{28}$	2281.35	761.5	571.3	457.3	761.5	571.4	457.3	96.5%	2281.7
$H3_{1-21}$	$K4me_1$	$C_{95}H_{174}N_{36}O_{28}$	2267.33	756.8	567.8	454.5	756.7	567.8	454.5	97.6%	2267.6
$H3_{1-21}$	$K4me_0$	$C_{94}H_{172}N_{36}O_{28}$	2253.31	752.1	564.3	451.7	752.0	564.3	451.6	97.3%	2253.6
$H3_{1-21}$	A4	$C_{91}H_{165}N_{35}O_{28}$	2196.26	733.1	550.1	440.3	733.0	550.0	440.2	98.1%	2196.5
$H3_{1-21}$	K4Ac	$C_{96}H_{174}N_{36}O_{29}$	2295.32	766.1	574.8	460.1	766.1	575.0	460.1	97.8%	2295.6
$H3_{1-21}$	K4pip	$C_{99}H_{180}N_{36}O_{28}$	2321.38	774.8	581.4	465.3	774.7	581.4	465.2	97.0%	2321.6

7.2 Microbiological methods and techniques

7.2.1 General materials and methods for microbiology

The pH was measured using glass calomel electrode (Aldrich) controlled by Jenway pH meter model 3305 calibrated with Buffer Solution pH 7 (yellow, cat#: CHE1402) and pH 10 (blue, cat#: CHE1364) Scientific Laboratory supplies. Centrifugation was achieved with Avanti J-26S XP Centrifuge (Beckman Coulter™) with rotors JLA-16.250 (8000 rpm) for harvesting cells and rotor JA-25.50 (20 000 rpm). Protein was concentrated using Allegra X-30R Centrifuge (Beckman Coulter™) carrying SX4400 rotor (4700 rpm). Centrifugations of small quantities was conducted using Microfuge 22R® (Beckman Coulter™). Sterilization of autoclave was carried out with FALCON LTE. Incubation of Petri dishes was conducted in Heraeus (D6450 Hanau), starter cultures in Model G25 (New Brunswick Scientific co. inc.) and large scale growth of media in Innova® 44 (New Brunswick Scientific). Samples were stored in U101 Innova® Ultra-Low Temperature Freezer (New Brunswick).

7.2.2 Production of competent cells

Agar plate was prepared following protocol (Chapter 7.2.3) without antibiotics and the plate was inoculated using inoculation loop streaked over the agar plate from competent *Escherichia coli* (*E. coli*) source stock. Plate was incubated (37 °C, overnight) and a single colony was picked to inoculate a starter culture (2TY media, 100 mL) while shaking (220 rpm, 37 °C). Fresh media (2TY, 100 mL) was inoculated (1.0 mL) and incubated with shaking (220 rpm, 37 °C) until an OD600: ~0.375 was reached. The media was split (2×50 mL) and cooled on ice (10 min). The falcon tubes were centrifuged (1600 × g, 7 min, 4.0 °C), supernatant was removed and the cells were suspended in calcium solution [$CaCl_2$ (100 mM), glycerol (15 v/v%) in MilliQ (filtered 0.2 µM)] (10 mL). The mixture was centrifuged (1100 × g, 5 min, 4.0 °C), supernatant was removed, calcium solution was added (10 mL) and incubated on ice (30 min). The mixture was centrifuged (1100 × g, 5 min, 4.0 °C), supernatant

was decanted of and the cells were suspended in calcium solution (2.0 mL), aliquoted and stored (−80 °C).

7.2.3 Preparation of antibiotics

Antibiotic solutions Kanamycin Sulphate (30 $g \cdot L^{-1}$) (Apollo Scientific Ltd, cat#: BIK0126) and Chloramphenicol (34 $g \cdot L^{-1}$) (Apollo Scientific Ltd, cat#: BIC0113) were prepared at 1000× the concentration in MilliQ. The solution was filtered (0.2 µm) in sterile Eppendorf tubes or falcon tubes and stored in the freezer (~ −24 °C).

7.2.4 Preparation of 2TY media or Terrific broth (TB)

2TY solution was prepared from Tryptone (16 $g \cdot L^{-1}$), Yeast extract (10 $g \cdot L^{-1}$), NaCl (5.0 $g \cdot L^{-1}$) in MilliQ (pH 7.0). Terrific broth was made from Terrific broth, Fisher Bioreagents™ (product #: 12891660) (47.6 $g \cdot L^{-1}$) or from [Tryptone (12 $g \cdot L^{-1}$), Yeast extract (24 $g \cdot L^{-1}$), Glycerol (4 $mL \cdot L^{-1}$), KH_2PO_4 (17 mM), K_2HPO_4 (72 mM) in MilliQ]. Solutions were sterilized in shaking flasks (1 L) equipped with stopper or filter.

7.2.5 Preparation of agar plate

To suspension of Agar (15 $g \cdot L^{-1}$), NaCl (5.0 $g \cdot L^{-1}$), Yeast Extract (10 $g \cdot L^{-1}$), Tryptone (16 $g \cdot L^{-1}$) in MilliQ (300 mL) was prepared in DURAN® bottle (500 mL) and sterilized (loose cap). The solution was cooled down and solidified. Using microwave irradiation the mixture was liquefied cooled by ambient temperatures to hand warm and appropriate antibiotics were added. The liquid were poured in sterile Petri dishes (~15 mL per Petri dish) and stored (~4 °C).

7.2.6 Transformation of plasmid in *E. coli*

Transformation of plasmids in competent *E. coli* cells was carried out using sterilized reagents and equipment. Competent cells from storage (−80 °C) were warmed up to liquid suspension by hand and transferred (~25–35 µL) using to an Eppendorf (1.5 mL). Plasmid (50 ng) was added followed by careful mixing and incubating on ice (30 min). The mixture was heat shocked for (42 sec, 45 °C) and recovered on ice (3.0 min). To the mixture was added 2TY media (950 µL), mixed using pipetting and incubated (37 °C, 1 hrs). If the cells were not cloudy, the solution was concentrated by centrifugation and supernatant (800 µL) was removed. The cells were suspended (if necessary) and pipetted on an agar plate (with appropriate antibiotics), spread using an inoculation loop and incubated (37 °C, overnight). Single colony was used for inoculation of the starter culture.

7.2.7 Preparation of HisTrap™ HP

Nickel affinity protein purification was achieved using HisTrap™ HP 5 mL (GE Health). Before every run the column was equilibrated using protocol column cleaning and after the run the protocol column storage was carried out. 1 column volume (CV) = 5 mL. Striping buffer [Na_2HPO_4 (20 mM), NaCl (500 mM), EDTA (50 mM) in MilliQ (pH 7.4, filtered (0.2 µm))], recharging buffer [$Ni(II)SO_4$ (50 mM) in MilliQ (filtered)] and storage solution [ethanol (20 v/v%) in MilliQ (80 v/v%)].

Column cleaning: MilliQ (3 CV), stripping buffer (5 CV), MilliQ (5 CV), recharge buffer (3 CV), MilliQ (3 CV), Binding buffer (3 CV).

Column storage: MilliQ (3 CV), stripping buffer (5 CV), MilliQ (5 CV), storage solution (3 CV).

7.2.8 SDS-PAGE preparation and analysis

SDS-PAGE gel was either used from pre-cast gels SDS-PAGE (NuPAGE™ 4-12% Bis-tris gel, cat#: NP0335BOX) or prepared in-house. Starting materials for in-house production were purchases from commercially available sources: Acrylamide / Bios-acrylamide (37.5:1), 30% solution (Sigma, cat#: A3699-100ML), *N,N,N',N'*-tetramethylethylenediamine (Sigma, cat#: T9281-25ML) (TEMED), Ammonium persulphate (Sigma, cat#: A3678-25G) (APS) and Sodium dodecyl sulphate (SDS) (Sigma- Aldrich, cat#: L4509-500G). SDS (Sodium dodecyl sulphate) PAGE (Poly acrylamide gel electrophoresis) gels were prepared by mixing resolving gel (Table 6.1.2) components which was rapidly added to the cast and set for 30 min. On top was placed the stacking gel (Table 7.2.1), added to the cast, lane comb (15 channels) was inserted and set for 30 min. The comb was gently removed, the gel was wetted (MilliQ) and stored (~4.0 °C).

Table 7.2.1: Component composition of resolving gel and stacking gel to produced SDS-PAGE.* Tris·HCl (0.5 M, pH 6.6).

	Resolving	Stacking
Tris HCl (1.5 M, pH 8.8)	1.25 mL	630 µL
MilliQ	1.15 mL	1.57 mL
Acrylamide (30%)	2.50 mL	250 µL
SDS (10 w/v%)	50 µL	25 µL
APS (10%)	50 µL	25 µL
TEMED	4.0 µL	2.5 µL

Gels were insert in the Bio-Rad Mini Protein II system and running buffer [Tris base (3.0 g), glycine (14.4 g), SDS (1.0 g) in MilliQ (pH ~8.3) (1.0 L)] was added. Precast gels were ran using Mini Gel tank (Invitrogen, ThermoFisher Scientific, cat#: A25977) and NuPAGE™ MOPS SDS Running buffer (Thermo Fisher, Scientific, cat#: NP0001). Samples were prepared from purified protein fractions (5.0 µL) and SDS-loading dye [Tris·HCl (100 mM), SDS (4 w/v%), Bromophenol blue (0.2 w/v%), glycerol (20 v/v%), 1,4-dithiothreitol (200 mM) in MilliQ (pH 6.8)] (5.0 µL) heated to 95 °C (5 min) and loaded on the SDS-PAGE together with 1 or 2 lanes of PageRuler™ (ThermoFisher scientific, cat#: 26617). Electrophoresis was performed at 180 V for ~55 min. Fixation of the gel was achieved

with Coomassie Brilliant Blue G-250 (ThermoFisher Scientific, cat#: 20279) [1 $g \cdot L^{-1}$, in methanol (50 v/v%), acetic acid (10 v/v%) in MilliQ] for 20–30 min followed by destaining solution [methanol (40 v/v%), acetic acid (10 v/v%) in MilliQ (50 v/v%)]. Precast gels were fixated with Instant*Blue*™ (Expedeon, cat#: ISB1L). Images of the gel were made using Gel Doc™ XR+ (Bio-Rad Laboratories) system controlled using Image Lab (v. 5.2.1 build 11). Images were edited using Adobe Photoshop CS6 (v. 13.0).

The molecular weight of the proteins and absorption coefficient was calculated by ExPASy ProtParam Tool (web.expasy.org/protparam).[359] The protein concentration was estimated using ND-1000 NanoDrop™ spectrophotometer (Thermo Scientific) or NanoDrop™ one (Thermo Scientific) measuring 280 nm.

Chapter 8. References

1. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* vol. 502 472–479 (2013).
2. Langemeijer, S. M. C. *et al.* Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat. Genet.* **41**, 838–842 (2009).
3. Lorsbach, R. B. *et al.* TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* **17**, 637–641 (2003).
4. Yang, H. *et al.* Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* **32**, 663–669 (2013).
5. Kooistra, S. M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. *Nat. Rev. Mol. Cell Biol.* **13**, 297–311 (2012).
6. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14**, 1025–1040 (2007).
7. Højfeldt, J. W., Agger, K. & Helin, K. Histone lysine demethylases as targets for anticancer therapy. *Nat. Rev. Drug Discov.* **12**, 917–930 (2013).
8. Wang, G. G. *et al.* Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **459**, 847–851 (2009).
9. Li, L. *et al.* Critical role of histone demethylase RBP2 in human gastric cancer angiogenesis. *Mol. Cancer* **13**, 81 (2014).
10. Cao, J. *et al.* Histone Demethylase RBP2 Is Critical for Breast Cancer Progression and Metastasis. *Cell Rep.* **6**, 868–877 (2014).
11. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **23**, 781–783 (2009).
12. Alberts, B, Johnson, A, Lewis, J, Raff, M, Roberts, K. and Walter, P. *Molecular Biology of The Cell. 4ed*. *New York: Garland Science;* (2005).
13. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
14. Satchwell, S. C., Drew, H. R. & Travers, A. A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191**, 659–675 (1986).
15. Schones, D. E. *et al.* Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell* **132**, 887–898 (2008).
16. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823–837 (2007).
17. Ammar, R. *et al.* Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife* **1**, (2012).
18. Hansen, J. C. Human mitotic chromosome structure: what happened to the 30-nm fibre? *EMBO J.* **31**, 1621–1623 (2012).
19. Marzluff, W. F., Gongidi, P., Woods, K. R., Jin, J. & Maltais, L. J. The Human and Mouse Replication-Dependent Histone Genes. *Genomics* **80**, 487–498 (2002).

20. Ahmad, K. & Henikoff, S. The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly. *Mol. Cell* **9**, 1191–1200 (2002).

21. Mito, Y., Henikoff, J. G. & Henikoff, S. Genome-scale profiling of histone H3.3 replacement patterns. *Nat. Genet.* **37**, 1090–1097 (2005).

22. Witt, O., Albig, W. & Doenecke, D. Testis-Specific Expression of a Novel Human H3 Histone Gene. *Exp. Cell Res.* **229**, 301–306 (1996).

23. Ueda, J. *et al.* Testis-Specific Histone Variant H3t Gene Is Essential for Entry into Spermatogenesis. *Cell Rep.* **18**, 593–600 (2017).

24. Urahama, T. *et al.* Histone H3.5 forms an unstable nucleosome and accumulates around transcription start sites in human testis. *Epigenetics and Chromatin* **9**, 2 (2016).

25. Schenk, R., Jenke, A., Zilbauer, M., Wirth, S. & Postberg, J. H3.5 is a novel hominid-specific histone H3 variant that is specifically expressed in the seminiferous tubules of human testes. *Chromosoma* **120**, 275–285 (2011).

26. Palmer, D. K., O'Day, K., Wener, M. H., Andrews, B. S. & Margolis, R. L. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J. Cell Biol.* **104**, 805–815 (1987).

27. Takahashi, K., Chen, E. S. & Yanagida, M. Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. *Science* **288**, 2215–2219 (2000).

28. Wiedemann, S. M. *et al.* Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. *J. Cell Biol.* **190**, 777–791 (2010).

29. Kujirai, T. *et al.* Structure and function of human histone H3.Y nucleosome. *Nucleic Acids Res.* **44**, 6127–6141 (2016).

30. Talbert, P. B. & Henikoff, S. Histone variants — ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* **11**, 264–275 (2010).

31. Globisch, D. *et al.* Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. *PLoS One* **5**, e15367 (2010).

32. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).

33. Stubbs, T. M. *et al.* Multi-tissue DNA methylation age predictor in mouse. *Genome Biol.* **18**, 68 (2017).

34. Chen, H.-F. & Wu, K.-J. Epigenetics, TET proteins, and hypoxia in epithelial-mesenchymal transition and tumorigenesis. *BioMedicine* **6**, 1 (2016).

35. Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet. Genome Res.* **14**, 9–25 (1975).

36. Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* **19**, 219–220 (1998).

37. Bestor, T. H. & Verdine, G. L. DNA methyltransferases. *Curr. Opin. Cell Biol.* **6**, 380–389 (1994).

38. Feinberg, A. P. Phenotypic plasticity and the epigenetics of human disease. *Nature* **447**, 433–440 (2007).

39. Zhang, W. & Xu, J. DNA methyltransferases and their roles in tumorigenesis. *Biomark. Res.* **5**, 1 (2017).

40. Yang, J., Lior-Hoffmann, L., Wang, S., Zhang, Y. & Broyde, S. DNA Cytosine Methylation: Structural and Thermodynamic Characterization of the Epigenetic Marking Mechanism. *Biochemistry* **52**, 2828–2838 (2013).

41. Kriaucionis, S. & Heintz, N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* **324**, 929–930 (2009).

42. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).

43. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).

44. Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8**, 1698–1710 (2009).

45. Hu, L. *et al.* Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **155**, 1545–1555 (2013).

46. Sudhamalla, B., Dey, D., Breski, M. & Islam, K. A rapid mass spectrometric method for the measurement of catalytic activity of ten-eleven translocation enzymes. *Anal. Biochem.* **534**, 28–35 (2017).

47. Nishio, K. *et al.* Thioether Macrocyclic Peptides Selected against TET1 Compact Catalytic Domain Inhibit TET1 Catalytic Activity. *ChemBioChem* **19**, 979–985 (2018).

48. Ko, M. *et al.* Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* **497**, 122–126 (2013).

49. Belle, R., Kawamura, A. & Arimondo, P. B. *Chemical Epigenetics, Chapter 9: Chemical coumpounds targeting DNA methylation and hydroxymethylation*. (In press, 2019). doi:10.1007/7355_2019_76.

50. Aik, W. *et al.* Structure of human RNA N6-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation. *Nucleic Acids Res.* **42**, 4741–4754 (2014).

51. 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.* **101**, 14051–14056 (2004).

52. Han, Z. *et al.* Crystal structure of the FTO protein reveals basis for its substrate specificity. *Nature* **464**, 1205–1209 (2010).

53. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Embryogenesis: Demethylation of the zygotic paternal genome. *Nature* **403**, 501–502 (2000).

54. He, Y.-F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **333**, (2011).

55. Weber, A. R. *et al.* Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat. Commun.* **7**, 10806 (2016).

56. Squires, J. E. *et al.* Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* **40**, 5023–5033 (2012).

57. Fu, L. *et al.* Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **136**, 11582–11585 (2014).

58. Iwan, K. *et al.* 5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo. *Nat.*

Chem. Biol. **14**, 72–78 (2018).

59. Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).

60. Nabel, C. S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8**, 751–758 (2012).

61. Hu, L. *et al.* Structural insight into substrate preference for TET-mediated oxidation. *Nature* **527**, 118–122 (2015).

62. Pais, J. E. *et al.* Biochemical characterization of a Naegleria TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. *Proc. Natl. Acad. Sci.* **112**, 4316–4321 (2015).

63. Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581 (2014).

64. Schröder, A. S. *et al.* 2′-(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes. *Chem. Commun.* **52**, 14361–14364 (2016).

65. DeNizio, J. E., Liu, M. Y., Leddin, E. M., Cisneros, G. A. & Kohli, R. M. Selectivity and Promiscuity in TET-Mediated Oxidation of 5-Methylcytosine in DNA and RNA. *Biochemistry* **58**, 411–421 (2019).

66. Hashimoto, H. *et al.* Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* **506**, 391–395 (2014).

67. Kizaki, S. & Sugiyama, H. CGmCGCG is a versatile substrate with which to evaluate Tet protein activity. *Org. Biomol. Chem.* **12**, 104–107 (2014).

68. Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* **6**, 1049–1055 (2014).

69. Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557 (2015).

70. Gackowski, D., Zarakowska, E., Starczak, M., Modrzejewska, M. & Olinski, R. Tissue-specific differences in DNA modifications (5-hydroxymethylcytosine, 5-formylcytosine, 5-carboxylcytosine and 5-hydroxymethyluracil) and their interrelationships. *PLoS One* **10**, e0144859 (2015).

71. Xu, Y. *et al.* Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. *Mol. Cell* **42**, 451–464 (2011).

72. Pastor, W. A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).

73. Raiber, E.-A. *et al.* 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22**, 44–49 (2015).

74. Hardwick, J. S. *et al.* 5-Formylcytosine does not change the global structure of DNA. *Nat. Struct. Mol. Biol.* **24**, 544–552 (2017).

75. Raiber, E.-A. *et al.* 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nat. Chem.* **10**, 1258–1266 (2018).

76. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14**, R119 (2013).

77. Zhang, H. *et al.* TET1 is a DNA-binding protein that modulates DNA methylation and gene

transcription via hydroxylation of 5-methylcytosine. *Cell Res.* **20**, 1390–1393 (2010).

78. Kellinger, M. W. *et al.* 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **19**, 831–833 (2012).

79. Ono, R. *et al.* LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res.* **62**, 4075–4080 (2002).

80. Huang, H. *et al.* TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proc. Natl. Acad. Sci.* **110**, 11994–11999 (2013).

81. Hsu, C.-H. *et al.* TET1 Suppresses Cancer Invasion by Activating the Tissue Inhibitors of Metalloproteinases. *Cell Rep.* **2**, 568–579 (2012).

82. Sun, M. *et al.* HMGA2/TET1/HOXA9 signaling pathway regulates breast cancer growth and metastasis. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9920–9925 (2013).

83. Kudo, Y. *et al.* Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Sci.* **103**, 670–676 (2012).

84. Rasmussen, K. D. & Helin, K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes and Development* vol. 30 733–750 (2016).

85. Weissmann, S. *et al.* Landscape of TET2 mutations in acute myeloid leukemia. *Leukemia* **26**, 934–942 (2012).

86. Quesada, V. *et al.* Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 47–52 (2012).

87. Quivoron, C. *et al.* TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. *Cancer Cell* **20**, 25–38 (2011).

88. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell* **18**, 553–567 (2010).

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

90. Zhao, S. *et al.* Glioma-Derived Mutations in IDH1 Dominantly Inhibit IDH1 Catalytic Activity and Induce HIF-1α. *Science* **324**, 261–265 (2009).

91. Greer, E. L. & Shi, Y. Histone methylation: A dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **13**, 343–357 (2012).

92. Fischle, W., Franz, H., Jacobs, S. A., Allis, C. D. & Khorasanizadeh, S. Specificity of the chromodomain Y chromosome family of chromodomains for lysine-methylated ARK(S/T) motifs. *J. Biol. Chem.* **283**, 19626–19635 (2008).

93. Rossetto, D., Avvakumov, N. & Côté, J. Histone phosphorylation: A chromatin modification involved in diverse nuclear events. *Epigenetics* **7**, 1098–1108 (2012).

94. Sakabe, K., Wang, Z. & Hart, G. W. β-N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proc. Natl. Acad. Sci.* **107**, 19915–19920 (2010).

95. Christophorou, M. A. *et al.* Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* **507**, 104–108 (2014).

96. García-Giménez, J. L. *et al.* Histone H3 Glutathionylation in Proliferating Mammalian Cells

Destabilizes Nucleosomal Structure. *Antioxid. Redox Signal.* **19**, 1305–1320 (2013).

97. Farrelly, L. A. *et al.* Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. *Nature* **567**, 535–539 (2019).

98. Wiśniewski, J. R., Zougman, A. & Mann, M. Nε-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**, 570–577 (2008).

99. Kebede, A. F. *et al.* Histone propionylation is a mark of active chromatin. *Nat. Struct. Mol. Biol.* **24**, 1048–1056 (2017).

100. 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).

101. Lu, Y. *et al.* Dynamics and functional interplay of histone lysine butyrylation, crotonylation, and acetylation in rice under starvation and submergence. *Genome Biol.* **19**, 144 (2018).

102. Xie, Z. *et al.* Metabolic Regulation of Gene Expression by Histone Lysine β-Hydroxybutyrylation. *Mol. Cell* **62**, 194–206 (2016).

103. Dai, L. *et al.* Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nat. Chem. Biol.* **10**, 365–370 (2014).

104. Xie, Z. *et al.* Lysine Succinylation and Lysine Malonylation in Histones. *Mol. Cell. Proteomics* **11**, 100–107 (2012).

105. Steven Stanley, J., Griffin, J. B. & Zempleni, J. Biotinylation of histones in human cells: Effects of cell proliferation. *Eur. J. Biochem.* **268**, 5424–5429 (2001).

106. Weake, V. M. & Workman, J. L. Histone Ubiquitination: Triggering Gene Activity. *Molecular Cell* vol. 29 653–663 (2008).

107. Shiio, Y. & Eisenman, R. N. Histone sumoylation is associated with transcriptional repression. *Proc. Natl. Acad. Sci.* **100**, 13225–13230 (2003).

108. Messner, S. *et al.* PARP1 ADP-ribosylates lysine residues of the core histone tails. *Nucleic Acids Res.* **38**, 6350–6362 (2010).

109. Borun, T. W., Pearson, D. & Paik, W. K. Studies of histone methylation during the HeLa S-3 cell cycle. *J. Biol. Chem.* **247**, 4288–4298 (1972).

110. Gershey, E. L., Haslett, G. W., Vidali, G. & Allfrey, V. G. Chemical studies of histone methylation. Evidence for the occurrence of 3-methylhistidine in avian erythrocyte histone fractions. *J. Biol. Chem.* **244**, 4871–4877 (1969).

111. Murray, K. The Occurrence of ϵ-N-Methyl Lysine in Histones. *Biochemistry* **3**, 10–15 (1964).

112. Paik, W. K. & Kim, S. E-N-dimethyllysine in histones. *Biochem. Biophys. Res. Commun.* **27**, 479–483 (1967).

113. Hempel, K., Lange, H. W. & Birkofer, L. [Epsilon-N-trimethyllysine, a new amino acid in histones]. *Naturwissenschaften* **55**, 37 (1968).

114. Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* **49**, e324–e324 (2017).

115. Allis, C. D. *et al.* New Nomenclature for Chromatin-Modifying Enzymes. *Cell* **131**, 633–636 (2007).

116. Byvoet, P., Shepherd, G. R. R., Hardin, J. M. M. & Noland, B. J. J. The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Arch. Biochem.*

Biophys. **148**, 558–567 (1972).

117. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).

118. Budhavarapu, V. N., Chavez, M. & Tyler, J. K. How is epigenetic information maintained through DNA replication? *Epigenetics Chromatin* **6**, 32 (2013).

119. Lee, B. M. & Mahadevan, L. C. Stability of histone modifications across mammalian genomes: implications for 'epigenetic' marking. *J. Cell. Biochem.* **108**, 22–34 (2009).

120. Zhao, B. *et al.* H3K14me3 genomic distributions and its regulation by KDM4 family demethylases. *Cell Res.* **28**, 1118–1120 (2018).

121. Jack, A. P. M. *et al.* H3K56me3 Is a Novel, Conserved Heterochromatic Mark That Largely but Not Completely Overlaps with H3K9me3 in Both Regulation and Localization. *PLoS One* **8**, e51765 (2013).

122. Voigt, P. *et al.* Asymmetrically Modified Nucleosomes. *Cell* **151**, 181–193 (2012).

123. Zee, B. M. *et al.* In vivo residue-specific histone methylation dynamics. *J. Biol. Chem.* **285**, 3341–3350 (2010).

124. Chin, H. G., Patnaik, D., Estève, P. O., Jacobsen, S. E. & Pradhan, S. Catalytic properties and kinetic mechanism of human recombinant Lys-9 histone H3 methyltransferase SUV39H1: Participation of the chromodomain in enzymatic catalysis. *Biochemistry* **45**, 3272–3284 (2006).

125. Guo, H.-B. & Guo, H. Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity. *Proc. Natl. Acad. Sci.* **104**, 8797–8802 (2007).

126. Yeates, T. O. Structures of SET domain proteins: Protein lysine methyltransferases make their mark. *Cell* vol. 111 5–7 (2002).

127. 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).

128. Deshpande, A. J. *et al.* AF10 Regulates Progressive H3K79 Methylation and HOX Gene Expression in Diverse AML Subtypes. *Cancer Cell* **26**, 896–908 (2014).

129. Okada, Y. *et al.* hDOT1L Links Histone Methylation to Leukemogenesis. *Cell* **121**, 167–178 (2005).

130. Bernt, K. M. *et al.* MLL-Rearranged Leukemia Is Dependent on Aberrant H3K79 Methylation by DOT1L. *Cancer Cell* **20**, 66–78 (2011).

131. Kerry, J. *et al.* MLL-AF4 Spreading Identifies Binding Sites that Are Distinct from Super-Enhancers and that Govern Sensitivity to DOT1L Inhibition in Leukemia. *Cell Rep.* **18**, 482–495 (2017).

132. Deshpande, A. J., Bradner, J. & Armstrong, S. A. Chromatin modifications as therapeutic targets in MLL-rearranged leukemia. *Trends in Immunology* vol. 33 563–570 (2012).

133. Cho, M. H. *et al.* DOT1L cooperates with the c-Myc-p300 complex to epigenetically derepress CDH1 transcription factors in breast cancer progression. *Nat. Commun.* **6**, 7821 (2015).

134. Kong, X. *et al.* Catalytic Mechanism Investigation of Lysine-Specific Demethylase 1 (LSD1): A Computational Study. *PLoS One* **6**, e25444 (2011).

135. Yu, T., Higashi, M., Cembran, A., Gao, J. & Truhlar, D. G. Concerted hydrogen atom and

electron transfer mechanism for catalysis by lysine-specific demethylase. *J. Phys. Chem. B* **117**, 8422–8429 (2013).

136. Akyüz, M. A. & Erdem, S. S. Computational modeling of the direct hydride transfer mechanism for the MAO catalyzed oxidation of phenethylamine and benzylamine: ONIOM (QM/QM) calculations. *J. Neural Transm.* **120**, 937–945 (2013).

137. Fitzpatrick, P. F. Oxidation of amines by flavoproteins. *Arch. Biochem. Biophys.* **493**, 13–25 (2010).

138. Forneris, F., Binda, C., Vanoni, M. A., Battaglioli, E. & Mattevi, A. Human histone demethylase LSD1 reads the histone code. *J. Biol. Chem.* **280**, 41360–41365 (2005).

139. Yang, M. *et al.* Structural Basis for CoREST-Dependent Demethylation of Nucleosomes by the Human LSD1 Histone Demethylase. *Mol. Cell* **23**, 377–387 (2006).

140. Wang, J. *et al.* The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat. Genet.* **41**, 125–129 (2009).

141. Kahl, P. *et al.* Androgen receptor coactivators lysine-specific histone demethylase 1 and four and a half LIM domain protein 2 predict risk of prostate cancer recurrence. *Cancer Res.* **66**, 11341–11347 (2006).

142. Hayami, S. *et al.* Overexpression of LSD1 contributes to human carcinogenesis through chromatin regulation in various cancers. *Int. J. Cancer* **128**, 574–586 (2011).

143. Kauffman, E. C. *et al.* Role of androgen receptor and associated lysine-demethylase coregulators, LSD1 and JMJD2A, in localized and advanced human bladder cancer. *Mol. Carcinog.* **50**, 931–944 (2011).

144. Lim, S. *et al.* Lysine-specific demethylase 1 (LSD1) is highly expressed in ER-negative breast cancers and a biomarker predicting aggressive biology. *Carcinogenesis* **31**, 512–520 (2010).

145. Wang, Y. *et al.* LSD1 Is a Subunit of the NuRD Complex and Targets the Metastasis Programs in Breast Cancer. *Cell* **138**, 660–672 (2009).

146. 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 (2011).

147. Clifton, I. J. *et al.* Structural studies on 2-oxoglutarate oxygenases and related double-stranded β-helix fold proteins. *J. Inorg. Biochem.* **100**, 644–669 (2006).

148. Hegg, E. L. & Que, L. *The 2-His-1-carboxylate facial triad - An emerging structural motif in mononuclear non-heme iron(II) enzymes*. *European Journal of Biochemistry* vol. 250 625–629 (John Wiley & Sons, Ltd (10.1111), 1997).

149. Horton, J. R., Upadhyay, A. K., Hashimoto, H., Zhang, X. & Cheng, X. Structural Basis for Human PHF2 Jumonji Domain Interaction with Metal Ions. *J. Mol. Biol.* **406**, 1–8 (2011).

150. Walport, L. J., Hopkinson, R. J. & Schofield, C. J. Mechanisms of human histone and nucleic acid demethylases. *Curr. Opin. Chem. Biol.* **16**, 525–534 (2012).

151. Flashman, E., Davies, S. L., Yeoh, K. K. & Schofield, C. J. Investigating the dependence of the hypoxia-inducible factor hydroxylases (factor inhibiting HIF and prolyl hydroxylase domain 2) on ascorbate and other reducing agents. *Biochem. J.* **427**, 135–142 (2010).

152. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature* **419**, 174–178 (2002).

153. Stubbe, J. Identification of two α-ketoglutarate-dependent dioxygenases in extracts of

Rhodotorula glutinis catalyzing deoxyuridine hydroxylation. *J. Biol. Chem.* **260**, 9972–9975 (1985).

154. Cardinale, G. J., Rhoads, R. E. & Udenfriend, S. Simultaneous incorporation of 18O into succinate and hydroxyproline catalyzed by collagen proline hydroxylase. *Biochem. Biophys. Res. Commun.* **43**, 537–543 (1971).

155. MYLLYLÄ, R., TUDERMAN, L. & KIVIRIKKO, K. I. Mechanism of the Prolyl Hydroxylase Reaction: 2. Kinetic Analysis of the Reaction Sequence. *Eur. J. Biochem.* **80**, 349–357 (1977).

156. Krebs, C. *et al.* Rapid freeze-quench 57 Fe Mössbauer spectroscopy: Monitoring changes of an iron-containing active site during a biochemical reaction. *Inorganic Chemistry* vol. 44 742–757 (2005).

157. Price, J. C., Barr, E. W., Tirupati, B., Bollinger, J. M. & Krebs, C. The first direct characterization of a high-valent iron intermediate in the reaction of an α-ketoglutarate-dependent dioxygenase: A high-spin Fe(IV) complex in taurine/α-ketoglutarate dioxygenase (TauD) from Escherichia coli. *Biochemistry* **42**, 7497–7508 (2003).

158. Proshlyakov, D. A., Henshaw, T. F., Monterosso, G. R., Ryle, M. J. & Hausinger, R. P. Direct Detection of Oxygen Intermediates in the Non-Heme Fe Enzyme Taurine/α-Ketoglutarate Dioxygenase. *J. Am. Chem. Soc.* **126**, 1022–1023 (2004).

159. Grzyska, P. K., Appelman, E. H., Hausinger, R. P. & Proshlyakov, D. A. Insight into the mechanism of an iron dioxygenase by resolution of steps following the FeIV=O species. *Proc. Natl. Acad. Sci.* **107**, 3982–3987 (2010).

160. Riggs-Gelasco, P. J. *et al.* EXAFS spectroscopic evidence for an Fe=O unit in the Fe(IV) intermediate observed during oxygen activation by taurine:α-ketoglutarate dioxygenase. *J. Am. Chem. Soc.* **126**, 8108–8109 (2004).

161. Markolovic, S. *et al.* Structure-function relationships of human JmjC oxygenases - demethylases versus hydroxylases. *Current Opinion in Structural Biology* vol. 41 62–72 (2016).

162. 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.* **107**, 9671–9676 (2010).

163. Del Rizzo, P. A., Krishnan, S. & Trievel, R. C. Crystal Structure and Functional Analysis of JMJD5 Indicate an Alternate Specificity and Function. *Mol. Cell. Biol.* **32**, 4044–4052 (2012).

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

165. Lu, T. *et al.* Regulation of NF- B by NSD1/FBXL11-dependent reversible lysine methylation of p65. *Proc. Natl. Acad. Sci.* **107**, 46–51 (2009).

166. Torres, I. O. *et al.* Histone demethylase KDM5A is regulated by its reader domain through a positive-feedback mechanism. *Nat. Commun.* **6**, 6204 (2015).

167. 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).

168. Tu, S. *et al.* The ARID domain of the H3K4 demethylase RBP2 binds to a DNA CCGCCC motif. *Nat. Struct. Mol. Biol.* **15**, 419–421 (2008).

169. Blackledge, N. P. *et al.* CpG Islands Recruit a Histone H3 Lysine 36 Demethylase. *Mol. Cell* **38**, 179–190 (2010).

170. Lin, C. *et al.* Recent advances in the ARID family: Focusing on roles in human cancer. *Onco. Targets. Ther.* **7**, 315–324 (2014).

171. 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**, 2631–2646 (2015).

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

173. Wang, J. *et al.* LSD1n is an H4K20 demethylase regulating memory formation via transcriptional elongation control. *Nat. Neurosci.* **18**, 1256–1264 (2015).

174. Ciccone, D. N. *et al.* KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* **461**, 415–418 (2009).

175. Frescas, D., Guardavaccaro, D., Bassermann, F., Koyama-Nasu, R. & Pagano, M. JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes. *Nature* **450**, 309–313 (2007).

176. He, J., Kallin, E. M., Tsukada, Y. I. & Zhang, Y. The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15Ink4b. *Nat. Struct. Mol. Biol.* **15**, 1169–1175 (2008).

177. Kang, J. Y. *et al.* KDM2B is a histone H3K79 demethylase and induces transcriptional repression via sirtuin-1–mediated chromatin silencing. *FASEB J.* **32**, 5737–5750 (2018).

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

179. Kim, J.-Y. *et al.* KDM3B Is the H3K9 Demethylase Involved in Transcriptional Activation of lmo2 in Leukemia. *Mol. Cell. Biol.* **32**, 2917–2933 (2012).

180. Klose, R. J. *et al.* The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* **442**, 312–316 (2006).

181. Walport, L. J. *et al.* Mechanistic and structural studies of KDM-catalysed demethylation of histone 1 isotype 4 at lysine 26. *FEBS Lett.* **592**, 3264–3273 (2018).

182. Whetstine, J. R. *et al.* Reversal of Histone Lysine Trimethylation by the JMJD2 Family of Histone Demethylases. *Cell* **125**, 467–481 (2006).

183. Fodor, B. D. *et al.* Jmjd2b antagonizes H3K9 trimethylation at pericentric heterochromatin in mammalian cells. *Genes Dev.* **20**, 1557–1562 (2006).

184. Trojer, P. *et al.* Dynamic histone H1 isotype 4 methylation and demethylation by histone lysine methyltransferase G9a/KMT1C and the jumonji domain-containing JMJD2/KDM4 proteins. *J. Biol. Chem.* **284**, 8395–8405 (2009).

185. Wissmann, M. *et al.* Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nat. Cell Biol.* **9**, 347–353 (2007).

186. Shin, S. & Janknecht, R. Diversity within the JMJD2 histone demethylase family. *Biochem. Biophys. Res. Commun.* **353**, 973–977 (2007).

187. Rose, N. R. *et al.* Inhibitor scaffolds for 2-oxoglutarate-dependent histone lysine demethylases. *J. Med. Chem.* **51**, 7053–7056 (2008).

188. Katoh, M. & Katoh, M. Identification and characterization of JMJD2 family genes in silico. *Int. J. Oncol.* **24**, 1623–1628 (2004).

189. Klose, R. J. *et al.* The Retinoblastoma Binding Protein RBP2 Is an H3K4 Demethylase. *Cell* **128**, 889–900 (2007).

190. Xiang, Y. *et al.* JARID1B is a histone H3 lysine 4 demethylase up-regulated in prostate cancer.

Proc. Natl. Acad. Sci. **104**, 19226–19231 (2007).

191. Iwase, S. *et al.* The X-Linked Mental Retardation Gene SMCX/JARID1C Defines a Family of Histone H3 Lysine 4 Demethylases. *Cell* **128**, 1077–1088 (2007).

192. Lee, M. G., Norman, J., Shilatifard, A. & Shiekhattar, R. Physical and Functional Association of a Trimethyl H3K4 Demethylase and Ring6a/MBLR, a Polycomb-like Protein. *Cell* **128**, 877–887 (2007).

193. Hong, S. *et al.* Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18439–18444 (2007).

194. Walport, L. J. *et al.* Human UTY(KDM6C) is a male-specific Nϵ-methyl lysyl demethylase. *J. Biol. Chem.* **289**, 18302–18313 (2014).

195. Tsukada, Y. I., Ishitani, T. & Nakayama, K. I. KDM7 is a dual demethylase for histone H3 Lys 9 and Lys 27 and functions in brain development. *Genes Dev.* **24**, 432–437 (2010).

196. Kleine-Kohlbrecher, D. *et al.* A Functional Link between the Histone Demethylase PHF8 and the Transcription Factor ZNF711 in X-Linked Mental Retardation. *Mol. Cell* **38**, 165–178 (2010).

197. Liu, W. *et al.* PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* **466**, 508–512 (2010).

198. Wen, H. *et al.* Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. *J. Biol. Chem.* **285**, 9322–9326 (2010).

199. Wagner, K. W. *et al.* KDM2A promotes lung tumorigenesis by epigenetically enhancing ERK1/2 signaling. *J. Clin. Invest.* **123**, 5231–5246 (2013).

200. Kottakis, F. *et al.* FGF-2 Regulates Cell Proliferation, Migration, and Angiogenesis through an NDY1/KDM2B-miR-101-EZH2 Pathway. *Mol. Cell* **43**, 285–298 (2011).

201. He, J., Nguyen, A. T. & Zhang, Y. KDM2b/JHDM1b, an H3K36me2-specific demethylase, is required for initiation and maintenance of acute myeloid leukemia. *Blood* **117**, 3869–3880 (2011).

202. Ramadoss, S. *et al.* Lysine-specific demethylase KDM3A regulates ovarian cancer stemness and chemoresistance. *Oncogene* **36**, 1537–1545 (2017).

203. Qi, J. *et al.* Siah2-Dependent Concerted Activity of HIF and FoxA2 Regulates Formation of Neuroendocrine Phenotype and Neuroendocrine Prostate Tumors. *Cancer Cell* **18**, 23–38 (2010).

204. Uemura, M. *et al.* Jumonji domain containing 1A is a novel prognostic marker for colorectal cancer: In vivo identification from hypoxic tumor cells. *Clin. Cancer Res.* **16**, 4636–4646 (2010).

205. Sáez, M. A. *et al.* Mutations in JMJD1C are involved in Rett syndrome and intellectual disability. *Genet. Med.* **18**, 378–385 (2016).

206. Italiano, A. *et al.* Molecular cytogenetic characterization of a metastatic lung sarcomatoid carcinoma: 9p23 neocentromere and 9p23∼p24 amplification including JAK2 and JMJD2C. *Cancer Genet. Cytogenet.* **167**, 122–130 (2006).

207. Patani, N., Jiang, W. G., Newbold, R. F. & Mokbel, K. Histone-modifier gene expression profiles are associated with pathological and clinical outcomes in human breast cancer. *Anticancer Res.* **31**, 4115–4126 (2011).

208. Liu, G. *et al.* Genomic amplification and oncogenic properties of the GASC1 histone

demethylase gene in breast cancer. *Oncogene* **28**, 4491–4500 (2009).

209. Ehrbrecht, A. *et al.* Comprehensive genomic analysis of desmoplastic medulloblastomas: Identification of novel amplified genes and separate evaluation of the different histological components. *J. Pathol.* **208**, 554–563 (2006).

210. Yang, Z. Q. *et al.* Identification of a novel gene, GASC1, within an amplicon at 9p23-24 frequently detected in esophageal cancer cell lines. *Cancer Res.* **60**, 4735–4739 (2000).

211. Vogt, T. *et al.* Deficiency of a novel retinoblastoma binding protein 2-homolog is a consistent feature of sporadic human melanoma skin cancer. *Lab. Investig.* **79**, 1615–1627 (1999).

212. Lu, P. J. *et al.* A novel gene (PLU-1) containing highly conserved putative DNA/chromatin binding motifs is specifically up-regulated in breast cancer. *J. Biol. Chem*. **274**, 15633–15645 (1999).

213. Hayami, S. *et al.* Overexpression of the JmjC histone demethylase KDM5B in human carcinogenesis: involvement in the proliferation of cancer cells through the E2F/RB pathway. *Mol. Cancer* **9**, 59 (2010).

214. Abidi, F. E. *et al.* Mutations in JARID1C are associated with X-linked mental retardation, short stature and hyperreflexia. *J. Med. Genet.* **45**, 787–793 (2008).

215. Santos-Rebouças, C. B. *et al.* A novel nonsense mutation in KDM5C/JARID1C gene causing intellectual disability, short stature and speech delay. *Neurosci. Lett.* **498**, 67–71 (2011).

216. Dalgliesh, G. L. *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360–363 (2010).

217. Perinchery, G. *et al.* Deletion of Y-chromosome specific genes in human prostate cancer. *J. Urol.* **163**, 1339–1342 (2000).

218. Jankowska, A. M. *et al.* Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A. *Blood* **118**, 3932–3941 (2011).

219. Van Haaften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat. Genet.* **41**, 521–523 (2009).

220. Agger, K. *et al.* The H3K27me3 demethylase JMJD3 contributes to the activation of the INK4A-ARF locus in response to oncogene- and stress-induced senescence. *Genes Dev.* **23**, 1171–1176 (2009).

221. Yamamoto, K. *et al.* Loss of histone demethylase KDM6B enhances aggressiveness of pancreatic cancer through downregulation of C/EBPα. *Carcinogenesis* **35**, 2404–2414 (2014).

222. Bloomer, L. D. S. *et al.* Male-Specific Region of the Y Chromosome and Cardiovascular Risk. *Arterioscler. Thromb. Vasc. Biol.* **33**, 1722–1727 (2013).

223. Lee, K.-H. *et al.* Histone demethylase KDM7A controls androgen receptor activity and tumor growth in prostate cancer. *Int. J. Cancer* **143**, 2849–2861 (2018).

224. Laumonnier, F. *et al.* Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. *J. Med. Genet.* **42**, 780–786 (2005).

225. Sinha, S. *et al.* Alterations in candidate genes PHF2, FANCC, PTCH1 and XPA at chromosomal 9q22.3 region: Pathological significance in early- and late-onset breast carcinoma. *Mol. Cancer* **7**, 84 (2008).

226. Ghosh, A. *et al.* Association of FANCC and PTCH1 with the development of early dysplastic lesions of the head and neck. *Ann. Surg. Oncol.* **19**, 528–538 (2012).

227. Hopkinson, R. J. *et al.* 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem. Sci.* **4**, 3110–3117 (2013).

228. Tumber, A. *et al.* Potent and Selective KDM5 Inhibitor Stops Cellular Demethylation of H3K4me3 at Transcription Start Sites and Proliferation of MM1S Myeloma Cells. *Cell Chem. Biol.* **24**, 371–380 (2017).

229. Kruidenier, L. *et al.* A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* **488**, 404–408 (2012).

230. Heinemann, B. *et al.* Inhibition of demethylases by GSK-J1/J4. *Nature* **514**, E1–E2 (2014).

231. Pieters, B., Belle, R. & Mecinovič, J. The effect of the length of histone H3K4me3 on recognition by reader proteins. *ChemBioChem* **14**, 2408–2412 (2013).

232. Woon, E. C. Y. *et al.* Linking of 2-oxoglutarate and substrate binding sites enables potent and highly selective inhibition of JmjC histone demethylases. *Angew. Chemie - Int. Ed.* **51**, 1631–1634 (2012).

233. Kawamura, A. *et al.* Highly selective inhibition of histone demethylases by de novo macrocyclic peptides. *Nat. Commun.* **8**, 14773 (2017).

234. Shi, X. *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**, 96–99 (2006).

235. van Ingen, H. *et al.* Structural Insight into the Recognition of the H3K4me3 Mark by the TFIID Subunit TAF3. *Structure* **16**, 1245–1256 (2008).

236. Peña, P. V. *et al.* Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**, 100–103 (2006).

237. Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–2083 (2002).

238. Lusser, A., Urwin, D. L. & Kadonaga, J. T. Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* **12**, 160–166 (2005).

239. Botuyan, M. V. *et al.* Structural Basis for the Methylation State-Specific Recognition of Histone H4-K20 by 53BP1 and Crb2 in DNA Repair. *Cell* **127**, 1361–1373 (2006).

240. Koga, H. *et al.* A human homolog of Drosophila lethal(3)malignant brain tumor (l(3)mbt) protein associates with condensed mitotic chromosomes. *Oncogene* **18**, 3799–3809 (1999).

241. Ma, J. C. & Dougherty, D. A. The cation-π interaction. *Chem. Rev.* **97**, 1303–1324 (1997).

242. Salonen, L. M., Ellermann, M. & Diederich, F. Aromatic rings in chemical and biological recognition: energetics and structures. *Angew. Chem. Int. Ed. Engl.* **50**, 4808–42 (2011).

243. Hughes, R. M., Wiggins, K. R., Khorasanizadeh, S. & Waters, M. L. Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11184–11188 (2007).

244. Kamps, J. J. A. G. *et al.* Chemical basis for the recognition of trimethyllysine by epigenetic reader proteins. *Nat. Commun.* **6**, 8911 (2015).

245. Richart, L., Real, F. X. & Sanchez-Arevalo Lobo, V. J. c-MYC partners with BPTF in human cancer. *Mol. Cell. Oncol.* **3**, e1152346 (2016).

246. Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* **14**, 341–356 (2013).

247. Song, C. X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic

priming. *Cell* **153**, 678–691 (2013).

248. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).

249. Laukka, T. *et al.* Fumarate and succinate regulate expression of hypoxia-inducible genes via TET enzymes. *J. Biol. Chem.* **291**, 4256–4265 (2016).

250. Szulwach, K. E. *et al.* 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* **14**, 1607–1616 (2011).

251. Shoichet, B. K. Interpreting steep dose-response curves in early inhibitor discovery. *J. Med. Chem.* **49**, 7274–7277 (2006).

252. Ohuchi, M., Murakami, H. & Suga, H. The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. *Curr. Opin. Chem. Biol.* **11**, 537–542 (2007).

253. Song, C. X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29**, 68–75 (2011).

254. Yeh, T. L. *et al.* Molecular and cellular mechanisms of HIF prolyl hydroxylase inhibitors in clinical trials. *Chem. Sci.* **8**, 7651–7668 (2017).

255. Hirsilä, M., Koivunen, P., Günzler, V., Kivirikko, K. I. & Myllyharju, J. Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *J. Biol. Chem.* **278**, 30772–30780 (2003).

256. Pergola, P. E., Spinowitz, B. S., Hartman, C. S., Maroni, B. J. & Haase, V. H. Vadadustat, a novel oral HIF stabilizer, provides effective anemia treatment in nondialysis-dependent chronic kidney disease. *Kidney Int.* **90**, 1115–1122 (2016).

257. Jin, C. *et al.* Chem-seq permits identification of genomic targets of drugs against androgen receptor regulation selected by functional phenotypic screens. *Proc. Natl. Acad. Sci.* **111**, 9235–9240 (2014).

258. Metzger, E. *et al.* KDM4 inhibition targets breast cancer stem–like cells. *Cancer Res.* **77**, 5900–5912 (2017).

259. Vinogradova, M. *et al.* An inhibitor of KDM5 demethylases reduces survival of drug-tolerant cancer cells. *Nat. Chem. Biol.* **12**, 531–538 (2016).

260. Wang, L. *et al.* A small molecule modulates Jumonji histone demethylase activity and selectively inhibits cancer growth. *Nat. Commun.* **4**, 2035 (2013).

261. Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447**, 407–412 (2007).

262. Rao, R. C. & Dou, Y. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat. Rev. Cancer* **15**, 334–346 (2015).

263. Subramanian, K. *et al.* Regulation of Estrogen Receptor α by the SET7 Lysine Methyltransferase. *Mol. Cell* **30**, 336–347 (2008).

264. Flanagan, J. F. *et al.* Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438**, 1181–1185 (2005).

265. Christensen, J. *et al.* RBP2 Belongs to a Family of Demethylases, Specific for Tri-and Dimethylated Lysine 4 on Histone 3. *Cell* **128**, 1063–1076 (2007).

266. Yamane, K. *et al.* PLU-1 Is an H3K4 Demethylase Involved in Transcriptional Repression and Breast Cancer Cell Proliferation. *Mol. Cell* **25**, 801–812 (2007).

267. Huang, F. *et al.* The JmjN domain of Jhd2 is important for its protein stability, and the plant homeodomain (PHD) finger mediates its chromatin association independent of H3K4 methylation. *J. Biol. Chem.* **285**, 24548–24561 (2010).

268. Horton, J. R. *et al.* Structural Basis for KDM5A Histone Lysine Demethylase Inhibition by Diverse Compounds. *Cell Chem. Biol.* **23**, 769–781 (2016).

269. Balciunas, D. & Ronne, H. Evidence of domain swapping within the jumonji family of transcription factors. *Trends Biochem. Sci.* **25**, 274–276 (2000).

270. Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* **29**, 2860–2874 (2001).

271. Scibetta, A. G. *et al.* Functional Analysis of the Transcription Repressor PLU-1/JARID1B. *Mol. Cell. Biol.* **27**, 7220–7235 (2007).

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

273. Longbotham, J. E. *et al.* Histone H3 binding to the PHD1 domain of histone demethylase KDM5A enables active site remodeling. *Nat. Commun.* **10**, 94 (2019).

274. Rasmussen, P. B. & Staller, P. The KDM5 family of histone demethylases as targets in oncology drug discovery. *Epigenomics* **6**, 277–286 (2014).

275. Banelli, B. *et al.* The histone demethylase KDM5A is a key factor for the resistance to temozolomide in glioblastoma. *Cell Cycle* **14**, 3418–3429 (2015).

276. Teng, Y. C. *et al.* Histone demethylase RBP2 promotes lung tumorigenesis and cancer metastasis. *Cancer Res.* **73**, 4711–4721 (2013).

277. Gildea, J. J., Lopez, R. & Shearn, A. A screen for new trithorax group genes identified little imaginal discs, the Drosophila melanogaster homologue of human retinoblastoma binding protein 2. *Genetics* **156**, 645–663 (2000).

278. Jensen, L. R. *et al.* Mutations in the JARID1C Gene, Which Is Involved in Transcriptional Regulation and Chromatin Remodeling, Cause X-Linked Mental Retardation. *Am. J. Hum. Genet.* **76**, 227–236 (2005).

279. Santos, C. *et al.* A novel mutation in JARID1C gene associated with mental retardation. *Eur. J. Hum. Genet.* **14**, 583–586 (2006).

280. Wagner, E. K., Nath, N., Flemming, R., Feltenberger, J. B. & Denu, J. M. Identification and Characterization of Small Molecule Inhibitors of a Plant Homeodomain Finger. *Biochemistry* **51**, 8293–8306 (2012).

281. Bhushan, B. *et al.* Investigations on small molecule inhibitors targeting the histone H3K4 tri-methyllysine binding PHD-finger of JmjC histone demethylases. *Bioorganic Med. Chem.* **26**, 2984–2991 (2018).

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

283. Hopkinson, R. J. *et al.* Is JmjC oxygenase catalysis limited to demethylation? *Angew. Chem. Int. Ed. Engl.* **52**, 7709–7713 (2013).

284. Hopkinson, R. J. *et al.* Human histone demethylase KDM6B can catalyse sequential oxidations. *Chem. Commun.* **54**, 7975–7978 (2018).

285. Zhang, Y. *et al.* The PHD1 finger of KDM5B recognizes unmodified H3K4 during the

demethylation of histone H3K4me2/3 by KDM5B. *Protein Cell* **5**, 837–850 (2014).

286. Usuki, T., Sugimura, T., Komatsu, A. & Koseki, Y. Biomimetic chichibabin pyridine synthesis of the COPD biomarkers and elastin cross-linkers isodesmosine and desmosine. *Org. Lett.* **16**, 1672–1675 (2014).

287. Jelesarov, I. & Bosshard, H. R. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit.* **12**, 3–18 (1999).

288. Vaughn, J. L., Goodwin, R. H., Tompkins, G. J. & McCawley, P. The establishment of two cell lines from the insect spodoptera frugiperda (lepidoptera; noctuidae). *In Vitro* **13**, 213–217 (1977).

289. Chance, B., Schoener, B., Oshino, R., Itshak, F. & Nakase, Y. Oxidation-Reduction Ratio Studies of Mitochondria in Freeze-trapped Samples. *J. Biol. Chem.* **254**, 4764–4771 (1979).

290. Kerr, R. G. & Kelly, K. An enzyme-based formaldehyde assay and its utility in a sponge sterol biosynthetic pathway. *J. Nat. Prod.* **62**, 201–202 (1999).

291. Couture, J.-F., Collazo, E., Ortiz-Tello, P. A., Brunzelle, J. S. & Trievel, R. C. Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat. Struct. Mol. Biol.* **14**, 689–695 (2007).

292. Johansson, C. *et al.* Structural analysis of human KDM5B guides histone demethylase inhibitor development. *Nat. Chem. Biol.* **12**, 539–545 (2016).

293. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

294. Albà, M. M. & Castresana, J. On homology searches by protein Blast and the characterization of the age of genes. *BMC Evol. Biol.* **7**, 53 (2007).

295. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

296. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539–539 (2014).

297. Hung, T. *et al.* ING4 Mediates Crosstalk between Histone H3 K4 Trimethylation and H3 Acetylation to Attenuate Cellular Transformation. *Mol. Cell* **33**, 248–256 (2009).

298. 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**, 785–796 (2006).

299. Ramon-Maiques, S. *et al.* The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proc. Natl. Acad. Sci.* **104**, 18993–18998 (2007).

300. Bramucci, E., Paiardini, A., Bossa, F. & Pascarella, S. PyMod: sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL. *BMC Bioinformatics* **13**, S2 (2012).

301. Johansson, C. *et al.* The roles of Jumonji-type oxygenases in human disease. *Epigenomics* **6**, 89–120 (2014).

302. Baumann, M. & Baxendale, I. R. An overview of the synthetic routes to the best selling drugs containing 6-membered heterocycles. *Beilstein J. Org. Chem.* **9**, 2265–2319 (2013).

303. Spannhoff, A., Hauser, A. T., Heinke, R., Sippl, W. & Jung, M. The emerging therapeutic potential of histone methyltransferase and demethylase inhibitors. *ChemMedChem* vol. 4 1568–

1582 (2009).

304. Gough, S. M. *et al.* NUP98-PHF23 Is a Chromatin-Modifying Oncoprotein That Causes a Wide Array of Leukemias Sensitive to Inhibition of PHD Histone Reader Function. *Cancer Discov.* **4**, 564–577 (2014).

305. Roussy, M. *et al. NUP98-BPTF* gene fusion identified in primary refractory acute megakaryoblastic leukemia of infancy. *Genes, Chromosom. Cancer* **57**, 311–319 (2018).

306. Rose, N. R. *et al.* Plant growth regulator daminozide is a selective inhibitor of human KDM2/7 histone demethylases. *J. Med. Chem.* **55**, 6639–6643 (2012).

307. Yap, K. L. & Zhou, M.-M. Keeping it in the family: diverse histone recognition by conserved structural folds. *Crit. Rev. Biochem. Mol. Biol.* **45**, 488–505 (2010).

308. Org, T. *et al.* The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO Rep.* **9**, 370–376 (2008).

309. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).

310. Miller, T. C. R. *et al.* Competitive Binding of a Benzimidazole to the Histone-Binding Pocket of the Pygo PHD Finger. *ACS Chem. Biol.* **9**, 2864–2874 (2014).

311. Ali, F., Yamaguchi, K., Fukuoka, M., Elhelaly, A. E. & Kuwata, K. Logical design of an anti-cancer agent targeting the plant homeodomain in Pygopus2. *Cancer Sci.* **107**, 1321–1328 (2016).

312. Amato, A., Lucas, X., Bortoluzzi, A., Wright, D. & Ciulli, A. Targeting Ligandable Pockets on Plant Homeodomain (PHD) Zinc Finger Domains by a Fragment-Based Approach. *ACS Chem. Biol.* **13**, 915–921 (2018).

313. Lee, J., Thompson, J. R., Botuyan, M. V. & Mer, G. Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nat. Struct. Mol. Biol.* **15**, 109–111 (2008).

314. Pieters, B. J. G. E., Meulenbroeks, E., Belle, R., Mecinović, J. & Jeltsch, A. The role of electrostatic interactions in binding of histone H3K4me2/3 to the Sgf29 tandem Tudor Domain. *PLoS One* **10**, (2015).

315. Shetty, D. *et al.* Synthesis and evaluation of macrocyclic amino acid derivatives for tumor imaging by gallium-68 positron emission tomography. *Bioorg. Med. Chem.* **18**, 7338–7347 (2010).

316. Qin, S. *et al.* Structural basis for histone mimicry and hijacking of host proteins by influenza virus protein NS1. *Nat. Commun.* **5**, 3952 (2014).

317. Zhao, S. *et al.* Kinetic and high-throughput profiling of epigenetic interactions by 3D-carbene chip-based surface plasmon resonance imaging technology. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E7245–E7254 (2017).

318. Yang, N. *et al.* Distinct mode of methylated lysine-4 of histone H3 recognition by tandem tudor-like domains of Spindlin1. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17954–9 (2012).

319. Li, H. *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91–95 (2006).

320. Zhang, Y. Investigating Substrate Selectivity and Inhibitor of the Human KDM7 Subfamily of JmjC Histone Lysine Demethylases. (St Cross College, University of Oxford, 2018).

321. Vermeulen, M. *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone

H3 lysine 4. *Cell* **131**, 58–69 (2007).

322. Bian, C. *et al.* Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO J.* **30**, 2829–2842 (2011).

323. Bada, J. L. [9] In Vivo racemization in mammalian proteins. *Methods Enzymol.* **106**, 98–115 (1984).

324. Csapó, J., Varga-Visi, É., Lóki, K. & Albert, C. The influence of manufacture on the free D-amino acid content of Cheddar cheese. *Amino Acids* **32**, 39–43 (2007).

325. E Chiavaro, A Caligani, G. P. Chiral indicators of aging in balsamic vinegars of Modena. *Ital J Food Sci* **10**, 327–329 (1998).

326. Watababe, A. *et al.* Role of lysine 39 of alanine racemase from Bacillus stearothermophilus that binds pyridoxal 5'-phosphate: Chemical rescue studies of Lys39 → Ala mutant. *J. Biol. Chem.* **274**, 4189–4194 (1999).

327. Wolosker, H. *et al.* Purification of serine racemase: Biosynthesis of the neuromodulator D-serine. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 721–725 (1999).

328. Schell, M. J., Brady, R. O., Molliver, M. E. & Snyder, S. H. D-serine as a neuromodulator: Regional and developmental localizations in rat brain glia resemble NMDA receptors. *J. Neurosci.* **17**, 1604–1615 (1997).

329. Takigawa, Y. *et al.* D-aspartate uptake into cultured rat pinealocytes and the concomitant effect on L-aspartate levels and melatonin secretion. *Biochem. Biophys. Res. Commun.* **248**, 641–647 (1998).

330. Yamada, R., Nagasaki, H., Nagata, Y., Wakabayashi, Y. & Iwashima, A. Administration of D-aspartate increases D-aspartate oxidase activity in mouse liver. *Biochim. Biophys. Acta - Gen. Subj.* **990**, 325–328 (1989).

331. Long, Z. *et al.* d-Aspartate in a Prolactin-Secreting Clonal Strain of Rat Pituitary Tumor Cells (GH3). *Biochem. Biophys. Res. Commun.* **276**, 1143–1147 (2000).

332. Sakai, K. *et al.* Emergence of d-aspartic acid in the differentiating neurons of the rat central nervous system. *Brain Res.* **808**, 65–71 (1998).

333. Hashimoto, A., Oka, T. & Nishikawa, T. Anatomical Distribution and Postnatal Changes in Endogenous Free D-Aspartate and D-Serine in Rat Brain and Periphery. *Eur. J. Neurosci.* **7**, 1657–1663 (1995).

334. Fujii, N. D-amino acid in elderly tissues. *Biol. Pharm. Bull.* **28**, 1585–1589 (2005).

335. Powell, J. T., Vine, N. & Crossman, M. On the accumulation of d-aspartate in elastin and other proteins of the ageing aorta. *Atherosclerosis* **97**, 201–208 (1992).

336. Shapiro, S. D., Endicott, S. K., Province, M. A., Pierce, J. A. & Campbell, E. J. Marked longevity of human lung parenchymal elastic fibers deduced from prevalence of D-aspartate and nuclear weapons-related radiocarbon. *J. Clin. Invest.* **87**, 1828–1834 (1991).

337. Fujii, N. *et al.* The presence of d-β-aspartic acid-containing peptides in elastic fibers of sun-damaged skin: a potent marker for ultraviolet-induced skin aging. *Biochem. Biophys. Res. Commun.* **294**, 1047–1051 (2002).

338. Ohtani, S., Yamamoto, T., Matsushima, Y. & Kobayashi, Y. Changes in the amount of D-aspartic acid in the human femur with age. *Growth, Dev. Aging* **62**, 141–148 (1998).

339. Structural alterations in the peptide backbone of beta-amyloid core protein may account for its deposition and stability in Alzheimer's disease. *ASBMB*.

340. Young, G. W. *et al.* Protein L-isoaspartyl methyltransferase catalyzes in vivo racemization of aspartate-25 in mammalian histone H2B. *J. Biol. Chem.* **280**, 26094–26098 (2005).

341. Qin, Z., Zhu, J. X. & Aswad, D. W. The d-isoAsp-25 variant of histone H2B is highly enriched in active chromatin: Potential role in the regulation of gene expression? *Amino Acids* **48**, 599–603 (2016).

342. Belle, R. *et al.* Investigating d-lysine stereochemistry for epigenetic methylation, demethylation and recognition. *Chem. Commun.* **53**, 13264–13267 (2017).

343. Poplawski, A. *et al.* Molecular Insights into the Recognition of N-Terminal Histone Modifications by the BRPF1 Bromodomain. *J. Mol. Biol.* **426**, 1661–1676 (2014).

344. Peng, Y. & Alexov, E. Cofactors-loaded quaternary structure of lysine-specific demethylase 5C (KDM5C) protein: Computational model. *Proteins: Struct. Funct. Bioinf.* **84**, 1797–1809 (2016).

345. Case, D. A. *et al. AMBER 12, University of California, San Francisco*. (2012).

346. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

347. Cortopassi, W. A., Simion, R., Honsby, C. E., França, T. C. C. & Paton, R. S. Dioxygen Binding in the Active Site of Histone Demethylase JMJD2A and the Role of the Protein Environment. *Chem. - A Eur. J.* **21**, 18983–18992 (2015).

348. Cortopassi, W. A. *et al.* Cation–π interactions in CREBBP bromodomain inhibition: an electrostatic model for small-molecule binding affinity and selectivity. *Org. Biomol. Chem.* **14**, 10926–10938 (2016).

349. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993).

350. Wang, H., Dommert, F. & Holm, C. Optimizing working parameters of the smooth particle mesh Ewald algorithm in terms of accuracy and efficiency. *J. Chem. Phys.* **133**, 034117 (2010).

351. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

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

353. Xiao, B. *et al.* Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* **421**, 652–656 (2003).

354. Wu, H. *et al.* Structural Biology of Human H3K9 Methyltransferases. *PLoS One* **5**, e8570 (2010).

355. Huang, Z.-P., Du, J.-T., Zhao, Y.-F. & Li, Y.-M. Synthesis of Site-Specifically Dimethylated and Trimethylated Peptides Derived from Histone H3 N-Terminal Tail. *Int. J. Pept. Res. Ther.* **12**, 187–193 (2006).

356. Christman, J. K. 5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: Mechanistic studies and their implications for cancer therapy. *Oncogene* vol. 21 5483–5495 (2002).

357. Asangani, I. A. *et al.* Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature* **510**, 278–282 (2014).

358. Wieneke, R. *et al.* Silica precipitation with synthetic silaffin peptides. *Org. Biomol. Chem.* **9**,

5482–6 (2011).

359. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. in *The Proteomics Protocols Handbook* 571–607 (2005). doi:10.1385/1-59259-890-0:571.

Chapter 9. Appendix

Table 9.1: Domain length of KDM5 proteins as determined using BLAST (NCBI) and Phyre2.

	KDM5A		KDM5B		KDM5C		KDM5D	
	Begin	End	Begin	End	Begin	End	Begin	End
JmjN	18	76	26	94	13	54	13	54
ARID	173	85	94	183	80	169	84	165
PHD1	295	343	311	356	326	372	316	361
JmjC	586	363	374	604	394	617	384	638
Zn	676	742	673	753	707	759	728	780
PLU1	742	1070	753	1086	772	1098	793	1116
PHD2	1163	1215	1179	1224	1188	1247	1205	1265
PHD3	1615	1658	1492	1538				

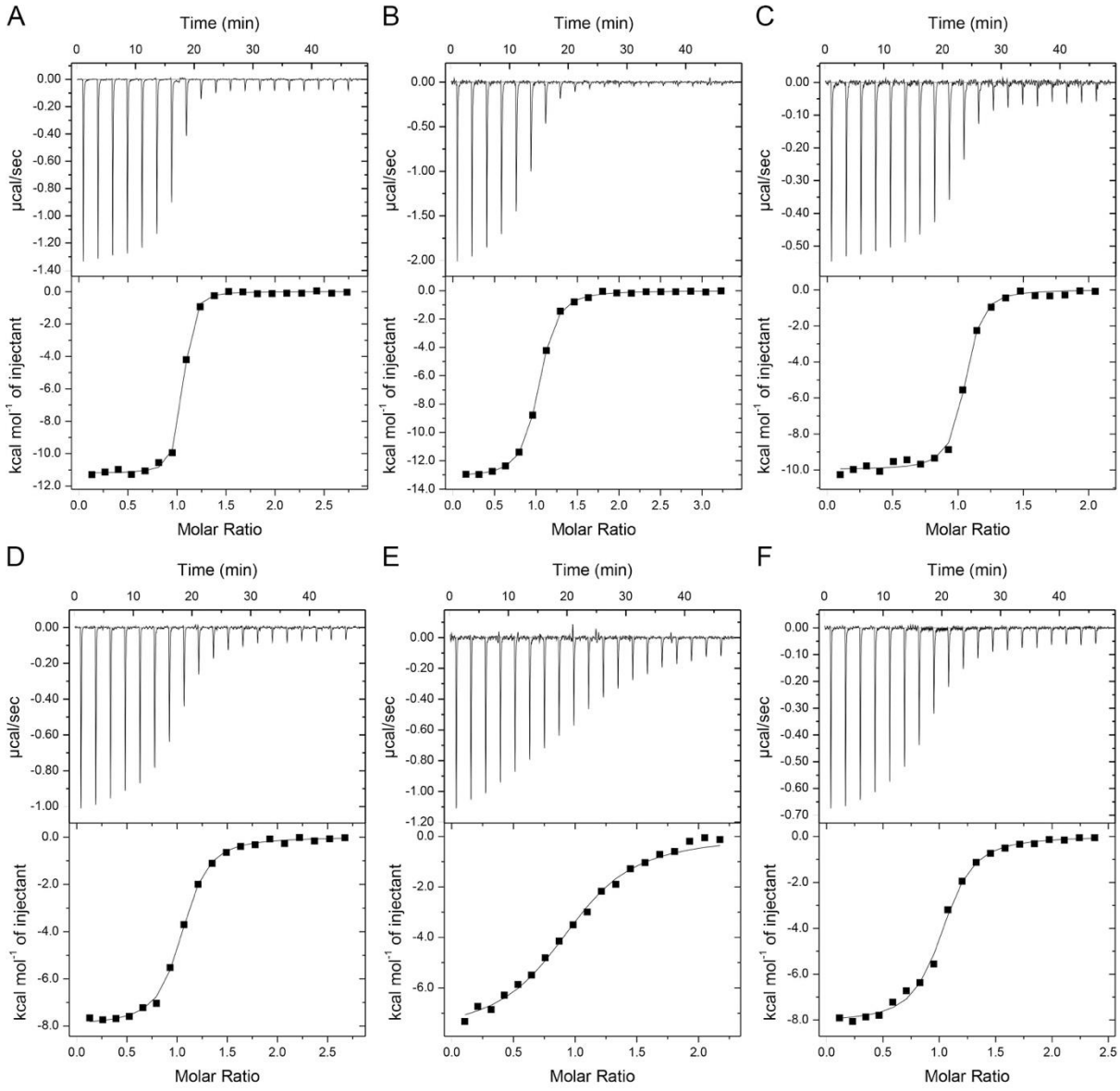


Figure 9.1: ITC binding analysis of KDM5A$_{PHD3}$ (A, D), BPTF$_{PHD}$ (B, E), TAF3$_{PHD}$ (C, F) against H3$_{1-10}$^{L}K4me$_3$ 5.1 (A-C) and H3$_{1-10}$^{D}K4me$_3$ 5.2 (D-F). The results demonstrate a decrease in binding strength for all binding complex containing ^{D}K4me$_3$ relative to ^{L}K4me$_3$.[342] Corresponding values are displayed in Table 5.4.2. This figure is adapted and reproduced with permission from the Royal Society of Chemistry.[342]

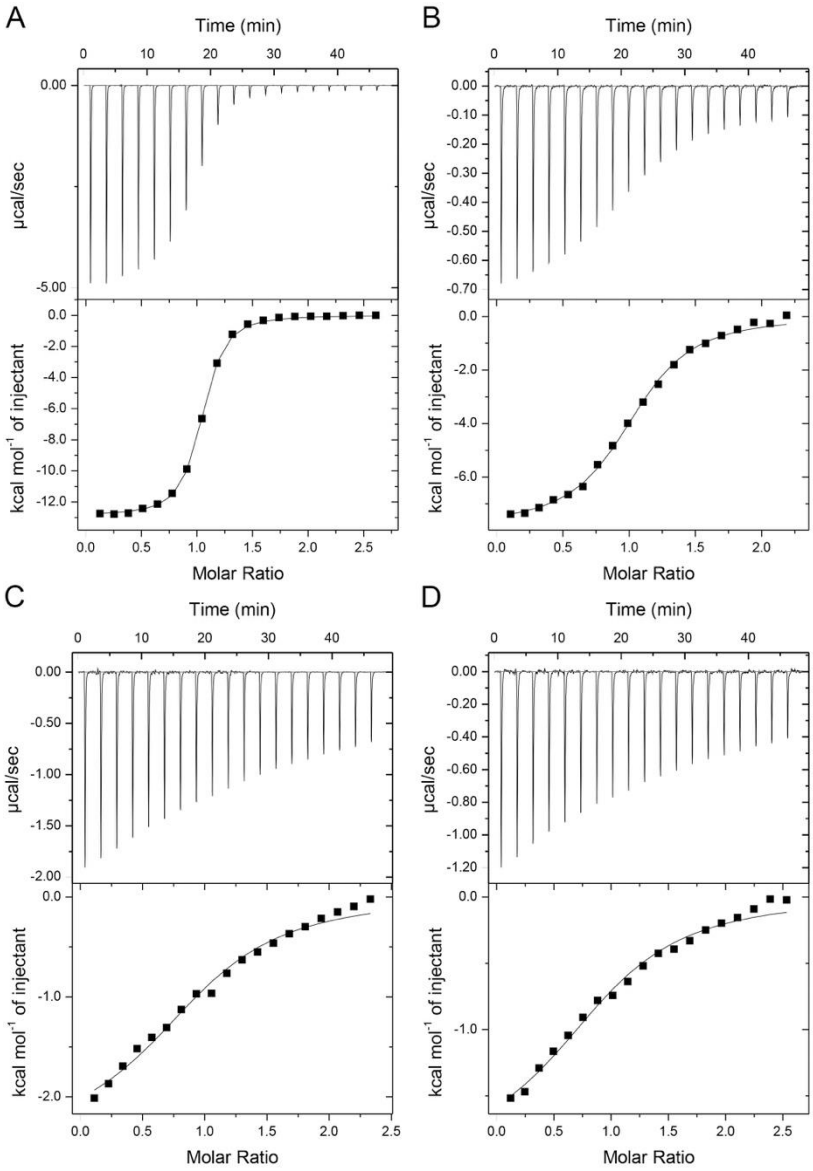


Figure 9.2: ITC binding analysis of KDM4A$_{TTD}$ (A, B) and SGF29$_{TTD}$ (B, D) with H3$_{1-10}$^{L}K4me$_3$ 5.1 (A, B) and H3$_{1-10}$^{D}K4me$_3$ 5.2 (C, D).[342] Corresponding values displayed in Table 5.4.2. This figure is adapted and reproduced with permission from the Royal Society of Chemistry.[342]

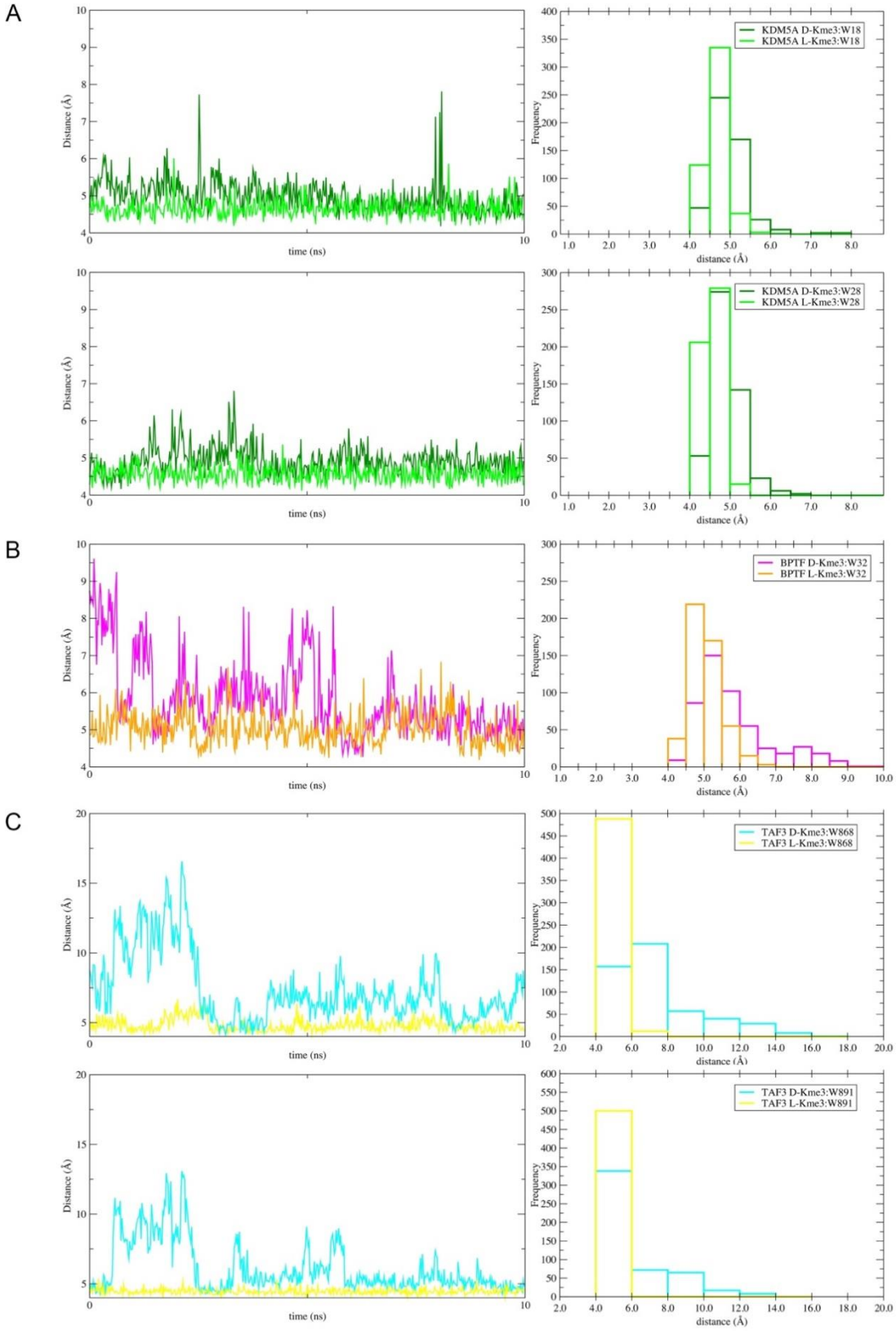


Figure 9.3: Molecular dynamics distance and bar plot of N^{ε}-Lysine relative to the corresponding aromatic residue of the reader aromatic cage. A): W1625 and W1635 of $KDM5A_{PHD3}$ in complex to $^{D}K4me_3$ (dark green) or $^{L}K4me_3$ (light green) (PDB: 2KGI); B) W32 of $BPTF_{PHD}$ in complex to $^{D}K4me_3$ (pink) or $^{L}K4me_3$ (orange) (PDB: 2F6J) and C) W868 and W891 of $TAF3_{PHD}$ in complex to $^{D}K4me_3$ (cyan) or $^{L}K4me_3$ (yellow) (PDB: 2K17). This figure is adapted and reproduced with permission from the Royal Society of Chemistry.[342]

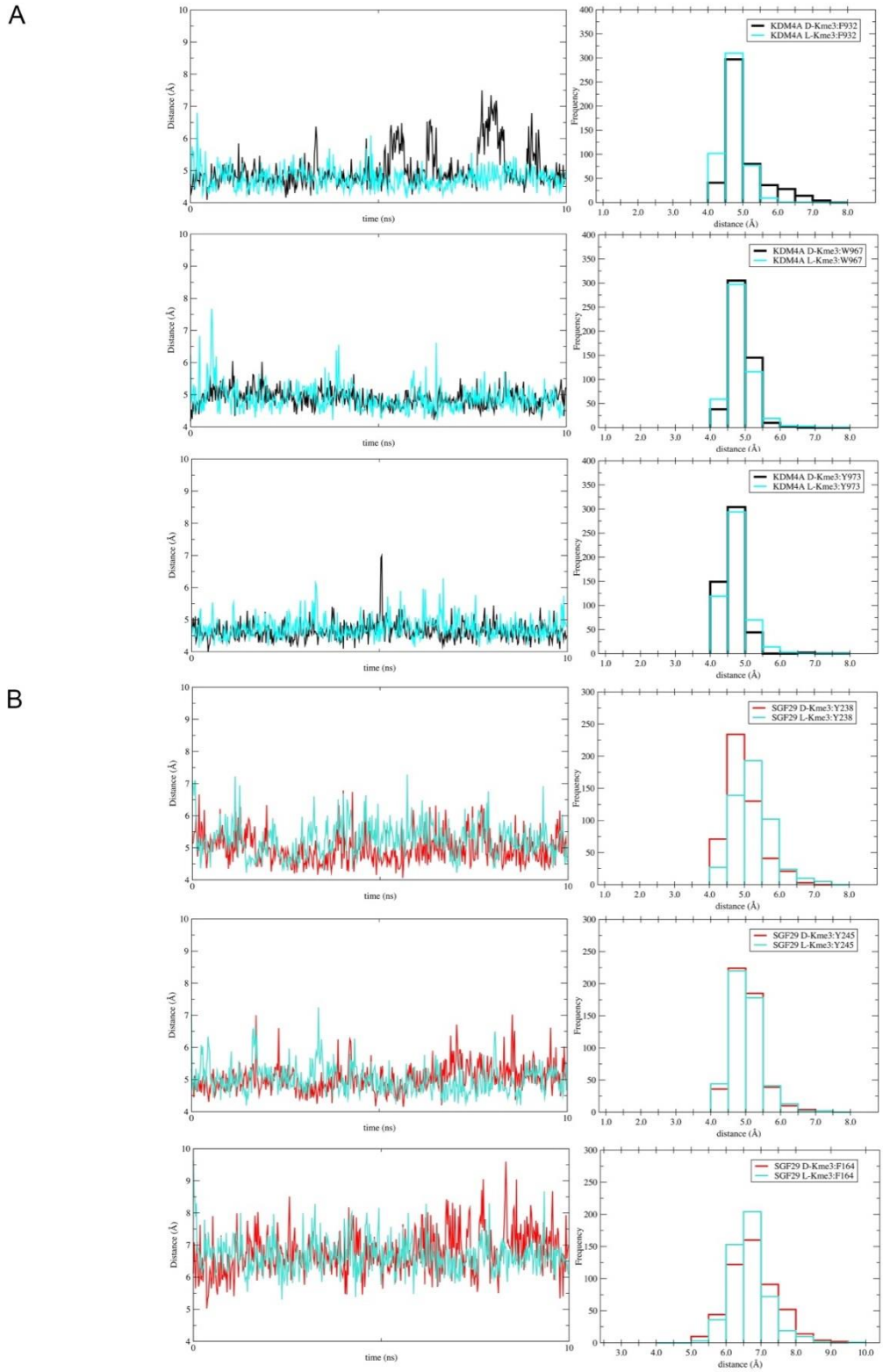


Figure 9.4: MD acquisition of TTD containing proteins displaying distance of *N*$^{\varepsilon}$-lysine towards the aromatic amino acids of the aromatic cage computer for 10 ns (20 ps / frame). A) computation for the KDM4A$_{TTD}$ residues F932, W967, Y973 in complex to ^{D}K4me$_3$ (black) peptide or ^{L}K4me$_3$ (cyan) (PDB: 2GFA) and B) SGF29$_{TTD}$ residues Y238, Y245, F164 in complex to ^{D}K4me$_3$ (black) peptide or ^{L}K4me$_3$ (cyan) (PDB: 2ME9). This figure is adapted and reproduced with permission from the Royal Society of Chemistry.[342]