

Louise Jane Walport

Chemistry Research Laboratory
Department of Chemical Biology
Lady Margaret Hall, University of Oxford

Trinity Term 2013

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

Louise Jane Walport
Lady Margaret Hall

DPhil Thesis
Trinity Term 2013

Structural and functional studies of chromatin modifying enzymes

Epigenetic regulation is a complex process involving the interplay of multiple different cellular factors. Work described in this thesis concerned the characterisation of proteins involved in the binding to, and demethylation of, histone 3 (H3) tails modified by *N*-methylation. Initial work focussed on the biophysical characterisation of the tandem plant homeodomains (PHD) of the chromatin remodeller CHD4. NMR spectroscopy was used to investigate the solution structure of the tandem PHDs. Studies on a more native-like construct including the *C*-terminal tandem chromodomains are also presented. Binding studies of the PHDs with H3 peptides reveal that the individual PHD fingers can independently bind a histone peptide.

The remainder of the work involved characterisation of JmjC histone demethylases (KDMs), enzymes that catalyse removal of N^{ϵ}-methyl groups from histone lysyl-residues. Initially, two members of the KDM7 subfamily, PHF8 and KIAA1718, were studied; a high-throughput screening assay for them was developed, which enabled identification of a selective inhibitor of the KDM2/7 subfamilies of KDMs, the plant growth regulator Daminozide. A disease-relevant mutation in PHF8 was studied and shown to cause mis-localisation of the enzyme to the cytoplasm, providing a potential explanation for the clinically observed phenotype.

Subsequent chapters describe unprecedented activities for the JmjC KDMs. 2OG oxygenases catalyse a wide range of oxidative reactions, predominantly mediated by initial substrate hydroxylation. The activity of PHF8 with lysine analogous was tested; the results demonstrated that PHF8, and other KDMs, can oxidatively remove N^{ϵ}-alkyl groups other than methyl groups, such as ethyl and isopropyl groups. The substrate scope of the JmjC KDMs thus has the potential to be wider than previously thought. Observation of β-hydroxylation of the N^{ϵ}-isopropyl group of a histone peptide including N^{ϵ}-methylisopropyllysine by JMJD2A/E supports the presumed mechanism of histone lysine demethylation as proceeding *via* initial hydroxylation. This work led to the discovery that JmjC KDMs can catalyse arginine demethylation. This novel arginine demethylase activity by JmjC KDMs was characterised and the work extended to encompass potential arginine demethylase activity in cells.

Biochemical characterisation of UTY, a homologue of the H3 K27 demethylases JMJD3 and UTX, which is reported to be inactive, was carried out; UTY was shown to catalyse demethylation at H3 trimethylated at K27 on peptidic substrates, albeit it at substantially lower rates than the other family members. To investigate the reason for this reduced activity, two variants were made, S1142G and P1214I; the latter variant was shown to be considerably more active than wildtype UTY, likely due to an increased peptide-binding interaction. Preliminary experiments in cells did not conclusively demonstrate histone demethylation, but a luciferase assay suggested that UTY may have catalytic activity in cells.

Overall the findings in the thesis suggest that the process of cellular epigenetic regulation is likely even more complex than previously thought, with the potential that JmjC KDMs carry out multiple, context-dependent functions.

Acknowledgments

Firstly I would like to thank my supervisors, Professor Chris Schofield and Professor Christina Redfield for their unfailing enthusiasm, ideas and encouragement. Chris, for allowing me the freedom to follow where my experiments led and for stopping me from getting lost along the way. Christina, for teaching me to think critically, being endlessly patient and for keeping me from despair when things weren't going to plan!

Much work in this thesis was only possible because of experimental help, advice and collaboration from colleagues in the lab and throughout the university. I am especially grateful for the direct experimental assistance provided by Matt Smith, Rok Sekirnik, Nikita Loik, Jacob Bush, David Yadin, Rosa Morra, Tony Tumber, Tobias Krojer, Shirley Li, Catrine Johansson, Akane Kawamura, Wei Ge, Tim Rooney, Rashed Chowdhury and James McCullagh. All specific contributions are acknowledged in the text. Thanks are due in particular to Richard Hopkinson for help, guidance and friendship throughout.

I have enjoyed all of my last four memorable years on the lower ground floor of the CRL and I am grateful to all members of the CJS group, past and present, for their help, suggestions and encouragement around the lab and for great tea-time conversation; especially, but in no particular order to Sarah, Jake, Rok, Sophie, Becky, Cyrille, Duncan, Olly, Eleanor, Akane, Martin, Ben, Nathan and Rich, amongst many others for lots of fun along the way. In addition, my special thanks go to Wendy and Adam for helping to keep the lab running smoothly and to Manso for keeping the fermentation room in order.

DPhils have their highs and lows and I'm grateful to friends and family for keeping me positive throughout. To the banter crew, thanks for sharing those moments with me; the last four years would have been a lot less fun without the banter list, PTCL Fridays and Sunday film club/thesis soup kitchen. To Guy and Vic, thank you for providing a distraction-free Suffolk haven for thesis writing. I am grateful to all my friends and family outside the lab for support and science-free diversions, in particular, to Coral, Alastair, Henry, Ruth, Suvi, Becca, Alex and Clive. Music has played an important part in my time here in Oxford, helping to put the science into perspective, and I am grateful to everyone who has shared this with me.

Most importantly, my heartfelt thanks go to my parents for their endless love, support and encouragement over the last four years and the twenty-two before them and to Robert, Emily and Fiona for sharing the journey with me.

Author Contibutions

All experiments described in this thesis were carried out by myself unless otherwise stated.

At the start of each chapter, a description of contributions from collaborators is summarised in italics and in addition their names are included in parentheses in the main text anywhere their experiments are described.

Table of contents

Acknowledgments ……..…….ii

Author contributions……..iii

Abbreviations ...……...……...xi

Chapter 1 Introduction 1

1.1 Epigenetics..... 1

1.1.1 DNA packaging 2

1.1.1.2 Structure of the histone tails.....4

1.1.1.3 Histone variants.....6

1.1.2 DNA Modifications 7

1.1.3 Post-translational modifications 9

1.1.3.2 Phosphorylation 11

1.1.3.3 Acetylation 12

1.1.3.4 Ubiquitylation.....12

1.1.3.5 Methylation.....13

1.1.4 2OG oxygenases 19

1.1.4.2 JmjC histone demethylases 22

1.1.5 Histone Readers 25

1.1.5.1 Bromodomains.....25

1.1.5.2 Tudor domains 26

1.1.5.3 Chromodomains.....27

1.1.5.4 Plant homeodomains 27

1.1.6 Chromatin Remodelers 28

1.1.7 Roles of epigenetics in disease.....30

1.2 Objectives..... 32

Chapter 2 Determining the structure of the tandem PHD domains of CHD4..... 36

2.1 CHD4..... 36

2.1.2 Objectives 37

2.2 Structure determination 38

2.3 Spectral assignment 40

2.3.2 Backbone assignments 42

2.3.3 Side chain assignments 46

2.4 Dihedral angles 48

2.4.2 The HMQC-J for measuring dihedral angles 49

2.4.3 Use of TALOS-N for dihedral angle predictions 49

2.4.4 Cis/Trans prolines within dPHD 50

2.5 Backbone dynamics 52

2.5.1 Heteronuclear NOE ratio 52

2.5.2 Measuring the heteronuclear NOE ratio of dPHD 53

2.6 dPHD NOE assignment and restraint generation 54

2.6.2 Defining restraints to maintain correct zinc-binding geometry 57

2.7 Simulated annealing calculations 57

2.7.2 Use of swapping to define stereospecific assignments 60

2.8 Refinement 61

2.8.1 Residual Dipolar Couplings 62

2.9 Generation of a final structural ensemble 64

2.10 Structure validation 66

2.11 Discussion 70

Chapter 3 Investigations into the binding interactions of the tandem PHD domains of CHD4..... 74

3.1 Introduction 74

3.1.2 Objectives 76

3.2 Investigating the histone-binding behaviour of dPHD 76

3.2.1 Measuring protein-protein interactions using NMR spectroscopy 76

3.2.2 Investigating the histone binding interactions of dPHD 77

3.2.3 Mapping the histone binding surface of dPHD 81

3.2.3.2 Confirming the binding of two histone peptides by mass spectrometry 85

3.2.4 Investigating the effect of varying the peptide length on dPHD binding 86

3.3 Investigating the interaction of dPHD with dCHD 89

3.4 Investigating the structure of dPHDdCHD 91

3.5 Investigating the binding interactions of dPHDdCHD 95

3.5.1 Investigating the DNA binding of dPHDdCHD 95

3.5.2 Investigating the histone binding of dPHDdCHD 96

3.5.2.2 Confirmation of the binding of two histone peptide to dPHDdCHD by MS 99

3.6 Discussion 100

Chapter 4 Investigations into the structure and function of the histone demethylase PHF8……. 104

4.1 The KDM7 family of histone demethylases 104

4.1.2 Objectives 106

4.2 Characterisation of PHF8 106

4.3 Characterisation of KIAA1718 109

4.4 Testing for demethylation of ARID5B 110

4.5 Investigating peptide binding by PHF8 and KIAA1718 111

4.6 Inhibitor assays 114

4.6.1 Kinetic characterisation of PHF8 115

4.6.2 Determination of PHF8 kinetic parameters 115

4.6.3 KIAA1718 kinetic parameters 116

4.6.4 IC_{50} value determination 117

4.6.5 ALPHA screen assay development for PHF8 117

4.6.5.1 Principles of the ALPHA screen 118

4.6.5.2 Assay development with PHF8 119

4.7 Inhibitor screening using the ALPHA screen 126

4.7.1 Daminozide is a specific inhibitor for the KDM2/7 subfamilies of JmjC demethylases 126

4.7.2 Daminozide structural biology 129

4.7.3 PHF8 crystallography 131

4.7.4 Inhibition by Daminozide in cells 133

4.7.5 Inhibition of PHF8 by hydroxamate based compounds 133

4.7.6 Autohydroxylation by PHF8 135

4.8 Exploring the substrate specificity of PHF8 136

4.9 Clinical mutations in PHF8 147

4.9.1 Bioinformatic analyses of PHF8 1306dupG 147

4.9.2 Production and testing of the clinical PHF8 mutant in cells 148

4.10 Discussion 150

Chapter 5 Investigations into arginine demethylation by 2OG oxygenases 155

5.1 Introduction 155

5.1.2 Objectives 157

5.2 Identification of arginine demethylation by JMJD2E 157

5.2.2 Demethylation of symmetric dimethylarginine by JMJD2E 159

5.3 Screening for arginine demethylation by other 2OG oxygenases 160

5.3.2 Screening for histone peptide arginine demethylation by 2OG oxygenases 162

5.3.3 Effect of histone acetylation of arginine demethylation 165

5.4 Characterisation of demethylation of H3 R2Me$_{2a}$ by JMJD2E 167

5.5 Investigating the differential sequence preference of KDMs for arginine and lysine demethylation 169

5.6 Characterisation of arginine demethylation of H3 K27RMe$_{2a}$ by JMJD3 171

5.7 JMJD2A crystallography with methylated arginine peptides 174

5.8 Arginine demethylation of full length histones 174

5.8.1 Principles of isotope ratio mass spectrometry 175

5.8.2 IRMS experiments with histones 175

5.9 Identification of histone arginine demethylation in cells 179

5.9.2 Immunofluorescence experiments using anti-dimethylarginine antibodies 183

5.10 Identification of non-histone substrates for the JmjC KDMs 185

5.10.1 Arginine methylation of p53 185

5.10.2 Peptide screening 186

5.10.3 Arginine demethylation of hnRNP K 188

5.10.4 Production of PRMT1 for *in vitro* protein methylation 190

5.10.5 Effect of arginine methylation on the cellular localisation of hnRNP K 192

5.10.6 *In vitro* demethylation of FL hnRNP K 194

5.10.7 Demethylation of FL hnRNP K by JMJD2E 196

5.10.8 Identification of hnRNP K demethylation by mass spectrometry 197

5.10.8.2 Coimmunoprecipitation of hnRNP K and JMJD2A 204

5.11 Discussion 205

Chapter 6 Characterisation of UTY 210

6.1 The KDM6 family of histone demethylases 210

6.1.2 Objectives 211

6.2 Cloning and expression of UTY 212

6.2.1 Large-scale expression of UTY 212

6.3 Characterisation of recombinant UTY 214

6.3.1 Detection of catalytic activity of UTY as uncoupled 2OG turnover 214

6.3.2 Determination of the substrate of UTY by peptide screening 215

6.3.3 Confirmation of histone demethylation by peptide NMR spectroscopy 216

6.3.4 Kinetic characterisation of UTY peptide demethylation 217

6.3.5 Peptide binding assays 220

6.3.6 Characterisation of inhibition of UTY by 2OG oxygenase inhibitors 221

6.4 Production and characterisation of UTY variants 222

6.4.2 Characterisation of UTY variants by MALDI MS and NMR spectroscopy 223

6.4.3 Kinetic characterisation of UTY variants 224

6.5 Structural studies of recombinant UTY 227

6.6 Investigating the cellular function of UTY 229

6.6.1 Studies of the global levels of H3 K27Me_3 by immunofluorescence 229

6.6.2 Effects of UTY on ANF expression 230

6.6.3 Studies on the expression levels of HOXA genes by qRT-PCR 234

6.7 Discussion 236

Chapter 7 Summary 240

Chapter 8 Materials and methods 245

8.1 General Procedures: 245

8.2 SDS-PAGE gel electrophoresis: 246

8.2.1 Reagents: 246

8.2.2 Method: 247

8.3 Agarose gel electrophoresis 247

8.3.1 Reagents: 247

8.3.2 Method: 247

8.4 Recombinant enzyme expression in *E. coli* 248

8.4.1 Reagents: 248

8.4.2 Methods: 250

8.5 Recombinant enzyme purification 252

8.5.1 Reagents: 252

8.5.2 Methods: 253

8.6 DNA techniques 255

8.6.1 Reagents 255

8.6.2 Methods 256

8.7 Activity assays: 259

8.7.1 MALDI-TOF MS assays: 259

8.7.2 FDH assay: 261

8.7.3 ALPHA screen activity assays: 261

8.7.4 ALPHA screen binding assays: 262

8.7.5 Fluorescence polarization assays: 262

8.7.6 Isothermal Titration Calorimetry: 262

8.7.7 Luciferase reporter assays 263

8.8 Western Blot: 264

8.9 Dot blots: 264

8.10 qRT PCR 265

8.11 Peptide synthesis: 266

8.12 Mammalian cell culture 268

8.12.1 Reagents 268

8.12.2 Methods 268

8.12.3 Immunofluoresence 269

8.12.4 Immunoprecipitation: 271

8.13 Mass spectrometry 271

8.13.1 Methanol/Chloroform extraction: 271

8.13.2 In-solution digest: 271

8.13.3 LCMS/MS: 272

8.14 Histone extraction 273

8.15 NMR Spectroscopy 273

8.15.2 NMR structure calculation 277

8.15.3 Average structure determination and ensemble RMSD calculations 278

8.15.4 Peptide titrations 278

Chapter 9 References 279

Appendix 1 308

Appendix 2 310

Appendix 3 312

Appendix 4 317

Abbreviations

1D	One-dimensional
2,4-PDCA	Pyridine-2,4-dicarboxylic acid
2D	Two-dimensional
2OG	2-Oxoglutarate
2TY	2 × Tryptone/yeast extract
3D	Three-dimensional
5caC	5-Carboxycytosine
5fC	5-Formylcytosine
5hmC	5-Hydroxymethylcytosine
5hmU	5-Hydroxymethyluracil
5mC	5-Methylcytosine
6×His-tag	Hexa-histidine tag
ABH	AlkB homologue
ADH	Aldehyde dehydrogenase
AdOx	Adenosine-2,3-dialdehyde
AlkB	Alkylation Repair Protein B
ALPHA	Amplified luminescent proximity homogeneous assay
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BER	Base excision repair
Bio	Biotin
BSA	Bovine Serum Albumin
C-terminus	Carboxy terminus
CHD	Chromodomain
CHD1-9	Chromodomain helicase DNA binding protein1-9
CHCA	α-Cyano-4-hydroxycinnamic acid
ChIP	Chromatin immunoprecipitation
CID	Collision-induced dissociation
CMV	Cytomegalovirus
COSY	Correlation spectroscopy
CV	Column volumes
Da	Dalton
dAFN	50 mM Ammonium formate, 500 mM sodium chloride in D_2O
dPHD	Tandem PHD domain construct of CHD4
dPHDdCHD	Tandem PHD and CHD domain construct of CHD4
dNTP	Deoxyribonucleotide
DAPI	4',6-Diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribosenucleic acid
DQF	Double-quantum filter
DSBH	Double-stranded β-helix
DTT	Dithiothreoitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
Et	Ethyl
ETD	Electron-transfer dissociation
FAD	Flavin adenine dinucleotide

FBS	Fetal bovine serum
FBXL	Fbox, Leucine-rich repeat protein
FDH	Formaldehyde dehydrogenase
FIH	Factor Inhibiting HIF
FITC	Fluorescein Isothiocyanate
FL	Full length
Fmoc	Fluorenylmethyloxycarbonyl
FP	Fluorescence polarisation
FPLC	Fast Protein Liquid Chromatography
GFP	Green fluorescent protein
H1	Histone H1
H2A	Histone H2A
H2B	Histone H2B
H3	Histone H3
H4	Histone H4
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HEK	Human embryonic kidney
HEPES	4-(2-Hydroxylethyl)-1-piperazineethanesulfonic acid
HIF	Hypoxia inducible factor
HMQC	Heteronuclear multiple quantum coherence
HMT	Histone methyltransferase
HOBt	1-Hydroxybenzotriazole
HP1	Heterochromatin protein 1
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single-quantum correlation spectroscopy
HRP	Horseradish peroxidase
^{i}Pr	Isopropyl
IOX1	5-Carboxy-8-hydroxyquinoline
IP	Immunoprecipitation
IPAP	In-phase / anti-phase
IPTG	Isopropyl β-*D*-thiogalactopyranoside
IRMS	Isotope ratio mass spectrometry
ITC	Isothermal titration calorimetry
JARID	Jumonji, AT rich interactive domain
JmjC	Jumonji-C
JMJD	Jumonji-C domain containing protein
K_D	Binding constant
KAc	N^{ε}-Acetyllysine
KDM	Lysine demethylase
KEt_2	N^{ε}-Diethyllysine
KFor	N^{ε}-Formyllysine
K^{i}Pr	N^{ε}-Isopropyllysine
KMe	N^{ε}-Methyllysine
KMe_2	N^{ε}-Dimethyllysine
KMe_3	N^{ε}-Trimethyllysine
KMeEt	N^{ε}-Methylethyllysine
KMe^{i}Pr	N^{ε}-Methylisopropyllysine
KMT	Lysine methyltransferase
K/O	Knockout

LCMS	Liquid chromatography mass spectrometry
LIC	Ligation independent cloning
LOXL	Lysyl oxidase-like
LSD	Lysine specific demethylase
MALDI	Matrix-assisted laser desorption / ionisation
MaxEnt	Maximum entropy
Me	Methyl
MS	Mass spectrometry
N-terminus	Amino terminal
NAD^+	Nicotinamide adenine dinucleotide
NMR	Nuclear Magnetic Resonance
NOE	Nuclear overhauser effect
NOESY	Nuclear overhauser effect spectroscopy
NOG	*N*-Oxalylglycine
NURD	Nucleosome remodeling deacetylase
OD	Optical density
OEt_2	N^{ε}-Diethylornithine
OMe_2	N^{ε}-Dimethylornithine
PAD4	Protein arginine deiminase 4
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	RCSB protein data bank
PHD	Plant homeodomain
PHD1-3	Prolyl hydroxylase 1-3
PHF	PHD finger protein
ppm	Parts per million
PTM	Post-translational modification
RAMA	Ramachandran-based torsion angle database potential
RCF	Relative centrifugal force
RDC	Residual dipolar coupling
RMe	N^{ω}-Monomethylarginine
RMe_{2a}	Asymmetric N^{ω}-dimethylarginine
RMe_{2s}	Symmetric N^{ω}-dimethylarginine
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RPM	Revolutions per minute
SAM	*S*-(5'-Adenosyl)-L-methionine
SAXS	Small angle X-ray scattering
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SILAC	Stable isotope labelling by amino acids in cell culture
SPPS	Solid phase peptide synthesis
TauD	Taurine dioxygenase
TAE	Tris-acetate-EDTA
TB	Terrific broth
TEMED	*N,N,N',N'*-Tetramethylethylenediamine
TET	Ten-eleven translocation
TEV	Tobacco etch mosaic virus
TFA	Trifluoroacetic acid
TOF	Time-of-flight
TOCSY	Total correlation spectroscopy

Tris	*Tris*-(hydroxylmethyl)aminomethane
UTX	Ubiquitously transcribed tetratricopeptide repeat, X chromosome
UTY	Ubiquitously transcribed tetratricopeptide repeat, Y chromosome
v/v	Volume for volume
w/v	Weight for volume
wt	Wild-type

Chapter 1
Introduction

1.1 Epigenetics

Despite the commonly held view that all diploid cells in an organism share the same deoxyribonucleic acid (DNA), each one is able to differentiate to form a wide variety of phenotypically distinct cell types. The central dogma of molecular biology, as originally defined by Francis Crick, states that nucleic acids transfer information into proteins, but proteins do not transfer information back to the DNA or RNA.[1,2] Simply put, "DNA makes RNA makes protein".[3] Given the variety of differential outcomes for cells, however, it is evident that the situation is more complex. The field of epigenetics involves the study of some of the processes that are responsible for the differentiation of cells and higher levels including whole organisms.

The term epigenetics has evolved over the years; it was originally coined by Conrad Waddington in 1942 as "the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being".[4] Waddington likened cell differentiation to a ball rolling down a hill into a series of valleys from which it cannot escape, over what he called the "epigenetic landscape" (Figure 1.1).[5] In

The image originally presented here cannot be made freely available via ORA because of copyright. The image can be viewed in reference 5, below.

Figure 1.1 Conrad Waddington's "epigenetic landscape". Reprinted with permission from The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology.[5]

this analogy, the ball is a cell, rolling from its undifferentiated state at the top of the valley into one of the separate valleys corresponding to different cell lineages, from which it is unable to become any other cell type.[6]

Use of the term epigenetics has since evolved, and is now considered to refer to the study of stable and mitotically heritable changes in cellular phenotype and gene expression that are not caused by alterations in the "ATCG" DNA sequence.[7] There are a number of biochemical mechanisms that are linked to epigenetic phenomena; these involve transcription factors, DNA modifications, non-coding RNAs, histone variants, nucleosome remodeling and post-translational histone modifications, some of which are described below. In general, these epigenetic phenomena become reset in the germline, where totipotency is again required.[8] To what extent epigenetic marks can be inherited across generations in higher organisms remains unclear.[9]

1.1.1 DNA packaging

Each cell contains around 2 metres of "linear" DNA.[10] In order to provide some structure for this within the cells, DNA is wound around positively charged histone proteins, to form a nucleosome.[11] Each nucleosome consists of a histone octamer, comprising two each of histone 3 (H3), histone 4 (H4), histone 2A (H2A) and histone 2B (H2B).[12] These form anti-parallel heterodimers (H3 with H4 and H2A with H2B), which then pair to form two tetramers (homodimers of the heterodimers, H3 / H4 and H2A / H2B) held together by interactions including hydrogen bonds and finally stack together to form the histone octamer. In general, 146 base pairs of DNA wrap around the core histone complex, stabilised by electrostatic interactions between positively charged histones and negatively charged DNA to form a nucleosome (Figure 1.1).

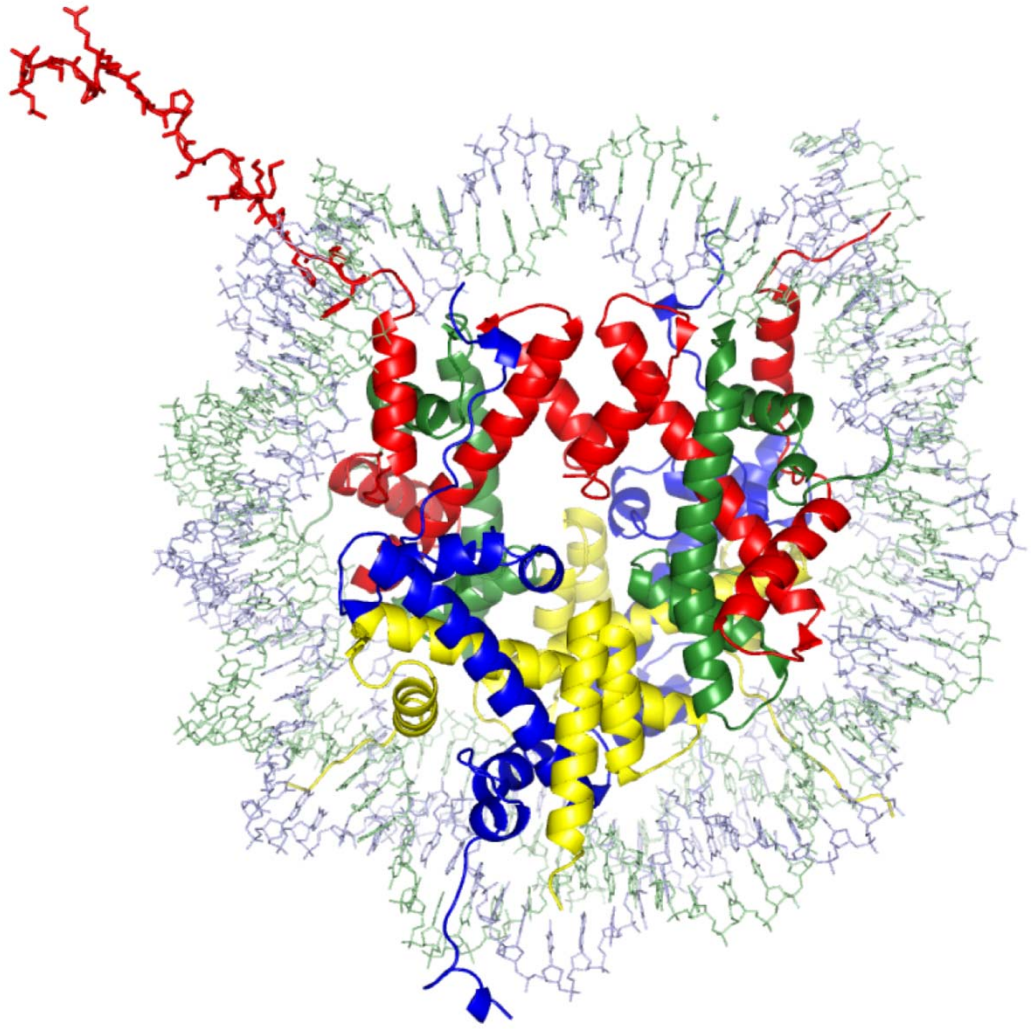


Figure 1.1 View from a crystal structure of a nucleosome (PDB ID: 1AOI)[12]. DNA is shown in pale blue and green. H3 are shown in red, H4 in green, H2A in yellow and H2B in blue. Due to flexibility, density is not observed for most of the *N*-terminal tails.

Each histone protein within the nucleosome consists of a histone fold domain (HFD), which is a globular core domain containing three α-helices separated by two loops, with a flexible, unstructured *N*-terminal tail that protrudes from the nucleosome structure beyond the DNA. These tails are thought to be important for the structure of individual nucleosomes, and tail deletions have been shown to reduce their stability.[13] Individual nucleosomes are joined together by linker regions of DNA to form what is known as the 10 nm fibre, or "beads on a string" (Figure 1.2).[14] Addition of linker histone 1 (H1) leads to further compaction into the 30 nm fibre.[15] Both of these

conformations can be clearly viewed in electron micrographs.[16] Beyond the 30 nm fibre, DNA can further condense to form a chromatin fibre, which is variably packed depending on the region of DNA present; euchromatin tends to adopt a more open structure, whilst heterochromatin tends to be highly compacted.[17] During metaphase, the chromatin forms its most highly condensed structures (DNA is condensed more than 1000 fold) to form mitotic chromosomes.[18]

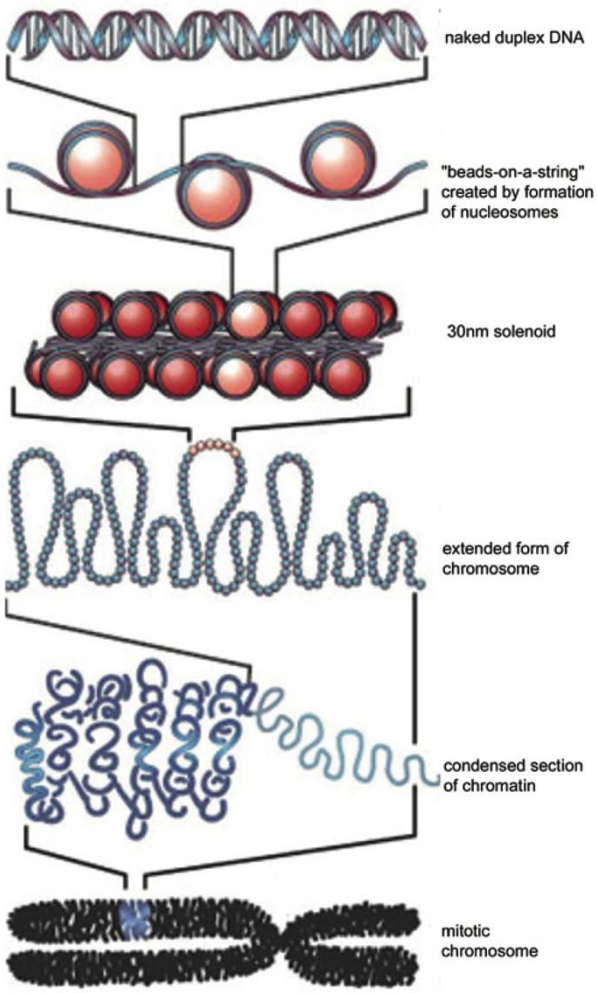


Figure 1.2 Diagram illustrating the different stages of DNA compaction within a cell. Figure reproduced from "http://beyondthedish.wordpress.com/tag/chromatin-fiber/".

1.1.1.2 Structure of the histone tails

The *N*-termini of histone proteins are considered to be intrinsically disordered. The tails are absent from the currently published crystal structures of nucleosomes (Figure

1.1),[12] suggesting that under these crystallographic conditions they have a high degree of disorder. A number of NMR studies have also shown the unmodified tails to be essentially unstructured, with amino acid residues corresponding to the histone tails found largely in the "random coil" region of the NMR spectra; for example in a recent study using magic angle spinning (MAS) NMR to study a 12mer nucleosomal array, Gao *et al.* demonstrate that in conditions that lead to a highly condensed chromatin array the tails remain highly dynamic.[19] Another study employing ion mobility mass spectrometry is consistent with the tails adopting multiple dynamic conformations in solution.[20] However, there is some evidence from computational studies and biophysical analyses that the histone tails may adopt some secondary structure in solution. Secondary structure predictions for the histone tails calculate that the H3, H4 and H2B tails have significant regions of α-helix.[21,22] Further, a number of molecular dynamics studies have identified regions of α-helix within the H3 and H2B tails and regions of β-sheet in the H4 tail.[23] These studies also suggest that post-translational modifications (Section 1.1.3) may alter the stability of different structurally distinct populations; for example, acetylation of H3 may increase its propensity to form β-hairpins, whilst acetylation of H4 is predicted to induce an increased α-helical content.[24,25] In addition, circular dichroism experiments suggest that between 17% and 35% of the α-helical content of the nucleosome as a whole is contributed by the histone tails. *Cis-trans* prolyl isomerisation by the yeast proline isomerase Fpr4 may also have dramatic effects on the histone tail structure. For example, isomerisation of H3 P38 alters the local structure of the tail, causing a kink in the peptide chain which alters its relative orientation within the nucleosome, blocking methylation at H3 K36 by Set2.[26] It is thus likely that the structures of the histone tails vary considerably depending on the biological context. As such, changes in local tail structure may have

roles in regulating protein binding and gene regulation.

1.1.1.3 Histone variants

The animal nucleosome typically includes the four canonical histone proteins, H3, H4, H2A and H2B.[12] These canonical histones are highly expressed in cells in a replication dependent manner, with a peak of expression during S phase, so as to provide a large pool of readily available histones to produce nucleosomes for newly synthesised DNA.[27] However, with the exception of H4, a number of histone variants also exist within human cells, which can be interchanged into the nucleosome structure (Table 1.1).

Table 1.1 Some of the known human histone variants and their regions of enrichment.

Histone	Category of histone	Histone variant	Region of enrichment
H3	Canonical H3	H3.1 H3.2	
	Replacement histone	H3.3	Active chromatin
	CenH3	CENP-A	Centromere
	H3t	H3t	Sperm
H4	Canonical H4	H4	
H2A	Canonical H2A	Multiple H2A variants	
	H2AZ	H2AZ1 H2AZ2	Flanking 'nucleosome free' regions
	H2AX	H2AX	DNA damage sites
	macroH2A	mH2A mH2A1.1 mH2A1.2	The inactive X chromosome
	H2A.Bbd	H2A.Bbd	Active chromatin - overlap with H4 acetylation
H2B	canonical H2B	multiple H2B variants	

The human histone variants all adopt the common structure of a histone fold domain, but have variations in their sequence and / or size.[28] Unlike the canonical histones, histone variants are deposited on DNA in a replication independent fashion.[29] Variation in the sequence and structure of histone variants allows them to regulate gene expression by altering the stability of nucleosomes, the amount of DNA that wraps around them, and the post-translational histone modifications that are possible.[30] For H3, the variants include H3.1 and H3.2 (together making up the

canonical H3), H3.3, H3t and CenH3 (known as CENP-A in mammals).[30] Which variant is included within any given nucleosome depends on the biological context. For example, CENP-A is included in nucleosomes found at the centromere,[31] whereas H3.3 marks active chromatin,[32] and H3t is found in the nucleosomes of sperm.[33] Mutation or dysregulation of histone variants is associated with cancers and other diseases.[34]

1.1.2 DNA Modifications

DNA methylation was the earliest discovered epigenetic modification.[35] It can be caused both by exogenous methylating agents, such as methyl methanesulfonate (MMS) and 4-(methylnitrosoamino)-1-(3-pyridyl)-1-butanone (NNK) or by DNA methyltransferases.[36–39] Whilst exogenous methylation tends to be random and must be actively detected and removed to avoid toxic effects, enzymatic methylation is directed and has important roles in transcriptional control and defining cellular identity.

Most human enzymatic DNA methylation identified to date occurs on cytosine to form 5-methylcytosine (5mC, Figure 1.3).[40] *De novo* DNA methylation is carried out by DNA methyltransferase 3A (DNMT3A) and DNMT3B,[41] whilst replication dependent methylation is carried out by DNMT1, which has a preference for hemi-methylated DNA.[42] In addition to 5mC, 5-hydroxymethylcytosine (5hmC),[43] 5-formylcytosine (5fC),[44] and 5-carboxycytosine (5caC) are also known to occur in humans (Figure 1.3). These other modified cytosine residues are formed by the action of the Ten-Eleven-Translocation (TET) enzymes.[45]

The TET enzymes form 5caC by successive oxidations of 5mC, forming 5hmC and 5fC as intermediates. 5caC can then be converted back to cytosine *via* base excision by thymine-DNA glycosylase (TDG) followed by base excision repair (BER).[46] This allows 5mC to be removed and replaced with cytosine. 5mC can also be deaminated by activation-induced deaminase (AID) to give thymine,[47,48] which is converted back to cytosine by TDG.[49] Similarly, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 1 (APOBEC1) or AID may deaminate 5hmC to 5-hydroxymethyluracil (5hmU) following which BER may regenerate cytosine,[50,51] although whether 5hmC is a substrate for these enzymes *in vivo* remains unclear.[52] Passive decarboxylation of 5caC is also possible.[53] Additionally, modifications of cytosine residues can also be lost via passive dilution during replication.[54,55]

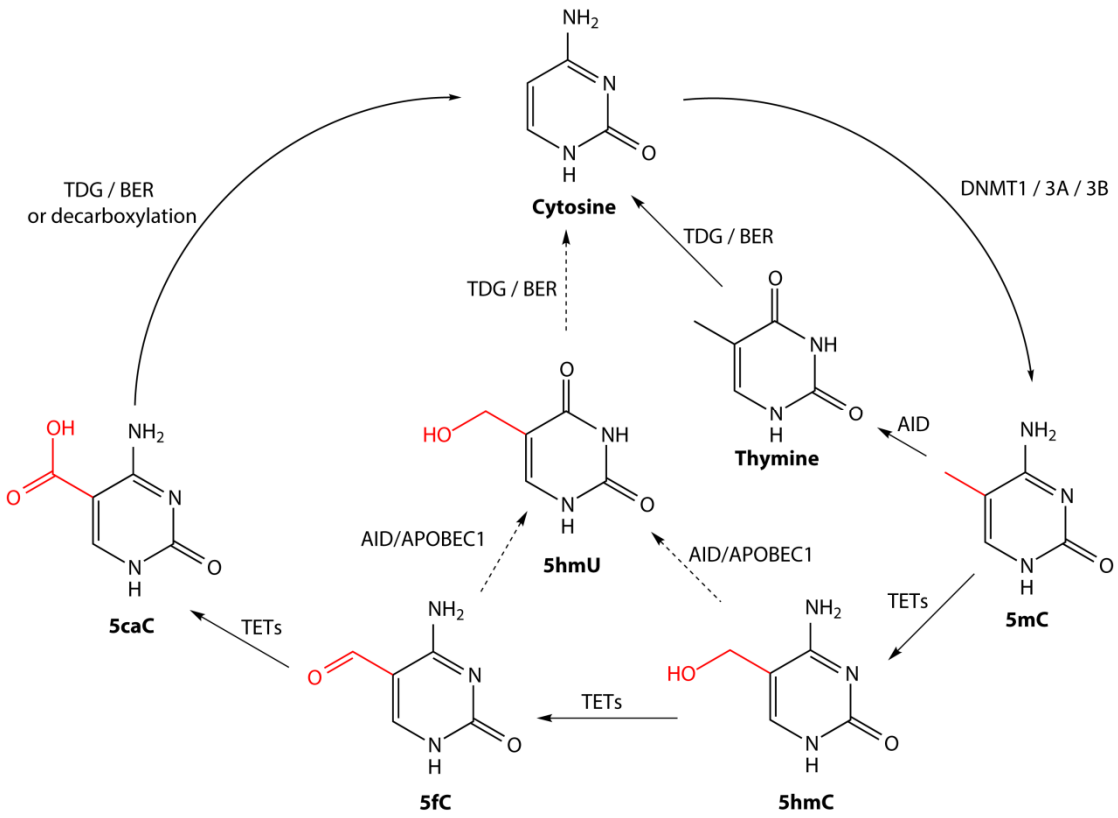


Figure 1.3 Scheme showing forms of modified cytosine found in humans and some of the pathways that interconvert them. Mechanisms of demethylation that are currently not fully validated are illustrated by dashed arrows. Cytosine modifications are shown in red.

DNA demethylation occurs at a global level during two stages of development: directly after fertilisation and then again during primordial germ cell development.[56,57] These events allow cellular totipotency to be regained. Following this, a new pattern of methylation is established by means of upregulation of DNMT3A and DNMT3B.[41] After these stages, demethylation only occurs in a targeted, gene specific manner.

DNA methylation primarily occurs at CpG dinucleotides, which are highly methylated within the human genome.[58] There are CpG rich regions in the genome, however, known as CpG islands, which are mostly found at gene promoters and are often hypomethylated.[59,60] DNA methylation generally marks regions of gene repression, and is involved in long-term silencing of genes, such as at imprinted genes and on the inactive X chromosome. However, it appears likely that DNA methylation is not the primary, or at least sole, cause of gene silencing, but rather the method of maintaining this silencing over many cellular generations.[61] In support of this hypothesis, there is evidence that targeted DNA methylation may be mediated by initial histone methylation, as catalysed by histone methyltransferases.[62] Aberrant DNA methylation is associated with many diseases, including psychiatric disorders, mid-line defects and cancers.[61,63–65]

1.1.3 Post-translational modifications

DNA encodes directly for the underlying amino acid sequence of all proteins in the genome, as defined by the ~20 commonly encoded residues.[1] These are synthesised *via* transcription and ribosomal translation.[66] After synthesis, newly made proteins can be modified in multiple ways, including cleavage,[67] addition of cofactors (eg. Haem)[68] or addition of small chemical moieties (eg. acetyl, hydroxyl groups).[69] These processes all constitute post-translational modifications (PTMs).[70]

Histone tails are amongst the most highly post-translationally modified proteins within the human proteome. Modifications are known to occur on lysyl-, arginyl-, threonyl-, tyrosyl- and seryl-residues, and these modifications are thought to account for many epigenetic phenomena.[71,72] Until recently this was mainly thought to be in the form of methylation, acetylation, phosphorylation and ubiquitylation, but with the increased power of mass spectrometry the library of known epigenetic marks is rapidly expanding to include modifications such as crotonylation and hydroxylation.[73,74] Figure 1.4 summarises the array of modifications that are currently known to occur on the *N*-terminal tails of the four core histones.

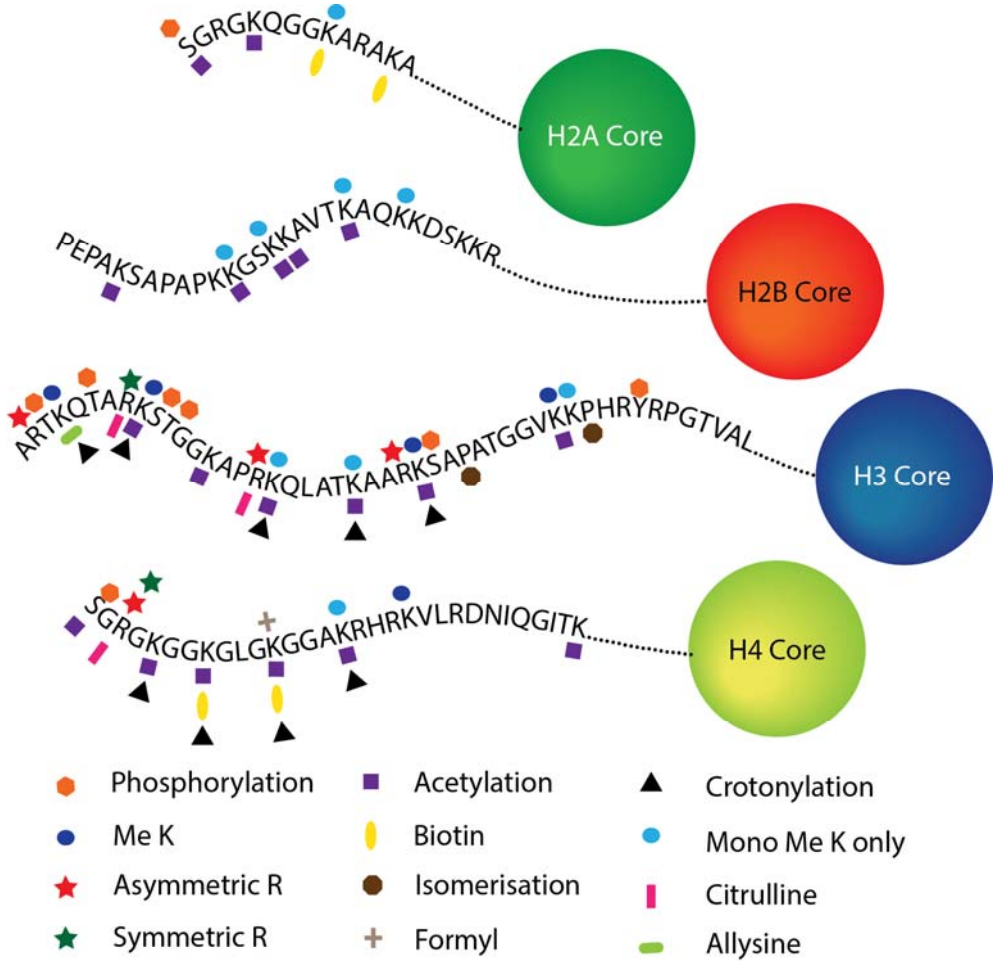


Figure 1.4 Post-translational modifications known to occur on the *N*-terminal tails of the four canonical core histones.

Multiple marks can occur in *cis* (on the same histone tail) or in *trans* (on different histone tails) in any one nucleosome. It has been proposed that different combinations of PTMs can induce different outcomes, in what has been named the "histone code".[75] Marks are placed by "writers", detected by "readers" and removed by "erasers", which all collectively modulate the chromatin environment to orchestrate gene transcription. This can be by direct alteration of the nucleosomal organisation, or by recruitment of transcription factors and regulatory complexes. Individual marks may antagonise others as for eg. H3 R2Me_{2a} and H3 K4Me_3 (where methylation of H3 R2 inhibits the activity of SET1, the enzyme that methylates H3 K4),[76] or directly block their placement by occupying the same amino acid residue. Below, current knowledge of the most common post-translational modifications is described.

1.1.3.2 Phosphorylation

Phosphorylation can occur on seryl-, threonyl- and tyrosyl-residues of histones (Figure 1.5). Histone phosphorylation is catalysed by protein kinases and removed by histone phosphatases. After DNA damage, the histone variant H2AX, present in nucleosomes around the break point, becomes hyperphosphorylated.[77] This phosphorylation is required to recruit the necessary DNA repair enzymes. After the event, the phosphorylation can be removed either by histone replacement or dephosphorylation. Interestingly, depending on the context, histone phosphorylation can also be involved in either chromatin compaction (during meiosis and mitosis) or gene activation by chromatin relaxation when associated with H3 acetylation.[78–82]

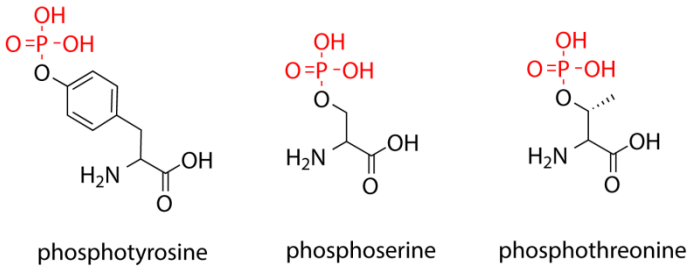


Figure 1.5 Forms of histone phosphorylation. Post-translational modifications are shown in red.

1.1.3.3 Acetylation

Acetylation of histones is known to occur on lysyl-residues (Figure 1.6). Acetyl groups can be added by histone acetyltransferases (HATs) and removed by histone deacetylases (HDACs).[83] In addition to being recognised by a number of histone readers, acetylation reduces the positive charge on lysyl-residues. This may thereby decrease the interaction of the histone tails with negatively charged DNA, causing a relaxation in DNA compaction and increased accessibility to other proteins and transcription factors, leading to an increase in gene transcription.[84,85]

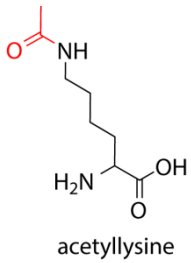


Figure 1.6 Structure of N^{ε}-acetylated lysine. The post-translational modification is shown in red.

1.1.3.4 Ubiquitylation

Ubiquitin is a 76 residue protein that can be attached to lysyl-residues by three

different enzymes through an enzymatic cascade reaction.[86] Once one ubiquitin molecule has been added this can itself be further ubiquitylated, often marking the histone protein for degradation.[87] Apart from degradation, monoubiquitylation can signal for protein recruitment or structural changes to the DNA, thereby altering gene transcription.[88,89] Monoubiquitylation can be reversed by deubiquitylation enzymes (DUBs).[90]

1.1.3.5 Methylation

Histone methylation occurs on both lysyl- and arginyl-residues. Unlike acetylation, it does not affect the charge of the lysyl- / arginyl-residues, and thus does not affect the structure of chromatin directly. Instead it can signal for both gene activation and gene repression,[91] depending on a number of factors, including the site and degree of the methylation.

Lysyl-residues can be N^{ε}-mono-, N^{ε}-di- or N^{ε}-trimethylated (Figure 1.7A), whilst arginine can be N^{ω}-mono- or N^{ω}-dimethylated (Figure 1.7B). In the case of arginine, this dimethylation can either be symmetric (Me_{2s}), with one methyl group on each terminal nitrogen of the guanidino group, or asymmetric (Me_{2a}), where both methyl groups are on the same terminal nitrogen.

Lysine and arginine methylation is catalysed by histone methyltransferases (HMTs). HMTs all use *S*-adenosyl methionine (SAM) as the methyl donor (Figure 1.7C), employing a nucleophilic substitution-displacement reaction.[92] Lysine methylation is catalysed by a large family of lysine methyltransferases (KMTs). With the exception of KMT4 (DOT1L) and a multiple subunit methyltransferase from the MLL1 core

complex,[93–95] these all possess an enzymatic SET domain that carries out the methylation reaction.[96] The family contains many members that all show high sequence and methylation-state selectivity and can also methylate non-histone substrates;[97] their selectivity can be further defined by binding partners *in vivo*.[98]

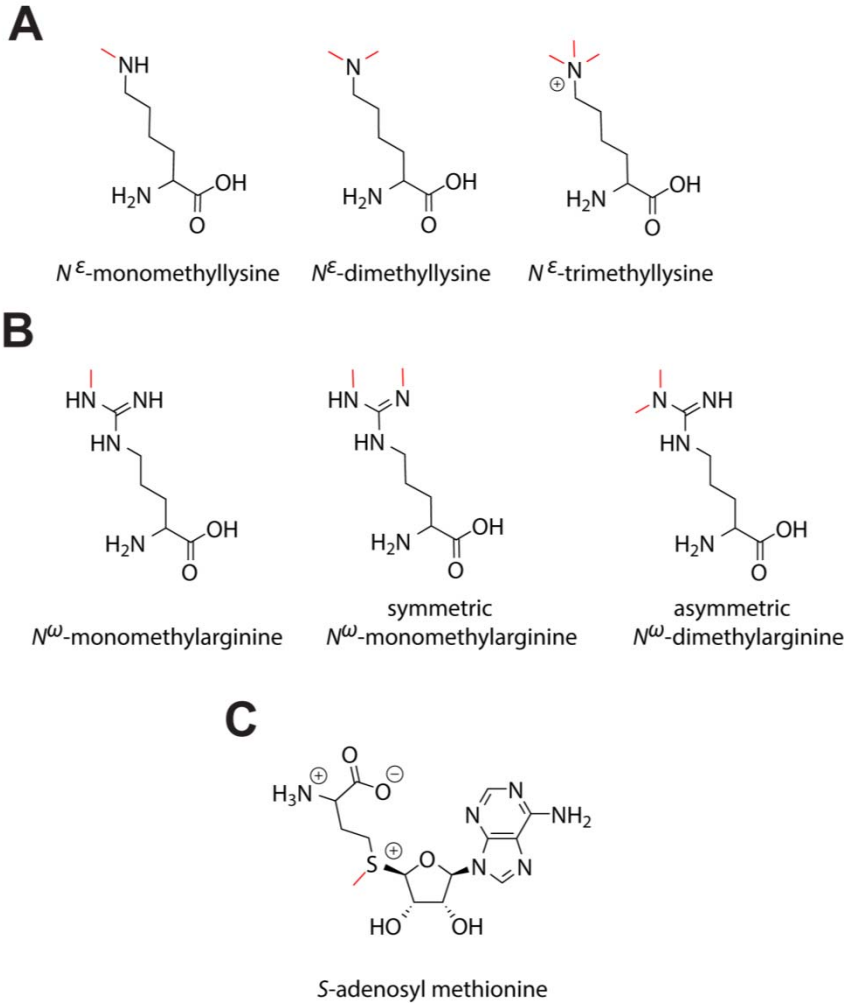


Figure 1.7 Forms of (**A**) lysine and (**B**) arginine methylation and (**C**) *S*-adenosyl methionine. Post-translational modifications of arginine and lysine and the donated methyl group of SAM are shown in red.

Lysine methylation has been identified on all four histone tails and has been ascribed a multitude of functions. H3 K4Me$_3$ is enriched at the transcription start sites of active genes, and the gene bodies of active genes are associated with enrichment of H4 K20Me and H3 K36Me$_3$, whilst the promoters of inactive genes are often enriched

in H3 K9Me$_3$ and H3 K27Me$_3$.[99–102] For example, high levels of H3 K27Me$_3$ are observed across the inactive X chromosome.[103,104] There are also a class of bivalent genes whose promoters are marked by both H3 K4Me$_3$ and H3 K27Me$_3$. In addition, methylation of a core residue on H3, K79 also occurs, mediated by DOT1L.[105–107] Unlike methylation of lysyl-residues on histone tail residues, no known demethylases for this mark are currently known.

Arginine methylation is catalysed by a family of 9 protein arginine methyltransferases (PRMT1-9). The product (mono / asymmetric / symmetric) and site of methylation varies between the different enzymes. As with the KMTs, in addition to arginine methylation of H3 and H4, the PRMTs also have a large number of non-histone substrates. Whilst some histone residues are only either symmetrically or asymmetrically methylated eg. H3 R8 and H3 R26, other residues can have either form of methylation eg. H4 R3. In these cases, the consequence of the methylation is dependent on which type occurs; whilst H4 R3Me$_{2a}$ recruits the P300/CBP-associated factor (PCAF) to histones, leading to increased acetylation at H3 K9 and H3 K14 and increased expression of beta-globin,[108] H4 R3Me$_{2s}$ increases binding of DNMT3A to chromatin, resulting in increased DNA methylation and reduced gene transcription.[109] Interestingly, for H3 R2 arginine methylation, the converse effect is observed, with asymmetric methylation leading to gene repression and symmetric methylation resulting in an increase in gene expression.[76,110,111]

N-Methylation was long thought to be an irreversible protein reaction. However, in 2004, lysine-specific demethylase 1 (LSD1), a nuclear member of the amine oxidase family was identified as a lysine demethylase.[112] LSD1 uses a flavin adenine

dinucleotide (FAD)-dependent amine oxidase domain to oxidatively demethylate lysyl-residues with concomitant reduction of FAD to FADH (Figure 1.8).[113,114] Due to the requirement for imine formation in the mechanistic pathway, LSD1 is only able to reverse mono- and dimethylation on lysyl-residues. As well as H3 K4 and H3 K9, it has also been reported to demethylate a number of non-histone substrates including p53 and DNMT1.[115,116] More recently, a second member of this family, LSD2 has been identified as another histone monomethyl- and dimethyl-demethylase.[117] At present H3 K4 is the only known substrate for LSD2.

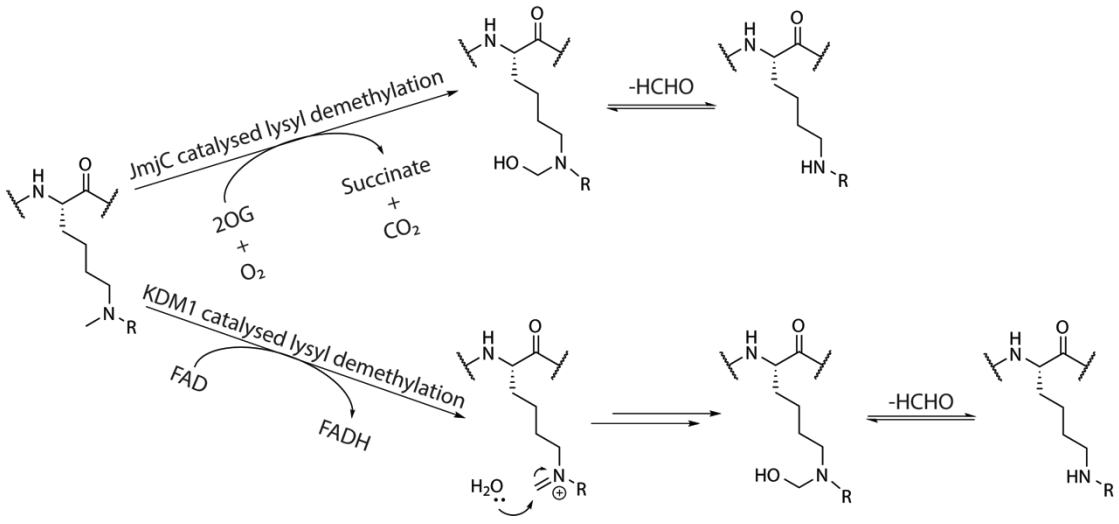


Figure 1.8 Scheme showing the known mechanisms of lysine demethylation.

A second family of histone lysine demethylases, the Jumonji-C histone demethylases (JmjC KDMs) was later identified. These are a subset of a much larger family of enzymes, the Fe(II)- and 2-oxoglutarate- (2OG) dependent dioxygenases (2OG oxygenases). The JmjC KDMs also employ an oxidative demethylation mechanism, but it does not require formation of an imine, and as such they are able to catalyse the removal of all three states of lysine methylation (Figure 1.8). The demethylation reaction is proposed to proceed *via* oxidation of the N^{ε}-methyl group to form a hemiaminal species, which subsequently fragments with loss of formaldehyde,

removing the methyl group from the lysyl-residue. Although this hemiaminal species has not been directly observed for any of the JmjC KDMs, a crystal structure of a related bacterial protein, AlkB, which demethylates DNA bases in an analogous fashion, has been solved with the hemiaminal intermediate trapped (Figure 1.9).[118]

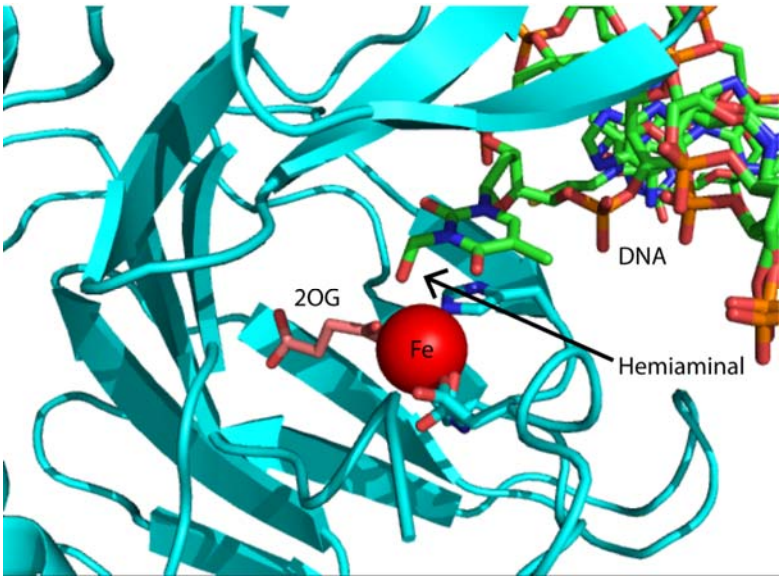


Figure 1.9 View from a crystal structure of AlkB (PDB ID: 3O1T)[118] showing the hemiaminal reaction intermediate. The DNA is shown in green and AlkB in cyan. The Fe atom is shown in red and 2OG in pink.

In addition to the LSD KDM family and JmjC KDM family, recently, a member of the copper containing lysyl oxidase (LOX) family,[119] LOXL2, has been described that can deaminate H3 K4Me_3 to allysine (Figure 1.10).[120] It uses the cofactor lysine tyrosylquinone (LTQ) and a copper(II) ion for the reaction, but the exact mechanism of deamination is currently unclear.

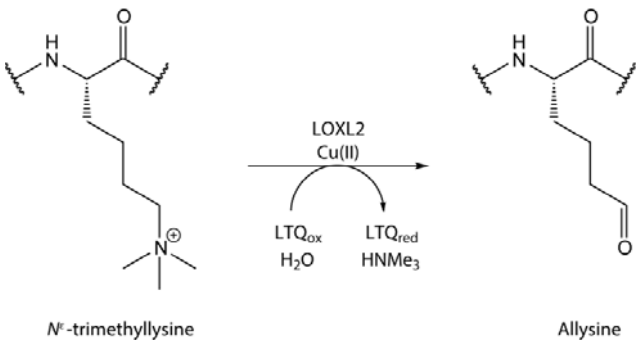


Figure 1.10 Scheme showing the conversion of N^{ε}-trimethyllysine to allysine catalysed by LOXL2.

There are currently no known arginine demethylases. The 2OG oxygenase, JMJD6, was initially characterised as an arginine demethylase of H3 R2 and H4 R3.[121] However, the arginine demethylase activity has never been replicated (to my knowledge), and JMJD6 has since been reassigned as a lysyl hydroxylase of U2AF65, a protein involved in RNA splicing (Figure 1.11).[122] A recent study has also implicated JMJD6 in hydroxylation of lysyl-residues within histones.[123,124]

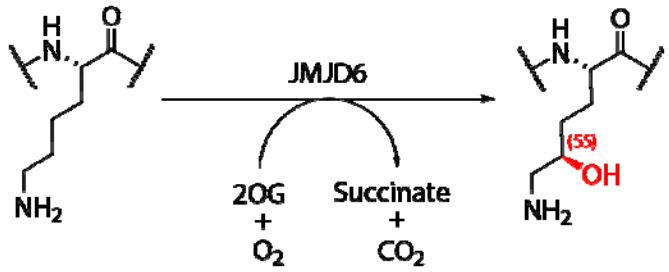


Figure 1.11 Scheme showing the hydroxylation of a lysyl-residue by JMJD6 to produce 5*S*-hydroxylysine.

Peptidylarginine deiminase 4 (PAD4) is able to catalyse the conversion of arginine and N^{ω}-monomethylarginine into citrulline (Figure 1.12).[125,126] PAD4 uses an active-site cystyl-residue to catalyse hydrolysis of the guanidino group *via* a nucleophilic addition-elimination reaction, to form citrulline.[127] Deimination proceeds *via* a hydrolytically labile thioamide intermediate. *In vitro* data suggests that demethylimination by PAD4 is much slower than deimination, and as such, in cells PAD4 may only accept non-methylated histone arginyl-residues.[128] Whether this is the case or not, PAD4 is not a true demethylase, as the non-methylated arginyl-residue is not recovered.

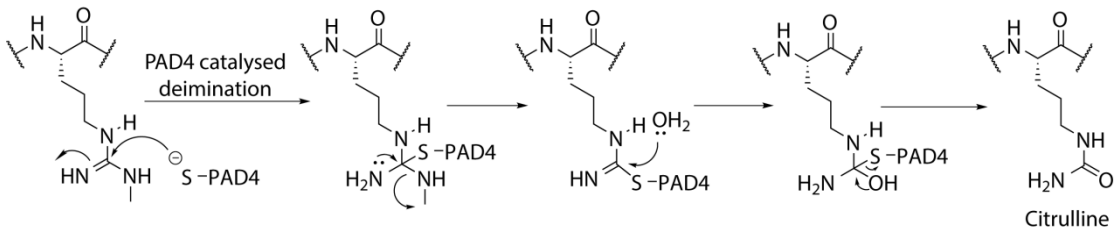


Figure 1.12 Outline mechanism of PAD4 demethylimination of N^{ω}-monomethylarginine

1.1.4 2OG oxygenases

The 2OG oxygenase enzymes are a superfamily of enzymes present in almost all organisms. They are able to catalyse a vast array of oxidative reactions, including epimerisation,[129] desaturation,[130] and ring closure (Figure 1.13).[131–137] In humans there are over 60 members that catalyse a huge variety of hydroxylation and demethylation reactions on both proteins and small molecules.[138]

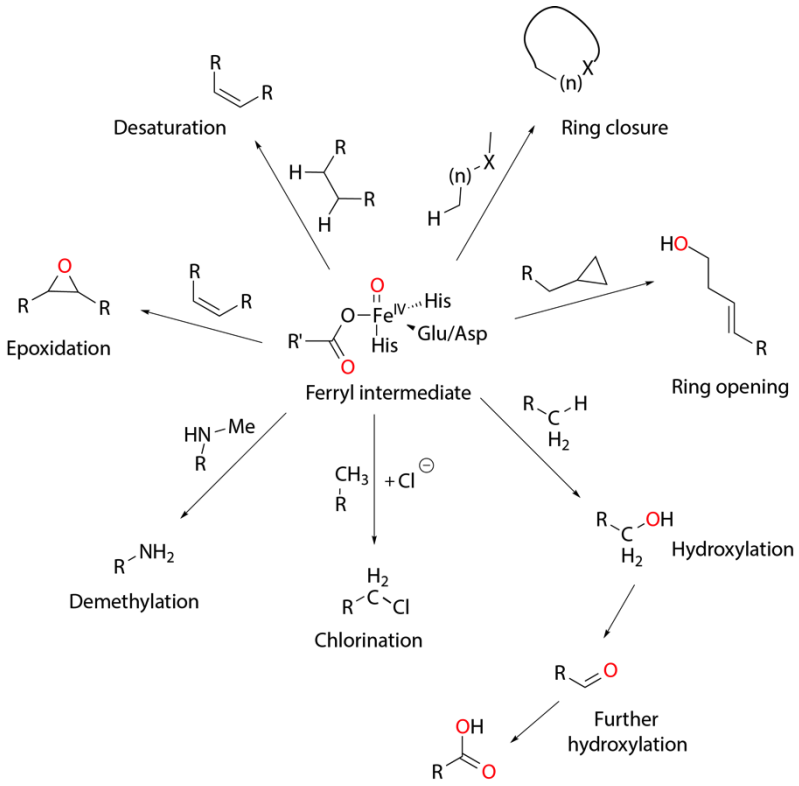


Figure 1.13 Scheme showing the variety of chemical reactions catalysed by 2OG oxygenase family members in both prokaryotes and eukaryotes, adapted from Clifton *et al.*.[131] R' represents the remainder of 2OG, oxygen atoms deriving from molecular oxygen are shown in red.

The consensus mechanism is based on mechanistic insights drawn from studies on a small number of 2OG oxygenases, including the small molecule, bacterial 2OG oxygenase, TauD, and the human oxygen sensing enzyme PHD2. Most structurally characterised 2OG oxygenases contain a conserved HXD/E…H iron-binding motif held within a double-stranded β-helix (DSBH, commonly known as a jellyroll motif)

(Figure 1.14).[139–141] In the resting state, the catalytic Fe(II) is octahedrally coordinated by these three amino acids and two or three water molecules. This resting state with the three water molecules bound has been observed experimentally in the crystal structure of the 2OG oxygenase, DAOCS.[142] Addition of 2OG displaces two of these water molecules, allowing one of its carboxylates to coordinate to the Fe(II) centre, whilst its other carboxylate is bound by a basic residue (arginine or lysine) in the active site.[143] This binding of 2OG to the Fe(II) centre produces a metal-to-ligand charge transfer band that can be observed using UV/vis spectroscopy. The final water molecule that binds to the Fe(II) centre is displaced on oxygen binding. It is thought that prime substrate binding weakens the binding of this metal-bound water, facilitating its displacement by oxygen.[144]

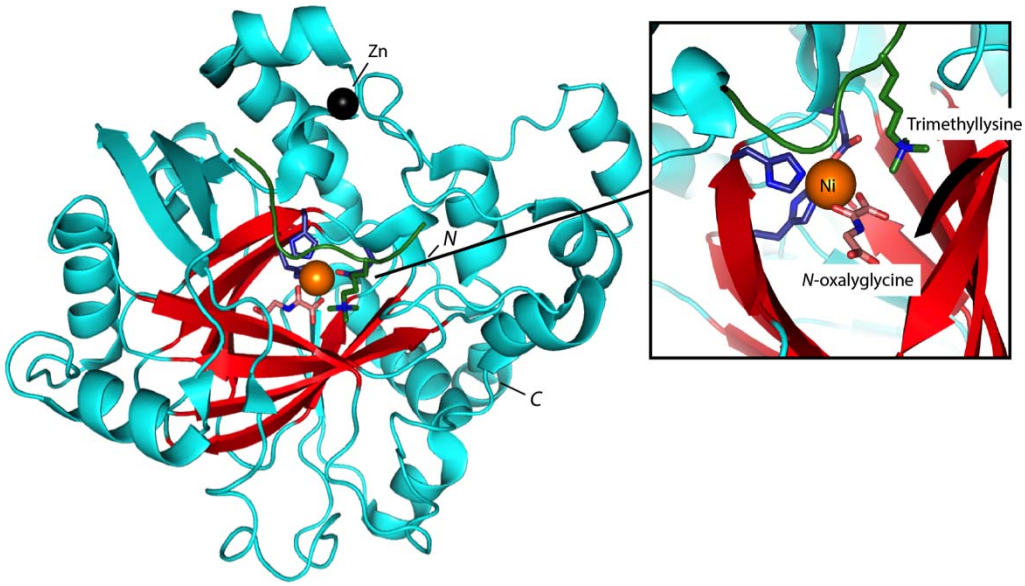


Figure 1.14 Views from a crystal structure of JMJD2A bound to an H3 K9Me$_3$ peptide (PDB ID 2OQ6)[145]. The β-barrel is shown in red. Nickel (substituted for iron) is shown in orange, iron-binding residues in purple, *N*-oxalylglycine in pink, zinc in black and histone peptide in green. The insert displays an enlarged image of the active site showing the octahedral coordination of the nickel.

Despite the diverse selection of reactions catalysed by the 2OG oxygenase family, they all undergo a remarkably conserved reaction mechanism, which is outlined in Figure 1.15.[146] Oxidation of the prime substrate (using molecular oxygen) is coupled

to conversion of the cosubstrate 2OG to succinate.[147] Electron transfer is mediated by a redox active Fe(II) ion. Initial reaction of Fe(II) with oxygen and concomitant decarboxylation of 2OG produces a reactive Fe(IV)-oxo species, which has been studied in some cases using spectroscopic techniques including UV-Vis, EPR and Raman spectroscopies.[148,149] The Fe(IV)-oxo species then abstracts a proton from the substrate C-H bond to form Fe(III)-OH and the hydroxyl group is subsequently transferred to the resultant substrate carbon radical. The hydroxyl group can further be lost, as in the case of the histone demethylases. The carbon radical can be quenched by other groups such as with a chloride radical, as in case of the halogenase CytC3. A number of 2OG oxygenases have been found to be more active in the presence of ascorbate, although the precise reason for this remains unknown.[150–152]

Due to their use of oxygen in the catalytic mechanism, members of the 2OG oxygenase family have previously been characterized as oxygen sensors that are involved in the hypoxic response.[149,153–155] This is best characterised for the human enzyme, PHD2; it has a weak $K_M(O2)$ and reacts slowly with oxygen. Thus, under hypoxic conditions, PHD2 becomes inactive, leading to accumulation of hypoxia inducible factor (HIF) and upregulation of hypoxia-related genes, such as EPO.[155] More recent work has also shown that the histone demethylase, JMJD2E, reacts slowly with oxygen,[156] and a second study has suggested that low oxygen availability reduces the activity of histone demethylases in macrophages,[157] raising the possibility that epigenetic transformations may be hypoxically regulated.

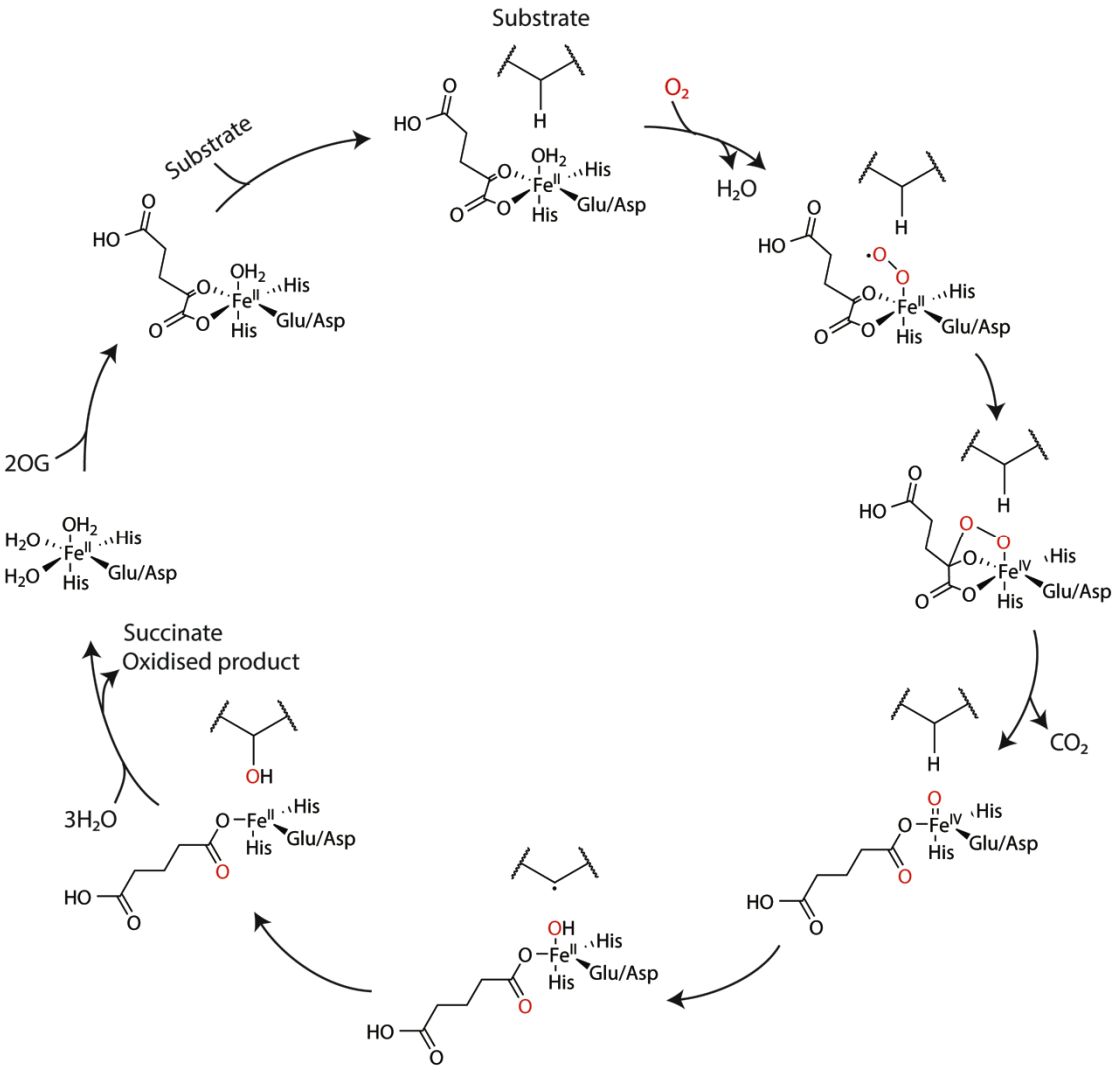


Figure 1.15 Scheme showing the consensus mechanism for oxidation by the 2OG oxygenase family, modified from Hausinger.[158]

1.1.4.2 JmjC histone demethylases

As the largest family of histone demethylases, the JmjC KDMs contain members that are able to remove most of the known histone lysine methylation marks.[159] However, there are currently no characterised enzymes capable of removing H4 K20Me_3 or H3 K79$Me_{1/2/3}$. For some JmjC KDMs, recent studies have also identified a number of non-histone substrates both on peptides *in vitro* and in cells.[160,161] Sequence and methylation state specificity is apparently dictated both by residues around the active site and those much further along the peptide binding surface.[145,162–164]

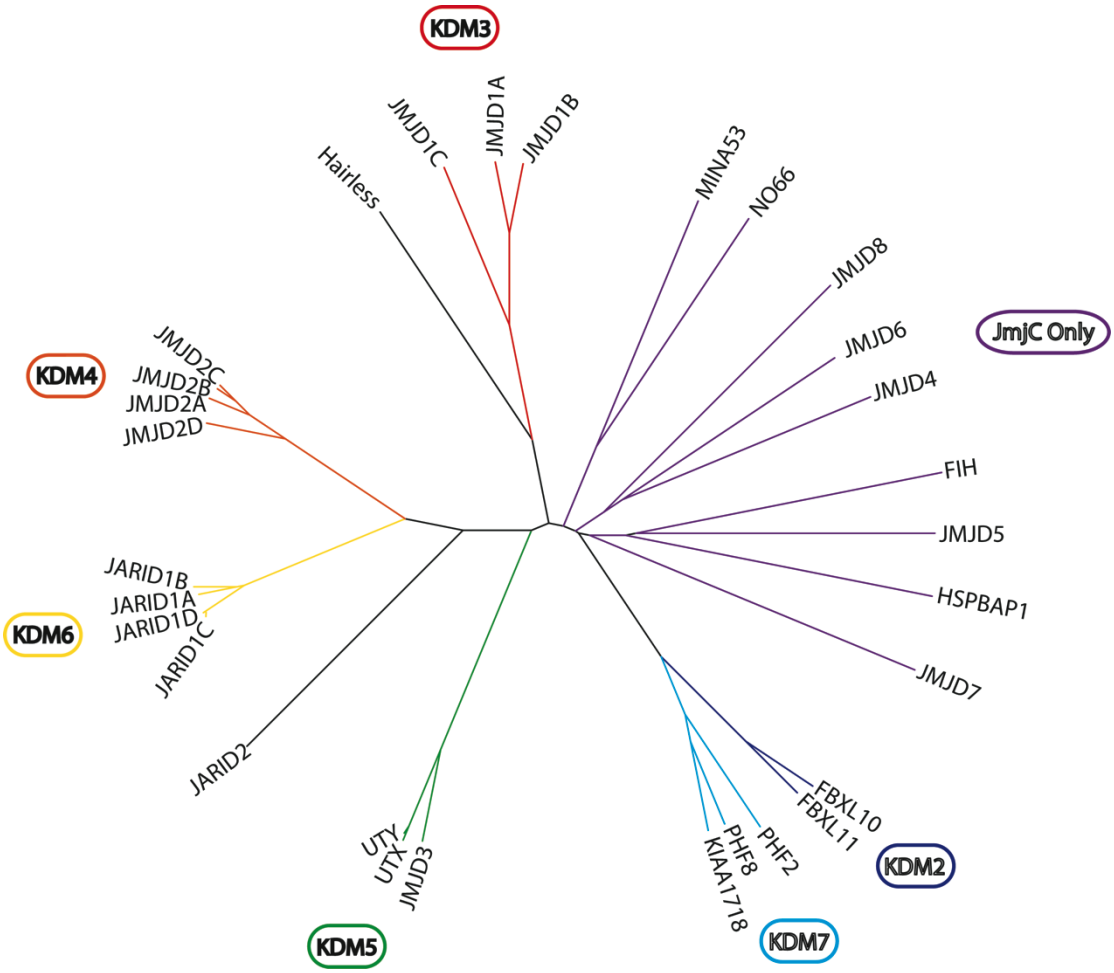


Figure 1.16 Phylogenetic tree of JmjC enzymes constructed from sequence alignments of the JmjC domains of each protein using ClustalW.

A detailed description of the mechanism of 2OG KDMs and the flavin-dependent monoamine oxidases can be found in "Mechanisms of human histone and nucleic acid demethylases" (Appendix 3).

The JmjC domain enzymes, as distinct from the more classical 2OG oxygenase domain (as found in the prolyl hydroxylases), have the basic 2OG-binding residue on the 4th β-strand of the DSBH core, rather than the 8th β-strand.[143] In addition, JmjC containing 2OG oxygenases often have a lysyl-residue binding to the 2OG cosubstrate rather than the arginyl-residue found in most other 2OG oxygenases. Phylogenetic analyses can be used to divide the JmjC family members into subgroups, which tend

to share common substrates (Figure 1.16).[165,166]

Although the majority of JmjC enzymes are histone demethylases, there are a number of JmjC domain containing enzymes that have alternative substrates. One example is FIH, which catalyses hydroxylation of the *C*-terminal region of the transcription factor, hypoxia inducible factor (HIF).[167] The reported reactions catalysed by JmjC enzymes are shown in Table 1.2.[168]

Table 1.2 Summary of the known substrates for characterised members of the JmjC family of 2OG oxygenases *in vivo*.[122,168–174] Subfamily colour coding matches that in Figure 1.16.

Subfamily	Enzyme	Modification(s) added / removed	Substrate(s)
KDM2	FBXL11	$KMe_{2/1}$	Histone (H3K36), p65
	FBXL10		Histone (H3K36)
KDM3	JMJD1A JMJD1B	$KMe_{2/1}$	Histone (H3K9)
KDM4	JMJD2A JMJD2B JMJD2C	$KMe_{3/2/1}$	Histone (H3K9, H3K36, H1.4K26)
	JMJD2D		Histone (H3K9, H1.4K26)
KDM5	JARID1A JARID1B JARID1C JARID1D	$KMe_{3/2/1}$	Histone (H3K4)
KDM6	UTX JMJD3	$KMe_{3/2/1}$	Histone (H3K27)
KDM7	KIAA1718	$KMe_{2/1}$	Histone (H3K9, H3K27)
	PHF2		Histone (H3K9), ARID5B
	PHF8		Histone (H3K9, H3K27, H4K20)
JmjC only	JMJD5	$Kme_{2/1}$, K-OH	Histone (H3K36), (NFATc1 hydroxylation proposed)
	MINA53	H-OH	Rpl27a
	NO66	$KMe_{3/2}$,H-OH	Histone (H3K4, H3K36), Rpl8
	FIH	N-OH, H-OH	HIF1α, ankryrin repeat proteins
	JMJD6	K-OH	U2AF65, histones

In addition to the catalytic JmjC domain, the majority of the KDMs also have multiple "reader" domains that help to target the catalytic domains to the correct targets. For example, the PHD domains of members of the KDM7 subfamily are known to bind to H3 tails trimethylated at K4, and the CXXC domains of FBXL11 target the enzyme to nonmethylated CpG DNA.[175–177] The binding domains possessed by characterised JmjC KDMs are shown in Table 1.3.

Table 1.3 Table showing the non-catalytic domains of JmjC KDMs.

Non-catalytic domain	JmjC KDMs containing the domain	Function
JmjN	JMJD2A-E, JARID1A-D	Unknown (required for catalytic activity)
Tudor	JMJD2A-C	Methyllysine Binding
Fbox	FBXL10/11	Protein-Protein Interactions
Tetratricopeptide Repeat	JMJD3, UTX, UTY	Protein-Protein Interactions
BRIGHT/ARID	JARID1A-D	DNA Binding
CXXC	FBXL10/11	DNA Binding
Leucine-rich Repeat	FBXL10/11	Protein-Protein Interactions
C5HC2 Zinc Finger	JARID1A-D	Unknown
PHD	FBXL10/11, JMJD2A-C, JARID1A-D, PHF8, KIAA1718, PHF2	Lysine/Methyllysine Binding

1.1.5 Histone Readers

Most chromatin-associating proteins contain one or more binding modules that interact with histone tails. The binding of these small domains helps to modulate and direct the activity of these proteins by localising them to the correct positions in the chromatin environment.[178] Descriptions of some of the histone readers relevant to this work are given below.

1.1.5.1 Bromodomains

The human proteome contains over 40 bromodomain-containing proteins.[179] Bromodomains consist of a left-handed 4 α-helix bundle, which recognises and binds to acetylated lysyl-residues (Figure 1.17A).[178] Variability in the loops joining these helices is thought to confer substrate specificity.[180] They are commonly found in chromatin-associated proteins, where their substrates are histone tails,[181] although there are also many examples of other substrates (eg. the bromodomain of CREB binding protein binds an acetylated lysyl-residue of p53)[182].

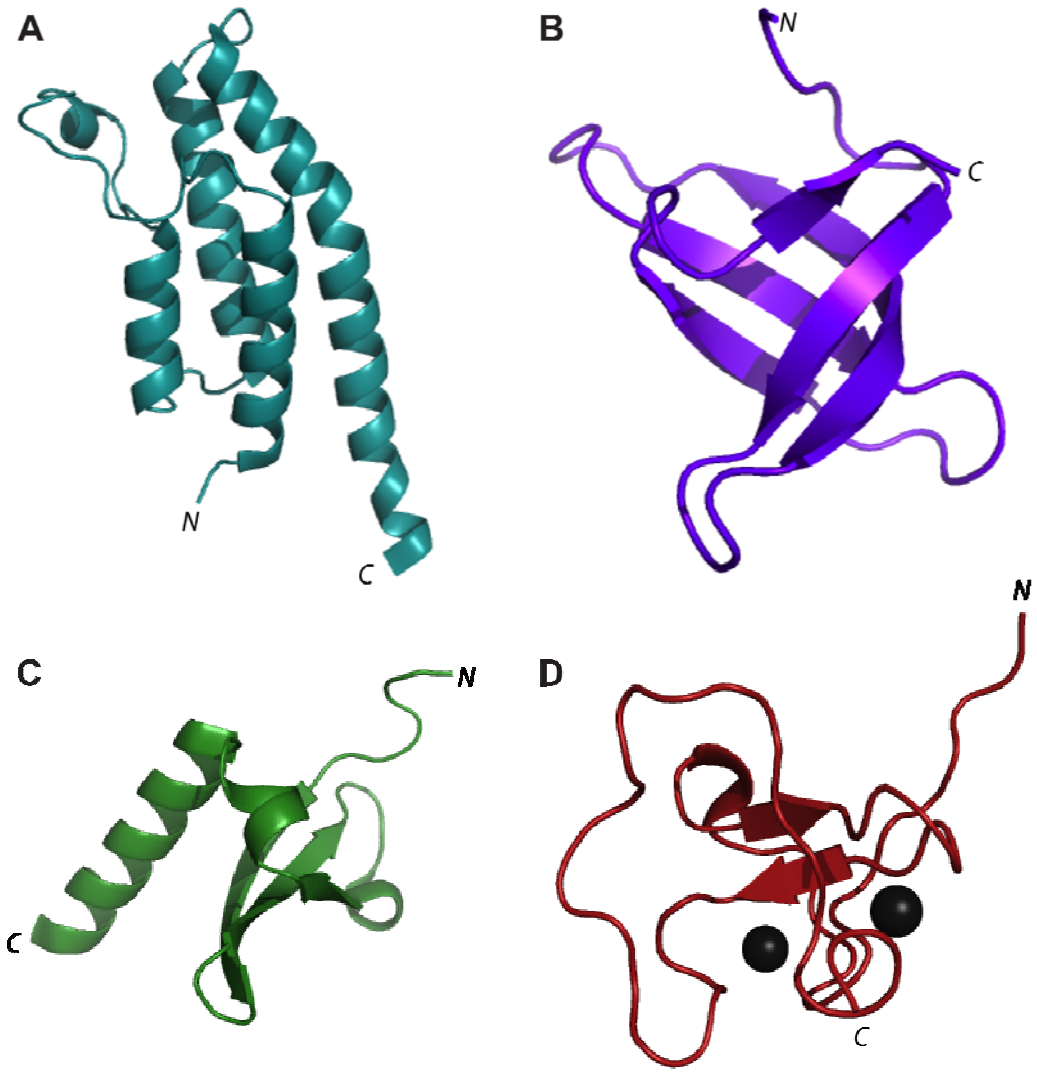


Figure 1.17 Representative structures of chromatin readers. **A)** View from a crystal structure of the BRD1 bromodomain (PDB ID 3RCW)[183] **B)** View from a crystal structure of the TDRD3 tudor domain (PDB ID 3PMT)[184] **C)** View from a crystal structure of the MPP8 chromodomain (PDB ID 3LWE)[185] **D)** View from a crystal structure of the UHRF1 PHD domain (PDB ID 3ZVZ)[186]. Zinc atoms are shown in black. *N*- and *C*-termini are marked.

1.1.5.2 Tudor domains

Tudor domains are the major characterised family of methylarginine binding domains.[187] The domain itself consists of a 5-stranded bent anti-parallel β-sheet, made up of around 60 amino acids, which bind the methylated residue in an aromatic cage of residues (Figure 1.17B).[188] They are able to bind either methylated arginyl- or lysyl-residues and it is thought that the openness of the aromatic cage defines this substrate specificity. They are involved in many cellular processes, including RNA

metabolism, histone modification and the DNA damage response.[189]

1.1.5.3 Chromodomains

Chromodomains were originally identified in two *Drosophila* proteins, Polycomb and HP1.[190] They have since been found in most plants and animals, and are frequently found in chromatin-associated proteins. Most commonly they bind methylated lysyl-residues,[191] although for some domains, DNA binding has also been reported.[192] Chromodomains contain 3 anti-parallel β-sheets packed against a single α-helix, amounting to around 50 amino acids (Figure 1.17C).

1.1.5.4 Plant homeodomains

Plant homeodomains (PHDs) are small protein domains of around 60 amino acids (Figure 1.17D). They contain two zinc ions bound by a C_4HC_3 ligand motif with cross-brace ligation topology (Figure 1.18).[193] They were discovered more than a decade ago,[194] and are ubiquitous in eukaryotic cells. PHDs bind variously modified histone tails and are frequently found in chromatin remodeling factors. Specificity for different histone tail modifications, including methylation, acetylation and phosphorylation, is induced by sequence variations amongst the loops in PHD fingers.[195] Commonly, however, they recognise either methylation or absence of methylation at H3 K4. Mutations within these domains have been linked to numerous diseases, such as cancer and X-linked mental retardation.[196]

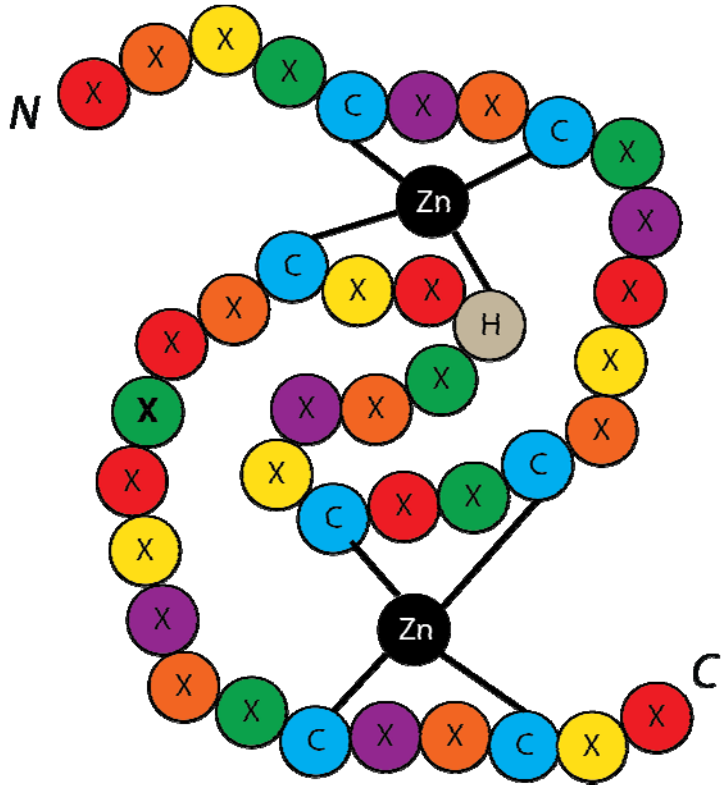


Figure 1.18 Cross-brace topology for zinc binding of PHD fingers. Circles represent amino acids, the key cysteine residues are shown in blue, the histidine in brown and Zn in black. X represents any amino acid.

1.1.6 Chromatin Remodelers

Nucleosomes are required to package DNA within a cell. However, for genes to be expressed, DNA must be accessible to other protein factors. This requires unpacking of heterochromatin at appropriate time points, for example during development or after DNA damage. The dynamic regulation of chromatin is carried out by chromatin remodeling proteins in cooperation with other chromatin-modifying enzymes.[197,198] Chromatin remodeling proteins utilise energy from ATP hydrolysis to package or unpackage chromatin by the movement, ejection or restructuring of nucleosomes on DNA.[199] They catalyse a number of nucleosomal processes, including nucleosome deposition, sliding and displacement (Figure 1.19).

There are four known families of chromatin remodelers, the SWI/SNF (switching

defective/sucrose nonfermenting) family,[200] the ISWI (imitation switch) family,[201] the INO80 (inositol requiring 80) family,[202] and the CHD (chromodomain, helicase, DNA binding) family.[203] All four families share five common features: 1) they contain a DNA-dependent ATPase domain; 2) they interact with nucleosomes; 3) they contain domains that recognise modified histone tails; 4) they contain regulatory domains; and 5) they contain binding domains that facilitate protein / protein interactions.[204] The specific function of these remodelers varies between the families; for example, ISWI family members are involved in nucleosome spacing, whereas the INO80 family is responsible for DNA repair and transcriptional activation.[205]

The complete mechanism of chromatin remodeling is currently unknown, but chromatin remodelers are thought to move DNA by anchoring to a fixed position on the nucleosome and then pumping DNA loops around the nucleosome.[206,207] Germline mutations in chromatin remodelers causes syndromes associated with errors in gene regulation, such as CHARGE syndrome,[208] and somatic mutations are often associated with cancer.[204]

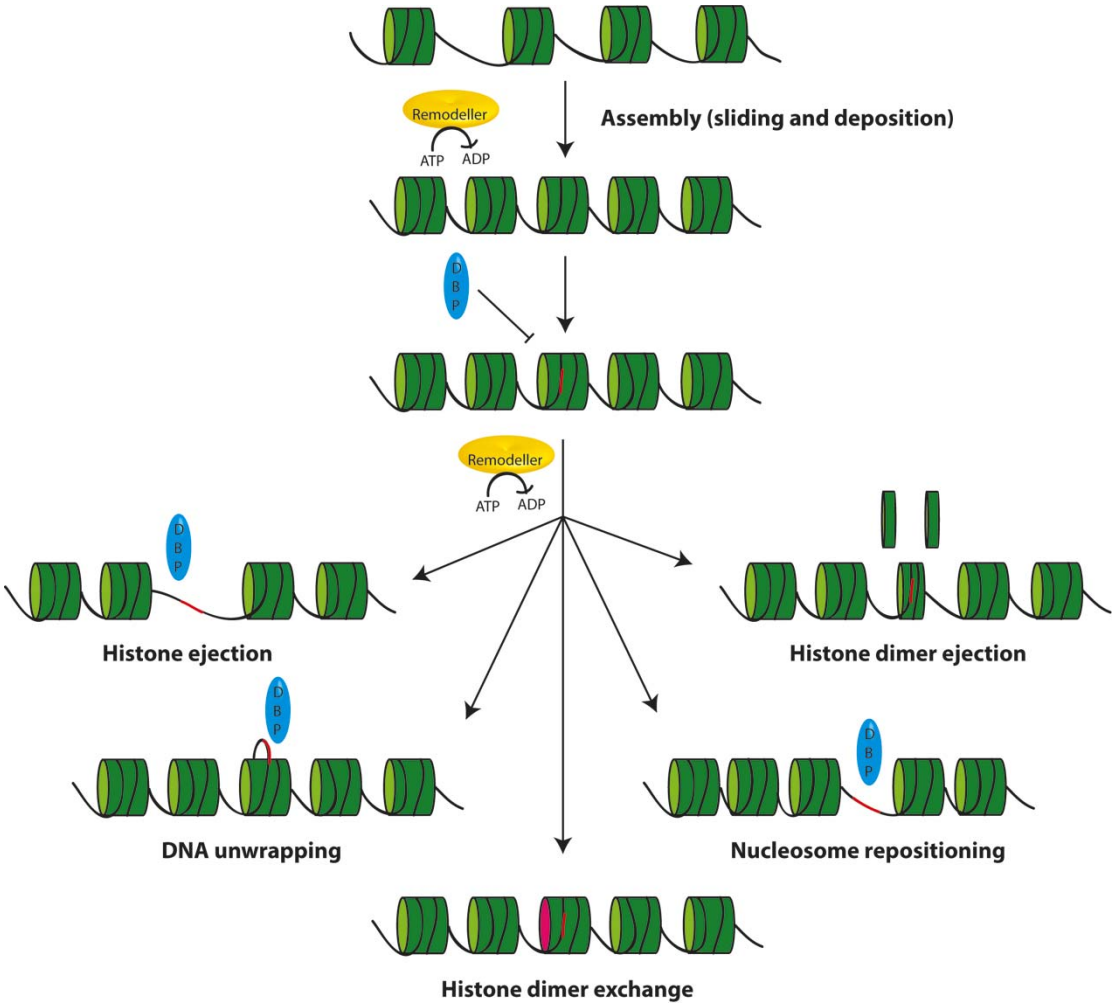


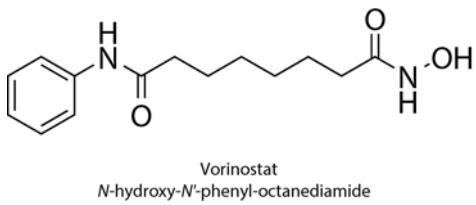
Figure 1.19 Scheme showing some of the functions of chromatin remodelers. Nucleosomes are represented by green cylinders and DNA by black lines. The occluded DNA is shown in red. The chromatin remodeler is shown as a yellow disk and a DNA binding protein as a blue ellipse. The pink nucleosome represents a nucleosome containing histone variants. Adapted from Cedric R. Clapier and Bradley R. Cairns.[199]

1.1.7 Roles of epigenetics in disease

Epigenetic modifications are crucial for cellular survival. Many mouse knockout (K/O) models of epigenetic writers and erasers have been made and these frequently result in embryo lethality or severe developmental phenotypes.[187,209,210] Epigenetic modulators are implicated in multiple diseases, including neuropsychiatric disorders, metabolic disorders and the immune response.[211] In addition, mutations and dysregulation of these proteins are also a hallmark of many cancers.[212] For example,

fusions of the *N*-terminus of the mixed lineage leukemia (MLL) protein to other protein binding domains frequently forms an oncoprotein that causes transcriptional activation by mistargeting of the SET domain, resulting in leukemia.[213,214]

Epigenetic modulations are dynamic and this makes them interesting targets for drug discovery. An HDAC inhibitor, Vorinostat (Figure 1.20), is already used in the clinic for treatment of cutaneous T-cell lymphoma,[215] and a number of other drugs targeting histone modifying proteins are in late stage clinical trials.[211]



Vorinostat
N-hydroxy-*N*'-phenyl-octanediamide

Figure 1.20 Structure of Vorinostat.

1.2 Objectives

The work described in this thesis concerns studies on the structures and functions of proteins involved in epigenetic regulation *via* histone binding and modifications (Figure 1.21). Histone binding / modifying proteins are mostly large multi-domain proteins whose functions in cells are likely regulated *via* multiple pathways. In many cases, there is evidence that individual protein domains have broad selectivity for different histone modifications. *In vivo* this specificity must be modulated so that catalysis only occurs in specific biological contexts.

The specificity of many chromatin interacting proteins is modulated by the presence of binding domains. The histone binding protein CHD4 contains two closely linked PHD fingers; No structures of the tandem CHD4 PHD fingers have been reported; however, structures of other tandem PHD fingers show them forming a single contiguous histone-binding domain.[216,217] The second PHD finger of CHD4 is reported to bind to H3 tails, although the nature of this binding and the importance of the other PHD finger remain unclear.[218] To address these questions, structural characterisation of the tandem PHD fingers of CHD4 was carried out to investigate relationships between the two PHD fingers and their interaction with histone proteins, using both an isolated CHD4 entity and in the context of more native-like protein constructs (Chapters 2 and 3).

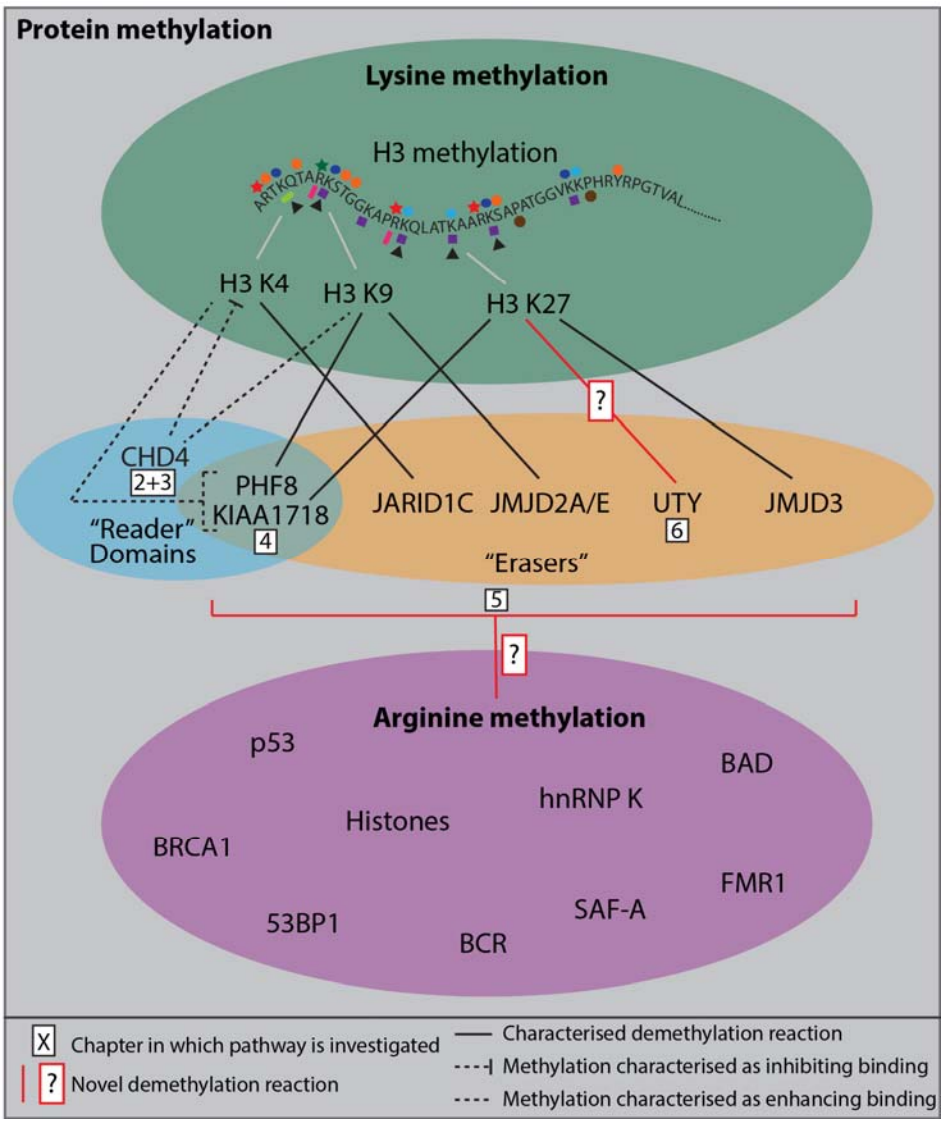


Figure 1.21 Summary of the pathways investigated throughout this thesis. Proteins shaded in green contain methylated lysyl-residues and those shaded in purple have methylated arginyl-residues. Proteins shaded in blue contain epigenetic "reader" domains that were investigated in work carried out in this thesis and those shaded in orange conatin epigenetic "eraser" domains.

The 2OG KDMs constitute the largest family of histone demethylases, of which many share similar specificities for *N*-methylated lysyl-residues in histones *in vitro*. Defining the function of these enzymes in cells can be complicated, especially as knockout/down of individual subfamily members can be compensated for by other related enzymes. Generation of these cell lines is also time-consuming and is limited to tractable cell types. Small molecule inhibition provides an alternative route to investigating 2OG KDM functions in cells; however, the development of selective

inhibitors and robust biochemical assays to identify them are essential for such studies. Biochemical analyses to investigate the activities, substrate selectivities and kinetics for the KDM7 subfamily of 2OG KDMs were carried out, as described in Chapter 4. The KDM7 subfamily, which contains three members, PHF8, KIAA1718 and PHF2, are of particular interest as their activities are likely regulated by the binding of post-translational modifications besides their methylated substrates.[176] Further, mutations of the KDM7 demethylase PHF8 have been linked to midline defects in both animals and humans;[219–222] however, the mechanisms by which such defects occur are unknown. It is also unclear whether the effects of PHF8 mutation may be alleviated by therapeutic inhibition. In order to help answer these questions, a robust assay that will facilitate high throughput inhibitor screening for the KDM7 subfamily was developed and used to identify KDM7 inhibitors (Chapter 4).

Unlike most other histone "writer" enzyme families and LSD1/2, the 2OG KDMs are presently thought to have only very few substrates besides histone lysyl residues. However, some members of the 2OG dependent oxygenase superfamily are known to have very broad substrate selectivities *in vivo* and are known to accept multiple proteins and / or residues as substrates (e.g. Factor Inhibiting HIF (FIH)).[137,173,223,224] Therefore, 2OG KDMs have the potential to react with substrates other than histones / *N*-methylated lysyl-residues. Investigations to determine whether 2OG KDMs can react with residues besides methylated lysines were therefore carried out (Chapters 4 and 5). It was intended that the results of these studies would give mechanistic insight into 2OG KDM catalysis and would inform on studies with 2OG KDM activities in cells.

UTY has previously been reported as a catalytically inactive member of the KDM6 subfamily of 2OG KDMs, despite having a fully competent Fe(II)-binding site and predicted 2OG-binding residues.[225–227] UTY is the Y-chromosome linked homologue of the X-chromosome linked enzyme, UTX, which is highly mutated in many cancers and is known to escape X-inactivation.[228] Therefore, the differential activities of UTY and UTX are of considerable interest, as such differences may be representative of physiological differences between males and females. It is known that male and female embryos have differential mortality rates upon knockdown of UTX,[209,229] potentially revealing that UTX catalysis/functions may be compensated by UTY. Studies on the biochemical characterisation of UTY are described, including attempts to identify its function(s) in cells (Chapter 6).

General questions that I aimed to investigate during the work described in this thesis were:

- Can *in vitro* experiments help dissect the functions of epigenetic proteins?
- Can selective, cell-permeable small molecule probes be developed which can provide a method for investigating the functions of these proteins in cells?
- What is the function of proteins containing multiple domains and do the domains within one protein affect the selectivities/functions of the protein as a whole?
- Is 2OG KDM catalysis limited to demethylation of histone lysyl-residues both *in vitro* and in cells?
- Is UTY a catalytically active 2OG KDM, or if not, what is the mechanistic explanation for this?

Chapter 2
Determining the structure of the tandem PHD domains of CHD4

2.1 CHD4

Chromodomain Helicase DNA binding protein 4 (CHD4, also known as Mi2β) is an ATP-dependent chromatin remodelling protein, which forms the main subunit of the Nucleosome Remodelling and Deacetylase complex, NuRD.[230] NuRD is a member of the CHD family of chromatin remodellers (Section 1.1.6, Chapter 1). The complex contains seven proteins that are associated with transcriptional repression.[230,231] There are 9 human CHD proteins that can be divided into 3 subfamilies based upon their domain composition.[203] CHD4 (and the closely related CHD3 / Mi2α and CHD5) contains double chromodomains (dCHD), an ATPase belonging to the SNF2 helicase family,[232] and double plant homeodomains (dPHD) (Figure 2.1).[233] The dCHD and dPHD modules are thought to target the ATPase motor to chromatin. The molecular mechanism of the remodelling activity of CHD4 and the complete process of targeting NuRD to specific chromatin sites are currently unknown.

NuRD catalyses a number of transcriptional regulatory processes including nucleosome remodelling and histone deacetylation, and is involved in developmental signalling and regulation of gene transcription.[234,235] CHD4 itself is required for T cell development,[236] and in patients with dermatomyositis it has been found to be an autoantigen.[237]

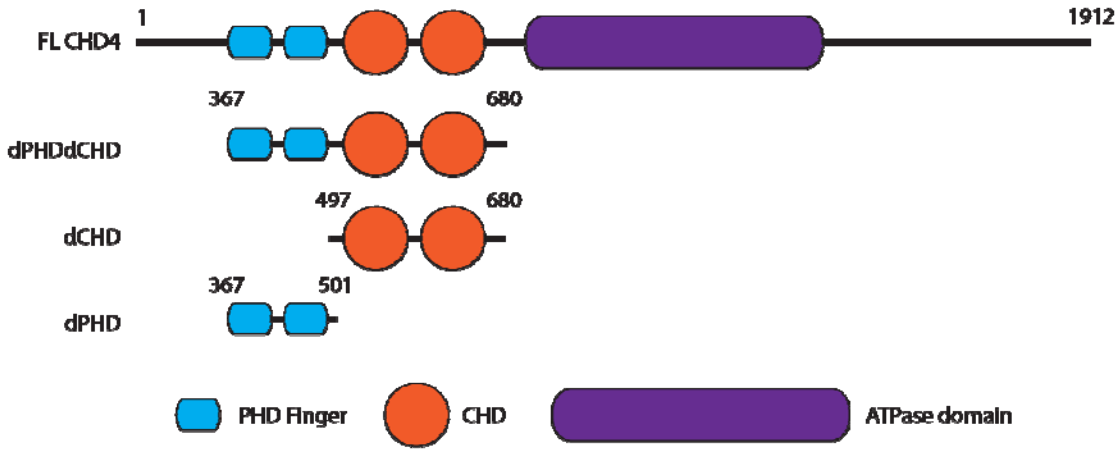


Figure 2.1 Domain organisation of full length (FL) CHD4. The structures of each construct used in this study are depicted underneath the full length protein schematic. Residue numbers are shown above each schematic.

The tandem CHDs of CHD4 bind to DNA in a sequence-independent fashion.[192] This is somewhat unusual, as the majority of characterised CHDs in the human proteome interact with histone tails.[238] The two canonical PHD fingers are separated by a 32 amino acid linker region that is rich in acidic amino acids. At the time this study was undertaken, the 2nd PHD domain of CHD4 had been reported to bind to the *N*-terminus of H3 tails.[218] The PHDs of CHD4 are important for its repressive function; mutations in either PHD of the dPHD module of CHD4 abrogate its repressive function, and expression of the double PHD fingers alone in cells disrupts pericentric chromatin in cells.[239]

2.1.2 Objectives

Although a solution structure of the 2nd PHD domain of CHD4 had already been solved, at the time that this project was undertaken no information on the structure of the 1st PHD finger, the connecting linker, or the conformation of the two domains together was available.[240] Previous attempts at crystallising CHD4 or subdomains of CHD4 have proved unsuccessful. Preliminary nuclear magnetic resonance (NMR) spectra (*collected by Professor Christina Redfield*), however, suggested that the

dPHD construct might be amenable to structure determination by NMR spectroscopy. The aim of the work described in this Chapter was to determine the solution structure of the tandem PHD domains of CHD4 by NMR spectroscopy. The relationship between the two domains and the structure of the connecting-linker was investigated.

Protein production was carried out by Dr. Rosa Morra and NMR spectra were collected by or in collaboration with Professor Christina Redfield, who also provided assistance with data processing and structure calculations.

2.2 Structure determination

Unlike X-ray crystal structures, where a single static protein conformation is observed, structures determined by NMR spectroscopy are normally represented by an ensemble of low energy structures that represent the variety of conformations that the protein can adopt in solution.[241,242] This makes NMR spectroscopy particularly suitable for proteins with large areas of flexibility, where a single static structure does not represent the native structure of the protein very well.

Structure calculations are carried out using molecular dynamics calculations that are restrained using theoretical (bond angles, lengths etc.) and experimentally determined restraints. Each calculation aims to minimise the energy of the generated structure, which is calculated based on how closely the structure obeys the theoretical and experimentally determined restraints. Structures that obey the restraints more tightly have lower energy and the calculations aim to reduce this as far as possible, reaching a global energy minimum that should best represent the structure of the protein being studied. Structure calculations are carried out in several stages; initial folding of the

structure begins from a randomly generated linear starting structure for the protein. Further rounds of refinement of the lowest energy structures from this initial stage are carried out after optimisation of the experimental restraints and addition of extra restraints. An outline of the procedure of structure determination by NMR spectroscopy is given in Figure 2.2.

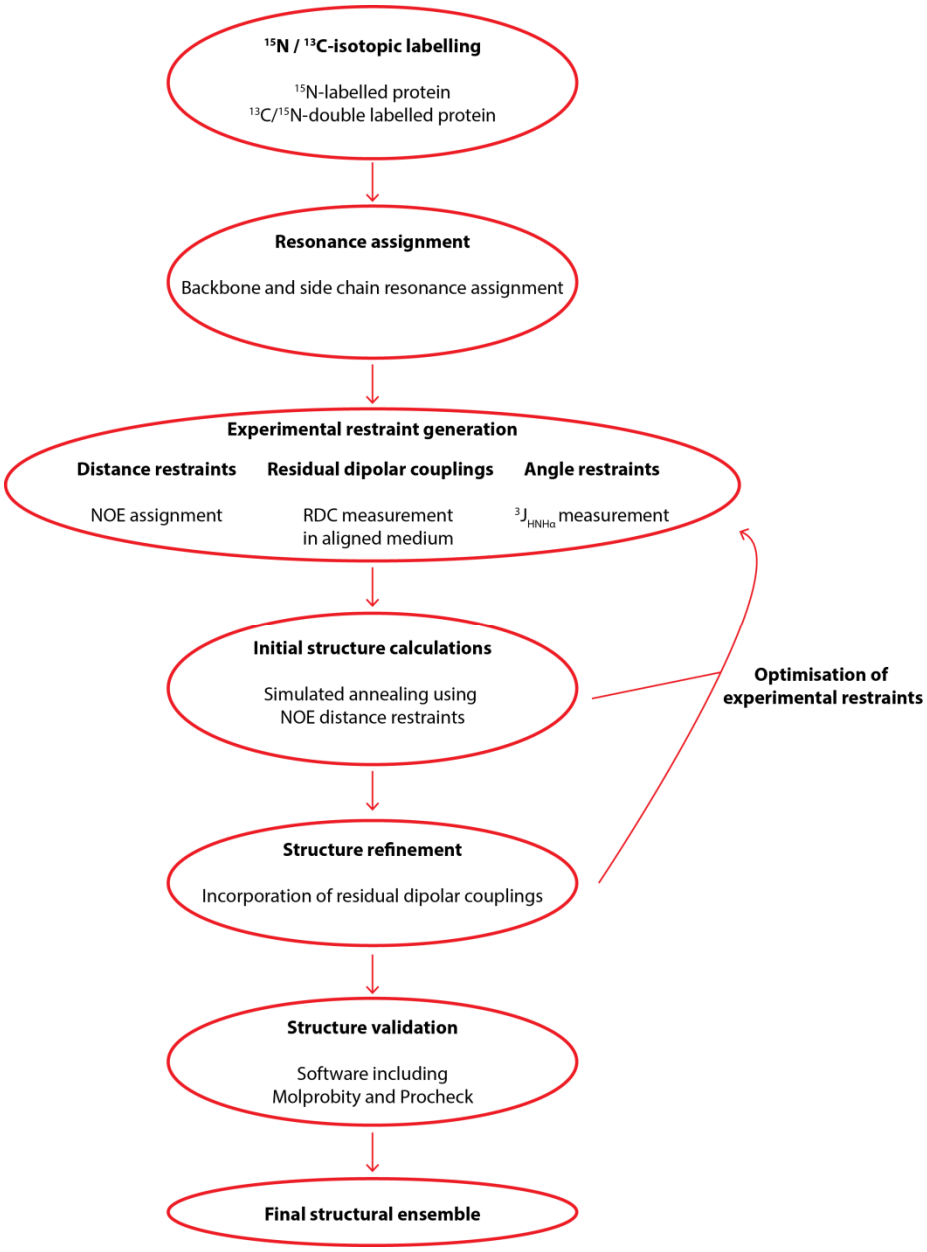


Figure 2.2 Outline of the procedure for generating a solution structure by NMR spectroscopy.

Structure determination by NMR spectroscopy initially requires complete backbone and side-chain assignments. Structural restraints can then be determined directly from

these chemical-shift values and by assignment of through-space interactions between atoms (intermolecular nuclear Overhauser effects (NOEs)[243]). The assignment procedure is simplified by use of multidimensional NMR, which spreads NMR peaks out over several dimensions and hence helps to reduce spectral overlap.[244]

To carry out multidimensional experiments on protein samples, incorporation of NMR-active forms of carbon (^{13}C) and nitrogen (^{15}N) (with a nuclear spin of ½) is required. To achieve this uniformly single- (^{15}N) and / or double- (^{15}N and ^{13}C) labelled protein is expressed in *Escherichia coli (E. coli)* cells. dPHD was expressed in M9 minimal media containing either $^{15}NH_4Cl$ and ^{12}C-glucose or $^{15}NH_4Cl$ and $^{13}C_6$-glucose (Cambridge Isotope Laboratories), to produce single- or double-labelled protein respectively) and purified first using Cobalt Talon™ resin, which binds specifically to the 8×His-tag of recombinant dPHD and then further on a 100 mL S75 size exclusion chromatography column (pre-equilibrated with buffer containing 20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT buffer) to homogeneity.[245] Protein was concentrated and either used directly for NMR experiments or flash frozen with liquid nitrogen in aliquots and stored at -80 °C (Section 8.5).

2.3 Spectral assignment

With labelled dPHD in hand, standard 2D and 3D NMR experiments were collected to allow complete assignment of the backbone and side-chain atoms of the protein (an outline of the experiments is shown in Figure 2.3) (Table 8.1). Assignments are made by a combination of spin system information gathered from experiments such as the ^{15}N-edited TOCSY-HSQC (Figure 2.3D) and by alignment of cross-peaks arising from experiments such as the HNCA and HNCO (Figure 2.3G and I respectively).[246]

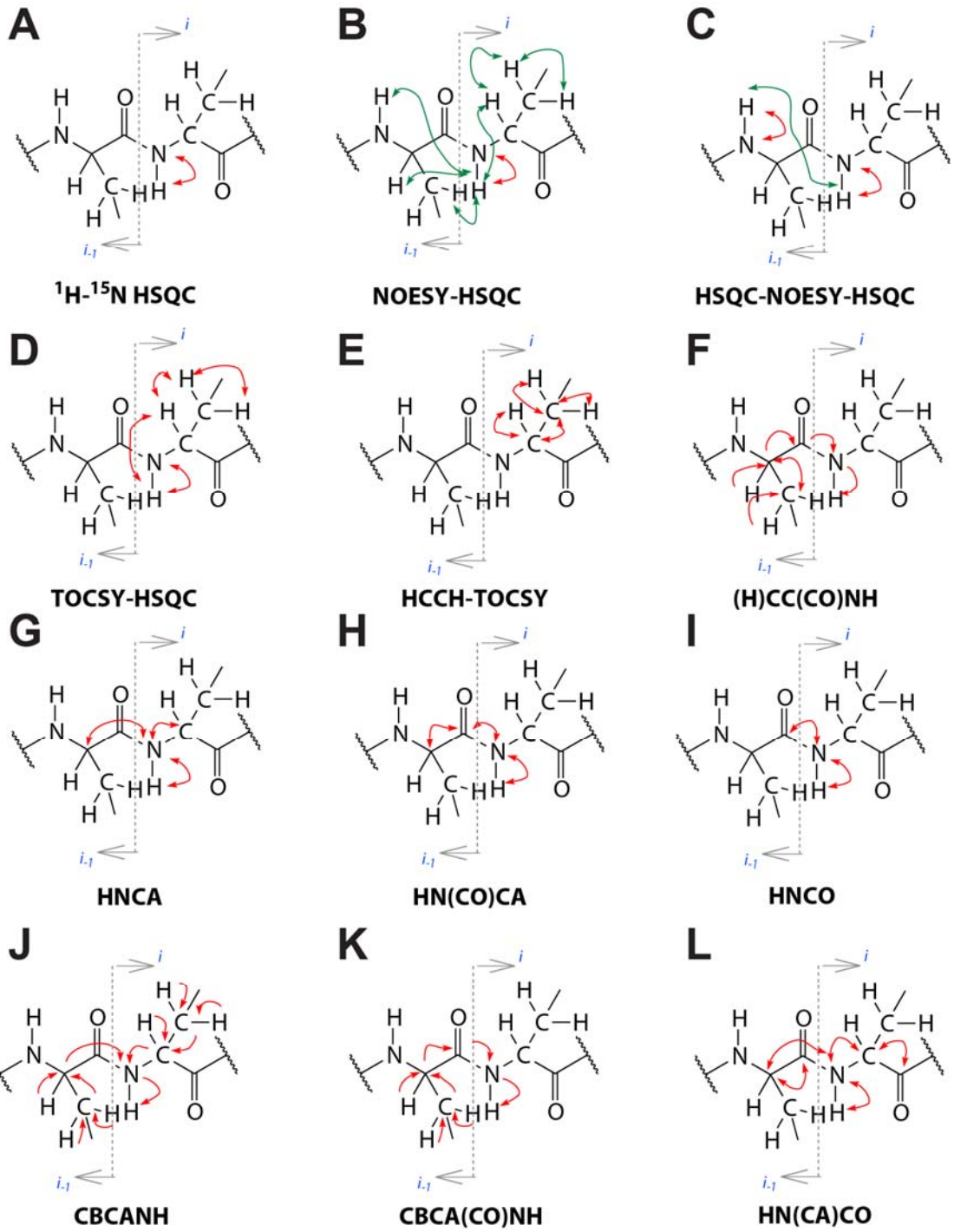


Figure 2.3 Illustration of the pulse transfer of standard 'triple resonance' NMR experiments. **A)** ^{1}H-^{15}N HSQC, **B)** NOESY-HSQC,[246] **C)** HSQC-NOESY-HSQC, **D)** TOCSY-HSQC,[246] **E)** HCCH-TOCSY,[247] **F)** (H)CC(CO)NH,[248] **G)** HNCA,[249] **H)** HN(CO)CA,[249] **I)** HNCO,[249] **J)** CBCANH, **K)** CBCA(CO)NH and **L)** HN(CA)CO.[250] Red arrows show through bond magnetisation transfer (J coupling) and green arrows through space transfer (cross-relaxation).

2.3.2 Backbone assignments

A 2D ^{1}H-^{15}N HSQC spectrum of dPHD was collected (Figure 2.4, *spectrum collected by Professor Christina Redfield*). This contained a single peak for the backbone NH of each residue (and thus does not have a peak for prolyl-residues), a peak for each tryptophan side-chain NH, usually found in the bottom left of the spectrum (green box, Figure 2.4) and a pair of peaks for the side-chain amide NH of each glutamyl- and asparigyl-residue, usually observed in the top right of the spectrum (blue box, Figure 2.4). Amino acids in unstructured regions of proteins adopt random-coil chemical shifts and are observed at the centre of the spectrum (red box, Figure 2.4). Protein structure alters the chemical environment experienced by the nuclei, causing observed chemical shifts to deviate from random-coil values, thus leading to spectral dispersion. Peak dispersion in the spectrum showed that dPHD was folded Figure 2.4.

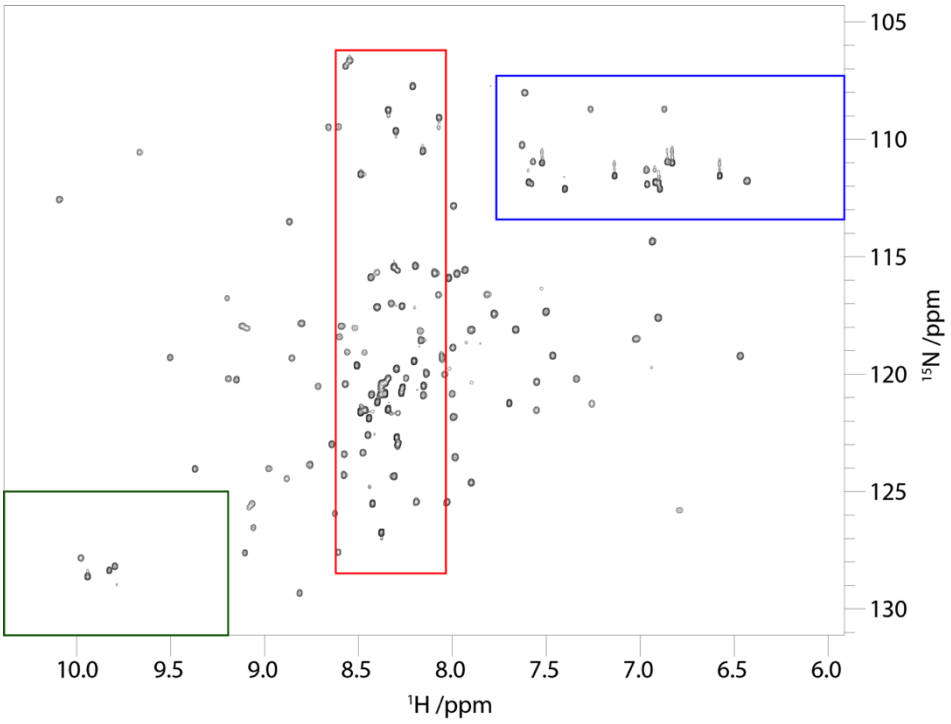


Figure 2.4 950 MHz ^{1}H-^{15}N HSQC spectrum of dPHD. The region of random-coil chemical shifts is marked by a red box, the region of tryptophan side chains by a green box and the region of side chain amide peaks by a blue box.

Backbone assignments were achieved using a combination of the standard "sequential assignment" and "triple-resonance" approaches. ^{15}N-edited 3D TOCSY-HSQC and NOESY-HSQC were used for initial assignments. The TOCSY-HSQC can be used to produce strips for each backbone H^N that display resonances corresponding to the side-chain protons of that residue. The chemical-shift values and number of peaks in each strip provide information on the identity of each amino acid.

Strips for G15, G16 and Q17 from the TOCSY-HSQC spectrum showed the α- and β-protons of three neighbouring residues (red peaks in Figure 2.5). The presence of two peaks in the region generally associated with α-protons suggested that the residues responsible for the first two strips were glycyl-residues. From the NOESY-HSQC experiment (black spectrum in Figure 2.5) NOEs between nearby backbone amides and side-chain protons can be identified, allowing them to be linked together. For example, NOEs to the α-protons of G15 can be observed from the strip corresponding to the backbone H^N of G16 and NOEs to the α-protons of G16 can be seen from the backbone H^N of E17. There were only three places in the dPHD sequence with two sequentially linked glycyl-residues and so this combination of information about the type of each amino acid and a number of residues that were linked together helped to localise groups of residues within the protein sequence.

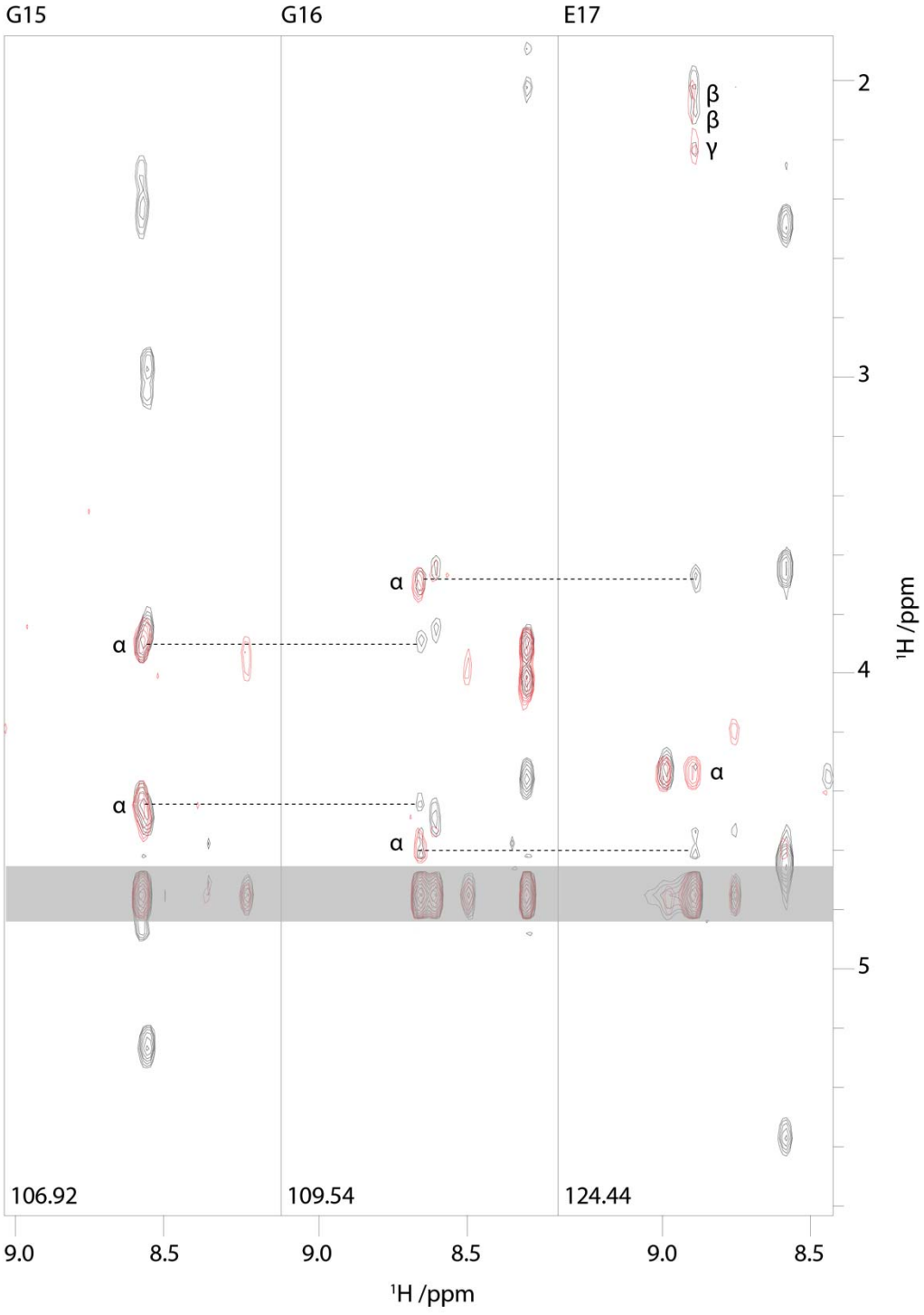


Figure 2.5 Strips from an overlay of a ^{15}N-edited 3D TOCSY-HSQC spectrum (red) and a ^{15}N-edited 3D NOESY-HSQC spectrum (black) for G15, G16 and E17. Dashed black lines show NOEs between the residues. The identity of each side-chain proton in the TOCSY-HSQC spectra is marked beside the peaks. The number in the bottom left corner of each strip gives the chemical shift of the backbone nitrogen in ppm. The grey area of the spectrum highlights peaks that are due to water.

Further assignments were made using pairs of spectra, one of which shows resonances only for the backbone or side chain of the previous residue, and a second, which gives resonances for its own backbone or side chain and that of the previous residue, allowing it to be connected with the strip for the neighbouring residue, which should contain peaks at the same chemical-shift values. This allowed residues to be paired up sequentially, and also provided a number of side-chain chemical-shift assignments. 3D HNCA and 3D HN(CO)CA spectra were used to connect residues *via* their alpha carbon (C^{α}) resonances (Figure 2.6A), and 3D HNCO and 3D HN(CA)CO spectra were used to connect residues *via* their backbone carbonyl resonances (Figure 2.6B).[249,250] Further, 3D CBCANH and 3D CBCA(CO)NH spectra were used to confirm the assignments and to assign beta carbon (C^{β}) resonances.

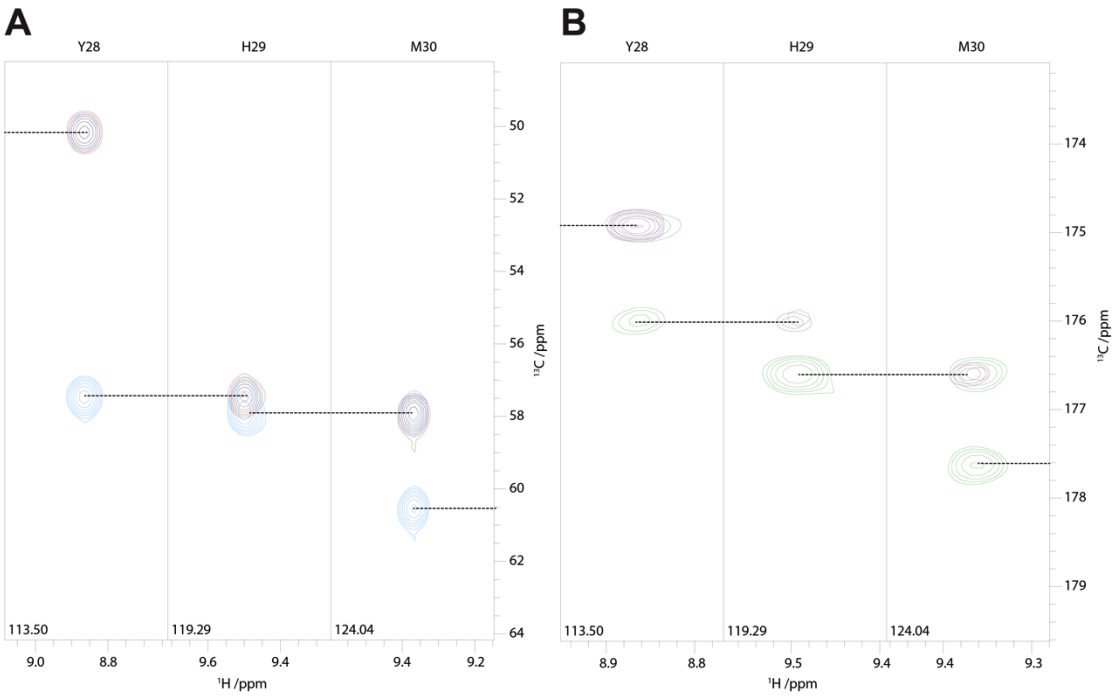


Figure 2.6 A) Overlay of an HNCA spectrum (blue) and an HN(CO)CA spectrum (red) for Y28, H29 and M30 and **B)** Overlay of an HNCO spectrum (green) and an HN(CA)CO (purple) for Y28, H29 and M30 The residue corresponding to each strip is shown above the spectra. Numbers in the bottom left of each strip give the backbone ^{15}N chemical shift in ppm.

Backbone $^{1}H^{N}$ and ^{15}N assignments were made for all residues within the two PHD fingers with the exception of *N*-terminal residues M1 and E2 and *C*-terminal residues

from K135 to K137, as well as all prolyl-residues in the construct, for which peaks are not observed in the ^{1}H-^{15}N HSQC spectrum (Figure 2.7). Within the linker, assignments for D80, D81 and H84 have not been obtained. Interestingly, peaks for residues in the linker region were mainly observed at the centre of the HSQC spectrum, with near random-coil chemical-shift values, suggesting that the linker may be unstructured.

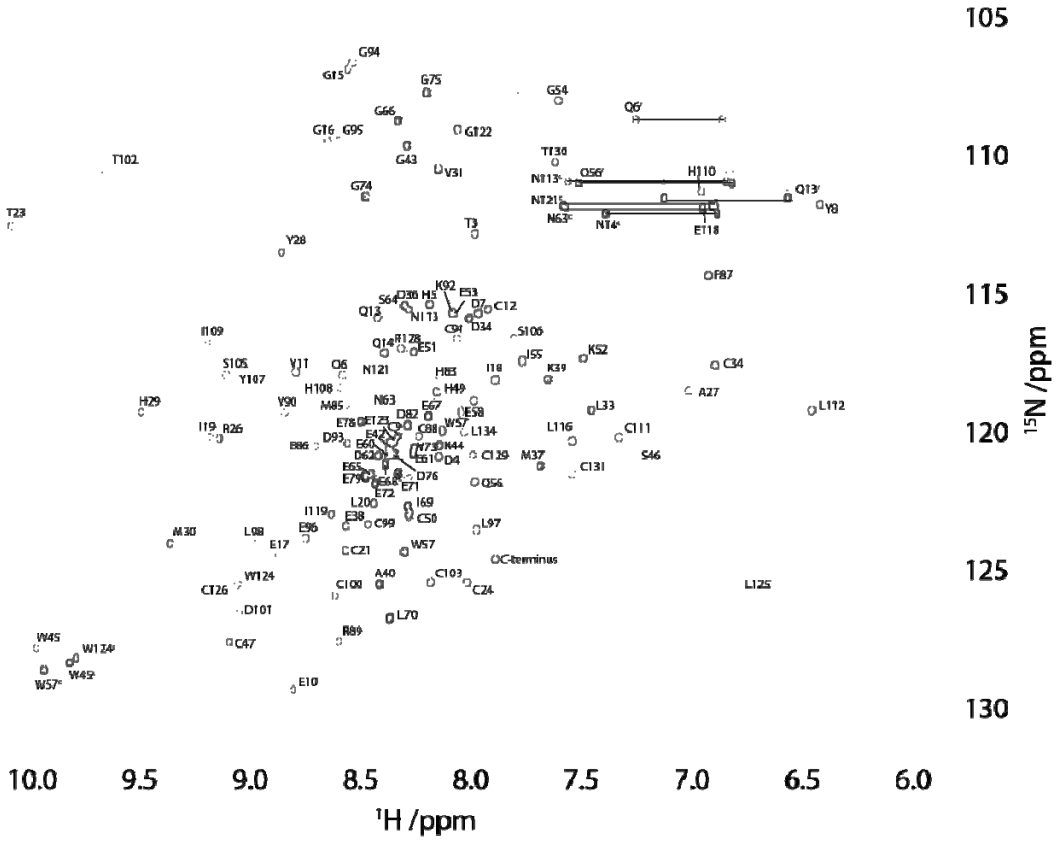


Figure 2.7 ^{1}H-^{15}N HSQC spectrum (950 MHz) of ^{15}N-labelled dPHD showing peak assignment for backbone amides. Black straight lines join the pairs of peaks observed for the side-chain amide NH of each glutamyl- and asparagyl-residue.

2.3.3 Side chain assignments

To produce restraints for structure determination calculations, full side-chain assignments are also required. ^{1}H side-chain assignments were made using data from the NOESY- and TOCSY-HSQC experiments and further ^{13}C side-chain assignments were made from the 3D (H)CC(CO)NH.[248] A 2D ^{1}H-^{13}C HSQC spectrum and

3D HCCH-TOCSY spectrum were used to correlate ^{1}H and ^{13}C assignments, which helped to extend the completeness of side-chain assignments.[247,251,252] Example strips from the 3D HCCH-TOCSY (purple spectrum) and the 3D ^{13}C-edited NOESY-HSQC (red spectrum) for the α- and β-protons of S46 are shown in Figure 2.8.

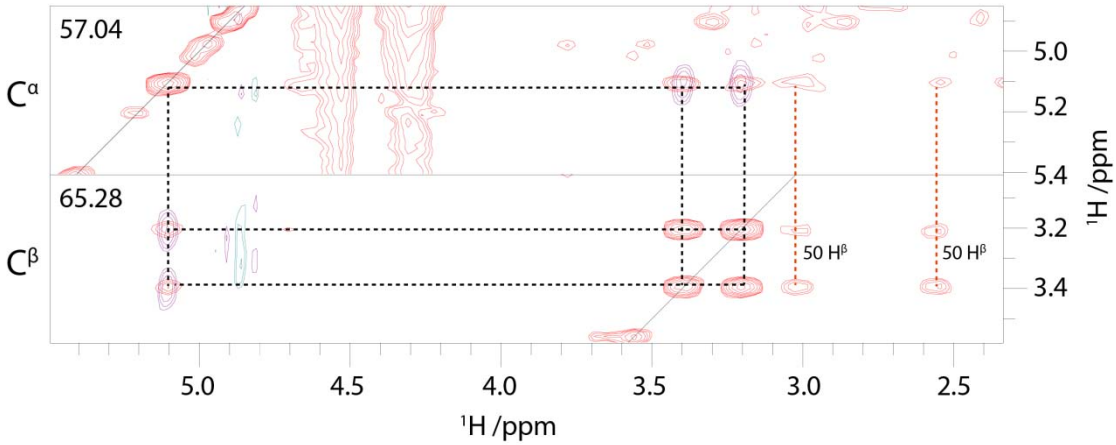


Figure 2.8 Strips from an overlay of a 3D HCCH-TOCSY spectrum (purple) and a 3D ^{13}C-edited NOESY-HSQC spectrum (red) for S46. The carbon that each strip corresponds to is marked to the left of the spectra and the ^{13}C chemical shift for each strip is marked in the top left of each strip in ppm. Dashed black lines show NOEs between the different atoms within the residue and dashed red lines show NOEs to atoms from different residues.

Aromatic protons were assigned using 2D ^{1}H experiments (2D ^{1}H-^{1}H NOESY and 2D ^{1}H-^{1}H TOCSY) carried out on an unlabelled dPHD sample (3 mM) (Table 8.1). The sample was exchanged into buffer made with D_2O so that backbone amide protons became deuterated and were no longer visible in the spectrum, reducing the spectral crowding.

The assignment statistics for dPHD are shown in Table 2.1 (full chemical-shift assignments are given in Appendix 1). Missing assignments are largely associated with residues in the linker region, which mainly showed chemical-shift values close to random-coil values and were thus hard to distinguish from each other, showing a high degree of overlap, which made it hard to assign them to a specific residue.

Table 2.1 Chart showing completeness of dPHD chemical-shift assignments. Statistics are for the 137 native residues of CHD4 included in the dPHD construct.

Category	Available	Assigned	% Assigned
All residues	137	133	90
Amide	161	131	97
Backbone	535	480	90
Backbone non-^{1}H	411	362	88
Side chain ^{1}H	683	586	86
Side chain non-^{1}H	410	240	59
Type $^{1}H^{N}$	124	118	95
Type $^{1}H^{\alpha}$	148	141	95
Type $^{1}H^{\beta}$	234	222	95
Type $^{1}H^{\gamma}$	157	120	76
Type $^{1}H^{\delta}$	88	66	75
Type $^{1}H^{\epsilon}$	46	28	61
Type $^{1}H^{\zeta}$	7	6	86
Type $^{1}H^{\eta}$	3	3	100
Type ^{15}N	137	118	86
Type $^{15}N^{\delta}$	10	3	30
Type $^{15}N^{\epsilon}$	17	7	41
Type $^{15}N^{\zeta}$	7	0	0
Type $^{15}N^{\eta}$	6	0	0
Type $^{13}C'$	137	116	85
Type $^{13}C^{\alpha}$	137	128	93
Type $^{13}C^{\beta}$	126	119	94
Type $^{13}C^{\gamma}$	112	70	63
Type $^{13}C^{\delta}$	88	38	43
Type $^{13}C^{\epsilon}$	28	3	43
Type $^{13}C^{\zeta}$	13	0	0
Type $^{13}C^{\eta}$	3	0	0

2.4 Dihedral angles

Torsion angles are used to define the geometries of the peptide backbone and side chain (Figure 2.9). NMR experiments can be used to gain valuable information on the angles adopted for each residue of the protein for use in structure calculations.

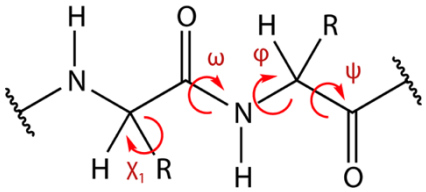


Figure 2.9 Illustration of four of the torsion angles defined for the peptide backbone and side chain. The angles are shown in red.

2.4.2 The HMQC-J for measuring dihedral angles

NMR coupling constants can be used to gain information on torsion angles. The Karplus equation (Equation 2.1) relates the ^{3}J-coupling constant directly to torsion angle.[253] An HMQC-J spectrum can be used to determine the $^{3}J_{HNH\alpha}$ coupling constants.[254] This experiment contains a cross-peak for each backbone NH in the protein, which is split into a doublet in the ^{15}N dimension (except for glycyl-residues that give doublets of doublets). The magnitude of this splitting can be measured and corresponds to the ^{3}J coupling. Splitting values less than 6 Hz are either characteristic of α-helices, with ϕ = -60° ± 30°, or a positive ϕ angle (ϕ = +60° ± 30°) and values greater than 8 Hz characteristic of β-sheet, with ϕ = -120° ± 40°. Due to averaging from intramolecular motion, dihedral angles cannot be assigned for splitting values between 6 Hz and 8 Hz. By use of an HMQC-J, ^{3}J-coupling constants were assigned (Section 8.15). For residues where the splitting was less than 6 Hz, dihedral restraints were not initially assigned due to ambiguity over whether they corresponded to +60° or -60°. After initial structure calculations, the ϕ angles for each residue in 15 low-energy structures were measured and the clustering of angles caused by the NOE restraints used to determine which angle was most likely. Where restraints already in use did not define any clustering of ϕ, no dihedral restraint was included in the calculations.

Equation 2.1 The Karplus equation. $J(\phi)$ is the ^{3}J coupling constant, ϕ the dihedral angle and A, B and C empirically determined constants that depend upon the atoms and substituents involved.

$$J(\emptyset) = A\cos^2\emptyset + B\cos\emptyset + C$$

2.4.3 Use of TALOS-N for dihedral angle predictions

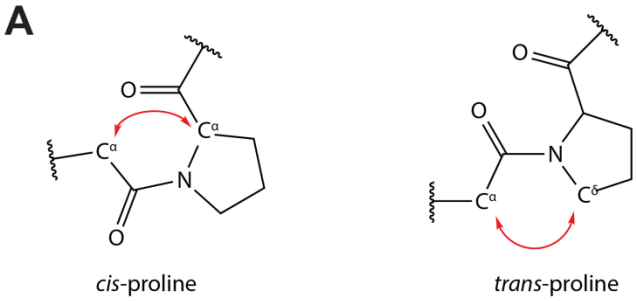
The absolute chemical-shift values for each residue of a protein can also provide

information on the Φ, Ψ and X_1 angles. Using H^N, H^α, C^α, C^β, CO and N chemical shifts and the sequence of the protein of interest, the programme TALOS-N uses artificial neural networks and a large database of structures with experimentally available chemical shifts to make angle predictions.[255] Assigned chemical shifts for dPHD were exported for use with TALOS-N, which produced angle restraints for all non-mobile residues (Section 8.15). Although clustered predictions were made for many residues, when these were used as dihedral restraints in structure calculations they produced a large number of violations. In a number of cases the values did not correlate with the values measured from the HMQC-J spectrum. TALOS-N makes predictions on the basis of similar structural motifs in structures within its database and is most reliable when there is secondary structure present. As PHD fingers do not have large regions of defined secondary structure it was suspected that the predictions might not be very reliable. In addition, each PHD contains two zinc ions, which will affect the chemical-shift values of the eight Zn(II)-ligating residues, but which are not included in the sequence considered by TALOS-N. These might, therefore, lead to inaccurate predictions. As a result, it was decided only to use HMQC-J dihedral restraints in the structure calculations described below.

2.4.4 Cis/Trans prolines within dPHD

The ω angle can adopt only two possible values, 0° (*cis*) or 180° (*trans*). Due to steric clashes between the amino acid side chains that occur if the bond is *cis*-, the *trans*-conformation is very significantly preferred. However, the presence of the ring in prolyl-residues means that for peptide bonds including a proline, the two conformations are of more similar energy and hence can adopt either a *cis*- or a *trans*-conformation. Although most peptide bonds containing prolyl-residues are

trans- about 5.2% of Xaa-Pro peptide bonds are *cis*-, compared to 0.02% of Xaa-Xaa (not Pro) peptide bonds.[256] *Cis*-prolyl-residues can be identified by inspection of their C^{β} and C^{γ} chemical shifts and observation of a characteristically strong H_i^{α}–H_{i+1}^{α} NOE, which is not observed for a *trans*-bond (Figure 2.10A).[257–259]



B

Proline residue number	C^{β} chemical shift	C^{γ} chemical shift
25	32.00	27.20
35	32.22	27.17
41	32.38	26.98
48	32.31	28.10
77	32.28	27.34
104	31.58	27.54
114	33.00	25.07
115	32.38	27.41
120	32.35	27.08
127	32.30	Not assigned
132	32.14	27.27

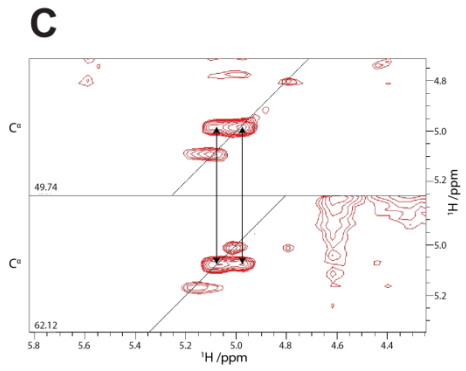


Figure 2.10 Identification of *cis*-prolyl-residues in dPHD. **A)** Two possible conformations for the ω angle of an Xaa-Pro peptide bond showing the close C^{α}–C^{α} or C^{α}–C^{δ} distance. **B)** C^{β} and C^{γ} chemical shifts of all prolyl-residues. Unusual chemical shifts are coloured red. **C)** Strips from a ^{13}C-edited NOESY-HSQC for the C^{α} of Asn113 and C^{α} of Pro114 showing the strong H^{α}–H^{α} NOE. The ^{13}C chemical shift for each strip is given in the bottom left corner in ppm.

By analysing the chemical-shift values for the assigned prolyl-residues of dPHD and then confirming by relating to the assigned NOEs, Pro114 was identified as being *cis*- (Figure 2.10B and C). All other prolyl-residues in dPHD adopt the *trans*-conformation. The published structure of the second PHD of CHD4 does not have the *cis*-conformation for this prolyl-residue.[240] Although the construct used in the previous study may differ in this region, it is also possible that the *cis* bond was missed, as the structure was determined using only proton NMR. Proton NMR does not allow determination of carbon chemical shifts and may have made the H_i^{α}–H_{i+1}^{α} NOE, which is close to the diagonal, difficult to observe in the spectra, which will

have been considerably more crowded as a result of being only 2D rather than 3D.

2.5 Backbone dynamics

To date, the only published structures of tandem PHD fingers are from proteins where the two domains have been found to be connected by a rigid linker, with a direct interface between the two PHD fingers so that they form a single contiguous binding surface.[216,217] Mostly random-coil chemical-shift values for linker residues already suggested that the dPHD linker was unstructured (Figure 2.7), and with backbone assignments in hand it was possible to further investigate the linker rigidity of dPHD.

2.5.1 Heteronuclear NOE ratio

The (^{1}H-^{15}N) heteronuclear NOE experiment allows investigation of the mobility of the backbone N-H bond on a fast timescale.[260] Unlike other experiments carried out in this Chapter, the pulse sequence begins with a pulse on the backbone nitrogen and the peak intensity thus relates to their equilibrium population. A pair of spectra is collected with and without pre-saturation of the ^{1}H. The saturation of the ^{1}H leads to perturbation of the ^{15}N population, and thus the peak intensity, as the ^{1}H population attempts to relax to its equilibrium state. The ratio of the peak intensities in the two spectra gives the heteronuclear NOE ratio ($\frac{I_{sat}}{I_{no\ sat}}$). Since the ^{15}N relaxation rate is related to bond mobility, this ratio provides an indication of the flexibility of each residue. Ratios greater than 0.75 correspond to highly restricted internal dynamics on a picosecond to nanosecond timescale, which can be equated to rigid backbone residues. Anything smaller than this implies some residue mobility; a ratio of less than 0.6 is often considered the threshold for a residue being mobile and a negative

value suggests that the given residues are extremely mobile.

2.5.2 Measuring the heteronuclear NOE ratio of dPHD

Heteronuclear ^{1}H-^{15}N NOE ratios for dPHD were measured by Professor Christina Redfield (Table 8.1). The measured heteronuclear NOE ratio was plotted against sequence number (Figure 2.11). This showed that the dPHD construct consists of two largely rigid domains separated by a 32 residue, very mobile linker, as demonstrated by the small and / or negative heteronuclear NOE ratios. The heteronuclear NOE plot also showed a small area of flexibility in both domains, from residue 38 - 44 in the first domain and residue 122 - 124 in the second domain. This likely corresponds to a flexible loop on the surface of each PHD finger.

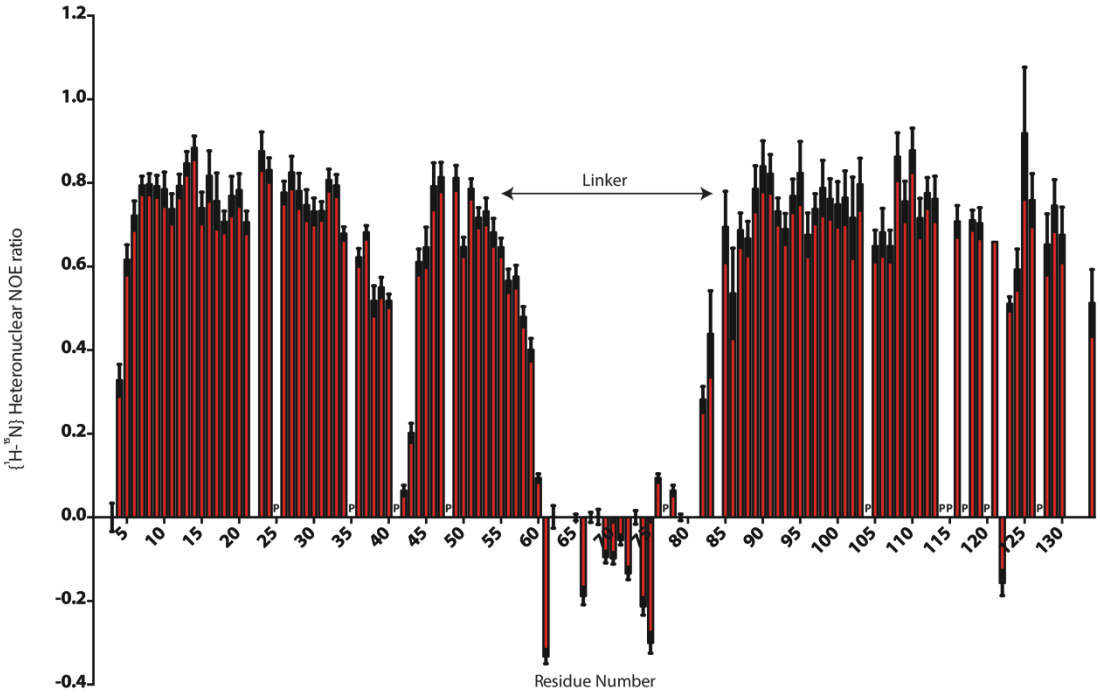


Figure 2.11 ^{1}H-^{15}N Heteronuclear NOE ratio plotted as a function of dPHD sequence.

Although this experiment demonstrated that the linker itself is flexible, there are still two possible conformations for the two PHD domains that are consistent with this

observation, depending on whether the two domains interact directly with each other (Figure 2.12A) or are separate distinct entities (Figure 2.12B). To determine which of these was the case, NOE assignment and structure calculations were required.

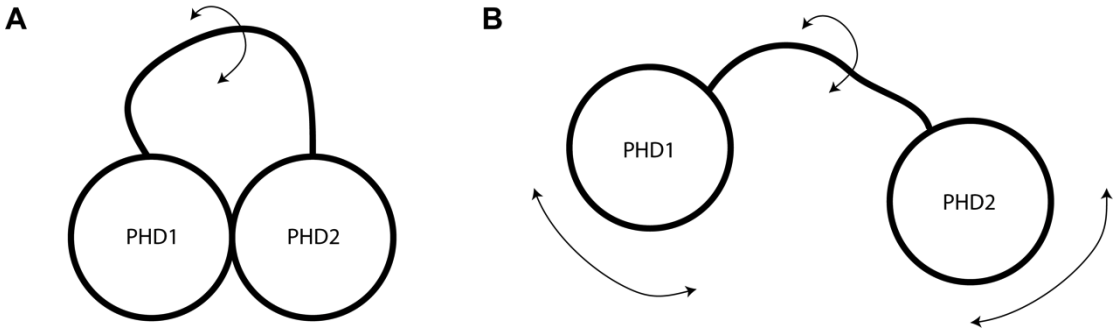


Figure 2.12 Two possible arrangements of the tandem PHD domains of CHD4 consistent with presence of a flexible linker. **A)** Directly interacting PHD domains. **B)** Independently oriented PHD domains. Arrows show the potential independent movement of the two PHD domains.

2.6 dPHD NOE assignment and restraint generation

The most significant set of structural restraints used in NMR structure calculations come from distance restraints created on the basis of assignment of intermolecular NOEs.[261] These NOEs are caused by dipole-dipole induced relaxation introduced by spins that are close in space (< ~5 Å for observation of significant NOEs). The intensity of the observed NOEs can be correlated with a distance between the two interacting protons responsible for the peak. The dipolar relaxation rate is directly proportional to the inverse sixth power of the distance separating the two interacting protons. Although in theory the intensity of the NOE should therefore correlate with an absolute internuclear distance, extra contributions to the intensity from dynamics and spin diffusion mean that in practice this direct correlation is not possible. Instead, NOEs are grouped by intensity and placed into distance bins. Knowledge of some relatively standard distances for which NOEs are observed, such as the $^{1}H^{N}_{i}$-$^{1}H^{\alpha}_{i-1}$ NOE is used to calibrate the NOE intensity thresholds for each distance bin. For the

calculations described below, the upper distance bounds for each bin were set at 2.8 Å, 3.5 Å, 5 Å and 8 Å. As NOEs were derived from several different spectra, the NOE intensity thresholds for each distance bin were determined individually for each spectrum. In addition, for unknown reasons, peaks from the 2nd PHD were weaker than those observed from the 1st PHD, so distances for NOEs from the two halves of the protein were calibrated separately.

For dPHD NOEs were assigned in the 2D ^{1}H-^{1}H NOESY-HSQC spectrum and both the ^{15}N- and ^{13}C-edited 3D NOESY-HSQC spectra (150 ms mixing time) by reference to the assigned backbone and side-chain chemical shifts (Section 8.15). Figure 2.13 shows strips from the ^{15}N-edited NOESY-HSQC for three neighbouring residues with the identity of the proton responsible for the observed NOEs labelled. In some cases, reciprocal NOE cross-peaks were observed in the spectra, for example between the backbone NH peaks of M30 and V31 (four NOEs connected by a square of dashed red lines in the bottom right of Figure 2.13), helping to confirm correct NOE assignment. Many of the NOEs detected are between residues that are close together in sequence, but folding of the protein also allows observation of NOEs between residues that are far apart in the sequence, but close in space, such as the NOEs from the backbone NH of H29 to the alpha- and beta-protons of C9 seen in Figure 2.13. These long-range NOE distance restraints are crucial for successful structure calculations as they help to create the global fold of the protein.

No NOEs were observed between residues in the two PHD fingers. If the two domains were in close proximity, with a fixed interface, NOEs corresponding to interactions between residues on these two joined surfaces would be expected. The

absence of these NOEs suggested that the two PHDs do not interact (Figure 2.12B).

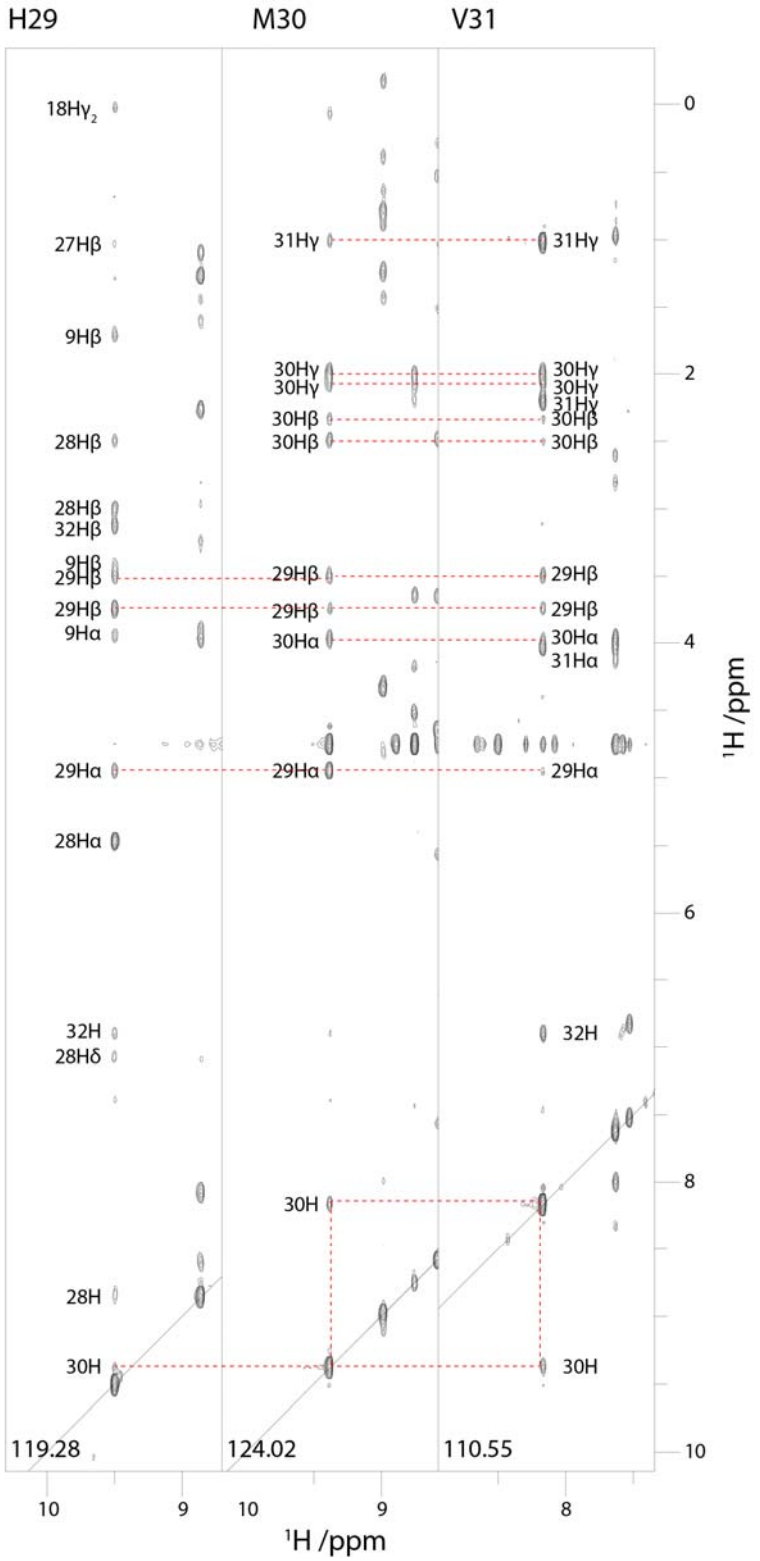


Figure 2.13 Strips from a ^{15}N-edited 3D NOESY-HSQC (950 MHz, 150 ms mixing time) for H29, M30 and V31. The residue corresponding to each strip is shown above the spectra. The chemical shift of the backbone ^{15}N is given in the bottom left of each strip in ppm. NOEs are between the backbone H^N and the proton labelled beside each peak in the spectrum. NOEs between residues of the neighbouring residues are marked by dashed red lines.

2.6.2 Defining restraints to maintain correct zinc-binding geometry

PHD fingers contain two structural zinc ions. As zinc ions are undetectable by NMR, structural restraints for them cannot be collected experimentally. However, in order to maintain these ions in the correct positions within the structure, the addition of restraints between the zinc-binding residues and the zinc ions is required. Zinc ions usually maintain tetrahedral coordination, and as the zinc coordination of PHD fingers is well established, a patch was generated to create bonds and angles between the zinc ions and the dPHD zinc-coordinating residues (C9, C12, H29 and C32 for the 1st Zn(II) ion and C21, C24, C47 and C50 for the 2nd Zn(II) ion in domain 1 and C88, C91, H108 and C111 for the 1st Zn(II) ion and C100, C103, C126 and C129 for the 2nd Zn(II) ion in domain 2) in order to maintain the correct geometry.

2.7 Simulated annealing calculations

To determine a structure by NMR spectroscopy a random extended protein structure is initially generated and simulated annealing is used to generate an energy-minimised, folded protein structure based upon covalent geometry, non-bonded contacts, the RAMA potential (which biases angles towards commonly observed angles from known structures) and experimentally-derived restraints. Simulated annealing uses a Monte Carlo approach to fold the protein, minimising its energy based on a given set of structural restraints.[261] Energy arises from differences between the restraints included in the calculations and the actual values for these terms observed in the final calculated protein structure. The energy for a given structure is calculated using Equation 2.2, where a different weighting is used to calculate the energy arising from each type of structural restraint (for example the

energy penalty for not obeying bond length and angle restraints is very severe, whilst the penalty for disobeying experimental NOE distance restraints is less).

Equation 2.2 Equation for calculating the energy of a given structure. E_x represents the energy derived from violations of x restraints.

$$E_{total} = E_{angle} + E_{bind} + E_{improper} + E_{vdW} + E_{NOE} + E_{dihedral} + \cdots$$

Xplor-NIH 2.29 was used for the structure calculations described below (Section 8.15.2).[262,263] Calculations were repeated multiple times to generate a family of similar, low-energy structures, which represent the structure of dPHD in solution. The overall scheme for the structure calculations is illustrated in Figure 2.14. For the first *ab initio* calculations, standard covalent geometry and non-bonded contacts were used, in addition to NMR derived distance restraints, bonds and angles to the Zn(II) ions for zinc-binding residues and a RAMA energy potential (which helps restrict dihedral angles to favourable regions of the Ramachandran diagram, improving the stereochemical quality of the NMR structures)[264]. The lowest-energy full-length structures generated using the simulated annealing protocol were then divided into the two separate domains (residues 1 – 77 and residues 76 – 137) and each domain refined individually with an increased van der Waals radius. The lowest-energy structure from each of these refinements was finally refined with RDC and dihedral angle restraints to produce 30 structures. For each domain the lowest-energy structures from these calculations were then chosen to form the final ensemble of 20 structures that represent the structure of the domain in solution.

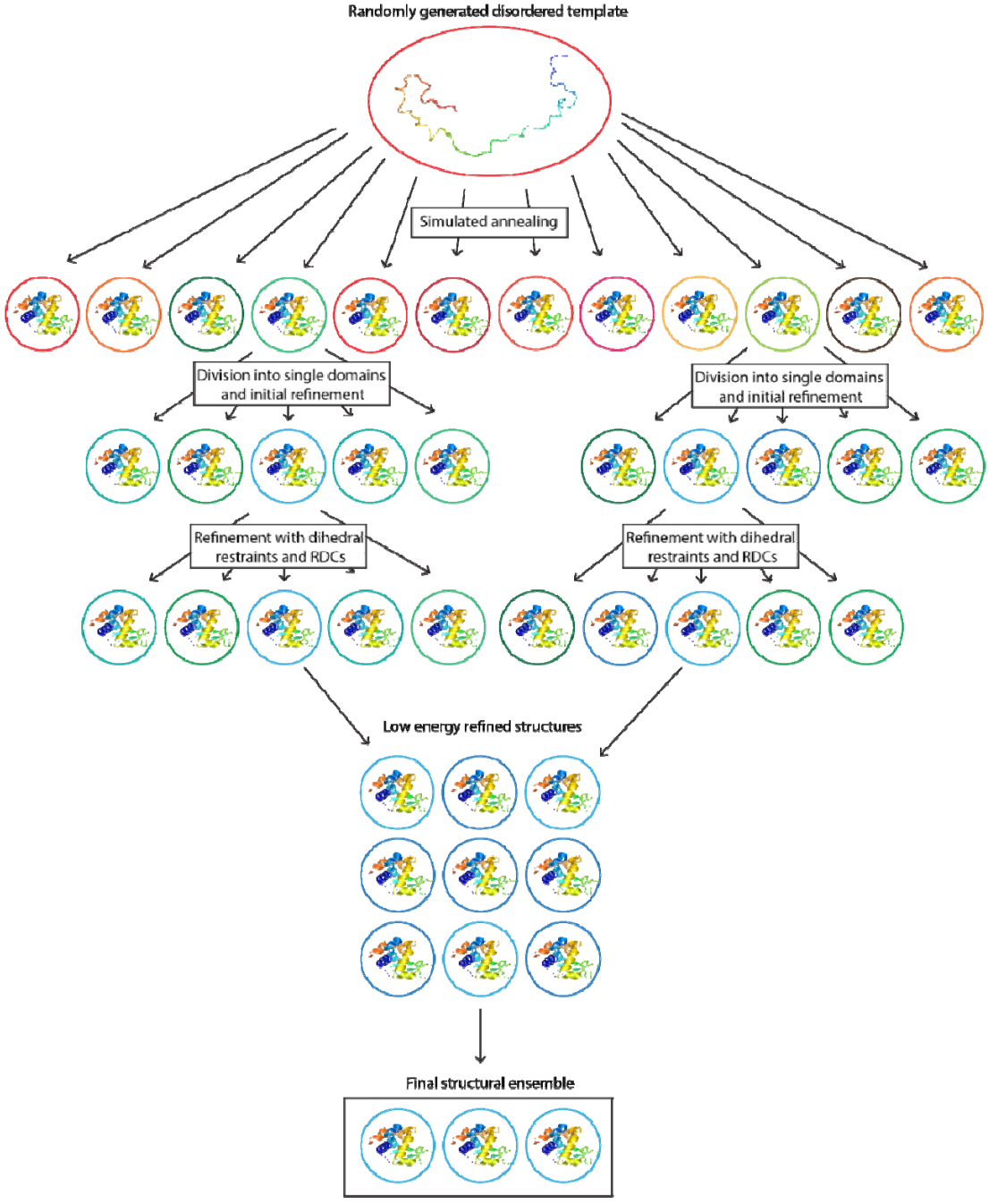


Figure 2.14 General scheme for generation of a structure ensemble by NMR spectroscopy. After each stage of structure calculation the lowest-energy structures were selected for use in the next round. After the simulated annealing calculations, dPHD structures were divided into the two separate PHD fingers for subsequent refinement. Red circles represent high-energy structures, green circles intermediate-energy structures and blue circles low-energy structures.

To optimise the experimental restraints, multiple initial rounds of simulated annealing and refinement calculations were carried out. Restraint violations for low-energy structures were analysed and where these occurred repeatedly the restraints were

checked for mistakes or misassignments in the initial analyses of the NMR spectra. In addition, the problem of stereospecific assignments was addressed and is described below.

2.7.2 Use of swapping to define stereospecific assignments

A number of amino acid residues contain prochiral atoms. Within the structure calculation this prochirality is maintained by improper angle restraints, but from NMR chemical shifts alone it is not possible to stereospecifically assign the groups. However, where the two prochiral protons are in different chemical environments, they can give rise to different chemical shifts and NOEs. One method of dealing with this is to assign a single ambiguous distance restraint to both protons. This removes the problem, but leads to a loss of information that can reduce the quality of the calculated structures. Therefore, to generate stereospecific restraints where possible, a swapping procedure was carried out utilising floating chirality (Section 8.15).[265] At each step of the simulated annealing procedure the improper restraints that maintain prochirality were removed to allow for random swapping of the two prochiral protons. If swapping resulted in a lower-energy structure this swapped configuration was retained, if it did not then the swapping was reversed. At the end of the calculation the final configuration of each prochiral proton was given in an output file.

Two NOE restraint files were generated for dPHD with the prochiral pairs initially stereochemically assigned opposite ways around in the two files. 100 simulated annealing structures were calculated with each set of restraints using this swapping procedure. The number of times each residue was swapped in the final structures of each set of 100 calculations was determined. For residues that were swapped more

than 70% of the time in one set of calculations and less than 30% of the time in the other calculation a stereochemical assignment was used based upon the assignment in the file that was not swapped frequently. However, where the residues were swapped between 30% and 70% of the time or where swapping occurred in neither calculation the restraints were merged into a single ambiguous restraint that had to be obeyed by either one of the two prochiral protons. On the basis of the swapping procedure, 57 prochiral pairs in dPHD were stereospecifically assigned in the NOE distance restraints for the final set of structure calculations.

2.8 Refinement

The 970 structures produced by simulated annealing were sorted according to their overall energy and the 73 lowest-energy structures used for further structural refinement. These 73 structures had no large violations of the theoretical restraints, indicating that the covalent geometry was not distorted by the experimentally derived restraints, and few NOE restraints consistently violated by > 0.3 Å. The refinement procedure was similar to the simulated annealing procedure except that it began from the coordinates of the low-energy structures generated previously and only allowed for small changes in the overall protein structure. Since the two domains have been shown to have no correlation with each other, in order to increase the speed of the calculations and to increase the likelihood of obtaining good, low-energy structures for each domain, refinement calculations were carried out for each PHD finger individually. Initial refinements of each domain from the simulated annealing NMR structures were carried out using an increased van der Waals potential (to reduce non-bonded clashes) and dihedral angle restraints, in addition to the restraints used in simulated annealing calculations. 30 structures for each domain were generated from

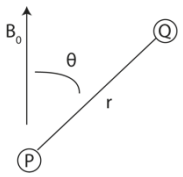
each low-energy simulated annealing structure and the lowest-energy structure from each of these calculations was then further refined by inclusion of residual dipolar coupling restraints.

2.8.1 Residual Dipolar Couplings

Dipolar couplings arise from the direct interaction of two spins (Equation 2.3). In a solution where the proteins tumble isotropically, this dipolar coupling averages to zero. By introducing weak alignment to a protein sample, small dipolar couplings can be retained; these are known as residual dipolar couplings (RDCs). The RDC_{NH} gives information on the orientation of each NH vector relative to the overall molecular alignment tensor.[266] The axial and rhombic components of this alignment tensor, D_a and D_R, can be approximated by plotting a histogram of the RDCs observed for the protein of interest. RDCs are included in structure refinements as a pseudo-energy term, which minimises the difference between the experimental and predicted RDC values. Unlike NOE and dihedral angle restraints, RDCs provide long-distance structural information, and their use within the structure calculations can help to improve the quality of the backbone of the generated structures.[267]

Equation 2.3 Dipolar couplings. D_{PQ} is the dipolar coupling between the two spins, P and Q, γ_P and γ_Q are the gyromagnetic ratio of P and Q respectively and θ is the angle between the spin systems and the magnetic field, B_0.

$$D_{PQ} \propto \gamma_P \gamma_Q r^{-3} < \frac{(3cos^2\theta - 1)}{2} >$$



Sample alignment was induced using Pf1 filamentous phage (10 $mg.mL^{-1}$), which contain nucleic acids that cause them to align within a magnetic field.[268] Presence of these aligned species causes partial alignment of the protein within the sample, allowing RDCs to be measured. Two ^{15}N-HSQC-type spectra are collected for the

sample without decoupling; one dataset is collected for an isotropic solution (without phage), where the observed splitting is caused only by one-bond scalar coupling ($^1J_{NH}$) and a second in the aligned solution, where the splitting is caused by a combination of the $^1J_{NH}$ and the RDC. By determining the difference between the splittings observed for the two samples, the RDC_{NH} were measured (Figure 2.15). To reduce the spectral crowding caused by the splitting of each peak into two, an IPAP (in-phase / antiphase) pulse sequence was used here to separate the pairs of peaks into two spectra, one with the upfield component of the doublet and a second with the downfield component of the doublet (Table 8.1).[269] RDC values for mobile residues (heteronuclear NOE ratio < 0.6) were excluded from the calculations, as the RDC depends upon the order parameter, resulting in some dynamic residues giving smaller RDC values than more rigid residues within the protein (Section 8.15).

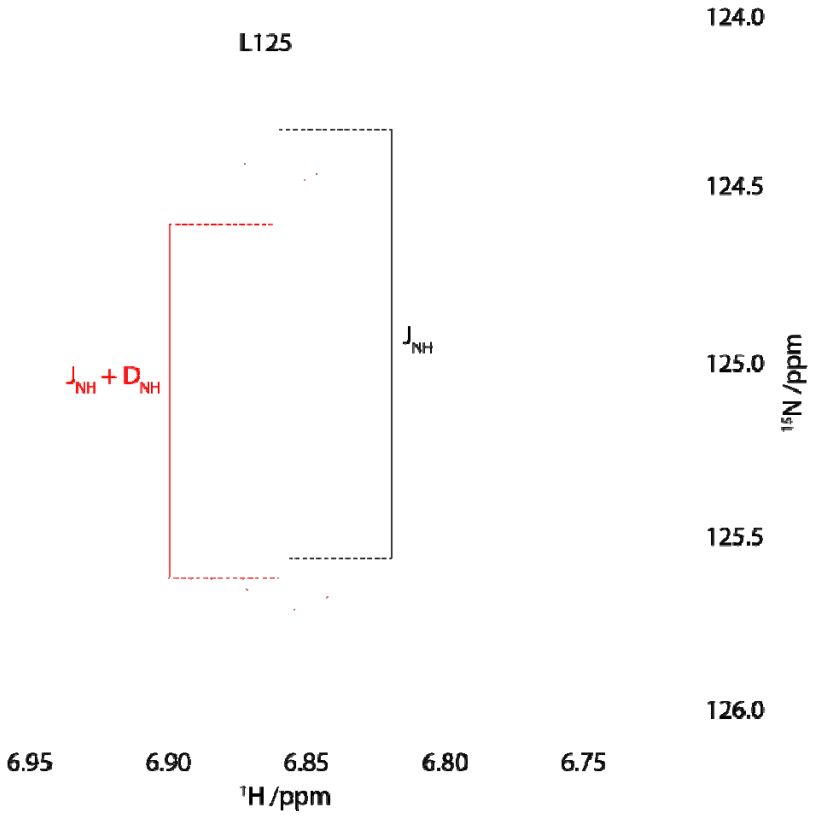


Figure 2.15 Peak L125 from ^{1}H-^{15}N IPAP-HSQC spectra (750 MHz) of dPHD collected without alignment (black spectrum) and with alignment created using a solution of 10 mg.mL^{-1} Pf1 filamentous phage (red spectrum). J_{NH} is the one bond scalar coupling and D_{NH} the RDC.

For use of the RDC restraints in the structure calculations, the alignment tensor must also be defined. Separate D_a and R values were determined for the two PHD fingers due to the flexible linker between them and the lack of observed interaction, which suggested that they can orient independently and may have different alignment tensors. When many RDC values are measured for a protein, the value of D_a and R ($R = \frac{D_a}{D_R}$) can be estimated from a histogram of the RDC values. However, when such histograms were plotted for each domain of dPHD the relatively small number of RDCs gave a sparse distribution from which it was difficult to extract reliable D_a and R values. To optimise the values obtained, a grid search was undertaken, which is commonly used to determine D_a and R values, but unfortunately the values did not converge. Instead D_a and R were estimated by fitting the experimentally-derived RDC values for each domain to the 20 models included in the previously published NMR structure of the 2nd PHD.[240] This gave a D_a value of 9.2 and an R value of 0.22 for the 1st PHD and a D_a value of 7.2 and an R value of 0.3 for the 2nd PHD. When these values were used in structure refinements, the values for the 1st domain gave good overall agreement, but the values for the 2nd domain resulted in large violations of the experimental RDC values that were larger than 14.4 Hz and thus could not be fitted using a D_a value of 7.2. The D_a value was therefore increased to 8 to allow the fitting of all experimentally determined RDC values.

2.9 Generation of a final structural ensemble

73 structures were refined for each domain with RDCs to produce 30 structures from each starting structure. This resulted in over 2000 fully-refined low-energy structures for each domain. The 20 lowest-energy structures for each domain were included in

the final ensemble, although no more than 2 structures from any given starting structure were used. In addition, structures from domain 1 with more than 12 NOE violations or 9 RDC violations and structures from domain 2 with more than 6 NOE violations or 4 RDC violations were excluded to ensure that the final ensemble matched the experimental restraints as closely as possible. Overlays of the final ensemble of structures for the 1st and 2nd PHD domains are shown in Figure 2.16A and B. From these families of structures, a single energy minimised average structure was also generated (Figure 2.16C and D) (Section 8.15.3).

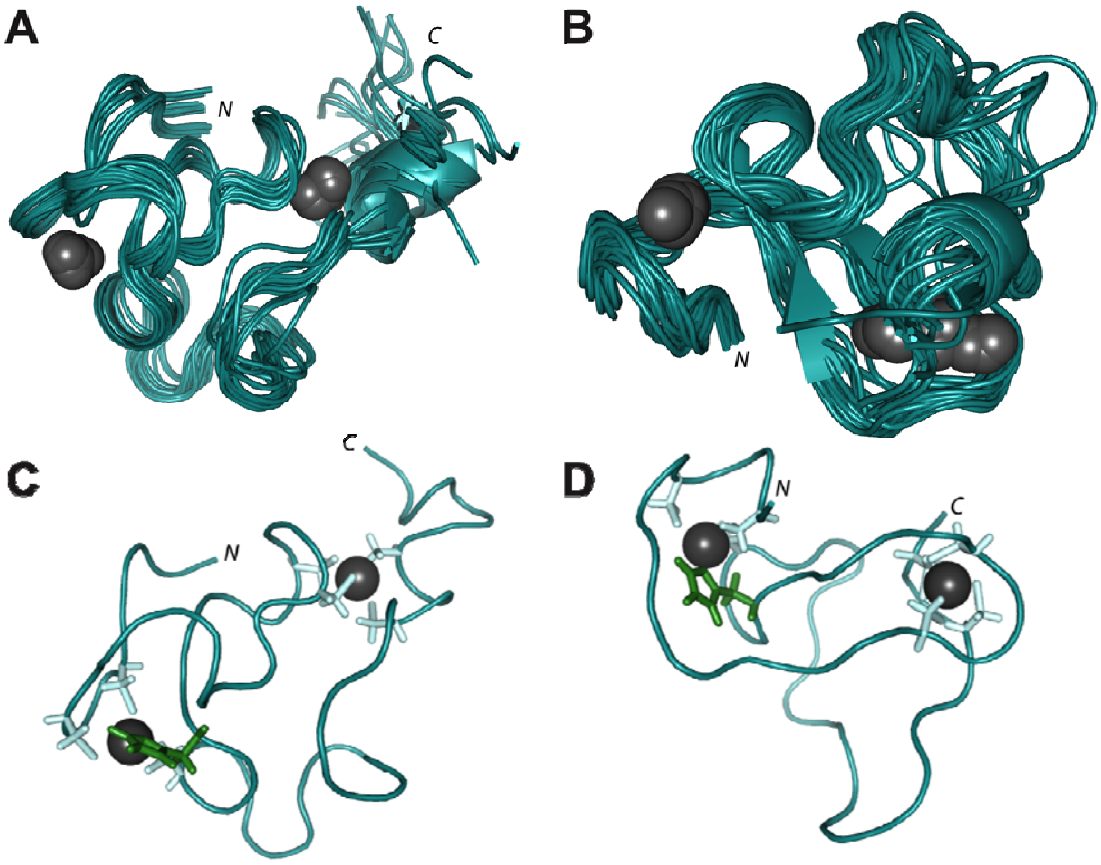


Figure 2.16 Final structures of dPHD **A)** Overlay of the ensemble of 20 structures of the 1st PHD. **B)** Overlay of the ensemble of 20 structures of the 2nd PHD. **C)** Energy-minimised average structure for the 1st PHD. **D)** Energy-minimised structure of the 2nd PHD. dPHD is shown as a cartoon in teal and Zn(II) ions are shown in grey. Zn(II) binding cysteine ligands are shown as sticks in pale blue and Zn(II) binding histidine ligands in green. *N*- and *C*-termini are marked.

Although in a few structures within the low-energy ensemble for domain 2 (Figure 2.16B) the characteristic short anti-parallel β-sheet of a PHD finger can be observed,

this is not maintained in all of the members of the ensemble and cannot be observed in any structures within the domain 1 ensemble. However, analysis of the region where this anti-parallel β-sheet would be expected showed that the two strands do come together in all structures and by analysing the distance between atoms within this region some of the required hydrogen bonds can be observed. Both structural ensembles show a small region of α-helix at their *C*-termini.

2.10 Structure validation

The quality of the final structure ensemble was assessed based on how well the structures obeyed the experimental and theoretical restraints and their stereochemical quality. The structure refinement statistics are given in Table 2.2. The final structures all showed good agreement with the experimentally-derived restraints. Ramachandran statistics and inter-atom clashes were analysed using Procheck-NMR and Molprobity using the Protein Structure Validation Software suite (http://psvs-1_4-dev.nesg.org/).

The average root-mean square deviation of each residue relative to the energy-minimised average structure of each domain was calculated (Figure 2.17A, note that the discontinuity in the plot between residues 76 and 77 comes where the structure was divided into two for refinement calculations) (Section 8.15.3). This data is also represented on the structures of each domain in Figure 2.17B and C. As expected, the backbone RMSD mirrored the plot observed for the heteronuclear NOE ratio (Figure 2.11) with highly flexible regions, such as the linker and residues around 41 and 123 showing the highest RMSD values. The overall RMSD value at C^{α} for the structured residues of domain 1 (heteronuclear NOE ratio > 0.6, 5-37 and 44-55) was 0.82 Å and the RMSD value at C^{α} for the structured residues of domain 2 (87-121 and

124-131) was 1.1 Å. The higher RMSD for the second domain is likely due to the decreased number of experimental restraints included in the calculations (Table 2.2).

Table 2.2 Structure refinement statistics for the two PHD domains of dPHD. Ramachandran statistics were produced by Procheck-NMR.

	Domain 1 (1-77)	Domain 2 (76-137)
NOE-derived distance restraints		
Total	3244	1794
Inter-residue	1011	508
Sequential (|i-j|=1)	986	583
Short-range (1<|i-j|<5)	490	212
Long-range (|i-j|≥5)	757	491
Dihedral angle restraints		
ϕ	23	12
Residual dipolar coupling restraints		
$^1D_{NH}$	42	34
Restraint violations		
Distance restraint violations > 0.3 Å	7.6	4.6
Dihedral angle violations > 5°	0.4	0
RDC violations > 1.5 Hz	7.2	2.2
RMSD from experimental restraints		
Distance restraints /Å	0.015	0.015
Dihedral angle restraints / °	1.5	0.35
RDC restraints /Hz	1.12	0.7
RMSD from idealised geometry		
Bonds /Å	0.0039	0.0023
Angles / °	0.79	0.48
Impropers / °	0.32	0.18
Ramachandran plot statistics		
Most favoured regions /%	43.5	71.3
Additionally allowed regions /%	37.6	25
Generously allowed regions /%	9.6	3.7
Disallowed regions /%	9.3	0

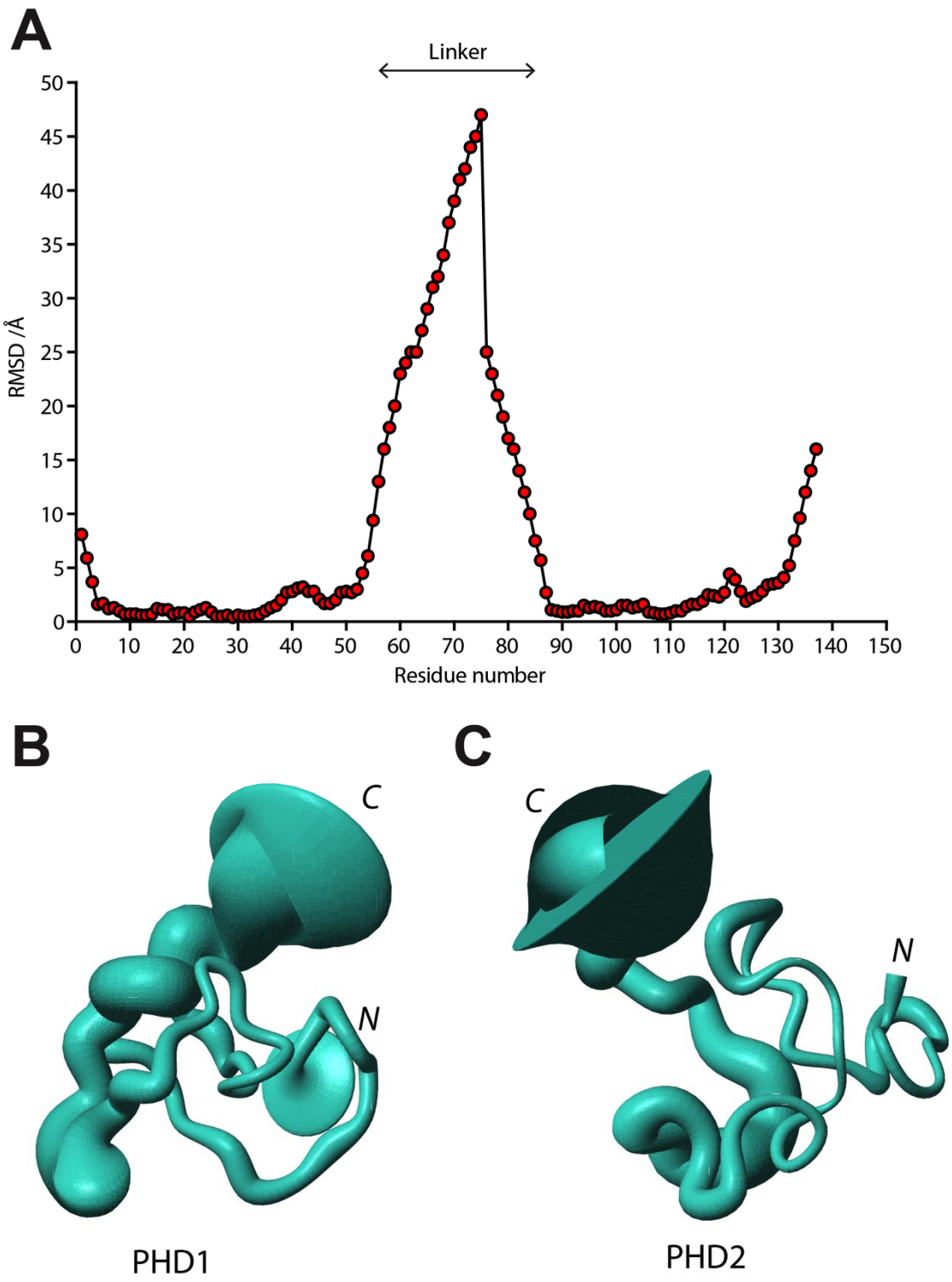


Figure 2.17 A) Graph showing the average root-mean square deviation (RMSD) for the backbone N, C^{α} and C' atoms from the energy-minimised average structure as a function of sequence for each ensemble of 20 structures. **B)** Representation of the ensemble of structures of the 1st PHD. The thickness of the line represents the RMSD of the ensemble in that region of the protein. **C)** Representation of the ensemble of structures of the 1st PHD. The thickness of the line represents the RMSD of the ensemble in that region of the protein.

In structure calculations of FL dPHD the relative orientation of the two PHDs was random. An overlay of the 1st PHD from 10 low-energy full-length structures showed the dispersion of the 2nd PHD relative to the 1st domain (Figure 2.18A). In addition to the orientation of the two fluctuating domains, the interdomain distance was also variable; Figure 2.18B shows an example of an extended structure, with an interdomain distance of 68 Å, and Figure 2.18C shows an example of a more compressed structure with a distance of 32 Å between the two domains.

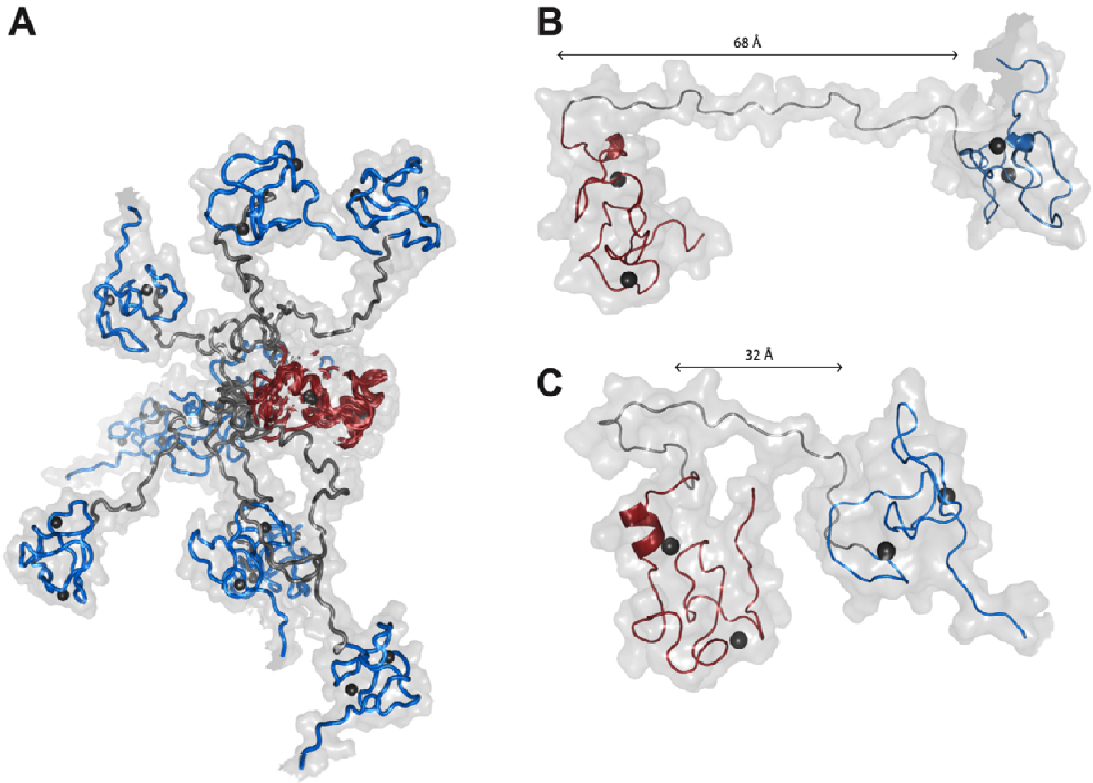


Figure 2.18 A) Overlay of the 1st PHD (red) from 10 low energy structures of the FL dPHD construct showing the dispersion of the 2nd PHD (blue). **B)** Example of an extended dPHD structure with the 1st PHD shown in red and the 2nd PHD shown in blue. **C)** Example of a compressed dPHD structure with the 1st PHD shown in red and the 2nd PHD shown in blue.

2.11 Discussion

Work in this Chapter described the determination of the solution structure of the tandem PHDs of the chromatin remodeller CHD4. The structure revealed that the two domains are structurally independent and separated by a highly mobile linker, thus representing the first example of a structure of a tandem PHD pair in which the two zinc domains are not directly interacting. This contrasts with the two published PHD pair structures from DPF3B and MOZ in which the two domains interact with one another.[216,217] The inherent flexibility in the structure probably helps to explain why previous crystallographic efforts have proved unsuccessful. Solution NMR spectroscopy provided an ideal alternative structural technique for determining the structure of this highly flexible protein. The resulting structural ensemble provides a representation of the various conformations that the two domains can adopt in solution. As a result, however, the atomic resolution of the structure is lower than would be achieved using X-ray crystallography.

The lack of regular secondary structure in PHD fingers has proven a challenge for structure determination. For example, the TALOS-N prediction software appears to make unreliable torsion angle predictions for the protein, presumably due to a lack of suitable models in its database, reducing the number of experimental restraints for inclusion in the structure calculations. The small number of torsion angle restraints included in the structure calculation may account for the relatively high number of angles observed in the generously allowed regions of the Ramachandran plot. Inclusion of more angle restraints might further reduce the RMSD values of the structure. Inclusion of further sets of RDC measurements, for example one-bond $^{13}C_{\alpha}$-

$^{1}H_{\alpha}$, could be used to further reduce the RMSD of the dPHD ensembles for each individual PHD finger.

The majority of structures in the final ensembles do not consistently show any anti-parallel β-sheet as expected for a classical PHD finger. However, analysis of the region where this anti-parallel β-sheet would be expected shows that the two strands do come together in all structures and by analysing the distance between atoms within this region some of the hydrogen bonds required for a β-sheet can be observed. To help to maintain this secondary structure in the ensemble, hydrogen-deuterium exchange experiments could be carried out to confirm if and where hydrogen bonds occur in this region, which could then be used as restraints in future structure calculations.

The heteronuclear NOE values (Figure 2.11) and structure calculations show that there is no fixed orientation of the two domains relative to each other in solution, and that they can sample a large range of conformations with variable interdomain distances (Figure 2.18). However, the 32-residue linker between the two domains is highly negatively charged and it is possible that, as a result, on average the two domains prefer to adopt a more extended conformation to reduce unfavourable electrostatic interactions. It is not possible to determine this using only the structural restraints acquired by NMR spectroscopy, which only provide relatively short distance restraints (< 8 Å) and do not include a full electrostatics force-field. Small-angle X-ray scattering (SAXS) is a technique that is used to obtain low-resolution information on the structure of multi-domain proteins or complexes in solution.[270] Although for a given protein it is possible to fit the collected data to a

single conformation of the protein, in the cases of proteins with regions of flexibility, it is important to consider that the data likely represents the average conformation of the protein in solution and is best represented by a weighted ensemble of different protein conformations. Two recent reports have used SAXS data to model the structure of CHD4 in solution, but in both cases the data has been used to fit a single structure.[245,271] In light of the data collected in this Chapter, which showed that the linker between the two PHDs is highly flexible, this single-structure approach to the analysis of SAXS data is unlikely to have provided a true representation of the solution structure of dPHD. SAXS data has been collected for dPHD and by fitting a large ensemble of refined NMR structures to this SAXS data it should be possible to select for a more restricted ensemble of interdomain conformations, weighted towards those that appear more frequently, that better represents the conformation of dPHD in solution.[272,273] Further experiments involving site-directed mutagenesis and addition of spin labels at different positions on the two domains could also be used to measure the average interdomain distance which could also be used to add additional refinement parameters for use in the structure calculations.

All structural characterisation has been carried out on the isolated protein. On binding to chromatin it is likely that the structure will be further modulated in order to avoid unfavourable interactions such as between the negatively charged linker region and the negatively charged DNA. Solution NMR spectroscopy cannot be used to investigate such large complexes, but solid state NMR and other lower resolution techniques such as electron microscopy could be used to look at the structure of CHD4 binding to nucleosomes.

Determination of the structure of this construct allowed for detailed studies into the binding interactions of it with histone tail peptide mimics, which are described in the following Chapter. The backbone assignments should allow for determination of binding constants between histone peptides and the PHDs, and for the binding site of the interactions to be mapped onto the protein structure.

Chapter 3
Investigations into the binding interactions of the tandem PHD domains of CHD4

3.1 Introduction

PHD fingers are chromatin reader domains commonly found in nuclear proteins. They are involved in molecular targeting and most characterised domains bind to H3 tails.[274] The H3 tail forms an induced β-strand along the surface of the PHD that extends the two-strand sheet already present at the core of the PHD (Figure 3.1). Aside from their conserved cross-brace zinc-binding motif (Figure 1.18, Chapter 1), PHD fingers have low sequence conservation, allowing for a variety of binding preferences for post-translational modifications. Sequence alignments have been used to classify PHD fingers into families, however, allowing some predictions about their sequence preferences to be made.[275]

The first PHDs to be functionally characterised were found to recognise H3 K4Me$_3$,[276–279] but subsequently PHD fingers that recognise non-methylated R2 and K4, methylated K9 and acetylation at K14 have also been described.[186,217,280–283] Recognition of K4Me$_3$ is usually *via* an aromatic, hydrophobic cage of residues. In PHDs that do not bind to histone tails with K4 trimethylation, some, or all, of these residues are replaced by acidic residues. The structurally characterised tandem PHDs of both MOZ and DPF3B have been shown to bind to a single histone peptide containing H3 K14 acetylation *via* a binding interface formed by both PHD

fingers.[216,217] Recently, a tandem PHD pair in MLL4, which contains 7 PHDs altogether, has been shown to interact with H4 tails rather than H3.[284]

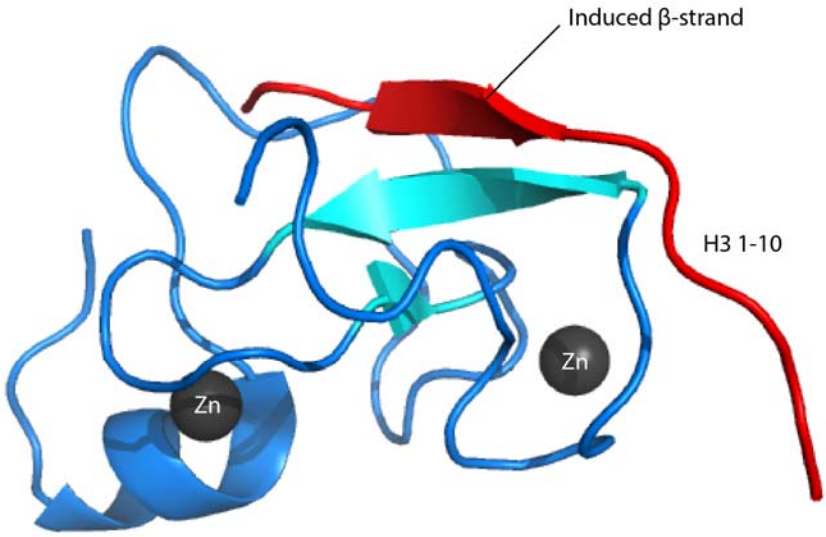


Figure 3.1 View from a crystal structure of the 1st PHD of AIRE (blue) bound to a histone peptide (H3 1-10, red) (PDB ID 2KFT).[283] The anti-parallel β-sheet of AIRE is shown in cyan and the zinc ions in grey.

Although most PHDs have been found to be histone-binding domains there are a few examples that have also been identified to have non-histone substrates, such as the PHD of MLL1, which also binds to CyP33 and the PHD of Pygo, which binds to the homology domain 1 (HD1) of BCL9.[285,286]

Musselman *et al.* have shown using pull-down experiments, tryptophan fluorescence and NMR spectroscopy that in isolation, the 2nd PHD of CHD4 binds to the *N*-terminal region of H3 tails.[218] It shows a preference for sequences methylated or acetylated at K9 (K_D of 0.9 μM for H3 1-21 K9Me$_3$), whilst trimethylation at K4 is found to abrogate the binding (K_D of 1.86 mM for H3 1-21 K4Me$_3$).

3.1.2 Objectives

In the previous Chapter, backbone $^{1}H^{N}$-^{15}N assignments for the tandem PHDs of CHD4 were determined. Work in this Chapter describes use of these chemical shifts for the characterisation of the binding interaction between the tandem PHDs of CHD4 and histone peptides; binding affinities were subsequently determined. The changes in chemical shift were used to map the peptide binding interactions onto the dPHD structure determined in Chapter 2. After investigating this construct, the work further extended to investigate the histone and DNA binding behaviour of a larger, more native CHD4 construct that also contained the two CHDs.

Most protein samples used in this Chapter were provided by Dr. Rosa Morra. The titration with H3 3-14 was carried out by Professor Christina Redfield, who also provided assistance with data collection and processing. K_D fitting programmes were written by Professor Christina Redfield. Non-denaturing mass spectra were collected by Ima-Obong Ebong. All peptides used in this Chapter were purchased from GL Biochem (Shanghai Ltd.) and were >95% pure.

3.2 Investigating the histone-binding behaviour of dPHD

3.2.1 Measuring protein-protein interactions using NMR spectroscopy

Chemical shifts are sensitive to the chemical environment around them. As such, a protein's backbone chemical shifts are altered when it binds to an interacting partner. In this way interactions can be observed by the alteration of peaks within the spectra in a dose-dependent manner.[287] Depending on the on- and off-rate of the binding this

can be observed as the gradual movement of the peak from the fully unbound position to the fully bound position (fast exchange), or by the decrease in intensity and broadening of the unbound peak until it disappears and the appearance of a corresponding new peak in the bound position (slow exchange). Addition of ligand affects the chemical shift of some peaks and by analysing which residues of the protein these peaks correspond to, the binding site on a protein can be determined. By titration of increasing concentrations of a known ligand into a protein and measuring the change in chemical shift observed at each ligand concentration, binding constants (K_Ds) can be calculated.[288,289]

3.2.2 Investigating the histone binding interactions of dPHD

Initially, the binding behaviour of dPHD with an unmodified H3 1-21 peptide (GLS Biochem >95% pure) was investigated. As the 2nd PHD has previously been shown to bind to H3 tails when in isolation, it was anticipated that peaks corresponding to this domain would move in the spectra.[218] In isolation, the binding affinity has previously been measured to be 18 μM, and it was of interest to see what effect the presence of the 1st domain would have on this interaction. Increasing concentrations of H3 peptide were added to labelled dPHD (210 μM) and a ^{1}H-^{15}N-HSQC spectrum collected after each peptide addition (Section 8.15.4). Interestingly, on addition of peptide, peaks from both domains were observed to move (Figure 3.2A), indicating that the environment of both domains was affected by the addition of peptide. Given that the structure of the tandem PHD construct showed no interaction between the two domains themselves when isolated in solution (Chapter 2), it was considered unlikely that changes in the structure of one domain would alter the structure of the other domain. The data thus suggested that both domains interact directly with the histone

peptide.

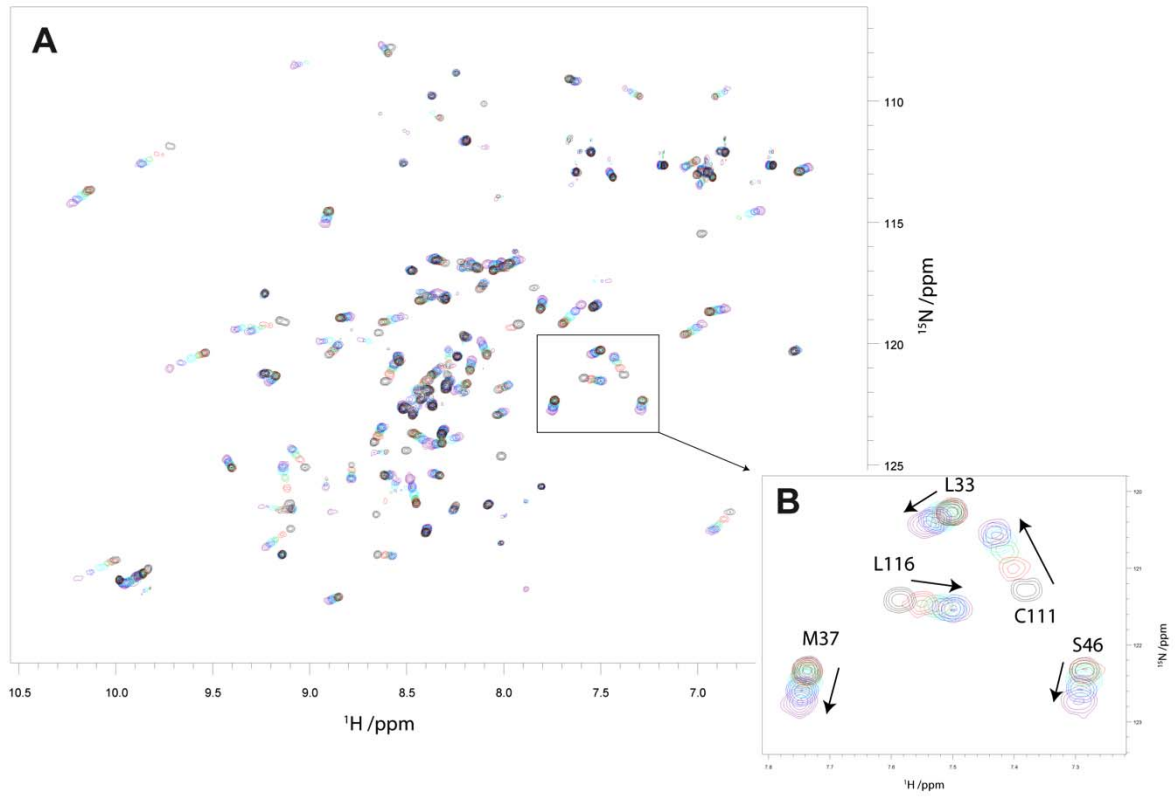


Figure 3.2 500 MHz ^{1}H-^{15}N-HSQC spectra of ^{15}N-labelled dPHD (210 µM) at pH 7.5, 293 K showing the effect of addition of an H3 1-21 peptide. Peak colour indicates the H3 peptide concentration: black, no peptide; red, 50 µM; green, 100 µM; cyan, 200 µM; blue, 400 µM; purple, 1 mM. **A)** Full spectrum. **B)** Expanded region of the spectrum. Labels give the residue number and black arrows show the direction of the peak movement on peptide addition.

The rate of signal movement was found to vary between residues. Peaks from residues in the 1st domain remained almost static at low concentrations of added peptide (the black, red and green spectra in Figure 3.2B almost overlay for M37, L33 and S46), whilst those in the 2nd domain moved from the first 50 µM peptide addition (for L116 and C111 the black and red spectra in Figure 3.2B show a large change in chemical shift). At higher concentrations of added peptide, when peaks in the 2nd domain began to stop moving (blue and purple spectra almost overlay in Figure 3.2B), indicating that they were approaching their fully peptide bound state, peaks in the 1st domain began to move more significantly (cyan, blue and purple spectra in Figure 3.2), suggesting that these residues were only beginning to bind to peptide at these

higher concentrations.

The chemical shifts for a well-resolved peak from each domain (H29 and F87) were extracted and plotted against added peptide concentration, which also illustrated the lag in the chemical shift movement of H29 from domain 1 relative to F87 from domain 2 (Figure 3.3). The difference in the movement of peaks from the two domains suggested that the protein was binding to two separate peptides, rather than forming a single contiguous binding surface and that the 2nd PHD bound to the histone tail peptide with a higher affinity than the 1st PHD. This bivalent binding behaviour was later confirmed by two subsequent reports that also demonstrate that both the 1st and 2nd PHD fingers of CHD4 can bind to H3 tails.[239,290]

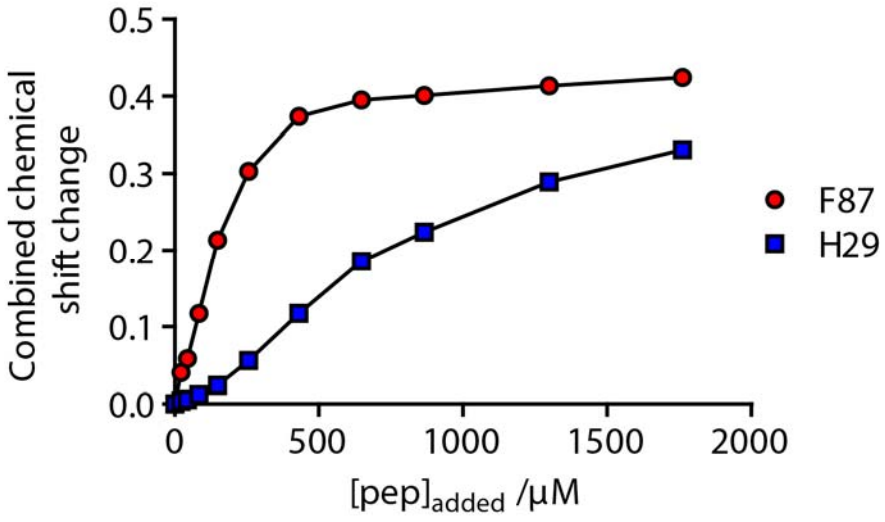


Figure 3.3 Graph showing the change in the combined chemical shift (calculated as $\sqrt{\Delta_H{}^2 + 0.1.\Delta_N{}^2}$, where Δ is the change in chemical shift from the spectrum of dPHD alone) for peaks from the 1st (H29) and 2nd (F87) domains of dPHD on addition of increasing concentrations of histone peptide.

Measurement of the change in chemical shift on addition of increasing concentrations of peptide allows the K_Ds for the two binding sites to be determined. For a protein with two independent binding sites the following equation applies to each site (Equation 3.1).

Equation 3.1 Equation for calculating a K_D from chemical shift changes, where Δ is the change in chemical shift from the starting chemical shift at a given concentration, Δ_0 is the chemical shift change between the fully bound and fully unbound state, $[pep]_{free}$ is the concentration of peptide free in solution and K_D is the binding constant.

$$\frac{\Delta}{\Delta_0} = \frac{[pep]_{free}}{[pep]_{free} + K_D}$$

The change in chemical shift (Δ) for each peak (Figure 3.2), was converted into the fraction of the way the peak had moved between its fully unbound and fully bound position ($\frac{\Delta}{\Delta_0}$). The chemical shifts of five well-resolved residues from each domain (T23, H29, W45, A27, A40 from domain 1 and T102, F87, D101, S105, W124 from domain 2) were extracted and used to determine the K_D values for each domain based on Equation 3.1. The K_D values were iteratively calculated using a 12 parameter fitting programme (Figure 3.4, *written by Prof. Christina Redfield*). Multiple parameters were required to include estimation of the fully bound peak position for each of the ten residues where complete saturation had not been reached as well as the two K_D values. The two binding sites were fitted together, as due to the high concentration of protein (210 μM) in the NMR samples, $[pep]_{free}$ could not be approximated to $[pep]_{added}$ and peptide depletion due to binding to both domains had to be accounted for. K_D values of 460 μM and 30 μM were determined for the 1st PHD and 2nd PHD respectively. The K_D value for the 2nd PHD was in reasonable agreement with the previously published value (18 μM), suggesting that presence of the 1st PHD does not affect peptide binding to the 2nd PHD finger.[218] Interestingly, the K_D value recently published for the 1st PHD finger is much lower than the value found for this tandem construct (reported as 3 μM for an H3 1-12 peptide *vs* 460 μM measured here for H3 1-21).[290] The construct of the 1st PHD used in this publication contains one extra native *N*-terminal residue and five extra residues prior to this, which could be responsible for the higher binding affinity. Alternatively, the presence

of the 2nd PHD or linker may affect the binding of the 1st PHD to the histone peptide. Consistent with this proposal, a more recent paper looking at a similar tandem construct of dPHD also found that the 1st PHD finger has a weaker K_D than the 2nd PHD finger.

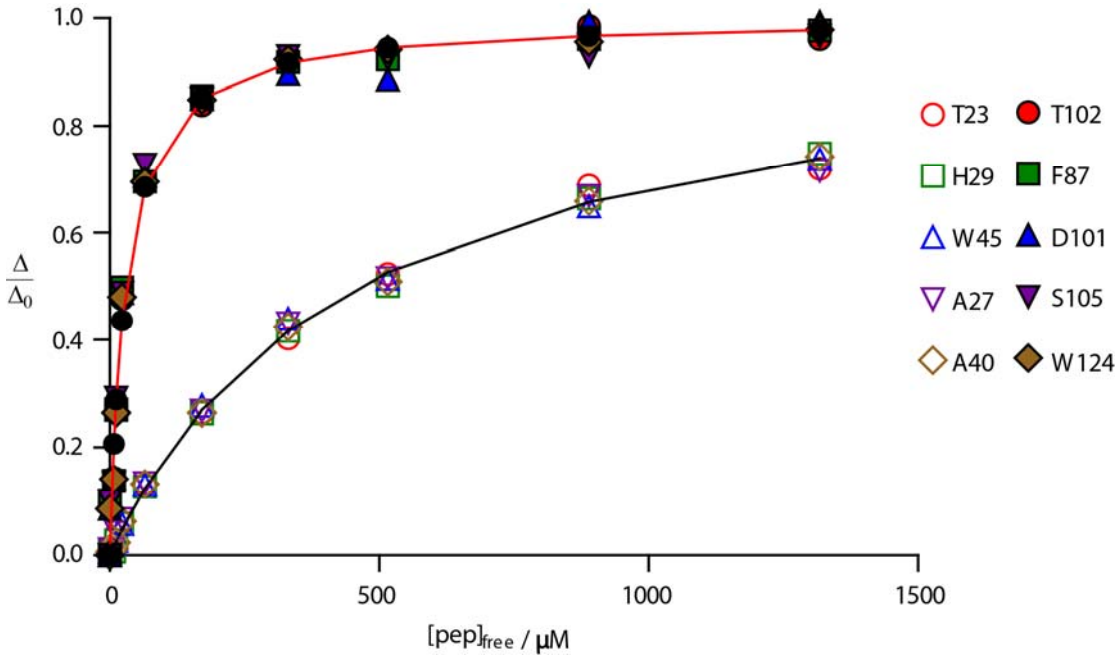


Figure 3.4 Graph showing the $\frac{\Delta}{\Delta_0}$ value for each of the ten residues fitted at increasing concentrations of H3 1-21 peptide. The black line shows the K_D fit for domain 1 and the red line the K_D fit for domain 2.

3.2.3 Mapping the histone binding surface of dPHD

Chemical shift changes give information about the binding site of the peptide as well as the binding affinity. The magnitude of the chemical-shift change can be loosely correlated with the degree to which the histone peptide binding perturbs the environment of a given residue. By mapping this onto the protein structure an approximation of the binding surface can be generated. Figure 3.5 shows the change in chemical shift for residues of dPHD on binding to the H3 21mer peptide.

Fully-bound chemical-shift positions were determined by reference to the calculated peptide K_Ds for each domain and Equation 3.1. Residues 1-76 were adjusted using the K_D for the first PHD (460 μM) and residues 77-137 using the K_D for the second PHD (30 μM). The individual chemical-shift changes in the nitrogen and proton dimensions give information on the direction of the shift in the peak, which is lost when the combined chemical-shift change is calculated. As the chemical shifts for nitrogen are ten-fold larger than the chemical-shift values for the proton dimension the chemical-shift values are scaled by 10-fold to make them more comparable.

Some residues, such as W124, were perturbed significantly in their nitrogen chemical-shift values, but very little in the proton chemical shift, whilst others, like I18, changed significantly in their proton chemical shift and very little in their nitrogen chemical shift (Figure 3.5A). The most significant chemical-shift changes were for residues within each of the PHD fingers, with only small changes for residues within the linker (Figure 3.5B), suggesting that the structure of the linker was not significantly perturbed on binding to the histone peptides and thus that the two domains remain as isolated binding domains separated by a flexible linker.

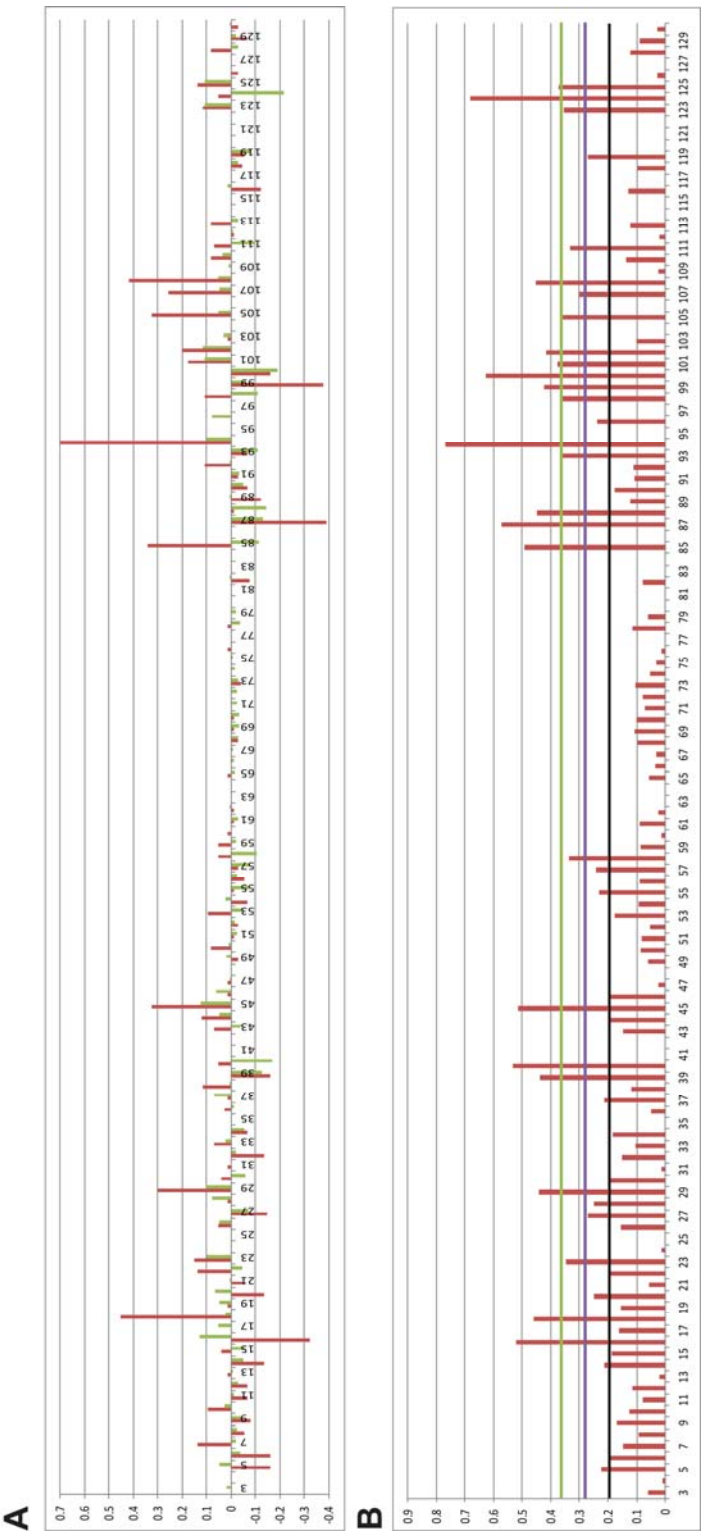


Figure 3.5
Plot showing the change in chemical shift on H3 1-21 peptide binding to dPHD as a function of sequence. **A)** Change in the proton dimension of the backbone N-H chemical shift (red) and change in the nitrogen dimension of the backbone N-H chemical shift (scaled by 0.1, green). **B)** Combined chemical shift change. Combined chemical shift is calculated as $\sqrt{{\Delta_H}^2 + 0.1.{\Delta_N}^2}$ for each residue, where Δ is the change in chemical shift from the chemical shift in dPHD alone. Green line shows 1 s.d. from the mean combined chemical shift change, purple line 0.5 s.d. from the mean combined chemical shift change and black line the mean chemical shift change.

To investigate the binding sites of the histone peptides the changes in mean chemical shift were mapped onto one of the dPHD structures from the ensemble generated in Chapter 2. Residues whose chemical shifts changed more than the mean chemical-shift change were considered to have been perturbed (above the green line in Figure 3.5B) with ones whose chemical shifts changed more than 1 standard deviation from than the mean (above the green line in Figure 3.5B) considered to have been most perturbed. The resultant binding sites are shown in Figure 3.6.

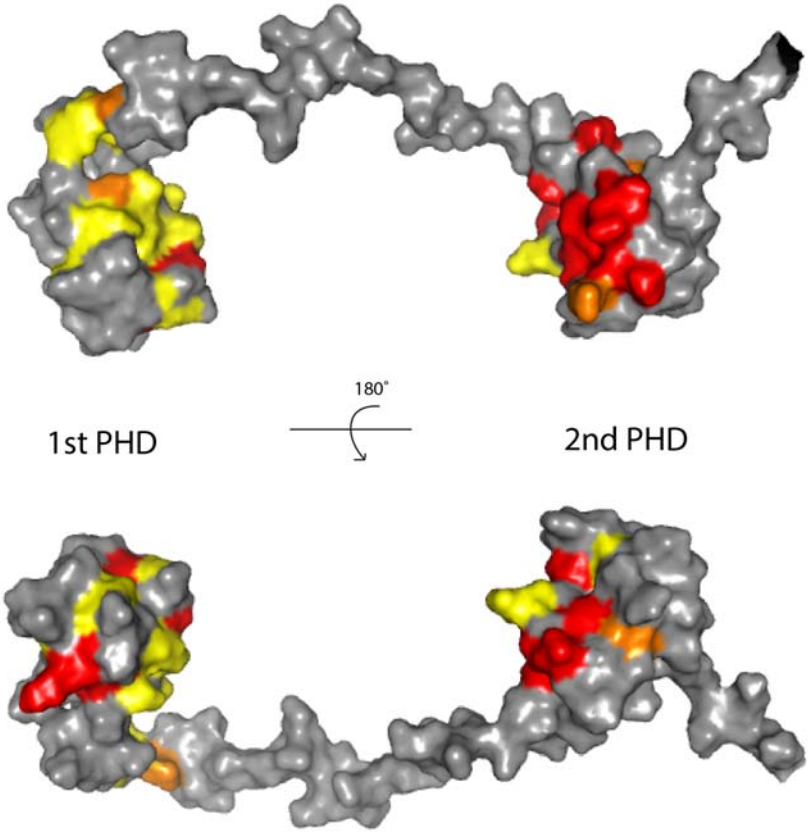


Figure 3.6 Binding sites for H3 1-21 mapped onto the structure of dPHD. Residues whose combined chemical shifts change more than 1 s.d. of the mean are shown in red, residues whose chemical shifts change more than 0.5 s.d. from the mean in orange and residues whose chemical shifts change more than 0.25 s.d. from the mean in yellow.

dPHD was found to have two well-defined histone-binding surfaces (Figure 3.6). The regions that shifted the most correspond to the flexible loops identified from the heteronuclear NOE ratio (Figure 2.11, Chapter 2) and the region expected to form the anti-parallel β-sheet in the PHD finger to which histone peptides are commonly found

to form a third β-strand (Figure 3.1, Chapter 3).

3.2.3.2 Confirming the binding of two histone peptides by mass spectrometry

To confirm that the behaviour of the chemical shifts was due to two peptides binding to the tandem construct, mass spectrometry (MS) experiments were carried out. Protein was buffer exchanged into 200 mM ammonium acetate and three samples were prepared, one with dPHD (20 μM), one with H3 1-21 peptide (20 μM) and a third with dPHD (20 μM) and H3 1-21 peptide (1 mM). *Samples were run by Ima-Obong Ebong (CRL, Oxford) using non-denaturing mass spectrometry*. The mass spectrum for the sample of both protein and peptide clearly showed mass shifts for the protein equal to addition of both one (~19.5 kDa) and two peptides (~21.8 kDa) to the protein (~17.2 kDa, Figure 3.7).

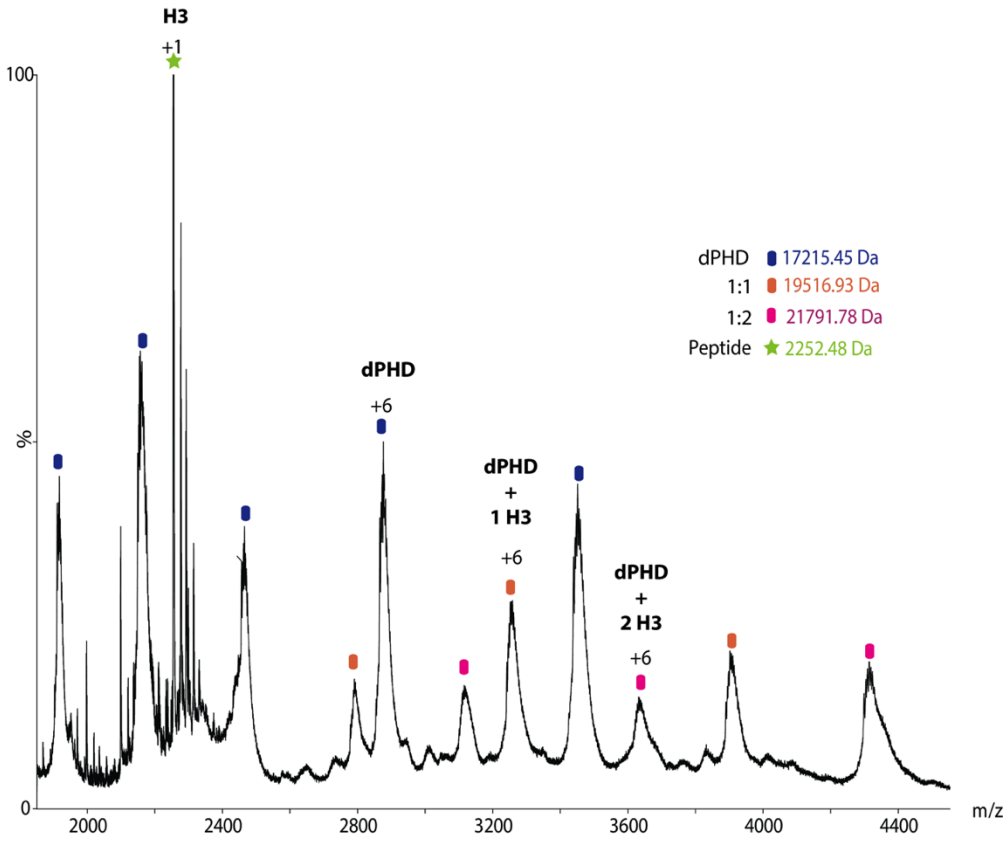


Figure 3.7 Native mass spectrum of dPHD (20 μM) binding to one and two equivalents of an H3 1-21 peptide (1 mM) in 200 mM ammonium acetate.

3.2.4 Investigating the effect of varying the peptide length on dPHD binding

Work by Musselman *et al.* demonstrates that the first two *N*-terminal residues of the histone tails are essential for PHD binding for the 2nd PHD of CHD4.[218] To confirm whether this was also true for the 1st PHD of the double PHD domain construct, H3 (3-14) was titrated into a sample of uniformly ^{15}N-labelled dPHD. ^{1}H-^{15}N-HSQC spectra were collected at increasing concentrations of histone peptide (*carried out by Prof. Christina Redfield*). No effect on the HSQC spectrum was observed at any concentration (data not shown), confirming that, as for the single PHD finger construct of the 2nd PHD, the first two *N*-terminal residues were essential for histone binding to the 1st PHD finger of CHD4.

The weak K_D of H3 1-21 with the first PHD suggested that the 21mer peptide might not be the optimal binding partner for the tandem PHDs. The effect of addition of a number of different length histone peptides was investigated to see whether this would alter the K_D values and if with a longer histone peptide, the domains would bind to a single peptide. Four more peptides of variable length were studied, each beginning at H3 A1 and having between 5 and 40 residues (H3 1-5, H3 1-12, H3 1-17 and H3 1-40 (GLS Biochem >95% pure)). As previously, ^{1}H-^{15}N-HSQC spectra of the labelled dPHD (210 μM) were collected with increasing concentrations of each peptide titrant. Chemical shifts of the same 10 residues were used to determine the K_D values for each peptide with each domain (Table 3.1). No binding was observed for the 5mer peptide. In all other cases, binding of two peptides was observed, even with the 40 residue H3 peptide that corresponded to the majority of the unstructured *N*-terminal tail of H3, suggesting each PHD always binds to a separate peptides.

Table 3.1 K_D values for H3 peptides of variable lengths binding to dPHD (210 μM) calculated by fitting of the change in chemical shift of T23, H29, W45, A27, A40 from domain 1 and T102, F87, D101, S105, W124 from domain 2 during NMR titration experiments.

Peptide	K_D /μM	
	1st PHD	2nd PHD
H3 1-5	-	-
H3 1-12	1500	120
H3 1-17	660	60
H3 1-21	460	30
H3 1-40	650	19

The K_D values demonstrated that increasing the length of the histone peptides increased their affinity for both domains of dPHD (more than two-fold for the 1st PHD and six-fold for the 2nd PHD). This may be due to an increasing number of interactions between the peptides and the protein or due to the presence of the negatively charged carboxyl-group at the *C*-terminus of the 12mer peptide where there is usually an amide bond causing unfavourable interactions with the peptide-protein binding site, or negatively-charged linker region, and reducing the peptide affinity. By looking at the degree of shift of each residue of dPHD individually for each peptide, the peptide-protein interaction behaviour was compared to see whether there were more changes to the protein with the longer peptides. As with the 21mer peptide, for peaks that did not reach saturation in all titrations, the calculated K_D for each peptide was used to determine the saturated positions for each peak using Equation 3.1. Figure 3.8 shows a comparison between the change in chemical shift for all residues from the two domains on binding to H3 1-12, H3 1-17 and H3 1-21 peptides. The pattern of shifts was similar for all three peptides, indicating that all three peptides bound to dPHD in a similar manner. Interestingly, the small chemical-shift changes observed for residues in the linker region (65-75), were considerably greater when dPHD bound to H3 1-21 and H3 1-17 peptides, compared to the H3 1-12 peptide. Although it is unclear in what way these residues

were interacting with the peptide, this may be a factor in the higher binding affinity of the construct for this peptide. These changes might also account for the difference in the K_D values for the 1st PHD when in isolation compared to when present in the tandem construct.

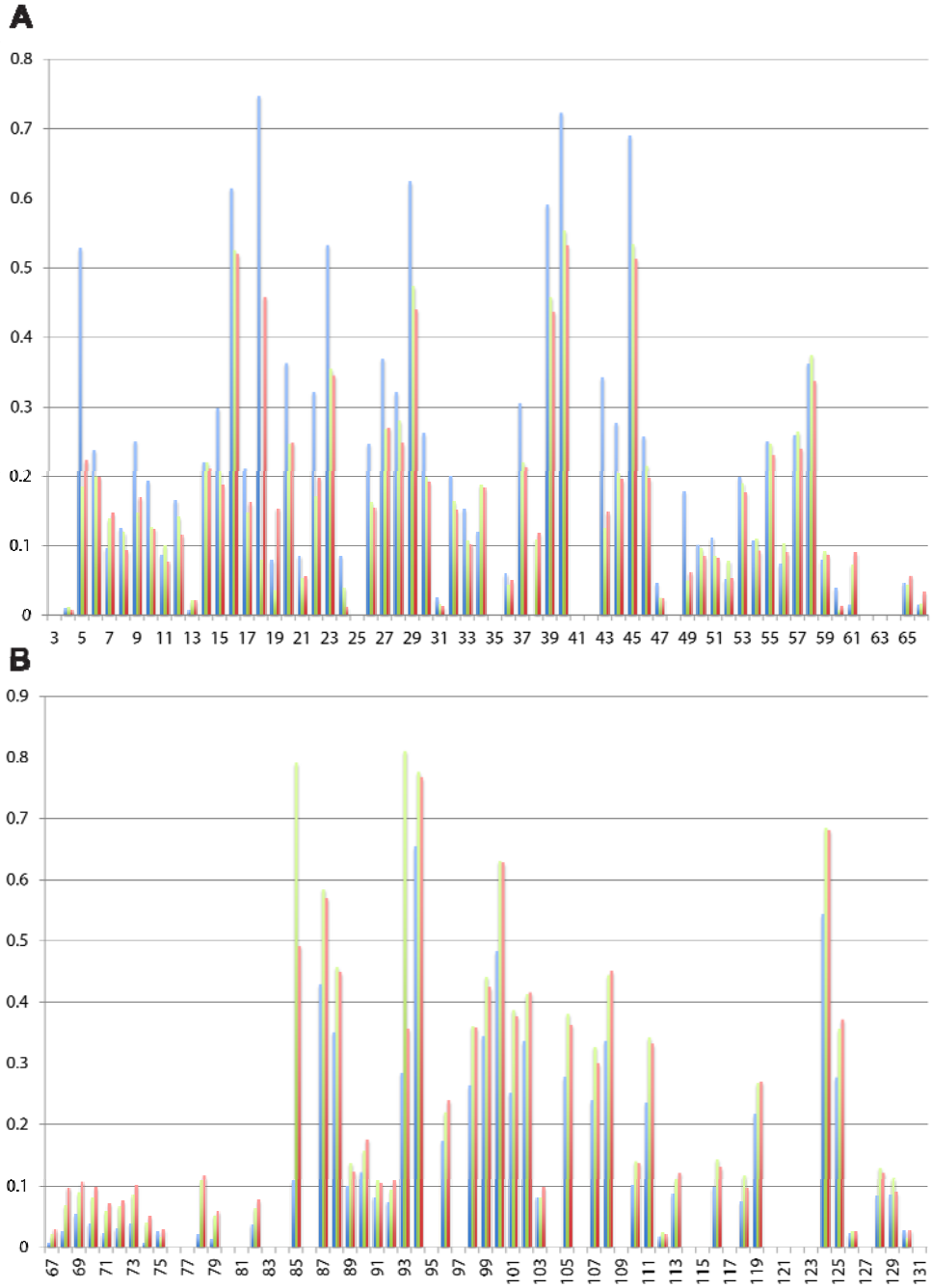


Figure 3.8 Graphs showing a comparison between the combined chemical shift changes for dPHD binding to H3 1-12 (red), H3 1-17 (green) and H3 1-21 (blue) as a function of amino acid sequence. **A)** Residues 3-66 of dPHD and **B)** residues 67-131 of dPHD.

Further titration experiments using a peptide trimethylated at H3 K4 or K9 were also carried out. Although much tighter binding was observed with H3 K9Me_3 and much weaker binding with H3 K4Me_3 it was not possible to fit exact K_D values.

3.3 Investigating the interaction of dPHD with dCHD

In FL CHD4 the two PHDs are *N*-terminal to two CHDs (Figure 2.1, Chapter 2). The structure of these four domains together is unknown. As there are multiple examples of PHDs that interact with other protein-binding domains, such as the PHD of PYGO, which interacts with the HD1 domain of BCL9, and the PHD of KAP1, which interacts with a neighbouring bromodomain, possible interactions between dPHD and dCHD were investigated.[285,291]

Initial experiments were carried out by titrating unlabelled recombinant dCHD (*provided by Dr Rosa Morra*) into uniformly ^{15}N-labelled dPHD. Two samples were prepared, one containing dPHD alone and the other containing a 1:1 ratio of dPHD:dCHD. An HSQC spectrum of each sample was collected. Following this, the two samples were mixed in varying amounts to alter the ratio of dPHD:dCHD. On mixing, some precipitation was observed. This was removed by centrifugation before each spectrum was collected. On addition of increasing concentrations of dCHD to dPHD, peaks in the HSQC spectrum began to disappear. The spectrum of 1:1 dPHD:dCHD is shown in Figure 3.9B, with a reference spectrum of dPHD alone in Figure 3.9A. The spectrum of the two proteins together showed that the majority of the peaks that disappear corresponded to peaks from the 2nd PHD (circled in red), whilst only a few peaks from the 1st PHD (circled in blue) were perturbed. In addition, most of the peaks corresponding to residues in the linker region remained

unperturbed, although due to the largely random-coil chemical-shift values of these residues, complete analysis was not possible.

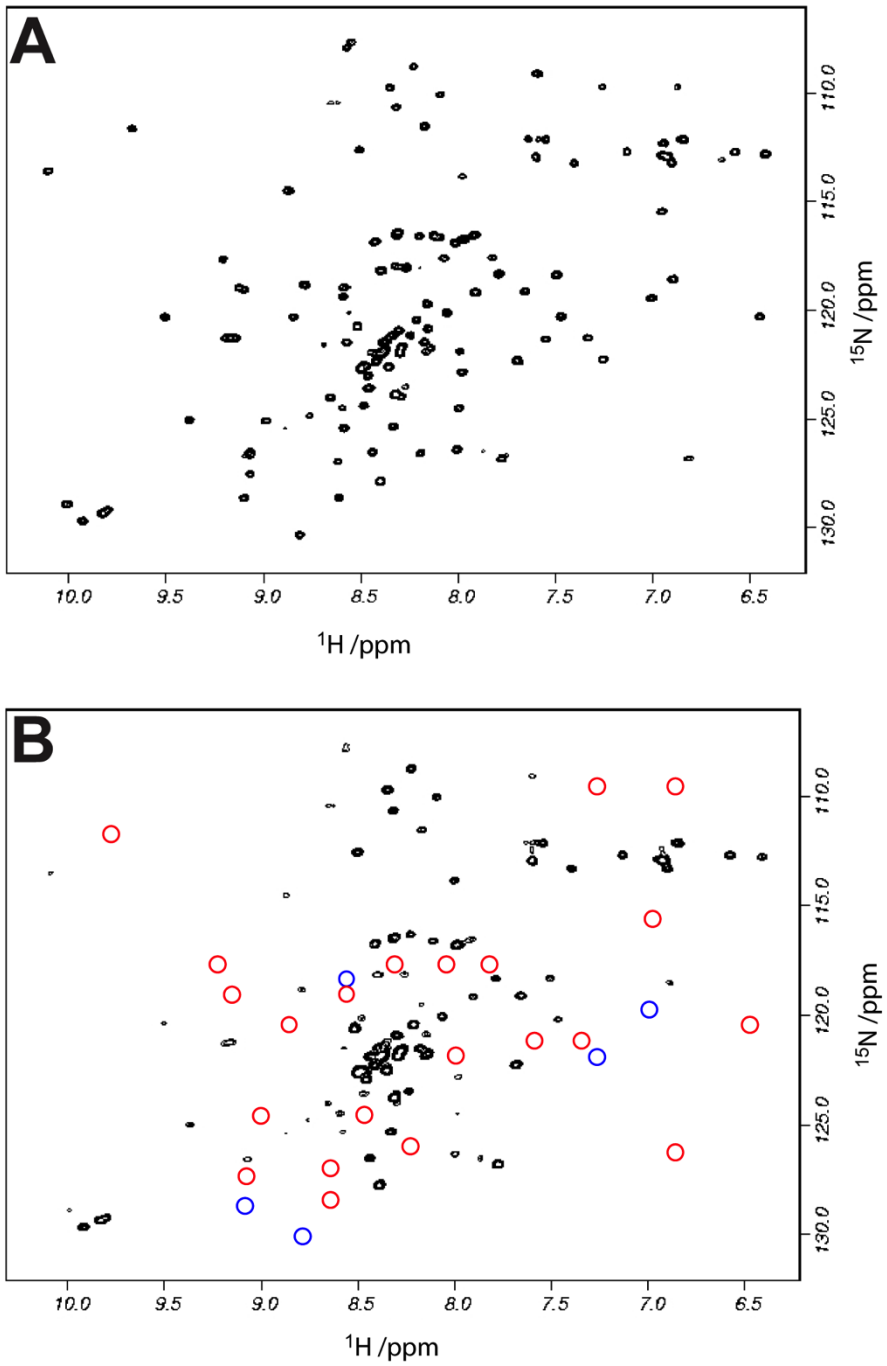


Figure 3.9 A) 500 MHz ^{1}H-^{15}N-HSQC of ^{15}N-labelled dPHD (190 µM). **B)** 500 MHz ^{1}H-^{15}N-HSQC of ^{15}N-labelled dPHD (190 µM) mixed with unlabelled dCHD (210 µM). Peaks that are missing from the first PHD domain are circled in blue; peaks that are missing from the 2nd domain are circled in red.

Loss of peaks during a titration due to peak broadening is usually due to the binding interaction having an intermediate on- and off-rate. Close to the point of saturated binding, however, the peaks would be expected to reappear at their fully-bound chemical-shift positions. As this was not achieved by addition of dCHD to dPHD, the two proteins were expressed separately and then mixed during purification and purified together on an S75 size exclusion column to try to isolate a fully bound dPHDdCHD complex (*complex prepared by Dr Rosa Morra*). Fractions corresponding to the mass of the two proteins together were collected for NMR experiments. Again, however, the spectrum showed the same number of missing peaks (data not shown). This might be because the complex slightly dissociated in solution and thus was still not fully bound in solution.

3.4 Investigating the structure of dPHDdCHD

Given that the dPHD and dCHD titrations did not provide conclusive results, a larger construct containing all four domains was studied directly (Figure 2.1, Chapter 2). Uniformly ^{15}N-labelled dPHDdCHD was produced recombinantly in *E. coli* and purified in an analogous manner to dPHD (*protein provided by Dr Rosa Morra*). This larger construct was studied by NMR spectroscopy. Initially, a ^{1}H-^{15}N-HSQC of the labelled dPHDdCHD (600 μM) was collected (Table 8.1). Despite being 38.4 kDa and containing 325 amino acids, very few peaks were observed in the spectrum (Figure 3.10). The peaks that could be observed showed significant chemical-shift dispersion, which suggested that the protein was correctly folded. One reason for the absence of peaks could be protein aggregation, so a second HSQC spectrum was collected at a lower dPHDdCHD concentration (200 μM), but this had no effect on the total number of peaks detected within the spectrum.

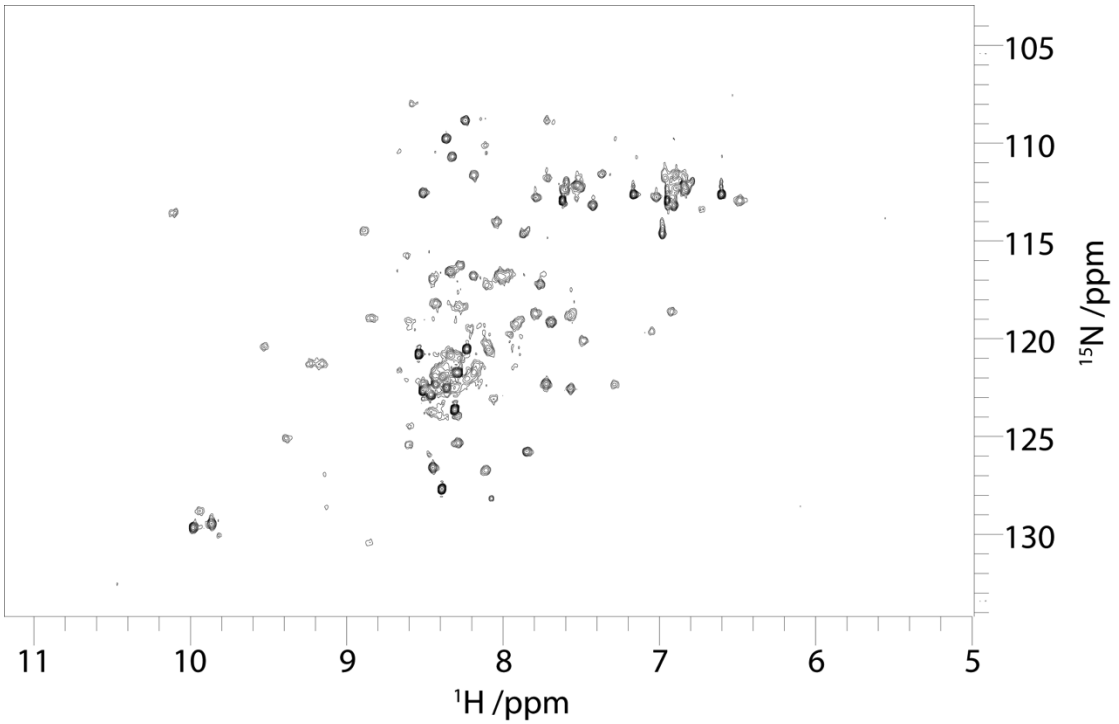


Figure 3.10 950 MHz ^{1}H-^{15}N-HSQC of ^{15}N-labelled dPHDdCHD (600 µM) at pH 7.5, 293 K

Interestingly, the majority of the peaks present, overlaid with the positions of peaks from the 1st PHD suggesting that the structure of this domain was not significantly perturbed in the larger dPHDdCHD construct (Figure 3.11). The absence of peaks corresponding to the 2nd PHD finger suggested that this domain was interacting in some way with the two CHDs also contained within the construct. In addition, the ^{15}N-labelled dPHDdCHD spectrum looked very similar to the spectrum of ^{15}N-labelled dPHD collected after addition of unlabelled dCHD (Figure 3.9B). There are a number of reasons why peaks can be missing from an HSQC spectrum. Large proteins often give peaks with broad line widths due to their slow tumbling rates; in some cases the peaks can become so broad that they cannot be observed in the spectrum. Internal dynamics can also lead to broadening of the peaks if it is on an intermediate timescale (millisecond to microsecond). Given that the loss of peaks was also observed during the dCHD titration as well as in the spectrum of dPHDdCHD, it was proposed that this dynamic situation was most likely the cause of the peak loss.

In some cases, changing the NMR frequency can overcome peak loss due to protein dynamics, so NMR spectra were collected at a variety of frequencies from 500 MHz to 950 MHz. Unfortunately, in no case did any extra peaks appear.

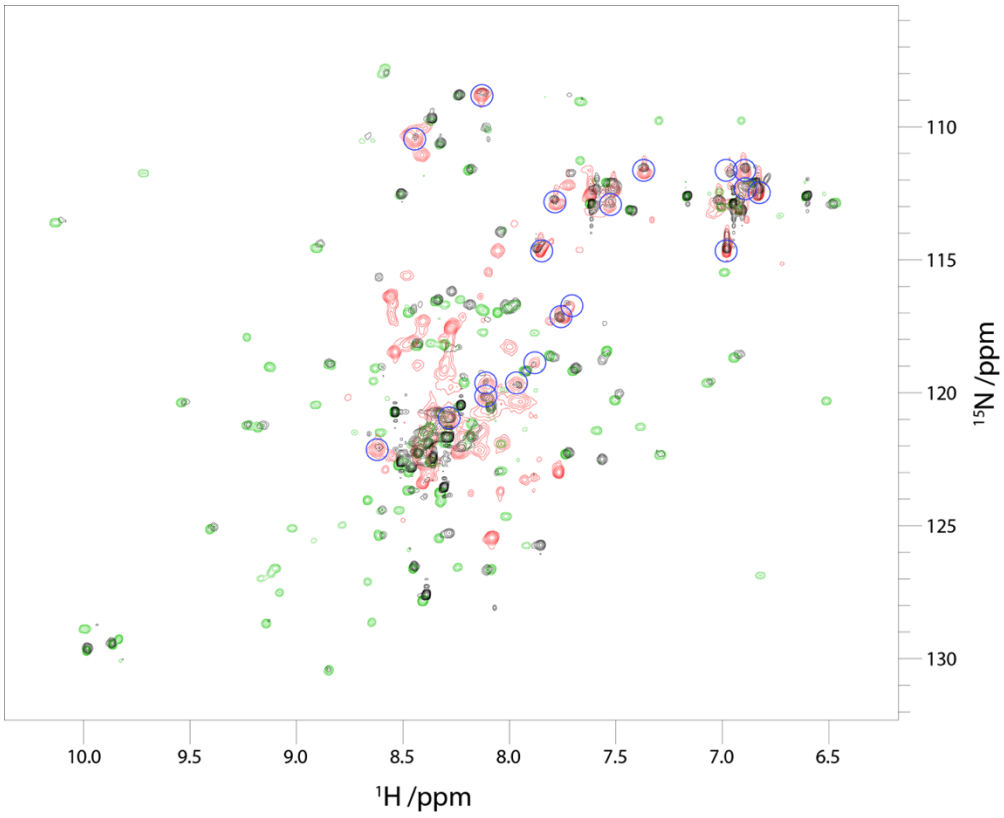


Figure 3.11 Overlay of a 750 MHz ^{1}H-^{15}N-HSQC spectrum of ^{15}N-labelled dPHDdCHD (black), a 950 MHz ^{1}H-^{15}N-HSQC spectrum of ^{15}N-labelled dPHD (green) and a 950 MHz ^{1}H-^{15}N-HSQC spectrum of ^{15}N labelled dCHD (red). Residues in the dPHDdCHD spectrum that overlay with ones from the dCHD spectrum are circled in blue.

In preliminary studies, a ^{1}H-^{15}N-HSQC spectrum of the ^{15}N labelled dCHD construct from CHD4 (Figure 2.1, Chapter 2) *was collected by Professor Christina Redfield*, which gave a poor spectrum that was considered unsuitable for NMR structure determination. Interestingly, however, an overlay of this spectrum with the spectrum of ^{15}N-labelled dPHDdCHD showed that a number of the peaks from the ^{1}H-^{15}N-HSQC spectrum of dPHDdCHD, which did not overlap with peaks from the dPHD ^{1}H-^{15}N-HSQC spectrum, overlapped with peaks in the ^{1}H-^{15}N-HSQC spectrum of dCHD (residues circled in blue, Figure 3.11).

The absence of many of the peaks expected for dPHDdCHD precluded full assignment and structure determination of this construct, but peaks that were present in the dPHDdCHD HSQC spectrum were assigned by comparison between it and the HSQC spectrum of dPHD (Figure 3.12). In some cases, such as for residues A40 and T3, the peaks overlapped exactly and assignment was unambiguous. For other residues, such as H5 and E53, the peaks had shifted slightly in the spectrum and assignment was less certain (assignment of these residues was assisted by the movement of these peaks on binding to H3 1-21 in a similar manner to that observed in dPHD alone, see below).

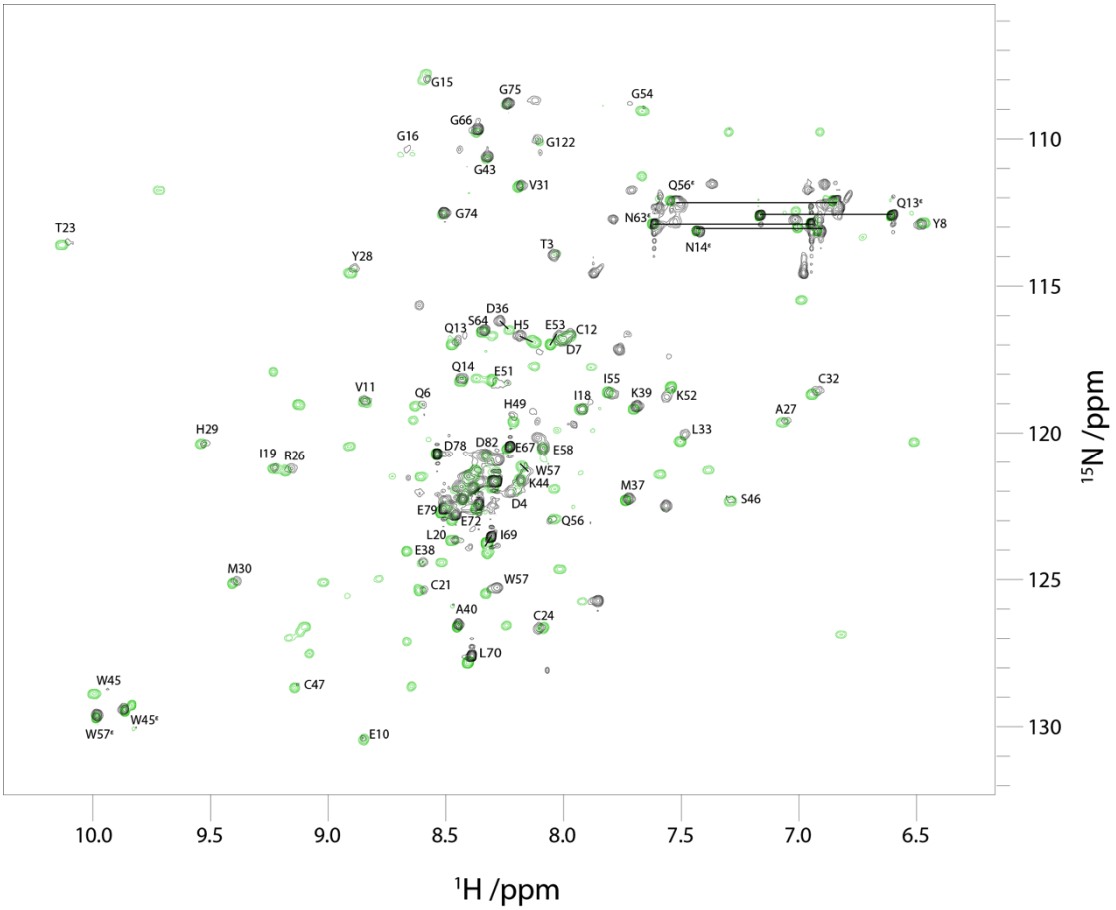


Figure 3.12 Overlay of a ^{1}H-^{15}N-HSQC spectrum of ^{15}N-labelled dPHDdCHD (black) and ^{1}H-^{15}N-HSQC spectrum of ^{15}N-labelled dPHD (green) showing the identity of peaks in the dPHDdCHD construct identified by comparison with the dPHD spectrum.

The differences in the chemical-shift values between dPHDdCHD and dPHD were analysed. Peaks that showed a high degree of variation (change in chemical shift greater than 0.1) were mapped onto the structure of dPHD (red regions in Figure 3.13). Interestingly these residues formed a single contiguous surface, suggesting that one face of the 1st PHD finger may interact with other domains from dPHDdCHD.

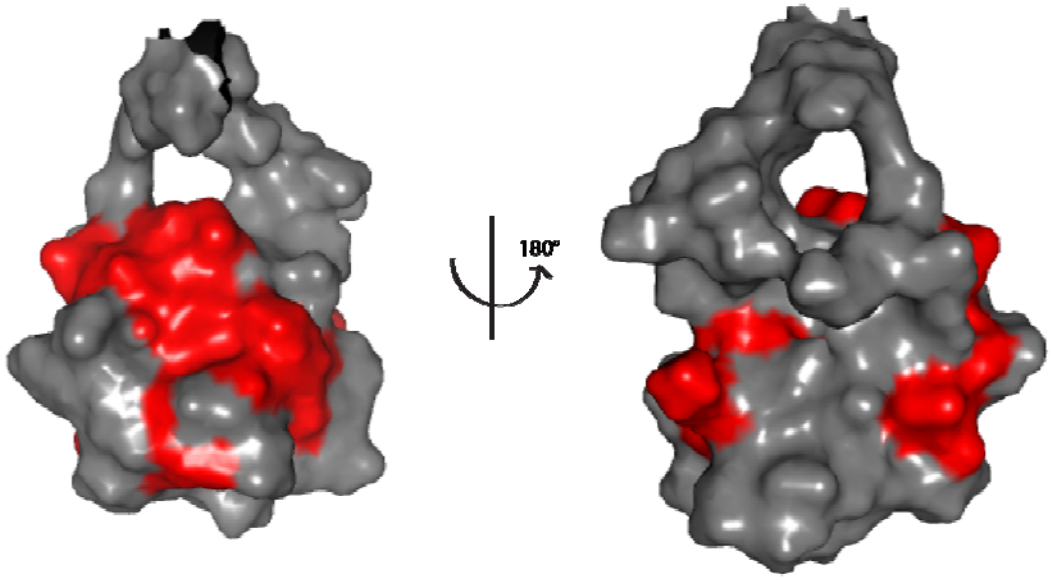


Figure 3.13 Structure of the 1st domain of dPHD (5-60) shown from two angles showing in red the residues whose peaks vary most significantly (change in combined chemical shift > 0.1) between the ^{1}H-^{15}N-HSQC spectrum of dPHD and the ^{1}H-^{15}N-HSQC spectrum of dPHDdCHD.

3.5 Investigating the binding interactions of dPHDdCHD

3.5.1 Investigating the DNA binding of dPHDdCHD

As the CHDs of CHD4 are known to bind to DNA, it was proposed that if the absence of peaks was due to fluctuations of the protein structure on a timescale that caused peak broadening, addition of DNA might stabilise the protein into a single conformation, causing these peaks to sharpen and reappear. A DNA titration was carried out with a 30mer oligonucleotide of random sequence (Sigma Aldrich). On addition of the DNA oligonucleotide to the dPHDdCHD (280 μM), a few peaks were observed to shift (Figure 3.14), but no additional peaks were observed.

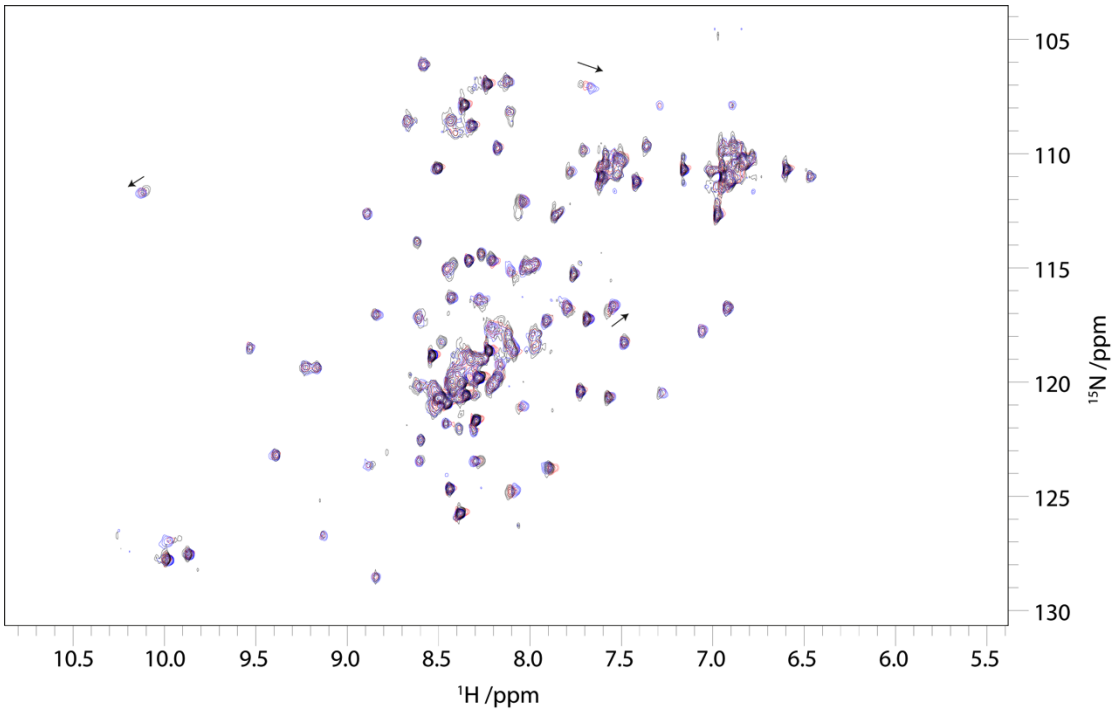


Figure 3.14 500 MHz ^{1}H-^{15}N-HSQC spectrum of ^{15}N-labelled dPHDdCHD (280 µM) at pH 7.5, 293 K showing the effect of addition of a 30mer oligonucleotide. Peak colour identifies the ratio of DNA oligonucleotide added: black, no oligonucleotide; red, 5:1 oligonucleotide:dPHDdCHD; blue, 10:1 oligonucleotide:dPHDdCHD. Black arrows indicate the movement of some of the peaks that shift the most significantly.

3.5.2 Investigating the histone binding of dPHDdCHD

There are examples where binding of a PHD finger to another protein alters its affinity for the histone tail.[285] As it appeared that the 2nd PHD finger of CHD4 was interacting in some way with the two CHDs in this larger dPHDdCHD construct, it was possible that the histone peptide binding behaviour of the domains would be affected by the presence of these two extra CHDs. A titration of the H3 1-21 peptide into dPHDdCHD (300 µM) was carried out to investigate whether the addition of the two CHDs altered the peptide binding properties of the two PHDs. ^{1}H-^{15}N-HSQC spectra were collected at increasing concentration of H3 peptide as was done for the dPHD alone. In initial experiments, although peaks from the first domain were observed to move, the chemical-shift change reached a maximum at 400 µM peptide and then the peaks began shifting back towards their starting position, so that by

900 µM the original peak positions were observed. The reason for this was unclear but was suspected to be due to protein degradation. To avoid this occurring during the experiment, a second titration was carried out using a new aliquot of dPHDdCHD (280 µM) for each peptide titration point. In this experiment peaks moved in a dose-dependent manner towards their saturated positions (Figure 3.15). Although it was hoped that peptide binding might stabilise one conformation of the dPHDdCHD construct, resulting in the appearance of some of the peaks from missing residues, as for the DNA titration (Figure 3.14), no new peaks were observed.

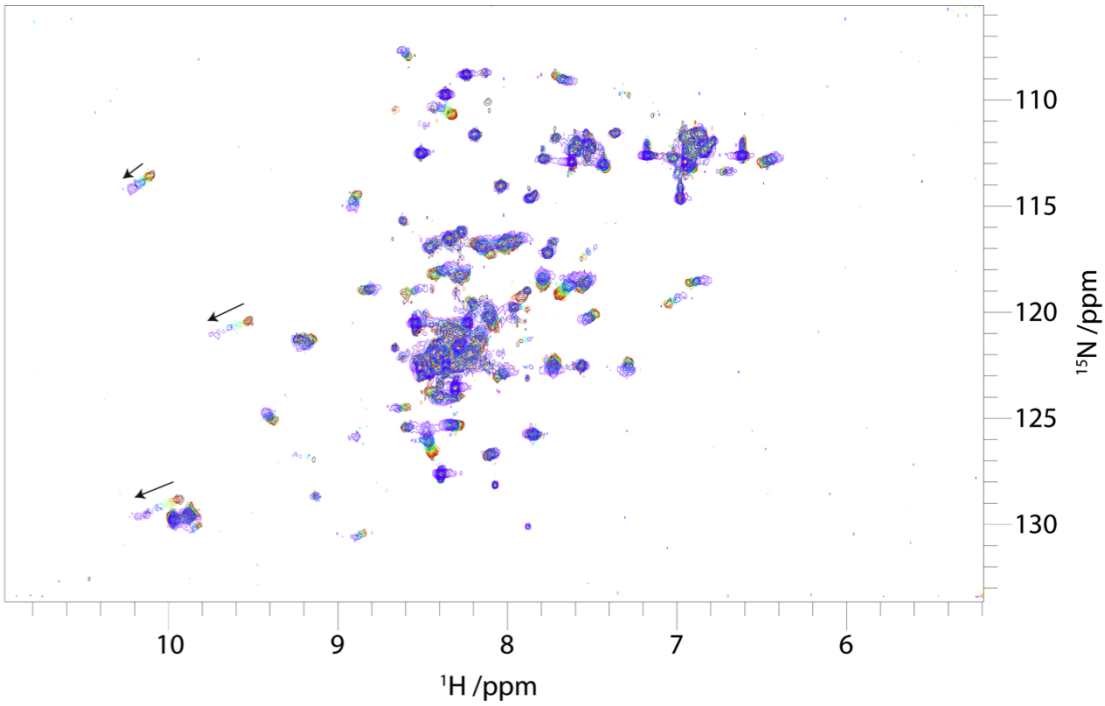


Figure 3.15 750 MHz ^{1}H-^{15}N-HSQC spectra of ^{15}N-labelled dPHDdCHD (280 µM) at pH 7.5, 293 K showing the effect of addition of an H3 1-21 peptide. Peak colour identifies the H3 peptide concentration: black, no peptide; red, 100 µM; green, 200 µM; cyan, 300 µM; blue, 450 µM; blue, 600 µM; blue, 800 µM; purple, 1.2 mM; purple, 1.8 mM; purple, 2.5 mM. Black arrows indicate the movement of some of the peaks that shift the most.

Plotting the chemical-shift change of A40 against the concentration of H3 1-21 peptide added, showed that a similar lag in peak movement occurred as for the dPHD construct alone (Figure 3.16A and B). This suggested that the larger construct was still binding to two peptides and thus that despite peaks from the 2nd PHD having

disappeared, suggesting that it was interacting with the CHDs, its conformation was still amenable to peptide binding. Without information from residues from the high affinity 2nd domain binding site it was not possible to fit exact K_D values, but given the high similarity between the peak movement in the two constructs it is likely that they were similar to the dPHD alone, with the 2nd domain having a higher affinity than the 1st domain (Figure 3.16B). Data in a recently published paper suggests that the binding affinities of histone peptides to this construct (measured by surface plasmon resonance), increase by over an order of magnitude from the tandem PHDs alone.[245] Although it is possible that the data in Figure 3.16 shows this construct to have slightly higher K_D values due to the slightly steeper gradient of the curve, it is difficult to compare the experiments between reports, since in the published work the binding is fitted to a single-site binding model, which from the data presented here is likely incorrect and may account for the difference to the values predicted here.

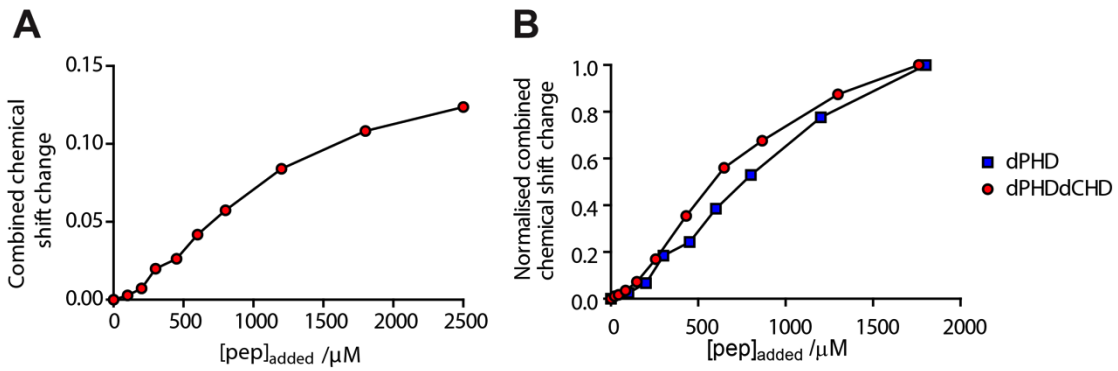


Figure 3.16 A) Graph showing the change in chemical shift for peaks corresponding to A40 of dPHDdCHD on addition of increasing concentrations of H3 1-21 peptide. **B)** Graph showing a comparison between the change in chemical shift for peaks corresponding to residue A40 for dPHDdCHD and H29 for dPHD normalised by dividing all changes in chemical shift by the change in chemical shift observed at the highest concentration of added peptide.

3.5.2.2 Confirmation of the binding of two histone peptide to dPHDdCHD by MS

Given that in this dPHDdCHD construct it was only possible to observe peaks from the first domain, and peptide binding to the second domain was only inferred from the shape of the binding curves generated from residues in PHD1, MS analysis was again carried out to confirm the binding of dPHDdCHD to two histone peptides. The protein was first buffer exchanged into 200 mM ammonium acetate and native MS collected *by Ima-Obong Ebong*. As was the case for the shorter dPHD construct (Figure 3.7), peaks were observed for addition of both one (~39.7 kDa) and two (~40 kDa) peptide molecules to the protein (~37.4 kDa, Figure 3.17), supporting the proposal that the two PHDs are still able to independently and simultaneously bind to a histone peptide.

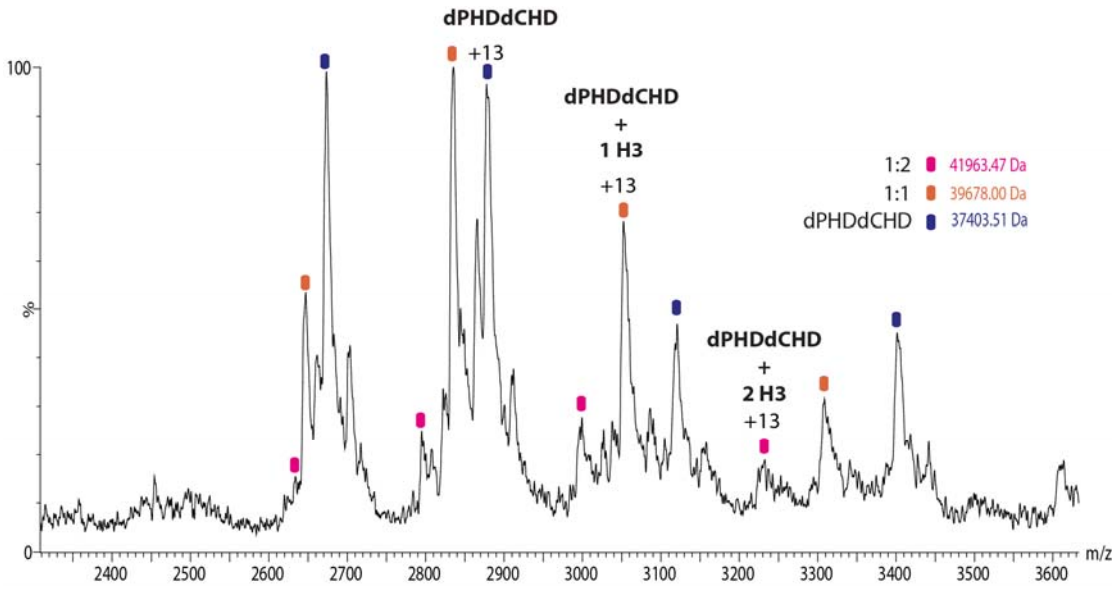


Figure 3.17 Native mass spectrum of dPHDdCHD (20 µM) bound to either one or two histone H3 1-21 peptides (1 mM).

3.6 Discussion

Chromatin remodellers are complex proteins that contain multiple different interaction domains in addition to their ATPase domain. The mechanism of targeting of these proteins to specific areas of chromatin is currently not well understood. Work in this Chapter demonstrated that CHD4 contains two tandem PHD fingers that are both able to bind independently to H3 tails. NMR spectroscopy provides a good method for measuring low affinity interactions and unlike other biophysical techniques can easily distinguish between one-site and multiple-site binding, and also gives additional information on the interacting regions of the protein and peptide. However, inherent errors on measurements of the chemical shift changes and sample concentrations, coupled to the requirement for high protein concentration of the samples make it less reliable for accurately determining high affinity interactions (<10 μM). It is also costly due to its high reagent usage and requirement for labelled protein. As a result, accurately measuring the K_D with the tightly binding H3 K9Me$_3$ peptide was not possible. Further experiments using an alternative biophysical technique such as isothermal titration calorimetry (ITC) or fluorescence polarisation could be used to measure these values.

The affinity of the first PHD finger for all the peptides studied is substantially lower than for the second PHD finger, of a magnitude that in isolation might be considered too weak to occur *in vivo*. The lower affinity of the first PHD for the histone tail in the full length protein, however, may be overcome by the increased local histone concentration it encounters as a result of binding by the second PHD to a histone tail and also by binding of the two neighbouring CHDs to nucleosomal DNA. Each

nucleosome contains two H3 proteins in the histone octamer and it is thus possible that the tandem PHD fingers of CHD4 are able to bridge the gap between them, anchoring the protein to the nucleosome more strongly than would occur with only a single PHD finger (Figure 3.18A). In a second model, the PHD fingers could bind to H3 tails in two neighbouring nucleosomes perhaps helping the chromatin remodeller to facilitate nucleosome sliding (Figure 3.18B). This ability of CHD4 to enhance chromatin compaction, which will reduce the availability of DNA for transcription, is consistent with the repressive role of CHD4 in cells. The abrogation of binding on trimethylation of H3 K4, which is often found at the promoters of "active" genes, is consistent with a model where CHD4 compacts chromatin by binding of both its PHD fingers to H3 tails on adjacent nucleosomes in regions of "silent" genes whilst not acting at "active" regions where the chromatin needs to be more open to allow for transcription.

Recent work by Musselman *et al.* using tryptophan fluorescence and gel shift assays has demonstrated that the tandem PHD domains can bind bivalently to tetrasomes *in vitro*, suggesting that the model in Figure 3.18A is more likely.[239] They further show that mutation or deletion of either PHD fingers from CHD4 reduces the repressive function of CHD4 in cells, although the effect is more pronounced for mutations in the 2nd PHD than the 1st PHD. It is nevertheless possible that in a nucleosomal array, where neighbouring nucleosomes are held close together by the joining DNA, the two PHD fingers bind to separate nucleosomes. A recent report has shown that *in vitro* the nucleosome sliding ability of CHD4 is reduced in the absence of the two PHD domains.[271] It would be interesting to investigate the effect of changing the length and amino acid composition of the linker between the two PHD fingers on the activity of

CHD4. Shortening the linker might prevent the two PHD fingers from simultaneously binding H3 tails and thus reduce the repressive function of CHD4 in a similar fashion to removal of one of the two PHDs.

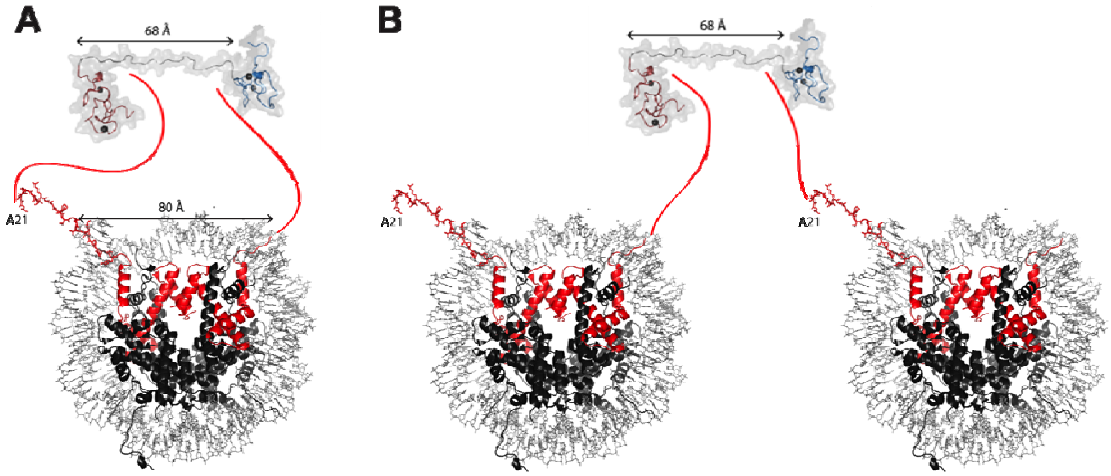


Figure 3.18 Possible models for the binding mode of the tandem PHD domains of CHD4. Nucleosome structure taken from PDB ID 1AOI.[12] **A)** The two PHD fingers bind separate H3 tails (red) on the same nucleosome. **B)** The two PHD fingers bind H3 tails on neighbouring nucleosomes. Red lines represent schematically the end of the H3 tails.

The absence of significant numbers of peaks in the dPHDdCHD spectrum made structural assignment difficult (Figure 3.10). It is possible that this was due to protein instability due to the position of the truncation produced. Further work could investigate production of a more stable protein construct that might be more amenable to NMR. Although the size of the protein is approaching the limit of solution phase NMR, the high modularity of the individual domains means that the protein should behave similarly to smaller proteins; although half the peaks are missing from the dPHDdCHD spectrum, those that can be obtained remain sharp and assignment should remain possible. Whether there is any flexibility between the two CHDs remains to be investigated. Using the current construct, cysteinyl-residues could be inserted by site-directed mutagenesis in relevant positions for attachment of spin-labels, which could be used to determine inter-domain distances and dynamics.

Tandem reader domains are a common feature of chromatin-associated proteins. In a number of cases the neighbouring domains act together to direct the binding of the protein. For example, the tandem PHD-Bromodomain of BPTF has been shown to bind with the PHD finger bound to an H3 tail and the bromodomain bound to an H4 tail on the same nucleosome.[277] Studies described in this Chapter on the tandem pairs of PHDs and CHDs in CHD4 suggest that there is some degree of interaction between the PHDs and the CHDs, but that the protein is still able to bind to two histone peptides with similar binding affinities. Although the absence of peaks in the spectra of dPHDdCHD preclude further structural studies, and the dCHD construct has been shown, so far, not to be amenable to study by NMR spectroscopy, in further work it would be interesting to look at a construct containing the 2nd PHD and 1st CHD of CHD4 and to look at the interaction of dPHD and dPHDdCHD with intact nucleosome arrays.

Chapter 4
Investigations into the structure and function of the histone demethylase PHF8

4.1 The KDM7 family of histone demethylases

The human KDM7 subfamily of histone demethylases consists of PHF8, KIAA1718 and PHF2.[166] These enzymes share significant sequence similarity and all accept only mono- and dimethyllysyl-residues, but PHF8 is an H3 K9$Me_{2/1}$ demethylase, whilst KIAA1718 acts on K27$Me_{2/1}$.[292–294] PHF8 has also been shown to demethylate H4 K20Me_1 both in cells and on nucleosomal substrates *in vitro*.[295,296] This activity has not been demonstrated with peptidic substrates, possibly due to the requirement for concomitant binding of the PHD finger to an H3 tail. Crystal structures of PHF8 show that the methylammonium binding pocket is tighter than that found in the JMJD2 subfamily, likely precluding trimethyllysine binding and demethylation due to steric hindrance.[145,297]

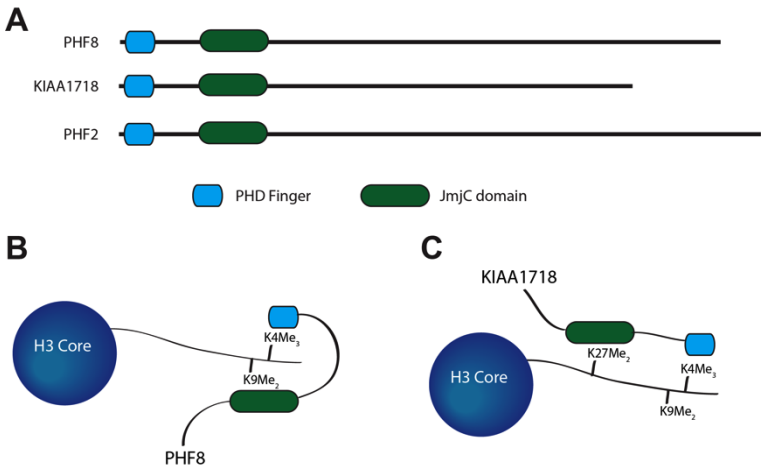


Figure 4.1 A) KDM7 subfamily domain architecture. **B)** Proposed binding mode of PHF8 to the H3 tail. **C)** Proposed binding mode of KIAA1718 to the H3 tail.

All three KDM7 members contain a PHD finger that binds the histone tail at H3 K4Me$_3$ and is thought to tune the selectivity of the JmjC domain (Figure 4.1).[176] For PHF8 the linker between the PHD finger and JmjC domain is flexible, such that binding of the PHD finger to H3 K4Me$_3$ targets the catalytic domain to K9 (Figure 4.1B). In contrast, the JmjC and PHD domains of KIAA1718 are connected by a rigid linker, such that when the PHD finger binds to H3 K4Me$_3$ the JmjC domain is held too far along the histone tail to demethylate at K9 and instead demethylates at K27 (Figure 4.1C).[176] Unlike PHF8 and KIAA1718, the activity of PHF2 remains questionable. PHF2 was initially thought to be inactive as it contains only two of the usual three Fe(II) binding residues, although it was subsequently shown to be able to bind Fe(II) using a tyrosyl-residue (Y321) in place of the second histidyl-residue.[177] Wen *et al.* demonstrated that the PHD finger of PHF2 binds to H3 K4Me$_3$, as for the other KDM7 members, and that *in vivo* PHF2 demethylates H3 K9Me$_1$ at rDNA promotors.[298] A more recent study suggests that PHF2 becomes active as an H3 K9Me$_2$ demethylase upon phosphorylation of a seryl-residue (S1056) by protein kinase A (PKA).[299] In this same study PHF2 was shown to associate with ARID5B (AT-rich interactive domain 5B), which helps direct PHF2 to its gene targets and is itself also a substrate for PHF2, undergoing demethylation of a lysyl-residue (K336).

Members of the KDM7 subfamily are implicated in development. Mutations in the human PHF8 gene in either the PHD finger or JmjC domain are linked to midline defects, such as cleft lips and palettes, and X-linked mental retardation.[220–222] Morpholino knockdown of the zebrafish PHF8 homologue causes brain and facial abnormalities in embryos consistent with the human phenotype.[295] Studies of KIAA1718 in chickens have shown that it promotes neural induction and PHF8 has

been shown to affect retinoic acid-induced neuronal differentiation in mouse embryonic carcinoma cells.[300,301] The *C. elegans* PHF8 homologue has been shown to be important for neuronal function, with worms bearing a PHF8 variant lacking the PHD finger showing impaired locomotion.[219]

4.1.2 Objectives

The work described in this Chapter involved the characterisation of two members of the KDM7 subfamily, PHF8 and KIAA1718. The work involved development of a high-throughput biochemical screening assay for PHF8, which allowed determination of small molecule IC_{50} values. Further work investigated the substrate scope of PHF8 with N^{ε}-lysine groups other than methyllysine. Finally immunofluorescence studies on a novel clinically-relevant mutation in PHF8 are described.

Modified lysine peptides used in this Chapter were synthesised by Dr. Richard Hopkinson. Inhibitor screening for enzymes other than PHF8 and KIAA1718 was carried out by Dr. Anthony Tumber (Structural Genomics Consortium, Oxford). Cell-based assays other than immunofluorescence with PHF8 were carried out by Dr. Shirley Li or collaborators in the group of Professor Naoki Miyata. The crystal structure of PHF8 with Daminozide was produced by Dr Tobias Krojer and the PHF8-Daminozide model was generated by Dr Nathan Rose.

4.2 Characterisation of PHF8

Three constructs of PHF8 were expressed recombinantly in *E. coli*; the PHD finger and JmjC domain (residues 37-483, FL PHF8 herein), the JmjC domain alone (residues 115-483) and the PHD finger alone (residues 37-116) (Figure 4.1A). These

were all purified as *N*-terminal 6×His-thioredoxin (Trx) tagged constructs, initially using a 5 mL HisTrap HP column, with tag cleavage using Tobacco Etch Virus (TEV) protease (overnight, 4 °C) and then a further size exclusion chromatography step (Section 8.5). Purified proteins (> 85% pure as judged by SDS-PAGE) were concentrated and flash frozen with liquid nitrogen before storage at -80 °C.

The activities of the FL construct and JmjC only construct were tested using a Matrix assisted laser desorption/ionisation-time of flight (MALDI-TOF) MS based assay (Section 8.7.1). This method gives an indication of enzyme activity by measuring the accurate mass of the substrate peptides, whereby demethylation may be observed as a decrease in the mass of the substrate peptide by increments of 14 Da. FL or JmjC only PHF8 (2 μM) were incubated with a histone H3 peptide methylated at H3 K4 and H3 K9 (ARTKMe_3QTARKMe_2STGGKA (GLS Biochem >95% pure), note that no demethylation of H3 K4 is expected) and cofactors (Fe(II) (100 μM), ascorbate (100 μM), 2OG (200 μM)). Reactions were quenched with methanol after 3 hours. Virtually no demethylation was observed with the JmjC only PHF8 construct (Figure 4.2A), but complete demethylation at H3 K9 was observed with FL PHF8 (Figure 4.2B).

The activity of recombinant PHF8 is significantly enhanced by the presence of the PHD finger. Though this may be due to the PHD finger anchoring the peptide to the JmjC domain, it is also possible that binding of the PHD to the peptide alters the conformation of the peptide making demethylation more favourable. To test whether the PHD finger of PHF8 can enhance the activity of the JmjC domain when "*in trans*", a MALDI assay was carried out including a 1:1 or 1:2 ratio of the JmjC only

construct and the PHD only construct together and incubating with the substrate peptide. However, no enhancement in the activity was observed (data not shown).

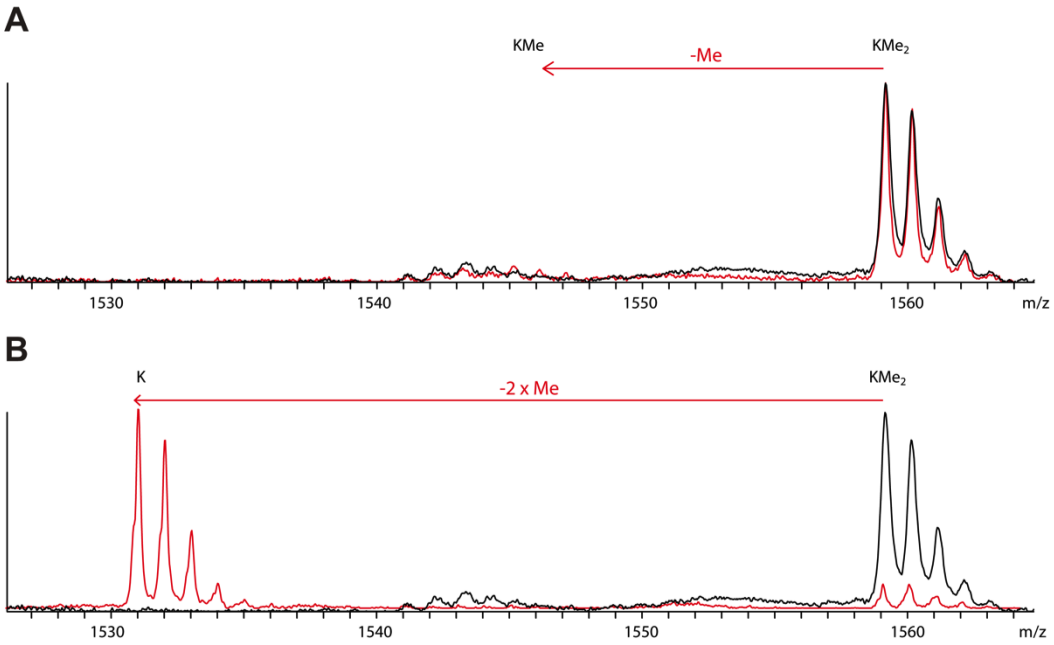


Figure 4.2 MALDI MS analysis of assays testing the activity of two PHF8 constructs. **A)** JmjC only. **B)** FL PHF8. Red trace shows reactions quenched after 3 hours, black trace shows the no enzyme control.

To investigate the activity of FL PHF8 in more detail, a timecourse with the same peptide was carried out over 90 minutes (Figure 4.3). An initial rapid increase in the amount of monomethylated peptide was observed. This appeared to reach a steady-state concentration, and a slower increase in the non-methylated peptide was detected over the course of the experiment.

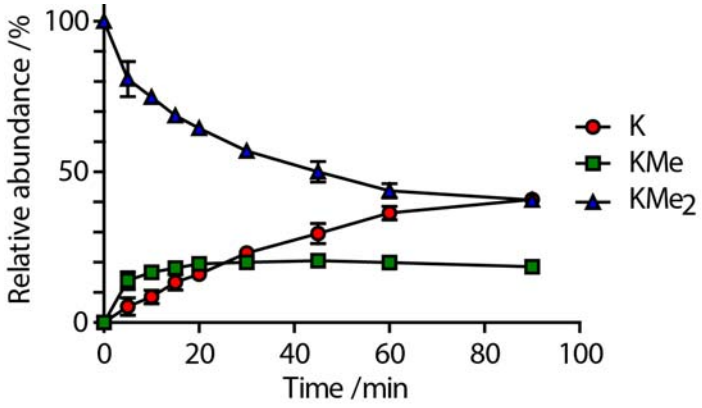


Figure 4.3 Demethylation of H3 K4Me$_3$K9Me$_2$ 1-15 (20 µM) by FL PHF8 (2 µM) over time, measured by MALDI MS.

4.3 Characterisation of KIAA1718

A KIAA1718 construct containing the PHD and JmjC domains (residues 38-480, FL KIAA1718 herein) was produced recombinantly. A MALDI assay was then carried out to examine the activity of this construct by incubating FL KIAA1718 with peptide substrates and cofactors for 1 hr at 37 °C. As for PHF8, the reaction was analysed by MALDI-TOF MS. Consistent with previous reports in the literature,[176,302] KIAA1718 catalysed demethylation of an H3 K27Me$_2$ peptide (GLS Biochem >95% pure) to the non-methylated state, but showed no demethylase activity on the PHF8 substrate peptide, H3 K4Me$_3$K9Me$_2$ (Figure 4.4).

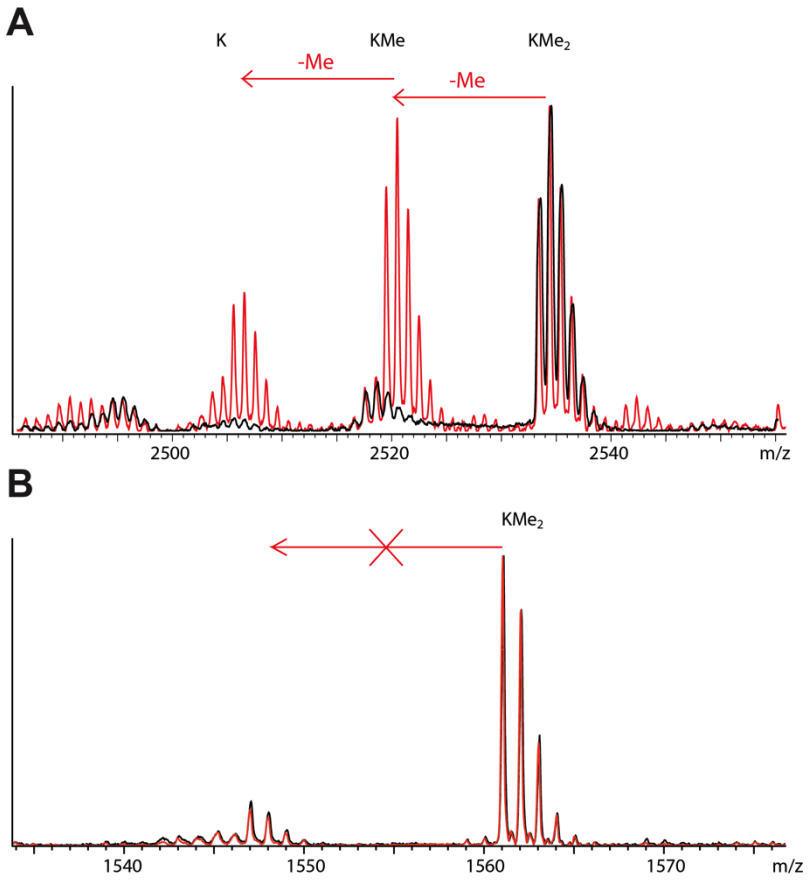


Figure 4.4 MALDI MS spectra showing the activity of recombinant FL KIAA1718. **A)** H3 K27Me$_2$ peptide substrate. **B)** H3 K4Me$_3$K9Me$_2$ peptide substrate. Red trace shows reaction after 1 hr, black trace shows the no enzyme control

4.4 Testing for demethylation of ARID5B

A recent paper has identified K336 of ARID5B as a substrate of the third KDM7 subfamily member, PHF2.[299] In this study, Baba *et al.* demonstrate PHF2-catalysed demethylation of full length ARID5B by immunofluorescence, using an antibody to the methylated residue. The authors identified the target demethylation site due to its similarity to the H3 sequence. On the basis of this alignment (Figure 4.5A), solid phase peptide synthesis (SPPS) was used to synthesise a peptide beginning at the alanine in ARID5B that aligns with the start of the H3 tail (ALYKYMKERKMe_2TPIERIP) to see whether ARID5B was also a substrate for PHF8 and / or KIAA1718 (Section 8.11). However, no demethylation was observed with either FL PHF8 or FL KIAA1718 1 hr after incubation with cofactors, suggesting that the dimethylated ARID5B peptide was not a substrate for either enzyme (Figure 4.5B). As the previous PHF2 study was conducted using full length ARID5B protein, it is possible that the lack of demethylase activity with PHF8 and KIAA1718 was due to the peptide adopting the wrong conformation in solution. However, given PHF8 and KIAA1718 are known to demethylate unstructured histone peptides, and thus do not require any secondary structure for binding, it is perhaps more likely that this substrate is demethylated specifically by PHF2.

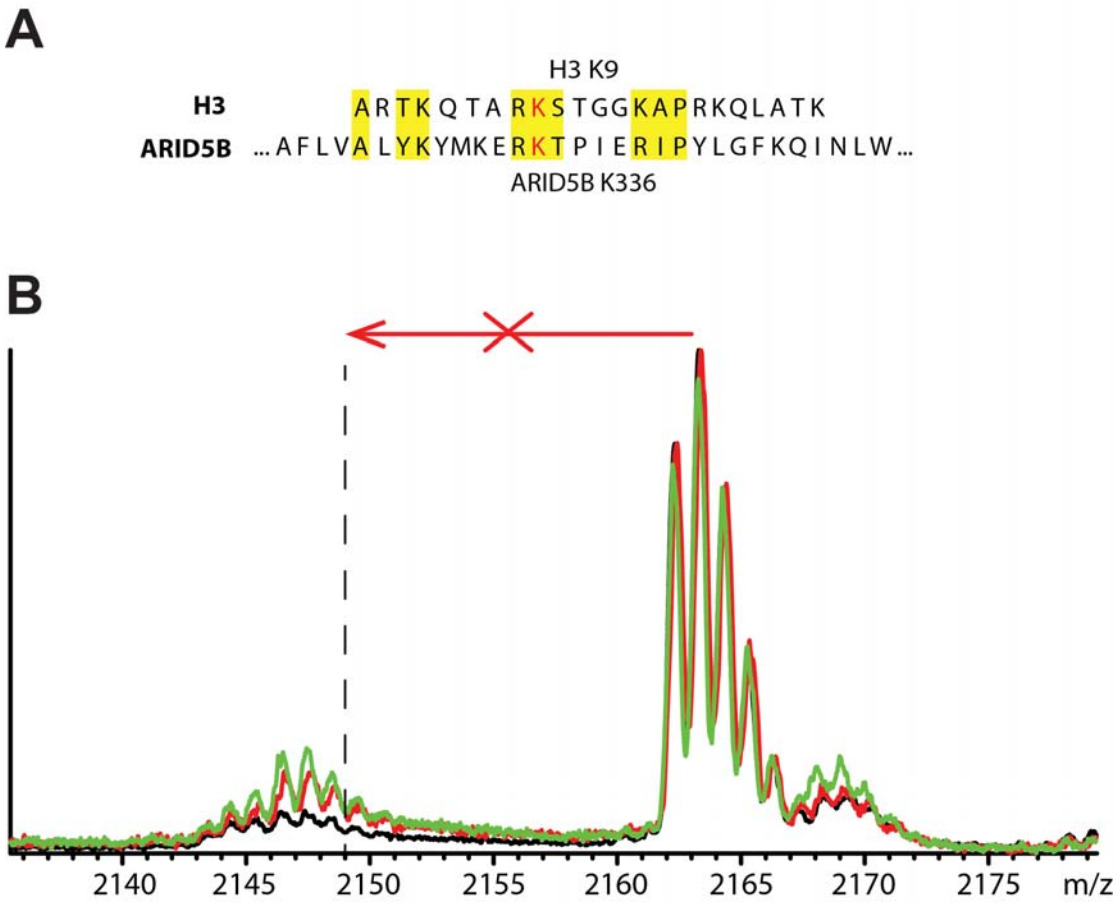


Figure 4.5 A) Alignment of histone H3 (above) and ARID5B (below) sequences. The target lysine residues are shown in red and residues that are similar between the two sequences are highlighted in yellow. **B)** MALDI spectra of ARID5B peptide (ALYKYMKERKMe_2TPIERIP) incubated with PHF8 (2 μM, red) or KIAA1718 (2 μM, green). Black spectrum shows the control reaction without enzyme.

4.5 Investigating peptide binding by PHF8 and KIAA1718

Work by Horton *et al.* has shown that KIAA1718 does not demethylate H3 K9 Me_2 when H3 K4Me_3 is present. Nevertheless, isothermal titration calorimetry (ITC) experiments indicate that KIAA1718 can bind to peptides containing both methylated residues.[176] To independently verify these results, the binding affinities of PHF8 (FL and PHD alone) and FL KIAA1718 for an H3 K4Me_3K9Me_2 peptide were determined.

Initially, ITC was attempted but did not prove successful (Appendix 2). Instead, a fluorescence polarisation (FP) assay (Section 8.7.5) was used to measure the K_D of PHF8 (FL, JmjC only and PHD alone) and FL KIAA1718 for an H3 K4Me$_3$K9Me$_2$ peptide conjugated to fluorescein isothiocyanate (FITC, Figure 4.6A) (purchased from GL Biochem Ltd., >95% pure).[303] FITC conjugation was carried out at the *C*-terminus of the peptide so as not to interfere with binding to the PHD finger of the proteins, which interacts with the *N*-terminal residues of H3 (Figure 4.6B).

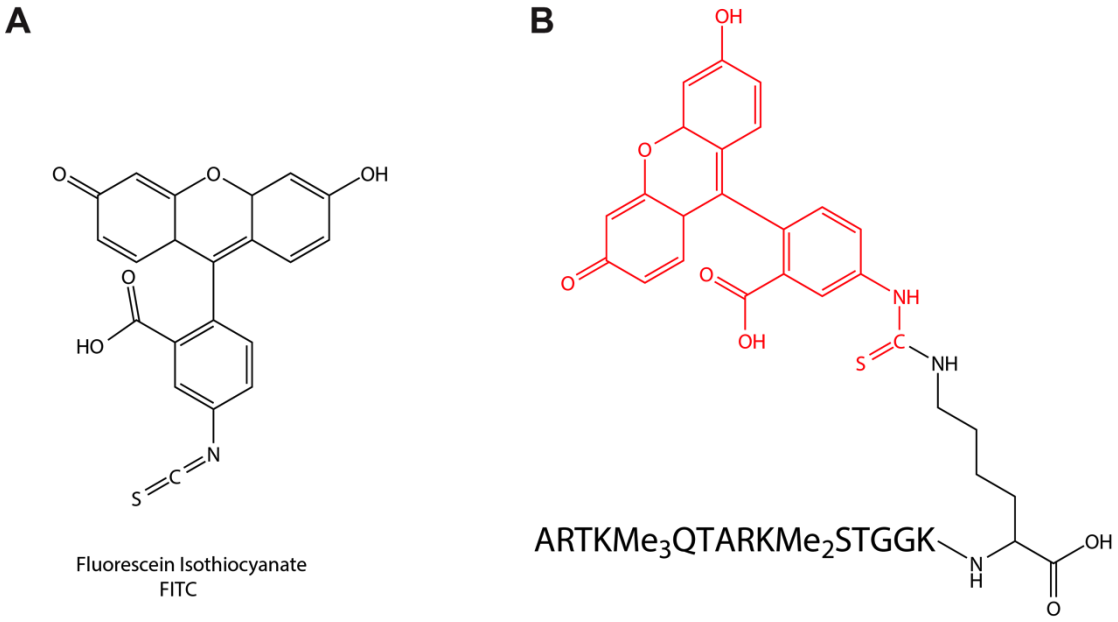


Figure 4.6 A) Structure of FITC. **B)** Structure of the H3 1-14 FP peptide probe. FITC is shown in red.

H3 1-14 K4Me$_3$K9Me$_2$-FITC peptide (50 nM) was incubated with increasing concentrations of protein (from 50 nM to 10 μM) for 15 min before the polarisation was measured and the change in polarisation calculated relative to the no protein controls (Figure 4.7). K_D values were then determined from the curves (Table 4.1). A lower K_D (corresponding to a stronger binding affinity) was observed for the FL PHF8 construct than for the FL KIAA1718 construct. This is consistent with the crystallographic observation that both domains of PHF8 (JmjC and PHD) bind to the

peptide leading to a larger binding surface with more protein-peptide interactions than for KIAA1718, where only the PHD finger (but not the JmjC domain) binds to the peptide.[176] These results imply that the lack of demethylase activity of KIAA1718 at H3 K9Me$_2$ is not due to an absence of binding to that region of the histone tail but due to the peptide being held in the wrong orientation to place the methylated lysyl-residue at the catalytic Fe(II) centre of the JmjC domain.

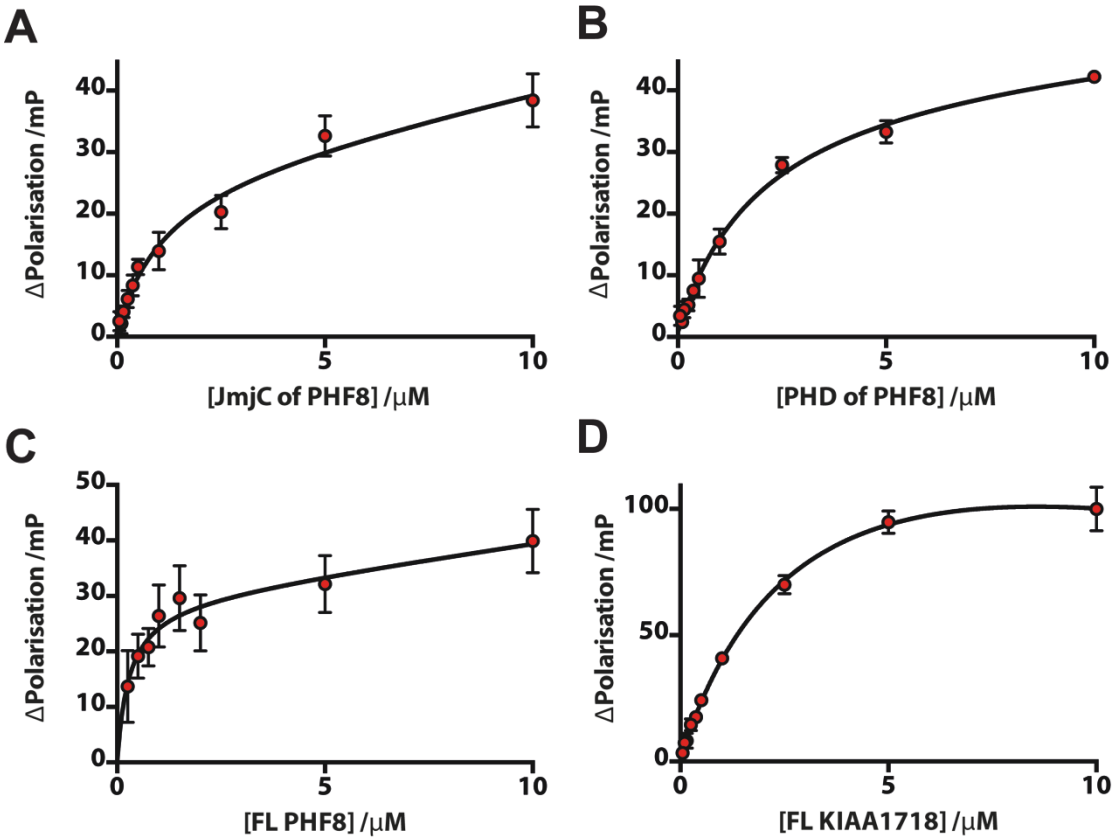


Figure 4.7 Fluorescence polarisation assay measuring the affinity of **A)** JmjC of PHF8, **B)** PHD of PHF8, **C)** FL PHF8 and **D)** FL KIAA1718 for an H3 K4Me3K9Me2-FITC peptide. Points are plotted as the mean of triplicate experiments with error bars showing ± 1 s.d. Data was fitted using a one-site binding model using GraphPad prism.

Table 4.1 Binding constants for PHF8 and KIAA1718 with H3 K4Me$_3$K9Me$_2$.

Enzyme	K_D /μM
JmjC of PHF8	1.1 ± 0.5
PHD of PHF8	2 ± 0.5
FL PHF8	0.28 ± 0.2
FL KIAA1718	3.5 ± 0.7

4.6 Inhibitor assays

Many of the proteins involved in epigenetic regulation play complex roles in cellular development and disease. As such they are interesting targets for drug discovery.[211] However, in many cases, apparent redundancy in their activities, as based upon their *in vitro* substrates, makes it difficult to dissect the role of each individual protein in any given disease. To address this challenge there is a drive to develop small molecule probes to investigate the function of individual proteins in cells. In order to be useful, inhibitors need to be cell permeable and highly selective for a given enzyme. As all 2OG oxygenases have the same basic catalytic mechanism and 2OG-binding site, designing selective inhibitors is difficult, but can be achieved by exploiting small differences in the active-site environment. There is already evidence that selective inhibition can be achieved; an inhibitor of the 2OG oxygenase BBOX1 is currently used in the clinic and drugs to treat anaemia by inhibiting the HIF prolyl-hydroxylases (PHD1-3) are also in clinical trials.[304]

In collaboration with the Structural Genomics Consortium (SGC) and a number of pharmaceutical companies, work in my group is directed towards developing chemical probes for each subfamily of the JmjC KDMs to allow study of their biological functions. In addition to inhibitor synthesis, a panel of inhibitor assays for each subfamily of the JmjC KDMs is required to evaluate the potency and selectivity of each inhibitor. Work described below aimed to develop a high-throughput assay for measuring IC_{50} values against PHF8 for use in the selectivity panel.

4.6.1 Kinetic characterisation of PHF8

In order to design an assay to measure IC_{50} values with inhibitors, knowledge of the enzyme of interest's kinetics are first required. Michaelis-Menten kinetics were applied to the PHF8 reaction to determine K_M and k_{cat} values for 2OG and peptide. As an alternative to mass spectrometry, a formaldehyde dehydrogenase (FDH) coupled assay that indirectly measures the production of formaldehyde over the course of the reaction, allowing direct measurement of the initial reaction rate was used. *Conditions for carrying out the FDH assay with PHF8 were previously optimised by Timothy Rooney*.

4.6.2 Determination of PHF8 kinetic parameters

The 2OG and peptide kinetic parameters for PHF8 were determined using the FDH assay (Figure 4.8, Section 8.7.2). To measure the $K_M(2OG)$ a saturating peptide concentration (100 μM) was used so that the concentration of 2OG would be the only factor limiting the reaction rate. The initial rate was measured at increasing concentrations of 2OG and the data fitted using the Michaelis-Menten equation to give a $K_M(2OG)$ of 16 μM. Having calculated a $K_M(2OG)$, a $K_M(peptide)$ was then measured. Here a concentration of 200 μM 2OG was used to ensure that this was no longer the component limiting the initial reaction rate. Kinetic parameters are shown in Table 4.2; a K_M value of 9.2 μM for the peptide was obtained.

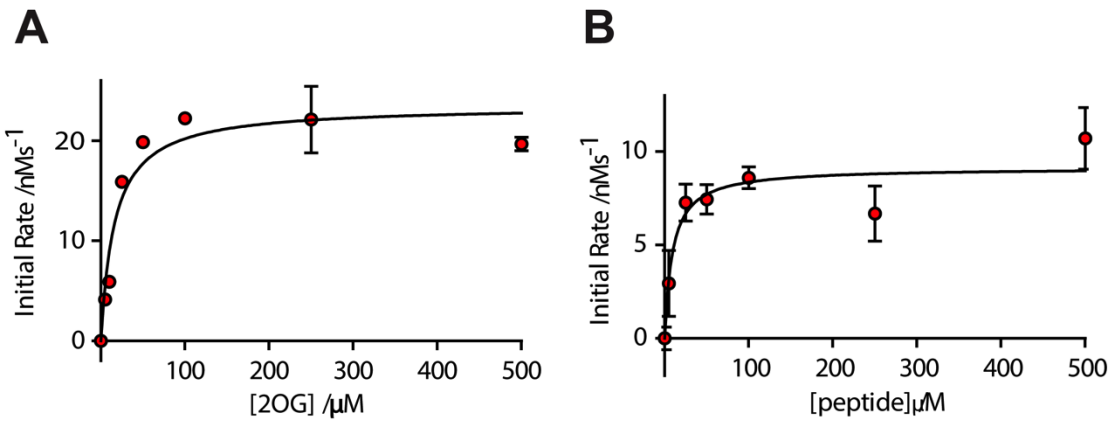


Figure 4.8 Kinetic parameters for PHF8. **A)** $K_M(2OG)$. **B)** $K_M(H3\ 1\text{-}15\ K4Me_3K9Me_2)$.

Table 4.2 Kinetic parameters for PHF8.

Substrate, X	$K_M(X)$ /μM	$k_{cat}(X)$ /μM.s^{-1}
2OG	16	0.046
H3 1-15 K4Me$_3$K9Me$_2$	9.2	0.018

4.6.3 KIAA1718 kinetic parameters

2OG and peptide K_M values were also determined for KIAA1718 (2 μM) using the MALDI MS assay (Figure 4.9, Section 8.7.1). For the $K_M(2OG)$, 75 μM peptide was used and for the $K_M(peptide)$, 200 μM 2OG. The kinetic parameters for KIAA1718 are shown in Table 4.3. Similar values were obtained to those for PHF8 as expected given the similarity of the catalytic domains. Kinetic experiments for 2OG showed low levels of substrate inhibition at high 2OG concentrations (250 μM and 500 μM, Figure 4.9).

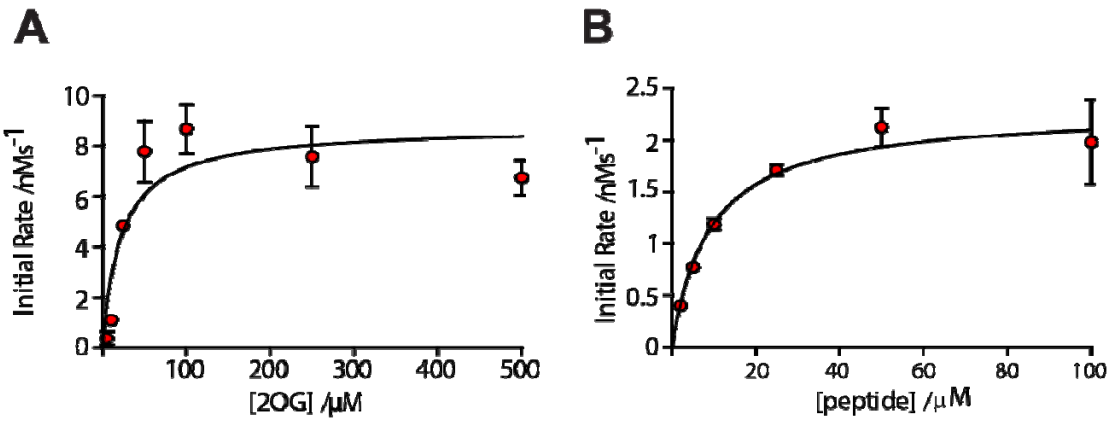


Figure 4.9 Kinetic parameters for KIAA1718. **A)** $K_M(2OG)$. **B)** $K_M(H3\ 21\text{-}44\ K27Me_2)$.

Table 4.3 Kinetic parameters for KIAA1718.

Substrate, X	K_M (X) /μM	k_{cat} (X) /μM.s^{-1}
2OG	22	0.0011
H3 21-44 K27Me$_2$	9.1	0.0029

4.6.4 IC_{50} value determination

Both the MALDI-TOF MS and FDH assays can be used to determine inhibitor IC_{50} values. However, both assays require relatively high concentrations of enzyme, precluding determination of IC_{50} values for potent inhibitors, because the measurable IC_{50} limit is half the concentration of the protein being inhibited in the assay. In addition, MALDI-TOF MS assays are laborious, preventing high-throughput screening. Thus, in order to carry out screening of large compound libraries, development of a more efficient assay was required.

4.6.5 ALPHA screen assay development for PHF8

The amplified luminescent proximity homogeneous assay (ALPHA) screen has previously been used for a number of high-throughput assays, including for histone demethylases.[305,306] As well as being high-throughput, ALPHA screen also uses low concentrations of enzyme, allowing for much lower IC_{50} values to be measured. Thus,

an ALPHA screen for measuring PHF8 activity was developed to facilitate compound screening.

4.6.5.1 Principles of the ALPHA screen

The ALPHA screen uses a pair of beads, which can be made to emit light when brought into close proximity.[305] Laser excitation of the donor bead at 680 nm causes a photosensitiser to convert triplet oxygen to its excited singlet state. The acceptor bead interacts with this oxygen molecule, emitting luminescence between 520 and 620 nm and concomitantly transferring the oxygen back to its ground singlet state.[307] For this to happen the two beads need to be in close proximity (< 200 nm). For the KDM assay the donor bead is coated with streptavidin, which binds to the biotinylated peptide.[306] The acceptor bead is coated with protein-A, which binds an antibody of choice. When the antibody binds to the biotinylated peptide, the two beads are brought together in close proximity enabling a luminescent signal to be observed. KDM assays employ an antibody that recognises the lysine methylation mark being studied; the antibody can recognise either the substrate mark (resulting in a signal-loss assay) or the product mark (resulting in a signal-gain assay). As the assay progresses, the amount of the mark changes due to demethylation and the resultant change in emission can be used to measure peptide demethylation.

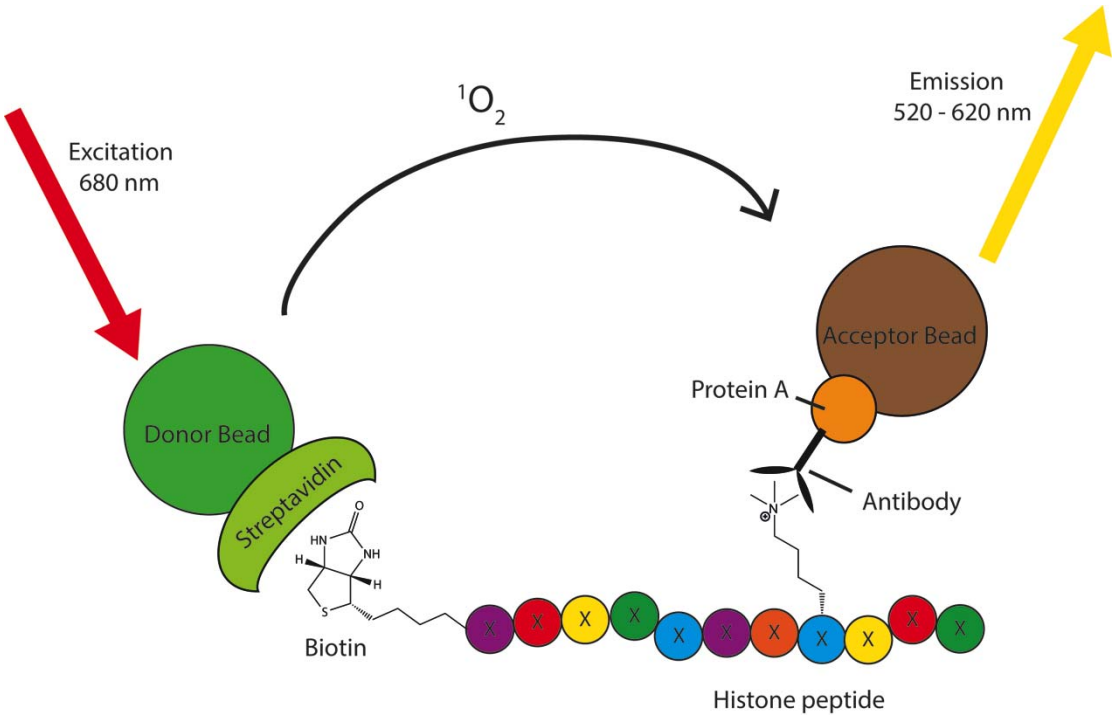


Figure 4.10 Setup of an ALPHA screen peptide turnover assay. On excitation, donor beads emit singlet oxygen, which is quenched by the acceptor bead if it is in sufficiently close proximity, leading to emission of light.

4.6.5.2 Assay development with PHF8

It was first necessary to determine which peptide would be suitable for the ALPHA screen. Both *N*- and *C*-terminally biotinylated peptides were tested in a MALDI MS assay. Activity was observed only for the *C*-terminally biotinylated peptide, not unsurprisingly given that PHF8 is known to interact with the *N*-terminus of H3 (data not shown).[176] Thus, H3 1-21 K4Me$_3$K9Me$_2$ *C*-terminal GGK-biotin was selected as the best peptide for the assay (GLS Biochem >95% pure).

There were two options for the assay design, either using an H3 K9Me$_2$ antibody or an H3 K9Me$_1$ antibody. Antibodies against both are commercially available, but whilst the H3 K9Me$_2$ antibody (ab1220, Abcam) had been tested and found to be effective at a 1:2000 dilution (*tested by Dr. Akane Kawamura*), the H3 K9Me$_1$ antibody (ab8896, Abcam) had not previously been evaluated. To evaluate selectivity,

four antibody dilutions were tested against increasing concentrations of 4 different peptides (H3 K9, H3 K9Me$_1$, H3 K9Me$_2$ and H3 K4Me$_3$K9Me$_2$). At higher concentrations the antibody was found to be selective for the dimethylated peptides and 1:2000 was selected as a suitable concentration for use in the assay (Figure 4.11).

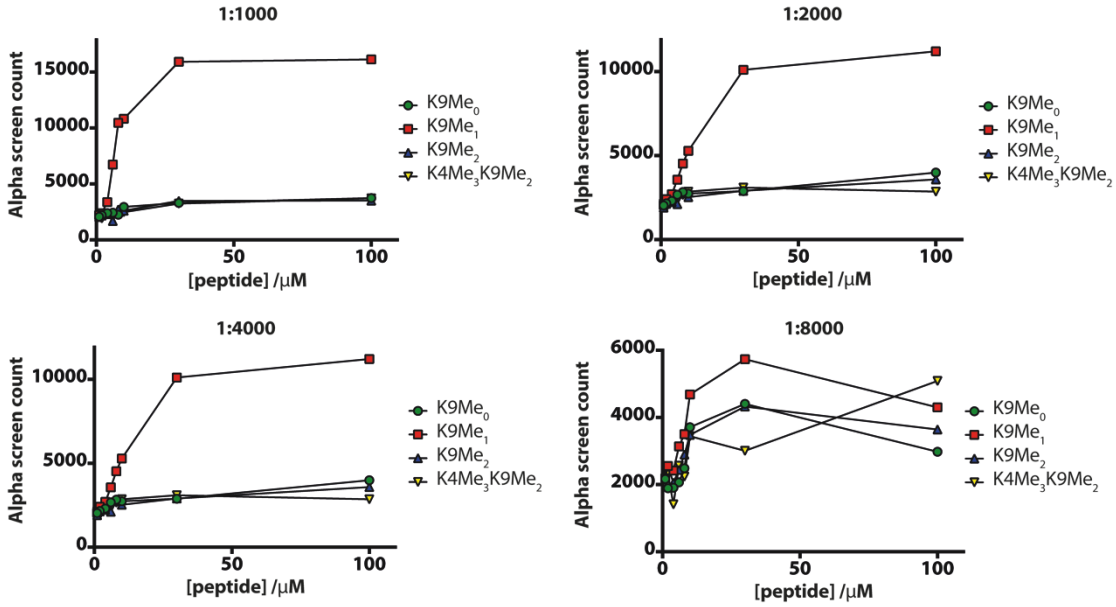


Figure 4.11 Evaluation of the selectivity of ab8896 (Abcam) for H3 K9Me$_1$ over K9Me$_{2/0}$ at four different concentrations by ALPHA screen.

Initially, pilot assays were performed using 6 concentrations of PHF8 and 300 nM peptide (Figure 4.12A, Section 8.7.3). Although a small gain in ALPHA screen signal was observed with increasing concentrations of enzyme in the assay, all values were extremely low and difficult to distinguish from background noise. In an attempt to rectify this, a number of assay components were varied, including peptide concentration / length / composition, protein concentration and assay length, but no increase in signal was observed (data not shown). Although high salt concentrations had previously been shown to improve enzyme activity in MALDI MS assays, it was decided to repeat the assay using a buffer without salt, in case this was having an inhibitory effect. Pleasingly, removal of salt from the buffer resulted in a significant signal being observed above background noise (Figure 4.12B).

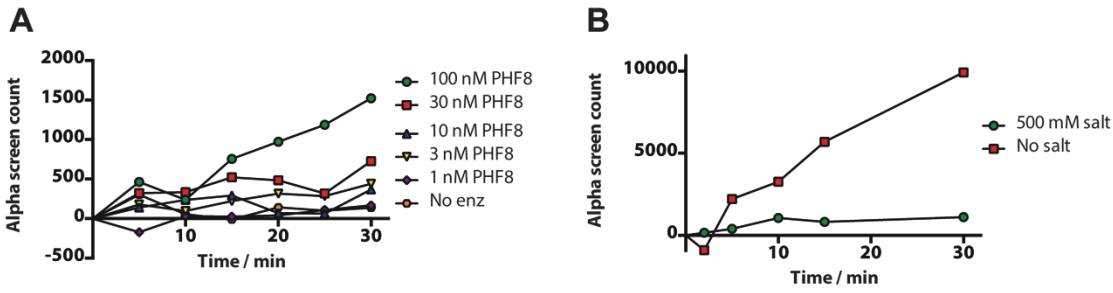


Figure 4.12 Pilot ALPHA screen assays using ab8896. **A)** Titration of PHF8 concentration in an assay containing H3 K4Me$_3$K9Me$_2$ (300 nM), 2OG (20 µM), Fe(II) (10 µM) and ascorbate (100 µM). **B)** Assay to test the effect of decreased salt concentration in the ALPHA screen buffer but otherwise using identical conditions to **A**. All data are shown with baseline subtraction.

Given the availability of antibodies for both a signal-gain assay (using ab8896 to measure the formation of the monomethylated H3 K4Me$_3$K9Me peptide, Figure 4.13) or a signal-loss assay (using ab1220 to measure the decrease in dimethylated H3 K4Me$_3$K9Me$_2$ peptide, Figure 4.13), both options were tested using two concentrations of enzyme (100 nM and 50 nM) and peptide (100 nM and 300 nM).

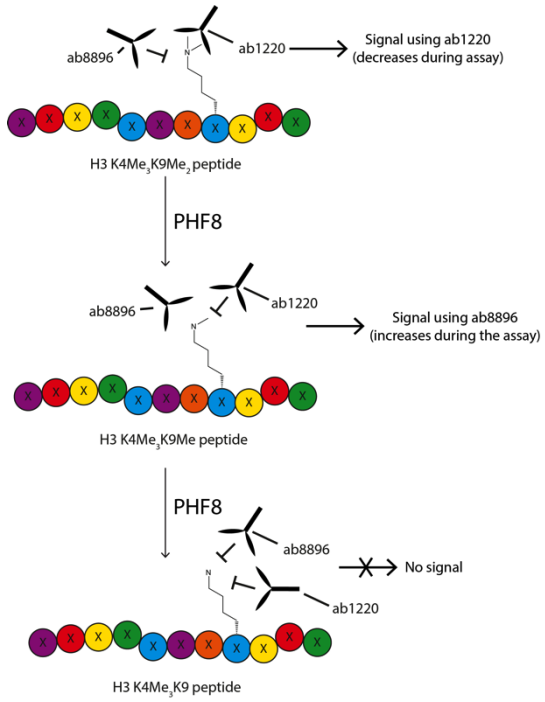


Figure 4.13 Outline of the principle of a signal-gain assay and a signal-loss assay.

Both assays showed demethylation relative to the no enzyme controls (Figure 4.14). Interestingly, for the signal-gain assay, a higher signal-to-noise ratio was observed with 50 nM PHF8 than for 100 nM PHF8 at both peptide concentrations (Figure 4.14A and B). It was suspected that this might be due to further demethylation of the monomethylated peptide to give the unmethylated peptide, which would no longer bind to the antibody and thus would not give a signal (Figure 4.13). For the signal-loss assay one experiment was carried out with biotinylated peptide (25 nM) and a second "spiking" assay with biotinylated peptide (25 nM) and a further non-biotinylated peptide (75 nM). No signal is produced from the non-biotinylated peptide in the assay. This allows the assay to have a higher overall peptide concentration (resulting in a higher peptide:protein ratio) without compromising the linear signal window, which would occur if more biotinylated peptide was included. Both reactions showed similar trends, with a decrease in signal over the course of the assay with increased PHF8 concentrations (Figure 4.14C and D), demonstrating that using a higher overall peptide concentration was viable. However, even with peptide spiking, the signal-loss assay still required concentrations of peptide similar to that of PHF8 to observe sufficient signal. In contrast, for the signal-gain assay it was possible to use higher concentrations of peptide relative to enzyme. It was therefore decided to pursue the signal-gain assay.

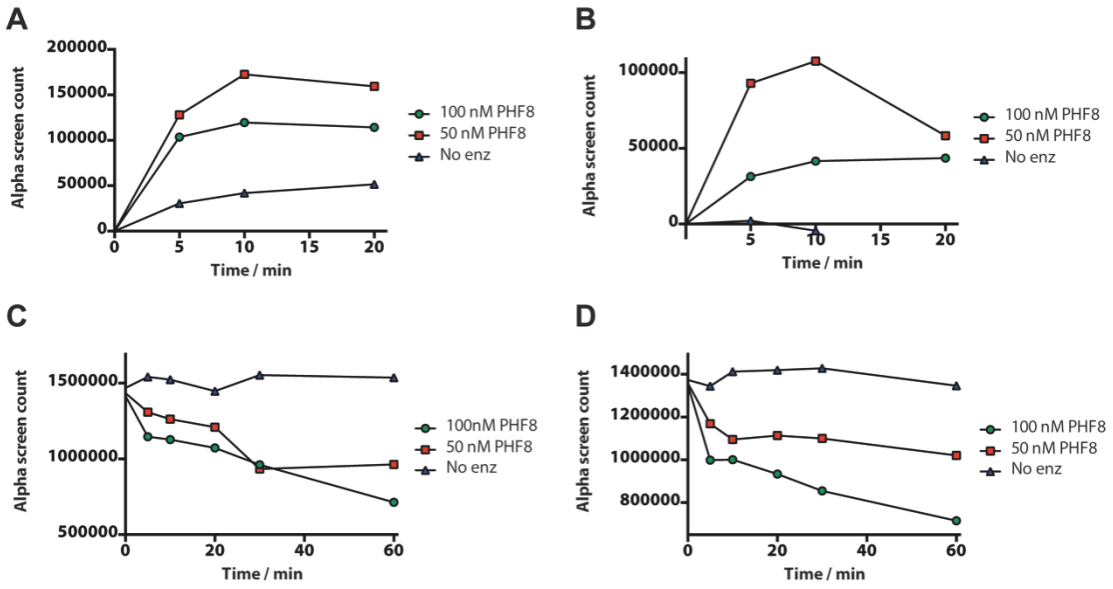


Figure 4.14 ALPHA screen comparing signal-gain (ab8896) and signal-loss (ab1220) assays. Signal-gain assays with **A)** H3 1-21 K4Me$_3$K9Me$_2$-Bio peptide (300 nM) or **B)** H3 1-21 K4Me$_3$K9Me$_2$-Bio peptide (100 nM). Signal loss assays with **C)** H3 1-21 K4Me$_3$K9Me$_2$-Bio peptide (25 nM) or **D)** H3 1-21 K4Me$_3$K9Me$_2$-Bio peptide (25 nM) and H3 1-21 K4Me$_3$K9Me$_2$ peptide (75 nM).

A PHF8 titration was carried out with peptide (at 300 or 500 nM) to determine (i) the optimal enzyme concentration for use in inhibitor assays and (ii) the linear range of enzyme activity (Figure 4.15). At high enzyme concentrations the phenomenon of decreased signal over time was observed again, highly suggestive of di-demethylation. Conditions for IC_{50} value measurements were selected (PHF8 (5 nM), peptide (300 nM), quenching after 10 min), which were in the linear range of the assay and provided sufficient signal-to-noise without using too much reagent, which is also an important consideration for an efficient high-throughput assay (Figure 4.15C).

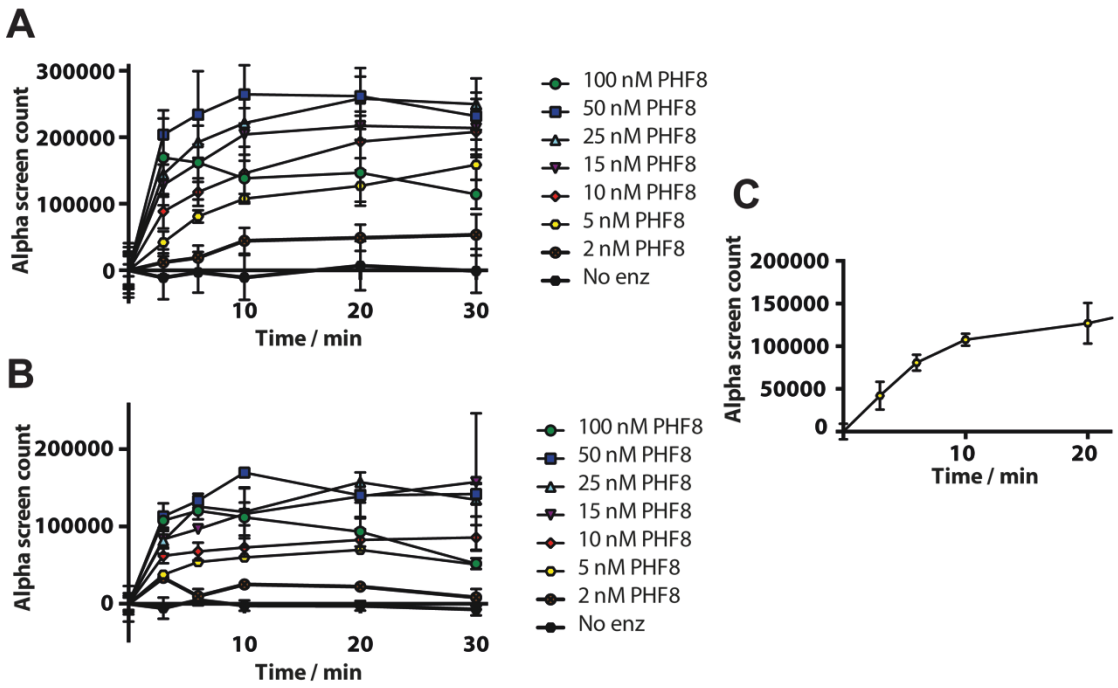


Figure 4.15 Titration of PHF8 concentration in a signal-gain ALPHA screen using ab8896, 2OG (20 μM), Fe(II) (10 μM) and ascorbate (100 μM). **A)** Using H3 1-21 K4Me$_3$K9Me$_2$-Bio peptide (300 nM). **B)** Using H3 1-21 K4Me$_3$K9Me$_2$-Bio peptide (500 nM). **C)** PHF8 (5 nM) with H3 1-21 K4Me$_3$K9Me$_2$-Bio peptide (300 nM). Points are plotted as the mean of duplicate experiments with error bars showing ± 1 s.d..

Using these conditions, determination of an IC_{50} value for the 2OG competitive inhibitor *N*-oxalylglycine (NOG) was attempted. Unfortunately, however, very low levels of luminescence signal were detected, suggesting that very little demethylation had occurred. Before starting the reaction, the enzyme mixture was pre-incubated with the inhibitor (NOG) for 15 min to allow equilibration. It was suspected that this pre-incubation time might be responsible for the loss of enzyme activity, which may also be further enhanced by the presence of 1% DMSO in the buffer. To investigate the effect of pre-incubation time on enzyme activity, PHF8 was incubated with Fe(II) (10 μM) and ascorbate (100 μM), but without inhibitor, at room temperature for variable lengths of time and its activity over a 20 min timecourse was measured. As proposed, activity was lost with increased pre-incubation time (Figure 4.16). As a pre-incubation period was considered essential to ensure that the reaction mixture had reached equilibrium even for slow-binding inhibitors, it was decided to increase the

enzyme concentration to 15 nM so that a measurable signal could be observed. Furthermore, pre-incubation was carried out without addition of Fe(II) to the reaction mixture, as this has been shown to help preserve enzyme activity (Dr Anthony Tumber, personal communication).

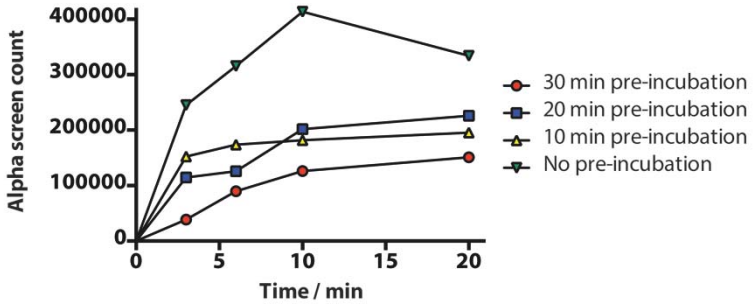


Figure 4.16 Effect of pre-incubation on PHF8 activity in an ALPHA screen.

A pilot IC_{50} assay was carried out with 2,4-pyridine dicarboxylic acid (2,4-PDCA), a known broad-spectrum inhibitor of 2OG oxygenases.[308] PHF8 (15 nM) was incubated with increasing concentrations of 2,4-PDCA (3 nM to 100 μM) for 15 min before the addition of peptide (300 nM) and cofactors. The reaction was quenched after 10 min, at which point antibody (ab8896), donor and acceptor beads were added, and the plate measured after 1 hour using an Envision plate reader (Perkin Elmer). Percentage inhibition was determined relative to a control reaction containing no inhibitor. By plotting this value against the log of the inhibitor concentration, an IC_{50} value was determined. For 2,4-PDCA an IC_{50} value of 270 μM was measured, in good agreement with the value determined using the MALDI MS assay (IC_{50}(MALDI) = 310 μM, *measured by Timothy Rooney*).

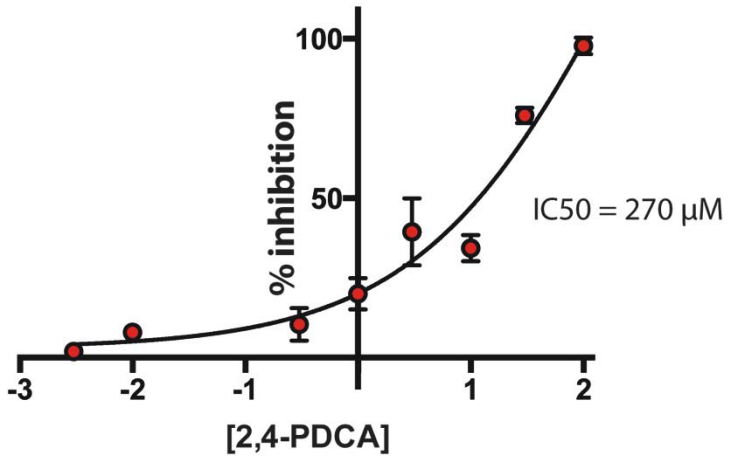
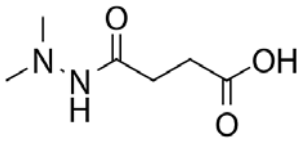


Figure 4.17 IC_{50} value for 2,4-PDCA against PHF8 (15 nM) using the ALPHA screen assay. Points are plotted as the mean of triplicate experiments and error bars give ± 1 s.d..

4.7 Inhibitor screening using the ALPHA screen

4.7.1 Daminozide is a specific inhibitor for the KDM2/7 subfamilies of JmjC demethylases

A screen of 2OG analogues identified Daminozide (*N*-(dimethylamino)succinamic acid) as an inhibitor of FBXL11, a KDM2 subfamily member (Figure 4.18). Daminozide, also commonly known as Alar, is a plant growth retardant that was used extensively in the 1960s to control plant size and fruit ripening.[309] Concern over its possible carcinogenicity, as well as that of its metabolite, 1,1-dimethylhydrazine, led to its withdrawal from use in food crops, although it is still used to control the growth of ornamental plants.[310,311] In plants Daminozide is thought to act by inhibition of plant 2OG oxygenases and other related oxidases involved in gibberellin and ethylene synthesis.[312,313]



Daminozide
N-(dimethylamino)succinamic acid

Figure 4.18 Structure of Daminozide.

As the KDM7 subfamily is closely related to the KDM2 subfamily,[166] inhibition assays were carried out with PHF8 and KIAA1718 to see whether they are also inhibited by Daminozide. PHF8 inhibition was determined using the ALPHA screen, whilst, the MALDI-TOF MS assay was used for KIAA1718, giving IC_{50} values of 0.55 μM and 2.2 μM respectively, and confirming that Daminozide also inhibited the KDM7 subfamily of KDMs (Figure 4.19). Direct comparison of IC_{50} values between different assay types is not possible and it is likely that the higher value measured with KIAA1718 was partially due to the higher concentration of enzyme used within the assay.

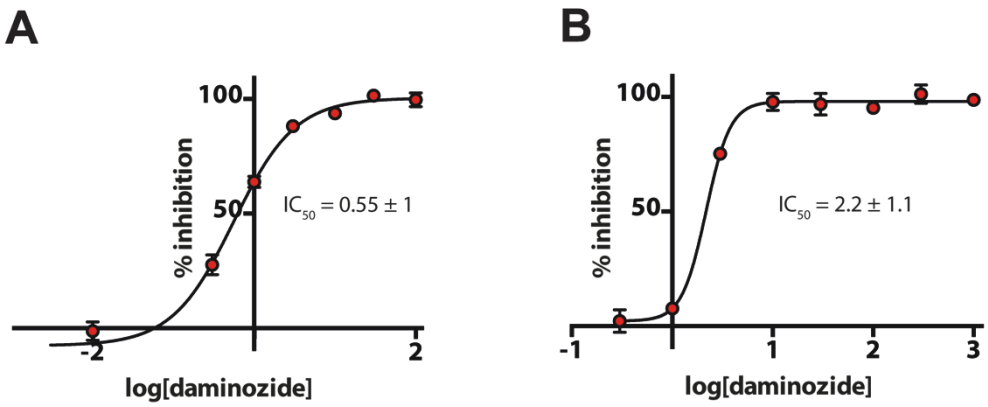


Figure 4.19 IC_{50} determination for Daminozide against **A)** PHF8 (15 nM) and **B)** KIAA1718 (2 μM) measured by ALPHA screen and MALDI-TOF MS respectively. Points are plotted as the mean of triplicate experiments and error bars give ± 1 s.d..

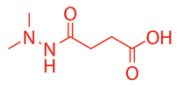
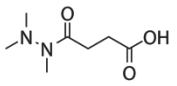
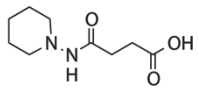
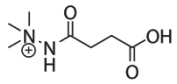
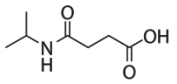
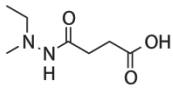
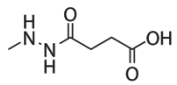
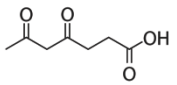
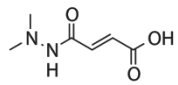
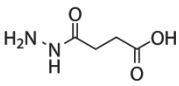
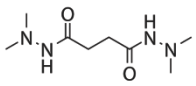
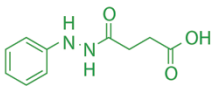
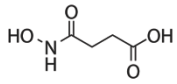
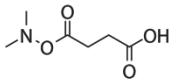
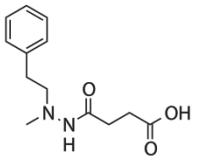
Screening of representative members from each of the other KDM subfamilies and several other 2OG oxygenases resulted in IC_{50} values of more than 100 μM for Daminozide (Table 4.4, *screening of other 2OG oxygenases carried out by Dr Anthony Tumber, SGC*).[314] Daminozide therefore appeared to be selective for the closely related KDM2 and KDM7 subfamilies. The high selectivity of Daminozide for these subfamilies was particularly interesting given that Daminozide has a low molecular weight (160 Da) and is achiral.

Table 4.4 IC_{50} values for Daminozide against a representative panel of 2OG oxygenases.

	FBXL11	PHF8	KIAA1718	JMJD1A	JMJD2E	JMJD3	JARID1B	FIH	PHD2	BBOX1
IC_{50} / μM	1.5	0.55	2.2	>100	>100	>100	>100	>100	>100	>100

A set of Daminozide analogues was synthesised in an attempt to increase the compound potency (*compounds synthesised by Dr Esther Woon*). These analogues were screened using the ALPHA screen against a panel of KDMs including PHF8. Data for PHF8 is shown in Table 4.5. Unfortunately, despite the extensive synthesis efforts it was not possible to find an analogue of Daminozide with increased potency for the KDM2 / 7 subfamilies, whilst retaining selectivity. Although compound **4** was slightly more potent against PHF8 than Daminozide (compound **1)**, it also inhibited JMJD2E and JARID1C with IC_{50} values of 0.2 μM and 0.48 μM respectively. Compound **6** remained selective for KDM2 / 7 but was less potent than the parent compound. Interestingly, it appeared that the space around the active site of PHF8 was larger than in FBXL11, as compounds **11** – **15** all inhibited PHF8 selectively over FBXL11. In particular, compound **14** was found to inhibit PHF8 with an IC_{50} value (1.1 μM) similar to that of Daminozide (0.55 μM), but with 10-fold less potency against FBXL11, which has an IC_{50} value of 19 μM. This result suggested that further selectivity between the closely related KDM2 and KDM7 families might be achievable.

Table 4.5 Table of IC_{50} values for Daminozide analogues against PHF8 (15 nM), measured using the ALPHA screen assay. Daminozide (**1**) is shown in red, a compound (**14**) selective for PHF8 over FBXL11 is shown in green.

	Inhibitor structure	IC_{50} /μM		Inhibitor structure	IC_{50}		Inhibitor structure	IC_{50}
1		0.55	6		11.5	11		5.5
2		35	7		>100	12		12.9
3		3.4	8		2.3	13		19
4		0.48	9		6.9	14		1.1
5		1.7	10		117	15		7.1

4.7.2 Daminozide structural biology

In order to investigate the structural basis for selectivity of Daminozide for the KDM2 / 7 subfamilies, Daminozide was modelled into the active site of a crystal structure of PHF8 by manual docking (Figure 4.20 based on PDB ID 3K3O,[315] *docking carried out by Dr Nathan Rose*). This model suggested that the dimethyl groups of Daminozide might be accommodated by a tight hydrophobic pocket adjacent to the Fe(II)-binding site in the active site of PHF8. The chemical composition of this pocket is likely similar in the case of FBXL11, but more polar in other KDM subfamilies, which mostly have a seryl- or threonyl-residue in place of V255 in PHF8 (shown in pink in Figure 4.20, V220 in FBXL11).

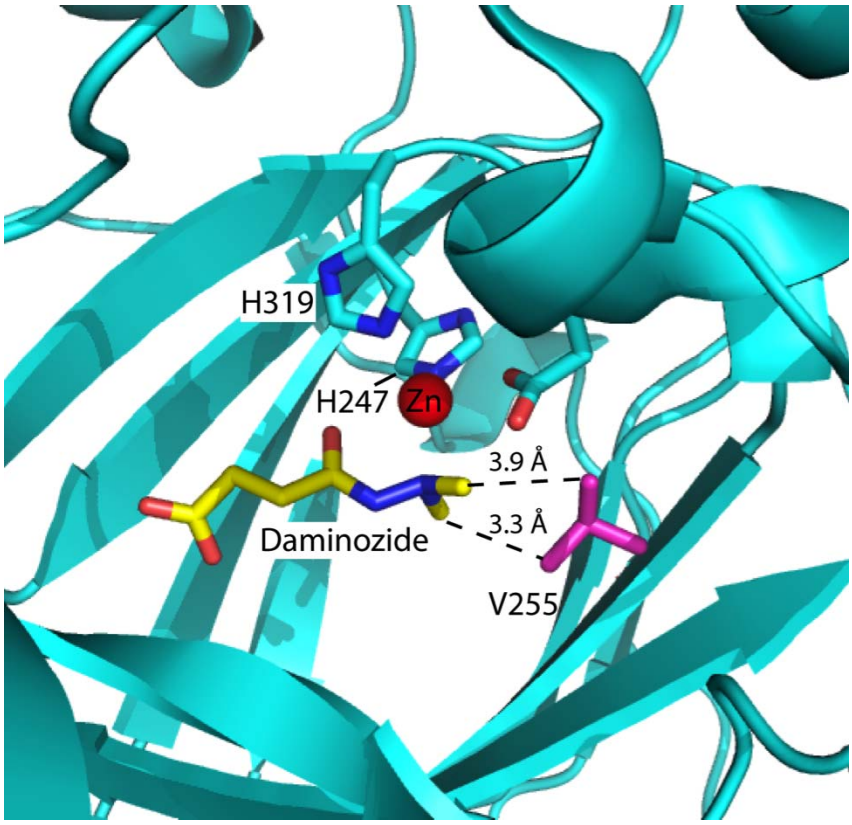


Figure 4.20 View from a model of Daminozide docked into the active site of PHF8 (*manual docking performed by Dr Nathan Rose*).

To determine whether the predicted binding mode was a factor in determining the selectivity of Daminozide for PHF8, site-directed mutagenesis was used to produce two variants of FL PHF8: V255S and V255T (Section 8.6.2). Both variants were successfully expressed in, and purified from *E. coli* cells. Their activity was initially assayed by MALDI-MS, which indicated that the V255T variant was inactive. The V255S variant also showed reduced activity relative to wt PHF8, making IC_{50} determination difficult. Nevertheless complete inhibition of activity was observed in the presence of 20 μM Daminozide, suggesting that the IC_{50} value was considerably lower than this concentration. It was thus unlikely that inhibition levels had been reduced to those found for other KDM subfamilies (all other KDM subfamily IC_{50} values > 100 μM) containing a seryl-residue at this position. As such, the selectivity of Daminozide for KDM2 / 7 cannot be explained by this amino acid substitution.

4.7.3 PHF8 crystallography

Although docking can provide some useful information on the potential binding orientations of inhibitors, the large number of possible binding conformations makes reliable structural predictions difficult. Where possible, biophysical analyses, such as employing co-crystallisation of the enzyme and inhibitor, are desirable. To this end, co-crystallisation of PHF8 and Daminozide was attempted, in order to investigate the reasons for the selectivity of Daminozide for the KDM2 / 7 subfamilies. Although structures of the catalytic core of PHF8 have been solved both with and without the PHD finger (PDB ID 3KV4, 3K3N, 3K3O and 2WWU),[176,297,315] the only FL PHF8 crystal structure includes a histone peptide, which may add rigidity to the structure. Consequently, it was decided to focus on crystallisation of the JmjC domain alone. Unfortunately no co-crystals of PHF8 and Daminozide were obtained (crystallography attempts are described in Appendix 3)

Whilst in-house attempts were being carried out, a co-crystal structure of PHF8 bound to Daminozide was obtained by collaborators at the SGC. The crystal structure was solved to 2.55 Å. As anticipated, Daminozide binds in the enzyme's 2OG binding pocket and coordinates the catalytic Fe(II) centre (Zn(II) in the crystal structure) in a bidentate manner (Figure 4.21). Co-crystallography with compound **14** (Table 4.5) was also attempted at the SGC but was unsuccessful.

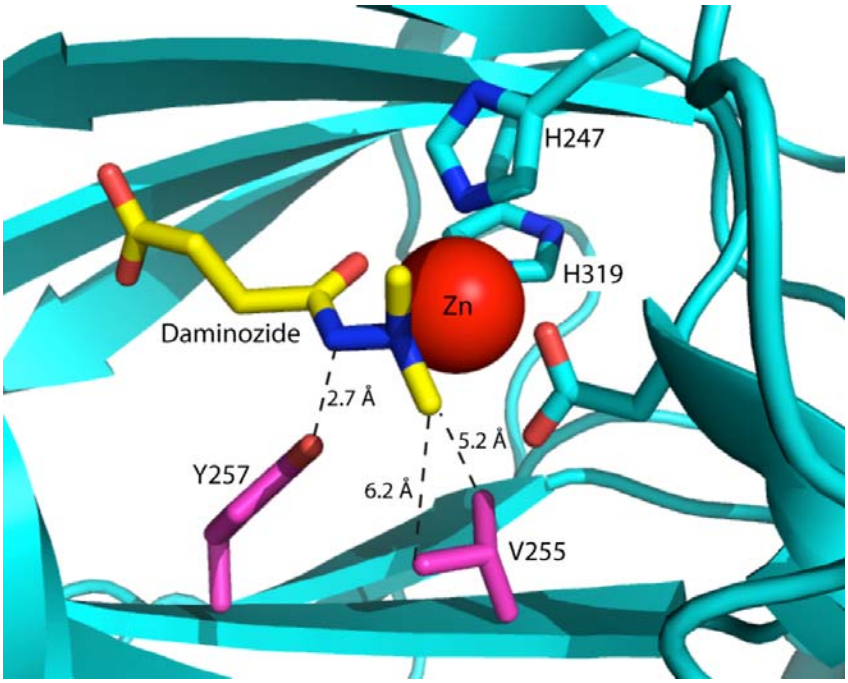


Figure 4.21 View from a crystal structure of the JmjC domain of PHF8 bound to Daminozide (PDB ID 4DO0)[314]. PHF8 is shown in turquoise, Zn in red and Daminozide in yellow. Residues studied by site directed mutagenesis are shown as sticks in pink.

As predicted, the crystal structure revealed Daminozide to coordinate to the Fe(II) centre (substituted for Zn(II) in this crystal structure) in a bidentate manner *via* its dimethylamino groups and acylhydrazide carbonyl. However, the orientation of the compound was revealed to differ from that predicted by docking (Figure 4.20), with the dimethylamino group of Daminozide binding to the zinc *trans* to H319 rather than to H247 (Figure 4.21). Interestingly, the co-crystal structure suggested that the valyl-residue (V255) predicted to be responsible for subfamily selectivity was located more than 5 Å from the methyl groups of Daminozide. Thus, it is unlikely to be a key determining feature in binding, consistent with the high degree of Daminozide inhibition observed with the recombinant PHF8 V255S variant (Section 4.7.2). Instead, a tyrosyl-residue (Y257, pink in Figure 4.21), which is not present in the other KDM subfamilies (replaced by, for example, N290 in JMJD2A), is located only 2.7 Å from the NH of the hydrazide, potentially forming favourable interactions.

4.7.4 Inhibition by Daminozide in cells

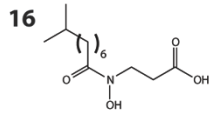
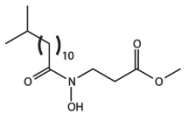
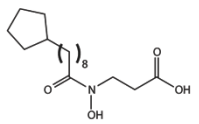
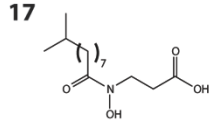
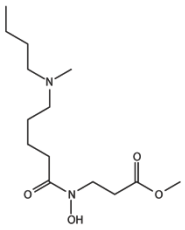
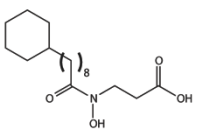
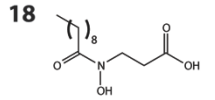
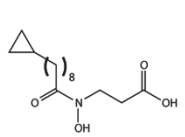
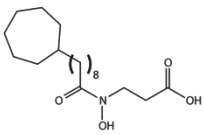
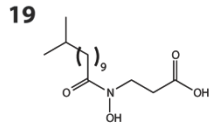
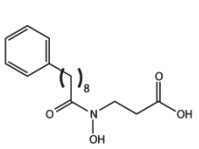
Inhibition of FBXL11 by Daminozide in cells was studied using an immunofluorescence assay looking at the methylation levels of H3 K36 (*carried out by Dr. Shirley Li*). Cells were treated with increasing concentrations of Daminozide, but unfortunately no changes in global methylation levels were observed. This was thought to be due to the acid moiety of Daminozide preventing cellular uptake across the lipid bilayer that forms the cell membrane. To circumvent this, acid groups can be masked as an ester to produce a non-charged prodrug, which is hydrolysed by esterases within the cell having crossed the lipid membrane.[316] Although a number of ester derivatives were synthesised, none were capable of causing inhibition within cells, either due to not being cell permeable or not being cleaved after entering the cell. It should be noted, however, that Daminozide was used as a plant growth regulator, and thus likely gets into plant cells.

This work is summarised in "Plant growth regulator daminozide is a selective inhibitor of human KDM2/7 histone demethylases" (Appendix 4).

4.7.5 Inhibition of PHF8 by hydroxamate based compounds

As described above, Daminozide is thought to be carcinogenic due to its breakdown product, 1,1-dimethylhydrazine. To combat this, collaborators in the group of Professor Naoki Miyata synthesised a series of hydroxamate inhibitors based upon the structure of Daminozide but without the 1,1-dimethylhydrazine substructure. These inhibitors were screened against PHF8 using the ALPHA screen (Section 8.7.3); IC_{50} values are given in Table 4.6.

Table 4.6 IC_{50} values for inhibition of PHF8 by hydroxyamate analogues measured by ALPHA screen.

Inhibitor structure	IC_{50} /μM	Inhibitor structure	IC_{50}	Inhibitor structure	IC_{50}
16 	3.3	**20** 	8.3	**23** 	9.7
17 	8.8	**21** 	3.2	**24** 	5.3
18 	4.6	**22** 	7.2	**25** 	3.8
19 	11.7			**26** 	5.9

These compounds were also screened against the panel of KDM subfamilies to investigate their selectivity. Compound **22** was shown to inhibit FBXL11 at similar levels to PHF8 (IC_{50} = 6.8 μM *vs* 7.2 μM), but showed more than seven-fold selectivity for these enzymes over JARID1A (IC_{50} = 55 μM), which was a member of the next most strongly inhibited subfamily. To check whether compound **22** was also an inhibitor of KIAA1718, inhibition was measured by MALDI-TOF MS giving an IC_{50} of 0.2 μM. Although, like Daminozide, these compounds still have an acid moiety, it was hoped that the lipophilicity of the long alkyl chains might help the compounds cross the cell membrane. The activity of compound **22** was tested in a cell-based assay (*carried out by collaborators in the group of Professor Naoki Miyata*). Western blots were used to analyse the levels of H3 K27Me_2 on treatment of cells with the inhibitor. Pleasingly, a dose-dependent increase in H3 K27Me_2 levels was observed on treatment of N2a cells with compound **22**, consistent with inhibition

of KIAA1718 demethylase activity.

This work is summarised in "Identification of the KDM2/7 histone lysine demethylase subfamily inhibitor and its antiproliferative activity" (Appendix 4)

4.7.6 Autohydroxylation by PHF8

In the absence of prime substrate, 2OG oxygenases can still turnover 2OG.[158] Whilst this sometimes occurs in an un-coupled manner, there are instances where this leads to autohydroxylation of aromatic residues close to the active site (Figure 4.22).[317–319]

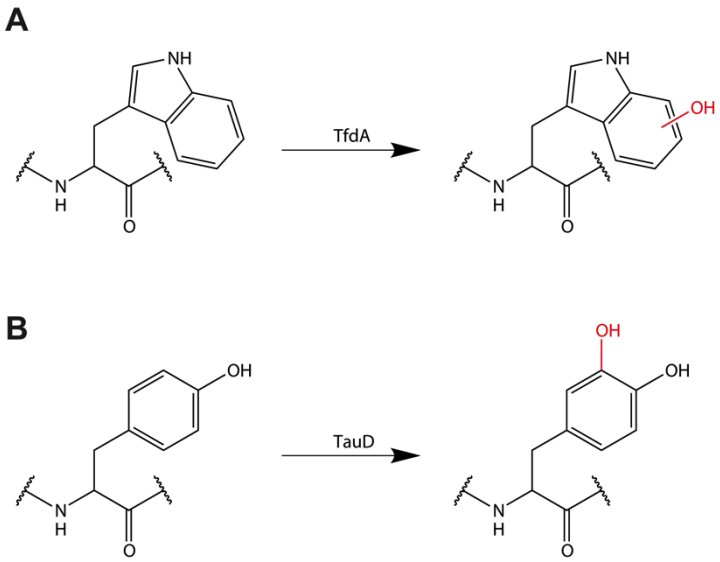


Figure 4.22 Possible products of auto-hydroxylation of aromatic residues by 2OG oxygenases. **A)** Hydroxylation of tryptophan by TfdA. **B)** Hydroxylation of tyrosine by TauD. Added hydroxyl-groups are shown in red.

The location of Y257 close to the catalytic Fe(II) centre (Figure 4.21) suggested that it might be a site of autohydroxylation for PHF8. Amino acid analysis was used to test whether autohydroxylation of PHF8 occurs *in vitro*. PHF8 (40 μM) was incubated with or without 2OG (100 μM) for 1 hour at 37 °C and quenched with one equivalent of methanol. The assay samples were then hydrolysed with 6 M HCl overnight to their constituent amino acids, and amino acid analysis performed. A

dihydroxytyrosine standard was included for comparison (*samples run by Dr Nikita Loik*). No evidence for autohydroxylation of PHF8 was obtained, as demonstrated by no appearance of any modified amino acids in the samples and in particular, no peak co-eluting with the dihydroxytyrosine standard (data not shown).

4.8 Exploring the substrate specificity of PHF8

Human 2OG oxygenases catalyse hydroxylations and *N*-methyl demethylations *via* hydroxylation. In addition to KDMs, members of the JmjC subfamilies include enzymes that catalyse hydroxylation reactions, such as the hypoxia inducible factor (HIF) hydroxylase FIH (Table 1.2, Chapter 1). To date, protein demethylation by KDMs has only been reported to occur on *N*-methyl groups, and it was therefore of interest to see whether PHF8 was able to react with any alkyl groups other than N^{ε}-methyllysine in the context of a histone peptide. A number of *N*-alkylated lysine analogues *were synthesised by Dr Richard Hopkinson* with *N*-alkyl groups including isopropyl or ethyl groups. The sensitivity of PHF8 to the length of the amino acid side chain was also investigated using ornithyl- as well as lysyl-residues. The lysine analogues investigated are shown in Figure 4.23. These analogues were inserted into the K9 position of the PHF8 peptide substrate, H3 1-15 $K4Me_3K9Me_2$ ($ARTKMe_3QTARXSTGGKA$, where X denotes the lysine/ornithine analogue, *synthesised by SPPS by Dr. Richard Hopkinson*, Section 8.11).

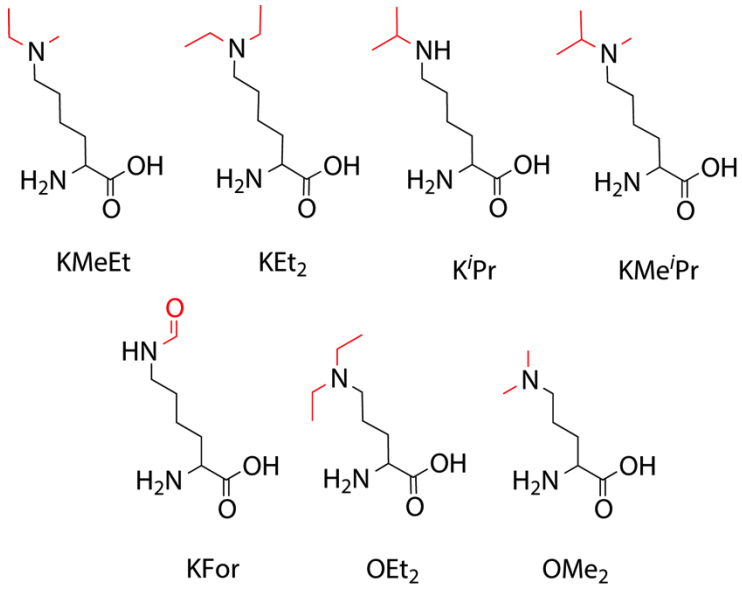


Figure 4.23 Lysine analogues inserted into position X (at the K9 position) of an H3 1-15 $K4Me_3K9Me_2$ peptide sequence, ART-KMe_3-QTAR-X-STGGKA.

All of the lysine analogues contained C-H bonds alpha to the side-chain nitrogen, allowing for potential abstraction by PHF8 in an analogous fashion to the demethylation reaction with dimethyllysine. Thus, it was anticipated that they might undergo similar oxidation reactions. The reaction of PHF8 with each of these peptides was tested using the MALDI MS assay under standard demethylation conditions (Section 8.7.1). Peptide (10 μM) and FL PHF8 (2 μM) were incubated together for 1 hr at 37 °C, then quenched with methanol and analysed. Interestingly, PHF8 was found to react with H3 1-15 peptides containing KMeEt, K^{i}Pr and OMe_2 (Figure 4.24). For the KMeEt-containing peptide, complete loss of both the methyl and ethyl groups to form the unalkylated peptide was observed after 1 hr (Figure 4.24A). For the K^{i}Pr-containing peptide, deisopropylation to the unalkylated peptide occurred (Figure 4.24B). This reaction was less efficient than the reaction with the KMeEt residue, resulting in around 50% deisopropylation after 1 hr. For the OMe_2 containing peptide, approximately 40% monodemethylation to the monomethylornithine peptide was observed, but there was no complete demethylation to the unalkylated product

(Figure 4.24C). In all cases, the extent of dealkylation was substantial, although for the peptides containing K^{i}Pr and OMe$_2$ the reaction was less efficient than in the control reaction with the wild type KMe$_2$ peptide, which like the KMeEt peptide was fully demethylated (data not shown). These results demonstrated that 2OG oxygenases can catalyse dealkylations other than demethylation and have some flexibility in positioning the amino acid side chain in the correct position for demethylation, allowing demethylation of the shorter dimethylated ornithyl-residue.

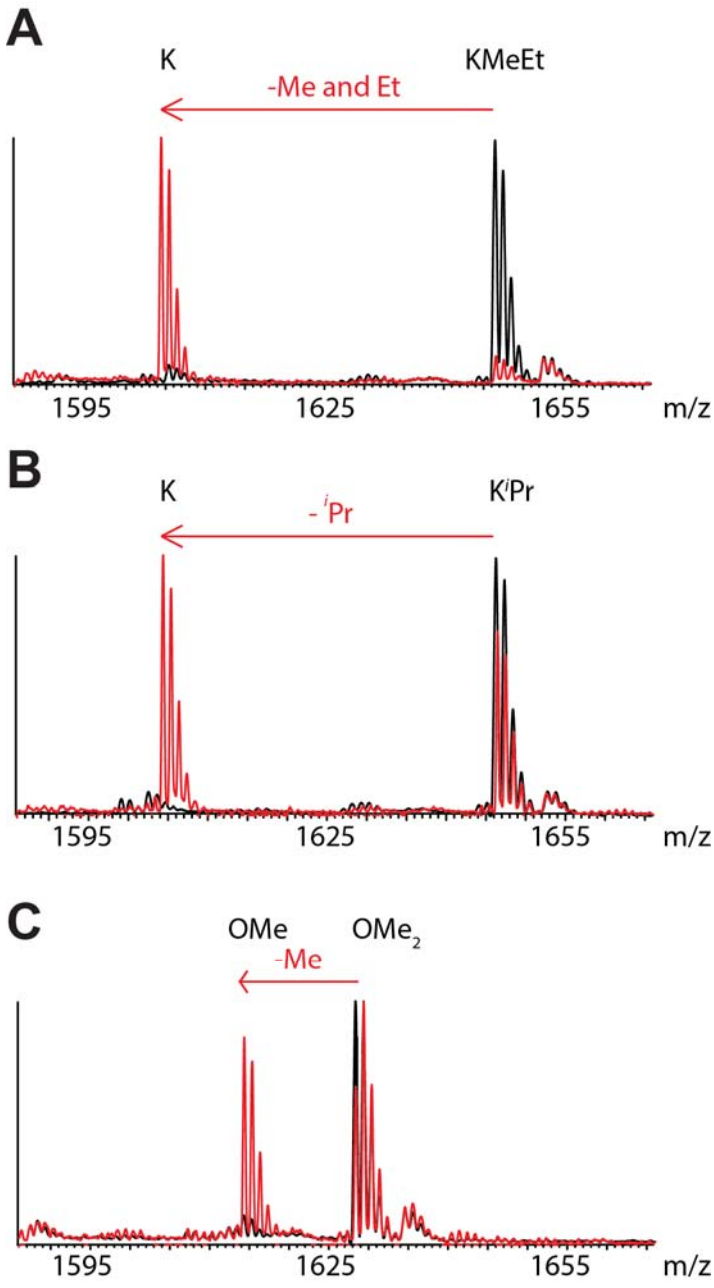


Figure 4.24 MALDI MS spectra showing the reaction of PHF8 (2 μM) with **A)** H3 1-15 K4Me$_3$K9MeEt, **B)** H3 1-15 K4Me$_3$K9^{i}Pr and **C)** H3 1-15 K4Me$_3$K9OMe$_2$. Control spectra are shown in black and enzyme reactions in red.

The reaction with the KMeEt-containing peptide was observed to go to completion over 1 hour (Figure 4.24A). A timecourse reaction was carried out to investigate whether the initial reaction was demethylation or deethylation or whether both occur simultaneously. The timecourse revealed initial accumulation of some

monoethyllysine peptide, but very little of the monomethyllysine peptide, suggesting that the peptide bound in the active site in a manner that favours initial demethylation (Figure 4.25A). This contrasted with the reaction of the same peptide with JMJD2E (*carried out by Dr Richard Hopkinson*) where initial deethylation appeared to be favoured (Figure 4.25B).

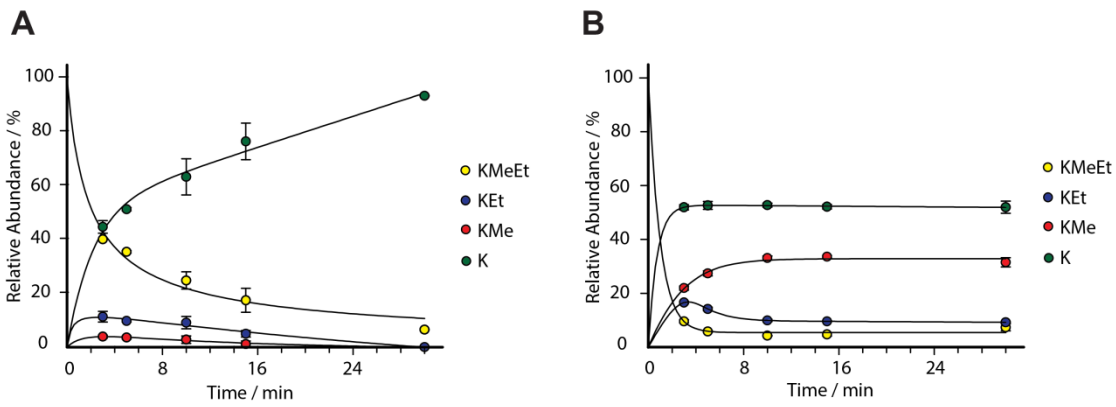


Figure 4.25 Timecourse reactions of **A)** PHF8 (2 μM) and **B)** JMJD2E (1 μM) with the H3 1-15 K4Me$_3$K9MeEt peptide (10 μM).

If deethylation occurred in the same manner as demethylation, with hydroxylation of the carbon alpha to the side-chain nitrogen, then production of acetaldehyde would be expected (Figure 4.26A). However, it is also possible to envisage a mechanism by which ethylene or ethylene oxide is lost, if the radical abstraction occurred on the carbon beta to the nitrogen. Analogous to FDH, there is an enzyme, aldehyde dehydrogenase (ADH) that is able to convert acetaldehyde to acetic acid, concomitantly reducing the cofactor NAD^+ to NADH. By coupling PHF8-catalysed deethylation to acetaldehyde oxidation by ADH, the production of acetalydehyde could be measured indirectly by quantifying the increase in fluorescent signal emitted due to the production of NADH in solution. Attempts were made to observe acetaldehyde production by this method. Initially a calibration of the fluorescence

levels observed at differing concentrations of acetaldehyde was carried out. Unfortunately, however, reaction of ADH with NAD^+ to produce NADH appeared to occur independently of the concentration of acetaldehyde present. As a result, the reaction could not be used to measure PHF8 activity. Instead, acetaldehyde production was demonstrated for the reaction with JMJD2E, using ^1H NMR spectroscopy (*experiment performed by Dr Richard Hopkinson*), where the appearance of two new peaks at δ_H 1.28 ppm and 2.19 ppm could be observed (Figure 4.26B). These peaks overlapped with those from a sample of authentic acetaldehyde that was added to the reaction mixture, helping to confirm their identity.

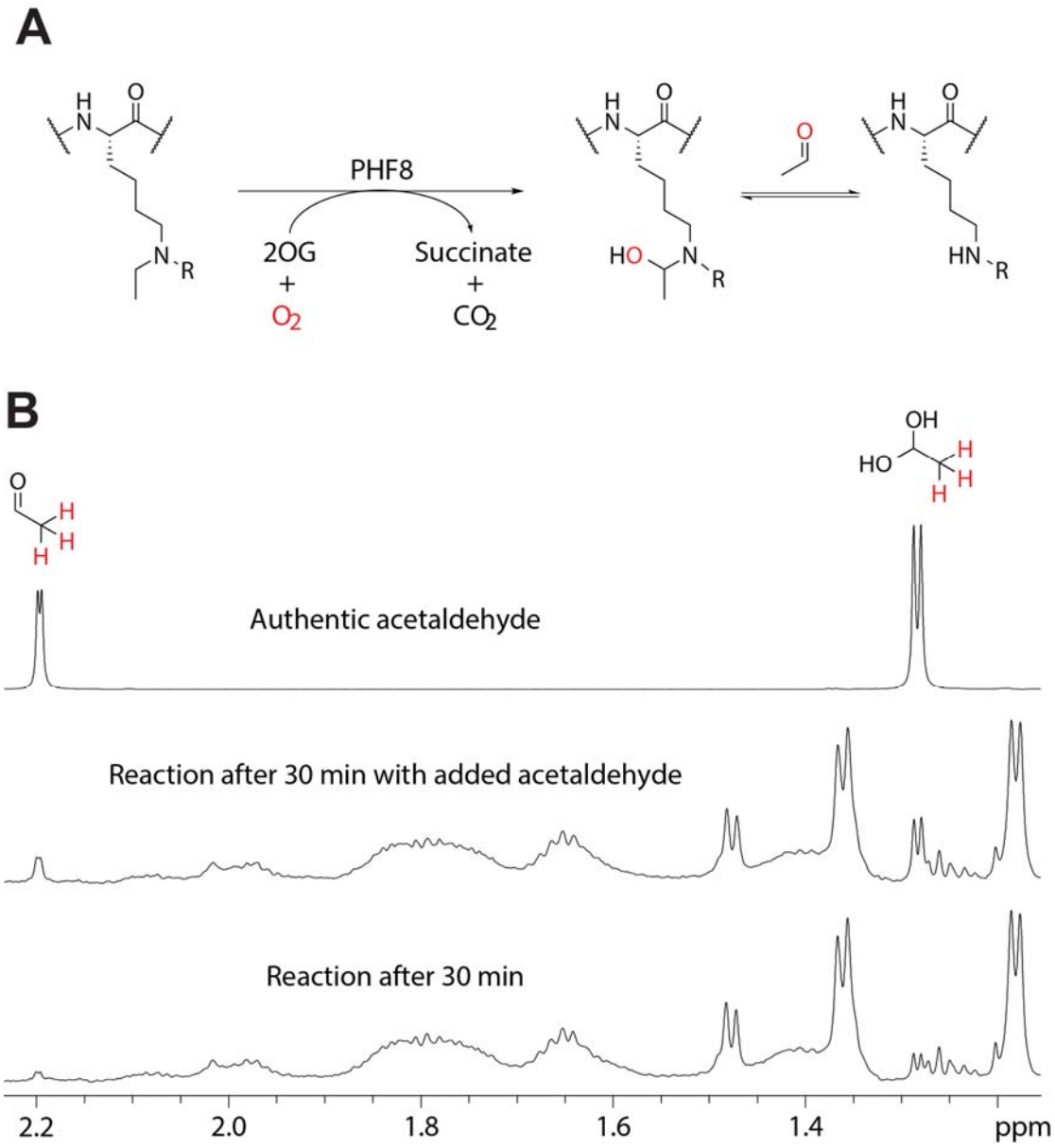


Figure 4.26 A) Scheme showing proposed deethylation reaction. Oxygen atoms involved in the oxidation are shown in red. **B)** ^{1}H NMR spectra showing the reaction of JMJD2E with a H3 1-15 K4Me$_3$K9MeEt peptide. Protons responsible for NMR peaks are shown in red. Production of acetaldehyde can be observed. The spectrum of authentic acetaldehyde is shown at the top of the panel.

For the reaction of PHF8 with the K^{i}Pr-containing peptide, deisopropylation was proposed to occur with production of acetone (Figure 4.27A). This was confirmed for JMJD2E by ^{1}H NMR spectroscopy (*performed by Dr Richard Hopkinson*) where the appearance of a singlet at δ_H 2.17 ppm, which overlapped with the peak from a sample of authentic acetone, could be observed (Figure 4.27B).

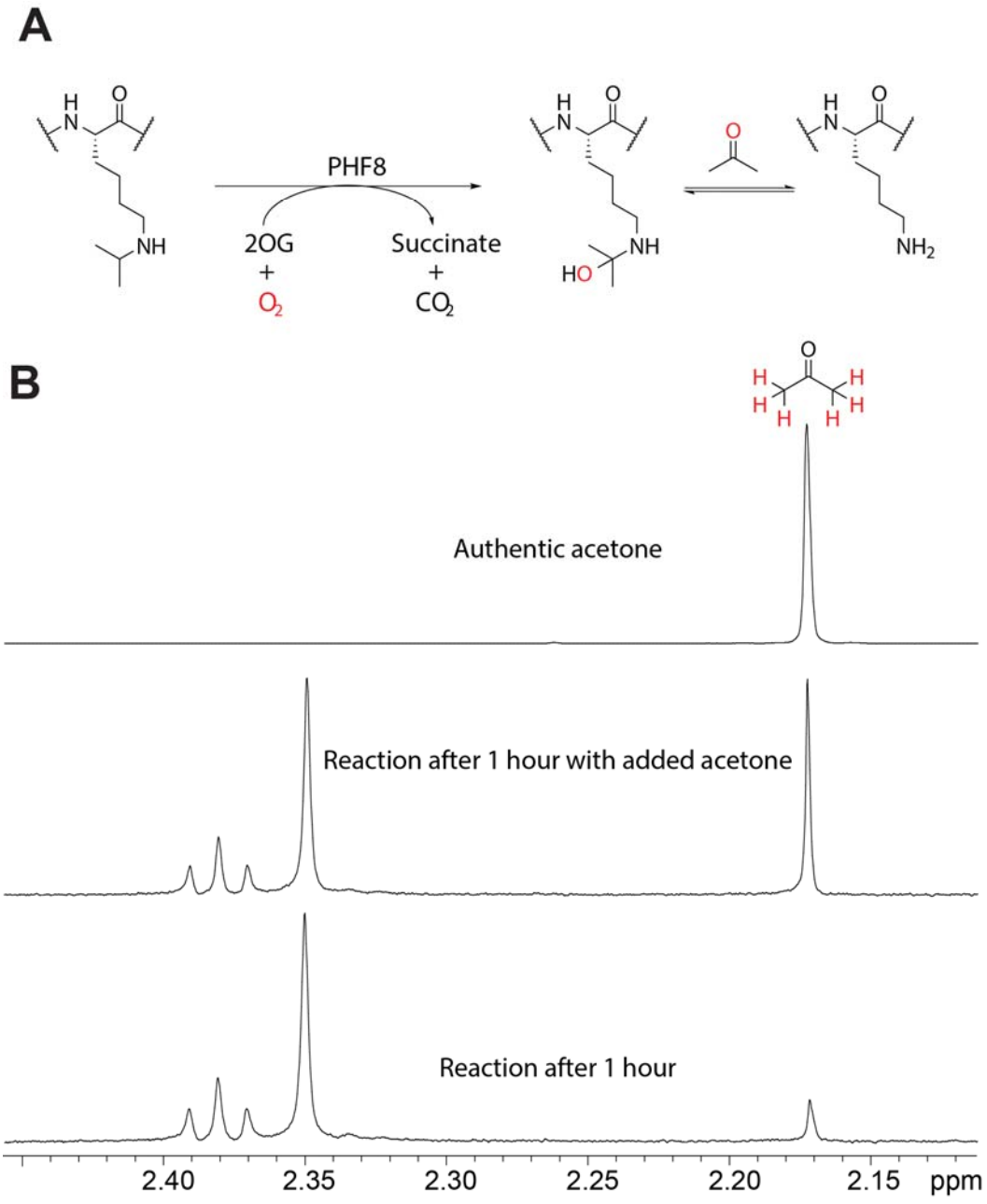


Figure 4.27 A) Scheme showing proposed deisopropylation reaction. Oxygen atoms involved in the oxidation are shown in red. **B)** ^{1}H NMR spectra showing the reaction of JMJD2E with a H3 1-15 K4Me$_3$K9^{i}Pr peptide. Protons responsible for the NMR peaks are shown in red. Production of acetone can be observed. The spectrum of authentic acetone is shown at the top of the panel.

For the reaction with the OMe$_2$-containing peptide, an increased concentration of PHF8 was used to see whether the reaction could be pushed to the point where removal of both methyl groups was observed. Despite using 6 μM PHF8 and obtaining around 80% monodemethylation to the monomethylornithine peptide, no fully dealkylated ornithine peptide product was observed (Figure 4.28).

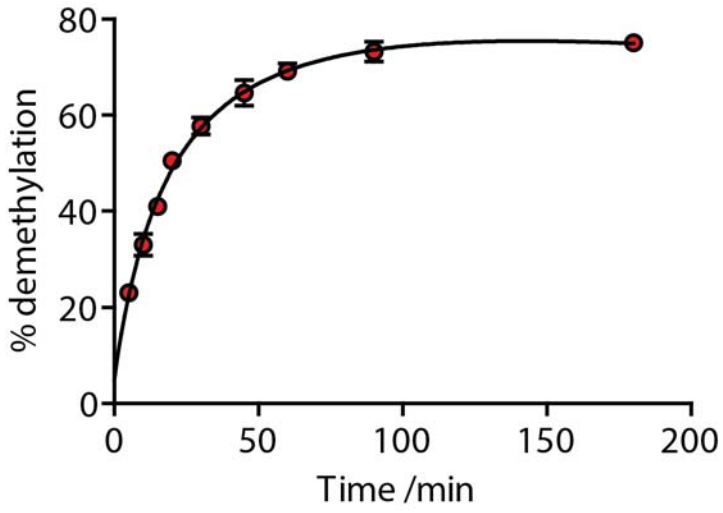


Figure 4.28 MALDI MS timecourse assay measuring the amount of monodemethyation of an H3 1-15 K4Me_3K9OMe_2 peptide (10 μM) by PHF8 (6 μM).

Interestingly, although demethylation could be observed when PHF8 was incubated with the dimethylornithine peptide, no reaction was observed for the same reaction with JMJD2E. This was consistent with crystallographic data for the two enzymes; when bound to PHF8, histone peptides penetrate deep into the catalytic site (yellow peptide in Figure 4.29), allowing scope for shortening of the lysine side chain (to ornithine) whilst still maintaining the alkyl group sufficiently close to the Fe(II) centre for activity. In contrast, with JMJD2A (which reacts similarly to JMJD2E) the histone peptide lysine side chain only just reaches the Fe(II) centre (pink peptide in Figure 4.29) and so likely with the shorter side chain of ornithine the methyl group cannot penetrate into the active site close enough to the Fe(II) centre for reaction. With the shorter side chain of ornithine it is possible that there would be space in the PHF8 active site for a trimethyllysine. As such it would be interesting to synthesise the trimethylornithine lysine analogue peptide to investigate whether the shortening of the alkyl side chain might convert PHF8 from a dimethyl-demethylase to a trimethyl-demethylase.

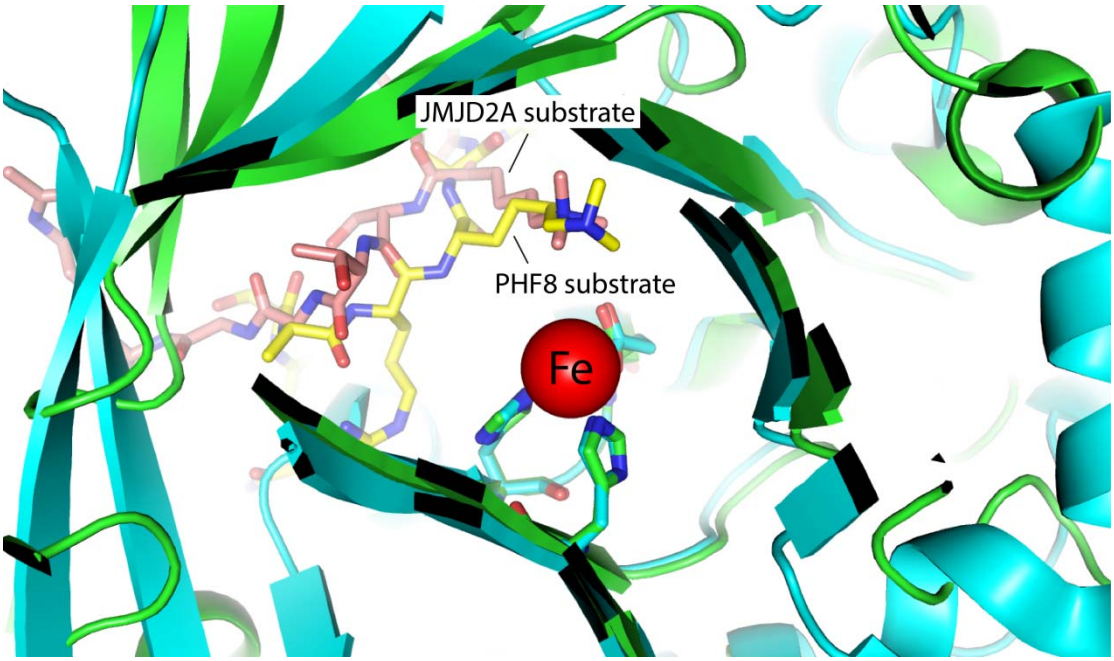


Figure 4.29 Overlay of crystal structures of PHF8 (PDB ID 3KV4)[176], shown in cyan, and JMJD2A (PDB ID 2OQ6)[145], shown in green, bound to histone peptides showing the different binding conformations of the substrate. The PHF8 substrate is shown in yellow and the JMJD2A substrate in pink. The PHF8 substrate can be seen to protrude further into the active site than the JMJD2A substrate.

Although no reaction was observed between PHF8 and the other lysine analogues, JMJD2E was observed to react with the KMe^{i}Pr peptide to form a hydroxylated product, in addition to small amounts of demethylated and deisopropylated products (Figure 4.30B, *experiment carried out by Dr Richard Hopkinson*). This was demonstrated by the appearance of a +16 peak in the MALDI MS spectrum (corresponding to the addition of a hydroxyl group), which was observed to be the predominant product. Given hydroxylation of the methyl group or alpha carbon of the ^{i}Pr group would lead to loss of these functional groups it was assumed that hydroxylation was occurring on the carbon beta to the nitrogen (Figure 4.30A). NMR analyses supported this assignment; NMR spectra showed a decrease in the peak corresponding to the unhydroxylated isopropyl-group (a in Figure 4.30C) and appearance of new peaks corresponding to the hydroxylated product (b, c and d in Figure 4.30C).

This provided the first experimental evidence that JmjC KDMs can catalyse hydroxylation, helping to support the proposed mechanism of histone demethylation occurring *via* hydroxylation.

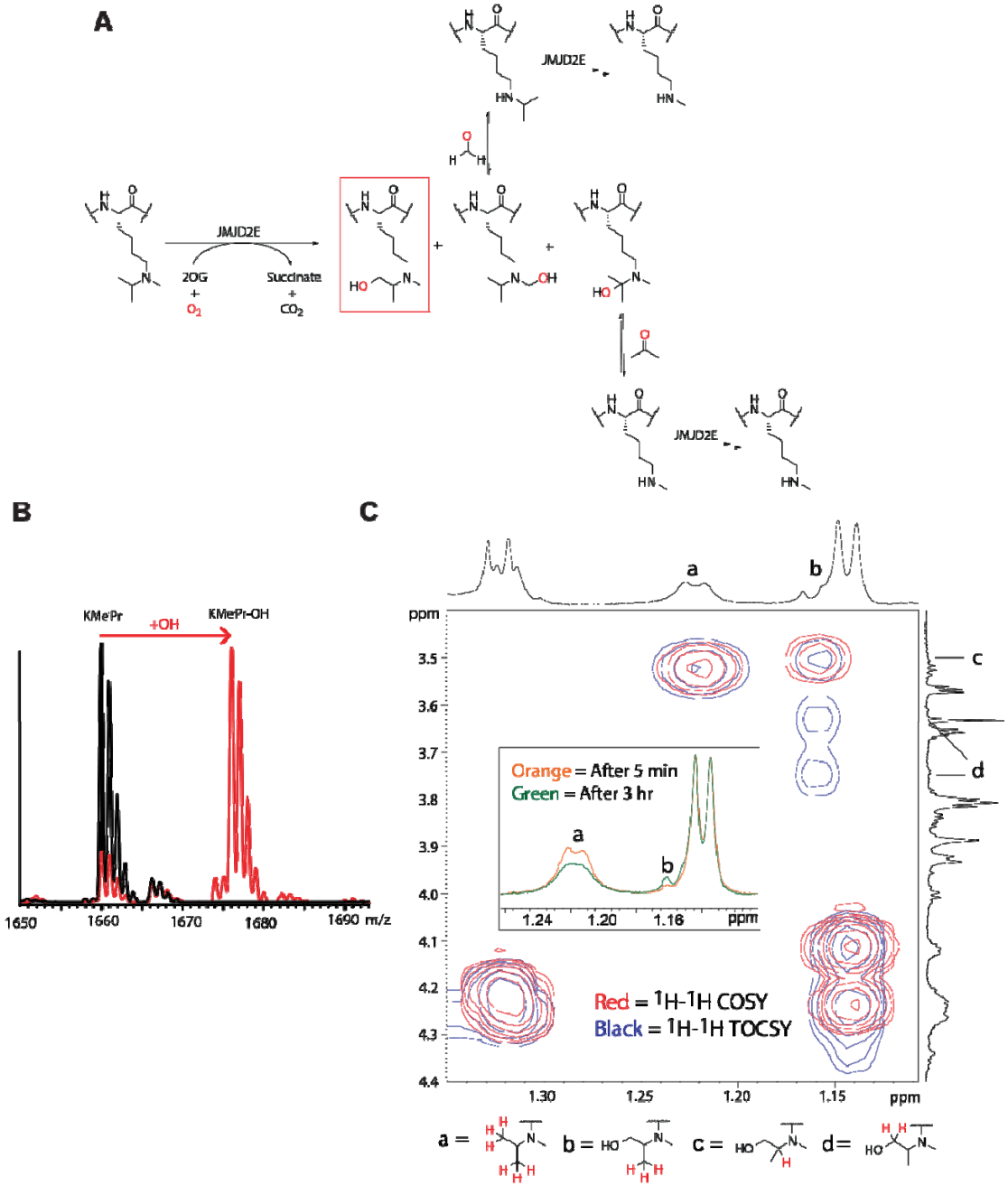


Figure 4.30 Reaction of JMJD2E with a H3 1-15 K4Me$_3$K9Me^{i}Pr peptide results in hydroxylation. **A)** Scheme showing the possible reaction outcomes. **B)** MALDI spectra of reaction after 1 hr. The red spectrum shows spectrum of reaction with JMJD2E (2 μM) and the black spectrum the control reaction in the absence of enzyme. **C)** NMR analyses of the resultant peptide.

This work is summarised in "Is JmjC oxygenase catalysis limited to demethylation?" (Appendix 4)

4.9 Clinical mutations in PHF8

Mutations in the PHF8 gene are associated with X-linked mental retardation and Siderius syndrome.[220,222,320] Recently, clinicians identified a family with a novel mutation in their PHF8 gene, which is found on the X-chromosome (personal communication, Lindsey Miller, Wayne State University). Phenotypic presentation appears to be variable between family members. The two sons show a number of dysmorphic features, including hypertelorism and developmental delay. In addition, the older brother also suffers from a number of psychiatric problems. Interestingly, female members of the family, who are heterozygous for this mutation, show some symptoms, which may also be associated with the mutation. The sister has pectus excavatum and abnormal body hair growth and the mother is reported to show exercise intolerance, muscle weakness, cramping and fatigue. Genetic tests identified an insertion in the gene, notated as c.1306dupG (accession NM_015107.2).

4.9.1 Bioinformatic analyses of PHF8 1306dupG

Bioinformatic analysis predicted the 1306dupG insertion to cause a frameshift that would result in a premature stop codon in the PHF8 sequence. The resulting protein would have only 522 amino acids rather than the expected 1024. This would leave the PHD finger and catalytic JmjC domain intact (as these are contained within the first 435 amino acids), but the following 93 amino acids are incorrect. Furthermore, the premature stop codon would remove all five annotated nuclear localization signals

(Figure 4.31).[292] This would be predicted to result in mislocalisation of the protein to the cytoplasm.

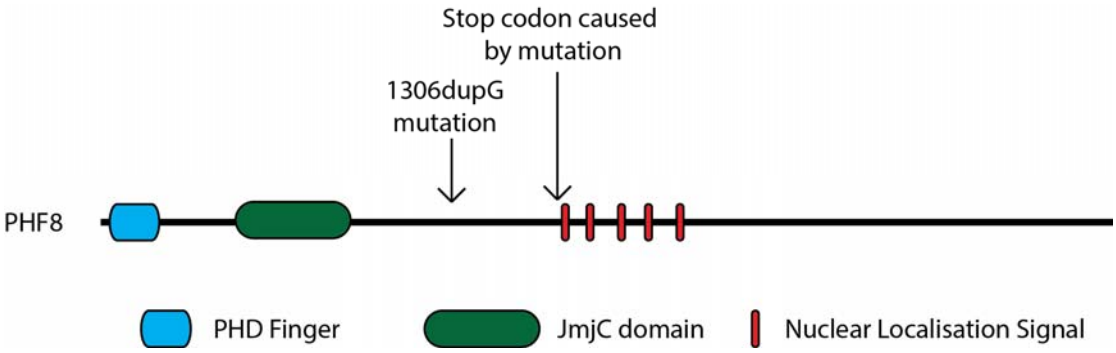


Figure 4.31 Domain organisation of PHF8 showing the location of the mutation and the site of the premature stop codon.

4.9.2 Production and testing of the clinical PHF8 mutant in cells

To test the outcome of the mutation in cells, site-directed mutagenesis was carried out on a plasmid encoding Hemagglutinin (HA)-tagged full-length PHF8 protein (gift from Dr. Christoph Loenarz) to produce the clinical variant (Section 8.6.2). Primers were designed to insert a single guanosine base (G) into the sequence at the desired location by site-directed mutagenesis. The FL PHF8 1306dupG variant was then overexpressed in three human cells lines (HeLa, HEK 293T and U2OS cells) and its localisation compared to that of wt PHF8 protein (Section 8.12.3). Unfortunately, all commercially available anti-PHF8 antibodies are raised to *C*-terminal epitopes and are thus unable to recognize the truncated protein. Consequently, an anti-HA antibody followed by a fluorescently labelled secondary antibody was used to visualize the exogenously expressed protein *via* its *N*-terminal HA tag. The nucleus itself was stained using 4',6-diamidino-2-phenylindole (DAPI), which binds strongly to regions of DNA rich in A-T nucleotides.[321] As predicted, the mutated PHF8 was found to be excluded from the nucleus in all three cell lines, whilst the wt protein was almost exclusively nuclear, as demonstrated by the overlap of red (anti-HA antibody) and

blue (DAPI) staining (Figure 4.32).

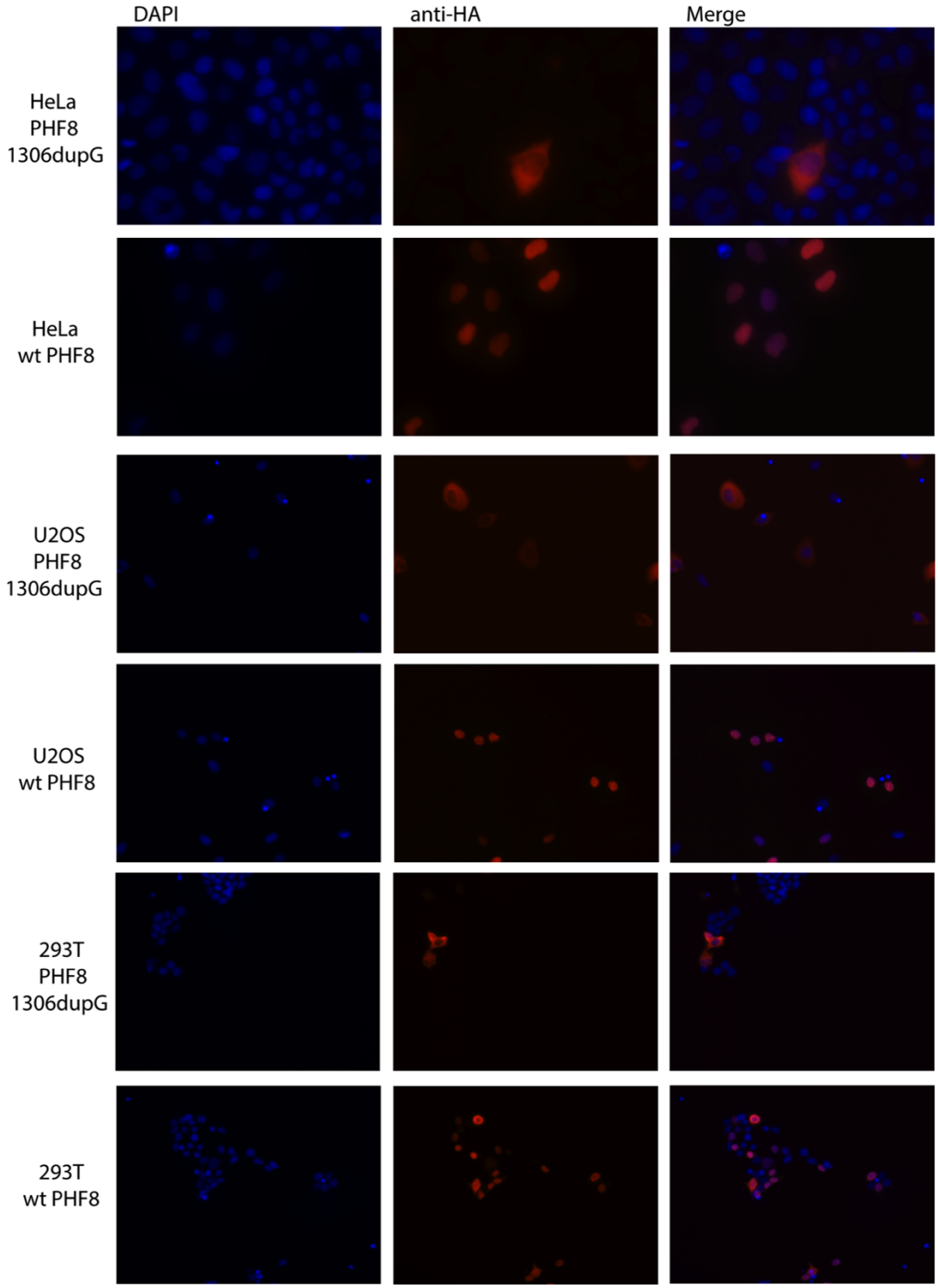


Figure 4.32 Immunofluorescence studies showing the sub-cellular localisation of HA-tagged wt PHF8 and PHF8 1306dupG. Nuclei were stained with DAPI and are shown in blue. PHF8 was visualised using an anti-HA antibody recognising its *N*-terminal HA tag and is shown in red.

A number of clinical mutations in the PHF8 gene have previously been

reported.[221,222,320] However, each of these is predicted to lead to expression of catalytically inactive forms of PHF8, either due to mutations or deletions in the JmjC domain itself or in the neighbouring PHD finger, which is essential for substrate binding (Figure 4.1B). In these cases, the lack of catalytic activity is thought to be directly responsible for the phenotype.[221,222,292] Therefore, the 1306dupG mutation represents the first example of a clinically relevant PHF8 mutation that possesses intact catalytic machinery, but appears mislocalised, and has been found to cause a phenotype. This is presumably due to the protein being unable to act on its intended histone substrate due to its altered sub-cellular localisation. It is also possible that expression of PHF8 within the cytoplasm allows it to erroneously demethylate an alternative protein that is not its intended substrate and that this too might cause unwanted side effects. Alternative splicing might also cause the mutation to be missed. To investigate whether any of these possibilities occur *in vivo,* determination of the expression levels and molecular weight of PHF8 in patient tissue samples would be required.

4.10 Discussion

The work described in this Chapter covered the biochemical characterisation of PHF8 and KIAA1718, two members of the KDM7 subfamily of histone demethylases. A high-throughput assay for measuring inhibitor IC_{50} values against PHF8 has been developed and corroboration of inhibition data for KIAA1718 can be further carried out using a MALDI-TOF MS assay. Robust, high-throughput assays for the screening of inhibitors are essential tools for the design of chemical probes and to facilitate drug development. This assay will enable the identification of KDM7 subfamily specific inhibitors. These should help elucidate the specific cellular functions of these

proteins, for example, the mechanisms by which PHF8 mutations lead to developmental disorders.

Daminozide, a novel inhibitor of the KDM2 and KDM7 subfamilies of histone demethylases has been characterised, and a slightly less selective, but cell-active variant screened against PHF8 and KIAA1718. The identification of an analogue of Daminozide that appears selective for PHF8 over FBXL11 could provide a tool for investigating the specific functions of the KDM7 subfamily. Further characterisation is required before this will provide a cell permeable chemical tool, however. Further work could be carried to develop an immunofluorescence or qPCR-based assay for measuring PHF8 activity in cells, which could be used to test the efficacy of PHF8 specific inhibitors. As described above, Daminozide was originally used as a plant growth retardant that was thought to act by inhibition of 2OG oxygenases involved in gibberellin and ethylene synthesis. The discovery that Daminozide is also an inhibitor of members of the human KDMs raises the question of whether its mode of action in plants is also caused by inhibition of related plant 2OG oxygenase histone demethylases. In high doses, Daminozide was shown to be carcinogenic in rats, and was therefore removed from use in food crops.[310] This toxic behaviour was largely ascribed to the breakdown metabolite, 1,1-dimethylhydrazine, but given the inhibition observed by Daminozide on KDM2 and KDM7 subfamily members it remains possible that inhibition of JmjC KDMs was also responsible for the toxicity. In either case, the extremely high levels of compound ingestion required to observe toxicity render it unlikely that harm was caused to humans through use of Daminozide on crops. Unfortunately, Daminozide has been demonstrated not to be cell active, likely due to its low cellular permeability. As cellular penetrability has been a significant

hindrance in the development of useful KDM inhibitors, future work should focus on alternative methods of compound delivery to cells, such as using nanoparticles or liposomes, or developing cell permeable prodrug forms of each inhibitor.[322,323] An alternative strategy may be to identify inhibitors without the acid moiety that is thought to be a barrier to cell permeability, for which high-throughput screening assays, such as the ALPHA screen described above will be invaluable. Inhibitors targeting the reader domains of the KDMs may also be useful and the FP assay described above for investigating the binding of peptide to the PHD finger of PHF8 could be adapted to identify inhibitors that block this interaction. The relatively high concentration of protein and peptide currently used in the assay render it unsuitable for a high-throughput assay without further modification. Miniaturisation is required to adapt this for use as a screening assay. In all the assays, small molecule inhibitors are only screened against truncated constructs of the full length KDMs. As such it is possible that they will not be so efficacious when binding to the full-length proteins in cells. With low molecular weight compounds such as Daminozide this is likely not a problem, but must be taken into account when larger inhibitors are investigated.

Mutations in the PHF8 gene are known to cause mid-line defects and X-linked mental retardation, and investigations revealed that a clinically relevant PHF8 mutation results in mislocalisation of the protein to the cytoplasm. As the experiments were not carried out on a confocal microscope, it is unclear whether in addition to the cytoplasmic localisation observed there also remains some partial nuclear localisation. Use of a confocal microscope for further localisation studies would help to determine this. In addition, western blots analysing the presence of flag-tagged protein in cytoplasmic and nuclear extracts could be carried out to further confirm the protein's

cellular localisation. Whether the resultant phenotype is due to a reduction in histone demethylation caused by PHF8 at crucial gene targets, or catalytic activity of PHF8 at unwanted sites in the cytoplasm is currently unclear. Coimmunoprecipitation studies could be used to investigate whether localisation to the cytoplasm causes PHF8 to demethylate unwanted substrates, or bind to proteins *via* interactions with its PHD finger. Although bioinformatic analysis would suggest that the protein remains active as it contains the complete PHD and JmjC domains, it will be crucial to demonstrate that the truncated protein is correctly folded and catalytically active by expressing the protein in *E. coli* and looking for demethylation of an H3 $K4Me_3K9Me_2$ peptide by MS.

The work described above also demonstrates that the substrate scope of PHF8 (and other KDMs) is potentially much broader than was previously thought. PHF8 has been shown to be able to remove both ethyl and isopropyl groups from the lysine side chain of histone peptides and to demethylate ornithyl-residues. It is interesting that in the case of ornithine, PHF8 is only able to monodemethylate this residue to form monomethylornithine. This is likely due to the fact that when the shorter side chain is bound in the active site it is further from the Fe(II) centre and once it is only singly methylated the most favourable binding conformation for the lysyl side chain points the methyl group away from Fe(II) centre as is proposed for JMJD2A.[145] Crystallography could be used to further investigate this and synthesis of the trimethylated ornithine peptide could be undertaken to investigate whether shortening the side-chain is able to convert a di-demethylase into a tri-demethylase.

JMJD2A was shown to catalyse a stable hydroxylation reaction, providing a link between the two different functions of JmjC enzymes, which are currently characterised as either hydroxylases or demethylases. Although the mechanism of demethylation by JmjC KDMs is thought to proceed *via* initial hydroxylation of the methyl group (Figure 1.8), the hemiaminal intermediate for the human enzymes has not yet been observed, presumably due to the fast collapse of this intermediate to produce formaldehyde and the demethylated peptide. Based upon the mechanism of other member of the 2OG oxygenase superfamily and the detection of formaldehyde on demethylation this is nonetheless considered the most likely reaction mechanism. Observation of hydroxylation of a lysine analogue by JMJD2A provides evidence that the KDMs can catalyse hydroxylation alone and thus that the proposed demethylation mechanism is possible. This evidence is still indirect, however, and it remains desirable to obtain a crystal structure in which the hemiaminal intermediate has been trapped (cf. AlkB (Figure 1.9)). In addition, synthesis of the all carbon side chain analogue of lysine could provide further evidence that an exact isostere of methylated lysine can be hydroxylated, although due to the loss of charge from the lysine side chain nitrogen, reaction with the KDMs may no longer be observed at all. Recent proteomic studies of histones have identified an increasing number of novel post-translational modifications, and the identification of new *in vitro* substrates for these enzymes also raises the question of whether the JmjC KDMs may have substrates other than N^{ε}-methyllysine within cells.[69,324]

Chapter 5
Investigations into arginine demethylation by 2OG oxygenases

5.1 Introduction

Like N^{ε}-lysine methylation, N^{ω}-methylation of arginyl-residues is an important post-translational modification on both histone tails and many other proteins. A study of arginine methylation levels in rat liver nuclei suggests that ~2% of all arginyl-residues are dimethylated,[325] and a number of more recent large-scale proteomic studies have identified arginine methylation sites on numerous proteins within the cell.[326,327] The guanidino group of arginine contains four potential hydrogen bond donors, and methylation of arginine reduces this number by one for each added methyl group. In addition these groups change its shape and hydrophobicity, all of which has effects on the interactions the arginyl-residue makes with other proteins / DNA (Figure 5.1).[328] Histone arginine methylation can cause both gene activation and gene repression (Section 1.1.3.5, Chapter 1), and arginine methylation of other proteins can cause many outcomes, including altering protein-protein or protein-DNA interactions, cellular localisation and protein stability.[329–332]

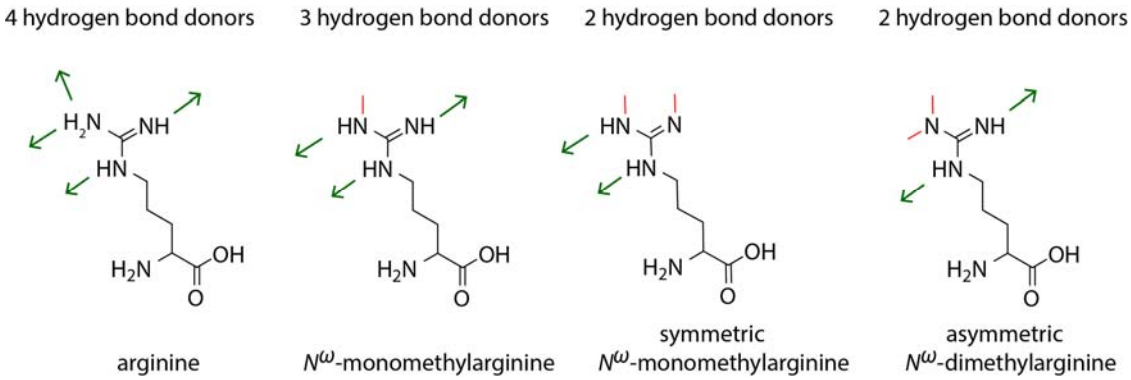


Figure 5.1 Identified forms of methylated arginine showing potential hydrogen bond donors with green arrows and methylation with red bonds.

The dynamic nature of the arginine methylation mark is currently questionable, however. Although the PRMT enzymes that catalyse addition of arginine methylation are well described, no true reversing demethylases are currently known. Deimination of monomethylarginine to citrulline by PAD4 has been characterised,[126,128] but this reaction does not return the unmethylated arginyl-residue (Figure 1.12, Chapter 1).

The 2OG oxygenase JMJD6 was initially identified as an arginine demethylase of H3 R2 and H4 R3, but this work has not yet been reproduced and JMJD6 has since been assigned as a lysine hydroxylase (Figure 1.11, Chapter 1).[121,122,333] There is some evidence, however, that protein arginine methylation levels do fluctuate within cells; for example, transient methylation of estrogen receptor α by PRMT1 occurs in response to estrogen, cyclical arginine methylation of H3 R17 and H4 R3 is observed by ChIP studies at the pS2 promotor after estrogen treatment of MCF-7 cells and global H3 R17Me_{2a} levels oscillate through the cell cycle, with maximum levels during mitosis.[334–336] In all these examples, the reduction in arginine methylation levels of the residue in question does not occur with concominant decrease in the levels of the protein in which the arginyl-residues are found, suggesting that the

arginine methylation is being removed by an active demethylation mechanism, rather than by whole protein degradation.

5.1.2 Objectives

The work described in this Chapter aimed to investigate the possibility of arginine demethylation catalysed by members of the 2OG oxygenase superfamily. Work initially focused on the histone demethylase JMJD2E and was then extended to screening of a panel of 2OG oxygenases against methylated arginine-containing peptides. Characterisation of the observed arginine demethylation was carried out. Work then addressed the identification of biologically relevant substrates and the possibility of arginine demethylation occurring in cells by a variety of different methods.

Initial MALDI experiments with JMJD2E were carried out by Dr. Richard Hopkinson, as were all NMR experiments. Orbitrap MS samples were run by Jacob Bush and histone MS samples by Rok Sekirnik. Peptides for screening were synthesised by Dr. Wei Ge. IRMS experiments were carried out in collaboration with Dr. Nikita Loik.

5.2 Identification of arginine demethylation by JMJD2E

The ability of the KDMs to react with lysine substrates modified with chemical moieties other than methyl groups (see Chapter 4) raises the possibility that they may also have other post-translationally modified substrates *in vivo*. Indeed, the structure of KMe^{i}Pr can be considered isosteric with that of asymmetric dimethylarginine, a mark known to be biologically relevant (Figure 5.2).

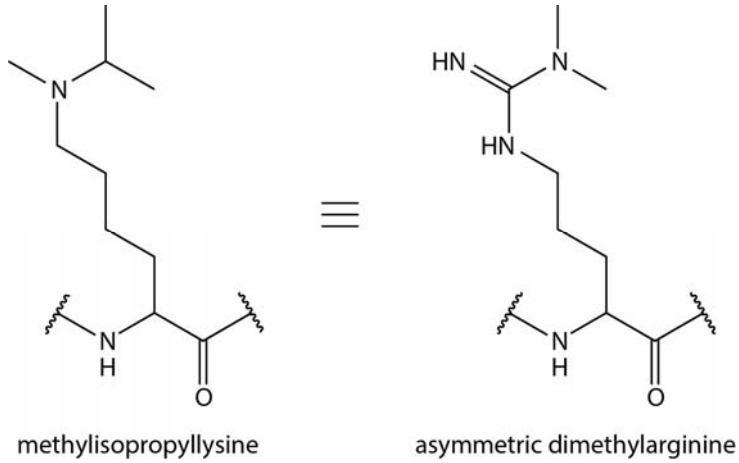


Figure 5.2 Methyisopropyllysine is isosteric with asymmetric dimethylarginine.

The ability of the JMJD2 subfamily to hydroxylate KMe^{i}Pr on the isopropyl group that "corresponds" to one of the methyl groups on arginine raises the possibility that they might also be able to demethylate arginyl-residues. To test this possibility, a histone peptide was synthesised where the target lysyl-residue (K9) for JMJD2E catalysed lysine demethylation was substituted with an asymmetrically dimethylated arginyl-residue (RMe_{2a}, ARTKQTARRMe_{2a}STGGKA, Section 8.11). Thus, the methylated arginyl-residue is in the same position as the methylisopropyllysyl-residue that is hydroxylated by JMJD2E (Chapter 4). It was hypothesised that placing the methylated arginyl-residue within a histone peptide sequence that is known to bind to JMJD2E should maximise the likelihood of demethylation occurring by directing the methylated residue towards the active site Fe(II). This peptide (10 μM) was incubated with JMJD2E (2 μM) and cofactors under standard demethylation conditions for 1 hr at 37 °C. After methanol quenching, the reaction mixture was analysed by MALDI-TOF MS (*experiment carried out by Dr Richard Hopkinson*, Section 8.7.1). Demethylation was observed to form the monomethylarginine species (peak at 1603 m/z, Figure 5.3A). Reactions were carried out in the absence of the essential cofactors, 2OG and Fe(II), to confirm that the demethylation reaction was being

carried out by JMJD2E catalysis, rather than an impurity in the enzyme purification. No reaction was observed in the absence of 2OG and very little in the absence of Fe(II) (Figure 5.3B and C); the low degree of activity detected in the absence of Fe(II) was ascribed to JMJD2E copurifying with some Fe(II) in the active site.

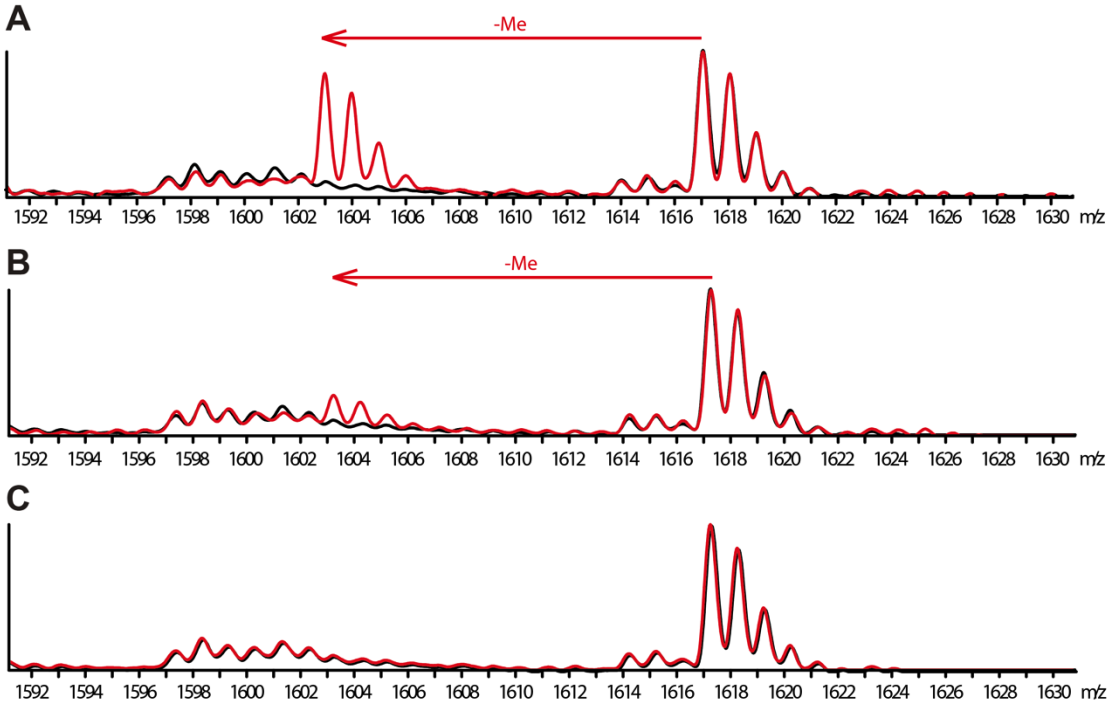


Figure 5.3 MALDI-TOF MS of demethylation of an H3 K9RMe$_{2a}$ 15mer peptide (10 µM) by JMJD2E (2 µM). **A)** Demethylation of H3 K9 RMe$_{2a}$ by JMJD2E. **B)** Very little demethylation is observed in the absence of Fe(II). **C)** No demethylation is observed in the absence of 2OG. All spectra are shown after incubation for 1 hr at 37 °C. The black spectra show control reactions in the absence of enzyme, the red spectra show the reactions including JMJD2E.

5.2.2 Demethylation of symmetric dimethylarginine by JMJD2E

Dimethylation of arginine can occur in two biologically relevant different forms, with both methyl groups on one of the guanidino nitrogen atoms, asymmetric methylation, or one methyl group on each of the terminal guanidino nitrogen atoms, symmetric methylation (Figure 5.1). These two forms are sterically different, so it is possible that an enzyme that can demethylate one form, may not accept the other as a substrate. To test whether JMJD2E is capable of removing symmetric arginine methylation, the same mutant peptide sequence was synthesised containing symmetric arginine

methylation (RMe_{2s}, ARTKQTARRMe_{2s}STGGKA, Section 8.11). Again JMJD2E (2 μM) was incubated with the peptide (10 μM) and cofactors for 1 hr at 37 °C before analysis by MALDI-TOF MS (*Experiment carried out by Dr Richard Hopkinson*). Peaks at -14 Da and -28 Da from the mass of the substrate peptide indicated that both mono- and di-demethylation had occurred (Figure 5.4).

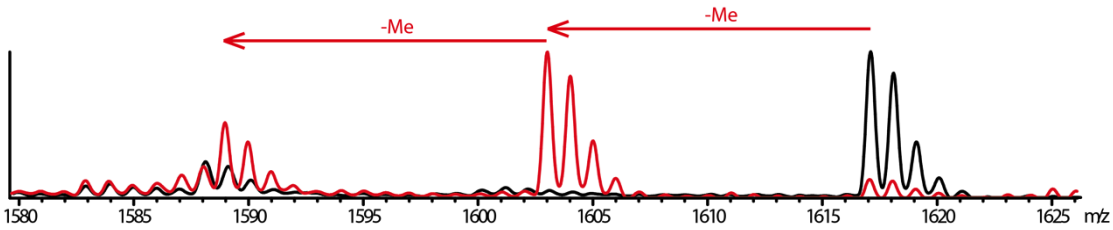


Figure 5.4 MALDI-TOF MS of demethylation of an H3 K9RMe_{2s} 15mer peptide (10 μM) by JMJD2E (2 μM) at 37 °C after 1 hr. A control reaction in the absence of enzyme is shown in black and the reaction containing JMJD2E in red.

5.3 Screening for arginine demethylation by other 2OG oxygenases

Given the ability of JMJD2E to demethylate methylated arginine, it was considered possible that other related JmjC KDMs might be able to carry out similar reactions. Eleven more peptides were synthesised, with K4, K9, K27 and K36 substituted for asymmetric dimethyl-, symmetric dimethyl- and monomethylarginine. In addition to the K9 mutant peptides, peptides containing K4Me_3 in addition to K9R were synthesised for testing with PHF8 (*synthesised by Dr. Wei Ge*). The sequences of the synthesised peptides and the recombinant enzymes with which they were screened are given in Table 5.1 (*recombinant enzymes other than PHF8, KIAA1718, UTX and JMJD3 were generously provided for testing by other members of the Schofield laboratory*).

Table 5.1 Table showing the methylated arginine peptides synthesised and the enzymes they were tested against. Target arginine residues are shown in bold.

Mutant residue	Peptide sequence	Enzymes tested
H3 K4R	ART**RMe$_{2a}$**QTARKSTGGKA ART**RMe$_{2s}$**QTARKSTGGKA ART**RMe**QTARKSTGGKA	JARID1C
H3 K9R	ARTKQTAR**RMe$_{2a}$**STGGKA ARTKQTAR**RMe$_{2s}$**STGGKA ARTKQTAR**RMe**STGGKA	JMJD2A/D/E JMJD1A
H3 (K4Me$_3$)K9R	ARTKQTAR**RMe$_{2a}$**STGGKA ARTKQTAR**RMe$_{2s}$**STGGKA ARTKMe$_3$QTAR**RMe**STGGKA	PHF8
H3 K27R	KQLATKAAR**RMe$_{2a}$**SAPAT KQLATKAAR**RMe$_{2s}$**SAPAT KQLATKAAR**RMe**SAPAT	JMJD3 UTX KIAA1718
H3 K36R	PATGGV**RMe$_{2a}$**KPHRY PATGGV**RMe$_{2s}$**KPHRY PATGGV**RMe**KPHRY	FBXL11

Although most of the JmjC KDMs tested did not display arginine demethylation activity with their mutant substrates, activity was detected for JMJD3 (Figure 5.5A), JARID1C (Figure 5.5B) and the other members of the KDM4 subfamily (JMJD2A and JMJD2D, *data not shown*). Like JMJD2A/D/E, JARID1C was observed to demethylate both symmetric and asymmetric dimethylated arginine, although the enzyme batch used likely contained protease impurities as the peptide was also degraded over the course of the experiment (JARID1C is not active when expressed recombinantly in *E. coli* and was acquired from collaborators at the Structural Genomics Consortium where it is expressed in insect cells). Unlike the other tested KDMs, found to be capable of removing arginine methylation, JMJD3 was unable to demethylate symmetrically methylated arginine. Interestingly, no demethylase activity was detected with the other KDM6 subfamily members, UTX or UTY (see Chapter 6).

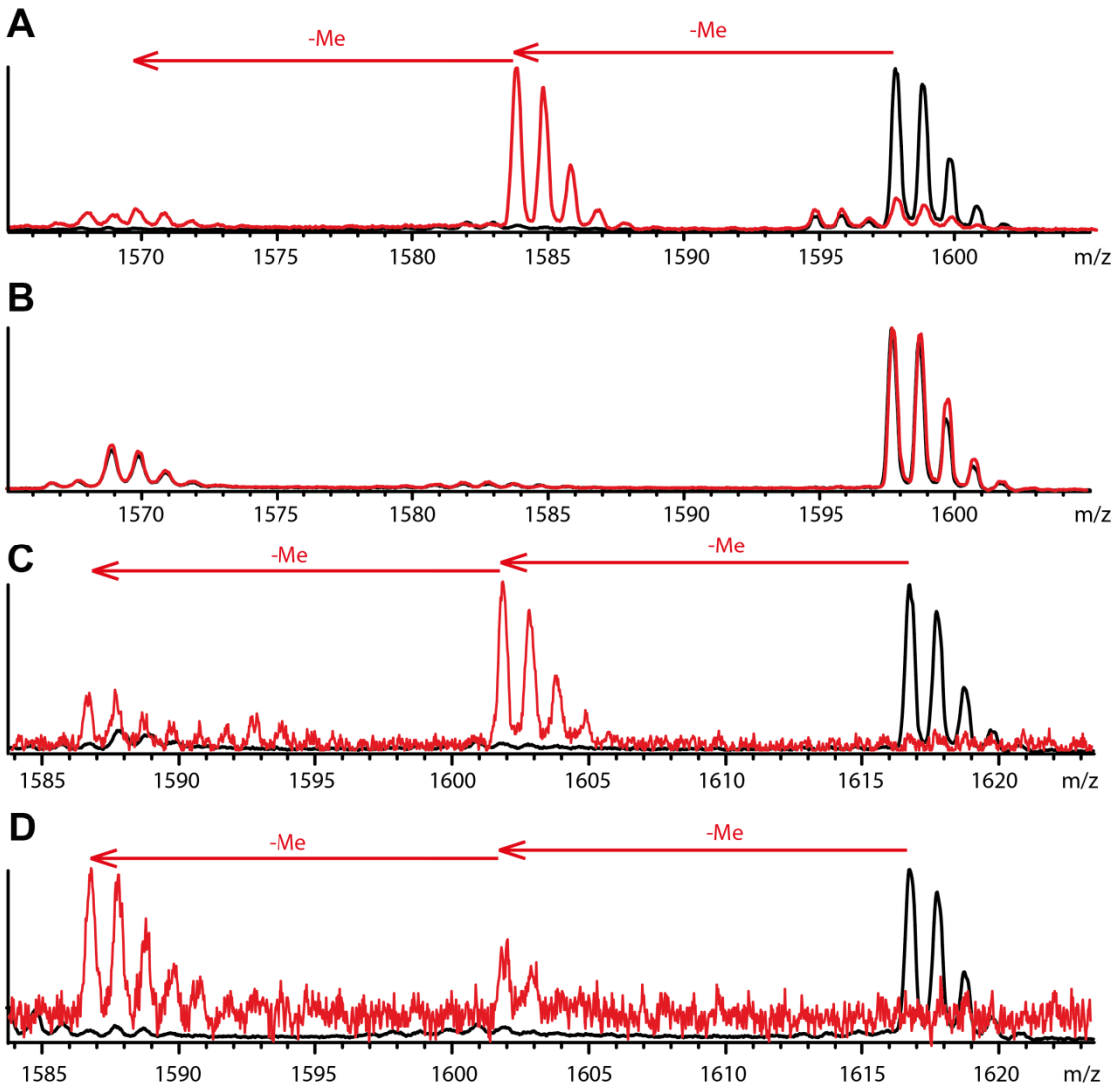


Figure 5.5 MALDI-TOF MS of demethylation of **A)** an H3 K27RMe$_{2a}$ 15mer peptide (10 μM) and **B)** an H3 K27RMe$_{2s}$ 15mer peptide (10 μM) by JMJD3 (2 μM) and **C)** an H3 K4RMe$_{2a}$ 15mer peptide (10 μM) and **D)** an H3 K4RMe$_{2s}$ 15mer peptide (10 μM) by JARID1C (2 μM) after 1 hr at 37 °C. Control reactions in the absence of enzyme are shown in black, spectra for reactions containing enzyme are shown in red. Note spectra for reaction with JARID1C showed a substantial extent of peptide substrate degradation in addition to demethylation.

5.3.2 Screening for histone peptide arginine demethylation by 2OG oxygenases

A panel of histone peptides methylated on arginyl-residues and the mutant peptides with lysyl-residues substituted for arginyl-residues was then created (Table 5.2, *peptides provided by Dr. Wei Ge*). The activity of the recombinant 2OG oxygenases

available in the laboratory was tested against these peptides. In addition to the enzymes listed in Table 5.1, the cytosine demethylase and DNA lyase ABH1 and the functionally unassigned 2OG oxygenase ABH5 were included in the screen.[337–339] JMJD6 was not included in the screen as it had previously been screened in the laboratory and shown to have no arginine demethylase activity.

Table 5.2 Arginine-methylated histone peptides screened against a recombinant 2OG oxygenase panel. Target arginyl-residues are shown in bold. KAc is N^{ε}-acetylated lysine.

Histone mark	Peptide name	Peptide sequence
H3 R2	H3 R2 1	A**RMe$_{2a}$**TKQTARKSTGGKAPRK
	H3 R2 2	A**RMe$_{2s}$**TKQTARKSTGGKAPRK
	H3 R2 3	A**RMe$_{2a}$**TKMe$_3$QTARKSTG
	H3 R2 4	A**RMe$_{2s}$**TKMe$_3$QTARKSTG
H3 R8	H3 R8 1	ARTKQTA**RMe$_{2a}$**KSTGGKAPRK
	H3 R8 2	ARTKQTA**RMe$_{2s}$**KSTGGKAPRK
	H3 R8 3	ARTKQTA**RMe**KSTGGKAPRK
	H3 R8 4	TKQTA**RMe$_{2a}$**KAcSTGGK
H3 R17	H3 R17 1	STGGKAP**RMe$_{2a}$**KQLATKAARK
	H3 R17 2	STGGKAP**RMe$_{2s}$**KQLATKAARK
	H3 R17 3	STGGKAP**RMe**KQLATKAARK
	H3 R17 4	TGGKAP**RMe$_{2a}$**KAcQLATK
	H3 R17 5	TGGKacAP**RMe$_{2a}$**KQLATK
	H3 R17 6	TGGKAP**RMe$_{2a}$**KAcQLATK
	H3 R17 7	TGGKAP**RMe$_{2s}$**KAcQLATK
H3 R26	H3 R26 1	QLATKAA**RMe$_{2a}$**KSAPATGGV
	H3 R26 2	QLATKAA**RMe$_{2s}$**KSAPATGGVK
	H3 R26 3	QLATKAA**RMe$_{2a}$**KSAPATGGV
H4 R3	H4 R3 1	SG**RMe$_{2a}$**GKGGKGLGKGGAK
	H4 R3 2	SG**RMe$_{2s}$**GKGGKGLGKGGAK
H2A R3	H2A R3 1	SG**RMe$_{2a}$**GKQGGKARAKTR

Although no new enzymes capable of demethylating arginine were detected, for both the JMJD2 subfamily and JARID1C, a number of biologically relevant histone peptide substrates were identified, in addition to the mutant histone peptides previously tested. The results provide the first evidence that arginine demethylation is possible on a biologically relevant peptide sequence and are summarised in Table 5.3.

Table 5.3 Table summarising the demethylation of methylated arginyl-residues on histone peptides (~10 μM) by JMJD2A (2 μM), JMJD2E (2 μM) and JARID1C (2 μM). Black crosses denote < 5% demethylation observed, red ticks denote > 5% demethylation.

Histone arginine methylation mark		Enzyme		
		JMJD2A	JMJD2E	JARID1C
H3 R2	Me_{2a}	✓	✓	✓
	Me_{2s}	✗	✓	✓
H3 R8	Me_{2a}	✗	✓	✓
	Me_{2s}	✗	✓	✗
H3 R17	Me_{2a}	✗	✗	✗
	Me_{2s}	✗	✗	✗
H3 R26	Me_{2a}	✗	✗	✗
	Me_{2s}	✗	✗	✗
H4 R3	Me_{2a}	✓	✓	✓
	Me_{2s}	✗	✓	✓
H2A R3	Me_{2a}	✗	✓	✗
	Me_{2s}	✗	✗	✗

JMJD2E was observed to demethylate symmetric and asymmetric methylation of H3 R2 and H3 R8 to the non-methylated species and symmetric and asymmetric methylation of H4 R3 only to the monomethylated species (Figure 5.6). JMJD2A showed more limited demethylation, with activity only for mono-demethylation of H3 R2Me_{2a} and H4 R3Me_{2a}, (likely due to its reduced catalytic activity relative to recombinant JMJD2E *in vitro*). JARID1C showed demethylase activity of symmetric and asymmetric methylation at H3 R2, H4 R3, the mutant H3 K9R and asymmetric methylation at H3 R8. The predominant products in all cases were the monomethylated arginine containing peptides, although small amounts of di-demethylation was observed at H3 R8 to form the non-methylated arginine peptide (H3 R8 3, Table 5.2). In peptides containing both arginine and lysine methylation in biologically relevant positions, such as peptide H3 R2 3 (Table 5.2) as many as five demethylation reactions could be observed resulting in the completely demethylated

peptide, due to lysine demethylation of H3 K4Me$_3$ and arginine demethylation of H3 R2Me$_{2a}$ by JARID1C. It was particularly interesting that JARID1C was able to demethylate the H3 K9RMe$_{2a/s}$ peptide as it is unable to demethylate the natural H3 K9Me$_3$ peptide, suggesting that the sequence preference for arginine demethylation may be different to that of lysine demethylation. For both the KDM4 subfamily and JARID1C, K9 acetylation appeared to block demethylation at R8 (no demethylation of peptide H3 R8 4). In the case of JMJD3, however, activity was only observed for the H3 K27R control peptide.

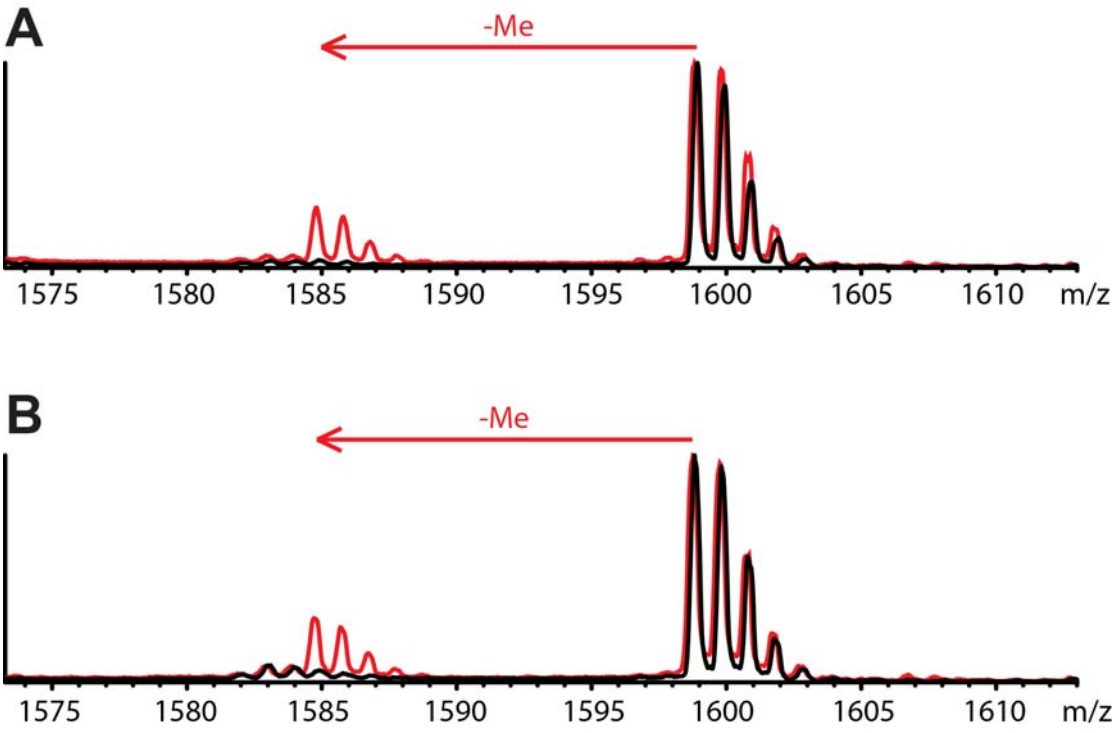


Figure 5.6 Example spectra from the 2OG oxygenase screen showing demethylation of **A)** H4R3 1 and **B)** H4 R3 2 to the monomethylated species by JMJD2E (2 μM) after 1 hr at 37 °C. Control reactions in the absence of enzyme are shown in black, spectra for reactions containing enzyme are shown in red.

5.3.3 Effect of histone acetylation of arginine demethylation

The effect of lysine acetylation in combination with arginine methylation on the demethylation behaviour of JMJD2E and JARID1C was investigated. Lysine

acetylation is known to occur on both H3 K4 and H3 K9 in cells. As these are both close to the target H3 R2 site it was proposed that acetylation of these residues might therefore affect the ability of JMJD2E and JARID1C to carry out demethylation. To test if this was the case, peptide sequences incorporating either H3 K4 or H3 K9 acetylation were synthesised (*synthesised by Dr. Wei Ge*, tested crude, Table 5.4).

Table 5.4 Peptide sequences with multiple post-translational modifications. Target arginyl-residues are shown in bold. KAc denotes N^{ε}-acetylated lysine.

Histone mark	Peptide name	Peptide sequence
H3 R2	Ac 1	A**RMe$_{2a}$**TKAcQTARKSTGGKAPRK
	Ac 2	A**RMe$_{2s}$**TKQTARKSTGGKAcAPRK

Acetylation of either H3 K4 or H3 K9 appeared to block arginine demethylation activity completely for JMJD2E. For JARID1C, acetylation of K4 (Ac 1) blocked all enzymatic activity, but although acetylation at K9 (Ac 2) reduced the demethylation activity a small amount of monomethylated peptide (< 10%) at 1615 m/z could be observed (Figure 5.7).

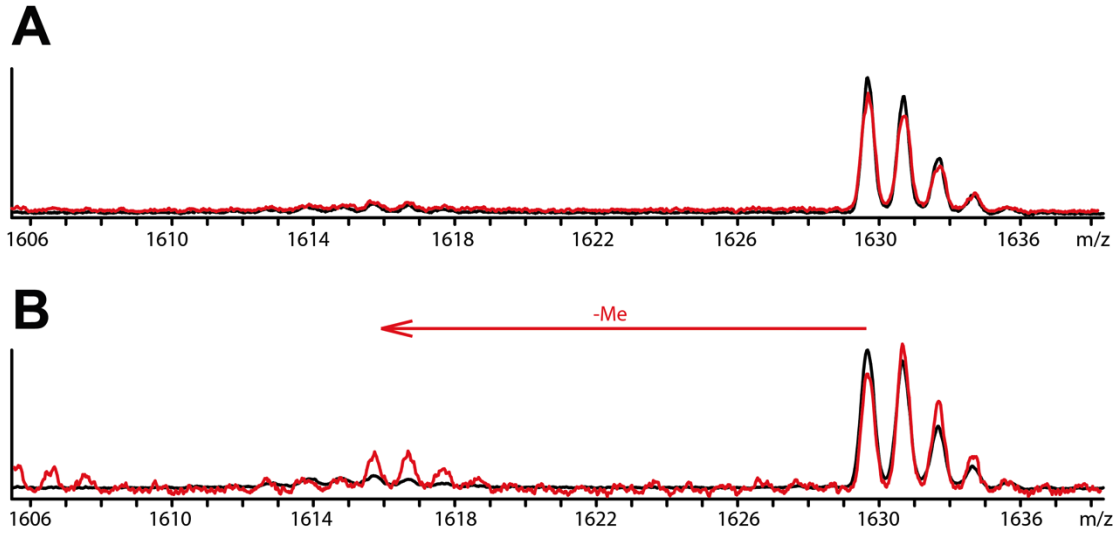


Figure 5.7 MALDI-TOF MS showing **A)** absence of demethylation of H3 R2Me$_{2a}$K4Ac 15mer peptide (10 μM) by JARID1C (2 μM) and **B)** slight demethylation of an H3 R2Me$_{2a}$K9Ac 15mer peptide (10 μM) by JARID1C (2 μM) after 1 hr at 37 °C. Control reactions in the absence of enzyme are shown in black, spectra for reactions containing enzyme are shown in red.

5.4 Characterisation of demethylation of H3 R2Me$_{2a}$ by JMJD2E

As the activity of JMJD2E on H3 R2Me$_{2a}$ was significantly greater than on its mutant H3 K9R peptide substrate, this peptide reaction was selected for characterisation by NMR and FDH assays (Section 8.7.2 and 8.15). JMJD2E (20 μM) was incubated with peptide (ARMe$_{2a}$TKQTARKSTGGKA, 100 μM) and cofactors, and a ^{1}H NMR spectrum was collected after 30 min. This demonstrated that monomethylarginine was being formed (formation of a small resonance at δ_H = 2.784 ppm) and that 2OG was being converted into succinate over the course of the reaction (decrease of the triplet resonance at δ_H = 2.39 and formation of a singlet resonance at δ_H = 2.35 ppm) (Figure 5.8, *experiment carried out by Dr. Richard Hopkinson*). For confident assignment of the monomethylarginine signal, a peptide standard is required.

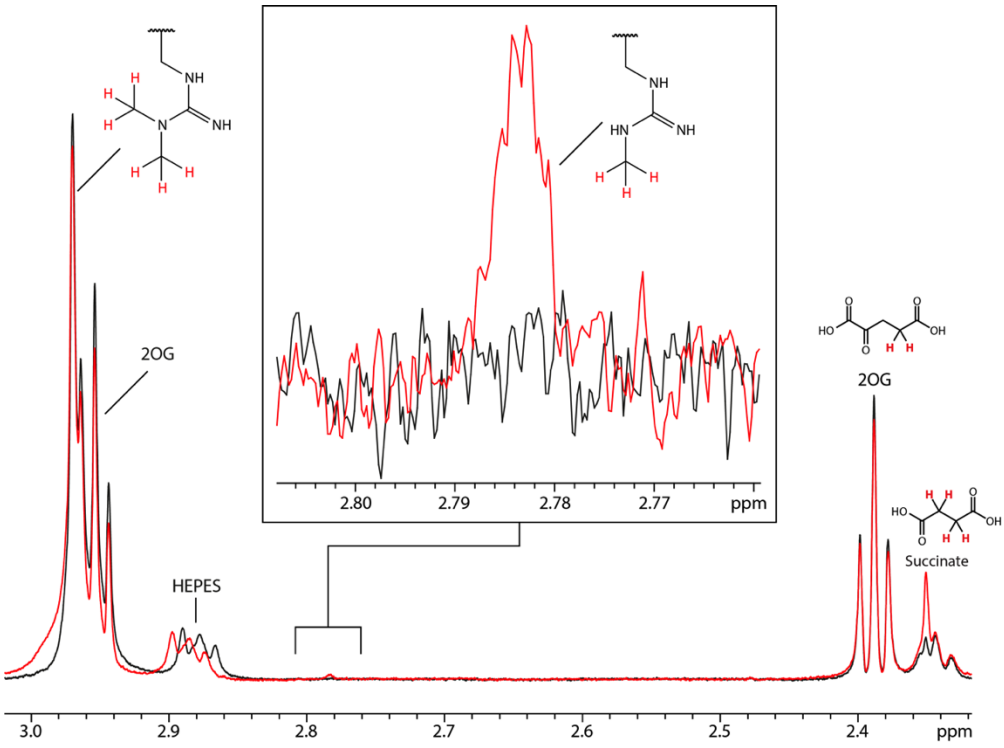


Figure 5.8 Arginine demethylation of an H3 R2Me$_{2a}$ 15mer peptide (100 μM) catalysed by JMJD2E (20 μM) studied by ^{1}H NMR spectroscopy (700 MHz). The black trace shows the reaction after 5 min and the red trace the reaction after 30 min. Protons responsible for the NMR signals are shown in red.

The FDH assay was then used to measure a K_M for the same H3 R2Me$_{2a}$ 15mer peptide with JMJD2E (2 μM) (Section 8.7.2). This gave a K_M*(peptide)* of 57 ± 16 μM and a k_{cat} of 0.023 ± 0.0014 s^{-1}. This K_M value is about three times higher than for the H3 K9Me$_3$ peptide (K_M = 17.4 ± 2.5 μM) and twice as high as the H3 K9Me$_2$ peptide (K_M = 32.2 ± 7.4 μM), suggesting that binding of this peptide to JMJD2E is not as efficient as binding of the published lysine methylated substrate. Likewise, the k_{cat} value determined is about three times lower than the k_{cat} value for the dimethylated lysine peptide (0.07 s^{-1}) suggesting the overall reaction with methylated arginine is slower than the reaction with lysine. This may be due to a less favourable binding orientation for oxidation or due to slower product release. In addition to providing kinetic parameters, this assay confirmed that formaldehyde is produced during the demethylation reaction.

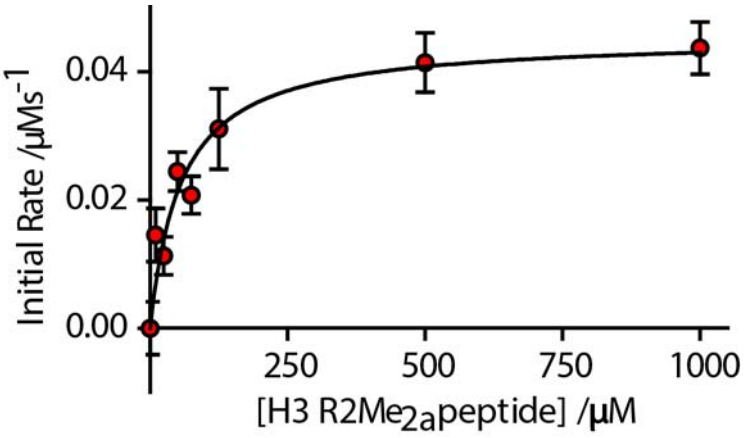


Figure 5.9 FDH assay measuring the kinetic parameters for an H3 R2Me$_{2a}$ 15 mer peptide with JMJD2E (2 μM). Points represent the mean of triplicate experiments and error bars show ± 1 s.d..

A competition assay was then carried out using a 1:1 mixture of the H3 R2Me$_{2a}$ 15mer peptide and an H3 K4Me$_3$K9Me$_2$ 15mer peptide (a peptide trimethylated at H3 K4 was used so that the two peptides had different masses and could be distinguished in the mass spectra) with JMJD2E (2 μM) and cofactors. The reaction was incubated at 37 °C and quenched with methanol at selected time points. Analysis of the

abundance of the different peptidic species by MALDI-TOF MS again showed that the rate of demethylation of the methylated lysine peptide was much faster than demethylation of the methylated arginine peptide (Figure 5.10, Section 8.7.1).

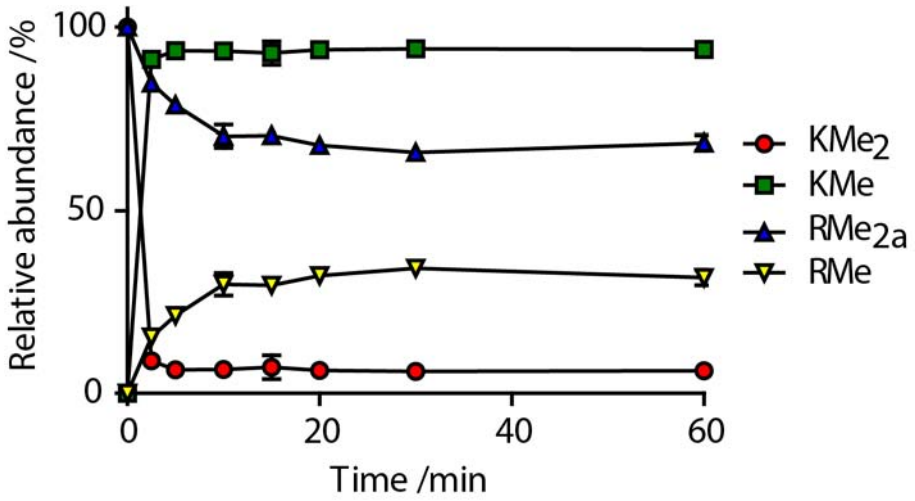


Figure 5.10 Competition between demethylation of an H3 $K4Me_3K9Me_2$ 15mer peptide and an H3 $R2Me_{2a}$ 15mer peptide by JMJD2E (2 μM) analysed by MALDI-TOF MS. Points represent the mean of triplicate experiments and error bars show ± 1 s.d..

5.5 Investigating the differential sequence preference of KDMs for arginine and lysine demethylation

Given that the H3 R2 methylated peptide appeared to be a much better substrate for JMJD2E than the mutant H3 K9R peptide based upon the sequence of its characterised lysine substrate peptide (H3 1-15 $K9Me_3$), it was envisaged that the sequence preference for the two different substrate residues (lysine *vs* arginine) might not be the same. The sequence preference for the two target residues also appeared to be different for JARID1C, which was observed to catalyse arginine demethylation of the H3 K9R peptide, where lysine demethylation of the biologically relevant H3 K9 peptide does not occur. To test this further, another variant peptide sequence was synthesised, substituting H3 R2 for a KMe_3 residue (*synthesised by Dr. Wei Ge,* $AKMe_3TKQTARKSTGGKA$). This peptide was incubated with JMJD2E (2 μM) or

JARID1C (2 μM) and cofactors for 1 hr at 37 °C and the quenched reaction analysed by MALDI-TOF MS. Although demethylation of the control H3 R2Me$_{2a}$ peptide occurred, no demethylation of the mutant H3 R2KMe$_3$ peptide was observed (Figure 5.11). This result again suggested that optimal binding of methylated lysyl- and arginyl-residues requires different neighbouring amino acid sequences.

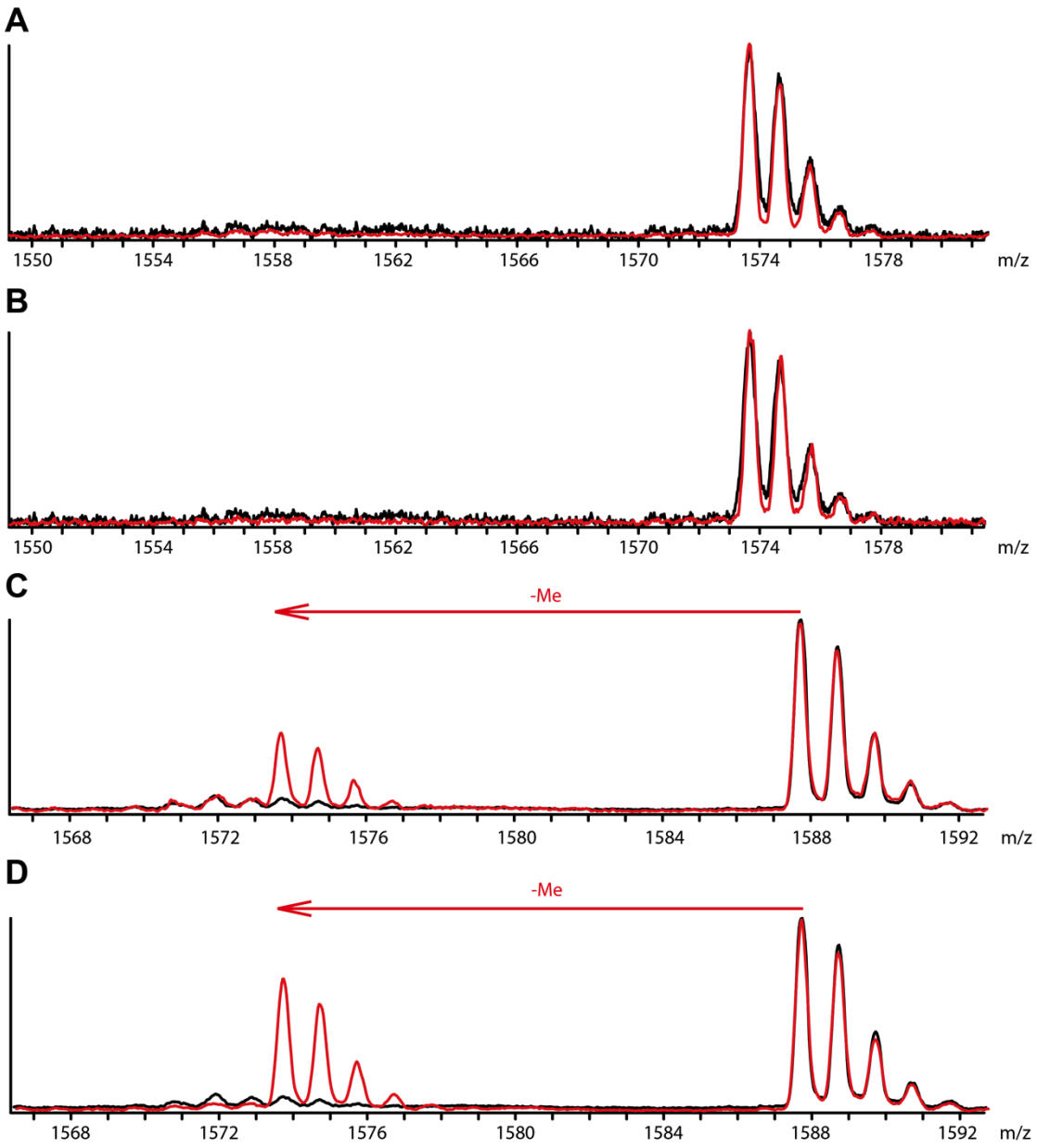


Figure 5.11 MALDI-TOF analysis of demethylation of an H3 R2KMe$_3$ 15mer peptide (10 μM, **A** and **B**) or an H3 R2Me$_{2a}$ 15mer peptide (10 μM, **C** and **D**) by JARID1C (2 μM, **A** and **C**) or JMJD2E (2 μM, **B** and **D**) after 1 hr at 37 °C. Control spectra in the absence of enzyme are shown in black and spectra containing enzyme in red.

5.6 Characterisation of arginine demethylation of H3 K27RMe$_{2a}$ by JMJD3

Characterisation of arginine demethylation of the H3 K27RMe$_{2a}$ 15mer peptide by JMJD3 was also carried out, as this appeared to be the most efficient of the demethylation reactions observed so far with any of the enzymes / peptide combinations. ^{1}H NMR spectroscopy was used to detect the demethylation product (*experiment carried out by Dr. Richard Hopkinson*, Section 8.15). JMJD3 (10 μM) was incubated with K27RMe$_{2a}$ 15mer peptide (1 mM) and cofactors at 25 °C and spectra collected every 3 min over 30 min. Production of a peak at δ_H = 2.78 ppm was observed to increase in a time-dependent manner and was assigned to the single methyl group of the monomethylarginine in the product peptide (Figure 5.12A). After 1 hour a multiplicity edited ^{1}H-^{13}C HSQC spectrum of the sample was collected. A peak was detected in the spectrum with δ_C = 27.5 ppm that was not observed in the spectrum of the starting peptide. This was similar to the predicted value for the monomethyl group of monomethylated arginine, further confirming that demethylation of the sample had occurred (prediction made using ACD prediction software).

To further confirm the production of formaldehyde in the NMR reaction, after 30 min, dimedone (1 mM) was added to the sample and a series of 1D spectra collected. Dimedone is known to form an adduct with formaldehyde, and the product can be detected by NMR spectroscopy (Figure 5.13A).[340,341] This showed loss of the dimedone signal at δ_H = 0.98 ppm and formation of signals around δ_H = 0.96 ppm and δ_H = 0.92 ppm, which can be assigned to the protons of the dimethyl groups of the

hydroxymethyldimedone adducts (Figure 5.13B).

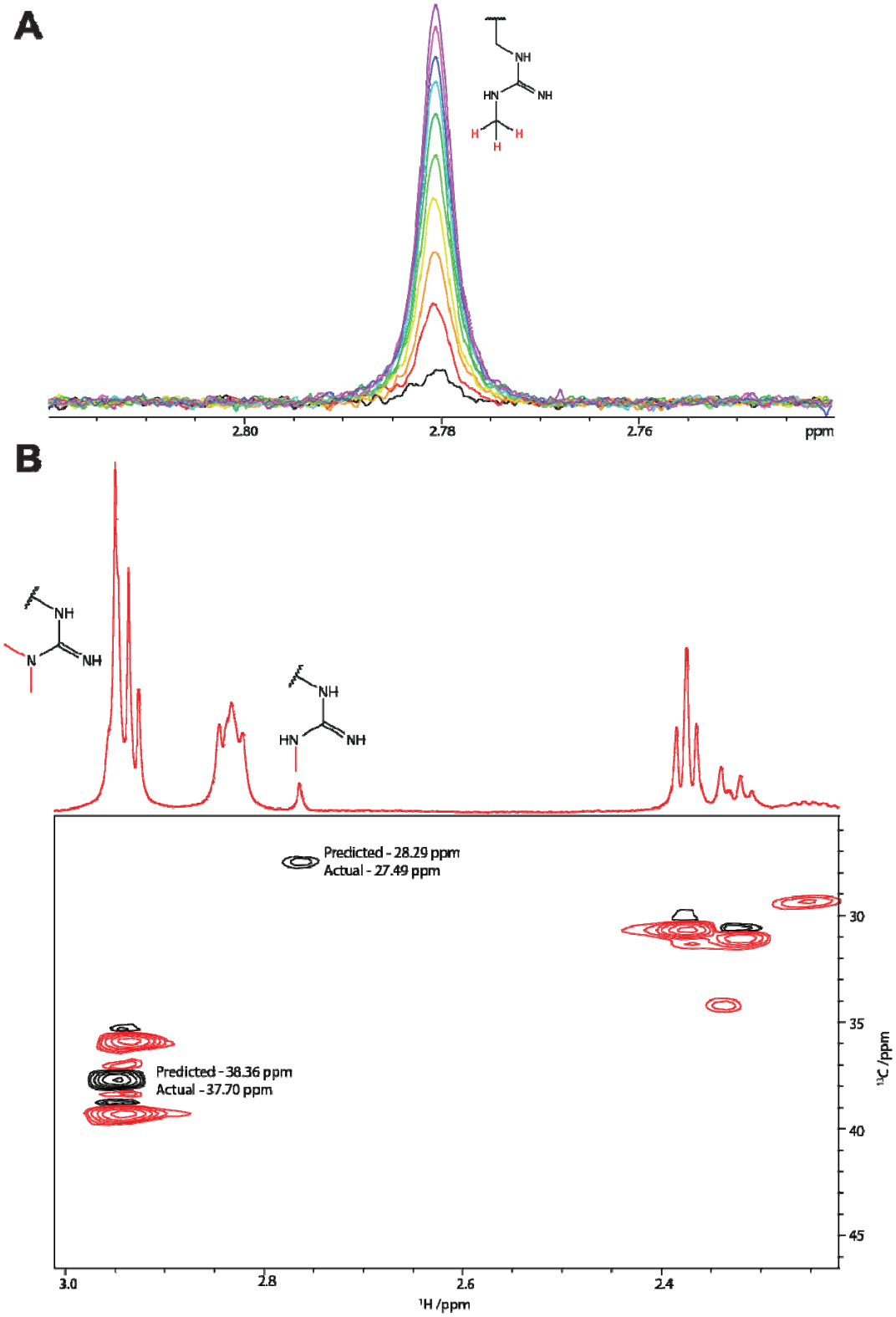


Figure 5.12 A) ^{1}H NMR spectra (700 MHz) of demethylation of an H3 K27RMe$_{2a}$ 15mer (1 mM) by JMJD3 (10 μM). Protons responsible for the NMR resonance are shown in red. Spectra were collected every 3 min over 30 min. The black spectrum shows the reaction after 90 s, and the final purple spectrum the reaction after 1710 s. **B)** 2D multiplicity edited ^{1}H-^{13}C HSQC of the same reaction after 1 hour. Carbon chemical shift values were predicted using ACD prediction software.

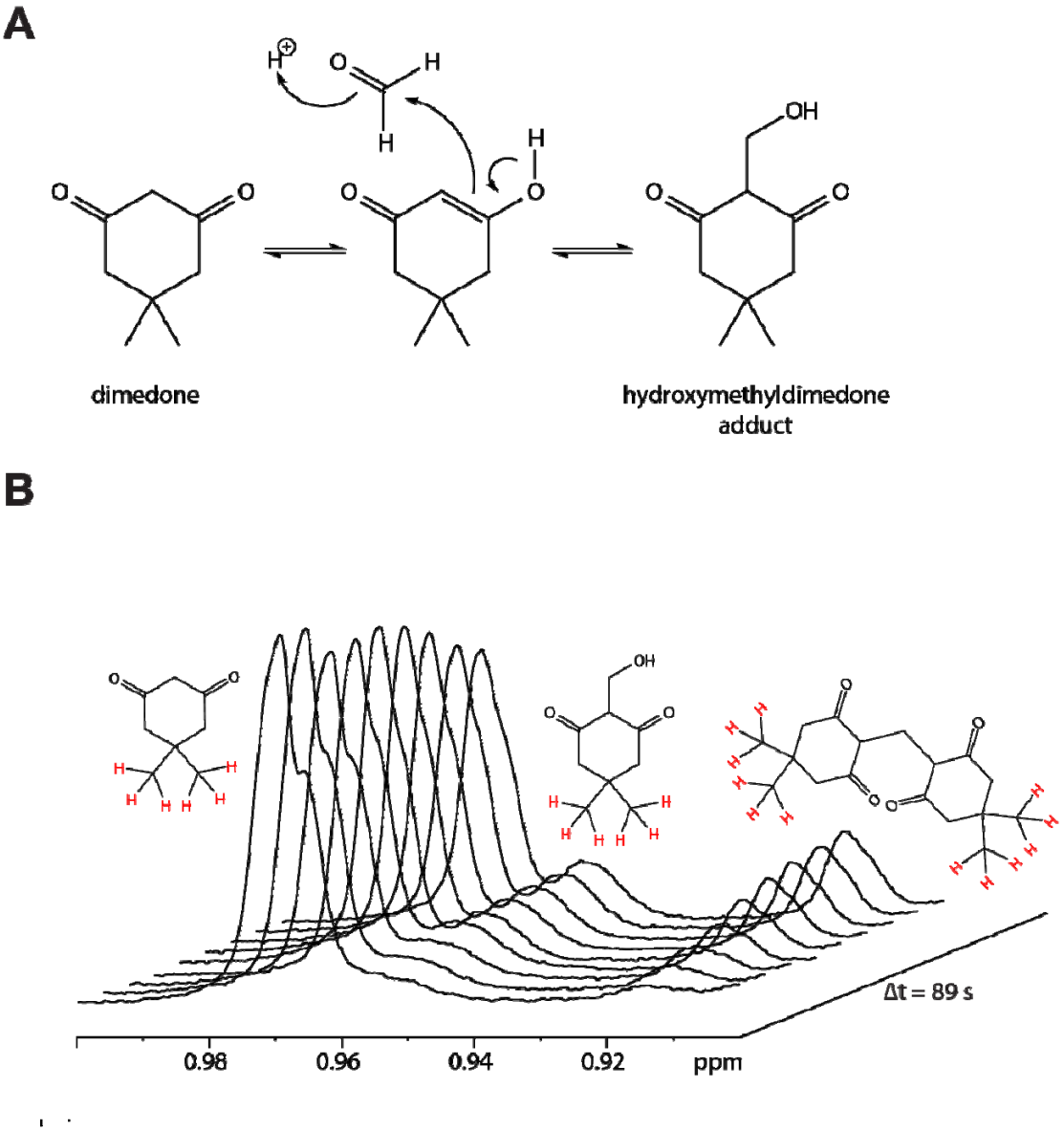


Figure 5.13 A) Reaction of dimedone with formaldehyde. **B)** ^{1}H NMR spectra (700 MHz) showing appearance of the hydroxylmethyldimedone adduct resonance in the reaction of JMJD3 (10 μM) with an H3 K27RMe$_{2a}$ 15mer peptide (1 mM) after 30 min incubation at 25 °C followed by addition of dimedone (1 mM).

Using the FDH assay, the kinetics of the demethylation reaction of JMJD3 with the

H3 K27RMe$_{2a}$ 15mer peptide were also determined (Section 8.7.2). A K_M of 68 ± 9.4 μM and a k_{cat} of 0.080 ± 0.0031 s^{-1} was measured using JMJD3 (2 μM) (Figure 5.14).

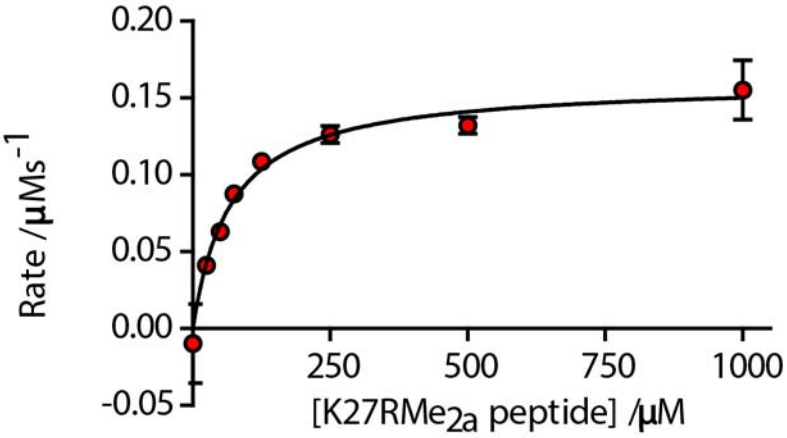


Figure 5.14 FDH assay measuring the kinetic parameters for an H3 K27RMe$_{2a}$ 15mer peptide with 2 μM JMJD3. Points represent the mean of triplicate experiments and error bars show ± 1 s.d..

5.7 JMJD2A crystallography with methylated arginine peptides

The variable sequence requirements for lysine demethylation compared to arginine demethylation suggested that the peptides might bind in different conformations. To investigate this, structural information on the binding of a methylated arginine peptide was required. JMJD2A was selected as a good target for peptide co-crystallography due to its propensity for crystallisation. Unfortunately, despite obtaining JMJD2A crystals, no co-crystals were observed (Appendix 3)

5.8 Arginine demethylation of full length histones

Although arginine demethylation has been demonstrated on histone peptides, which make good mimics for the unstructured tails of histone proteins, it would be desirable to confirm that demethylation of full length histone proteins is also possible. A method that does not require a specific antibody involves use of amino acid analysis.

Isotope ratio mass spectrometry (IRMS) can be used to accurately measure the quantity of labelled amino acids within a sample.[342]

5.8.1 Principles of isotope ratio mass spectrometry

IRMS is used to measure the abundance of different isotopes within a sample. Samples are separated by HPLC and then oxidised to form dinitrogen, sulphur dioxide, carbon dioxide and water. The work described here focused only on detection of carbon isotopes within samples. The IRMS uses three detectors to simultaneously measure the ion count for three masses, which for CO_2 are 44, 45 and 46. 44 corresponds to the mass of CO_2 containing the most abundant isotopes of carbon and oxygen (12 and 16 respectively), trace 45 corresponds to CO_2 containing one heavy isotope, ^{13}C or ^{17}O, and trace 46 corresponds to CO_2 containing two heavy isotopes, two ^{17}O or ^{13}C and ^{17}O. The ratio of the ion count for each trace gives the isotope ratio, which is a measure of the relative abundance of these different isotopes within a sample. IRMS is often used for carbon dating of archaeological samples, where the ^{13}C ratio is measured relative to an internationally defined standard, the Pee Dee Belemnite Fossil.[343]

5.8.2 IRMS experiments with histones

For histone demethylation experiments, HEK 293T cells were grown in Dulbecco's modified eagle's media (DMEM) containing ^{13}C labelled arginine or lysine in place of unlabelled arginine or lysine so that these were incorporated into all proteins produced by the cells and histones labelled with "heavy" lysine or arginine could be extracted (Section 8.12). Cells were maintained in labelled media for five cell doublings before

protein was harvested to ensure full incorporation of the labelled amino acids.

HEK 293T cells labelled with either heavy lysine or arginine were then harvested and the histones enriched by acid extraction following a literature procedure (Section 8.14);[344] After lysis, 0.4 M HCl was added to the nuclear pellet, which was incubated for 1 hr at 4 °C. This acted as an enrichment step, as basic histone proteins are soluble in the acid solution where most other proteins precipitate. The resulting mixture was centrifuged (14000 r.p.m., 20 min, 4 °C) and ice-cold acetone added to the supernatant. This was left at -20 °C overnight, causing the histones to precipitate. The resulting precipitate was run on an 18% SDS PAGE gel to visualise the histone enrichment (Figure 5.15). A high degree of enrichment of the four canonical histones and the linker H1 histone can be observed.

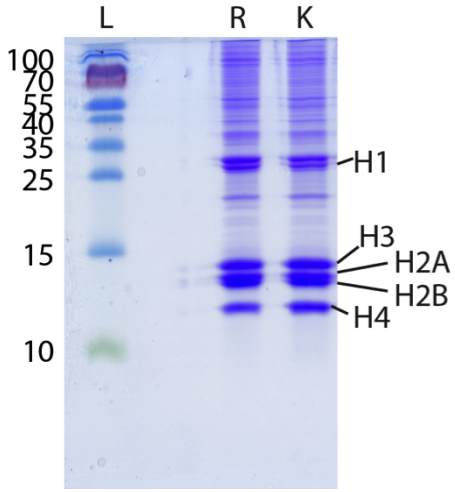


Figure 5.15 18% SDS PAGE gel showing enriched histones extracted from HEK 293T cells (10 cm plates) labelled with ^{13}C lysine or ^{13}C arginine. The contents of each lane of the gel are marked above it, with L denoting the ladder, R the sample containing "heavy" arginine-labelled histones and K the sample containing "heavy" lysine-labelled histones. Numbers to the left of each gel give the size of each molecular weight marker in kDa.

For analysis by IRMS, samples were hydrolysed to their constituent amino acids using 6 M HCl and injected onto the IRMS (*samples run by Dr. Nikita Loik*). Full labelling of the samples was confirmed by the presence of peaks in trace 45 without

corresponding peaks in trace 44 for modified forms of arginine in the "heavy" arginine-labelled samples and modified forms of lysine in the "heavy" lysine-labelled samples (Figure 5.16). The low natural abundance of ^{13}C (~1%) and ^{17}O (~0.04%) means that for all other non-labelled amino acids a 1:1 ratio of trace 44:trace 45 (the intensity of trace 45 is displayed multiplied by 100) is observed and trace 46 can effectively be ignored. The identification of each modified amino acid can be determined by reference to the elution time relative to standards of each residue (*Nikita Loik, unpublished data*). The acid hydrolysis causes loss of acetylation and phosphorylation, so that only methylated and hydroxylated post-translationally modified amino acids are observed.

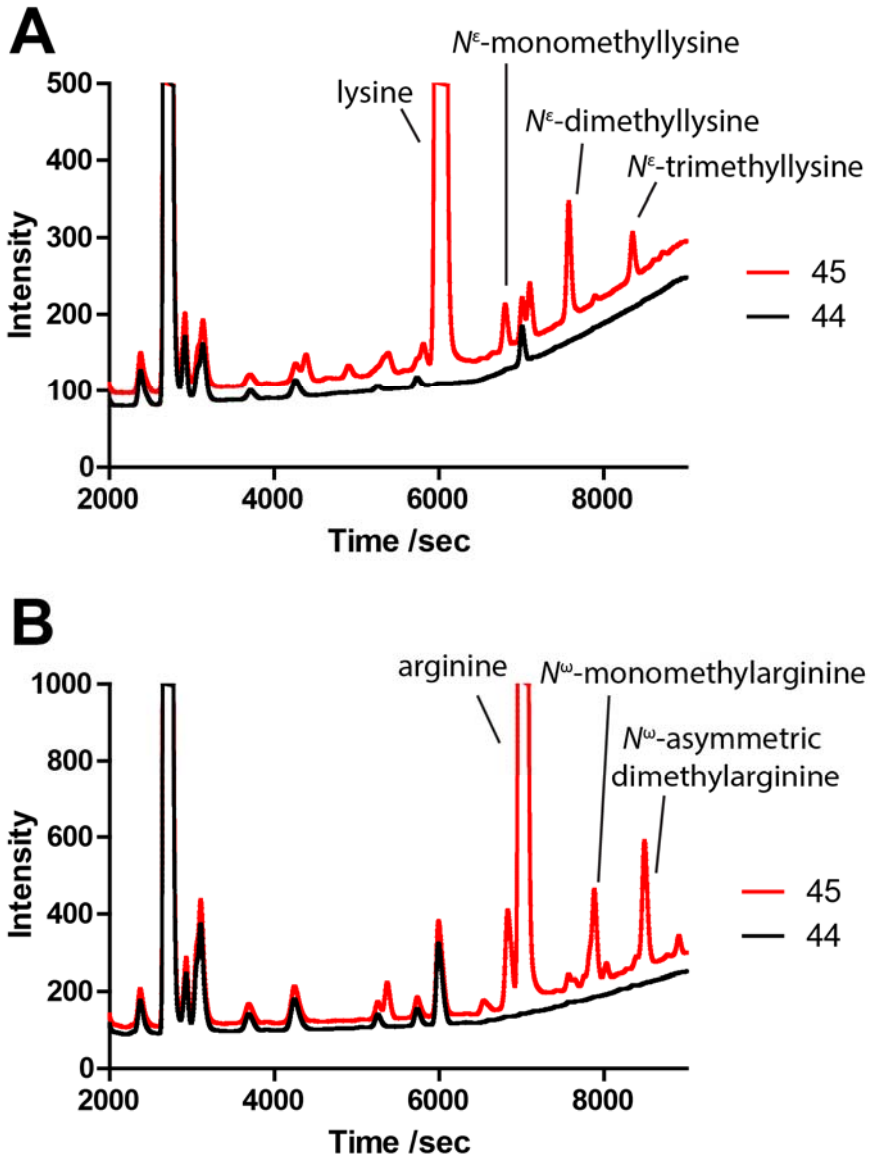


Figure 5.16 IRMS spectrum of bulk histones acid extracted from **A) "**heavy" lysine-labelled HEK 293T cells and **B)** "heavy" arginine-labelled HEK 293T cells. Trace 44 is shown in black and trace 45 multiplied by 100 in red. The identity of the amino acids is marked above each trace. Traces are shown here only for elution times between 2000 and 9000 s and not to the full intensity observed for unmodified lysine / arginine to allow visualisation of the modified amino acid peaks.

Having determined that modified lysyl- and arginyl-residues can be identified in histone samples extracted from cells it was next necessary to see whether demethylation of these samples by recombinant enzymes could be observed. Approximately 50 μg of "heavy" arginine-labelled histones were incubated with JMJD2E (10 μM) and cofactors (2OG (100 μM), Fe(II) (100 μM) and ascorbate (1 mM)). The samples were quenched by freezing in liquid nitrogen and then

hydrolysed to their constituent amino acids using 6 M HCl. The samples were analysed by IRMS (*samples run by Dr. Nikita Loik*). A change in the ratio of the methylated forms of arginine could be observed (before treatment of the histones with JMJD2E there is more RMe_{2a} in the cells than RMe, after treatment with JMJD2E there is less RMe_{2a} than RMe, Figure 5.17). The results indicated that demethylation of full length histones by JMJD2E was possible.

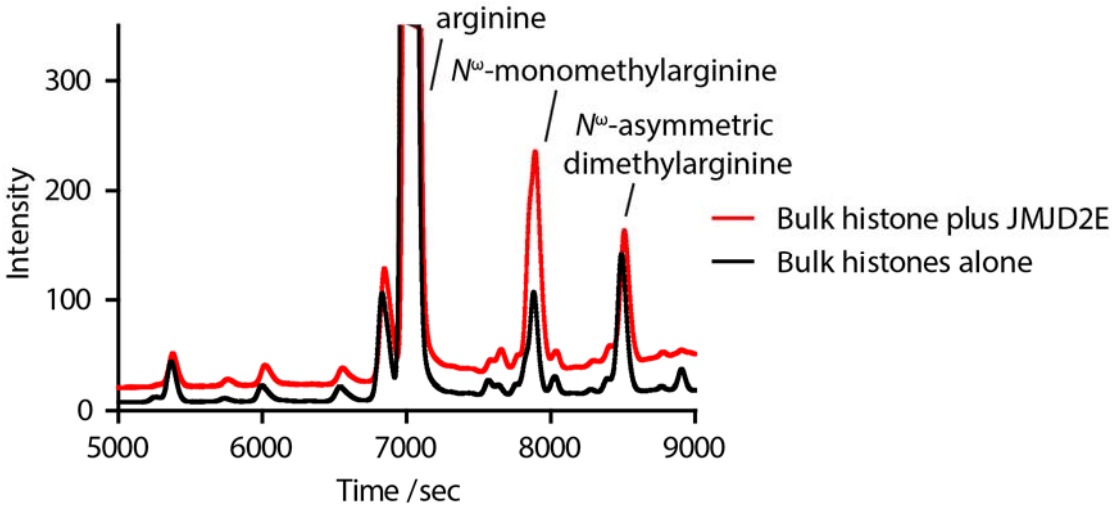


Figure 5.17 IRMS spectra of acid extracted "heavy" arginine-labelled histones (50 µg) from HEK 293T cells treated with JMJD2E (10 µM). Identity of peaks corresponding to methylated / unmodified arginine is marked above the traces. Only trace 45 is shown. The baseline is corrected by subtracting trace 44 from trace 45 and the data normalised between runs to the intensity of the unmodified arginine peak.

5.9 Identification of histone arginine demethylation in cells

The next goal was to investigate whether histone arginine demethylation can also occur within cells. The known activity of these enzymes as histone lysine demethylases complicates this. The intricate crosstalk of different histone modifications means that it may be difficult to prove conclusively that any change in arginine methylation levels are not caused by an indirect effect of changes in lysine methylation levels.

Initially attempts were made to extend the IRMS histone analysis to cellular samples. It was envisaged that labelled cells could be treated with a variety of conditions, such as overexpression or inhibition of JMJD2A or JARID1C and then histones could be extracted and the ratio of differently modified amino acids within the samples analysed. To this end, a method of accurately identifying the quantity of histones within each sample was required to allow comparison between different samples. As the histone extraction protocol only enriches for histones rather than purifying them exclusively a further purification step was required. Excision of bands corresponding to each histone from an SDS PAGE gel (as in Figure 5.15) followed by in-gel trypsin digest before amino acid hydrolysis was attempted but this led to significant sample loss and a high degree of variability between samples, precluding detection of small differences in methylation levels between different cellular treatments (data not shown).

Instead, histone separation by high performance liquid chromatography (HPLC) was attempted, follow a literature procedure (*HPLC optimisation carried out by Dr. Nikita Loik and Rok Sekirnik*).[344,345] Total ion chromatograms appeared to provide baseline separation of the histone proteins, however, when eluted fractions were separated by gel electrophoresis, fractions were shown to contain a mixture of different histones, precluding comparison between samples. To identify each peak on the HPLC trace, a mass spectrum was collected for each one. Interestingly, the mass spectra for each of the histone isoforms showed a series of peaks mostly separated by 14 Da, which corresponded to addition of methyl groups (Figure 5.18). Peaks separated by 42 Da corresponded either to addition of three methyl groups to the histones or one acetyl group, as these both have the same mass. No histone phosphorylation was observed.

As this mark is acid labile, it is possible the acid-based histone extraction resulted in its removal. The vast combination of post-translational marks possible on multiple residues of the H3 tail mean that each peak in the mass spectrum likely corresponds to many different combinations of histone marks on the tail. Dissection of the specific combinations of marks in each peak will be complex.

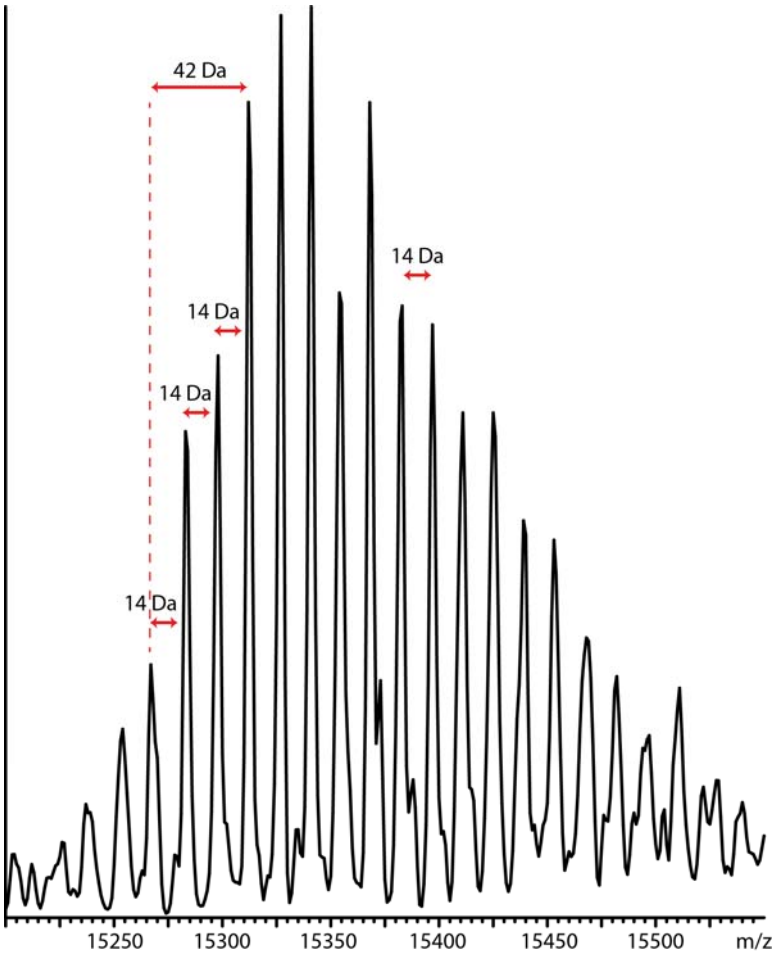


Figure 5.18 Mass spectrum of H3 from histones acid extracted from HEK 293T cells and separated by HPLC. Data was extracted using the MaxEnt feature of MassLynx v4.0.

It has previously been shown by whole protein MS that the pattern of histone modifications is altered on treatment of cells with HDAC inhibitors.[346] To see whether the global pattern of histone modifications was also altered upon treatment of cells with an inhibitor of histone demethylation, cells were treated with increasing concentrations of the pan-JmjC KDM inhibitor IOX1.[308] 24 hours after IOX1

treatment the cells were harvested. Histones were acid extracted from each cell sample and the samples analysed by MS (*samples run by Rok Sekirnik*). Interestingly, a concentration dependent shift in the modification pattern of H3 towards higher masses, indicative of higher methylation / acetylation states was observed with increasing concentrations of IOX1 (Figure 5.19). This was as expected for inhibition of demethylation by the JmjC KDMs. This technique could be extended to create an assay for measuring the global effect of inhibitors of chromatin modifying enzymes on the pattern of histone modifications.

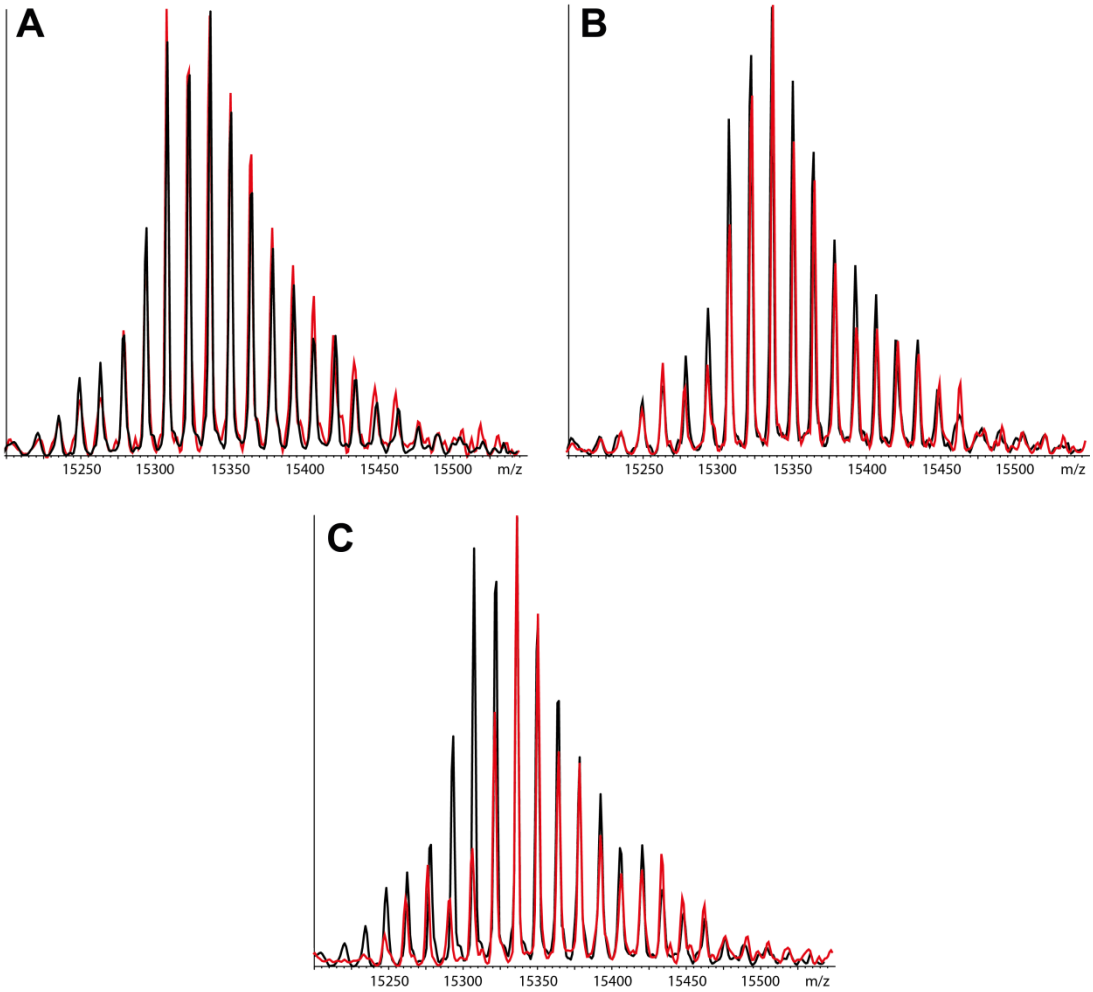


Figure 5.19 Mass spectra of H3 after UPLCMS of histones extracted from HEK 293T cells treated with increasing concentrations of IOX1. **A)** Overlay of DMSO treated cells (black) and 10 µM IOX1 treated cells (red). **B)** Overlay of 30 µM IOX1 treated cells (black) and 100 µM IOX1 treated cells (red). **C)** Overlay of MS from DMSO treated cells (black) and 300 µM IOX1 treated cells (red).

5.9.2 Immunofluorescence experiments using anti-dimethylarginine antibodies

Due to the difficulties of extending the IRMS method to analysis of cellular samples, immunofluorescence experiments with antibodies that recognise dimethylarginine were attempted as an alternative method for demonstrating histone demethylation of histones within cells (Section 8.12.3). Experiments were carried out with two antibodies, one specific for H3 R2Me_{2a} (ab 80075, Abcam) and a second specific for H4 R3Me_{2s} (ab5823, Abcam). Experiments were carried out in HeLa cells. Cells transfected with either wt or catalytically inactive FL Flag-JARID1A were fixed and probed with one of the anti-methylated arginine antibodies (or an anti-H3 K4Me_3 antibody as a positive control) and an anti-Flag antibody that recognised the Flag-JARID1A. Unfortunately no evidence for demethylation was observed in either of the samples probed with antibodies recognising the methylated arginine, as would have been evidenced by a decrease in the red staining, although the positive control with the anti-H3 K4Me_3 antibody did show a decrease of staining in the cells that were exogenously overexpressing wt FL Flag-JARID1A, indicative of K4 demethylation (some example images are shown in Figure 5.20).

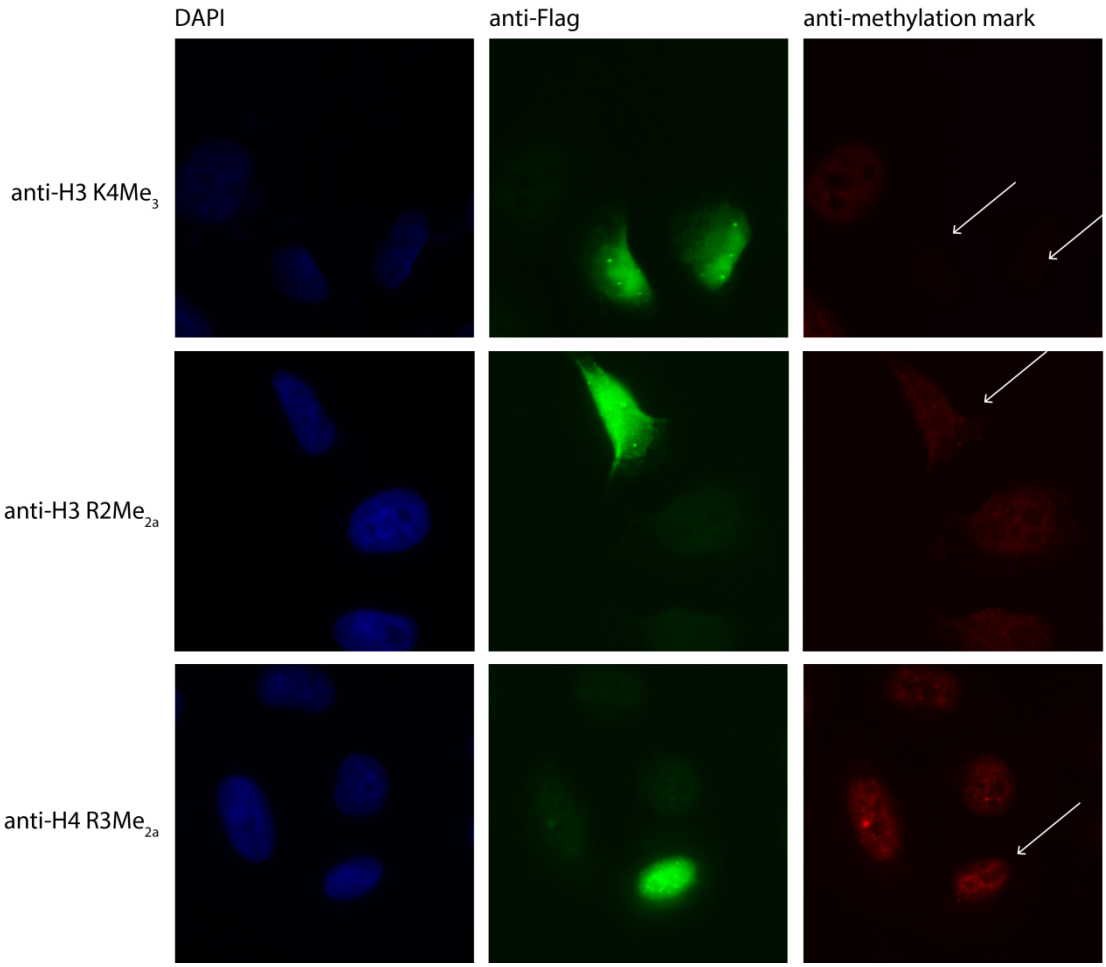


Figure 5.20 Immunofluorescence of HeLa cells overexpressing wt FL Flag-JARID1A and probed with antibodies to arginine or lysine methylation. Identity of the primary antibody to methylation is marked to the left of each row of images and the identitiy of the staining above each image. White arrows show cells that are overexpressing FL Flag-JARID1A.

Though no demethylation was observed using these antibodies, it was possible that this was due to cross-reaction with other proteins / marks within the cell. To validate if they could be used to measure changes in methylation levels, HeLa cells were seeded and treated with AMI-1 (either 50 μM or 100 μM), a known inhibitor of the PRMT enzymes.[347] After 24 hours these samples were fixed and stained in the same manner as the cells transfected with histone demethylases. As for JARID1C, no change in methylated arginine levels was observed (as indicated by levels of red staining, data not shown), suggesting that the antibodies tested were unsuitable for detecting histone demethylation by immunofluorescence.

5.10 Identification of non-histone substrates for the JmjC KDMs

5.10.1 Arginine methylation of p53

JMJD3 interacts directly with the tumour suppressor protein p53.[348,349] Although JMJD3 increases the nuclear localisation of p53 in a demethylase-activity-dependent fashion, the mechanism of this effect is currently unclear. The DNA damage response of p53 is modulated by its arginine methylation status, and it was therefore hypothesised that JMJD3 might accept p53 as a substrate for arginine and / or lysine demethylation.[350] Peptides corresponding to the known sites of arginine and lysine methylation on p53 were synthesised and tested as substrates of JMJD3.

Table 5.5 Peptide sequences synthesised from p53.

Peptide name	Peptide sequence
p53 1	APRMPEAAP**RMe$_{2a}$** VAPAP
p53 2	EDPGPDEAP**RMe$_{2a}$**MPEAA
p53 3	KTYQGSYGF**RMe$_{2a}$**LGFLH
p53 4	VDSTPPPGT**RMe$_{2a}$**VRAMA
p53 5	STPPPGTRV**RMe$_{2a}$** AMAIY
p53 6	SQHMTEVV**RMe$_{2a}$**RCPHHE
p53 7	SQHMTEVVR**RMe$_{2a}$**CPHHE
p53 8	FTLQI**RMe$_{2s}$**GRERFE
p53 9	LQIRG**RMe$_{2s}$**ERFEMF
p53 10	IRGRE**RMe$_{2s}$**FEMFRE
p53 11	MGGMN**RMe$_{2a}$**RPILTI
p53 12	MGGMN**RMe$_{2s}$**RPILTI
p53 13	AHSSHLKS**KMe$_3$**KGQSTSRH
p53 14	AHSSHLKS**KMe$_2$**KGQSTSRH

The 14 peptides (10 μM) were incubated with JMJD3 (1 μM) and cofactors for 1 hr at 37 °C and demethylation analysed by MALDI-TOF MS (Section 8.7.1). No demethylation was observed with any peptides screened against JMJD3. Using the same conditions the p53 peptides were also incubated with JMJD2E (more recently JMJD2A and JMJD2D have also been shown to interact with p53)[351,352] and JARID1C, but neither of these proteins demethylated any of the peptides.

5.10.2 Peptide screening

In an attempt to identify other substrates for the identified KDM arginine demethylases, methylated peptides from a number of abundant nuclear proteins or proteins known to interact with the relevant enzymes were synthesised including peptides from proteins such as BRCA1, 53BP1 and p300 (Table 5.6).

All peptides (10 μM) were incubated with the KDM (JMJD2E (1 μM) / JARID1C (1 μM) / JMJD3 (1 μM)) and cofactors for 1 hr at 37 °C and demethylation analysed by MALDI-TOF MS (Section 8.7.1). No demethylation was observed with any peptides screened against JMJD3. However, demethylation of peptides from BRCA1 (BRCA1 1, Table 5.6), hnRNP K (hnRNP K 8 and low levels of hnRNP K 9, Table 5.6) and fragile X mental retardation protein 1 (FMR1 1,2 and 3, Table 5.6) was observed with JMJD2E and from 53BP1 (53BP1 1,2 and 3, Table 5.6) with JARID1C. Example spectra for demethylation of hnRNP K 8 and BRCA1 1 are shown in Figure 5.21

Table 5.6 Peptides synthesised for arginine demethylation screening. The target arginine residues are shown in bold.

Protein	Peptide name	Peptide sequence
NF-κB	NF-κB 1	RSDLLEMAR**RMe$_{2a}$**SGPAL
	NF-κB 2	KTLLLDMAR**RMe$_{2a}$**SSPAA
BRCA1	BRCA1 1	APKKN**RMe$_{2a}$**LRRKSSTRH
BCR	BCR 1	SMYEDIS**RMe$_{2a}$**GLQGTY
BAD	BAD1	GAVEI**RMe$_{2a}$**SRHSSYPAG
	BAD2	GAVEIRS**RMe$_{2a}$**HSSYPAG
	BAD3	GEEPSPF**RMe$_{2a}$**GRSRSAP
	BAD4	GEEPSPFRG**RMe$_{2a}$**SRSAP
SAF-A	SAF-A 1	NRGNYNQNF**RMe$_{2a}$**GRGNNRG
	SAF-A 2	NRGNYNQNFRG**RMe$_{2a}$**GNNRG
	SAF-A 3	NRGNYNQNFRGRGNN**RMe$_{2a}$**G
	SAF-A 4	N**RMe$_{2a}$**GNYNQNFRGRGNNRG
FMR1	FMR1 1	GGRGQGG**RMe$_{2a}$**GRGGGFK
	FMR1 2	GGRGQGGRG**RMe$_{2a}$**GGGFK
	FMR1 3	RRPGG**RMe$_{2a}$**GRSVSGGRG
E2F1	E2F1 1	LAESSGPARG**RMe$_{2s}$**GRHPGKGVKS
	E2F1 2	LAESSGPARGRG**RMe$_{2s}$**HPGKGVKS
	E2F1 3	LAESSGPARG**RMe$_{2s}$**G**RMe$_{2s}$**HPGKGVKS
53BP1	53BP1 1	GKAPVTP**RMe$_{2a}$**GRGRRGRPPSRTTG
	53BP1 2	GKAPVTPRG**RMe$_{2a}$**GRRGRPPSRTTG
	53BP1 3	GKAPVTPRGRG**RMe$_{2a}$**RGRPPSRTTG
HOXA9	HOXA9 1	IKPEPLSA**RMe$_{2s}$**RGDCPT
hnRNP K	hnRNP K 1	RPMPPS**RMe$_{2a}$**RDYDD
	hnRNP K 2	RPMPPS**RMe$_{2s}$**RDYDD
	hnRNP K 3	RPMPPS**RMe**RDYDD
	hnRNP K 4	GRGGRPMPPS**RMe$_{2a}$**RDYDDMSPRR
	hnRNP K 5	RMPPG**RMe$_{2a}$**GGRPMPPSR
	hnRNP K 6	VGFPMRG**RMe$_{2a}$**GGFDRMP
	hnRNP K 7	VGRPM**RMe$_{2a}$**GRGGFDRMP
	hnRNP K 8	G**RMe$_{2a}$**GGRGGSRARNLPL
	hnRNP K 9	GRGG**RMe$_{2a}$**GGSRARNLPL

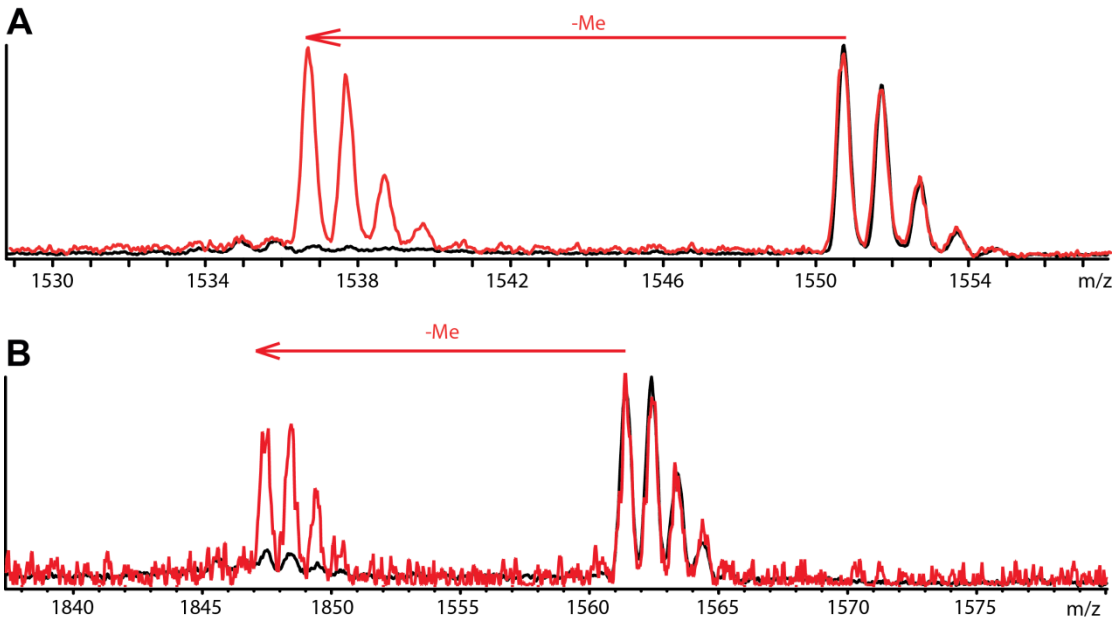


Figure 5.21 MALDI spectra of demethylation of **A)** hnRNP K 8 peptide (10 µM) by JMJD2E (1 µM) and **B)** BRCA1 1 peptide (10 µM) by JARID1C (1 µM). Red spectra show reaction with enzyme and black spectra control reactions in the absence of enzyme.

Although the peptide screening identified arginine demethylation on a number of peptide sequences, it was not clear whether demethylation of the same residues on full-length proteins could also occur. Unlike histone tails which are intrinsically disordered *in vivo* and where peptides thus make good substrate mimics, the secondary structure of most proteins means that although demethylation may occur on peptide fragments, when folded within a protein the correct conformation for active site binding and activity may not be possible. Chen *et al.* have demonstrated that the arginine methylation levels of hnRNP K decrease after DNA damage, although the protein levels do not, suggesting that active demethylation may be taking place.[353] As such, hnRNP K was chosen as the initial target for full-length protein work.

5.10.3 Arginine demethylation of hnRNP K

hnRNP K (heterogenous nuclear ribonucleoprotein K) is a member of a large family of hnRNPs.[354] hnRNPs are RNA binding proteins which all contain RNA binding motifs and other functional domains, many of which facilitate protein-protein

interactions. These form large complexes containing multiple hnRNPs in addition to RNA. hnRNP K was the first hnRNP to be functionally characterised.[355] It was determined to be a transcription factor, which binds preferentially to single-stranded DNA. As for a number of the hnRNPs, asymmetric arginine methylation has been identified on multiple residues of hnRNP K.[356–359]

To test whether demethylation of FL hnRNP K occurs, the DNA encoding for the full length sequence of hnRNP K was amplified (from image clone 2964383) and inserted into both a 6xHis-tagged vector for expression in *E. coli* and an *N*-terminally GFP-tagged vector for expression in mammalian cells (Section 8.6). Recombinant hnRNP K was expressed in BL21(DE3) *E. coli* and purified initially using a 5 mL HisTrap column (Section 8.4 and 8.5). A significant amount of protein was observed to elute in the wash step with buffer containing 40 mM imidazole, so in future purifications the column was washed extensively with lysis buffer only (containing only 20 mM imidazole) (Figure 5.22A). Pooled fractions (1-7) were concentrated and loaded onto a 300 mL S200 size exclusion column (Figure 5.22B). Fractions 38 – 44 were pooled and concentrated before being flash frozen using liquid nitrogen and stored at -80 °C for future use.

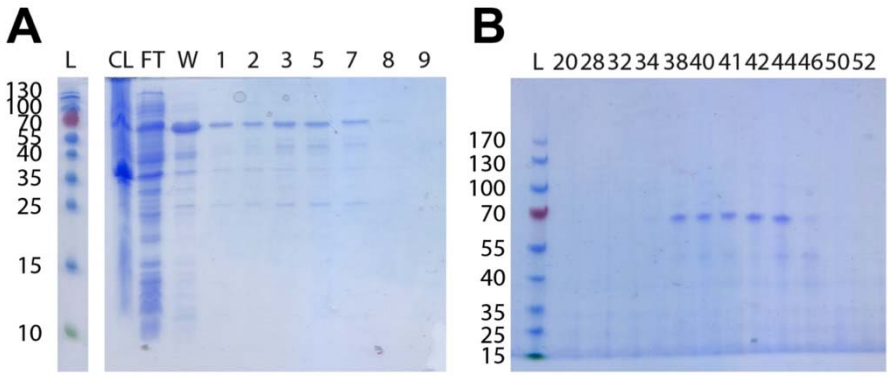


Figure 5.22 SDS PAGE gels of fractions from the purification of recombinant hnRNP K using **A)** a 5 mL HisTrap column, followed by **B)** a 300 mL S200 size exclusion column. The contents of each lane are labelled above the gels. L denotes the ladder, CL the cell lysate, FT the flowthrough, W the wash and numbers the fraction numbers. Numbers to the left of each gel give the size of each molecular weight marker in kDa.

5.10.4 Production of PRMT1 for *in vitro* protein methylation

PRMT1 is an asymmetric arginine methyltransferase that is responsible for approximately 85% of the arginine methylation within a human cell.[360,361] In order to facilitate *in vitro* substrate methylation reactions, DNA encoding for PRMT1 (residues 1 – 361) was amplified from a GFP-tagged vector (gift from Dr Frank Fackelmayer) and inserted into pNIC-BSA1 (including an *N*-terminal 6×His tag) to allow recombinant expression in *E. coli* (Section 8.6). The enzyme was successfully expressed in BL21 DE3 cells and purified using a 5 mL HisTrap column followed by a 300 mL S200 size exclusion column (Figure 5.23). Fractions 17-29 from the size exclusion column were pooled, concentrated, flash frozen and stored at -80 °C for future use (Section 8.5).

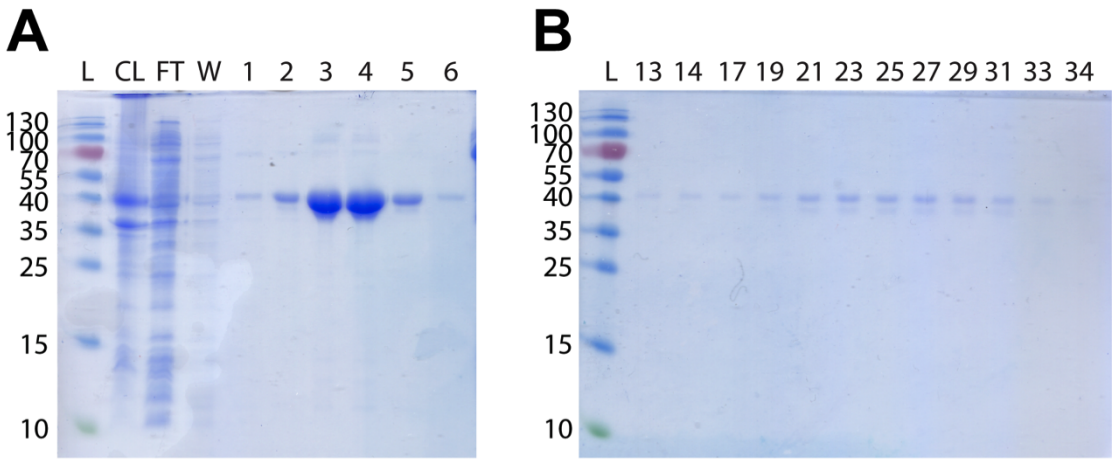


Figure 5.23 SDS PAGE gels from a purification of His-PRMT1 by **A)** 5 mL HisTrap column, followed by **B)** 300 mL S200 size exclusion column. Fractions run on the gel are marked above each gel, with L denoting the ladder, CL the cell lysate, FT the flow through, W the wash and numbers the fraction numbers. The size of each molecular weight marker in kDa is given to the left of the gels.

Activity of PRMT1 is most commonly determined by use of a radioactivity-based assay;[362] radiolabelled SAM is used as the PRMT cofactor so that methylation causes substrate peptides or proteins to become radioactive, and the level of radioactivity can be measured. To avoid use of radiation, if possible, a MALDI-TOF assay was attempted using peptide substrates and non-labelled SAM (S-(5'-adenosyl)-L-methionine *p*-toluenesulfonate salt, Sigma Aldrich) as the cofactor. Recombinant PRMT1 (1 μM or 0.1 μM) was incubated with SAM (50 μM or 5 μM, respectively) and a H3 1-19 or a H4 1-18 peptide (20 μM) in buffer containing 50 mM HEPES, pH 8.0, 0.1 mM EDTA and 150 mM NaCl (*peptides were provided by Dr. Wei Ge* from a library of unpurified peptides and thus contained a number of impurity peaks). Control reactions were carried out in the absence of SAM. As SAM is quite unstable at 37 °C the assay was carried out at 30 °C for 1 hr. Observation of peaks in the mass spectrum at +14 Da and +28 Da relative to the substrate peaks indicated that the recombinant PRMT1 was active and a MALDI assay could be used to detect methylation activity (Figure 5.24, Section 8.7.1). Significant activity was observed

with the H4 peptide substrate and small amounts with the H3 peptide.

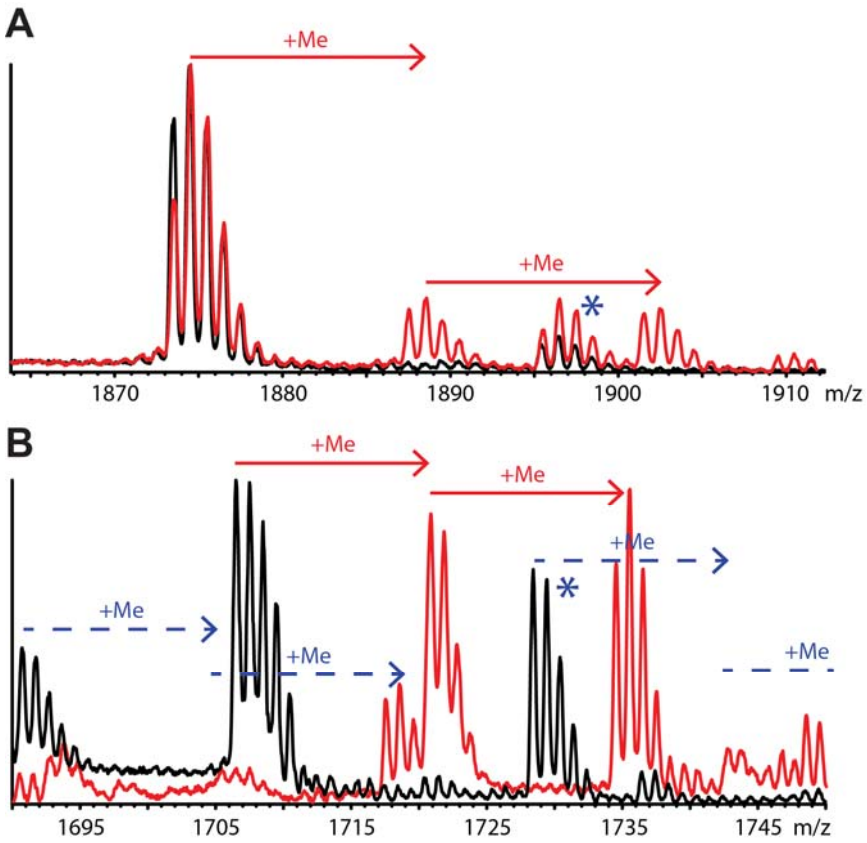


Figure 5.24 Methylation of peptides by recombinant PRMT1 (1 μM) after 1 hr incubation at 30 °C studied by MALDI-TOF MS. **A)** H3 1-19 peptide (20 μM) and **B)** H4 1-18 peptide (20 μM). Spectra of control reactions carried out without SAM are shown in black and spectra containing PRMT1 are shown in red. Blue stars mark impurity peaks and blue dashed lines show methylation of impurity peaks by PRMT1.

5.10.5 Effect of arginine methylation on the cellular localisation of hnRNP K

hnRNP K is methylated by PRMT1 on a number of residues. Arginine methylation levels have been reported to affect the localisation of the protein within the cell, causing shuttling between the nucleus and the cytoplasm.[363] In a second report, however, no such effect was observed.[359] If arginine methylation levels of hnRNP K do alter its cellular localisation, this phenomenom could be exploited as an assay to identify demethylation of the protein by JMJD2A.

Using immunofluorescence, the localisation of exogenously overexpressed hnRNP K in the presence of overexpressed FL wt Flag-JMJD2A, FL H188A catalytically inactive Flag-JMJD2A or an inhibitor of arginine methylation, adenosine-2,3-dialdehyde (AdOx, which is an inhibitor of AdoHyc hydrolyase, thus resulting in elevated levels of AdoHyc (S-adenosylhomocysteine)[364] in cells, which inhibits the PRMTs), was studied (Section 8.12.3). Cells were transfected with the relevant proteins, or treated with AdOx (100 μM) for 24 hours and then fixed, permeabilised and probed with primary and secondary antibodies. Although the GFP-tag of hnRNP K could be detected directly, an anti-Flag antibody was required to identify which cells were also overexpressing FL JMJD2A. Unfortunately, in all cases, nuclear localisation of GFP-hnRNP K was observed with no change in localisation under any of the treatment conditions (Figure 5.25). As the methylation inhibitor, AdOx, did not affect localisation (suggesting that under these experimental conditions localisation of hnRNP K was not being affected by its arginine methylation levels), it was not possible to conclude anything about the occurrence of arginine demethylation by JMJD2A in cells.

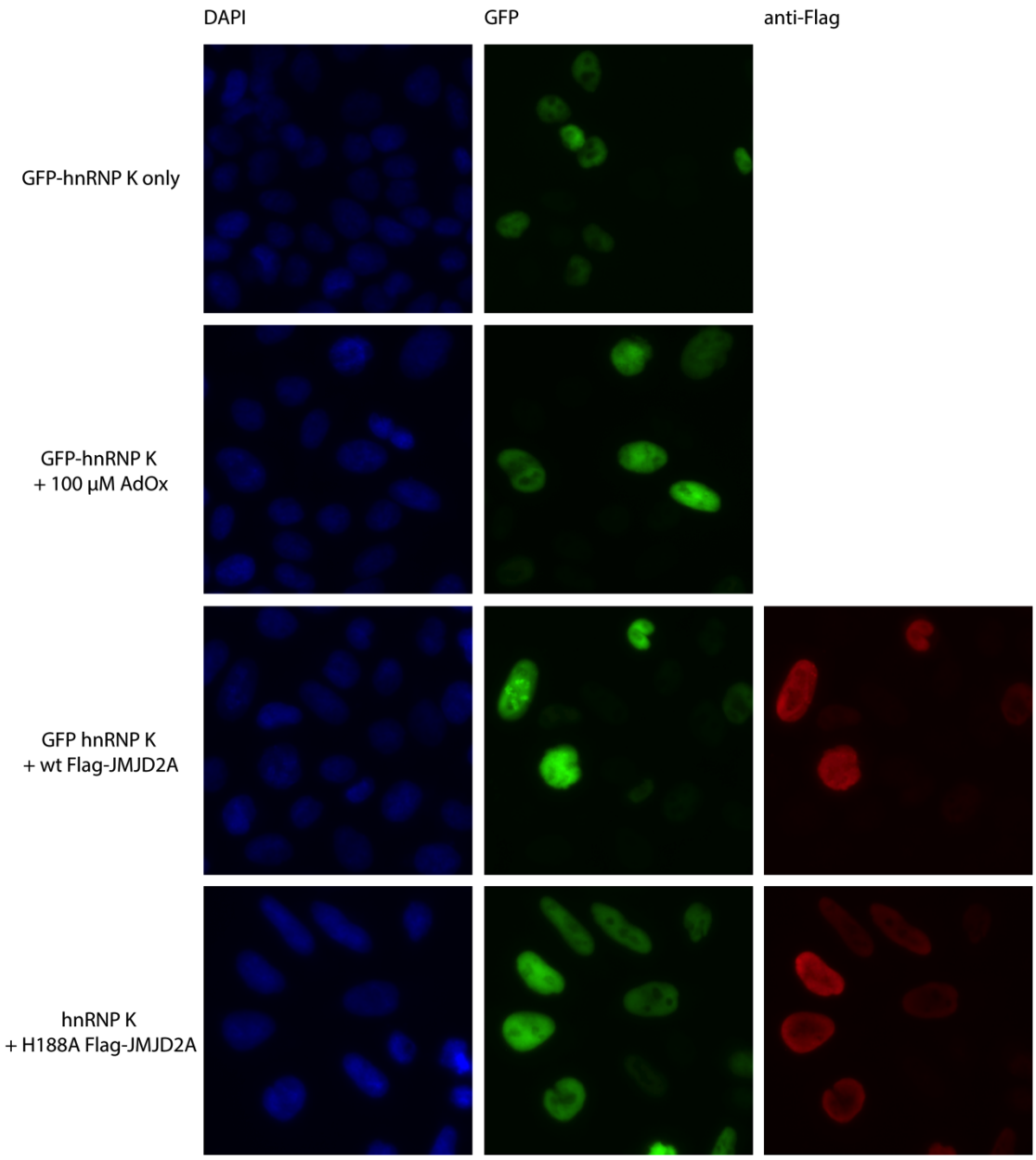


Figure 5.25 Immunofluorescence experiments showing the localisation of hnRNP K in HEK 293T cells under a variety of treatment conditions (described to the left of each row of images). Identity of the staining is marked above each column of images.

5.10.6 *In vitro* demethylation of FL hnRNP K

Next, *in vitro* methylation of recombinant hnRNP K was attempted using recombinant PRMT1. hnRNP K (20 µM) was incubated with PRMT1 (2 µM) and SAM (50 µM) for 30 min at 30 °C. Analysis of the reaction was carried out using a dot blot probed

with one of two antibodies that should recognise arginine methylation (ab107514, pan-RMe_2 and ab80075, anti-H3 $R2Me_{2a}$, Section 8.9). The membrane was spotted with the reaction mixture and each protein alone (PRMT1 only and hnRNP K only) at three different concentrations. In the case of ab107514 a signal was observed for the sample incubated with PRMT1 that was not present in any of the control samples, suggesting that it was recognising methylation of the hnRNP K and that methylation of the recombinant hnRNP K had been successfully acheived (Figure 5.26A). The other antibody, ab80075, could not be used for this purpose as the dot blot indicated that it also recognised PRMT1 even in the absence of hnRNP K (Figure 5.26B). The detection of PRMT1 by ab80075, which should be selective for H3 $R2Me_{2a}$ provided further confirmation that it was unsuitable for measuring H3 R2 methylation levels by immunofluorescence.

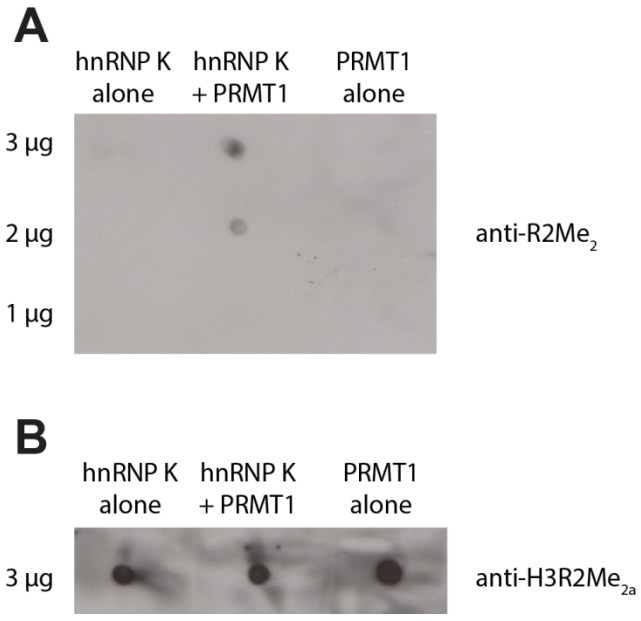


Figure 5.26 Dot blot analysis of methylation of recombinant hnRNP K by recombinant PRMT1 probed with **A)** ab107514 or **B)** ab80075. Reactions spotted in each column are marked above each membrane, amount of material spotted for each dot is given on the left of each gel and the identity of the antibody on the right.

5.10.7 Demethylation of FL hnRNP K by JMJD2E

To test for demethylation of FL hnRNP K by recombinant JMJD2E, the same experiment was repeated using twice the concentration of PRMT1 and hnRNP K. After methylation by PRMT1, the methylated hnRNP K was buffer exchanged into 50 mM HEPES, pH 7.5, 200 mM NaCl (hnRNP K precipitated in the absence of NaCl), to remove the SAM and EDTA. The resulting protein solution was incubated with recombinant JMJD2E (2 μM), ascorbate (100 μM), Fe(II) (10 μM) and 2OG (20 μM) for 1 hr at 37 °C to allow demethylation of the recombinantly methylated hnRNP K by JMJD2E. Control reactions were carried out in the absence of JMJD2E. A dot blot was again carried out, including a sample of JMJD2E alone to confirm that the antibody does not recognise it. Unfortunately, no decrease in signal was observed on incubation with recombinant JMJD2E (Figure 5.27).

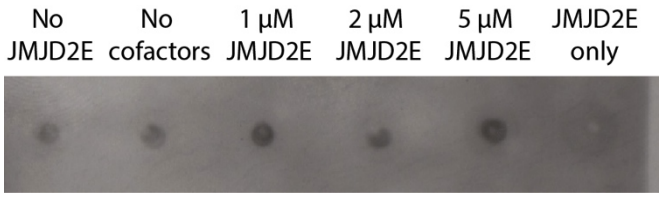


Figure 5.27 Dot blot analysis of demethylation reaction of hnRNP K by JMJD2E probed with ab107514. Reaction spotted in each position of the membrane is shown above the blot.

Although these results may be indicative of an absence of demethylation activity of JMJD2E for the FL folded protein, PRMT1 can methylate hnRNP K at a number of sites and it was also possible that the methylation detected by the antibody and observed on the dot blot was not at the residue that can be demethylated by JMJD2E. In addition, the antibody may also recognise the monomethylated protein (which is a much poorer substrate for JMJD2E), so that a signal change may not be observable even if demethylation has taken place.

5.10.8 Identification of hnRNP K demethylation by mass spectrometry

Quantitative mass spectrometry was then used to investigate potential demethylation of R296 in full-length hnRNP K by JMJD2A/E. As is the case for much arginine methylation within the cell, the peptide of interest is located in a RGG motif, which is cut repeatedly by trypsin.[365] Chiou *et al*. have demonstrated that peptides containing the target arginyl-residue can be detected after digestion with trypsin, as a result of missed cleavages, likely due to the reduced activity of trypsin at methylated arginyl-residues.[357] Using recombinant hnRNP K, digestion conditions were tested with three different digestion enzymes, trypsin, proteinase K and Asp-N (Section 8.13.2). In the case of trypsin, variable incubation times were used to maximise the likelihood of achieving the missed cleavages required for detection of the peptide of interest. The predicted digestion sites for each enzyme around the target arginine residue are shown in Figure 5.28.

1 METEQPEETFPNTETNGEFGKRPAEDMEEEQAFKRSRNTDEMVELRILLQSKNAGA

57 VIGKGGKNIKALRTDYNASVSVPDSSGPERILSISADIETIGEILKKIIPTLEEGLQLPSPT

119 ATSQLPLESDAVECLNYQHYKGSDFDCELRLLIHQSLAGGIIGVKGAKIKELRENTQT

177 TIKLFQECCPHSTDRVVLIGGKPDRVVECIKIILDLISESPIKGRAQPYDPNFYDETYDY

237 GGFTMMFDDRRGRPVGFPM**R**G**R**GGFDRMPPG**R**GGRPMPPSRRDYDDMSPR**R**G

289 PPPPPPG**R**GG**R**GGS**R**ARNLPLPPPPPPRGGDLMAYDRRGRPGDRYDGMVGFSAD

343 ETWDSAIDTWSPSEWQMAYEPQGGSGYDYSYAGGRGSYGDLGGPIITTQVTIPKDL

399 AGSIIGKGGQRIKQIRHESGASIKIDEPLEGSEDRIITITGTQDQIQNAQYLLQNSVKQYA

460 DVEGF

Figure 5.28 Amino acid sequence of hnRNP K. Residue numbers are shown to the left of the figure. Arginine residues known to be methylated are shown in bold and the target arginine for demethylation by JMJD2A is shown in red. Sites of predicted cleavage by Asp-N are shown by dashed purple lines, sites for proteinase K by dashed orange lines and sites for cleavage by trypsin as dashed green lines. The most suitable peptide for analysis found by partial trypsin digest is underlined in red.

In order to produce methylation, the hnRNP K was first incubated with recombinant PRMT1. Following this, digestion reactions were carried out following the manufacturers' protocols for each enzyme. Peptides containing the methylated arginine of interest (R296, Figure 5.28) could be detected after digestion with proteinase K and the shorter incubations with trypsin (Section 8.13.3). As increased trypsin digestion time appeared to slightly decrease the intensity of the desired peptide peak, 30 min was chosen as the optimal digestion time.

It was then necessary to test whether the same peptide could be detected from protein immunoprecipitated from cells, and if so, whether endogenous PRMT1 levels were high enough to methylate exogenously overexpressed hnRNP K. HEK 293T cells were seeded (15 cm dish) and transfected with GFP-hnRNP K (Section 8.12.2). After 48 hours the cells were harvested, lysed, and immunoprecipitated with GFP-Trap®

beads (Section 8.12.4). Protein was eluted from the beads using 0.2 M glycine, pH 2, and the sample digested in-solution for 30 min with trypsin (1:200). As for the recombinant protein, the desired peptide could be detected by LCMS/MS analysis and peptide intensities implied high levels of methylation at both R296 and R298 (Figure 5.29).

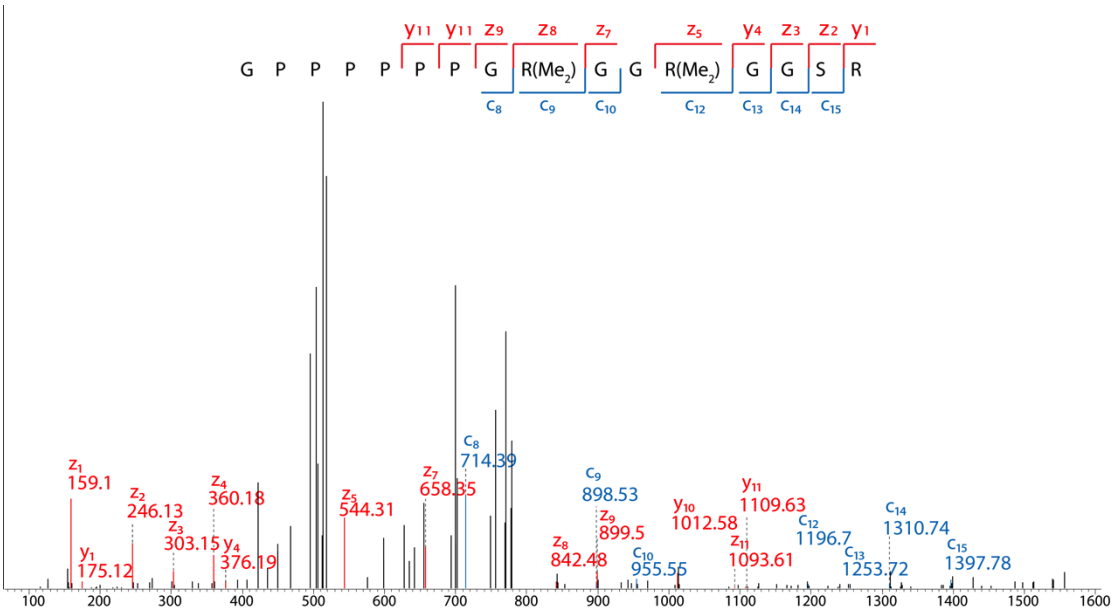


Figure 5.29 MS/MS spectrum of a peptide fragment of GFP-hnRNP K showing dimethylation of the target arginyl-residue. The peptide fragment is shown above the spectrum with red lines marking z and y ions and blue lines marking c ions.

SILAC (stable isotope labelling by amino acids in cell culture) is a technique used for performing quantification by proteomics.[366] As for the IRMS described above, cells are grown as two separate populations, one of which has some standard amino acids (usually lysine and arginine) replaced by isotopically-labelled amino acids so that they have a different, heavier mass. These amino acids are incorporated into proteins. Cells are passaged for at least five cell doublings to allow for complete labelling of proteins within the cells. Lysine and arginine are commonly used for labelling because peptides for the final proteomic analyses are most often produced by sample cleavage with trypsin, which cleaves at the *C*-terminal ends of lysine and arginine, meaning that all peptides will contain at least one of these two residues and will thus

all be "heavy". Peptides from the two cell populations are mixed and then processed and analysed together as the matching peptides will be separated by a mass shift corresponding to the number of lysyl- / arginyl-residues in that peptide. Where levels of the peptides are the same between samples, the ratio of these pairs of peaks will be one. However, if treatment of one of the two cell populations has altered the protein levels, this will be detectable by a change in the peptide ratio (allowing for relative quantification of protein levels or, when looking at specific peptides, the levels of post-translational modifications in cells after a variety of cellular treatments).[69,326] Dialysed fetal bovine serum (FBS) is used in the media to avoid contamination with unlabelled protein present in the FBS.

HEK 293T cells were labelled using $^{13}C_6{}^{15}N_4$-arginine and $^{13}C_6{}^{15}N_2$-lysine for five cell doublings (Section 8.12). After this time cells (10 cm dish) were harvested and lysed to check for complete labelling. Peptides produced by in-solution trypsinolysis were analysed by LCMS/MS (*sample run by Jacob Bush*) and the ratio of heavy:light peptides was extracted (Section 8.13). As expected, complete labelling (> 99%) was observed with no light peptides detected in the labelled cells except for contaminants from sample processing (such as keratin).

Next, experiments were carried out to look for demethylation of hnRNP K in cells. Initially a pair of 15 cm dishes were seeded with labelled or non-labelled HEK 293T cells transfected with GFP-tagged hnRNP K and the generic PRMT inhibitor AdOx (100 μM) was added to the labelled cells. Cells were harvested 24 hours after inhibitor treatment and lysed. A BCA assay was used to quantify the protein levels, and the pairs of samples were mixed in a 1:1 ratio (Section 8.5.2).

Immunoprecipitation of GFP-hnRNP K was carried out using GFP-Trap® beads as for the test digestion, and after elution of the protein from the beads, a 30 min in-solution trypsin digest carried out and the samples analysed by LCMS/MS (*samples run by Jacob Bush*, Section 8.13). The methylated peptide occurred at a considerably lower heavy:light ratio than any other peptide from hnRNP K, consistent with arginine methylation by PRMT1 having been inhibited by AdOx in the "heavy" sample (Figure 5.30A). This suggested that the methodology could be used to identify demethylation of hnRNP K by JMJD2A.

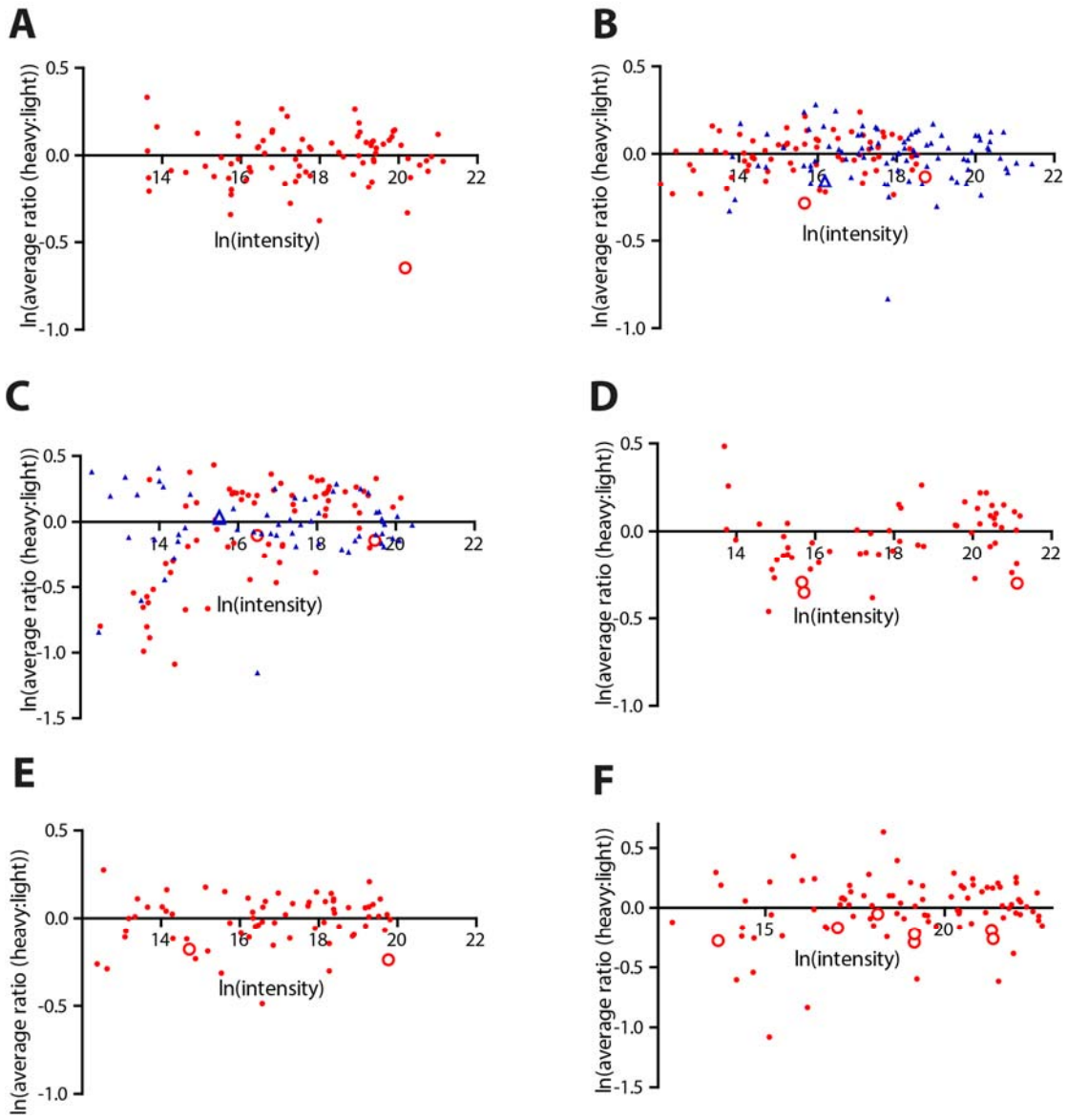


Figure 5.30 SILAC MS experiments measuring methylation levels of GFP-hnRNP K in HEK 293T cells after various cell treatments. **A)** "Heavy" cell population treated with AdOx (100 μM). **B)** "Heavy" cell population overexpressing FL wt Flag-JMJD2A, "light" cell population overexpressing FL H188A Flag-JMJD2A. **C)** "Heavy" cell population treated with IOX1 (100 μM). **D)** "Light" cell population overexpressing FL wt Flag-JMJD2A, "heavy" cell population overexpressing FL H188A Flag-JMJ2A. **E)** "Heavy" cell population overexpressing FL wt Flag-JMJD2A, "light" cell population overexpressing FL H188A Flag-JMJ2A. Both cell populations treated with IOX1 (100 μM). **F)** Average intensity and ratio of hnRNP K peptides from all experiments showing that peptides containing six prolyl-residues occur at a heavy:light ratio below the mean. In all cases peptides from GFP-hnRNP K are represented by small red dots (and small blue triangles for biological repeats) except for peptides containing the dimethylated target arginyl-residue (R296), or peptides containing six prolyl-residues in **F**, which are represented by larger red circles (or larger blue triangles for biological repeats).

Next, the effect of overexpression of JMJD2A on hnRNP K methylation levels was investigated. A pair of plates were seeded and transfected with GFP-hnRNP K and

either FL wt Flag-JMJD2A in the labelled cells or the catalytic mutant FL H188A Flag-JMJ2A in the non-labelled cells. Biological duplicates were carried out. Although all peptides methylated at R296 were observed at a heavy:light ratio below the mean of all peptides from hnRNP K in both replicates, as expected if wt JMJD2A was catalysing demethylation, the low ratio was less pronounced than for the cells treated with AdOx (Figure 5.30B). To investigate whether this slight decrease in heavy:light ratio was due to demethylation by JMJD2A, further experiments were carried out. In one experiment the "heavy" cells were treated with IOX1 (100 μM) to inhibit endogenous JMJD2A,[308] where the ratio might be expected to become larger than the mean due to decreased methylation in the "heavy" cells, in a second experiment overexpression of wt JMJD2A in the "heavy" cells was repeated but treating them with IOX1 (100 μM) to inhibit the activity of the exogenously overexpressed protein and in a third experiment the labelling was inverted, with wt JMJD2A overexpressed in the "light" cells. Unfortunately, all of these experiments also showed a slight decrease in the heavy:light ratio of the methylated peptides (Figure 5.30C-E). This suggested that the "heavy" labelling procedure affected hnRNP K methylation. Arginine can interconvert with proline in cells. This would result in some prolyl-residues becoming "heavy" labelled when "heavy" arginine is added to the media.[367,368] If this occurred, any peptide containing "heavy" proline would not be identified by the search programme (MaxQuant),[369] resulting in a lower intensity of this peptide in the "heavy" cells and thus a decreased heavy:light ratio for these peptides. Comparing the ratio of all hnRNP K peptides containing six prolyl-residues to all other peptides from hnRNP K demonstrated that these peptides always occurred at ratios below the mean. This suggested that arginine to proline interconversion might account for the lower ratio observed for the methylated peptide

in the JMJD2A experiments (Figure 5.30F). Overall these experiments were inconclusive, but suggested that under the assayed conditions, significant demethylation of FL hnRNP K by JMJD2A was not occurring. Further experiments in which proline is supplemented into the media, in addition to arginine and lysine, are required to reduce the interconversion of arginine and proline.

5.10.8.2 Coimmunoprecipitation of hnRNP K and JMJD2A

Attempts were also made to confirm whether there was an interaction between JMJD2A and hnRNP K by coimmunoprecipitation (Section 8.12.4). Cells were transfected with GFP-hnRNP K and an empty Flag vector, FL Flag-JMJD2A and an empty GFP vector or both GFP-hnRNP K and FL Flag-JMJD2A. The three samples were harvested and lysed. With one half of each sample an IP was carried out using anti-Flag beads, and with the other half an IP with GFP-Trap® beads (ChromoTek). Western blots of the samples were probed either with an anti-GFP or an anti-Flag primary antibody (Section 8.8). Unfortunately, no band corresponding to JMJD2A was observed after GFP IP of hnRNP K, and although a very small band could be observed for hnRNP K after the Flag IP, this band was also in the hnRNP K only control, suggesting that it may bind non-specifically to the anti-Flag beads (Figure 5.31). These results provided further evidence that demethylation of hnRNP K by JMJD2A in cells likely does not occur, at least under the assayed conditions.

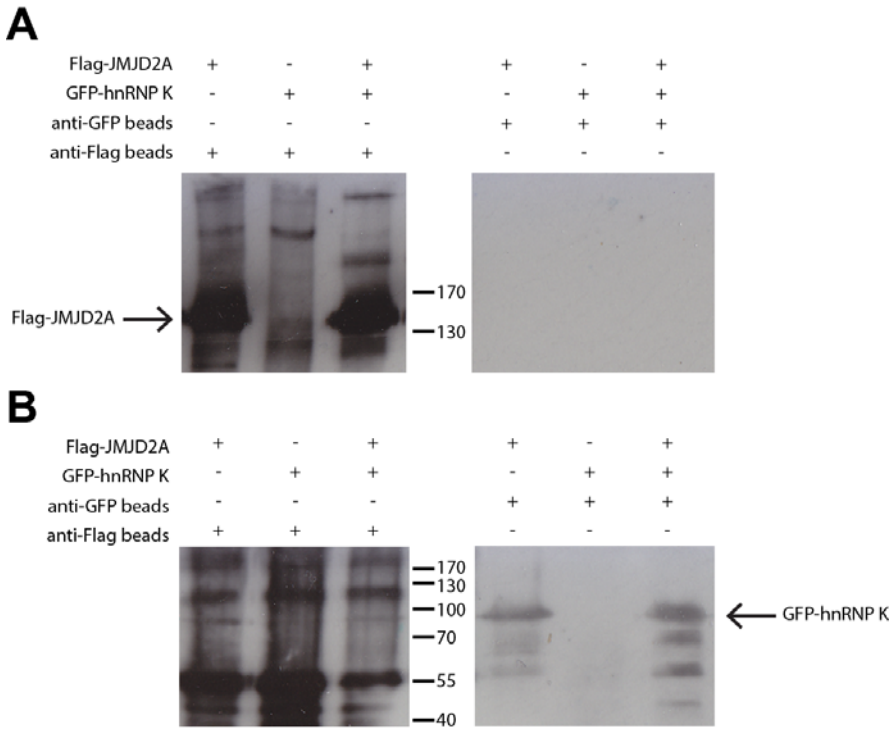


Figure 5.31 Coimmunoprecipitation of exogenously overexpressed GFP-hnRNP K and Flag-JMJD2A using anti-Flag or anti-GFP beads. **A)** Western blots probed with anti-Flag antibody and **B)** western blots probed with anti-GFP antibody. The contents of each lane are marked above the blots and the molecular weight ladder is marked between each blot with numbers representing the mass in kDa.

5.11 Discussion

Work in this Chapter describes the identification of arginine demethylation by three subfamilies of JmjC KDMs. These enzymes represent the first *bona fide* arginine demethylases. Interestingly, the sequence specificity of the enzymes for lysine and arginine has been shown to differ. This suggests that the residues bind in a different conformation in the active site and that selectivity for each residue may be tuned depending on the cellular context. It also adds to the complexity of identifying novel biological substrates as searching for sequences with similarity to the characterised histone sequences may not identify arginine substrates.

Although it has not been possible to crystallise JMJD2A with a methylated arginyl-residue in its active site, inhibitor work in the laboratory using a DNA-encoded cyclic-peptide library to select for peptide inhibitors of JMJD2A produced a potent peptide inhibitor of JMJD2A (IC_{50} = 42 nM).[370] Interestingly, this peptide was successfully co-crystallised with JMJD2A and the crystal structure revealed that an arginyl-residue in the peptide binds in the active site, supporting the idea that binding of this residue is favourable (Figure 5.32, Dr. Oliver King, *unpublished data*). The tight binding of an arginyl-residue in the active site of JMJD2A might account for the low k_{cat} value for demethylation of the H3 $R2Me_{2a}$ peptide by JMJD2A due to slow release of the demethylated product peptide (Section 5.6).

Arginine demethylase activity on histone tail residues has been observed *in vitro* for both JARID1C and the KDM4 subfamily, although as yet it has not been possible to demonstrate that the same reactions also occurs within cells. Work here has been hampered by a lack of specific antibodies. The interplay of multiple post-translational histone modifications also complicate the analysis and make it hard to demonstrate conclusively that any changes in arginine methylation levels are caused by direct demethylation and not an indirect effect of changing levels of histone lysine methylation.

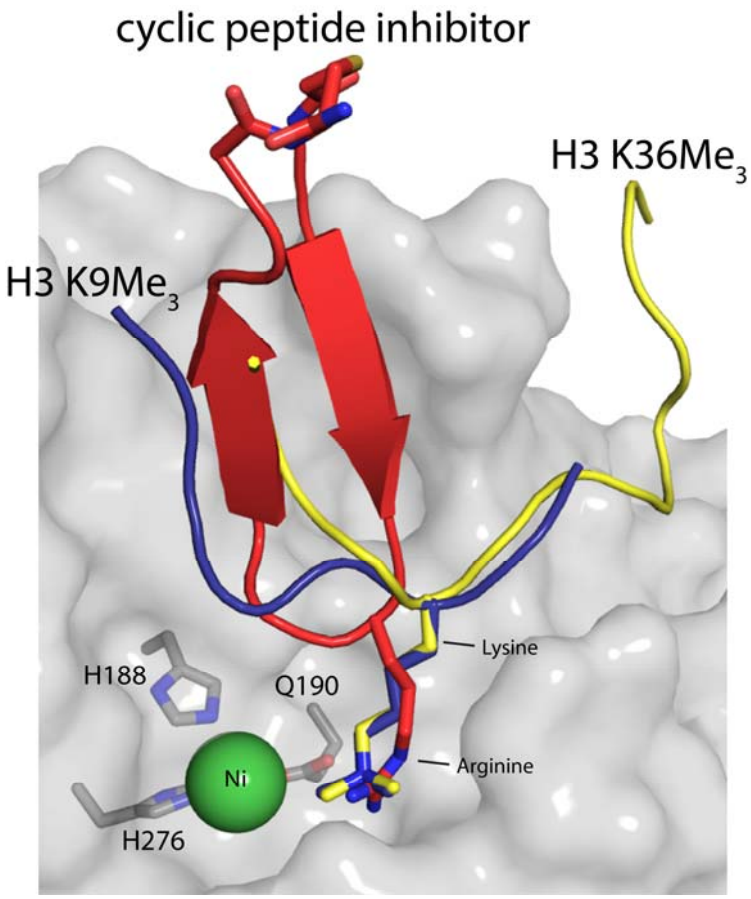


Figure 5.32 View from a crystal structure of JMJD2A (shown as a grey space-fill model) bound to a cyclic peptide inhibitor (red). Overlay of histone peptide conformation of peptide containing H3 $K9Me_3$ (blue, from PDB ID 2OQ6) and H3 $K36Me_3$ (yellow, from PDB ID 2OS2).[145] Adapted from a figure made by Dr. Martin Münzel.

Although IRMS experiments with extracted histones were used to demonstrate that arginine demethylation can occur on full length histones, the acid enrichment for histone proteins from cells will also have enriched for other abundant highly basic proteins such as ribosomes. As such it is possible that the decrease in ratio of dimethylated arginine as compared to monomethylated arginine may be due to demethylation of one of these other proteins rather than the histones themselves. This contamination with other non-histone proteins is not a problem for the whole protein MS assays, however, where UPLC-MS is used to further purify the histones.

The identification of multiple non-histone peptide substrates for JMJD2A/E and JARID1C suggests that multiple biological substrates may be identified. There is precedent for this in other families of epigenetic modifiers, such as the PRMTs and KMTs, which all have many cellular substrates in addition to histones. Substrate targeting is often mediated by other binding domains within the protein or other non-substrate binding partners; for example, the selectivity of the histone demethylase LSD1 for non-histone substrates in addition to histone substrates is modulated by a number of binding partners.[371]

Although no biologically relevant peptidic substrates have yet been identified for JMJD3, it remains the most active of the KDMs tested on its mutant K27R substrate peptide. This result suggests that JMJD3 may have other non-histone methylated arginine substrates that are yet to be identified. Though synthesis of libraries of methylated peptides has provided a number of potential substrates for the other enzymes, this method of substrate identification is expensive and labour intensive, and, as a peptide is often unlikely to mimic the conformation of the sequence in a folded protein risks false substrate identifications. Unfortunately, studies on full-length hnRNP K suggest in this case the peptide data was misleading and that demethylation of hnRNP K likely does not occur *in vivo.* It is possible, however, that this lack of activity on full-length protein is an artefact of having the GFP tag which may block binding by JMJD2E. To investigate whether this is the case experiments could be performed with an alternative tag such as a Myc tag. Identification of novel substrates of "eraser" proteins by standard whole protein techniques is hampered by the difficulty in identifying the correlated lost methylation. In addition, classical techniques for substrate hunting, such as yeast two-hybrid cannot be used as the

modified amino acids cannot easily be incorporated into proteins made within cells. A more expedient and reliable method might be to identify substrates directly in cells, using SILAC experiments like the one described above but using enrichment of peptides by one of the non-specific methyl-arginine antibodies instead of immunoprecipitation of a specific protein.

Chapter 6
Characterisation of UTY

6.1 The KDM6 family of histone demethylases

In 2007, the KDM6 subfamily of JmjC KDMs, UTX (KDM6A) and JMJD3 (KDM6B), were identified as H3 K27Me$_3$ demethylases.[225–227,372] However, in the same studies, the third Y-linked family member, UTY, was reported not to have catalytic demethylase activity in *in vitro* demethylation assays or in cells, despite sharing over 88% sequence similarity with its X-linked homologue, UTX. Sequence alignments show both UTX and UTY to have similar domain organisation (Figure 6.1), with *N*-terminal tetratricopeptide repeat (TPR) domains of unknown function and *C*-terminal JmjC and zinc binding domains.



Figure 6.1 Domain organisation of the KDM6 subfamily of JmjC KDMs.

Within the JmjC domains, UTX and UTY share over 98% similarity and UTY contains the same complete triad of required Fe(II)-binding residues (HXE…H) and the same 2OG-binding lysyl-residue present in JMJD3 and UTX (Figure 6.2). UTY itself has been shown to have multiple splice isoforms both with and without the JmjC domain and a region in the *N*-terminus of many of the isoforms has been demonstrated to be a male-specific minor histocompatibility antigen that has been linked to transplant rejection.[373–375]

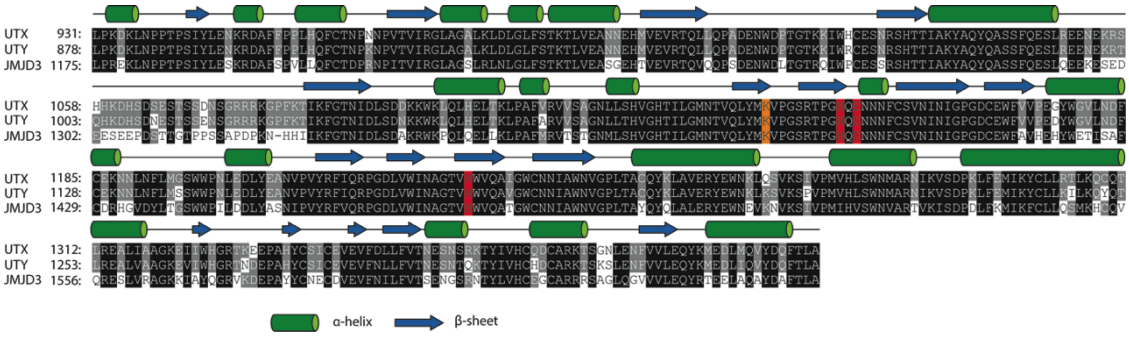


Figure 6.2 Sequence alignments of the JmjC domains of UTX, UTY and JMJD3. Secondary structure for UTX is illustrated above the alignment with green cylinders representing α-helices and blue arrows representing β-sheets. Numbering for UTY corresponds to isoform 3. JMJD3 / UTX and predicted UTY Fe(II)-binding residues are coloured in red and the 2OG binding lysyl-residue in orange. Black shading shows residues that are identical in all three family members, grey shading shows positions where only two members share the same amino acid.

Recent studies have also shown that both UTX and UTY associate with similar sets of proteins in complexes in cells.[376] Work in UTX knockout (K/O) embryonic stem (ES) cells and UTX K/O mice has shown a sex-dependent effect on development; whilst many of the male UTX deleted littermates survive to birth, loss of UTX in the female mice is embryo lethal in all cases.[209,229,376] Interestingly, UTX is also known to escape X-inactivation, which results in existence of a dosage imbalance between sexes.[228]

6.1.2 Objectives

The combined sequence analyses and recent cellular results suggested that like UTX, UTY might be a functional demethylase. In light of work described in Chapter 5 showing that JMJD3 can demethylate arginine residues, it was also considered possible that UTY might be an arginine demethylase rather than a lysine demethylase, hence its having been ascribed no original catalytic function. Work in this Chapter described the cloning, purification and characterisation of UTY, a 2OG oxygenase that was previously thought not to have catalytic activity. Following substrate peptide screening and kinetic characterisation of catalytically active UTY, preliminary experiments investigating its function in a cellular context are also described.

NMR experiments in this Chapter were carried out either in collaboration with or by Dr. Richard Hopkinson. The crystal structure of UTY was produced by Dr. Catrine Johansson.

6.2 Cloning and expression of UTY

A recently published paper containing a crystal structure of UTX shows it to contain both JmjC and zinc-binding domains.[377] Initially, the residues of UTY matching this construct were amplified from a clone kindly provided by Kai Ge (Addgene plasmid 17439)[226]. DNA encoding for UTY residues P818 - S1347 were cloned using ligation independent cloning into a pNIC28-Bsa4 vector that encodes for an *N*-terminal 6×His tag (Section 8.6.2). Unfortunately, initial expression of this construct provided very little soluble recombinant protein. Hence, UTY was cloned into a second vector, pNH-Trx, encoding for an *N*-terminal 6×His tag followed by a thioredoxin (Trx) fusion partner. Trx is known to improve folding and protein solubility and has been shown to increase soluble protein production in *E. coli* for other JmjC demethylases (unpublished data).[378] Expression trials with this new construct showed high levels of expression in the soluble fraction under all temperature and IPTG conditions tested (Section 8.5). Due to the larger number of cells obtained by overnight incubations at 18 °C compared to 4 hr at 37 °C, the overnight induction using 0.2 M IPTG was chosen for future large-scale growths.

6.2.1 Large-scale expression of UTY

Using the conditions determined above, a 7.2 L *E. coli* growth of recombinant His-tagged UTY was undertaken. The protein was first purified from crude cell

lysates using a 5 mL Ni HisTrap column. Fractions 3 – 7 were pooled and concentrated (Figure 6.3A). At this point the protein was divided into two, and half was loaded directly onto a 300 mL S200 size exclusion column pre-equilibrated with buffer containing 10 mM HEPES (pH 7.5), 200 mM NaCl, 5% glycerol. 5 mL fractions were collected and analysed by SDS-PAGE (Figure 6.3B). Pure fractions (22-25) were pooled and concentrated before flash freezing in liquid nitrogen and storage at -80 °C.

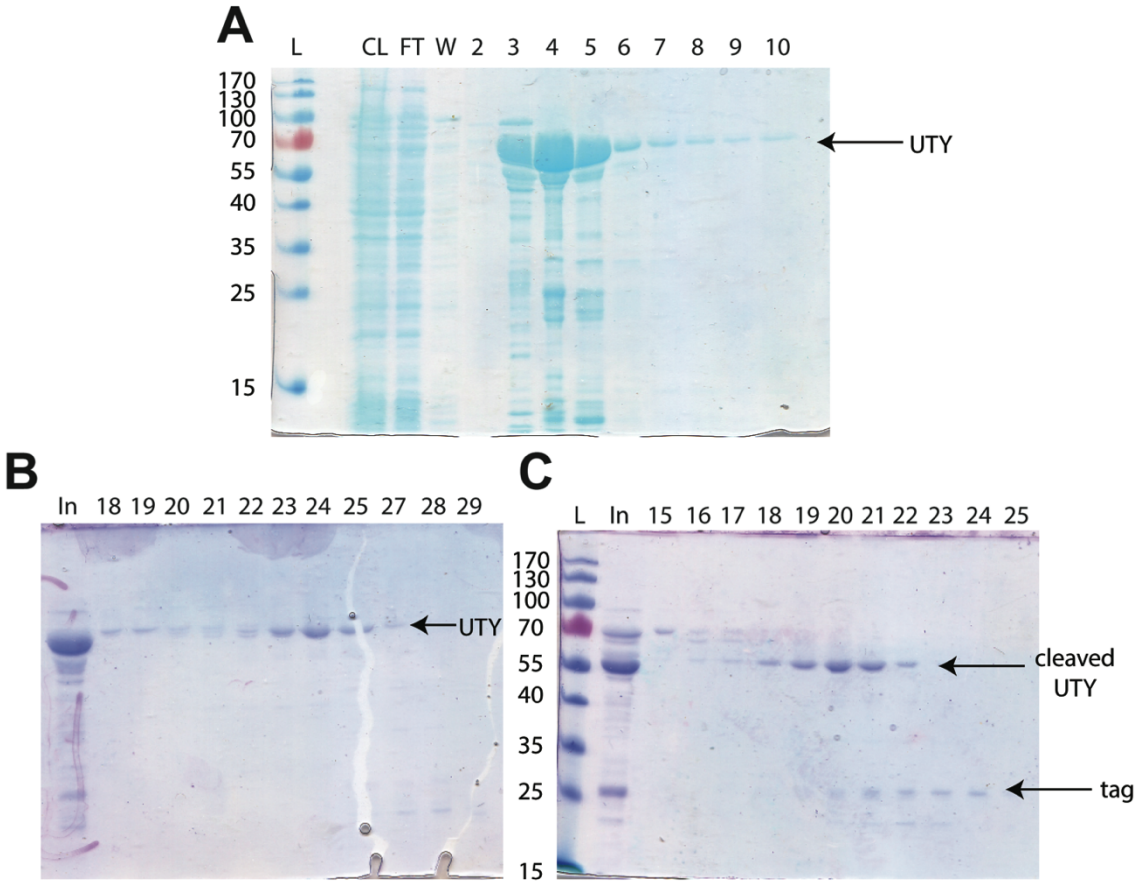


Figure 6.3 A) Fractions from 5 mL Ni HisTrap column. **B)** Fractions after size exclusion chromatography using a 300 mL S200 column. **C)** Cleaved fractions after size exclusion chromatography. L denotes ladder, CL the cell lysate, FT the flow-through, W the wash, In the input and numbers indicate the fraction number loaded in each lane. Numbers on the left of the gels give the molecular weight of each standard in kDa.

The other batch of enzyme was incubated overnight at 4 °C with Tobacco Etch Virus (TEV) protease to cleave the 6×His-Trx tag before application to an S200 300 mL size exclusion column. Again 5 mL fractions were collected and analysed by

SDS-PAGE (Figure 6.3C). Fractions 19 – 22 were pooled and concentrated using a 30 kDa cut-off spin concentrator, to allow further removal of the 25 kDa tag. The concentrated protein was then flash frozen with liquid nitrogen before storage at -80 °C.

6.3 Characterisation of recombinant UTY

6.3.1 Detection of catalytic activity of UTY as uncoupled 2OG turnover

2OG oxygenases couple the oxidative decarboxylation of 2OG to oxidation of their prime substrate (Figure 1.15, Chapter 1). For most of these enzymes, 2OG decarboxylation can also occur in the absence of prime substrate (uncoupled turnover), albeit at a significantly reduced level.[158] Using the recombinant His-Trx-tagged UTY, initial experiments were conducted to look for evidence of uncoupled 2OG turnover, which would be indicative of the protein being correctly folded and potentially active. ^{1}H NMR spectroscopy was used to observe conversion of 2OG to succinate (Section 8.15). UTY (10 µM) was incubated with 2OG (500 µM), Fe (100 µM) and ascorbate (1 mM) in a 2 mm MATCH NMR tube and spectra were collected every 168 sec. 2OG turnover was observed in the ^{1}H NMR analyses as indicated by consumption of 2OG (decrease in the triplet resonance at δ_H = 2.39 ppm) and production of succinate (formation of a singlet resonance at δ_H = 2.35 ppm), demonstrating that the recombinant UTY was active (Figure 6.4A). Identical conditions were used to measure the rate of uncoupled 2OG turnover with recombinant UTX and JMJD3. Interestingly, the rate of uncoupled 2OG turnover for all three recombinant enzymes was found to be comparable, with UTY the most active of the three enzymes (Figure 6.4B).

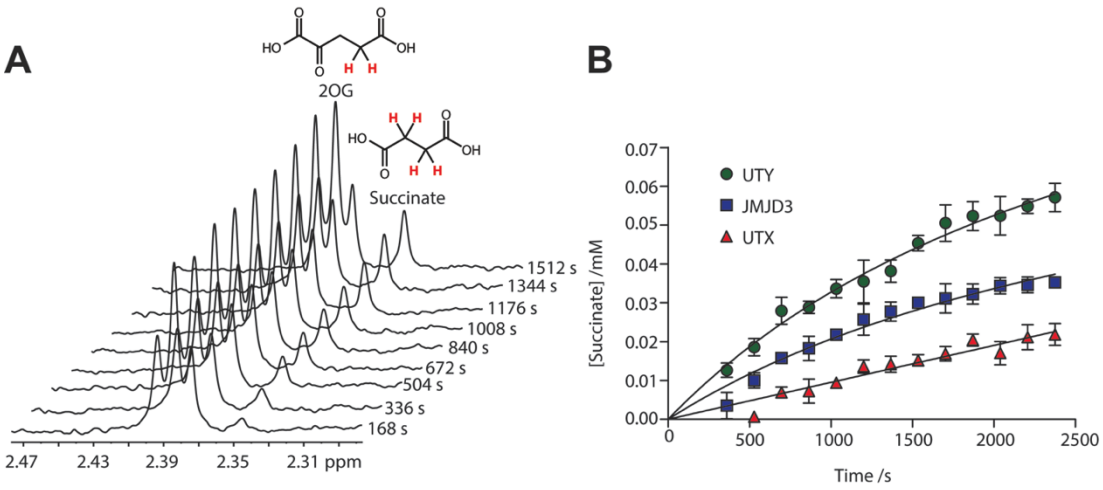


Figure 6.4 A) ^{1}H NMR spectra showing uncoupled conversion of 2OG to succinate by UTY. Protons responsible for the observed NMR peaks are shown in red. **B)** Comparison of the rate of conversion of 2OG to succinate by UTY, JMJD3 and UTX as measured by ^{1}H NMR. Data is plotted as the mean of triplicate experiments and error bars show ± 1 s.d..

6.3.2 Determination of the substrate of UTY by peptide screening

Having determined that the recombinant UTY was active, the potential of UTY to act as a histone demethylase was then investigated. UTY (5 μM) was incubated with cofactors (100 μM 2OG, 100 μM Fe(II) and 1 mM ascorbate) and histone peptides (between 15 and 21 residues in length) corresponding to many of the known lysine and arginine methylation sites of H3 (at K4, K9, K18, K27 and K36 and R2, R8, R17, R26), H2A (at K5 and R3) and H4 (at K20 and R3) in all methylation states (KMe_3, KMe_2, KMe, RMe_{2a}, RMe_{2s}, RMe) for an hour at 37 °C. Demethylation was assessed by MALDI-MS with mass shifts of -14 Da and -28 Da corresponding to mono- and di-demethylation, respectively (Section 8.7.1). As for the other subfamily members, UTX and JMJD3, with UTY significant demethylation was observed only for the H3K27Me_3 peptide. Both mono- and di-demethylation of the trimethylated peptide were observed, with a preference for the trimethylated state (Figure 6.5). Demethylation was shown to be dependent on the presence of Fe(II) (removed from the enzyme using EDTA) and 2OG.

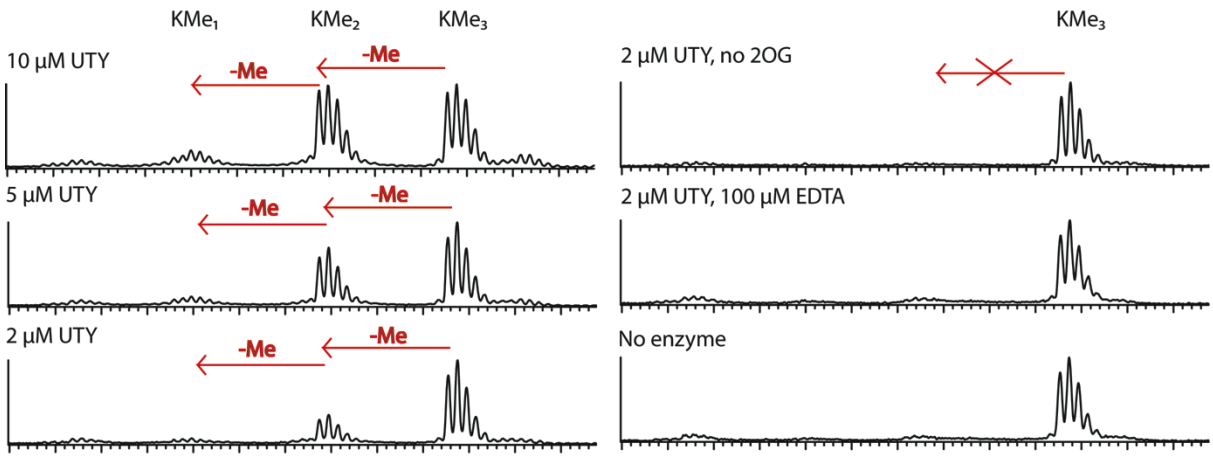


Figure 6.5 MALDI assay showing demethylation of a H3 K27Me$_3$ 11-34 peptide (10 µM) by recombinant UTY. -14 Da mass shifts are indicated by red arrows.

No demethylation was observed for the K4, K9, K18 and K36 methylated peptides, any peptides from H2A or H4 or any of the methylated arginine peptides. In addition, no activity was observed with the cleaved UTY protein, likely due to the increased amount of time required for the purification.

6.3.3 Confirmation of histone demethylation by peptide NMR spectroscopy

To further confirm *bona fide* lysine demethylation, the UTY reaction was also studied using ^{1}H NMR spectroscopy to assay for formation of a dimethyllysine peak (Section 8.15). Spectra were collected for UTY, UTX, and JMJD3 to allow comparison of reaction rates and confirm that the same product peptide was being formed. As for the experiments looking at uncoupled turnover, enzyme (10 µM) was mixed with 2OG (500 µM), Fe(II) (100 µM) and ascorbate (1 mM) and H3 K27Me$_3$ peptide (1 mM) and spectra were collected every 168 sec for 50 min and then again after 16 hr. In the case of UTX and JMJD3, the dimethyl peak could be observed at δ_H = 2.89 ppm (Figure 6.6A). Using 10 µM UTY the presence of a demethylated peak was ambiguous, so the concentration was increased to 44 µM UTY. With the increased

enzyme concentration, it was possible to observe the dimethyl lysine peak after 16 hr (Figure 6.6B).

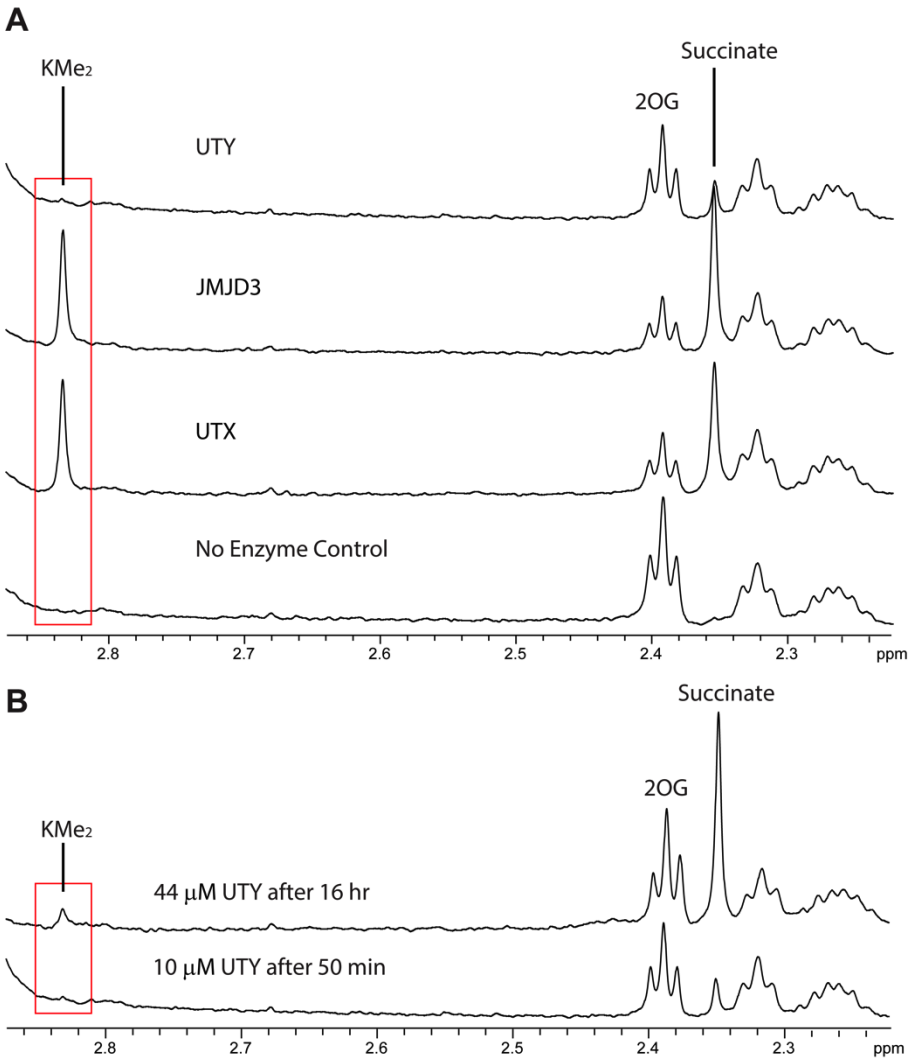


Figure 6.6 A) ^{1}H NMR spectra showing demethylation of H3 K27Me$_3$ 11-34 (1 mM) by UTY (10 µM), JMJD3 (10 µM) and UTX (10 µM). Bottom spectrum shows a no enzyme control. **B)** ^{1}H NMR spectra showing demethylation of H3 K27Me$_3$ 11-34 (1 mM) by UTY (10 µM) at an increased enzyme concentration after overnight incubation. The product dimethyllysine peak is marked by a red box.

6.3.4 Kinetic characterisation of UTY peptide demethylation

Histone demethylation can also be studied by the FDH assay, which provides a convenient method for measuring enzyme kinetics (Section 4.6.1, Chapter 4). Initially a pilot assay was carried out to assay for formaldehyde production. Increasing concentrations of UTY (0 - 4 µM) were mixed with H3K27Me$_3$ peptide (100 µM),

2OG (20 μM), Fe(II) (5 μM), ascorbate (100 μM), FDH (0.025 units) and NAD^+ (50 μM) and the initial rate of demethylation was measured (Section 8.7.2). As expected for a histone demethylase, increasing rates of formaldehyde production were observed with increasing concentrations of UTY, in a dose-dependent manner (Figure 6.7A).

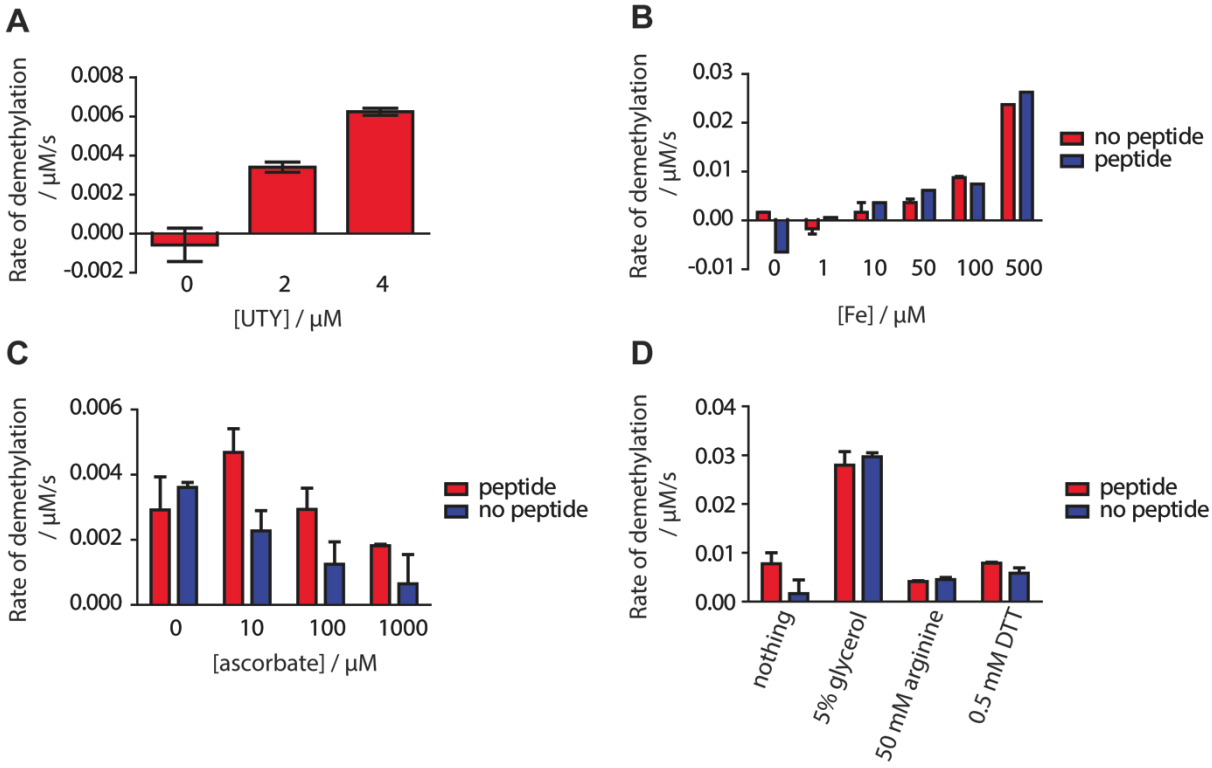


Figure 6.7 FDH assays showing the rate of formaldehyde production from demethylation of H3 $K27Me_3$ 12-34 by UTY. **A)** Effect of varying concentrations of UTY. **B)** Effect of Fe(II) concentration. **C)** Effect of varying ascorbate concentration. **D)** Effect of buffer additives.

Before carrying out any kinetic experiments it was important to optimise the assay conditions. Using UTY (2 μM) and fixed concentrations of all reaction components except the one being optimised, the effects of changing Fe(II) (Figure 6.7B), ascorbate (Figure 6.7C) and buffer (Figure 6.7D) conditions were tested. The addition of glycerol (5%) has been shown to improve enzyme stability,[379] whilst DTT (0.5 mM) and arginine (50 mM) have each been shown to generally increase the activity of enzymes.[380] Unfortunately, even in the absence of peptide substrate, where

no formaldehyde should be produced, a substantial amount of NAD^+ reduction to NADH was observed. A similar conversion of NAD^+ to NADH uncoupled to histone demethylation has been observed for PHF8 (T.Rooney, Rotation Report). For PHF8, removal of the 6×His-Trx tag removed this uncoupled reduction, but here this was not possible due to the loss of activity on cleavage of the tag (Section 6.3.2). The largest signal-to-noise window was observed with Fe(II) (10 μM), ascorbate (10 μM) and without buffer additives, but due to the background reaction it was decided that the FDH assay was unsuitable for determining kinetic parameters of UTY.

Instead, K_M and k_{cat} values for 2OG were determined using the MALDI MS based assay with a H3 11-34 K27Me_3 peptide as the substrate (Section 8.7.1). Kinetic parameters were measured for UTY, as well as UTX and JMJD3 for comparison. In all cases, peptide and Fe(II) were added in excess at 25 μM and 100 μM respectively (based on previously determined data for UTX), to ensure 2OG was the only rate-limiting reaction component. The initial rate of reaction was measured at a variety of 2OG concentrations and the data fitted to the Michaelis-Menten equation to extract kinetic parameters (Figure 6.8). The K_M values for 2OG of all three members were found to be similar (Table 6.1), but UTY had a k_{cat} value more than five-fold lower than either JMJD3 or UTY. This was consistent with the lower levels of demethylation observed by NMR and MALDI (Section 6.4.2). These results suggested that the decreased enzymatic activity was not due to reduced 2OG binding (also consistent with the high rate of uncoupled 2OG turnover (Section 6.3.1)), but more likely due to a weaker peptide binding affinity, so that the peptide concentration (25 μM) used in the reaction was not above the peptide K_M.

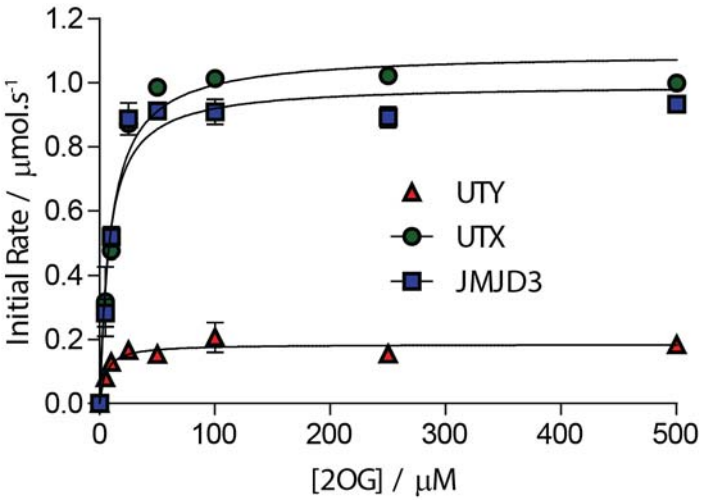


Figure 6.8 MALDI assay measuring the rate of reaction of KDM6 subfamily members with variable concentrations of 2OG. Data were fitted to the Michaelis-Menten equation using GraphPad Prism.

Table 6.1 Table of kinetic parameters for 2OG measured for KDM6 subfamily members JMJD3, UTX and UTY.

	K_M / µM
JMJD3	8.2 ± 1.0
UTX	10 ± 1.3
wt UTY	5.1 ± 1.4

6.3.5 Peptide binding assays

In addition to measuring enzyme activity (Section 4.6.5, Chapter 4), the ALPHA screen can be used to measure the binding of peptides to proteins.[306] For this assay, the donor bead is conjugated to an anti-His antibody, which recognises the protein's 6×His-tag, and the acceptor bead is conjugated to streptavidin so that it binds to the biotinylated peptide of interest. Where protein and peptide interact a signal is measured due to the proximity of the two beads. The magnitude of the signal is proportional to the strength of the binding interaction. This assay was used to look at the interaction of the KDM6 family with a library of *N*-terminally biotinylated 21 residue histone peptides purchased from Alta Bioscience. Due to variability in the exact concentration of the peptides in the library, the assay cannot be used to measure absolute affinities of the enzymes for the peptides, but it can be used to compare the

relative binding affinities within the subfamily for different peptides. Reagents were mixed to a final concentration of enzyme (200 nM), NOG (1 μM) and peptide (200 nM) and incubated for 15 min before ALPHA screen donor and acceptor beads were added. Luminescence readings were taken after 1 hr (Section 8.7.4). The data did not show the predicted maximal binding for the H3 peptide containing K27Me$_3$ (H3 20-40 K27 Me$_3$). This was likely due to the absence of H3 R17 that has previously been demonstrated to be important for substrate binding.[377] However, Figure 6.9 shows the similarity between the binding heat maps for each of the three members of the KDM6 subfamily for a representative row of H3 1-21 peptides.

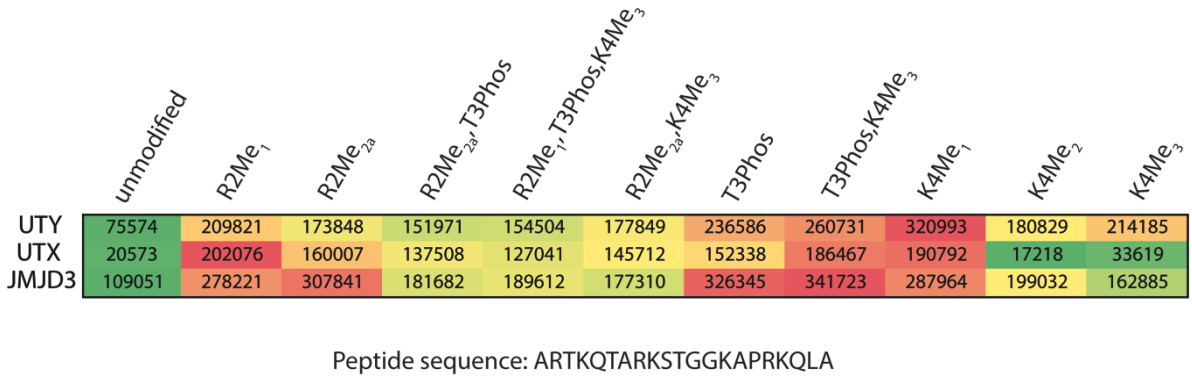


Figure 6.9 ALPHA screen binding assay looking at binding of UTY, UTX and JMJD3 with a series of H3 1-21 peptides from Alta Bioscience. The numbers give the ALPHA screen count. The baseline reading for each histone peptide without any enzyme added was subtracted from each reading. Colouring shows the magnitude of the ALPHA screen count, with red marking the highest reading and green the lowest.

6.3.6 Characterisation of inhibition of UTY by 2OG oxygenase inhibitors

The MALDI MS assay was used to compare the behaviour of the different KDM6 subfamily members to a variety of inhibitors (Section 8.7.1). Due to the requirement for a high concentration of UTY to allow reliable detection of peptide demethylation with this assay, it was not possible to measure IC_{50} values, which can only be determined down to concentrations half that of the enzyme concentration used for the assay. Consequently, a panel of 9 histone demethylase inhibitors was selected and

each compound (10 μM) ranked in order of its potency as an inhibitor for each of the three enzymes (UTX (2 μM), UTY (5 μM) and JMJD3 (2 μM)). As expected for closely related enzymes, the inhibitor profiles were similar (Figure 6.10). The UTX and JMJD3 specific inhibitor GSK-J1 also targeted UTY and was found to be the most potent inhibitor tested.[381] To date this inhibitor has shown selectivity for JMJD3 and UTX over other KDM subfamilies tested and so its inhibition of UTY was consistent with UTY being closely related structurally and functionally to the other two KDM6 subfamily members.

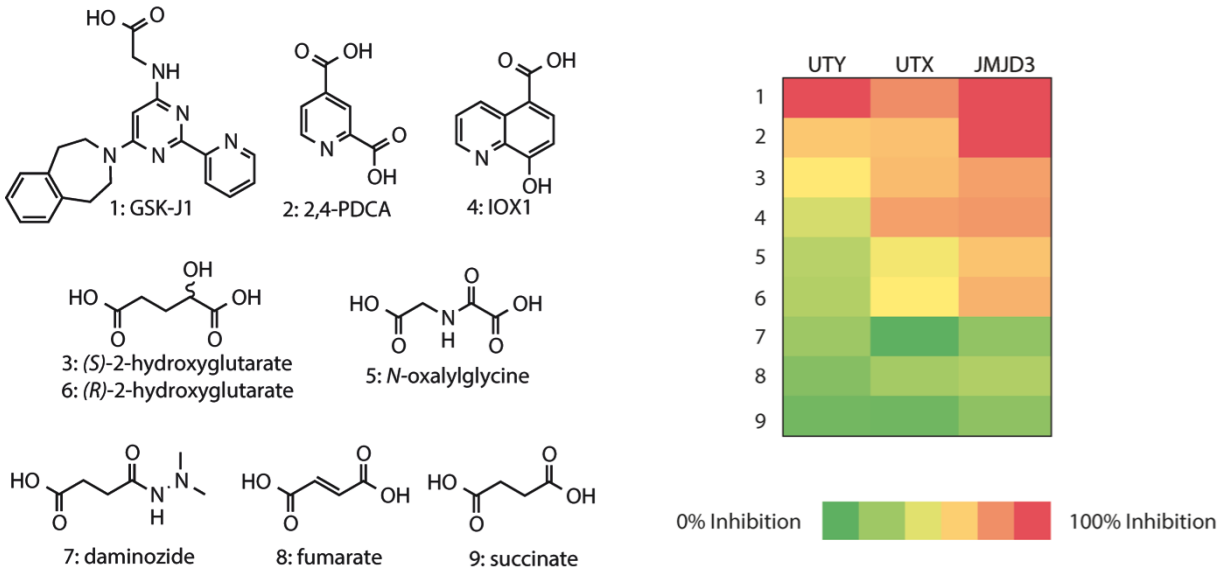


Figure 6.10 Response of KDM6 subfamily members to 9 known 2OG oxygenase inhibitors. Inhibitors are ranked by their potency at 10 μM. Structures of the inhibitors are shown on the left.

6.4 Production and characterisation of UTY variants

Although the catalytic domains of UTY and UTX share 88% sequence identity (Figure 6.2), the observed activity of UTY appeared to be much reduced relative to UTX, as measured by NMR spectroscopy (Figure 6.6A). Sequence alignments were used to identify residues in UTX that might account for this difference in activity. Two residues in particular, S1138 and P1214, were identified that differed both

structurally and chemically from those present in UTX and JMJD3 (where the equivalent residues are a glycine and isoleucine respectively). Based on a recent co-crystal structure of UTX in complex with an H3 K27Me$_3$ peptide (PDB ID 3AVR), they also appeared potentially important for peptide binding (Figure 6.11).[377]

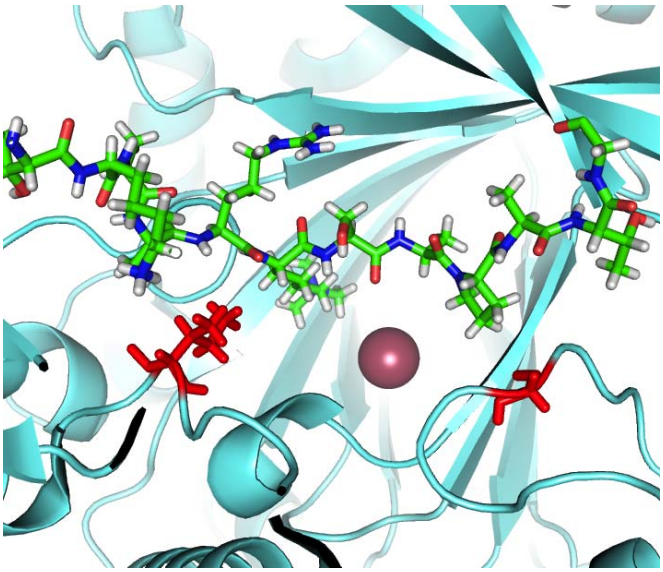


Figure 6.11 View from a crystal structure of UTX complexed with a 22 residue H3 K27Me$_3$ peptide (PDB ID 3AVR).[377] UTX is shown in blue, peptide in green and Ni (substituting for Fe(II)) in dark red. The two residues selected for mutagenesis in UTY are shown as sticks in red.

To test whether increased UTY activity could be induced by substitution of these residues, site-directed mutagenesis was used to convert the prolyl- or seryl-residue individually in UTY to the corresponding isoleucyl- or glycyl-residue present within UTX and JMJD3 (Section 8.6.2). Two variants, UTY S1138G and UTY P1214I were successfully produced recombinantly in *E. coli* and purified in a manner analogous to wt UTY (Section 6.2.1).

6.4.2 Characterisation of UTY variants by MALDI MS and NMR spectroscopy

NMR spectroscopy was used to compare the activity of these variants to the wt UTY protein (Section 8.15). Whereas UTY S1138G showed only slightly elevated activity

levels relative to that of wt UTY, the P1214I variant was found to be considerably more active (Figure 6.12).

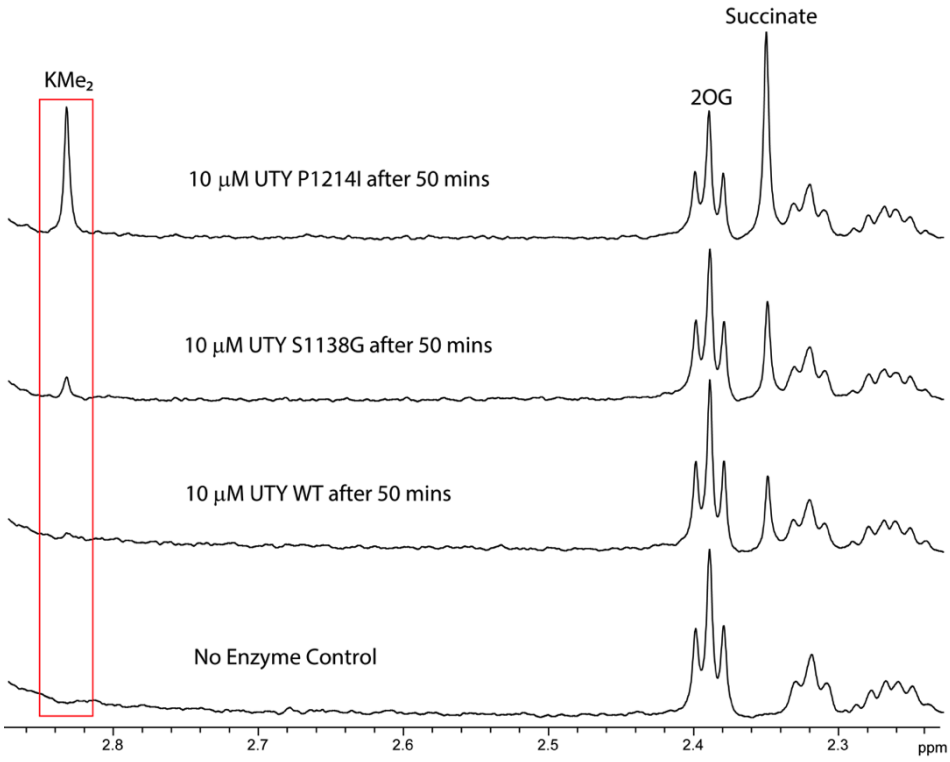


Figure 6.12 ^{1}H NMR spectra showing demethylation of H3 K27Me$_3$ 11-34 (1 mM) by wt UTY (10 μM), UTY S1138G (10 μM) and UTY P1214I (10 μM) after 50 min. Bottom spectrum shows a no enzyme control. The dimethyllysine peak is shown in a red box.

6.4.3 Kinetic characterisation of UTY variants

2OG K_M values for the UTY variants were determined using a MALDI MS assay with a 23 residue H3 K27Me$_3$ peptide as the substrate (Section 8.7.1). The rate of reaction of the UTY variants with increasing concentrations of 2OG was fitted to the Michaelis-Menten equation to determine kinetic parameters (Figure 6.13). The K_M values for 2OG for the wt UTY protein and UTY variants were found to be similar to each other and to those for UTX and JMJD3 (Table 6.2). However, the maximum rate of reaction for UTY P1214I was found to be considerably higher (approximately six-fold higher than for wt UTY) suggesting that peptide binding was stronger for the UTY P1214I variant than for wt UTY, and more similar to JMJD3 and UTX.

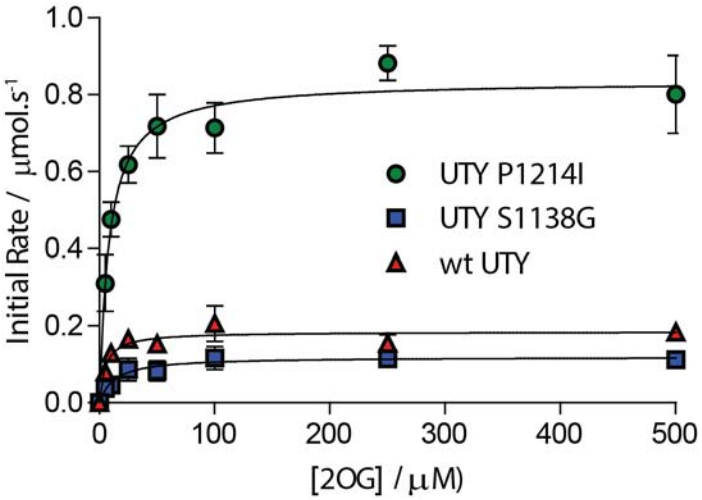


Figure 6.13 MALDI MS assay measuring the rate of reaction of UTY variants with variable concentrations of 2OG. Data were fitted to the Michaelis-Menten equation using GraphPad Prism. Points represent the mean reaction rate and error bars show ± s.d. of triplicate experiments.

Table 6.2 Kinetic parameters measured for UTY variants with the kinetic parameters for other KDM6 subfamily members shown below for reference.

	K_M / µM
wt UTY	5.1 ± 1.4
UTY P1214I	8.3 ± 1.2
UTY S1138G	13 ± 3.8
JMJD3	8.2 ± 1.0
UTX	10 ± 1.3

Although the activity of wt UTY was too low to successfully measure a K_M for the peptide substrate, the increased activity observed with the P1214I variant enabled the determination of a K_M of 295 µM (Figure 6.14). This is still about seven-fold higher than the peptide K_M determined for UTX (43 µM, data not shown), suggesting that there were other residues that were also responsible for the reduced activity of wt UTY.

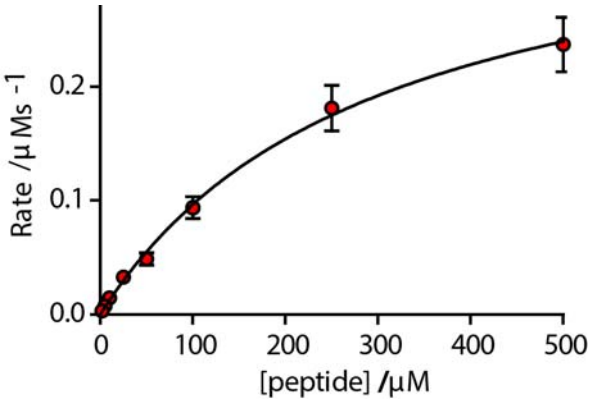


Figure 6.14 MALDI MS assay measuring the rate of reaction of UTY variant P1214I with variable concentrations of H3 14-34 K27Me$_3$. Data were fitted to the Michaelis-Menten equation using GraphPad Prism. Points represent the mean reaction rate and error bars show ± s.d. of triplicate experiments.

The increased activity of the P1214I UTY variant suggested that the isoleucyl-residue present in UTX and JMJD3 at that position was important for correct substrate binding, either due to protein conformation or enzyme-peptide binding interactions. *In vivo*, however, this single amino acid alteration may not play such a significant role in catalysis; all three enzymes possess additional domains (Figure 6.1), which likely also contribute to histone substrate binding and may outweigh the effect of the reduction in binding interaction from this single amino acid substitution. In addition, like many other chromatin modifying enzymes, both UTY and UTX associate in protein complexes and it is possible that additional co-factors are needed for optimal demethylase activity. A study has recently shown that conversion of these residues in FL UTX and JMJD3 to the prolyl- and seryl-residues present in UTY reduces, but does not abolish, their activity in cells as measured by immunofluorescence of global H3 K27Me$_3$ levels, consistent with these not being the only differences between UTY and UTX / JMJD3 responsible for its reduced activity.[376] They show that two further amino acid substitutions (UTX Y1135C and T1143I) that are present in mouse UTY (in human UTY the residue is the same as in UTX and JMJD3) abolished global demethylation activity of UTX and JMJD3 in cells, although they do not investigate

the demethylation activity of the proteins *in vitro*, and it is thus possible that the enzyme is still slightly active, but below the threshold that can be observed by immunofluorescence.

6.5 Structural studies of recombinant UTY

Whilst these studies were being undertaken, a crystal structure of the *C*-terminal region of UTY (L878-S1347, isoform 3) in complex with 2OG was solved by Dr. Catrine Johansson (Structural Genomics Consortium, Oxford) at 1.8 Å resolution. As for UTX, the structure showed the *C*-terminal construct to contain three conserved domains: a JmjC domain (881-1188, blue), a helical domain (1193-1258 and 1327-1344, orange) and a zinc-binding domain (1263-1326, green) (Figure 6.15A). Although there were minor differences between the structures, superimposition of the structure of UTY with that of UTX (PDB ID 3AVS) showed them to be extremely similar (overall RMSD value at C^{α} = 0.93). As for UTX, the Zn(II) is coordinated by 4 cystyl-residues, C1278, C1281, C1305 and C1308. As for UTX and JMJD3 the 2OG carboxylate-binding lysyl-residue was found on the 1st strand of the β-barrel, rather than on the 4th strand, which is more characteristic of JmjC domain enzymes. Interestingly, the overall conformation around P1214 was almost identical, confirming that the increased activity of P1214I was likely due to increased interaction with A25 of the histone peptide rather than a change in the overall protein conformation (Figure 6.15B).

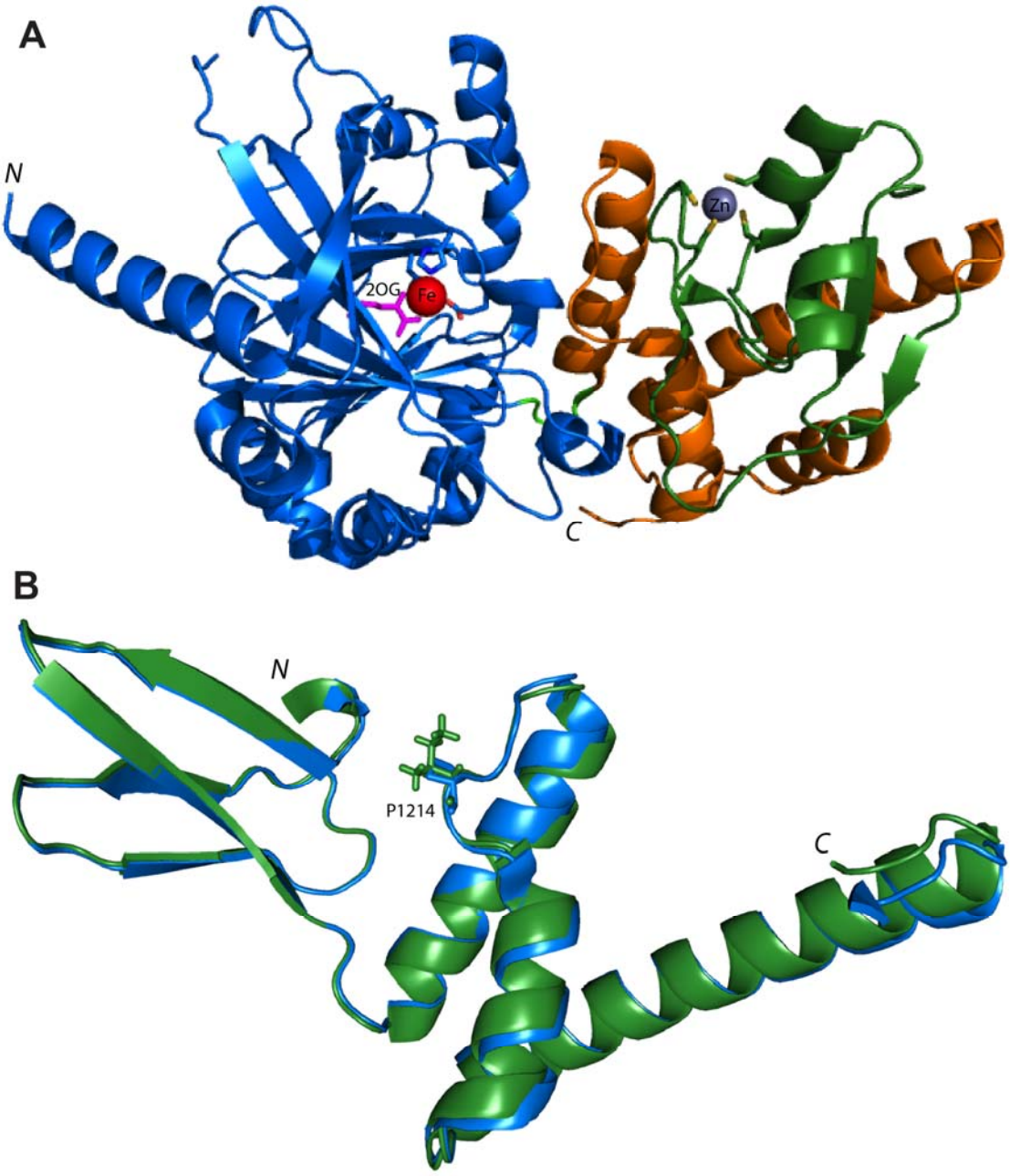


Figure 6.15 A) View from a crystal structure of UTY (PDB ID 3AVS). The JmjC domain is shown in blue, the helical domain in orange and the zinc-binding domain in green. 2OG is shown in pink, Fe(II) in red and Zn in grey. Metal chelating residues are shown as sticks. **B)** Overlay of the structures of UTX (green) and UTY (blue) in the region of UTY P1214. *N*- and *C*-termini are labelled.

6.6 Investigating the cellular function of UTY

6.6.1 Studies of the global levels of H3 K27Me$_3$ by immunofluorescence

To determine whether UTY was active as a demethylase in cells, the full length protein was amplified and inserted into a Flag-tagged vector by LIC, and site-directed mutagenesis employed to produce the likely catalytically inactive variant, H1095A (missing one of the Fe(II) binding residues) (Section 8.6.2). Immunofluorescence was then used to assay for effects of exogenous overexpression of FL UTY or the catalytically inactive mutant FL H1095A UTY on global H3 K27Me$_3$ levels in HeLa cells (Section 8.12.3). HeLa cells were transfected with DNA encoding for wt UTY or UTY H1095A and left for 24 hours before being fixed with formaldehyde, permeabilised with Triton X-100 and stained with one antibody to the FLAG tag of UTY and a second complementary antibody to H3 K27Me$_3$ (Millipore 07-449). These were visualized with two dye-conjugated secondary antibodies. Nuclei were stained using DAPI. As observed previously, exogenous expression of FL wt UTY caused no detectable decrease in the global levels of H3 K27Me$_3$ (Figure 6.16A). A positive control expressing FL wt JMJD3 showed a decrease in H3 K27Me$_3$ as demonstrated by a decrease in the intensity of the red staining (Figure 6.16B). Overexpression of the catalytically inactive JMJD3 mutant did not cause the same decrease in methylation levels (Figure 6.16C). These results suggested that the cellular demethylase activity of UTY may be context dependent, limited to specific targets or outweighed by functional redundancy with the more activity UTX (present endogenously in HeLa cells), so that its demethylase activity cannot be observed as a global change in H3 K27Me$_3$ levels.

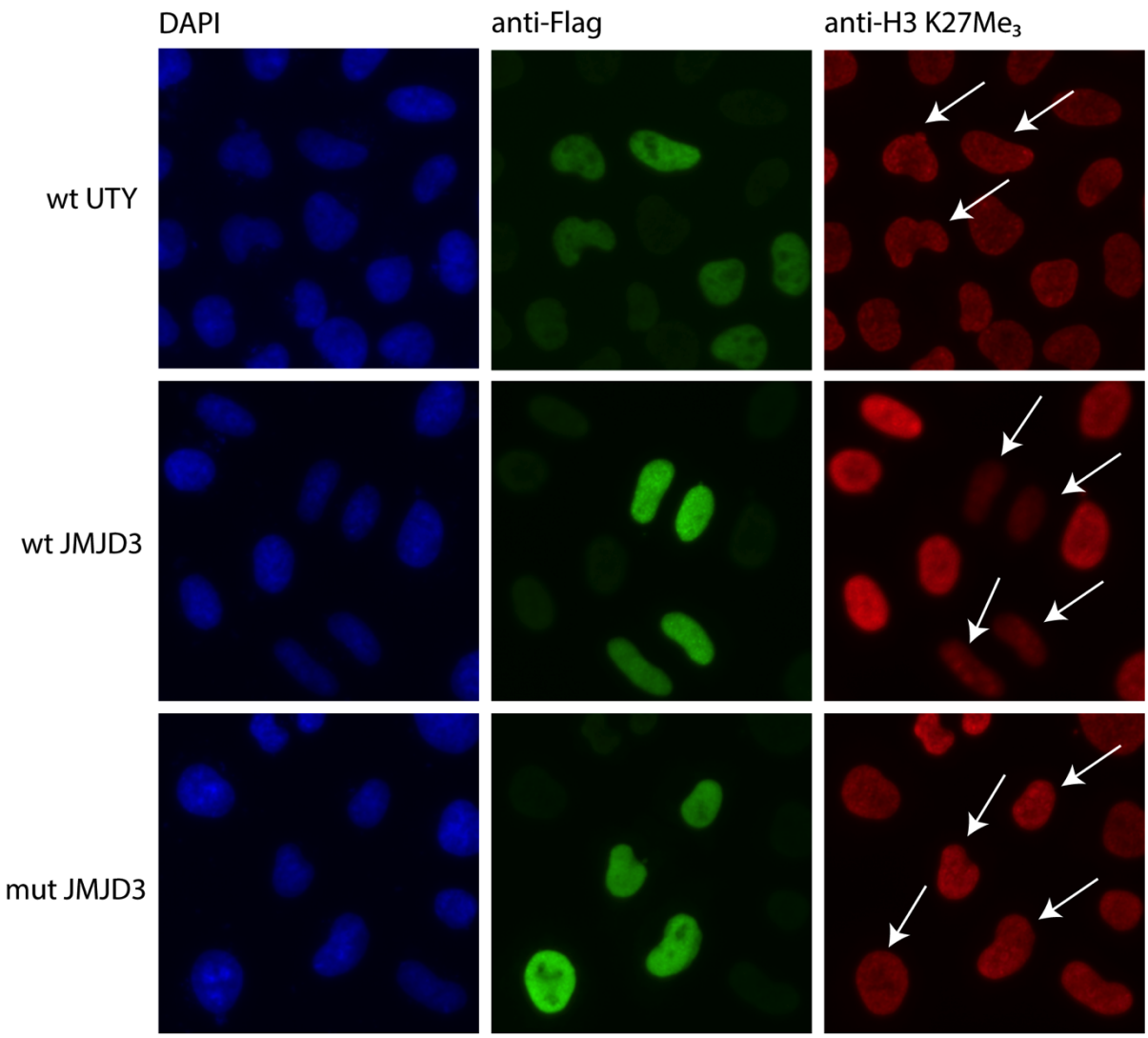


Figure 6.16 Immunofluorescence measuring H3 K27Me$_3$ levels in Hela cells. Nuclei are shown in blue, FLAG-tagged protein in green and H3 K27Me$_3$ levels in red. The antibody responsible for each set of staining is shown above each column and the exogenously overexpressed protein beside each row. White arrows identify some of the cells overexpressing the indicated KDMs.

6.6.2 Effects of UTY on ANF expression

Recently, UTX has been shown to be involved in the differentiation of cells towards cardiac lineages.[229] In this study, Lee *et al*. used a dual luciferase assay to measure the expression from the promoter of atrial natriuretic factor (ANF), a marker of differentiating chamber myocardium. Enhanced ANF-promotor coupled luciferase expression was observed in the presence of UTX and the transcription factor Nkx2.5.

Shpargel *et al.* have independently shown that increased ANF expression was also observed with exogenous expression of UTY but to a lesser extent than for UTX, which they ascribe to a non-catalytic effect of the enzyme.[376] Given the discovery that UTY is an active histone demethylase *in vitro,* these experiments were repeated also with the catalytically inactive UTY variant to investigate whether the catalytic activity of UTY is required for this expression enhancement.

All luciferase and transcription factor plasmids were generously provided by Benoit Bruneau.[382] The initial experiments aimed to verify that ANF expression could be enhanced by overexpression of FL wt UTY and investigate the effect of the catalytically inactive (H1095A) UTY variant. Equivalent numbers of HEK 293T cells were seeded in 96-well plates and transfected at sub-confluency after 24 hours with both luciferase plasmids (ANF-Firefly luciferase and CMV-*Renilla* luciferase) and varying amounts of either FL wt or H1095A UTY. After a further 24 hours, cells were lysed and the expression of the firefly and *Renilla* luciferases quantified. Each experiment was run with triplicate independent transfections, and data is shown for the mean of three biological replicates (Section 8.7.7). Overexpression of wt UTY caused an increase in ANF expression, which was not observed with exogenous expression of the catalytic mutant (Figure 6.17). Transfection of an increased amount of wt FL UTY DNA caused a further increase in ANF-coupled luciferase expression. This pointed towards UTY having a catalytic function in cells, although it did not directly link it to histone demethylase activity.

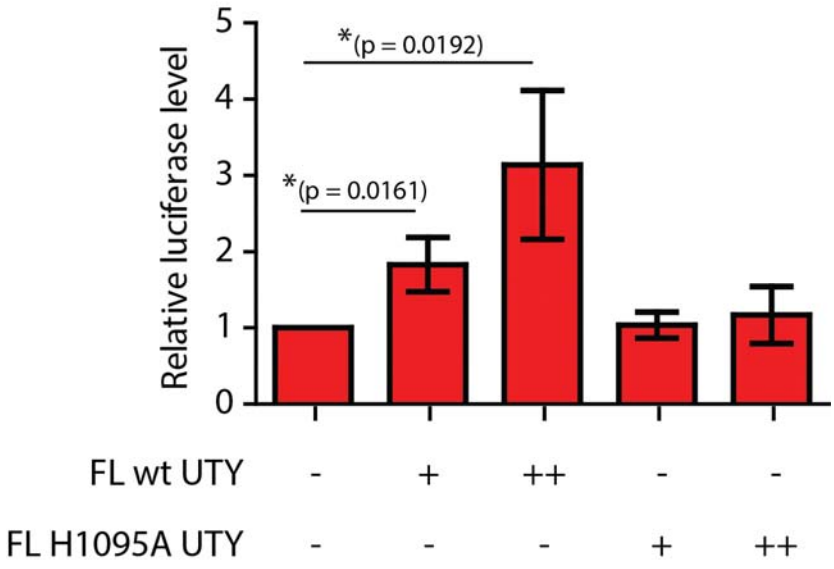


Figure 6.17 Dual luciferase assay in HEK 293T cells measuring the expression levels of ANF-promotor coupled luciferase in the presence of FL wt or H1095A UTY. Data is plotted as the mean of three biological replicates and error bars show ± s.d.. * gives p values calculated using an unpaired *t* test.

In the previous studies, enhancement of ANF expression by UTX was observed when coupled to expression of the transcription factor, Nkx2.5.[229,376] As such, the two luciferase plasmids and a plasmid encoding for Nkx2.5 were all transfected into HEK 293T cells with or without either FL wt or H1095A UTY to investigate what effect this had on the luciferase reporter. Initially, the experiment was carried out with titration of the Nkx2.5 plasmid alone to optimize the amount of DNA required at two different concentrations of ANF-luciferase plasmid (Figure 6.18). Since the enhancement caused by UTY was expected to be much smaller than that caused by Nkx2.5, it was desirable to use a quantity of plasmid that only gave a small enhancement of expression so as not to outweigh the smaller effect caused by UTY. As shown in Figure 6.18, 0.05 µg Nkx2.5 plasmid and 0.1 µg ANF-luciferase plasmid was selected for transfections in further experiments as they seemed to provide sufficient enhancement to be measured reliably, but without being so large that it would be impossible to see a further smaller enhancement caused by UTY.

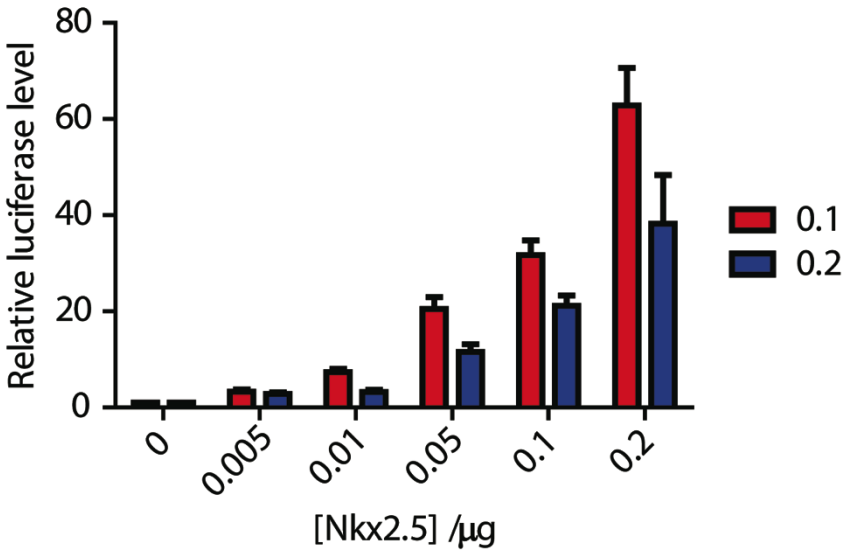


Figure 6.18 Dual luciferase assay in HEK 293T cells measuring the expression levels of ANF-promotor coupled luciferase at different concentrations of transfected Nkx2.5. 0.1 µg (red) or 0.2 µg ANF-luciferase plasmid were was used. Data is plotted as the mean of triplicate transfections and error bars show ± s.d..

Using 0.05 µg Nkx2.5 plasmid and 0.1 µg ANF-luciferase plasmid, the effect of exogenous UTY overexpression on ANF-coupled luciferase expression was investigated. Unfortunately, unlike in previously published studies,[376] a significant further increase in luciferase expression on exogenous overexpression of wt FL UTY was not observed (Figure 6.19).

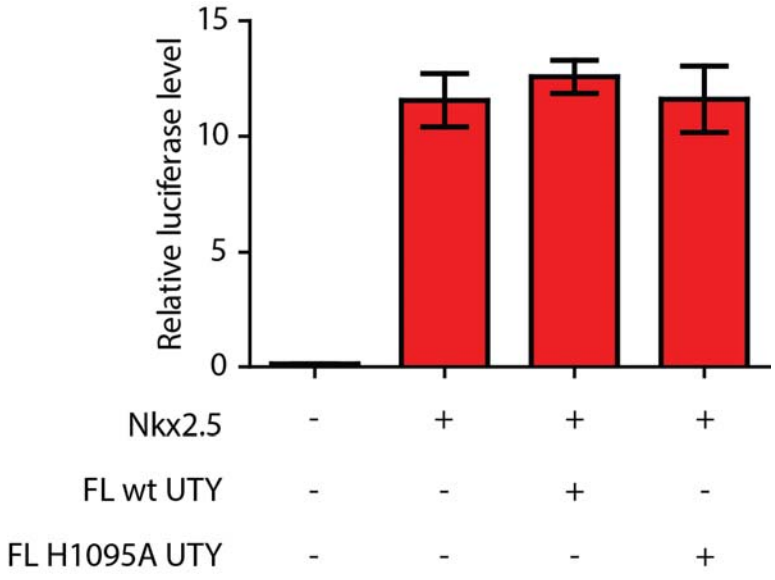


Figure 6.19 Dual luciferase assay in HEK 293T cells measuring the expression levels of ANF-promotor coupled luciferase in the presence of Nkx2.5 and FL wt UTY or FL H1095A UTY. Data is plotted as the mean of triplicate transfections and error bars show ± s.d..

6.6.3 Studies on the expression levels of HOXA genes by qRT-PCR

Previous ChIP experiments with UTX have shown that it occupies the promoters of homeobox (HOX) gene clusters and furthermore, knockdown or overexpression of UTX alters the mRNA levels of a number of these genes in several cell types.[225] ChIP experiments have shown that UTX binding occurs concomitantly with decreases in H3 $K27Me_3$ at these gene promoters. Given the similar demethylation activity of UTY *in vitro* the effect of overexpression of UTY on HOX gene expression in cell culture was tested.

The HOXA gene cluster contains 11 genes that are involved in body patterning during development.[383] Primers were designed to measure the expression of the HOXA gene cluster and housekeeping genes in HEK 293T cells. Initially, linearity of the primers was confirmed in untreated HEK 293T cells and the dissociation curves were inspected to check for a single amplification product, which was further confirmed by agarose gel electrophoresis of the PCR products. Next cells were cultured with transient overexpression of either FL wt UTY or an empty Flag control vector for 24 hours. To allow comparison of gene expression between experiments, gene expression levels must be normalised to a housekeeping gene whose expression is known not to change as a result of the cell treatment conditions being investigated. As epigenetic regulators can modify the expression levels of a large number of genes, it is particularly important to check that the enzyme of interest does not affect the expression of any gene chosen as a "housekeeping" gene.[384] Four commonly used housekeeping genes, GAPDH, PolR2a, β-actin and B2M, were selected and, using identical quantities of input cDNA, their expression levels in the cells expressing either FL UTY or Flag control were compared (Section 8.10). All four housekeeping

genes appeared unaffected by UTY overexpression as can be seen by the similar Ct value (Figure 6.20). UTY overexpression in these cells was also confirmed using primers to the UTY sequence (Figure 6.20). It is desirable to have a housekeeping gene that is expressed at a similar cycle number to your genes of interest. As the HOXA genes are expressed at low levels in HEK 293T cells and thus appear at relatively high cycle numbers, PolR2a was selected as the most suitable housekeeping gene.

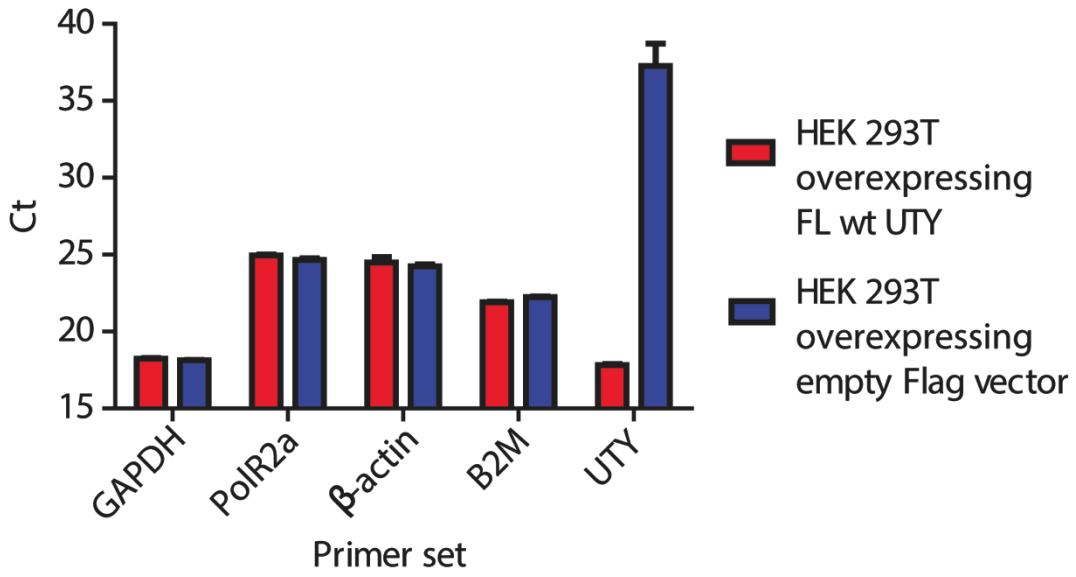


Figure 6.20 qRT-PCR measuring the expression of UTY and housekeeping genes from HEK 293T cells. Experiments were carried out in triplicate and data is shown as the mean of two biological replicates. Error bars show ± s.d. of the mean.

Expression of HOXA genes in the two samples was then quantified by qRT-PCR, relative to PolR2a. qRT-PCR experiments were carried out in triplicate and the data is shown for two biological replicates. Unfortunately no changes in expression levels were observed for any of the HOX genes investigated (Figure 6.21).

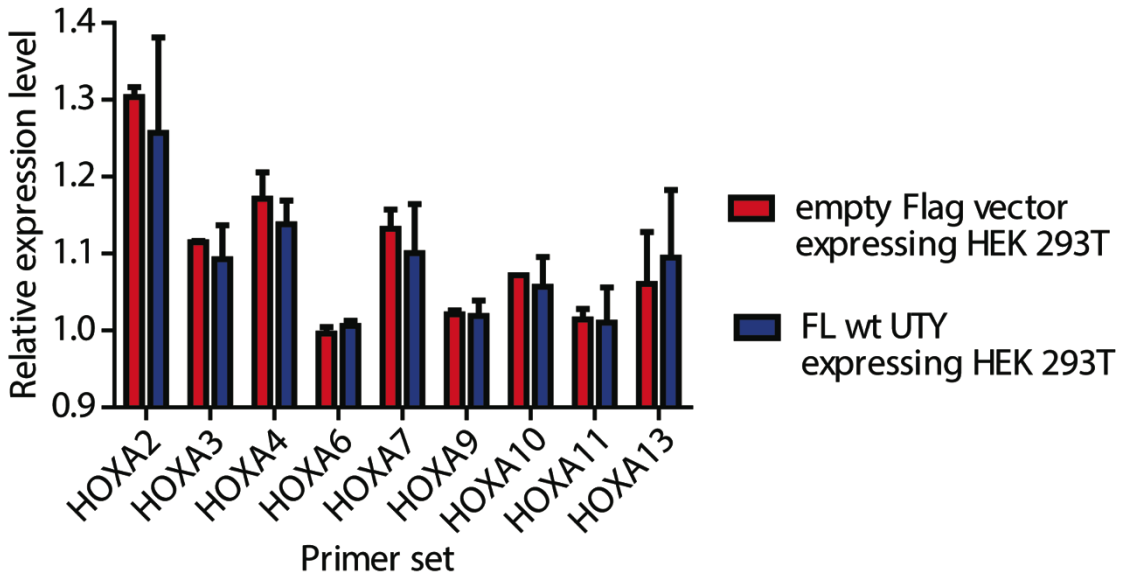


Figure 6.21 qRT-PCR experiments measuring the relative expression levels of HOXA genes in HEK 293T cells expressing exogenous FL wt UTY or an empty Flag vector. Experiments were carried out in triplicate with data plotted as the mean of biological duplicates. Error bars show ± s.d. of the mean.

6.7 Discussion

In this Chapter work towards the characterisation of UTY has been described. UTY was found to be an active member of the KDM6 subfamily of JmjC demethylases. Having identified UTY as a novel histone H3 K27Me$_3$ demethylase, the activity of UTY was compared to the other members of the KDM6 subfamily, UTX and JMJD3. Despite having similar levels of prime substrate uncoupled 2OG turnover and K_M*(2OG)* values, the demethylation rate of UTY was found to be substantially lower than for the other enzymes. *In vitro*, mutagenesis studies identified this to be at least partly due to substitution of an isoleucyl-residue in JMJD3 and UTX for a prolyl-residue in UTY. A crystal structure of UTY was found to be very similar to that of UTX (overall RMSD value at C_α = 0.93), with no overall structural differences. It is thus likely that this difference in activity between UTX and UTY is

partly due to a reduction in the hydrophobic interaction between UTY and its histone peptide substrate. This could be confirmed by carrying out a binding experiment, such as fluorescence polarisation or ITC to measure the peptide K_D values. Even with this mutation, the K_M*(peptide)* for the variant UTY was still measured to be higher than for UTX and JMJD3 suggesting that it is a combination of a number of small changes in the amino acid sequence of UTY that account for the reduced activity and that the singly modified H3 K27Me$_3$ peptide alone may not be the optimal substrate for UTY, at least *in vitro*. The binding partner of the TPR domains of UTY (and UTX and JMJD3) are currently unknown, but in many other chromatin modifying proteins these "reader" domains help with targeting proteins to their substrates. These TPR domains may anchor UTY close to the histone tail *in vivo*, helping to overcome the reduced binding affinity of the JmjC domain itself for this substrate. In addition, UTY has been shown to associate with the multi-protein MLL3/MLL4 H3 K4 methyltransferase complex.[376,385] This complex interacts with histones, and so the presence of UTY within this complex may help to position UTY in a favourable position for catalysing demethylation under specific circumstances. Purely *in vitro* experiments are unable to investigate these possibilities. If additional non-histone substrates exist, coimmunoprecipitation experiments to look for cellular binding partners that may be methylated and proteomics experiments looking at changes in protein lysine methylation on knockdown / overexpression of UTY could be used to investigate the possibility of further non-histone substrates of UTY.

Immunofluorescence experiments in HeLa cells were unable to show any direct global histone demethylation. However, a dual luciferase assay looking at the effect of UTY on ANF promotor expression in HEK 293T cells showed an enhancement of

expression when wt UTY was present, but not when the catalytically inactive H1095A UTY was present. This points towards UTY having a catalytic effect in cells, although does not directly link this to H3 K27 methylation levels as luciferase assays only give a transcription / translation readout. ChIP on the ANF-luciferase plasmid could now be carried out using an anti-H3 K27Me_3 antibody to measure potential changes in methylation level catalysed by UTY. Using cardiomyocytes or other cells known to express ANF, the expression levels from the ANF gene itself could also be investigated to look directly at the effect of UTY levels on its expression.

qRT-PCR experiments were unable to find an effect of UTY on HOXA gene expression. This experiment was limited by the small sample of genes measured and having looked at only a very few genes, this does not eliminate the possibility that UTY is acting as a demethylase in some contexts. In addition, HOX gene expression is low in differentiated cell lines, and it is thus possible that although no effects on expression levels was observed in differentiated HEK 293T cells, in ES cells this situation might be different. To confirm that UTY is a histone demethylase in cells and to determine its cellular gene targets it will be necessary to undertake further experiments including ChIP-seq-type studies. Both the HeLa and HEK 293T cell lines used in cell-based experiments were female-derived. As such they contained no endogenous UTY and will also have lacked other male-specific gene products. Some of these may be essential for the endogenous functioning of UTY and it is thus possible that this contributed to the lack of catalytic function observed in the cells used in experiments described in this thesis. Further experiments could be carried out in male derived embryonic stem cells to address this possibility.

The work detailed in this Chapter has demonstrated UTY to have functional redundancy with UTX *in vitro*. It remains to be confirmed whether this is the case in cells and whole organisms. It has been proposed that to prevent dosage imbalance caused by X-inactivation, Y-homologues of proteins on the X chromosome often evolve to lack activity.[386] In cases where the X-homologue escapes inactivation a dosage imbalance between sexes exists if there is not an active Y-homologue. The evidence that UTX escapes X-inactivation, might suggest that UTY has some functional redundancy with UTX to avoid any consequences of this dosage imbalance. In addition, the evidence that loss of UTX in female mice has more severe consequences than loss in male mice that will also have UTY further suggests that these proteins have some functional redundancy *in vivo* as well (although mouse UTY differs more significantly from UTX and JMJD3 than human UTY, so it is unclear whether recombinant mouse UTY would be active).[209,228,229,376] Although in the literature this redundancy has so far only been demonstrated to be non-catalytic functional overlap, work in this Chapter suggests that it is likely to include overlap of demethylase activity as well. In mice, UTX has been shown to activate expression of female specific RHOX genes involved in sexual reproduction.[387] It is possible, therefore, given the ability of UTY to demethylate H3 K27Me_3 and thus activate repressed genes, that UTY may be required in males for sex determination during development. Further studies are required to elucidate the full role of UTY in histone demethylation in cells.

The work described in this Chapter is summarised in "Human UTY is a Male-Specific N^{ε}-Methyl Lysyl-Demethylase" (Appendix 4).

Chapter 7
Summary

The human epigenetic response is a complex process, the full regulation of which is only just beginning to be elucidated. Work described in this thesis has focused on the structural and biochemical characterisation of a number of proteins involved in the human epigenetic response.

Experiments in the first two Chapters were directed towards study of the "reader" domains of the nucleosome remodeler CHD4. Initially, the solution structure of the tandem PHD fingers of CHD4 was solved using NMR spectroscopy. This revealed the domains to be individually structured with a flexible linker between them, allowing adoption of a large range of different conformations of the two domains relative to each other and a variable inter-domain distance (Figure 2.16). Interestingly, work in Chapter 3 showed that binding to histone peptides does not induce any inter-domain contacts, unlike for the tandem PHDs of DPF3B and MOZ.[216,217] These binding experiments identified both PHD fingers as histone-binding domains that interact with the same region of the *N*-terminus of the H3 tail. NMR studies of a larger construct of CHD4 containing the tandem PHDs and CHDs demonstrated that, despite structural changes, the larger construct was still able to bind to two separate H3 peptides, showing similar affinities to the tandem PHD construct in isolation.

SAXS provides structural data on the size and shape of macromolecules in solution. Future work could make use of SAXS data already collected for the dPHD construct, to select for an ensemble of domain conformations from those calculated in the NMR

structure calculations that best represent the structure of the construct as a whole in solution. At their most extended the two PHD fingers can span a large enough distance to bind to the two H3 tails contained in a single nucleosome (Figure 2.18). The dual histone-binding ability of these domains is thus consistent with two different models of CHD4's mode of action; either the two domains bind to individual H3 tails on the same nucleosome, or each binds to an H3 tail on neighbouring nucleosomes, possibly helping to mediate nucleosome sliding catalysed by the ATPase domain (Figure 3.18). A recent report suggests that the former may be more likely, but further work is required to conclusively elucidate which of these possibilities occurs *in vivo*.[239]

Work carried out in Chapters 4 and 6 addressed the biochemical characterization of three different JmjC 2OG oxygenases. Initial biophysical experiments with PHF8 and KIAA1718, two JmjC histone lysine demethylases from the KDM7 subfamily of KDMs, provided kinetic parameters for use in inhibitor assay development. These were used to develop a high-throughput screening assay to identify inhibitors of PHF8 activity, using the ALPHA screen, and a complementary MALDI MS-based assay for measuring inhibition of KIAA1718. The ALPHA screen was used to identify two inhibitors that are selective for the KDM2 / 7 subfamilies over other JmjC KDM subfamilies. Although the first, Daminozide, was not active in cells, experiments with the second demonstrated a dose-dependent increase in global H3 K27Me$_2$ levels, consistent with inhibition of the demethylase activity of KIAA1718. These inhibitors will be useful for elucidating the roles of the KDM2 / 7 subfamilies in cells and in development, including using studies in zebrafish.

Work described in Chapter 6 involved studies of UTY, the Y-linked member of the KDM6 subfamily of JmjC KDMs, which was previously thought to be catalytically inactive.[225,226] However, experiments described in this Chapter identified UTY to be catalytically active *in vitro* and capable of demethylating H3 peptides trimethylated at K27. A crystal structure of the JmjC and zinc-binding domains of UTY showed it to have a similar overall structure to JMJD3 and UTX, suggesting that subtle differences are responsible for the reduced demethylase activity observed, rather than global structural changes (Figure 6.15). In keeping with this proposal, a UTY variant, P1214I, showed much enhanced activity relative to wt UTY, likely due to tighter peptide binding. Immunofluorescence experiments in cells using exogenously overexpressed FL UTY did not show any changes in global H3 K27Me_3 levels, but a luciferase assay looking at expression from the promotor of ANF showed enhanced expression in the presence of wt UTY, but not a catalytically-inactive mutant, H1095A UTY, suggesting that UTY has catalytic activity in cells (Figure 6.17). Future ChIP-seq studies could be used to identify UTY gene targets in cells. The discovery that UTY is an active KDM will be an important consideration in interpretation of studies looking at the function of the KDM6 subfamily members in cells and whole organisms.

In Chapter 4 a novel clinical frame-shift mutation in the PHF8 gene was investigated. The mutation was demonstrated to cause mislocalisation of exogenously overexpressed PHF8 to the cytoplasm in cells, probably due to loss of its five annotated nuclear localisation signals. The observed patient phenotype is likely caused by the truncated cytoplasmic PHF8 being unable to catalyse demethylation of nuclear H3 K9Me_2 at its specific gene targets, resulting in altered protein expression

during development. Mislocalisation of PHF8 to the cytoplasm will allow it to interact with many proteins that it does not usually encounter due to its nuclear localisation. In addition to a reduction in histone demethylation, binding of PHF8 to some of these proteins *via* its JmjC domain or PHD finger may also contribute to the observed phenotype.

The remainder of Chapter 4 and Chapter 5 addressed novel reactions of JmjC KDMs. The reaction of PHF8 with a series of lysine analogues with N^{ε}-post-translational modifications other than methyl groups was studied. MALDI MS-based experiments demonstrated that PHF8 can oxidatively remove ethyl and isopropyl groups from the lysine side chain and can demethylate dimethylornithine when it replaces lysine at the K9 position in a H3 peptide sequence (Section 4.8). The reaction of JMJD2E with methylisopropyllysine was shown to form a stably hydroxylated product, helping to validate the proposed hydroxylation-mediated mechanism of lysine demethylation (Figure 4.30). The structural similarity between this residue and asymmetric dimethylarginine led to the discovery that JMJD2E can demethylate methylated arginyl-residues. Following from this discovery, work in Chapter 5 described the screening of a panel of recombinant 2OG oxygenases for potential arginine demethylase activity on peptide substrates. Members of the KDM4 (JMJD2A, JMJD2D and JMJD2D), KDM5 (JARID1C) and KDM6 (JMJD3) subfamilies of JmjC KDMs were found to catalyse arginine demethylation *in vitro*, although the reaction was less efficient than the demethylation of histone lysyl-residues. Both JARID1C and JMJD2E appeared to show different sequence preferences for the residues around the target arginyl-residue compared to around the lysyl-residues and it is thus possible that the reaction will be more efficient when a biologically relevant

substrate is identified. In addition, many of the JmjC KDMs contain auxiliary binding domains that direct substrate targeting and the presence of these domains may improve the efficiency of the demethylation reaction. Future experiments will seek to identify biologically relevant substrates directly from cells using a SILAC-based MS assay developed in Chapter 6 (Section 5.10.8).

The JmjC KDMs are considered to be interesting drug targets and there is currently much focus on developing selective compounds to inhibit these enzymes. The detrimental effects of loss or mutation of these enzymes during development, such as occurs with PHF8, suggest that availability of agonists may also be valuable. For successful therapeutic intervention a detailed understanding of the role of epigenetic proteins in development and disease is desirable. The discovery that arginine methylation can be actively removed will also have important consequences for the role of this post-translational mark in signalling and disease. Together, the work described in this thesis has characterised novel activities for the JmjC KDMs suggesting that their functions may be more diverse than previously thought. This broad activity of JmjC KDMs is consistent with other chromatin-modifying enzymes, such as the PRMTs, which have multiple cellular targets.[328] As such, pan-inhibition of any given KDM's activity on all targets by inhibition at the active site may not always be desirable. If target regulation is mediated by protein-protein interactions with "reader" domains contained alongside the JmjC domains in KDMs, inhibition of these contacts may be useful for target specific inhibition.

Chapter 8
Materials and methods

All chemicals were purchased from Sigma-Aldrich unless otherwise stated. Tryptone, yeast extract and agar were obtained from Oxoid. Water was purified using a Millipore Elix® 10 system and then passed through a 0.22 μm filter by a Millipore Milli-Q system. Throughout, standard sterile practices were followed. Equipment and media were sterilised by autoclaving for 20 minutes at 121 °C. For heat-sensitive solutions sterilisation was performed using a 0.22 μm filter.

8.1 General Procedures:

pH measurement:

pH measurements were made using a Hanna Instruments HI 9321 microprocessor pH meter fitted with a 5 mm diameter electrode. Before use the probe was calibrated in either the range pH 4-7 or pH 7-10, as appropriate, using standard calibration solutions (Fisher Scientific) and stored in 4 M KCl solution between uses. pH of solutions was adjusted with stock solutions of HCl or NaOH (or DCl and NaOD for NMR samples in D_2O) as necessary.

Waterbaths:

A Grant W14 waterbath was used for transformations at 42 °C, and a Grant Y14 waterbath was used for incubations at 37 °C.

8.2 SDS-PAGE gel electrophoresis:

8.2.1 Reagents:

SDS-PAGE gels (1 gel):

Reagent	Resolving gel (18% / 15 % / 12 %)	Stacking gel
Tris-HCl (1 M, pH 7.5)	1.25 mL	-
Tris-HCl (0.5 M, pH 6.8)	-	0.63 mL
Milli-Q water	0.65 mL / 1.15 mL / 1.65 mL	1.57 mL
30% (w/v) acrylamide / 1.034% (w/v) bis-acrylamide stock	3 mL / 2.5 mL / 2 mL	0.25 mL
SDS (10% (w/v))	50 µL	25 µL
10% (w/v) APS	50 µL	25 µL
TEMED	4 µL	2.5 µL

TEMED and APS were added just prior to gel casting. Resolving gel % was determined based on the size of the proteins to be separated.

10 × SDS-PAGE running buffer (1 L):

Reagent	Amount /g
Tris-Base	30
Glycine	144
SDS	10

SDS-PAGE gel destain:

Reagent	Amount
Acetic acid	10% (w/v)
Methanol	30% (w/v)

SDS-PAGE gel stain:

As for gel destain with addition of 0.25% (w/v) Coomassie® brilliant blue R

SDS-PAGE sample loading buffer (100 mL):

Reagent	Amount
Tris-HCl (0.5 M, pH 6.8)	10 mL
Bromophenol blue (0.2% (w/v))	0.2 g
SDS (10% (w/v))	20 mL
Glycerol	12 mL
β-mercaptoethanol	5 mL

8.2.2 Method:

Samples were heated for 5 min at 100 °C before use. Polyacrylamide gels were run under denaturing conditions for 40-60 min at RT, 180 V using a Bio-Rad Mini-PROTEAN II system. Gels were stained with SDS-PAGE gel stain for 20 min and then destained with SDS-PAGE destain overnight.

8.3 Agarose gel electrophoresis

8.3.1 Reagents:

50 x TAE (100 mL):

Reagent	Amount
Tris-Base	24.2 g
Glacial acetic acid	5.71 mL
EDTA	1.86 g

5 x DNA sample loading buffer (5 mL):

Reagent	Amount
TAE (1 ×)	1.5 mL
Glycerol	3.5 mL
Bromophenol blue (0.25% (w/v))	12.5 mg

8.3.2 Method:

1% agarose gels were used for DNA visualisation. They were formed by heating 0.4 g agarose (Bioline) in 40 mL TAE buffer until dissolved, using a SuperWave 750 microwave (Tecnolec). This was allowed to cool to around 60 °C, before addition of 4 µl SYBR Safe™ DNA gel stain (Invitrogen). Gels were cast in an H2-SET gel tray (Anachem) to a depth of 0.7 cm with a comb inserted at one end of the gel to enable sample loading and allowed to cool at room temperature for the gel to set. Gels were run on a Bio-Rad system at a constant potential of 120 V for approximately 35 min.

Bands were visualised by UV light illumination using a Gel Logic 200 imaging system (Kodak) and a 535 nm emission filter. Samples were run against a 2-log DNA Ladder 0.1-10 kb (Invitrogen) allowing molecular weight determination.

8.4 Recombinant enzyme expression in *E. coli*

8.4.1 Reagents:

Vectors:

Vector name	Antibiotic resistance	Tag	Tag cleavage site	Promotor	Proteins
pET28a	Kanamycin	*N*-terminal 6×His tag	Thrombin	T7	TEV
pNIC28-BSA4	Kanamycin	*N*-terminal 6×His tag	TEV	T7	hnRNP K, PRMT1, UTY, KIAA1718
pNH-TrxT	Kanamycin	*N*-terminal 6×His-Trx tag	TEV	T7	PHF8, UTY, JMJD3
pcDNA3-N-Flag-LIC	Ampicillin	*N*-Flag-tag	-	CMV	UTY
pcDNA3-N-GFP-LIC	Ampicillin	*N*-GFP-tag	-	CMV	hnRNP K
pEGFP-C1	Ampicillin	*N*-GFP-tag	-	CMV	PRMT1
pEF6	Ampicillin	*N*-HA-tag	-	CMV	PHF8
pTriEx2	Kanamycin	*C*-terminal 8×His tag	-	T7	dPHD, dPHDdCHD

Antibiotics:

Antibiotics were used at a final concentration of 100 $\mu g.mL^{-1}$ of Ampicillin or 30 $\mu g.mL^{-1}$ of Kanamycin.

Incubations:

Large-scale incubations were carried out in a New Brunswick Scientific Innova 44 incubator at 37 °C with 200 rpm shaking. Small-scale incubations (100 mL or less) were carried out in a New Brunswick Series 25 incubator.

2 × Tryptone-Yeast (2-TY) media (1 L):

2-TY media used for large-scale growths was prepared by Mr H. M. Jubier.

Reagent	Amount /g
Bacto-tryptone	16
Yeast extract	10
NaCl	5

To produce solid medium for plates, 20 g agar was added to the media before autoclaving.

Terrific Broth (TB) media (1 L):

UTY and FL PHF8 were expressed in TB media.

Reagent	Amount
Bacto-tryptone	12 g
Yeast extract	24 g
Glycerol	4 mL

Solids were dissolved in 900 mL water. After autoclaving, 100 mL filter sterilized 0.17 M KH_2PO_4, 0.72 M K_2HPO_4 solution was added.

Minimal media for NMR spectroscoy (1L):

Reagent	Amount /g
Na_2HPO_4	6
KH_2PO_4	3
NaCl	0.5

After autoclaving, to each 500 mL was added 0.5 mL $CaCl_2$ (0.1M), 1 mL $MgSO_4$ (1M), 10 mL glucose (20% w/v), 10 mL Yeast Nitrogen Base and 7.5 mL NH_4Cl, all of which were first filter sterilised. For labelling $^{14}NH_4Cl$ and glucose were replaced by their labeled counterparts, $^{15}NH_4Cl$ (99% labeled, Cambridge Isotope Laboratories) and ^{13}C-glucose (99% labeled, Cambridge Isotope Laboratories).

8.4.2 Methods:

Optical density measurements (OD_{600}):

OD_{600} values were measured in a 1.6 mL cuvette, against a reference sample of initial growth medium in a Novaspec® II spectrophotometer (Pharmacia).

Glycerol stocks:

Samples of un-induced cells were preserved by gently mixing 750 µL of culture with 250 µL sterile glycerol and flash freezing in liquid nitrogen before storage at -80 °C.

Transformations:

XL10-Gold ultracompetent cells were used for production of DNA and BL21 (DE3) cells for bacterial expression of His-tagged proteins. Competent cells were thawed on ice. 1 µL of plasmid DNA (100 ng. μL^{-1}) was added to 25 µL cells in a pre-chilled 1.6 mL sterile 1.6 mL Eppendorf tube. Tubes were left on ice for 20 minutes before being "heat shocked" in a water bath at 42 °C for the 30 s. Cells were left on ice for a further 2 minutes before 0.4 mL 2-TY media was added. The mixture was incubated for 1 hr at 37 °C with shaking. 250 µL of the mixture was plated onto agar plates containing kanamycin or ampicillin. For LIC cloning plates also contained 5 % sucrose. The plates were incubated at 37 °C overnight before use.

Production of competent cells:

Competent cells were prepared from commercial stocks (Stratagene). Bacteria were streaked onto antibiotic-free agar plates and grown overnight. A single colony from the freshly-prepared plates was used to inoculate 100 mL sterile antibiotic-free 2TY medium and grown overnight (37 °C, 220 rpm). Four 100 mL 2TY cultures were inoculated with 200 µL starter culture and the medium grown (37 °C, 220 rpm) to an OD_{600} of 0.6. Pre-chilled centrifuge tubes were used to pellet cells (750 × g, 5 min,

4 °C). Pellets were resuspended in 200 mL ice-cold 100 mL $CaCl_2$, 20 % glycerol and incubated on ice for 20 min. Cells were again pelleted (750 × g, 5 min, 4 °C) and resuspended in 8 mL of the same solution. 50 µL aliquots of cells were flash frozen in liquid nitrogen and stored at -80 °C for future use.

Starter cultures:

100 mL 2-TY media containing the appropriate antibiotic was inoculated with a single colony from an agar plate or a scraping from a glycerol stock and then incubated overnight in a New Brunswick Scientific G25 environmental shaker at 37 °C and 250 rpm shaking.

Expression trials:

500 mL flasks containing 100 mL 2-TY media were inoculated with the appropriate antibiotic and 1 mL starter culture. At OD_{600} ~ 0.6, cultures were separated into incubators held at a variety of temperatures (typically 18 °C and 37 °C) and each induced with variable concentrations of IPTG (typically 0.2 mM, 0.5 mM or 1 mM). Cultures were incubated for four hours at 37 °C or at 18 °C overnight. 1.6 mL of each culture was transferred to a 1.6 mL Eppendorf tube and pelleted. The spent media was removed and the cells resuspended in 300 µL His column binding buffer. 25 µL was removed for analysis as whole cell extract and the remainder was lysed by sonication (2 × 15 s with 20 s interval) on ice. Cell debris was removed by centrifugation and the supernatant removed for analysis as cell lysate. Soluble expression levels were assessed by SDS-PAGE gel electrophoresis using 15 µL samples of the whole cell extract and cell lysate.

Large scale growths:

12 × 2 L flasks containing 600 mL 2-TY media and the appropriate antibiotic were each inoculated with 6 mL starter culture. At OD_{600} ~ 0.6, protein expression was induced using an appropriate amount of IPTG (given below, Melford Laboratories Ltd) as determined during expression trials. Cultures were incubated for four hours or overnight as appropriate. Cells were then pelleted by centrifugation and stored in a resealable plastic bag at –80 °C.

Construct	[IPTG] /mM	Induction temperature /°C	Induction length /hr
FL PHF8	0.3	18	16
JmjC of PHF8	0.1	18	16
PHD of PHF8	0.3	18	16
FL KIAA1718	0.2	18	16
UTY	0.2	18	16
PRMT1	0.1	18	16
hnRNP K	0.1	18	16

8.5 Recombinant enzyme purification

8.5.1 Reagents:

Ni His-trap purification buffers:

Reagents	Charge	Binding	Wash	Elution	Strip
HEPES	-	50 mM	50 mM	50 mM	-
NaCl	-	0.5 M	0.5 M	0.5 M	-
Imidazole	-	20 mM	40 mM	250 mM	-
$NiSO_4$	50 mM	-	-	-	-
EDTA	-	-	-	-	100 mM
Glycerol		5%	5%	5%	

1 mM DTT was added to buffers immediately prior to purification of some proteins.

Size exclusion chromatography buffers:

All buffers contained 10 mM HEPES, pH 7.5, 5% glycerol and a variable concentration of NaCl, dependent on the protein being purified (500 mM PHF8, KIAA1718, 200 mM JMJD3, UTY, PRMT1, hnRNP K). 1 mM DTT was added to buffers immediately prior to purification of some proteins.

Ion exchange chromatography buffers:

Protein was diluted into buffer containing 50 mM HEPES, pH 7.5, 50 mM NaCl. Elution was carried out using buffer containing 50 mM HEPES, pH 7.5, 1 M NaCl

8.5.2 Methods:

All purifications were carried out at 4 °C using AKTA^{TM} FPLC systems. All buffers were filtered with a 0.2 μm filter and buffers for size exclusion chromatography were degassed before use.

Cell pellets from large-scale growths were resuspended in column equilibration buffer (volume equal to 5 × mass used) with 10 μg.mL^{-1} DNase1 and protease inhibitor cocktail tablets (Roche) and lysed on ice by sonication (4 min of 3 s on, 4 s off) using a Vibra Cell VCX 500 with a 13 mm probe. Cell debris was removed by centrifugation (20 min, 14000 rpm) and the supernatant filtered though a 0.4 μm filter before application to a purification column.

6×His tagged protein purification:

Purification was carried out using a 5 mL nickel affinity His-Trap column (Novagen). Buffers were run at 1 mL.min^{-1}. Resin was prepared by washing with 3 CV MilliQ water, then 3 CV charge buffer, then 3 CV MilliQ water, then 4 CV binding buffer before the cell lysate was applied to the column. The column was washed with 6 CV wash buffer before protein elution with 5 CV elution buffer in 2 mL fractions or gradient elution over 10 CV in 5 mL fractions. Sample fractions were analysed by UV spectroscopy and SDS PAGE to determine which fractions contained pure protein and these were pooled and concentrated to between 2 and 10 mL for further purification.

His-tag cleavage by TEV:

His-TEV was stored at 6 mg.mL^{-1}. For tag cleavage TEV was added to His-tagged protein after Ni-affinity chromatography at a ratio of 1:20 (w/w) and incubated overnight at 4 °C. To remove the TEV, cleaved protein was diluted into Ni His-trap binding buffer without any Imidazole to dilute the Imidazole concentration to 20 mM and reapplied to a 5 mL His-Trap column. Cleaved protein was collected in the flow-through and concentrated for further purification.

Size exclusion chromatography:

Up to 10 mL of previously purified protein was loaded using a 10 mL superloop at 0.5 mL.min^{-1}onto a 300 mL size exclusion column containing S75 (for PHD of PHF8 and dPHD) or S200 (for all other proteins) size exclusion resin (Amersham Biosciences), pre-equilibrated with 1.1 CV water followed by 1.1 CV gel filtration buffer. The protein was eluted at a flow rate of 1 mL.min^{-1} using 1 CV gel filtration buffer. 5 mL fraction were collected and fractions analysed by SDS-PAGE on the basis of the UV trace. Pure protein was concentrated and frozen at -80 °C for future use.

Ion exchange chromatography:

Protein from size exclusion chromatography was diluted into ion exchange buffer to a final NaCl concentration of 50 mM and applied to a 10 mL QSepharose column at 5 mL.min^{-1}. Protein was washed with 5 CV binding buffer and eluted with a gradient up to 500 mM NaCl over 10 CV. 5 mL fraction were collected and analysed by SDS-PAGE on the basis of the UV trace. Pure protein was concentrated and frozen in aliquots at -80 °C for future use.

Protein concentration determination:

Protein concentration was determined using 1.5 µL protein solution on a NanoDrop ND-1000 UV/Vis spectrophotometer to measure absorbance at 280 nm. ε was calculated using the ProtParam tool on the ExPASY Proteomics Server (www.expasy.org) and concentration calculated using the equation, A = ε.c.l.

Bicinchoninic acid assay assay:

Protein concentration of cell lysates or mixtures of proteins was determined using the bicinchoninic acid (BCA) assay. A standard curve was measured each time using variable concentrations of bovine serum albumin (BSA, 20 $\mu g.mL^{-1}$ – 2 $mg.mL^{-1}$). 200 µL of working reagent (50:1 reagent A:reagent B, Thermoscientific) was added to 25 µL of each sample in a black, clear bottomed 96-well plate. Plates were incubated at 37 °C for 30 min, cooled to room temperature and the absorbance at 562 nm measured. Protein concentration was estimated by comparison to the standard curve.

8.6 DNA techniques

8.6.1 Reagents

Oligonucleotide primers:

Oligonucleotide primers were purchased from Sigma Life Science and resuspended in sterile Milli-Q water to give a concentration of 100 µM. The sequence of all primers used in this thesis are given below.

DNA sequencing:

Plasmid sequencing was carried out by Source BioScience Lifesciences.

Oligonucleotide primer sequences:

Construct	Name	Sequence
PHF8 1306dupG	Fwd	AAGATGTGGCCAGGG**G**AGATCCGCCTGG
	Rev	AGGCGGATCT**C**CCCTGGCCAGAGTCTTTAATG
JmjC and Zn UTY	Fwd	tacttccaatccatgCCTTTAGCTAGCCACAGATCTATTC
	Rev	tatccacctttactgTCAAGATGAGGATGATAATGAAAGAG
FL UTY	Fwd	tacttccaatccatgAAATCCTGCGCAGTGTCGCTCAC
	Rev	same as JmjC and Zn UTY Rev
FL H1095A UTY	Fwd	AGTCGGACACCAGGT**GCC**CAAGAAAATAACAAC
	Rev	GAAGTTGTTATTTTCTT**GGC**ACCTGGTGTCCGACTCCC
P1214I UTY	Fwd	AAGTGTGAAGTCA**ATA**GTACCCATGGTGCATC
	Rev	ATGCACCATGGGTAC**TAT**TGACTTCACACTTTTC
S1138G UTY	Fwd	ATTTTTTAATG**GGT**TCTTGGTGGCCCAAC
	Rev	GGGCCACCAAGAA**CC**CCATTAAAAAATTC
hnRNP K	Fwd	tacttccaatccatgGAAACTGAACAGCCAGAAG
	Rev	tatccacctttactgTTAGAATCCTTCAACATC
PRMT1	Fwd	tacttccaatccatgGCGGCAGCCGAGCCGAGGCCGCGAACTGC
	Rev	tatccacctttactgTCAGCGCATCCGGTAGTCGGTGGAGCAGGACAG
PHF8 V255S	Fwd	CTTTGGTGGCACCTCT**AGC**TGGTACCATGTACTC
	Rev	GTACATGGTACCA**GCT**AGAGGTGCCACCAAAGTC
PHF8 V255T	Fwd	CTTTGGTGGCACCTCT**ACC**TGGTACCATGTACTC
	Rev	GTACATGGTACCA**GGT**AGAGGTGCCACCAAAGTC
PHF8 Y257F	Fwd	ACCTCTGTCTGG**TTC**CATGTACTCAAGGGTG
	Rev	CCTTGAGTACATG**GAA**CCAGACAGAGGTGAC
PHF8 Y257A	Fwd	ACCTCTGTCTGG**GCC**CATGTACTCAAG
	Rev	TTGAGTACATG**GGC**CCAGACAGAGGTGAC

All DNA bases are shown in capitals except for LIC overhangs. Alterations from the native DNA sequences are shown in bold.

8.6.2 Methods

DNA concentration determination:

DNA concentration was determined using a NanoDrop ND-1000 UV/Vis spectrophotometer to measure absorbance at 260 nm. The purity of the DNA was assessed by reference to the A_{260}/A_{280} ratio.

Plasmid DNA purification:

Plasmid DNA was purified from 5 mL cell culture using a GeneJET Plasmid Miniprep Kit (Thermo Scientific) following the manufacturer's protocol. DNA was

eluted with 50 µL sterile MilliQ water. A GeneJET Plasmid Maxiprep kit (Thermo Scientific) was used for large scale transfection-grade DNA purification.

DNA amplification:

DNA sequences for LIC were amplified by the PCR.

Reagent	Concentration	Volume / µL
*Pfu*Turbo® reaction buffer	10 ×	5
DNA template	0.1 $\mu g.\mu L^{-1}$	1
Forward primer	10 µM	2.5
Reverse primer	10 µM	2.5
dNTP mix	10 mM	1
Sterile Milli-Q water	-	37
PfuTurbo® DNA polymerase		1

A Techne Genius thermal cycler was programmed to hold at 95 °C for 30 s to denature the template DNA.

40 cycles followed:

Denaturing at 95 °C for 30 s
Annealing at 5 °C below primer T_m for 30 s
Extension at 72 °C for x min (1 min per kb of plasmid)

At the end, reactions were held at 72 °C for a 10 min final extension before holding at 4 °C.

Site-directed mutagenesis:

Site-directed mutagenesis was carried out following the procedure of the QuikChange™ site-directed mutagenesis kit from Stratagene. Primers containing the desired mutation were used to amplify the full vector and insert sequence.

Reagent	Concentration	Volume / µL
*Pfu*Turbo® reaction buffer	10 ×	5
DNA template	0.1 $\mu g.\mu L^{-1}$	1
Forward primer	10 µM	1
Reverse primer	10 µM	1
dNTP mix	10 mM	1
Sterile Milli-Q water	-	40

A Techne Genius thermal cycler was programmed to hold at 95 °C for 30 s to denature the template DNA. 1 µL PfuTurbo® DNA polymerase was then added.

18 cycles followed:

Denaturing at 95 °C for 30 s
Annealing at 55 °C for 1 min
Extension at 72 °C for x min (1 min per kb of plasmid)

At the end, samples were held at 4 °C before 1 µL Dpn1 was added. Tubes were left at 37 °C for 1 hour to digest the parental DNA before samples were transformed into competent *E.coli* for growth and sequencing.

Ligation Independent Cloning (LIC):

LIC provides a framework for cloning the same PCR product into multiple different vectors.[388] Oligonucleotides including LIC sequences at either end were used to amplify the required DNA insert with addition of LIC overhangs by PCR as described above. Where a single PCR product was observed, PCR reactions were purified using a GeneJET PCR Purification kit (Thermo Scientific) according to the manufacturer's protocol before further use. Where multiple products were observed the band corresponding to the correctly sized amplification product was excised and DNA was extracted using a GeneJET Gel Extraction kit (Thermo Scientific), according to the manufacturer's protocol. First, bands were excised from agarose gels using a clean scalpel. DNA was eluted using 30 µL sterile MilliQ water.

Vector DNA was digested with BSA1 in NEB buffer 3 with 0.1 µg BSA for 3 hours at 37 °C. The linearised DNA was purified using a GeneJET PCR Purification kit (Thermo Scientific). Eluted DNA (50 µL) was treated with 1 µL T4 DNA polymerase buffer, 2.5 µL 100 mM dGTP, 1 µL 10 $mg.mL^{-1}$ BSA, 5 µL 100 mM DTT, 5 µL T4 DNA polymerase (Novagen) and 26.5 µL sterile MilliQ water and incubated at 22 °C

for 30 min to generate sticky ends before heat inactivation of the polymerase at 75 °C for 20 min. PCR products were treated in an analogous fashion but with 2.5 µL 100 mM dCTP instead of dGTP to produce complementary sticky ends. 2 µL of the DNA insert was incubated with 1 µL of the linear vector DNA for 10 min at RT and the mixture transformed into competent *E. coli* and plated onto agar plates containing 5% sucrose in addition to the required antibiotic to select for clones that had been correctly linearized (and thus had lost the insert that makes them susceptible to sucrose).

Colony PCR:

Single colonies on transformation plates from LIC reactions were picked and added to a PCR mix containing forward and reverse primers that bind to the vector backbone. The PCR cycle used was the same as described for DNA amplification.

Reagent	Concentration	Volume added / µL
Biomix Red™ (Bioline)	2 ×	10
Forward primer	10 µM	0.5
Reverse primer	10 µM	0.5
Sterile MilliQ water	-	9

PCR products were visualised directly by agarose gel electrophoresis and colonies shown to contain an insert of the required size were grown overnight to allow plasmid extraction for DNA sequencing.

8.7 Activity assays:

8.7.1 MALDI-TOF MS assays:

Assays were initiated by mixing of two stocks. One contained enzyme and $(NH_4)_2Fe(SO_4)_2$, the other contained all other reagents. 50 mM HEPES, pH 7.5 was used to make up the final volume for all enzymes except for UTY where 5% glycerol

was included in the buffer, JMJD3 where 200 mM NaCl was included in the buffer and PHF8 where 100 mM HEPES, pH 7.5, 500 mM NaCl was used. The two solutions were pipetted into sample wells of a 96-well plate to a final volume of 10 µL without mixing the two reagents. Final concentrations varied between experiments, but unless otherwise described were: 2-OG 100 µM; Fe(II) 10 µM; sodium ascorbate 1 mM. All sample wells were mixed simultaneously and incubated at 37 °C with shaking for 1 hr unless otherwise specified. Samples were quenched with 10 µL methanol. 1 µL assay mixture was spotted onto a 12 × 8 target plate and 1 µL CHCA matrix (Laser BioLabs) added. Samples were left to air dry before being analysed by MALDI-TOF MS. Negative controls without enzyme and with no incubation time were included in each experiment. All assays were carried out in triplicate unless otherwise stated.

For inhibition assays, enzymes were pre-incubated with inhibitor for 15 min before initiation of the reaction with all other reagents. Where necessary for inhibitor solubilisation, 1% DMSO was included in the assay buffer. Data were normalised to a no enzyme control.

MALDI matrix CHCA solution:

Reagent	Amount
CHCA (LaserBioLabs)	5 µg
Acetonitrile	350 µL
Trifluoroacetic acid	0.5 µL
Milli-Q water	150 µL

MALDI-TOF MS:

MALDI-TOF MS was carried out using a Waters® MALDI micro MX™ mass spectrometer (Waters Corporation) in positive ion reflectron mode. A laser energy of 220, pulse voltage of 1250 V and detector voltage of 2750 V were used. The mass

window collected was determined based upon the peptides in the sample. Spectra were processed using MassLynx^{TM} v4.0. Background subtraction was applied before analysis. % demethylation was determined by reference to the intensity of the methylated and non-methylated peptide peaks.

8.7.2 FDH assay:

Assays were initiated by mixing of two stocks. The first contained enzyme (2 µM), Fe(II) (10 µM) and NAD^+ (500 µM), the other contained FDH (0.025 units), ascorbate (100 µM), 2OG (assay dependent concentration) and peptide substrate (assay dependent concentration). 50 mM HEPES, pH 7.5, 0.01% v/v Tween-20 was used to make up the final volume for all enzymes except for PHF8 where 100 mM HEPES, pH 7.5, 500 mM NaCl was used. Fluorescence was monitored (355 nm excitation, 460 nm emission) every 30 seconds using an EnVision™ 2104 Multilabel Reader (Perkin Elmer).

8.7.3 ALPHA screen activity assays:

Assays were initiated by mixing of two stocks. One contained enzyme and $(NH_4)_2Fe(SO_4)_2$, the other contained all other reagents. Buffer was supplemented with 0.1% w/v BSA and 0.01% v/v Tween-20 to a final volume of 10 µL in 384-well white Proxiplates (Perkin Elmer). Reactions were quenched with 5 µL 30 mM EDTA and 5 µL ALPHA screen donor and acceptor beads (Protein A donor beads, streptavidin acceptor beads, pre-incubated with the required antibody, final bead concentration 0.02 $mg.mL^{-1}$) added (Perkin Elmer). The sample was left in the dark for 1 hour before analysis using an EnVision™ 2104 Multilabel Reader (Perkin Elmer). For inhibition assays, enzyme and inhibitor were pre-incubated together in buffer for

15 min before initiation of the reaction with all other reagents. Where necessary for inhibitor solubilisation, 1% DMSO was included in the assay buffer. Data were normalised to a no enzyme control.

8.7.4 ALPHA screen binding assays:

5 µL enzyme (400 nM) was incubated with 5 µL biotinylated peptide (400 nM) for 15 min at RT in 384-well white Proxiplates (Perkin Elmer). 10 µL ALPHA screen donor and acceptor beads (anti-His donor bead, streptavidin acceptor bead, final bead concentration 0.02 mg.mL^{-1}) were added and the plate incubated in the dark for 1 hr and analysed using an EnVision™ 2104 Multilabel Reader (Perkin Elmer).

8.7.5 Fluorescence polarization assays:

50 nM FITC-conjugated peptide (H3 1-14 K4Me$_3$K9Me$_2$-FITC peptide) was incubated with a serial dilution of enzyme (50 nM to 10 µM) in buffer containing 100 mM HEPES, 500 mM NaCl, 0.1% Tween-20 for 15 min at RT in a black low-binding 96-well plate (Corning) The fluorescence was measured using a POLARstar OPTIMA (BMG LABTECH). Excitation was carried out at 485 nm and emission at 520 nm using two cross-polarised filters. Each well was read 5 times and the average of these readings used. Binding curves were fit using GraphPad Prism.

8.7.6 Isothermal Titration Calorimetry:

ITC measurements were carried out using an iTC$_{200}$ Microcalorimeter (MicroCal). Before experimental runs, protein was dialysed against buffer (100 mM HEPES, pH 7.5 500 mM NaCl), which was also used to dissolve peptide and metal solids for use as the titrant to ensure matched buffers. Experiments were performed using

50 μM FL PHF8 or PHD of PHF8 in the cell. A 40 μL syringe was filled with titrant (peptide or Mn(II)) and the reference cell filled with degassed buffer. 1.3 μL aliquots of titrant were added to the cell on each injection with stirring at 1000 rpm (the initial injection was only 0.6 μL and was excluded from analysis to compensate from 'backslash' of the syringe injector mechanism on initial use). Data was analysed using the customized ORIGIN v7.0 software provided with the calorimeter by the manufacturer. Automatic baseline correction was applied to the raw power-time curve and a further baseline correction used to correct for heats of dilution. Values for the binding constant were determined by non-linear least squares fitting to a two-state single-site binding model included in the ORIGIN software.

8.7.7 Luciferase reporter assays

The ANF-promoter luciferase reporter was a kind gift from Benoit Bruneau.[382] Luciferase activity was measured using the Promega Dual Luciferase Reporter Assay System on a Promega Glomax 20/20 Luminometer. Assays were carried out as previously described.[229,376] 8000 cells were seeded into 96-well plates and left for 24 hours. Cells were then transfected with required plasmids as described above and left for a further 24 hours. Cells were washed with PBS followed by addition of 20 μL passive lysis buffer, after which cells were left to lyse at RT for 20 min. 10 μL lysate was transferred to a 1.6 mL Eppendorf tube containing 20 μL LAR II reagent and the Firefly luciferase activity measured after 5 s using a 10 s integration time. 30 s after addition of LAR II, 20 μL Stop & Glo reagent was added and the *Renilla* luciferase activity measured 5 s after addition of the second reagent using a 10 s integration time. Readings were normalized to a Renilla Luciferase control that was co-transfected with all samples.

8.8 Western Blot:

Protein samples were separated by SDS-PAGE electrophoresis including a lane with PageRuler™ prestained protein molecular weight ladder. The gel was blotted onto Hybond ECL membrane (0.2 µm pore size, GE Healthcare) in transfer buffer (2.5 mM Tris, 192 mM glycine, 20% methanol) for 2 hours at a constant current of 200 mA on ice, using a Mini Trans-Blot Electrophore. Ponceau S dye was used to visualize successful protein transfer after blotting. The membrane was destained using MilliQ water and then blocked overnight in blocking buffer (either 5% milk or FBS in Tris Buffered Saline, 0.1% Tween 20 (TBS-T; 25 mM Tris, 137 mM NaCl, 2.68 mM KCl, 0.1 % Tween 20, pH 8.8)) at 4 °C. Membranes were probed with primary antibody at appropriate dilutions in 1 mL blocking buffer and HRP-conjugated secondary antibodies at 1:15000 in 10 mL blocking buffer, each for 1 hour at room temperature. After each antibody incubation membranes were washed with TBS-T (3 × 10 min with shaking). Bands were visualised using ECL™ detection reagents (GE Healthcare), using 0.5 mL each of reagent 1 and 2 per gel blot. Blots were incubated with the combined detection reagent for 1 min and exposed to Amersham™ Hyperfilm ECL film (GE Healthcare). Film was developed using a Compack X4 (Xograph) film developer.

8.9 Dot blots:

Dot blots were carried out in the same manner as western blots except that instead of carrying out a transfer, samples were spotted directly onto the membrane and left to dry before blocking.

8.10 qRT PCR

RNA was extracted from cells using an RNeasy® Plus Mini Kit (Quiagen) according to the manufacturer's procedure. 1 µg extracted RNA was analysed for purity by gel electrophoresis using a 1% agarose gel. cDNA was synthesised from 2 µg extracted RNA using an Affinity Script qPCR cDNA synthesis kit (Agilent) according to the manufacturer's protocol, using both oligo(dT) and random hexamers. Oligonucleotide primers were purchased from Sigma and the sequences are given below.

Gene	Primer Name	Sequence
HOXA1	hHOXA1_f	TCCAGCGCAGACTTTTGACT
	hHOXA1_r	GTAGTGAAGTTGGTGCGCAC
HOXA2	hHOXA2_f	AGAAGAAGGCGGCCAAGAAA
	hHOXA2_r	CATCGGCGATTTCCAGGGAT
HOXA3	hHOXA3_f	CTGCTCAACTCACCCACAGT
	hHOXA3_r	CTCTTGTCGCCAGCGCAG
HOXA4	hHOXA4_f	GGTGTACCCCTGGATGAAGAA
	hHOXA4_r	CTCCAGCTCCAAGACCTGC
HOXA5	hHOXA5_f	CGCAAGCTGCACATAAGTCA
	hHOXA5_r	GGAACTCCTTCTCCAGCTCC
HOXA6	hHOXA6_f	CAGGGCAAAGCACTCCATGA
	hHOXA6_r	ATGGCTCCCATACACAGCAC
HOXA7	hHOXA7_f	TGAGGCCAATTTCCGCATCT
	hHOXA7_r	GGAACTCCTTCTCCAGCTCC
HOXA9	hHOXA9_f	AGGCGCCTTCTCTGAAAACA
	hHOXA9_r	CACCGCTTTTTCCGAGTGGA
HOXA10	hHOXA10_f	CAAAGCCTCGCCGGAGAA
	hHOXA10_r	CAGTGTCTGGTGCTTCGTGT
HOXA11	hHOXA11_f	CCACACTGAGGACAAGGCC
	hHOXA11_r	CGCTGAAGAAGAACTCCCGT
HOXA13	hHOXA13_f	GGCCAAATGTACTGCCCCAA
	hHOXA13_r	ATAGGAGCTGGCATCCGAGG
Polr2a	hPolr2a_F	CCATCAAGAGAGTCCAGTTCG
	hPolr2a_R	ACCCTCCGTCACAGACATTC
GAPDH	hGAPDH_F	CCATGGGGAAGGTGAAGGTC
	hGAPHDH_r	TGAAGGGGTCATTGATGGCA
B2M	hB2M_f	GAGATGTCTCGCTCCGTGG
	hB2M_r	CCATTCTCTGCTGGATGACGT
UTY	hUTY_cat_f	AAGTTCCAGGGAGTCGGACA
	hUTY_cat_r	ACCATTCACAATCTCCTGGACC

qPCR reactions were carried out with Brilliant II SYBR Green Mastermix using a MiniOpticon Real-Time PCR System (Bio-Rad). Primer "linearity" was confirmed by carrying out a serial dilution of input cDNA and products were analysed by gel electrophoresis using a 2% agarose gel to ensure a single product was amplified. Reactions were carried out in white 96-well PCR plates (STARLAB) Each reaction contained the components given below.

Reagent	Amount
cDNA	1 µL
Forward primer (10 µM)	0.4 µL
Reverse primer (10 µM)	0.4 µL
SYBR Mastermix	10 µL
RNase-free water	8.2 µL

The thermal cycler was programmed to hold at 95 °C for 10 min.

40 cycles followed:

Denaturing at 95 °C for 3 s
Annealing at 60 °C for 30 s
Extension at 72 °C for 1 min
Fluorescence reading (excitation at 497 nm, detection at 523 nm)

At the end, samples were held at 95 °C for 1 min. Melt curves were then measure between 45 °C and 90 °C at 0.5 °C intervals. Data was exported to Excel and analysis carried out using GraphPad prism. Sigmoidal curves were fitted to the data and the Ct value determined to be the midpoint of each curve. Data were normalised between experiments to a GAPDH housekeeping gene.

8.11 Peptide synthesis:

Peptides were either purchased from GLS Biochem at >95% purity or synthesized in-house by solid phase peptide synthesis (SPPS) from their *C*-termini using a CS-Bio CS336S or CS336X automated peptide synthesiser, using PL-AMS resin and a Rink

amide linker (4-(2',4'-dimethoxyphenyl-fmoc-aminomethyl)-phenoxyacetamido-norleucyl-MBHA resin). Amino acids were purchased from Novabiochem or CS Bio and were used in large excess (10 eq.) except in the case of expensive modified amino acids (such as methylated arginyl-residues) where for cost efficiency lower excesses were used (3-5 eq.) and the coupling time doubled to six hours. N^{α}-FMOC-protected amino acids were activated using diisopropylcarbodiimide and hydroxybenzotriazole (AGTC Bioproducts) before addition to the resin. Resin and amino acids were incubated with shaking for 3 hours before the solution-phase reagents were drained and the N^{α}-FMOC removed by treating the resin with 20% piperidine (Merck Chemicals) in DMF (Rathburn Chemicals). The process was then repeated with the other amino acids. Final cleavage of the peptides from the resin was achieved by incubation for three hours with trifluoroacetic acid (TFA):triethylsilane (97.5:2.5), which also removed the side-chain protecting groups. The peptide was precipitated with diethyl ether from the cleavage mixture and the resultant solid dissolved in MilliQ water and lyophilized overnight. Peptides were purified by HPLC and the mass and purity confirmed by MALDI-TOF MS.

Lyophilisation:

Samples were frozen at -80 °C before being lyophilised at RT and 4 mBar using a Micro Modulyo freeze drier. Samples remained on the machine until only a powder remained.

8.12 Mammalian cell culture

8.12.1 Reagents

All procedures were carried out in a Class II Biological Safety Cabiner (Nuaire Nu 437-400E). *H. Sapiens* HEK 293T, HeLa and U2OS cells were grown in Dulbecco's Modified Eagle's Medium (DMEM, Lonza) supplemented with glutamax and 10% fetal bovine serum (FBS) in a Binder BD 53 incubator with a humidified 5% CO_2 atmosphere at 37 °C unless otherwise described.

Vessel	Surface area /cm^2	No. cells at 100% confluence	Media volume /mL
96 well plate	0.3	2.8×10^4	0.1
25 cm^2 flask (T25)	25	2.5×10^6	5
10 cm diameter Petri dish	55	5.5×10^6	10
15 cm diameter Petri dish	148	1.48×10^7	20

8.12.2 Methods

Cell passaging:

Strongly adherent HeLa cells and U2OS cells were washed with PBS, then trypsinised from the plate using 0.2 mg.mL^{-1} Trypsin (Lonza) at 37 °C until cells were released from the surface of the plate. The reaction was quenched by addition of fresh media. Cells were fully resuspended by pipetting and the appropriate volume added to new plates containing pre-warmed media.

Semi-adherent 293T were washed with PBS and then removed from the plate surface by pipetting in PBS to form a suspension. An appropriate volume of cells was added to new plates containing pre-warmed media.

For labelling cells were grown in DMEM (SILAC DMEM, Sigma) supplemented with the relevant amino acids labelled with ^{13}C and ^{2}H. Cells were passaged 5 times and stocks of fully-labelled cells frozen for future use.

Cell Storage

Cells were frozen in DMEM containing 20% FBS and 10% DMSO and stored in a -80 °C freezer for short-term storage (less than 1 year) and in a liquid nitrogen dewar for longer term storage.

Transfection

Cells were transiently transfected with plasmid DNA using linear PEI (1 $mg.mL^{-1}$ poly(ethylenimine), pH 7.2) or Lipofectamine® 2000. Transfection was carried out with cells at ~50% confluency as judged using a Motic AE20 (Ted Pella) microscope. In both cases 2 hours prior to transfection media was exchanged for fresh media. DNA for transfection and transfection reagent were separately diluted in OptiMEM® and incubated at room temperature for 5 min. The two reagents were then mixed and left at room temperature for 10 min before adding dropwise to the cells. 0.1 µg plasmid DNA was used for transfections in 96 well plates and 20 µg DNA for transfections in 15 cm dishes. Linear PEI was used at a ratio of 12:1 w/w and for transfections in 96 well plates 0.2 µL Lipofectamine® 2000 was used.

8.12.3 Immunofluoresence

8000 cells per well were plated into 96 well plates one day prior to transfection. Transfection was carried out as described above. After 24 hours cells were washed with PBS and then fixed with 4% formaldehyde in PBS for 15 min at room

temperature. Formaldehyde was produced by heating paraformaldehyde at 60 °C in water until it dissolved. After 15 min the formaldehyde was removed and the cells washed twice with PBS. Cells were permeabilised by incubation for 6 - 8 min at room temperature with 0.2% Triton-X100. The Triton-X100 was removed and the cells washed twice with PBS before being blocked for 30 min at room temperature with 5% FBS in PBS. Primary antibodies were made up to the appropriate dilution in blocking buffer and added to the cells before incubation for 1 hr at room temperature or overnight at 4 °C. After incubation cells were washed with PBS (3 × 10 min). Secondary antibodies were added, again at the appropriate dilution in blocking buffer, for 1 hr at room temperature in the dark. Cells were washed with PBS (3 × 10 min) and stained with DAPI (0.2 $\mu g.mL^{-1}$) before visualization. Cells were visualized using a Zeiss Axioobserver epifluorescence microscope with a 20 × objective.

Fluorophore	**Excitation wavelength /nm**	**λ_{max} absorption /nm**	**λ_{max} emission /nm**
EGFP	525	507	488
Cy3	543	550	570
DAPI	445	358	461
FITC	525	490	525
AlexaFluor 488	525	496	519
AlexaFluor 594	543	590	617

Cell lysis:

Cells were lysed for 1 hr on ice in buffer containing 10 mM HEPES, pH 7.5, 300 mM NaCl, 0.2% NP40, 1 $\mu g.mL^{-1}$ DNase1, protease inhibitors (Roche) and phosphatase inhibitors (Sigma). After lysis cell debris was pelleted by centrifugation at 14000 rpm for 20 min at 4 °C.

8.12.4 Immunoprecipitation:

Cell lysate was added to beads (pre-equilibrated in lysis buffer without NP40) and incubated at 4 °C for 2 hours. Beads (GFP-Trap or Anti-Flag M2 affinity gel) were washed 5 times with lysis buffer. Between washes beads were isolated by centrifugation at 1.5 rcf for 4 min at 4 °C. Proteins were eluted from the beads by addition of 0.2 M glycine, pH 2 or addition of SDS-PAGE loading buffer and incubation for 10 min at 90 °C. Proteins bound to the beads were detected by SDS-PAGE separation followed by western blotting or digestion followed by mass spectrometry.

8.13 Mass spectrometry

8.13.1 Methanol/Chloroform extraction:

Proteins were diluted to 200 µL in MilliQ water and 600 µL methanol and 150 µL chloroform added. The sample was mixed by vortexing and a further 450 µL MilliQ water added. Following mixing the sample was centrifuged at 14000 rpm for 1 min at RT. The upper aqueous phase was removed without disturbing the precipitate at the phase interface and 450 µL MilliQ added. Following mixing by vortex the sample was centrifuged for 2 min at 14000 rpm and the supernatant removed leaving the protein precipitate for use in further procedures.

8.13.2 In-solution digest:

All solutions were made up in 50 mM ammonium bicarbonate. 9 mM DTT was added to samples for digestion (direct from an immunoprecipitation or following methanol/chloroform extraction). These were incubated at 56 °C for 15 min to reduce

all cysteinyl-residues. 17 mM iodoacetamide was then added and the samples left to alkylate for 15 min in the dark. A further 9 mM DTT was added to quench the remaining iodoacetamide and the samples left at RT for 15 min. To this the required digestion enzyme was added and the volume made up to 100 µL with MilliQ water. Samples were incubated at 37 °C overnight for complete digestion, or for shorter times for partial digests, and quenched with 10 % formic acid (FA). Samples were purified using a SepPak column according to the manufacturers procedure before mass spectrometric analysis.

8.13.3 LCMS/MS:

Samples were analyzed using an Orbitrap Elite (Thermo Fisher Scientific, DE) connected to a UHPLC Proxeon EASY-nLC 1000 and an EASY-Spray nano-electrospray ion source. Peptides were trapped on an Acclaim PepMap® trapping column (100 µm internal diameter x 20 mm, 5 µm C18) and separated on an EASY-spray Acclaim PepMap® analytical column (75 µm internal diameter x 500 mm, RSLC C18, 2 µm, 100 Å). Solvent A consisted of 0.1% formic acid and solvent B of 0.1% formic acid in acetonitrile. Peptides were separated using a 3 hour gradient of 7% to 30% solvent B at a flow rate of 200 $nL.min^{-1}$. Full scan MS spectra were acquired in the Orbitrap (350-1500 m/z, resolution 120000, AGC target 1e6, maximum injection time 250 ms). CID and ETD spectra were acquired in the Ion Trap (resolution 7500, AGC cation target 3e4, AGC Anion target 2e5, maximum injection time 100 ms). The 20 most intense peaks in the full MS scan were selected for fragmentation using a data-dependent decision tree (DDDT). ETD fragmentation was used for charge states 3, 4 and 5 with m/z less than 750 and for all charge states

greater than 5. CID fragmentation was used for all other peptides.

The raw data files generated were processed using MaxQuant software (Version 1.4.1.2), integrated with the Andromeda search engine as described elsewhere.[369,389] The maximum number of missed cleavages was set to be 4. Methylation of arginine was allowed as a variable modification, and "heavy" lysine and arginine as a modification in the SILAC samples. All spectra were manually validated.

8.14 Histone extraction

To cell pellets from cell lysis was added 0.4 M HCl. The resuspended pellet was incubated at 4 °C for 1 hour. Samples were centrifuged for 20 min at 14000 rpm at 4 °C. The supernatant was added to ice-cold acetone and left at -20 °C for 16 hours. Histones were pelleted by centrifugation for 20 min at 14000 rpm at 4 °C and the pellets washed three times with ice-cold acetone. Finally pellets were air dried and resuspended in MilliQ water for analysis.

8.15 NMR Spectroscopy

NMR involving JmjC proteins was performed on a Bruker Avance AV111 700 MHz spectrometer equipped with an inverse TCI CryoProbe, optimised for ^{1}H observation, and installed with Topspin 2 software. Samples were made up in deuterated ammonium formate buffer. Experiments were carried out in 2 mm MATCH NMR tubes (Wilgenberg) and the solvent deuterium signal was used as an internal lock signal.

CHD4 NMR experiments were performed on several spectrometers: a Bruker Avance 500 MHz spectrometer with a Cryoplatform equipped with a TCI CryoProbe and three home-built spectrometers (600 MHz, 750 MHz or 950 MHz) with GE/Omega consoles and triple-resonance probes. All experiments were carried out at 25 °C in BMS-3 Shigemi tubes (Shigemi Inc.) using a minimum sample volume of 270 µL. Samples were prepared in 20 mM Tris pH 7.5, 200 mM NaCl, 10 mM DTT prepared in 5% D_2O / 95% H_2O (v/v), except where otherwise described.

Data acquisition and processing:

Data acquired from the Bruker spectrometer were processed directly using the NMRPipe package.[390] Data acquired on the spectrometers with GE/Omega consoles was first converted into binary format and then processed using NMRPipe. Data was then analysed using Analysis2.1 (CCPN).[391]

NMR parameter for each experiment are described below. Sweep widths were scaled according to the ^1H operating frequencies of the different spectrometers. When processing, Gaussian and sine-bell window functions were applied to the direct and indirect dimensions respectively. Zero-filling of the indirect dimensions of resulting spectra was used.

2D ^1H experiments were carried out after sample exchange into buffer made with 100% D_2O for assignment of aromatic side-chains. 2D heteronuclear experiments were performed using single- or double-labelled protein. 3D double resonance experiments were carried out on single- or double-labelled protein samples and 3D triple resonance experiments were carried out on double-labelled protein samples. Parameters for each experiment are given below.

Experimental parameters for 2D homonuclear ^{1}H experiments

Experiment	Number of scans	Direct dimension			Indirect dimension			Mixing time /ms
		^{1}H freq. /MHz	^{1}H sweep width /Hz	^{1}H points	^{1}H freq. /MHz	^{1}H sweep width /Hz	^{1}H points	
NOESY	96	749.99403	9009.01	1024	749.99403	9009.01	1024	150
TOCSY	64	749.99403	9009.01	1024	749.99403	9009.01	1024	56
DQF-COSY	48	749.99403	9009.01	1024	749.99403	9009.01	1024	-

Experimental parameters for 2D heteronuclear experiments

Experiment	Number of scans	^{1}H freq. /MHz	^{1}H sweep width /Hz	^{1}H points	^{15}N freq. /MHz	^{15}N sweep width /Hz	^{15}N points	^{13}C freq. /MHz	^{13}C sweep width /Hz	^{13}C points
^{15}N HSQC	16	749.99403	9345.79	1024	76.004926	3861	128	-	-	-
^{15}N HMQC-J	128	598.5595403	7518.8	1024	60.658638	3086.42	1024	-	-	-
^{15}N-{^{1}H} Heteronuclear NOE	96	749.99403	9345.79	1024	76.004926	3846.15	128	-	-	-
IPAP HSQC	160	598.5595403	7518.8	1024	60.658638	3086.42	251	-	-	-
^{13}C HSQC	8	749.99403	9345.79	1024	-	-	-	188.592696	23255.81	128

Experimental parameters for 3D experiments

Experiment	Number of scans	^{1}H freq. /MHz	^{1}H sweep width /Hz	^{1}H points	^{15}N freq. /MHz	^{15}N sweep width /Hz	^{15}N points	^{1}H freq. /MHz	^{1}H sweep width /Hz	^{1}H points	Mixing time /ms
^{15}N-edited NOESY-HSQC	8	749.99403	9345.79	512	76.004926	3861	64	749.99403	8264.46	256	75
HSQC-NOESY-HSQC	16	749.99403	9345.79	512	76.004926	3846.15	64	749.99403	9345.79	128	150

Experiment	Number of scans	^{1}H freq. /MHz	^{1}H sweep width /Hz	^{1}H points	^{13}C freq. /MHz	^{13}C sweep width /Hz	^{13}C points	^{1}H freq. /MHz	^{1}H sweep width /Hz	^{1}H points	Mixing time /ms
^{13}C-edited NOESY-HSQC	8	749.99403	9345.79	512	188.592696	23255.81	128	749.99403	16949.15	256	75
HCCH-TOCSY	8	598.5595403	7518.8	512	150.5249623	9090.91	128	598.5595403	7518.8	256	-

Experiment	Number of scans	^{1}H freq. /MHz	^{1}H sweep width /Hz	^{1}H points	^{15}N freq. /MHz	^{15}N sweep width /Hz	^{15}N points	^{13}C freq. /MHz	^{13}C sweep width /Hz	^{13}C points
HNCA	16	500.012	6250	1024	50.672	1420.455	56	125.727	3521.127	100
HN(CO)CA	24	500.012	6250	1024	50.672	1420.455	56	125.727	3521.127	100
HNCO	20	500.012	6250	1024	50.672	1420.455	56	125.727	3521.127	100
HN(CA)CO	40	500.012	6250	1024	50.672	1420.455	56	125.727	1886.24	100
HBHACONH	16	500.012	6250	1024	50.672	1420.455	56	125.727	4761.905	178
CBCANH	36	500.012	6250	1024	50.672	1420.455	56	125.727	7462.687	100
CBCA(CO)NH	20	500.012	6250	1024	50.672	1420.455	56	125.727	7462.687	100

Table 8.1 NMR acquisition parameters

Measuring of J-couplings from HMQC-J spectra

The ^{15}N dimension was zero-filled to 2048 complex points and analysed in Felix 2.3 (Accelrys). Each doublet peak was fit to two curves using the Oxford HMQC-J fitting tool. J-coupling values were defined as the difference between the chemical shift values of the two peaks. Peaks were only analysed where the splitting was resolvable. For values < 6 Hz Φ angle restraints were defined as either -60 or +60 ± 30° and for > 8 Hz as -120 ± 40°.

Heteronuclear NOE experiment

IPAP-HSQC spectra were collected with and without ^{1}H pre-saturation (4 s) in an interleaved manner and then separated into two data sets. The peak-picking, calculation of the NOE ratio and noise estimation were performed using software written by Professor Christina Redfield. The intensity of a block of points in the blank spectrum without any peaks was used to estimate the noise, which was then used to calculate the error on the NOE ratio for each peak. The delay period between the 90° ^{15}N pulse and the 180° ^{1}H pulse in the INEPT portion of the pulse sequence was 2.75 ms.

Generation of distance restraints

Distance restraints were generated from peak intensities in 2D and 3D NOESY spectra using CCPN Analysis. For each distance bin a peak height threshold was defined for NOEs from each spectrum. Restraints were divided into two files accounting for unambiguous and ambiguous restraints and were exported in CNS format using the Format Converter.

Residual dipolar coupling (RDC) measurements

Samples were aligned by addition of 10 mg.mL^{-1} Pf1 filamentous phage. IPAP-HSQC spectra were first acquired without phage addition to obtain the isotropic peak splitting values and then again after addition of the alignment media. RDC values were calculated as the difference between the two splitting values. The experimental RDC values were compared to RDC values predicted for a published structure of the 2nd PHD domain of CHD4 using a programme written by Professor Christina Redfield in order to determine D_a and R.

8.15.2 NMR structure calculation

Structure calculations were carried out using Xplor 3.8 for prochiral swapping calculations and then Xplor-NIH 2.29 for later stages of the structure calculations.[262,263] The initial structures were generated from a randomly generated extended template by simulated annealing. After prochiral-swapping, all restraints with arbitrary stereospecific assignments were modified to remove these and the restraints for these residues were merged to include only the shorter distance restraint. Zinc-ligation was maintained by bonds and angles between the zinc ligands and the zinc ions. Structures were generated using a two-step procedure, initially by *ab initio* simulated annealing using NOE restraints and a RAMA force-field, followed by division of the structure into the two PHD domains and a round of refinement with an increased van der Waals radius. A second refinement of the lowest energy structure for each domain was carried out using RDC and dihedral angle restraints. The 20 lowest energy structures from this ensemble were grouped as an ensemble and validated using MolProbity and PROCHECK.

8.15.3 Average structure determination and ensemble RMSD calculations

An energy-minimised average structure representing each ensemble of the 20 lowest-energy structures was produced using a script in Xplor-NIH 2.29. This script also determined the RMSD of each atom of the structures in the ensembles from the average structure.

8.15.4 Peptide titrations

NMR peptide titration experiments were carried out using 210 μM ^{15}N-labelled dPHD or 300 μM ^{15}N-labelled dPHDdCHD. Increasing concentrations of unlabelled peptides (10 mM stocks in MilliQ water) were added to the sample. A ^{1}H-^{15}N-HSQC was collected after each peptide addition and the change in chemical shift analysed. K_D calculations accounted for the dilution in protein concentration on addition of peptide. K_D values were fit using five well-resolved peaks from each domain using a 12-parameter fitting programme written by Professor Christina Redfield.

Chapter 9
References

1. Crick, F. Ideas on Protein Synthesis. in *Symp. Soc. Exp. Biol. XII* 139–163 (1956). at <http://profiles.nlm.nih.gov/ps/retrieve/ResourceMetadata/SCBBFT>

2. Crick, F. Central Dogma of Molecular Biology. *Nature* **227,** 561–563 (1970).

3. Leavitt, S. A. & Marshall, N. Deciphering the Genetic Code: Marshall Nirenberg. *Off. NIH Hist.* (2010).

4. Waddington, C. H. The epigenotype. *Endeavour* **1,** 18–20 (1942).

5. Waddington, C. H. *The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology*. (Allen & Unwin, 1957).

6. Hemberger, M., Dean, W. & Reik, W. Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nat. Rev. Mol. Cell Biol.* **10,** 526–37 (2009).

7. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: a landscape takes shape. *Cell* **128,** 635–8 (2007).

8. Seisenberger, S., Peat, J. R., Hore, T. A., Santos, F., Dean, W. & Reik, W. Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20110330 (2013).

9. Gill, M. E., Erkek, S. & Peters, A. H. F. M. Parental epigenetic control of embryogenesis: a balance between inheritance and reprogramming? *Curr. Opin. Cell Biol.* **24,** 387–96 (2012).

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

11. Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184,** 868–71 (1974).

12. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389,** 251–60 (1997).

13. Iwasaki, W., Miya, Y., Horikoshi, N., Osakabe, A., Taguchi, H., Tachiwana, H., Shibata, T., Kagawa, W. & Kurumizaka, H. Contribution of histone N-terminal tails to the structure and stability of nucleosomes. *FEBS Open Bio* **3,** 363–9 (2013).

14. Greulich, K. O., Wachtel, E., Ausio, J., Seger, D. & Eisenberg, H. Transition of chromatin from the "10 nm" lower order structure, to the "30 nm" higher order structure as followed by small angle X-ray scattering. *J. Mol. Biol.* **193,** 709–721 (1987).

15. Li, G. & Reinberg, D. Chromatin higher-order structures and gene regulation. *Curr. Opin. Genet. Dev.* **21,** 175–86 (2011).

16. Robinson, P. J. J., Fairall, L., Huynh, V. A. T. & Rhodes, D. EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 6506–11 (2006).

17. Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N. P. & Bickmore, W. A. Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* **118,** 555–66 (2004).

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

19. Gao, M., Nadaud, P. S., Bernier, M. W., North, J. A., Hammel, P. C., Poirier, M. G. & Jaroniec, C. P. Histone H3 and H4 N-terminal tails in nucleosome arrays at cellular concentrations probed by magic angle spinning NMR spectroscopy. *J. Am. Chem. Soc.* **135,** 15278–81 (2013).

20. Saikusa, K., Fuchigami, S., Takahashi, K., Asano, Y., Nagadoi, A., Tachiwana, H., Kurumizaka, H., Ikeguchi, M., Nishimura, Y. & Akashi, S. Gas-phase structure of the histone multimers characterized by ion mobility mass spectrometry and molecular dynamics simulation. *Anal. Chem.* **85,** 4165–71 (2013).

21. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292,** 195–202 (1999).

22. Du Preez, L. L. & Patterton, H.-G. Secondary structures of the core histone N-terminal tails: their role in regulating chromatin structure. *Subcell. Biochem.* **61,** 37–55 (2013).

23. Vellore, N. A. & Baron, R. Epigenetic Molecular Recognition: A Biomolecular Modeling Perspective. *ChemMedChem* **9,** 484–94 (2014).

24. Liu, H. & Duan, Y. Effects of posttranslational modifications on the structure and dynamics of histone H3 N-terminal Peptide. *Biophys. J.* **94,** 4579–85 (2008).

25. Potoyan, D. A. & Papoian, G. A. Regulation of the H4 tail binding and folding landscapes via Lys-16 acetylation. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 17857–62 (2012).

26. Nelson, C. J., Santos-Rosa, H. & Kouzarides, T. Proline isomerization of histone H3 regulates lysine methylation and gene expression. *Cell* **126,** 905–16 (2006).

27. Marzluff, W. F., Wagner, E. J. & Duronio, R. J. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat. Rev. Genet.* **9,** 843–854 (2008).

28. Arents, G. & Moudrianakis, E. N. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 11170–11174 (1995).

29. Henikoff, S. & Ahmad, K. Assembly of variant histones into chromatin. *Annu. Rev. Cell Dev. Biol.* **21,** 133–53 (2005).

30. Talbert, P. B. & Henikoff, S. Histone variants--ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* **11,** 264–275 (2010).

31. Earnshaw, W. C. & Rothfield, N. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* **91,** 313–21 (1985).

32. Ahmad, K. & Henikoff, S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* **9,** 1191–1200 (2002).

33. Witt, O., Albig, W. & Doenecke, D. Testis-specific expression of a novel human H3 histone gene. *Exp. Cell Res.* **229,** 301–306 (1996).

34. Skene, P. J. & Henikoff, S. Histone variants in pluripotency and disease. *Development* **140,** 2513–2524 (2013).

35. Robertson, K. D. DNA methylation: past, present and future directions. *Carcinogenesis* **21,** 461–467 (2000).

36. Wyatt, M. D. & Pittman, D. L. Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chem. Res. Toxicol.* **19,** 1580–94 (2006).

37. Hang, B. Formation and repair of tobacco carcinogen-derived bulky DNA adducts. *J. Nucleic Acids* **2010,** 709521 (2010).

38. Cheng, X. DNA modification by methyltransferases. *Curr. Opin. Struct. Biol.* **5,** 4–10 (1995).

39. Jin, B. & Robertson, K. D. DNA methyltransferases, DNA damage repair, and cancer. *Adv. Exp. Med. Biol.* **754,** 3–29 (2013).

40. Fu, Y. & He, C. Nucleic acid modifications with epigenetic significance. *Curr. Opin. Chem. Biol.* **16,** 516–24 (2012).

41. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99,** 247–57 (1999).

42. Jeltsch, A. On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics* **1,** 63–6

43. Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L. & Rao, A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–5 (2009).

44. Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C. A., Hagemeier, C. & Carell, T. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50,** 7008–12 (2011).

45. Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C. & Zhang,

Y. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science (80-.).* DOI: 10.1126/science.1210597 (2011).

46. He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C. & Xu, G.-L. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333,** 1303–7 (2011).

47. Bhutani, N., Brady, J. J., Damian, M., Sacco, A., Corbel, S. Y. & Blau, H. M. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463,** 1042–7 (2010).

48. Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.* **279,** 52353–60 (2004).

49. Zhu, J.-K. Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* **43,** 143–66 (2009).

50. Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E. & Reik, W. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463,** 1101–5 (2010).

51. Guo, J. U., Su, Y., Zhong, C., Ming, G. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145,** 423–34 (2011).

52. Nabel, C. S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H. L., Stivers, J. T., Zhang, Y. & Kohli, R. M. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8,** 751–8 (2012).

53. Schiesser, S., Hackner, B., Pfaffeneder, T., Müller, M., Hagemeier, C., Truss, M. & Carell, T. Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angew. Chem. Int. Ed. Engl.* **51,** 6516–20 (2012).

54. Inoue, A. & Zhang, Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334,** 194 (2011).

55. Inoue, A., Shen, L., Dai, Q., He, C. & Zhang, Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res.* **21,** 1670–6 (2011).

56. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14,** 204–20 (2013).

57. Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A. & Meissner, A. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484,** 339–44 (2012).

58. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11,** 204–20 (2010).

59. Straussman, R., Nejman, D., Roberts, D., Steinfeld, I., Blum, B., Benvenisty, N., Simon, I., Yakhini, Z. & Cedar, H. Developmental programming of CpG island methylation profiles in the human genome. *Nat. Struct. Mol. Biol.* **16,** 564–71 (2009).

60. Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A. & Cedar, H. Sp1 elements protect a CpG island from de novo methylation. *Nature* **371,** 435–8 (1994).

61. Bergman, Y. & Cedar, H. DNA methylation dynamics in health and disease. *Nat. Struct. Mol. Biol.* **20,** 274–81 (2013).

62. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10,** 295–304 (2009).

63. Robertson, K. D. DNA methylation and human disease. *Nat. Rev. Genet.* **6,** 597–610 (2005).

64. Pogribny, I. P. & Beland, F. A. DNA hypomethylation in the origin and pathogenesis of human diseases. *Cell. Mol. Life Sci.* **66,** 2249–61 (2009).

65. Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R. A., Niveleau, A., Cedar, H. & Simon, I. Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat. Genet.* **38,** 149–53 (2006).

66. Haselkorn, R. & Rothman-Denes, L. B. Protein synthesis. *Annu. Rev. Biochem.* **42,** 397–438 (1973).

67. Bailey, C. A. Processing at the Carboxyl Terminus of Nascent Placental Alkaline Phosphatase in a Cell-Free System: Evidence for Specific Cleavage of a Signal Peptide. *Proc. Natl. Acad. Sci.* **86,** 22–26 (1989).

68. Mavridou, D. A. I., Stevens, J. M., Mönkemeyer, L., Daltrop, O., di Gleria, K., Kessler, B. M., Ferguson, S. J. & Allen, J. W. A. A pivotal heme-transfer reaction intermediate in cytochrome c biogenesis. *J. Biol. Chem.* **287,** 2342–52 (2012).

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

70. Walsh, C. T. *Posttranslational modification of proteins : Expanding nature's inventory*. (Roberts and Co. Publishers, 2006).

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

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

73. Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., Lu, Z., Ye, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B. & Zhao, Y. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146,** 1016–28 (2011).

74. Tweedie-Cullen, R. Y., Brunner, A. M., Grossmann, J., Mohanna, S., Sichau, D., Nanni, P., Panse, C. & Mansuy, I. M. Identification of combinatorial patterns of post-

translational modifications on individual histones in the mouse brain. *PLoS One* **7,** e36980 (2012).

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

76. Kirmizis, A., Santos-Rosa, H., Penkett, C. J., Singer, M. A., Vermeulen, M., Mann, M., Bahler, J., Green, R. D. & Kouzarides, T. Arginine methylation at H3R2 controls deposition of H3K4 trimethylation. *Nat. Lett.* **449,** 928–934 (2007).

77. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273,** 5858–68 (1998).

78. Cheung, P., Tanner, K. G., Cheung, W. L., Sassone-Corsi, P., Denu, J. M. & Allis, C. D. Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Mol. Cell* **5,** 905–15 (2000).

79. Lau, P. N. I. & Cheung, P. Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 2801–6 (2011).

80. Wei, Y., Mizzen, C. A., Cook, R. G., Gorovsky, M. A. & Allis, C. D. Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 7480–4 (1998).

81. Wei, Y., Yu, L., Bowen, J., Gorovsky, M. A. & Allis, C. D. Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. *Cell* **97,** 99–109 (1999).

82. Rossetto, D., Avvakumov, N. & Cote, J. Histone phosphorylation A chromatin modification involved in diverse nuclear events. *Epigenetics* **7,** 1098–1108 (2012).

83. Katan-Khaykovich, Y. & Struhl, K. Dynamics of global histone acetylation and deacetylation in vivo: rapid restoration of normal histone acetylation status upon removal of activators and repressors. *Genes Dev.* **16,** 743–52 (2002).

84. Hebbes, T. R., Thorne, A. W. & Crane-Robinson, C. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.* **7,** 1395–402 (1988).

85. Struhl, K. Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* **12,** 599–606 (1998).

86. Braun, S. & Madhani, H. D. Shaping the landscape: mechanistic consequences of ubiquitin modification of chromatin. *EMBO Rep.* **13,** 619–30 (2012).

87. Hochstrasser, M. Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30,** 405–39 (1996).

88. Wyce, A., Xiao, T., Whelan, K. A., Kosman, C., Walter, W., Eick, D., Hughes, T. R., Krogan, N. J., Strahl, B. D. & Berger, S. L. H2B ubiquitylation acts as a barrier to Ctk1 nucleosomal recruitment prior to removal by Ubp8 within a SAGA-related

complex. *Mol. Cell* **27,** 275–88 (2007).

89. Fradet-Turcotte, A., Canny, M. D., Escribano-Díaz, C., Orthwein, A., Leung, C. C. Y., Huang, H., Landry, M.-C., Kitevski-LeBlanc, J., Noordermeer, S. M., Sicheri, F. & Durocher, D. 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature* **499,** 50–4 (2013).

90. Kessler, B. M. & Edelmann, M. J. PTMs in conversation: activity and function of deubiquitinating enzymes regulated via post-translational modifications. *Cell Biochem. Biophys.* **60,** 21–38 (2011).

91. Lachner, M., O'Sullivan, R. J. & Jenuwein, T. An epigenetic road map for histone lysine methylation. *J. Cell Sci.* **116,** 2117–24 (2003).

92. Le, D. D. & Fujimori, D. G. Protein and nucleic acid methylating enzymes: mechanisms and regulation. *Curr. Opin. Chem. Biol.* **16,** 507–15 (2012).

93. Van Leeuwen, F., Gafken, P. R. & Gottschling, D. E. Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* **109,** 745–56 (2002).

94. 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,** 711–23 (2003).

95. Patel, A., Vought, V. E., Dharmarajan, V. & Cosgrove, M. S. A novel non-SET domain multi-subunit methyltransferase required for sequential nucleosomal histone H3 methylation by the mixed lineage leukemia protein-1 (MLL1) core complex. *J. Biol. Chem.* **286,** 3359–69 (2011).

96. Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D. & Jenuwein, T. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406,** 593–9 (2000).

97. Black, J. C., Van Rechem, C. & Whetstine, J. R. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol. Cell* **48,** 491–507 (2012).

98. Schneider, J., Wood, A., Lee, J.-S., Schuster, R., Dueker, J., Maguire, C., Swanson, S. K., Florens, L., Washburn, M. P. & Shilatifard, A. Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. *Mol. Cell* **19,** 849–56 (2005).

99. Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S. & Zhang, Y. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298,** 1039–43 (2002).

100. Hansen, K. H., Bracken, A. P., Pasini, D., Dietrich, N., Gehani, S. S., Monrad, A., Rappsilber, J., Lerdrup, M. & Helin, K. A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* **10,** 1291–300 (2008).

101. Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C. & Kouzarides, T. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat. Cell Biol.* **6,** 73–7 (2004).

102. Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C. T., Schreiber, S. L., Mellor, J. & Kouzarides, T. Active genes are tri-methylated at K4 of histone H3. *Nature* **419,** 407–11 (2002).

103. Trojer, P. & Reinberg, D. Facultative heterochromatin: is there a distinctive molecular signature? *Mol. Cell* **28,** 1–13 (2007).

104. Heard, E. Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. *Curr. Opin. Genet. Dev.* **15,** 482–9 (2005).

105. Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. & Zhao, K. High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–37 (2007).

106. Mellor, J. Linking the cell cycle to histone modifications: Dot1, G1/S, and cycling K79me2. *Mol. Cell* **35,** 729–30 (2009).

107. Nguyen, A. T. & Zhang, Y. The diverse functions of Dot1 and H3K79 methylation. *Genes Dev.* **25,** 1345–58 (2011).

108. Li, X., Hu, X., Patel, B., Zhou, Z., Liang, S., Ybarra, R., Qiu, Y., Felsenfeld, G., Bungert, J. & Huang, S. H4R3 methylation facilitates b-globin transcription by regulating histone acetyltransferase binding and H3 acetylation. *Blood* **115,** 2028–2037 (2010).

109. Zhao, Q., G., R., Tan, Y. T., Li, H., Moritz, R. L., Simpson, R. J., Cerruti, L., Curtis, D. J., Patel, D. J., Allis, C. D., Cunningham, J. M. & Jane, S. M. . PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat. Struct. Mol. Biol.* **16,** 304–311 (2009).

110. Migliori, V., Müller, J., Phalke, S., Low, D., Bezzi, M., Mok, W. C., Sahu, S. K., Gunaratne, J., Capasso, P., Bassi, C., Cecatiello, V., De Marco, A., Blackstock, W., Kuznetsov, V., Amati, B., Mapelli, M. & Guccione, E. Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat. Struct. Mol. Biol.* **19,** 136–44 (2012).

111. Yuan, C.-C., Matthews, A. G. W., Jin, Y., Chen, C. F., Chapman, B. A., Ohsumi, T. K., Glass, K. C., Kutateladze, T. G., Borowsky, M. L., Struhl, K. & Oettinger, M. A. Histone H3R2 symmetric dimethylation and histone H3K4 trimethylation are tightly correlated in eukaryotic genomes. *Cell Rep.* **1,** 83–90 (2012).

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

113. Scrutton, N. S. Chemical aspects of amine oxidation by flavoprotein enzymes. *Nat. Prod. Rep.* **21,** 722–30 (2004).

114. Forneris, F., Binda, C., Dall'Aglio, A., Fraaije, M. W., Battaglioli, E. & Mattevi, A. A highly specific mechanism of histone H3-K4 recognition by histone demethylase LSD1. *J. Biol. Chem.* **281,** 35289–95 (2006).

115. Huang, J., Sngupta, R., Exspejo, A. B., Lee, M. G., Dorsey, J. A., Richter, M., Oprvil,

S., Shiekhattar, R., Bedford, M. T., Jenuwein, T. & Berger, S. L. p53 is regulated by the lysine demethylase LSD1. *Nature* **449,** 125–129 (2007).

116. Wang, J., Hevi, S., Kurash, J. K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G., Gaudet, F., Li, E. & Chen, T. The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat. Genet.* **41,** 125–129 (2009).

117. Karytinos, A., Formeris, F., Profumo, A., Ciossani, G., Battaglioli, E., Binda, C. & Mattevi, A. A Novel Mammalian Flavin-dependent Histone Demethylase. *J. Biol. Chem.* **284,** 17775–17782 (2009).

118. Yi, C., Jia, G., Hou, G., Dai, Q., Zhang, W., Zheng, G., Jian, X., Yang, C. G., Cui, Q. & He, C. Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature* **468,** 330–333 (2010).

119. Smith-Mungo, L. I. & Kagan, H. M. Lysyl oxidase: properties, regulation and multiple functions in biology. *Matrix Biol.* **16,** 387–98 (1998).

120. Herranz, N., Dave, N., Millanes-Romero, A., Morey, L., Díaz, V. M., Lórenz-Fonfría, V., Gutierrez-Gallego, R., Jerónimo, C., Di Croce, L., García de Herreros, A. & Peiró, S. Lysyl oxidase-like 2 deaminates lysine 4 in histone H3. *Mol. Cell* **46,** 369–76 (2012).

121. Chang, B., Chen, Y., Zhao, Y.-F. & Bruick, R. K. JMJD6 is a histone arginine demethylase. *Science (80-.).* **318,** 444–447 (2007).

122. Webby, C. J., Wolf, A., Gromak, N., Dreger, M., Kramer, H., Kessler, B. M., Nielsen, M. L., Schmitz, C., Butler, D. S., Yates 3rd., J. R., Delahunty, C. M., Lengeling, A., Mann, M., Proudfoot, N. J., Schofield, C. J. & Bottger, A. Jmjd6 catalyses lysyl hydroxylation of U2AF65, a protein associated with RNA splicing. *Science (80-.).* **325,** 90–93 (2009).

123. Unoki, M., Masuda, A., Dohmae, N., Arita, K., Yoshimatsu, M., Iwai, Y., Fukui, Y., Ueda, K., Hamamoto, R., Shirakawa, M., Sasaki, H. & Nakamura, Y. Lysyl 5-hydroxylation, a novel histone modification, by Jumonji domain containing 6 (JMJD6). *J. Biol. Chem.* **288,** 6053–62 (2013).

124. Hong, X., Zang, J., White, J., Wang, C., Pan, C.-H., Zhao, R., Murphy, R. C., Dai, S., Henson, P., Kappler, J. W., Hagman, J. & Zhang, G. Interaction of JMJD6 with single-stranded RNA. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 14568–72 (2010).

125. Cuthbert, G. L., Daujat, S., Snowden, A. W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P. D., Tempst, P., Bannister, A. J. & Kouzarides, T. Histone Deimination Antagonizes Arginine Methylation. *Cell* **118,** 545–553 (2004).

126. Wang, Y., Wysocka, J., Savegh, J., Lee, Y. H., Perlin, J. R., Leonelli, L., Sonbuchner, L. S., McDonald, C. H., Cook, R. G., Roeder, R. G., Clarke, S., Stallcup, M. R., Allis, C. D. & Coonrod, S. A. Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science (80-.).* **306,** 279–283 (2004).

127. Wei, Y., Zhou, H., Sun, Y., He, Y. & Luo, Y. Insight into the catalytic mechanism of arginine deiminase: functional studies on the crucial sites. *Proteins* **66,** 740–50 (2007).

128. Thompson, P. R. & Fast, W. Histone citrullination by protein arginine deiminase: is arginine methylation a green light or a roadblock? *ACS Chem. Biol.* **1,** 433–41 (2006).

129. Watanabe, M., Sumida, N., Murakami, S., Anzai, H., Thompson, C. J., Tateno, Y. & Murakami, T. A phosphonate-induced gene which promotes Penicillium-mediated bioconversion of cis-propenylphosphonic acid to fosfomycin. *Appl. Environ. Microbiol.* **65,** 1036–44 (1999).

130. Zhang, Z., Ren, J., Stammers, D. K., Baldwin, J. E., Harlos, K. & Schofield, C. J. Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat. Struct. Biol.* **7,** 127–33 (2000).

131. Clifton, I. J., McDonough, M. A., Ehrismann, D., Kershaw, N. J., Granatino, N. & Schofield, C. J. Structural studies on 2-oxoglutarate oxygenases and related double-stranded [beta]-helix fold proteins. *J. Inorg. Biochem.* **100,** 644–669 (2006).

132. Vaillancourt, F. H., Yin, J. & Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme FeII alpha-ketoglutarate- and O2-dependent halogenase. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 10111–6 (2005).

133. Flashman, E. & Schofield, C. J. The most versatile of all reactive intermediates? *Nat Chem Biol* **3,** 86–87 (2007).

134. Baldwin, J. E., Adlington, R. M., Coates, J. B., Crabbe, M. J., Crouch, N. P., Keeping, J. W., Knight, G. C., Schofield, C. J., Ting, H. H. & Vallejo, C. A. Purification and initial characterization of an enzyme with deacetoxycephalosporin C synthetase and hydroxylase activities. *Biochem. J.* **245,** 831–41 (1987).

135. Lloyd, M. D., Lee, H. J., Harlos, K., Zhang, Z. H., Baldwin, J. E., Schofield, C. J., Charnock, J. M., Garner, C. D., Hara, T., Terwisscha van Scheltinga, A. C., Valegård, K., Viklund, J. A., Hajdu, J., Andersson, I., Danielsson, A. & Bhikhabhai, R. Studies on the active site of deacetoxycephalosporin C synthase. *J. Mol. Biol.* **287,** 943–60 (1999).

136. Schneider, J. & Shilatifard, A. Histone demethylation by hydroxylation: chemistry in action. *ACS Chem Biol* **1,** 75–81 (2006).

137. Linke, S., Stojkoski, C., Kewley, R. J., Booker, G. W., Whitelaw, M. L. & Peet, D. J. Substrate requirements of the oxygen-sensing asparaginyl hydroxylase factor-inhibiting hypoxia-inducible factor. *J. Biol. Chem.* **279,** 14391–7 (2004).

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

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

140. Branden, C. I. & Tooze, J. *Introduction to Protein Structure*. (Garland Science, 1998). at <http://www.amazon.co.uk/Introduction-Protein-Structure-Carl-Branden/dp/0815323050>

141. Aik, W., McDonough, M. A., Thalhammer, A., Chowdhury, R. & Schofield, C. J.

Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr. Opin. Struct. Biol.* **22,** 691–700 (2012).

142. Valegård, K., van Scheltinga, A. C., Lloyd, M. D., Hara, T., Ramaswamy, S., Perrakis, A., Thompson, A., Lee, H. J., Baldwin, J. E., Schofield, C. J., Hajdu, J. & Andersson, I. Structure of a cephalosporin synthase. *Nature* **394,** 805–9 (1998).

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

144. Zhou, J., Gunsior, M., Bachmann, B. O., Townsend, C. A. & Solomon, E. I. Substrate Binding to the α-Ketoglutarate-Dependent Non-Heme Iron Enzyme Clavaminate Synthase 2: Coupling Mechanism of Oxidative Decarboxylation and Hydroxylation. *J. Am. Chem. Soc.* **120,** 13539–13540 (1998).

145. Ng, S. S., Kavanagh, K. L., McDonough, M. A., Butler, D., Pilka, E. S., Lienard, B. M., Bray, J. E., Savitsky, P., Gileadi, C., Von Delft, F., Rose, N. R., Offer, J., Scheinost, J. C., Borowski, T., Sundstrom, M., Schofield, C. J. & Oppermann, U. Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **448,** 87–91 (2007).

146. Hewitson, K. S., Granatino, N., Welford, R. W. D., McDonough, M. A. & Schofield, C. J. Oxidation by 2-oxoglutarate oxygenases: non-haem iron systems in catalysis and signalling. *Philos. Trans. A. Math. Phys. Eng. Sci.* **363,** 807–28; discussion 1035–40 (2005).

147. Costas, M., Mehn, M. P., Jensen, M. P. & Que, L. Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem. Rev.* **104,** 939–86 (2004).

148. Hoffart, L. M., Barr, E. W., Guyer, R. B., Bollinger, J. M. & Krebs, C. Direct spectroscopic detection of a C-H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 14738–43 (2006).

149. Flashman, E., Hoffart, L. M., Hamed, R. B., Bollinger, J. M., Krebs, C. & Schofield, C. J. Evidence for the slow reaction of hypoxia-inducible factor prolyl hydroxylase 2 with oxygen. *FEBS J.* **277,** 4089–99 (2010).

150. Murad, S. Regulation of Collagen Synthesis by Ascorbic Acid. *Proc. Natl. Acad. Sci.* **78,** 2879–2882 (1981).

151. Myllyla, R., Majamaa, K., Gunzler, V. A., Hanauske-Abel, H. M. & Kivirikko, K. I. Ascorbate Is Consumed Stoichiometrically in the Uncoupled Reactions Catalyzed by Prolyl 4-Hydroxylase and Lysyl Hydroxylase. *J. Biol. Chem.* **259,** 5403–5405 (1984).

152. Kuttan, R., Parrott, D. P., Kaplan, S. R. & Fuller, G. C. Effect of ascorbic acid on prolyl hydroxylase activity, collagen hydroxylation and collagen synthesis in human synovial cells in culture. *Res. Commun. Chem. Pathol. Pharmacol.* **26,** 337–45 (1979).

153. Semenza, G. L. Hypoxia-inducible factor 1: oxygen homeostasis and disease pathophysiology. *Trends Mol. Med.* **7,** 345–50 (2001).

154. Loenarz, C., Chowdhury, R., Schofield, C. J. & Flashman, E. Oxygenases for oxygen sensing. *Pure Appl. Chem.* **80,** 1837–1847 (2008).

155. Chowdhury, R. (University O. O., Hardy, A. (University O. O. & Schofield, C. J. (University O. O. The human oxygen sensing machinery and its manipulation. *Chem. Soc. Rev.* **37,** 1308–1319 (2008).

156. Sánchez-Fernández, E. M., Tarhonskaya, H., Al-Qahtani, K., Hopkinson, R. J., McCullagh, J. S. O., Schofield, C. J. & Flashman, E. Investigations on the oxygen dependence of a 2-oxoglutarate histone demethylase. *Biochem. J.* **449,** 491–6 (2013).

157. Tausendschön, M., Dehne, N. & Brüne, B. Hypoxia causes epigenetic gene regulation in macrophages by attenuating Jumonji histone demethylase activity. *Cytokine* **53,** 256–62 (2011).

158. Hausinger, R. P. Fe(II)/α-Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Crit. Rev. Biochem. Mol. Biol.* **39,** 21–68 (2004).

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

160. Lu, T., Jackson, M. W., Wang, B., Yang, M., Chance, M. R., Miyagi, M., Gudkov, A. V & Stark, G. R. Regulation of NF-kappaB by NSD1/FBXL11-dependent reversible lysine methylation of p65. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 46–51 (2010).

161. Ponnaluri, V. K., Vavilala, D. T., Putty, S., Gutheil, W. G. & Mukherji, M. Identification of non-histone substrates for JMJD2A-C histone demethylases. *Biochem Biophys Res Commun* (2009). at <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19799855>

162. Chen, Z., Zang, J., Whetstine, J., Hong, X., Davrazou, F., Kutateladze, T. G., Simpson, M., Mao, Q., Pan, C. H., Dai, S., Hagman, J., Hansen, K., Shi, Y. & Zhang, G. Structural insights into histone demethylation by JMJD2 family members. *Cell* **125,** 691–702 (2006).

163. Hillringhaus, L., Yue, W. W., Rose, N. R., Ng, S. S., Gileadi, C., Loenarz, C., Bello, S., Bray, J. E., Schofield, C. J. & Oppermann, U. Structural and evolutionary basis for the dual substrate electivity of the human KDM4 histone demethylase family. *J. Biol. Chem.* **286,** 41616–41625 (2011).

164. 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).

165. Zhou, X. & Ma, H. Evolutionary history of histone demethylase families: distinct evolutionary patterns suggest functional divergence. *BMC Evol. Biol.* **8,** 294 (2008).

166. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* **7,** 715–727 (2006).

167. Elkins, J. M., Hewitson, K. S., McNeill, L. A., Seibel, J. F., Schlemminger, I., Pugh, C. W., Ratcliffe, P. J. & Schofield, C. J. Structure of factor-inhibiting hypoxia-

inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J. Biol. Chem.* **278,** 1802–6 (2003).

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

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

170. Hsia, D. A., Tepper, C. G., Pochampalli, M. R., Hsia, E. Y., Izumiya, C., Huerta, S. B., Wright, M. E., Chen, H. W., Kung, H. J. & Izumiya, Y. KDM8, a H3K36me2 histone demethylase that acts in the cyclin A1 coding region to regulate cancer cell proliferation. *Proc Natl Acad Sci U S A* **107,** 9671–9676 (2010).

171. Youn, M.-Y., Yokoyama, A., Fujiyama-Nakamura, S., Ohtake, F., Minehata, K., Yasuda, H., Suzuki, T., Kato, S. & Imai, Y. JMJD5, a Jumonji C (JmjC) domain-containing protein, negatively regulates osteoclastogenesis by facilitating NFATc1 protein degradation. *J. Biol. Chem.* **287,** 12994–3004 (2012).

172. Cockman, M. E., Webb, J. D. & Ratcliffe, P. J. FIH-dependent asparaginyl hydroxylation of ankyrin repeat domain-containing proteins. *Ann. N. Y. Acad. Sci.* **1177,** 9–18 (2009).

173. Yang, M., Chowdhury, R., Ge, W., Hamed, R. B., McDonough, M. A., Claridge, T. D. W., Kessler, B. M., Cockman, M. E., Ratcliffe, P. J. & Schofield, C. J. Factor-inhibiting hypoxia-inducible factor (FIH) catalyses the post-translational hydroxylation of histidinyl residues within ankyrin repeat domains. *FEBS J.* **278,** 1086–97 (2011).

174. Sinha, K. M., Yasuda, H., Coombes, M. M., Dent, S. Y. R. & de Crombrugghe, B. Regulation of the osteoblast-specific transcription factor Osterix by NO66, a Jumonji family histone demethylase. *EMBO J.* **29,** 68–79 (2010).

175. Blackledge, N. P., Zhou, J. C., Tolstorukov, M. Y., Farcas, A. M., Park, P. J. & Klose, R. J. CpG Islands Recruit a Histone H3 Lysine 36 Demethylase. *Mol. Cell* **38,** 179–190 (2010).

176. Horton, J. R., Upadhyay, A. K., Qi, H. H., Zhang, X., Shi, Y. & Cheng, X. Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat. Struct. Mol. Biol.* **17,** 38–43 (2010).

177. 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 (2010).

178. De la Cruz, X., Lois, S., Sánchez-Molina, S. & Martínez-Balbás, M. A. Do protein motifs read the histone code? *Bioessays* **27,** 164–75 (2005).

179. Filippakopoulos, P. & Knapp, S. The bromodomain interaction module. *FEBS Lett.* **586,** 2692–704 (2012).

180. Mujtaba, S., Zeng, L. & Zhou, M.-M. Structure and acetyl-lysine recognition of the bromodomain. *Oncogene* **26,** 5521–7 (2007).

181. Muller, S., Filippakopoulos, P. & Knapp, S. Bromodomains as therapeutic targets. *Expert Rev. Mol. Med.* **13,** e29 (2011).

182. Mujtaba, S., He, Y., Zeng, L., Yan, S., Plotnikova, O., Sachchidanand, Sanchez, R., Zeleznik-Le, N. J., Ronai, Z. & Zhou, M.-M. Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Mol. Cell* **13,** 251–63 (2004).

183. Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.-P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., Gingras, A.-C., Arrowsmith, C. H. & Knapp, S. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149,** 214–31 (2012).

184. Liu, K., Guo, Y., Liu, H., Bian, C., Lam, R., Liu, Y., Mackenzie, F., Rojas, L. A., Reinberg, D., Bedford, M. T., Xu, R.-M. & Min, J. Crystal structure of TDRD3 and methyl-arginine binding characterization of TDRD3, SMN and SPF30. *PLoS One* **7,** e30375 (2012).

185. Li, J., Li, Z., Ruan, J., Xu, C., Tong, Y., Pan, P. W., Tempel, W., Crombet, L., Min, J. & Zang, J. Structural Basis for Specific Binding of Human MPP8 Chromodomain to Histone H3 Methylated at Lysine 9. *PLoS One* **6,** e25104 (2011).

186. Lallous, N., Legrand, P., McEwen, A. G., Ramón-Maiques, S., Samama, J.-P. & Birck, C. The PHD finger of human UHRF1 reveals a new subgroup of unmethylated histone H3 tail readers. *PLoS One* **6,** e27599 (2011).

187. Yang, Y. & Bedford, M. T. Protein arginine methyltransferases and cancer. *Nat. Rev. Cancer* **13,** 37–50 (2013).

188. Pek, J. W., Anand, A. & Kai, T. Tudor domain proteins in development. *Development* **139,** 2255–66 (2012).

189. Chen, C., Nott, T. J., Jin, J. & Pawson, T. Deciphering arginine methylation: Tudor tells the tale. *Nat. Rev. Mol. Cell Biol.* **12,** 629–42 (2011).

190. Paro, R. & Hogness, D. S. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 263–7 (1991).

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

192. Bouazoune, K. The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. *EMBO J.* **21,** 2430–2440 (2002).

193. Bienz, M. The PHD finger, a nuclear protein-interaction domain. **31,** 35–40 (2006).

194. Schindler, U., Beckmann, H. & Cashmore, A. R. HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. *Plant J.* **4,** 137–150 (1993).

195. Garske, A. L., Oliver, S. S., Wagner, E. K., Musselman, C. A., LeRoy, G., Garcia, B. A., Kutateladze, T. G. & Denu, J. M. Combinatorial profiling of chromatin binding

modules reveals multisite discrimination. *Nat. Chem. Biol.* **6,** 283–290 (2010).

196. Baker, L. A., Allis, C. D. & Wang, G. G. PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. *Mutat. Res.* **647,** 3–12 (2008).

197. Mellor, J. The dynamics of chromatin remodeling at promoters. *Mol. Cell* **19,** 147–57 (2005).

198. Mellor, J. Dynamic nucleosomes and gene transcription. *Trends Genet.* **22,** 320–9 (2006).

199. Clapier, C. R. & Cairns, B. R. The Biology of Chromatin Remodeling Complexes. *Annu. Rev. Biochem.* **78,** 273–304 (2009).

200. Tang, L., Nogales, E. & Ciferri, C. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. *Prog. Biophys. Mol. Biol.* **102,** 122–8

201. Erdel, F. & Rippe, K. Chromatin remodelling in mammalian cells by ISWI-type complexes--where, when and why? *FEBS J.* **278,** 3608–18 (2011).

202. Bao, Y. & Shen, X. INO80 subfamily of chromatin remodeling complexes. *Mutat. Res.* **618,** 18–29 (2007).

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

204. Nair, S. S. & Kumar, R. Chromatin remodeling in cancer: a gateway to regulate gene transcription. *Mol. Oncol.* **6,** 611–9 (2012).

205. Watanabe, S. & Peterson, C. L. The INO80 family of chromatin-remodeling enzymes: regulators of histone variant dynamics. *Cold Spring Harb. Symp. Quant. Biol.* **75,** 35–42 (2010).

206. Zhang, Y., Smith, C. L., Saha, A., Grill, S. W., Mihardja, S., Smith, S. B., Cairns, B. R., Peterson, C. L. & Bustamante, C. DNA translocation and loop formation mechanism of chromatin remodeling by SWI/SNF and RSC. *Mol. Cell* **24,** 559–68 (2006).

207. Strohner, R., Wachsmuth, M., Dachauer, K., Mazurkiewicz, J., Hochstatter, J., Rippe, K. & Längst, G. A "loop recapture" mechanism for ACF-dependent nucleosome remodeling. *Nat. Struct. Mol. Biol.* **12,** 683–90 (2005).

208. Vissers, L. E. L. M., van Ravenswaaij, C. M. A., Admiraal, R., Hurst, J. A., de Vries, B. B. A., Janssen, I. M., van der Vliet, W. A., Huys, E. H. L. P. G., de Jong, P. J., Hamel, B. C. J., Schoenmakers, E. F. P. M., Brunner, H. G., Veltman, J. A. & van Kessel, A. G. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat. Genet.* **36,** 955–7 (2004).

209. Welstead, G. G., Creyghton, M. P., Bilodeau, S., Cheng, A. W., Markoulaki, S., Young, R. A. & Jaenisch, R. X-linked H3K27me3 demethylase Utx is required for embryonic development in a sex-specific manner. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 13004–9 (2012).

210. Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E. & Sasaki, H. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429,** 900–3 (2004).

211. Arrowsmith, C. H., Bountra, C., Fish, P. V, Lee, K. & Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Discov.* **11,** 384–400 (2012).

212. Cohen, I., Poręba, E., Kamieniarz, K. & Schneider, R. Histone modifiers in cancer: friends or foes? *Genes Cancer* **2,** 631–47 (2011).

213. Slany, R. K. When epigenetics kills: MLL fusion proteins in leukemia. *Hematol. Oncol.* **23,** 1–9 (2005).

214. Ayton, P. M. & Cleary, M. L. Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. *Oncogene* **20,** 5695–707 (2001).

215. Marks, P. A. & Breslow, R. Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug. *Nat. Biotechnol.* **25,** 84–90 (2007).

216. Qiu, Y., Liu, L., Zhao, C., Han, C., Li, F., Zhang, J., Wang, Y., Li, G., Mei, Y., Wu, M., Wu, J. & Shi, Y. Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ reveals a regulatory mechanism for HOXA9 transcription. *Genes Dev.* **26,** 1376–91 (2012).

217. Zeng, L., Zhang, Q., Li, S., Plotnikov, A. N., Walsh, M. J. & Zhou, M.-M. Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* **466,** 258–62 (2010).

218. Musselman, C. A., Mansfield, R. E., Garske, A. L., Davrazou, F., Kwan, A. H., Oliver, S. S., O'Leary, H., Denu, J. M., Mackay, J. P. & Kutateladze, T. G. Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochem. J.* **423,** 179–187 (2009).

219. Kleine-Kohlbrecher, D., Christensen, J., Vandamme, J., Abarrategui, I., Bak, M., Tommerup, N., Shi, X., Gozani, O., Rappsilber, J. & Salcini, A. E. A Functional Link between the Histone Demethylase PHF8 and the Transcription Factor ZNF711 in X-Linked Mental Retardation. *Mol. Cell* **38,** 165–178 (2010).

220. Laumonnier, F., Holbert, S., Ronce, N., Faravelli, F., Lenzner, S., Schwartz, C. E., Lespinasse, J., Van Esch, H., Lacombe, D., Goizet, C., Phan-Dinh Tuy, F., van Bokhoven, H., Fryns, J.-P., Chelly, J., Ropers, H.-H., Moraine, C., Hamel, B. C. J. & Briault, S. Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. *J. Med. Genet.* **42,** 780–6 (2005).

221. Koivisto, a M., Ala-Mello, S., Lemmelä, S., Komu, H. a, Rautio, J. & Järvelä, I. Screening of mutations in the PHF8 gene and identification of a novel mutation in a Finnish family with XLMR and cleft lip/cleft palate. *Clin. Genet.* **72,** 145–9 (2007).

222. Abidi, F. E., Miano, M. G., Murray, J. C. & Schwartz, C. E. A novel mutation in the PHF8 gene is associated with X-linked mental retardation with cleft lip/cleft palate. *Clin. Genet.* **72,** 19–22 (2007).

223. Coleman, M. L., McDonough, M. A., Hewitson, K. S., Coles, C., Mecinovic, J., Edelmann, M., Cook, K. M., Cockman, M. E., Lancaster, D. E., Kessler, B. M., Oldham, N. J., Ratcliffe, P. J. & Schofield, C. J. Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J. Biol. Chem.* **282,** 24027–38 (2007).

224. Cockman, M. E., Lancaster, D. E., Stolze, I. P., Hewitson, K. S., McDonough, M. a, Coleman, M. L., Coles, C. H., Yu, X., Hay, R. T., Ley, S. C., Pugh, C. W., Oldham, N. J., Masson, N., Schofield, C. J. & Ratcliffe, P. J. Posttranslational hydroxylation of ankyrin repeats in IkappaB proteins by the hypoxia-inducible factor (HIF) asparaginyl hydroxylase, factor inhibiting HIF (FIH). *Proc. Natl. Acad. Sci. U. S. A.* **103,** 14767–72 (2006).

225. Agger, K., Cloos, P. A. C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A. E. & Helin, K. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449,** 731–734 (2007).

226. Hong, S., Cho, Y.-W., Yu, L.-R., Yu, H., Veenstra, T. D. & Ge, K. Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 18439–44 (2007).

227. Lan, F., Bayliss, P. E., Rinn, J. L., Whetstine, J. R., Wang, J. K., Chen, S., Iwase, S., Alpatov, R., Issaeva, I., Canaani, E., Roberts, T. M., Chang, H. Y. & Shi, Y. A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* **449,** 689–94 (2007).

228. Greenfield, A. The UTX gene escapes X inactivation in mice and humans. *Hum. Mol. Genet.* **7,** 737–742 (1998).

229. Lee, S., Lee, J. W. & Lee, S.-K. UTX, a histone H3-lysine 27 demethylase, acts as a critical switch to activate the cardiac developmental program. *Dev. Cell* **22,** 25–37 (2012).

230. Xue, Y., Wong, J., Moreno, G. T., Young, M. K., Côté, J. & Wang, W. NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol. Cell* **2,** 851–61 (1998).

231. Bowen, N. J., Fujita, N., Kajita, M. & Wade, P. A. Mi-2/NuRD: multiple complexes for many purposes. *Biochim. Biophys. Acta* **1677,** 52–7 (2004).

232. Eisen, J. A., Sweder, K. S. & Hanawalt, P. C. Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res.* **23,** 2715–23 (1995).

233. Flanagan, J. F., Blus, B. J., Kim, D., Clines, K. L., Rastinejad, F. & Khorasanizadeh, S. Molecular implications of evolutionary differences in CHD double chromodomains. *J. Mol. Biol.* **369,** 334–42 (2007).

234. Denslow, S. A. & Wade, P. A. The human Mi-2/NuRD complex and gene regulation. *Oncogene* **26,** 5433–8 (2007).

235. Ho, L. & Crabtree, G. R. Chromatin remodelling during development. *Nature* **463,**

474–84 (2010).

236. Williams, C. J., Naito, T., Arco, P. G.-D., Seavitt, J. R., Cashman, S. M., De Souza, B., Qi, X., Keables, P., Von Andrian, U. H. & Georgopoulos, K. The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development. *Immunity* **20,** 719–33 (2004).

237. Seelig, H. P., Moosbrugger, I., Ehrfeld, H., Fink, T., Renz, M. & Genth, E. The major dermatomyositis-specific Mi-2 autoantigen is a presumed helicase involved in transcriptional activation. *Arthritis Rheum.* **38,** 1389–99 (1995).

238. Bottomley, M. J. Structures of protein domains that create or recognize histone modifications. *EMBO Rep.* **5,** 464–9 (2004).

239. Musselman, C. A., Ramírez, J., Sims, J. K., Mansfield, R. E., Oliver, S. S., Denu, J. M., Mackay, J. P., Wade, P. A., Hagman, J. & Kutateladze, T. G. Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 787–92 (2012).

240. Kwan, a, Gell, D., Verger, a, Crossley, M., Matthews, J. & Mackay, J. Engineering a Protein Scaffold from a PHD Finger. *Structure* **11,** 803–813 (2003).

241. Nilges, M. Structure calculation from NMR data. *Curr. Opin. Struct. Biol.* **6,** 617–23 (1996).

242. Wüthrich, K. NMR - this other method for protein and nucleic acid structure determination. *Acta Crystallogr. D. Biol. Crystallogr.* **51,** 249–70 (1995).

243. Anderson, W. A. & Freeman, R. Influence of a Second Radiofrequency Field on High-Resolution Nuclear Magnetic Resonance Spectra. *J. Chem. Phys.* **37,** 85 (1962).

244. Kanelis, V., Forman-Kay, J. D. & Kay, L. E. Multidimensional NMR methods for protein structure determination. *IUBMB Life* **52,** 291–302 (2001).

245. Morra, R., Lee, B. M., Shaw, H., Tuma, R. & Mancini, E. J. Concerted action of the PHD, chromo and motor domains regulates the human chromatin remodelling ATPase CHD4. *FEBS Lett.* **586,** 2513–21 (2012).

246. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to . *Biochemistry* **28,** 6150–6 (1989).

247. Bax, A., Clore, M. G. & Gronenborn, A. M. H-1-H-1 correlation via isotropic mixing of C-13 magnetization, a new 3-dimensional approach for assigning H-1 and C-13 spectra of C-13- enriched proteins. *J. Magn. Reson.* **88,** 425–431 (1990).

248. Grzesiek, S., Anglister, J. & Bax, A. Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in C-13/N-15-Enriched Proteins By Isotropic Mixing of C-13 Magnetization. *J. Magn. Reson. Ser. B* **101,** 114–119 (1993).

249. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance

NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* **89,** 496–514 (1990).

250. Clubb, R., Thanabal, V. & Wagner, G. A Constant-Time 3-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue H-1(N), N-15, and C-13(') Chemical-Shifts in N-15-C-13-Labeled Proteins. *J. Magn. Reson.* **97,** 213 – 217 (1992).

251. Vuister, G. W. & Bax, A. *Resolution enhancement and spectral editing of uniformly 13C-enriched proteins by homonuclear broadband 13C decoupling*. *J. Magn. Reson.* **98,** 428–435 (1992).

252. Santoro, J. & King, G. C. A constant-time 2D overbodenhausen experiment for inverse correlation of isotopically enriched species. *J. Magn. Reson.* **97,** 202–207 (1992).

253. Minch, M. J. Orientational dependence of vicinal proton-proton NMR coupling constants: The Karplus relationship. *Concepts Magn. Reson.* **6,** 41–56 (1994).

254. Wishart, D. S. & Wang, Y. Facile measurement of polypeptide JHNH alpha coupling constants from HMQC-J spectra. *J. Biomol. NMR* **11,** 329–36 (1998).

255. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56,** 227–41 (2013).

256. Vitagliano, L., Berisio, R., Mastrangelo, A., Mazzarella, L. & Zagari, A. Preferred proline puckerings in cis and trans peptide groups: implications for collagen stability. *Protein Sci.* **10,** 2627–32 (2001).

257. Wüthrich, K. *NMR of proteins and nucleic acids*. (Wiley, 1986).

258. Shen, Y. & Bax, A. Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. *J. Biomol. NMR* **46,** 199–204 (2010).

259. Siemion, I. Z., Wieland, T. & Pook, K. H. Influence of the distance of the proline carbonyl from the beta and gamma carbon on the 13C chemical shifts. *Angew. Chem. Int. Ed. Engl.* **14,** 702–3 (1975).

260. Kay, L. E., Torchia, D. A. & Bax, A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28,** 8972–9 (1989).

261. Nilges, M., Gronenborn, A. M., Brünger, A. T. & Clore, G. M. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng.* **2,** 27–38 (1988).

262. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160,** 65–73 (2003).

263. Schwieters, C. D., Kuszewski, J. J. & Clore, G. M. Using Xplor-NIH for NMR Molecular Structure Determination. *ChemInform* **37,** (2006).

264. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Protein Sci.* **5,** 1067–80 (1996).

265. Folmer, R. H., Hilbers, C. W., Konings, R. N. & Nilges, M. Floating stereospecific assignment revisited: application to an 18 kDa protein and comparison with J-coupling data. *J. Biomol. NMR* **9,** 245–58 (1997).

266. Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 9279–83 (1995).

267. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Mol. Biol.* **4,** 732–8 (1997).

268. Clore, G. M., Starich, M. R. & Gronenborn, A. M. Measurement of Residual Dipolar Couplings of Macromolecules Aligned in the Nematic Phase of a Colloidal Suspension of Rod-Shaped Viruses. *J. Am. Chem. Soc.* **120,** 10571–10572 (1998).

269. Yao, L., Ying, J. & Bax, A. Improved accuracy of 15N-1H scalar and residual dipolar couplings from gradient-enhanced IPAP-HSQC experiments on protonated proteins. *J. Biomol. NMR* **43,** 161–70 (2009).

270. Mertens, H. D. T. & Svergun, D. I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172,** 128–41 (2010).

271. Watson, A. A. *et al.* The PHD and chromo domains regulate the ATPase activity of the human chromatin remodeler CHD4. *J. Mol. Biol.* **422,** 3–17 (2012).

272. Bernadó, P. & Svergun, D. I. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* **8,** 151–67 (2012).

273. Bertini, I., Giachetti, A., Luchinat, C., Parigi, G., Petoukhov, M. V, Pierattelli, R., Ravera, E. & Svergun, D. I. Conformational space of flexible biological macromolecules from average data. *J. Am. Chem. Soc.* **132,** 13553–8 (2010).

274. Li, Y. & Li, H. Many keys to push: diversifying the "readership" of plant homeodomain fingers. *Acta Biochim. Biophys. Sin. (Shanghai).* **44,** 28–39 (2012).

275. Slama, P. & Geman, D. Identification of family-determining residues in PHD fingers. *Nucleic Acids Res.* 1–14 (2010). doi:10.1093/nar/gkq947

276. Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M., Tackett, A. J., Chait, B. T., Badenhorst, P., Wu, C. & Allis, C. D. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442,** 86–90 (2006).

277. Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D. & Patel, D. J. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442,** 91–5 (2006).

278. Shi, X. *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442,** 96–9 (2006).

279. Peña, P. V, Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V, Gozani, O., Zhao, R. & Kutateladze, T. G. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442,** 100–3 (2006).

280. Eustermann, S., Yang, J.-C., Law, M. J., Amos, R., Chapman, L. M., Jelinska, C., Garrick, D., Clynes, D., Gibbons, R. J., Rhodes, D., Higgs, D. R. & Neuhaus, D. Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. *Nat. Struct. Mol. Biol.* **18,** 777–82 (2011).

281. Wang, C., Shen, J., Yang, Z., Chen, P., Zhao, B., Hu, W., Lan, W., Tong, X., Wu, H., Li, G. & Cao, C. Structural basis for site-specific reading of unmodified R2 of histone H3 tail by UHRF1 PHD finger. *Cell Res.* **21,** 1379–82 (2011).

282. Lan, F., Collins, R. E., De Cegli, R., Alpatov, R., Horton, J. R., Shi, X., Gozani, O., Cheng, X. & Shi, Y. Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448,** 718–22 (2007).

283. Chakravarty, S., Zeng, L. & Zhou, M.-M. Structure and site-specific recognition of histone H3 by the PHD finger of human autoimmune regulator. *Structure* **17,** 670–9 (2009).

284. Dhar, S. S., Lee, S.-H., Kan, P.-Y., Voigt, P., Ma, L., Shi, X., Reinberg, D. & Lee, M. G. Trans-tail regulation of MLL4-catalyzed H3K4 methylation by H4R3 symmetric dimethylation is mediated by a tandem PHD of MLL4. *Genes Dev.* **26,** 2749–62 (2012).

285. Miller, T. C. R., Rutherford, T. J., Johnson, C. M., Fiedler, M. & Bienz, M. Allosteric remodelling of the histone H3 binding pocket in the Pygo2 PHD finger triggered by its binding to the B9L/BCL9 co-factor. *J. Mol. Biol.* **401,** 969–84 (2010).

286. Wang, Z., Song, J., Milne, T. A., Wang, G. G., Li, H., Allis, C. D. & Patel, D. J. Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. *Cell* **141,** 1183–94 (2010).

287. Clarkson, J. & Campbell, I. D. Studies of protein-ligand interactions by NMR. *Biochem. Soc. Trans.* **31,** 1006–9 (2003).

288. Fielding, L. NMR methods for the determination of protein-ligand dissociation constants. *Curr. Top. Med. Chem.* **3,** 39–53 (2003).

289. Rajagopal, P., Waygood, E. B., Reizer, J., Saier, M. H. & Klevit, R. E. Demonstration of protein-protein interaction specificity by NMR chemical shift mapping. *Protein Sci.* **6,** 2624–7 (1997).

290. Mansfield, R. E., Musselman, C. A., Kwan, A. H., Oliver, S. S., Garske, A. L., Davrazou, F., Denu, J. M., Kutateladze, T. G. & Mackay, J. P. Plant Homeodomain (PHD) Fingers of CHD4 Are Histone H3-binding Modules with Preference for Unmodified H3K4 and Methylated H3K9. *J Biol Chem* **286,** 11779–11791 (2011).

291. Zeng, L., Yap, K. L., Ivanov, A. V, Wang, X., Mujtaba, S., Plotnikova, O., Rauscher,

F. J. & Zhou, M.-M. Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. *Nat. Struct. Mol. Biol.* **15,** 626–33 (2008).

292. Loenarz, C., Ge, W., Coleman, M. L., Rose, N. R., Cooper, C. D. O., Klose, R. J., Ratcliffe, P. J. & Schofield, C. J. PHF8, a gene associated with cleft lip/palate and mental retardation, encodes for an Nepsilon-dimethyl lysine demethylase. *Hum. Mol. Genet.* **19,** 217–22 (2010).

293. Qiu, J., Shi, G., Jia, Y., Li, J., Wu, M., Li, J., Dong, S. & Wong, J. The X-linked mental retardation gene PHF8 is a histone demethylase involved in neuronal differentiation. *Cell Res.* **20,** 908–918 (2010).

294. Yokoyama, A., Okuno, Y., Chikanishi, T., Hashiba, W., Sekine, H., Fujiki, R. & Kato, S. KIAA1718 is a histone demethylase that erases repressive histone methyl marks. *Genes Cells* **15,** 867–73 (2010).

295. Qi, H. H., Sarkissian, M., Hu, G.-Q., Wang, Z., Bhattacharjee, A., Gordon, D. B., Gonzales, M., Lan, F., Ongusaha, P. P., Huarte, M., Yaghi, N. K., Lim, H., Garcia, B. a, Brizuela, L., Zhao, K., Roberts, T. M. & Shi, Y. Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. *Nature* **466,** (2010).

296. Liu, W., Tanasa, B., Tyurina, O. V, Zhou, T. Y., Gassmann, R., Liu, W. T., Ohgi, K. a, Benner, C., Garcia-Bassets, I., Aggarwal, A. K., Desai, A., Dorrestein, P. C., Glass, C. K. & Rosenfeld, M. G. PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* **466,** (2010).

297. Yue, W. W., Hozjan, V., Ge, W., Loenarz, C., Cooper, C. D. O., Schofield, C. J., Kavanagh, K. L., Oppermann, U. & McDonough, M. a. Crystal structure of the PHF8 Jumonji domain, an Nepsilon-methyl lysine demethylase. *FEBS Lett.* **584,** 825–30 (2010).

298. Wen, H., Li, J., Song, T., Lu, M., Kan, P.-Y., Lee, M. G., Sha, B. & Shi, X. Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. *J. Biol. Chem.* **285,** 9322–6 (2010).

299. Baba, A., Ohtake, F., Okuno, Y., Yokota, K., Okada, M., Imai, Y., Ni, M., Meyer, C. a, Igarashi, K., Kanno, J., Brown, M. & Kato, S. PKA-dependent regulation of the histone lysine demethylase complex PHF2-ARID5B. *Nat. Cell Biol.* **13,** 1–10 (2011).

300. Tsukada, Y., 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,** 432–7 (2010).

301. Huang, C., Xiang, Y., Wang, Y., Li, X., Xu, L., Zhu, Z., Zhang, T., Zhu, Q., Zhang, K., Jing, N. & Chen, C. D. Dual-specificity histone demethylase KIAA1718 (KDM7A) regulates neural differentiation through FGF4. *Cell Res.* **20,** 154–65 (2010).

302. Suganuma, T. & Workman, J. L. Features of the PHF8/KIAA1718 histone demethylase. *Cell Res.* **20,** 861–2 (2010).

303. Rossi, A. M. & Taylor, C. W. Analysis of protein-ligand interactions by fluorescence polarization. *Nat. Protoc.* **6,** 365–87 (2011).

304. Liepinsh, E., Vilskersts, R., Loca, D., Kirjanova, O., Pugovichs, O., Kalvinsh, I. & Dambrova, M. Mildronate, an inhibitor of carnitine biosynthesis, induces an increase in gamma-butyrobetaine contents and cardioprotection in isolated rat heart infarction. *J. Cardiovasc. Pharmacol.* **48,** 314–9 (2006).

305. Eglen, R. M., Reisine, T., Roby, P., Rouleau, N., Illy, C., Bossé, R. & Bielefeld, M. The use of AlphaScreen technology in HTS: current status. *Curr. Chem. Genomics* **1,** 2–10 (2008).

306. Kawamura, A., Tumber, A., Rose, N. R., King, O. N. F., Daniel, M., Oppermann, U., Heightman, T. D. & Schofield, C. Development of homogeneous luminescence assays for histone demethylase catalysis and binding. *Anal. Biochem.* **404,** 7 (2010).

307. Ullman, E. F. Luminescent Oxygen Channeling Immunoassay: Measurement of Particle Binding Kinetics by Chemiluminescence. *Proc. Natl. Acad. Sci.* **91,** 5426–5430 (1994).

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

309. Riddell, J. A., Hageman, H. A., J'anthony, C. M. & Hubbard, W. L. Retardation of Plant Growth by a New Group of Chemicals. *Science* **136,** 391 (1962).

310. Hasegawa, R., Cabral, R., Hoshiya, T., Hakoi, K., Ogiso, T., Boonyaphiphat, P., Shirai, T. & Ito, N. Carcinogenic potential of some pesticides in a medium-term multi-organ bioassay in rats. *Int. J. Cancer* **54,** 489–93 (1993).

311. B Toth, L. W. Induction of tumors in mice with the herbicide succinic acid 2,2-dimethylhydrazide. *Cancer Res.* **37,** 3497 – 500 (1977).

312. Brown, R. G. S., Kawaide, H., Yang, Y.-Y., Rademacher, W. & Kamiya, Y. Daminozide and prohexadione have similar modes of action as inhibitors of the late stages of gibberellin metabolism. *Physiol. Plant.* **101,** 309–313 (1997).

313. Rademacher, W. GROWTH RETARDANTS: Effects on Gibberellin Biosynthesis and Other Metabolic Pathways. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51,** 501–531 (2000).

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

315. Yu, L., Wang, Y., Huang, S., Wang, J., Deng, Z., Zhang, Q., Wu, W., Zhang, X., Liu, Z., Gong, W. & Chen, Z. Structural insights into a novel histone demethylase PHF8. *Cell Res.* **20,** 166–73 (2010).

316. Huttunen, K. M., Raunio, H. & Rautio, J. Prodrugs--from serendipity to rational design. *Pharmacol. Rev.* **63,** 750–71 (2011).

317. Ryle, M. J., Liu, A., Muthukumaran, R. B., Ho, R. Y. N., Koehntop, K. D., McCracken, J., Que, L. & Hausinger, R. P. O2- and alpha-ketoglutarate-dependent tyrosyl radical formation in TauD, an alpha-keto acid-dependent non-heme iron dioxygenase. *Biochemistry* **42,** 1854–62 (2003).

318. Liu, A., Ho, R. Y. N., Que, L., Ryle, M. J., Phinney, B. S. & Hausinger, R. P. Alternative Reactivity of an α-Ketoglutarate-Dependent Iron(II) Oxygenase: Enzyme Self-Hydroxylation. *J. Am. Chem. Soc.* **123,** 5126–5127 (2001).

319. Mantri, M., Zhang, Z., McDonough, M. A. & Schofield, C. J. Autocatalysed oxidative modifications to 2-oxoglutarate dependent oxygenases. *FEBS J.* **279,** 1563–75 (2012).

320. Siderius, L. E., Hamel, B. C., van Bokhoven, H., de Jager, F., van den Helm, B., Kremer, H., Heineman-de Boer, J. a, Ropers, H. H. & Mariman, E. C. X-linked mental retardation associated with cleft lip/palate maps to Xp11.3-q21.3. *Am. J. Med. Genet.* **85,** 216–20 (1999).

321. Kapuscinski, J. DAPI: a DNA-specific fluorescent probe. *Biotech. Histochem.* **70,** 220–33 (1995).

322. Douglas, S. J., Davis, S. S. & Illum, L. Nanoparticles in drug delivery. *Crit. Rev. Ther. Drug Carrier Syst.* **3,** 233–61 (1987).

323. Maham, A., Tang, Z., Wu, H., Wang, J. & Lin, Y. Protein-based nanomedicine platforms for drug delivery. *Small* **5,** 1706–21 (2009).

324. Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montallier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., Lu, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B. & Zhao, Y. Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* **146,** 1016–1028 (2011).

325. Boffa, L. C., Karn, J., Vidali, G. & Allfrey, V. G. Distribution of NG, NG,-dimethylarginine in nuclear protein fractions. *Biochem. Biophys. Res. Commun.* **74,** 969–76 (1977).

326. Ong, S.-E., Mittler, G. & Mann, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods* **1,** 119–26 (2004).

327. Uhlmann, T., Geoghegan, V. L., Thomas, B., Ridlova, G., Trudgian, D. C. & Acuto, O. A method for large-scale identification of protein arginine methylation. *Mol. Cell. Proteomics* **11,** 1489–99 (2012).

328. Bedford, M. T. & Clarke, S. G. Protein Arginine Methylation in Mammals: Who, What, and Why. *Mol. Cell* **33,** 1–13 (2009).

329. Dillon, M. B. C., Rust, H. L., Thompson, P. R. & Mowen, K. A. Automethylation of Protein Arginine Methyltransferase 8 Regulates Activity by Impeding AdoMet Sensitivity. *J. Biol. Chem.* (2013). doi:10.1074/jbc.M113.491092

330. Blackwell, E. & Ceman, S. Arginine methylation of RNA-binding proteins regulates cell function and differentiation. *Mol. Reprod. Dev.* **79,** 163–75 (2012).

331. Lee, Y.-H. & Stallcup, M. R. Minireview: protein arginine methylation of nonhistone proteins in transcriptional regulation. *Mol. Endocrinol.* **23,** 425–33 (2009).

332. Wei, H., Wang, B., Miyagi, M., She, Y., Gopalan, B., Huang, D.-B., Ghosh, G., Stark, G. R. & Lu, T. PRMT5 dimethylates R30 of the p65 subunit to activate NF-κB. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 13516–21 (2013).

333. Mantri, M., Loik, N. D., Hamed, R. B., Claridge, T. D. W., McCullagh, J. S. O. & Schofield, C. J. The 2-Oxoglutarate-Dependent Oxygenase JMJD6 Catalyses Oxidation of Lysine Residues to give 5S-Hydroxylysine Residues. *ChemBioChem* **12,** 531–534 (2011).

334. Métivier, R., Penot, G., Hübner, M. R., Reid, G., Brand, H., Kos, M. & Gannon, F. Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115,** 751–63 (2003).

335. Le Romancer, M., Treilleux, I., Leconte, N., Robin-Lespinasse, Y., Sentis, S., Bouchekioua-Bouzaghou, K., Goddard, S., Gobert-Gosse, S. & Corbo, L. Regulation of Estrogen Rapid Signaling through Arginine Methylation by PRMT1. *Mol. Cell* **31,** 212–221 (2008).

336. Sakabe, K. & Hart, G. W. O-GlcNAc transferase regulates mitotic chromatin dynamics. *J. Biol. Chem.* **285,** 34460–8 (2010).

337. Thalhammer, A., Bencokova, Z., Poole, R., Loenarz, C., Adam, J., O'Flaherty, L., Schödel, J., Mole, D., Giaslakiotis, K., Schofield, C. J., Hammond, E. M., Ratcliffe, P. J. & Pollard, P. J. Human AlkB homologue 5 is a nuclear 2-oxoglutarate dependent oxygenase and a direct target of hypoxia-inducible factor 1α (HIF-1α). *PLoS One* **6,** e16210 (2011).

338. Westbye, M. P., Feyzi, E., Aas, P. A., Vågbø, C. B., Talstad, V. A., Kavli, B., Hagen, L., Sundheim, O., Akbari, M., Liabakk, N.-B., Slupphaug, G., Otterlei, M. & Krokan, H. E. Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *J. Biol. Chem.* **283,** 25046–56 (2008).

339. Müller, T. A., Meek, K. & Hausinger, R. P. Human AlkB homologue 1 (ABH1) exhibits DNA lyase activity at abasic sites. *DNA Repair (Amst).* **9,** 58–65 (2010).

340. Spencer, D. & Henshall, T. The Kinetics and Mechanism of the Reaction of Formaldehyde with Dimedone. Part I. *J. Am. Chem. Soc.* **77,** 1943–1948 (1955).

341. Hopkinson, R. J., Hamed, R. B., Rose, N. R., Claridge, T. D. W. & Schofield, C. J. Monitoring the Activity of 2-Oxoglutarate Dependent Histone Demethylases by NMR spectroscopy: Direct Observation of Formaldehyde. *ChemBioChem* **11,** 506–510 (2010).

342. Godin, J.-P., Fay, L.-B. & Hopfgartner, G. Liquid chromatography combined with mass spectrometry for 13C isotopic analysis in life science research. *Mass Spectrom. Rev.* **26,** 751–74

343. Werner, R. A. & Brand, W. A. Referencing strategies and techniques in stable isotope ratio analysis. *Rapid Commun. Mass Spectrom.* **15,** 501–19 (2001).

344. Shechter, D., Dormann, H. L., Allis, C. D. & Hake, S. B. Extraction, purification and analysis of histones. *Nat. Protoc.* **2,** 1445–57 (2007).

345. Gurley, L. R., Valdez, J. G., Prentice, D. A. & Spall, W. D. Histone fractionation by high-performance liquid chromatography. *Anal. Biochem.* **129,** 132–144 (1983).

346. Naldi, M., Calonghi, N., Masotti, L., Parolin, C., Valente, S., Mai, A. & Andrisano, V.

Histone post-translational modifications by HPLC-ESI-MS after HT29 cell treatment with histone deacetylase inhibitors. *Proteomics* **9,** 5437–45 (2009).

347. Cheng, D., Yadav, N., King, R. W., Swanson, M. S., Weinstein, E. J. & Bedford, M. T. Small molecule regulators of protein arginine methyltransferases. *J. Biol. Chem.* **279,** 23892–9 (2004).

348. Solá, S., Xavier, J. M., Santos, D. M., Aranha, M. M., Morgado, A. L., Jepsen, K. & Rodrigues, C. M. P. p53 interaction with JMJD3 results in its nuclear distribution during mouse neural stem cell differentiation. *PLoS One* **6,** e18421 (2011).

349. Ene, C. I. *et al.* Histone demethylase Jumonji D3 (JMJD3) as a tumor suppressor by regulating p53 protein nuclear stabilization. *PLoS One* **7,** e51407 (2012).

350. Jansson, M., Durant, S. T., Cho, E.-C., Sheahan, S., Edelmann, M., Kessler, B. M. & La Thangue, N. B. Arginine methylation regulates the p53 response. *Nat. Cell Biol.* **10,** 1431–1439 (2008).

351. Kim, T.-D., Shin, S., Berry, W. L., Oh, S. & Janknecht, R. The JMJD2A demethylase regulates apoptosis and proliferation in colon cancer cells. *J. Cell. Biochem.* **113,** 1368–76 (2012).

352. Kim, T.-D., Oh, S., Shin, S. & Janknecht, R. Regulation of Tumor Suppressor p53 and HCT116 Cell Physiology by Histone Demethylase JMJD2D/KDM4D. *PLoS One* **7,** e34618 (2012).

353. Chen, Y., Zhou, X., Liu, N., Wang, C., Zhang, L., Mo, W. & Hu, G. Arginine methylation of hnRNP K enhances p53 transcriptional activity. *FEBS Lett.* **582,** 1761–5 (2008).

354. Krecic, A. M. & Swanson, M. S. hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.* **11,** 363–71 (1999).

355. Michelotti, E. F., Michelotti, G. A., Aronsohn, A. I. & Levens, D. Heterogeneous nuclear ribonucleoprotein K is a transcription factor. *Mol. Cell. Biol.* **16,** 2350–60 (1996).

356. Chen, Y., Zhou, X., Liu, N., Wang, C., Zhang, L., Mo, W. & Hu, G. Arginine methylation of hnRNP K enhances p53 transcriptional activity. *FEBS Lett.* **582,** 1761–5 (2008).

357. Chiou, Y.-Y., Lin, W.-J., Fu, S.-L. & Lin, C.-H. Direct Mass-Spectrometric Identification of Arg296 and Arg299 as the Methylation Sites of hnRNP K Protein for Methyltransferase PRMT1. *Protein J.* **26,** 87–93 (2006).

358. Chan, J. Y.-H., Hsieh, T.-Y., Liu, S.-T., Chou, W.-Y., Chung, M.-H. & Huang, S.-M. Physical and functional interactions between hnRNP K and PRMT family proteins. *FEBS Lett.* **583,** 281–6 (2009).

359. Ostareck-Lederer, A., Ostareck, D. H., Rucknagel, K. P., Schierhorn, A., Moritz, B., Huttelmaier, S., Flach, N., Handoko, L. & Wahle, E. Asymmetric arginine dimethylation of heterogeneous nuclear ribonucleoprotein K by protein-arginine methyltransferase 1 inhibits its interaction with c-Src. *J. Biol. Chem.* **281,** 11115–25

(2006).

360. Pawlak, M. R., Scherer, C. A., Chen, J., Roshon, M. J. & Ruley, H. E. Arginine N-Methyltransferase 1 Is Required for Early Postimplantation Mouse Development, but Cells Deficient in the Enzyme Are Viable. *Mol. Cell. Biol.* **20,** 4859–4869 (2000).

361. Tang, J. PRMT1 Is the Predominant Type I Protein Arginine Methyltransferase in Mammalian Cells. *J. Biol. Chem.* **275,** 7723–7730 (2000).

362. Lakowski, T. M., Zurita-Lopez, C., Clarke, S. G. & Frankel, A. Approaches to measuring the activities of protein arginine N-methyltransferases. *Anal. Biochem.* **397,** 1–11 (2010).

363. Chang, Y.-I., Hsu, S.-C., Chau, G.-Y., Huang, C.-Y. F., Sung, J.-S., Hua, W.-K. & Lin, W.-J. Identification of the methylation preference region in heterogeneous nuclear ribonucleoprotein K by protein arginine methyltransferase 1 and its implication in regulating nuclear/cytoplasmic distribution. *Biochem. Biophys. Res. Commun.* **404,** 865–9 (2011).

364. Chen, D.-H., Wu, K.-T., Hung, C.-J., Hsieh, M. & Li, C. Effects of adenosine dialdehyde treatment on in vitro and in vivo stable protein methylation in HeLa cells. *J. Biochem.* **136,** 371–6 (2004).

365. Thandapani, P., O'Connor, T. R., Bailey, T. L. & Richard, S. Defining the RGG/RG motif. *Mol. Cell* **50,** 613–23 (2013).

366. Ong, S.-E. & Mann, M. Stable isotope labeling by amino acids in cell culture for quantitative proteomics. *Methods Mol. Biol.* **359,** 37–52 (2007).

367. Van Hoof, D., Pinkse, M. W. H., Oostwaard, D. W.-V., Mummery, C. L., Heck, A. J. R. & Krijgsveld, J. An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nat. Methods* **4,** 677–8 (2007).

368. Bendall, S. C., Hughes, C., Stewart, M. H., Doble, B., Bhatia, M. & Lajoie, G. A. Prevention of amino acid conversion in SILAC experiments with embryonic stem cells. *Mol. Cell. Proteomics* **7,** 1587–97 (2008).

369. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–72 (2008).

370. Goto, Y., Ohta, A., Sako, Y., Yamagishi, Y., Murakami, H. & Suga, H. Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. *ACS Chem. Biol.* **3,** 120–9 (2008).

371. Nicholson, T. B. & Chen, T. No Title. *Epigenetics* **4,** 129–132 (2009).

372. Lee, M. G., Villa, R., Trojer, P., Norman, J., Yan, K.-P., Reinberg, D., Di Croce, L. & Shiekhattar, R. Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* **318,** 447–50 (2007).

373. Laaser, I., Theis, F. J., de Angelis, M. H., Kolb, H.-J. & Adamski, J. Huge splicing frequency in human Y chromosomal UTY gene. *OMICS* **15,** 141–54 (2011).

374. Vogt, M. H., Goulmy, E., Kloosterboer, F. M., Blokland, E., de Paus, R. A., Willemze, R. & Falkenburg, J. H. UTY gene codes for an HLA-B60-restricted human male-specific minor histocompatibility antigen involved in stem cell graft rejection: characterization of the critical polymorphic amino acid residues for T-cell recognition. *Blood* **96,** 3126–32 (2000).

375. Simpson, E. & Roopenian, D. Minor histocompatibility antigens. *Curr. Opin. Immunol.* **9,** 655–61 (1997).

376. Shpargel, K. B., Sengoku, T., Yokoyama, S. & Magnuson, T. UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. *PLoS Genet.* **8,** e1002964 (2012).

377. Sengoku, T. & Yokoyama, S. Structural basis for histone H3 Lys 27 demethylation by UTX/KDM6A. *Genes Dev.* **25,** 2266–77 (2011).

378. LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F. & McCoy, J. M. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. *Biotechnology. (N. Y).* **11,** 187–93 (1993).

379. Bradbury, S. L. & Jakoby, W. B. Glycerol as an enzyme-stabilizing agent: effects on aldehyde dehydrogenase. *Proc. Natl. Acad. Sci. U. S. A.* **69,** 2373–6 (1972).

380. Golovanov, A. P., Hautbergue, G. M., Wilson, S. A. & Lian, L.-Y. A simple method for improving protein solubility and long-term stability. *J. Am. Chem. Soc.* **126,** 8933–9 (2004).

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

382. Bruneau, B. G., Nemer, G., Schmitt, J. P., Charron, F., Robitaille, L., Caron, S., Conner, D. A., Gessler, M., Nemer, M. & Seidman, C. E. A Murine Model of Holt-Oram Syndrome Defines Roles of the T-Box Transcription Factor Tbx5 in Cardiogenesis and Disease. *Cell* **106,** 709–721 (2001).

383. Pearson, J. C., Lemons, D. & McGinnis, W. Modulating Hox gene functions during animal body patterning. *Nat. Rev. Genet.* **6,** 893–904 (2005).

384. She, X., Rohl, C. A., Castle, J. C., Kulkarni, A. V, Johnson, J. M. & Chen, R. Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. *BMC Genomics* **10,** 269 (2009).

385. Cho, Y.-W., Hong, T., Hong, S., Guo, H., Yu, H., Kim, D., Guszczynski, T., Dressler, G. R., Copeland, T. D., Kalkum, M. & Ge, K. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J. Biol. Chem.* **282,** 20395–406 (2007).

386. Jegalian, K. & Page, D. C. A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature* **394,** 776–80 (1998).

387. Berletch, J. B., Deng, X., Nguyen, D. K. & Disteche, C. M. Female bias in Rhox6 and 9 regulation by the histone demethylase KDM6A. *PLoS Genet.* **9,** e1003489 (2013).

388. Savitsky, P., Bray, J., Cooper, C. D. O., Marsden, B. D., Mahajan, P., Burgess-Brown, N. A. & Gileadi, O. High-throughput production of human proteins for crystallization: the SGC experience. *J. Struct. Biol.* **172,** 3–13 (2010).

389. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V & Mann, M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10,** 1794–805 (2011).

390. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–93 (1995).

391. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59,** 687–96 (2005).

392. Pierce, M. M., Raman, C. S. & Nall, B. T. Isothermal titration calorimetry of protein-protein interactions. *Methods* **19,** 213–21 (1999).

393. Woon, E. C. Y., Tumber, A., Kawamura, A., Hillringhaus, L., Rose, N. R., Ma, J. H. Y., Chan, M. C., Walport, L. J., Che, K. H., Ng, S. S., Marsden, B. D., McDonough, M. A. & Schofield, C. J. Linking of 2-Oxoglutarate and Substrate Binding Sites Enables Potent and Highly Selective Inhibition of JmjC Histone Demethylases. *Angew. Chemie Int. Ed.* (2012).

Appendix 1

Chemical shift values for dPHD

		H	N	N^{δ}	N^{ε}	H^{α}	H^{β}	H^{γ}	H^{δ}	H^{ε}	H^{ζ}	H^{η}	C	C^{α}	C^{β}	C^{γ}	C^{δ}
1	Met	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	Glu	-	-	-	-	4.3	1.89, 2.01	2.22	-	-	-	-	175.54	56.93	30.35	36.26	-
3	Thr	7.99	112.83	-	-	4.27	4.13	0.98	-	-	-	-	173.23	61.23	69.92	-	-
4	Asp	8.16	120.82	-	-	4.62	2.66, 2.57	-	-	-	-	-	174.9	54.49	41.52	-	-
5	His	8.19	115.42	-	-	4.91	3.31, 2.86	-	7.65	8.77	-	-	174.74	56.61	32.73	-	-
6	Gln	8.6	118	-	108.72	4.37	1.83, 1.63	2.05, 2.41	-	6.88, 7.27	-	-	173.9	54.74	28.29	32.48	179.09
7	Asp	7.98	115.66	-	-	4.5	2.85, 2.57	-	-	-	-	-	173.72	52.6	40.58	-	-
8	Tyr	6.44	111.7	-	-	4.47	2.26, 2.60	-	6.88	6.75	-	-	171.28	53.68	40.65	-	-
9	Cys	8.34	120.22	-	-	3.97	3.43, 1.74	-	-	-	-	-	175.96	57.22	31.69	-	-
10	Glu	8.81	129.41	-	-	4.13	1.66, 1.41	2.40, 2.40	-	-	-	-	175.9	57.57	31.63	33.97	-
11	Val	8.81	117.89	-	-	4	2.3	1.14, 1.3	-	-	-	-	176.65	65.42	33.2	20.84, 21.71	-
12	Cys	7.94	115.57	-	-	5.02	3.38, 2.96	-	-	-	-	-	175.29	58.85	31.69	-	-
13	Gln	8.44	115.88	-	111.54	4.08	2.16, 2.36	2.12, 2.19	-	7.14, 6.58	-	-	173.7	58.35	25.77	34.67	179.57
14	Gln	8.4	117.18	-	111.89	4.86	2.33, 2.45	2.51, 2.5	-	7.4, 6.91	-	-	175.68	54.35	31.36	34	179.78
15	Gly	8.57	106.92	-	-	3.91, 4.45	-	-	-	-	-	-	173.58	44.07	-	-	-
16	Gly	8.66	109.54	-	-	3.7,4.6	-	-	-	-	-	-	172.36	44.81	-	-	-
17	Glu	8.89	124.44	-	-	4.34	2.11, 2.03	2.23, 2.23	-	-	-	-	174.58	56.63	28.49	36.3	-
18	Ile	7.9	118.1	-	-	4.65	1.44	1.06, 1.18 ,0.04	0.39	-	-	-	174.21	59.37	40.83	17.39, 23.55	12.51
19	Ile	9.2	120.2	-	-	4.34	1.63	1.29, 0.99, 0.55	0.11	-	-	-	174.43	59.77	39.43	26.59, 17.67	13.12
20	Leu	8.45	122.61	-	-	4.67	0.31, 1.54	1.55	0.58, 0.55	-	-	-	175.88	53.93	41.33	27	23.06, 24.71
21	Cys	8.59	124.31	-	-	4.61	2.50, 3.66	-	-	-	-	-	176.02	61.45	30.3	-	-
22	Asp	9.13	125.9	-	-	5.12	2.82, 2.93	-	-	-	-	-	177.13	58.08	42.73	-	-
23	Thr	10.1	112.53	-	-	4.56	4.75	1.04	-	-	-	-	171.84	61.29	69.62	21.93	-
24	Cys	8.04	125.51	-	-	3.91	2.91, 2.74	-	-	-	-	-	170.91	58.92	30.04	-	-
25	Pro	-	-	-	-	4.28	1.96, 2.24	1.88, 1.85	3.28, 3.34	-	-	-	175.75	64.28	32	27.2	50.27
26	Arg	9.15	120.27	-	-	3.78	1.82, 2.02	1.43, 1.42	-	-	-	-	172.58	58.37	31.79	29.7	-
27	Ala	7.03	118.48	-	-	5.59	1.06	-	-	-	-	-	174.83	50.12	24.48	-	-
28	Tyr	8.88	113.59	-	-	5.49	3.02, 2.52	-	7.09	6.91	-	-	176.02	57.43	44.52	-	-
29	His	9.51	119.28	-	-	4.97	3.76, 3.51	-	7.41	7.41	-	-	176.6	57.86	31.39	-	-
30	Met	9.38	124.02	-	-	3.99	2.51, 2.35	1.99, 2.07	-	2.02	-	-	177.66	60.48	32.71	-	-
31	Val	8.17	110.55	-	-	4.05	2.22	1.04, 1.02	-	-	-	-	176.1	64.2	31.55	20.10, 21.1	-
32	Cys	6.91	117.6	-	-	4.22	3.14, 3.11	-	-	-	-	-	175.45	62.32	29.72	-	-
33	Leu	7.47	119.21	-	-	4.32	1.99, 1.57	1.69	0.71, 0.7	-	-	-	176.57	56.41	43.23	26.82	25.15, 24.11
34	Asp	8.02	115.94	-	-	4.79	2.77, 2.66	-	-	-	-	-	173.62	52.21	42.46	-	-
35	Pro	-	-	-	-	4.43	1.99, 2.37	2.10, 2.01	3.75, 3.92	-	-	-	176.01	64.49	32.22	27.17	51.02
36	Asp	8.31	115.45	-	-	4.73	2.56, 2.77	-	-	-	-	-	175.46	54.3	41.35	-	-
37	Met	7.71	121.23	-	-	4.16	2.05, 2.02	2.51, 2.48	-	-	-	-	174.88	57.07	33.28	32.13	-
38	Glu	8.58	123.43	-	-	4.26	1.92, 2.01	2.25, 2.16	-	-	-	-	175.03	57.04	31.64	36.49	-
39	Lys	7.67	118.1	-	-	4.37	1.60, 1.74	1.29, 1.31	1.57	2.89	-	-	173.63	54.86	34.7	24.14	29.15
40	Ala	8.42	125.54	-	-	4.28	1.27	-	-	-	-	-	174.64	50.79	17.27	-	-
41	Pro	-	-	-	-	4.22	1.66, 2.15	1.21, 1.35	3.22, 3.43	-	-	-	- -	63.35	32.28	26.98	50.2
42	Glu	8.37	120.41	-	-	4.37	1.92, 2.04	2.25	-	-	-	-	- -	56.33	30.93	36.45	-
43	Gly	8.3	109.66	-	-	3.91, 4.03		-	-	-	-	-	172.67	44.79	-	-	-
44	Lys	8.16	120.47	-	-	4.44	1.56, 1.66	1.29, 1.31	1.46	2.85	-	-	175.82	56.27	32.68	24.73	-
45	Trp	9.97	127.78	-	128.29	4.68	3.31, 2.95	-	7.29	7.57, 9.82	7.09, 6.62	6.47	172.66	59.09	31	-	-
46	Ser	7.26	121.25	-	-	5.1	3.39, 3.20	-	-	-	-	-	171.76	57.01	65.21	-	-
47	Cys	9.1	127.58	-	-	3.44	2.31, 3.09	-	-	-	-	-	172.37	57	31.2	-	-
48	Pro	-	-	-	-	4.19	2.37, 1.79	1.81, 1.92	-	-	-	-	178.49	65.39	32.31	28.1	-
49	His	8.17	118.54	-	-	4.14	3.15, 3.13	-	-	-	-	-	176.68	60.88	30.67	-	-
50	Cys	8.29	123.02	-	-	3.77	2.55, 3.02	-	-	-	-	-	178.03	64.8	28.28	-	-
51	Glu	8.26	117.12	-	-	3.86	1.86, 1.87	2.12, 1.89		-	-	-	177.53	59.22	30.12	36.63	-
52	Lys	7.51	117.38	-	-	4.1	1.82, 1.79	1.46, 1.37	1.58	2.94, 3.00	-	-	177.21	58.37	33.22	24.98	29.22
53	Glu	8.1	115.75	-	-	4.24	2.06, 1.60	2.18, 1.95	-	-	-	-	176.42	56.53	30.37	35.9	-
54	Gly	7.62	107.95	-	-	3.77, 4.19	-	-	-	-	-	-	173.52	44.92	-	-	-
55	Ile	7.78	117.48	-	-	4.26	1.87	1.30, 1.13, 0.85	0.81	-	-	-	175	61.73	38.89	26.41, 17.77	13.41
56	Gln	8	121.84	-	110.96	4.39	1.84, 2.28	2.29	-	6.83, 7.52	-	-	175.12	56.19	29.96	34.16	179.29
57	Trp	8.15	119.99	-	128.66	4.64	3.27, 3.13	-	7.13	7.47, 9.94	7.18, 7.03	6.84	174.12	56.67	29.53		-
58	Glu	8.06	119.33	-	-	4.18	1.79, 1.93	2.15, 2.16	-	-	-	-	175.19	56.17	30.9	36.29	-
59	Ala	8.3	124.37	-	-	4.3	1.35	-	-	-	-	-	176	52.25	19.17	-	-
60	Lys	8.38	120.8	-	-	4.24	1.79, 1.73	1.41, 1.40	-	-	-	-	- -	-	-	-	-
61	Glu	8.27	120.71	-	-	4.29	1.91, 2.08	2.24, 2.24	-	-	-	-	175.29	56.49	30.59	36.37	-
62	Asp	8.44	120.83	-	-	4.61	2.71, 2.60	-	-	-	-	-	175.18	54.37	41.44	-	-
63	Asn	8.47	119.13	111.9	-	4.62	2.85, 2.79	-	7.59, 6.91	-	-	-	174.71	53.43	39.12	-	-
64	Ser	8.4	115.67	-	-	4.42	3.91, 3.88	-	-	-	-	-	174.02	58.97	63.86	-	-
65	Glu	8.46	121.56	-	-	4.32	1.95, 2.09	2.26	-	-	-	-	176.2	56.94	30.42	36.42	-
66	Gly	8.35	108.73	-	-	3.96, 3.96		-	-	-	-	-	173.21	45.36	-	-	-
67	Glu	8.21	119.48	-	-	4.31	2.02, 2.04	2.22, 2.20	-	-	-	-	175.61	56.43	30.86	36.36	-
68	Glu	8.4	121.18	-	-	4.27	2.03, 1.89	2.24, 2.24	-	-	-	-	175.28	56.55	30.67	36.38	-

		H	N	N^{δ}	N^{ε}	H^{α}	H^{β}	H^{γ}	H^{δ}	H^{ε}	H^{ζ}	H^{η}	C	C^{α}	C^{β}	C^{γ}	C^{δ}	C^{ε}
69	Ile	8.29	122.78	-	-	4.14	1.84	1.47, 1.17, 0.87	0.85	-	-	-	175.21	60.84	38.57	17.58, 27.27	12.75	-
70	Leu	8.38	126.76	-	-	4.41	1.58, 1.63	1.56	0.93, 0.86	-	-	-	176.22	54.93	42.58	27.12	23.7, 24.78	-
71	Glu	8.34	121.57	-	-	4.27	1.90, 1.99	2.22	-	-	-	-	175.35	56.42	30.78	36.39	-	-
72	Glu	8.44	121.92	-	-	4.3	2.03, 1.92	2.22	-	-	-	-	- -	56.46	30.62	36.47	-	-
73	Val	8.27	120.68	-	-	4.16	2.11, 0.94	0.95	-	-	-	-	175.96	62.4	32.97	21.27, 20.47	-	-
74	Gly	8.49	111.43	-	-	3.97, 3.98	-	-	-	-	-	-	173.61	45.5	-	-	-	-
75	Gly	8.21	107.77	-	-	3.93, 3.99	-	-	-	-	-	-	172.73	44.91	-	-	-	-
76	Asp	8.36	120.68	-	-	4.89	2.48, 2.72	-	-	-	-	-	- -	52.6	41.05	-	-	-
77	Pro	-	-	-	-	4.42	2.30, 1.94	2.02, 2.02	3.86, 3.73	-	-	-	176.4	63.4	32.28	27.34	50.79	-
78	Glu	8.51	119.67	-	-	4.23	1.91, 2.04	2.29, 2.23	-	-	-	-	175.94	56.51	33.37	36.42	-	-
79	Glu	8.49	121.59	-	-	4.28	1.91, 1.99	2.23, 2.23	-	-	-	-	- -	56.51	30.59	36.42	-	-
81	Asp	-	-	-	-	4.56	2.59, 2.64	-	-	-	-	-	- -	-	-	-	-	-
82	Asp	8.29	119.83	-	-	4.55	2.65, 2.60	-	-	-	-	-	- -	54.28	41.3	-	-	-
83	His	8.19	118.21	-	-	4.61	3.09, 3.16	-	-	-	-	-	- -	56	30.16	-	-	-
84	His	-	-	-	-	-	-	-	-	-	-	-	176.5	57.29	30.26	-	-	-
85	Met	8.56	118.96	-	-	4.23	1.81, 1.86	2.54	-	-	-	-	175.94	56.06	33.45	31.85	-	-
86	Glu	8.7	120.52	-	-	4.12	1.99, 1.78	2.06, 2.16	-	-	-	-	173.26	56.25	29.76	36.41		-
87	Phe	6.94	114.42	-	-	4.47	2.46, 2.70	-	6.93	7.26	-	-	171.28	54.57	41.37	-	-	-
88	Cys	8.25	120.08	-	-	3.89	1.86, 3.33	-	-	-	-	-	176.86	57.62	31.55	-	-	-
89	Arg	8.61	127.58	-	-	3.92	1.62, 1.46	1.70, 1.63	3.1	4, 3.25	-	-	175.97	58.58	30.99	27.45	43.67	-
90	Val	8.86	119.34	-	-	3.99	2.28	1.30, 1.12	-	-	-	-	176.31	65.52	33.29	20.70, 21.89	-	-
91	Cys	8.08	116.69	-	-	4.91	3.38, 2.98	-	-	-	-	-	175.44	59.4	31.5	-	-	-
92	Lys	8.1	115.7	-	-	4.16	2.19, 1.82	-	-	-	-	-	173.64	58.29	28.92	-	-	-
93	Asp	8.57	120.49	-	-	5.28	3.07, 2.98	-	-	-	-	-	176.46	53.32	43.48	-	-	-
94	Gly	8.55	106.78	-	-	3.87, 4.49	-	-	-	-	-	-	173.5	44.01	-	-	-	-
95	Gly	8.6	109.51	-	-	4.53, 3.65	-	-	-	-	-	-	171.77	44.8	-	-	-	-
96	Glu	8.75	123.87	-	-	4.19	2.07, 2.02	2.19, 2.19	-	-	-	-	174.25	56.68	29.26	36.47	-	-
97	Leu	7.98	123.54	-	-	4.35	1.45,0.90	1.22	0.43, -0.15	-	-	-	174.67	53.64	44.41	25.75	22.43, 25.56	-
98	Leu	8.99	124	-	-	4.35	0.80, 1.26	0.66	0.39, -0.59	-	-	-	174.69	53.82	43.48	26.16	23.32, 25.54	-
99	Cys	8.49	123.33	-	-	4.95	2.43, 2.25	-	-	-	-	-	173.33	58.08	28.83	-	-	-
100	Cys	8.63	125.99	-	-	4.65	3.82, 2.73	-	-	-	-	-	175.59	61.12	30.78	-	-	-
101	Asp	9.06	126.58	-	-	4.85	2.71, 2.79	-	-	-	-	-	177.46	57.91	43.01	-	-	-
102	Thr	9.69	110.7	-	-	4.62	4.6	1.04	-	-	-	-	172.83	62.12, 58.44	69.99, 30.98	21.95	-	-
103	Cys	8.2	125.53	-	-	5.02	3.23, 2.69	-	-	-	-	-	- -	58.44	30.98	-	-	-
104	Pro	-	-	-	-	4.47	2.13, 1.90	-	3.47, 3.68	-	-	-	175.87	63.61	31.58	27.54	-	-
105	Ser	9.12	117.99	-	-	4.39	3.81, 4.24	-	-	-	-	-	171.6	61.38	65.16	-	-	-
106	Ser	7.82	116.68	-	-	5.73	3.36, 3.53	-	-	-	-	-	171.38	56.83	66.28	-	-	-
107	Tyr	9.09	118.01	-	-	5.53	2.97, 2.37	-	7.11	6.87	-	-	175.66	57.08	44.39	-	-	-
108	His	8.61	118.43	-	-	4.85	3.28, 2.65	-	7.45	7.39	-	-	178.22	57.68	31.19	-	-	-
109	Ile	9.2	116.8	-	-	3.85	2.08	1.46, 0.79, 1.29	0.65	-	-	-	176.1	65.94	36.44	19.16	13.26	-
110	His	6.98	111.22	-	-	4.78	3.05, 3.41	-	-	-	-	-	175.55	56.62	30.16	-	-	-
111	Cys	7.35	120.22	-	-	4.43	3.26, 2.84	-	-	-	-	-	174.37	61.36	30.41	-	-	-
112	Leu	6.47	119.23	-	-	4.16	1.45, 1.70	0.8	0.82, 0.53	-	-	-	173.93	54.74	44.44	26.92	27.07, 23.16	6,27.07
113	Asn	8.28	115.54	110.95	-	4.98	2.61, 2.89	-	6.87, 7.56	-	-	-	172.22	49.74	41.03	176.92	-	-
114	Pro	-	-	-	-	5.07	2.45, 2.07	1.94	3.68, 3.79	-	-	-	-	62.12	33	25.07	50.22	-
115	Pro	-	-	-	-	4.43	2.29, 1.95	2.03	3.87, 3.77	-	-	-	176.89	62.91	32.38	27.41	50.36	-
116	Leu	7.55	120.34	-	-	4.82	1.53, 1.44	1.74	0.69, 0.92	-	-	-	175.76	51.76	42.98	-	25.64, 22.92	-
117	Pro	-	-	-	-	4.38	2.25, 2.01	-	3.78, 3.88	-	-	-	175.58	63.74	32.19	19	50.25	-
118	Glu	6.97	111.87	-	-	4.39	1.82, 2.02	-	-	-	-	-	173.84	53.87	32.62	34.97	-	-
119	Ile	8.64	122.95	-	-	3.93	1.78	1.59, 1.07, 0.9		0.78	-	-	173.54	59.76	38.13	16.77	12.72	-
120	Pro	-	-	-	-	4.47	1.97, 2.32	1.86, 1.81	3.41, 3.67	-	-	-	175.24	62.9	32.35	27.08	51.28	-
121	Asn	8.52	118.03	111.73	-	4.78	2.78, 2.76	-	6.92, 7.59	-	-	-	174.81	53.1	39.41	-	-	-
122	Gly	8.07	109.07	-	-	3.96, 4.02	-	-	-	-	-	-	- -	44.84	-	-	-	-
123	Glu	8.36	120.48	-	-	4.35	1.97, 1.89	2.22	-	-	-	-	175.21	57.4	30.66	36.72	-	-
124	Trp	9.07	125.56	-	128.19	4.52	2.96, 3.22	-	6.87	9.81, 7.29	7.12, 6.69	6.53	171.61	59.98	30.53	-	-	-
125	Leu	6.8	125.89	-	-	4.89	1.15, 0.88	1.57	0.81, 0.74	-	-	-	174.41	52.23	44.32	27.59	24.16	25.27
126	Cys	9.08	125.71	-	-	3.48	2.44, 3.28	-	-	-	-	-	- -	57.09	29.8	-	-	-
127	Pro	-	-	-	-	4.17	2.39, 1.82	-	-	-	-	-	177.86	65.3	32.36	-	-	-
128	Arg	8.33	117.04	-	-	4.14	1.89, 1.63	1.77	-	-	-	-	177.43	58.34	30.9	28.11	43.53	-
129	Cys	8	120.92	-	-	4	2.62, 2.82	-	-	-	-	-	175.66	63.14	30.05	-	-	-
130	Thr	7.62	110.34	-	-	3.99	4.04	0.99	-	-	-	-	173.28	62.96	69.45	-	-	-
131	Cys	7.55	121.5	-	-	4.57	2.76, 2.91	-	-	-	-	-	-	-	27.6	-	-	-
132	Pro	-	-	-	-	-	-	-	-	-	-	-	175.86	62.87	32.14	27.27	-	-
133	Ala	8.22	122.34	-	-	4.26	1.38	-	-	-	-	-	-	52.67	19.24	-	-	-
134	Leu	8	118.88	-	-	4.27	1.60,1.67	-	0.9	-	-	-	-	-	-	-	-	-
135	Lys	8.05	119.99	-	-	4.23	-	-	-	-	-	-	-	-	-	-	-	-

Appendix 2

Isothermal Titration Calorimetry

K_D values for peptide binding to the PHD of PHF8 were also measured using ITC (Section 8.7.6).[392] The advantage of ITC over FP is that modifications such as FITC (which may affect the binding of the peptide to the target protein) do not need to be made to either binding partner to monitor the binding. Instead the calorimeter measures the change in temperature that occurs on addition of one binding partner to the other due to the binding event. Peptide is titrated into a cell containing the protein of interest, allowing a K_D to be calculated. In order to avoid contributions to the temperature change from mixing of different buffers, binding components are dialysed into the same buffer prior to the experiment. ITC was attempted with PHF8 to provide a complementary method of measuring histone peptide affinities and as an assay that could potentially also be extended to measure the K_D values of PHF8 inhibitors.

Firstly FL PHF8 was incubated with 200 μM EDTA to remove any metal bound in the active site that had co-purified with the enzyme during purification. FL PHF8 and the PHD finger of PHF8 were then dialysed against buffer containing 100 mM HEPES, pH 7 and 500 mM NaCl. The H3 K4Me$_3$K9Me$_2$ 15mer peptide was dissolved in the same buffer as used for dialysis. An initial titration of peptide (500 μM) into PHD finger (25 μM) was carried out at 25 °C (Figure 9.1). A curve was plotted to allow measurement of a K_D for the histone peptide (4.2 μM), which was comparable to the value of 2 μM determined by the FP assay (Table 4.1).

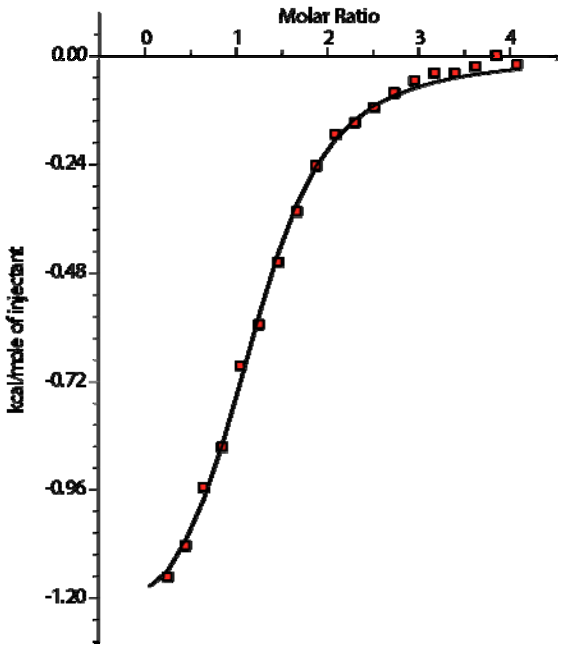


Figure 9.1 ITC experiment titrating H3 K4Me$_3$K9Me$_2$ 15mer (500 µM) into the PHD only construct of PHF8 (25 µM).

ITC was then attempted with FL PHF8. Initially, measurement of a K_D for metal binding was undertaken by titrating Mn(II) (750 µM) into FL PHF8 (50 µM). Manganese was used in place of Fe(II) due to the propensity of Fe(II) to oxidise in solution. Unfortunately under all conditions attempted the addition of Mn(II) resulted in precipitation of the protein over the timecourse of the experiment, precluding K_D determination. In subsequent experiments, FL PHF8 was dialysed against fresh buffer containing manganese to ensure that the protein was fully metal bound. Titration of peptide (500 µM) into FL PHF8 (50 µM) was carried out. Unfortunately, despite a number of attempts, reliable data were not obtained, precluding measurement of a K_D by ITC. Similar difficulties were encountered for ITC experiments with JMJD2A within the laboratory (*Dr Nathan Rose, unpublished data*).

Appendix 3

Crystallography of PHF8

Initially the PHF8 protein was purified to homogeneity using a His-trap column, tag-cleavage with TEV, rebinding to a His-trap column and then size exclusion chromatography. Fractions 13-15 from the size exclusion column were pooled and concentrated to 14 mg.mL^{-1} before flash freezing in liquid nitrogen and storage in aliquots at -80 °C.

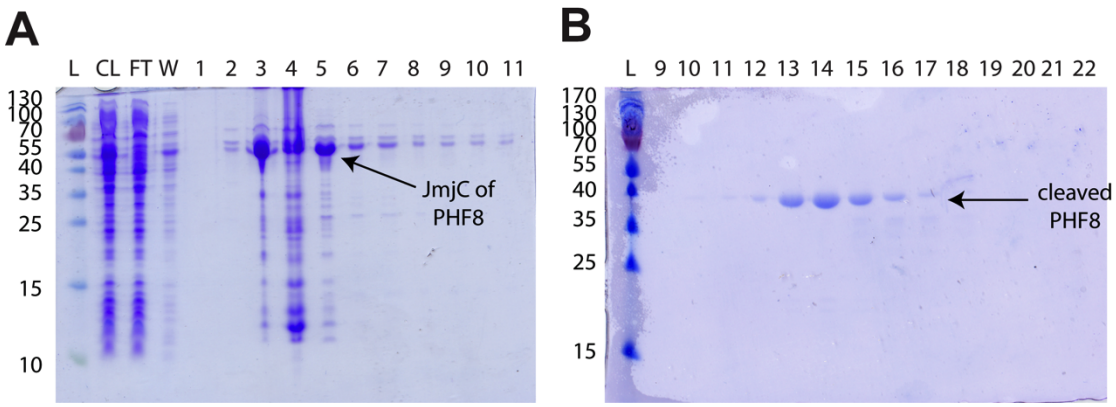


Figure 9.2 Purification of the JmjC domain of PHF8 using **A)** a His-trap column followed by **B)** an S200 300 mL size exclusion column. 5 mL fractions are labelled, L = molecular weight ladder, CL = cell lysate, FT = flow through and W = wash.

Several 96-well crystal trays were seeded containing PHF8 (11 mg.mL^{-1}) and a variety of cofactors / inhibitors (1 – 4, Table 9.1). These were left to incubate at either 4 °C or 20 °C. Conditions are given in Table 9.1. Unfortunately no crystals were observed from these crystal trays. A second batch of PHF8 was purified and crystal trays seeded directly after purification by size exclusion chromatography, without prior freezing (5 – 9, Table 9.1). Once again no crystals were observed. Using this same batch of PHF8, three more crystal trays were seeded (10 – 12, Table 9.1). This time, small crystals were observed in several conditions from trays 10 and 11 after 1 week. However, although a protein diffraction pattern to around 3 Å was observed

when the crystals were screened using the in-house diffractometer, the diffraction pattern was found to be extremely mosaic. Using the same batch of purified protein, more crystal trays were seeded with and without Daminozide (13 – 16, Table 9.1). The best crystals were obtained in a condition containing $LiSO_4$ and PEG400, so an optimisation screen was carried out around these conditions. Two trays were seeded using the optimisation screen (17 and 18, Table 9.1). Unfortunately, no crystals were observed.

Table 9.1 Crystallisation conditions. Plates in which crystals were obtained are shown in red. Mn = manganese, NOG = *N*-oxalylglycine, IEX = ion exchange chromatography.

Plate ID	[PHF8] /mg.mL^{-1}	Metal	Inhibitor	Screen	Incubation temperature /°C	Extra purification	Crystals obtained?
1	10.2	-	-	JCSG	4	-	No
2	10.2	-	-	JCSG	20	-	No
3	10.2	-	2 mM daminozide	JCSG	20	-	No
4	10.2	-	4 mM daminozide	JCSG	20	-	No
5	11	-	2 mM NOG	JCSG	4	-	No
6	11	2 mM Mn	2 mM NOG	JCSG	4	-	No
7	11	2 mM Mn	2 mM daminozide	JCSG	4	-	No
8	11	2 mM Mn	4 mM daminozide	JCSG	4	-	No
9	11	-	2 mM daminozide	JCSG	4	-	No
10	11	-	2 mM NOG	JCSG	4	-	Yes
11	11	-	2 mM NOG	JCSG	20	-	Yes
12	11	-	2 mM NOG	Salt RX	4	-	No
13	11	-	-	JCSG	4	-	No
14	11	-	-	JCSG	20	-	No
15	11	-	2 mM daminozide	JCSG	4	-	Yes
16	11	-	2 mM daminozide	JCSG	20	-	No
17	11	-	2 mM NOG	Optimisation screen	4	-	No
18	11	-	2 mM NOG	Optimisation screen	20	-	No
19	10	-	2 mM NOG	JCSG	4	IEX	No
20	10	-	2 mM NOG	Salt RX	4	IEX	No

The purification protocol used to produce one of the published crystal structures included a final ion exchange chromatography step.[297] To see whether this would help with crystallisation, a third batch of PHF8 was prepared and ion exchange chromatography carried out after the size exclusion chromatography step. The protein eluted in two main peaks, which were pooled separately and concentrated. Protein from the second peak was used to seed two more crystal trays (19 and 20, Table 9.1). Once again no crystals were obtained.

Crystallography of JMJD2A with methylated arginine peptides

Initial crystal trays were seeded using JMJD2A (11 mg.mL^{-1}, *purified by Dr Oliver King*) with NOG (2 mM), Ni (1 mM) and either a H3 R2Me$_{2a}$ 15mer peptide (1 mM) or a H3 K9RMe$_{2a}$ 15mer peptide (1 mM) (1-4, Table 9.2). Two trays were also seeded without Ni or NOG (5 and 6, Table 9.2). After 1 week, crystals were observed in all plates using the PACT Premier broad screen (1, 3 and 5, Table 9.2), although considerably more crystals were observed in the tray containing the H3 K9RMe$_{2a}$ 15mer peptide and Ni and NOG (plate 1). Screening of some crystals using the in-house diffractometer gave good diffraction. Unfortunately, although on initial refinement some peptide density was observed in the active site, further refinement showed that the peptide occupancy was too low to solve the structure (*refinement carried out by Dr Rasheduzzan Chowdhury*). Further trays were seeded containing a higher ratio of peptide:protein (2 mM peptide, 9 mg.mL^{-1} JMJD2A) in an attempt to increase the peptide occupancy, but in this case no crystals were obtained (7 and 8, Table 9.2).

Table 9.2 Crystallisation conditions used to obtain a co-crystal structure between JMJD2A and an arginine peptide. Trays in which crystals were observed are shown in red. H3 R9 denotes the H3 K9RMe$_{2a}$ 15mer histone peptide and H3 R2 denotes the H3 R2Me$_{2a}$ 15mer histone peptide.

Plate ID	[JMJD2A] /mg.mL^{-1}	Metal	Inhibitor	Peptide	Screen	Crystals obtained?
1	11	1 mM Ni	2 mM NOG	1 mM H3 R9	PACT Premier	Yes
2	11	1 mM Ni	2 mM NOG	1 mM H3 R9	JCSG	No
3	11	1 mM Ni	2 mM NOG	1 mM H3 R2	PACT Premier	Yes
4	11	1 mM Ni	2 mM NOG	1 mM H3 R2	JCSG	No
5	11	-	-	1 mM H3 R9	PACT Premier	Yes
6	11	-	-	1 mM H3 R9	JCSG	No
7	9	1 mM Ni	2 mM NOG	2 mM H3 R2	PACT Premier	No
8	9	1 mM Ni	2 mM NOG	2 mM H3 R9	PACT Premier	No
9	11	-	-	2 mM 21 mer H3 R9	PACT Premier	No
10	11	-	-	2 mM 8 mer H3 R9	PACT Premier	No

One possibility for the difficulty in obtaining a co-crystal structure was that the peptide was not of an optimal length, so that there were either residues that were not

binding to the protein and instead interfering with crystal packing or insufficient residues to form a strong enough interaction with the protein. As most crystals were obtained when JMJD2A was incubated with the H3 K9RMe$_{2a}$ 15mer co-crystallisation efforts were focussed on this sequence. To investigate either of these possibilities, an 8 residue H3 K9RMe$_{2a}$ peptide (ARRMe$_{2a}$STGGK) and a 21 residue H3 K9RMe$_{2a}$ peptide (ARTKQTARRMe$_{2a}$STGGKAPRKQLA) were synthesised by SPPS and purified by HPLC. Trays containing these peptides were seeded (9 and 10, Table 9.2). Unfortunately, however, no crystals containing peptide were obtained with either peptide.

Previous work in the Schofield laboratory has shown that substrate-enzyme co-crystals can be obtained by linking the peptide by a disulphide bond to *N*-oxalyl-*D*-cysteine (NOC), which can bind in the 2OG-binding pocket (Figure 9.3).[393] To do this, a cysteinyl-residue must be substituted into the substrate peptide. With JMJD2A it was shown that substitution of H3 T11 for a cysteinyl-residue in the native histone peptide sequence allowed binding of the peptide-NOC disulphide in the active site in a conformation very similar to that of the native peptide alone. A peptide was synthesised with K9 substituted for asymmetric dimethylarginine and T11 substituted for a cysteine (ARRMe$_{2a}$SCGGK). Two peptides based around H3 R2 methylation with T3 or K4 substituted for a cysteine were also synthesised (ARMe$_{2a}$CKQTARKSTGGKA and ARMe$_{2a}$TCQTARKSTGGKA, *experiments carried out in collaboration with Matthew Smith*).

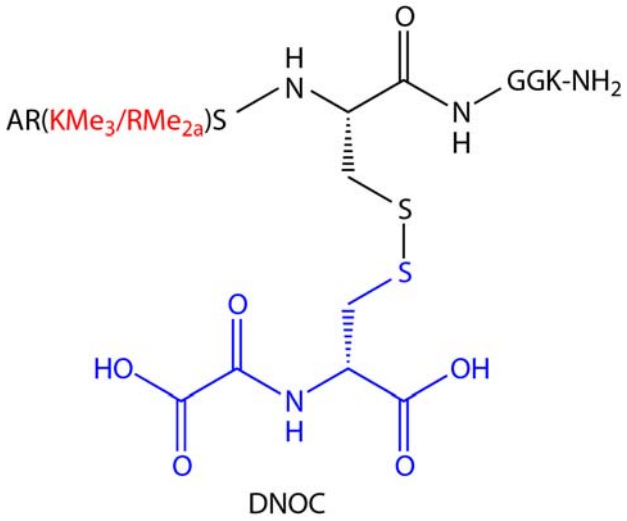


Figure 9.3 Structure of H3 K9R peptide-DNOC disulphide. DNOC is shown in blue, the active site target residue is shown in red. KMe_3 in the published work, RMe_{2a} here.

JMJD2A (11 $mg.mL^{-1}$) was incubated with DNOC (10 mM) and peptide (5 mM) and crystal trays seeded and left to grow at 4 °C. Unfortunately although once again crystals were obtained, no occupancy of the disulphide peptide could be observed.

Appendix 4

Publications arising from work described in this thesis

10) UTY is an N^{ϵ} histone 3 lysine-27 demethylase
Walport LJ, Hopkinson RJ, Vollmar M, Madden SK, Gileadi C, Oppermann U, Schofield CJ, Johansson C
J Biol Chem, accepted

9) Total Synthesis of the Antitumor Antibiotic (±)-Streptonigrin: First and Second Generation Routes for Efficient de-novo Pyridine Formation using Ring Closing Metathesis
Donohoe T, Jones C, Kornahrens A, Barbosa LC, Walport L, Tatton M, O'Hagan M, Rathi A, Baker D
J Org Chem, **78**, 12338-12350 (2013).

8) Identification of the KDM2/7 Histone Lysine Demethylase Subfamily Inhibitor and its Antiproliferative Activity
Suzuki T, Ozasa H, Itoh Y, Zhan P, Sawada H, Mino K, Walport L, Ohkubo R, Kawamura A, Yonezawa M, Tsukada Y, Tumber A, Nakagawa H, Hasegawa M, Sasaki R, Mizukami T, Schofield CJ, Miyata N
J Med Chem. **56,** 7222-7231 (2013)

7) The Ugi four-component reaction enables expedient synthesis and comparison of photoaffinity probes
Bush JR, Walport LJ, McGouran JF, Leung IKH, Berridge G, van Berkel SS, Basak A, Kessler BM, Schofield CJ
Chem. Sci., **4**, 4115-4120 (2013)

6) 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation
Hopkinson RJ, Tumber A, Yapp C, Chowdhury R, Aik W, Che KH, Li XS, Kristensen JBL, King ONF, Chan MC, Yeoh KK, Choi H, Walport LJ, Thinnes CC, Bush JT, Lejeune C, Rydzik AM, Rose NR, Bagg EA, McDonough MA, Krojer TJ, Yue WW, Ng SS, Olsen L, Brennan PE, Oppermann U, Müller S, Klose RJ, Ratcliffe PJ, Schofield CJ, Kawamura A
Chem. Sci., **4**, 3110-3117 (2013)

5) Is JmjC Oxygenase Catalysis Limited to Demethylation?
Hopkinson RJ, Walport LJ, Münzel M, Rose NR, Smart TJ, Kawamura A, Claridge TD, Schofield CJ
Angew Chem Int Ed Engl. **52**, 7709-7713 (2013)

4) Mechanisms of human histone and nucleic acid demethylases
Walport LJ, Hopkinson RJ, Schofield CJ
Curr Opin Chem Biol. **16**, 525-34 (2012)

3) Plant growth regulator daminozide is a selective inhibitor of human KDM2/7 histone demethylases
Rose NR[‡], Woon EC[‡], Tumber A[‡], Walport LJ, Chowdhury R, Li XS, King ON, Lejeune C, Ng SS, Krojer T, Chan MC, Rydzik AM, Hopkinson RJ, Che KH, Daniel M, Strain-Damerell C, Gileadi C, Kochan G, Leung IK, Dunford J, Yeoh KK, Ratcliffe PJ, Burgess-Brown N, von Delft F, Muller S, Marsden B, Brennan PE, McDonough MA, Oppermann U, Klose RJ, Schofield CJ, Kawamura A
J Med Chem. **55**, 6639-43 (2012)

2) Dynamic combinatorial mass spectrometry leads to inhibitors of a 2-oxoglutarate-dependent nucleic acid demethylase
Woon EC, Demetriades M, Bagg EA, Aik W, Krylova SM, Ma JH, Chan M, Walport LJ, Wegman DW, Dack KN, McDonough MA, Krylov SN, Schofield CJ.
J Med Chem. **55**, 2173-84 (2012)

1) Linking of 2-oxoglutarate and substrate binding sites enables potent and highly selective inhibition of JmjC histone demethylases
Woon EC, Tumber A, Kawamura A, Hillringhaus L, Ge W, Rose NR, Ma JH, Chan MC, Walport LJ, Che KH, Ng SS, Marsden BD, Oppermann U, McDonough MA, Schofield CJ.
Angew Chem Int Ed Engl. **51**, 1631-4 (2012)