Development, Synthesis and Characterisation of Chemical Probes for PHD Fingers







Oliver Douglas Coleman

Merton College, University of Oxford

Michaelmas Term 2018

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

Abstract

Development, Synthesis and Characterisation of Chemical Probes for PHD Fingers

Oliver D. Coleman **DPhil Thesis (Synthesis for Biology and Medicine CDT)**
Merton College, University of Oxford **Michaelmas Term, 2018**

The KDM7 proteins are a sub-family of histone lysine demethylases (KDMs) which play an important role in regulating the post-translational modification (PTM) state of histones. Histone proteins are responsible for packaging of DNA, forming a vital element of epigenetic regulation of gene transcription. Therefore, the KDMs play crucial biological roles and have been linked to a variety of important health and disease states, including those as part of neurological, immunological and oncological processes. This makes it important to develop tools to aid study of their roles and improve our understanding and control of fundamental biology. The human KDM7 sub-family (A, B, C) are each composed of an epigenetic reader domain (plant homeodomain finger, [PHD]) and a catalytic demethylase domain (Jumonji-C, [JmjC]). To date, chemical probe development for the KDM7s has encountered challenges when targeting their well-conserved JmjC domains. PHD fingers are an important class of epigenetic reader domain, but no potent or selective probes for them are reported.

This thesis describes work towards the development of potent and selective chemical probes for the human KDM7 sub-family and its PHD finger. An mRNA display based approach was used to screen a randomised library of ~10^{12} cyclic peptides and enrich them against KDM7B. Candidate cyclic peptides were assessed for binding affinity and domain selectivity with biophysical techniques: BioLayer Interferometry (BLI), Isothermal Titration Calorimetry (ITC) and hydrogen-deuterium exchange mass spectrometry (HDxMS). Three cyclic peptide hits were identified, one (OC9) with nanomolar affinity for the KDM7 PHD finger and two others (OC3, OC6) with nanomolar affinity for the KDM7B combined PHD and JmjC domains. In-vitro functional assays, primarily by MALDI-TOF-MS and AlphaScreen, were used to characterise the impact of cyclic peptide binding on displacement of cognate histone peptide ligand H3K4me3 and consequences on JmjC demethylase activity. Similar assays were used to profile selectivity against related reader domains and KDM proteins. OC9 was identified to be a potent and selective inhibitor of the KDM7 PHD finger, able to displace H3K4me3, which resulted in enhancement of KDM7A, and inhibition of KDM7B, demethylation against the H3K9me2 mark. OC3 was found to be an effective inhibitor of KDM7B, whilst OC6 was found to display broad-KDM inhibition.

The binding behaviour of OC9 was further characterised by protein NMR and intrinsic tryptophan fluorescence to reveal effects on a key tryptophan residue in the KDM7 PHD finger, located within the aromatic cage essential for binding of cognate H3K4me3. Structure activity relationships for OC9 were explored by alanine scanning and built upon by targeted substitutions to optimise OC9 towards increased compatibility with membrane environments, supported by computational techniques. Peptide NMR studies revealed OC9 existed as an interconverting mixture of isomers on the NMR timescale. A click-chemistry compatible derivative of OC9 was synthesised and used to create a biotinylated version of OC9. This affinity probe successfully pulled-down overexpressed full-length KDM7B, via its PHD finger, from nuclear cell lysate.

In summary, the work described herein has identified a cyclic peptide as the first potent and selective binding ligand of both a PHD finger and the human KDM7 sub-family, and demonstrated its suitability for use as a chemical probe to study KDM7 biology.

Table of Contents

Abstract i
Table of Contents ii
Acknowledgements ix
Abbreviations xi
List of Contributions from Others xiii
Chapter 0. Introduction 1
0.1. Working Definition of Epigenetics 1
0.2. The Histone Machinery 2
0.3. The Histone Lysine Demethylases (KDMs) 5
0.4. The KDM7 Sub-Family 9
0.5. Plant Homeodomain (PHD) Fingers 12
0.6. Chemical Tools for KDMs and PHD Fingers 16
0.7. Cyclic Peptides and Selection Platforms 21
0.8. Project Objectives 29
Chapter 1. Cyclisation and Lysine PTMs in Peptide Libraries 30
1.1. Introduction 30
1.2. Making Cyclic Peptide Libraries 34
1.3. Synthesis of Modified Lysines for tRNA Acylation 37
1.4. Incorporating Modified Lysines into Peptides 41
1.5. Conclusions and Future Work 45
Chapter 2. Cyclic Peptide Selection against KDM7B 46
2.1. Introduction 46
2.2. Design of the mRNA-Display Selection 48
2.3. KDM7B Selection Results 52
2.4. Solid-Phase Synthesis of Cyclic Peptides 55
2.5. Validation of Cyclic Peptide Binding 57
2.5.1. BioLayer Interferometry 57
2.5.2. Isothermal Titration Calorimetry 66
2.6. Conclusions and Future Work 70
Chapter 3. Cyclic Peptide Function vs KDM7s 72
3.1. Introduction 72
3.2. Cyclic Peptide Displacement vs KDM7$_{[PHD+JmjC]}$ 74
3.2.1. Displacement of H3K4me3 Peptide 74

3.2.2. Displacement of Methylstat 77
3.3. Cyclic Peptide Displacement vs KDM7$_{[PHD]}$ 78
3.4. Cyclic Peptide Effects on KDM7 Activity vs Histone Peptides 80
3.4.1. Cyclic Peptides vs KDM7B Demethylase Activity 82
3.4.2. Cyclic Peptides vs KDM7A Demethylase Activity 87
3.5. Summary of Peptide Displacement and Activity Assay Results 91
3.6. Cyclic Peptide Effects on KDM7 Activity vs Histone Extracts 94
3.7. Conclusions and Future Work 98
Chapter 4. Cyclic Peptide Selectivity Studies 100
4.1. Introduction 100
4.2. Binding vs Non-KDM7 Reader Domains 101
4.3. Displacement vs Non-KDM7 Reader Domains 106
4.4. Effect on Non-KDM7 Enzyme Activity 109
4.5. Conclusions and Future Work 116
Chapter 5. Peptide Interactions with KDM7B by HDxMS 118
5.1. Introduction 118
5.2. HDxMS on KDM7B$_{[PHD+JmjC]}$ 120
5.3. HDxMS on KDM7B$_{[PHD]}$ 126
5.4. Conclusions and Future Work 130
Chapter 6. Studies on the Interaction of OC9 with KDM7$_{[PHD]}$ 131
6.1. Introduction 131
6.2. KDM7B$_{[PHD]}$ Solution NMR 132
6.3. KDM7$_{[PHD]}$ Tryptophan Intrinsic Fluorescence Assay 139
6.4. Conclusions and Future Work 142
Chapter 7. OC9 Structure-Activity Relationships (SAR) 143
7.1. Introduction 143
7.2. Alanine Scanning: Displacement and Functional Activity 145
7.3. Second Generation Variants: Residue Selection 149
7.4. Second Generation Variants: Binding, Displacement, Selectivity 156
7.5. OC9 Solution NMR Peptide Studies 167
7.6. Conclusions and Future Work 179
Chapter 8. OC9 vs KDM7s in a Cellular Context 182
8.1. Introduction 182
8.2. NanoBRET Cell Assay Preparation 185
8.3. Nuclear Localisation of N-terminal HaloTag-KDM7B 187
8.4. KDM7B NanoBRET Assay 188

8.5. Affinity Pull-Down of KDM7B from Nuclear Lysate 192
8.6. Conclusions and Future Work 199
Chapter 9. Project Summary 201
Chapter 10. Materials and Methods 207
10.1. Small Molecule Synthesis 207
10.1.1. Boc-Lys(Me2)-OH 209
10.1.2. Boc-Lys(Me2)-ODNB 210
10.1.3. H-Lys(Me2)-ODNB 211
10.1.4. Boc-Lys(Me3)-ODNB 212
10.1.5. H-Lys(Me3)-ODNB 214
10.1.6. Boc-Lys(Ac)-OH 215
10.1.7. Boc-Lys(Ac)-ODNB 216
10.1.8. H-Lys(Ac)-ODNB 217
10.1.9. Boc-Lys(Bn,Me)-OH 218
10.1.10. Boc-Lys(Boc,Me)-OH 219
10.1.11. Boc-Lys(Boc,Me)-ODNB 221
10.1.12. H-Lys(Me)-ODNB 222
10.1.13. Boc-Lys(Et2)-OH 223
10.1.14. Boc-Lys(Et2)-ODNB 224
10.1.15. H-Lys(Et2)-ODNB 226
10.1.16. Boc-Lys(Et3)-ODNB 227
10.1.17. H-Lys(Et3)-ONB 228
10.1.18. N-succinimidyl 2-chloroacetate 229
10.1.19. ClAc-^{L}Tyr-CME 230
10.2. List of Bacterial Plasmid Vectors 232
10.3. Polymerase Chain Reaction (PCR): General Protocol 233
10.4. Ligation Independent Cloning (LIC): Primers 234
10.5. Ligation Independent Cloning (LIC): General Protocol 234
10.6. Miniprep and Midiprep 236
10.7. Bacterial Cell Types 236
10.8. Bacterial Transformations: General Protocol 236
10.9. Protein Production Materials 237
10.9.1. Media 237
10.9.2. Antibiotics 237
10.9.3. Starter Cultures 237
10.10. SDS-PAGE Gel Analysis 238

10.11. Protein Production and Purification Protocols 238
10.11.1. General Protocol for His-Tagged Proteins 238
10.11.2. KDM7B$_{[PHD+JmjC]}$-Biotin 241
10.11.3. ^{13}C,^{15}N-Labelled KDM7B$_{[PHD]}$ 242
10.11.4. KDM7B$_{[PHD]}$ 245
10.11.5. KDM7B$_{[PHD+JmjC]}$ 245
10.11.6. KDM7A$_{[PHD+JmjC]}$ 245
10.11.7. DIDO1$_{[PHD]}$ 245
10.11.8. AIRE$_{[PHD1]}$ 246
10.12. Protein Mass Spectrometry 247
10.13. Protein Crystallography Trials 248
10.14. Preparation of Materials for IVTT and mRNA Display Selection 249
10.14.1. Acylation of tRNA/Microhelix 249
10.15. In-Vitro Transcription/Translation (IVTT) Reactions 250
10.16. Magnetic Bead Loading Optimisation for Selection 251
10.17. Procedure for mRNA Display Selection on KDM7B$_{[PHD+JmjC]}$ 252
10.17.1. mRNA Ligation 252
10.17.2. In-vitro Translation 252
10.17.3. Reverse Transcription (RT) 252
10.17.4. Magnetic Bead Preparation 253
10.17.5. Negative Selections (NS) 253
10.17.6. Loading of Magnetic Beads with the Target Protein 254
10.17.7. Positive Selection and DNA Recovery 254
10.17.8. Quantitative PCR (qPCR) 255
10.17.9. PCR Library Amplification 256
10.17.10. PCR and qPCR Primers 256
10.17.11. DNA Transcription to mRNA 256
10.17.12. Next Generation Sequencing and Data Analysis 257
10.18. DNA Agarose Gel Analysis 258
10.19. Phenol-Chloroform Purification of DNA/RNA 258
10.20. RNA Urea Acrylamide Gel Analysis 259
10.21. Solid Phase Peptide Synthesis (SPPS): General Protocol 260
10.21.1. Synthesis 260
10.21.2. Modification and Deprotection & Cleavage 260
10.21.3. Cyclisation 261
10.22. HPLC Peptide Purification: General Protocol 262

10.23. Histone Peptide Sourcing 263
10.24. MALDI-TOF-MS Spectra Acquisition 263
10.25. Cyclic Peptide High-Resolution ESI-MS (HRMS) 264
10.26. NMR Protocols 264
10.26.1. Calibration of Peptide Concentration 264
10.26.2. $KDM7B_{[PHD]}$ Protein Solution NMR 266
10.26.3. OC9 Solution NMR 267
10.27. BioLayer Interferometry (BLI): General Protocol 268
10.27.1. BLI: Association then Dissociation Binding Model 269
10.28. Tryptophan Intrinsic Fluorescence Assay 270
10.29. Isothermal Titration Calorimetry (ITC): General Protocol 271
10.29.1. ITC: Experiment Conditions 272
10.30. HDxMS Protocol 272
10.31. Displacement Assay (AlphaScreen): General Protocol 275
10.31.1. Linear Range Determination AlphaScreen Displacement Assay 275
10.31.2. Dose-Response AlphaScreen Displacement Assay 276
10.31.3. Table of AlphaScreen Displacement Assay Conditions 277
10.32. ECHO Dispensing of Compounds 277
10.33. General Compounds for Demethylase Activity Assays 278
10.34. Activity Assay (AlphaScreen): General Protocol 278
10.34.1. Table of AlphaScreen Activity Assay Conditions 279
10.35. Activity Assays (MALDI-TOF-MS): General Protocol 280
10.35.1. Single Point Concentration MALDI-TOF-MS Activity Assay 280
10.35.2. Dose-Response MALDI-TOF-MS Activity Assay 281
10.35.3. Table of MALDI Assay Conditions 282
10.36. Activity Assay (RapidFire): General Protocol 283
10.37. Histone Extract Demethylase Activity Assay: General Protocol 284
10.38. Affinity Pull-Down of KDM7B from HEK293T Nuclear Lysate 285
10.39. Western Blot Procedure 286
10.39.1. Western Blot Antibodies 287
10.40. Synthesis of OC9-Biotin 288
10.41. NanoBRET Experiments 290
10.41.1. Cellular Assay 290
10.41.2. Site-Directed Mutagenesis 292
10.42. Peptide Stability Testing 293
10.43. Physicochemical Property Testing 296

Chapter 11. Appendix....298
11.1. Human KDM7 Canonical Domain Sequences (UniProt)....298
11.1.1. KDM7A (Isoform 1, Q6ZMT4-1)....298
11.1.2. KDM7B (Isoform 2, Q9UPP1-2)....298
11.1.3. KDM7C (Isoform 1, O75151-1)....298
11.1.4. KDM7B (Isoform 1, Q9UPP1-1)....299
11.2. Protein Construct Information....299
11.2.1. $KDM7B_{[PHD+JmjC]}$-Biotin....301
11.2.2. $KDM7A_{[PHD]}$....301
11.2.3. $KDM7A_{[PHD+JmjC]}$....302
11.2.4. $KDM7B_{[PHD]}$....302
11.2.5. $KDM7B_{[JmjC]}$....303
11.2.6. $KDM7B_{[PHD+JmjC]}$....303
11.2.7. $KDM7C_{[PHD]}$....304
11.2.8. $KDM7C_{[PHD+JmjC]}$....304
11.2.9. $DIDO1_{[PHD]}$....304
11.2.10. $AIRE_{[PHD1]}$....305
11.2.11. $ING2_{[PHD]}$....305
11.2.12. $TAF3_{[PHD]}$....305
11.2.13. $SPIN1_{[TTD]}$....305
11.2.14. $KDM2A_{[Full-Length]}$....305
11.2.15. $KDM2B_{[Full-Length]}$....305
11.2.16. $KDM3A_{[JmjC+ZincFinger]}$....305
11.2.17. $KDM4A_{[JmjC+JmjN]}$....306
11.2.18. $KDM5C_{[JmjC+PHD+ZincFinger+JmjN+ARID]}$....306
11.2.19. $KDM6B_{[JmjC]}$....306
11.2.20. $EGLN1_{[Fe-2OG-Dioxygenase]}$....306
11.3. Protein Mass Spectrometry Data....306
11.4. List of Protein Contributors....307
11.5. Cyclic Peptide High-Resolution Mass Spectrometry....308
11.5.1. HRMS Summary Table....308
11.5.2. Example HRMS Spectra....309
11.6. Cyclic Peptide MALDI-TOF-MS....310
11.6.1. MALDI-TOF-MS Summary Table....310
11.6.2. MALDI-TOF-MS Spectra....311
11.7. IVTT MALDI-TOF-MS Spectra....319

11.8. Small Molecule ^{1}H and ^{13}C 1D-NMR Spectra 321
11.9. Crystallography Conditions Trialled for OC9 and KDM7s 327
11.10. ITC Control Data 327
11.11. HDxMS Deuterium Uptake Plots 328
11.11.1. $KDM7B_{[PHD+JmjC]}$ 328
11.11.2. $KDM7B_{[PHD]}$ 336
11.12. Example MALDI-TOF Spectra for OC9 vs KDM7 Activity Assays 338
11.13. MALDI-TOF Activity Assay Timecourses 339
11.14. RapidFire-MS Timecourse 340
11.15. OC3 and OC6 Sequence WebLogo Plots from NGS Data 340
11.16. List of Publications and Prizes (Presentations, Conferences) 341
Chapter 12. References 342

Acknowledgements

It has been thoroughly enjoyable to work under the supervision of Dr Akane Kawamura whose support, guidance and mentorship have been crucial. I have had opportunities to lead the project, to supervise other students and to learn new and varied techniques. It has been challenging, exciting and worthwhile, and I have felt able to improve as a scientist throughout. I am tremendously grateful to Akane. I am also very grateful to Prof Chris Schofield for his supervision, time and encouragement, giving me the freedom to follow the research in the most interesting directions as they developed.

The SBM CDT, with the EPSRC and associated Industrial Partners, have provided the platform and funding to pursue this DPhil adventure, with thanks to Prof Martin Smith and Prof Darren Dixon who gave me the opportunity to join the program. Lara, Maitane and Laura at the SBM CDT have been wonderful in their support, both personal and professional, over the last few years. It has also been brilliant to share this journey with my friends and colleagues Kilian & Emilie, Joe, Yao, James, Steve and others through the SBM CDT.

Within the Schofield and Kawamura labs, working in collaboration with Yijia Zhang has been immense in producing KDM7 proteins and in being someone with whom I could discuss the practical problems we faced. Dr Tom McAllister taught me how to operate as a better molecular biologist and has been a valued source of discussion throughout. Other friends and colleagues in the group have been great sources of personal support, laughs, digression and professional assistance, so thank you to Joanna, Martine, Rich, Illaria, Bhaskar, John-Paul, Roman, Ben, Louise, Tony, Becky, Leah, Gareth, Adam and the rest of the AK and CJS groups. Extra thanks to Akane, Chris, Tom, Martine, and Becky for proof-reading. I was also privileged to supervise two diligent and productive Masters students, Karen and Shane, with work from both of them contributing to this thesis. Thank you to all my friends outside of the Oxford bubble, Mark, George, Franky, Katie, and others for keeping me sane and playing so many excellent board-games.

Thank you to all of the support staff in the Wellcome Trust Centre for Human Genetics, the CRL and the NMR department, including Dr Nader Amin and Prof Tim Claridge, and the Mass Spectrometry department, including James Wickens and Victor. Thanks also to the SGC for

the donated resources and machine time. All of their collective contributions have been instrumental in helping me to properly characterise my work.

My industrial collaborators have made some highly appreciated contributions to this work, providing access to experiments that would not have been possible without them. From AstraZeneca, Dr Hitesh Sanganee has encouraged and pushed me to make progress, Dr Christopher Stubbs gave dedicated time to performing, analysing and explaining the hydrogen-deuterium exchange mass spectrometry experiments, and Dr William McCoull facilitated access to physicochemical property testing. From Vertex Pharmaceuticals, Sarah Skerratt enabled collaboration with Ewa Chudyk and Ron Knegtel who provided expert computational assistance for peptide conformation analysis.

I would also like to thank my college, Merton, for all of their support and Prof Veronique Gouverneur for the opportunity to teach at Merton, which has been an unexpected benefit to me, my research and hopefully my students!

Most importantly, I would be nowhere at all without the support of my family. It has been a challenging road to get to this point, with many visits to the hospital over the years prior to and during this DPhil, and I could not have remained so determined to succeed without their unyielding support. Thank you to Max for being a companion and brother on our shared adventures, to Dad for your willingness to support me, and to Kev and Mum for all the sacrifices you have made, all of the hours by the hospital bed-side, putting up with my grumpiness, supporting me, making life more enjoyable and just always being there when called upon. I could never do all their influence justice in words, but will always strive to be dedicated and useful in what I do.

Abbreviations

2OG	2-Oxoglutarate
95 % C.I.	The 95 % Confidence Interval Limits
aaRS	Amino acyl tRNA synthetase
Ac	Acetyl
Asc	Ascorbate
BCA	Bicinchoninic acid
BLI	BioLayer Interferometry
BME	β-Mercaptoethanol
BSA	Bovine Serum Albumin
ClAc	Chloro-acetyl
Cr	Crotonyl
Da	Dalton (Mass Unit)
DMEM	Dulbecco's Modified Eagle Medium
DTT	Dithiothreitol
EC_{50}	Half maximal efficacy concentration
Et	Ethyl
FAD	Flavin adenine dinucleotide
FIT	Flexible In-Vitro Translation
fMet	Formyl-Methionine
FPLC	Fast Protein Liquid Chromatography
H3	Histone 3
HDxMS	Hydrogen-Deuterium exchange mass spectrometry
HEK293T	Human Embryonic Kidney 293T cells
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence
IC_{50}	Half maximal inhibitory concentration
IOX1	5-Carboxy-8-hydroxyquinoline
IPTG	Isopropyl β-D-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
IVTT	In-Vitro Transcription and Translation
JmjC	Jumonji-C
JmjC-KDM	Jumonji-C containing Histone Lysine Demethylase
K_D	Dissociation (affinity) constant
KDM	Histone Lysine Demethylase
Kme	Lysine methylated on the $^{\varepsilon}$N position
LSD	Lysine Specific Demethylase
MALDI	Matrix Assisted Laser Desorption Ionisation
Me	Methyl
mRNA	Messenger RNA
MS	Mass Spectrometry
Mw	Molecular Weight
NGS	Next generation sequencing
NMR	Nuclear Magnetic Resonance

NOG	N-oxalylglutarate
PBS	Phosphate Buffered Saline (pH 7.4)
PBS-T	Phosphate Buffered Saline (+ 0.1 % Tween 20)
PCR	Polymerase Chain Reaction
PHD	Plant Homeodomain Finger
PHD+JmjC	The PHD Finger natively linked to the JmjC Domain
Pip	Piperidinyl
ppm	Parts per million
Pr	Propionyl
PTM	Post Translation Modification
Pyr	Pyrrolidinyl
qPCR	Quantitative PCR
RAPID	Random nonstandard Peptide Integrated Discovery
ROESY	Rotating-frame Nuclear Overhauser effect Correlation Spectroscopy
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SPPS	Solid Phase Peptide Synthesis
TAE	Tris-acetate-EDTA buffer (1X = 40 mM Tris, 20 mM Acetic acid, 1 mM EDTA)
TBE	Tris-borate-EDTA buffer (1X = 89 mM Tris, 89 mM boric acid, 2 mM EDTA)
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic Acid
THPTA	Tris(benzyltriazolylmethyl)amine
TLC	Thin Layer Chromatography
TOCSY	Total Correlation Spectroscopy
TOF	Time-of-Flight
tRNA	Transfer RNA
TTD	Tandem Tudor Domain
uAA	Unnatural amino acid
UV	Ultraviolet
WT	Wild Type

List of Contributions from Others

Individual contributions are noted appropriately within each Chapter. This list summarises the people and context of their contributed work described in this thesis.

Protein Production: Yijia Zhang (donated several KDM7 proteins used throughout the project). Finn Wolfreys (Protein NMR, Chapter 6). For cross-screening studies in Chapter 4, proteins were donated by: The SGC, Dr Tom McAllister, Dr Bhaskar Bhushan, Eidaraus Salah, Dr Rachel Schamo Schiller.

NMR: Dr Martine Abboud (Protein NMR, Chapter 6) and Dr Nader Amin (Peptide NMR, Chapter 7).

Peptide Synthesis and Purification: Shane Clark (Chapter 7).

Assays: Yijia Zhang (Fluorescence Polarisation, Chapter 3), Dr Anthony Tumber (KDM cross-screening, Chapter 7), Dr John-Paul Bukowski (Immunofluorescence assay for KDM7B, Chapter 8).

Synthesis and IVTT Testing: Karen Heathcote (Synthesis of some of the lysine derivatives, Chapter 1).

HDxMS: Dr Christopher Stubbs (AstraZeneca, Chapter 5)

Physicochemical Property Testing of Peptides: Dr William McCoull and Dr Hitesh Sanganee (AstraZeneca, Chapter 7).

Computational Modelling: Dr Ewa Chudyk and Dr Ron Knegtel (Vertex Pharmaceuticals, Chapter 7).

Next-Generation Sequencing and Analysis: Dr Tom McAllister (Kawamura Lab) and Dr Ed Hookway (Chapter 2).

Chapter 0. Introduction

0.1. Working Definition of Epigenetics

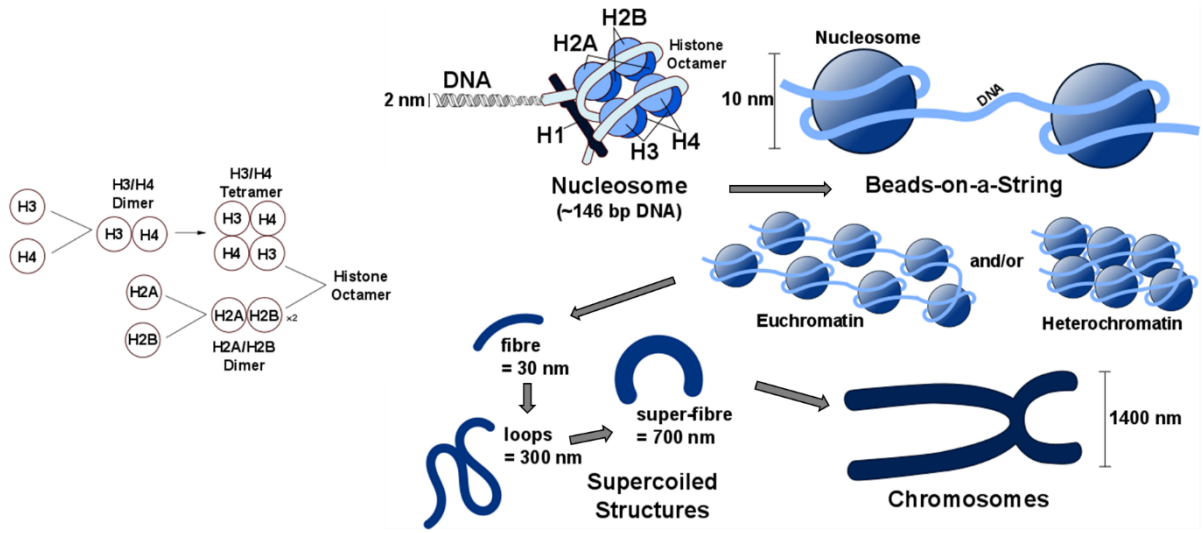
Epigenetics in its broadest sense can be considered as the control and communication of biological change. It is the level of regulation applied on top ('*epi*') of the basic ACGT DNA sequence which affects the interchange of the 'central dogma' of information transfer in biology: DNA to RNA to Protein. In reality, all three elements can and do affect each other which is arguably a key principle of epigenetics.

Epigenetic changes can be initiated by both environmental [1–3] and genetic [4,5] factors, but their heritability remains a subject of fascination and scrutiny, with levels varying between DNA replication and cell division [6] and across generations [7,8]. Exploiting epigenetics to reprogram and control cells is becoming increasingly important in the burgeoning era of stem cells and precision medicine [9]. Some of the most important epigenetic control mechanisms identified thus far include: modification of the fundamental DNA base pairs to non-canonical bases (methylation [10–12], demethylation [13], oxidation [14,15]); histone proteins and reversible post-translational modification (PTM) thereof [16–19]; modification of RNAs [20,21]; regulatory RNAs [22–24] and regulatory proteins [25,26]. These mechanisms hold much promise for understanding the fundamental processes associated with health and disease, which is particularly exciting for the study of change and memory with relevance to the brain [27], immunology [28–30] and immunotherapy [31]. However, all of these epigenetic processes are also multi-layered, interlinking and almost overwhelmingly complex, requiring many more years of dedicated efforts to continue improving understanding for engineering of control in therapeutic interventions. The focus of this thesis centres around (chemical modulation of) epigenetic proteins that regulate histone PTMs.

0.2. The Histone Machinery

Encoding an entire genome with several billion base-pairs results in multi-metre lengths of DNA [32]. To fit inside the microscopic nuclear contents of a cell, the DNA must be packaged and compacted. This starts with wrapping of the 'string-like' double stranded DNA around a 'bead-like' histone octamer consisting of two heterodimeric units made from four proteins (H2A, H2B, H3, H4), with H1 providing structural integrity along the chain (**F.0.2.i.**). The chained nucleosome units progressively coil up into higher order structures, compressing the volumetric space each time, until arriving at the densely packed recognisable chromosome structure [33]. Highly compacted nucleosome units (heterochromatin) are associated with transcriptional inactivation and reduced accessibility [34], whereas less tightly-packed nucleosomes (euchromatin) are associated with increased accessibility and differing levels of active gene transcription [35].

Histone lysine and arginine residues are crucial for stabilising the interactions with DNA, with their often positively charged side chains electrostatically binding the negatively charged DNA phosphate backbone [36] and also reversible chemical bond formation of side-chain functionality with modified DNA bases (e.g. lysine Schiff base with 5-formyl-cytosine) [37]. Altering the strength of interactions between DNA and histones, by reversibly post-translationally modifying key histone residues, provides an important way of controlling the DNA compaction state and hence transcription.

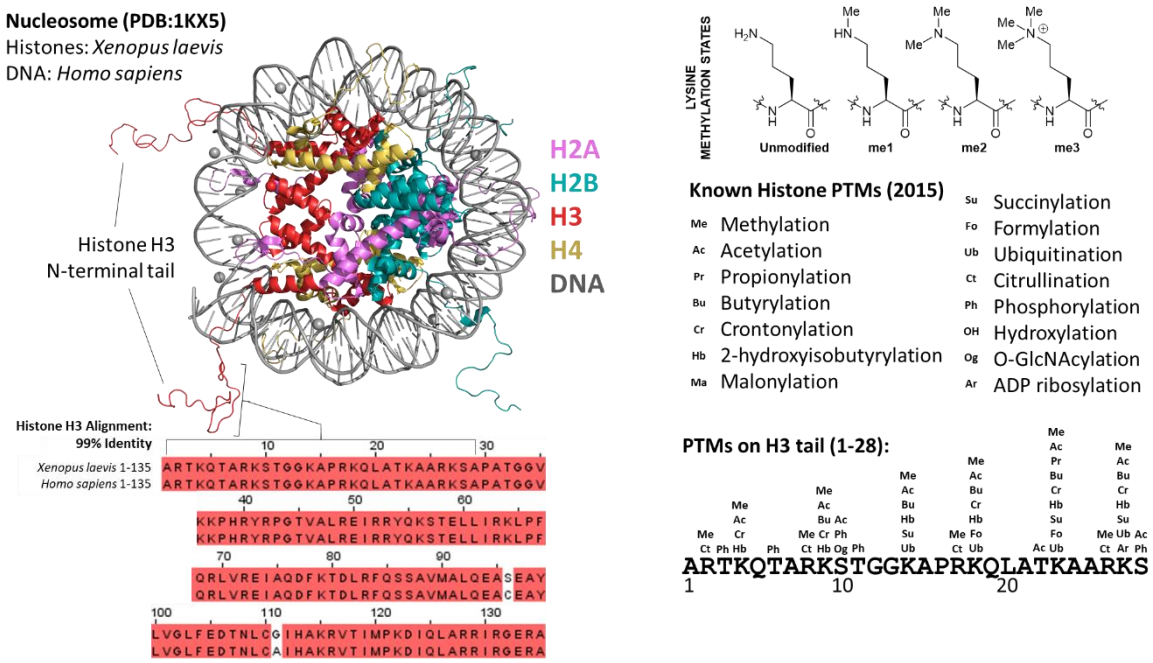


F.0.2.i. DNA is packaged into a cell through progressive structures. DNA wraps around a core histone octamer, composed of histone proteins (two dimers from H2A-H2B and an H3-H4 tetramer, supported by protein H1), to form a nucleosome unit. The nucleosomes are thus connected like 'beads-on-a-string' which winds around itself to form chromatin, leading to the formation of thicker fibres – their degree of compaction signifies the chromatin state (looser winding = euchromatin, tighter winding = heterochromatin). These compacted-chromatin fibres then supercoil to eventually become a recognisable chromosome, able to fit within a cell nucleus.

Discovery efforts continue apace to reveal new and varied post-translational modifications (PTMs) on histones and their contributions to epigenetic regulation [38]. The histone N-terminal tails are the most frequently modified section, with their relatively unstructured 'tail-like' appearance often exposed away from the well organised nucleosome core but remaining essential to function [39] (**F.0.2.ii.**). The N-terminal tail of H3 in particular has been noted for its propensity to be heavily modified with a variety of PTMs [40] – especially on lysines (e.g. K4, K9, K14, K18, K23, K27, K36) (**F.0.2.ii.**) – recently reviewed extensively in 2016 by Huang and co-workers identifying PTMs on all four histone proteins [41]. Methylation is a particularly common PTM for lysines within histone tails and can affect transcriptional activation and deactivation dependent on the degree of methylation (me0/1/2/3) and lysine location within the amino acid sequence (e.g. H3 = histone 3, K4 = lysine at position 4, me3 = 3 methyl groups on ε-N). By way of example, H3K4me3 is often located at active transcription sites [42], whereas H3K9me3 is more frequently associated with transcriptional repression [43]. Much remains to

be elucidated surrounding the roles of these histone marks and the consequences of epigenetic 'cross-talk' where multiple modifications are co-localised.

The transcriptional state of nucleosomes can be reversibly altered since most known PTMs are dynamic. Epigenetic proteins that recognise and modify PTMs are therefore commonly alluded to as 'readers', 'writers' and 'erasers'. For lysine methylation, the 'writers' are histone lysine methyltransferases (KMTs) and 'erasers' the histone lysine demethylases (KDMs), both of which also possess 'reader' capabilities. The work described in this thesis focuses on the KDMs and reader domains.

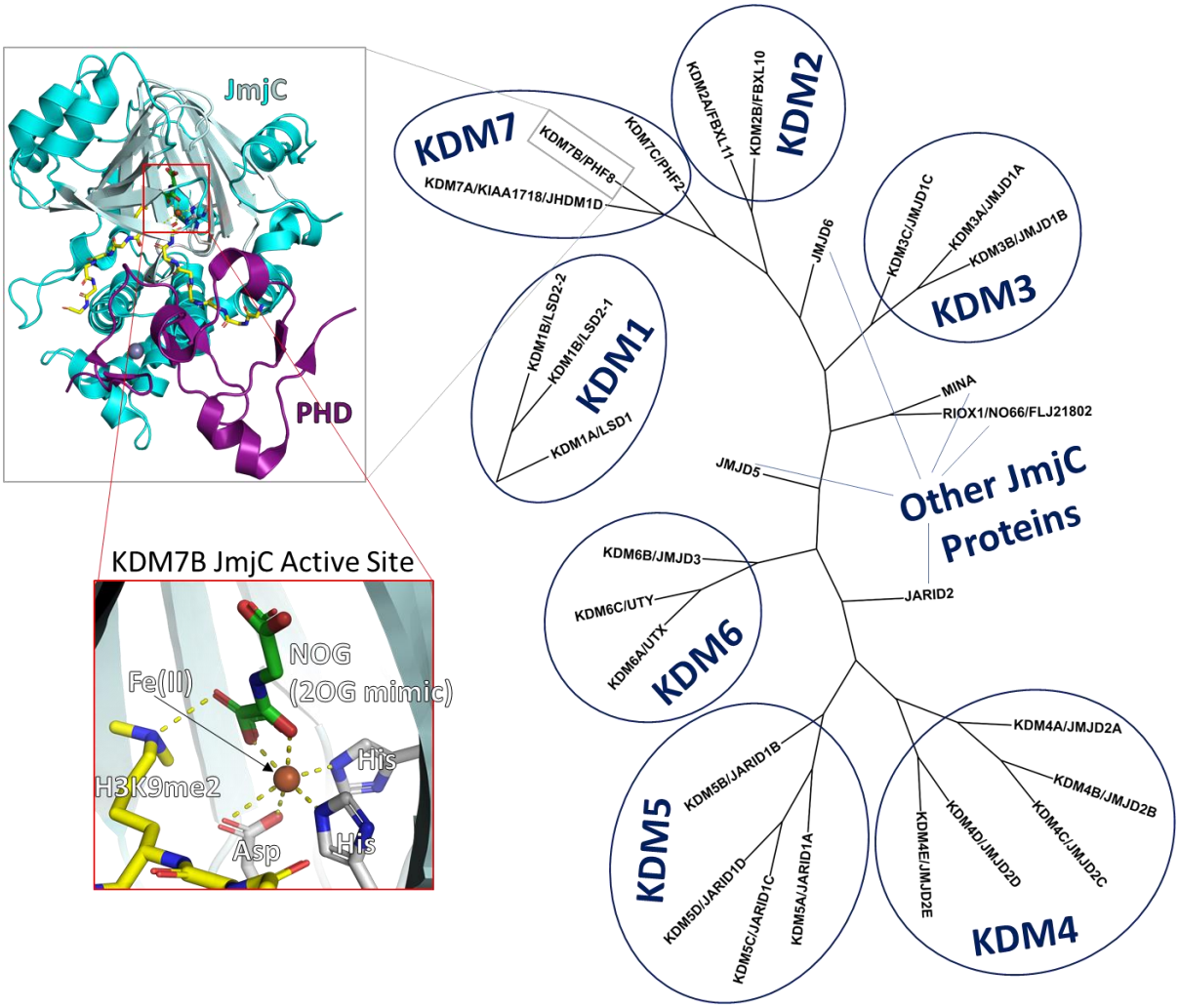


F.0.2.ii. View from a crystal structure of a nucleosome particle (PDB: 1KX5) shows DNA (~147 bp) wrapped twice around a core histone octamer [44]. The histone H3 sequence is highly conserved across eukaryotic species, as shown by the H3 sequence alignment for *Xenopus laevis* (frog) and *Homo sapiens* (human). The protruding histone H3 N-terminal tail is highlighted, as well as its large number of post-translational modifications (PTMs) for amino acid residues 1-28. The four possible lysine PTM methylation states (0, 1, 2, 3 methyl groups) are depicted for clarity, with the methyl group residing on the N^{ε} position.

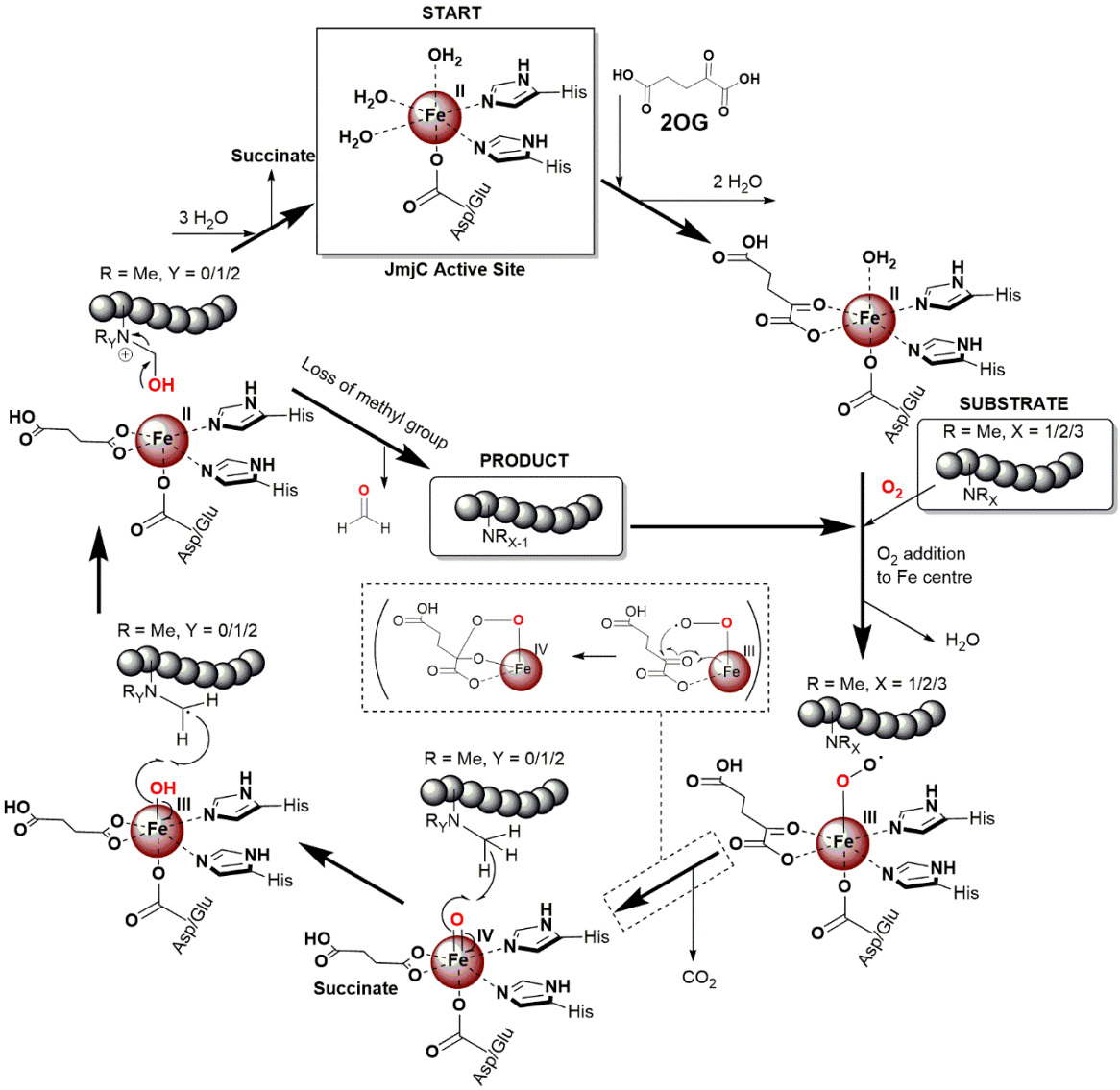
0.3. The Histone Lysine Demethylases (KDMs)

KMTs were first discovered in 2000, but the sequences of their KDM counterparts were only identified in 2003 and 2006 as two mechanistic classes [45]. Individual KDM enzymes are categorised into families first according to their mechanism of action. The KDM1 sub-family operates through a flavin adenine dinucleotide (FAD) dependent monoamine oxidase mechanism which can remove mono- and di-methyl groups only. The other KDM sub-families (2, 3, 4, 5, 6, 7) are divided based on the relatedness of their core catalytic Jumonji-C (JmjC) domain (**F.0.3.i.**) which operates through the proposed consensus mechanism of non-haem iron(II) and 2-oxoglutarate (2OG) dependent oxygenases [46] (**F.0.3.ii.**). As a result of this mechanism, JmjC-KDMs are capable of removing mono-, di- and tri-methyl groups.

The JmjC domain (**F.0.3.i.**) is characterised by eight β-strands arranged in two four-stranded sheets that form a β-helix (aka: jelly roll, cupin barrel, or JmjC fold) which is well conserved across JmjC-2OG oxygenases [47]. The JmjC domain hosts a conserved active site pocket with two histidine residues and a glutamate/aspartate coordinating iron(II), which is in turn used to enable oxidation of 2OG co-substrate with aqueous dissolved oxygen (O_2) to hydroxylate histone lysine N^{ε}-methyl groups, giving a hemiaminal which collapses to release the methyl group as formaldehyde. KDM substrate specificities vary, with different preferences for lysine positions and methylation states within the histone sequence [48] (**F.0.3.iii.**). Some JmjC-KDMs have also been reported to accept methylated arginine histone substrates [49] which opens the possibility of additional catalytic epigenetic roles beyond simply histone lysine demethylation, especially as the wider 2OG dependent oxygenase superfamily of enzymes have been reported to catalyse other chemical transformations (e.g. epoxidation, halogenation, ring expansion, ring closure, epimerisation, desaturation) [46,47,50].



F.0.3.i. Human KDM family of enzymes. The histone lysine demethylases are categorised according their mechanism of action and catalytic domain relatedness. KDM1s operated through an FAD-dependent mechanism, whereas JmjC containing KDMs 2-7 operate through a Fe(II) and 2OG dependent mechanism. The phylogenetic tree classifies KDM families and JmjC-containing putative demethylases according to their JmjC domains. A crystal structure of KDM7B is shown (PDB:3KV4), highlighting its cupin-barrel (silver-blue) within its JmjC domain (blue) and adjacent PHD finger (purple). A close-up of the JmjC active site pocket for KDM7B shows Fe(II) coordinated by two histidine residues and an aspartate residue, and NOG (N-oxalylglycine – a mimic of cofactor 2OG that does not permit catalytic turnover) coordinating to Fe(II) in a bidentate manner. H3K9me2 peptide substrate lies at the active site, near Fe(II) and NOG.



F.0.3.ii. The proposed consensus mechanism of the Fe(II) dependent 2OG oxygenases applied to histone lysine demethylation. 2OG co-substrate enters coordination with Fe(II), followed by methylated lysine substrate in concert with binding of aqueous dissolved O_2, whereupon Fe(II) is oxidised during reduction of 2OG to succinate – via loss of CO_2 – generating a reactive Fe(IV) species, which in turn oxidises the lysine methyl group by radical H-abstraction and radical oxygen transfer, regenerating Fe(II). The newly formed methyllysine hemiaminal species is unstable and collapses to release the methyl group as formaldehyde. Formed succinate is released from the metal centre and the catalytically active Fe(II) species can re-join the cycle.

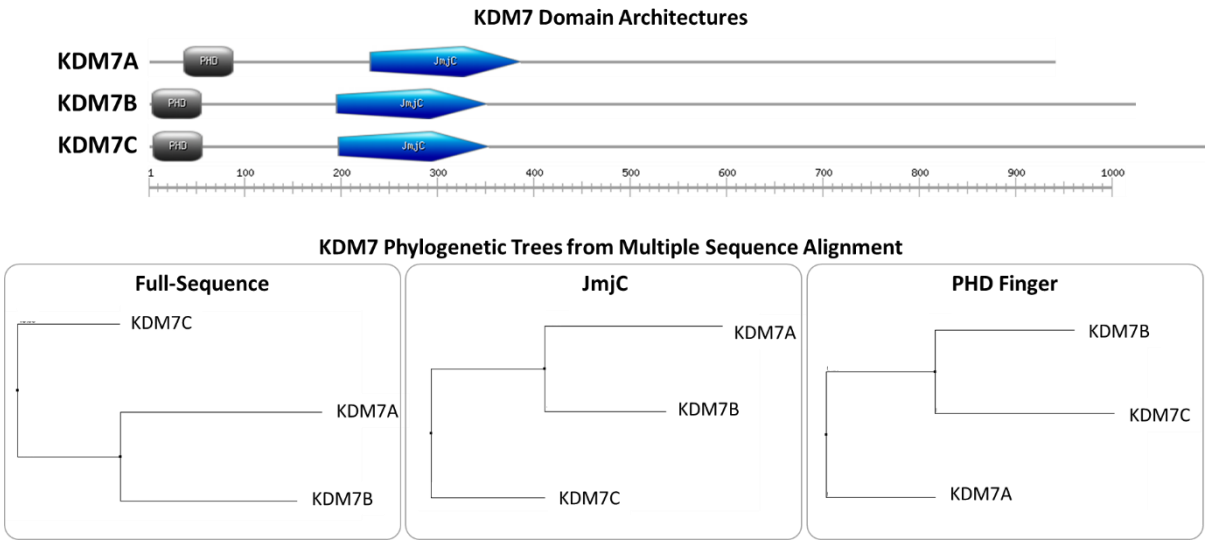
The fundamental biological nature of DNA accessibility means regulation of histone PTMs have wide-ranging consequences for health and disease in many organisms. Knockout mouse phenotype and RNAi knockdown studies have linked KDMs to various cancers [51–53], developmental defects [54], embryonic development & lethality [55], neural development [56,57] and immunological regulation [58] among other biological phenomena. The KDMs are potential therapeutic targets, but a survey in 2016 recorded only four small molecule entities against KDMs in clinical development, all of which targeted the FAD-KDM1 sub-family, and only one in pre-clinical development, which targeted the JmjC-KDM5 sub-family [59]. Chemical tools [60] (*vide infra* 0.6.) are needed to help understand the biological functions of the KDMs, and the catalytic and non-catalytic roles they play. The KDM7 sub-family presents the focus for study in this work.

KDM	Confirmed JmjC Demethylation Substrate Histone Marks	Reader Domain Binding to Histone Marks
KDM2A	H3K36me2/1	
KDM2B	H3K36me2/1	
KDM3A	H3K9me2/1	
KDM3B	H3K9me2/1	
KDM3C		
KDM4A	H3K9me3/me2, H3K36me3/2	TTD (H3K4me3, H3K23me3, H4K20me3)
KDM4B	H3K9me3/me2, H3K36me3/2	TTD (H3K23me3)
KDM4C	H3K9me3/me2, H3K36me3/2	TTD (H3K4me3)
KDM4D	H3K9me3/me2	
KDM5A	H3K4me3/2	PHD1 (H3K4me0), PHD3 (H3K4me3)
KDM5B	H3K4me3/2	PHD1 (H3K4me0), PHD3 (H3K4me3)
KDM5C	H3K4me3/2	PHD1 (H3K9me3)
KDM5D	H3K4me3/2	
KDM6A	H3K27me3/2	
KDM6B	H3K27me3/2	
KDM7A	H3K9me2/1, H3K27me2/1	PHD (H3K4me3)
KDM7B	H3K9me2/1	PHD (H3K4me3)
KDM7C		PHD (H3K4me3)

F.0.3.iii. Table of KDM JmjC domain demethylase substrate specificities that have been verified in vitro or in cells by more than one report from at least two different labs. Reader domain binding preference for histone marks is stated where known. Data adapted from a review by C.Højrup, O.D.Coleman, J-P.Bukowski, R.P.Clausen & A.Kawamura [61]. KEY: TTD = tandem tudor domain, PHD = plant homeodomain finger.

0.4. The KDM7 Sub-Family

The human KDM7 sub-family consists of three members: KDM7A (aka KIAA1718), KDM7B (aka PHF8) and KDM7C (aka PHF2). The KDM7 domain architecture (**F.0.4.i**) consists of an N-terminal plant homeodomain (PHD) finger (hereafter referred to as $KDM7_{[PHD]}$), responsible for driving KDM7 binding to the histone H3K4me3 mark, linked to the larger catalytic JmjC demethylase domain (hereafter referred to as $KDM7_{[JmjC]}$). The sequence C-terminal to the JmjC domain is yet undefined, with no classifiable functional domains beyond nuclear localisation signals (NLS). However, the poorly characterised C-terminus of KDM7B has been reported to be important for association to RNA Polymerase I & II [62,63]. Additionally, several post-translational phosphorylation sites are reported experimentally and computationally in the latter half of the KDM7 sequences [64].



F.0.4.i. The human KDM7 sub-family domain architectures. Each KDM7 contains a PHD finger and JmjC domain. The relatedness of the full-length human KDM7 sequences, their JmjC domains and PHD fingers are shown by phylogenetic tree (generated with JalView). KDM7A and KDM7B are more closely related overall and in their JmjC domains, however KDM7B and KDM7C are more similar in their PHD fingers. All domains were classified using the canonical sequences for KDM7A-1, KDM7B-2, KDM7C-1 (UniProt) and domain maps were generated with ExPASy Prosite.

KDM7A$_{[JmjC]}$ demethylation primarily targets H3K9me1/2, but also H3K27me1/2 [65]. KDM7B$_{[JmjC]}$ primarily demethylates H3K9me1/2 [65]. Both KDM7A and KDM7B are reported to demethylate H4K20me1 [66,67] dependent on the histone or nucleosome context, although this assignment remains controversial. The KDM7A$_{[PHD]}$ and KDM7B$_{[PHD]}$ regulates their JmjC demethylase activity against H3K9me2 differently, dependent on the presence of H3K4me3 (Chapter 3) [65]. KDM7C has been reported to demethylate H3K9me1/2 in cells [68,69] and H4K20me3 on mono-nucleosomes [70]. However, these KDM7C activities remain controversial as none have yet been recapitulated in recombinantly-purified protein assays [65]. KDM7C possesses a tyrosine at position 321 to chelate active-site Fe(II) instead of the conventional histidine, but even Y321H mutation fails to restore in-vitro demethylation [71]. It may be the case that some KDM7 demethylase activities are only observable in cells when the enzyme forms part of a larger complex or has itself become post-translationally modified [69,71]. Further study would be aided by KDM7 selective chemical probes and inhibitors; however, unfortunately none are yet reported (*vide infra* 0.6.).

Wide-ranging biological roles of the KDM7 sub-family are increasingly being described (references are reported in the table - **F.0.4.ii.**). KDM7A, KDM7B and KDM7C have all been linked to progression of various types of cancer, whilst KDM7A and KDM7B have been linked to playing important roles in neural development in mouse knockout and cell-based studies. A mouse knockout study of KDM7C suggested a role in neonatal development and control of adipogenesis. Other putative biological roles for the KDM7s include control of immunological responses and hypoxia signalling, amongst others.

The detailed molecular mechanism by which the KDM7s regulate these varied processes requires further investigation. Further questions also remain regarding roles the KDM7 sub-

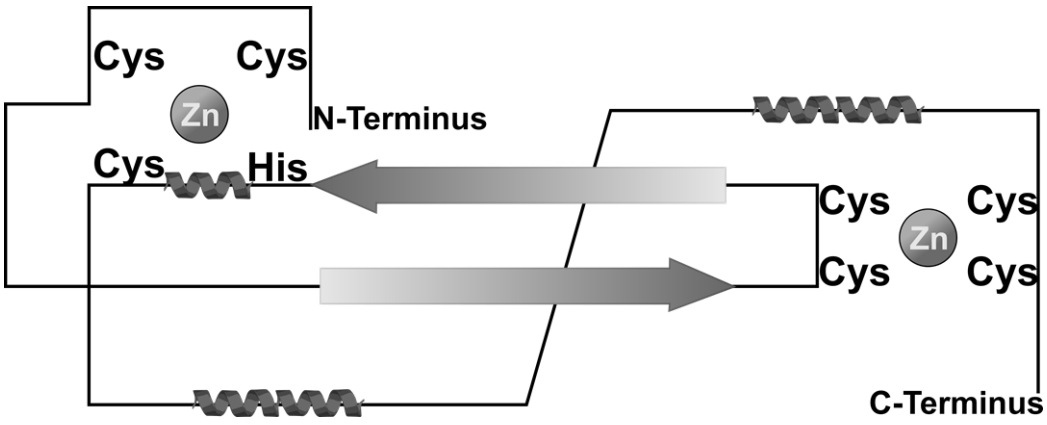
family may play that are independent of their demethylase activity. For example, KDM7$_{[PHD]}$ binds to the histone H3K4me3 mark which is often located at transcription start sites [72], in turn leading to the association of KDM7B playing a role with transcriptional activation of poised RNA polymerase I & II [63,73,74].

Protein	Neural Biology	Cancer Biology	Other Potential Biological Associations
KDM7A	Linked to control of mouse embryonic stem cell differentiation to neural cell types by transcriptional activation of the FGF4 gene in knockdown and rescue experiments [56].	RNAi knockdown reduced LNCaP (androgen-sensitive human prostate cancer) proliferation [75].	RNAi knockdown reduced ICAM1 protein expression and leukocyte adhesion after TNF-α stimulation, KDM7A stabilises ICAM1 in lysosomes [76].
KDM7B	Knockouts of KDM7B in mice: mis-regulation of serotonin levels observed alongside increased resistance to anxiety and depressive-like behaviours [57]; impaired learning and memory [77]. Other studies have linked KDM7B and its gene deletions to impaired mental function with an association to ZNF711 in neural development deficiencies [78]; brain and jaw development in zebrafish [67] and neuronal cell differentiation [79].	Overexpressed and correlated to poor survival in NSCLC by promoting miR-21 expression [80]; acts as an oncogene in breast cancer [81,82]; promotes progression of leukaemia (ALL and APL) [83,84]; regulates miRNAs in and development of prostate cancer [53,85–87]; linked to progression of esophageal squamous cell carcinoma [88]; role in gastric cancer metastasis [89] and hepatocellular carcinoma metastasis [90].	Regulating endothelial cell migration [91]; becoming down-regulated in cardiac hypertrophy [92]; hypoxia signalling [93]; promotion of innate immunity signalling to activate adaptive immunity responses (evidenced by proteomic studies on macrophages) [94]; silencing of IFNγ-induced inflammatory genes [95]; a role in cell cycle progression [66,96].
KDM7C		The KDM7C C-terminal region is believed to associate to the tumour suppressor protein p53 and expression of KDM7C is downregulated in colon cancer [97].	Neonatal development and adipogenesis in a mouse knockout study [98]; cell-based study suggesting a link with adipogenesis by binding to transcription factors C/EBPδ and C/EBPα [99]; control of pro-inflammatory genes [70]; bone formation [100]; and direct interaction with ARID5B [69,101].

F.0.4.ii. Table outlining selected publications that have linked KDM7A, KDM7B and KDM7C to specific biological outcomes, among the most prevalent of which for all three is in cancer.

0.5. Plant Homeodomain (PHD) Fingers

PHD fingers are a sub-set of zinc finger proteins. The first PHD finger was discovered in 1993 and was named for its origin of discovery as part of the plant homeodomain HAT3.1 [102]. PHD fingers now constitute one class of the many epigenetic reader domains which together enable recognition of histone PTMs [103], including: methyllysine, methylarginine, acetyllysine, phosphoserine and unmodified H3 [104] – all of which are of broad importance to biology [105]. PHD fingers are small (typically 40-80 amino acids and less than 10 kDa [106]), flexible domains that bind two zinc ions through a Cys_4-His-Cys_3 motif (**F.0.5.i.**) [107]. Their amino acid content and sequences are otherwise variable, resulting in a great diversity of structures, but they generally possess several short α-helical regions and bind to their target through use of a short antiparallel β-sheet. PHD fingers are present in more than 100 proteins across the human proteome [107] and have been found in proteins that play roles in: transcriptional regulation, cell cycle control, autoimmune regulation, DNA methylation, chromatin remodelling and nuclear signalling [107,108]. Pertaining to KDMs, PHD fingers are present in the KDM2, KDM4, KDM5 and KDM7 sub-families, with a binding partner currently identified only for $KDM7_{[PHD]}$ and $KDM5_{[PHD1/PHD3]}$ [109] (**F.0.5.ii.**).



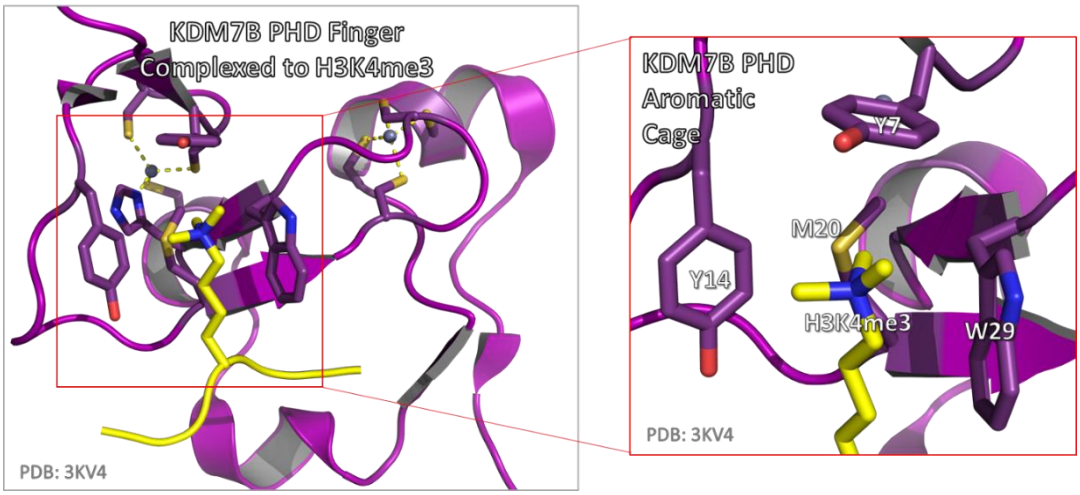
F.0.5.i. Common structural features of a PHD finger. The PHD finger is a short domain, typically 40 to 80 amino acids, consisting of a central β-sheet (integral for binding of methylated lysine) and several short α-helical regions. Two zinc ions are bound, one by a Cys_3His motif and the other by a Cys_4 motif. Structural information adapted from Sanchez & Zhou [107].

A common function for many functionally assigned PHD fingers is 'reading' of lysine methylation states on the histone H3 N-terminal tail with high nanomolar to low micromolar affinity. The largest subset of PHD fingers identified so far, which includes the KDM7s, favour binding to H3K4me3 by engaging the trimethyllysine in aromatic and hydrophobic cages [107,110] (**F.0.5.ii.**). Others, such as autoimmune regulator (AIRE) PHD1, favour binding to the first four unmodified H3 N-terminal residues through formation of ionic interaction salt bridges [111,112]. Fewer PHD fingers have been found to bind to other histone marks (e.g. H3K9me3 and H3K14ac) [113] and fewer still for non-histone targets (DNA, Pin1, B9L HP1, Ubc9 and CyP33 RRM) [113]. Additionally, some PHD fingers have been shown not to interact with common histone PTMs [114], with the precise target and function of many still to be uncovered.

Target	PHD Finger
H3K4me3	BPTF
	DIDO
	ING1
	ING2
	ING3
	ING4
	ING5
	KDM5A-PHD3
	KDM5B-PHD3
	KDM7A
	KDM7B
	KDM7C
	MLL1
	MLL5
	PHF13
	PHF23
	PYGO
	RAG2
	TAF3
	YNG1

Target	PHD Finger
H3-unmodified 1-4	AIRE-PHD1
	BAZ2A
	BAZ2B
	BHC80
	BRPF2-PHD1
	CHD4-PHD1
	CHD4-PHD2
	CHD5-PHD2
	DNMT3L
	JADE1-PHD1
	KDM5A-PHD1
	KDM5B-PHD1
	PFI-PHD1
	TRIM24
	TRIM33
	UHRF1

Target	PHD Finger
H3K9me3	CHD4-PHD1
	CHD4-PHD2
	KDM5C-PHD1
	TRIM33
	NSD3-PHD5



F.0.5.ii. (Top): Three main classes of PHD finger target binding marks and their known associated PHD fingers. Binding to H3K4me3 is the most frequent function identified to date. **(Bottom):** The KDM7B$_{[PHD]}$ in complex with its target mark H3K4me3 [65] (PDB: 3KV4). The zinc binding $Cys_3HisCys_4$ motifs are shown. The inset zoom highlights the aromatic cage residues responsible for binding the lysine N$^{\varepsilon}$-trimethyl group through pi-cation interaction.

It is worth noting that some histone-reading PHD fingers are simultaneously sensitive to other PTMs surrounding the target lysine, providing a level of combinatorial histone PTM recognition even within a PHD finger. For example, H3R2 methylation is known to disrupt binding a subset of H3-unmodified and H3K4me3 PHD fingers [115], whilst phosphorylation of nearby serine and threonine residues can either disrupt [116] or enhance [113] H3 engagement for other subsets. PHD fingers can even be found acting in concert with other histone reader domains (e.g. tandem-PHD [117], Tudor [118] or Bromodomains [119]) to further increase the inter-domain cross-talk and the capacity to recognise and discriminate between complex combinatorial PTMs. More generally, the role of the histone-interacting PHD finger is perhaps best considered as a type of biological homing system, serving to respond to a dynamically changing PTM landscape and guide its attached parent protein to its required location.

Whilst progress has been made through use of biochemical approaches with arrays of modified histone peptides to elucidate the function of PHD fingers and their target specificity [120–122], there currently exists a dearth of suitable chemical tools to conduct further detailed study on the consequences of PHD finger functions to wider biology (*vide infra 0.6.*). There is a strong case to be made for the development of such tools, as they provide a temporal-control alternative to genetic approaches (e.g. knockdown and mutations) which risk disrupting protein integrity and inducing un-natural effects.

PHD fingers have been directly linked to biological outcomes. For example, the first PHD finger of AIRE ($AIRE_{[PHD1]}$) has been found to be crucial to promiscuous gene expression for successful education of nascent T-cells in the thymus [123–125]. The third PHD finger of KDM5A ($KDM5A_{[PHD3]}$) has been implicated in acute myeloid leukaemia (AML), reported as a fusion oncoprotein with NUP98 to induce aberrant activation of genes in AML [126]. PHD fingers in

NSD2 (aka MMSET) H3K36 methyltransferase are reported to be important for cellular activity and transcriptional activation of oncogenes in multiple myeloma [127]. Thus, PHD fingers are emerging as potential targets in oncology. PHD fingers have also been reported to regulate the catalytic activities of their associated enzymes. For KDMs, some of the PHD fingers in KDM5A were reported to regulate demethylase activity [109] and there is strong evidence for the single PHD finger in each KDM7 also being able to allosterically control JmjC demethylase activity [65,68], with consequence on biological phenotypes [74]. It is therefore crucial to properly understand how PHD fingers influence epigenetic control, and chemical probes are needed to elucidate the functions of PHD fingers in biology. The capacity of PHD fingers to wield such influence may also make them more than objects of academic curiosity, as cases could conceivably arise where PHD finger-probes become useful in validating desirable targets for future therapeutic development [115]. Yet, remarkably no selective or potent PHD finger probes or inhibitors are currently reported.

0.6. Chemical Tools for KDMs and PHD Fingers

A chemical tool for biological research can be considered a molecular entity that serves as either a functional probe or an inhibitor. A probe is designed to bind to a target(s) in order to either study or manipulate the biological system in which it residues. An inhibitor can be a chemical probe, but additionally acts to disrupt a normal function of the target. Probes are frequently modified to possess chemical 'handles' which allow their observation within cells, or direct retrieval to recover bound proteins. They are designed to serve a fundamentally different purpose to that of a drug, as a drug must be designed to exert a beneficial effect in a living system whereas a probe must deliver precise target interrogation for biological understanding of health and disease. Chemical probes offer a complementary approach to RNAi and genetic mutation alternatives which can help to build evidence for and validate biological processes.

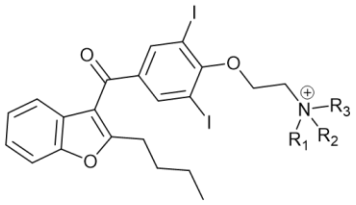
Chemical probes have been the subject of much discussion in the literature, with a series of excellent reviews in recent years [128–134] delving into their impacts and shortcomings. The Structural Genomics Consortium (SGC) has suggested a set of criteria for a molecule to meet before it can be considered a useful probe: an in-vitro potency less than 100 nM; at least 30-fold selectivity versus similar proteins and target engagement within cells at 1 μM. Of these three criteria, for the KDMs in general, potency has been readily achievable whilst selectivity and cell activity have been major challenges [61,135]. Without well-defined selectivity profiles, probes can be unreliable as it becomes difficult to discern whether observed effects are truly associated to the target of interest or the result of off-target behaviours.

Development of probes for PHD fingers has proven particularly challenging, with fewer than five publications to-date of compounds targeting PHD fingers. A series of amiodarone

derivatives assessed by Wagner et al. [136] identified two compounds (di-N-desethylamiodarone and WAG-003) reported to possess micromolar potency for displacement of H3K4me3 peptide from $KDM5A_{[PHD3]}$, but suffering from poor and unpredictable selectivity against other reader domains (micromolar potency against $KDM4A_{[TTD]}$ and $ING2_{[PHD]}$, no effect against $RAG2_{[PHD]}$, $AIRE_{[PHD1]}$, $BHC80_{[PHD]}$, $UHRF1_{[TTD]}$). Bhushan et al [137] could not reproduce the efficacies seen for these compounds, nor identify rational structure-activity relationships despite extensive chemical derivatisation. Further, they found this series of compounds to also bind non-specifically to JmjC domains and inhibit KDM5 and KDM7 demethylase activities, suggesting they do not hold potential for development of genuine PHD finger probes (**F.0.6.i.**). Other reports of compounds with millimolar affinity for $PYGO_{[PHD]}$ and no selectivity profiles [138–140] do not present viable probe molecules, but do highlight the difficulty in finding small molecules capable of binding to a domain that is inherently flexible and without a traditional 'binding-pocket' (*vide supra* 0.5.).

Development of small molecule probes against other epigenetic reader domains has been more successful, with high nanomolar to low micromolar potency as well as selectivity achievable for the SPIN1 tandem-tudor domain (TTD) (H3K4me3 binding) [141] and a series of BET-bromodomain (KAc binding) inhibitors [142], some of which have seen use in cancer studies [143]. A potent and cell active small molecule inhibitor has been reported for L3MBTL3, a Kme1/2 reader belonging to the malignant brain tumor (MBT) family of chromatin-interacting transcriptional repressors, which was able to displace its methyllysine reading function and facilitated discovery of new protein-protein interactions [144]. At least two potent small molecule inhibitors are reported to bind within the H3K27me3-reading pocket of EED, a subunit of the PRC2 complex involved in epigenetic regulation, with implications for understanding cancer progression [145,146]. In addition, a picomolar affinity small molecule

enabled discovery that inhibition of the 'WIN' site (an 'ARA' motif binder) in the protein WDR5 led to its dissociation from chromatin [147], where WDR5 is believed to be important in organising epigenetic writer proteins and contribute to oncogenesis. Hence, development of probes for PHD fingers represents a significant unmet need [148,149] and would be likely to present great opportunity for increasing understanding of epigenetic processes.



Amiodarone: $R_{1,2}$ = Et, R_3 = H
di-N-desethylamiodarone: $R_{1,2,3}$ = H
WAG-003: $R_{1,2,3}$ = Me

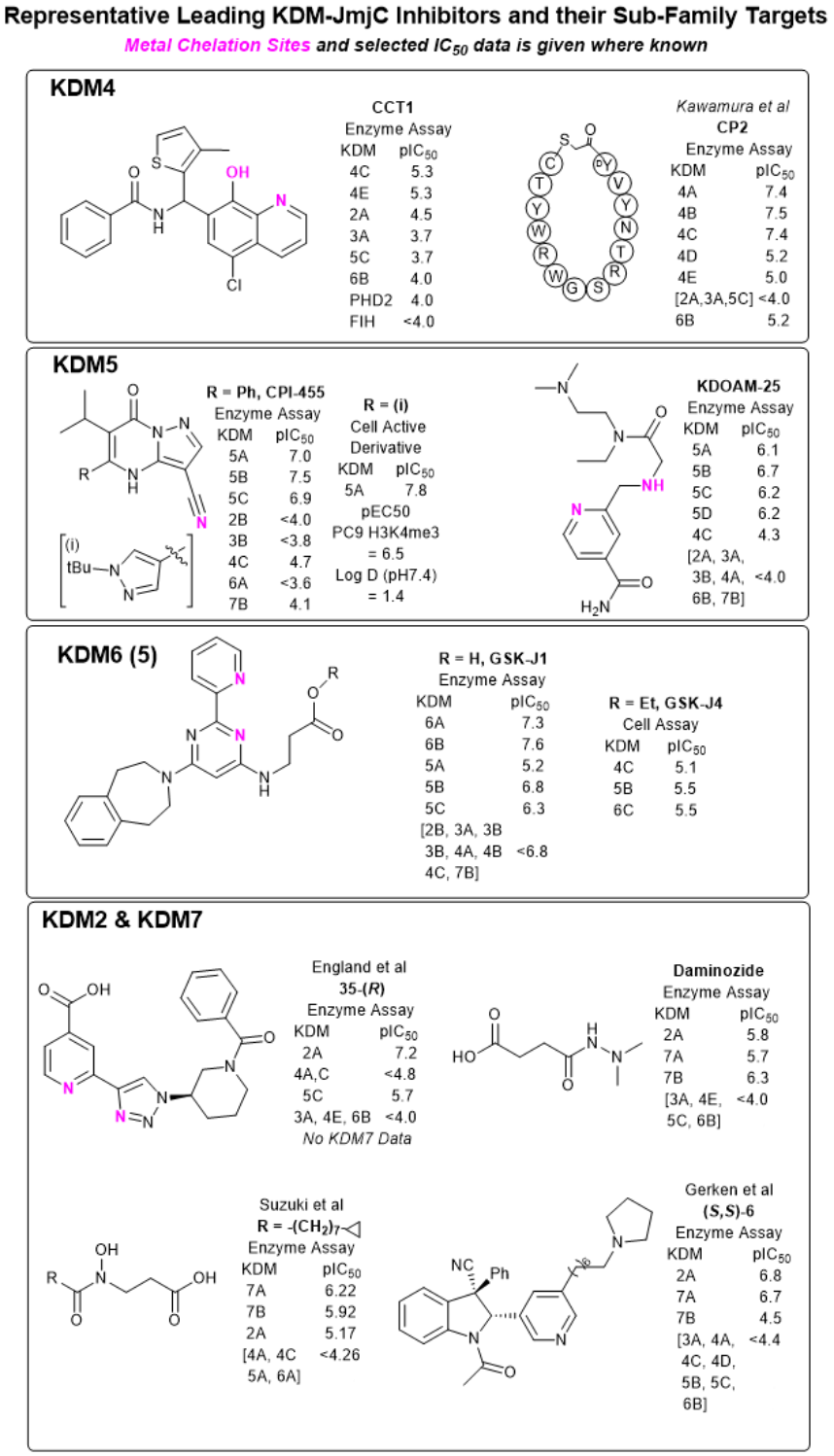
			Displacement IC_{50} (μM)						
	Reader	Cognate	Wagnet et al. 2011			Bhushan et al. 2018			
Protein	Domain	Binder	1	2	3	Self	1	3	Compound List
KDM5A	PHD3	H3K4me3	NM	26 ±15	30 ±14	0.22 ±0.02	99 ±23	1658 ±490	**Self** Cognate Binder
KDM4A	DTD	H3K4me3	-	72 ±26	34 ±14	-	-	-	**1** Amiodarone
ING2	PHD	H3K4me3	-	97 ±41	105 ±33	-	-	-	**2** di-N-desethylamiodarone
KDM7A	PHD	H3K4me3	-	-	-	1.9 ±0.1	156 ±12	2232 ±790	**3** WAG-003
KDM7B	PHD	H3K4me3	-	-	-	1.7 ±0.2	252 ±53	307 ±76	
KDM7C	PHD	H3K4me3	-	-	-	2.7 ±0.2	84 ±14	976 ±146	

*NM = not measurable, - = not assessed

F.0.6.i. A summary of small molecule compounds assessed as possible PHD finger inhibitors by Wagner et al.[136] and Bhushan et al.[137] None of the compounds displayed potent or selective inhibition of target PHD finger binding to its cognate histone peptide. Bhushan et al. performed extensive derivatisation of amiodarone for structure-activity studies and did not identify any substantially more potent or selective inhibitors in the compound series.

More progress has been made with probe development against the KDMs through targeting of their JmjC domains [135,150]. However, the well-conserved nature of the JmjC domain (*vide supra* 0.3.) has presented continual and significant problems concerning sub-family selectivity. Many of the current KDM small molecule probes act through binding within the JmjC active site, chelating to iron(II) and competitively mimicking the 2OG co-substrate. This generic mechanism of action in part accounts for the fundamental selectivity issues observed with this approach, but has also led to issues in cellular systems with the high native concentrations of 2OG (~2 mM [151]) outcompeting the probe. Furthermore, the common use of carboxylic acid-like functionalities to achieve 2OG mimicry has meant that cell permeability

can sometimes present an obstacle [152]. Small molecule probes for KDMs with in-vitro potency and selectivity have only been developed with substantial success against the KDM4 and KDM5 sub-families (**F.0.6.ii.**) [153].



F.0.6.ii. Selected examples of inhibitors targeting the KDM JmjC domain for inhibition of demethylase activity. Selected pIC_{50} data is shown where available to give an indication of compound selectivity. No KDM3 or KDM7 selective inhibitors are yet reported. CCT1 [154]. CP2 [155]. CPI-455 [156,157]. KDOAM-25 [158]. GSK-J1/4 [152]. 35-(*R*) [159]. Daminozide [160]. Suzuki et al [161]. Gerken et al [162].

Several potent small molecule probes targeting the KDM7 sub-family JmjC domain have been developed (**F.0.6.ii.**), but all have suffered with off-target efficacy against the JmjC-similar KDM2 sub-family [160,162], and poor cell permeability. The histone H3K4me3 N-terminal peptide, cyclised and conjugated to TAT (the HIV transactivator of transcription), was reported as an inhibitor for KDM7B wherein it displayed low micromolar affinity and inhibition of demethylase activity. However, this peptide also exhibited poor selectivity, with equal magnitude efficacy of inhibition against KDM5B and KDM4C, and high micromolar inhibition of KDM2A, KDM3B and KDM6A [163]. These results are perhaps unsurprising given the generic origin of the sequence, and the knowledge that poly-arginine sequences alone (as found in TAT) can be sufficient to inhibit the KDMs with nanomolar potency [155].

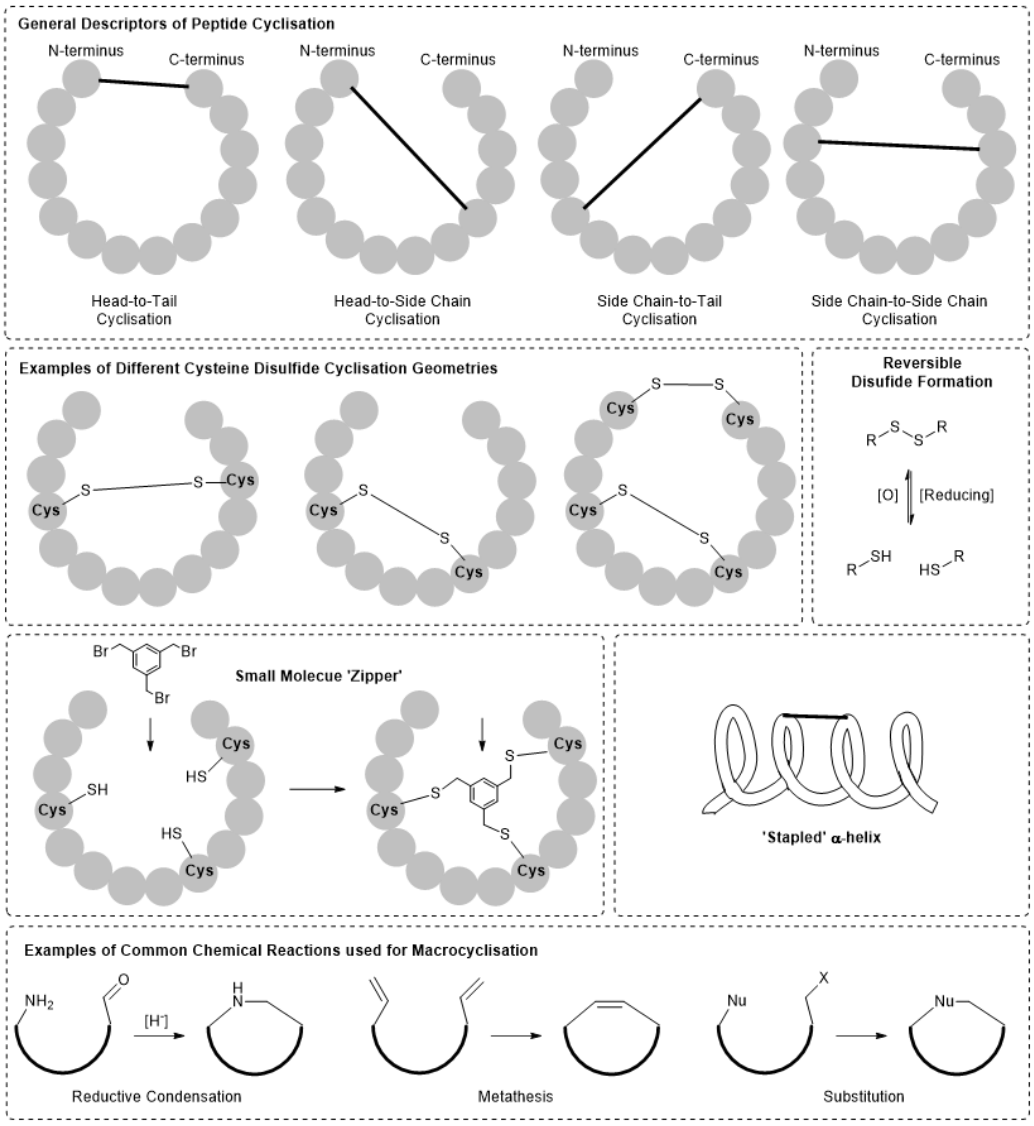
However, novel cyclic peptides discovered through other means have recently begun to show exceptional promise as chemical probes against the KDMs. The first KDM4 sub-family isoform selective inhibitors were reported by Kawamura and co-workers as 12mer cyclic peptides (CP2 (**F.0.6.ii.**) & CP4), derived from mRNA display [155]. They possessed low nanomolar binding affinity and IC_{50} inhibition of KDM4A,B,C JmjC demethylase activity, but no inhibition of other 2OG oxygenases (KDM2A, KDM3A, KDM5C, KDM6B, EGLN1, FIH) up to micromolar concentrations [155]. Crystallography identified the cyclic peptide binding site to be on the surface surrounding and within the channel leading to the JmjC active site, which prevented histone H3K9/K36me3 insertion and did not rely on metal chelation or mimicry of 2OG. This novel mode of action giving rise to unprecedented potency and selectivity suggested a new way forwards using cyclic peptides to target KDMs.

0.7. Cyclic Peptides and Selection Platforms

Cyclic peptides represent an intermediate macromolecular chemical space between small molecules (< 500 Da in size) and proteins (> 5 kDa). As a result, they share traits of both classes being capable of possessing the exquisite target selectivity and affinity comparable to antibodies [164] whilst retaining the synthetic manipulability of small molecules. Their larger size enables them to make more contacts with their target and they are capable of tackling challenging protein-protein interactions [165]. These properties make them an excellent candidate class of molecules for chemical probes. The most accessible targets for cyclic peptides are generally acknowledged to be extracellular, since reliably achieving cell permeability remains problematic. However the challenge of cell permeability can be surmounted for cyclic peptides, as proven by their development against intracellular targets with increasing success [166,167].

Peptide cyclisation is important for providing conformational restraint, which is beneficial for potency by reducing the entropic cost associated with binding and for boosting resistance to degradation from proteases and in some cases also improving cell permeability [168]. Cyclisation can take several forms, classifiable in four general ways: head-to-tail, side chain-to-side chain, head-to-side chain and side chain-to-tail (**F.0.7.i.**) [169]. Head-to-tail cyclisation connects the N-terminal α-NH_2 position to the C-terminal α-COOH position. Cyclisations involving side chains can be much more varied. A classic approach is disulfide formation between two free cysteine thiols – their position within the peptide can be varied to create differently sized macrocycles, with use of orthogonal oxidation chemistries possible to give multiple cyclisation events (**F.0.7.i.**). However, disulfides are not often resistant to the reducing conditions encountered in biological systems, leading to unstable cyclisation. Other approaches to stably constrain peptides have typically included use of unnatural amino acids to provide orthogonal side

chain reactivity giving cyclisations via substitution [170], condensation [171], and other chemistries [169]. Small molecule 'zipper' type molecules have been shown to serve well as chemical connectors by linking cysteine residues or using click chemistry, enabling formation of mono, bi and polycyclic structures (**F.0.7.i.**) [172,173]. 'Stapling' linkages have been applied to short alpha-helical peptides to lock in their conformation, often via use of metathesis or 'click' reactions, creating so-called 'stapled peptides' [174] which have found use as agents targeting protein-protein interactions [175] effective in cells [176].



F.0.7.i. Peptides are cyclised in one of four general ways: head-to-tail, head-to-side chain, side chain-to-tail, side chain-to-tail. Cysteines have been commonly used to achieve this, but the disulfide bond is vulnerable to reducing conditions leading to loss of cyclisation. Small molecule 'zipper' molecules with electrophile handles to react with cysteine have been employed to irreversibly cyclise peptides. Other irreversible cyclisations are increasingly common that do not rely on cysteine at all, including condensations (e.g. reductive amination), metathesis reactions and nucleophilic substitutions.

There is renewed enthusiasm around the use of cyclic peptides for therapeutic development. In addition to their excellent potency and selectivity, they possess several further advantages. Their size, complex conformational geometries and ability to exploit naturally occurring amino acid interactions means they are useful for binding to targets that are normally considered 'undruggable', being able to bind to allosteric sites as well as within pockets. They can be adapted to carry payloads [177] and are readily evolved commercially with other technological developments (e.g. Bicycle Therapeutics). Additionally, peptides of length 15 residues or fewer [178] are less likely than small molecules to be broken down into products with unwanted immunogenicity, since they are constituted mostly from naturally occurring amino acids. This is traded-off against higher clearance rates, since the body is well equipped to breakdown peptide-based molecules – however, clearance rates and other pharmacokinetic properties are now increasingly more manageable through new synthetic modifications [179], formulations and patient-delivery options [177]. Several recent reviews have explored the resurgence in interest in the pharmaceutical sector [180], highlighting increased investment and opportunities [178] with 155 peptides in active clinical development as of March 2017 [177] and cyclic peptides proving worthy for targeting a wide range of disease indications including: cancer, immunology, infection, respiratory disease, blood disorders and other areas. New discovery platforms have even been developed to address the current need for new antimicrobial peptides [181].

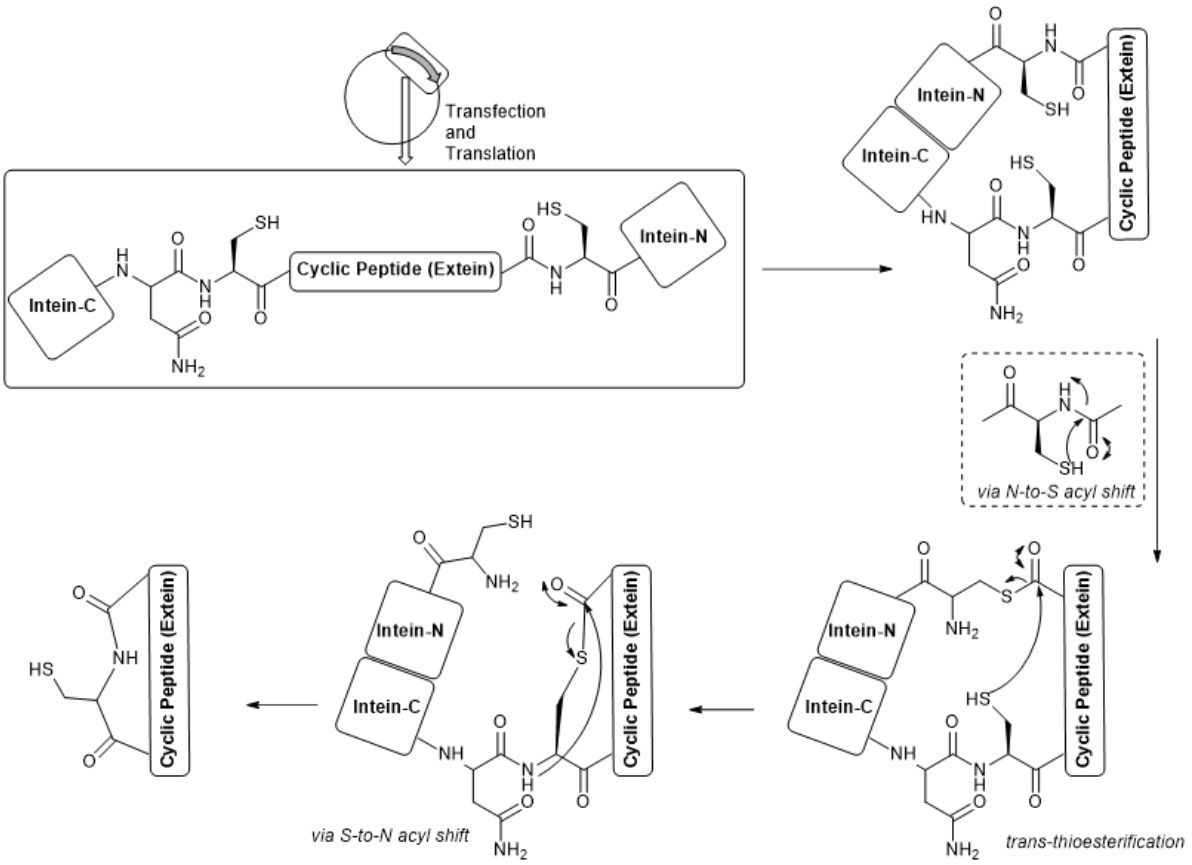
Improvements in technology have been a crucial factor in refining the ability to successfully develop cyclic peptides against a target of interest. Cyclic peptide therapeutics have historically been most frequently adapted from natural products (e.g. the anti-cancer HDAC inhibitor Romidepsin) [182], but a variety of high-throughput methodologies are now providing easier access to novel and unpredictable scaffolds for both probes and therapeutics. Whilst

cyclic peptide libraries are accessible through traditional synthetic methods [169,183], peptides possess an inherent advantage in their ability to be genetically encoded by DNA for high-throughput ribosomal synthesis. DNA encoding means library length and amino acid composition can be easily controlled, with discovery platforms commonly adopting randomised peptide libraries through use of NNK type codons (N = A/C/G/T, K = G/T) – able to represent all 20 amino acids for fewer than the normal 64 possible codons to maximise synthesis efficiency. The various discovery platforms also each ensure a given peptide is traceable to its encoding sequence for later information retrieval. Together, this enables rapid creation of diverse peptide libraries that are much larger (typically 10^6 to 10^{12} members) [184] than classical small molecule libraries, which can be iteratively screened for enrichment against a target. Several reviews elsewhere in recent years have explored available peptide library discovery platforms in depth [184–188], as well as the emergence of DNA encoded small molecule libraries for drug discovery [189,190]. Highlights of selected cyclic peptide discovery platforms are discussed here.

A well-established approach for peptide discovery is that of phage display [191–193], which relies on the creation of cyclic peptide libraries (up to 10^{10} members) [184] through phagemids; virus-compatible plasmids that encode for a poly peptide attached to a viral coat protein. Transfection of *E. coli* with phagemid libraries and an appropriate bacteriophage leads to the excretion of phage particles with their coats displaying multiple copies of a single peptide from the library, which can then be collected and screened for affinity against a target. Peptide cyclisation is typically achieved through incorporation of multiple cysteine residues to enable disulfide formation or addition of a small molecule ‘zipper’ [194]. However, phage display is limited to the 20 proteinogenic amino acids and the interaction of displayed peptides can be interfered with by the viral coat proteins, with hits typically possessing target

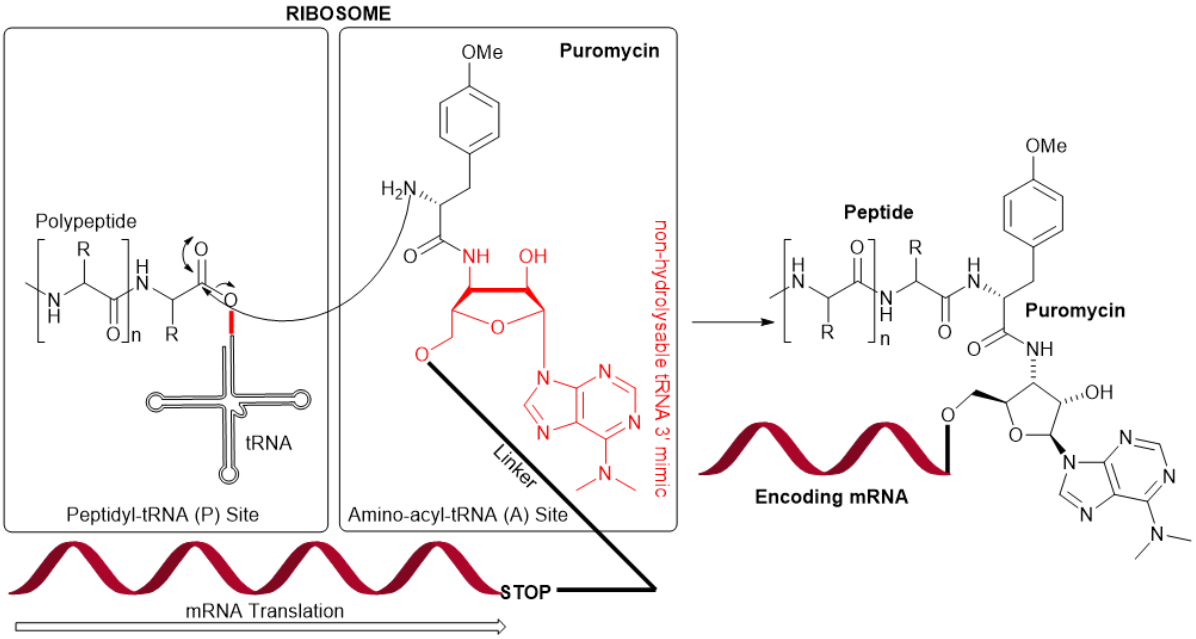
affinity in the high nanomolar to low micromolar regions. In addition, selection cycles are often time intensive. Nonetheless, it is a valuable technique and many cyclic peptides have been discovered this way with utility against membrane receptors [195], cancer targets [196] and in other drug development programs [194,197].

A further significant cyclic peptide discovery platform developed by Benkovic and Tavassoli is split-intein circular ligation of peptides and proteins (SICLOPPS) [198], which stands out for its ability to create cyclic peptide libraries within cells and couple the library selection to a functional readout [199]. SICLOPPS generates cyclic peptides through exploitation of the intein phenomenon: the ability of particular protein sequences (inteins) to excise themselves from their surrounding polypeptide (exteins). SICLOPPS-based selections function through creation of a split-intein-cyclic peptide fusion construct as a DNA sequence in a plasmid [200] which, once transfected into a target cell assay (eukaryotic or prokaryotic based), is translated to give a protein where the split-intein ends associate to form an active intein and the extein (cyclic peptide) sequence is released with a head-to-tail amide cyclisation (**F.0.7.ii.**). This approach carries caveats in that it relies on intracellular ribosomal synthesis of the cyclic peptide and therefore cannot normally incorporate unnatural amino acids, and the cyclisation mechanism places restrictions on the amino acids surrounding the cyclisation site; necessitating a cysteine which remains in the final cyclic peptide. SICLOPPS has been demonstrated to be an adaptable platform, as it can be coupled to a reverse two-hybrid system which elicits a phenotypic effect in response to disruption of a target protein-protein interaction (e.g. cell death/survival) [201]. Limiting the maximum cyclic peptide length to around six amino acids (10^6 member libraries) has yielded efficacious sequences with low micromolar affinity which are amenable to conversion into more classically ‘drug-like’ small molecules [202].



F.0.7.ii. The mechanism of peptide cyclisation for the SICLOPPS approach to cyclic peptide library generation and screening. A split-intein sandwiches a randomizable cyclic peptide sequence (extein), which associates to form an active intein capable of excising the extein and cyclising the peptide in the process. Typically, a cysteine residue is used to enable cyclisation and as a result remains present within the final cyclic peptide.

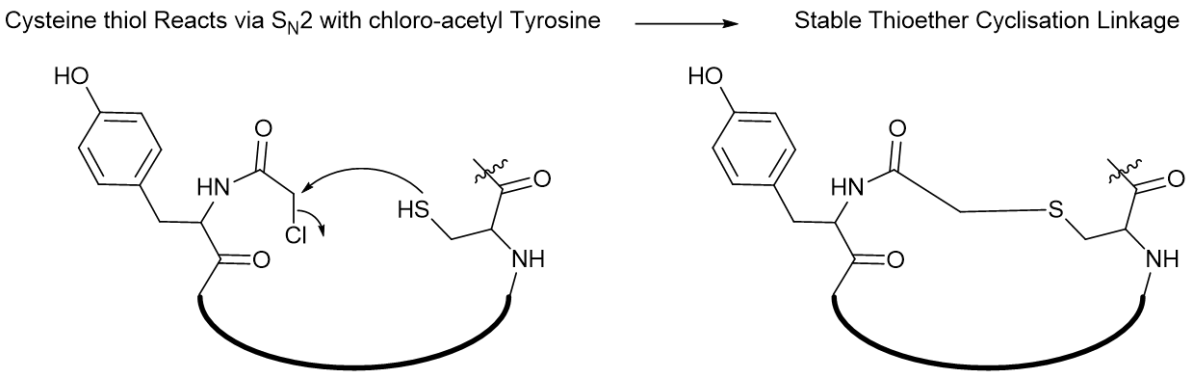
Cyclic peptide discovery can also be carried out entirely in-vitro in a cell-free environment. One of the most successful and advanced in-vitro screening platforms is mRNA-display – pioneered by Jack Szostak and others [203] – and makes use of the concept of coupling a ribosomally synthesised peptide to its encoding mRNA sequence [204]. This is usually accomplished by pre-ligating the mRNA to puromycin, an acyl-tRNA mimic, which can covalently link the peptide C-terminus to the mRNA sequence during ribosomal synthesis to ensure peptide sequence information can later be recovered (**F.0.7.iii**). Translation of the mRNA is accomplished in-vitro with a ribosomal 'kit' – the barebones extracted and purified components of the ribosomal machinery, often from *E. coli*, necessary for peptide synthesis.



F.0.7.iii. mRNA-peptide linkage utilised in mRNA display. Puromycin acts as a mimic of the 3' end of amino-acylated tRNA, with a non-hydrolysable amide linkage to the nucleoside, and can insert itself into the ribosome when no other acylated tRNAs are present. A puromycin-armed PEG-DNA linker is ligated to the mRNA prior to translation. The elongated peptide reacts with puromycin and becomes attached to its encoding sequence as part of the mRNA display selection process.

The Suga lab have developed an mRNA display platform coupled with Flexible In-vitro Translation (FIT), termed Random nonstandard Peptides Integrated Discovery (RaPID) [205]. This system is capable of generating highly customisable and impressively diverse cyclic peptide libraries (> 10^{12} members). The platform has arisen from the Suga lab's development of a flexizyme technology (Chapter 1.1.) to enable pre-acylation of specific tRNAs with pre-synthesised unnatural amino acids, which are in turn readily incorporated into peptides [170] through genetic reprogramming (FIT) [206,207]. The flexizymes have been used to generate cyclic peptide libraries by reprogramming initiator tRNA with chloro-acetyl tyrosine to spontaneously cyclise via S_N2 with a cysteine side chain in the polypeptide sequence (**F.0.7.iv.**). Since the selection process can be carried out entirely in-vitro, it is possible to complete one round of selection (library generation, panning against a target and enriched library recovery) within a day.

Typically, several rounds are required to generate libraries sufficiently enriched to contain cyclic peptides with low nanomolar to high picomolar affinity against their target [204], capable of achieving selectivity even between protein isoforms [208]. However, the selection process ordinarily can only guarantee binding affinity, with additional characterisation work required to uncover any functional effects. Even so, cyclic peptides originating from this platform have regularly been shown to possess potent inhibitory activities against a variety of challenging protein-protein interactions [209,210]. The ease with which large numbers of potential hits can be generated against diverse targets has facilitated the application of this technology across challenging drug targets in the pharmaceutical industry via PeptiDream, a spin-out company of the Suga lab. Further details of the mRNA display technology platform are expounded on in Chapter 1 and Chapter 2, as it forms the basis for cyclic peptide discovery in this project.



F.0.7.iv. Mechanism of peptide cyclisation commonly employed by Suga and co-workers for generation of cyclic peptides during mRNA-display.

0.8. Project Objectives

There is strong need for the development of chemical probes targeting challenging epigenetic protein families, to aid the study of epigenetics. The work in this thesis describes the identification and the development of cyclic peptides against the KDM7s and their PHD fingers.

The research discussed herein concerns the following objectives:

1) To use flexizyme technology to create cyclic peptide libraries for screening by mRNA display (Chapter 1)

2) To use flexizyme technology and chemical synthesis to expand the repertoire of relevant post-translationally modified amino acids that can be incorporated into peptides by in-vitro translation (Chapter 1)

3) To develop potent and selective cyclic peptide probes through mRNA display targeting the KDM7 sub-family (Chapter 2, 3, 4)

4) To demonstrate PHD fingers are tractable targets for epigenetic study by developing a potent and selective cyclic peptide probe targeting $KDM7_{[PHD]}$ (Chapter 2, 3, 4)

5) To use cyclic peptide probes to investigate the effect of the $KDM7_{[PHD]}$ on the $KDM7_{[JmjC]}$ catalytic demethylase activity (Chapter 3)

6) To investigate the mode of action and properties of any cyclic peptides found to interact primarily with the $KDM7_{[PHD]}$ (Chapter 5, 6, 7)

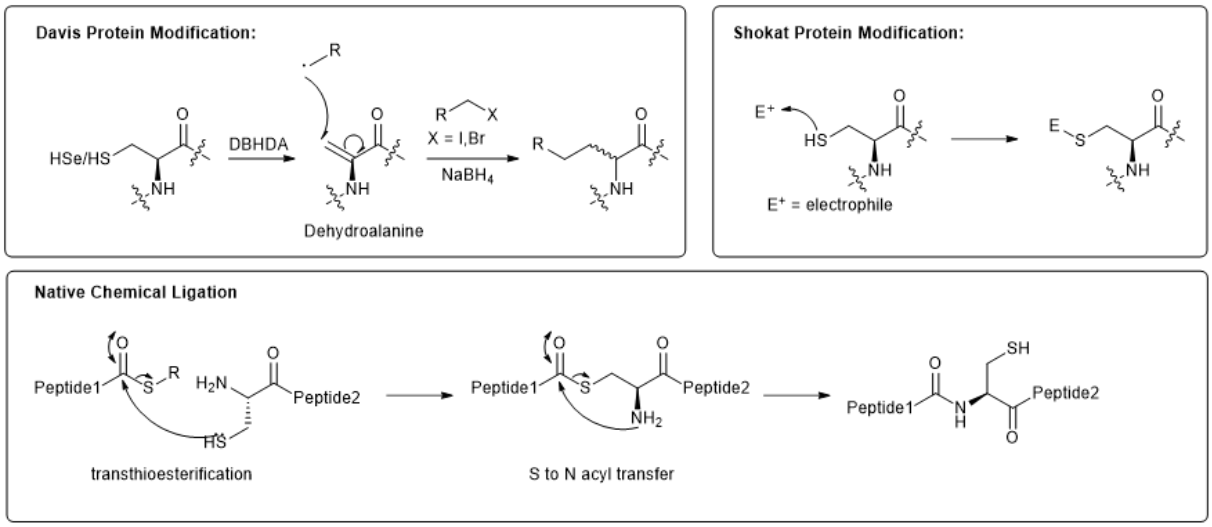
7) To demonstrate the applicability of such developed probes to studying the importance of the $KDM7_{[PHD]}$ to overall KDM7 function in a cellular context (Chapter 8)

Chapter 1. Cyclisation and Lysine PTMs in Peptide Libraries

1.1. Introduction

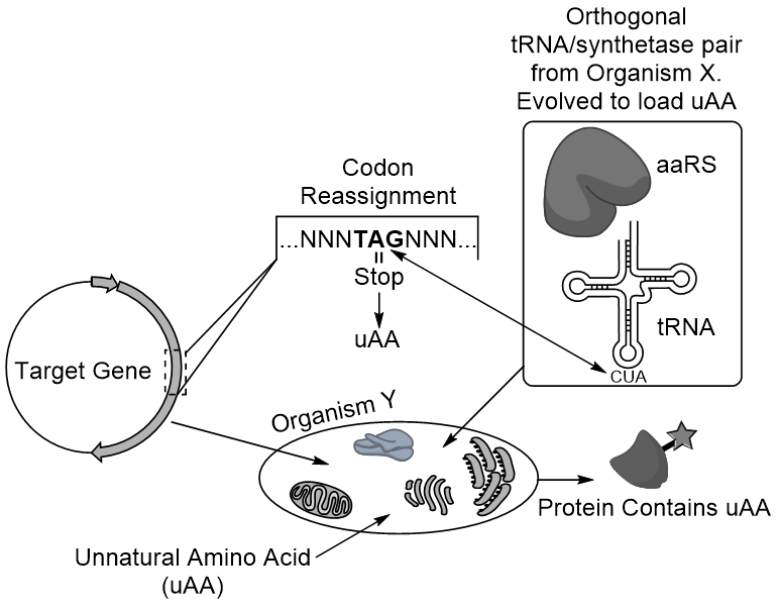
The ability to incorporate unnatural amino acids into peptides and proteins is valuable for creation of novel chemical entities with customised properties, and recreation of naturally occurring post-translational modifications (PTMs) to aid study of their function. Unnatural amino acids (uAAs) can be chemically synthesised to provide desired functionalities, but incorporating them into peptides and proteins via high-throughput ribosomal synthesis can be challenging. Several approaches have been developed to install unnatural functionality.

Chemical approaches have been used to circumvent the ribosome, which include post-ribosomal protein modification (**F.1.1.i.**). For example, a method reported by the Davis lab converts exposed (seleno)cysteine residues into dehydroalanine and reacts the latter with an appropriate nucleophile [211]. Their use of Luche-like radical chemistry was shown to permit installation of a large variety of PTMs with carbon-carbon sp^3 linkages, but with uncontrolled stereochemistry. This method is limited to installing as many modifications as there are ways to orthogonally install the dehydroalanine. The Shokat method involves cysteine alkylation [212], but is limited by potential simultaneous reaction of all exposed cysteines with an electrophile to create mimicked, and not exact, PTMs. Synthetic control of proteins is possible through native chemical ligation of peptides/proteins [213] and has been used successfully for many years, but often suffers from leaving chemical scars at joining sites and can sometimes require unreliable post-synthesis protein folding, all of which are time and resource intensive processes.



F.1.1.i. An overview of common methods for post-ribosomal protein modification. All of these methods are limited to modifications of primarily cysteine side-chains, and cannot completely recapitulate natural linkages.

Genetic reprogramming approaches that instead seek to make use of the ribosome for synthesis have commonly required the evolving of an orthogonal tRNA / amino acyl tRNA synthetase pair (e.g. pyrrolysyl tRNA synthetases) which rely on use of stop, or non-triplet, codons and expression in bacterial systems [214] (**F.1.1.ii.**). This is a labour-intensive process and is limited in the number and efficiency of achievable modifications.



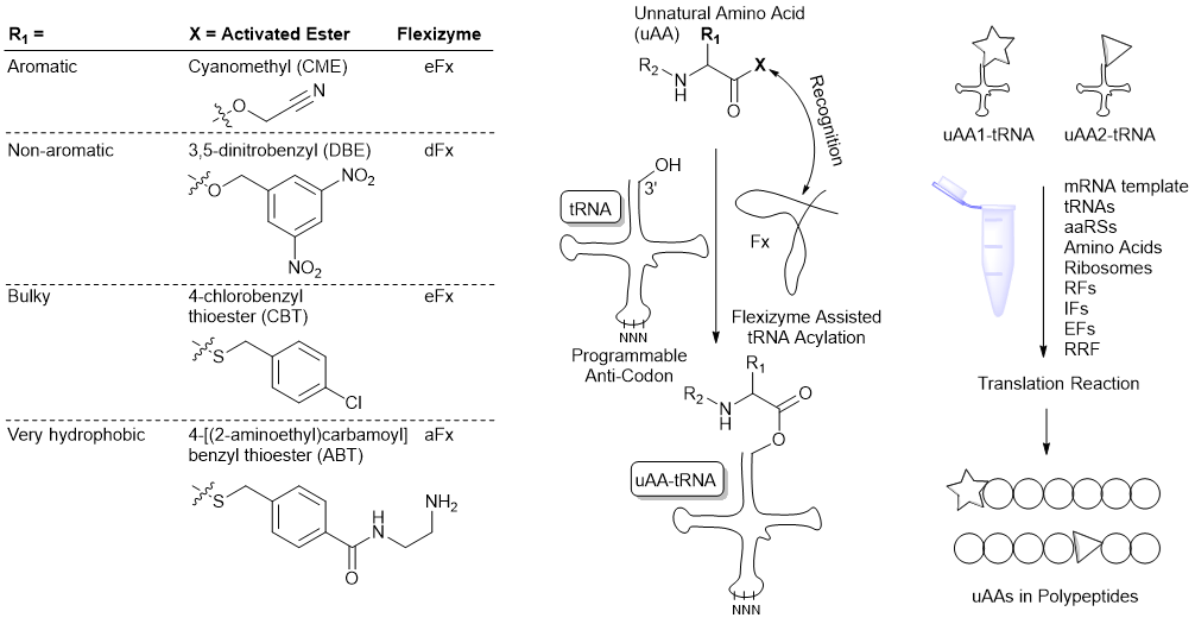
F.1.1.ii. An overview of the process involved in the use of orthogonal tRNA Synthetase pairs to express proteins containing unnatural amino acids in bacteria.

The Suga lab have pioneered the use of the flexizyme system (FIT) [206,215] which is capable of acylating isolated customised tRNAs with uAAs and incorporating them into peptides and proteins in cell-free in-vitro transcription/translation (IVTT) systems (making use of only the essential ribosomal machinery), assuming the uAA is compatible with passage through the ribosome. The flexizyme is a custom ribozyme (RNA enzyme), analogous to an amino acyl tRNA synthetase (aaRS) that has been evolved to accelerate acylation of an adapted tRNA with uAAs. The uAA is made as one of a select set of activated esters which is matched to its appropriate flexizyme (**F.1.1.iii.**). One or more tRNAs are then chosen for pre-acylation to enable their codons to become genetically reprogrammed to incorporate one or more uAAs in any subsequent in-vitro translation reaction, provided the codon's cognate amino acid is withheld because of its much higher natural incorporation efficiency. It has been demonstrated that it is theoretically possible to reassign all 64 codons to uAAs [216], however each additional uAA engenders progressive drop-offs in ribosomal efficiency which limits the maximum number of incorporable uAAs, but nonetheless this remains more effective than alternative methods. Moreover, the whole process from tRNA acylation to protein synthesis can be run within a day. This platform therefore enables high-throughput and high-quality site-specific incorporation of uAAs into peptides and proteins.

The flexizyme approach has proven exceptionally valuable for the rapid and reliable generation of large cyclic peptide libraries (>10^{12} members) for use in mRNA display, via incorporation of a chloro-acetyl amino acid residue at the initiator site which can react with a cysteine thiol (Chapter 0). These large cyclic peptide libraries are powerful for the development of potent ligands against challenging targets [208], which is pertinent to the objective of developing probes against the KDM7 sub-family and PHD fingers. The ability to use uAAs in cyclic peptides is further advantageous for customisation of physical properties

[217], loading of 'warhead' functionalities to exploit PTM mimicry [218] and covalent inhibition [219]. The applicability of this platform has been demonstrated for installation of a large variety of uAAs, reviewed elsewhere [170]. This includes expression of histone peptides that incorporate modified lysines with PTMs (methylation and acetylation) [220].

This Chapter describes the synthesis of uAA esters for acylating tRNA using flexizymes, and their applications in peptide cyclisation for mRNA display and site-specific uAA installation in peptides and proteins.

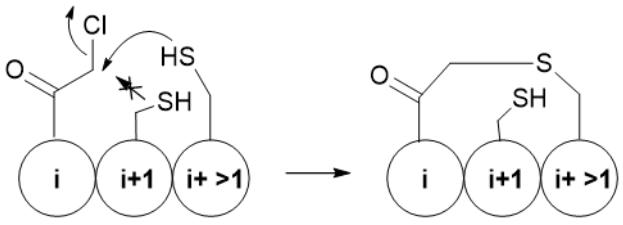


F.1.1.iii. The flexizyme system was developed by the Suga lab to enable acylation of individually reprogrammed tRNAs with suitably activated amino acids. Each amino acid activated ester is paired with a variant of the flexizyme ribozyme. Multiple tRNAs can be acylated to incorporate one or more unnatural amino acids into peptides and proteins during in-vitro ribosomal synthesis [206].

1.2. Making Cyclic Peptide Libraries

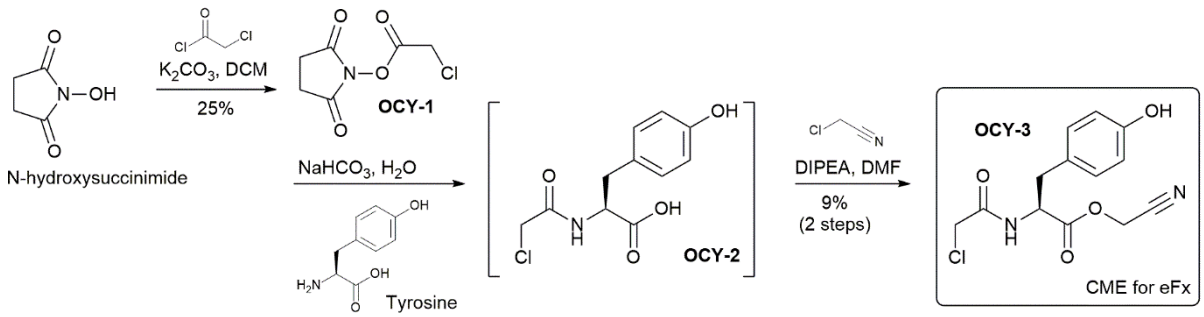
Making cyclic peptide libraries by in-vitro ribosomal translation is important in order to be able to screen them against protein targets by mRNA display. To achieve this, work by Suga and co-workers has established a system that reprograms the ATG initiator by omission of Met and introduction of pre-acylated initiator fMet-tRNA^{ATG} derivative with an N-chloro-acetylated amino acid compatible with flexizyme. Use of unnaturally acylated tRNA at the ATG initiator position results in higher uAA-incorporation efficiencies compared to incorporation at an elongator position, since the initiator position has greater tolerance of modifications compared to peptide elongation sites [170].

An N-terminal chloro-acetyl amino acid residue enables peptide cyclisation via spontaneous S_N2 reaction with another nucleophilic residue, preferentially a cysteine thiol, to irreversibly form a thioether linkage. The closest cysteine at any position from i+2 onwards can result in cyclisation (**F.1.2.i.**) [221]. Cyclic peptide library amino acid content is often randomised as part of a selection (Chapter 0,2), but fixing of a cysteine at the end of the randomised region enables the possibility of cyclisation for each peptide. It is also noteworthy that if only the initiator fMet-tRNA$_{ATG}$ is reprogrammed then any ATG codons subsequently located within the peptide sequence effectively function as stop codons, since the initiator fMet-tRNA cannot be used to elongate the peptide and methionine must be withheld from the translation reaction to prevent its greater native efficiency outcompeting the desired uAA incorporation [222,223].



F.1.2.i. A cysteine at the first residue in position i+ >1, where i is the chloro-acetylated residue, is able to cyclise the peptide effectively to form a thioether linkage.

Chloro-acetyl tyrosine (ClAcTyr) has been previously shown to function effectively as the N-terminal residue for cyclisation in this approach. Thus, the synthesis of the cyanomethyl ester of ClAcTyr was required for compatibility with eFx flexizyme to acylate tRNA (**F.1.2.ii.**). Initially, chloro-acetyl chloride was converted to its N-hydroxysuccinimidyl ester **OCY-1**, before reaction with unprotected tyrosine in the presence of base which favoured chloro acetylation of the nucleophilic amino group. The chloro-acetylated tyrosine **OCY-2** was telescoped immediately to the next step, with addition of chloroacetonitrile for esterification of the carboxylate. Column chromatography isolated the desired cyanomethyl ester **OCY-3** in low overall yield. This yield would benefit from optimisation, however the availability of all reagents made it a low-cost process with sufficient product still isolated for molecular biology / tRNA acylation purposes. The final compound **OCY-3** was dissolved as a 200 mM DMSO solution for long term (>2 years) storage at -20 °C.



F.1.2.ii. Synthesis of the chloro-acetyl tyrosine cyanomethyl ester for use with eFx flexizyme to acylate initiator tRNA $fMet_{ATG}$.

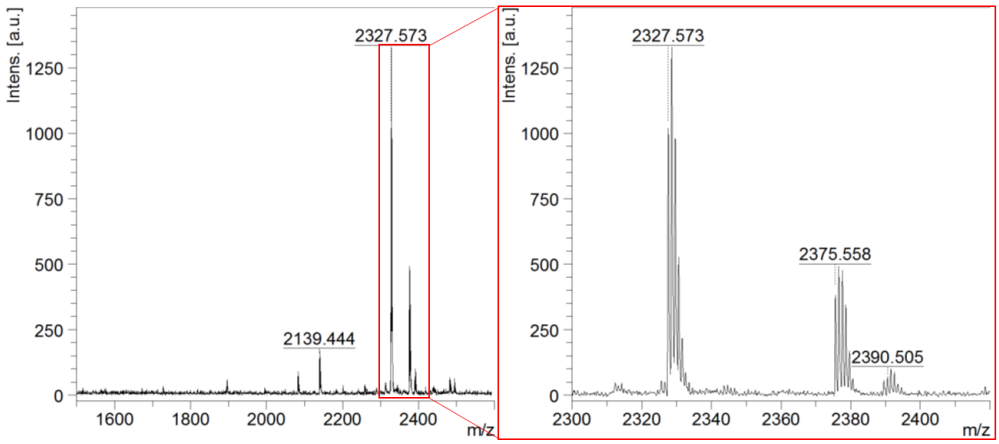
The initiator tRNA, $fMet_{ATG}$, (generated by PCR and transcription) was then acylated with the chloro-acetyl tyrosine cyanomethyl ester according to the protocol of Goto et al [224] (see Materials and Methods). Proof of acylation, and therefore of the ability to create cyclic peptides by in-vitro translation, was established by successful incorporation in an IVTT synthesis of a test peptide sequence from a known DNA template (**F.1.2.iii.**). The cyclic peptide was extracted from its IVTT reaction mixture by use of C_{18} ZipTip (Millipore) and analysis by MALDI-TOF-MS (see Materials and Methods). The test sequence, CP2, was a KDM4A,B,C inhibitory cyclic peptide derived from a previous mRNA display selection by Kawamura and co-workers [155]. The reprogramming of the initiator ATG codon meant that elongator ATG codons would be left available for testing incorporation of other uAAs. Expanding the available library of uAAs with modified lysines was desirable for their relevance to the study of epigenetics (Chapter 0) and represented a secondary objective for the project.

CP2 ATGGTGTATAATACGCGTTCTGGTTGGCGTTGGTATACGTGCGGCAGCGGCAGCGGCAGCTAG

CP2 ClAcYVYNTRSGWRWYTCGSGSGS*

- Initiator ATG is reprogrammed to incorporate chloro-acetyl tyrosine pre-loaded onto initiator tRNA $fMet_{ATG}$
- Cyclisation occurs between the chloro-acetyl group and the underlined cysteine residue's side chain
- GSGSGS is a linker sequence used in mRNA display

In-vitro Translation	Exact Mass [M+H]	
	Calculated	Observed (MALDI-TOF-MS)
CP2 (cyclised)	2326.993	2327.573



F.1.2.iii. Confirmation of cyclic peptide synthesis by IVTT. The initiator tRNA $fMet_{ATG}$ was pre-acylated with chloro-acetyl tyrosine, which was then successfully incorporated into an in-vitro translation of a test cyclic peptide sequence, CP2, which gave spontaneous cyclisation to its cysteine residue. Peptides were extracted from the in-vitro translation reaction mixture by C_{18} zip-tip and analysed by MALDI-TOF-MS. (* denotes stop codon, encoded by TGC).

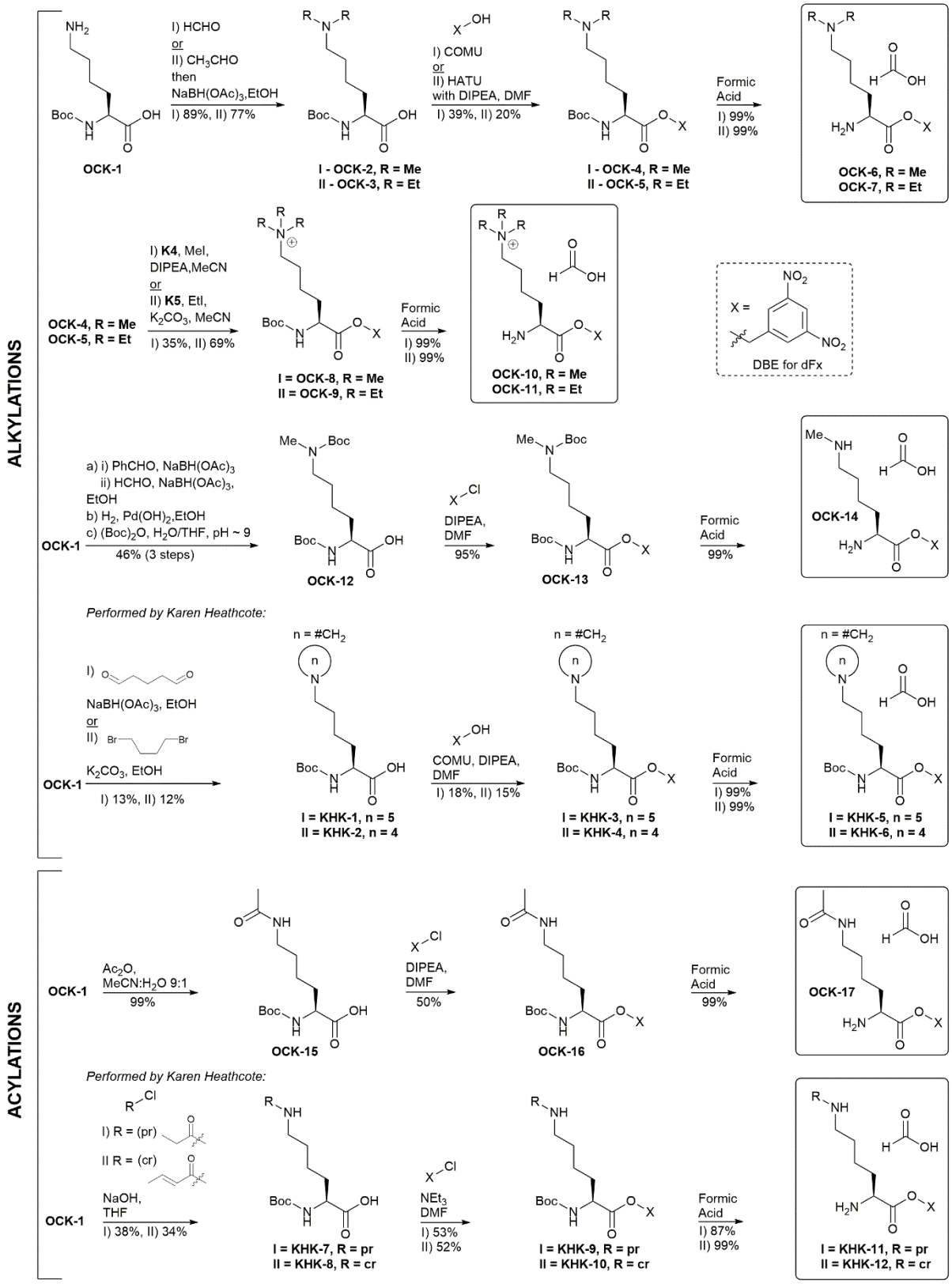
1.3. Synthesis of Modified Lysines for tRNA Acylation

Direct supervision was provided to Karen Heathcote (Part II MChem Student, University of Oxford) who synthesised and characterised the crotonoyl, propionyl, pyrrolidyl and piperidinyl lysines (denoted KHK).

The study of epigenetic regulatory proteins and their targets often requires fine control of PTMs to better understand their interactions. The number of biologically relevant PTMs being discovered is expanding [38], putting increasing demand on the ability to make and study them [225].

The KDMs and their reader domains are sensitive to changes in the site and degree of methylation of lysine residues on the histone tails (Chapter 0). The ability to precisely introduce methylated and other modified lysines in peptides and histones is highly valued for enabling study of their effect on enzymatic activities. Solid phase synthesis is routinely used to make short peptide fragments (up to ~ 40 amino acids) containing modified residues, but can be slow and inefficient. The FIT system has been successfully applied to creating histone H3 tail 36-mer peptides with single and multiple methylated lysines by the Suga lab to investigate epigenetic cross-talk [220]. Therefore, the FIT system provides a potential platform to rapidly make large libraries of PTM containing peptides and even full-length histones. This is especially important considering the combinatorial nature of histone PTMs, exemplified by the histone H3 sequence possessing at least 7 lysine sites known to commonly undergo a variety of modifications; the number of combinations possible for these positions with methylation states (4) alone is $4^7 = 16384$.

Several modifications of interest to lysine were targeted for synthesis (**F.1.3.i.**). The N^{ε} methylated (Kme1, Kme2, Kme3), acetylated (Kac), crotonoylated (Kcr) and propionylated (Kpr) forms of lysine were selected based on their biological relevance due to their naturally occurring presence in histones (**F.0.2.ii.**). The ability to customise the lysine residue with chemical synthesis prior to incorporation made it attractive to consider non-naturally occurring modifications, with a view towards exploiting PTM mimicry. Unpublished work within the Kawamura and Schofield groups by Roman Belle has identified several epigenetic reader domains with the capability of binding to non-naturally alkylated lysine (di or tri-ethyl at the N^{ε} position, aka Ket2 and Ket3) and there are reports of small molecules containing pyrrolidyl (Kpyr) or piperidinyl (Kpip) lysine motifs capable of inhibiting methyltransferases & demethylases [226] and methyl-lysine binding proteins [227]. Therefore, these modified lysines were also targeted for synthesis, with a view to future work that would seek to combine their useful PTM mimicry with the benefits of cyclic peptides for more targeted epigenetic probe development.



F.1.3.i. Synthetic transformations employed in the synthesis of the activated dinitrobenzyl ester derivatives of lysines, needed to match with dFx flexizyme for tRNA acylation. Various lysine derivatives were modified to contain natural post-translational modifications at the N^{ε} position, and non-natural mimics thereof.

All modified lysines were made as their activated 3,5-dinitrobenzyl esters to match with dFx for tRNA acylation (**F.1.3.i.**). All syntheses started from Boc-Lysine (**OCK-1**), with modification of the ε-NH_2 group using standard alkylation and acylation transformations as the first step. Late stage esterification of all compounds was designed to place the most effective purification by column chromatography prior to final deprotection, and minimise 'flask-time' of the base-sensitive activated ester to increase overall yield. Quaternization of the ε-amino position to make the trimethyl (**OCK-10**) or triethyl lysine (**OCK-11**) proceeded cleanly with the appropriate alkyl-iodide and could be performed after purification of the protected activated ester, thus reducing the number of steps involving problematic handling of the amphiphilic charged compound. Neat formic acid was employed to quantitatively remove Boc protecting groups at room temperature which mitigated against activated ester degradation and safety concerns associated with heating of high energy dinitrobenzyl groups. No final purification was necessary beyond vacuum removal of the formic acid, with the isolated final compound often an oil with an excess of formate – shown by NMR – ensuring the liberated α-NH_2 remained protonated to prevent intermolecular oligomerisation of the activated ester. Final compounds (**OCK-6,7,10,11,14,17** & **KHK-5,6,11,12**) were dissolved to a final concentration of 200 mM into DMSO as stock solutions for long-term storage at -80 °C. Full characterisation and final compound NMR spectra are given in the Materials and Methods and the Appendix.

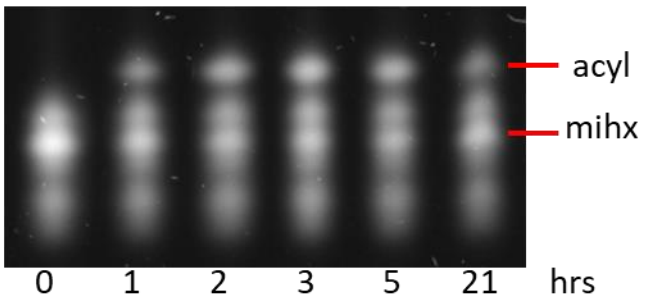
1.4. Incorporating Modified Lysines into Peptides

Testing of crotonoyl, propionyl, pyrrolidinyl and piperidinyl lysines and generation of the histone plasmid library was performed by Karen Heathcote (Part II MChem Student, University of Oxford) under supervision as part of the overall research project.

Incorporation of the modified lysines into peptides and histones first required acylation of tRNA. Conditions for tRNA acylation with the modified lysines were optimised using microhelix and dFx flexizyme [228], a smaller mimic of tRNA with the stem loop and 3' acylation site (see Materials and Methods). The smaller size of microhelix compared to tRNA (~7 kDa versus ~25 kDa) enabled acylated and un-acylated forms to be distinguished by acrylamide-urea gel, with acylation yield estimated by relative band intensities (**F.1.4.i.**). All of the synthesised lysines were able to acylate microhelix – albeit with relatively low yields, not unexpected given previous reports [206] (typically around 30 %).

Microhelix Lysine Acylation		
ε-N Modification	Optimised Time / hr	Yield %
me1	3	57
me2	3	43
me3	2	31
et2	3	9
et3	4	42
pyr	3	20
pip	3	22
ac	3	41
pr	3	30
cr	4	27

Example Microhelix Acylation Timecourse:
Kme1



$$\text{Yield \%} = 100x \frac{\text{relative intensity acyl}}{\text{relative intensity (acyl + mihx)}}$$

F.1.4.i. Acylation of microhelix RNA with modified lysines. A small tRNA mimic, microhelix, was used to test dFx flexizyme catalysed acylation with the dinitrobenzyl activated ester modified lysines. Each lysine acylation was quenched at different time points to give an optimised duration for the microhelix acylation, used to inform the length of time that should be used for acylation of full tRNA. An example of the microhelix acylation with Kme1 shows analysis by 20% acrylamide ACID-PAGE urea gel, imaged with SYBR Gold. Percentage yield was calculated using the relative intensities of the bands as determined by Image Lab (BioRad). N^{ε} modification refers to compounds synthesised in **F.1.3.i.**

Selected modified lysines were tested for incorporation into peptides. The cyclic peptide CP2 was used as a test sequence, with added relevance for modified lysines due to the finding of Kawamura and co-workers that substitution of CP2 arginine-6 with trimethyl lysine allowed CP2 R6Kme3 to function as a substrate for demethylation by KDM4A [155]. Incorporating differently modified lysines to this R6 position could therefore be of future interest in studying KDM4 JmjC catalytic activity promiscuity. Trimethyl, triethyl, acetyl, crotonoyl, propionyl, pyrrolidinyl and piperidinyl lysines were acylated to a glutamate elongator tRNA derivative, reprogrammed to the ATG codon (see Materials and Methods for sequence details), to enable their use alongside initiator-tRNA ClAcTyr-fMet$_{ATG}$ for incorporation into cyclic peptides.

The CP2 R6M DNA encoding sequence was used to provide the necessary ATG codon for recognition of the reprogrammed elongator tRNA-Glu$_{ATG}$. Preliminary IVTT syntheses with pre-acylated Glu-based tRNA had indicated a tendency towards glutamate leakage, with the un-acylated tRNA-Glu derivative picking up glutamate amino acid in solution and installing it at the reprogrammed codon. Therefore, glutamate amino acid was excluded from these IVTT syntheses. The modified lysines were subsequently successfully installed into the CP2 R6M cyclic peptide (**F.1.4.ii.**). In some cases, a minor peak with mass at +48 Da from the target mass was observed (see also **F.1.2.iii.**) – the cause of which is suspected to be a component of the IVTT mix reacting with the chloro-acetyl group due to the consistency of the mass difference between differently modified peptides.

CP2 ATGGTGTATAATACGCGTTCTGGTTGGCGTTGGTATACGTGCGGCAGCGGCAGCGGCAGCTAG

CP2 R6M ATGGTGTATAATACGATGTCTGGTTGGCGTTGGTATACGTGCGGCAGCGGCAGCGGCAGCTAG

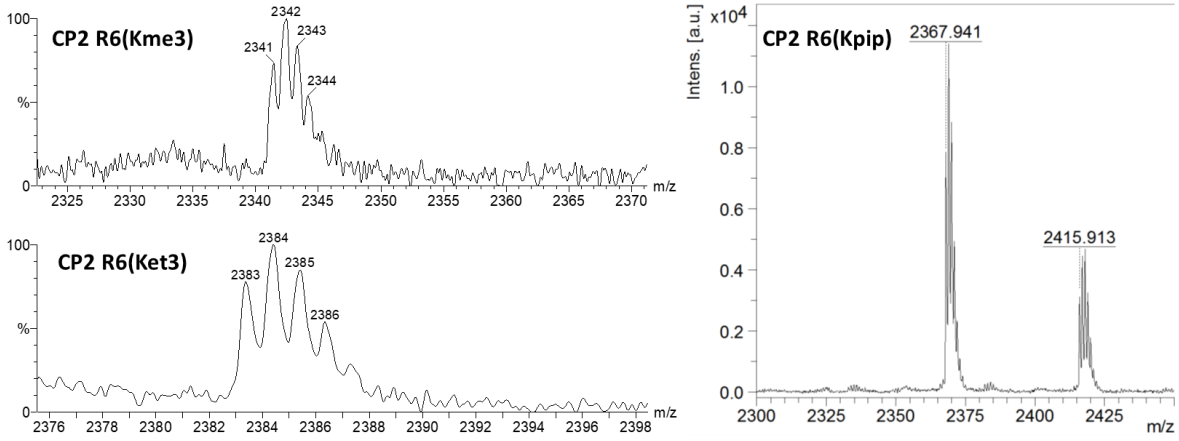
CP2 ClAcYVYNTRSGWRWYTCGSGSGS*

CP2 R6M ClAcYVYNTMSGWRWYTCGSGSGS*

- Initiator ATG is reprogrammed to incorporate chloro-acetyl tyrosine pre-loaded onto initiator tRNA $fMet_{ATG}$
- Cyclisation occurs between the chloro-acetyl group and the underlined cysteine residue's side chain
- Elongator ATG is reprogrammed to incorporate a modified lysine pre-loaded onto elongator tRNA Glu_{ATG}
- GSGSGS is a linker sequence used in mRNA display

In-vitro Translation	Peptide Sequence	Exact Mass [M+H]	
		Calculated	Observed (MALDI-TOF-MS)
CP2 (cyclised)	YVYNTRSGWRWYTCGSGSGS	2326.993	2327.573
CP2 R6M (linear, Met control)	formylMVYNTMSGWRWYTCGSGSGS	2258	2258
CP2 R6Kme3 (cyclised)	YVYNT(Kme3)SGWRWYTCGSGSGS	2341	2341
CP2 R6Ket3 (cyclised)	YVYNT(Ket3)SGWRWYTCGSGSGS	2383	2383
CP2 R6Kac (cyclised)	YVYNT(Kac)SGWRWYTCGSGSGS	2341	2341
*CP2 R6Kpr (cyclised)	YVYNT(Kpr)SGWRWYTCGSGSGS	2355.014	2355.982
*CP2 R6Kcr (cyclised)	YVYNT(Kcr)SGWRWYTCGSGSGS	2367.014	2368.076
*CP2 R6Kpyr (cyclised)	YVYNT(Kpyr)SGWRWYTCGSGSGS	2353.034	2354.021
*CP2 R6Kpip (cyclised)	YVYNT(Kpip)SGWRWYTCGSGSGS	2367.050	2367.941

Example MALDI-TOF-MS Spectra:



F.1.4.ii. A selection of the modified lysines were incorporated into a test cyclic peptide sequence, CP2 R6M. The initiator ATG was reprogrammed to chloro-acetyl tyrosine and the elongator ATG reprogrammed to the target modified lysine, using fMet and Glu derived tRNAs respectively. Peptides were extracted from crude IVTT reaction mixtures by C_{18} ZipTip and analysed by MALDI-TOF-MS. * = Performed by Karen Heathcote. See Appendix for all MALDI-TOF-MS spectra.

Other lysines (monomethyl, dimethyl and acetyl) were acylated to tRNA-Glu_{TGG} and demonstrated to be incorporable into a histone H3 peptide [1-21] with the K14 position reprogrammed to the TGG codon (cognate tryptophan) (**F.1.4.iii.**). The human histone H3 sequence does not otherwise contain TGG codons, making it useful for reprogramming uAA incorporation into H3 peptides as the tryptophan amino acid can be withheld without impact

on H3 synthesis integrity (**F.0.2.ii.**). However, it was still necessary to also remove glutamate from the translation reaction to maximise desired uAA incorporation. Histone peptides were extracted from their IVTT reaction mixture by SCX ZipTip (solid-phase support within a pipette tip for cation exchange purification) with analysis by MALDI-TOF-MS (see Materials and Methods). The formyl methionine initiating residue remained within the final peptide, however previous work (Cassandra Kennedey, Part II MChem, University of Oxford) demonstrated that addition of peptide deformylase (PDF) and methionine amino peptidase (MAP) post-translation was sufficient to generate the native N-terminal H3 sequence.

Further work by Karen Heathcote (Part II Thesis, University of Oxford, unpublished) has demonstrated single incorporation of modified lysines into a full-length histone H3 sequence (e.g. H3K4me3 & H3K9ac), with identification by Western Blot antibody detection.

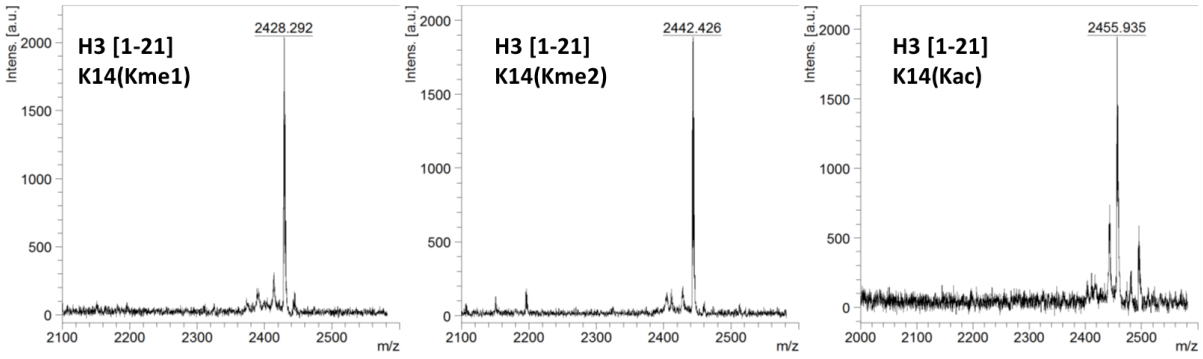
Histone H3 [1-21] ARTKQTARKSTGGKAPRKQLA*

Histone H3 [1-21] K14W ARTKQTARKSTGGWAPRKQLA*

- Lysine codon at position 14 is converted to TGG (tryptophan)
- Elongator TGG is reprogrammed to incorporate a modified lysine pre-loaded onto elongator tRNA-Glu_{TGG}
- The initiating formyl-Methionine (^{f}M) remains present in the IVTT translated peptide as no enzyme is present to remove it as would usually occur in-vivo

In-vitro Translation	Peptide Sequence	Exact Mass [M+H]	
		Calculated	Observed (MALDI-TOF-MS)
H3K14Kme1 [1-21]	^{f}MARTKQTARKSTGG(Kme1)APRKQLA	2427.373	2428.292
H3K14Kme2 [1-21]	^{f}MARTKQTARKSTGG(Kme2)APRKQLA	2441.388	2442.426
H3K14Kac [1-21]	^{f}MARTKQTARKSTGG(Kac)APRKQLA	2455.367	2455.935

Example MALDI-TOF-MS Spectra:



F.1.4.iii. IVTT synthesis of histone H3 peptide containing modified lysine residues. Modified lysines containing post-translational modifications relevant to epigenetics were installed to a histone H3 [1-21] peptide sequence by IVTT synthesis by reprogramming the K14 position codon to TGG achieving incorporation through pre-acylation of an elongator tRNA-Glu_{TGG}. Peptides were extracted from crude IVTT reaction mixtures by SCX ZipTip and analysed by MALDI-TOF-MS.

1.5. Conclusions and Future Work

The flexizyme system was used to establish reliable in-vitro ribosomal translation of cyclic peptides. Cyclisation relied on a chloro-acetyl tyrosine residue, an activated ester derivative of which was synthesised to enable acylation of initiator tRNA $fMet_{ATG}$ with eFx flexizyme, in-turn enabling reprogramming of the ATG initiator position. The chloro-acetyl group could then cyclise with a downstream cysteine thiol residue. This approach to making cyclic peptides is applicable to the generation of large (>10^{12} member) libraries for mRNA display which will be used for development of cyclic peptides against the KDM7 sub-family (Chapter 2 and forwards).

Progress was made towards aiding development of a platform for study of epigenetic post-translational modifications. A set of modified lysine activated esters, containing biologically relevant natural and non-natural PTMs, were synthesised and shown to acylate a microhelix mimic of tRNA. Several lysines (Kme3, Ket3, Kac, Kcr, Kpr, Kpyr, Kpip) were successfully incorporated alongside chloro-acetyl tyrosine by IVTT synthesis into a cyclic peptide test sequence, CP2 R6M, with possible relevance to future study of KDM4 JmjC interactions. Additional modified lysines (Kme1, Kme2, Kac) were singly incorporated into a shortened histone H3 peptide [1-21] and full-length H3 sequence by IVTT. Future work would seek to address issues surrounding the efficiency of reprogramming glutamate derived tRNAs, with a need to minimise undesired glutamate amino acid incorporation. Future work would also further develop the system to maximise efficiency of multiple uAA incorporations, particularly concerning differently modified lysines, to rapidly synthesise large and diverse histone PTM libraries to facilitate study of epigenetic cross-talk on enzymatic activities.

Chapter 2. Cyclic Peptide Selection against KDM7B

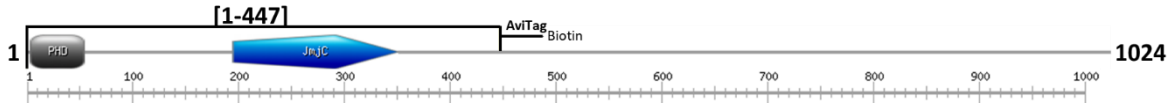
2.1. Introduction

The primary objectives of the project included generating potent ligands for the KDM7 sub-family, and more specifically to its PHD finger. Previous small molecule efforts had failed to deliver either potency or selectivity (Chapter 0), and so a new approach was required. The mRNA display cyclic peptide discovery platform (Chapter 0, 1) provides a powerful, adaptable and cutting-edge technology – that is an alternative to small molecule high-throughput screening – to pursue development of a cyclic peptide probe for this challenging target.

Of the three KDM7 sub-family members, KDM7B possessed an interesting reliance on epigenetic cross-talk, as its PHD finger binding to H3K4me3 is reported to be near essential to its JmjC H3K9me2 demethylase activity [65]. KDM7B therefore provided an excellent target for the development and testing of both general KDM7 and $KDM7_{[PHD]}$ specific probes, as any PHD finger inhibition could hopefully be linked to an additional catalytic functional readout beyond disruption of the protein-protein interaction with cognate H3K4me3. The $KDM7B_{[PHD+JmjC]}$ construct (**F.2.1.i.**) was used for mRNA display selection. The $KDM7_{[PHD]}$ by itself is only 7 kDa, and such small proteins have proven problematic for selection (unpublished work within the Kawamura lab). However, $KDM7B_{[PHD+JmjC]}$ is ~55 kDa, and by immobilising its C-terminus provides both a target for selection and a scaffold to present the $KDM7_{[PHD]}$ more effectively in solution for potential binding to peptides during the mRNA display selection process.

Selection against the $KDM7B_{[PHD+JmjC]}$ target would not control for the location of any cyclic peptide binding. Any hits resulting from selection would require further characterisation to identify the ligand binding site.

KDM7B$_{[PHD+JmjC]}$-biotin construct for mRNA Display Selection:

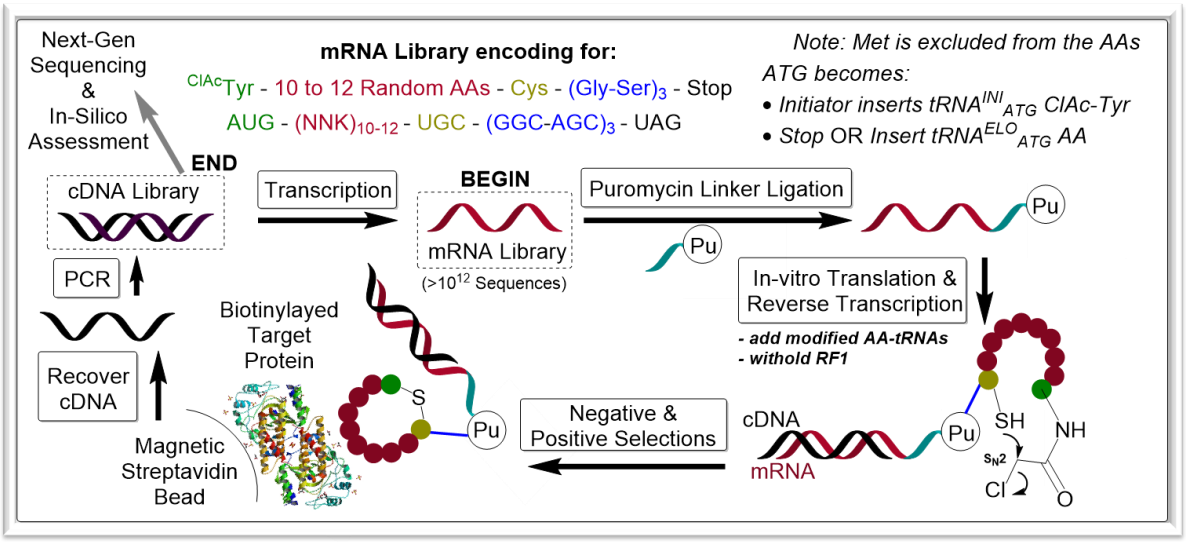


F.2.1.i. The domain map for KDM7B is shown. The full-length KDM7B contains a PHD finger and JmjC domain towards its N-terminus. The protein region covered by the KDM7B$_{[PHD+JmjC]}$ construct (AAs 1-447) used for mRNA display selection is highlighted, containing both the PHD and JmjC domains with additional amino acids towards the C-terminus. A C-terminal AviTag was added to enable biotinylation during production. Image generated using ExPASy PROSITE 'MyDomains' and information for KDM7B isoform 2 (Q9UPP1, UniProt).

2.2. Design of the mRNA-Display Selection

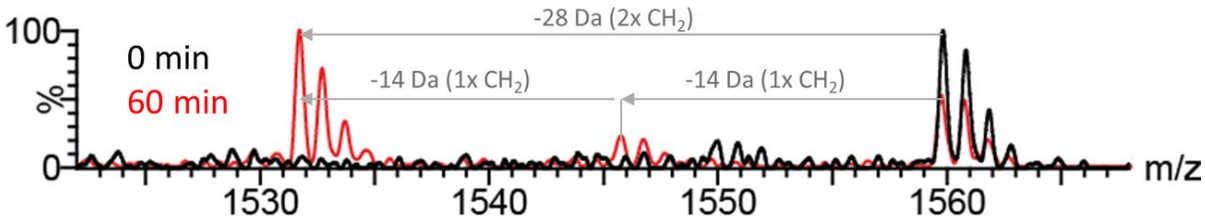
A key feature of the mRNA display platform is the ability to customise the cyclic peptide library composition. Given the challenging nature of the PHD finger containing KDM7 target, it was deemed likely that longer peptide sequences would be needed to maximise the opportunity for finding binding sequences. Previous work within the group indicated that library compositions consisting of a wide range of peptide lengths tended to result in the biased enrichment of the longer peptides. Thus, a longer length of peptide library with a tighter variable range (10, 11 or 12 amino acids) was designed (**F.2.2.i.**). The variable region content was randomised at each position by use of an NNK codon, permitting any of the four bases for N, and either G or T for K. The NNK codon is commonly employed in 'saturation mutagenesis' to enable coverage of all 20 amino acids with use of only 32 codons and a 3.1% probability of a stop codon (TAG only), as opposed to NNN providing 64 codons with a 4.7% probability of a stop codon (TAG, TAA, TGA). The use of NNK also helps to reduce, but not eliminate, the codon bias of the library as some amino acids are encoded by multiple codons and may therefore occur more frequently through simple statistical inclusion instead of genuine benefit to binding.

The initiator codon, ATG, was reprogrammed to incorporate chloro-acetyl tyrosine instead of methionine. An initiator-tRNA $fMet^{ATG}$ derivative was pre-acylated using flexizyme (Chapter 1) and methionine was excluded from the translation reaction. A cysteine residue was fixed at the C-terminus of the variable region, followed by a GSGSGS linker to puromycin. This meant the cyclic peptide library would consist mostly of natural amino acids, apart from the chloro-acetyl tyrosine initiator, and any Met codons in the variable region would either function as a stop codon, cause a mis-incorporation or likely be a mutation from PCR library amplification.

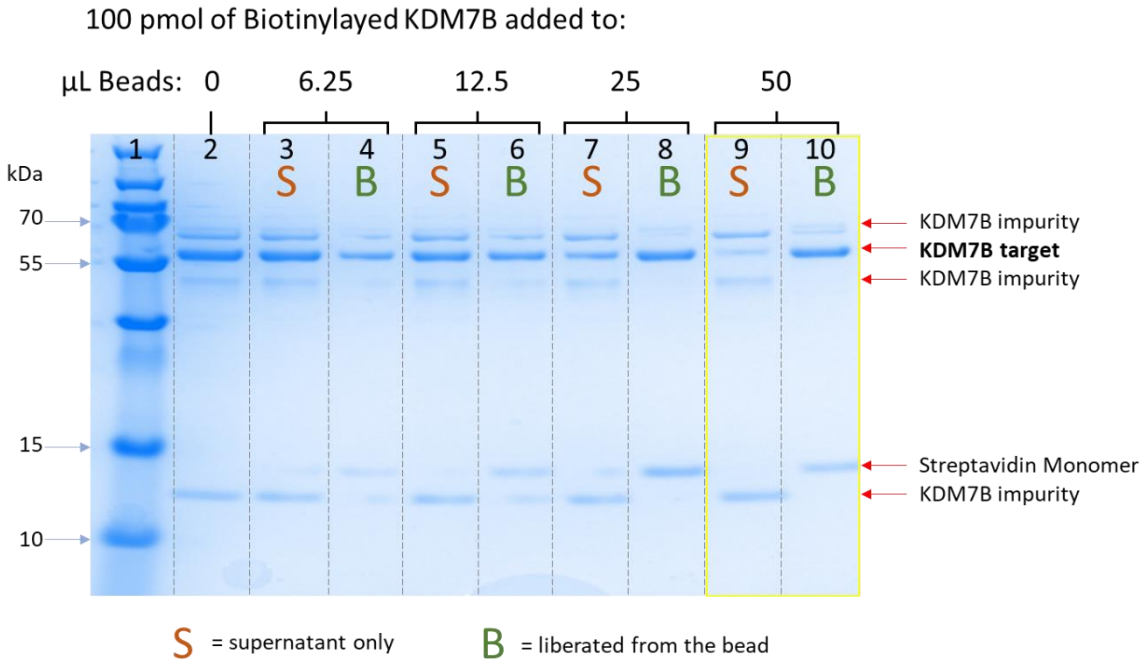


F.2.2.i. The mRNA display library design and selection process. The library consisted of a variable region of 10-12 random amino acids, flanked by a reprogrammed ATG-initiator codon to incorporate chloro-acetyl tyrosine and a fixed cysteine residue to ensure cyclic thioether formation was possible. Methionine was excluded from the translation, meaning any non-initiator ATG codons would be most likely to result in early termination of the peptide. A 'GSGSGS' linker sequence provided space between the peptide and the site of mRNA-puromycin ligation. Five rounds of selection were performed.

For the selection to proceed, the protein target had to be appropriately immobilised on magnetic beads. KDM7B$_{[PHD+JmjC]}$ was produced recombinantly in *E. coli* with a C-terminal 15 residue AviTag [229] (plasmid was donated by Finn Wolfreys, Brennan Lab) – which allows site-specific biotinylation by co-expressing biotin ligase during production. This enabled use of the extremely tight (sub-nanomolar affinity) naturally occurring biotin-streptavidin interaction for immobilisation. Demethylase activity of the protein was confirmed via a MALDI-TOF-MS based assay (**F.2.2.ii.**), indicating that the PHD and JmjC domains were appropriately folded. The capability of this KDM7B$_{[PHD]}$ containing construct to bind its cognate H3K4me3 partner whilst immobilised, and by implication its retention of proper folding, was confirmed with BioLayer Interferometry (BLI) (*vide infra* 2.5.). Prior to use in the selection, the quantity of the biotinylated-KDM7B$_{[PHD+JmjC]}$ needed for loading onto the magnetic streptavidin beads was optimised (**F.2.2.iii.**). This established the correct ratio of beads to protein necessary to achieve the desired concentration of target during the selection.



F.2.2.ii. Demethylation Activity of Biotinylated KDM7B$_{[PHD+JmjC]}$. The enzyme (4 µM) was incubated with H3K4me3**K9me2** [1-14] (20 µM) substrate with 10 µM Fe(II), 100 µM 2OG, 100 µM ascorbate in 40 mM HEPES pH 7.5 with 200 mM NaCl at 37 °C. MALDI-TOF-MS analysis showed the loss of two CH_2 units (14 Da each) over time, as expected for two-demethylation events associated to the well-established H3K9me2 target mark, confirming that KDM7B catalytic activity was retained after biotinylation.

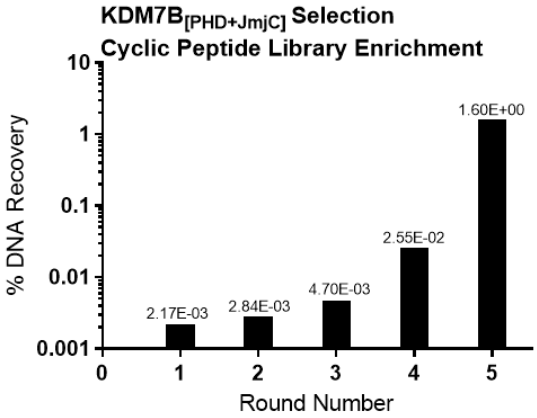


F.2.2.iii. Optimisation of magnetic bead loading with biotinylated KDM7B$_{[PHD+JmjC]}$. The ratio of magnetic streptavidin coated beads to biotinylated KDM7B$_{[PHD+JmjC]}$ protein quantity was optimised to maximise loading of the target protein. 50 µL of beads were found to be sufficient for nearly complete binding of biotinylated protein in solution (100 pmol), making the final optimal bead loading ratio: 2 pmol of protein per 1 µL of beads.

The preparation and sourcing of all other selection materials and the procedure for carrying out the selection itself were optimised within the Kawamura lab and are described in more detail in the Materials and Methods. Briefly, the puromycin-ligated mRNA library was translated in-vitro with use of a PURExpress In-Vitro Protein Synthesis kit (ΔAAs ΔtRNAs ΔRF1,2,3, New England Biolabs) and addition of chloro-acetyl tyrosine-tRNA, RF1 and amino acids minus methionine. Reverse transcription non-covalently partnered and stabilised every mRNA sequence with its cDNA sequence. Three negative selection steps against streptavidin-magnetic beads alone removed non-specifically binding cyclic peptides, and one positive selection step against the immobilised $KDM7B_{[PHD+JmjC]}$ target was followed by stringent washing of the beads to remove weakly target-binding peptides. DNA sequences of cyclic peptides binding to $KDM7B_{[PHD+JmjC]}$ were recovered by heat elution, quantified by qPCR, PCR amplified and transcribed into mRNA to begin a new round of selection with an enriched library (**F.2.2.i.**).

2.3. KDM7B Selection Results

The selection against KDM7B$_{[PHD+JmjC]}$ was performed over five rounds to progressively enrich the library, with the percentage recovery of DNA quantitated by qPCR and recorded from round to round (**F.2.3.i**). As binding peptides are enriched, they are expected to represent a greater proportion of the library and hence the proportion of DNA recovered compared to the input is expected to increase. Therefore, the jump to 1% DNA recovery after round five indicated substantial enrichment of the library with binding sequences, and so in-depth next-generation sequencing (NGS) was performed to identify them. The raw NGS data was processed with a proprietary software script developed by Dr Ed Hookway to provide a list of peptides and a count of their occurrence. A further analysis developed by Dr Tom McAllister was used to rank the peptide sequences by their total proportion of the overall library (**F.2.3.ii.**). One immediately noticeable result was the presence of some sequences with a variable region longer than the predetermined 10-12mer size. These originated from stop-codon readthrough, via a mutation or frameshift in the DNA sequence. Since the frequency of these longer mutated peptides was low compared to unmutated peptides, they were discounted from further analysis and candidacy for future synthesis given the likelihood of their possessing greater handling difficulties and undesirable physical properties.



F.2.3.i. Cyclic peptide library enrichment during selection against KDM7B$_{[PHD+JmjC]}$. The round-to-round recovery of the cyclic peptide DNA library from the positive selection against KDM7B$_{[PHD+JmjC]}$ is shown. A significant 100-fold enrichment in recovery is seen at round 5, suggesting a substantial proportion of the library encoded for peptides that bound to the target. The percentage of DNA recovery is relative to the amount of input DNA.

	Total Peps:	5000													
	total reads	668875	Enrichment	Length											
Rank	rawPeptide	Freq	%age												
1	PVLLVYPRRLI	105814	15.531	11	P	V	L	L	V	Y	P	R	R	L	I	
2	VYGTYYLSWIP	70396	10.332	11	V	Y	G	T	Y	Y	L	S	W	I	P	
3	CPLYIWPRTSLP	41716	6.123	12	C	P	L	Y	I	W	P	R	T	S	L	P
4	LVRIYPNKLIWW	30548	4.484	12	L	V	R	I	Y	P	N	K	L	I	W	W
5	CLIETRALYP	28374	4.165	10	C	L	I	E	T	R	A	L	Y	P		
6	PYILYPSVYTTN	26183	3.843	12	P	Y	I	L	Y	P	S	V	Y	T	T	N
7	GSFYVAFACLRQRQRQLGRGAE	22324	3.277	22	G	S	F	Y	V	A	F	A	C	L	R	Q
8	YLTHPYPSVFRT	20679	3.035	12	Y	L	T	H	P	Y	P	S	V	F	R	T
9	IVPSPIWYHWVA	19726	2.895	12	I	V	P	S	P	I	W	Y	H	W	V	A
10	CPLFLYPRQVIP	15959	2.342	12	C	P	L	F	L	Y	P	R	Q	V	I	P
11	CDILTRLEYR	13536	1.987	10	C	D	I	L	T	R	L	E	Y	R		
12	CPLFLNPRIHLP	12990	1.907	12	C	P	L	F	L	N	P	R	I	H	L	P
13	LHSNCFTFSALRQRQRQLGRGAE	11372	1.669	23	L	H	S	N	C	F	T	F	S	A	L	R
14	CLISTRLAYP	11190	1.642	10	C	L	I	S	T	R	L	A	Y	P		
15	GNTLCVYQVPCT	10899	1.600	12	G	N	T	L	C	V	Y	Q	V	P	C	T
16	VYKSYPTVFKTN	10686	1.568	12	V	Y	K	S	Y	P	T	V	F	K	T	N
17	PVLLIYPRRLI	9061	1.330	11	P	V	L	L	I	Y	P	R	R	L	I	
18	CLIATRVVYP	7846	1.152	10	C	L	I	A	T	R	V	V	Y	P		
19	ILLNQRNLLWP	6365	0.934	11	I	L	L	N	Q	R	N	L	L	W	P	
20	FYRKFPSVYKT	6038	0.886	11	F	Y	R	K	F	P	S	V	Y	K	T	

F.2.3.ii. Analysis of Next-Generation-Sequencing (NGS) output of recovered DNA from the 5th round of selection. The NGS results for the library at the end of the fifth-round selection against KDM7B$_{[PHD+JmjC]}$ are shown, with the top 20 peptide variable region sequences ranked by their % of the total peptide population. KEY: Freq = number of NGS reads. Enrichment = % of overall library. Length = length of variable region. Only the peptide variable region is shown, each sequence also has a fixed Tyr (Y) *N*-terminus and Cys (C) *C*-terminus.

The choice of hit sequences for synthesis and validation was primarily considered on their percentage enrichment as a proportion of the whole library, but the abundance of data that came with the NGS results presented opportunity for additional analysis. Multiple hit sequences had an enrichment of 1-10 %, with none greater than 16 %. Progressive rounds of enrichment might eventually be expected to push the library towards consensus sequences for different binding sites, but in doing so reduce the hit diversity. Sequencing the library at the first major enrichment step retained diversity in the hit sequences to improve the chances of discovering multiple peptides with different binding modes, as the selection against KDM7B$_{[PHD+JmjC]}$ did not control for the location of cyclic peptide binding. The diversity of the hit sequences was made clear by extracting peptide 'families' from the NGS data. Large library selections of this nature can sometimes result in hit sequences becoming dominated by peptides that are closely related, either by substitutions or common motif. A computational analysis (developed by Dr Tom McAllister) categorised the relatedness of every hit cyclic

peptide sequence based on a Hamming distance (number of allowable variable positions). All mutant peptides with variable regions >12 amino acids were excluded from analysis. Amongst the remaining hit peptides, 8 of the top 10 enriched sequences by NGS were classed to belong to different families at Hamming distance 6 (set to allow up to 6 substitutions within a peptide family, which was half of the maximum variable region size). Many amino acids share similar properties (e.g. isoleucine and leucine) which can be exploited in so-called 'fuzzy-matching' to account for like-for-like substitutions, but even once this was applied 8 of the top 10 hits were classifiable as different families (**F.2.3.iii.**). The top 8 most enriched cyclic peptide sequences were selected for solid-phase peptide synthesis (SPPS) and binding validation.

Rank	Peptide	Count	Proportion	Peptide Length	Family No. Hamming_6	Distance from Parent	FUZZY Family No. Hamming_6	FUZZY Distance from Parent
1	PVLLVYPRRLI	105814	0.155308256	11	1	0	1	0
2	VYGTYYLSWIP	70396	0.103323568	11	2	0	2	0
3	CPLYIWPRTSLP	41716	0.061228564	12	3	0	3	0
4	LVRIYPNKLIWW	30548	0.044836757	12	4	0	4	0
5	CLIETRALYP	28374	0.041645874	10	5	0	5	0
6	PYILYPSVYTTN	26183	0.038430038	12	6	0	6	0
8	YLTHPYPSVFRT	20679	0.030351555	12	8	0	8	0
9	IVPSPIWYHWVA	19726	0.028952791	12	9	0	9	0
10	CPLFLYPRQVIP	15959	0.023423786	12	3	6	1	5
11	CDILTRLEYR	13536	0.019867433	10	5	5	5	5
12	CPLFLNPRIHLP	12990	0.019066043	12	3	5	3	3
14	CLISTRLAYP	11190	0.016424097	10	5	3	5	3
15	GNTLCVYQVPCT	10899	0.015996982	12	11	0	11	0
16	VYKSYPTVFKTN	10686	0.015684352	12	6	6	6	5
17	PVLLIYPRRLI	9061	0.013299262	11	1	1	1	0
18	CLIATRVVYP	7846	0.011515949	10	5	3	5	2
19	ILLNQRNLLWP	6365	0.009342214	11	12	0	12	0
20	FYRKFPSVYKT	6038	0.008862261	11	6	6	6	5
21	CTLPKATTRLVL	5944	0.008724292	12	13	0	13	0
22	FKTLYPSVWKT	5795	0.008505598	11	6	6	6	6
23	LLTYSHHRELTP	5707	0.008376436	12	14	0	14	0
24	YPYSKYPNVYKT	5664	0.008313323	12	15	0	8	5
25	LIRYANVHLHFA	5629	0.008261952	12	16	0	15	0
26	WEFLPTLCLYIT	5120	0.007514868	12	17	0	16	0
27	RYALFLNVFKT	5109	0.007498723	11	18	0	6	6
28	YLLIHSSVWRT	4920	0.007221319	11	19	0	17	0

Key

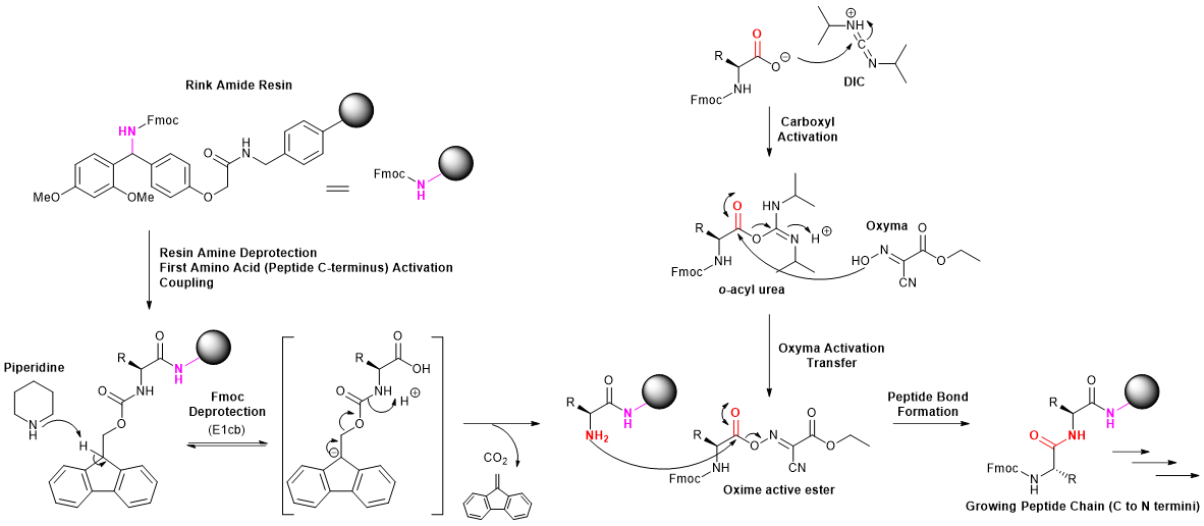
= Family Leader in exact and fuzzy analysis
= Family Leader in exact analysis and not in fuzzy
= Sub-Member of a Family

F.2.3.iii. Hit cyclic peptide families classification from KDM7B selection. Hit cyclic peptide variable region sequences were grouped into families based on their relatedness (Hamming distance = 6). KEY: Count = number of NGS reads. Proportion = % of the overall library. Peptide length = length of variable region. Family no. = family identifier based on exact matching of amino acids. Distance from parent = number of amino acids changed compared to family parent sequence. Fuzzy = family classification accounting for amino acids with similar properties being treated as equivalent (E:D, I:L, K:R, Q:N, V:L, Y:F).

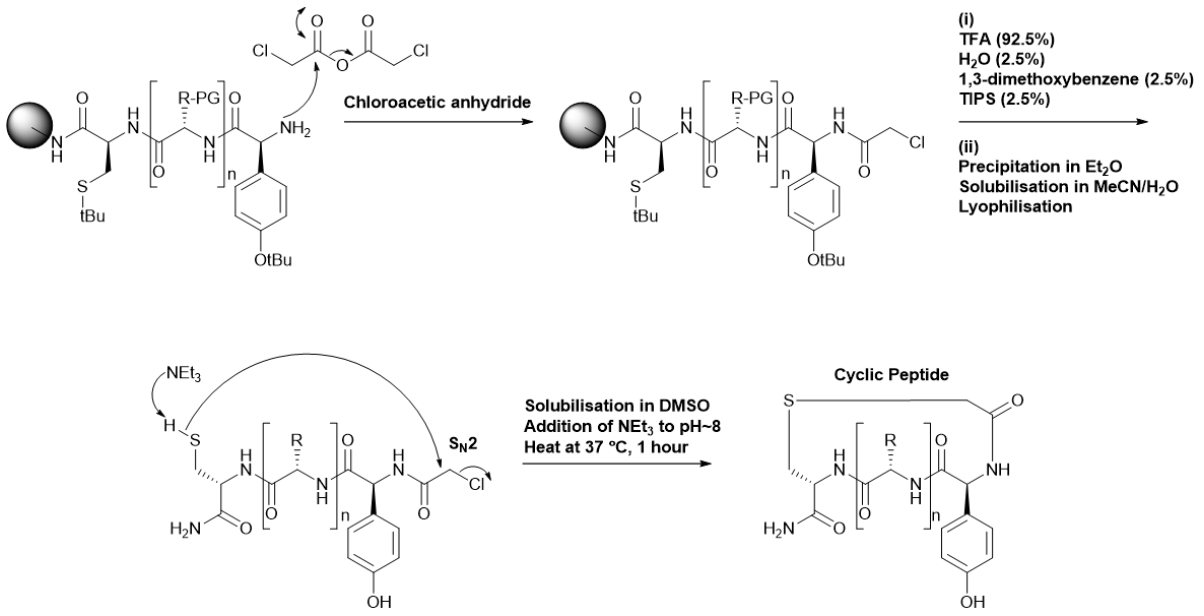
2.4. Solid-Phase Synthesis of Cyclic Peptides

The cyclic peptides hit sequences were synthesised on a 0.05 mmol scale using traditional Fmoc-based SPPS (**F.2.4.i.**) on a Liberty Blue machine. DIC was used as the coupling agent, with Oxyma used as an activating base to minimise α-CH stereocentre racemisation. Oxyma undergoes an acyl exchange with the carbodiimide activated amino acid, to form the activated oxime ester, which achieves a favourable balance of reactivity towards intermolecular peptide bond formation over possible intra and intermolecular side reactions that could lead to racemisation (e.g. azlactone formation or enolization [230,231]).

The N-terminal chloro-acetyl modification was added post-synthesis to the Rink amide resin-bound peptide, with subsequent deprotection and liberation followed by cyclisation at pH ~8 (**F.2.4.ii.**). The SPPS derived cyclic peptides differed from their mRNA-display counterparts as the common GSGSGS linker was not included and the cysteine C-terminus ended as a primary amide. MALDI-TOF-MS was used to confirm the success of each synthetic stage, and each cyclic peptide was purified by HPLC for hit validation. OC3, OC4, OC6, OC8, OC9 and OC10 were made successfully; however, OC5 and OC7 were not able to be made in sufficient quantity/purity for testing at the time. Testing was progressed with those that could be made and purified. Full SPPS and HPLC protocols are available in Materials and Methods.



F.2.4.i. The mechanism for peptide chain growth during the deprotection, activation and coupling cycle of Fmoc based solid-phase-peptide-synthesis (SPPS).



F.2.4.ii. The resin-bound side-chain-protected linear peptides resulting from solid-phase synthesis were reacted with chloroacetic anhydride through their unprotected N-terminal amino group. Subsequent cleavage from the resin and removal of side-chain protecting groups enabled cyclisation with the cysteine thiol under basic conditions. All cyclic peptides were purified by HPLC.

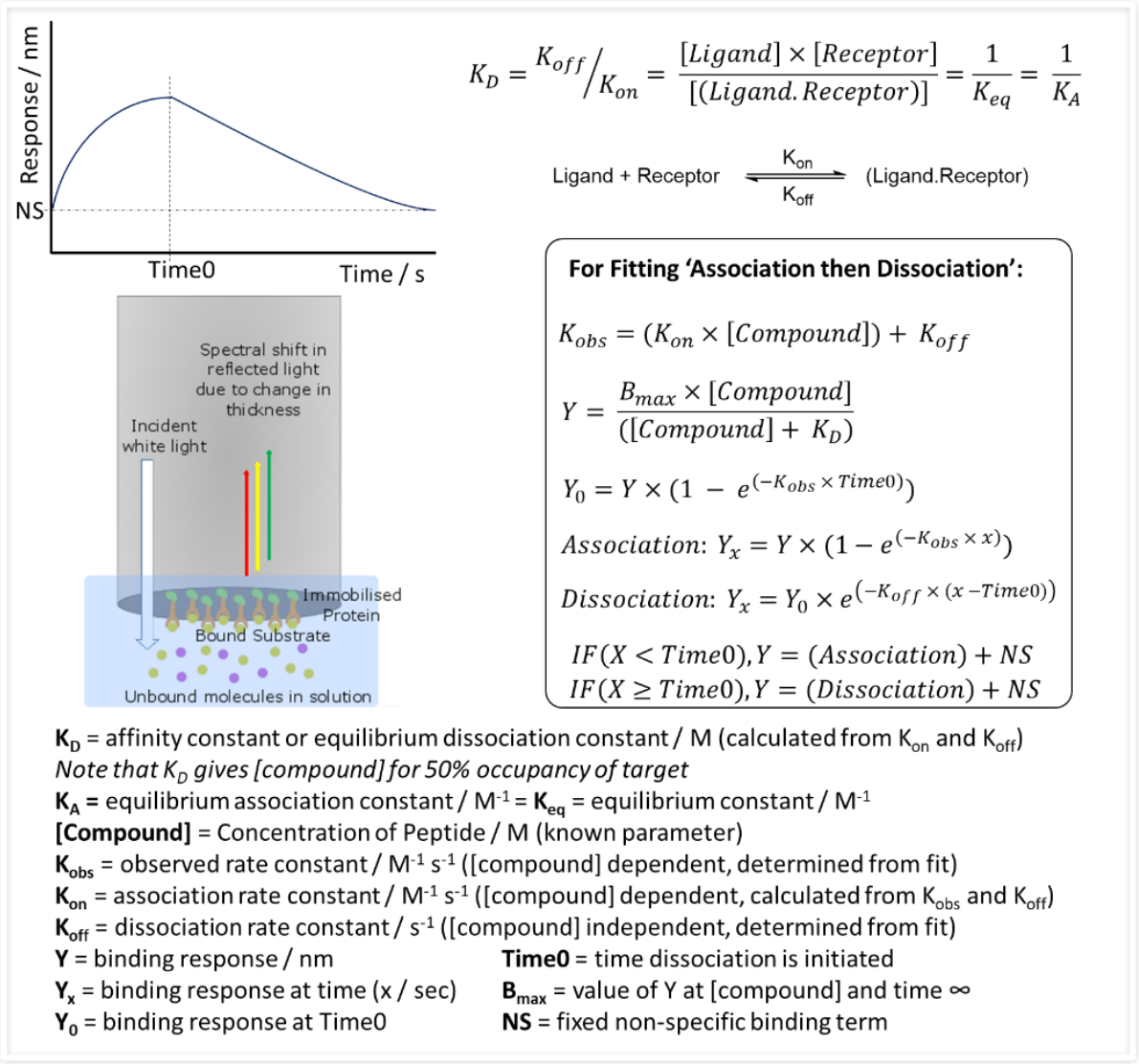
2.5. Validation of Cyclic Peptide Binding

Several of the KDM7 proteins (His-KDM7A$_{[PHD]}$, His-KDM7C$_{[PHD]}$, His-KDM7A$_{[PHD+JmjC]}$, His-KDM7B$_{[PHD+JmjC]}$, His-KDM7C$_{[PHD+JmjC]}$) were gratefully received from Yijia Zhang (Kawamura and Schofield Labs).

2.5.1. BioLayer Interferometry

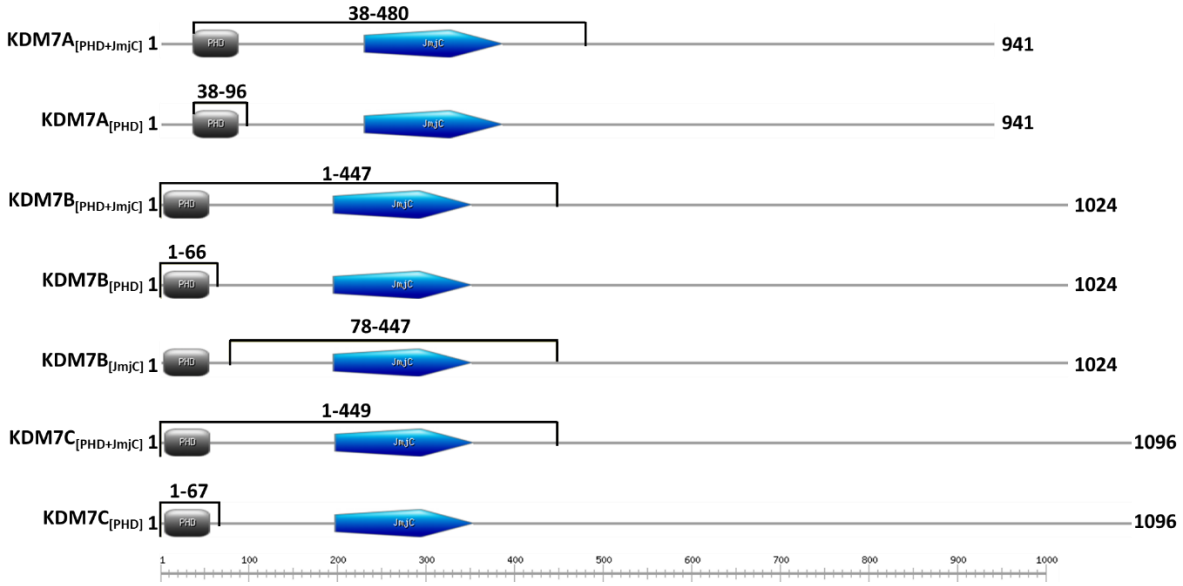
Bio-Layer Interferometry (BLI) on an OctetRed-384 (FortèBio) system was used to assess the synthesised cyclic peptides (OC3, OC4, OC6, OC8, OC9, OC10) for binding to KDM7s. BLI is capable of directly measuring rates of binding in real-time to determine dissociation (affinity, K_D) constants (**F.2.5.i.**) [232]. In BLI, a biomolecule is immobilised on the surface of a sensor and dipped into a solution of its binding partner. Commonly the protein is immobilised and dipped into the ligand. A range of functionalised-sensors are available, making immobilisation possible via streptavidin-biotin, nickel-histidine, proteinA-IgG and other partnered interactions. BLI is a technique with a relatively high throughput (e.g. 384-well plate), enabling rapid characterisation of binding for multiple compounds against multiple targets. A typical procedure for characterisation of the cyclic peptides included immobilisation of a KDM7 protein of interest on a row of BLI sensors, association to a peptide dilution series in solution, then transfer to buffer-only wells for dissociation. Throughout the following BLI results within this thesis, the top concentration of peptide is noted on its respective graph following a 2-fold dilution series thereafter. Vigorous shaking of the BLI apparatus to mix solutions and non-saturative loading of protein to the BLI sensor were used to minimise the impact of diffusion/mass-transfer on observed rates of binding for compounds up to high picomolar affinity [232]. Full experimental details are available in Materials and Methods.

Use of BLI enabled identification of binding cyclic peptides, with the 'Association then Dissociation' model used to fit the data where appropriate to determine binding constants. Full details of the models are given in the Materials and Methods.



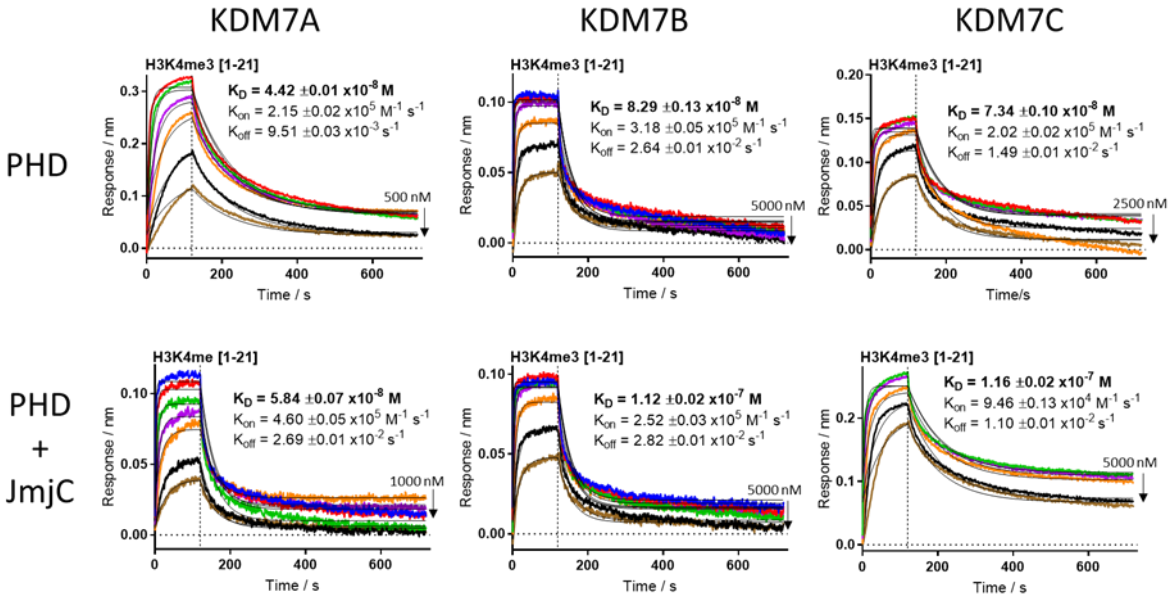
F.2.5.i. Principle of BLI and the 'Association then Dissociation' model (Prism v7) used to fit data. BLI measures the real-time binding interaction between two partners through optical monitoring of the surface when the sensor, pre-loaded with partner A, is dipped into a solution of partner B. Partner A can be loaded to a fixed amount (nm) on multiple sensors (up to 8) in parallel and dipped into multiple concentrations of partner B to give accurate determination of the affinity constant, K_D, through measurement of K_{obs} and K_{off} rates and fitting of the data to the 'Association then Dissociation' model in Prism v7 (GraphPad).

The cyclic peptides were investigated for which of the KDM7B domains (PHD and/or JmjC) they bound to, and some were progressed to assessment for selectivity within the KDM7 sub-family. To achieve this, the KDM7 domains (KDM7A,B,$C_{[PHD]}$, $KDM7B_{[JmjC]}$, $KDM7A,B,C_{[PHD+JmjC]}$, **F.2.5.ii**) were recombinantly produced with N-terminal His tags (in collaboration with Yijia Zhang, see Materials and Methods for full construct details). These proteins were compatible with Ni-NTA immobilisation for BLI. The use of the N-terminally located His immobilisation tag also provided a check against whether cyclic peptide binding was reliant on the immobilisation interface or biotin-AviTag. The Ni-NTA sensors were amenable to regeneration between samples which was important for increasing assessment throughput.



F.2.5.ii. The full-length KDM7 domain maps are represented (generated using ExPASy Prosite MyDomains) with the sequence spanned by their respective recombinant constructs ($KDM7_{[PHD]}$ or $KDM7_{[JmjC]}$ or $KDM7_{[PHD+JmjC]}$) annotated in black.

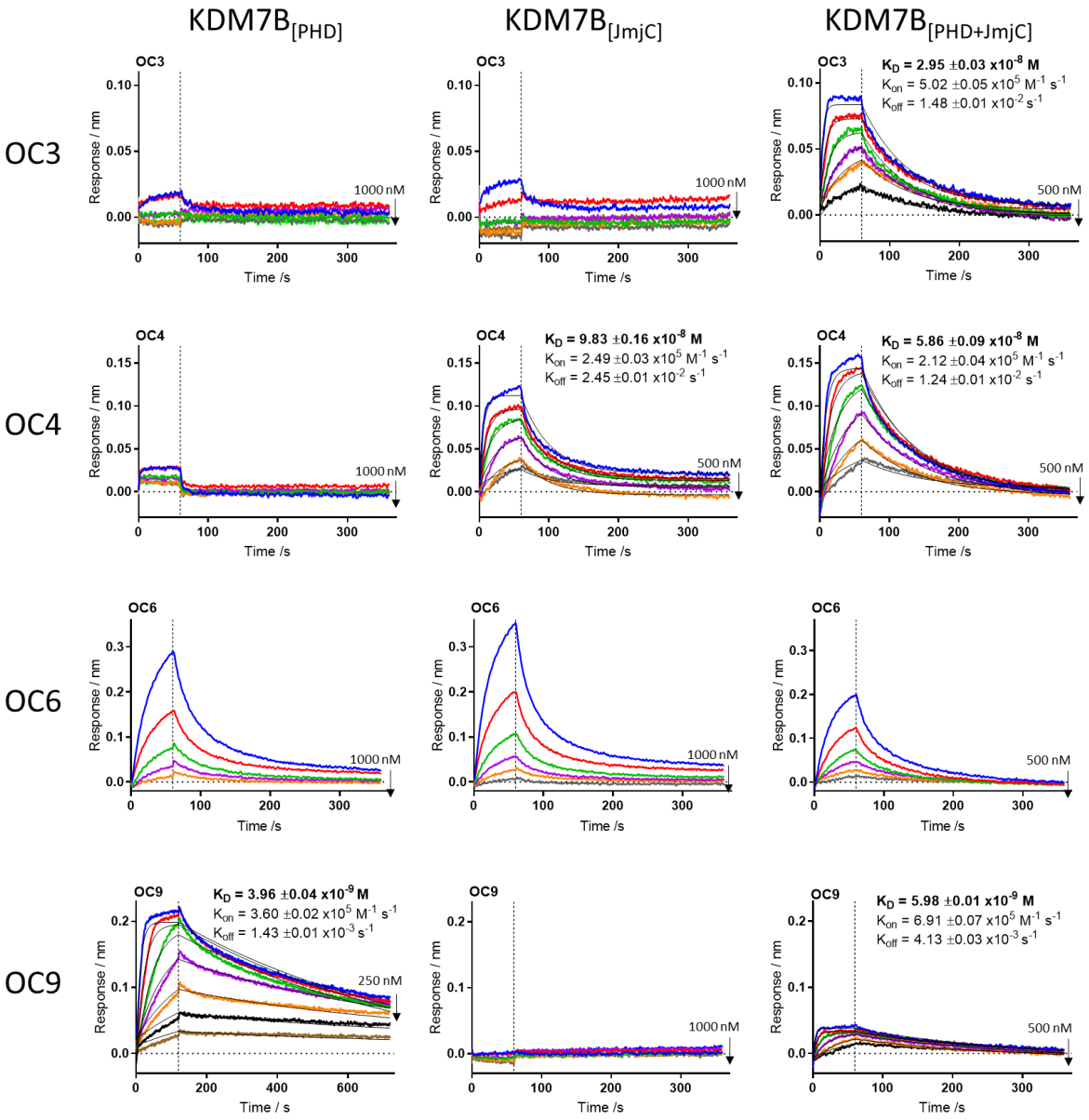
The His-KDM7$_{[PHD] \text{ and } [PHD+JmjC]}$ proteins were first validated using BLI with cognate histone H3K4me3 peptide (**F.2.5.iii.**). H3K4me3 [1-21] had affinities of K_D (nM) = 44 [KDM7A$_{[PHD]}$], 83 [KDM7B$_{[PHD]}$], 73 [KDM7C$_{[PHD]}$], similar to those reported elsewhere [65,116]. The His-KDM7$_{[PHD+JmjC]}$ binding affinities (K_D (nM) = 58 [KDM7A$_{[PHD+JmjC]}$], 112 [KDM7B$_{[PHD+JmjC]}$], 116 [KDM7C$_{[PHD+JmjC]}$]) were comparable to those for His-KDM7$_{[PHD]}$, which highlighted the significant extent to which KDM7$_{[PHD]}$ drives the interaction with the histone H3K4me3 peptide tail for the KDM7s. The KDM7$_{[JmjC]}$ was more challenging to validate for BLI, as it does not by itself possess a significant level of affinity for any histone H3 peptide, even for its substrate H3K9me2 (K_D >10 µM) [65], meaning attempts to determine affinity for any cognate histone peptide partner by BLI were confounded by significant interference from non-specific interactions at high ligand concentrations. Nonetheless, KDM7$_{[JmjC]}$ was still included as part of the cyclic peptide screening process.



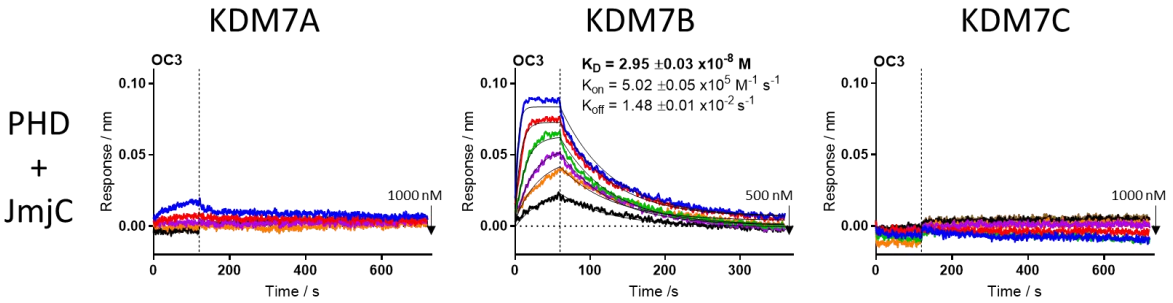
F.2.5.iii. Cognate H3K4me3 [1-21] peptide binding affinities to KDM7A,B,C$_{[PHD]}$ and KDM7A,B,C$_{[PHD+JmjC]}$ were determined by BLI, found to be in the mid to high nanomolar range. Proteins were immobilised by N-terminal His-tag on Ni-NTA sensors. Each coloured trace represents the response for a peptide concentration (2-fold dilution series). The fitted data, overlaid in black where appropriate fitting with the 'Association then Dissociation' model could be performed, were used to determine the binding affinity, K_D, and K_{on} and K_{off}. The start of dissociation is indicated with a dashed vertical line.

Each of the cyclic peptide hits was tested against KDM7B$_{[PHD] \text{ or } [JmjC] \text{ or } [PHD+JmjC]}$ (**F.2.5.iv.**). OC8 and OC10 did not appear to bind these proteins and were removed from the process. OC4 was found to bind KDM7B$_{[JmjC]}$ (K_D = 98 nM) and had similar affinity for KDM7B$_{[PHD+JmjC]}$ (K_D = 59 nM), but was not selected for further characterisation due to the desire to focus on PHD-finger interacting cyclic peptides. OC6 also appeared to bind KDM7B$_{[PHD+JmjC]}$, but gave a peculiarly large response and similar appearing binding traces against KDM7B$_{[PHD]/[JmjC]}$, which was suggestive of OC6 interacting more substantially with the sensor coating itself, rather than the immobilised protein, as it was highly unlikely to retain the same binding response profile against all three constructs. OC6 was therefore discontinued from further Ni-NTA based BLI studies. OC3 bound to KDM7B$_{[PHD+JmjC]}$ (K_D = 30 nM), but neither of the individual domains, indicating it required the PHD finger and JmjC domain in some capacity for binding. OC9 bound to KDM7B$_{[PHD+JmjC]}$ (K_D = 6 nM) and excitingly to KDM7B$_{[PHD]}$ (K_D = 4 nM), but not to KDM7B$_{[JmjC]}$. Focus for further characterisation was placed on OC3 and OC9 for their having possible PHD finger interactions.

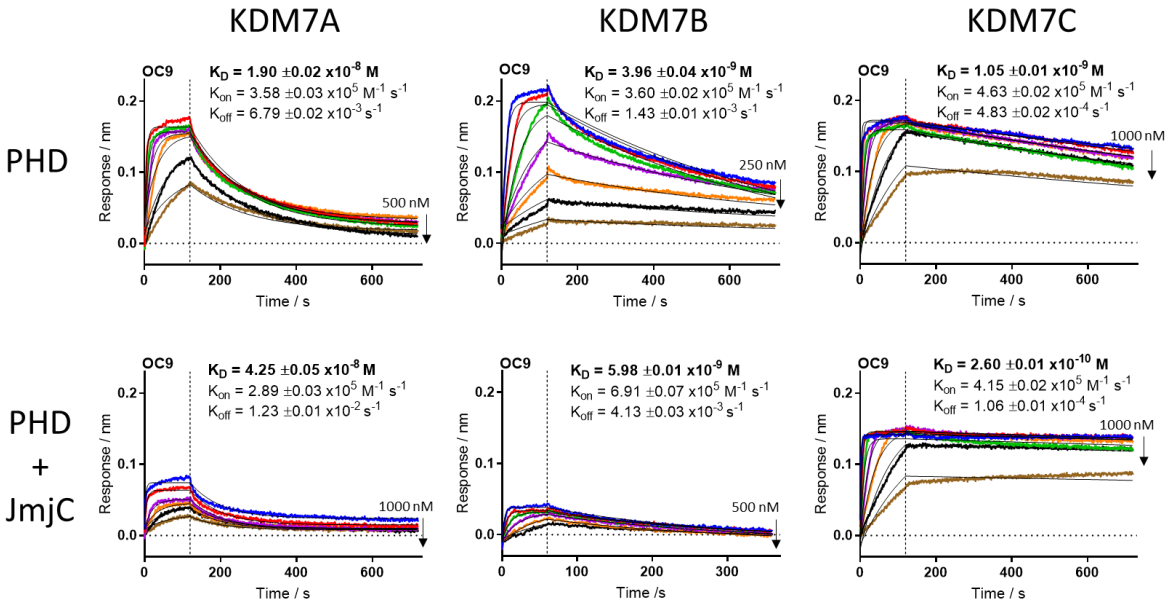
Profiling of OC3 against KDM7A,B,C$_{[PHD+JmjC]}$ found, remarkably, that it bound only to KDM7B$_{[PHD+JmjC]}$ (**F.2.5.v.**). This promised potential selectivity even within the KDM7 sub-family. A more extensive KDM7 profiling of OC9 revealed that it possessed low nanomolar affinity against each KDM7A,B,C$_{[PHD]}$ and a similar level of affinity for each KDM7A,B,C$_{[PHD+JmjC]}$ (**F.2.5.vi.**). This strongly suggested that OC9 interaction with KDM7$_{[PHD]}$ was the major contribution to its binding.



F.2.5.iv. Cyclic peptides were screened for binding to KDM7B$_{[PHD]}$, KDM7B$_{[JmjC]}$, KDM7B$_{[PHD+JmjC]}$ by BLI. Proteins were immobilised by N-terminal His-tag on Ni-NTA sensors. Each coloured trace represents the response for a peptide concentration (2-fold dilution series). The fitted data, overlaid in black where appropriate fitting with the 'Association then Dissociation' model could be performed, were used to determine the binding affinity, K_D, and K_{on} and K_{off}. The start of dissociation is indicated with a dashed vertical line.



F.2.5.v. Cyclic peptide OC3 binding affinities to KDM7A,B,C$_{[PHD+JmjC]}$ were determined by BLI. Binding was only observed against KDM7B$_{[PHD+JmjC]}$ and to be of nanomolar affinity. Proteins were immobilised by N-terminal His-tag on Ni-NTA sensors. Each coloured trace represents the response for a peptide concentration (2-fold dilution series). The fitted data, overlaid in black where appropriate fitting with the 'Association then Dissociation' model could be performed, were used to determine the binding affinity, K_D, and K_{on} and K_{off}. The start of dissociation is indicated with a dashed vertical line.



F.2.5.vi. Cyclic peptide OC9 binding affinities to KDM7A,B,C$_{[PHD]}$ and KDM7A,B,C$_{[PHD+JmjC]}$ were determined by BLI. Low nanomolar affinity was observed against each target. Proteins were immobilised by N-terminal His-tag on Ni-NTA sensors. Each coloured trace represents the response for a peptide concentration (2-fold dilution series). The fitted data, overlaid in black where appropriate fitting with the 'Association then Dissociation' model could be performed, were used to determine the binding affinity, K_D, and K_{on} and K_{off}. The start of dissociation is indicated with a dashed vertical line.

The comparative BLI results are summarised in Table **F.2.5.vii.** which highlights the off-rate (K_{off}) as being the primary determinant in overall affinity. The cyclic peptide and H3K4me3 [1-21] on-rates (K_{on}) all fell within an order of magnitude around 1x 10^5 M^{-1} s^{-1}, consistent with a literature report of computational modelling work from the lab of David Baker [233] which assessed typical protein-protein interaction rates to be in the order of 10^4 to 10^6 M^{-1} s^{-1}. The cyclic peptide off-rates varied by up to two orders of magnitude (1x 10^{-2} s^{-1} for the weakest binders, 1x 10^{-4} s^{-1} for the strongest). In considering a simple system of [P] + [L] in equilibrium with [PL], as is observed for this case of cyclic peptide binding to KDM7 domains, the K_{off} value can be converted to a value for the half-life ($t_{1/2}$) or residence time (t_r), which serve as a measure of the lifetime of the protein-ligand complex [PL], by [234]:

$$t_{1/2} = \ln(2) / K_{off} \qquad t_r = 1 / K_{off}$$

Therefore, the residence time (in seconds) for each ligand with $KDM7B_{[PHD+JmjC]}$ can be estimated to be: 240 s [OC9]; 90 s [OC6]; 70 [OC3]; 40 s [H3K4me3]. Binding potency (K_D) is related to the off-rate (K_{off}) by:

$$K_D = K_{off} / K_{on}$$

and on-rates are typically diffusion-rate limited to 10^4 to 10^6 M^{-1} s^{-1} for protein-protein type interactions [233] (the type of interaction that could reasonably be supposed to be the case for larger ~14mer cyclic peptides [235]), or ~10^8 M^{-1} s^{-1} for small molecules [236]. Therefore, ligands with high binding potency (K_D < 10^{-8} M) will often, but not always, have slow off rates and hence longer residency times. The residence time of the ligand is independent of its concentration and it can sometimes be considered a more useful predictive metric of the functional efficacy of a ligand in-vivo than binding potency, due to a high residence time

leading to a high occupancy of receptors even with low free ligand concentrations [236]. Residence times have been used to guide medicinal chemistry lead optimisation and structure-activity relationship analysis [237], sometimes more effectively than using potency (K_D) alone [238–240]. The cyclic peptides (OC3 and more-so OC9) possess slow off rates, therefore long residence times, which are slower/longer than the cognate H3K4me3 and thus potentially a beneficial property for efficacy in future in-vivo settings.

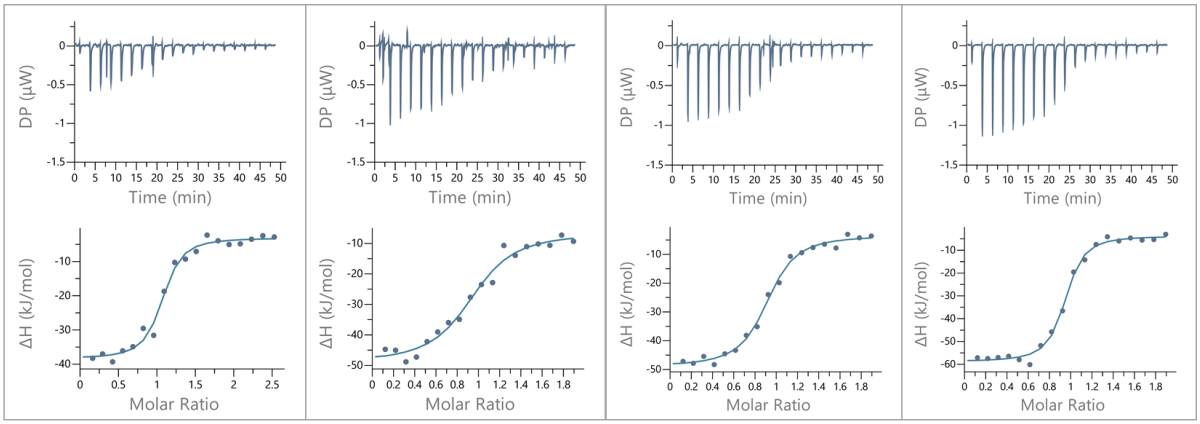
Peptide	KDM Protein	K_D (nM)	± nM	pK_D	K_{on} (M^{-1} s^{-1})	K_{off} (s^{-1})	Global Fit (R^2)
OC9	7A PHD	**19.0**	0.2	**7.72**	3.58E+05	6.79E-03	0.98
	7B PHD	**4.0**	0.1	**8.40**	3.60E+05	1.43E-03	0.98
	7C PHD	**1.1**	0.1	**8.98**	4.63E+05	4.85E-04	0.98
	7A PHD+JmjC	**42.5**	0.5	**7.37**	2.89E+05	1.23E-02	0.98
	7B PHD+JmjC	**6.0**	0.1	**8.22**	6.91E+05	4.13E-03	0.98
	7C PHD+JmjC	**0.3**	0.1	**9.59**	4.15E+05	1.06E-04	0.98
H3K4me3 [1-21]	7A PHD	**44.2**	0.1	**7.35**	2.15E+05	9.51E-03	0.98
	7B PHD	**82.9**	1.4	**7.08**	3.18E+05	2.64E-02	0.97
	7C PHD	**73.4**	1.0	**7.13**	2.02E+05	1.49E-02	0.97
	7A PHD+JmjC	**58.4**	0.7	**7.23**	4.60E+05	2.69E-02	0.98
	7B PHD+JmjC	**112**	2.0	**6.95**	2.52E+05	2.82E-02	0.98
	7C PHD+JmjC	**116**	2.0	**6.94**	9.46E+04	1.10E-02	0.96
OC3	7B PHD+JmjC	**295**	0.3	**6.53**	5.02E+05	1.48E-02	0.99
OC4	7B JmjC	**98.3**	2.1	**7.01**	2.49E+05	2.44E-02	0.98
	7B PHD+JmjC	**58.6**	0.1	**7.23**	2.12E+05	1.24E-02	0.98

F.2.5.vii. The BLI results are summarised where protein binding was observed for H3K4me3 [1-21] and the cyclic peptides. Tabulated parameters were calculated with the 'Association then Dissociation' model.

2.5.2. Isothermal Titration Calorimetry

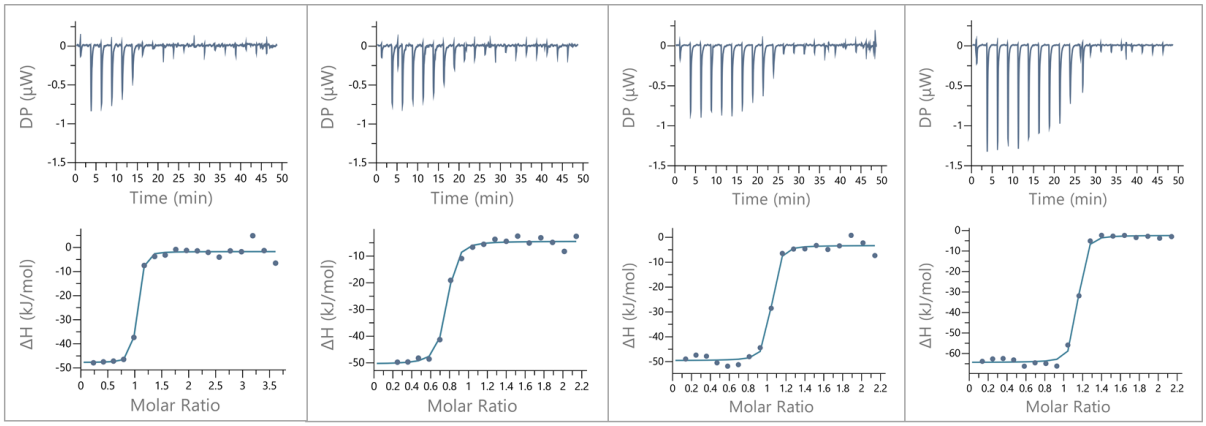
Corroboration of the OC3 and OC9 binding affinities towards their apparent KDM7 target without immobilisation of either partner was performed by ITC – often considered a 'gold standard' technique for qualifying the strength of protein-ligand interactions. OC6 could not be assessed by ITC due to its poor solubility in aqueous buffer and OC4 was not assessed by ITC. The recombinant KDM7A,B,C$_{[PHD]}$ and KDM7B$_{[PHD+JmjC]}$ proteins had their purification/immobilisation tags cleaved during production to exclude the possibility of their contributing to binding affinity. The KDM7s were first validated for use with ITC by testing H3K4me3 [1-21] peptide: K_D (nM) = 448 [KDM7A$_{[PHD]}$]; 205 [KDM7B$_{[PHD]}$]; 81 [KDM7C$_{[PHD]}$]; 168 [KDM7B$_{[PHD+JmjC]}$] (**F.2.5.viii.**). These affinities were in line with those previously reported [65].

ITC confirmed OC9 bound with low nanomolar affinity to KDM7s: KDM7A$_{[PHD]}$ (K_D = 33 nM), KDM7B$_{[PHD]}$ (K_D = 12 nM) and KDM7C$_{[PHD]}$ (K_D = 10 nM) as well as KDM7B$_{[PHD+JmjC]}$ (K_D = 11 nM) (**F.2.5.ix.**). ITC also confirmed OC3 bound to KDM7B$_{[PHD+JmjC]}$ (K_D = 87 nM) (**F.2.5.x.**) and OC3, OC9 and H3K4me3 [1-21] all demonstrated binding in a 1:1 ratio (sites, N) with their targets.



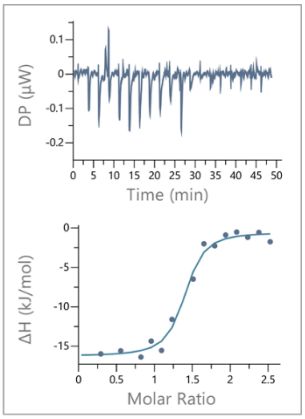
KDM7B$_{[PHD+JmjC]}$ vs H3K4me3 [1-21]	**KDM7A$_{[PHD]}$ vs H3K4me3 [1-21]**	**KDM7B$_{[PHD]}$ vs H3K4me3 [1-21]**	**KDM7C$_{[PHD]}$ vs H3K4me3 [1-21]**
K_D = 168 ± 71 nM	K_D = 448 ± 182 nM	K_D = 205 ± 46 nM	K_D = 81 ± 19 nM
pK_D = 6.77	pK_D = 6.35	pK_D = 6.69	pK_D = 7.09
N = 1.02 ± 0.03	N = 0.93 ± 0.04	N = 0.89 ± 0.02	N = 0.90 ± 0.01
ΔH = -36.7 ± 2.1 kJ mol^{-1}	ΔH = -42.7 ± 3.6 kJ mol^{-1}	ΔH = -45.6 ± 1.54 kJ mol^{-1}	ΔH = -55.1 ± 1.4 kJ mol^{-1}
TΔS = 2.0 kJ mol^{-1}	TΔS = -6.4 kJ mol^{-1}	TΔS = -7.4 kJ mol^{-1}	TΔS = -14.6 kJ mol^{-1}
ΔG = -38.7 kJ mol^{-1}	ΔG = -36.3 kJ mol^{-1}	ΔG = -38.2 kJ mol^{-1}	ΔG = -40.5 kJ mol^{-1}

F.2.5.viii. H3K4me3 [1-21] bound to all three KDM7 PHD fingers and KDM7B$_{[PHD+JmjC]}$ with nanomolar affinity by ITC. See Materials and Methods for experimental details.



$KDM7B_{[PHD+JmjC]}$ vs OC9	**$KDM7A_{[PHD]}$ vs OC9**	**$KDM7B_{[PHD]}$ vs OC9**	**$KDM7C_{[PHD]}$ vs OC9**
K_D = 11 ± 6 nM	K_D = 33 ± 29 nM	K_D = 12 ± 7 nM	K_D = 10 ± 3 nM
pK_D = 7.98	pK_D = 7.48	pK_D = 7.93	pK_D = 8.02
N = 0.95 ± 0.02	N = 0.71 ± 0.01	N = 0.99 ± 0.01	N = 1.09 ± 0.01
ΔH = -46.0 ± 1.5 kJ mol^{-1}	ΔH = -46.0 ± 1.3 kJ mol^{-1}	ΔH = -46.2 ± 1.2 kJ mol^{-1}	ΔH = -62.0 ± 1.0 kJ mol^{-1}
TΔS = -0.4 kJ mol^{-1}	TΔS = -3.3 kJ mol^{-1}	TΔS = -0.9 kJ mol^{-1}	TΔS = -16.1 kJ mol^{-1}
ΔG = -45.6 kJ mol^{-1}	ΔG = -42.7 kJ mol^{-1}	ΔG = -45.3 kJ mol^{-1}	ΔG = -45.8 kJ mol^{-1}

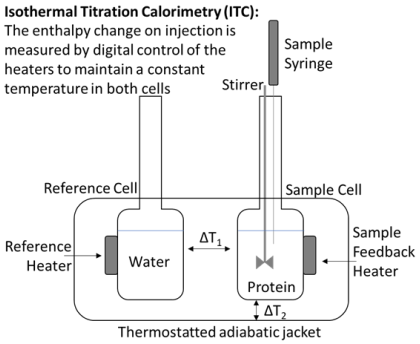
F.2.5.ix. OC9 bound to all three KDM7 PHD fingers and $KDM7B_{[PHD+JmjC]}$ with low nanomolar affinity by ITC. See Materials and Methods for experimental details.

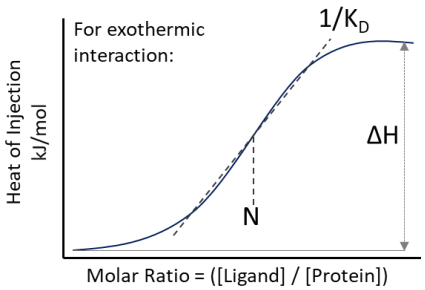


$KDM7B_{[PHD+JmjC]}$ vs OC3
K_D = 87 ± 37 nM
pK_D = 7.06
N = 1.34 ± 0.03
ΔH = -15.7 ± 0.9 kJ mol^{-1}
TΔS = 24.6 kJ mol^{-1}
ΔG = -40.3 kJ mol^{-1}

F.2.5.x. OC3 bound to $KDM7B_{[PHD+JmjC]}$ with nanomolar affinity by ITC. See Materials and Methods for experimental details.

Precise thermodynamic measurements would require further experiments with additional controls; however, the values determined in these binding assays (**F.2.5.xi.**) gave an indication of what the primary driving forces for binding of each cyclic peptide may be. OC3 binding appeared to be driven by a favourable mixture of exothermic enthalpy change (ΔH = -15.7 kJ/mol) and increase in entropy (TΔS = 24.6 kJ/mol) contributing to a free energy term (ΔG = -40.3 kJ/mol) strongly favouring the bound form of the equilibrium. OC9 gave similar free energy terms (ΔG = -42.7 to -45.8 kJ/mol) but instead appeared driven by larger exothermic enthalpy changes (ΔH = -46.0 to -60.0 kJ/mol), with smaller entropic decrease penalties (TΔS = -0.4 to -16.1 kJ/mol). Binding of H3K4me3 [1-21] gave lesser exothermic enthalpy changes compared to OC9 (ΔG = -36.7 to -55.1 kJ/mol) but similar entropy terms (TΔS = 2.0 to -14.6) that contributed to favourable free energies (ΔG = -36.3 to -40.5 kJ/mol) being reduced in magnitude, accounting for their affinity difference. H3K4me3 [1-21] is well characterised to bind to the KDM7s exclusively via the PHD finger [65]. The similarity of the measured thermodynamic parameters for OC9 compared to H3K4me3 [1-21], combined with similar values for OC9 binding to both $KDM7B_{[PHD]}$ and $KDM7B_{[PHD+JmjC]}$, was supportive of the conclusion from BLI that OC9 binding is primarily to $KDM7_{[PHD]}$.





For Sample, S, injection to Protein, P: S + P ⇌ SP

$$\Delta G = \Delta H - T\Delta S = -RTln(K_{eq}) = \mathrm{RTln}(K_D) \qquad K_{eq} = {}^{1}/_{K_D} = \frac{[SP]_{eq}}{[S]_{eq}\,[P]_{eq}}$$

K_D = affinity constant / M (determined from fit) **T** = temperature / K (known)
ΔG = change in Gibbs free energy / kJ/mol^{-1} (calculated) **ΔS** = change in entropy / kJ/mol^{-1} (calculated)
ΔH = change in enthalpy / kJ/mol^{-1} (measured) **R** = universal gas constant

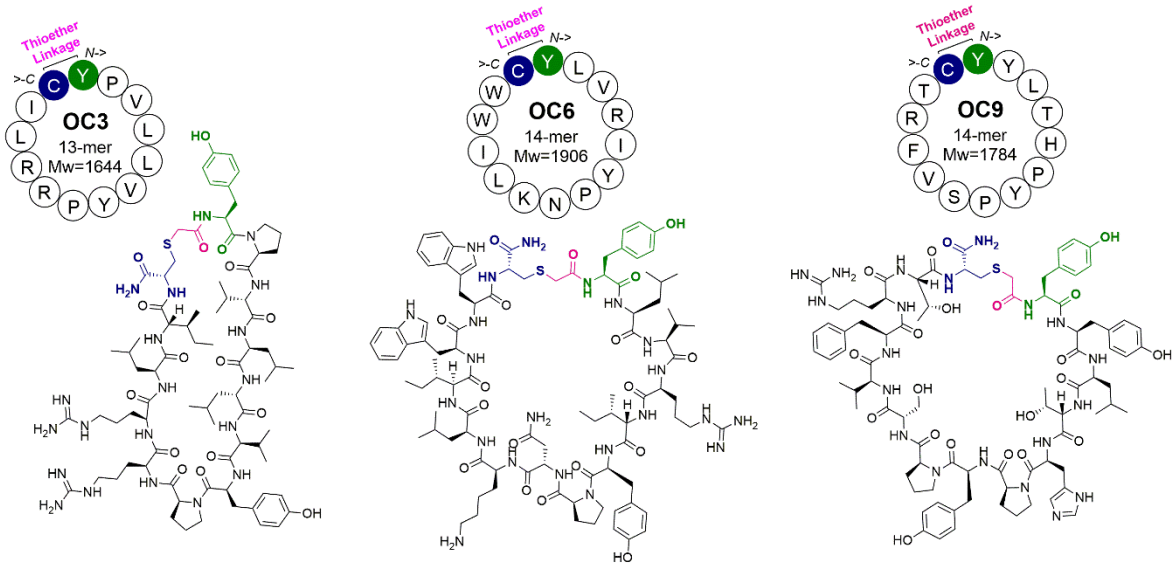
Peptide	KDM Protein	K_D (nM)	± nM	pK_D	N (sites)	± N	kJ/mol			
							ΔH	± ΔH	-TΔS	ΔG
OC9	7A PHD	**33**	29	**7.48**	0.71	0.01	-46.0	1.3	3.3	-42.7
	7B PHD	**12**	7	**7.93**	0.99	0.01	-46.2	1.2	0.9	-45.3
	7C PHD	**10**	4	**8.02**	1.09	0.01	-62.0	1.0	16.1	-45.8
	7B PHD+JmjC	**11**	6	**7.98**	0.95	0.02	-46.0	1.5	0.4	-45.6
H3K4me3 [1-21]	7A PHD	**448**	182	**6.35**	0.93	0.04	-42.7	3.6	6.4	-36.3
	7B PHD	**205**	46	**6.69**	0.89	0.02	-45.6	1.5	7.4	-38.2
	7C PHD	**81**	19	**7.09**	0.90	0.01	-55.1	1.4	14.6	-40.5
	7B PHD+JmjC	**168**	71	**6.77**	1.02	0.03	-36.7	2.1	-2.0	-38.7
OC3	7B PHD+JmjC	**87**	37	**7.06**	1.34	0.03	-15.7	0.9	-24.6	-40.3

F.2.5.xi. The principle and experimental set-up underpinning the ITC experiment are described. ITC results for OC9, OC3 and H3K4me3 [1-21] against their respective KDM7 targets are summarised, with data fitted to determine affinity and thermodynamic parameters using Malvern MicroCal PEAQ-ITC Analysis.

2.6. Conclusions and Future Work

The successful enrichment of a cyclic peptide library against the KDM7B$_{[PHD+JmjC]}$ by mRNA display screening produced a variety of hit 12-14mer cyclic peptide sequences. Cyclic peptides were synthesised by SPPS and successfully screened by BLI for binding to KDM7A, KDM7B and KDM7C targets. OC3 bound to KDM7B$_{[PHD+JmjC]}$ with nanomolar affinity but did not appear to bind to KDM7A,C$_{[PHD+JmjC]}$, KDM7B$_{[PHD]}$ or KDM7B$_{[JmjC]}$, indicating its binding requires some dual interaction with the KDM7B PHD finger and JmjC domain. OC9 bound to each of KDM7A,B,C$_{[PHD]}$ and KDM7A,B,C$_{[PHD+JmjC]}$ with low nanomolar affinity, but did not bind KDM7B$_{[JmjC]}$, indicating OC9 binding was principally to the KDM7$_{[PHD]}$. ITC was used to confirm the binding affinities of OC3 and OC9 for their respective targets, which was at least one magnitude more potent than was observed for the cognate histone H3K4me3 peptide. Thus, these results illustrate the power of the mRNA display approach for discovering potent cyclic peptides against challenging targets.

Confirmed hits OC3 and OC9 were progressed for further profiling of functional effects against the KDM7 sub-family (Chapter 3). The cyclic peptide OC6 could not be assessed as extensively as OC3 and OC9 for binding selectivity by BLI or ITC. Future work would seek to assess this more thoroughly; however, since OC6 could not be excluded at this stage it warranted inclusion for further investigation. Functional assays would be important in assessing the utility of OC3, OC6 and OC9 (**F.2.6.i.**) as potential chemical probes against the criteria and objectives set out in Chapter 0. Future work with new mRNA display selections could apply the principle learned with the KDM7B$_{[PHD+JmjC]}$ selection to use a scaffold protein for display of PHD fingers, or other under-explored smaller protein targets. Further selections against KDM7s could also make use of modified lysines in a linear histone peptide-based library (Chapter 1) to help investigate natural binding partners and substrates.

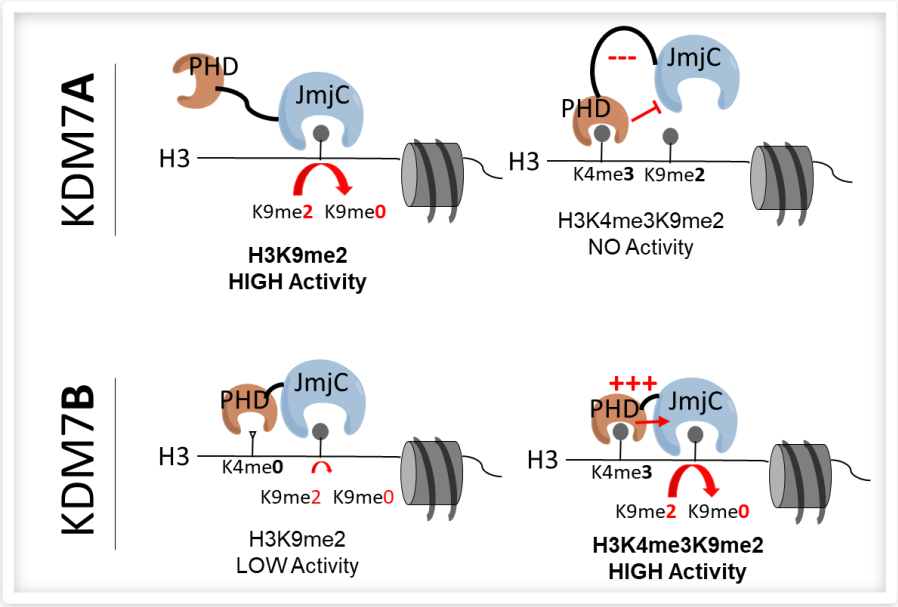


F.2.6.i. The structures of the lead cyclic peptide hits (OC3, OC6, OC9) selected for further study against the KDM7s.

Chapter 3. Cyclic Peptide Function vs KDM7s

3.1. Introduction

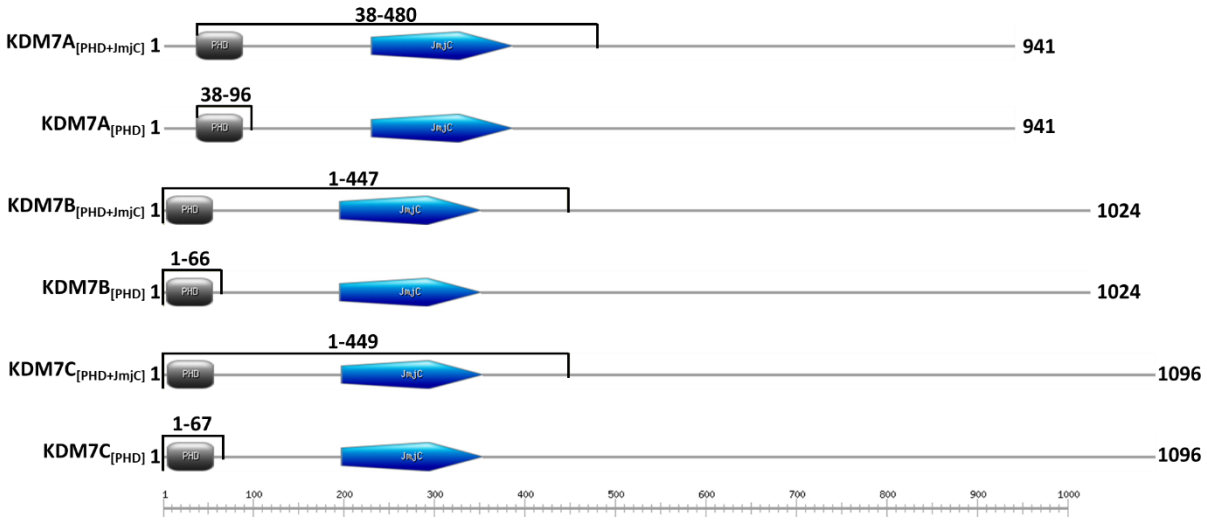
Further studies were carried out on the three cyclic peptides (OC3, OC6, OC9) to determine whether their nanomolar potent binding affinity to the KDM7$_{[PHD]}$ or KDM7$_{[PHD+JmjC]}$ resulted in functional inhibition of the 'reading' action of the PHD finger for the histone H3K4me3 peptide. It was also important to investigate consequences of cyclic peptide binding on the KDM7 JmjC domain demethylase activity, given the different effects of epigenetic cross-talk on KDM7A,B substrate preferences; PHD finger binding to co-localised H3K4me3 activates H3K9me2 demethylation by the JmjC for KDM7B, but represses it for KDM7A (**F.3.1.i.**).



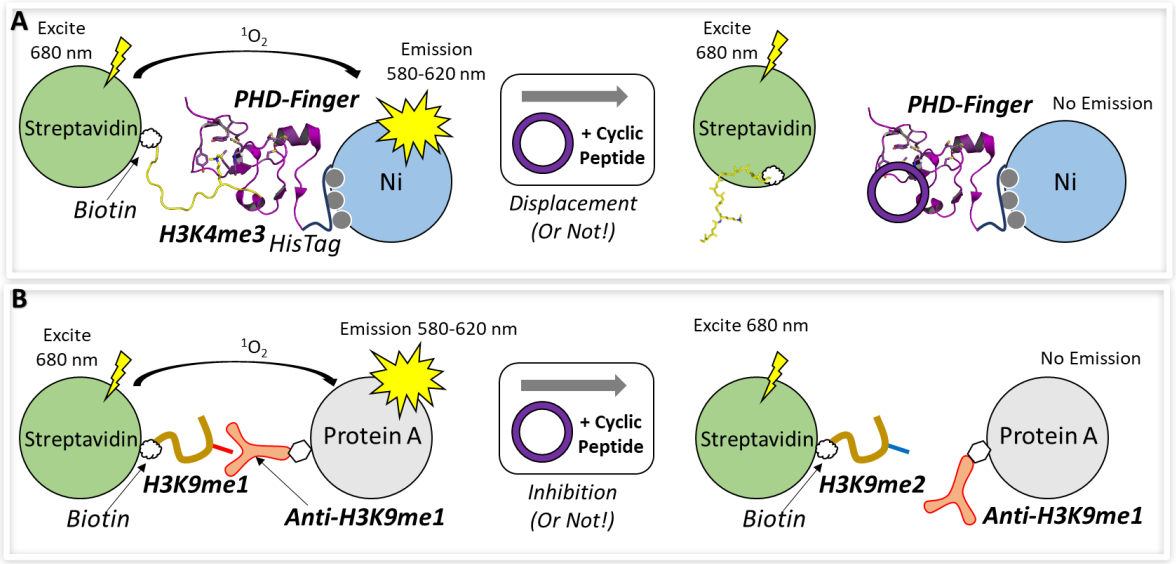
F.3.1.i. The JmjC demethylase activity of KDM7A and KDM7B depends on multiple histone H3 tail methylation states. KDM7A demethylase activity against H3K9me2 is repressed by its PHD finger engagement with co-located H3K4me3, whereas KDM7B demethylase activity on H3K9me2 is promoted by such engagement.

In this Chapter, multiple in-vitro functional assays were used to investigate the cyclic peptide effects on the recombinant KDM7s: displacement of histone H3K4me3 from the PHD finger and assays on JmjC demethylase activity. Displacement assays measured cyclic peptide competition with H3K4me3 for binding to KDM7$_{[PHD]}$ or KDM7$_{[PHD+JmjC]}$ constructs (**F.3.1.ii.**) and were conducted by AlphaScreen assay (**F.3.1.iii.A**). This provided a sensitive (down to low

nanomolar cyclic peptide concentration) and robust method to detect disruption of the cognate protein-protein interaction. Activity assays sought to measure cyclic peptide effects on KDM7$_{[PHD+JmjC]}$ demethylation, primarily concerning the H3K9me2 mark. For H3 peptide substrates this was assessed using various orthogonal assays: antibody-based AlphaScreen assay (**F.3.1.iii.B**), mass spectrometry-based methods (MALDI-TOF or RapidFire) and Western Blot analysis for assays on full-length histone extracts.



F.3.1.ii. The full-length KDM7 domain maps are represented (generated using ExPASy Prosite MyDomains) with the sequence spanned by their respective recombinant constructs (KDM7$_{[PHD]}$ or KDM7$_{[PHD+JmjC]}$) annotated in black.

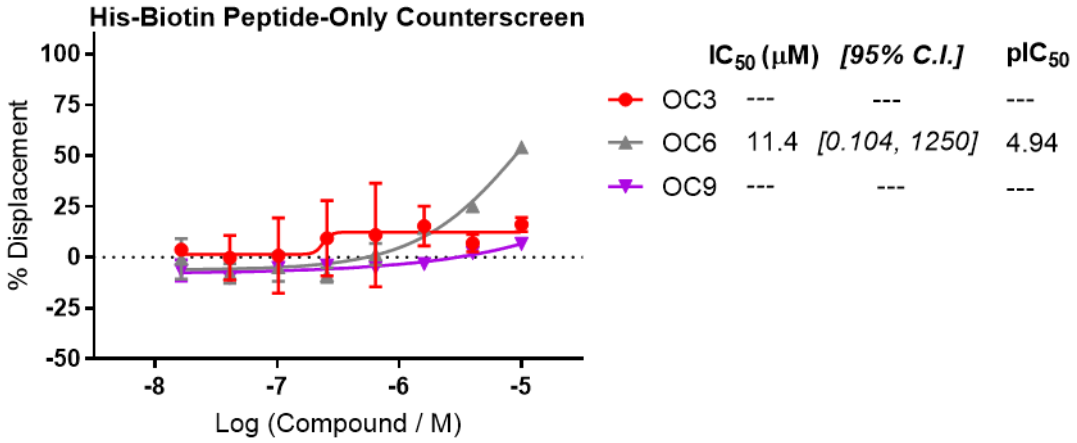


F.3.1.iii. AlphaScreen configurations for assessing binding and enzymatic activity. **A)** The AlphaScreen Histidine (Nickel Chelate) Streptavidin system is used to detect disruption of a protein-histone protein fragment interaction between two suitably tagged targets. **B)** The AlphaScreen IgG (Protein A) Streptavidin system is used to detect the turnover of substrate by an enzyme through antibody recognition of product formation.

3.2. Cyclic Peptide Displacement vs KDM7$_{[PHD+JmjC]}$

3.2.1. Displacement of H3K4me3 Peptide

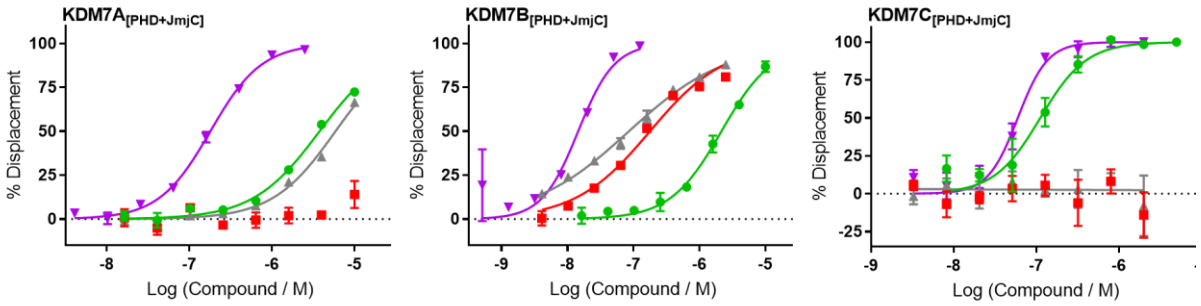
The cyclic peptides (OC3, OC6, OC9) were characterised for their ability to compete with histone H3K4me3 peptide [1-21] for binding to KDM7$_{[PHD+JmjC]}$ using the nickel-histidine based AlphaScreen displacement assay (**F.3.1.ii.A**) [137]. Unlabelled H3K4me3 [1-21] was included as a control to enable comparison to the strength of the cognate interaction, and a counter-screening for non-specific AlphaScreen signal assay interference was performed for the three cyclic peptides (**F.3.2.i.**). OC3 and OC9 did not exhibit any signal interference up to the micromolar concentrations tested, but substantial disruption of AlphaScreen signal was observed for 10 µM OC6.



F.3.2.i. AlphaScreen counterscreen assay with OC3,6,9 cyclic peptides. The cyclic peptide compounds OC3, OC6, OC9 were assessed for non-specific disruption of AlphaScreen assay signal by titration against a fixed concentration of His-Biotin peptide (6760303M, PerkinElmer), capable of binding both AlphaScreen beads at once to mimic a protein-protein interaction. The IC_{50} and 95 % confidence interval values for disruption of the signal are summarised. Data were fitted using Prism 7: log(inhibitor) vs. normalized response -- Variable slope.

The three cyclic peptides were tested for displacement of biotinylated-H3K4me3 [1-21-GGK-biotin] from KDM7A,B,C$_{[PHD+JmjC]}$ (**F.3.2.ii.**). OC9 was found to be the most potent competitor against each, with IC_{50} (nM) = 180 [KDM7A$_{[PHD+JmjC]}$], 14.5 [KDM7B$_{[PHD+JmjC]}$], 59.8 [KDM7C$_{[PHD+JmjC]}$] making it up to two orders of magnitude more effective than H3K4me3 peptide self-displacement. OC6 appeared able to effectively disrupt KDM7B$_{[PHD+JmjC]}$ (IC_{50} =

85.2 nM), and did not displace KDM7C$_{[PHD+JmjC]}$, but did weakly disrupt H3K4me3 binding to KDM7A$_{[PHD+JmjC]}$ (IC_{50} = 5880 nM). However, OC6 had exhibited non-specific interference of AlphaScreen control signals at higher (micromolar) concentrations, thus it is likely that non-specific effects were contributing towards the observed KDM7A IC_{50}. OC3 also effectively displaced H3K4me3 from KDM7B$_{[PHD+JmjC]}$ (IC_{50} = 178 nM), but did not displace KDM7A$_{[PHD+JmjC]}$ or KDM7C$_{[PHD+JmjC]}$ – supported by previous BLI results – indicating potential selectivity even within the KDM7 sub-family.

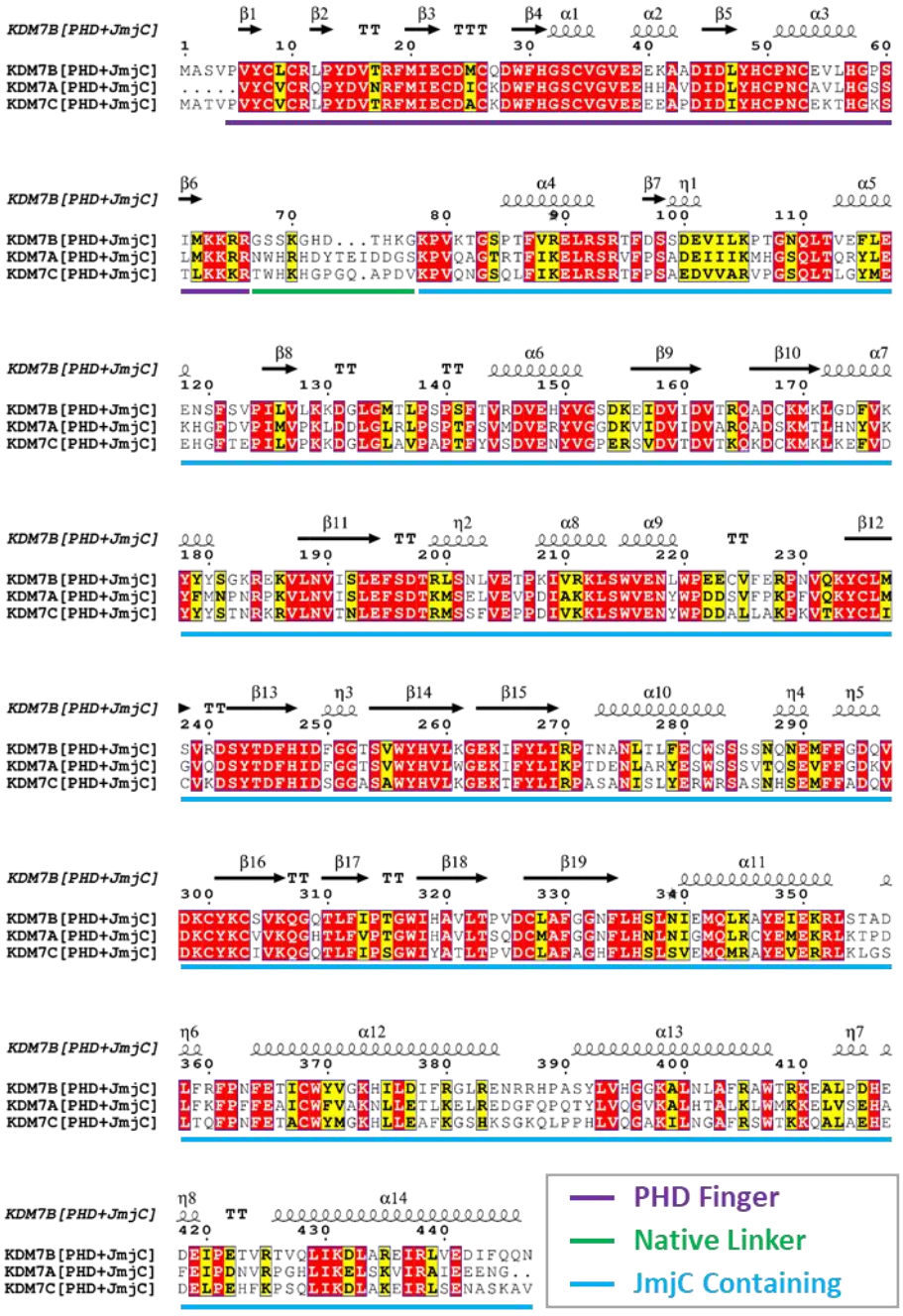


[PHD+ JmjC]	IC$_{50}$ (nM) + *[95% C.I.]* (pIC$_{50}$) OC3	OC6	OC9	H3K4me3 [1-21]
KDM7A	**>10,000** *[---]* (<5.00)	**5880** *[5390, 6420]* (5.23)	**180** *[168, 192]* (6.74)	**3770** *[3410, 4170]* (5.42)
KDM7B	**178** *[147, 215]* (6.75)	**85.2** *[76.9, 94.3]* (7.07)	**14.5** *[10.8, 19.5]* (7.84)	**2160** *[1930, 2410]* (5.67)
KDM7C	**>2000** *[---]* (<5.70)	**>2000** *[---]* (<5.70)	**59.8** *[52.7, 67.8]* (7.22)	**112** *[90.4, 139]* (6.95)

F.3.2.ii. Cyclic peptide assessment by AlphaScreen peptide displacement assay against KDM7$_{[PHD+JmjC]}$. The compounds OC3, OC6, OC9 and H3K4me3 [1-21] were titrated against a fixed concentration of each KDM7A,B,C$_{[PHD+JmjC]}$ with H3K4me3[1-21]-GGK-Biotin cognate binding partner in an AlphaScreen displacement assay. The IC_{50} and 95 % confidence interval values for displacement of the protein-protein interaction are summarised. Data were fitted using Prism 7: log(inhibitor) vs. normalized response -- Variable slope. Full conditions are given in Materials and Methods.

Protein sequence alignments for the three KDM7$_{[PHD+JmjC]}$ domain constructs (**F.3.2.iii.**) clearly show the conservation of residues in the KDM7$_{[PHD]}$, accounting for the ability of OC9 to similarly bind to and functionally affect all three KDM7 sub-family members. The KDM7$_{[JmjC]}$ domains are also generally well conserved, with the greatest differences in amino acid

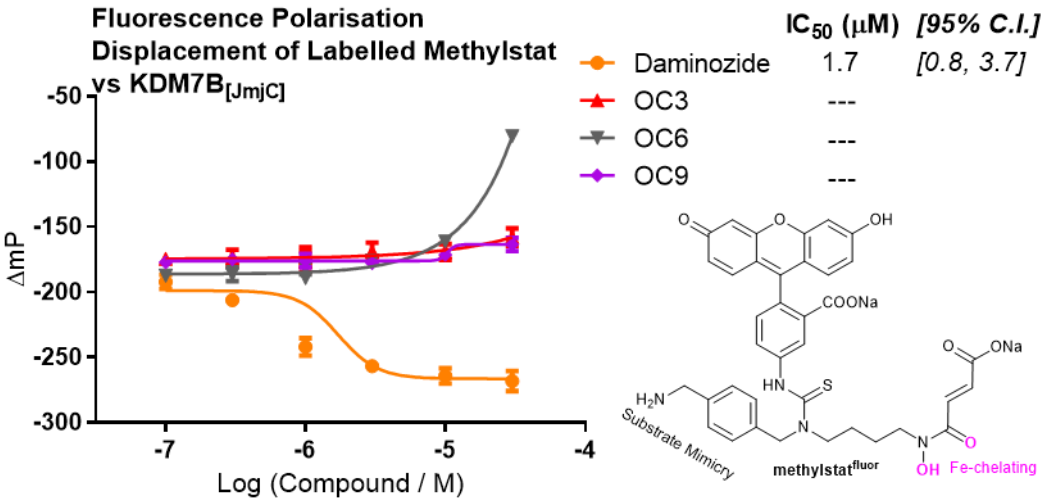
content towards its C-terminus, making it difficult to conclude any immediate explanation for the apparent favouring of KDM7B$_{[PHD+JmjC]}$ by OC3 and OC6.



F.3.2.iii. KDM7 sub-family sequence alignment for PHD+JmjC domain constructs. The KDM7A,B,C$_{[PHD+JmjC]}$ construct sequences (**F.3.1.i**) were parsed through multiple sequence alignment with Clustal Omega (EMBL) [241]. Secondary structure annotation based on a KDM7B crystal structure (PDB:3KV4) was added using ESPript 3.0 [242].

3.2.2. Displacement of Methylstat

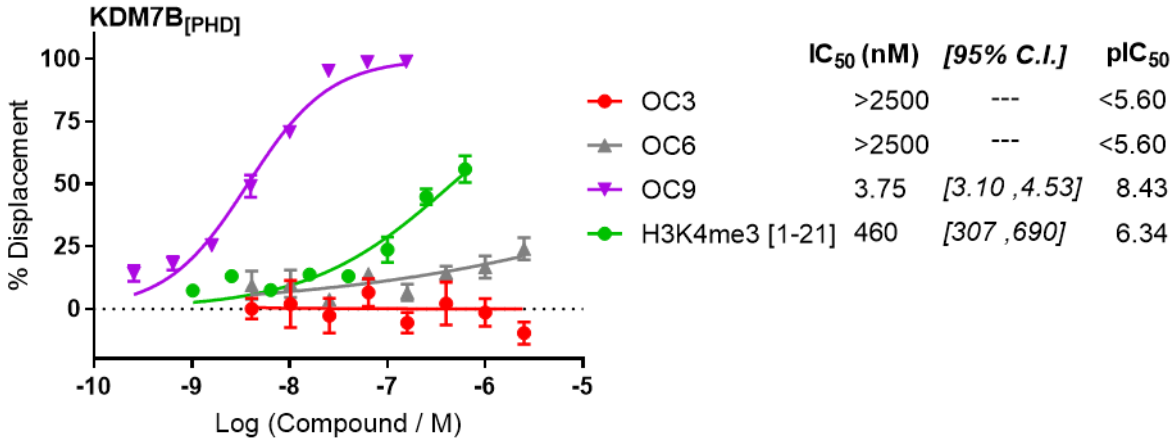
Methylstat is a KDM inhibitor known to bind within the JmjC-KDM active site by competing with peptide substrate and 2OG co-substrate [243]. The fluorescence probe derivative is useful for identification of JmjC-KDM inhibitors that compete with methylstat binding in fluorescence polarisation assays, and has been previously validated for metal-chelating small molecule KDM inhibitors and histone peptides for KDM2A [244] and KDM7s (Yijia Zhang, DPhil Thesis, submitted). A methylstat displacement assay was conducted against $KDM7B_{[JmjC]}$ by Yijia Zhang. A fluorescently labelled version of methylstat was displaced by the 2OG-competitive Daminozide (IC_{50} = 1.7 µM) (**F.3.2.iv.**), but none of the three cyclic peptides. This suggested the cyclic peptides did not interact with the $KDM7B_{[JmjC]}$ active site pocket in a manner similar to methylstat, i.e. that they did not displace a metal-chelating or peptide binding motif at the JmjC active site.



F.3.2.iv. Fluorescence polarisation assay for assessing methylstat displacement. A fluorescence polarisation experiment (performed by Yijia Zhang) tested for cyclic peptide displacement of a labelled-methylstat molecule from the $KDM7B_{[JmjC]}$. The small molecule 2OG-competitor Daminozide, known to inhibit KDM7s, was included as a control and was the only compound to displace methylstat. Assay conditions are given by Yijia Zhang, DPhil Thesis. The IC_{50} and 95 % confidence interval values for displacement of methylstat interaction are summarised. Data were fitted using Prism 7: log(inhibitor) vs. response -- Variable slope.

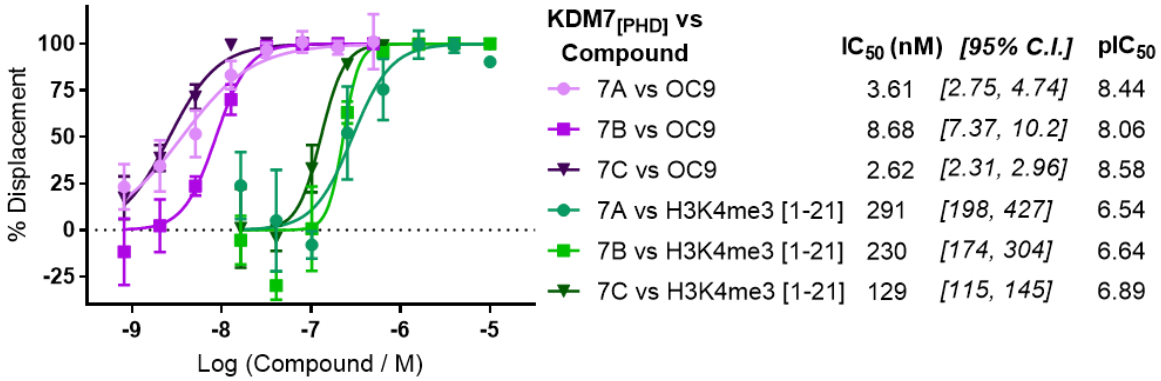
3.3. Cyclic Peptide Displacement vs KDM7$_{[PHD]}$

The major component of KDM7 binding affinity for the H3K4me3 peptide is driven by interaction with the PHD finger [65]. Testing of KDM7B$_{[PHD]}$ using the AlphaScreen displacement assay was used to elucidate whether the major component of cyclic peptide competition was mediated through the PHD finger, or whether the JmjC domain was also required (**F.3.3.i.**). OC3 did not displace H3K4me3 [1-21] peptide from KDM7B$_{[PHD]}$, as might be expected from its previously observed lack of binding by BLI (Chapter 2). OC6 also did not displace KDM7B$_{[PHD]}$. This suggested that both OC3 and OC6 required the presence of the JmjC domain with the PHD finger in order to bind and exert displacement of H3K4me3 peptide. In contrast, OC9 was able to potently displace H3K4me3 peptide from KDM7B$_{[PHD]}$ (IC_{50} = 3.75 nM) suggesting it did not require interaction with the JmjC domain to be able displace H3K4me3 peptide. This observation was supported by BLI and ITC results (Chapter 2), and evidences binding of OC9 to be driven primarily, if not entirely, by interaction with the PHD finger.



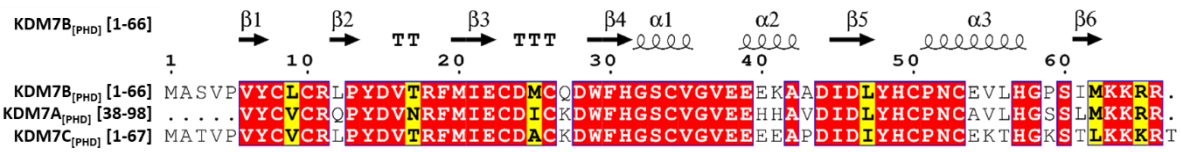
F.3.3.i. Cyclic peptide assessment by AlphaScreen peptide displacement assay against KDM7B$_{[PHD]}$. The compounds OC3, OC6, OC9 and H3K4me3 [1-21] were titrated against a fixed concentration of KDM7B$_{[PHD]}$ with H3K4me3[1-21]-GGK-Biotin cognate binding partner in an AlphaScreen displacement assay. The IC_{50} and 95 % confidence interval values for displacement of the protein-protein interaction are summarised. Data were fitted using Prism 7: log(inhibitor) vs. normalized response -- Variable slope. Full conditions are given in Materials and Methods.

OC9 was subsequently also profiled against KDM7A$_{[PHD]}$ and KDM7C$_{[PHD]}$ and found to possess a low nanomolar displacement potency against all three KDM7A,B,C$_{[PHD]}$ (**F.3.3.ii.**). This builds a robust case for its binding and competition with the H3K4me3 [1-21] peptide being dependent only on the PHD finger, and is logically sound given the high KDM7A,B,C$_{[PHD]}$ construct sequence homology (>80% similarity) (**F.3.3.iii.**). It was also notable that displacement of KDM7$_{[PHD]}$ by OC9 was again approximately two orders of magnitude more effective than self-displacement by H3K4me3 peptide. Displacement by OC9 was more effective against KDM7$_{[PHD]}$ than KDM7$_{[PHD+JmjC]}$, suggesting the large JmjC domain may sterically hinder access to the PHD finger, but OC9 nonetheless remains highly effective against both constructs. Next, the effect on KDM7$_{[PHD+JmjC]}$ catalytic activity by displacement of the histone peptide with the cyclic peptides was examined.



KDM7$_{[PHD]}$ vs Compound	IC_{50} (nM)	*[95% C.I.]*	pIC_{50}
7A vs OC9	3.61	*[2.75, 4.74]*	8.44
7B vs OC9	8.68	*[7.37, 10.2]*	8.06
7C vs OC9	2.62	*[2.31, 2.96]*	8.58
7A vs H3K4me3 [1-21]	291	*[198, 427]*	6.54
7B vs H3K4me3 [1-21]	230	*[174, 304]*	6.64
7C vs H3K4me3 [1-21]	129	*[115, 145]*	6.89

F.3.3.ii. OC9 assessment by AlphaScreen peptide displacement assay against KDM7A,B,C$_{[PHD]}$. The compounds OC9 and H3K4me3 [1-21] were titrated against a fixed concentration of KDM7A,B,C$_{[PHD]}$ with H3K4me3[1-21]-GGK-Biotin cognate binding partner in an AlphaScreen displacement assay. The IC_{50} and 95 % confidence interval values for displacement of the protein-protein interaction are summarised. Data were fitted using Prism 7: log(inhibitor) vs. normalized response -- Variable slope. Full conditions are given in the Materials and Methods.



F.3.3.iii. Human KDM7 sub-family sequence alignment for the PHD finger constructs. The KDM7A,B,C$_{[PHD]}$ constructs were parsed through multiple sequence alignment with Clustal Omega (EMBL) [241]. Secondary structure annotation is based on a KDM7B crystal structure (PDB:3KV4) and was added using ESPript 3.0 [242].

3.4. Cyclic Peptide Effects on KDM7 Activity vs Histone Peptides

Work by the Horton lab [65] had previously concluded that KDM7A and KDM7B demonstrated catalytic demethylase activity against histone H3K9me2 peptide in-vitro, however KDM7C did not. Work within the Kawamura and Schofield labs by Yijia Zhang (unpublished) also found no in-vitro demethylase capability for various KDM7C constructs; KDM7C was therefore not included in the following activity assays. The work by Horton and co-workers had demonstrated effective demethylation of H3K9me2 by KDM7A was strongly favoured by absence of combinatorial H3K4me3, whereas KDM7B demethylation of H3K9me2 depended strongly on the presence of H3K4me3. This was posited to be due to their PHD finger engagement with the H3K4me3 mark exerting a different allosteric-regulatory effect (**F.3.1.i.**). Reasons for this are not yet completely understood, but a potential explanation exists in the difference between the KDM7A and KDM7B native PHD-to-JmjC linker sequence, as their amino acid content differs significantly and the KDM7A sequence is slightly longer (**F.3.4.i.**). This is believed to result in PHD finger binding of H3K4me3 leading to conformational restriction of KDM7A JmjC domain access to H3K9me2, but conversely conformational activation of the KDM7B JmjC domain.

Protein	PHD-to-JmjC Linker	#AAs
KDM7A	NWHRHDYTEIDDGS--	14
KDM7C	-WHKHGPGQAPDV---	12
KDM7B	-----GSSKGHDTHKG	11
	. : *	

Key: '*' = conserved ':' = highly similar
'.' = weakly similar ' ' = dissimilar

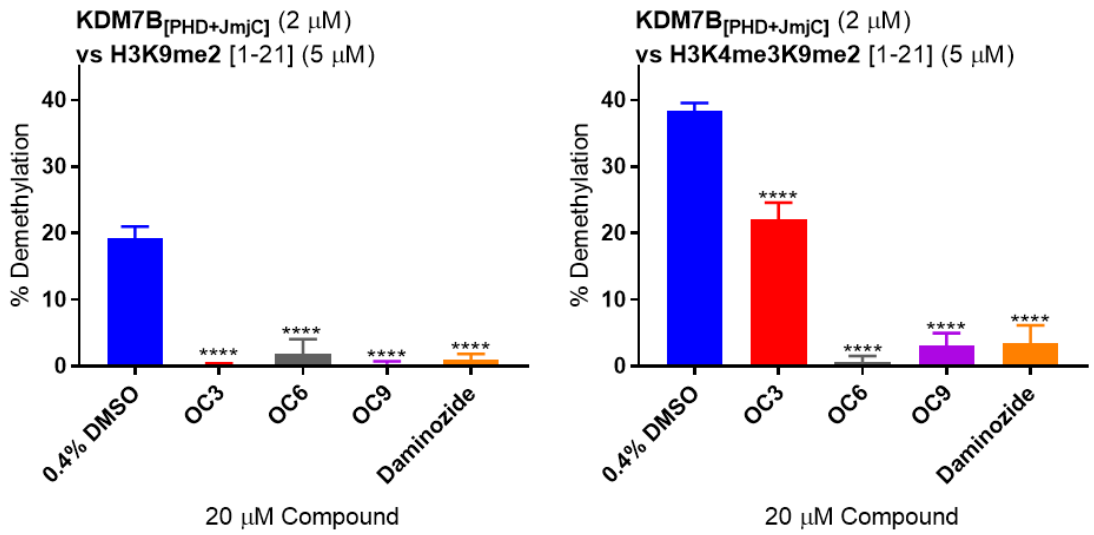
F.3.4.i. Comparison of the native amino acid sequences linking the PHD finger to the JmjC domain for the KDM7 sub-family. Alignment performed by multiple sequence alignment (Clustal Omega [241]).

Therefore, multiple assays were developed to assess the effects of the cyclic peptides (OC3, OC6, OC9) on the KDM7A,B$_{[PHD+JmjC]}$ demethylase activities, with comparison to the established KDM7$_{[JmjC]}$ active-site inhibitor Daminozide (a small molecule 2OG-competitor). Unlike Daminozide, the cyclic peptides were determined not to be competitive with methylstat, a 2OG-competitive inhibitor (**F.3.2.iv.**). The K_M^{app} value for co-substrate 2OG with KDM7B was previously determined to be ~17 µM [160]. Therefore, the following in-vitro assays used 2OG at a concentration that was similar to, or in excess of, its K_M^{app} concentration to minimise 2OG restriction of enzyme activity, to better evaluate the cyclic peptide effects [245]. 2OG concentrations in cells are often found to be >100 µM, also making higher 2OG concentrations more biologically relevant [153]. The sodium ascorbate, Fe(II), and histone peptide substrate were used at concentrations that allowed for a signal window of ≥20% demethylation within the linear range of activity, whilst being comparable to literature precedent [160,246], and other work (unpublished) within the Schofield and Kawamura labs for KDM7s and JmjC-KDMs.

3.4.1. Cyclic Peptides vs KDM7B Demethylase Activity

Initial assessment of the cyclic peptides against the primary selection target ($KDM7B_{[PHD+JmjC]}$) demethylase activity was carried out by a MALDI-TOF-MS based assay at a single concentration of compound (20 µM, 10-fold excess over enzyme) against H3K9me2 [1-21] and H3K4me3K9me2 [1-21] substrates (**F.3.4.ii.**). Consistent with literature [65], the presence of H3K4me3 enhanced KDM7B catalytic activity towards H3K9me2 demethylation for this protein construct (Yijia Zhang, DPhil Thesis, submitted). Therefore, when assessing the effects of the cyclic peptides, different assay conditions were optimised for $KDM7B_{[PHD+JmjC]}$ against both substrates to achieve optimal signal windows (conditions are given in figure legends and in full in the Materials and Methods).

The compounds OC6, OC9 and Daminozide all demonstrated near complete inhibition of $KDM7B_{[PHD+JmjC]}$ demethylation against both substrates. OC3 gave near complete inhibition against the less favoured H3K9me2 [1-21] substrate and partial inhibition against the more favoured H3K4me3K9me2 [1-21] substrate – this was consistent with AlphaScreen results suggesting OC3 was the weakest cyclic peptide for displacement of H3K4me3. The fact that OC9 was found to be an effective inhibitor against the disfavoured H3K9me2 substrate suggested the importance of even weakened $KDM7B_{[PHD]}$ affinity to the unmethylated H3K4 position (*vide infra 3.5.*) being important to JmjC activation. This highlights the critical importance of epigenetic cross-talk for KDM7B, making 'reading' of the H3K4 position an imperative condition for 'erasing' of H3K9me2.



F.3.4.ii. Initial assessment of cyclic peptide inhibition of KDM7B$_{[PHD+JmjC]}$. A MALDI-TOF-MS assay was used to determine KDM7B$_{[PHD+JmjC]}$ demethylation of H3K9me2 (with 100 μM Fe(II), 1 mM NaAsc, 200 μM 2OG) or H3K4me3K9me2 [1-21] (with 50 μM Fe(II), 100 μM NaAsc, 50 μM 2OG) substrates and assess inhibition by different compounds at a single excess concentration. Different conditions were optimised for the different enzyme batches used against the two substrates. An ordinary one-way ANOVA with Dunnett's multiple comparison was used to test for statistical significance of the compound means relative to the 0.4 % DMSO control (multiplicity adjusted p values = *0.0332, **0.0021, ***0.0002, ****<0.0001). Full conditions are given in Materials and Methods.

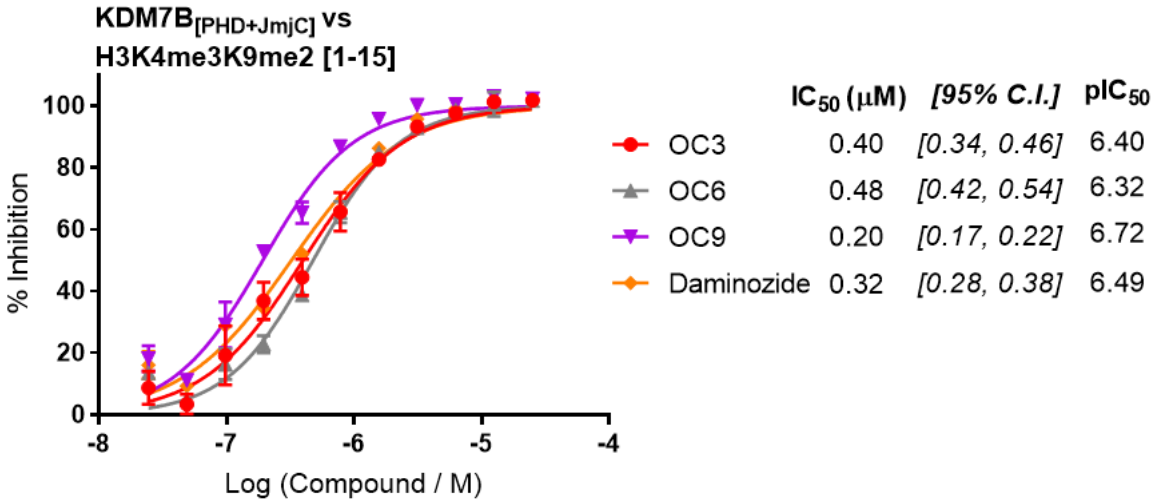
The dose dependence for cyclic peptide inhibition of KDM7B$_{[PHD+JmjC]}$ against the favoured H3K4me3K9me2 peptide substrate was investigated by RapidFire-MS, MALDI-TOF-MS and AlphaScreen assays to provide a robust assessment of cyclic peptide in-vitro efficacy against KDM7B catalytic activity.

In the RapidFire-MS assay, shorter H3K4me3K9me2 [1-15] substrate was used to enable C_{18} cartridge purification of the peptide from the reaction mixture. All three cyclic peptides were found to possess sub-micromolar potency of inhibition (**F.3.4.iii.**). Using RapidFire-MS: IC_{50} (μM) = 0.4 [OC3]; 0.5 [OC6]; 0.2 [OC9]. In the MALDI-TOF-MS assay, H3K4me3K9me2 [1-21] substrate was used. Low micromolar potency of inhibition was observed for each cyclic peptide (**F.3.4.iv.**). Using MALDI-TOF-MS: IC_{50} (μM) = 5.5 [OC3]; 2.1 [OC6]; 1.4 [OC9]. In the AlphaScreen activity assay, OC9 again demonstrated a high level of inhibitory efficacy (IC_{50} = 0.42 μM) (**F.3.4.v.**) and OC3 also appeared inhibitory (IC_{50} = 0.52 μM) (**F.3.4.vi.**). OC6 failed to demonstrate clear inhibition of KDM7B by AlphaScreen activity assay under the tested

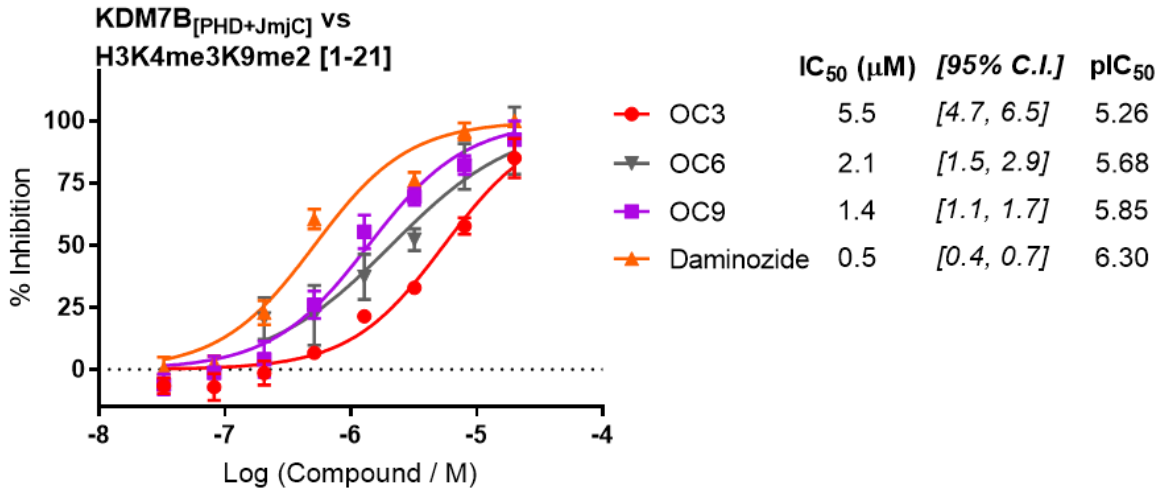
conditions (**F.3.4.vii.**). This was potentially due to assay interference at higher micromolar concentrations, as indicated by the observed very large and unusual increase in AlphaScreen signal (shown as 'negative' inhibition values on the graph). This could be a dose-dependent increase in non-specific aggregation of the AlphaScreen beads, which may reflect OC6 incompatibility with the assay.

It was notable that OC9 ranked as the most potent cyclic peptide inhibitor in each KDM7B activity assay, since all indications were that it bound only to the KDM7 PHD finger. OC9 (IC_{50} (μM) = 0.2 [RapidFire], 0.42 [AlphaScreen] or 1.4 [AlphaScreen]) proved similarly effective to Daminozide (found here with IC_{50} (μM) = 0.3 [RapidFire]; 0.5 [MALDI-TOF]; 1.4 [AlphaScreen] and previously reported with IC_{50} = 0.55 μM by AlphaScreen [160]). The similar potency of OC9 and Daminozide afforded an opportunity for a unique combination experiment; if the action of each compound was truly independent from the other, (PHD finger vs JmjC), then a synergistic effect on inhibition was expected by administering both at once to the KDM7B system. Reassuringly this was observed in the AlphaScreen assay (**F.3.4.v.**), with the 'dual compound' potency (IC_{50} = 0.12 μM) increased by a factor of ~3.5 relative to OC9 alone.

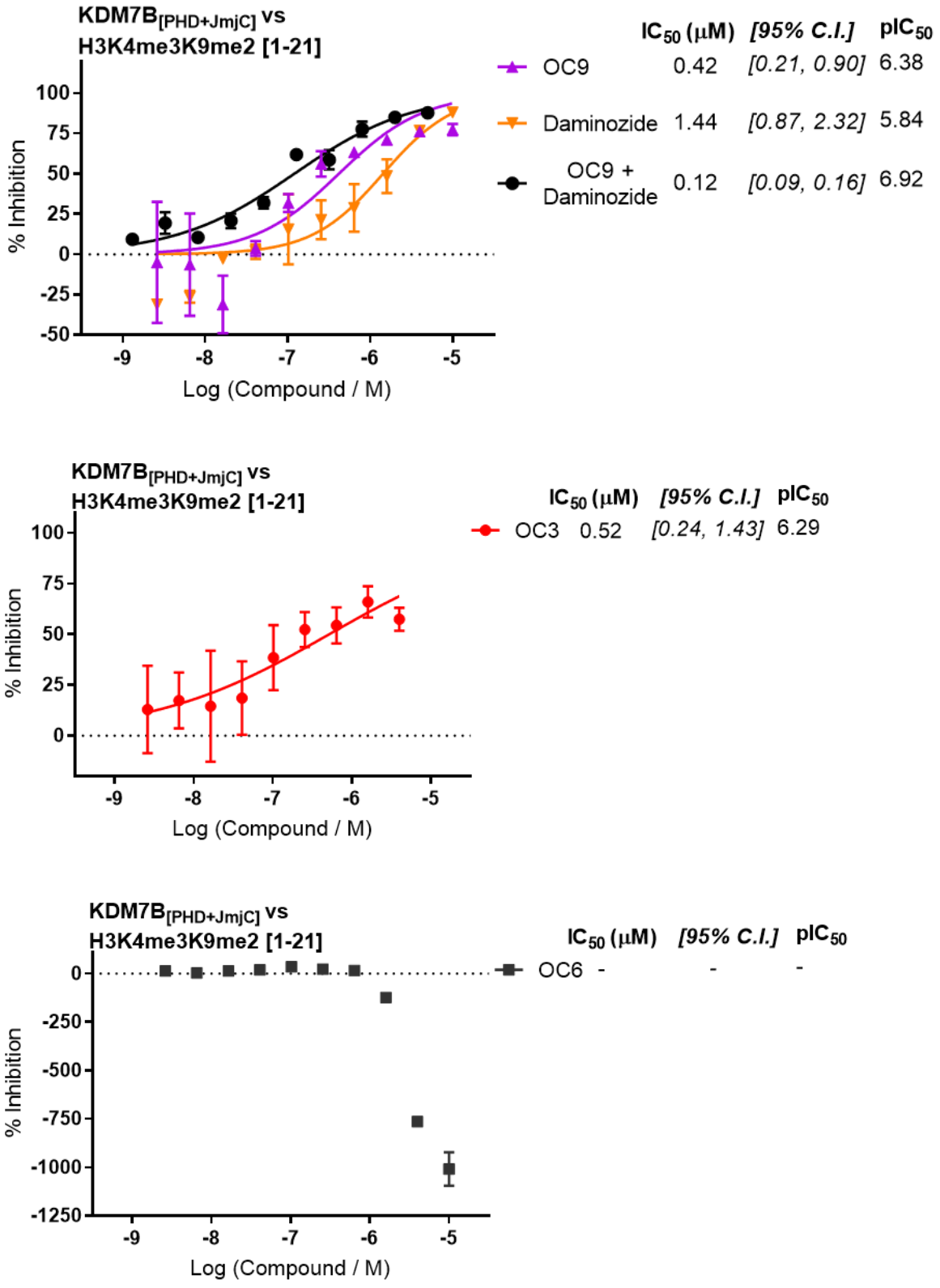
The IC_{50} values for all four compounds varied within an order of magnitude between the different assays, which may be accounted for by the difference in length of peptide substrates (either 15mer or 21mer), the difference in the detection methodologies (C_{18} peptide purification injection to ESI-MS vs direct MALDI-TOF detection vs antibody-H3K9me1 detection) and the differently optimised assay conditions. Together, these results provide a good indication that the in-vitro potency of inhibition by each cyclic peptide (OC3, OC6, OC9) of $KDM7B_{[PHD+JmjC]}$ demethylation of H3K4me3K9me2 resides in the low micromolar range.



F.3.4.iii. RapidFire-MS dose-response assessment of cyclic peptide inhibition of KDM7B. A RapidFire-MS assay was used to assess compound dilution series for inhibition of $KDM7B_{[PHD+JmjC]}$ (0.5 μM) demethylation of H3K4me3K9me2 [1-15] substrate (5 μM) in the presence of 100 μM ascorbate, 20 μM 2OG, 5 μM Fe(II). The total proportions of K9me2/1/0 containing peptides in the +4 charge state were recorded by ESI-MS detection after C18 cartridge purification on an Agilent RapidFire system. IC_{50} values and 95 % confidence intervals are summarised. Analysis of inhibition curves with Prism 7.0 fitted by: log(inhibitor) vs. normalized response -- Variable slope. Full conditions are given in Materials and Methods.



F.3.4.iv. MALDI-TOF-MS dose-response assessment of cyclic peptide inhibition of KDM7B. A MALDI-TOF-MS assay was used to assess compound dilution series for inhibition of $KDM7B_{[PHD+JmjC]}$ (0.35 μM) demethylation of H3K4me3K9me2 [1-21] substrate (1 μM) in the presence of 200 μM ascorbate, 100 μM 2OG, 100 μM Fe(II). IC_{50} values and 95 % confidence intervals are summarised. Analysis of inhibition curves with Prism 7.0 fitted by: log(inhibitor) vs. normalized response -- Variable slope. Full conditions are given in Materials and Methods.

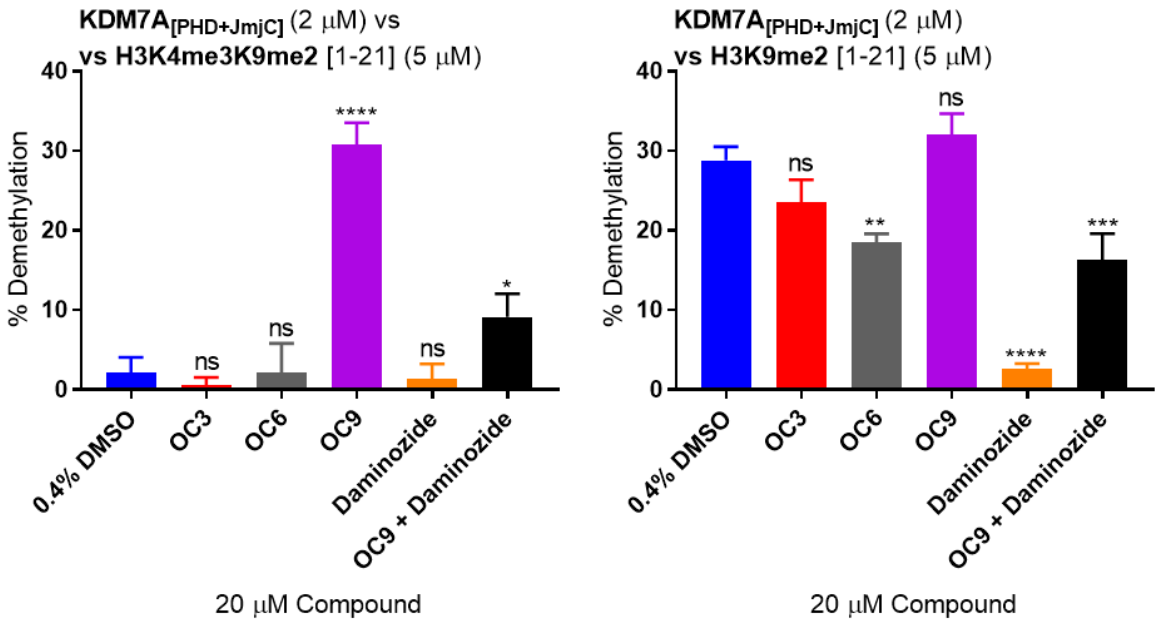


F.3.4.v & vi & vii. AlphaScreen dose-response assessment of cyclic peptide inhibition of KDM7B$_{[PHD+JmjC]}$. An AlphaScreen assay was used to assess compound dilution series for inhibition of KDM7B$_{[PHD+JmjC]}$ (20 nM) demethylation of H3K4me3K9me2 [1-21-GGK-biotin] substrate (150 nM) by antibody detection of generated H3K9me1 in the presence of 100 μM ascorbate, 20 μM 2OG, 10 μM Fe(II). IC_{50} values and 95 % confidence intervals are summarised. Analysis of inhibition curves with Prism 7.0 fitted by: log(inhibitor) vs. normalized response -- Variable slope. Full conditions are given in Materials and Methods.

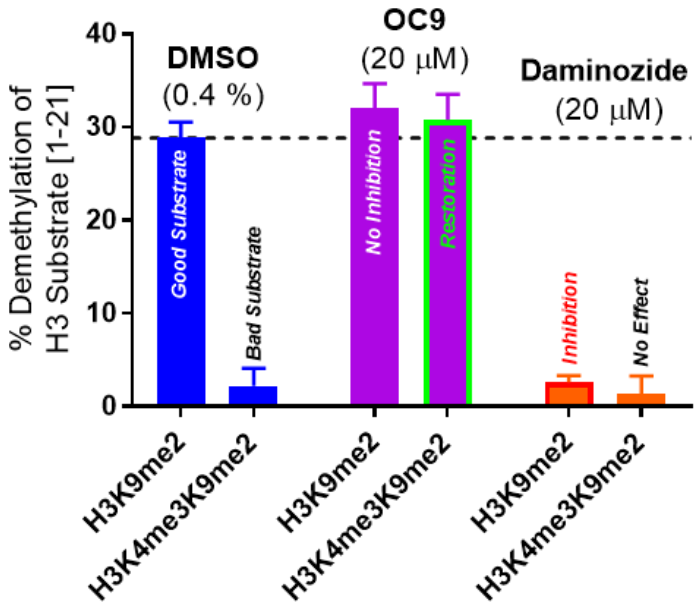
3.4.2. Cyclic Peptides vs KDM7A Demethylase Activity

The cyclic peptides were next assessed against $KDM7A_{[PHD+JmJC]}$ demethylase activity, initially at a single concentration (20 µM) using a MALDI-TOF-MS assay against H3K9me2 [1-21] and H3K4me3K9me2 [1-21] substrates (**F.3.4.viii.**). Consistent with literature [65], $KDM7A_{[PHD+JmjC]}$ demethylation activity was observed against H3K9me2, but was essentially abolished by the presence of H3K4me3 (DMSO-only controls). Against the favoured H3K9me2 substrate, Daminozide displayed the expected high level of inhibition whilst OC6 gave a lesser level of inhibition and OC3 gave only non-significant inhibition – supported by a lack of substantial OC3 interaction with $KDM7A_{[PHD+JmjC]}$ in previous binding and displacement studies. OC3, OC6 and Daminozide had no discernible effect on $KDM7A_{[PHD+JmjC]}$ demethylation of the disfavoured H3K4me3K9me2 substrate due to the very low level of demethylation seen even for the DMSO-only control.

More excitingly, OC9 was able to 'reactivate' and restore $KDM7A_{[PHD+JmjC]}$ demethylation against the normally disfavoured H3K4me3K9me2 substrate to the level of demethylation seen for the favoured H3K9me2 substrate (**F.3.4.ix.**). OC9 did not appear to induce significant enhancement of $KDM7A_{[PHD+JmjC]}$ demethylation on the favoured H3K9me2 substrate under the same conditions. Concurrent inhibition of the KDM7A JmjC active-site with 'blinding' of the PHD finger, by co-dosing of Daminozide with OC9, partially nullified OC9 enhancement of KDM7A against H3K4me3K9me2. This observation further supported these two compounds acting through separate modes of action. The evidence for enhancement of KDM7A activity by OC9 also provides excellent support for OC9 exerting influence only on the KDM7 PHD finger, as any direct inhibition of the JmjC domain would be likely to prevent such an effect.



F.3.4.viii. Initial assessment of cyclic peptide inhibition of KDM7A. A MALDI-TOF-MS based assay was used to determine KDM7B$_{[PHD+JmjC]}$ demethylation of H3K9me2 or H3K4me3K9me2 [1-21] substrates (with 100 μM Fe(II), 1 mM NaAsc, 200 μM 2OG) and assess inhibition by an excess of different compounds at a single concentration. An ordinary one-way ANOVA with Dunnett's multiple comparison was used to test for statistical significance of the compound means relative to the 0.4 % DMSO control (multiplicity adjusted p values = ns>0.0332, *0.0332, **0.0021, ***0.0002, ****<0.0001). Full conditions are given in Materials and Methods.

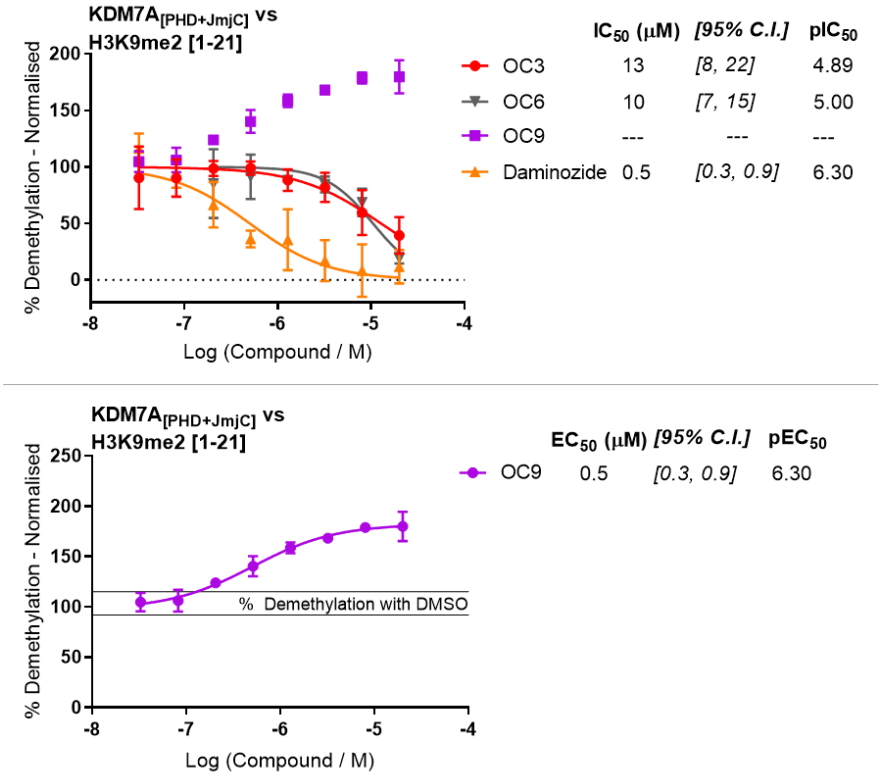


F.3.4.ix. Comparison of the effects of OC9 and Daminozide on KDM7A$_{[PHD+JmjC]}$ demethylation of H3K9me2 and H3K4me3K9me2 [1-21] substrates. Adapted from **F.3.4.vii.** By single concentration testing in a MALDI-TOF-MS assay, OC9 restores KDM7A activity versus the disfavoured H3K4me3K9me2 substrate to levels of activity seen against the favoured H3K9me2 substrate.

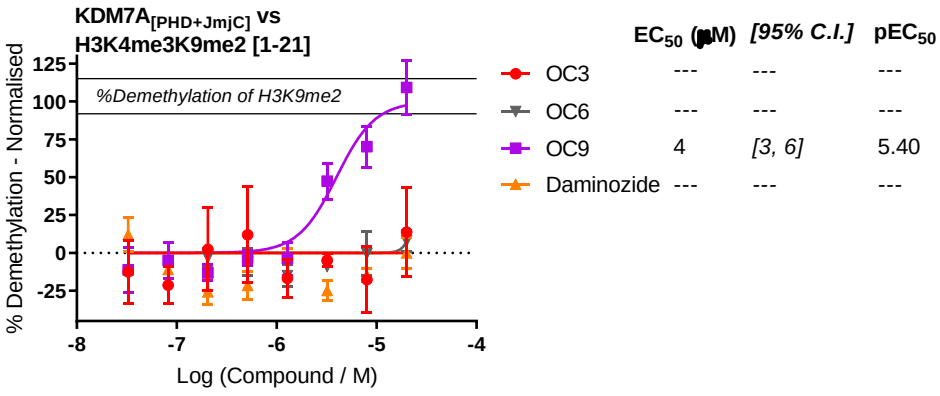
The dose dependence of cyclic peptide effects on $KDM7A_{[PHD+JmjC]}$ demethylation of both H3K9me2 and H3K4me3K9me2 peptide substrates were established by MALDI-TOF-MS assay. Relative to the favoured H3K9me2 [1-21] substrate (**F.3.4.x.**), Daminozide was the most potent inhibitor (IC_{50} = 0.5 μM, previously reported IC_{50} = 2.1 μM in a MALDI-TOF-MS assay [160]). OC9 gave a measurable dose-dependent enhancement (EC_{50} = 0.5 μM), supporting the idea of weak residual $KDM7_{[PHD]}$ affinity for H3K4me0 still being of consequence to JmjC activity. Despite a lack of efficacy in displacement and binding studies, OC3 and OC6 were found to inhibit $KDM7A_{[PHD+JmjC]}$ (OC3 IC_{50} = 13 μM and OC6 IC_{50} = 10 μM) perhaps suggesting a previously undetected weaker level of inhibitory interaction with the KDM7A JmjC domain.

Relative to the disfavoured H3K4me3K9me2 substrate (**F.3.4.xi.**), only OC9 was seen to exhibit a dose-dependent behaviour, enhancing demethylation by $KDM7A_{[PHD+JmjC]}$ (EC_{50} = 4 μM). The ~10-fold weakening of OC9 EC_{50} potency between H3K9me2 and H3K4me3K9me2 is supportive of the idea of more facile displacement of H3K4me0 from the KDM7A PHD finger. The magnitude of OC9 potency against $KDM7A_{[PHD+JmjC]}$ was comparable to that observed against $KDM7B_{[PHD+JmjC]}$, which was consistent with similar potency displacement by OC9 of their PHD fingers, despite the opposite effects on demethylation activity.

The lack of $KDM7A_{[PHD+JmjC]}$ enhancement by OC3 and OC6 was important to note, as it suggested that their displacement of histone peptide, if any, was not functioning through the PHD finger alone akin to OC9. This argues for it being likely that any interaction that either OC3 or OC6 might possess with the KDM7A JmjC domain serves to prevent reactivation of catalytic turnover.



F.3.4.x. MALDI-TOF-MS dose-response assessment of cyclic peptide effects on KDM7A vs H3K9me2. A MALDI-TOF-MS assay was used to assess compound effects on KDM7A$_{[PHD+JmjC]}$ (0.35 µM) demethylation of H3K9me2 [1-21] substrate (2 µM) in the presence of 200 µM ascorbate, 100 µM 2OG, 100 µM Fe(II). IC/EC$_{50}$ values and 95 % confidence intervals are summarised. Analysis of inhibition curves with Prism 7.0 fitted by: log(inhibitor) vs. response or normalised response -- Variable slope. Full conditions are given in Materials and Methods.



F.3.4.xi. MALDI-TOF-MS dose-response assessment of cyclic peptide effects on KDM7A vs H3K4me3K9me2. A MALDI-TOF-MS assay was used to assess compound effects on KDM7A$_{[PHD+JmjC]}$ (0.35 µM) demethylation of H3K4me3K9me2 [1-21] substrate (2 µM) in the presence of 200 µM ascorbate, 100 µM 2OG, 100 µM Fe(II). IC/EC$_{50}$ values and 95 % confidence intervals are summarised. Analysis of inhibition curves with Prism 7.0 fitted by: log(inhibitor) vs. normalised response -- Variable slope. Full conditions are given in Materials and Methods.

3.5. Summary of Peptide Displacement and Activity Assay Results

All displacement and activity assay IC/EC_{50} values are summarised (**F.3.5.i.**).

				Compound XC_{50} (μM)					
				OC3	OC6	OC9		H3K4me3[1-21]	Daminozide
Assay	Peptide	Protein	Variant	IC_{50}	IC_{50}	IC_{50}	EC_{50}	IC_{50}	IC_{50}
Displacement (AlphaScreen)	H3K4me3 [1-21-GGK-Biotin]	KDM7A	PHD+JmjC	>10	5.88	0.180	---	3.77	NM
			PHD	NM	NM	0.004	---	0.291	NM
		KDM7B	PHD+JmjC	0.178	0.085	0.015	---	2.16	NM
			PHD	---	---	0.009	---	0.230	NM
		KDM7C	PHD+JmjC	>2	>2	0.060	---	0.112	NM
			PHD	NM	NM	0.003	---	0.129	NM
Demethylase Activity	H3K4me3 K9me2 [1-21]	KDM7A	MALDI	---	---	---	4.0	NM	---
		KDM7B	AlphaScreen	---	---	0.42	---	NM	1.44
			MALDI	5.5	2.1	1.4	---	NM	0.5
	H3K9me2 [1-21]	KDM7A	MALDI	13	10	---	0.5	NM	0.5

NM = not measured --- = no effect

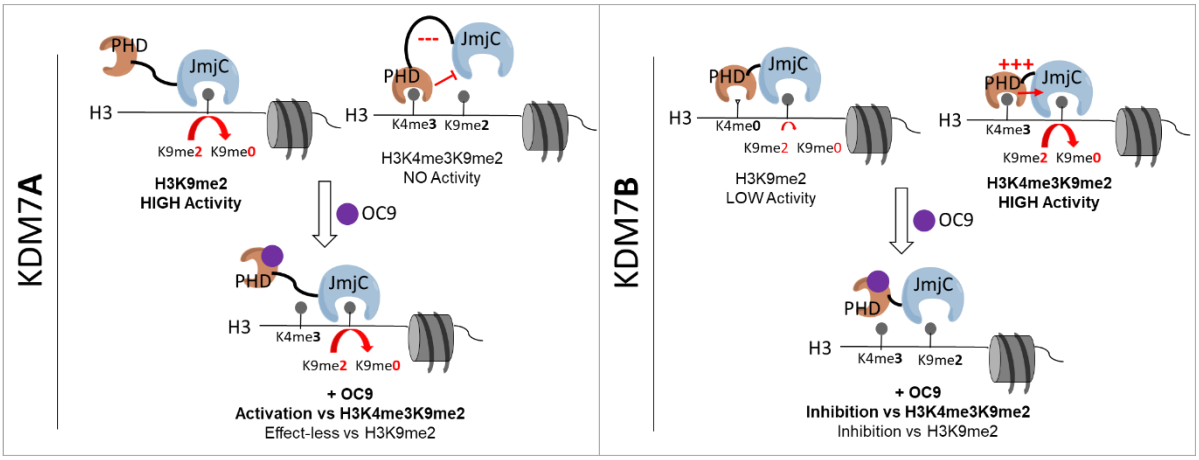
F.3.5.i. A summary of results for the effects of the cyclic peptides (OC3, OC6, OC9), H3K4me3 [1-21] and Daminozide on the KDM7 sub-family in AlphaScreen displacement assays, and the listed KDM7A and KDM7B demethylase activity assays for demethylation of the H3K9me2 mark.

In assessing the effects of the cyclic peptides on KDM7 catalytic activities, it is helpful to consider the epigenetic cross-talk occurring within KDM7A,$B_{[PHD+JmjC]}$. The two substrates under consideration are H3K4me3K9me2 (favoured by KDM7B, disfavoured by KDM7A) and H3K9me2 (favoured by KDM7A, disfavoured by KDM7B). Inhibition only of $KDM7_{[PHD]}$ binding to H3K4me3 (i.e. by OC9) would serve to effectively render either histone substrate as H3K9me2 in appearance to the JmjC domain of both KDM7A and KDM7B (**F.3.5.ii.**). This would therefore explain the ability of OC9 to 'reactivate' KDM7A demethylation on normally disfavoured H3K4me3K9me2 and explain OC9 inhibition of KDM7B against normally favoured H3K4me3K9me2.

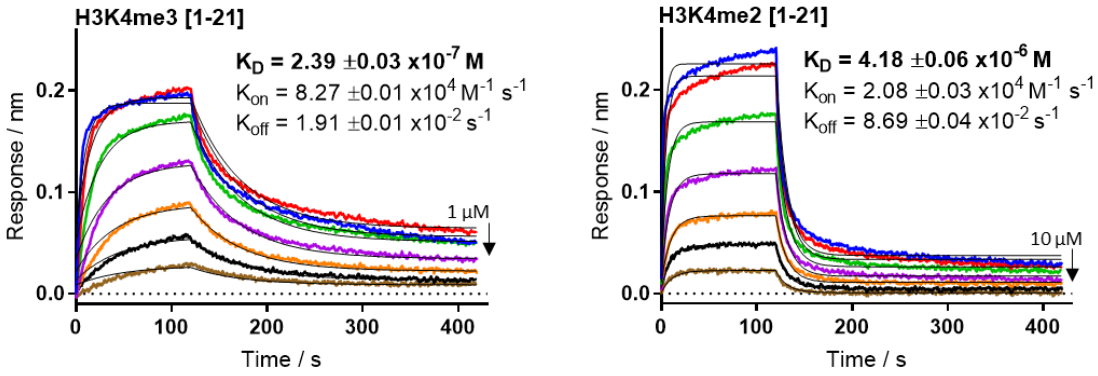
However, OC9 also appeared to activate KDM7A even against its favoured H3K9me2 substrate, and inhibit KDM7B against its already disfavoured H3K9me2 substrate. To explain this, it is important to consider that the $KDM7_{[PHD]}$ could retain weaker levels of affinity for the lower/un-methylated states of H3K4 [1-21]. BLI experiments with $KDM7B_{[PHD]}$ against

H3K4x [1-21] gave K_D (μM) = 0.2 [x = me3], 4.2 [x = me2], >10 μM [x = me1/0] (**F.3.5.iii.**). Therefore, one could argue for the KDM7 PHD finger providing a graded regulation of the attached JmjC domain, and thus inhibition of the PHD finger could still be expected to influence JmjC activity against H3K9me2 substrates containing lower methylation states on H3K4. On the other hand, if displacement of the histone peptide from $KDM7_{[PHD+JmjC]}$ is due to disruption of both PHD and JmjC domains (i.e. by OC3 or OC6), this would explain inhibition of KDM7A and KDM7B catalytic activity against either substrate. Direct inhibition of the JmjC active site should also result in inhibition for each substrate, as was observed for the 2OG-competitive small molecule Daminozide.

The displacement efficacy (IC_{50}) of OC9 against the $KDM7_{[PHD+JmjC]}$ appeared generally to be 10- to 100-fold more effective than the potency of OC9 for inhibiting or enhancing $KDM7_{[PHD+JmjC]}$ demethylation activity. This could reflect the difference in the assay types, as the displacement assay assessed the efficacy of OC9 for competing the binding of H3K4me3 [1-21] peptide to the PHD finger, whereas the activity assays assessed the effect of OC9 on demethylation by the JmjC domain of a different histone peptide substrate, H3K4me3K9me2 [1-21]. Furthermore, the activity assays required higher concentrations of substrate peptide and, as it appears that OC9 can compete with the histone peptide for binding to the protein, then a higher substrate concentration would compete more with OC9 for binding and therefore negatively impact observed OC9 potency.



F.3.5.ii. The proposed effects of OC9 on KDM7A and KDM7B JmjC catalytic activities through inhibition of their respective PHD fingers. Both KDM7A and KDM7B effectively 'see' only H3K9me2 in the presence of OC9, as OC9 'blinds' their PHD finger to any co-localised H3K4me3 mark. H3K9me2 alone is favoured by KDM7A and disfavoured by KDM7B, and thus OC9 exerts its different allosteric regulatory effects.



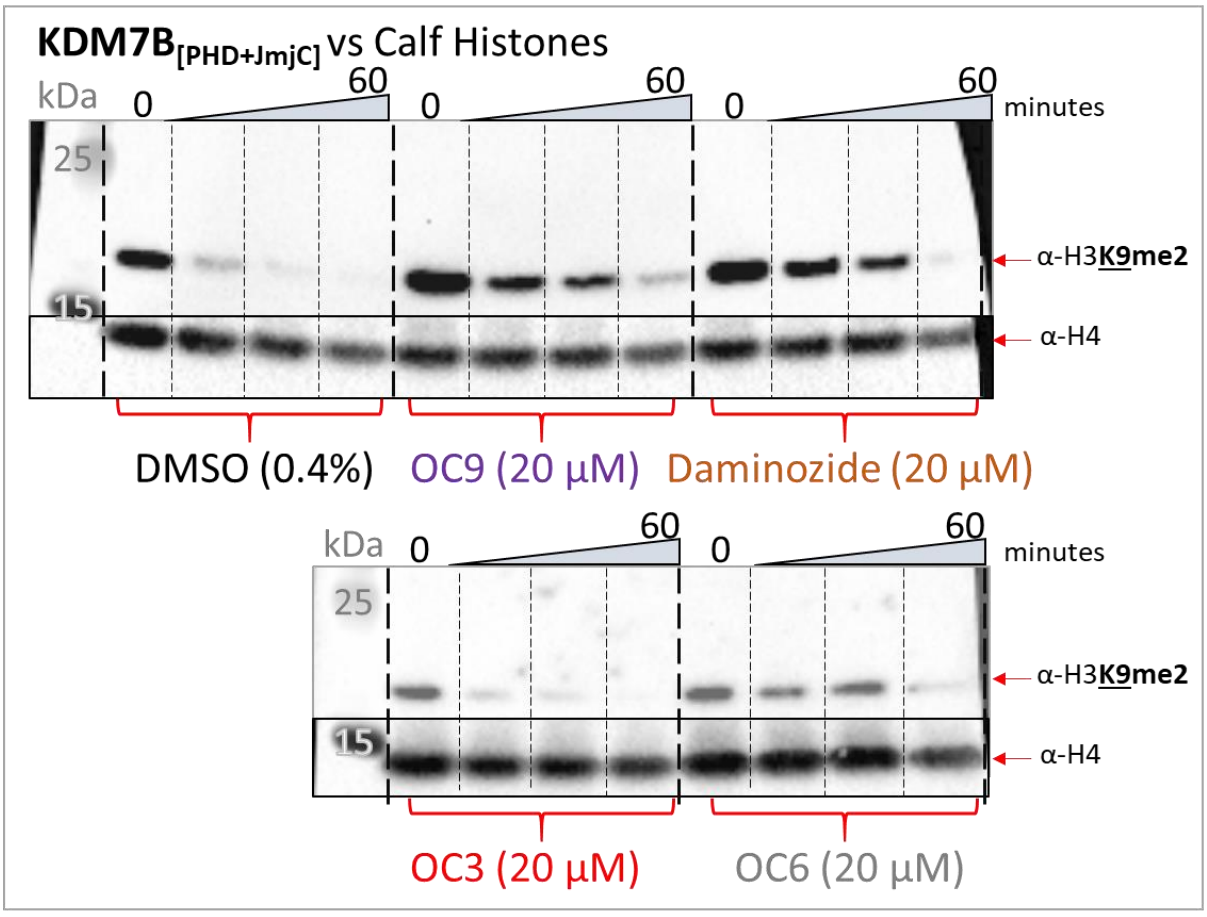
F.3.5.iii. H3K4meX peptide binding affinity for KDM7B$_{[PHD]}$. The histone H3 peptides H3K4meX [1-21] (where X = 0,1,2,3) were assessed by BLI for binding affinity to KDM7B$_{[PHD]}$ immobilised on Ni-NTA sensors. Data were fitted using the Association then Dissociation model. For the H3K4me1/0 peptides, their lower affinities necessitated higher concentrations (> 10 µM) to generate a substantial response, whereupon non-specific binding was evident and fitting was not considered reliable.

3.6. Cyclic Peptide Effects on KDM7 Activity vs Histone Extracts

In-vitro assays to this point had focused on use of shortened H3 peptides as surrogates for the full-length histone protein. Next, cyclic peptide effects on full-length histones were investigated using calf-thymus histone extracts. The histone extracts were purchased commercially (SigmaAldrich) and not controlled for levels of different post-translational modifications, or their relative combinatorial PTMS (e.g. H3K4me3K9me2 relative to H3K9me2). Thus, these histones represented a more 'real-world' unpredictable scenario than the highly controlled peptide substrates.

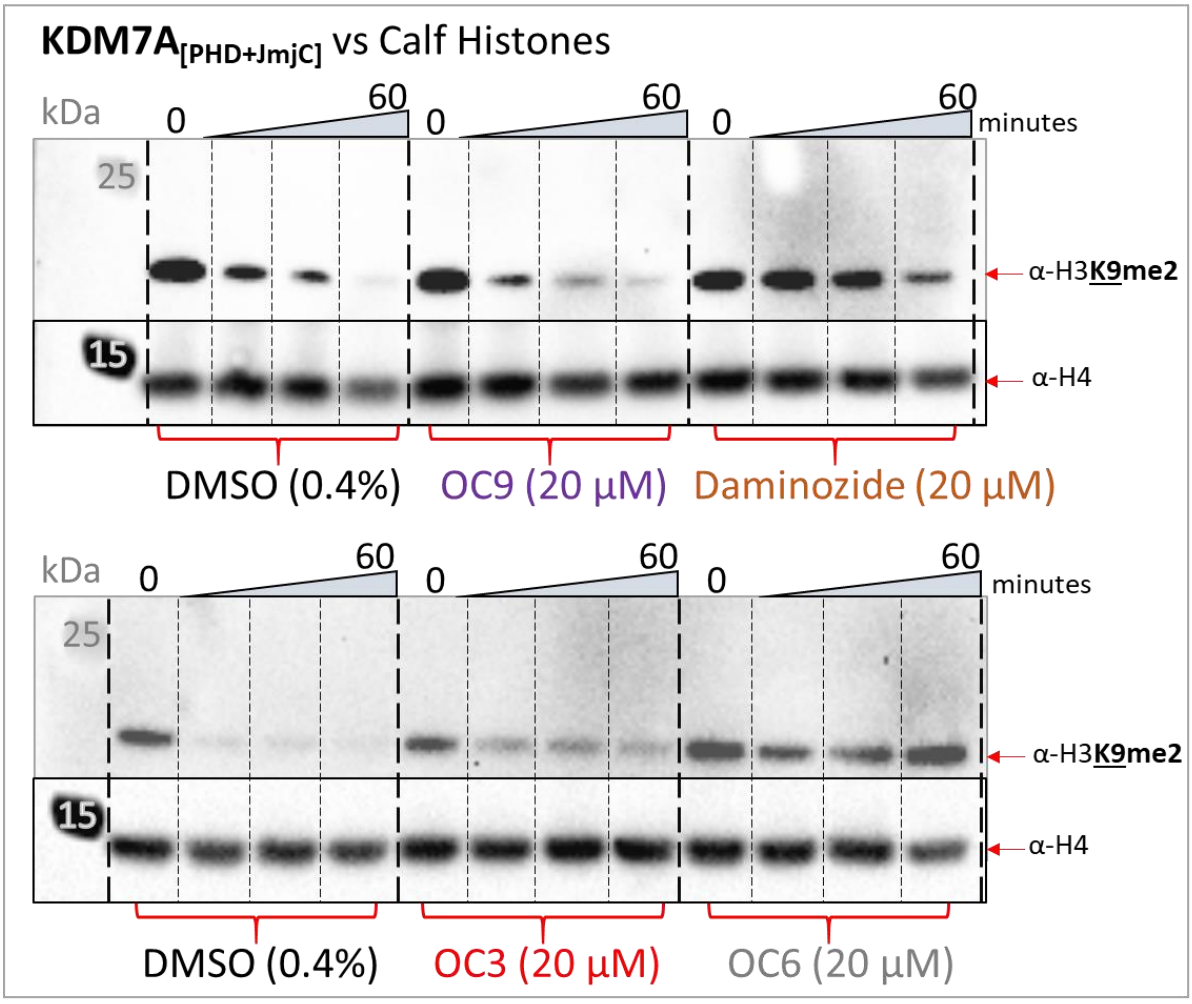
Histone extract demethylation assays were optimised and conducted in assay conditions analogous to the MALDI-TOF-MS assays (i.e. a single concentration of compound in >10-fold excess of the enzyme was used). Assay samples were analysed by SDS-PAGE, and subsequent transfer to a nitrocellulose membrane for antibody detection by Western Blot analysis of any changes in the level of the target H3 lysine methyl mark over time. Each blot was also then analysed for levels of histone H4 as a loading control for comparison, to support any observation of changes in the levels of histone H3-based marks being due to genuine effects of the compounds. Elevated levels of 2OG were used to better mimic conditions within a cell [153] (see Materials and Methods for full details).

Incubation of KDM7B$_{[PHD+JmjC]}$ with calf-thymus histone extracts and detection of H3K9me2 levels over time evidenced recapitulation of OC9 inhibition of demethylation, qualitatively judged to again be similarly effective to Daminozide (**F.3.5.iii.**). OC6 also appeared to retain an inhibitory effect, however OC3 failed to demonstrate significant inhibition of KDM7B$_{[PHD+JmjC]}$ under these conditions. Further work is required to verify whether OC3 efficacy against KDM7B$_{[PHD+JmjC]}$ translates to a full histone context.



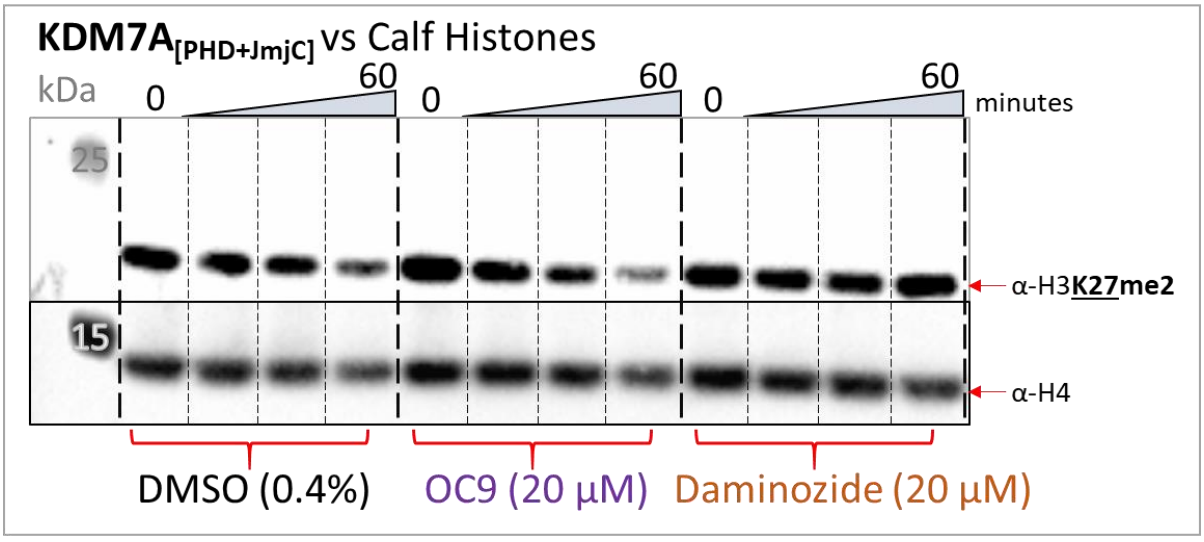
F.3.6.i. Assessment of cyclic peptide effects on KDM7B$_{[PHD+JmjC]}$ demethylation of H3K9me2 on full-length calf-histones. KDM7B$_{[PHD+JmjC]}$ (1 µM) demethylation of calf-thymus histone extracts (50 µg/mL) was challenged with a single concentration of compound (20 µM) to investigate compound effects on demethylase activity in the presence of 100 µM Fe(II), 1 mM 2OG and 1 mM NaAsc. Western Blot analysis was used to monitor the substrate H3K9me2 mark levels over time and the membrane was subsequently analysed for H4 levels as a loading control. Full conditions are given in Materials and Methods.

Incubation of KDM7A$_{[PHD+JmjC]}$ with calf-thymus histone extracts and detection of H3K9me2 levels over time showed that OC9 remained an activator of KDM7A$_{[PHD+JmjC]}$ demethylation of H3K9me2 even on full-length histones, with Daminozide continuing to serve as a functional inhibition control (**F.3.6.ii.**). This suggested that OC9 disruption of the PHD-finger to mediate inhibition of JmjC turnover remains relevant in a more complex epigenetic scenario, bolstering the prospects of OC9 serving as a useful chemical probe in a genuine biological context in future experiments. OC6 appeared to inhibit KDM7A$_{[PHD+JmjC]}$ demethylation of H3K9me2 on histones, and OC3 gave possible minor inhibition by qualitative assessment.



F.3.6.ii. Assessment of cyclic peptide effects on KDM7A$_{[PHD+JmjC]}$ demethylation of H3K9me2 on full-length calf-histones. KDM7A$_{[PHD+JmjC]}$ (0.5 μM) demethylation of calf-thymus histone extracts (50 μg/mL) was challenged with a single concentration of compound (20 μM) to investigate compound effects on demethylase activity in the presence of 100 μM Fe(II), 1 mM 2OG and 1 mM NaAsc. Western Blot analysis was used to monitor the substrate H3K9me2 mark levels over time and the membrane was subsequently analysed for H4 levels as a loading control. OC9 and Daminozide were assessed in a separately performed assay to OC3 and OC6. Full conditions are given in Materials and Methods.

KDM7A has also been reported as an H3K27me2 demethylase in cells [56]. Given the ability of OC9 to modulate KDM7A$_{[PHD+JmjC]}$ activity against H3K9me2, OC9 was also investigated for modulation of KDM7A$_{[PHD+JmjC]}$ demethylation of H3K27me2 on histone extracts (**F.3.6.iii.**). Demethylation of H3K27me2 appeared much weaker in the DMSO-only control than for H3K9me2 under the same conditions. Nevertheless, OC9 appeared to slightly boost the level of demethylation towards the end of the assay time-window (60 minutes) and Daminozide appeared inhibitory; future experiments would seek to refine this for a more definitive answer. Use of H3K4me3K27me2 peptide would also be desired for proof-of-principle, as it would provide a more controlled cross-talk than is possible with the histone extracts; however, the length of this peptide (> 27mer) makes it more challenging to synthesise and so was not investigated within the timeframe of this work.



F.3.6.iii. Assessment of cyclic peptide effects on KDM7A$_{[PHD+JmjC]}$ demethylation of H3K27me2 on full-length calf-histones. KDM7A$_{[PHD+JmjC]}$ (0.5 µM) demethylation of calf-thymus histone extracts (50 µg/mL) was challenged with a single concentration of compound (20 µM) to investigate compound effects on demethylase activity in the presence of 100 µM Fe(II), 1 mM 2OG and 1 mM NaAsc. Western Blot analysis was used to monitor the substrate H3K27me2 mark levels over time and the membrane was subsequently analysed for H4 levels as a loading control. Full conditions are given in Materials and Methods.

3.7. Conclusions and Future Work

The binding of each cyclic peptide (OC3, OC6, OC9) to the KDM7s (Chapter 2) translated into in-vitro histone peptide displacement that was more effective than H3K4me3 self-displacement, which had different effects on KDM7A,B$_{[PHD+JmjC]}$ demethylation activities. OC9 was able to displace H3K4me3 [1-21] peptide binding to all three KDM7A,B,C$_{[PHD]}$ and KDM7A,B,C$_{[PHD+JmjC]}$ with low nanomolar IC_{50}. This resulted in OC9 possessing low micromolar potency EC_{50} enhancement of KDM7A$_{[PHD+JmjC]}$ and IC_{50} inhibition of KDM7B$_{[PHD+JmjC]}$ demethylation against both H3K4me3K9me2 and H3K9me2 [1-21] peptide substrates. These catalytic impacts of OC9 were also demonstrated against full-length, unpredictably modified, calf-thymus histone extracts. Thus, OC9 serves to disrupt the KDM7$_{[PHD]}$-H3K4me3 protein-protein interaction, which is important in epigenetic regulation, to access a novel mechanism for allosteric control of KDM7A,B demethylation. Accordingly, OC9 represents an achievement as – to the best of our knowledge – it is the first potent PHD finger inhibitor.

OC3 was shown to displace H3K4me3 [1-21] peptide from KDM7B$_{[PHD+JmjC]}$ with nanomolar IC_{50} potency, and to function as an inhibitor with low micromolar potency IC_{50} of KDM7B$_{[PHD+JmjC]}$ demethylation of H3K4me3K9me2 and H3K9me2 [1-21] peptide substrates. Further work is needed to resolve the efficacy of OC3 on KDM7B$_{[PHD+JmjC]}$ against full-length histones. Site-specific installation of H3K4me3 and H3K9me2 on full-length histones using flexizyme (Chapter 1) may provide useful reagents for such studies. OC6 displaced H3K4me3 [1-21] peptide from KDM7B$_{[PHD+JmjC]}$ with nanomolar potency IC_{50}, apparently having a low micromolar potency IC_{50} inhibition of KDM7B$_{[PHD+JmjC]}$ demethylation against H3K4me3K9me2 and H3K9me2 [1-21] peptides, and inhibiting KDM7B$_{[PHD+JmjC]}$ against full-length histones. The peculiar AlphaScreen assay behaviour exhibited by OC6 at higher concentrations would require further investigation to understand the cause (see Chapter 7, 9). Further work is also

needed to fully assess the selectivity of OC3 and OC6 towards inhibition of KDM7A$_{[PHD+JmjC]}$ demethylation of H3K9me2 in both peptide and full-length histone contexts, since conflicting observations were made regarding the levels of inhibition.

Future work could include detailed kinetic studies to attempt classify the inhibitory mode of each cyclic peptide (e.g. competitive, uncompetitive, non-competitive or mixed [245]). Currently, the cyclic peptides are not believed to directly bind to the JmjC active site (no displacement of methylstat was observed), so they may be classifiable as allosteric inhibitors. However, the binding interaction of H3K4me3 with the PHD finger also contributes to JmjC activity, and it is not yet known precisely how the cyclic peptides effect the displacement of H3K4me3; i.e. whether it is through direct competition for the aromatic cage, or an indirect consequence of conformational change. It is thus a complex scenario.

These functional investigations meant that OC9 was designated as the lead chemical probe candidate for its inhibition of the KDM7 PHD finger, notable for meeting one of the project's primary objectives, and OC9 forms the focus of the following work described within this thesis. The next phase of research assessed the wider selectivity profile of OC9, and also OC3 and OC6 (Chapter 4).

Chapter 4. Cyclic Peptide Selectivity Studies

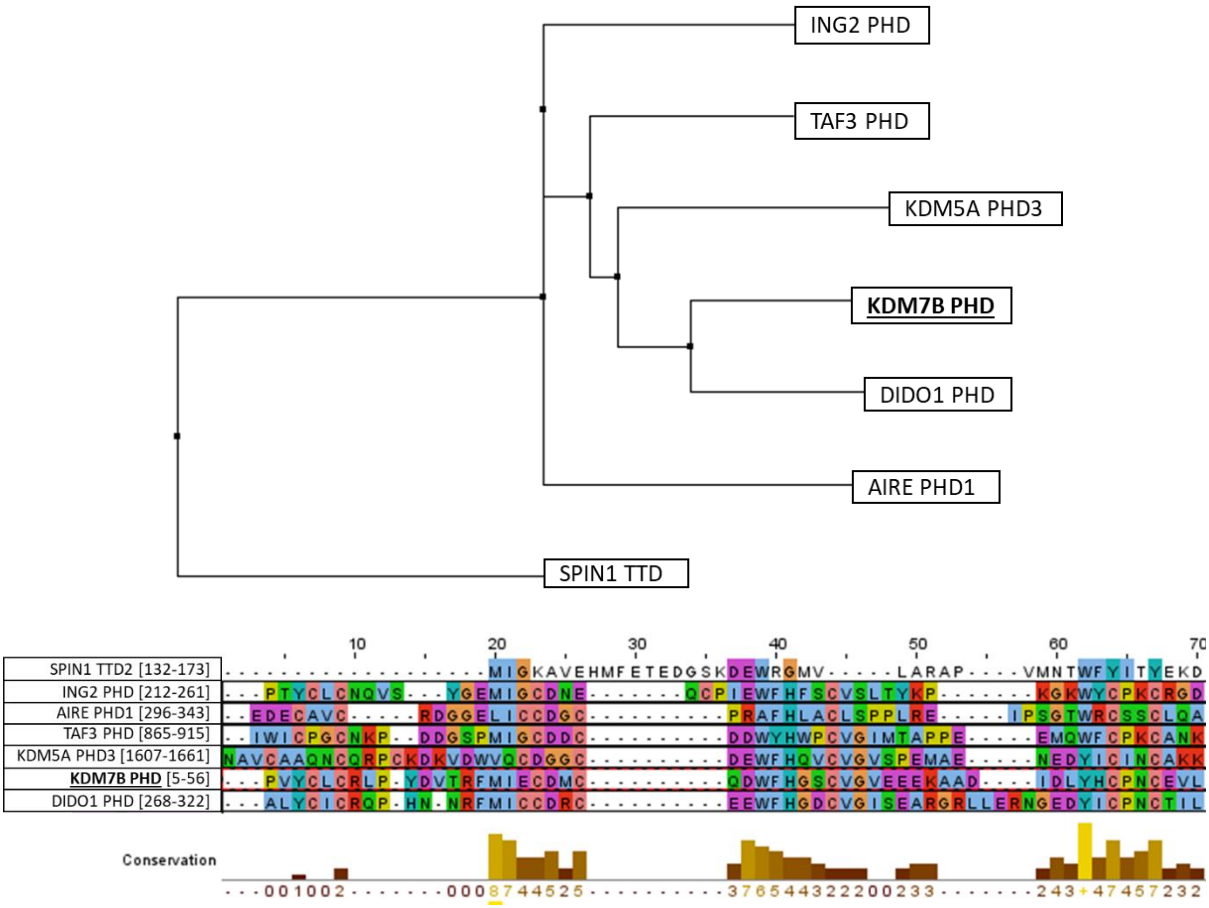
4.1. Introduction

The cyclic peptides described in Chapters 2 & 3 represented promising candidates as chemical probes for the KDM7 sub-family, meeting and exceeding the suggested affinity criteria as set out by the Structural Genomics Consortium (SGC): K_D or IC_{50} < 100 nM (Chapter 0). Another important criterion for a chemical probe is selectivity for its intended target [134]. Therefore, it was essential to build a picture of the cyclic peptide selectivity profiles by binding (K_D), histone peptide displacement, and KDM catalytic activity studies. The selectivity profiling focused more extensively on OC9 than OC3 and OC6, as OC9 represented the first unique PHD finger binding ligand; previous attempts at creating PHD finger specific chemical probes had resulted in very poor selectivity (Chapter 0). Significant challenges have also been encountered in the development of KDM7 specific chemical probes (Chapter 0). Thus, the objective of the work described in this chapter was to establish a panel of PHD-fingers and KDMs for cross-screening, and to characterise the cyclic peptides for PHD finger and KDM selectivity. Targeting PHD fingers in particular posed a significant challenge to attaining selectivity and, as the cyclic peptides could potentially represent an adaptable scaffold for targeting them, it was important to thoroughly characterise off-target target binding effects against related non-KDM7 proteins.

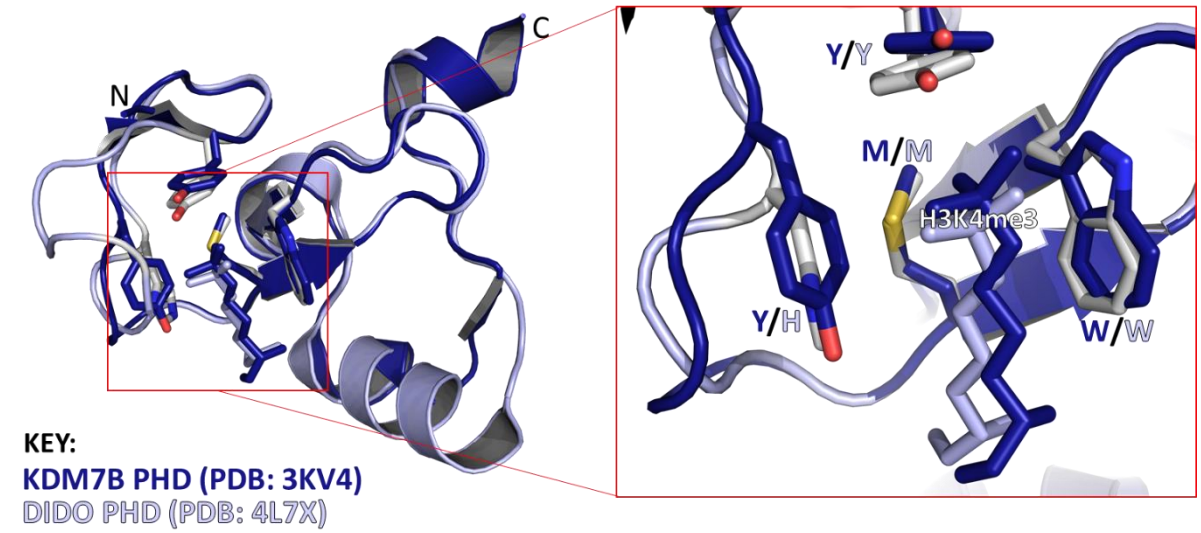
4.2. Binding vs Non-KDM7 Reader Domains

Several of the reader domain proteins used in cross screening (ING2, SPIN1, TAF3) were gratefully received from the SGC. KDM5A$_{[PHD3]}$ was donated by Dr Bhaskar Bhushan.

Several reader domains were selected for cyclic peptide cross screening, including five PHD fingers (DIDO1, KDM5A, TAF3, ING2, AIRE) and one Tandem Tudor domain (TTD) (SPIN1$_{[TTD]}$). All of the reader domains selected, except for AIRE$_{[PHD1]}$, are known to bind to H3K4me3 – as does KDM7B$_{[PHD]}$. Their canonical sequence alignments to KDM7B$_{[PHD]}$ are shown (**F.4.2.i.**).The H3K4me3-binding DIDO1$_{[PHD]}$ is notable for being the most closely homologous known PHD finger to the KDM7$_{[PHD]}$, (canonical sequence 46 % identity, 66 % similarity). The aromatic cage of the DIDO1$_{[PHD]}$ differs to that of the KDM7$_{[PHD]}$ only by a tyrosine to histidine switch (**F.4.2.ii.**). Testing of these reader domains ensured a robust challenge to the PHD finger selectivity of OC9, since any simple mimicry of the cognate H3K4me3 binding interaction (e.g. through an arginine side chain [155]) would be more likely to be detected as an off-target interaction.

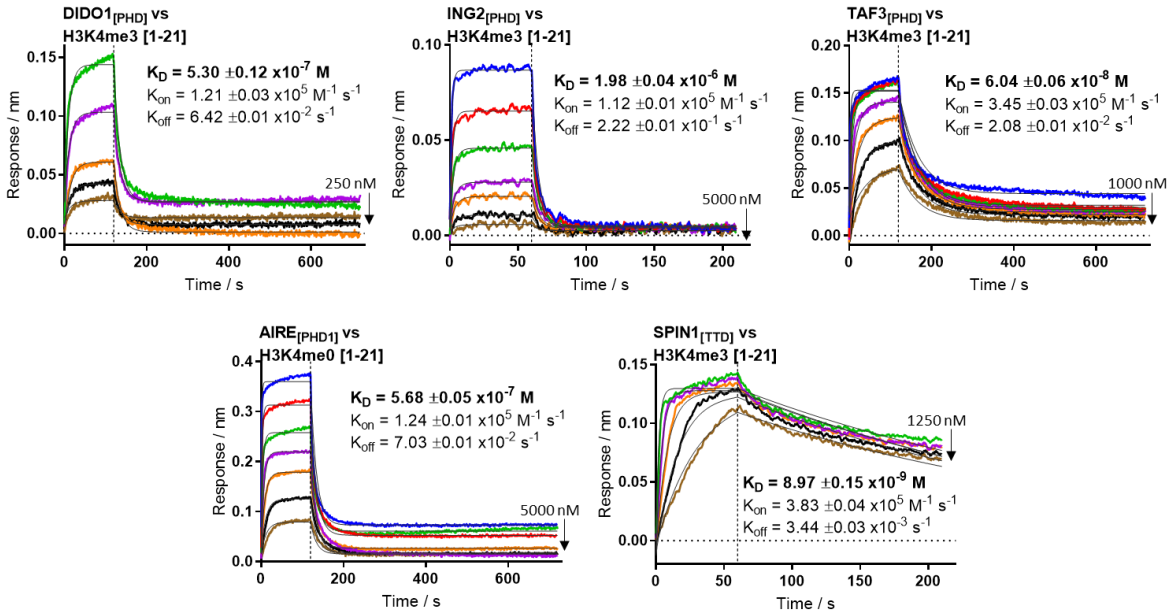


F.4.2.i. Human reader domains used for cross-screening of cyclic peptide probes. The canonical domain sequences (UniProt) were aligned using Clustal Omega [241] and visualised using JalView. DIDO1$_{[PHD]}$ is the most closely related non-KDM7 PHD finger to that of the KDM7s. Amino acids are coloured by similar physical properties.

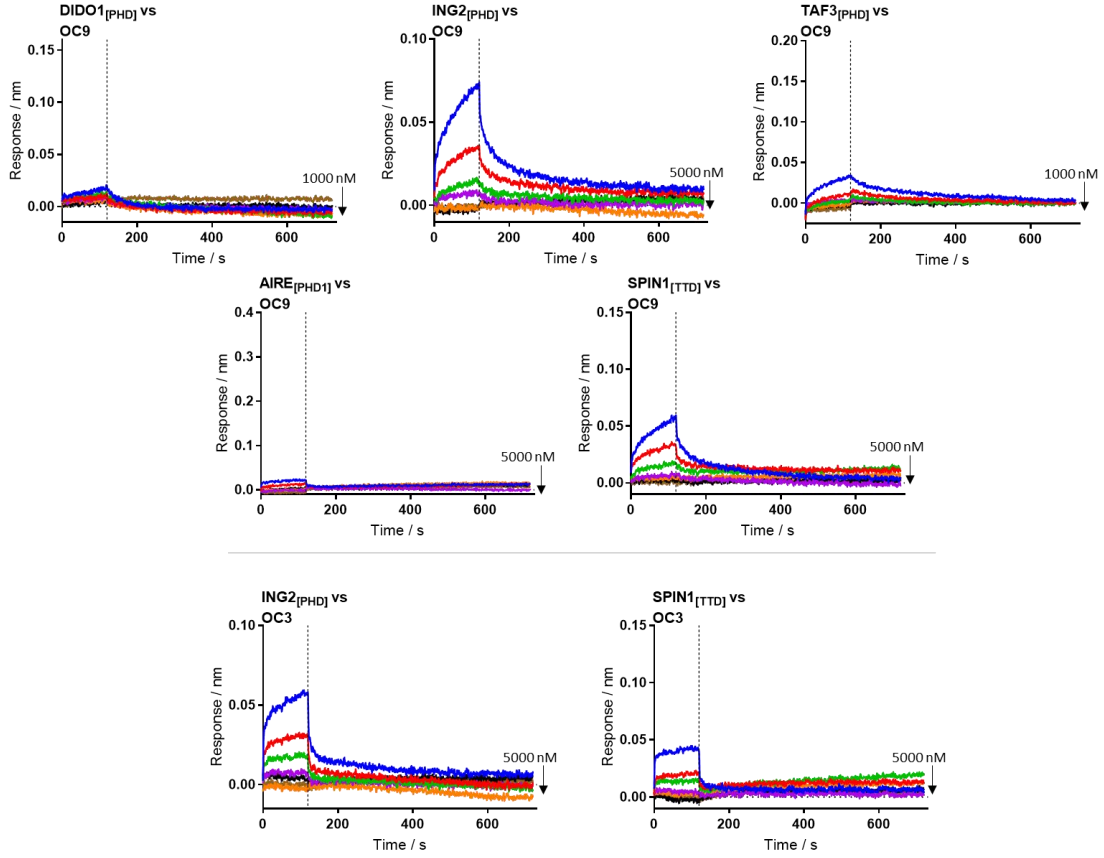


F.4.2.ii. Overlay of views from the crystal structures of the PHD fingers of KDM7B (PDB:3KV4) and DIDO (PDB:4L7X) when in complex with their cognate HK4me3 peptide binding partner. The structures overlay closely, especially around the region of the trimethyllysine-binding aromatic cage. DIDO possesses a histidine residue on one side of its aromatic cage whereas KDM7B instead possesses a tyrosine.

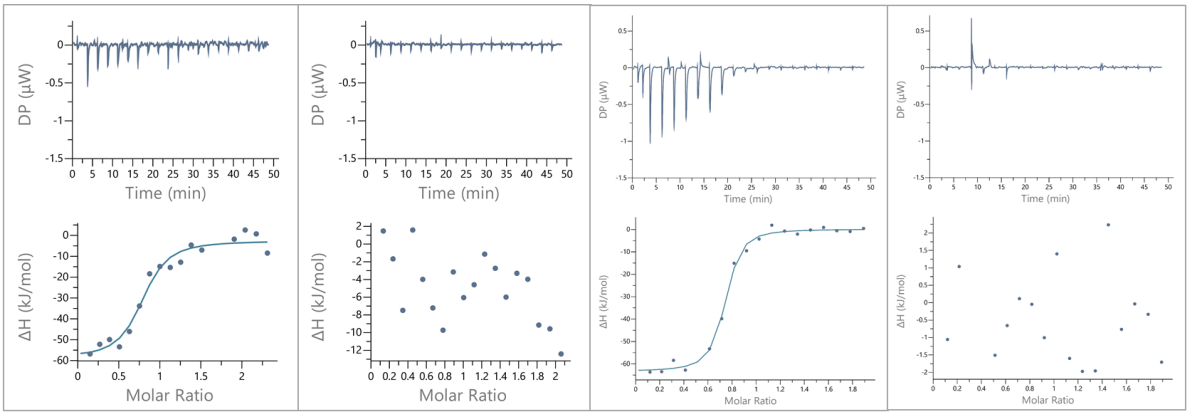
BLI was used to characterise binding of OC9 to four of the PHD fingers and the TTD, and OC3 to one PHD finger and the TTD. Each of the reader domain proteins possessed their expected high nanomolar to low micromolar binding affinity towards their cognate histone peptide partner (**F.4.2.iii.**), indicating they were functioning correctly whilst immobilised. Neither OC3 nor OC9 displayed substantial binding interactions with the panel of reader domains that could be meaningfully fitted by the Association then Dissociation model (**F.4.2.iv.**). However, at the highest concentrations of OC3 and OC9 tested (in the low micromolar region) they began to exhibit a level of increased response signal in some cases (against $ING2_{[PHD]}$, $SPIN1_{[TTD]}$, $TAF3_{[PHD]}$), but this was considered to be most likely due to non-specific binding. This was supported by the lack of observed binding by OC9 to the $DIDO1_{[PHD]}$ being confirmed by ITC (**F.2.4.v.**), and OC9 also did not bind by ITC to $KDM5A_{[PHD3]}$ (**F.2.4.v.**), a representative H3K4me3-binder from the JmjC-KDMs. Similar BLI data for OC6 is not reported due to its propensity for interacting with the sensor apparatus fouling attempted observations (Chapter 2). The lack of binding to these reader domains was further examined by AlphaScreen displacement assay (section 4.3.).



F.4.2.iii. Cognate histone peptide binding affinities to reader domains were determined by BLI. They were found to be in the low nanomolar to low micromolar range. Proteins were immobilised by His-tag on Ni-NTA sensors. Each coloured trace represents the response for a peptide concentration (2-fold dilution series). The fitted data, overlaid in black where appropriate fitting with the 'Association then Dissociation' model (Prism v7) could be performed, were used to determine the binding affinity, K_D, and K_{on} and K_{off}. The start of dissociation is indicated with a dashed vertical line.



F.4.2.iv. OC9 and OC3 were screened for binding to non-KDM7 reader domains by BLI. They were found not to exhibit specific binding to any of the proteins up to the low micromolar concentrations tested. Proteins were immobilised by His-tag on Ni-NTA sensors. Each coloured trace represents the response for a peptide concentration (2-fold dilution series). Data could not be fitted with the 'Association then Dissociation' and were instead by the One-Site Total Binding model (Prism v7) (**F.4.2.v.**). The start of dissociation is indicated with a dashed vertical line.



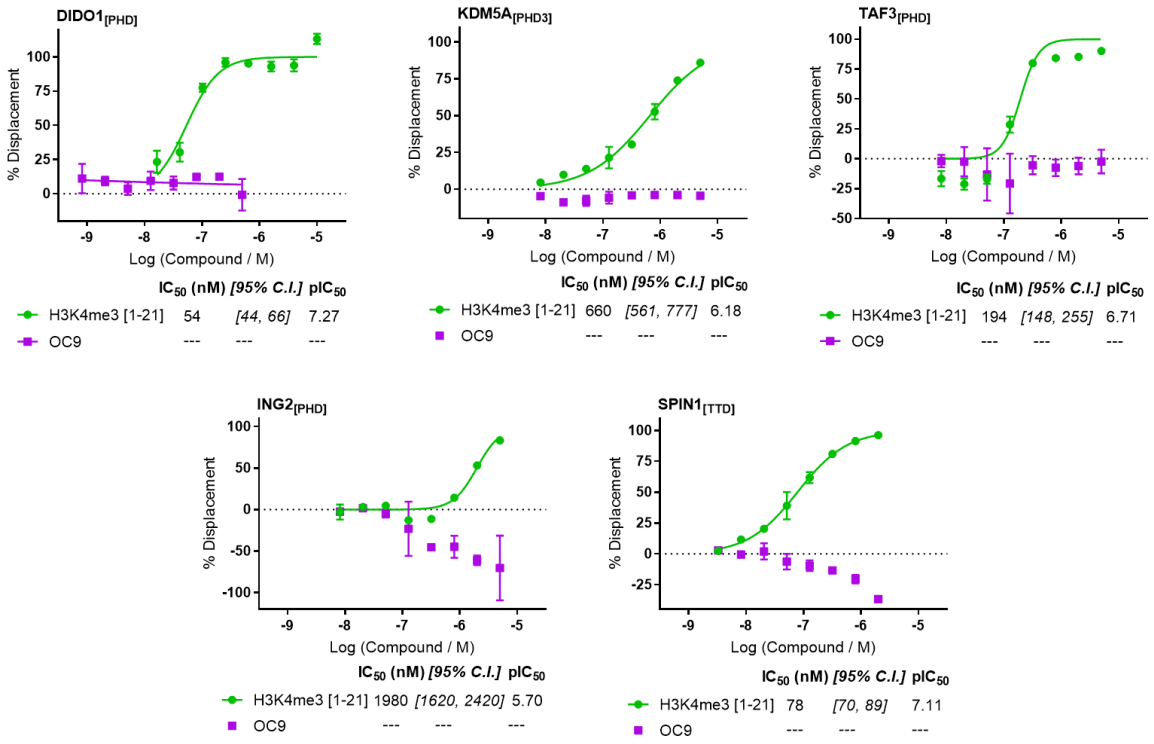
KDM5A$_{[PHD3]}$ vs H3K4me3 [1-21]	**KDM5A$_{[PHD3]}$ vs OC9**	**DIDO1$_{[PHD]}$ vs H3K4me3 [1-21]**	**DIDO1$_{[PHD]}$ vs OC9**
K_D = 145 ± 69 nM	K_D = ---	K_D = 57.3 ± 13.9 nM	K_D = ---
pK_D = 6.84		pK_D = 7.24	
N = 0.76 ± 0.04		N = 0.70 ± 0.01	
ΔH = -57.3 ± 4.9 kJ mol^{-1}		ΔH = -63.5 ± 1.5 kJ mol^{-1}	
TΔS = -18.4 kJ mol^{-1}		TΔS = -22.1 kJ mol^{-1}	
ΔG = -39.0 kJ mol^{-1}		ΔG = -41.4 kJ mol^{-1}	

F.4.2.v. OC9 was tested for binding to KDM5A$_{[PHD3]}$ and DIDO1$_{[PHD]}$ by ITC. The KDM5A$_{[PHD3]}$ and DIDO1$_{[PHD]}$ bound to its cognate H3K4me3 binding partner with high nanomolar affinity, but did not show any discernible affinity of binding to OC9. See Materials and Methods for experimental details.

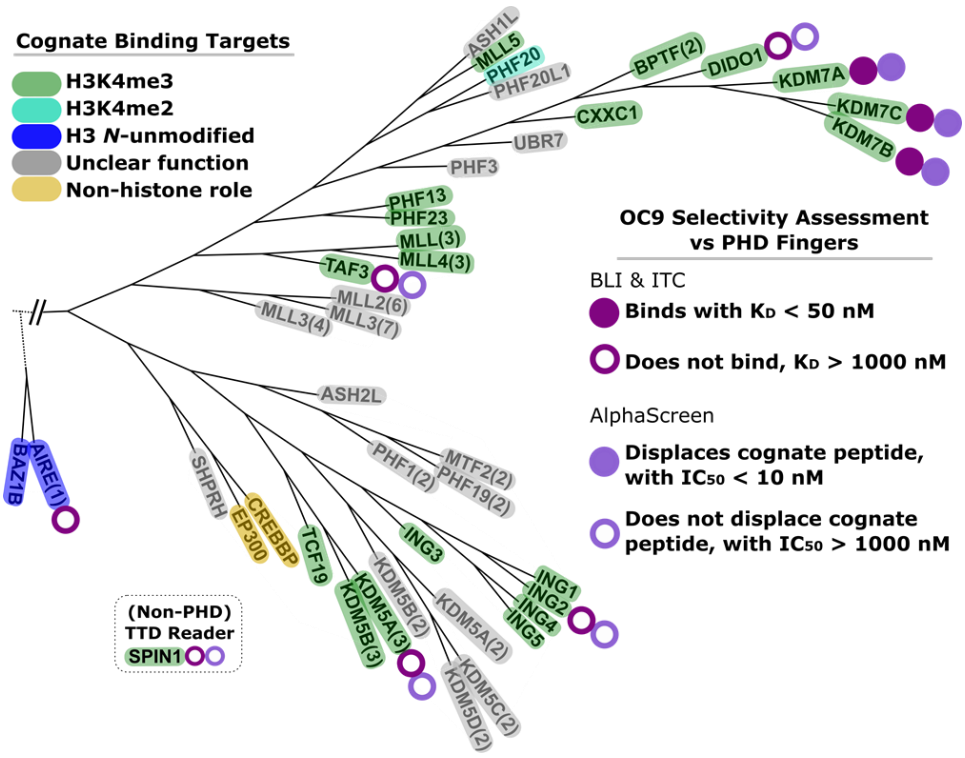
4.3. Displacement vs Non-KDM7 Reader Domains

The apparent binding selectivity demonstrated by OC9 was next corroborated by testing in a displacement assay against a panel of H3K4me3-binding reader domains ($DIDO1_{[PHD]}$, $KDM5A_{[PHD3]}$, $TAF3_{[PHD]}$, $ING2_{[PHD]}$, $SPIN1_{[TTD]}$) by AlphaScreen (described in Chapter 3). The assay set-up was validated using a non-biotinylated H3K4me3 [1-21] self-displacement control. The non-biotinylated-H3K4me3 peptide displaced the biotinylated peptide with expected IC_{50} potencies in all cases. OC9 did not displace H3K4me3 [1-21] peptide from $DIDO1_{[PHD]}$ or any of the other three H3K4me3-binding PHD fingers and the TTD, up to the low micromolar concentrations tested (**F.4.3.i.**). This observation, together with the BLI results (section 4.2.), substantiated an excellent reader domain selectivity profile for OC9 (**F.4.3.ii.**) and suggested that OC9 was not interacting with $KDM7_{[PHD]}$ by simple mimicry of the H3K4me3 interaction.

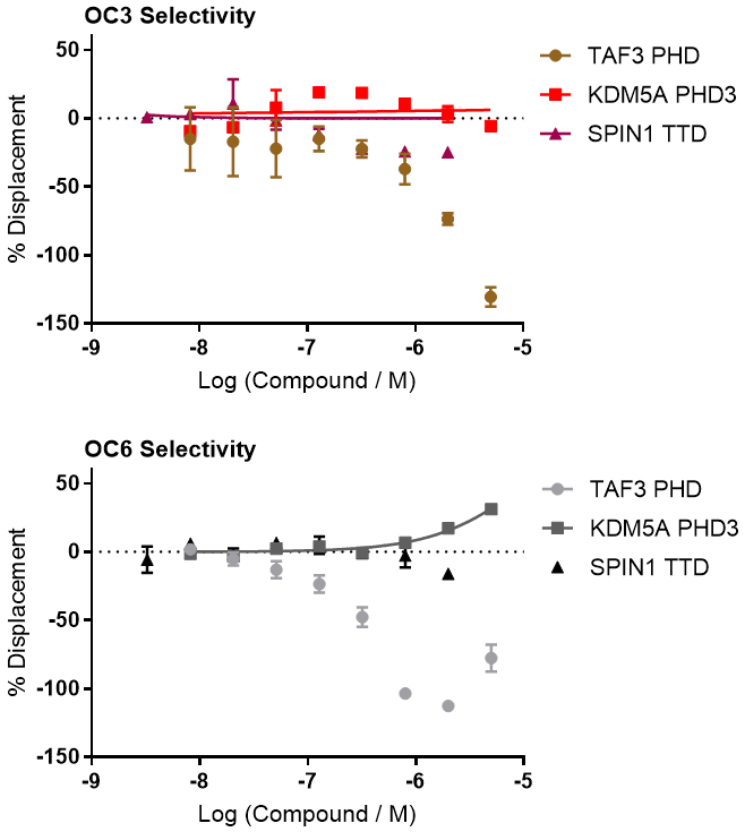
A peculiar observation was noted at higher concentrations of OC9 against $ING2_{[PHD]}$ and $SPIN1_{[TTD]}$, as the AlphaScreen signal appeared to increase (shown as 'negative' displacement on the graphs). This would ordinarily indicate an 'enhancement' of histone-peptide binding; however, such an effect would seem unlikely in this case given the lack of specific binding demonstrated by OC9 to either protein by BLI. It may instead be an effect of non-specific protein-interaction at higher peptide concentrations, as both OC9 and OC3 exhibited non-specific BLI signal increase for these two reader domain proteins around micromolar cyclic peptide concentrations. Indeed OC3, and OC6, were also found to exhibit a similar magnitude 'enhancement' effect on $SPIN1_{[TTD]}$, as well as $TAF3_{[PHD]}$, binding signal by the AlphaScreen displacement assay, but neither displaced $KDM5A_{[PHD3]}$ (**F.4.3.iii.**).



F.4.3.i. OC9 selectivity towards several H3K4me3-binding reader domains was assayed by AlphaScreen displacement. Non-biotinylated H3K4me3 [1-21] served as a positive control for displacement of its biotinylated counterpart. OC9 did not displace biotinylated cognate histone peptide from the tested domains. Conditions are given in Materials and Methods.



F.4.3.ii. Phylogenetic tree of human PHD-fingers and the selectivity profile of OC9. The OC9 binding and displacement selectivity profile is annotated on a PHD finger phylogenetic tree, derived from PHD finger domain alignments, adapted from SGC ChromoHub [247].



F.4.3.iii. OC3 and OC6 selectivity towards several H3K4me3-binding reader domains was assayed by AlphaScreen displacement. Non-biotinylated H3K4me3 [1-21] served as a positive control for displacement of its biotinylated counterpart. OC3 and OC6 did not displace biotinylated cognate histone peptide from the tested domains. Conditions are given in Materials and Methods.

4.4. Effect on Non-KDM7 Enzyme Activity

Several of the KDMs used in cross screening (KDM3, KDM4, KDM5, KDM6) were gratefully received from other members of the Schofield and Kawamura labs (Eidarus Salah, Dr Rachel Schiller) and EGLN1 was from Dr Tom McAllister (Kawamura Lab).

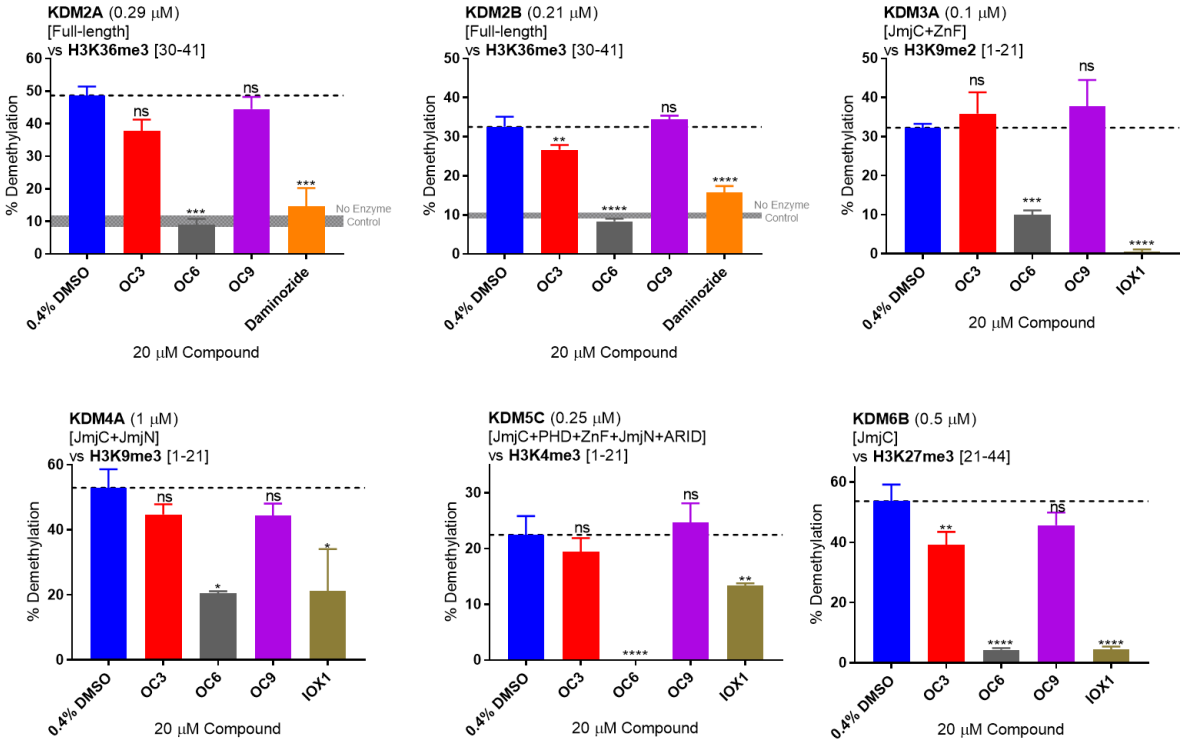
A further important assessment of cyclic peptide selectivity was to determine their effect, if any, on the demethylase activity of a representative member from each JmjC-KDM sub-family (KDM2A, KDM2B, KDM3A, KDM4A, KDM5C, KDM6B, KDM7A, KDM7B). Developing compounds with selectivity between JmjC-KDM families in general has proven challenging (Chapter 0). Previous attempts to develop a KDM7 specific inhibitor had fallen short with off-target efficacy most commonly against the KDM2 sub-family [160,162]. The KDM selectivity panel herein sought to address these concerns for the cyclic peptides by testing them against the JmjC and surrounding functional domains (**F.4.4.i.**), with some constructs including PHD and zinc fingers.



F.4.4.i. KDM domain architectures and constructs used for cross-screening assays. The JmjC-KDMs and the domains contained within their recombinant constructs (indicated with a red-line) used for cross screening by MALDI-TOF-MS assay are labelled. Domain maps are adapted from InterPro [248] and verified with UniProt (Human KDM isoforms). Amino acid limits for the constructs are given in the Appendix.

The primary cross-screening of JmjC-KDM demethylation activities was performed by MALDI-TOF-MS assay at a single concentration of cyclic peptide (20 μM). A low concentration of KDM enzyme was tested against a high concentration of cyclic peptide to ensure the best chance of uncovering any propensity for off-target activity (**F.4.4.ii.**). Importantly, the panel included the full-length KDM2A and KDM2B proteins, the closest sub-family to the KDM7s. To date, all small molecules that inhibit the KDM7s also inhibit the KDM2s (Chapter 0). Targeting the $KDM7_{[PHD]}$ provides a potentially more selective target for inhibition than the JmjC domain – the $KDM2_{[PHD]}$ are highly dissimilar to $KDM7_{[PHD]}$ (**F.4.4.iii.**) and are not believed to play as crucial a role in their JmjC activity. This dissimilarity was important, as it meant the approach used in this work for cyclic peptide development against KDM7s enabled the well-established KDM2/7 inhibitor selectivity issues to be tackled.

Pleasingly, the $KDM7_{[PHD]}$ targeting OC9 did not inhibit either KDM2, or any of the other KDMs tested. The $KDM7B_{[PHD+JmjC]}$-targeting OC3 also proved highly selective for the KDM7 sub-family over the other KDMs, giving only weak inhibition against KDM2B and KDM6B under the tested conditions. The issue of KDM2/7 selectivity was emphasised by the near-complete inhibition of both KDM2A,B proteins, as well as the KDM7s, in the same MALDI-TOF-MS based assay by small molecule Daminozide. On the other hand, the $KDM7B_{[PHD+JmjC]}$-targeting OC6 displayed a near complete inhibition of all panel members except KDM4A, signifying either a non-specific general protein disruption or a potential capability to disrupt a well-conserved site on the JmjC domain.



F.4.4.ii. A panel of JmjC-KDM enzymes were tested by MALDI-TOF-MS based assay (see Materials and Methods) for demethylase activity against their respective histone H3 peptide substrates. Each enzyme was challenged with 20 μM of a cyclic peptide (OC3, OC6, OC9) and a control small molecule inhibitor (Daminozide or IOX1). An ordinary one-way ANOVA with Dunnett's multiple comparison was used to test for statistical significance of the compound means relative to the 0.4 % DMSO control (multiplicity adjusted p values = ns>0.0332, *0.0332, **0.0021, ***0.0002, ****<0.0001).

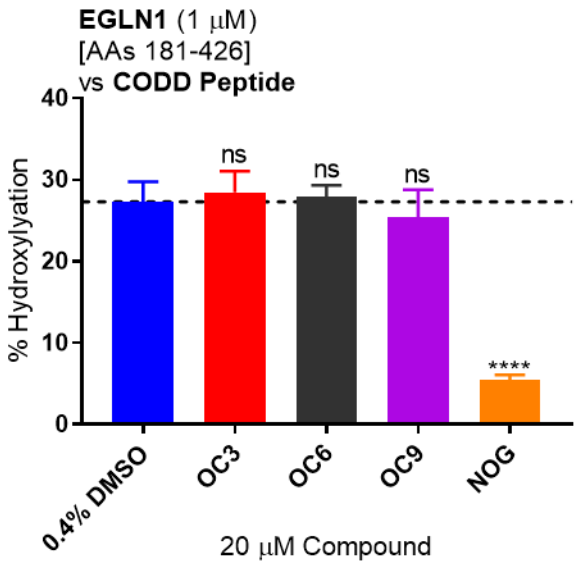


F.4.4.iii. Alignment of the KDM7B and KDM2A,B canonical PHD finger sequences. The canonical PHD finger sequences for KDM7B, KDM2A, KDM2B were aligned using Clustal Omega [241] and visualised with JalView. The KDM7B PHD finger amino acid sequence is highly dissimilar to those of KDM2A and KDM2B.

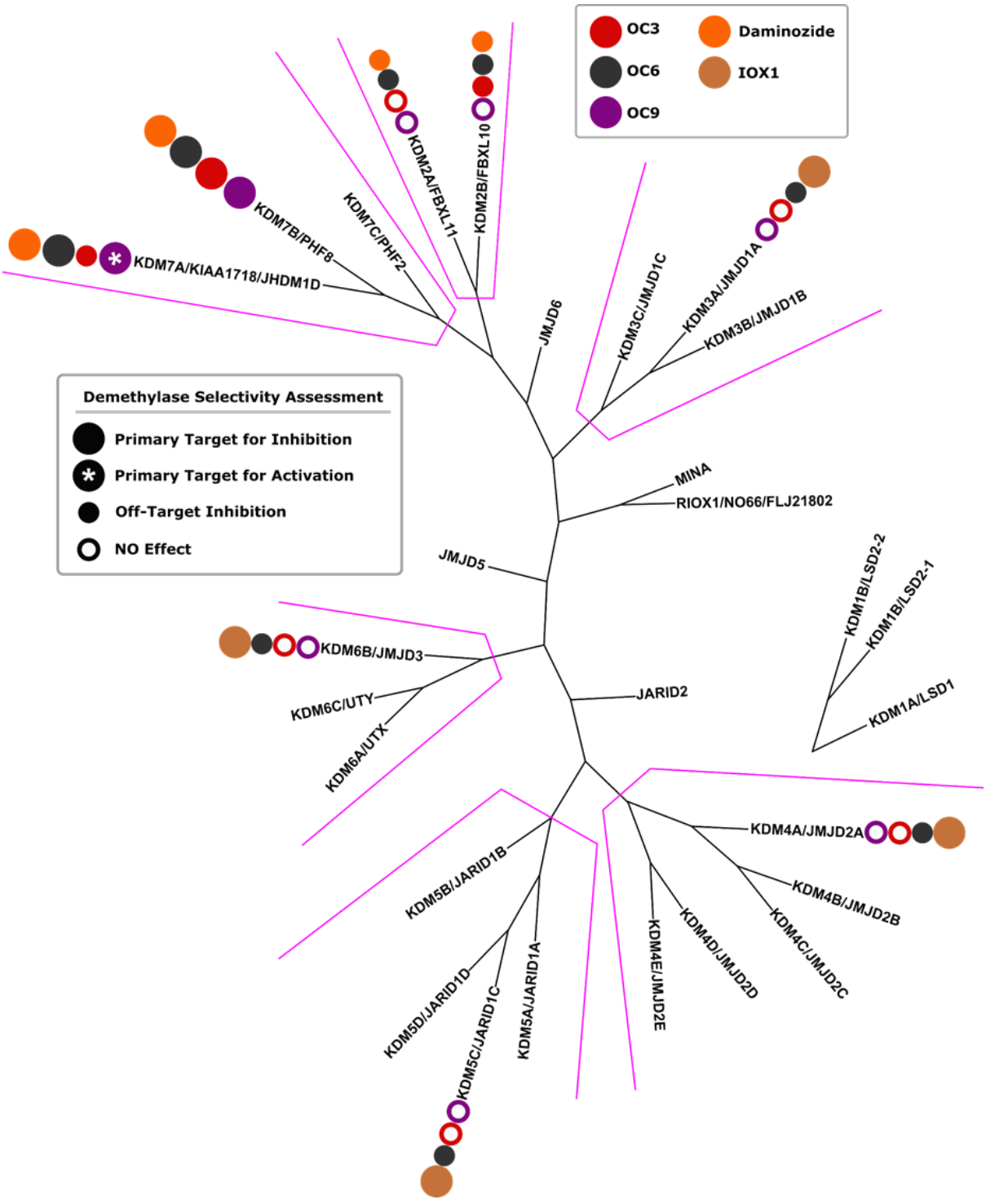
To test whether inhibition by OC6 was due to a non-specific mechanism of action, all three cyclic peptides were assayed against a different 2OG oxygenase enzyme (prolyl-hydroxylase isoform 2, aka EGLN1), using the same MALDI-TOF-MS assay as for the JmjC-KDM cross-screening. EGLN1 is important in hypoxia-response regulation and operates through an Fe(II), 2OG dependent mechanism, similar to the JmjC-KDMs, to hydroxylate key proline residues on its HIF1-α protein substrate [164]. The MALDI-TOF-MS assay (**F.4.4.iv.**) made use of an established modified-peptide substrate, C-terminal oxygen-dependent degradation domain

(CODD – sequence given in the Materials and Methods) of HIF1-α. None of the cyclic peptides (OC3, OC6, OC9) inhibited EGLN1 activity, indicating that OC6 inhibition of the JmjC-KDMs was less likely to be due to a non-specific mechanism of protein disruption.

The overall in-vitro JmjC-KDM demethylase selectivity profiles of the cyclic peptides are summarised (**F.4.4.v.**). It is notable that OC9 possesses robust selectivity for the KDM7 sub-family and has a superior selectivity profile to those previously demonstrated by other cyclic peptide and small molecule efforts against KDM7s and PHD fingers (Chapter 0).

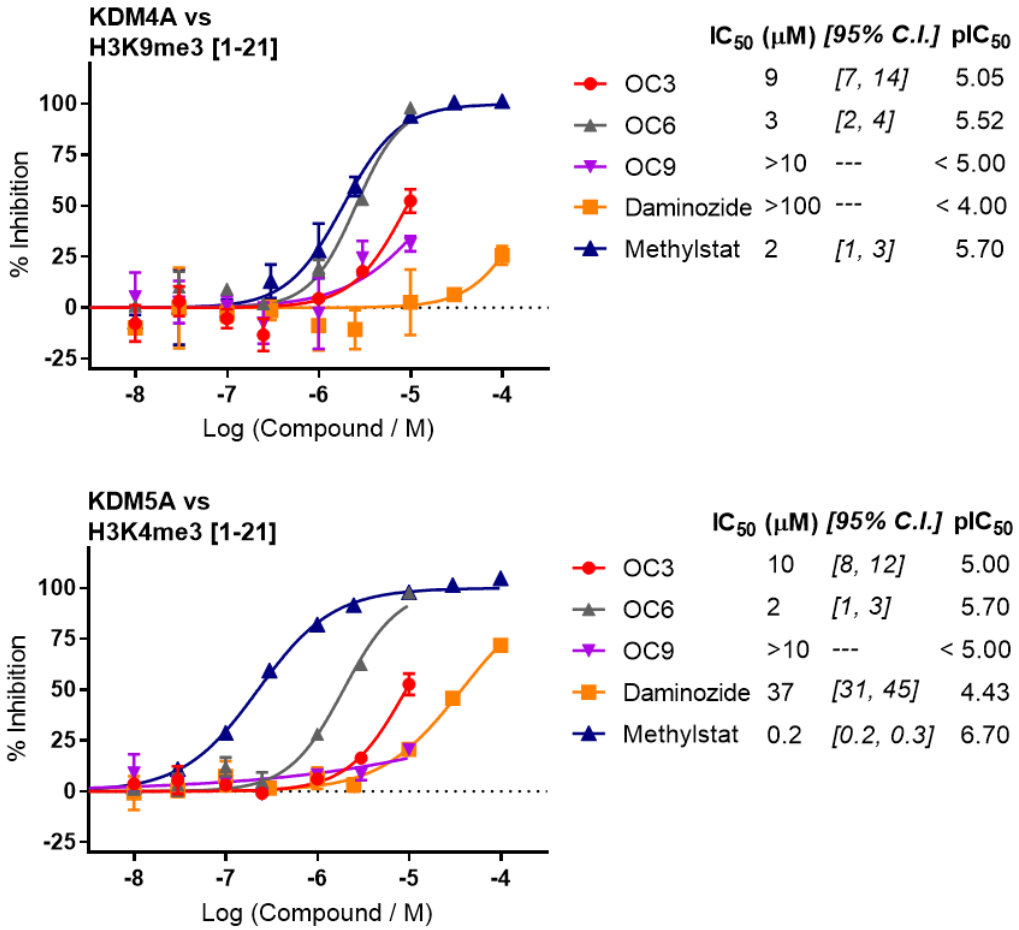


F.4.4.iv. Cyclic peptide selectivity assessment against a non-JmjC-KDM 2OG oxygenase. Prolyl-hydroxylase isoform 2 (EGLN1, 1 μM) was tested by MALDI-TOF-MS assay for hydroxylation of HIF1-α CODD peptide substrate fragment (see Materials and Methods) in the presence of an excess of each cyclic peptide and a broad-spectrum 2OG oxygenase inhibitor, NOG (N-oxalylglycine). An ordinary one-way ANOVA with Dunnett's multiple comparison was used to test for statistical significance of the compound means relative to the 0.4 % DMSO control (multiplicity adjusted p values = ns>0.0332, *0.0332, **0.0021, ***0.0002, ****<0.0001).



F.4.4.v. Phylogenetic tree of human JmjC-KDMs and the OC3, OC6, OC9 selectivity profiles based on MALDI-TOF-MS assays. The cyclic peptide and small molecule (Daminozide, IOX1) selectivity profiles versus the JmjC-KDMs by MALDI-TOF-MS demethylation assays, with compounds at 20 μM, are annotated on a phylogenetic tree derived from JmjC domain alignments, adapted from SGC ChromoHub [247].

All three cyclic peptides were additionally profiled by AlphaScreen activity assay (as described in Chapter 3) against KDM4A and KDM5A demethylase activity (**F.4.4.vi.**). A known small molecule inhibitor (methylstat [243], Chapter 3) was included as a positive control compound. OC9 had little to no inhibitory activity against either enzyme up to the tested 10 μM concentration, corroborating the MALDI-TOF-MS cross-screen assay result. OC3 appeared to effect ~50 % inhibition of KDM4A and KDM5A at its highest concentration (10 μM), but little to no inhibition below this concentration. This result was in conflict with the previous MALDI-TOF-MS assay, where OC3 at 20 μM did not give any significant inhibition of KDM4A or KDM5C. Therefore, the dose-dependent result for OC3 requires further validation. On the other hand, OC6 had inhibited KDM4A and KDM5C by MALDI-TOF-MS assay, and the AlphaScreen activity assay suggested an IC_{50} for OC6 inhibition in the low micromolar regime against both sub-families. However, given the previous observation of OC6 assay interference in AlphaScreen at higher concentrations (Chapter 3), this should also be confirmed with an alternative approach.



F.4.4.vi. Cyclic peptide selectivity screening against JmjC-KDMs by AlphaScreen. An AlphaScreen activity assay was used to assess compound inhibition of KDM4A (30 nM) and KDM5A (100 nM) demethylation of [H3-GGK-biotin] substrate by antibody detection of generated product, in the presence of ascorbate, 2OG, Fe(II). For full conditions see Materials and Methods. IC_{50} values and 95 % confidence intervals are summarised. Analysis of inhibition curves with Prism 7.0 fitted by: log(inhibitor) vs. normalized response -- Variable slope.

4.5. Conclusions and Future Work

Selectivity profiles were established for the three KDM7-targeting cyclic peptides. OC9 did not exhibit any significant off-target binding or displacement of any non-KDM7 PHD finger or TTD reader domains so far tested. The panel included human reader domains that bind the same histone mark as KDM7$_{[PHD]}$ (H3K4me3), and importantly the PHD finger most similar to the KDM7s (DIDO1$_{[PHD]}$). OC9 also did not inhibit the demethylase activity of any of the JmjC-KDM sub-families, thus helping to overcome the historical challenges associated with KDM2/7 selectivity in chemical probe development. Reasons for the exceptional selectivity of OC9 were explored in further studies (Chapter 5, 6). OC3 did not significantly inhibit the activity of the majority of the JmjC-KDM panel, with only minor off-target inhibition of KDM2B and KDM6B, and did not bind to or displace several other H3K4me3-reader domains. OC6 significantly inhibited a member of every JmjC-KDM sub-family in the cross-screening panel, though whether this was due to non-specific protein disruption was not yet entirely clear, but OC6 did not inhibit a non-JmjC 2OG oxygenase (EGLN1) and did not appear to displace H3K4me3 peptide from other non-KDM7 reader domains.

These selectivity profiles have implications for the future use of each cyclic peptide. OC9 is well suited for use as a KDM7 selective probe, now meeting two important criteria, as it is capable of potently disrupting KDM7B$_{[PHD]}$-H3K4me3 interaction whilst leaving other reader domains and KDMs unchallenged, thus permitting study of the role of KDM7$_{[PHD]}$. OC3 appears potentially useful as a KDM7 selective probe, capable of potent disruption of the KDM7B$_{[PHD+JmjC]}$ histone H3K4me3 interaction without appearing likely to disrupt other reader domains. OC3 may also be useful as an inhibitor of KDM7B demethylase activity; however, caution is advised against use at higher micromolar concentrations until the observed weaker off-target inhibition of several KDMs have been further explored. Whilst OC6 did not impact

the binding of several other reader domains, it is arguably not suited to use as a KDM7 selective probe due to its high inhibition efficacy against most JmjC-KDMs and concern surrounding a potential non-specific mechanism of protein disruption.

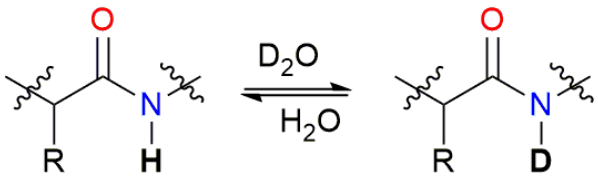
The cyclic peptides could also serve as starting points for future small molecule development. Elucidation of OC3 binding may offer a path towards small molecules that are specific for the KDM7 sub-family and, perhaps more importantly, study of the OC9 interaction may help to understand how it is possible to selectively target PHD fingers. Therefore, investigation of the mechanism of action of OC9 and its unique ability to selectively disrupt $KDM7_{[PHD]}$ formed the focus for the next phase of research (Chapter 5, 6).

Chapter 5. Peptide Interactions with KDM7B by HDxMS

Experimental procedures, raw-data analysis and data plots for HDxMS in this Chapter were performed by Dr Christopher J. Stubbs at AstraZeneca as part of the SBM CDT collaboration.

5.1. Introduction

The amide backbones in proteins often readily exchange at their N-H in protic solution (**F.5.1.i.**). Depending on the level of solvent exposure and hydrogen bonding for a given residue, the rate of this exchange is altered [249]. The difference in exchange rates is advantageous in the context of a deuterated solution (e.g. D_2O based) and can be exploited to monitor binding, as the D nucleus is a full 1 Da heavier than the H nucleus, which allows for different levels of deuteration uptake across the protein to be monitored by mass spectrometry, following proteolytic digestion under acidic aqueous conditions to minimise the rate of backbone amide proton back-exchange. Greater solvent exposure of the side chain residues (e.g. Asn, Gln) often results in their H-D exchange upon quenching in H_2O-based solution being too rapid for reliable observation by hydrogen-deuterium exchange mass spectrometry (HDxMS) [250].



F.5.1.i. The amide nitrogen undergoes proton exchange in equilibrium with the surrounding protic solvent, often rapidly. On addition of excess deuterated solvent, this equilibrium results in deuteration of the backbone.

Protein (HDxMS) [251–254] can be applied in the context of a ligand-protein interaction to find protein residues which are affected, directly or indirectly, by addition of a ligand [255–257]. The protein is incubated with a ligand in un-deuterated buffer prior to dilution into a D_2O based solution, proteolytic digestion and analysis. Peptide residues are deemed 'protected' from

the solvent when they exhibit a level of deuteration that is statistically significantly lower relative to exchange with no ligand present. Occlusion of a given amino acid residue, either directly by the ligand itself or indirectly through induced conformational change, can reduce its backbone amide N-H/D rate of exchange. Deuterium uptake could also feasibly be increased in places where a previously hidden surface became exposed. It is important to make the distinction that HDxMS cannot reveal the definitive protein-ligand interaction surface, but instead only reveal those residues whose backbone amides are affected upon ligand binding. This is still a useful result, giving an indication of the overall consequence of ligand binding on a protein. It would be unlikely for an important protein-ligand interaction surface to be missed entirely with the HDxMS technique [258,259].

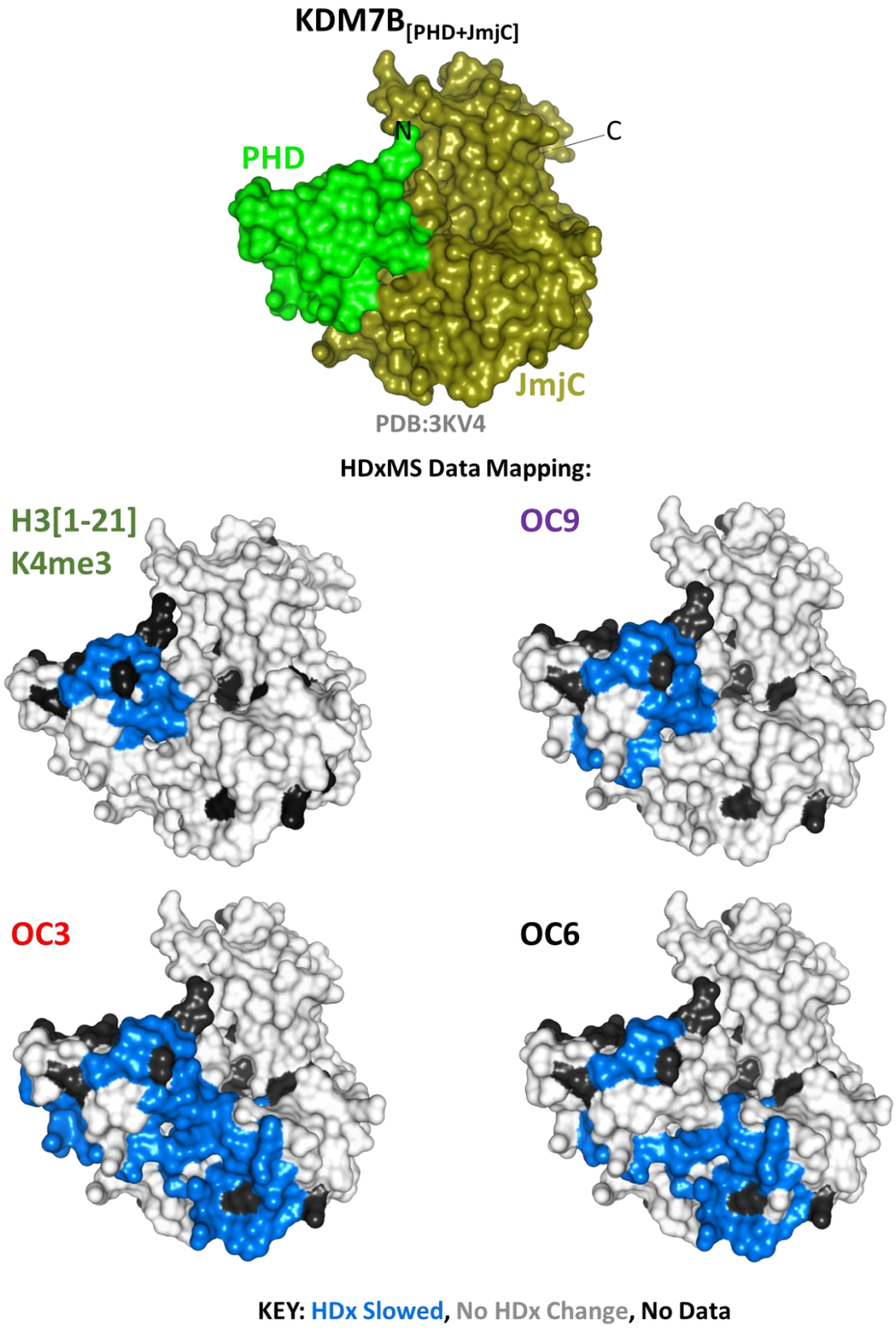
HDxMS possesses several advantages as a technique, requiring low amounts of material (low micromolar concentrations of microlitre quantities for both protein and ligand), and being relatively straightforward to set-up (simple incubation of the protein with the ligand in a D_2O based buffer) and is amenable to automation (in-line proteolytic digest, UPLC peptide separation, ESI-MS fragment detection and library-fingerprint software aided analysis). In this way, a result can often be obtained much faster than would be expected from alternative methods, such as de-novo protein NMR assignment or crystallography trials.

However, it is not yet commonly possible to guarantee analytical coverage of every individual amino acid of the protein [260], as this is dependent on the peptide fragments that are produced upon enzymatic digestion. Nonetheless, the technique enables several samples to be analysed in quick succession, with use of multiple recombinant protein constructs facilitating increased peptide resolution. Rapid insight can be gained into effects on protein surfaces and thus shape a picture of the dynamic interactions for a given ligand.

5.2. HDxMS on KDM7B$_{[PHD+JmjC]}$

Previous studies using BLI, ITC and AlphaScreen displacement indicated the cyclic peptides OC3 and OC6 interacted with KDM7B$_{[PHD+JmjC]}$, whereas OC9 interacted with KDM7A,B,C$_{[PHD]}$. It was important to confirm this PHD vs PHD+JmjC domain-binding selectivity through an orthogonal approach.

For HDxMS, KDM7B$_{[PHD+JmjC]}$ (purification tags cleaved) was incubated separately with one of OC3, OC6, OC9 or H3K4me3 [1-21] initially in un-deuterated, then deuterated buffer, then digested. The digested-protein peptide-fragments were analysed for statistically significant changes in their uptake of deuterium over time, relative to a DMSO-control (see HDxMS Materials and Methods and Appendix for full details and dataset). A visual summary of the results was achieved by mapping the data onto a pre-existing crystal structure (PDB:3KV4, KDM7B$_{[PHD+JmjC]}$ in complex with H3K4me3K9me2), by colour to indicate whether peptide fragments had changed their deuterium uptake behaviour in the presence of the compound (**F.5.2.i.**). It is important to note that this crystal structure does not represent the actual (unknown) structure adopted on binding of one of the cyclic peptides, and that the PHD-to-JmjC linker region (AAs 64-77) was not resolved in the reported structure so data could not be mapped to this region.

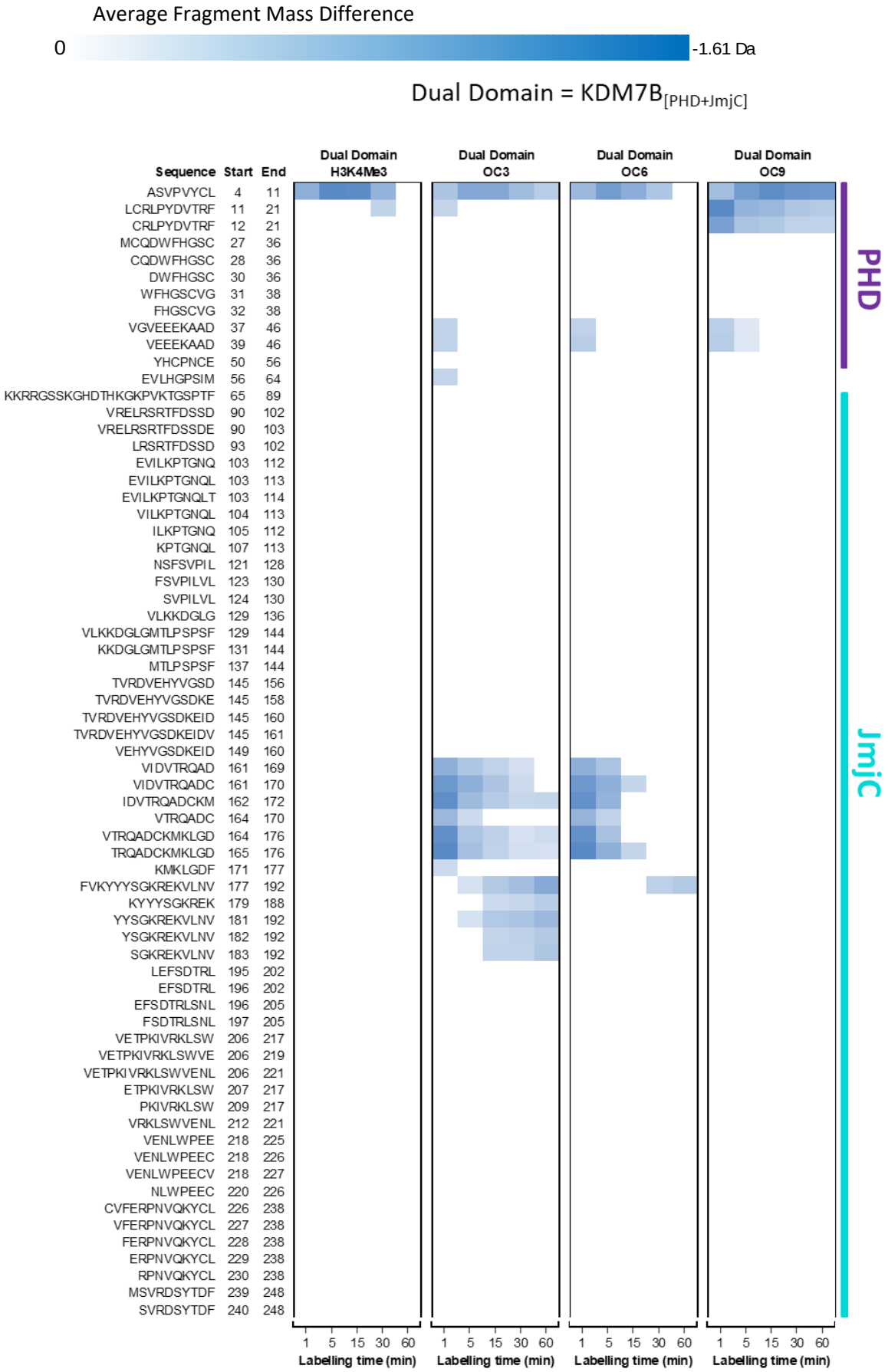


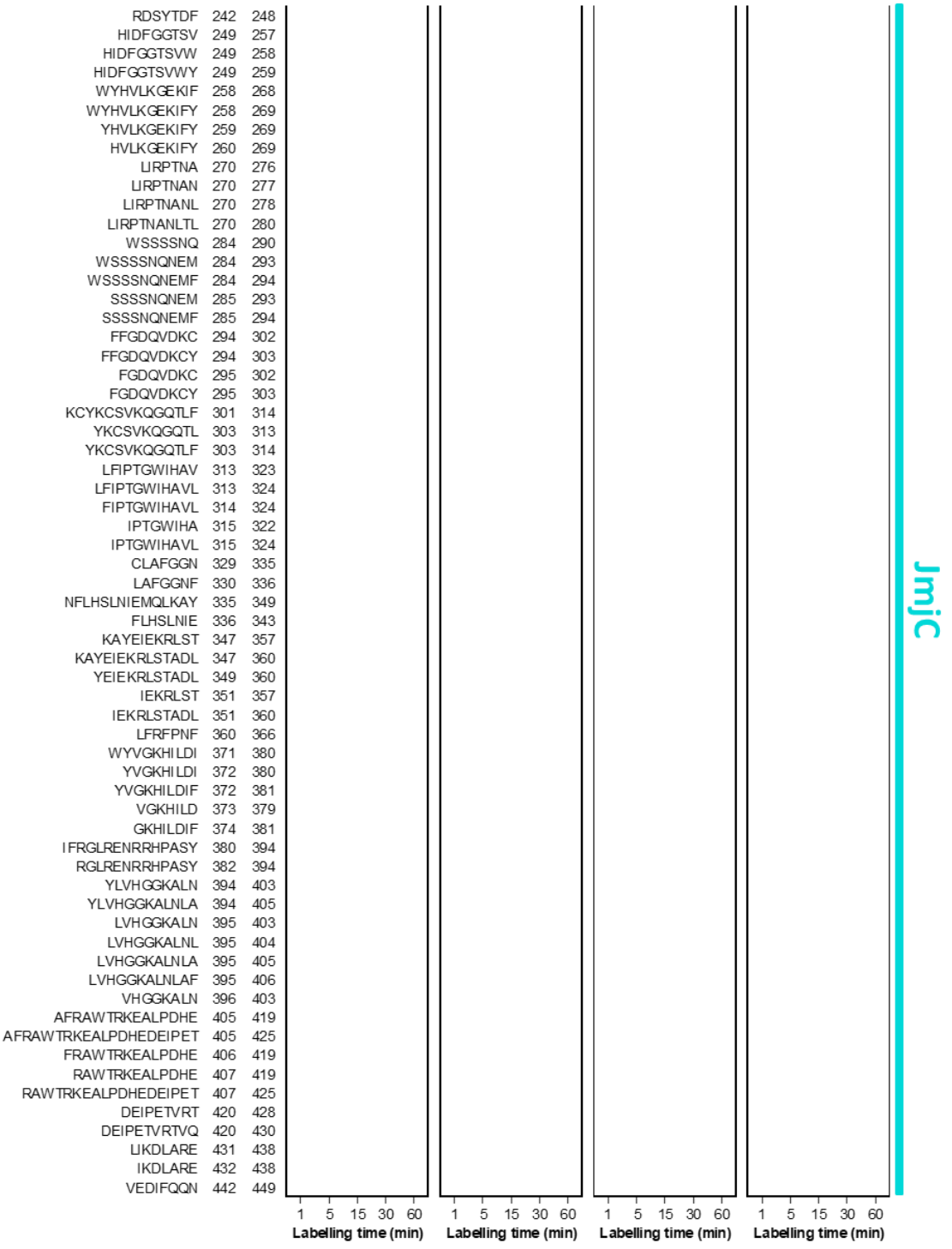
F.5.2.i. HDxMS data for KDM7B$_{[PHD+JmjC]}$ incubation with compound (H3K4me3 [1-21] / OC3 / OC6 / OC9) were mapped to a pre-existing crystal structure (Top, KDM7B$_{[PHD+JmjC]}$ with H3K4me3K9me2, PDB:3KV4). Deuterium exchange was analysed by protein digestion (under acidic conditions) and peptide fragment mass spectrometry. Residues slower to exhibit H-D exchange in the presence of the compound, relative to a DMSO control, are shown in blue. Those with no change are shown in white, and those with insufficient or no data for a decision are black. The PHD finger aromatic cage residue side chains are expanded for the H3K4me3 structure only. **Key:** Blue = slowed HDx, White = no HDx change, Black = No data. Peptides with a p-value <0.01 were considered significant, see Materials and Methods for details of statistical analysis.

The apparent similarity in HDx profiles for OC3 and OC6 was somewhat unexpected given their different amino acid make-ups and differing in vitro selectivity profiles (Chapter 3,4). Inspection of the peptide-fragment deuterium uptake plots over time (**F.5.2.ii.**) shows that OC3 affected a greater number of residues on both PHD finger and JmjC domains compared to OC6, with OC3 also affecting a peptide fragment near the beginning of the KDM7B-unique linker region (EVLHGPSIM). However, the fact that a relatively similar HDx profile was observed for OC6 compared to OC3 supported the prospect of OC6 having a genuine specific interaction with $KDM7B_{[PHD+JmjC]}$.

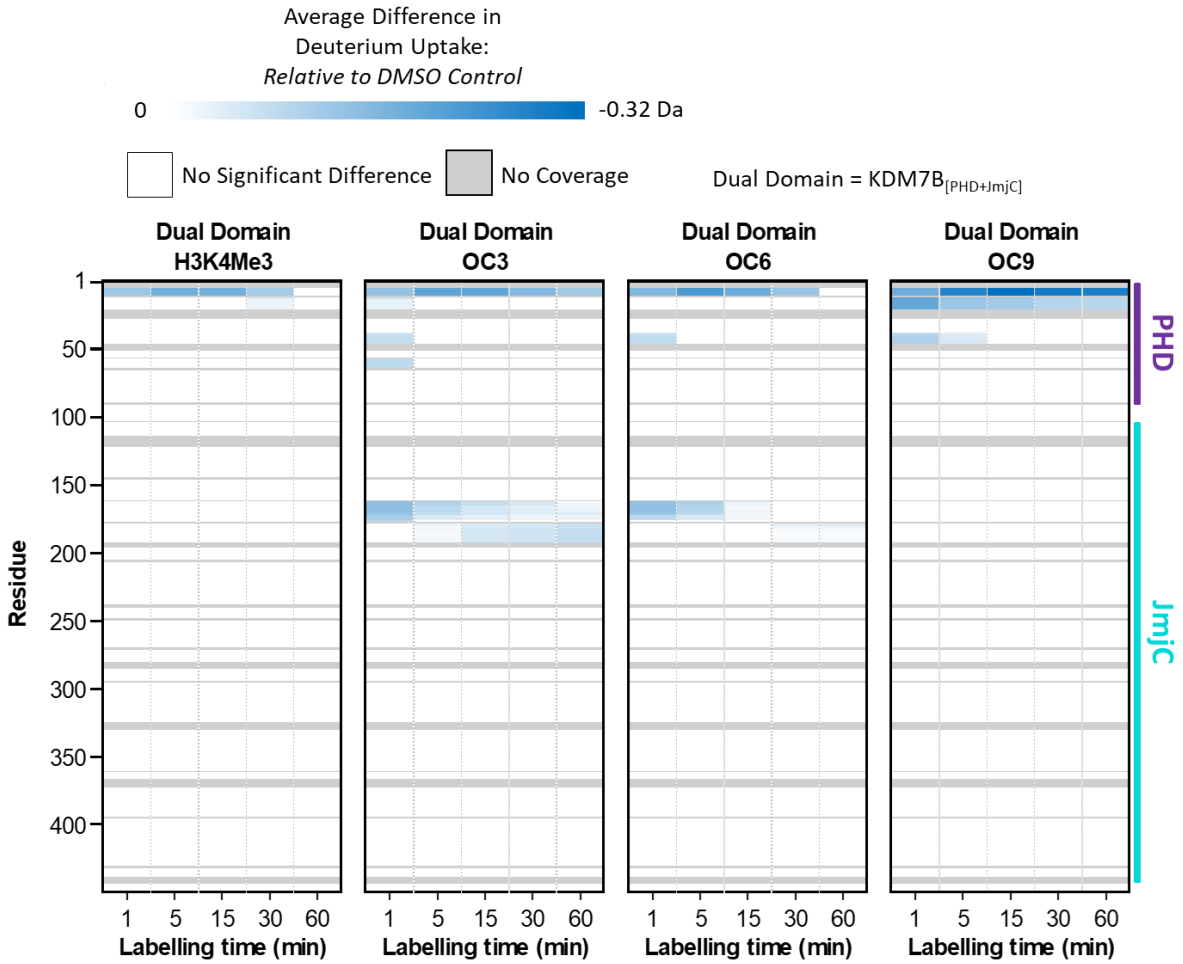
The HDx effects of H3K4me3 [1-21] and OC9 on $KDM7B_{[PHD+JmjC]}$ were confined only to the KDM7B PHD finger (**F.5.2.iii**). It could be argued that relatively few of the PHD finger residues were affected by H3K4me3 [1-21], and no effect was observed on the JmjC domain (the H3K9me2 mark responsible for weaker interaction with the JmjC active site is not present). Further HDxMS experiments were carried out on $KDM7B_{[PHD]}$ with H3K4me3 [1-21] and OC9 to better understand their impacts on the PHD finger relative to one-another.

The relative intensity of the blue shading in these plots (**F.5.2.ii & iii**) does not denote a stronger interaction with a particular set of residues, only the relative mass difference between ligand-sample and no-ligand control – all shaded blue areas can be considered equally significant.


Average Fragment Mass Difference
0
-1.61 Da
Dual Domain = KDM7B[PHD+JmjC]
Sequence Start End
Dual Domain H3K4Me3
Dual Domain OC3
Dual Domain OC6
Dual Domain OC9
ASVPVYCL 4 11
LCRLPYDVTRF 11 21
CRLPYDVTRF 12 21
MCQDWFHGSC 27 36
CQDWFHGSC 28 36
DWFHGSC 30 36
WFHGSCVG 31 38
FHGSCVG 32 38
VGVEEEKAAD 37 46
VEEEKAAD 39 46
YHCPNCE 50 56
EVLHGPSIM 56 64
KKRRGSSKGHDTHKGKPVKTGSPTF 65 89
VRELRSRTFDSSD 90 102
VRELRSRTFDSSDE 90 103
LRSRTFDSSD 93 102
EVILKPTGNQ 103 112
EVILKPTGNQL 103 113
EVILKPTGNQLT 103 114
VILKPTGNQL 104 113
ILKPTGNQ 105 112
KPTGNQL 107 113
NSFSVPIL 121 128
FSVPILVL 123 130
SVPILVL 124 130
VLKKDGLG 129 136
VLKKDGLGMTLPSPSF 129 144
KKDGLGMTLPSPSF 131 144
MTLPSPSF 137 144
TVRDVEHYVGSD 145 156
TVRDVEHYVGSDKE 145 158
TVRDVEHYVGSDKEID 145 160
TVRDVEHYVGSDKEIDV 145 161
VEHYVGSDKEID 149 160
VIDVTRQAD 161 169
VIDVTRQADC 161 170
IDVTRQADCKM 162 172
VTRQADC 164 170
VTRQADCKMKLGD 164 176
TRQADCKMKLGD 165 176
KMKLGDF 171 177
FVKYYYSGKREKVLNV 177 192
KYYYSGKREK 179 188
YYSGKREKVLNV 181 192
YSGKREKVLNV 182 192
SGKREKVLNV 183 192
LEFSDTRL 195 202
EFSDTRL 196 202
EFSDTRLSNL 196 205
FSDTRLSNL 197 205
VETPKIVRKLSW 206 217
VETPKIVRKLSWVE 206 219
VETPKIVRKLSWVENL 206 221
ETPKIVRKLSW 207 217
PKIVRKLSW 209 217
VRKLSWVENL 212 221
VENLWPEE 218 225
VENLWPEEC 218 226
VENLWPEECV 218 227
NLWPEEC 220 226
CVFERPNVQKYCL 226 238
VFERPNVQKYCL 227 238
FERPNVQKYCL 228 238
ERPNVQKYCL 229 238
RPNVQKYCL 230 238
MSVRDSYTDF 239 248
SVRDSYTDF 240 248
PHD
JmjC
1 5 15 30 60
Labelling time (min)



F.5.2.ii. HDxMS data for peptide fragments from KDM7B$_{[PHD+JmjC]}$ incubation with compounds. The sequences of the peptide fragments and their terminal residue amino acid numbers relating to their location within KDM7B$_{[PHD+JmjC]}$ are shown on the left. Blue regions indicate a statistically significant slowing of HDx – the intensity of the blue colour should not be taken to mean a more significant change, as it only relates to the average observed mass change. See Materials and Methods for details of raw data analysis.

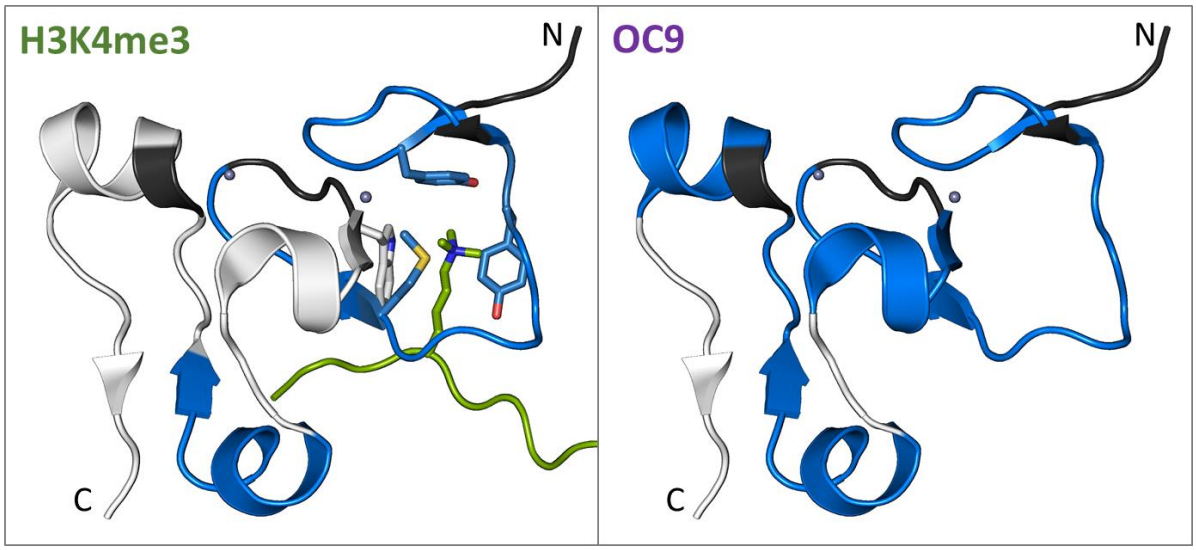


F.5.2.iii. HDxMS data for peptide fragments from KDM7B$_{[PHD+JmjC]}$ incubation with compounds. Data from peptide fragments have been processed to indicate likely changes in individual amino acids (left y axis). OC9 and H3K4me3 [1-21] affected residues only within the PHD finger region, whereas OC3 and OC6 affected residues within the PHD finger and JmjC domains. See Materials and Methods for details of raw data analysis.

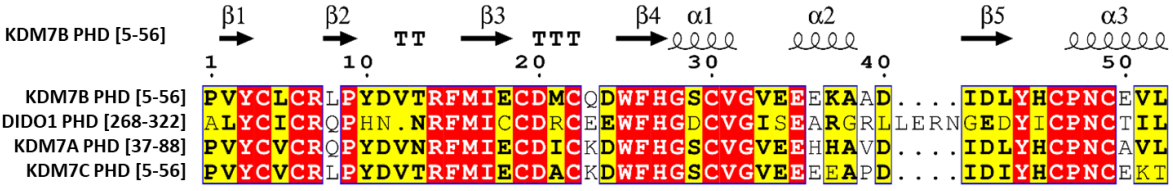
5.3. HDxMS on KDM7B$_{[PHD]}$

Use of KDM7B$_{[PHD]}$ in HDxMS experiments gave better opportunity for improved peptide coverage and resolution of the PHD finger interaction with OC9 and H3K4me3 [1-21]. Experiments were carried out similarly to those for KDM7B$_{[PHD+JmjC]}$.

OC9 was found to slow amide backbone HDx across a large proportion of KDM7B$_{[PHD]}$ – more than was observed for H3K4me3 [1-21] which had its HDx effects confined to primarily to the site of the aromatic cage where it is known to bind (**F.5.3.i.**), suggesting OC9 does not rely on simple mimicry of the cognate interaction. The regions of the PHD finger affected by OC9 include α-helical (α1, AAs 28-32) and loop structures which support the positioning of the aromatic cage and are likely to be important for overall KDM7B$_{[PHD]}$ integrity. Both H3K4me3 [1-21] and OC9 affected a small α-helix (α2, AAs 35-39) consisting of residues where the amino acid sequence is most divergent from other closely related PHD fingers (e.g. DIDO1) (**F.5.3.ii.**). The fact that OC9 affected HDx for a further string of residues (AAs 38-57) in the region where the DIDO1 and KDM7B PHD finger sequences most differ perhaps contributes to explaining to the remarkable binding selectivity demonstrated by OC9 (Chapter 4).



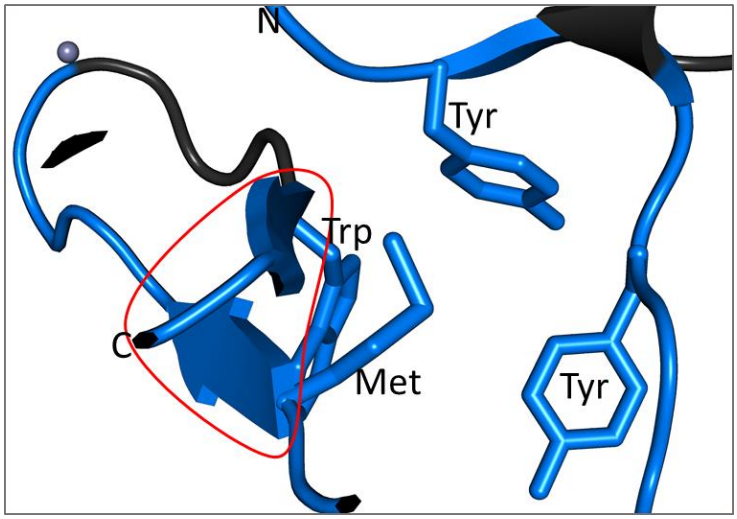
F.5.3.i. HDxMS data for KDM7B$_{[PHD]}$ (residues 3-64 shown) incubation with compound (H3K4me3 [1-21] or OC9) were mapped to a pre-existing crystal structure (KDM7B$_{[PHD+JmjC]}$ with H3K4me3K9me2, PDB:3KV4). Deuterium exchange was analysed by protein digestion (under acidic conditions) and peptide fragment mass spectrometry. Residues slower to exhibit H-D exchange in the presence of the compound, relative to a DMSO control, are shown in blue. Those with no change are shown in white, and those with insufficient or no data for a decision are black. The PHD finger aromatic cage residue side chains are expanded for the H3K4me3 structure only.



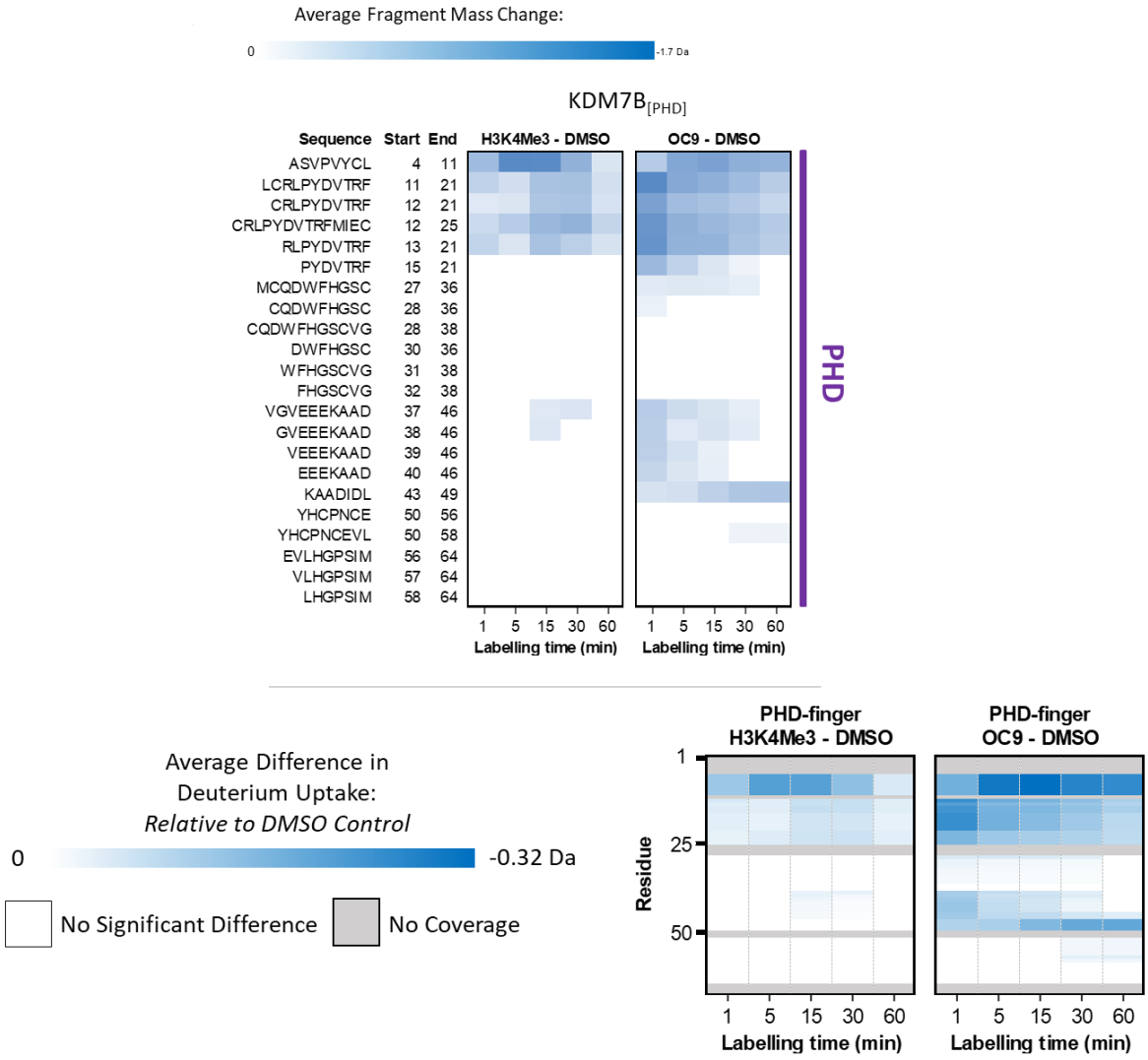
F.5.3.ii. Amino acid sequence alignments for the canonical KDM7$_{[PHD]}$ and DIDO1$_{[PHD]}$ sequences. The region from AAs 34-45 demonstrates the most significant difference between the families of PHD finger. OC9 exhibited effects on KDM7B$_{[PHD]}$ HDx for residues within the AAs 30-50 region. Secondary structure annotations from KDM7B (PDB:3KV4) were applied using ESPript3 [242], with alignments conducted using Clustal Omega [241].

Explanation for the ability of OC9 to inhibit KDM7$_{[PHD]}$ binding to H3K4me3 begins to emerge from observation of slowed H-D exchange on peptide fragments containing the aromatic cage residues. Whilst this KDM7B$_{[PHD]}$ HDxMS result cannot conclusively show that OC9 binds directly to the aromatic cage, it does reveal a change to the cage is effected upon binding of OC9 to make it less solvent accessible. (**F.5.3.iii.**). The integrity of the aromatic cage is vital to proper 'reading' function (Chapter 0, 8), therefore slowing of HDx within this area means access for H3K4me3 is also likely to be restricted. Further to this, analysis of the peptide-level

deuterium uptake plots over time (**F.5.3.iv.**) made it clear that the extent of backbone amide occlusion by OC9 was high compared to H3K4me3 [1-21].



F.5.3.iii. The KDM7B$_{[PHD]}$ aromatic cage residues (Trp, Met, Tyr, Tyr) experience slowed H-D exchange in the presence of OC9, suggesting they are less solvent accessible than in the unbound state. The supporting β-sheet residues (circled in red) and surrounding loop regions are similarly affected, making it more likely that access for H3K4me3 would also be restricted and thus is likely to contribute to OC9 inhibition of KDM7$_{[PHD]}$ function.



F.5.3.iv. (Top): HDxMS data for peptide fragments from KDM7B$_{[PHD]}$ incubation with compounds. The sequences of the peptide fragments and their terminal residue amino acid numbers relating to their location within KDM7B$_{[PHD]}$ are shown on the left. Blue regions indicate a statistically significant slowing of HDx – the intensity of the blue colour should not be taken to mean a more significant change, as it only relates to the average observed mass change. **(Bottom):** Data from peptide fragments have been processed to indicate likely changes in individual amino acids (left y axis). OC9 binding serves to slow HDx across a greater proportion of the PHD finger than H3K4me3 [1-21]. See Materials and Methods for details of raw data analysis.

5.4. Conclusions and Future Work

HDxMS experiments on $KDM7B_{[PHD+JmjC]}$ and $KDM7B_{[PHD]}$ are consistent with binding of OC3 and OC6 affecting both the PHD and JmjC domains, and that binding of OC9 affects only the PHD finger. These data are supported by previous binding analyses (Chapter 2, 3). H3K4me3 [1-21] appears to slow HDx for residues only within the PHD finger and in the immediate vicinity of its known binding location, whereas OC9 elicited widespread HDx effects in regions containing 52 of the 64 amino acid residues within $KDM7B_{[PHD]}$.

Strong clues towards the mechanism of action of OC9 have emerged, with evidence of OC9 slowing deuterium uptake around the crucial aromatic cage of $KDM7B_{[PHD]}$, meaning OC9 may restrict H3K4me3 access to the aromatic cage. In contrast, the exact mechanism through which OC3 and OC6 compete with H3K4me3 binding to the $KDM7_{[PHD+JmjC]}$ remains unknown, with neither significantly affecting deuterium exchange within the aromatic cage. It may be that their shared disruption of the face of the PHD finger adjoining the JmjC domain is sufficient to disengage PHD finger binding, but more work is needed to elucidate this.

Future HDxMS work could include extension of the experiments to complete cyclic peptide profiling against KDM7A and KDM7C. Their high sequence homology with KDM7B increases the probability of gaining similar quality data, which would enable identification of common trends to help narrow down more precise binding interfaces. HDxMS could also be applied to aid cyclic peptide selectivity profiling, for example with OC9 and non-$KDM7_{[PHD]}$ reader domains to further investigate its selectivity profile. The robust evidence-base for OC9 interaction with the $KDM7_{[PHD]}$ represented an important proof-of-principle that PHD fingers are tractable targets for probe development in epigenetic study. The next phase of research (Chapter 6) sought to progress understanding of how OC9 interacts with $KDM7_{[PHD]}$.

Chapter 6. Studies on the Interaction of OC9 with KDM7$_{[PHD]}$

6.1. Introduction

OC9 represents the first reported potent and selective PHD finger inhibitor, making it important to understand the mechanism through which it binds. HDxMS studies had provided some insight into putative OC9-KDM7B$_{[PHD]}$ interaction surfaces, but did not provide sufficient detail alone to understand the specific amino acid interactions. In this Chapter, the work described was trialled to attempt to gain more information. Some basic crystallography trials were performed with KDM7B$_{[PHD]}$ and KDM7B$_{[PHD+JmjC]}$ but proved unsuccessful, and so solution NMR was explored instead.

6.2. KDM7B$_{[PHD]}$ Solution NMR

Protein NMR experiments and data analysis were assisted Dr Martine Abboud (Schofield Lab, University of Oxford)

Protein NMR is a valuable tool for studying protein-ligand interactions in solution and guiding drug discovery [261–265]. It can reveal dynamic and time-dependent interactions, including conformational changes and protein-complex stability. The KDM7B/C$_{[PHD]}$ have previously been proven to be amenable to solution NMR studies [68] and therefore this approach was employed to help to investigate the interaction of OC9 with KDM7B$_{[PHD]}$.

Previous work within the Brennan/Schofield groups by Finn Wolfreys (DPhil Thesis, unpublished) had assigned solution NMR spectra for the uniformly labelled ^{15}N-KDM7B$_{[PHD]}$, a sample of which was donated for use in this study. Titration of sub-stoichiometric to eventual excess of OC9 into a solution of ^{15}N-KDM7B$_{[PHD]}$ monitored by ^{1}H-^{15}N HSQC NMR revealed a substantial number of chemical shift changes for the observable backbone amide and side chain N-H correlations of the KDM7B$_{[PHD]}$ residues in the overlaid spectra (**F.6.2.i.**) For example, the resonances of an unassigned but distinct isolated residue moved from 6.3939 (^{1}H), 112.694 (^{15}N) in the free protein to 6.3316 [δ+0.0623] (^{1}H), 112.279 [δ+0.415] (^{15}N) in the saturated complex. This suggested binding of OC9 induced widespread conformational and/or structural changes in KDM7B$_{[PHD]}$ and supported results from earlier HDxMS studies, which together contend for OC9 affecting a large proportion of the KDM7$_{[PHD]}$ structure. This can be reasonably explained given their similar magnitude of size (OC9 at 14 residues and 1764 Da, and the ^{15}N-KDM7B$_{[PHD]}$ construct at 67 residues and ~8 kDa) likely enabling interaction over a large binding surface relative to the size of the PHD finger.

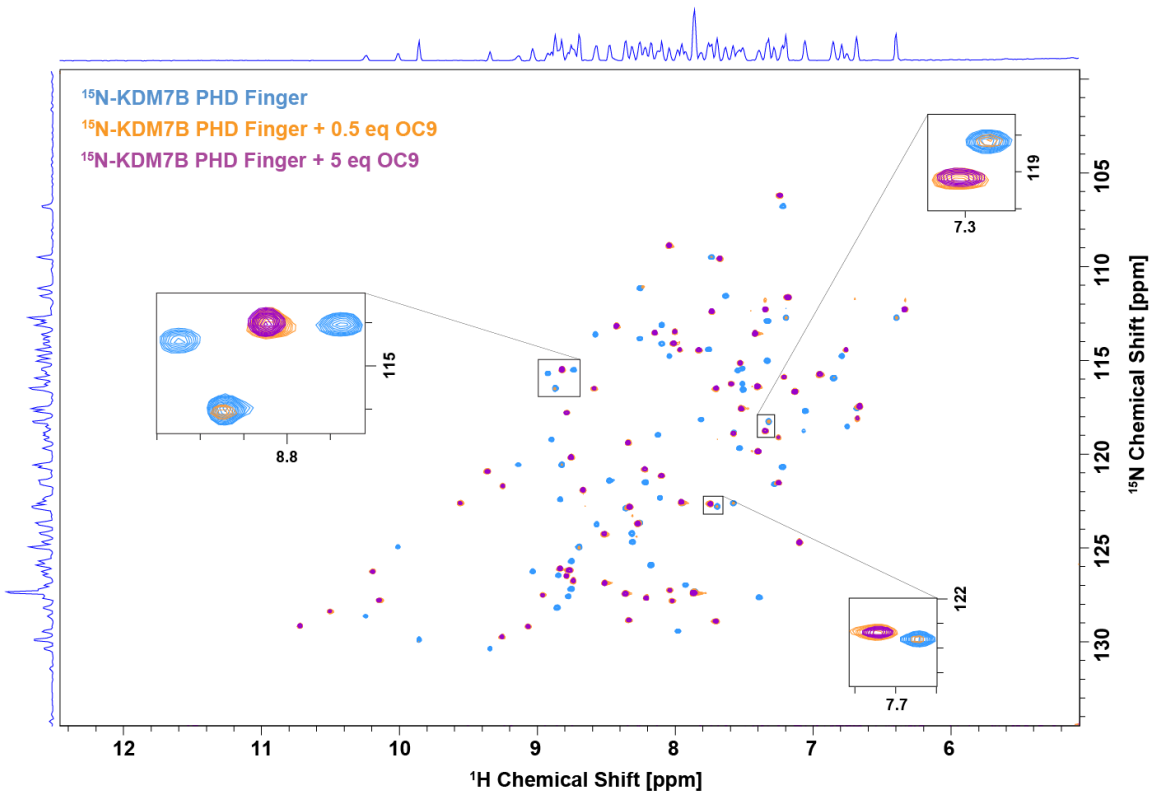
The affinity of the OC9 interaction with $KDM7B_{[PHD]}$ (K_D ~10 nM by ITC/BLI) is characterised by a slow off rate (K_{off} = 1.43 x10^{-3} s^{-1} by BLI, Chapter 2) which gives a longer protein-ligand complex lifetime ($1/K_{off}$) and made a slow-exchange regime on the NMR-timescale more likely [266]. An NMR slow exchange regime is often expected when the off rate for a protein-ligand interaction is much slower than the typical difference in a chemical shift change (Δω (Hz = s^{-1})) for a given resonance between the free (ω_P) and bound (ω_{P+L}) protein states [267]. This can be summarised as estimation of a slow exchange regime when: K_{off} (s^{-1}) << Δω (s^{-1}) [268]. For example, with the observed ^{1}H Δppm = 0.1, on a 700 MHz instrument, ΔHz = 70 s^{-1}. The K_{off} for OC9 with $KDM7B_{[PHD]}$ was previously measured to be on the order of 10^{-3} s^{-1}, which is much slower.

A slow exchange regime was indeed observed to be the case; residues for the OC9-PHD complex appeared at their final chemical shift synchronous with disappearance from the free PHD finger (**F.6.2.i.**). This behaviour negated the prospect of using the previously assigned data for the free PHD finger to track a gradual chemical shift of selected residues with titration of ligand for quicker assignment of the new complex – something that can be performed (aka: the nearest neighbour approach) for ligand complexes in fast-to-intermediate NMR exchange [267].

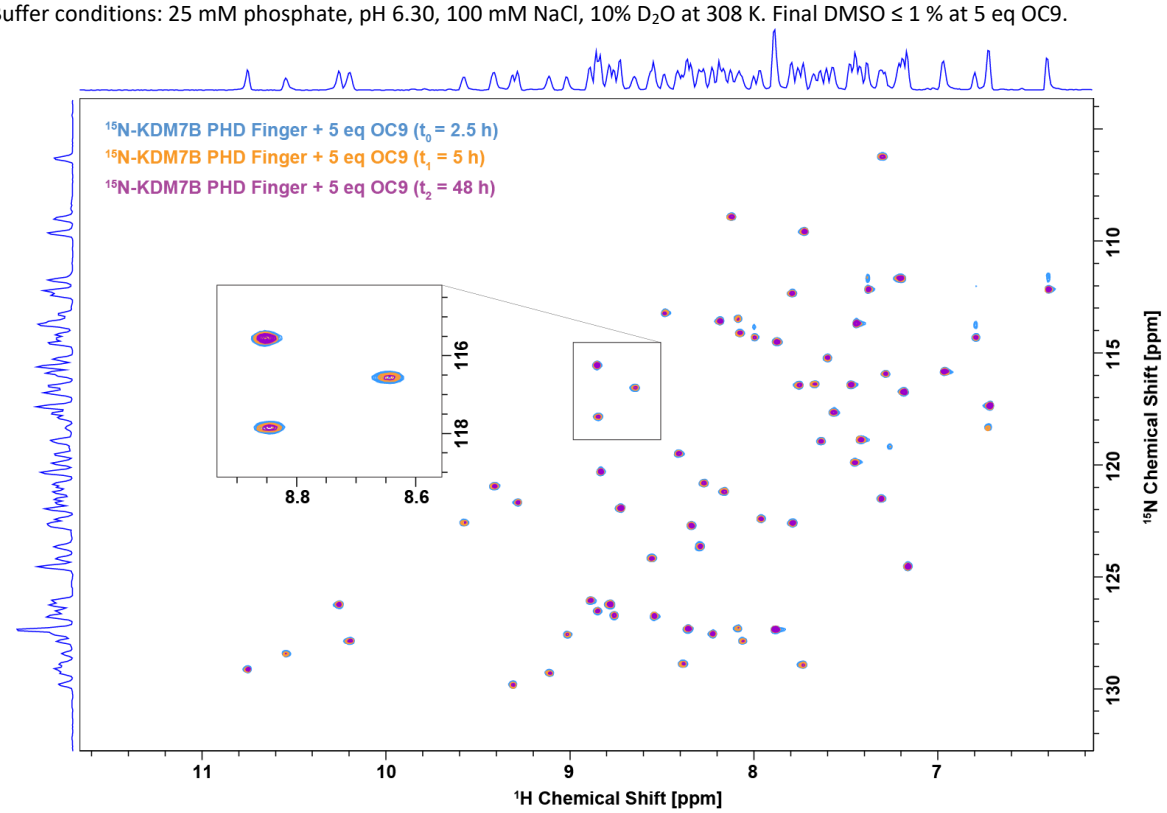
Nevertheless, these spectra provided valuable data. Firstly, the PHD finger remained well dispersed and folded upon binding of OC9, with no collapsing of the residues' resonances into a narrow range which would indicate unfolding or denaturation [269]. Additional experiments showed excellent stability of the OC9-$KDM7B_{[PHD]}$ complex, with little change in residue chemical shifts in the overlaid ^{1}H-^{15}N HSQC spectra at 2.5, 5 and 48 hours at 308 K (35 °C) (**F.6.2.ii.**). This further strengthens the future use-case for OC9 in a more protein-rich

biological context as there appears to be a low risk of causing protein aggregation which could lead to domino effects giving unpredictable results.

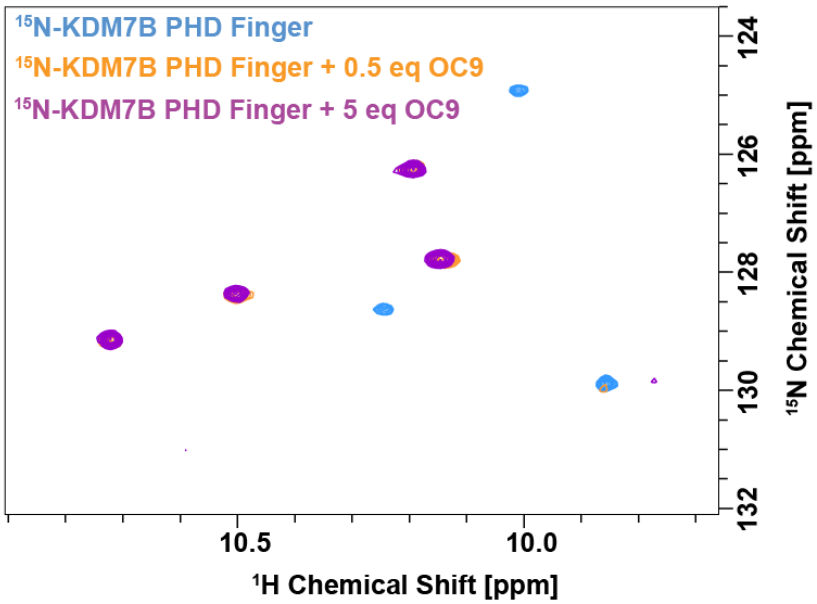
The NMR spectra are unassigned; however, even without full assignment some potential insight could still be gained. For example, the tryptophan indole N^{ϵ}-H^{ϵ} is normally characterised by typical shifts of ^{1}H ~10 ppm and ^{15}N ~129 ppm [270] which often stand distinct from the N-H shifts of other amino acids. The region of the ^{1}H-^{15}N spectrum containing these resonance ppm values is sparsely populated, as expected, and experiences some significant chemical shift changes upon binding of OC9 (**F.6.2.iii.**). Only one tryptophan residue is present in KDM7B$_{[PHD]}$ and it is part of the aromatic cage that is essential for H3K4me3 binding (**F.6.2.iv.**). The NMR data therefore suggest that, if the tryptophan is indeed located within this chemical shift window, the tryptophan could find itself in a changed environment in the presence of OC9, which may help account for the ability of OC9 to displace binding of H3K4me3.



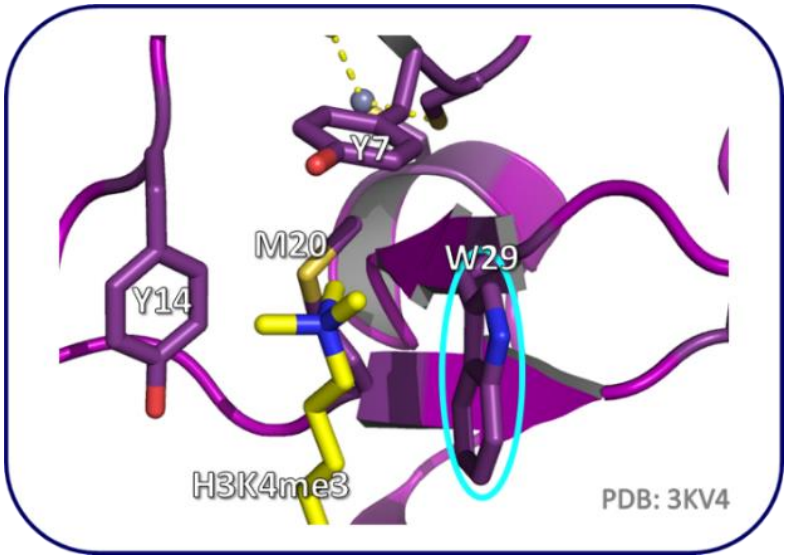
F.6.2.i. ^{1}H-^{15}N HSQC NMR spectra (AV700) are overlaid for titration of OC9 (from a concentrated DMSO stock) into a buffered solution of 50 µM ^{15}N-labelled KDM7B$_{[PHD]}$ (AAs 1-66, no tags). Highlighted close-ups show residues presenting evidence of the interactions of OC9 with the PHD finger inducing chemical shift changes in a slow exchange system on the NMR timescale. Buffer conditions: 25 mM phosphate, pH 6.30, 100 mM NaCl, 10% D_2O at 308 K. Final DMSO ≤ 1 % at 5 eq OC9.



F.6.2.ii. Stability of the OC9-KDM7B$_{[PHD]}$ complex was assessed by NMR (AV700). Overlaid ^{1}H-^{15}N HSQC spectra were recorded over the first 2.5 hours, at 5 hours and 48 hours. Close-ups indicate the stability of the complex over time – peaks are stable and did not revert to an unbound or denatured chemical shift. OC9 (from a concentrated DMSO stock) was added to a final concentration of 1 mM with 200 µM ^{15}N-labelled KDM7B PHD finger (AAs 1-66, no tags). Buffer conditions: 25 mM phosphate, pH 6.30, 100 mM NaCl, 10% D_2O at 308 K. Final DMSO ≤ 2 % at 5 eq OC9.

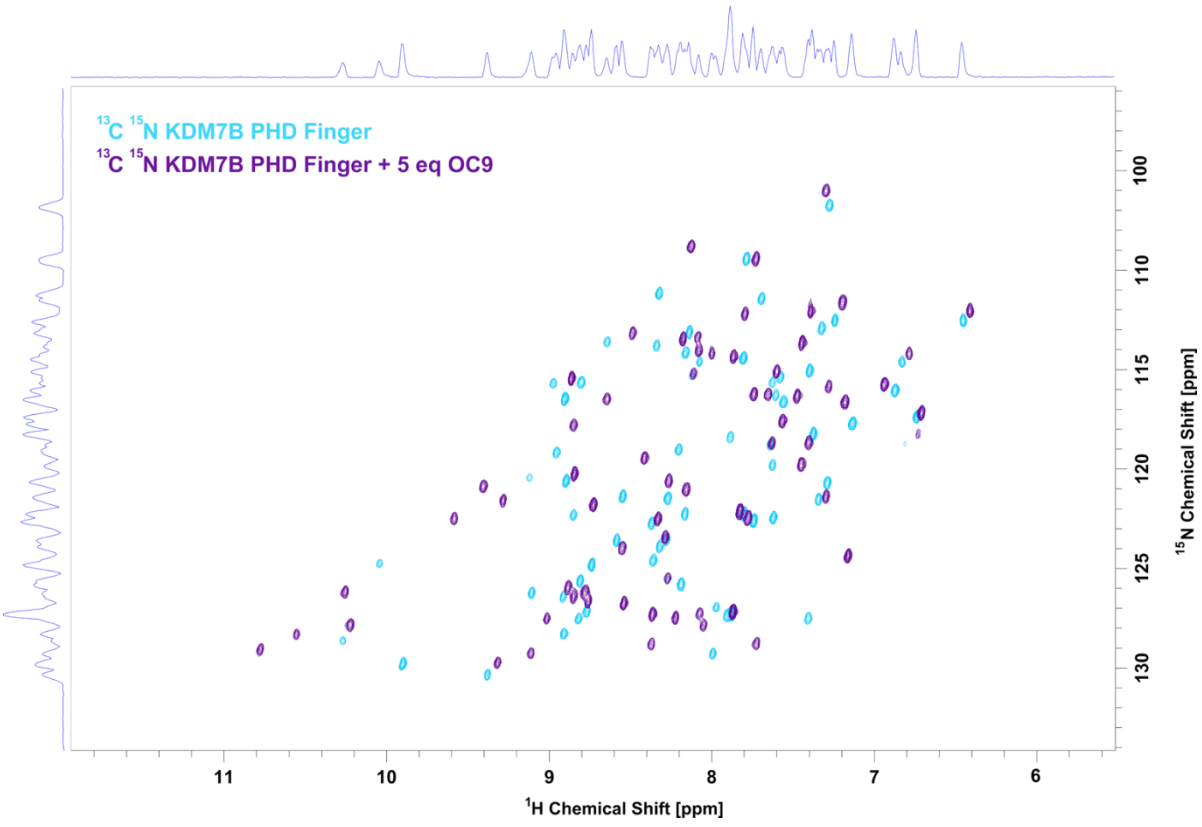


F.6.2.iii. The ^{1}H-^{15}N HSQC spectra (AV700) chemical shift region likely to contain the ^{15}N-labelled KDM7B$_{[PHD]}$ tryptophan indole N$^{\epsilon}$-H$^{\epsilon}$ is shown. There is a substantial chemical shift of all residues within this region on addition of OC9, indicating the tryptophan residue could be significantly impacted by binding of OC9. 50 µM ^{15}N-labelled KDM7B$_{[PHD]}$ (AAs 1-66, no tags). Buffer conditions: 25 mM phosphate, pH 6.30, 100 mM NaCl, 10% D_2O at 308 K. Final DMSO ≤ 1 % at 5 eq OC9.

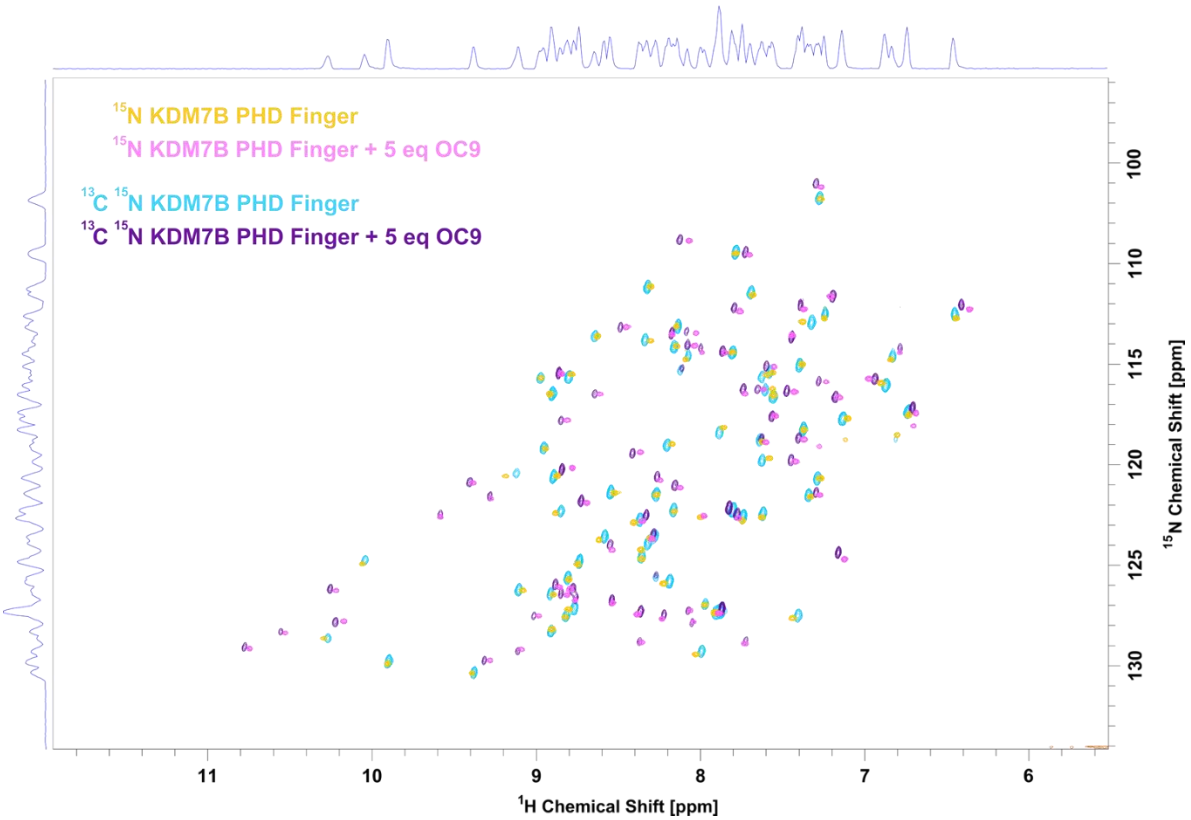


F.6.2.iv. The four labelled residues (Y7, Y14, M20, W29) constitute the aromatic cage in KDM7B$_{[PHD]}$ and are responsible for the major binding interaction with H3K4me3. The tryptophan residue, W29, is highlighted in blue. Structure from PDB:3KV4.

Further NMR experiments are needed to solve a 3D-NMR solution structure of the OC9-PHD complex, and so production of recombinant KDM7B$_{[PHD]}$ was subsequently optimised in minimal media to successfully perform uniform $^{13}C,^{15}N$ labelling (see Materials and Methods). This doubly isotopically labelled protein was used to conduct further 2D and 3D-NMR experiments with OC9 that are necessary for assignment of a ligand-complex solution structure. The $^{13}C,^{15}N$ labelled KDM7$_{[PHD]}$ also evidenced a substantial number of chemical shift changes by ^{1}H-^{15}N-HSQC NMR upon binding of OC9 (**F.6.2.v.**). Both the free protein and the OC9-complex spectra were found to overlay well with the previously acquired data for the donated ^{15}N labelled KDM7B$_{[PHD]}$ (**F.6.2.vi.**). Work to complete the assignments will be part of future work.



F.6.2.v. $^{13}C,^{15}N$ labelled KDM7$_{[PHD]}$ was produced to enable further 2D and 3D NMR experiments for investigation and future assignment of the structure of the complex with OC9. ^{1}H-^{15}N HSQC NMR (AV700) spectra are overlaid for the free protein (200 µM $^{13}C,^{15}N$-labelled KDM7B$_{[PHD]}$ [AAs 1-66, no tags]) and after addition of OC9 (from a concentrated DMSO stock) to a final concentration of 1 mM. Buffer conditions: 25 mM phosphate, pH 6.30, 100 mM NaCl, 10% D_2O at 308 K. Final DMSO = 5 % at 5 eq OC9.



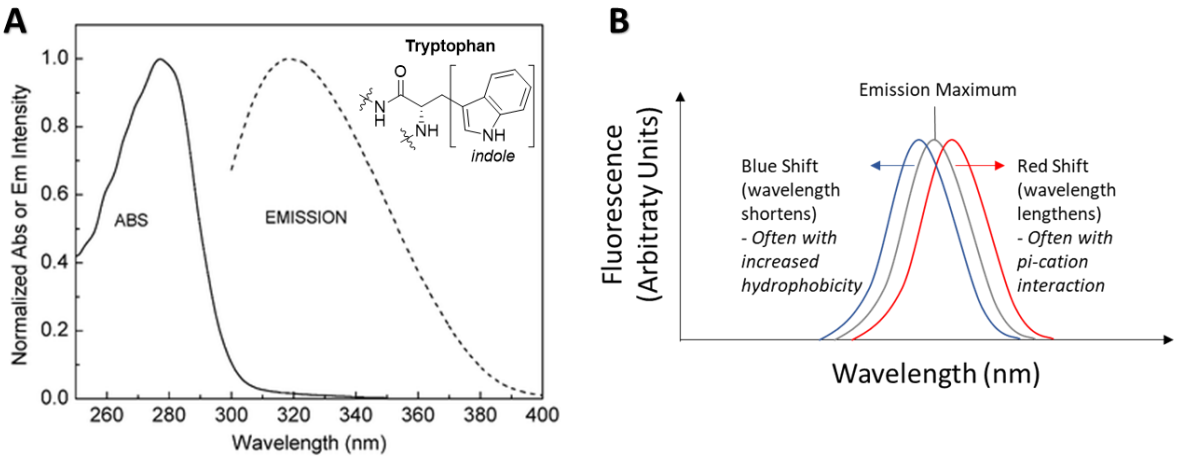
F.6.2.vi. The ^{1}H-^{15}N HSQC NMR (AV700) spectra for the ^{15}N-only and ^{13}C,^{15}N labelled KDM7B$_{[PHD]}$ overlay well both for the free protein (gold and aquamarine) and in the presence of excess OC9 (purple and pink), evidencing the reproducibility of the impact of OC9 binding to the PHD finger and the suitability of the protein for use in further NMR characterisation work. For KDM7B$_{[PHD]}$ [AAs 1-66, no tags], the ^{13}C,^{15}N-labelled version = 200 µM and ^{15}N-labelled version = 50 µM. OC9 was added from a concentrated DMSO stock to a final concentration equivalent to a 5-fold excess over protein. Buffer conditions: 25 mM phosphate, pH 6.30, 100 mM NaCl, 10% D_2O at 308 K. Final DMSO ≤ 5 % at 5 eq OC9.

6.3. KDM7$_{[PHD]}$ Tryptophan Intrinsic Fluorescence Assay

Given the potential NMR evidence for direct disruption of the essential tryptophan residue of the KDM7B$_{[PHD]}$ aromatic cage, further support for this was sought with the use of a tryptophan fluorescence assay. The indole moiety possessed by tryptophan gives it an intrinsic fluorescence, with a typical light absorbance maximum at ~280 nm and emission maximum at ~350 nm in aqueous environment [271] (**F.6.3.i.A**). Other aromatic residues, tyrosine and phenylalanine, exhibit this property much more weakly making the measurement of such spectra typically sensitive to tryptophan [272]. The emission intensity and wavelength are dependent on its microenvironment [273], affected by a multitude of factors, among them: local hydrophobicity, solvent exposure and adjacent residues.

The protein folding state influences the local hydrophobicity and degree to which the tryptophan is 'hidden' from the solvent, where tryptophan exposure to polar solvent can induce a *red shift* – and therefore reduced exposure a *blue shift* – of the fluorescence wavelength maximum [273] (**F.6.3.i.B**). A complicating factor can be fluorescence quenching by adjacent amino acids through excited-state proton transfer (Lys or Tyr), or by excited-state electron transfer (Gln, Asn, Glu, Asp, Cys, His) [272]. Tryptophan is globally rare across most proteins, with PHD fingers typically possessing only one or two tryptophan residues. Its singular location within the aromatic cage of KDM7B$_{[PHD]}$ (AAs 1-66) gives additional insight into the state of the aromatic cage upon binding of H3K4me3 and OC9. Pi-cation interactions have also been shown to *red shift* and increase tryptophan fluorescence intensity for amino-cation to tryptophan pairings [274,275], which is a key interaction responsible for KDM7$_{[PHD]}$ binding to H3K4me3. The pi-cation effect on the emission spectra can be measured by scanning multiple wavelengths, or picking a fixed wavelength and monitoring intensity.

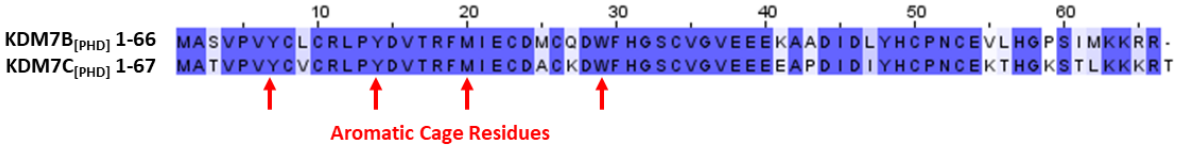
Variants of the assay exploiting this fact have been used to study histone peptide binding affinities to PHD fingers [116].



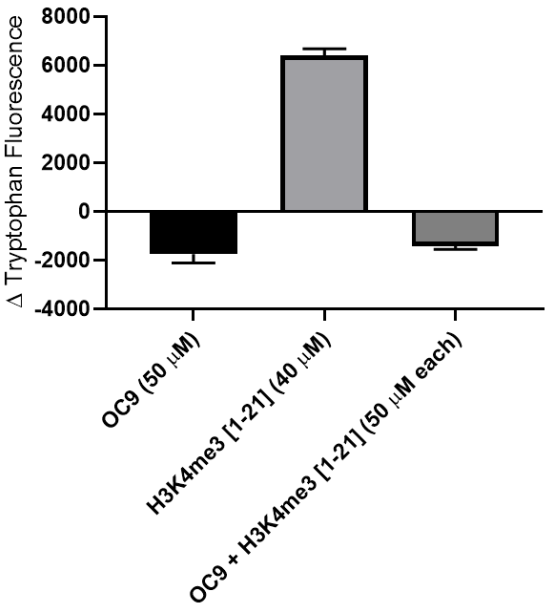
F.6.3.i. Principle of intrinsic tryptophan fluorescence experiments. **A)** An example tryptophan absorbance / emission spectrum of thermophile protein TH0593. Adapted from Luchansky and co-workers [276] with permission from The Royal Society of Chemistry. **B)** Diagrammatic representation of emission spectrum shifts that can occur dependent on changes on the tryptophan microenvironment.

The KDM7C[PHD] construct was similar to that of KDM7B[PHD] (80% identity, 90% similarity, **F.6.3.ii.**), with an identical aromatic cage, and was used to conduct tryptophan fluorescence studies due to it possessing the highest affinity of the three KDM7[PHD] for both H3K4me3 peptide and OC9 by ITC (Chapter 2). Addition of the H3K4me3 [1-21] peptide substantially increased the measured tryptophan fluorescence intensity at 350 nm (**F.6.3.iii.**), indicating the expected tryptophan pi-cation lysine trimethyl interaction was influencing the emission spectrum. Addition of OC9 instead induced a lower magnitude decrease in tryptophan fluorescence intensity, arguing for the exclusion of pi-cation interaction mimicry by OC9 as a mode of binding. The intensity decrease with OC9 could be explained by the action of its binding creating, directly or indirectly, a more hydrophobic environment for tryptophan serving to suppress total emission intensity or shorten the wavelength of the emission maximum. Alternatively, insertion of one of the OC9 residues into the aromatic cage may be quenching tryptophan fluorescence. Co-addition of both OC9 and H3K4me3 gave a

tryptophan fluorescence intensity very close to that observed for OC9 alone, consistent with OC9 outcompeting H3K4me3 for binding to KDM7[PHD].



F.6.3.ii. Alignment of KDM7B[PHD] and KDM7C[PHD] constructs with Clustal Omega [241], processed by JalView to colour identical residues blue and residues with similar properties in grey. The aromatic cage residues are indicated with red arrows.



F.6.3.iii. Tryptophan fluorescence data were recorded on a Pherastar (BMG) for addition of OC9 or H3K4me3 (from concentrated DMSO stock) into 25 μM KDM7C[PHD] (AAs 1 to 67, no tags). Buffer conditions: 1xPBS, 1 % DMSO. Δ Fluorescence was measured with respect to a solution of protein and DMSO only. Excitation performed at λ = 280 nm and emission detection recorded at λ = 350 nm. Experiment conducted in technical duplicate.

6.4. Conclusions and Future Work

Protein ^{1}H-^{15}N HSQC NMR studies have revealed that KDM7B$_{[PHD]}$ experiences substantial changes upon binding of OC9, as evidenced by the global changes in chemical shifts. The stability of this protein-ligand complex is maintained over 48 hours with no denaturation observed at 35 °C. The complex is in slow-exchange on the NMR timescale (micro to milliseconds) which supports the low nanomolar affinity of interaction. The NMR and additional tryptophan fluorescence studies have provided some indication of possible disruption of the key KDM7$_{[PHD]}$ aromatic cage tryptophan residue by OC9 – in the presence and absence of H3K4me3 – potentially representing a plausible mechanism of action for disrupting binding of H3K4me3.

Further experiments to investigate the protein secondary structures, such as the α-helices and β-sheets, whilst in the unbound and OC9-bound states could be performed (e.g. by circular dichroism). Future work could also seek to build on data sets acquired in the course of these experiments, primarily undertaking to attempt assignment of the solution NMR data for the OC9-KDM7B$_{[PHD]}$ complex, and conducting further crystallography trials to build a picture of the precise interactions responsible for binding. More insight into how OC9 binds to its target was gained by synthetic studies on the OC9 cyclic peptide sequence itself, the focus of the next phase of research (Chapter 7).

Chapter 7. OC9 Structure-Activity Relationships (SAR)

7.1. Introduction

To make synthetic modifications to the cyclic peptide and tailor its properties towards desired ends, an understanding of the relationship between its chemical structure and activity is useful. The amino acid side chains provide the richest and most variable source of chemical functionality. The classical starting point for structure-activity-relationship (SAR) studies in peptide chemistry is an alanine scan. Alanine is used for its being the simplest chiral proteinogenic amino acid, with the methyl group providing chiral integrity alongside minimal functionality. Every amino acid in the peptide is thus switched individually one-at-a-time to Ala, and subsequent effects on peptide properties determined to identify crucial residues.

As a first approximation, if the activity of the peptide is not adversely affected by an Ala substitution then it is of low functional criticality and identified as a suitable site for more drastic modification. Substitution of some residues can result in near loss of activity which highlights key motifs. The extent to which activity is adversely affected informs on the importance of that particular residue, which may then warrant further investigation, through alternate substitutions, of where the tolerance for alteration lies. The amide backbone can also be targeted for modification, with α-N-methylation a well-established approach [277,278] in addition to use of alternate β-peptide backbones [279], depsipeptide linkages [280] and amide isosteres [281,282] to improve metabolic stability.

The overall aim of the work described in this Chapter was to explore the SAR of OC9. Whilst the unaltered OC9 cyclic peptide may already serve well as a chemical tool for many biochemical studies, the knowledge gained from SAR analysis will allow bestowment of more favourable physicochemical properties for cellular applications [283]. Increasing the metabolic

resistance and membrane permeability, whilst maintaining respectable target potency and selectivity [284], is essential for use in cellular and in vivo contexts – giving strong justification to pursuit of their improvement.

7.2. Alanine Scanning: Displacement and Functional Activity

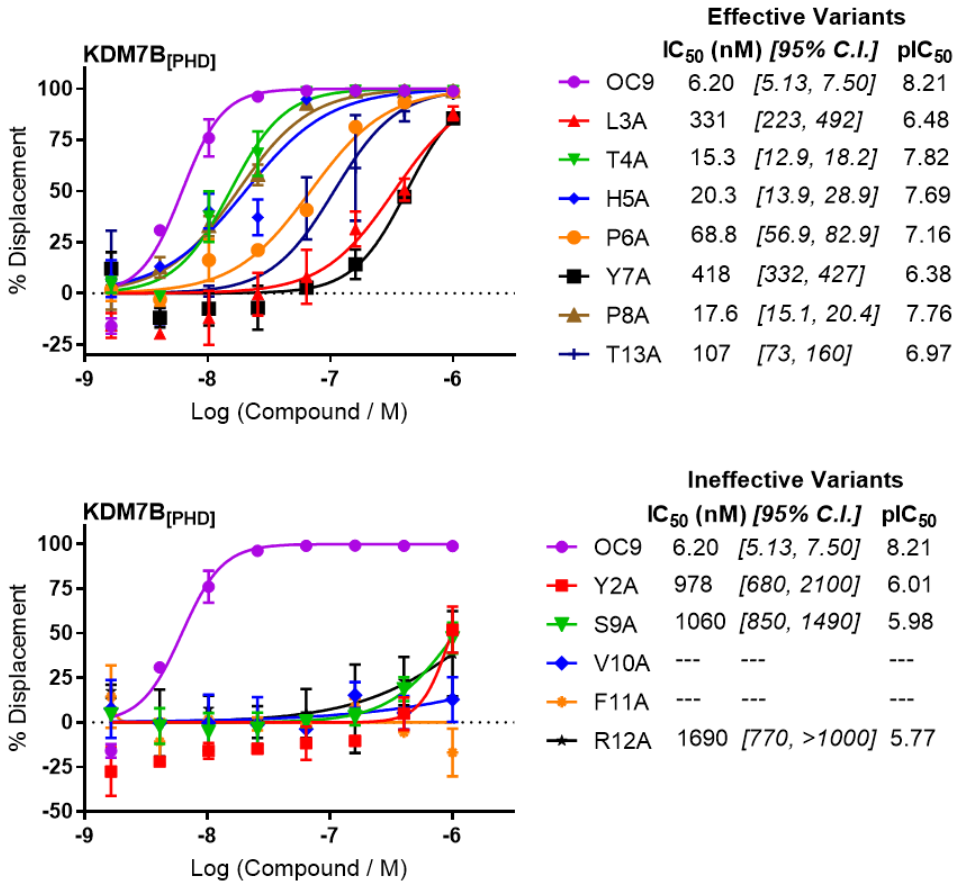
All OC9 alanine variants were assessed by AlphaScreen assay for displacement of biotinylated H3K4me3 [1-21] peptide from $KDM7B_{[PHD]}$ (**F.7.2.i.**), as well as testing by AlphaScreen for inhibition of $KDM7B_{[PHD+JmjC]}$ demethylase activity against H3K4me3K9me2 substrate (**F.7.2.ii.**).

Three OC9 alanine variants (T4A, H5A, P8A) were very well tolerated, having minimal negative impact on displacement efficacy compared to OC9. Considering the loss of functional groups on being converted to Ala, these residue side chains are unlikely to play major roles in either PHD finger binding or critical internal-structuring of OC9. The fact that P8A is well tolerated is informative, as alanine is often predominantly a *trans*-amide, indicating that the proline at this position could favour the *trans*-amide (*vide infra* 7.6.).

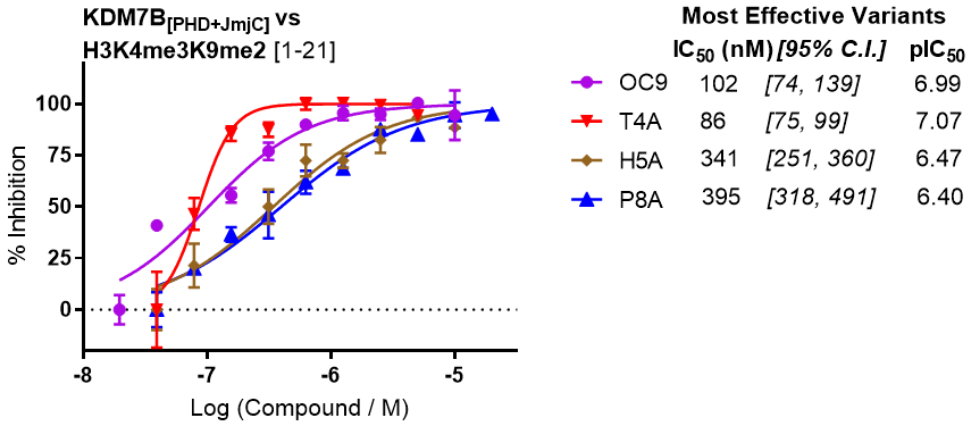
Four alanine variants were moderately tolerated (L3A, P6A, Y7A, T13A) weakening displacement efficacy by 1-2 orders of magnitude. Five alanine variants were poorly tolerated (Y2A, S9A, V10A, F11A, R12A) weakening displacement efficacy by ≥2 orders of magnitude, with the adjacency of residues 9-12 indicating them to be a potential critical interaction surface. Notably, V10A and F11A showed no indication of any displacement up to 1 µM, indicating they were particularly important residues. For the inhibition assay, only the three least detrimental alanine variants (T4A, H5A, P8A) appeared to retain any discernible inhibitory efficacy, again correlating well with their displacement potency remaining very similar to OC9 (**F.7.2.iii.**).

Interestingly, several of the amino acid sites within OC9 that were found empirically to tolerate substitution to alanine could be predicted from the original NGS data. Using the peptide family analysis with fuzzy amino acid matching and isolating peptides within the OC9

family at Hamming distance 6 (Chapter 2) had given 133 sequences. This data could be visually represented with a WebLogo plot [285] (**F.7.2.iv.**) which makes clear the variability of positions T4 and H5 and shows a highly conserved SVX_1X_2T (X_1 = F/Y, X_2 = R/K) motif, with its essential nature for binding to $KDM7B_{[PHD]}$ borne out experimentally in the alanine scan results. This computational analysis did not make obvious the importance of leucine at position 3, or the lack of penalty to substitution of proline at position 8. The visualisation does, however, suggest alternate substitutions instead of alanine which could be made at some positions and would be interesting to explore in future work. Similar logo-plot analysis was performed for OC3 and OC6 (see Appendix) and will be useful in future work.

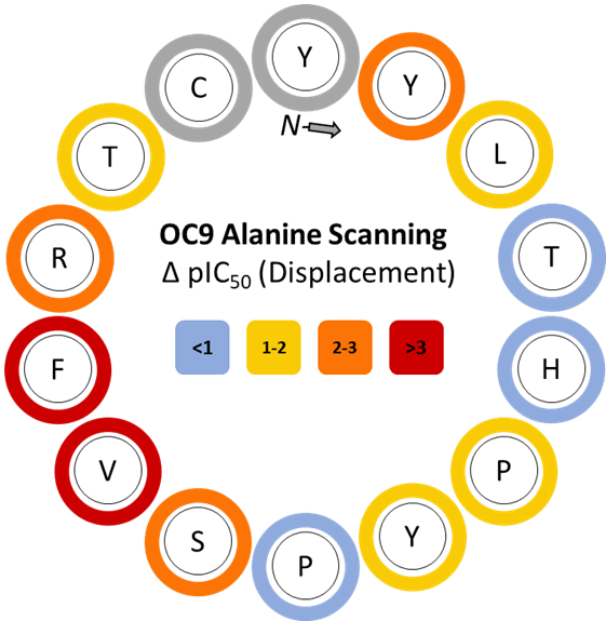


F.7.2.i. OC9 alanine scan peptides assessment by AlphaScreen peptide displacement assay against $KDM7B_{[PHD]}$. The compounds were titrated against a fixed concentration of $KDM7B_{[PHD]}$ with H3K4me3[1-21]-GGK-Biotin cognate binding partner in an AlphaScreen displacement assay. The IC_{50} and 95 % confidence interval values for displacement of the protein-protein interaction are summarised. Data were fitted using Prism 7: log(inhibitor) vs. normalized response -- Variable slope. Full conditions are given in Materials and Methods.

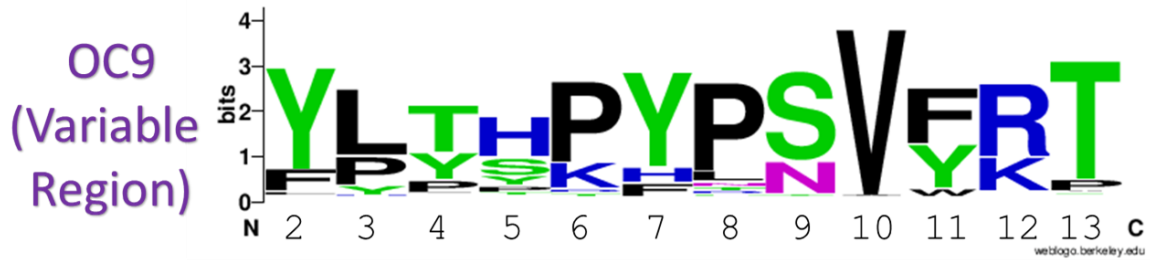


F.7.2.ii. OC9 alanine scan peptides assessment by AlphaScreen assay against KDM7B$_{[PHD+JmjC]}$ demethylation activity. Only the alanine variants that possessed inhibitory potency are displayed. An AlphaScreen assay was used to assess OC9 alanine variants for inhibition of KDM7B$_{[PHD+JmjC]}$ demethylation of H3K4me3K9me2 [1-21-GGK-biotin] substrate by antibody detection of generated H3K9me1 in the presence of ascorbate, 2OG, Fe(II). IC_{50} values and 95 % confidence intervals are summarised. Analysis of inhibition curves with Prism 7.0 fitted by: log(inhibitor) vs. normalized response -- Variable slope. Full conditions are given in Materials and Methods.

OC9	Displacement pIC_{50}	Δ(WT-Ala) Dis. pIC_{50}	Inhibition pIC_{50}
WT	8.21		6.99
Y2A	6.01	2.20	< 5.00
L3A	6.48	1.73	< 5.00
T4A	7.82	0.39	7.07
H5A	7.69	0.51	6.47
P6A	7.16	1.05	< 5.00
Y7A	6.38	1.83	< 5.00
P8A	7.76	0.45	6.40
S9A	5.98	2.23	< 5.00
V10A	< 4.00	8.21	< 5.00
F11A	< 4.00	8.21	< 5.00
R12A	5.77	2.43	< 5.00
T13A	6.97	1.24	< 5.00



F.7.2.iii. The OC9 alanine scan peptide displacement and inhibition potencies against KDM7B are summarised. The alanine substitutions that gave little change in binding potency (T4, H5, P8) also effected little change in displacement and inhibition potency. Those with the weakest binding affinities also exhibited the weakest displacement affinities. The valine-10 and phenylalanine-11 residues stood out for being crucial to displacement efficacy.



F.7.2.iv. A WebLogo plot showing the variations in amino acids in NGS-selection data at the variable positions of OC9. The represented data are from all members of the OC9 peptide 'family', derived from the original KDM7B$_{[PHD+JmjC]}$ selection results (Chapter 2) with fuzzy-amino acid matching at a Hamming distance of 6. The positions 9-13 (SVFRT) are highly conserved in terms of amino acid properties, with valine-10 appearing to be almost entirely unchanged. These observations are supported by alanine scanning results from BLI binding experiments.

The fact that F11A proved crucial to displacement possibly signified an important role for the aromatic phenyl ring in competing with the cognate histone peptide, and it would be worthwhile investigating further whether this is the case. The phenylalanine residue could reasonably be supposed to have the potential to engage in a pi-pi interaction with the aromatic cage of the PHD finger to effect displacement. Previous tryptophan fluorescence studies on OC9 interaction with the KDM7 PHD finger (Chapter 6) had shown a decrease in observed fluorescence at higher OC9 concentrations – given the known ability of phenylalanine residues to quench tryptophan fluorescence, direct insertion of F11 to the aromatic cage is possible.

The criticality of valine-10 to displacement was also interesting, and is supported by valine-10 being essentially constant in the NGS results from original mRNA display selection (**F.7.2.iv.**). The valine-10 could be posited to be vital either to OC9 active-structure conformation or to a key hydrophobic protein-protein interaction. A reported survey of 1654 globular proteins [286] revealed similar behaviour between valine and alanine in terms of their number and environment context of protein contacts, but alanine is approximately only half as hydrophobic as valine. It would therefore seem unlikely that weakening of a single hydrophobic protein-peptide interaction to a 14-mer cyclic peptide would be the lynchpin to binding, but critical valine to alanine mutations in proteins are not entirely without precedent [287]. Computational reports have demonstrated that a single valine residue can be highly influential on cyclic peptide conformational populations due to internal steric clashing [288]. On balance, this argues mostly in favour of valine-10 being more important to OC9 adopting an active conformation. Regardless of the precise cause, the near-total lack of displacement exhibited by OC9 V10A and F11A makes them useful negative control compounds for future experiments.

7.3. Second Generation Variants: Residue Selection

Further alteration of the OC9 sequence was sought to investigate the importance of specific functional groups and to attempt to improve physicochemical properties for cell activity. A mixture of computational and empirical approaches were taken, since the prediction of cyclic peptide properties in general is difficult [289], even for log P values [290]. There is a significant need in the field of cyclic peptide research for high throughput methods to guide development of desirable properties like cell permeability. Unfortunately, accurate high-throughput computational prediction of the membrane permeability of peptides and cyclic peptides remains a challenging pursuit. Moreover, peptides and other high-weight macrocycles can possess dynamic secondary structure interactions giving rise to chameleonic behaviours [291] which aid both passive and active transport – as evidenced by cyclosporin [292,293]. The ability to compute all of these complex internal and external peptide interactions simultaneously, let alone the biological pathways that may supersede them, is not currently available. Ultimately, the dynamic behaviours of individual peptides often prove the most important determinants for cell permeability and other physicochemical properties on a case-by-case basis.

Basic physicochemical properties of OC3, OC6 and OC9 were tested at AstraZeneca as part of the SBM CDT collaboration (**F.7.3.i.**). Aqueous solubility was assessed by incubation in pH 7.4 phosphate buffer for 24 hours and analysis of supernatant by LC-UV quantification. OC9 and OC3 possessed nearly millimolar levels of aqueous solubility, making them highly compatible with biochemical testing environments. OC6 was still soluble in aqueous conditions but apparently only up to low micromolar concentrations; this may help account for some of the unusual assay behaviours observed at higher micromolar OC6 concentrations [294] (Chapter 3, 4). The logD value, a measure of compound concentration distribution between pH 7.4 buffer

and octanol, was experimentally determined and found to be between 0 and -1 for each cyclic peptide which indicated that each favoured the aqueous phase. This would reasonably be expected for compounds possessing high total polar surface areas (PSAs), as was computationally predicted (> 600 Å) for these cyclic peptides. A Caco-2 assay, using a human colorectal adenocarcinoma HTB-37 epithelial cell line, was used as a model system for preliminary assessment of permeability across a gut-like cell membrane. None of OC3, OC6, or OC9 possessed appreciable passive membrane permeability under the tested conditions.

The stability of each cyclic peptide was tested by incubation separately with either rat hepatocyte cells (RHC) or human liver microsomes (HLM – vesicles with no cell membrane that contain important metabolizing enzymes e.g. CYPs and UGTs). These incubations give an approximation for intrinsic clearance (CL_{int}) – the susceptibility of the cyclic peptide to metabolism by liver enzymes without blood flow or blood-protein binding restrictions. In general, for human microsomal assays, CL_{int} (µL/min/mg) is considered: low < 10, medium 10 – 55, high > 55, and for rat hepatocyte assays CL_{int} (µL/min/10^6 cells) is considered: low < 7, medium 7 – 39, high > 39 [295]. By these criteria, OC3 has high HLM and RHC clearance whilst OC6 and OC9 have medium HLM and RHC clearance.

	MW	logD	Aqueous Solubility µM	Caco-2 Papp (x10^{-6} cm/s)	PSA Å	Rat Hepatocyte CLint (µL/min/x10^6 cells)	Human Liver Microsome CLint (µL/min/mg)
OC3	1644	-0.80	990	<0.90	623	48	69
OC6	1906	-0.24	3.6	---	706	16	31
OC9	1786	-1.00	>1000	---	705	26	34

F.7.3.i. An assessment of basic physicochemical properties for the cyclic peptides was conducted by AstraZeneca, as well as a Caco-2 permeability assay. PSA was computationally predicted, the rest of the parameters were experimentally determined. See Materials and Methods for details.

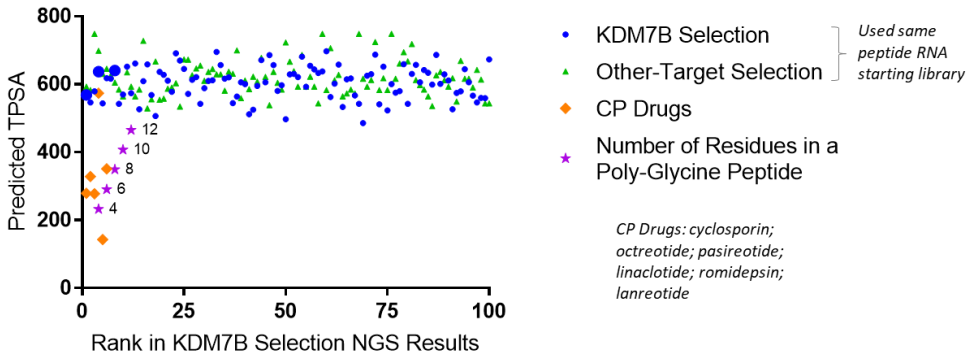
One of the most likely future use cases for the cyclic peptides was considered to be study of the KDM7s in a cellular context, requiring delivery across a cell membrane. The logD < 0 and lack of detectable permeability by Caco-2 assay suggested that it would be desirable to try and increase cyclic peptide compatibility with membrane environments, and thus the likelihood of utility in cellular assays. A higher throughput method was sought to aid development towards this end. The intrinsic clearance results were less immediately demanding of attention, as they were sufficient for such an intended use, but would likely require further optimisation in any future in-vivo work [296].

Whilst computational prediction of cyclic peptide behaviours carries known limitations, useful analyses have been reported which make links between the number of hydrogen bond donors and acceptors, PSA, molecular weight, and the likelihood of some form of membrane permeability [297–299]. Typically, peptides greater than 1000 Da and/or with a PSA greater than 200 Å will struggle for passive permeability / oral bioavailability, not accounting for chameleonic behaviour, but these guidelines are continuously reviewed and updated in the literature. Recently Villar et al [300] suggested a molecular weight up to 1300 Da and PSA up to 500 Å may be permissible for non-oral delivery of macrocycle drugs. Thus, it was decided to adopt a computationally guided approach to increasing membrane compatibility by lowering PSA and molecular weight and reducing hydrogen bonding as a reasonable path for useful development of the OC9 sequence.

Physicochemical property prediction for the cyclic peptides was conducted using the built-in tools with ChemDraw Professional v16 (**F.7.3.ii.**). The software compatibility with Microsoft Excel enabled automated analysis of large sets of cyclic peptides to track the effect of any chemical changes on predicted PSA. The predicted properties of the top 100 most abundant

sequence hits from the original KDM7B$_{[PHD+JmjC]}$ mRNA display selection did not show any peptides with major PSA deviations suitable for additional investigation. This is largely due to the PSA being inextricably linked to size of the peptide – as evidenced by its predicted linear increase with sequential addition of glycine residues to a peptide and the similar overlaid PSA values of hits from another mRNA display selection using the same starting library against an unrelated target. Example commercial cyclic peptide therapeutics are also given for comparison, with their predicted PSAs being notable for typically being much lower (< 400 Å) than the hit sequences.

Cyclic Peptide Name	Amino Acid Sequence	MW (Da)	# H DONORS	#H ACCEPTORS	cLogP	MolNet cLogP	Diameter (Å)	TPSA (Å)
OC3	YPVLLVYPRRLIC	1644	20	22	6.47	3.85	30	568
OC6	YLVRIYPNKLIWWC	1920	23	24	5.18	4.27	33	637
OC9	YYLTHPYPSVFRTC	1786	23	26	1.64	-0.32	31	641



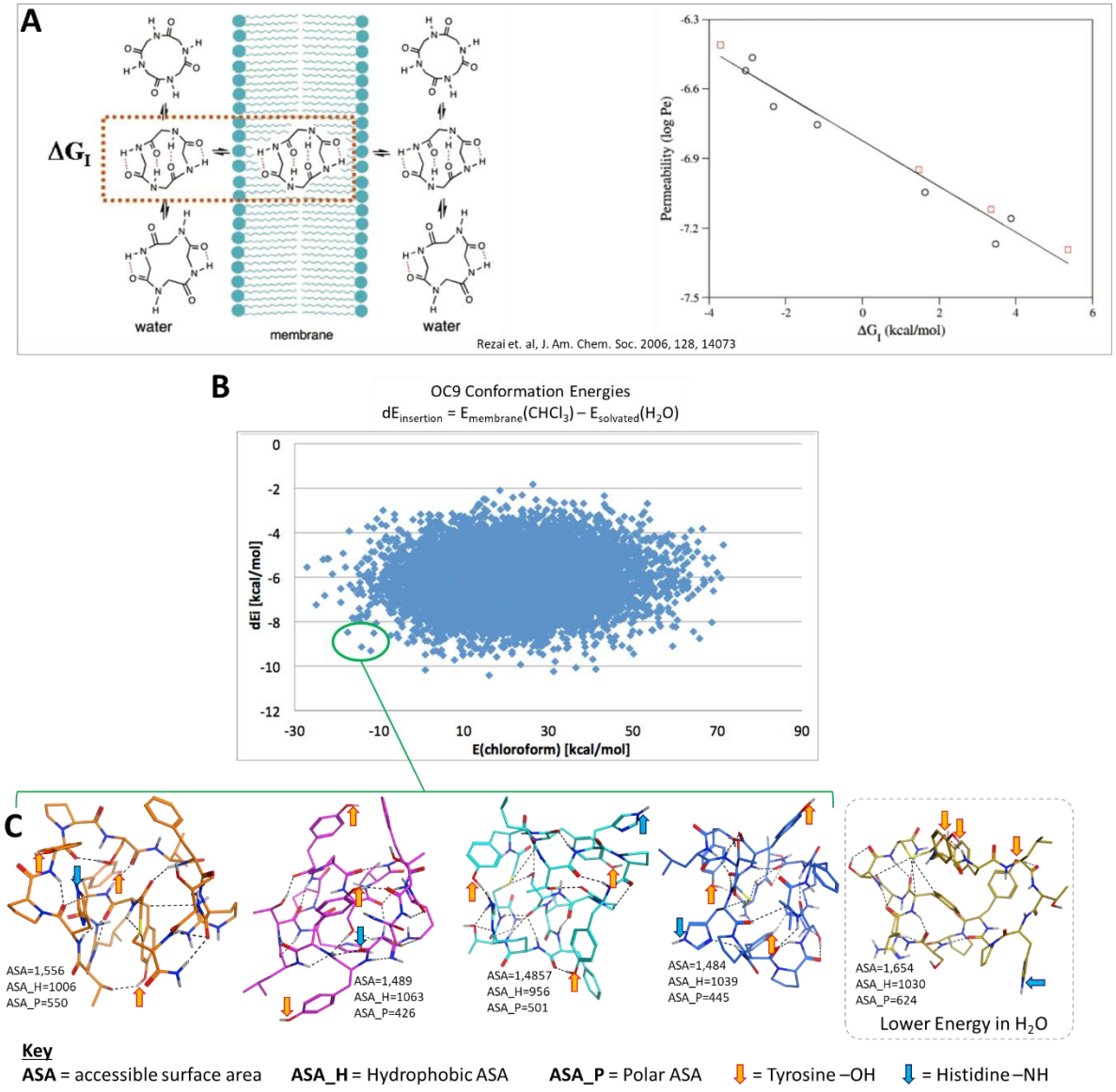
F.7.3.ii. Basic physicochemical properties were computationally predicted (using ChemDraw Profession v16) for the three lead cyclic peptide hits (OC3, OC6, OC9) resulting from selection against KDM7B. The total polar surface area (TPSA) for each of the top 100 most enriched peptides resulting from selection against KDM7B were plotted against those for peptides derived from selection against a different target, but using the same starting library. Some commercial cyclic peptide drugs are also plotted for comparison, alongside the predicted TPSA for a poly-glycine peptide increasing in size. The TPSAs of the peptides are strongly linked to their size, with most hits from a selection varying within 100 TPSA units. OC3, OC6 and OC9 are indicated with larger blue circles.

In addition to PSA, prediction of the free energy of insertion (ΔGi) of peptides into a membrane (from a high to a low dielectric medium) has been reported to correlate well with an empirical approach – the parallel artificial membrane permeability assay (PAMPA) [301]. To enlist further computational expertise and help identify the most time-economical route for peptide modification, collaboration was established with Vertex Pharmaceuticals through the SBM CDT. Dr Ewa Chudyk and Dr Ron Knegtel from Vertex conducted modelling, using in-house software, of OC9 conformational energies in high dielectric (water) and low dielectric (chloroform) environments (**F.7.3.iii.**).

The modelling sampled 10,000 conformations and was not able to guarantee sampling of all available conformational space, but nonetheless provided a large data set for guidance. A low population overlap was observed between the desired low energy chloroform-conformations (stable in membrane) and negative ΔGi values (spontaneous phase transition). For the conformations that did overlap, a common feature was solvent exposure of the histidine and three tyrosine side-chains. The polar tyrosine hydroxyl moiety would be well solvated in water through hydrogen-bonding and present an enthalpic energy penalty for transition from high to low dielectric media, making investigation of its removal a priority in new OC9 variants. Additional suggestions for improving membrane permeability included replacement of the arginine with a non-ionisable residue (difficult given its importance in alanine scanning results), methylation of solvent exposed backbone amide N-H groups and lowering of the overall molecular weight.

A set of second generation OC9 variants (**F.7.3.iv.**) were thus designed based on: the toleration of a given site for alanine substitution, computationally predicted parameters (e.g.

lowering TPSA), reduction of hydrogen bonding, reduction of molecular weight and introduction of non-natural residues (e.g. D-alanine) for increased metabolic stability.



F.7.3.iii. A) Schematic representation of peptide conformational rearrangements moving from water to membrane, and the associated free energy equilibria. Figure reprinted with permission from Rezai et al. [301] Copyright 2006 American Chemical Society. **B)** Modelling of OC9 conformational energies in different dielectric environments (ΔGi on the y-axis, ΔG[chloroform] on the x-axis) – performed by Vertex Pharmaceuticals. **C)** Example OC9 conformations with low membrane (chloroform) energy and negative ΔGi energy (favourable for membrane insertion). The exposed tyrosine and histidine polar side chain residues are highlighted.

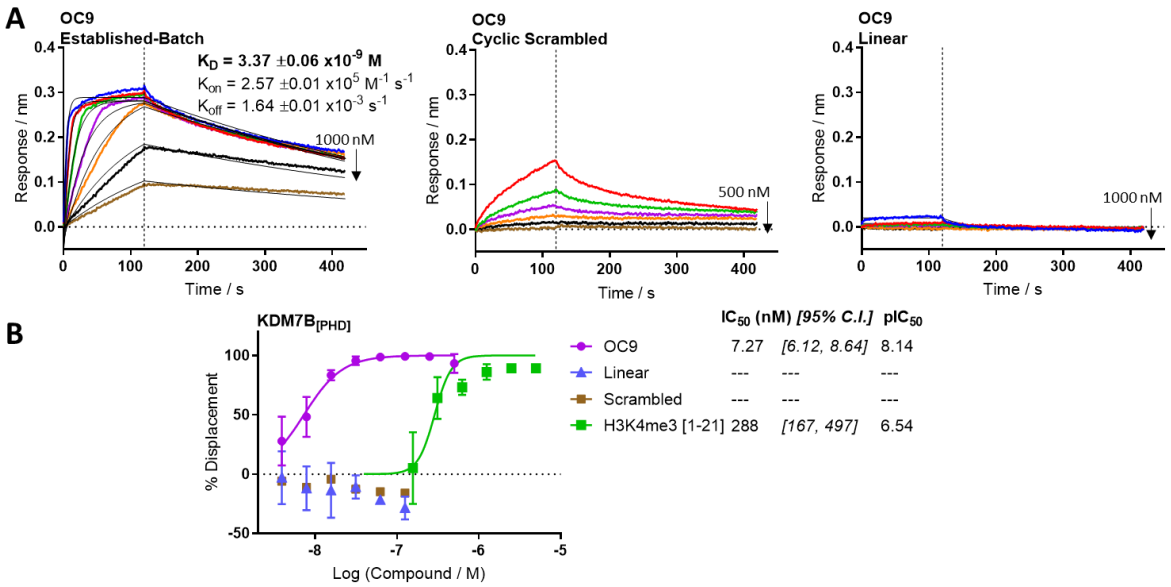
OC9 Variant	Amino Acid Sequence
Wild-Type	YYLTHPYPSVFRTC
Y1F	**F**YLTHPYPSVFRTC
Y2F	Y**F**LTHPYPSVFRTC
Y7F	YYLTHP**F**PSVFRTC
Y1F Y2F Y7F	**FF**LTHP**F**PSVFRTC
T4A H5A P8A	YYL**AA**PY**A**SVFRTC
OC9.2.0	**FF**L**AA**P**F**PSVFRTC
OC9 T4-dAla	YYL**a**HPYPSVFRTC
OC9 H5-dAla	YYLT**a**PYPSVFRTC
OC9 P8-dAla	YYLTHPY**a**SVFRTC

F.7.3.iv. Table of 'Second Generation' variants of OC9. Variants were designed based on substituting sites that had shown tolerance to alanine substitution, in order to to attempt to reduce the total polar surface area, number of hydrogen bond donors / acceptors and introduce non-natural residues.

7.4. Second Generation Variants: Binding, Displacement, Selectivity

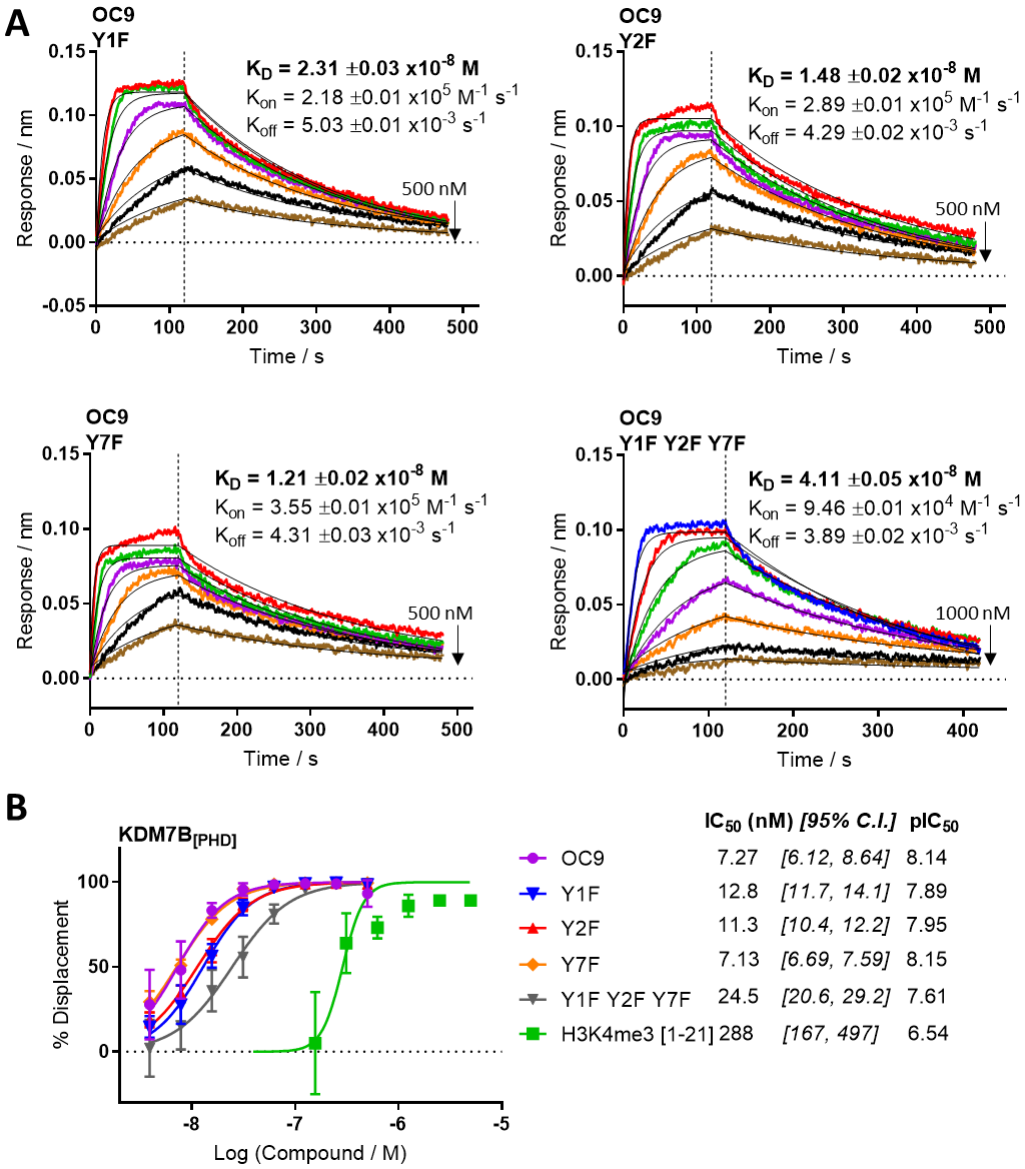
New OC9 variants were synthesised in stages, using information from one set to inform the next. Synthesis and purification of the variant peptides was carried out by Shane Clark (MSc Student, Oxford Pharmacology). The new OC9 variants were assessed for binding to $KDM7B_{[PHD]}$ by BLI and displacement of biotinylated H3K4me3 [1-21] peptide by AlphaScreen (with H3K4me3 [1-21] control).

The first variant set established the importance of the cyclisation (**F.7.4.i.**). The linear OC9 sequence (acetyl capped at the N-terminus) had no detectable binding or displacement of $KDM7B_{[PHD]}$. Thus, cyclisation is likely essential for pre-organisation of the peptide into the proper conformation for binding – the high entropic cost associated with organising a 14-mer linear peptide to a similar state, given the large number of accessible conformations, makes it otherwise unlikely. A scrambled cyclic OC9 sequence gave an unexpectedly large BLI response, but it did not exhibit any displacement of $KDM7B_{[PHD]}$ in an AlphaScreen displacement assay and so was considered to be most likely due to non-specific interaction. This supports OC9 interaction (K_D = 3.37 nM, IC_{50} = 7.27 nM) being sequence-dependent rather than due to amino acid content alone.



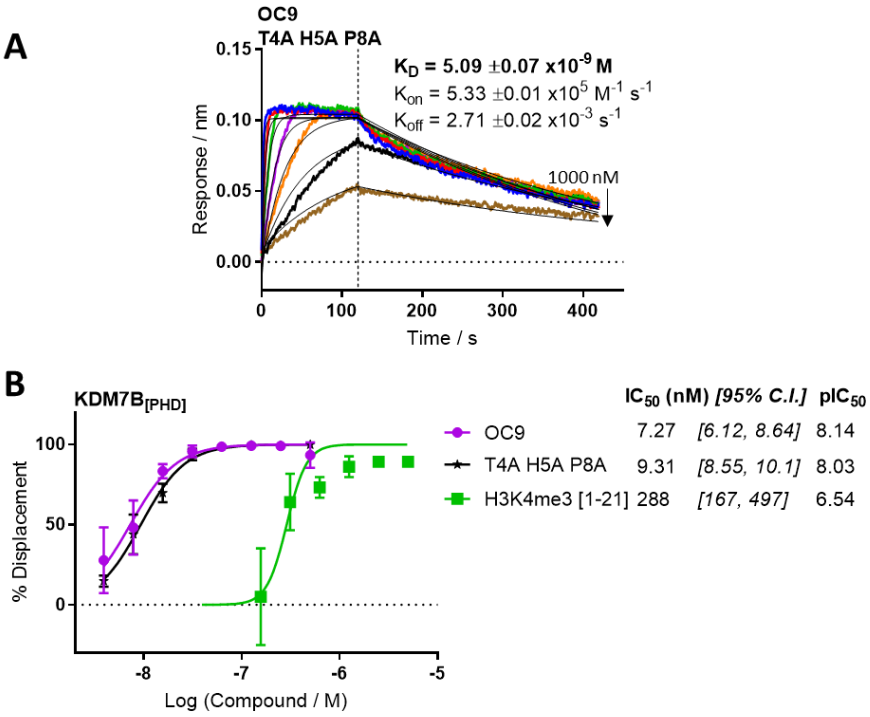
F.7.4.i. OC9 variants (Linear with acetyl cap on N-terminus, and a Scrambled Cyclic version) were tested against KDM7B$_{[PHD]}$ for **A)** binding by BLI and **B)** displacement of biotinylated H3K4me3 [1-21] peptide by AlphaScreen. The linear peptide gave no binding response or displacement, and the scrambled sequence was considerd most likely to bind only non-specifically (**A**) as it gave no displacement (**B**). Experiment conditions are given in Materials and Methods.

Substitution of each tyrosine to phenylalanine, removing hydroxyl functionality but retaining aromaticity, did not significantly adversely affect binding to or displacement of KDM7B$_{[PHD]}$, either by individual substitution or in tandem at all three positions: K_D (nM) = 23.1 [Y1F], 14.8 [Y2F], 12.1 [Y7F], 41.1 [Y1F, Y2F, Y7F]; IC_{50} (nM) = 12.8 [Y1F], 11.3 [Y2F], 7.13 [Y7F], 24.5 [Y1F, Y2F, Y7F]) (**F.7.4.ii.**). This was supported by the NGS data WebLogo plots (**F.7.2.iv.**). The tyrosine hydroxyl groups are therefore unlikely to be essential for interaction with KDM7$_{[PHD]}$, with their side chain bulk and aromaticity the more important features (OC9 efficacy is abrogated as Y2A and weakened as Y7A).



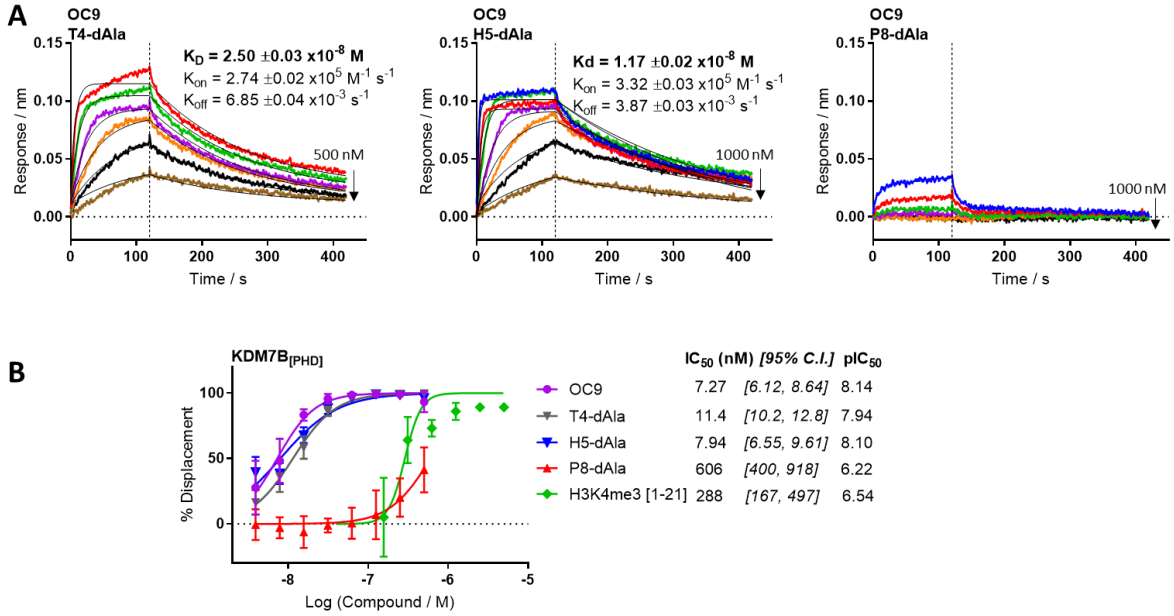
F.7.4.ii. OC9 variants substituted with phenylalanine for tyrosine were tested against KDM7B$_{[PHD]}$ for **A)** binding by BLI and **B)** displacement of biotinylated H3K4me3 [1-21] peptide by AlphaScreen. All tyrosine positions tolerated the substitution well individually, and the substitutions in tandem in a single cyclic peptide, to give interactions comparable to unmodified OC9. Experiment conditions are given in Materials and Methods.

The OC9 alanine scan had identified three sites (T4, H5, P8) that tolerated individual substitution. The combination of all three presented a facile route to lowering molecular weight and pleasingly this remained similarly effected to unsubstituted OC9 in terms of binding potency by BLI (K_D = 5.09 nM) and displacement of KDM7B$_{[PHD]}$ by AlphaScreen (IC_{50} = 9.31 nM) (**F.7.4.iii.**).



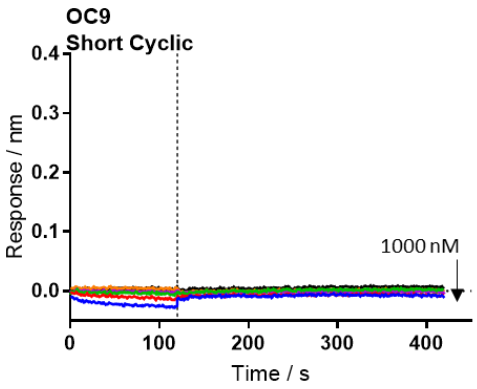
F.7.4.iii. An OC9 variant combining several low-impact alanine substitutions (T4A, H5A, P8A) to lower the overall molecular weight was tested against KDM7B$_{[PHD]}$ for **A)** binding by BLI and **B)** displacement of biotinylated H3K4me3 [1-21] peptide by AlphaScreen. Its efficacy in both assays remained comparable to unmodified OC9. Experiment conditions are given in Materials and Methods.

A straightforward route to improving peptide protease resistance is introduction of non-natural chemical matter [179]. Each of the three sites in OC9 that previously tolerated alanine substitution with little to no penalty (positions 4, 5, 8) were therefore substituted with *D*-alanine, well-precedented to confer increased proteolytic stability [302]. Testing by BLI and AlphaScreen against KDM7B$_{[PHD]}$ (**F.7.4.iv.**) revealed this substitution had no adverse effect on binding or displacement efficacy in place of T4 or H5, (K_D = 25 nM, 11.7 nM and IC_{50} = 11.4, 7.94 nM respectively) but was unfavourable to both binding and displacement in place of P8 (K_D = non-specific, IC_{50} = 606 nM) which indicated the side-chain geometry at position 8 was consequential to OC9 efficacy.



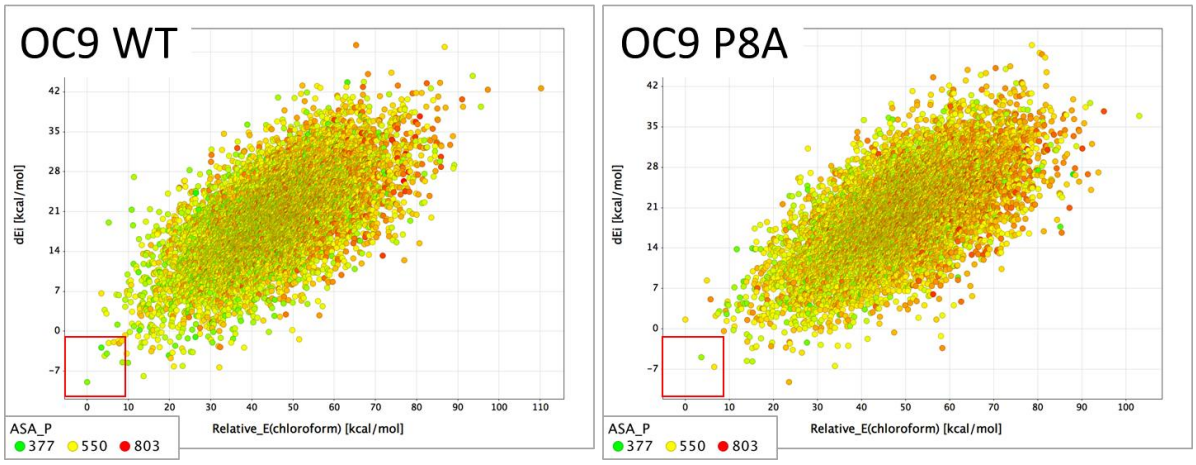
F.7.4.iv. OC9 variants substituted with *D*-alanine were tested against KDM7B$_{[PHD]}$ for **A)** binding by BLI and **B)** displacement of biotinylated H3K4me3 [1-21] peptide by AlphaScreen. The T4 and H5 positions tolerated the substitution and gave interactions comparable to unmodified OC9, but at position P8 it had a severe detrimental impact on efficacy. Experiment conditions are given in Materials and Methods.

A key motif identified by alanine scanning, SVFRT, was made as a short cyclic peptide (AASVFRTAAC) and tested by BLI to ascertain whether it retained any affinity to KDM7B$_{[PHD]}$, and thus whether it may serve as a smaller scaffold more suited to small molecule conversion (**F.7.4.v.**). No binding was observed up to 1 µM and so further testing of different peptide derivatives that include the hotspot sequence will be needed to explore what potential, if any, it possesses as a binding motif.

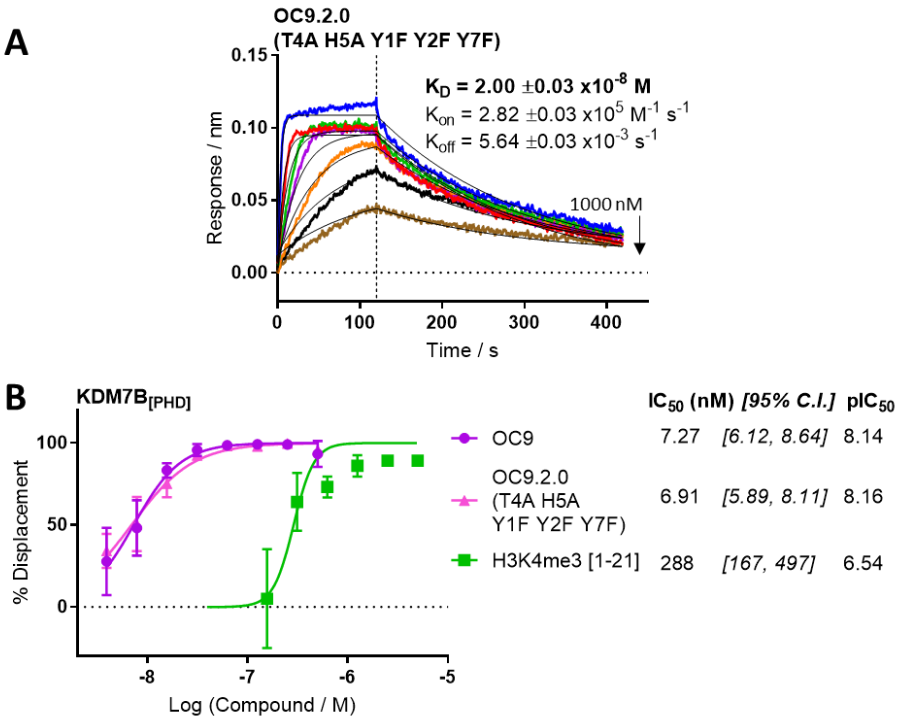


F.7.4.v. A possible OC9 key motif (SVFRT) in cyclised sequence (AASVFRTAAC) was tested against KDM7B$_{[PHD]}$ for binding by BLI, but did not show any indication of binding response up to 1 µM.

Several well-tolerated OC9 substitutions (Y1F, Y2F, T4A, H5A, Y7F) were incorporated simultaneously to create OC9.2.0, which sought to lower molecular weight and minimise hydrogen bond acceptors/donors. The P8A substitution was not included as modelling by Vertex had indicated the hydrophobic proline side chain offered access to more low-energy low-polar surface area conformations and its loss would decrease the population probability of membrane-insertion (**F.7.4.vi.**). Remarkably, combination of these substitutions into OC9.2.0 was very well tolerated, with similar BLI binding potency (K_D = 20 nM) and AlphaScreen displacement efficacy (IC_{50} = 6.91 nM) to OC9 against $KDM7B_{[PHD]}$ (**F.7.4.vii.**).

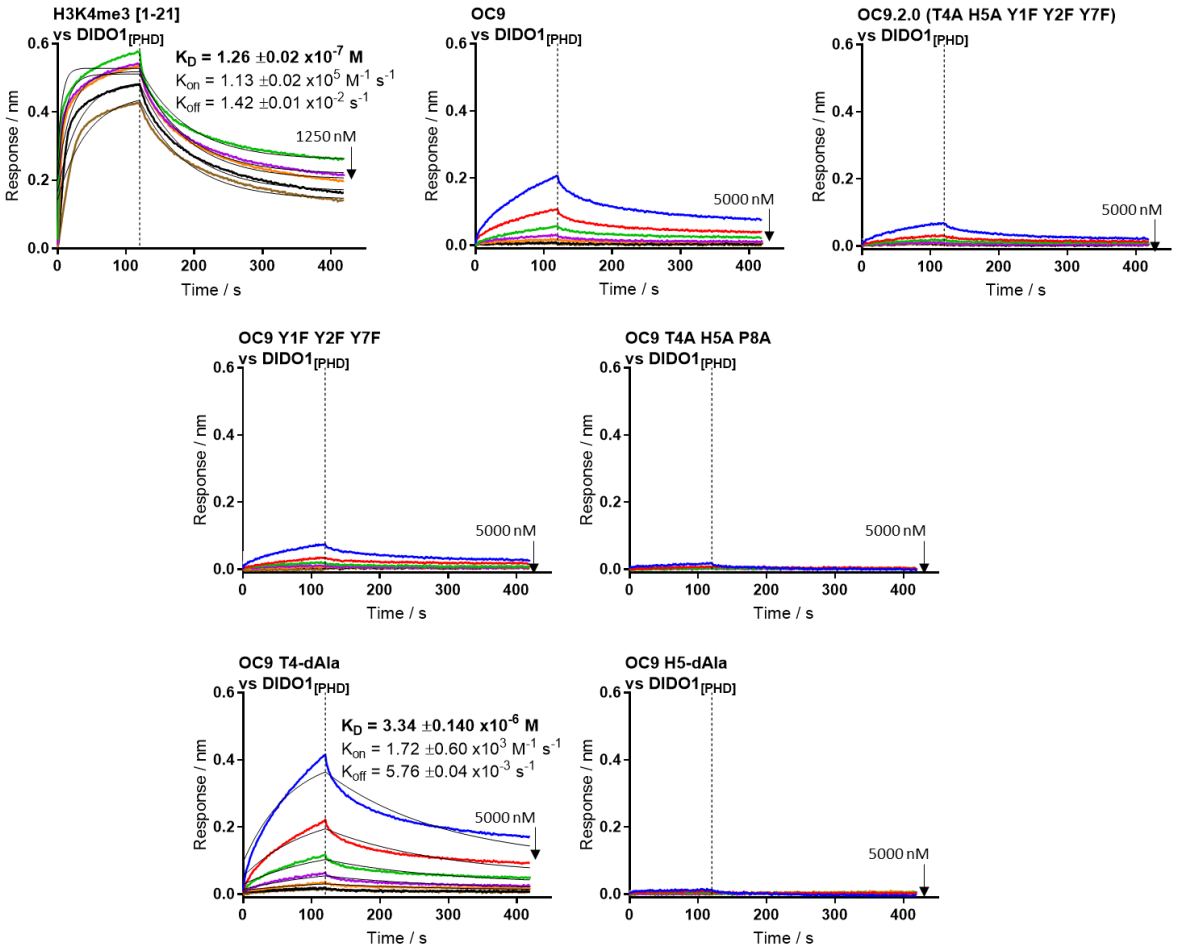


F.7.4.vi. Computational modelling of OC9 and OC9_P8A conformation energies was carried out by Ewa Chudyk and Ron Knegtel from Vertex Pharmaceuticals. The population of energy states in chloroform are plotted against the membrane-transition energy, with each data point coloured according to its predicted accessible polar surface area (ASA_P). Few conformations are predicted to possess clearly favourable membrane compatibility in either peptide (red-box), variant P8A is predicted to decrease this population further.



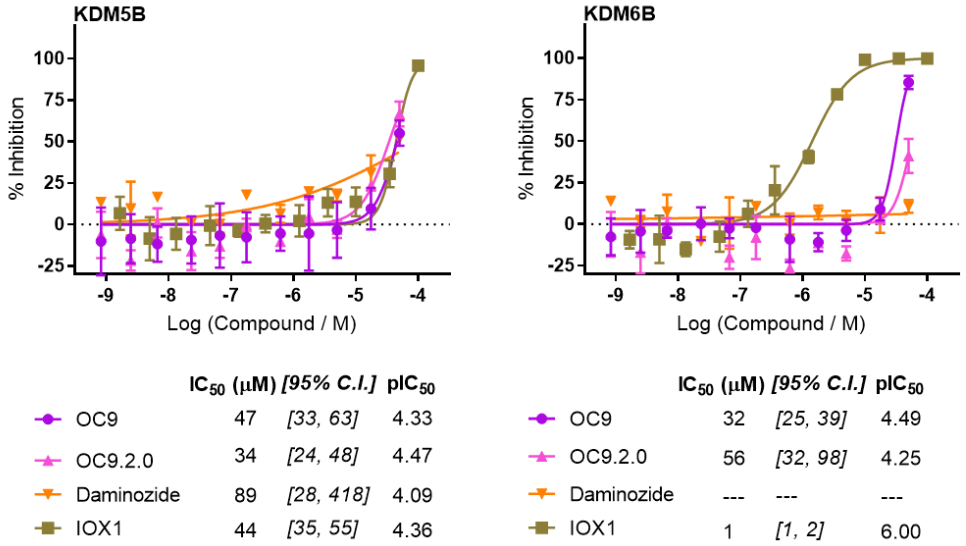
F.7.4.vii. An OC9 variant (OC9.2.0) combining several low-impact substitutions (T4A, H5A, Y1F, Y2F, Y7F) was tested against KDM7B$_{[PHD]}$ for **A)** binding by BLI and **B)** displacement of biotinylated H3K4me3 [1-21] peptide by AlphaScreen. Its efficacy in both assays remained comparable to unmodified OC9. Experiment conditions are given in Materials and Methods.

Several of the OC9 variants, ([OC9.2.0]; [OC9 Y1F Y2F, Y7F]; [OC9 T4A H5A P8A]; [OC9 T4-dAla]; [OC9 H5-dAla]), had their selectivity assessed against the closely related DIDO1$_{[PHD]}$ by BLI (**F.7.5.viii.**). H3K4me3 [1-21] was included as a positive functional control, and OC9 itself was re-tested as a negative control. Only the OC9 T4-dAla variant showed binding responses which could be meaningfully fitted with the Association then Dissociation model. Thus, in addition to previous ITC data confirming lack of OC9 binding to DIDO1$_{[PHD]}$, this data supported the likely only non-specific nature of the observed responses. The T4-dAla variant should be confirmed for non-specific interaction in an alternative assay. Overall, these selected substitutions of OC9 did not appear to render any significant loss of binding selectivity over DIDO1$_{[PHD]}$.



F.7.4.viii. OC9 substitution variants were tested for binding selectivity against $DIDO1_{[PHD]}$ by BLI. Data were examined with the Association then Dissociation model and only the cognate H3K4me3 could be sensibly fitted with the Association then Dissociation model. Experiment conditions are given in Materials and Methods.

In addition, OC9 and OC9.2.0 were tested by AlphaScreen assay against $KDM5B_{[JmjC]}$ and $KDM6B_{[JmjC]}$ demethylase activity (performed by Anthony Tumber) (**F.7.4.ix.**). Neither OC9 nor OC9.2.0 displayed inhibition of either enzyme up to low micromolar quantity of cyclic peptide, before a sudden jump at in inhibition at 50 µM of cyclic peptide. This observation indicated possible off-target inhibition at very high concentration (> 1000-fold selectivity over the KDM7 sub-family), but this would need verification in an orthogonal assay. Importantly it indicated that OC9.2.0 did not appear more promiscuous than OC9 even with its increased hydrophobicity (**F.7.4.x.**).



	IC_{50} (μM)	*[95% C.I.]*	pIC_{50}
OC9	47	*[33, 63]*	4.33
OC9.2.0	34	*[24, 48]*	4.47
Daminozide	89	*[28, 418]*	4.09
IOX1	44	*[35, 55]*	4.36

	IC_{50} (μM)	*[95% C.I.]*	pIC_{50}
OC9	32	*[25, 39]*	4.49
OC9.2.0	56	*[32, 98]*	4.25
Daminozide	---	---	---
IOX1	1	*[1, 2]*	6.00

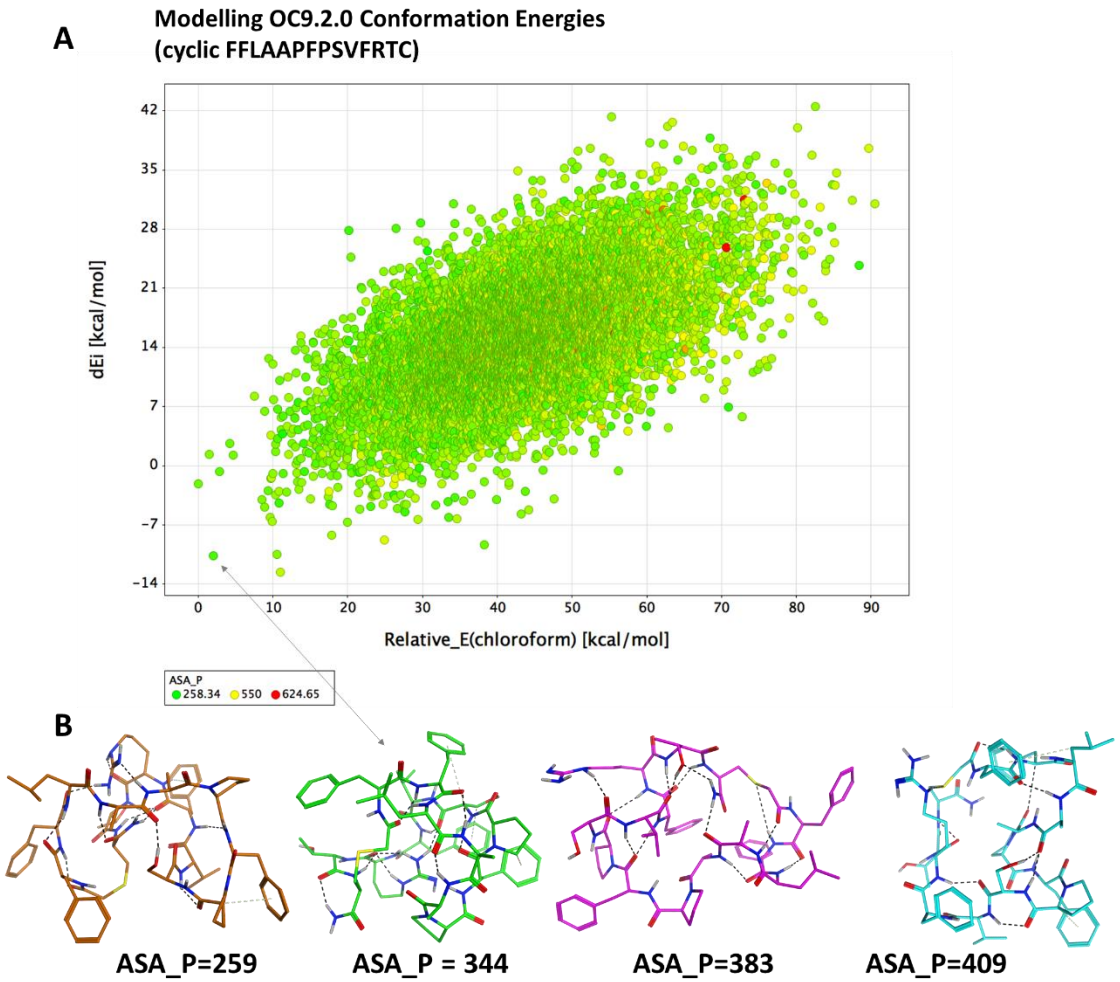
F.7.4.ix. OC9 and variant OC9.2.0 were tested for selectivity against non-KDM7 KDMs in an AlphaScreen activity demethylation assay, performed by Anthony Tumber (Schofield Lab) – assay conditions are previously reported [160,303]. Neither cyclic peptide substantially inhibited either KDM5B or KDM6B at up to micromolar concentrations. Data were fitted with Prism v7 using the log(inhibitor) vs. normalized response -- Variable slope model.

Physicochemical properties of all OC9 variants were computationally predicted (**F.7.4.x.**) to assess how far they could be considered to meet the stated objectives for this phase of research, primarily to lower polar surface area (PSA) and molecular weight whilst retaining efficacy and selectivity. Each of the variant peptides achieved one or more of these goals. Importantly, the cyclic peptide which incorporated several substitutions whilst retaining very similar efficacy to OC9, dubbed OC9.2.0, was predicted to have a PSA lowered by over 100 Å, a reduction in molecular weight from 1768 to 1642 Da and a reduction in the number of H-bond donors and acceptors by 5 and 6 respectively – making it a major step towards increasing compatibility with membrane environments. It is noteworthy that OC9.2.0 remained appreciably soluble under all encountered aqueous conditions, indicating that its increased hydrophobicity was not detrimental to handling. The increased hydrophobicity may lead to increased clearance in vivo, but this could be optimised in a future phase of work.

OC9 Variant	Amino Acid Sequence	MW (Da)	# H DONORS	#H ACCEPTORS	cLogP	MolNet cLogP	Diameter (Å)	TPSA (Å)
Wild-Type	YYLTHPYPSVFRTC	*1786*	*23*	*26*	*1.64*	*-0.32*	*31*	*641*
Y1F	**F**YLTHPYPSVFRTC	1770	22	25	2.31	0.09	31	620
Y2F	Y**F**LTHPYPSVFRTC	1770	22	25	2.31	0.09	31	620
Y7F	YYLTHP**F**PSVFRTC	1770	22	25	2.31	0.09	31	620
Y1F Y2F Y7F	**FF**LTHP**F**PSVFRTC	1738	20	23	3.64	0.91	31	580
T4A H5A P8A	YYL**AA**PY**A**SVFRTC	1664	22	23	2.85	0.30	30	605
OC9.2.0	**FF**L**AA**P**F**PSVFRTC	1642	18	20	5.34	1.73	29	535

F.7.4.x. Basic physicochemical properties of OC9 variants were computationally predicted (using ChemDraw Profession v16) to assess the impact of substitutions on their molecular weight, hydrogen bonding capability, hydrophobicity (cLogP, higher positive value indicates increased hydrophobicity) and total polar surface area (TPSA). OC9.2.0 (Y1F, Y2F, Y7F, T4A, H5A) succeeded in significantly reducing molecular weight, hydrogen bonding and TPSA by making substitutions which were also predicted to substantially increase hydrophobicity.

Computational modelling of OC9.2.0 cyclic peptide conformations, and their association to aqueous and membranous (chloroform) environments, was performed by Ewa Chudyk (Vertex Pharmaceuticals) (**F.7.4.xi**). This made it clear that the substitutions present in OC9.2.0 gave substantial benefits towards increased membrane compatibility. The accessible polar surface area was predicted to be drastically reduced, compared to OC9 (**F.7.4.vi.**), across the global set of modelled conformations, and there was an increased population of peptides predicted to have favourable membrane insertion energies (ΔGi) alongside lower energy in hydrophobic environments. Selected low-chloroform energy conformations indicated that those with low accessible polar surface areas were capable of masking their side chains from solvent exposure through internal hydrogen bonding. Together, this work supports the new variant (OC9.2.0) having improved prospects for membrane permeability.



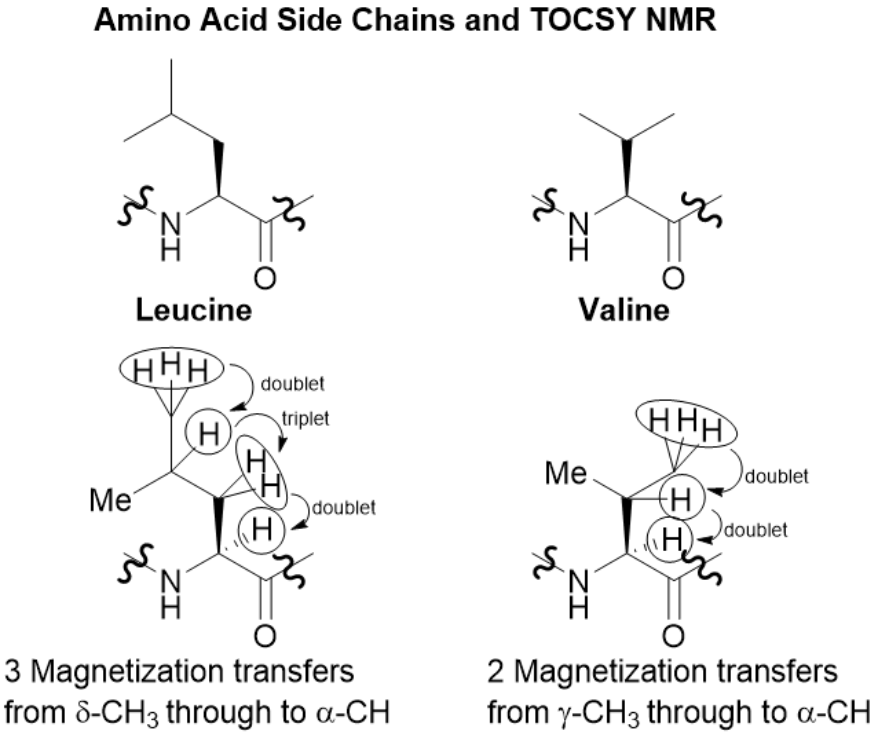
F.7.4.xi. A) Modelling of OC9.2.0 conformational energies in different dielectric environments (ΔGi on the y-axis, ΔG[chloroform] on the x-axis) – modelling performed by Ewa Chudyk, Vertex Pharmaceuticals. **B)** Example OC9.2.0 conformations with a range of accessible polar surface areas (ASA_P). The lower ASA_P conformations are able to mask their side chains with internal hydrogen bonding.

7.5. OC9 Solution NMR Peptide Studies

Experiments and analysis were performed in collaboration with Dr Nader Amin (Department of Chemistry, NMR Core Facility, University of Oxford).

Systematic synthetic work on OC9 proved effective but demanding of time. It also did not necessarily help discern whether a given alteration was responsible for correct peptide conformation or docking with the PHD finger. The computational assistance from Vertex had started with little constraint on possible cyclic peptide structure, with the resulting conformational space too large to fully sample. More precise progress could be made on both fronts if an empirical OC9 solution structure was known and solution NMR is an effective technique for guiding cyclic peptide structure development [304–306].

Solution NMR was performed on isotopically unlabelled OC9 in phosphate buffer, pH 6.3. 2D ^{1}H-^{1}H TOtal Correlated SpectroscopY (TOCSY) was used as the primary experiment for a feasibility assessment of peptide structure NMR-assignment. This is a 2D NMR technique that looks at through-bond magnetization transfer, analogously to a traditional COSY experiment, to read correlations transferred further down a chain instead of only to the nearest neighbour. It is particularly useful when studying peptides because each amino acid is considered to form an isolated spin system (the amide bond halts any magnetization transfer) with each having a diagnostic number of cross-peaks and pattern of shifts for the coupling network in its unique side-chain (**F.7.5.i.**).

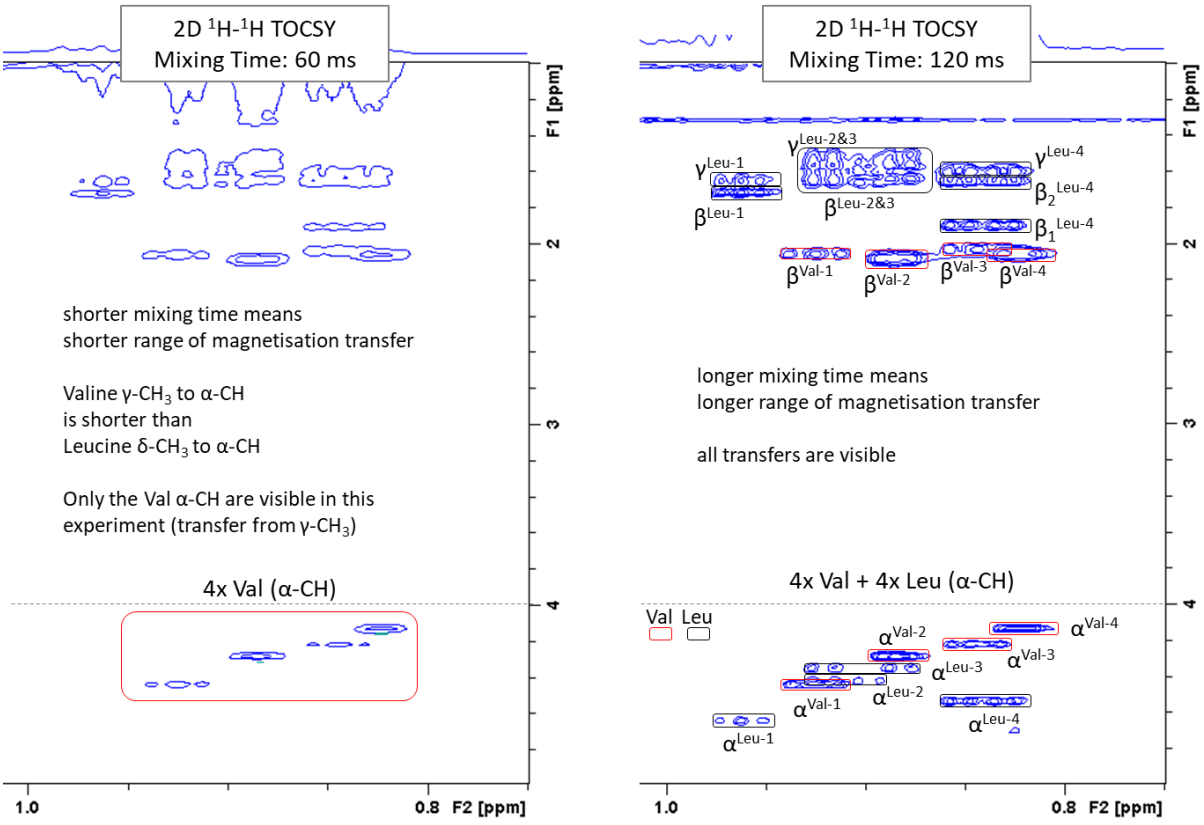


F.7.5.i. The principle of 2D ^{1}H-TOCSY NMR. Magnetization is transferred down the side chain of the amino acid, giving a characteristic number and pattern of shifts to enable identification. Magnetization transfer is halted by the amide bond, leaving each side chain as an isolated spin system.

OC9 contains the following number of residues: 3x Tyr, 1x Leu, 2x Thr, 1x His, 2x Pro, 1x Ser, 1x Val, 1x Phe, 1x Arg, 1x Cys. A single Leu and Val are the only residues present with a saturated all-carbon side chain, with a ^{1}H chemical shift for their terminal methyl (CH_3) groups expected between 0.5 to 1 ppm. Therefore, it was expected that only two sets of TOCSY cross-peaks would be observed in OC9 around or below 1 ppm; one for Leu, one for Val. In fact, four sets of signals were observed for each (**F.7.5.ii.**). The α-CH in amino acids has a characteristic shift around 4 ppm and was helpful in determining how many residue patterns were likely to be overlaid. Four signals for each residue suggested up to four OC9 conformations were present.

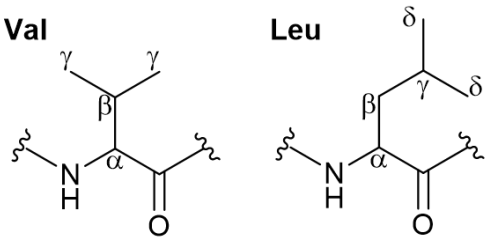
OC9 (YYLTHPYPSVFRTC)

Counting Conformations by Examining Methyl Group Side-Chain Correlations
within Leucine and Valine
(terminal -CH_3 group has characteristic ^1H signal at 0.5 to 1 ppm in both)



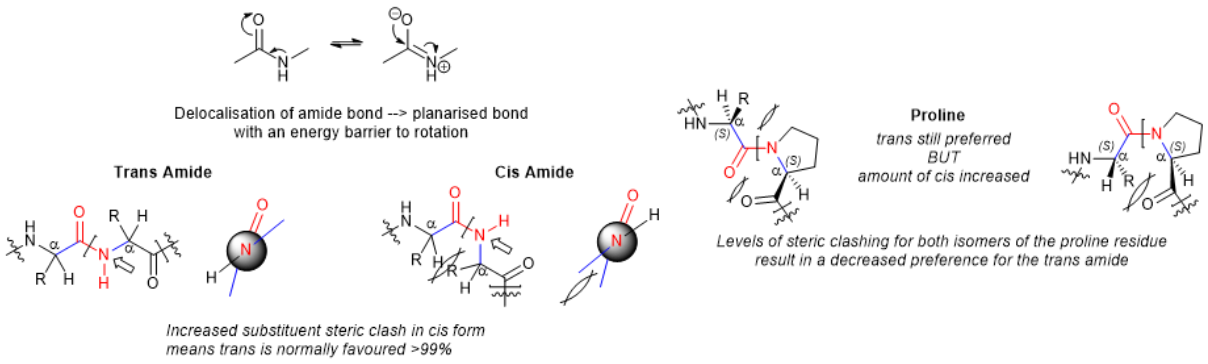
Amino Acid Representation in NMR
Predicted: Val = 1, Leu = 1
Observed: Val = 4, Leu = 4
Indicates 4x OC9 conformations

Val γ-CH_3 / Leu δ-CH_3 groups expected to give doublet coupling to β-CH / γ-CH, respectively.
In reality, peptide chemical environment can make the terminal CH_3 groups non-equivalent and give a doublet of doublets (overlapping). The Leu β-CH_2 can also behave non-equivalently.



F.7.5.ii. Comparison of ^{1}H-^{1}H 2D TOCSY NMR spectra (AV700) for OC9 (400 µM) with different mixing times (60 and 120 ms). Examination of correlations associated to the leucine and valine terminal methyl groups to help count the number of conformations visible on the NMR timescale. Buffer conditions: 25 mM phosphate, pH 6.30, 100 mM NaCl, 10% D_2O at 308 K. Final DMSO ≤ 5 %.

Whilst detrimental to the prospects of quickly determining a solution structure, this somewhat unexpected result opened a window to the impact of the proline residues in OC9. The delocalisation of electrons across an amide bond results in its planarization to either a *cis* or *trans* geometry. Ordinarily, peptide bonds are predominantly (> 95 %) *trans*, with the *cis* conformation disfavoured to minimise any steric clashing of their substituents (**F.7.5.iii.**). However, the proline side chain is cyclised onto the backbone amino group which can reduce its preference for *trans* over *cis*, with the amount of *cis*-proline-amide in proteins ranging from 5 to 30% [307–309] depending on the local chemical environment. Both *trans*- and *cis*-proline N-amides are capable of inducing β-turn features, especially useful in cyclic peptides, but levels of *cis*-proline-amide are well documented to increase at such conformationally restricted sites; a study by Vitagliano et al. on the Cambridge Structural Database in 2001 found 57% of proline amides to be *cis* in cyclic peptides, compared to just 6% for linear [310].

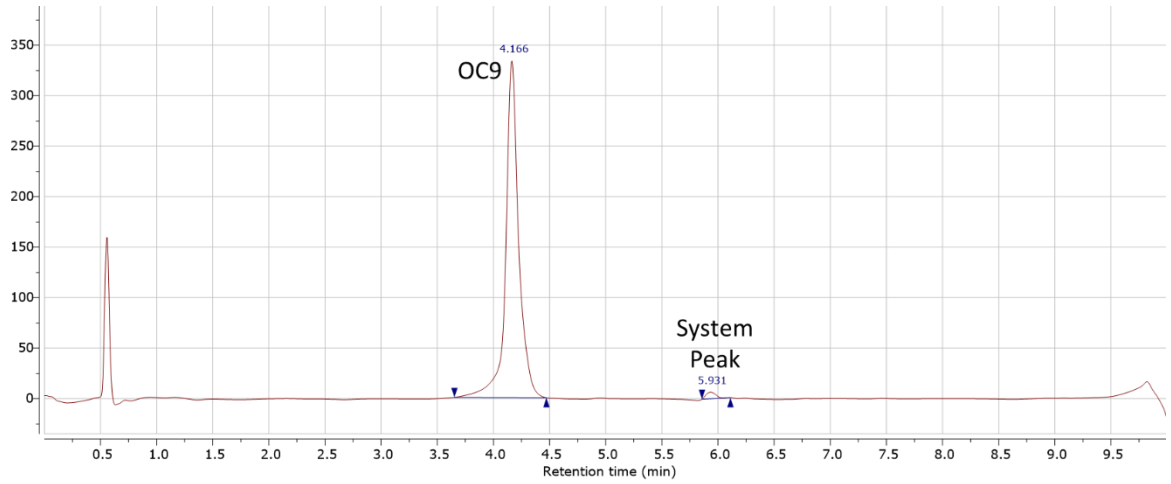


F.7.5.iii. Proline amide geometries. Proline typically still favours the *trans*-amide, but has a much reduced preference over the alternative *cis* amide compared to other amino acids due to increased steric clashing in both of the proline amide conformations.

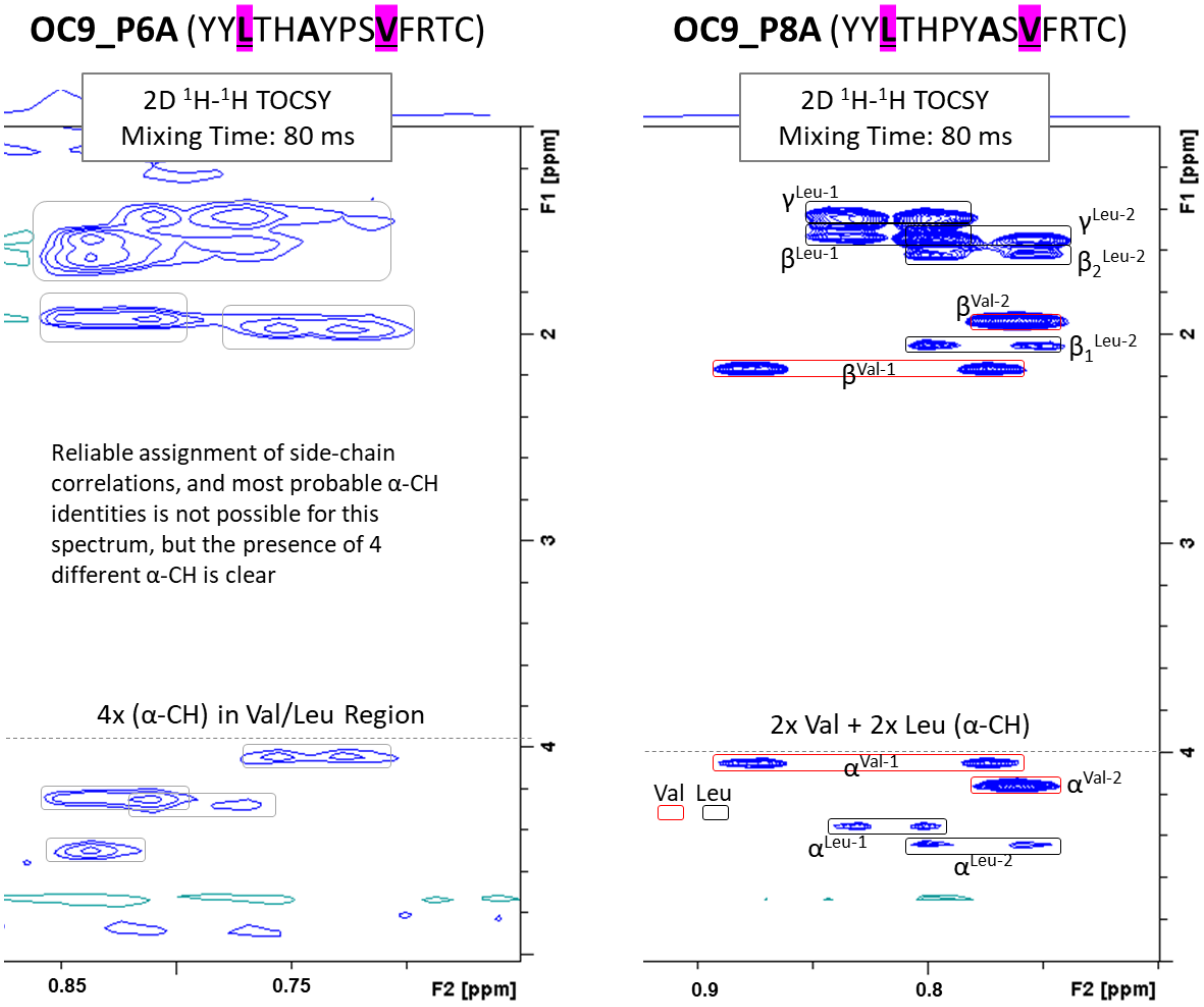
It follows that for a peptide containing *n* proline residues there are 2^n possible *cis* and *trans* conformations of the N-prolyl amide bonds. OC9 possesses two prolines, giving four theoretical conformations. A total of eight signal sets for Leu/Val residues (four signals for each) were observed by TOCSY, when only one signal set for each would ordinarily be expected. Proline-induced *cis*/*trans* isomerism is a good candidate explanation for these

observations. An alternate explanation to consider is the presence of multiple diastereomers through partial or total epimerisation of one or more amino acid stereocentres. Analogously by the 2^n rule, only two amino acid stereocentres would have needed substantial loss of chiral integrity to give four possible conformations. A small level of epimerisation during synthesis is nearly unavoidable, however the chosen activation reagents for synthesis (Oxyma, DIC) are well evidenced with Fmoc-amino acid solid phase synthesis using Rink amide resin to reduce this to very low (< 0.5% per coupling) levels [230] and it would have required that no more than 2 stereocentres had suffered epimerisation. LCMS analysis of OC9 gave one distinct peak (**F.7.5.iv.**), indicating that multiple diastereomers were unlikely to be present.

Strong support for the proline hypothesis as the main cause of multiple leucine and valine signals was found by logical extension of the TOCSY experiments to the two OC9 proline-to-alanine variants (P6A, P8A) (**F.7.5.v.**). Alanine is commonly conformationally *trans* and cannot normally offer the same access to the *cis*-amide conformation. Each individual proline-to-alanine variant possessed only two signals for Leu/Val, as predicted for a single *cis*/*trans* isomerism event with only one proline residue now present.



F.7.5.iv. Liquid Chromatography (LC) analysis of OC9 peptide. Only one peak is visible on the 210 nm wavelength indicating only one peptide species is present. Column: Vydac 218TP C18 4.6x50 mm. Solvents: A (MilliQ 0.01 % TFA), B (MeCN 0.01 % TFA) Gradient: 0 min 2 % B, 0.5 min 2 % B, 5 min 50 % B, 7.5 min 75 % B, 8.5 min 100 % B, 9 min 2 % B, 10 min 2 % B.



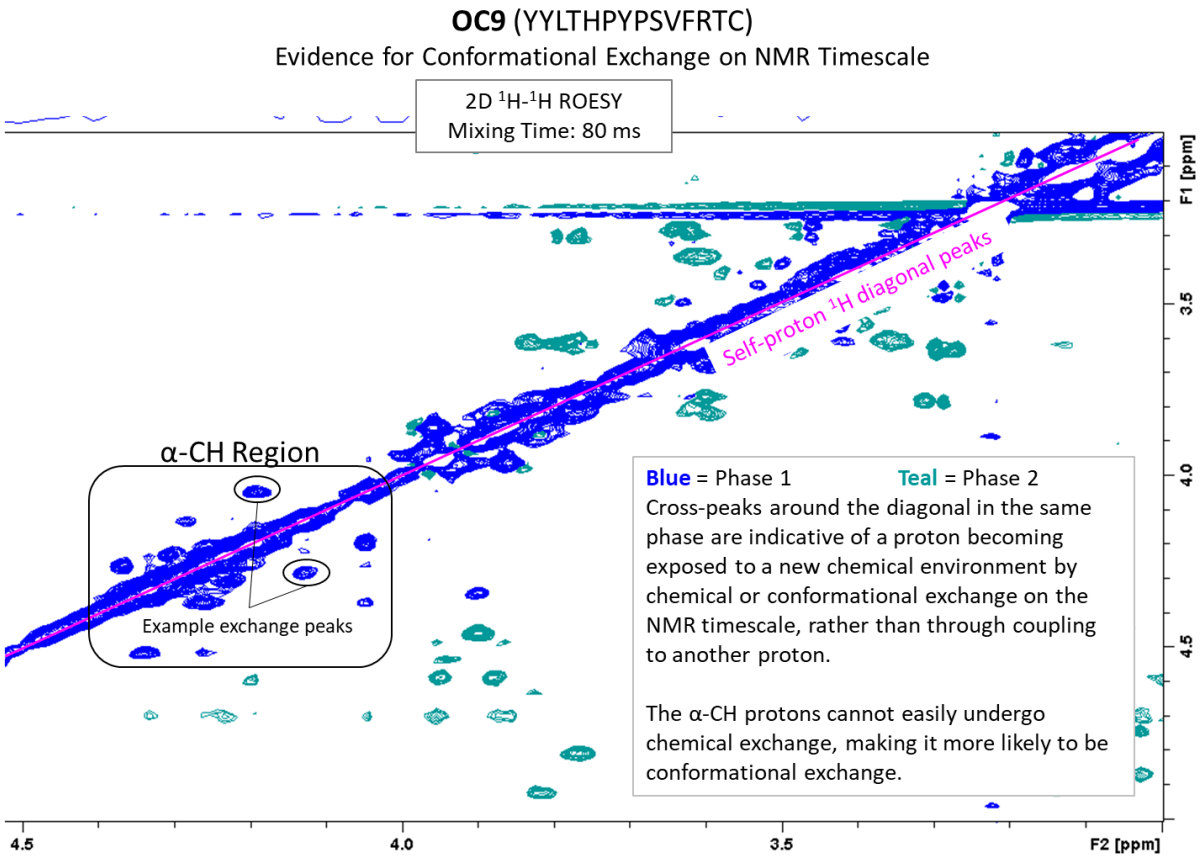
Amino Acid Representation in NMR
Predicted: Val = 1, Leu = 1
Observed: Val = 2, Leu = 2
Indicates 2x OC9_PxA conformations

In each OC9 proline variant, the number of conformations is halved compared to wild-type, supporting the proline amide *cis-/trans-* isomerism being responsible for the different conformations on NMR timescale.

Different chemical shifts and splitting patterns are seen between both, indicating a difference between chemical environments in available conformations.

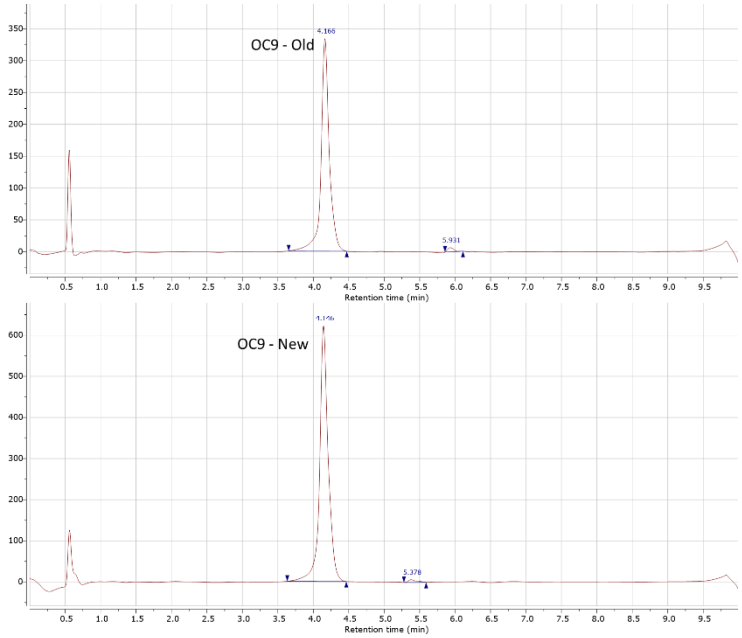
F.7.5.v. Comparison of ^{1}H-^{1}H 2D TOCSY NMR spectra (AV600) for OC9 P6A and P8A (100 µM each) with the same mixing times (80 ms). Examination of correlations associated to the leucine and valine terminal methyl groups to help count the number of conformations visible on the NMR timescale. Buffer conditions: 25 mM phosphate, pH 6.30, 100 mM NaCl, 10% D_2O at 308 K. Final DMSO ≤ 1 %.

One concern arising from knowledge of multiple conformations is that one or more of them may be inactive, or much more active, towards protein-interaction than others. If the conformers cannot interconvert easily this would effectively render it a solution of multiple compounds, as each conformer would possess different binding characteristics. However, cyclic peptides have been reported to demonstrate unexpectedly fast proline amide *cis*/*trans* isomerisation rates through suspected intramolecular catalysis [311]. Reassuringly, evidence was found in a 2D ^{1}H-^{1}H Rotating frame Overhauser Effect SpectroscopY (ROESY) NMR experiment for interconversion between conformers on the NMR timescale (**F.7.5.vi.**), considered to be in the 'intermediate' regime; sufficiently slow to be observed but sufficiently fast to readily occur [312,313].

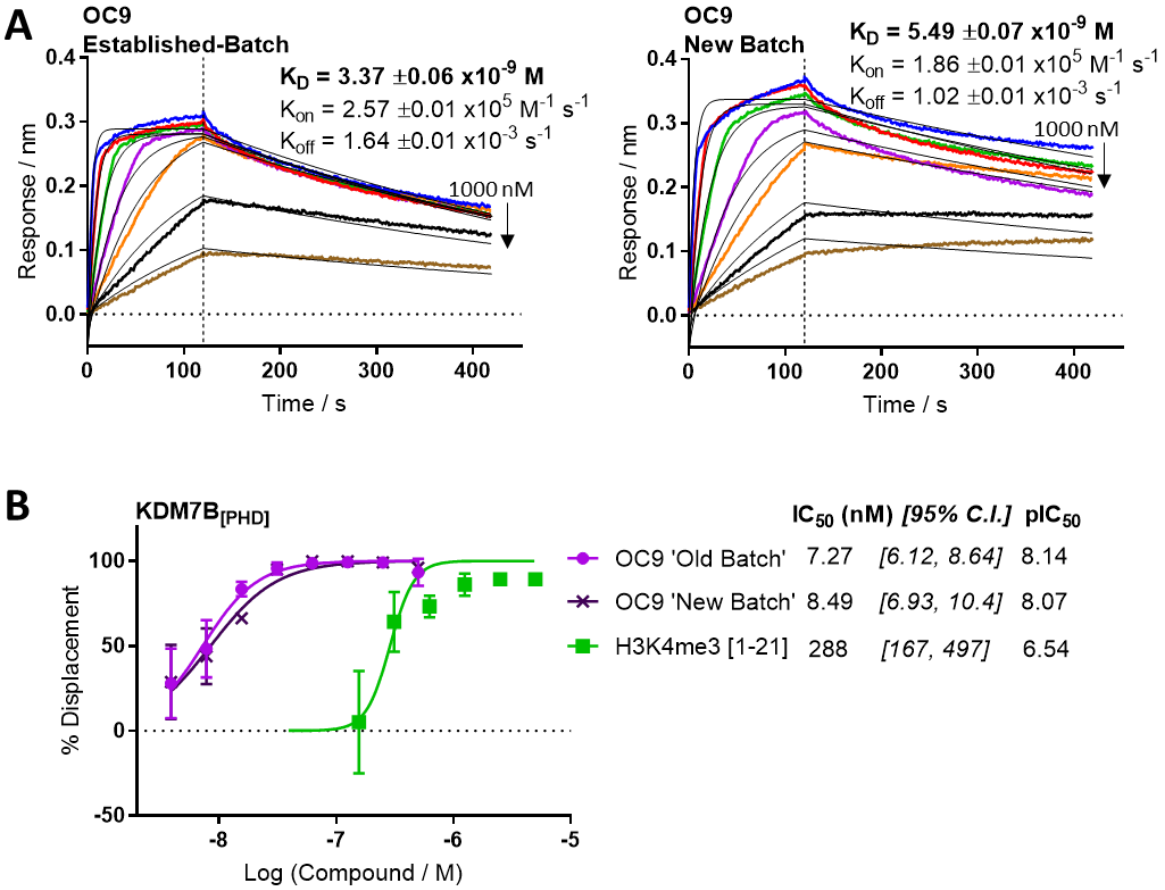


F.7.5.vi. A ^{1}H-^{1}H 2D ROESY NMR spectra (AV700) for OC9 (400 µM) with mixing time = 80 ms. Examination of correlations and their phase around the diagonal found same-phase exchange peaks around the α-CH region, providing evidence for conformational exchange on the NMR timescale. Buffer conditions: 25 mM phosphate, pH 6.30, 100 mM NaCl, 10% D_2O at 308 K. Final DMSO ≤ 5 %.

A new batch of OC9 was synthesised under 'gentler' conditions to the original synthesis, to see if a difference in potency could be measured. Cyclisation originally called for use of a water/acetonitrile mix at 60 °C, pH 8, for up to 30 minutes. The heating accelerated cyclisation but would also have provided more energy for isomer interconversion. An updated synthesis methodology used DMSO at 37 °C, with NEt_3 additive to pH 8, for 60 minutes. This gave excellent conversion to cyclic product, but reduced energy input for isomer interconversion. If the final ratio of isomers differed between methods, and interconversion was subsequently not possible, it would be reasonable to assume a difference in observed potency. This was not found, as both old and new methodologies provided one peak by LCMS (**F.7.5.vii.**) and equally potent OC9 (K_D = 3.37 and 5.49 nM for old and new batches respectively) (**F.7.5.viii.**). This further supports the assertion that, despite multiple isomers being observable on the timescale of the NMR experiment, they are readily able to interconvert for functional purposes. This is rationalised even if only one isomer is able to bind; with all four rapidly interconverting in equilibrium, as one is removed through protein binding the isomer equilibrium is continually re-established until all of the peptide is bound.



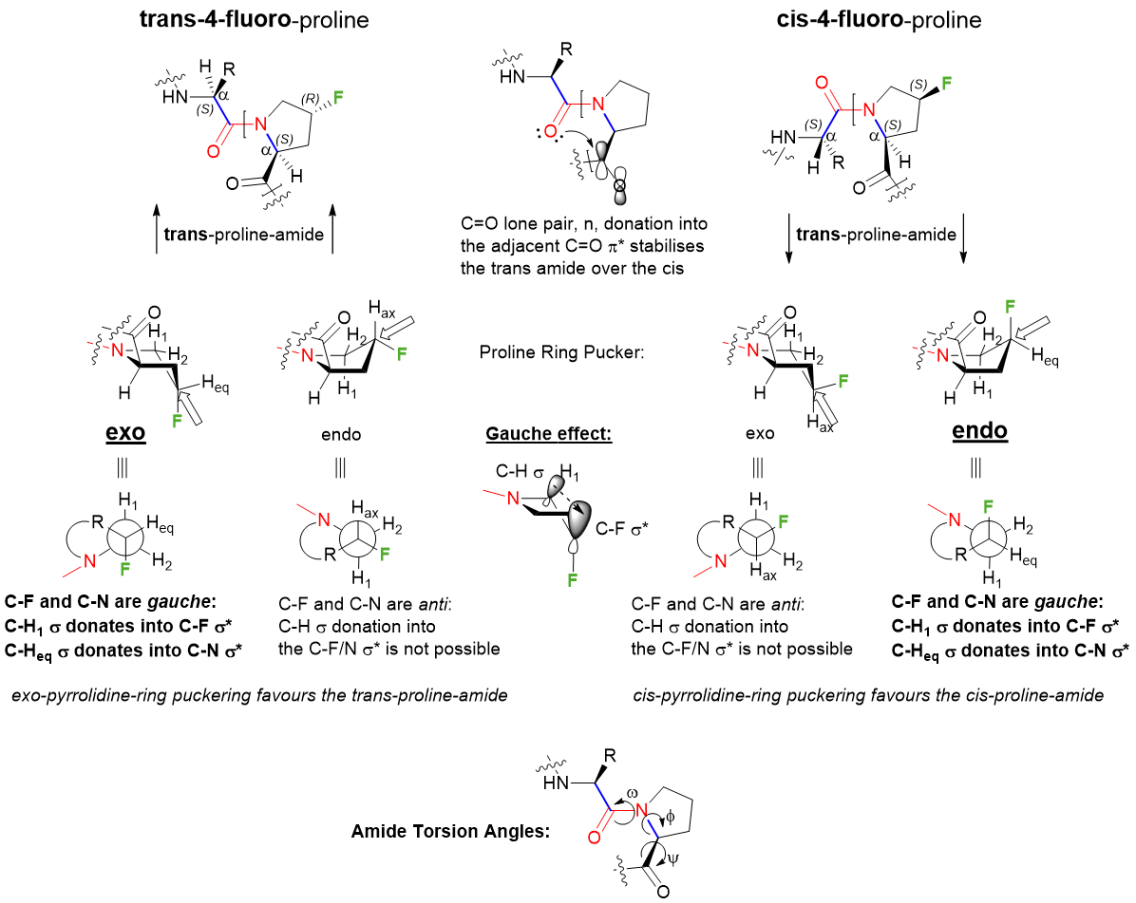
F.7.5.vii. Liquid Chromatography (LC) traces for OC9 batches. 'Old' represents the higher temperature synthesis compared to 'New'. One major peak is observed at the same retention time when monitoring at 210 nm in both cases.



F.7.5.viii. Different batches of OC9 were tested against KDM7B$_{[PHD]}$ for **A)** binding by BLI and **B)** displacement of biotinylated H3K4me3 [1-21] peptide by AlphaScreen. The established 'old' batch of OC9 was synthesised with use of higher temperatures than the 'new' batch. Both batches gave equivalent binding and displacement potencies, indicating little batch-to-batch variability and no consequence of the different synthesis temperatures on final populations of accessible conformations. Experiment conditions are given in Materials and Methods.

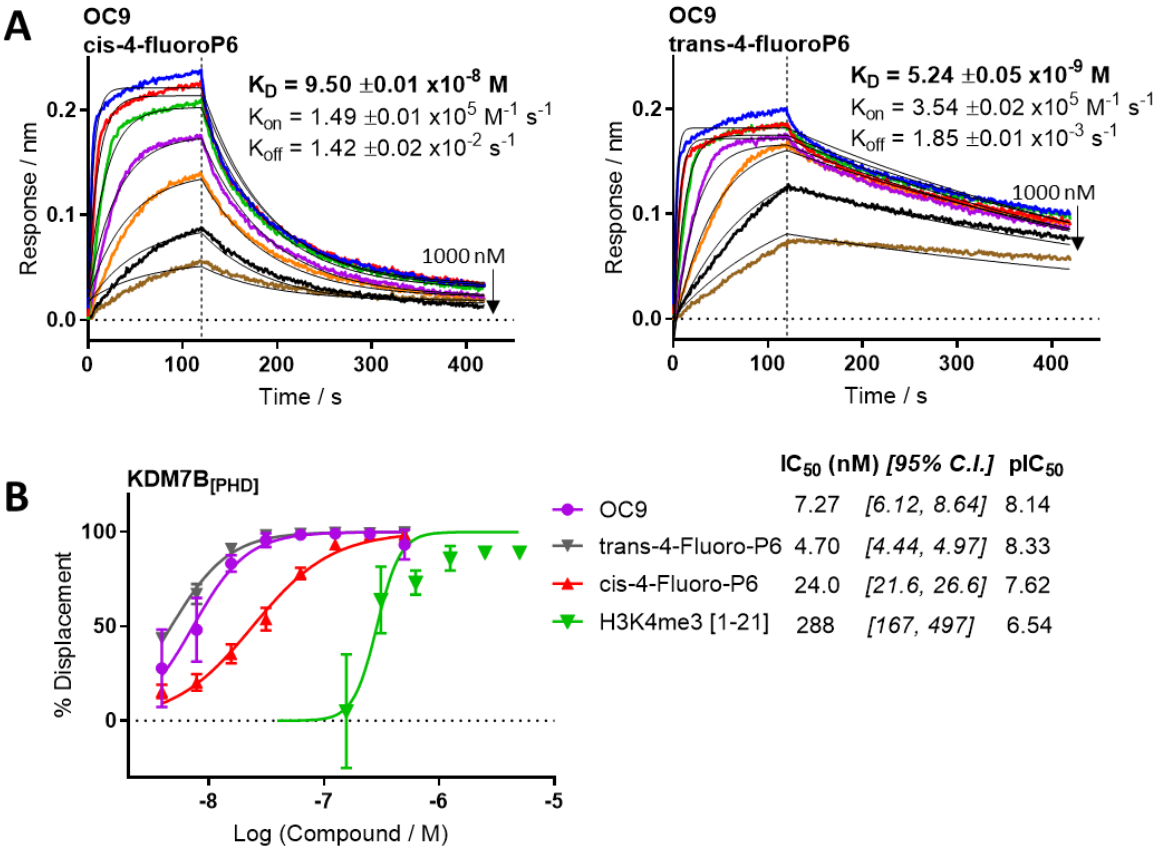
The subject of proline conformation exerting an effect on OC9 efficacy was further tested with modification of proline-6, previously determined to be the more important of the two prolines by alanine scanning. Commercially available *trans* or *cis*-4-fluoro-proline was substituted at position 6 in the synthesis of two new OC9 variants. Fluorination of the proline ring at carbon-4 influences both the proline amide *cis*/*trans* isomerism and accompanying puckering (*exo*/*endo*) of the pyrrolidine ring [314,315] (**F.7.5.ix.**). Placing the fluorine *cis*-4 (*S*) on the proline ring favours the *cis*-proline-amide, whilst placing the fluorine *trans*-4 (*R*) favours the *trans*-proline-amide. The lab of Ronald T. Raines has performed extensive computational and NMR investigation [316–318], concluding that a major contribution of fluorine at this position

is towards biasing the pyrrolidine ring puckering through gauche effects [319] (**F.7.5.ix.**). Raines et al. further explain this contributes to the *cis*/*trans* amide ratio through alteration of the amide torsion angles, with *exo* puckering pre-organisation inducing *trans*-amide and *endo* pre-organisation increasing levels of *cis*-amide. Stabilisation of the proline *trans*-amide is boosted by the overlap between the proline amide carboxyl lone pair (*i-1*) n donation into the proline (*i*) carboxyl π* (**F.7.5.ix.**), analogously to the Burgi-Dunitz 'angle of attack' invoked in nucleophilic attack [320]. In the *trans*-fluoro isomer this overlap is promoted, increasing population of the *trans*-amide. The *cis*-fluoro isomer destabilises such overlap in the *trans*-amide, which allows the population of the *cis*-amide to increase. Despite the complex and developing understanding of the underlying reasons behind such phenomena, it is undoubtedly useful in a practical sense to probe proline conformational effects.



F.7.5.ix. Fluorination of proline at position-4 of its side chain can lead to biasing of the proline amide conformation, dependent on whether the fluorine is *trans* or *cis*. The *trans* fluorination likely favours the *trans* amide, whereas the *cis* fluorination favours increasing the population of the *cis* amide, through stereoelectronic effects.

Binding assessment of both fluoroproline-isomers to KDM7B$_{[PHD]}$ by BLI (**F.7.5.x.**) revealed the *trans*-fluoro (increased *trans*-proline-amide) gave an equivalent K_D (5.24 nM) to OC9, whereas the *cis*-fluoro (increased *cis*-proline-amide) K_D (95.0 nM) was weakened by a factor of ~5. The most significant negative impact of the *cis*-fluoro isomer was hastening of the off rate by an order of magnitude – suggesting higher levels of the *cis*-proline-amide may be preventing OC9 from 'locking-in' as effectively once bound to KDM7B$_{[PHD]}$. This trend followed in AlphaScreen displacement assays (**F.7.5.x.**) as the *trans*-fluoro isomer was similarly potent (IC_{50} = 4.70 nM) to OC9 (IC_{50} = 7.27 nM), and the *cis*-fluoro approximately three-fold weaker (IC_{50} = 24.0 nM). These data point towards the *trans*-proline-amide at residue 6 being the most bioactive isomer. Considering the subtlety of the change in placing a single fluorine on a side-chain in a 1786 Da peptide, it emphasises the importance of this proline conformation and the need for the OC9 isomers to be able to readily interconvert. This also indicates the energy barrier to amide interconversion must be reasonably low, especially compared to the small molecule DMF where it is so high (88 kJ/mol [321]) that it is essentially un-rotatable at 298K.



F.7.5.x. OC9 variants, fluorinated *cis* or *trans* at position-4 on proline 6, were tested against KDM7B$_{[PHD]}$ for **A)** binding by BLI and **B)** displacement of biotinylated H3K4me3 [1-21] peptide by AlphaScreen. The *trans*-fluoro variant proved more effective than the *cis*, indicating that proline-6 is likely to favour a *trans*-amide conformation. Experiment conditions are given in Materials and Methods.

7.6. Conclusions and Future Work

Alanine scanning, targeted substitutions and solution NMR have been used to build a picture of the structure-activity relationship for OC9 interaction with the KDM7 PHD finger (**F.7.6.i.**). The OC9 T4 and H5 residues tolerate side chain stereocentre inversion and substantial functional group change – their hydroxyl and imidazole moieties appeared almost inconsequential to binding and displacement – allowing for future synthetic customisation at these positions. A crucial motif (SVFRT) has been identified, which may serve as the beginnings of a convertible pharmacophore against the KDM7 PHD finger. All three OC9 tyrosine residues possess side chains important for efficacy, but this does not depend on their hydroxyl functionality. OC9 exists in solution as an interconverting mixture of conformers on an NMR timescale, which the proline amide geometries are responsible for controlling. The proline-*trans*-amide at position 6 favours bioactivity, as does a *trans*-amide and L-stereochemistry at position 8. Assignment of the solution NMR structure of OC9 peptide in isolation would be useful to better understand how its 3D conformation is supported.

Computational modelling was used to inform OC9 sequence development towards desired physicochemical properties; however, the project came to an end before the modified OC9 peptides could be tested in physicochemical assays. Nonetheless, a second-generation cyclic peptide (OC9.2.0) has been developed incorporating several substitutions whilst maintaining potency and selectivity with lowered polar surface area, molecular weight and hydrogen bonding, which may be better suited to membrane environments. Further elaboration of the OC9 amino acid content might necessitate higher throughput peptide production (e.g. Syro [322] and SPOT synthesis [323]). However, it must be acknowledged that seeking alternative approaches to achieving cell-penetration with cyclic peptides may be a prudent way forward given the high-difficulty bar that has already been overcome to selectively and potently target

the KDM7s, and the new difficulty associated with re-optimising physicochemical properties through extensive chemical modification. Advancements have been made in recent years in understanding cell membrane transport of peptides [180,324] and the development of improved cell-penetrating motifs which can be appended to existing cyclic peptides [325–328], alongside the acceptance of the need for new approaches [329] to aid cell penetrance. Classically, cell-penetrating peptide motifs have involved use of poly-cationic TAT/poly-Arg sequences [330] which themselves exhibit often overlooked effects against a broad range of proteins [331,332], including inhibition of KDM demethylase activity (PolyR vs $KDM4A_{[JmjC]}$ IC_{50} = 40 nM [155]). It may then be wise to also consider linking across disciplines to incorporate new methods for generalised intracellular peptide delivery which are already being applied in the oligonucleotide field (e.g. nanoparticle [333] and liposomal [334]).

Progress was made in identifying compatible locations in OC9 for un-natural D-alanine installation, these modifications will be tested in future proteolysis assays to determine the best site for boosting degradation resistance, pertinent to most applications. Further stability testing of new cyclic peptide variants will be pursued as part of ongoing collaborations with industrial partners. In addition, the mRNA display system could be further adapted to include protease resistance as a selection step [335,336], or specialised stability-enhancing tags appended to the peptides [337]. Future work will also include testing of OC3 and OC6 alanine scan peptides, which have been synthesised (Shane Clark). The finding that the NGS data could be linked to the empirical findings for suitable substitution sites should help inform further modifications in future work [338]. The suitability of the existing OC9 sequence as a chemical probe in cell-based experiments was next investigated in Chapter 8.

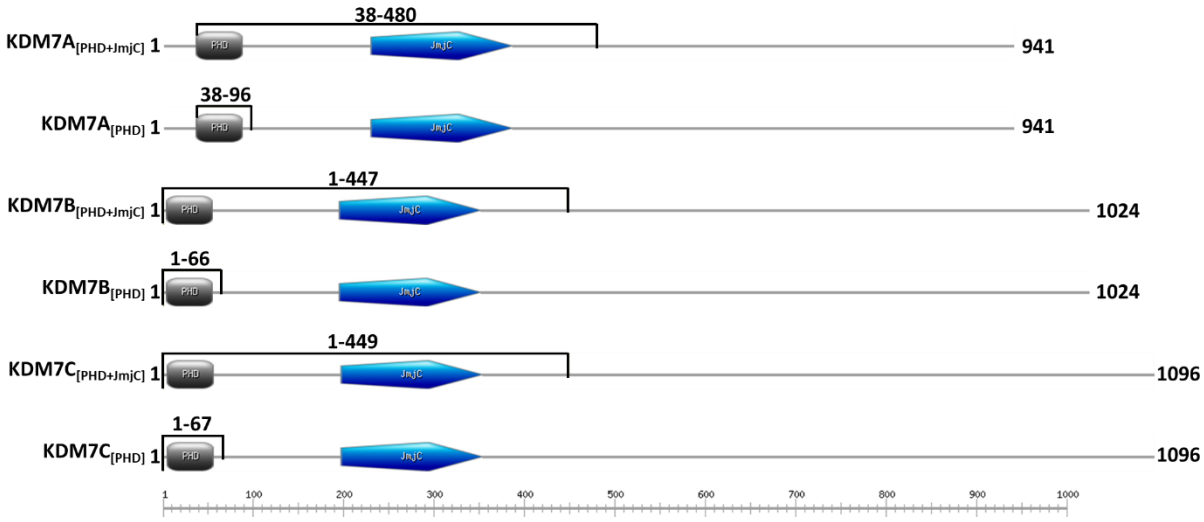
OC9	K_D (nM)	pK_D	Δ(WT-Var) pK_D	Displacement pIC_{50}	Δ(WT-Var) Dis. pIC_{50}
WT	3.37	8.47		8.14	
Linear	> 1000	< 6.00		< 6.30	
Scrambled	> 1000	< 6.00		< 6.30	
Y1F	23.1	7.64	0.83	7.89	0.25
Y2F	14.8	7.83	0.64	7.95	0.19
Y7F	12.1	7.92	0.55	8.15	-0.01
Y1F Y2F Y7F	41.1	7.39	1.08	7.61	0.53
T4A H5A P8A	5.09	8.29	0.18	8.03	0.11
T4-dAla	25.0	7.60	0.87	7.94	0.20
H5-dAla	11.7	7.93	0.54	8.10	0.04
P8-dAla	> 1000	< 6.00		6.22	1.92
OC9.2.0 (T4A H5A Y1F Y2F Y7F)	20	7.70	0.77	8.16	-0.02
WT-NewBatch	5.49	8.26	0.21	8.07	0.07
cis-4-fluoro-P6	95.0	7.02	1.45	7.62	0.52
trans-4-fluoro-P6	5.24	8.28	0.19	8.33	-0.19

F.7.6.i. Summary of the BLI (K_D) and AlphaScreen (IC_{50}) displacement results for second generation OC9 variant peptides.

Chapter 8. OC9 vs KDM7s in a Cellular Context

8.1. Introduction

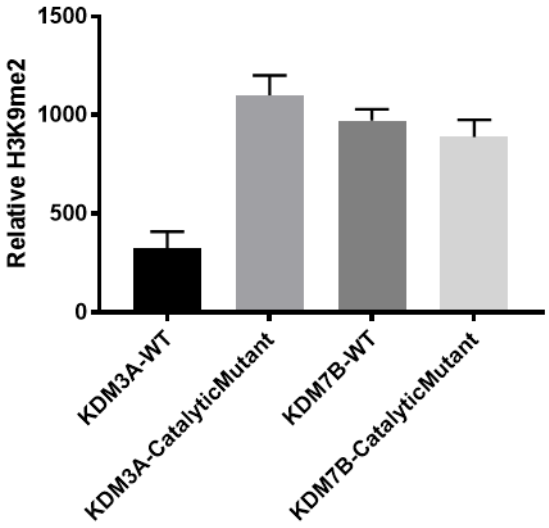
The experiments to this point had been performed using shortened versions of the KDM7 proteins. Recombinant production of the full-length KDM7 protein in bacteria is problematic and so constructs containing the domains of interest had been used instead (**F.8.1.i.**). This was found to be a successful approach in terms of enabling cyclic peptide development and testing, but to be truly useful a probe must be shown to engage and act on its endogenous full-length target in a cellular environment.



F.8.1.i. The full-length KDM7 domain maps are represented (generated using ExPASy Prosite MyDomains) with the sequence spanned by their respective recombinant constructs annotated in black.

Despite strong evidence for significant whole-organism phenotypes with KDM7 knockouts (Chapter 0), the development of reliable in-vitro cell-based phenotypic assays for KDM7s has not yet been achieved. For many KDMs, ectopic expression of the full-length KDMs can significantly alter the global nuclear histone methylation status in transfected cells, and these changes are routinely used in KDM inhibitor assessment in cells [153]. However, while overexpression of KDM7s has been reported to alter histone methylation status in cells, these reports have been somewhat controversial. Sometimes only weak demethylation of the established histone target H3K9me2 is observed with full-length KDM7B in cells [78], whilst

other labs report conflicting KDM7 target marks including H3K27me2, H3K36me2 and H4K20me1 that vary dependent on whether the experiment assesses extracted histones or full nucleosomes [66,67]. These cellular demethylation results have not been reproduceable in our laboratory (Dr John-Paul Bukowski, unpublished, **F.8.1.ii.**). Several factors may contribute towards these varied reports: the epigenetic cross-talk requirement for the KDM7s complicating their catalytic behaviour, the unknown roles of protein super-complexes that they may be involved with, and the intricate overlapping roles of all KDM/KMTs providing unexpected redundancies.



F.8.1.ii. Overexpression of KDM7B in HEK293T cells does not lead to obvious increased demethylation of H3K9me2, especially when compared to another H3K9me2 demethylase (KDM3A). (Unpublished results from work by Dr John-Paul Bukowski, Kawamura Lab).

Full-length KDM7 was assessed herein using HEK293T cells. The cells are derived from HEK293, a human embryonic kidney cell line, but stably express a mutant SV40 large T antigen protein capable of binding SV40 promoters and boosting transfected protein expression. Use of HEK293T cells is not without concern regarding relevance of results to 'real-world' scenarios, with literature reports highlighting reproducibility problems with high levels of genetic variation in some common laboratory cell lines [339]. Nonetheless, such cell lines

continue to be used in academic testing. It is hoped the fundamental nature of histone epigenetic regulation affords a nuclear environment that remains reasonably representative of the 'real-world' [340].

Placing full-length KDM7 in a cellular environment provides additional challenges for target engagement over and above cell-penetration and intracellular localisation, as the target protein is likely to be occluded in part or in whole by endogenous complexes. In some cases this has been a stumbling block for drug discovery and chemical probe development, with the complexity of the cell making it difficult to discern if results are due to the desired specific interaction or other confounding factors [341]. The attachment of either small molecule or protein labels to aid visualisation is valuable, but can alter the true behaviour of the probe or target protein and give misleading results, which must be borne in mind [342,343].

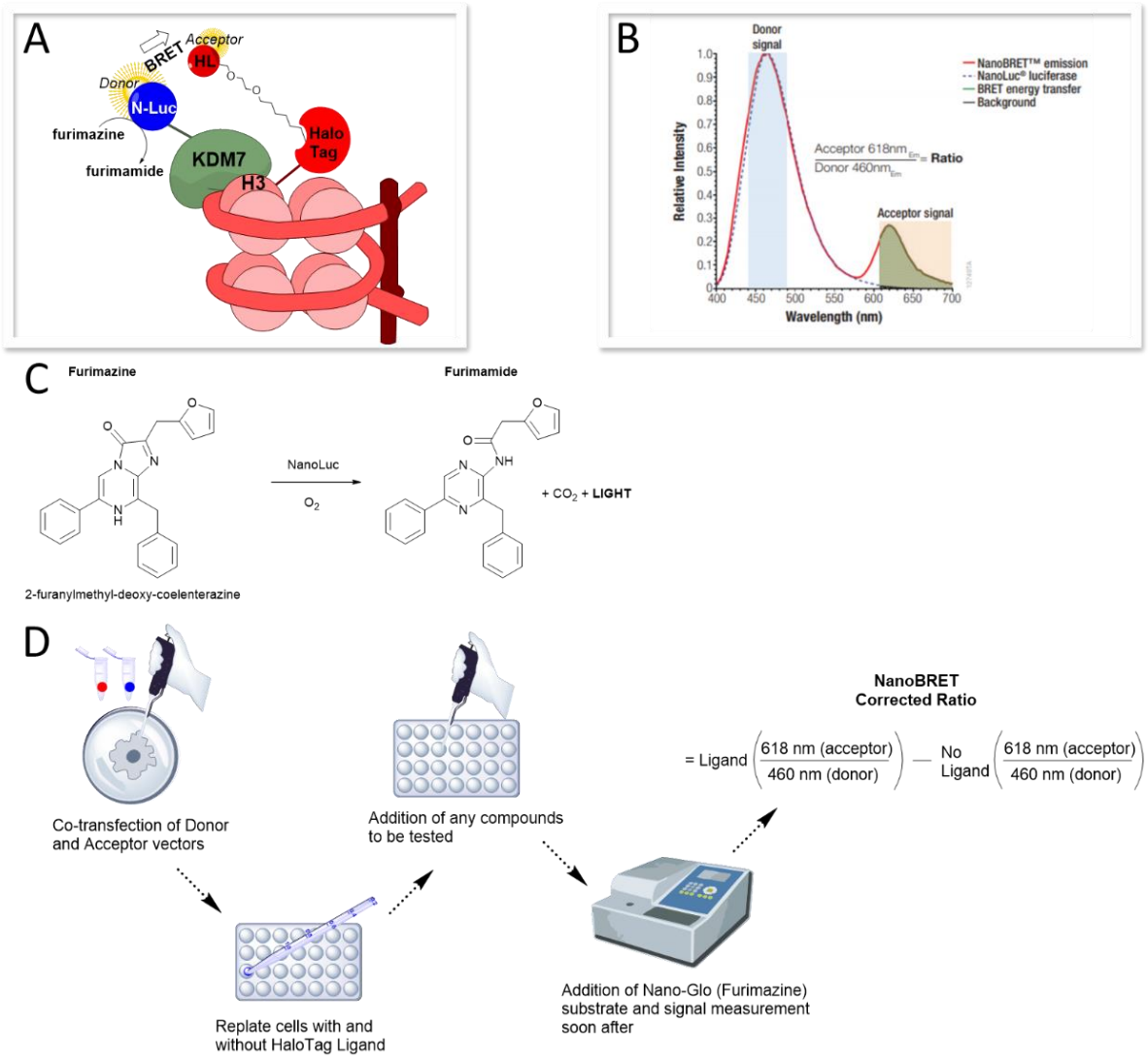
The overall objective of the work described in this Chapter was to investigate the utility of the cyclic peptide, OC9, by testing for engagement with the PHD finger of a full-length KDM7 protein in a cellular context. To assess this, assay development was aimed at first enabling investigation of the cellular protein-protein interactions of $KDM7_{[PHD]}$ to in-turn enable subsequent testing of inhibitors for this interaction. Two complementary cell-based approaches were used, a NanoBRET assay and an affinity purification experiment.

8.2. NanoBRET Cell Assay Preparation

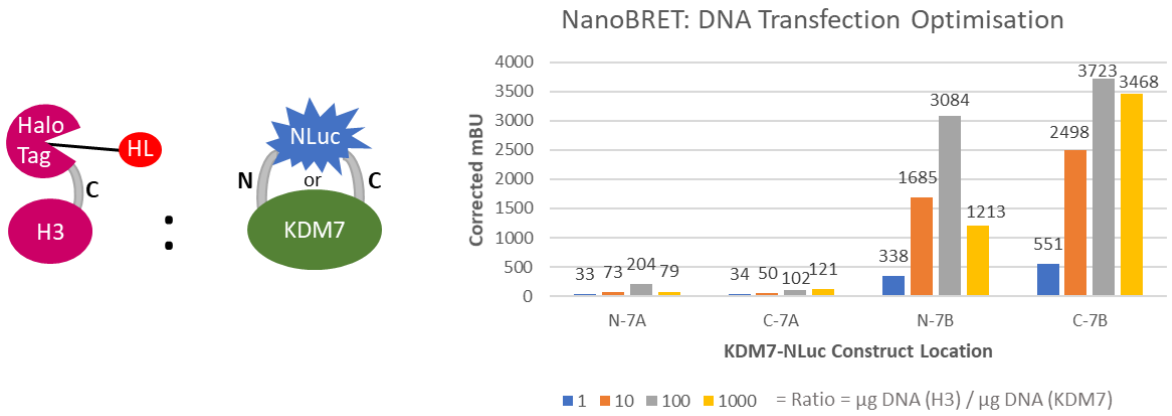
The NanoBRET assay (**F.8.2.i.**) is designed to measure real-time protein-protein interaction levels in live cells through bioluminescence resonance energy transfer (BRET) [344]. A HaloTag construct is appended to one of the proteins of interest, and a NanoLuc (NLuc) tag appended to the other. The HaloTag is a 33 kDa modified dehalogenase protein, developed to covalently attach to a halogenated ligand bearing a fluorescent moiety (HL). The NLuc tag is a 19 kDa nano-luciferase, derived from the deep-sea shrimp *Oplophorus gracilirostris*, that is ATP-independent and highly bioluminescent when paired with its novel furimazine substrate [345]. When the NLuc furimazine substrate is added to the system, the NLuc protein-tag luminesces at 460 nm (donor) and if the HL bearing partner (acceptor) is nearby, energy transfer occurs and fluorescence at 618 nm is observed. The ratio of this acceptor/donor fluorescence is calculated, corrected for the no-HL control, and the abundance of the protein-protein interaction in question can be analysed quantitatively [344].

The NanoBRET assay requires optimisation of the HaloTag/NLuc pairing and plasmid DNA levels for transfection. Many epigenetic NanoBRET assays have been configured with the histone H3.3 C-terminal HaloTag vector and the smaller NLuc fusion on the target epigenetic protein (BRD4/Histone H3.3 Interaction Assay, Promega). The C-terminal location of the HaloTag makes it less likely to interfere with reading of the crucial H3 N-terminal tail and H3.3 is often incorporated into open chromatin [346], making this configuration and H3 isoform the most suitable for study of epigenetic protein-protein interactions. Therefore, the same approach was adopted in these studies. N- and C-terminal NLuc fusion full-length KDM7A and KDM7B constructs were tested for interaction with C-terminal HaloTag H3.3 (**F.8.2.ii.**). Both constructs were transfected into HEK293T cells at differing DNA ratios and the NanoBRET assay was carried out. The C-terminal NLuc KDM7B construct at a DNA ratio of 100:1 (H3.3 :

KDM7B) had the highest signal compared to background NanoBRET giving the largest potential signal window. The NLuc KDM7A construct gave a lower NanoBRET ratio in both configurations and was not pursued further at the time.



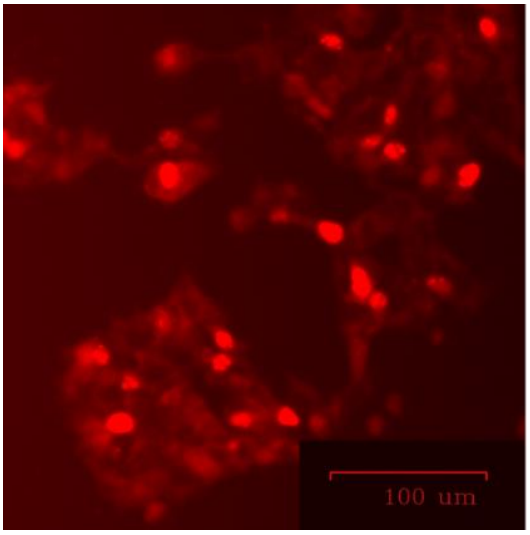
F.8.2.i. A: The principle behind detection of protein-protein interaction in a NanoBRET assay. **B**: Representative luminescence signal trace and calculation of acceptor/donor ratio signals (adapted from Promega TM349). **C**: The substrate used for generation of bioluminescence by NanoLuc. **D**: The approximate workflow used for running for the NanoBRET assay and calculation used to determine the BRET Units (BU). Start to finish is < 48 hours.



F.8.2.ii. NanoBRET signal optimisation. The KDM7A/B constructs were trialled with the NLuc tag either at the N- or C-terminus against the Histone H3 construct with a C-terminal HaloTag. The ratio of the amount of DNA used for transfection (~ 2 µg total DNA per well of 8 x10^5 cells) was varied simultaneously. Y axis is in milli-BRET Units (mBU).

8.3. Nuclear Localisation of N-terminal HaloTag-KDM7B

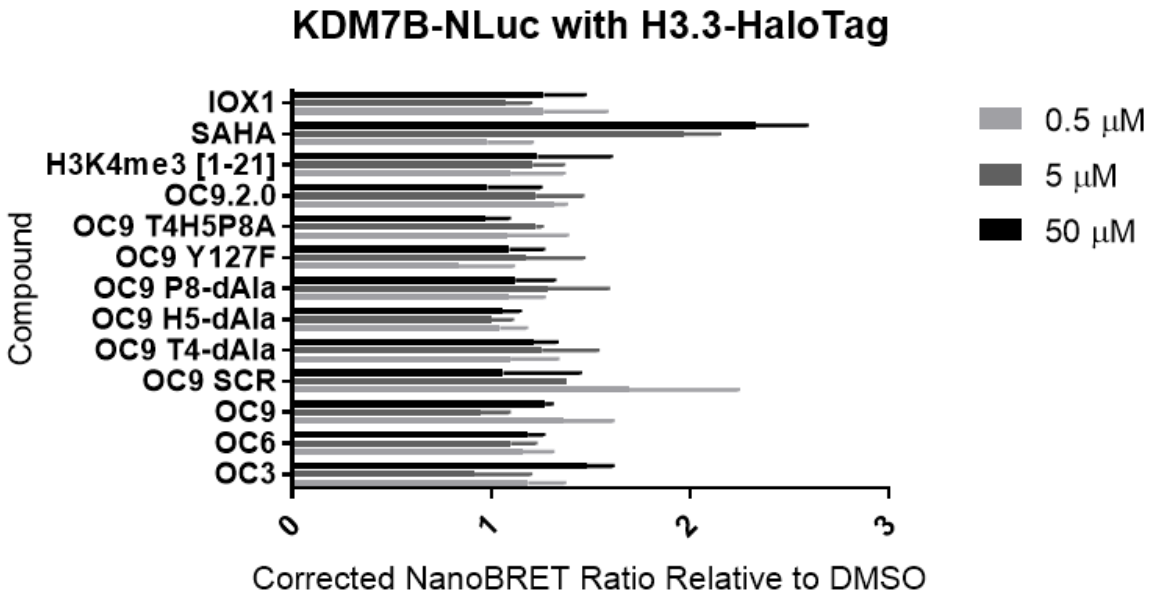
In order to confirm that the full-length KDM7B was overexpressed in the cell nucleus, an N-terminal HaloTag KDM7B construct was used to transfect HEK293T cells, and the fluorescent ligand HL was used to image and confirm nuclear localisation of KDM7B in live cells (**F.8.3.i.**).



F.8.2.iii. HEK293T cells were transfected with full-length HaloTag-KDM7B.v.1. and imaged after dosing with HaloTag Ligand to show the nuclear localisation of full-length KDM7B expression.

8.4. KDM7B NanoBRET Assay

The C-terminal NLuc KDM7B and C-terminal HaloTag H3.3 overexpressing cells were dosed with compounds for 16 hours at 50, 5, 0.5 µM (0.5 % DMSO). Relative NanoBRET ratios for cyclic peptide-containing wells remained within error of the DMSO-only control (**F.8.4.i.**), indicating little evidence for disruption of the KDM7B-H3.3 interaction. However, the small molecule histone deacetylase (HDAC) inhibitor (SAHA aka Vorinostat [347]) was tested simultaneously and showed a dose-dependent increase (up to two-fold) in the NanoBRET ratio compared to DMSO. Higher levels of histone acetylation are believed to increase euchromatin levels [35], potentially opening access to H3 and thereby increasing total interactions with KDM7B. The SAHA-induced NanoBRET ratio increase supported the idea that the measured BRET signal was due to genuine interaction between KDM7B and histone H3.3.

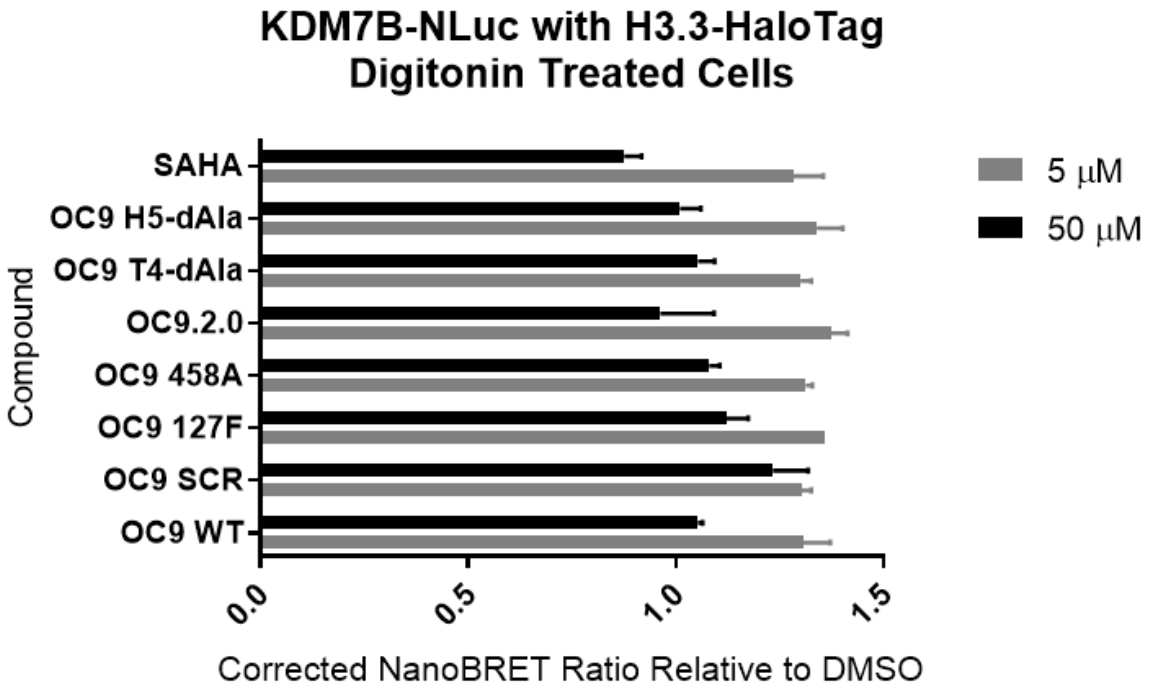


F.8.4.i. HEK293T cells were transfected with KDM7B-NLuc and Histone H3.3-HaloTag, then dosed with compounds (0.5, 5, 50 µM, 0.5% DMSO final) for 16 hours prior to addition of NanoGlo substrate and NanoBRET reading. Corrected NanoBRET ratios were normalised with respect to the DMSO-only compound control. Only SAHA showed an effect, indicating increase protein-protein interaction due to inhibition of HDACs.

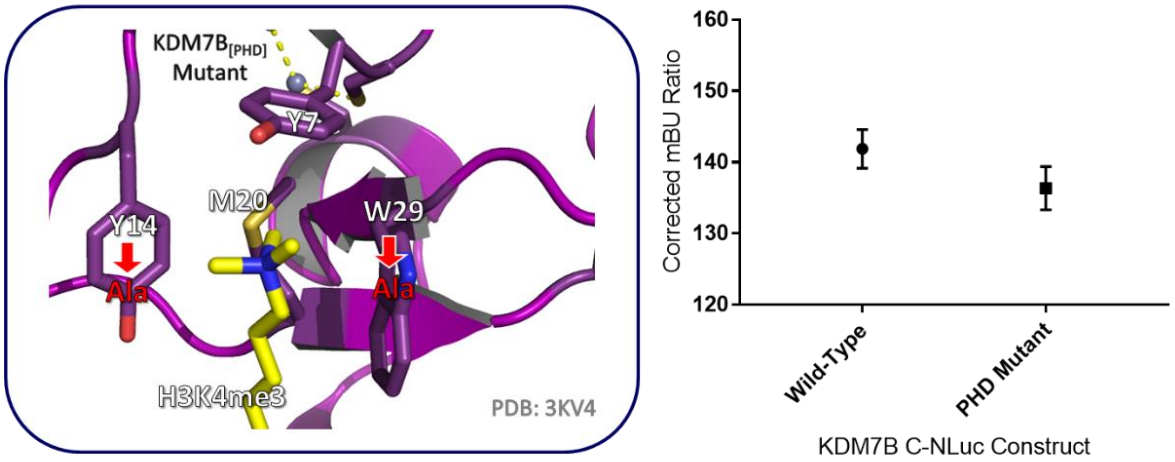
It was possible the lack of cyclic peptide efficacy was due to poor cell permeability. To attempt to overcome this, a modified approach was trialled with dilution of the compounds into a solution containing digitonin (a steroidal saponin capable of solubilising lipids and membranes). Digitonin has been used previously in NanoBRET assays to aid compound cell penetration [348]; however, it can be toxic to cells with extended incubation times. Since the assay objective was measurement of a protein-protein interaction, a shorter incubation time could be used to try and minimise risk of unwanted deleterious effects. The cells were incubated with compounds (5, 50 μM) in digitonin-containing media for 2 hours, but this also failed to demonstrate efficacy of the cyclic peptides to disrupt the targeted protein-protein interaction, with little change in the NanoBRET ratio compared to DMSO-control (**F.8.4.ii.**). SAHA did not appear to dose-dependently increase the NanoBRET ratio under these conditions, potentially due to the shorter incubation precluding a detectable increase in histone acetylation.

As a cellular control, a KDM7B$_{[PHD]}$ mutant was created, with conversion of the essential tyrosine-14 and tryptophan-29 within the aromatic cage to alanine to weaken binding to the histone H3K4me3 mark [62], serving as a positive-control for PHD finger inhibition. Interestingly, transfection of the KDM7B$_{[PHD]}$ mutant in parallel with wild-type showed very similar NanoBRET ratios (**F.8.4.iii.**). The no-HL negative control for both, (i.e. no BRET acceptor present), provided little-to-no signal. Together, this suggested the KDM7B$_{[PHD]}$ mutant was still in close enough proximity to the HaloTag H3.3 for BRET to occur, despite mutation of the aromatic cage. Reasons for this are as yet inconclusive, with possibilities speculated to be the C-terminal region of KDM7B providing some unknown association to the H3.3 protein, or the overexpression of both protein partners causing high local concentrations leading to unnatural association mechanisms. The failure of the KDM7B$_{[PHD]}$ mutant to show a

significantly decreased NanoBRET ratio made further testing of cyclic peptides at this stage illogical, as further assay development was needed to ensure PHD finger inhibition would be detectable.



F.8.4.ii. HEK293T cells were transfected with KDM7B-NLuc and histone H3.3-HaloTag, then dosed with compounds (5, 50 μM, and digitonin 0.5% DMSO final) for 2 hours prior to addition of NanoGlo substrate and NanoBRET reading. Corrected NanoBRET ratios were normalised with respect to the DMSO-only compound control. No compounds showed a substantial effect on the protein-protein interaction.

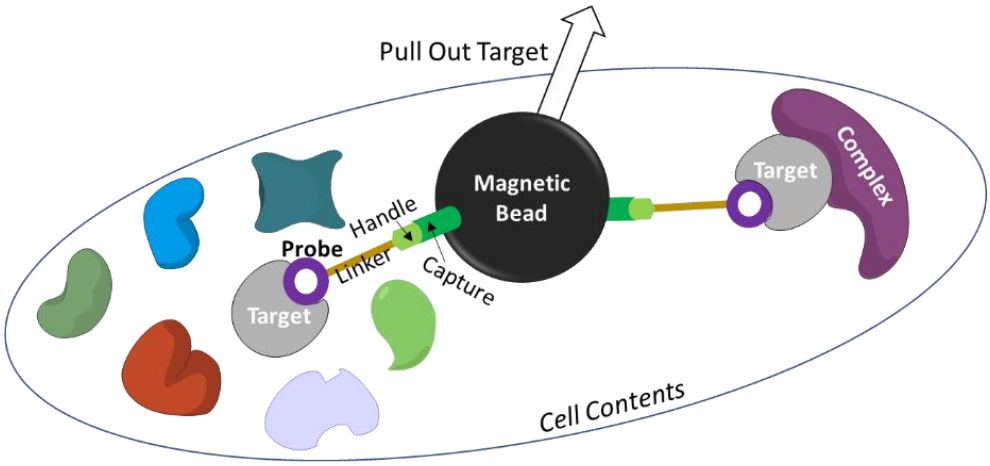


F.8.2.v. HEK293T cells were co-transfected with Histone H3.3-HaloTag and a full-length KDM7B-NLuc construct, either wild-type or PHD finger mutant, and assessed by NanoBRET for interaction. The PHD finger mutant did not show a decrease in mBU ratio compared to wild-type, indicating the mutations were insufficient to completely abrogate the target protein-protein interaction under the tested conditions. The experiment was run in technical sextuplicate.

Significant optimisation will likely be required to determine the KDM7A and KDM7B constructs and assay conditions necessary to achieve reliable testing of cyclic peptides with the NanoBRET system. Future work could involve developing NanoBRET assays using shortened KDM7 constructs, similar to those used successfully for in-vitro assays. Once optimised, the assay has the potential to provide a robust and high-throughput methodology for analysing protein-protein interactions in cells [349,350]. Importantly, the ability to express full-length KDM7 protein in the cell nucleus remained useful by providing the requisite target for an affinity purification (pull-down) assay.

8.5. Affinity Pull-Down of KDM7B from Nuclear Lysate

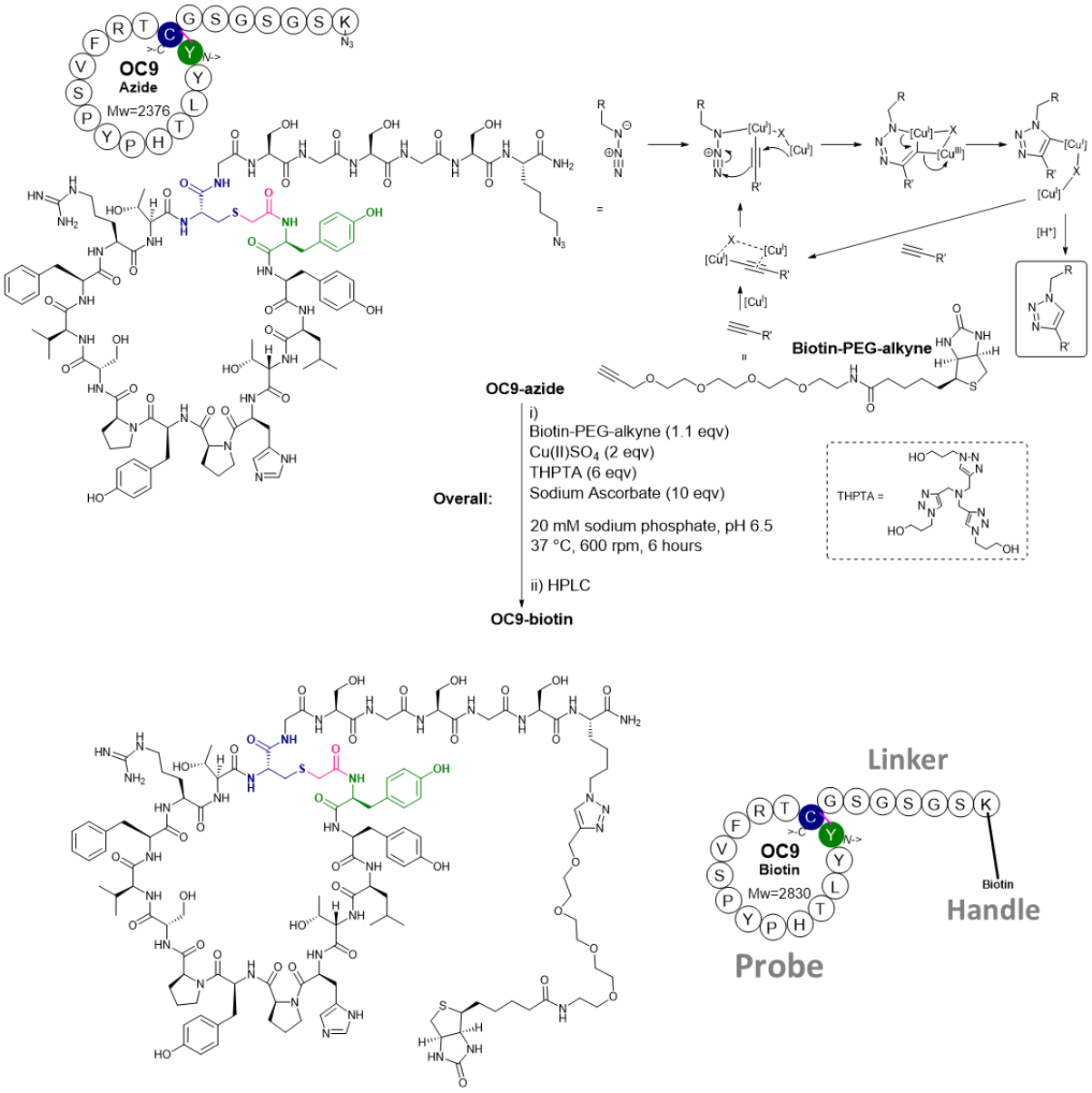
Affinity reagents are useful tools to extract (pull-down) target proteins and complexes from their native context to investigate protein-protein interactions and build a picture of their biological roles (**F.8.5.i.**). Pull-down can be coupled to other analytical techniques to identify the intended target (e.g. by Western Blot) and other molecules that the target is bound to (e.g. mass spectrometry for proteomics, or sequencing for DNA/RNA). Antibodies are commonly employed as affinity probes in (co)immunoprecipitation experiments, but small molecules can also be adapted to suit this role. The probe must be modified to contain a handle minimally detrimental to binding, with the handle used to recover the probe and in-turn anything it is bound to. Typically, the very high affinity biotin-streptavidin interaction is exploited for probe-handle recovery. For the probe itself, a tight (nanomolar or better) affinity interaction with a slow off-rate between the probe and target can help with efficient capture, and survival through the stringent washing steps that form part of most pull-down protocols – requirements that OC9 fulfils (Chapter 2, residence time). Thus, OC9 was converted into an affinity probe with additional functionality to enable KDM7 capture and recovery.



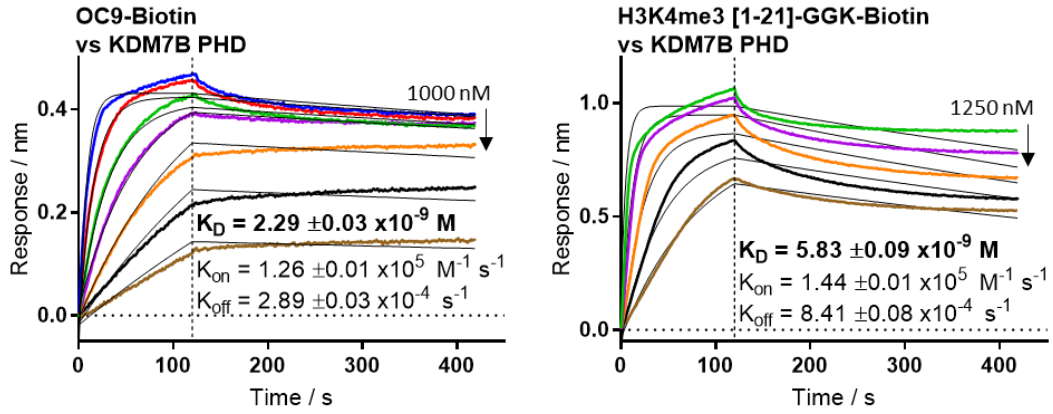
F.8.5.i. The principle of an affinity pull-down of a target from its native context is depicted. The probe can be added to the cell contents before or after membrane lysis. Rigorous wash steps are performed after recovery of the magnetic beads to minimise non-specific recovery of unrelated proteins.

One advantage of deriving the cyclic peptide hits from the mRNA display selection is the built-in placement of a linker location; all successfully selected peptides must be able to bind their target whilst possessing a C-terminal GSGSGS spacer sequence connected to their puromycin-PEG-RNA. Thus, the GSGSGS linker was reinstated to the OC9 'warhead' and azidolysine was added at the C-terminus via solid-phase-synthesis, with cyclisation and purification as normal. The azide provided orthogonal reactivity for 'click' conjugation of any desired handle. A biotin-PEG(4)-alkyne handle was purchased (Iris Biotech) and reacted under copper catalysis for efficient conversion using methodology adapted from the Spring Lab [351] (**F.8.5.ii.**).

HPLC purification of the reaction mixture yielded the desired derivative, OC9-biotin. This approach afforded a high purity affinity probe that retained potent binding of $KDM7_{[PHD]}$ (K_D = 2 nM by BLI, **F.8.5.iii.**) possessing a biotin handle for easy capture by magnetic streptavidin beads and a spacer between 'warhead' and 'handle' to minimise functional interference. The much larger size (2830 Da) of OC9-biotin made membrane penetration unlikely. To circumvent this, the affinity probe was added after cell membrane lysis.



F.8.5.ii. Solid-phase peptide synthesis was used to make a derivative of OC9 containing the original C-terminal GSGSGS linker from mRNA display and an added azido-lysine, which permitted facile conversion into probes containing additional functionality by click chemistry. Addition of a biotin was used to create the affinity probe OC9-biotin. The biotin moiety provides a handle for capture from solution, enabling use in pull-down experiments.



F.8.5.iii. OC9-biotin and H3K4me3 [1-21-GGK-biotin] were tested for binding to KDM7B$_{[PHD]}$ by BLI. The protein was immobilised by N-terminal His-tag on Ni-NTA sensors. Each coloured trace represents the response for a peptide concentration (2-fold dilution series). The fitted data, overlaid in black, from the 'Association then Dissociation' model (prism v7) were used to determine the binding affinity, K_D, and K_{on} and K_{off}. The start of dissociation is indicated with a dashed vertical line.

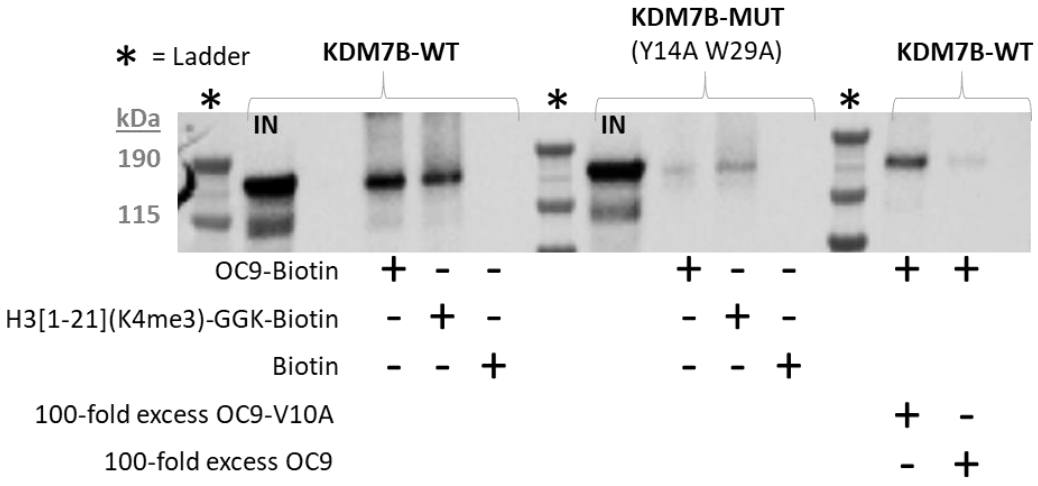
A pull-down of full-length KDM7B by OC9 was performed. In brief (see Materials and Methods for full details), as follows: HEK293T cells were transiently transfected with a C-terminally NLuc tagged full-length KDM7B construct, either wild-type or PHD finger mutant (Y14A, W29A), and cultured overnight. Frozen cell pellets were lysed using a commercially available kit (Nuclear Extraction Kit ab219177, Abcam) which made it possible to isolate proteins belonging to three different lysate fractions (cytosolic, soluble nuclear and insoluble nuclear) and retain intact protein-complexes and active enzymes within the soluble nuclear extract [352]. This was important as the KDM7s localise to the nucleus and thus would aid their remaining in as close to native context as possible post-lysis, providing the necessary level of additional challenge to the binding of OC9. A bicinchoninic acid (BCA) assay was used for quantification of total-protein in the lysate. Both the wild-type and the mutant soluble nuclear lysates were diluted to an equal final concentration prior to pull-down. OC9-biotin was added and incubated with each lysate before extraction by addition of streptavidin coated magnetic beads. Three stringent bead wash/vessel transfer steps were used to minimise non-specific recovery and heat denaturation of the beads was used for elution of recovered proteins. An

H3K4me3 [1-21-GGK-biotin] peptide was used for a positive control pull-down and a biotin small molecule used for a negative control pull-down.

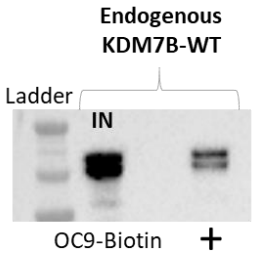
Western Blot analysis using an anti-KDM7B (AAs 800-900) antibody confirmed the OC9 and H3K4me3 affinity probes successfully pulled-down overexpressed wild-type full-length KDM7B from soluble nuclear lysate (**F.8.5.iv.**). This is evidence that OC9 is able to bind to a full-length KDM7 even in the presence of a plethora of unknown nuclear protein partners. The mutation of the $KDM7B_{[PHD]}$ aromatic cage reduced the pull-down yield for the H3K4me3 affinity probe, as expected due to its binding being known to be dependent on the aromatic cage. The presence of a weaker but still notable band for H3K4me3 based pull-down of the PHD finger mutant KDM7B could be due to H3K4me3 binding to proteins that also bind KDM7B. OC9 pull-down of the KDM7B PHD finger mutant was more drastically reduced, which clearly suggests the integrity of the aromatic cage is also necessary for OC9 binding; either for direct interaction or general PHD finger structural stability. It was unknown prior to the experiment whether this would be the case. A band just below the target band is visible in all successful pull-downs – it is not yet known whether this represents a degradation product, splice variant or off-target antibody reactivity.

To further support the PHD finger dependence for OC9-biotin pull-down of KDM7B wild-type, a competition experiment was performed with a 100-fold excess of non-biotinylated OC9 or OC9_V10A. The alanine variant OC9_V10A was previously (Chapter 7) demonstrated to be a much weaker binder than OC9 (> 3 orders of magnitude) and to be much less effective at displacement of $KDM7B_{[PHD]}$ than OC9. The competition pull-down (**F.8.5.iv.**) revealed that OC9-biotin was almost entirely outcompeted by a 100-fold excess of OC9, but pull-down by OC9-biotin was still observed for the competition with OC9_V10A. These observations

indicate that high-occupancy of the PHD finger of KDM7B by OC9 prevents OC9-biotin from binding the target, whereas the reduced PHD finger competitiveness of OC9_V10A enabled OC9-biotin to remain functionally effective as a pull-down probe, thus supporting the PHD finger dependence of the pull-down. This result also served to further support the criticality of the valine-10 residue for binding of OC9. OC9-Biotin was additionally able to pull-down endogenously expressed KDM7B from HEK293T nuclear lysate (**F.8.5.v.**).



F.8.5.iv. Western Blot analysis using an anti-KDM7B antibody (Ab36068) of an affinity probe (OC9-biotin, H3K4me3[1-21-GGK-biotin], biotin) pull-down of overexpressed full-length KDM7B.v.1. wild-type or PHD finger mutant (both with C-terminal NLuc tag) from nuclear lysate of transiently transfected HEK293T cells. A competition experiment was also performed with un-biotinylated OC9 or un-biotinylated OC9_V10A (non-functional variant).



F.8.5.v. Western Blot analysis using an anti-KDM7B antibody (Ab36068) of OC9-biotin pull-down of endogenous KDM7B from nuclear lysate of HEK293T cells.

Together these results endorse the ability of the OC9-biotin affinity probe to engage with full-length KDM7B in a cellular context, and that the engagement is through specific interaction with the PHD finger of KDM7B. However, whilst the described analysis by Western Blot is useful in demonstrating this, its reliance on use of a single antibody for detection means only the chosen target can be revealed. H3K4me3 will certainly interact with and likely pull-down other epigenetic proteins [116], but it is as yet unknown whether OC9 would also interact with unknown partners in a native context. The selectivity profiling (Chapter 4) suggests that OC9 should be highly selective for recovery of KDM7s and their associated proteins, but further work would seek to address this through methods such as total-protein staining of SDS-PAGE gel samples, to optimise the procedure for subsequent mass spectrometry-based proteomic analysis of the OC9-biotin pull-down.

8.6. Conclusions and Future Work

Initial efforts were unsuccessful in establishing a live cell-based NanoBRET assay for real-time analysis of the PHD finger dependent KDM7 protein-protein interaction with histone H3.3. Therefore, compound efficacy in disrupting this interaction in cells cannot yet be determined, nor can definitive assessment of cell permeability of OC9 or other cyclic peptides yet be made. Re-optimisation of the NanoBRET assay will be necessary, possibly with new KDM7 constructs. The use of a system with a large milliBRET ratio may not be optimal, and lower BRET ratios should be tested with protein-mutant controls to see if they offer improved sensitivity. It may be necessary to develop alternative whole-cell based assays with appropriate target-engagement readouts. For example, cellular thermal shift assays (CETSA) may be able to offer indirect assessment of target engagement by thermal stabilisation and thus also cell permeability. Implementation of other cell-biology techniques such as siRNA and CRISPR should also be considered for creation of cell lines with improved assay-appropriate KDM7 dependent phenotypes (e.g. transcriptomic, proteomic, epigenetic mark levels).

An azide derivative of OC9 was readily synthesised to provide orthogonal functionality for click chemistry derivatisation of OC9 into multi-functional probes. This was used to create the affinity probe OC9-biotin, without impacting target-binding potency, which was demonstrated by Western Blot analysis to serve as an effective pull-down probe for engagement and capture of overexpressed full-length KDM7B protein from soluble nuclear lysate extract. Capture was evidenced to be due to specific interaction with the PHD finger of KDM7B, as pull-down efficiency was greatly reduced by either mutation of the aromatic cage within the PHD finger, or competition with unlabelled OC9 for wild-type KDM7B, but pull-down was still effective when in competition for wild-type KDM7B with the weakly binding variant OC9_V10A.

Further experiments are necessary to explore the ability of OC9-biotin to pull-down endogenously expressed KDM7s from different cell-lines. Its proven pull-down efficacy indicates that it should nonetheless be useful for proteomics studies to help profile the nuclear interactome of KDM7s, and thus give clues to unexplored biological roles. This could be combined with traditional antibody-based pull-down approaches (e.g. IP, ChIP, ChemSeq), with simultaneous use of unlabelled OC9 to block the PHD finger of the KDM7s, to gain insight into PHD finger dependence of the KDM7 interactomes. This represents a unique use-case prospect for OC9 and is an approach for PHD finger investigation that is as yet unexplored in the field of epigenetics.

Chapter 9. Project Summary

Insight into the landscape of epigenetics and histone protein regulation has evolved rapidly in recent years to reveal a complex network of interlinking post-translational marks (PTMs) [225]. Multiple classes of proteins are involved in the dynamic regulation of these changes. Many of the epigenetic regulatory proteins possess reader domains to be able to interact in a targeted way with this shifting-PTM milieu [104]. Lysine methylation stands out as being a prevalent modification, principally on the histone H3 tail, and is associated to biological consequences [353]. Hence, it is important to properly understand how methyllysine reader domains function [354]. However, there is a dearth of chemical tools available for studying important protein-protein interactions in epigenetics in general, and especially methyllysine reader domains [148,355]. Chemical tools are needed to assist with in-vitro and in-vivo studies, concerning both fundamental biological processes and any validation of potential medicinal chemistry targets [132–134].

Many PHD fingers, a type of zinc-finger protein, are known to act as methyllysine reader domains [107,356]. No potent or selective PHD finger probes have been reported. PHD fingers are present in hundreds of human proteins [107], including several of the lysine histone demethylases (KDMs), and they are often crucial for proper protein function. The KDM7 sub-family are heavily reliant on their PHD finger recognition of H3K4me3 for guidance of their JmjC domain demethylase function [65]. The KDM7s have been connected to neurological, oncological and immunological processes (see Chapter 0 for references), with others undoubtedly yet to be discovered, but no KDM7-selective probes are reported either. The work described in this thesis has explored the application of cyclic peptide mRNA display to

discover novel ligands against the KDM7A,B,C sub-family, and investigated selective and potent targeting of the PHD finger domain – previously considered an 'intractable' target [357].

Three cyclic peptides (OC3, OC6, OC9) were generated by mRNA display against $KDM7B_{[PHD+JmjC]}$. Characterisation by BLI, ITC and HDxMS found OC9 to have low nanomolar binding affinity (K_D) for $KDM7A,B,C_{[PHD]}$ and $KDM7A,B,C_{[PHD+JmjC]}$, but no apparent interaction with $KDM7A,B,C_{[JmjC]}$. OC9 gave low nanomolar potency IC_{50} (AlphaScreen) displacement of the cognate $KDM7_{[PHD]}$ ligand (H3K4me3 [1-21] peptide) from $KDM7A,B,C_{[PHD]}$ and $KDM7A,B,C_{[PHD+JmjC]}$. OC9 interaction with $KDM7_{[PHD]}$ was further studied in solution by protein NMR and intrinsic tryptophan fluorescence assay, which suggested disruption of a key tryptophan residue within the PHD finger aromatic cage as a plausible reason for OC9 being able to prevent PHD finger binding to H3K4me3. Characterisation by a mixture of BLI, ITC and HDxMS found OC3 and OC6 to bind across the natively joined domains of $KDM7B_{[PHD+JmjC]}$ to give nanomolar nanomolar IC_{50} displacement of H3K4me3 [1-21] peptide.

The PHD finger regulates the JmjC demethylase activities differently in KDM7s: binding to H3K4me3 by $KDM7A_{[PHD]}$ is inhibitory, binding to H3K4me3 by $KDM7B_{[PHD]}$ is activating. No demethylase activity is exhibited in-vitro by KDM7C. In-vitro characterisation of cyclic peptide functional effects on KDM7A,B was performed using demethylase assays on histone peptides and full-length histone extracts. OC9 inhibition of $KDM7B_{[PHD]}$ resulted in effective inhibition of $KDM7B_{[PHD+JmjC]}$ demethylation of the H3K9me2 mark, with low micromolar IC_{50} against H3K4me3K9me2 [1-21] (MALDI-TOF-MS, AlphaScreen). This efficacy ranked similarly to that of a leading $KDM7_{[JmjC]}$ targeting small molecule, Daminozide. OC9 inhibition of $KDM7A_{[PHD]}$ resulted in enhancement of $KDM7A_{[PHD+JmjC]}$ demethylation of the H3K9me2 mark in both H3K4me3K9me2 and H3K9me2 [1-21] substrates, with a low micromolar EC_{50} (MALDI-TOF-

MS). The inhibition of KDM7B, and enhancement of KDM7A, by OC9 was recapitulated in-vitro on full-length histone extracts by Western Blot analysis. This result was notable, as the assay provided a more challenging and 'real-world' scenario due to the histones containing multiple and varied post-translational modifications. OC3 and OC6 served to inhibit $KDM7B_{[PHD+JmjC]}$ demethylation of H3K4me3K9me2 with low micromolar IC_{50}, but were less effective than OC9 (MALDI-TOF-MS). OC6 appeared to inhibit $KDM7A,B_{[PHD+JmjC]}$ demethylase activities on full-length histone extracts, whereas OC3 did not appear to significantly inhibit either (Western Blot). However, results from solubility testing suggesting low micromolar aqueous solubility of OC6 meant that a non-specific mechanism of disruption may explain some of the assay behaviours exhibited by OC6 in this concentration range [294]. Therefore, further work remains to fully understand the KDM7-related behaviours of OC3 and OC6.

In-vitro selectivity profiles for the cyclic peptides were assessed by binding assays (BLI & ITC), cognate peptide binding-partner displacement (AlphaScreen) and JmjC-KDM demethylase activity (MALDI-TOF-MS, AlphaScreen, RapidFire-MS). OC9 was revealed to possess unprecedented selectivity for $KDM7(A,B,C)_{[PHD]}$ over other reader domains (five PHD fingers, one tandem-tudor domain) tested at micromolar concentrations. Profiling included the H3K4me3-binding $DIDO1_{[PHD]}$ which has the highest known human PHD finger sequence identity compared to the KDM7s. OC9 at 20 µM did not affect the demethylase activity of representative members from other JmjC-KDM sub-families, including KDM2A and KDM2B (the closest sub-family to the $KDM7_{[JmjC]}$) which are often difficult to achieve selectivity over with small molecule inhibition. OC3 did not displace or bind to any tested reader domains, and at 20 µM did not substantially inhibit other JmjC-KDM sub-family members. OC6 at 20 µM was found to possess a broad inhibition profile against all JmjC-KDM sub-family members tested, but the possibility of non-specific protein disruption must be borne in mind. The three

cyclic peptides did not inhibit HIF prolyl-hydroxylase 2 (EGLN1), a non-JmjC member of the 2OG oxygenase superfamily of therapeutic importance [358], under the same cross-screening conditions.

Preliminary characterisation of physicochemical properties of OC3, OC6 and OC9 indicated that none are likely to be passively membrane permeable. Structure-activity studies were conducted on OC9 to identify key residues that are critical for binding and function, which informed the design of targeted amino acid substitutions. This led to the development of a second-generation OC9 variant sequence (OC9.2.0), which was computationally predicted to have more favourable properties for cell permeability, including reduced accessible polar surface area and an increased compatibility with membrane environments. OC9 was evidenced to exist as a mixture of conformers in solution on the NMR timescale, likely to be due to the proline residues and potential *cis*/*trans* amide interconversion, but this was found not to be consequential to OC9 binding and PHD finger displacement potency.

OC9 was derivatised with a non-obstructive azide-linker, facilitating click-chemistry to install a biotin handle to create an affinity probe. Affinity pull-down from nuclear lysate using the biotinylated-OC9 probe was able to capture full-length KDM7B, which had been overexpressed in HEK293T cells. The pull-down by OC9 was found to be dependent on binding to the $KDM7_{[PHD]}$. Mutation of two key residues within the PHD finger aromatic cage, or competition with unlabelled OC9, drastically reduced pull-down yield, whereas pull-down by OC9-biotin remained effectual when in competition with the unlabelled weakly-binding OC9 variant, OC9_V10A. It is hoped that use of OC9 in affinity pull-downs, combined with proteomic analysis, will help to elucidate new biological roles for the KDM7 PHD finger. Initial attempts to develop a cell-based NanoBRET assay to test cyclic peptide permeabilities and

efficacies proved inconclusive. Optimisation of cell-based assays for exploration of the biological roles of the KDM7s will be required to further assess the cyclic peptides.

The principle of targeting of KDM reader domains, which has been successfully demonstrated in this work, represents a new strategy for modulating KDM activities. This could represent a valuable alternative approach to targeting of the JmjC domains, given the high-conservation of the JmjC domains resulting in frequent selectivity issues. Further, whilst recent work by Brennan and co-workers has found success in developing small molecules for some previously untargeted epigenetic reader domains ('acyllysine' readers: PCAF Bromodomain [359] and YEATS [360]), they have also suggested domain-domain interface targeting as an attractive strategy [361] for tackling other reader domains (e.g. PHD fingers). The success of OC3 in disrupting the KDM7B PHD finger reading action through multi-domain targeting (PHD+JmjC) demonstrates that the pursuit of such an approach could be worthwhile.

Another objective of the research was to develop methodologies to aid study of epigenetic histone modifications, and the effects of combinatorial marks. To this end, a series of modified lysine derivatives were synthesised containing matching or mimicked post-translational modifications relevant to those found in histones. Flexizyme technology was used to acylate tRNA with these non-canonical amino acids, and successful incorporation into test peptide sequences using ribosomal synthesis in an IVTT system was achieved. Future work will be aimed towards expanding the 'histone code' with this technology, increasing the number of incorporations, to study combinatorial histone modifications with an ambition to link the PTMs to chromatin structure and transcriptional regulation.

Overall, the work reported herein has identified cyclic peptide ligands for the KDM7 sub-family with unique modes of inhibition and binding properties. OC9 is the first potent and

selective KDM7 probe and PHD finger inhibitor, and OC3 is a potential KDM7A,B-selective inhibitor. OC9 and OC3 should be able to serve as useful tools in the study of the KDM7s, and potentially even provide scaffolds for adaptation into probes against similar intractable targets. Importantly, this work has demonstrated that it is feasible to develop a selective ligand for a PHD finger domain. Targeting of PHD fingers will provide new ways to study epigenetics, through disruption of important protein-protein interactions and allosteric regulation of epigenetic cross-talk and catalytic activities.

A final summary of the binding, displacement and inhibition potencies of OC9, and an overview of its selectivity profile, is given below (**F.9.i.**):

A

OC9 Efficacy Summary:		K_D (nM)		IC_{50} (nM)	IC/EC_{50} (nM) *
		BLI	ITC	AlphaScreen Displacement of H3K4me3 [1-21]	Demethylase Activity vs K9me2 [1-21] Peptides
KDM7A	PHD	19	33	3.6	
	PHD+JmjC	43		180	EC: 500 (vs H3K9me2) & 4000 (vs H3K4me3K9me2) (MALDI)
KDM7B	PHD	4	12	8.7	
	PHD+JmjC	6	11	15	IC: 1400 (MALDI) & 420 (AlphaScreen) (vs H3K4me3K9me2)
KDM7C	PHD	1.1	10	2.6	
	PHD+JmjC	0.3		60	

*IC denotes inhibitory, EC denotes enhancement

B OC9 Selectivity Profiling Summary:

No Displacement (up to at least 1 µM) of other H3K4me3 Reader Domains:

DIDO1 PHD
KDM5A PHD3
TAF3 PHD
ING2 PHD
SPIN TTD

by AlphaScreen

No Binding by ITC

DIDO1 PHD
KDM5A PHD3

No Inhibition (at 20 µM) of Demethylation by other JmjC-KDMs:

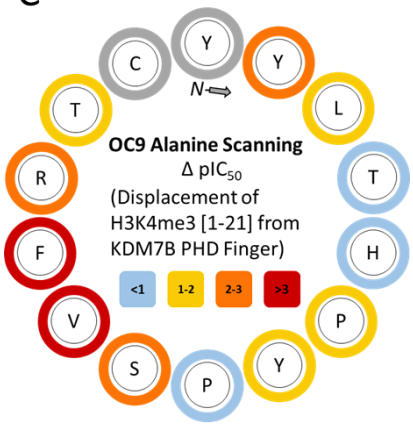
KDM2A
KDM2B
KDM3A
KDM4A
KDM5A *
KDM5B *
KDM5C
KDM6B

or Oxidation by other 2OG-Oxygenase

EGLN1

** by AlphaScreen, others by MALDI-TOF*

C



F.9.i. A) Summary table for the measured binding, displacement and inhibitory potencies of OC9 against the KDM7 proteins across the project. The BLI binding constants are taken from the 'Association then Dissociation' fits. **B)** Summary table for the selectivity profiling of OC9 across multiple reader domains and KDMs. **C)** Alanine scanning effect on KDM7B PHD finger displacement.

Chapter 10. Materials and Methods

10.1. Small Molecule Synthesis

Chemicals:

Chemicals for synthesis were purchased from Sigma Aldrich, Alfa Aesar, and Fluorochem and used without further purification unless otherwise stated.

Thin Layer Chromatography:

Plastic 0.20 mm Polygram Sil g/UV254 plates were used for TLC. Plates were dried using a heat gun, then visualised under UV (254 nm) or by use of an appropriate staining agent (e.g. ninhydrin, potassium permanganate, sulfuric acid) and further heating.

Chromatography:

Chromatography purification of small molecules was performed using a Biotage Isolera One system and Biotage SNAP KP Sil columns, unless otherwise stated. *Note: CV refers to column volume throughout the following described purification procedures.*

IR Spectroscopy:

IR spectra were recorded using a Bruker Tensor 27 Fourier Transform spectrophotometer using a KBr disk or attenuated total reflectance (ATR) attachment. Compounds were analysed from solutions, oils, or solids, with each spectrum corrected using background measurements. Only peaks of interest are reported.

Small Molecule Mass Spectrometry:

High-resolution mass spectra were recorded using a Bruker MicroTOF ESI mass spectrometer.

Optical Rotation:

Optical rotation was recorded using a Schmidt Haensch Unipol Polarimeter. Solvents were spectrophotometric grade $CHCl_3$ (Sigma-Aldrich) or MeOH (HPLC Grade, Sigma Aldrich). Background measurements with the corresponding solvent were used to zero readings before use.

NMR Spectroscopy:

NMR spectra were processed using MestReNova v11.0.1.

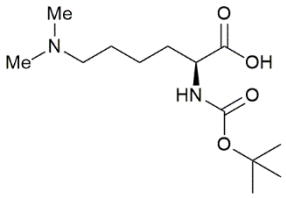
^1H NMR spectra were recorded using a 400 Bruker (AVIII NanoBay) (400 MHz). Protons, multiplicity and coupling constants (reported to the nearest 0.5 Hz) were determined using the MestReNova software. Chemical shift information (δH) is given in parts per million (ppm) relative to the residual solvent signal ($CDCl_3$ 7.26 ppm, MeOH-d_4 3.31 ppm). Where compounds are reported as having an excess of formate, ^1H-NMR integral ratios were used to estimate the formate content and yield of product.

^{13}C NMR spectra were recorded using a 400 Bruker (AVIII NanoBay) (100 MHz). Analysis was performed with MestReNova software. Carbon chemical shift (δC) is given in parts per million (ppm) relative to the solvent signal ($CDCl_3$ 77.16 ppm, MeOH-d_4 49.00). Spectra were recorded with decoupling from ^1H nuclei.

Additional 2D spectra (COSY, HSQC & HMBC) were used to support assignments where necessary.

10.1.1. Boc-Lys(Me2)-OH

(N2-(*tert*-Butoxycarbonyl)-N6,N6-dimethyl-L-lysine)



To a 100 mL round bottomed flask was added Boc-Lys-OH (1000 mg, 4.07 mmol), formaldehyde (37 w/w % solution in H_2O, 0.990 mL, 12.2 mmol, 3 eqv) and 20 mL ethanol. The white suspension was stirred at room temperature under an air atmosphere for 5 minutes, then sodium triacetoxyborohyride ($NaBH(OAc)_3$, STAB) (3010 mg, 14.25 mmol, 3.5 eqv) was added. Effervescence was observed and the reaction was monitored by TLC (100% MeOH, ninhydrin) for disappearance of starting material. After 2 hours the reaction was complete. The resulting white solid precipitate was collected by filtration; ^{1}H-NMR analysis determined this was not the desired product. The remaining liquor was concentrated in vacuo to give an oil, which was re-dissolved in several mL of MeOH and dried using a rotary evaporator onto TELOS diatomaceous earth. This was used to dry load a 30 g C_{18} KP-Sil Biotage column, eluted using a gradient of H_2O in MeCN (0 to 50% (v/v) over 15 CV). The product containing fractions were identified by TLC, combined, and the solvent was evaporated in vacuo until a thick oil remained, which was lyophilised to yield the desired compound as a hygroscopic white sticky solid (988 mg, 89% yield). This procedure was adapted from Baba et al [362]. Spectroscopic data were previously reported in a different solvent.

^{1}H NMR (400 MHz, methanol-d_4) δ 3.96 (dd, *J* = 7.0, 5.1 Hz, 1H, -αC**H**), 3.06 (t, *J* = 8.0 Hz, 2H, -C**H₂**-NMe_2), 2.83 (s, 6H, -N(C**H₃**)₂), 1.86 – 1.63 (m, 4H, -C**H₂**-CH_2-NMe_2, -αCH-C**H₂**-), 1.44 (s, 11H, -C**H₃**x3, CH_2-C**H₂**-CH_2);

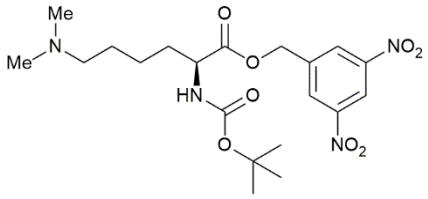
^{13}C NMR (101 MHz, methanol-d_4) δ 177.73 (-**C**OOH), 156.22 (Boc **C**=O), 78.68 (Me$_3$**C**-), 57.43 (Me$_2$N-**C**H$_2$-), 55.05 (α**C**), 41.95 ((**C**H$_3$)$_2$N-), 32.17 (-**C**H$_2$-CH$_2$-NMe$_2$), 27.36 (-CH$_2$-**C**H$_2$-CH$_2$-), 23.82 (-αCH-C**H$_2$**-), 21.97 (3x**C**H$_3$);

IR (cm^{-1}): 3402, 2975, 2935, 1697 (C=O), 1596, 1484, 1390, 1366;

HRMS [M+H] = calc 275.19653, obs 275.19623; $\alpha_D^{25} = +21, MeOH, c = 0.62.$

10.1.2. Boc-Lys(Me2)-ODNB

(3,5-Dinitrobenzyl N2-(*tert*-butoxycarbonyl)-N6,N6-dimethyl-L-lysinate)



To an oven dried 25 mL round-bottomed flask was added Boc-Lys(me2)-OH (489 mg, 1.78 mmol) and DMF (6 mL). With vigorous stirring, DIPEA (1550 µL, *p* = 0.742 mg/mL, 8.74 mmol, 4.9 eqv) was added until total dissolution was observed. Then, COMU [(1-Cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate] (764 mg, 1.78 mmol, 1 eqv) and 3,5-dinitrobenzyl alcohol (353 mg, 1.78 mmol, 1 eqv) were added at room temperature under N_2. The resulting orange solution turned to deep red within several minutes. The reaction was monitored by LCMS and TLC (DCM/MeOH 85/15 (v/v), product R_f = 0.20, ninhydrin stain) for disappearance of starting material. After 18 hours the reaction was diluted with 200 mL EtOAc and washed 5 x 200 mL saturated bicarbonate solution. LCMS indicated only trace amount of product in the aqueous layer. The organic phase was dried over $MgSO_4$ and concentrated in vacuo. The resulting brown oil was purified by re-dissolution in DCM, then evaporation of the solvent using a rotary evaporator onto TELOS diatomaceous earth and dry loading of a Biotage column (25g KP-Sil) eluted in DCM

with a gradient of MeOH (3% to 5% to 15% (v/v) over 28 CV). The product containing fractions were identified by TLC, combined, and concentrated in-vacuo to yield the desired compound as a yellow/brown oil (314 mg, 39% yield). This procedure was adapted from Kang et al [220] and Murakami et al [363]. Spectroscopic data were not previously reported for this compound.

^{1}H NMR (400 MHz, chloroform-*d*) δ 9.02 (t, *J* = 2.0 Hz, 1H, *p*-Ar-**H**), 8.56 (d, *J* = 2.0 Hz, 2H, *o*-Ar-**H**), 5.36 (d, *J* = 3.0 Hz, 2H, Ar-C**H$_2$**-), 5.27 (d, *J* = 8.0 Hz, 1H, -N**H**-), 4.38 – 4.25 (m, 1H, αC**H**), 2.27 (t, *J* = 7.0 Hz, 2H, Me$_2$NC**H$_2$**-), 2.22 (s, **Me$_2$**N-), 1.86 (m, 1H, αCH-C**H**'H''), 1.70 (dq, *J* = 14.0, 8.0 Hz, 1H, αCH-CH'**H''**), 1.50 (p, *J* = 6.5 Hz, 2H, Me$_2$N-CH$_2$-C**H$_2$**-), 1.43 (s, 11H, 3xC**H$_3$**, -CH$_2$-C**H$_2$**-CH$_2$-);

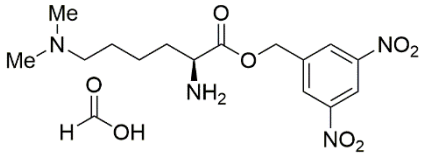
^{13}C NMR (101 MHz, chloroform-*d*) δ 172.60 (-**C**O$_2$-Ar), 155.51 (Boc **C**=O), 148.65 (**C**-NO$_2$), 140.15 (-CO$_2$-CH$_2$-Ar**C**-), 127.79 (-C-**C**H-C(NO$_2$)-), 118.55 (-C(NO$_2$)-**C**H-), 80.18 (Me$_3$**C**-), 64.29 (Ar-**C**H$_2$-), 58.90 (Me$_2$N-**C**H$_2$-), 53.65 (α**C**), 45.25 ((**C**H$_3$)$_2$N-), 31.64 (-αCH-C**H$_2$**-), 28.24 (3x**C**H$_3$), 26.84 (-**C**H$_2$-CH$_2$-NMe$_2$), 23.12 (-CH$_2$-**C**H$_2$-CH$_2$-);

IR (cm^{-1}): 2976, 2942, 1748 (C=O, ester), 1709 (C=O, carbamate), 1544 (C-NO$_2$), 1344 (C-NO$_2$);

HRMS [M+H] = calc 455.21364, obs 455.21146; $\alpha_D^{25} = -1.2, CHCl_3, c = 1.2.$

10.1.3. H-Lys(Me2)-ODNB

3,5-Dinitrobenzyl N6,N6-dimethyl-L-lysinate



To a 10 mL flat bottomed glass vial was added Boc-Lys(Me2)-ODNB (67.1 mg, 0.148 mmol) and formic acid (1 mL). The resulting clear dark orange solution was stirred at room temperature, with monitoring for disappearance of starting material by TLC and LCMS. After

7 hours the formic acid was removed using a rotary evaporator under high vacuum to yield the desired product as a yellow oil (80 mg, >99% apparent yield), with an excess of formate present by NMR. This procedure was adapted from Mathew et al [364] and Kang et al [220]. Spectroscopic data were not previously reported for this compound.

^{1}H NMR (400 MHz, methanol-d_4) δ 9.00 (t, *J* = 2.0 Hz, 1H, *p*-Ar-**H**), 8.74 – 8.69 (m, 2H, *o*-Ar-**H**), 8.26 (s, formate), 5.57 – 5.51 (s, 2H, Ar-C**H$_2$**-), 4.20 (t, *J* = 6.5 Hz, 1H, αC**H**), 3.15 – 3.06 (m, 2H, Me_2N-C**H$_2$**-), 2.86 (s, 6H, **Me$_2$**N-), 2.12 – 1.90 (m, 2H, αCH-C**H$_2$**-), 1.86 – 1.72 (m, 2H, Me_2N-CH_2-C**H$_2$**-), 1.67 – 1.41 (m, 2H, -CH_2-C**H$_2$**-CH_2-);

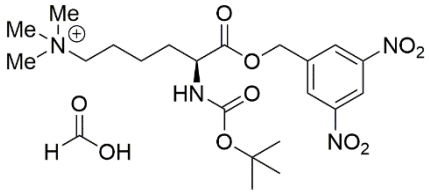
^{13}C NMR (101 MHz, methanol-d_4) δ 169.12 (-**C**O_2-Ar), 165.51 (formate), 148.66 (**C**-NO_2), 139.61 (-CO_2-CH_2-Ar**C**-), 127.93 (-C-**C**H-C(NO_2)-), 118.19 (-C(NO_2)-**C**H-), 65.33 (Ar-**C**H_2), 56.78 (Me_2N-**C**H_2-), 52.17 (α**C**), 41.83 ((**C**H_3)$_2$N-), 29.72 (-αCH-C**H$_2$**-), 23.63 (-**C**H_2-CH_2-NMe_2), 21.76 (-CH_2-**C**H_2-CH_2-);

IR (cm^{-1}): 1752 (C=O ester), 1596, 1544 (C-NO_2), 1345 (C-NO_2);

HRMS [M+H] = calc 355.16121, obs 355.16097; $\alpha_D^{25} = +3.5, MeOH, c = 0.67.$

10.1.4. Boc-Lys(Me3)-ODNB

(*S*)-5-((*tert*-Butoxycarbonyl)amino)-6-((3,5-dinitrobenzyl)oxy)-N,N,N-trimethyl-6-oxohexan-1-aminium **or** (3,5-Dinitrobenzyl N2-(*tert*-butoxycarbonyl)-N6,N6,N6-trimethyl-L-lysinate)



To a 10 mL flat bottomed glass vial was added Boc-Lys(Me2)-ODNB (50 mg, 0.11 mmol), and methyl iodide (9 µL, *p* = 2.28, 1.3 eqv, 0.14 mmol) and DIPEA (20 µL, *p* = 0.742, 1 eqv, 0.11 mmol) and DCM (1 mL). The vial was wrapped in tin foil and sealed, and the resulting yellow

solution was stirred at room temperature for 1 hour. LCMS indicated consumption of starting material, at which point 180 mg of diatomaceous earth was added and the slurry was concentrated in vacuo using a rotary evaporator. The earth was loaded to a Biotage SNAP KP-SIL 10g column and eluted in $CHCl_3$ with a gradient of MeOH (containing 5% (v/v) formic acid) (0 to 70% over 30 CV). The product containing fractions were identified by TLC, combined, filtered and concentrated in vacuo to yield the desired product as a yellow oil (25 mg, 35% apparent yield), with an excess of formate present by NMR. This procedure was adapted from Baba et al [362]. Spectroscopic data were not previously reported for this compound.

^{1}H NMR (400 MHz, methanol-d_4) δ 8.95 (t, *J* = 2.0 Hz, 1H, *p*-Ar-**H**), 8.67 (d, *J* = 2.0 Hz, 2H, *o*-Ar-**H**), 8.38 (s, formate), 5.41 (s, 2H, Ar-C**H$_2$**-), 4.22 (dd, *J* = 9.0, 5.0 Hz, 1H, αC**H**), 3.35 (d, *J* = 8.5 Hz, 2H, Me_3N-C**H$_2$**-), 3.13 (s, 9H, **Me$_3$**-N-), 2.01 – 1.74 (m, 4H, αCH-C**H$_2$**, Me_3N-CH_2-C**H$_2$**-), 1.42 (s, 11H, 3xC**H$_3$**, -CH_2-C**H$_2$**-CH_2-);

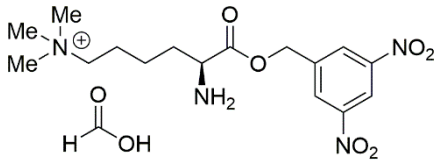
^{13}C NMR (101 MHz, methanol-d_4) δ 172.21 (-**C**O_2-Ar), 166.82 (formate), 156.80 (Boc **C**=O), 148.60 (**C**-NO_2), 140.70 (-CO_2-CH_2-Ar**C**-), 127.54 (-C-**C**H-C(NO_2)-), 117.78 (-C(NO_2)-**C**H-), 79.38 (Me_3-**C**-), 66.07 (Me_3N-**C**H_2-), 64.22 (Ar-**C**H_2-), 53.44 (α**C**H), 52.14 (**Me$_3$**N-), 30.37 (-αCH-C**H$_2$**-), 27.25 (3x**C**H_3), 22.40 (-CH_2-**C**H_2-CH_2-), 22.01 (-**C**H_2-CH_2-NMe_3);

IR (cm^{-1}): 1743 (C=O ester), 1700 (C=O carbamate), 1591, 1542 (C-NO_2), 1344(C-NO_2);

HRMS [M+H] = calc 469.22929, obs 469.22934; $\alpha_D^{25} = -6.3, MeOH, c = 0.85.$

10.1.5. H-Lys(Me3)-ODNB

(*S*)-5-Amino-6-((3,5-dinitrobenzyl)oxy)-N,N,N-trimethyl-6-oxohexan-1-aminium **or** 3,5-Dinitrobenzyl N6,N6,N6-trimethyl-L-lysinate



To a 10 mL round bottomed glass vial was added Boc-Lys(Me3)-ODNB (25 mg, 0.038 mmol) and formic acid (0.5 mL). The vial was wrapped in tin foil and the solution was stirred at room temperature for 7 hours. LCMS indicated reaction completion and the formic acid was removed using a rotary evaporator under high vacuum, then dried overnight under high vacuum. This gave the desired compound as a yellow oil (20 mg, 95% apparent yield), with an excess of formate present by NMR. This procedure was adapted from Mathew et al [364] and Kang et al [220]. Spectroscopic data were not previously reported for this compound.

^{1}H NMR (400 MHz, methanol-d_4) δ 8.99 (t, *J* = 2.0 Hz, 1H, *p*-Ar-**H**), 8.71 (d, *J* = 2.0 Hz, 2H, *o*-Ar-**H**), 8.24 (s, formate), 5.53 (s, 2H, Ar-C**H$_2$**-), 4.25 (s, 1H, αC**H**), 3.37 (dt, *J* = 11.5, 7.5 Hz, 2H, Me_3NC**H$_2$**-), 3.14 (s, 9H, **Me$_3$**-N-), 2.06 (d, *J* = 13.0 Hz, 2H, αCH-C**H$_2$**), 1.87 (s, 2H, Me_3NCH_2-C**H$_2$**-), 1.70 – 1.41 (m, 2H, -CH_2-C**H$_2$**-CH_2-);

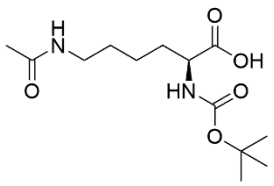
^{13}C NMR (101 MHz, methanol-d_4) δ 168.96 (-**C**O_2-Ar), 165.08 (formate), 148.65 (**C**-NO_2), 139.64 (-CO_2-CH_2-Ar**C**-), 127.95 (-C-**C**H-C(NO_2)-), 118.18 (-C(NO_2)-**C**H-), 65.79 (Me_3N-**C**H_2-), 65.35 (Ar-**C**H_2-), 52.22 (**Me$_3$**N-), 52.12 (α**C**H), 29.65 (-αCH-C**H$_2$**-), 22.06 (-**C**H_2-CH_2-NMe_3), 21.63 (-CH_2-**C**H_2-CH_2-);

IR (cm^{-1}): 1748 (C=O ester), 1707(C=O formic acid), 1672, 1582, 1540 (C-NO_2), 1344 (C-NO_2);

HRMS [M+H] = calc 369.17686, obs 369.17676; $\alpha_D^{25} = +3.8, MeOH, c = 0.45.$

10.1.6. Boc-Lys(Ac)-OH

N6-Acetyl-N2-(*tert*-butoxycarbonyl)-L-lysine



To a 25 mL round bottomed flask was added Boc-Lys-OH (200 mg, 0.813 mmol) and 5 mL of MeCN/H_2O (9:1 v/v). With vigorous stirring at room temperature, acetic anhydride, Ac_2O, (152 µL, 1.79 mmol, 2.2 eqv) was added dropwise to the white suspension and no notable exotherm was observed by a thermometer submerged in the reaction solution. The reaction turned to a colourless, clear, solution over the next hour. The reaction was monitored by TLC (100% MeOH, ninhydrin) for the disappearance of starting material. After 3 hours the reaction solution was frozen using LN_2 and then lyophilised overnight to yield the desired product as a hygroscopic white solid (232 mg, >99% yield). This procedure was adapted from Gupta et al [365]. Spectroscopic data were previously reported in a different solvent.

^{1}H NMR (400 MHz, methanol-d_4) δ 4.06 (dd, J = 9.0, 5.0 Hz, 1H, α-C**H**), 3.16 (t, J = 7.0 Hz, 2H, AcNH-C**H$_2$**-), 1.92 (d, J = 3.0 Hz, 3H, C**H$_3$**CONH-), 1.88 – 1.75 (m, 1H, α-CH-**H**'H''), 1.71 – 1.59 (m, 1H, α-CH-H'**H''**), 1.56 – 1.47 (m, 2H, AcNH-CH_2-C**H$_2$**-), 1.44 (s, 11H, -CH_2-C**H$_2$**-CH_2-, 3x CH_3);

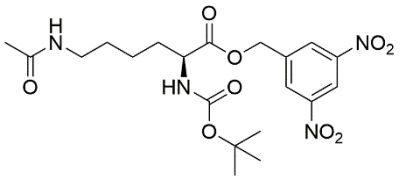
13C NMR (101 MHz, methanol-d_4) δ 174.79 (-**C**OOH), 171.79 (Me-**C**(O)-NR), 156.75 (Boc **C**=O), 79.04 (Me_3**C**-), 53.37 (α-**C**), 38.79 (AcNH-**C**H_2-), 31.04 (αC-**C**H_2-), 28.47 (AcNH-CH_2-**C**H_2-), 27.31 (3x**C**H_3-), 22.89 (-CH_2-**C**H_2-CH_2-), 21.11 (**C**H_3-CONH-);

IR (cm^{-1}): 1694 (C=O acid), 1654 (C=O carbamate), 1626 (C=O amide);

HRMS [M+Na] = calc: 311.15774, obs: 311.15768; $\alpha_D^{25} = -4.3, MeOH\ c = 0.43.$

10.1.7. Boc-Lys(Ac)-ODNB

3,5-Dinitrobenzyl N6-acetyl-N2-(*tert*-butoxycarbonyl)-L-lysinate



To a 10 mL round bottomed flask was added Boc-Lys(Ac)-OH (116 mg crude, 0.406 mmol) and 3 mL DMF. With stirring at room temperature, COMU [(1-Cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate] (261 mg, 0.609 mmol, 1 eqv) and 3,5-dinitrobenzyl alcohol (80 mg, 0.406 mmol, 1 eqv) were added. The formation of product was followed by TLC (EtOAc, product R_f = 0.40). After 18 hours the reaction was diluted with 100 mL of EtOAc, washed twice with 100 mL saturated bicarbonate, twice with 5 w/w % citric acid and finally 100 mL brine. The organic layer was then dried over $MgSO_4$, concentrated in vacuo, and purified by column chromatography (10 g SNAP Biotage, 50% (v/v) using a gradient of EtOAc in cyclohexane (0 to 100 % over 15 CV). The product containing fractions were identified by TLC and concentrated in vacuo to yield the desired product as an amber oil (79 mg, 50% yield). This procedure was adapted Kang et al [220] and Murakami et al [363]. Spectroscopic data were not previously reported for this compound.

^{1}H NMR (400 MHz, chloroform-*d*) δ 9.01 (t, *J* = 2.0 Hz, 1H, *p*-Ar-**H**), 8.63 – 8.53 (m, 2H, *o*-Ar-**H**), 5.59 (s, 1H, N**H**), 5.38 – 5.34 (m, 2H, -C**H$_2$**-Ar), 5.11 – 5.05 (m, 1H, N**H**), 4.34 (td, *J* = 8.0, 5.0 Hz, 1H, αC**H**), 3.23 (ddd, *J* = 12.0, 7.0, 5.5 Hz, 2H, AcNH-C**H$_2$**-), 1.96 (s, 3H, -HNCOC**H$_3$**), 1.91 – 1.80 (m, 1H, αCH-C**H'**H''-), 1.76 – 1.65 (m, 1H, αCH-CH'**H''**), 1.60 – 1.50 (m, 2H, AcNH-CH$_2$-C**H$_2$**-), 1.43 (s, 11H, -CH$_2$-C**H$_2$**-CH$_2$-, 3xC**H$_3$**);

^{13}C NMR (101 MHz, chloroform-*d*) δ 172.43 (-O$_2$**C**-Ar), 170.25 (Me**C**ONH-), 155.55 (Boc **C**=O), 148.63 (**C**-NO$_2$), 140.07 (-CO$_2$-CH$_2$-Ar**C**-), 127.83 (-C-**C**H-C(NO$_2$)-), 118.58 (-C(NO$_2$)-**C**H-), 80.27

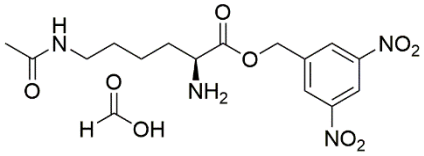
(Me_3**C**-), 64.38 (Ar-**C**H_2-), 53.37 (α**C**), 38.94 (AcNH-**C**H_2-), 31.78 (αC-**C**H_2-), 29.04 (AcNH-CH_2-**C**H_2-), 28.23 (3x**C**H_3-), 23.26 (-CH_2-**C**H_2-CH_2-), 22.64 (**C**H_3-CONH-);

IR (cm^{-1}): 1747 (C=O ester), 1705 (C=O carbamate), 1656 (C=O amide), 1543 (C-NO_2), 1457, 1366, 1344 (C-NO_2), 1251, 1159;

HRMS [M+H] = calc: 469.19290, obs: 469.19302; $\alpha_D^{25} = -1.2, CHCl_3, c = 0.58.$

10.1.8. H-Lys(Ac)-ODNB

3,5-Dinitrobenzyl N6-acetyl-L-lysinate



To a 10 mL flat bottomed glass vial was added Boc-Lys(Ac)-ODNB (54 mg, 0.11 mmol) and 1 mL of formic acid. The solution was stirred at room temperature and the disappearance of starting material was monitored by TLC and LCMS. After 5 hours the formic acid was removed by using a rotary evaporator under high vacuum, then the residue was further dried overnight under high vacuum, to yield the desired product as a yellow oil (63 mg, >99% apparent yield), with an excess of formate present by NMR. This procedure was adapted from Mathew et al [364] and Kang et al [220]. Spectroscopic data were not reported for this compound.

^{1}H NMR (400 MHz, methanol-d_4) δ 9.00 (t, *J* = 2.0 Hz, 1H, *p*-Ar-**H**), 8.72 – 8.69 (m, 2H, *o*-Ar-**H**), 8.16 (s, formate), 5.55 – 5.51 (m, 2H, Ar-C**H₂**-), 4.19 (t, *J* = 6.5 Hz, 1H, αC**H**), 3.20 – 3.11 (m, 2H, AcNH-C**H₂**-), 2.09 – 1.94 (m, 2H, αCH-C**H₂**-), 1.92 (s, 3H, -HNCOC**H₃**), 1.60 – 1.34 (m, 4H, AcNH-CH_2-C**H₂**-, -CH_2-C**H₂**-CH_2-);

^{13}C NMR (101 MHz, methanol-d_4) δ 171.97 (-O_2**C**-Ar), 169.09 (Me**C**ONH-), 164.19 (formate), 148.67 (**C**-NO_2), 139.56 (-CO_2-CH_2-Ar**C**-), 128.01 (-C-**C**H-C(NO_2)-), 118.22 (-C(NO_2)-**C**H-), 65.34

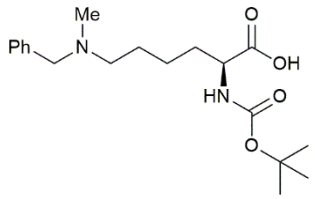
(-**C**H$_2$-Ar), 52.43 (α**C**), 38.39 (AcNH-**C**H$_2$-), 29.75 (αC-**C**H$_2$-), 28.51 (AcNH-CH$_2$-**C**H$_2$-), 21.89 (**C**H$_3$-CONH-), 21.15 (-CH$_2$-**C**H$_2$-CH$_2$-);

IR (cm^{-1}): 3408, 1750 (C=O ester), 1655 (C=O amide), 1543 (C-NO$_2$), 1346 (C-NO$_2$);

HRMS [M+H] = calc: 369.14048, obs: 369.14084; $\alpha_D^{25} = +4.6, MeOH, c = 0.92.$

10.1.9. Boc-Lys(Bn,Me)-OH

N6-Benzyl-N2-(*tert*-butoxycarbonyl)-N6-methyl-L-lysine



To a 100 mL round bottomed flask was added Boc-Lys-OH (495 mg, 2.01 mmol), and freshly distilled benzaldehyde (240 µL, 2.4 mmol, 1.2 eqv) in 18 mL EtOH. The white suspension was stirred at room temperature for 30 minutes, gradually becoming clearer. Sodium triacetoxy borohydride (STAB, NaBH(OAc)$_3$) (636 mg, 3.01 mmol, 1.5 eqv) was then added and the reaction was stirred overnight. After 16 hours, formaldehyde (37.5 w/w % in H$_2$O) (450 µL, 8.04 mmol, 4 eqv) and NaBH(OAc)$_3$ (637 mg, 3.01 mmol, 1.5 eqv) were added. The reaction was monitored by TLC, (DCM/MeOH, 90/10 (v/v)), for consumption of the starting material. After a further 1.5 hours the reaction was deemed complete and the white precipitate was collected by filtration. The liquor was reduced to dryness onto diatomaceous earth in vacuo and subjected to flash chromatography (25g SNAP Biotage) using a gradient of MeOH in DCM (1% to 50% (v/v) over 25 CV) to give the desired product as a hygroscopic solid (374 mg, 53% yield). This procedure was adapted from Anderson et al [366] and Huang et al [367]. Spectroscopic data were not previously reported for this compound.

^{1}H NMR (400 MHz, methanol-d_4) δ 7.50 (qd, *J* = 5.5, 4.5, 2.0 Hz, 2H, Ar-**H**), 7.47 – 7.42 (m, 3H, Ar-**H**), 4.20 (s, 2H, Ar-C**H₂**-), 3.97 (dd, *J* = 7.5, 5.0 Hz, 1H, αC**H**), 3.05 – 2.95 (m, 2H, (Bn,Me)-N-C**H₂**-), 2.66 (s, 3H, C**H₃**-N(Bn)-R), 1.79 (ddd, *J* = 24.0, 11.0, 6.5 Hz, 3H, αCH-C**H'**H'', BnMeN-CH_2-C**H₂**-), 1.67 (ddd, *J* = 14.0, 8.0, 6.5 Hz, 1H, αCH-CH'**H''**), 1.43 (s, 11H, -CH_2-C**H₂**-CH_2-, 6xC**H₃**);

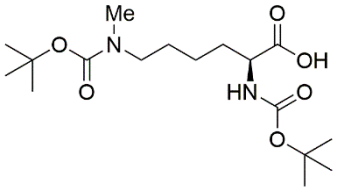
^{13}C NMR (101 MHz, methanol-d_4) δ 178.01 (-**C**OOH), 156.27 (Boc **C**=O), 130.75 (Ar-**C**), 130.64 (Ar-**C**), 129.25 (Ar-**C**), 128.76 (Ar-**C**), 78.61 (Me_3**C**-), 59.34 (Ar-C**H₂**-), 55.43 (BnMeN-**C**H_2-), 55.21 (α**C**H), 38.61 (**C**H_3-N(Bn)-R), 32.31 (αCH-**C**H_2-), 27.41 (3x**C**H_3), 23.64 (BnMeN-CH_2-**C**H_2-), 22.53 (CH_2-**C**H_2-CH_2);

IR (cm^{-1}): 3393 (br), 2977 (br), 1690 (br, C=O), 1571, 1402, 1365, 1167;

HRMS [M+H] = calc: 351.22783, obs: 351.22791; $\alpha_D^{25} = +10, MeOH, c = 3.4.$

10.1.10. Boc-Lys(Boc,Me)-OH

N2,N6-bis(*tert*-Butoxycarbonyl)-N6-methyl-L-lysine



Boc-Lys(Bn,Me)-OH (374 mg, 1.07 mmol) was added to 15 mL EtOH and $Pd(OH)_2$ (75 mg, 20 w/w %) and stirred at room temperature under an H_2 atmosphere. The reaction was monitored by TLC and LCMS for consumption of starting material. After 3 hours the H_2 atmosphere was replaced with N_2 and the catalyst was collected by filtration through celite. The product-containing residue was subjected to several repeated evaporations by dilution in a small volume of methanol and using a rotary evaporator. The residue was then placed under high vacuum to dry overnight. The dried residue was dissolved in 5 mL pH 10 H_2O (adjusted with 5 M NaOH) and 5 mL THF. On ice, boc anhydride (Boc_2O), (300 µL, 1.38 mmol,

1.29 eqv) was added and the pH was readjusted to > 9. The reaction was monitored by TLC (MeOH) for consumption of starting material. The reaction was monitored regularly to ensure pH > 9, adjusting when necessary with 1M NaOH. After 1 hour, a further 150 µL Boc_2O was charged. After 2 hours, TLC indicated completion of the reaction and the reaction was adjusted to pH 12. The solution was washed twice with 35 mL of diethyl ether, then the aqueous layer was acidified to pH 4 using 1 M HCl (noting the formation of a white precipitate), and extracted using DCM (3x 35 mL). The aqueous layer became clear. The combined DCM extractions were dried with $MgSO_4$ and concentrated in vacuo to yield the desired product as a colourless, clear oil (335 mg, 87% yield). This procedure was adapted from Anderson et al [366] and Huang et al [367]. Spectroscopic data were not previously reported for this compound.

^{1}H NMR (400 MHz, methanol-d_4) δ 4.11 – 3.92 (m, 1H, αC**H**), 3.23 (d, *J* = 7.0 Hz, 2H, , -NC**H₂**-), 2.83 (s, 3H, **Me**NBoc-), 1.90 – 1.75 (m, 1H, αCH-C**H'**H''), 1.67 (qd, *J* = 9.0, 4.5 Hz, 1H, αCH-CH'**H''**), 1.62 – 1.50 (m, 2H, BocMeN-CH_2-C**H₂**-), 1.45 (d, *J* = 5.0 Hz, 18H, 6xC**H₃**), 1.40 – 1.26 (m, 2H, -CH_2-C**H₂**-CH_2-);

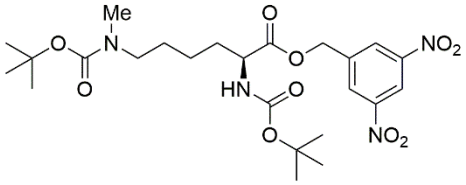
^{13}C NMR (101 MHz, methanol-d_4) δ 174.79 (-**C**OOH), 156.71 (Boc **C**=O), 156.24 (-O-**C**(O)-NHR-), 79.46 (Me_3**C**-), 79.02 (Me_3**C**-), 53.33 (α**C**-H), 48.19 (BocMeN-**C**H_2-), 33.00 (**C**H_3-N(Boc)-R), 31.08 (αCH-**C**H_2-), 27.35 (3x**C**H_3), 26.64 (BocNHMe-CH_2-**C**H_2-), 22.67 (-CH_2-**C**H_2-CH_2-);

IR (cm^{-1}): 2977, 2932, 1696 (br, C=O);

HRMS [M-H]: calc = 359.21876, obs = 359.21863; $\alpha_D^{25} = -1.8, MeOH, c = 0.20.$

10.1.11. Boc-Lys(Boc,Me)-ODNB

3,5-Dinitrobenzyl N2,N6-bis(*tert*-butoxycarbonyl)-N6-methyl-L-lysinate



To a 5 mL round bottomed flask was added Boc-Lys(Boc,Me)-OH (160 mg, 0.444 mmol), 1 mL DMF, 3,5-dinitrobenzyl chloride (96.3 mg, 0.4 mmol, 0.9 eqv), and NEt_3 (124 µL, 0.888 mmol, 2 eqv). The initially yellow solution was heated at 45°C for 4 hours and monitored by TLC (30% (v/v) EtOAc in cyclohexane, product R_f = 0.20), with the reaction turning pink/purple over the course of the first hour. The reaction was stirred overnight at 35 °C, and after 24 hours the DMF was removed using a rotary evaporator under high vacuum. The residue was then dissolved in MeOH and reduced to dryness onto diatomaceous earth using a rotary evaporator. This was purified by chromatography (Biotage 25g SNAP) using a gradient of EtOAc in cyclohexane (0% to 25% (v/v) over 25 CV), to give the desired product as a yellow oil (228 mg, 95% yield). This procedure was adapted from Kang et al [220] and Murakami et al [363]. Spectroscopic data were not previously reported for this compound.

Restricted rotation gave rotamer peaks in the NMR spectra

^{1}H NMR (400 MHz, chloroform-*d*) δ 9.00 (t, *J* = 2.0 Hz, 1H, *p*-Ar-**H**), 8.56 (d, *J* = 2.0 Hz, 2H, *o*-Ar-**H**), 5.35 (d, *J* = 7.0 Hz, 2H, Ar-C**H$_2$**-), 4.94 (s, 1H, -N**H**-), 4.29 (s, 1H, αC**H**), 3.16 (dt, *J* = 13.5, 6.5 Hz, 2H, -NC**H$_2$**-), 2.80 (s, 3H, C**H$_3$**-(Boc)N-), 1.91 – 1.62 (m, 2H, αCH-C**H$_2$**-), 1.60 – 1.46 (m, 1H, BocMeN-CH$_2$-C**H'**H''), 1.42 (d, *J* = 8.5 Hz, 19H, BocMeN-CH$_2$-CH'**H''**, 6xC**H$_3$**), 1.38 – 1.23 (m, 2H, -CH$_2$-C**H$_2$**-CH$_2$-);

Rotamer peaks arbitrarily assigned A or B

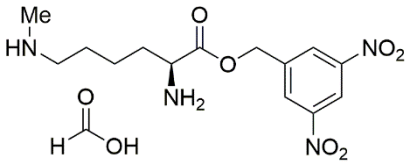
^{13}C NMR (126 MHz, chloroform-*d*) δ 172.54 (-**C**O_2-Ar), 156.15 (Boc **C**=O), 155.69 (-O-**C**(O)-NHR-), 148.64 (**C**-NO_2), 140.23 (-CO_2-CH_2-Ar**C**-), 127.80 (-C-**C**H-C(NO_2)-), 118.54 (-C(NO_2)-**C**H-), 80.33 ([A] Me_3**C**), 80.08 ([B] Me_3**C**), 79.34 (Me_3**C**), 64.27 (Ar-**C**H_2-), 53.42 (α**C**H), 48.42 ([A] BocNHMe-**C**H_2-), 47.07 ([B] BocNHMe-**C**H_2-), 34.07 (BocNH(**C**H_3)-, 32.04 ([A] αCH-**C**H_2-), 31.03 ([B] αCH-**C**H_2-), 28.46 (3x**C**H_3), 28.24 (3x**C**H_3), 27.46 ([A] BocNHMe-CH_2-**C**H_2-), 26.77 ([B] BocNHMe-CH_2-**C**H_2-), 22.70 ([A] -CH_2-**C**H_2-CH_2-), 22.14 ([B] -CH_2-**C**H_2-CH_2-);

IR (cm^{-1}): 3331, 2978, 2932, 1753 (C=O ester), 1714 (br, C=O carbamate), 1547 (C-NO_2), 1456, 1393, 1366, 1344 (C-NO_2), 1160;

HRMS [M+Na] = calc: 563.23236, obs: 563.23187; $\alpha_D^{25} = -1.1, CHCl_3, c = 2.5.$

10.1.12. H-Lys(Me)-ODNB

3,5-Dinitrobenzyl N6-methyl-L-lysinate



To a 10 mL flat bottomed glass vial was added Boc-Lys(Boc,Me)-ODNB (111 mg, 0.206 mmol) and 1 mL of formic acid. The solution was stirred at room temperature for 8 hours, then the formic acid was removed using a rotary evaporator under high vacuum. The resulting residue was dried overnight under high vacuum to give the desired product as an orange oil (110 mg, >99% apparent yield), with an excess of formate present by NMR. This procedure was adapted from Mathew et al [364] and Kang et al [220]. Spectroscopic data were not previously reported for this compound.

^{1}H NMR (400 MHz, methanol-d_4) δ 8.99 (t, J = 2.0 Hz, 1H, *p*-Ar-**H**), 8.73 – 8.68 (m, 2H, *o*-Ar-**H**), 8.27 (s, formate), 5.53 (s, 2H, Ar-C**H$_2$**-), 4.19 (t, J = 6.5 Hz, 1H, αC**H**), 3.03 – 2.94 (m, 2H, MeNH-C**H$_2$**-), 2.69 (s, 3H, **Me**NH-), 2.12 – 1.91 (m, 2H, αCH-C**H$_2$**-), 1.80 – 1.68 (m, 2H, MeNH-CH$_2$-C**H$_2$**-), 1.67 – 1.44 (m, 2H, CH$_2$-C**H$_2$**-CH$_2$-);

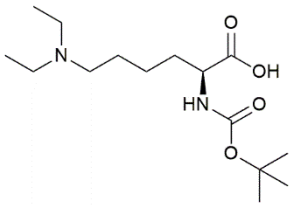
^{13}C NMR (126 MHz, methanol-d_4) δ 169.18 (-**C**O$_2$-Ar), 165.57 (formate), 148.69 (-**C**-NO$_2$), 139.64 (-CO$_2$-CH$_2$-**C**-), 127.96 (C-**C**H-C(NO$_2$)-), 118.20 (C(NO$_2$-**C**H-), 65.35 (-CO$_2$-**C**H$_2$-), 52.24 (α**C**H), 48.28 (MeNH-**C**H$_2$-), 32.08 (**C**H$_3$-NH-), 29.79 (αCH-**C**H$_2$-), 25.17 (MeNH-CH$_2$-**C**H$_2$-), 21.81 (-CH$_2$-**C**H$_2$-CH$_2$-);

IR (cm^{-1}): 1754 (C=O ester), 1596, 1543 (C-NO$_2$), 1345 (C-NO$_2$);

HRMS [M+H] = calc: 341.14556, obs: 341.14554; $\alpha_D^{25} = +6.1, MeOH, c = 0.92.$

10.1.13. Boc-Lys(Et2)-OH

N2-(*tert*-Butoxycarbonyl)-N6,N6-diethyl-L-lysine



To a 100 mL RBF was added BocLysOH (1.0 g, 4.1 mmol, 1 eqv) and 20 mL EtOH on ice. Acetaldehyde (1 mL, 28.8 mmol, 7.1 eqv) that had been pre-chilled on ice was added and the solution turned brown. NaBH(OAc)$_3$ (STAB) (3.01 g, 14.2 mmol, 3.5 eqv) was added in a single portion and the reaction was stirred on ice for 1 hour, turning to an orange suspension. The reaction was monitored for consumption of starting material by LCMS. The solid was collected by filtration and the liquor was concentrated in vacuo to a brown oil, then dried onto diatomaceous earth using a rotary evaporator. Purification was performed by chromatography (Biotage SNAP KP Sil 25g) using a gradient of DCM/MeOH (0 – 85% (v/v) 15

CV) to yield the desired product as a hygroscopic brown solid (957 mg, 77% yield). This procedure was adapted from Baba et al [362]. Spectroscopic data were not previously reported for this compound.

^{1}H NMR (400 MHz, methanol-d_4) δ 3.97 (dd, *J* = 7.0, 5.0 Hz, 1H, αC**H**), 3.18 (q, *J* = 7.5 Hz, 4H, -N(C**H₂**CH₃)₂), 3.10 – 3.02 (m, 2H, Et₂N-C**H₂**-), 1.88 – 1.58 (m, 4H, αCH-C**H₂**- and Et₂N-CH₂-C**H₂**-), 1.44 (s, 11H, -CH₂-C**H₂**-CH₂- and 3x Boc C**H₃**), 1.29 (t, *J* = 7.5 Hz, 6H, -N(CH₂**CH₃**)₂);

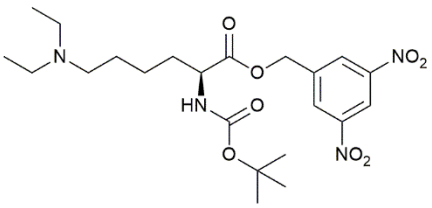
^{13}C NMR (101 MHz, methanol-d_4) δ 177.85 (**C**OOH), 156.21 (Boc **C**=O), 78.62 (-**C**Me₃) 55.14 (α**C**H), 51.49 (Et₂N-**C**H₂-), 46.72 (-N(**C**H₂CH₃)₂), 32.35 (αCH-**C**H₂-), 27.39 (3x Boc **C**H₃), 23.16 (Et₂N-CH₂-**C**H₂-), 22.28 (-CH₂-**C**H₂-CH₂-), 7.80 (-N(CH₂**C**H₃)₂);

IR (cm^{-1}): 1695 (br, C=O), 1590, 1388, 1364;

HRMS [M+H] = calc: 303.22783, obs: 22784; $\alpha_D^{25} = +16, MeOH, c = 2.0.$

10.1.14. Boc-Lys(Et2)-ODNB

3,5-Dinitrobenzyl N2-(*tert*-butoxycarbonyl)-N6,N6-diethyl-L-lysinate



Boc-Lys(Et2)-OH (815 mg, 2.70 mmol, 1 eqv) was added to 5 mL DMF, followed by DIPEA (977 µL, 5.4 mmol, 2 eqv) then 3,4-dinitrobenzyl alcohol (552 mg, 2.78 mmol, 1.03 eqv) then HATU ((1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate)) (1283 mg, 3.38 mmol, 1.2 eqv). The vessel containing the deep brown solution was covered in tin foil and stirred at room temperature for 17 hours, with consumption of starting material monitored by LCMS. On completion, the reaction was diluted with 200 mL of EtOAc and washed with saturated bicarbonate (3x 150 mL) and brine

(1x 150 mL). The organic layer was dried over $MgSO_4$, concentrated in vacuo onto diatomaceous earth and purified by chromatography (Biotage SNAP KP Sil 50g) using a gradient of DCM/MeOH 0-15% (v/v) 28 CV) to yield the desired product as a dark brown oil (256 mg, 20% yield). This procedure was adapted from Kang et al [220] and Murakami et al [363]. Spectroscopic data were not previously reported for this compound.

^{1}H NMR (400 MHz, chloroform-*d*) δ 9.01 (t, *J* = 2.0 Hz, 1H, *p*-Ar-**H**), 8.56 (d, *J* = 2.0 Hz, 2H, *o*-Ar-**H**), 5.37 (d, *J* = 3.5 Hz, 2H, Ar-C**H$_2$**-), 5.21 (d, *J* = 8.0 Hz, 1H, N**H**), 4.33 (dd, *J* = 8.0, 5.0 Hz, 1H, αC**H**), 2.86 – 2.73 (m, 4H, -N(C**H$_2$**CH_3)$_2$), 2.68 (s, 2H, Et_2N-C**H$_2$**-), 1.88 (dtd, *J* = 13.5, 8.5, 5.5 Hz, 1H, αCH-C**H'**H''), 1.73 (td, *J* = 13.0, 10.5, 6.0 Hz, 3H, αCH-CH'**H''** and Et_2N-CH_2-C**H$_2$**-), 1.42 (s, 11H, -CH_2-C**H$_2$**-CH_2- and 3x Boc C**H$_3$**), 1.19 (q, *J* = 9.0, 8.0 Hz, 6H, -N(CH_2C**H$_3$**)$_2$);

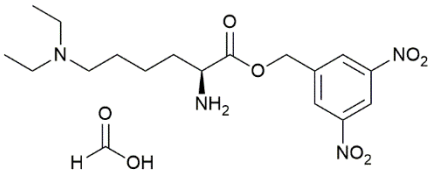
^{13}C NMR (101 MHz, chloroform-*d*) δ 172.35 (**C**O_2CH_2), 155.49 (Boc **C**=O), 148.66 (**C**-NO_2), 140.05 (CO_2-CH_2-Ar**C**-), 127.90 (-C-**C**H-C(NO_2)-), 118.65 (-C(NO_2)-**C**H-), 80.35 (Me_3**C**-), 64.49 (Ar-**C**H_2-), 53.45 (α**C**H), 51.61 (Et_2N-**C**H_2-), 46.51 (-N(**C**H_2CH_3)$_2$), 31.79 (-αCH-**C**H_2-), 28.26 (3x **C**H_3), 24.59 (Et_2N-CH_2**C**H_2-), 23.14 (-CH_2-**C**H_2-CH_2-), 9.91 (-N(CH_2**C**H_3)$_2$);

IR (cm^{-1}): 1746 (C=O ester), 1705 (C=O carbamate), 1597, 1543 (C-NO_2), 1342 (C-NO_2);

HRMS [M+H] = calc: 483.24494, obs: 483.24442; $\alpha_D^{25} = +1.0, CHCl_3, c = 0.6.$

10.1.15. H-Lys(Et2)-ODNB

3,5-Dinitrobenzyl N6,N6-diethyl-L-lysinate



Boc-Lys(Et2)-ODNB (28.5 mg, 0.059 mmol) was added to a flat bottomed glass vial in 1 mL formic acid, wrapped in tin foil and stirred at room temperature for 5 hours. Starting material consumption was monitored by LCMS. On completion, the formic acid was removed using a rotary evaporator under high vacuum to yield the desired product as an orange oil (35 mg, 80% apparent yield), with an excess of formate present by NMR. This procedure was adapted from Mathew et al [364] and Kang et al [220]. Spectroscopic data were not previously reported for this compound.

^{1}H NMR (400 MHz, methanol-d$_4$) δ 8.98 (t, J = 2.0 Hz, 1H, *p*-Ar-**H**), 8.71 (d, J = 2.0 Hz, 2H, *o*-Ar-**H**), 8.22 (s, 8H, formate), 5.53 (s, 2H, Ar-**C**H$_2$-), 4.23 (t, J = 6.5 Hz, 1H, α**C**H), 3.22 (q, J = 7.5 Hz, 4H, -N(C**H$_2$**CH$_3$)$_2$), 3.17 – 3.07 (m, 2H, Et$_2$N-C**H$_2$**-), 2.13 – 1.91 (m, 2H, αCH-C**H$_2$**-), 1.79 (dtd, J = 11.5, 8.0, 5.5 Hz, 2H, Et$_2$N-CH$_2$-C**H$_2$**-), 1.68 – 1.45 (m, 2H, CH$_2$-C**H$_2$**-CH$_2$-), 1.31 (t, J = 7.5 Hz, 6H, -N(CH$_2$C**H$_3$**)$_2$);

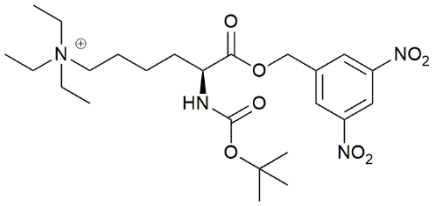
^{13}C NMR (101 MHz, methanol-d$_4$) δ 168.96 (**C**O$_2$CH$_2$), 164.87 (formate), 148.65 (**C**-NO$_2$), 139.62 (CO$_2$-CH$_2$-Ar**C**-), 127.98 (-C-**C**H-C(NO$_2$)-), 118.18 (-C(NO$_2$)-**C**H-), 65.38 (Ar-**C**H$_2$-), 52.18 (α**C**H), 51.02 (Et$_2$N-**C**H$_2$-), 46.82 (-N(**C**H$_2$CH$_3$)$_2$), 29.66 (-αCH-**C**H$_2$-), 23.04 (Et$_2$N-CH$_2$**C**H$_2$-), 21.95 (-CH$_2$-**C**H$_2$-CH$_2$-), 7.65 (-N(CH$_2$**C**H$_3$)$_2$);

IR (cm^{-1}): 2919 (br), 1748 (C=O ester), 1710 (C=O formate), 1576, 1541 (C-NO$_2$), 1344 (C-NO$_2$);

HRMS [M+H] = calc: 383.19251, obs: 383.19263; $\alpha_D^{25} = +7.0, MeOH, c = 0.1.$

10.1.16. Boc-Lys(Et3)-ODNB

(*S*)-5-((*tert*-Butoxycarbonyl)amino)-6-((3,5-dinitrobenzyl)oxy)-N,N,N-triethyl-6-oxohexan-1-aminium **or** 3,5-Dinitrobenzyl N2-(*tert*-butoxycarbonyl)-N6,N6,N6-triethyl-L-lysinate



Boc-Lys(Et2)-ODNB (108 mg, 0.224 mmol) was added to a flat bottomed glass vial with 0.5 mL MeCN. Ethyl iodide (40 µL, 0.5 mmol, *p* = 1.95, 2.2 eqv) was added with anhydrous potassium carbonate (31 mg, 0.224 mmol, 1 eqv). The reaction vessel was wrapped in tin foil and heated to 60 °C for 6 hours, then 40 °C for 10 hours. Consumption of starting material was monitored by LCMS. On completion, the mixture was concentrated onto diatomaceous earth and purified by chromatography (Biotage 10g SNAP KP Sil) using a gradient of $CHCl_3$/MeOH (5% formic acid (v/v)) 0-35% (v/v) 18 CV) to yield the desired product as a yellow oil (103 mg, 69% apparent yield), with an excess of formate present by NMR. This procedure was adapted from Baba et al [362]. Spectroscopic data were not previously reported for this compound.

^{1}H NMR (400 MHz, chloroform-*d*) δ 8.96 (t, *J* = 2.0 Hz, 1H, *p*-Ar-**H**), 8.59 (d, *J* = 2.0 Hz, 2H, *o*-Ar-**H**), 8.08 (s, 4H formate), 5.57 (d, *J* = 8.0 Hz, 1H, N**H**), 5.42 (s, 2H, Ar-C**H$_2$**-), 4.41 – 4.29 (m, 1H, αC**H**), 3.37 (q, *J* = 7.5 Hz, 6H, -N(C**H$_2$**CH$_3$)$_3$), 3.33 – 3.19 (m, 2H, Et$_3$N-C**H$_2$**-), 1.95 (dd, *J* = 13.5, 6.5 Hz, 1H, αCH-C**H'**H''), 1.90 – 1.71 (m, 3H, αCH-CH'**H''**, Et$_3$N-CH$_2$-C**H$_2$**-), 1.63 – 1.45 (m, 2H, -CH$_2$-C**H$_2$**-CH$_2$-), 1.43 – 1.33 (m, 18H, 3x Boc C**H$_3$** and -N(CH$_2$C**H$_3$**)$_3$);

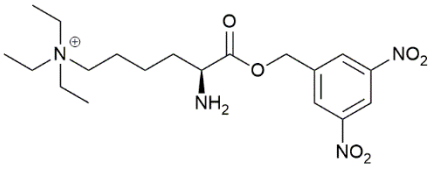
^{13}C NMR (101 MHz, chloroform-*d*) δ 172.11 (**C**O$_2$CH$_2$), 164.35 (formate), 156.04 (Boc **C**=O), 148.55 (**C**-NO$_2$), 140.11 (CO$_2$-CH$_2$-Ar**C**-), 128.19 (-C-**C**H-C(NO$_2$)-), 118.60 (-C(NO$_2$)-**C**H-), 80.75 (Me$_3$**C**-), 64.85 (Ar-**C**H$_2$-), 57.33 (Et$_3$N-**C**H$_2$-), 53.57 (-N(**C**H$_2$CH$_3$)$_3$), 53.15 (α**C**H), 31.52 (-αCH-**C**H$_2$-), 28.26 (Boc 3x **C**H$_3$), 22.57 (-CH$_2$-**C**H$_2$-CH$_2$-), 21.48 (Et$_3$N-CH$_2$**C**H$_2$), 8.05 (-N(CH$_2$**C**H$_3$)$_3$);

IR (cm^{-1}): 1748 (C=O ester), 1707 (C=O carbamate), 1543 (C-NO_2), 1345 (C-NO_2);

HRMS [M+H] = calc: 511.27624, obs: 511.27588; $\alpha_D^{25} = +3.2, CHCl_3, c = 0.15.$

10.1.17. H-Lys(Et3)-ONB

(*S*)-5-Amino-6-((3,5-dinitrobenzyl)oxy)-N,N,N-triethyl-6-oxohexan-1-aminium **or** 3,5-Dinitrobenzyl N6,N6-diethyl-L-lysinate



Boc-Lys(Et3)-ODNB (55 mg, 0.079 mmol) was added to a flat bottomed glass vial in 1 mL formic acid, wrapped in tin foil and stirred at room temperature for 5 hours. Starting material consumption was monitored by LCMS. On completion, the formic acid was removed using a rotary evaporator under high vacuum to yield the desired product as an orange oil (54 mg, 88% apparent yield), with an excess of formate present by NMR. This procedure was adapted from Mathew et al [364] and Kang et al [220]. Spectroscopic data were not previously reported for this compound.

^{1}H NMR (400 MHz, methanol-d_4) δ 8.98 (t, *J* = 2.0 Hz, 1H, *p*-Ar-**H**), 8.71 (d, *J* = 2.0 Hz, 2H, *o*-Ar-**H**), 8.22 (s, 8H, formate), 5.53 (s, 2H, Ar-C**H₂**-), 4.23 (t, *J* = 6.5 Hz, 1H, αC**H**), 3.22 (q, *J* = 7.5 Hz, 4H, -N(C**H₂**CH_3)$_3$), 3.17 – 3.07 (m, 2H, Et_3N-C**H₂**-), 2.13 – 1.91 (m, 2H, αCH-C**H₂**-), 1.79 (dtd, *J* = 11.5, 8.0, 5.5 Hz, 2H, Et_3N-CH_2-C**H₂**-), 1.68 – 1.45 (m, 2H, -CH_2-C**H₂**-CH_2-), 1.31 (t, *J* = 7.5 Hz, 6H, -N(CH_2C**H₃**)$_3$);

^{13}C NMR (101 MHz, methanol-d_4) δ 168.86 (**C**O_2CH_2), 163.98 (formate), 148.65 (**C**-NO_2), 139.55 (CO_2-CH_2-Ar**C**-), 128.10 (-C-**C**H-C(NO_2)-), 118.23 (-C(NO_2)-**C**H-), 65.51 (Ar-**C**H_2-), 56.22

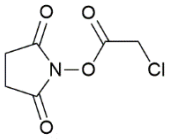
(Et_3N-**C**H_2-), 52.69 (-N(**C**$H_2CH_3)_3$), 52.28 (α**C**H), 29.78 (-αCH-**C**H_2-), 21.87 (-CH_2-**C**H_2-CH_2-), 21.05 (Et_3N-CH_2**C**H_2-), 6.46 (-N(CH_2**C**$H_3)_3$);

IR (cm^{-1}): 1748 (C=O ester), 1582, 1539 (C-NO_2), 1344 (C-NO_2);

HRMS [M+H] = calc: 411.22381, obs: 411.22418; $\alpha_D^{25} = +4.1, MeOH, c = 1.6.$

10.1.18. N-succinimidyl 2-chloroacetate

2,5-Dioxopyrrolidin-1-yl 2-chloroacetate



Chloroacetyl chloride (1 mL, 12.5 mmol, 1 eqv) was added to 50 mL DCM and stirred on ice. Anhydrous potassium carbonate (K_2CO_3) (1.9 g, 13.8 mmol, 1.1 eqv) was then added in a single portion. Then, with vigorous stirring, N-hydroxy succinimide (1.43 g, 12.5 mmol, 1 eqv) was added giving a white suspension. Stirring on ice was continued for 1 hour and then at room temperature for 2 hours at which point TLC (100% EtOAc) indicated starting material consumption. The reaction was quenched with 40 mL H_2O and stirred to give two clear phases. The organic layer was separated, dried over $MgSO_4$ and reduced to dryness in vacuo to yield the desired compound as a yellow solid in high (>90 %) purity (601 mg, 25% yield). This procedure was performed according to literature precedent and spectroscopic data were consistent with reported data [368].

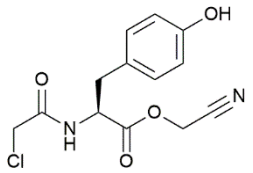
^{1}H NMR (400 MHz, chloroform-*d*) δ 4.37 (s, 2H, -C**H$_2$**-Cl), 2.87 (s, 4H, 2x C**H$_2$**);

^{13}C NMR (101 MHz, chloroform-*d*) δ 168.38 (**C**=O), 163.18 (**C**=O), 37.84 (-**C**H_2-Cl), 25.54 (2x **C**H_2);

IR (cm^{-1}): 1781, 1699 (C=O, br), 1213; Melting Point: 113 ±2 °C.

10.1.19. ClAc-^{L}Tyr-CME

Cyanomethyl (2-chloroacetyl)-L-tyrosinate



L-tyrosine (410 mg, 2.27 mmol, 1.1 eqv) was added to 9 mL H_2O in a 50 mL RBF and $NaHCO_3$ (480 mg, 5.71 mmol, 2.8 eqv) was charged in a single portion, then (N-succinimidyl 2-chloroacetate) (383 mg, 2.00 mmol, 1 eqv) was added in 2.5 mL dioxane. With vigorous stirring the white suspension frothed initially, and the reaction was then stirred at room temperature for 3 hours. Product formation was monitored by LCMS. On completion, the reaction was washed with EtOAc (2x 20 mL), the aqueous layer was acidified to pH 1 using 1 M HCl and extracted with EtOAc (3x 20 mL). The three organic extractions were combined, dried over $MgSO_4$ and concentrated in vacuo to give a yellow oil (287 mg, containing (2-chloroacetyl)-L-tyrosine). This was immediately dissolved in 2 mL DMF. To this was first added chloroacetonitrile (450 µL, p = 1.19), then DIPEA (720 µL), and the resulting clear yellow solution was stirred at room temperature for 17 hours. LCMS indicated consumption of starting material, at which point the reaction was diluted with 40 mL Et_2O to give a cloudy white suspension. The organic layer was washed using 1 M HCl (3x 10 mL) and saturated bicarbonate solution (3x 10 mL) and brine (1x 15 mL). The organic layer was dried over $MgSO_4$ and concentrated in vacuo to yield an orange oil, then wet loaded to a 10 g SNAP KP Sil Biotage column and eluted using a gradient of EtOAc in cyclohexane (0-80% (v/v) 15 CV). The product containing fractions were identified by TLC, combined, and concentrated in vacuo to yield the desired product as a colourless oil (45 mg, 9% yield). This procedure was performed according

to the literature report from Hayashi et al [208]. Spectroscopic data were previously reported in a different solvent.

^{1}H NMR (400 MHz, chloroform-*d*) δ 6.98 (dd, *J* = 9.0, 2.5 Hz, 2H, *o*-Ar-**H**), 6.81 – 6.74 (m, 2H, *m*-Ar-**H**), 4.83 (dt, *J* = 8.0, 6.0 Hz, 1H, αC**H**), 4.80 – 4.65 (m, 2H, C**H$_2$**-CN), 4.02 (d, *J* = 2.0 Hz, 2H, C**H$_2$**-Cl), 3.08 (d, *J* = 6.0 Hz, 2H, αCH-C**H$_2$**-);

^{13}C NMR (101 MHz, chloroform-*d*) δ 169.67 (**C**=O), 166.24 (**C**=O), 155.84 (Ar**C**-OH), 130.30 (o-Ar-**C**), 125.73 (Ar**C**-CH_2-), 115.93 (*m*-Ar-**C**), 113.72 (**C**N), 53.36 (α**C**H), 49.04 (**C**H_2CN), 42.24 (**C**H_2Cl), 36.64 (αCH-**C**H_2-);

IR (cm^{-1}): 3478, 3330, 2356 (CN), 2329, 1750 (C=O ester), 1653 (C=O amide), 1516, 1189;

HRMS [M+H] = calc: 297.06366, obs: 297.06384; $\alpha_D^{25} = -4.3, CHCl_3, c = 1.9.$

10.2. List of Bacterial Plasmid Vectors

pCDF-BirA (SGC, Accession: JF914075): This is an expression vector for *E. coli* biotin ligase (BIR). The vector also carries the CloDF13-derived CDF replicon, *lacI* gene and streptomycin/ spectinomycin resistance gene (SmR).

pNH-TrxT (SGC, Accession: GenBank GU269914): This is a pET expression vector with His_6-Trx (Trx = *E. coli* thioredoxin) in an N-terminal fusion sequence, with a TEV protease cleavage site. The vector includes sites for Ligation-Independent Cloning (LIC), and a "stuffer" fragment that includes the SacB gene, allowing negative selection on 5% sucrose LB agar plates. Antibiotic resistance: Kanamycin.

N-terminal fusion sequence:

MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQG
KLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGTENL
YFQ*SM

(* = TEV cleavage site)

pNIC28-Bsa4 (SGC, Accession: GenBank EF198106): This is a pET expression vector with His_6 tag in 22-aa N-terminal fusion peptide, with TEV protease cleavage site. The vector includes sites for LIC cloning, and a "stuffer" fragment that includes the SacB gene, allowing negative selection on 5% sucrose LB agar plates. Antibiotic resistance: Kanamycin.

N-terminal fusion sequence: MHHHHHHSSGVDLGTENLYFQ*SM (* = TEV cleavage site)

pNIC-Bio3 (SGC, Accession: JN792439): This is a pET expression vector with His_6 tag in 22-aa N-terminal fusion peptide, with TEV protease cleavage site, and a biotinylation sequence in C-terminal fusion. The vector includes sites for LIC cloning, and a "stuffer" fragment that includes the SacB gene, allowing negative selection on 5% sucrose. Antibiotic resistance: Kanamycin.

N-terminal fusion sequence: `MHHHHHHSSGVDLGTENLYFQ*SM` (* = TEV cleavage site).

C-terminal fusion sequence: `GSKGGYGLNDIFEAQKIEWHE`

pNIC-ZB (SGC, Accession: GenBank GU452710): This is a pET expression vector with His6 and Z-basic (ZB) tags followed by a TEV protease cleavage site. The vector includes sites for LIC cloning, and a "stuffer" fragment that includes the SacB gene, allowing negative selection on 5% sucrose. Antibiotic resistance: Kanamycin.

N-terminal fusion sequence:

`MHHHHHHSSGVDNKFNKERRRARREIRHLPNLNREQRRAFIRSLRDDPSQSANLLAEAKKLN`

`DAQPKGTENLYFQ*SM` (* = TEV cleavage site)

10.3. Polymerase Chain Reaction (PCR): General Protocol

Equipment: DNA Engine Tetrad 2 Peltier Thermal Cycler (BioRad).

PCR reactions were performed in standard thin wall 100 µL PCR tubes, with tailored reaction concentrations and parameters according to the specified conditions for a given experiment. General polymerases used were: BioTaq (Bioline) for general DNA amplifications in the production of RNAs; Q5 (NEB) for target gene cloning and mRNA display related PCR; PfuUltraII Fusion (Agilent) for site directed mutagenesis and longer (> 1kb) target gene cloning.

10.4. Ligation Independent Cloning (LIC): Primers

Forwards primer LIC tag 5'-TACTTCCAATCCATG-3' (appended to the 5'-end of the forwards primer matching gene of interest)

Reverse primer LIC tag 5'-TATCCACCTTTACTG-3' (appended to the 5'-end of the reverse primer matching gene of interest)

DIDO1$_{[PHD]}$: Forwards (with LIC tag): 5'-TACTTCCAATCCATGGAATGTGAGGGTTACGACCC-3'. Reverse (with LIC tag): 5'-TATCCACCTTTACTGCTGCACTTGCAGAATGGTGC-3'

AIRE$_{[PHD1]}$: Forwards (with LIC tag): 5'-TACTTCCAATCCATGCAGCTCCACCAGAAG-3'. Reverse (with LIC tag): 5'-TATCCACCTTTACTGCTCTCCTCTGCCCGGGG-3'

10.5. Ligation Independent Cloning (LIC): General Protocol

Construct inserts were cloned from appropriate template plasmids. LIC primer tags were designed according to SGC vector information and target gene overlap was designed to be at least 15 base pairs. The procedure was adapted from the SGC protocol [369].

Agarose 1% gel recipe = 0.5 g agarose, 50 mL 1xTBE (40 mM tris 20 mM borate 1 mM EDTA) buffer, microwaved to dissolve agarose, then cooled, 5 µL SYBR Safe added, and set.

Insert Preparation: For one 25 µL PCR reaction: 2.5 µL ~5 ng/µL template (< 1000 ng), 1.25 µL of 10 µM forward primer, 1.25 µL of 10 µM reverse primer, 5 µL of 5X Q5 Buffer, 0.5 µL of 10 mM dNTPs, 0.25 µL of Q5 polymerase (NEB), MilliQ to 25 µL (14.25 µL). PCR program used: 98 °C 30s, 5x[98 °C 7s, 60 °C 15s, 72 °C 20s], 25x[98 °C 7s, 72 °C 35s], 72 °C 2 min. The DNA was purified by 2% agarose gel (80 V for 30 minutes), and the band of the target weight was excised and purified using a gel extraction kit (Qiagen) as per manufacturer's instructions.

Cohesive ends were generated by reaction assembled in a PCR tube containing: PCR product DNA (~5 ng/µL) 7.15 µL, 10X T4 polymerase buffer 1 µL, dCTP 25 mM 1 µL, 100 mM DTT 0.5 µL, BSA (NEB, 10 mg/ml) 0.1 µL, T4 DNA polymerase (Novagen or NEB) 0.25 µL. The reaction was incubated for 30 minutes at 22 °C, then 20 minutes at 75 °C to inactivate the enzyme.

Vector Preparation: For a 100 µL reaction volume with pNIC-Bio3/pNIC28-Bsa4 and restriction enzyme BsaI: MilliQ (volume adjustable according to DNA concentration), 10 µL NEB Cutsmart buffer, 5 µg plasmid (volume determined according to DNA concentration), 3 µL (30 units) BsaI restriction enzyme. The reaction was incubated for 2 hours at 50 °C. Verification of the DNA cut was analysed by a 0.8% agarose gel (0.4 g agarose, 50 mL 1X TAE (89 mM tris 89 mM acetate 2 mM EDTA) buffer, 5 µL SYBR Safe). Samples (1 µL cut purified vector, 4 µL MilliQ, 1 µL DNA loading dye) were run at 80 V for 1 hour, then imaged under UV. Cut vector DNA was purified using a PCR purification kit (Qiagen) and eluted in 50 µL MilliQ. Cohesive ends were generated by reaction assembled in a PCR tube containing: MilliQ 11 µL, purified cut plasmid 25 µL, 10X T4 DNA polymerase buffer 5 µL (Novagen or NEB buffer 2), dGTP (25 mM) 5 µL, BSA (NEB, 20 mg/ml) 0.25 µL, 100 mM DTT (filtered) 2.5 µL, T4 DNA polymerase (LIC qualified) 1.25 µL (Novagen or NEB). The reaction was incubated for 30 minutes at 22 °C, then 20 minutes at 75 °C to inactivate the enzyme.

Annealing and Transformation: 1 µL of T4 treated vector DNA was mixed with 2 µL of T4 treated insert DNA for 20 minutes at 20 °C prior to general bacterial transformation procedure. The transformed bacterial cells were plated onto 5% sucrose LB agar plates containing the appropriate antibiotics. A colony PCR was performed to confirm insert presence in selected colonies. Corresponding plasmids were obtained by 5-10 mL miniprep and verified by Sanger sequencing.

10.6. Miniprep and Midiprep

Mini and Midi preps were performed with a NucleoSpin Plasmid (Mini) or NucleoBond Xtra (Midi) according to the manufacturer's (Macherey-Nagel) instructions.

10.7. Bacterial Cell Types

Plasmid Replication: XL-10 Gold, TOP10, NEB 10-beta, NEB 5-alpha

Protein Production: BL21(DE3), BL21(DE3) pLysS, Rosetta2 (DE3) pLysS, Rosetta pBirA

Notes from the manufacturer (New England Biolabs):

Rosetta2 cells are BL21 derivatives designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli.* These strains supply tRNAs for 7 rare codones (AGA, AGG, AUA, CUA, GGA, CCC, and CGG) on a compatible chloramphenicol-resistant plasmid. The tRNA genes are driven by their native promoters.

DE3 indicates that the host is a lysogen of λDE3, and therefore carries a chromosomal copy of the T7 RNA polymerase gene under control of the lacUV5 promoter. Such strains are suitable for production of protein from target genes cloned in pET vectors by induction with IPTG.

10.8. Bacterial Transformations: General Protocol

25 µL of competent *E. coli* cells were incubated with desired plasmid DNA for 20 minutes on ice, then subjected to heat shock at 42°C by water bath for 30 s, returned to ice for 30 s and 200 µL of 2TY media added, then immediate transfer to 37°C for 1 hour recovery at 180 rpm. Cells were then plated on appropriate antibiotic plates.

10.9. Protein Production Materials

A New Brunswick Scientific G25 Environmental Shaker was used for incubations.

10.9.1. Media

All media was autoclaved prior to use for bacterial growth.

2TY (1 L): Tryptone (16 g), Yeast Extract (10 g), NaCl (5 g), pH 7.0 (adjust 5 M NaOH), MilliQ water

LB (1 L): Tryptone (10 g), Yeast Extract (5 g), NaCl (10 g), MilliQ water

LB Agar (300 mL): Tryptone (3 g), Yeast Extract (1.5 g), NaCl (3 g), Agar (4.5 g), MilliQ water

5 % Sucrose LB Agar: as for LB agar, but make up to 90 % volume, autoclave, then add the final 10 % volume with a solution of 50 w/w % sucrose in MilliQ water (0.22 um filter sterilised).

10.9.2. Antibiotics

Antibiotics were prepared to a stock concentration (1000x final concentration) and filter sterilised (0.2 µm) before use. Aliquots were frozen for long term storage. Stock concentrations: Kanamycin = 50 mg/mL, Ampicillin = 50 mg/mL, Spectinomycin = 50 mg/mL, chloramphenicol = 34 mg/mL (in EtOH).

10.9.3. Starter Cultures

100 mL of 2TY media containing the appropriate antibiotics was inoculated with a pipette tip scraping from either the agar plate culture or glycerol stock. Cells were shaken overnight at 37°C at 180 rpm.

10.10. SDS-PAGE Gel Analysis

Samples were prepared by addition of NuPAGE LDS Sample Buffer (4X) and DTT (to a final concentration of 1 mM), then heated at 95 °C for 5 minutes.

Samples were separated by gel electrophoresis on commercially purchased pre-cast gels, loading up to 20 µL per well of a 10-well gel, or up to 15 µL per well of a 15-well gel. Commercially purchased gels that were used were either a NuPAGE Novex 12% Bis-Tris Protein Gel 1.0 mm (for target molecular weights > 50 kDa) or a NuPAGE Novex 4-12% Bis-Tris Protein Gel 1.0 mm (for target molecular weights < 50 kDa). PageRuler Prestained Plus Ladder (10 to 250 kDa, Thermo Scientific 26619) was used as a size marker. Gels were run in a Novex XCell SureLock Mini-Cell tank according to the manufacturer's instructions, in either MES or MOPS buffer at 180 V for 45 minutes, and analysed with Instant Blue (Expedeon).

10.11. Protein Production and Purification Protocols

10.11.1. General Protocol for His-Tagged Proteins

Equipment: ÄKTA Pure FPLC System and Fraction Collector F9-C (GE Healthcare), HisTrap (5 mL, GE Healthcare) or IMAC Sepharose 6 Fast Flow resin (GE Healthcare) in gravity flow column (Econo-Pac, BioRad). Size Exclusion Purification Column (HiLoad 16/600 Superdex 75 pg - S75 120 mL or 300 mL, GE Healthcare). NanoDrop ND-1000 UV/Vis spectrophotometer (Thermo Scientific). **Software:** Unicorn 6.4.1

Expression: Rosetta2 or BL21 (DE3) *E. coli* cells were transformed with the desired KDM7 plasmid. A 100 mL starter culture was inoculated and grown with appropriate antibiotics for 16 hours. Every 1 L of 2TY expression media was inoculated with 10 mL of the overnight starter culture and grown at 37 °C with appropriate antibiotics at 180 RPM until an $OD_{600\,nm}$ of 0.6 to 0.8 was reached. Expression was induced with IPTG (Isopropyl β-D-1-thiogalactopyranoside)

(0.1 to 0.25 mM final) and $ZnCl_2$ added (10 µM final) and continued for 16 hours at 18 °C. Cells were harvested by centrifugation (4,000 g) and pellets stored at -80 °C.

Purification:

Binding Buffer: (25 mM HEPES, 500 mM NaCl, 5 mM imidazole, pH 7.4, 1 mM TCEP) with DNaseI and Protease Inhibitor Cocktail Tablet (Sigma Aldrich).

Elution Buffer: (25 mM HEPES, 500 mM NaCl, 250 mM imidazole, pH 7.4, 1 mM TCEP)

Gel Filtration (GF) buffer: (25 mM HEPES, 200 mM NaCl, pH 7.4, 1 mM TCEP)

Buffers were prepared with MilliQ and filtered (0.2 µm) prior to use, and stored at 4 °C. The S75 120 mL GF column was washed in 0.1 M NaOH, then MilliQ, then GF buffer prior to use. All steps were performed at 4 °C unless otherwise stated.

For ~10 g of frozen cell pellet: the cell pellet was re-suspended in 60 mL of binding buffer with vigorous magnetic stirring at 4 °C for 30 to 60 minutes until a homogenous light brown suspension was observed. Cells were lysed by sonication for 8 mins total time, 3s on 4s off, (60 % AMP, Sonics VibraCell). Cell debris was separated from lysate by centrifugation (10,000 g for 20 mins at 18 °C) and the recovered lysate was filtered (0.45 µm). For HisTag proteins, either manual or automated affinity chromatography was used.

Manual HisTrap: Resin (IMAC Sepharose 6 Fast Flow, GE Healthcare) was prepared and charged with Nickel according to the manufacturer's instructions. The lysate was applied to 2 mL of resin and washed 4 times with 15 mL of binding buffer, with a final wash of binding buffer with 40 mM imidazole. The target protein was eluted in 4x 4 mL of elution buffer (each elution mixed with resin and incubated for 10 minutes) and concentrated to 2 mL with a 10 kDa molecular-weight cut-off (MWCO) centrifugal concentrator (Amicon Ultra, Millipore).

Automated HisTrap: (Performed on a ÄKTA Pure FPLC System with automated fraction collection). Sample was applied to a 5 mL HisTrap HP (GE Healthcare), washed with binding buffer and eluted with increasing gradient of elution buffer with automatic fraction collection and concentration of those containing target protein as determined by SDS PAGE gel.

Optional TEV Cleavage of Purification Tag: Tobacco etch virus (TEV) protease (donated by Dr Ilaria Pettinati, Schofield Group) was added to the concentrated protein elution fraction in a 10:1 ratio of Protein:TEV (mg) and incubated at 4 °C for 16 hours. Cleavage was assessed by SDS-PAGE gel. The cleaved protein solution was buffer exchanged into binding buffer using a PD-10 column (GE Healthcare) and re-applied to Ni-charged IMAC-Resin or a HisTrap to bind and remove any un-cleaved protein and TEV protease. The target protein was eluted in the binding buffer and cleavage of the purification tag was assessed by SDS-PAGE. All proteins (except $KDM7B_{[PHD+JmjC]}$-biotin, which was deemed > 90 % purity post-rebind) were further purified by SEC.

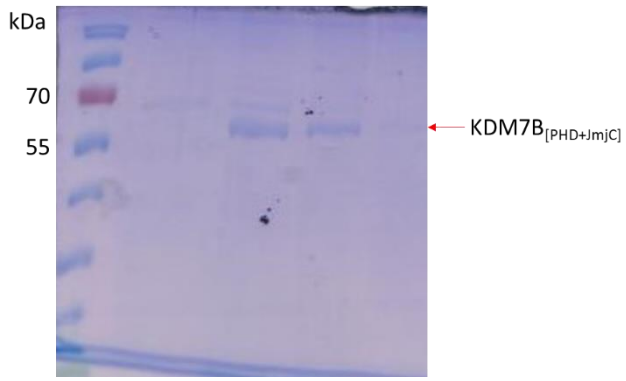
Size Exclusion Chromatography (SEC) Purification: The protein sample was applied to the gel filtration column in a 5 mL loop and run in GF buffer. Fractions (4 mL) were automatically collected and then analysed by SDS-PAGE gel. Instant Blue staining was performed for 15 minutes and a visual assessment used to ascertain fractions containing the target band (> 90%) for pooling. The pooled fractions were centrifugally concentrated (3 kDa cut-off for cleaved PHD finger, 10 kDa otherwise), and the final protein concentration was estimated by an average of three NanoDrop readings. Concentrated protein was aliquoted on ice into PCR tubes that were immediately snap frozen in liquid nitrogen for long-term storage at -80 °C.

10.11.2. KDM7B$_{[PHD+JmjC]}$-Biotin

His-Thioredoxin-KDM7B$_{[PHD+JmjC]}$ (1-447) in pNIC-Bio3 was donated by Finn Wolfreys, TDI. This construct was transformed into pBirA-containing Rosetta *E.coli* cells. A single colony was used to inoculate a starter culture in 2TY supplemented with chloramphenicol (34 µg/mL), spectinomycin (50 µg/mL) and kanamycin (30 µg/mL), and grown overnight (37 °C, 160 rpm). The starter culture (1 mL) was diluted 1000-fold in 1 L LB supplemented with spectinomycin (50 µg/mL) and kanamycin (30 µg/mL) and grown (37 °C, 160 rpm). The expression of KDM7B and BirA (biotin ligase) was induced with 0.5 mM IPTG at OD600nm ~1.00 and the media was further supplemented with 0.5 mM biotin solution. Cells were further grown for 18 hrs (20 °C, 160 rpm) and harvested (4,000 g at 4 °C). Cell pellets were stored at -80 °C.

Cell pellet was thawed and re-suspended in 100 mL binding buffer (500 mM NaCl, 5% Glycerol, 50 mM HEPES pH 7.5, 5 mM Imidazole) containing DNaseI and PMSF, 1 mM $MgCl_2$, and lysed by sonication. Cell debris was removed by centrifugation (10,000 g, 4 °C, 35 min) and the supernatant was filtered (0.45 µm) and loaded onto His-Trap column (CV = 5mL) on AKTA Purifier (GE Healthcare). The column was washed with binding buffer (3 CV) and further washed with wash buffer (500 mM NaCl, 5% Glycerol, 50 mM HEPES pH 7.5, 30 mM Imidazole) (3 CV). His-tagged KDM7B was eluted with increasing imidazole gradient using elution buffer (500 mM NaCl, 5% Glycerol, 50 mM HEPES pH 7.5, 250 mM Imidazole). His-KDM7B containing fractions were analysed by SDS-PAGE and pooled and concentrated to 14 mg/mL. The concentrated protein was then treated with TEV protease (at 50:1 protein:TEV weight/weight) at 4 °C overnight for cleavage of the thioredoxin and HisTag. HisTag cleavage was confirmed by SDS-PAGE and the cleaved product was reapplied to the His-Trap column to remove any uncleaved KDM7B and His-tagged TEV protease. The unbound flowthrough containing the cleaved KDM7B was collected and buffer exchanged into gel filtration buffer

(10 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol) and concentrated to 2 mg/mL. Protein aliquots were snap frozen in liquid nitrogen for long term storage. Biotinylation was confirmed by mass spectrometry (Molecular weight = 54237, Observed = 54240). Concentrations were measured with NanoDrop and estimated extinction coefficients, ε, (ExPASY ProtParam) (uncleaved ε = 91300, cleaved ε = 73000).



F.10.11.2.i. *SDS-PAGE gel showing the purified fractions of KDM7B$_{[PHD+JmjC]}$-biotin.*

10.11.3. $^{13}C,^{15}N$-Labelled KDM7B$_{[PHD]}$

Isotopically labelled reagents ($^{15}NH_4Cl$ and $^{13}C_6$-glucose) were purchased from CK Isotopes. For buffer recipes, see General KDM7 protocol. Rosetta2 *E. coli* cells used for expression.

A 2 L growth was performed according to the following labelled minimal media recipe:

Reagent	g (per L of growth)		**Separately, combine:**	volume (per L of growth)
Na_2HPO_4	6		1 M $MgSO_4.7H_2O$	500 µL
KH_2PO_4	3		20 % w/v $^{13}C_6$-glucose	20 mL
$^{15}NH_4Cl$	1		0.5 % w/v vitamin B1	50 µL
NaCl	0.5		Filter (0.22 um)	
$CaCl_2$	0.003			
MilliQ	up to 1 L			
Autoclave the growth solution				
Add filtered solution of Mg, $^{13}C_6$-glucose, B1				

The autoclaved solution was prepared the day before growth, and the separate glucose-containing solution prepared immediately before use on the day.

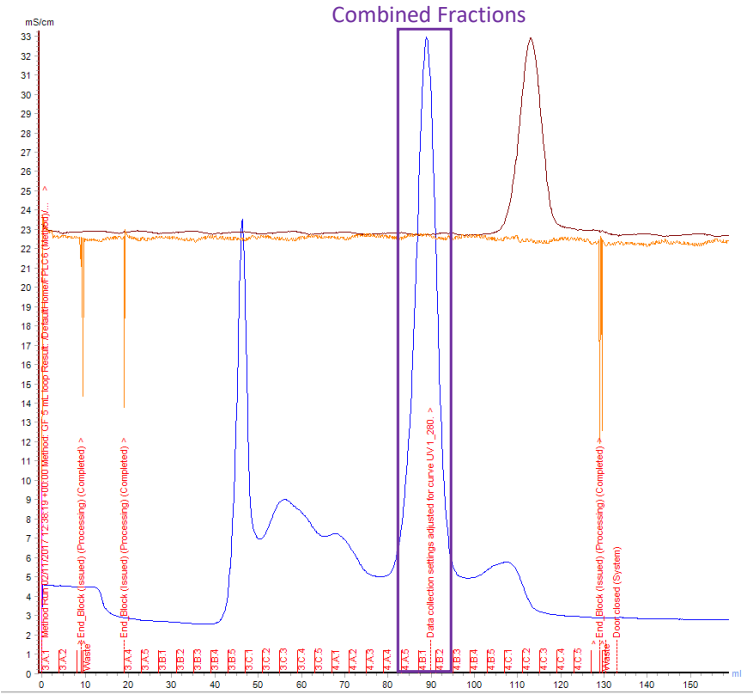
Each 1 L solution for protein production was inoculated with 12 mL from an overnight starter culture (pNHTrxT-KDM7B$_{[PHD]}$ in 100 mL unlabelled 2TY media with kanamycin) and kanamycin and grown at 37 °C with shaking (180 rpm) for 5 hours until $OD_{600\ nm}$ ~0.6 to 0.8. Expression was initiated with IPTG (0.25 mM final concentration), and $ZnCl_2$ was added (final concentration 5 µM) and growth proceeded at 20 °C for 16 hours with shaking (180 rpm).

Cells were harvested by centrifugation (4,000 g) and the collected cell pellet (7.7g wet) was frozen at -80 °C whilst the IMAC resin was prepared. 2 mL of loose IMAC resin was prepared and charged with nickel according to manufacturer's instructions. The cell pellet was resuspended in 50 mL of binding buffer over 30 mins. Re-suspended cells were lysed by sonication (3s on, 4s off for ~12 mins total time, 60% AMP, Sonics VibraCell) and the lysate was separated from debris by centrifugation (15,000 g for 20 mins at 18 °C). All further purification steps were conducted at 4 °C. The recovered lysate was filtered (0.45 um) and loaded onto the gravity column containing pre-prepared Ni-charged IMAC resin. The resin was washed with 40 mL of binding buffer and then 20 mL of binding buffer with 10 mM imidazole. The protein was eluted with 3x 5 mL of elution buffer (each elution was incubated for 10 mins, disturbing the resin each time). Elution fractions were combined and TEV protease was added (~10:1 Protein:TEV mg) for 16 hours. Cleavage of the purification tag was confirmed by SDS-PAGE. The protein was buffer exchanged into binding buffer using a PD-10 column (GE Healthcare) before re-application to freshly regenerated and Ni-charged IMAC resin. The solution was incubated with the resin for 15 minutes, with regular agitation until final settling. The buffer containing the desired protein was eluted and collected, with an additional elution from the resin of 3 mL of binding buffer. The protein was concentrated (3 kDa MWCO, Amicon Ultra, Millipore) to a volume of 2.5 mL and loaded onto an S75 120 mL size exclusion column using a 5 mL loop. Protein was eluted at 1 mL/min, appearing around

80 minutes. SDS-PAGE gel was used to confirm KDM7B$_{[PHD]}$ containing fractions and they were combined and concentrated down to 0.5 mL. A G-25 gravity column (GE Healthcare) was used to buffer-exchange the final solution into NMR buffer (25 mM phosphate pH 6.3, 100 mM NaCl), before final centrifugal concentration (4 mL, 3 kDa MWCO, Amicon Ultra, Millipore) to 0.5 mL to give a final yield of 2.29 mg/mL (~282 µM). The solution was aliquoted (50 µL) and flash frozen in liquid nitrogen for long-term storage at -80 °C.

Protein concentration was determined using NanoDrop (ExPASy estimated extinction co-efficient for cleaved protein: 9970).

Labelling was confirmed by protein mass spectrometry: Molecular Weight (fully isotopically labelled, TEV cleaved): 8153. Observed [M+H]: 8140 (99.8% labelled).



F.10.11.3. *Gel Filtration Chromatogram for the SEC purification of the desired $^{13}C,^{15}N$-Labelled KDM7B$_{[PHD]}$. The blue line represents the UV absorbance. The brown line indicates conductivity.*

10.11.4. KDM7B$_{[PHD]}$

The cell pellet containing pNH-TrxT-KDM7B$_{[PHD]}$ expressed in BL21(DE3) *E. coli* cells was donated by Yijia Zhang. Expression and purification were performed according to the General Protocol (manual, uncleaved).

10.11.5. KDM7B$_{[PHD+JmjC]}$

pNH-TrxT-KDM7B$_{[PHD+JmjC]}$ was transformed into BL21(DE3) *E. coli* cells. Starter cultures were grown in 100 mL 2TY with kanamycin and chloramphenicol. Expression was performed on a 2 L scale growth in 2TY with kanamycin, with a 10 mL starter culture inoculation, grown at 37 °C to an $OD_{600\ nm}$ ~ 0.8 before induction with IPTG (0.1 mM final), with growth continued for 16 hours at 18 °C. Cell pellets were harvested by centrifugation (4,000 g) and stored at -80 °C. The purification followed the General Protocol (manual, cleaved).

10.11.6. KDM7A$_{[PHD+JmjC]}$

pNH-TrxT-KDM7A$_{[PHD+JmjC]}$ was transformed into Rosetta2 *E. coli* cells. Starter cultures were grown in 100 mL 2TY with kanamycin and chloramphenicol. Expression was performed on a 2 L scale growth in 2TY with kanamycin, with a 10 mL starter culture inoculation, grown at 37 °C to an $OD_{600\ nm}$ ~ 1 before induction with IPTG (0.25 mM final) and addition of $ZnCl_2$ (10 µM final), with growth continued for 16 hours at 18 °C. Cell pellets were harvested by centrifugation (4,000 g) and stored at -80 °C. Purification was assisted by Yijia Zhang, and performed according to the General Protocol (automated, cleaved and uncleaved).

10.11.7. DIDO1$_{[PHD]}$

pNH-TrxT-DIDO1$_{[PHD]}$ was transformed into Rosetta2 *E. coli* cells. Starter cultures were grown in 100 mL 2TY with kanamycin and chloramphenicol. Expression was performed on a 2 L scale growth in 2TY with kanamycin, with a 10 mL starter culture inoculation, grown at 37 °C to an $OD_{600\ nm}$ ~ 1 before induction with IPTG (0.25 mM final) and addition of $ZnCl_2$ (10 µM final),

with growth continued for 16 hours at 18 °C. Cell pellets were harvested by centrifugation (4,000 g) and stored at -80 °C. The purification followed the General Protocol (manual, uncleaved).

10.11.8. AIRE$_{[PHD1]}$

Rosetta pBirA cells were transformed with pNIC-Bio3-AIRE$_{[PHD1]}$. A 3 L scale grwoth was performed in 2TY media inoculated with 10 mL of starter culture and kanamycin and spectinomycin antibiotics. Growth was performed at 37 °C until $OD_{600\,nm}$ = 0.55. Induction was performed with IPTG (1 mM final) and addition of biotin (1 mM final) and $ZnCl_2$ (10 µM final) and expression continued at 18^{o}C for 16 hours. Cell pellets were harvested by centrifugation (4,000 g) and stored at -80 °C. Buffers used for purification were: Binding Buffer (25 mM HEPES pH 7.5, 500 mM NaCl, 5 mM imidazole); Elution Buffer (25 mM HEPES pH 7.5, 500 mM NaCl, 500 mM imidazole); Gel Filtration Buffer (25 mM HEPES pH 7.5, 200 mM NaCl, 5% glycerol). Cells were re-suspended in binding buffer (with added DNaseI and Protease inhibitor cocktail tablet) and biotin (1 mM final) was added. Purification was then carried out according to the General Protocol (automated, uncleaved) with the specified buffer modifications.

10.12. Protein Mass Spectrometry

Low resolution positive ion electrospray ionisation mass spectrometry was used to analyse intact protein masses using the instrumental set-ups as below:

1) LCT Premier XE ionisation (Micromass) mass spectrometer connected to an Acquity Ultra Performance liquid chromatography (UPLC, Waters) system using an Acquity UPLCR BEH300 C18 column at 50 °C. The protein (~ 0.1 mg/mL, 0.5 µL) was injected and eluted using a gradient system from Solvent A (95 % water, 0.1 % (v/v) formic acid) to Solvent B (95 % acetonitrile, 0.1 % (v/v) formic acid):

Time / min	% Solvent A	% Solvent B
0	95	5
8	5	95
8.1	95	5
10	95	5

Parameters: Capillary 3,000 V; Sample cone 35 V; Desolvation 250 °C; Source 80 °C; Cone gas flow 100 L/h; Desolvation gas flow (N_2) 400 L/h. Sodium formate was used as the internal calibrant.

OR

2) Xevo GS-XS QToF mass spectrometer connected to an Acquity Ultra Performance liquid chromatography (UPLC, Waters) system using a ProSwift RP-1S Analytical Phenyl column (Thermo Fisher). The protein (~ 1 µM, 5 µL) was injected and eluted using a gradient system from Solvent A (100 % water, 0.1 % (v/v) formic acid) to Solvent B (100 % acetonitrile, 0.1 % (v/v) formic acid):

Time / min	% Solvent A	% Solvent B
0	95	5
1	95	5
7	5	95
8	5	95
8.1	95	5
10	95	5

Parameters: Source 3,000 V; Cone 30 V; Source 100 °C; Desolvation 350 °C; Cone gas flow 30 L/h; Desolvation gas flow (N_2) 600 L/h. A standard solution of Lecuine EnK (4 µg/mL) was used as the internal lock-spray calibrant.

All spectra were processed using MassLynx v4.0 and v4.1 (Waters) with the Maximum Entropy method (MaxEnt1). A table of acquired data for produced proteins is given in the Appendix.

10.13. Protein Crystallography Trials

The proteins KDM7C$_{[PHD+JmjC]}$ and KDM7B$_{[PHD]}$ used for crystallography trials were donated by Yijia Zhang.

Concentrated KDM7B$_{[PHD+JmjC]}$ protein (246 µM, pNH-TrxT expression with purification tags cleaved) was thawed slowly on ice. Concentrated stocks of $MnCl_2$ and N-oxalylglycine (NOG) were prepared in buffer (20 mM HEPES, pH 7.4, 200 mM NaCl, 1 mM TCEP). A concentrated stock of OC9 was prepared in 50 mM HEPES, pH 7.4, 200 mM NaCl, 1 mM TCEP. Protein (optionally first added to Mn^{II} and NOG), was added to OC9. Final concentrations in the mixture: KDM7B (157 µM); Mn^{II} (450 µM); NOG (450 µM); OC9 (301 µM).

A similar approach was used in the trial of tag-cleaved KDM7C$_{[PHD+JmjC]}$: KDM7C (228 µM); Mn^{II} (500 µM); NOG (1 mM); OC9 (500 µM, from DMSO stock, 0.4 % DMSO final)

A similar approach was also used for tag-cleaved KDM7B$_{[PHD]}$: KDM7B (420 µM); OC9 (420 µM).

Solutions were dispensed to plates (Crystalquick plates, 96-well, Greiner 609171) with the three-drop method using a Crystal Phoenix (Art Robbins). Plates were incubated at either 4 °C (KDM7B) or 20 °C (KDM7C) and visually inspected by microscope for crystal formation daily for the first week, weekly thereafter over 3 months, and then once a month for 6 months.

10.14. Preparation of Materials for IVTT and mRNA Display Selection

DNA templates for flexizyme, microhelix and tRNA were prepared by multi-stage PCR according to the published protocol by Goto et al [228]. RNA Transcription reactions (1 mL) were performed for 16 hours at 37 °C and contained the following final concentrations of reagents (KOH was to ensure pH 8-9, by pH paper):

Microhelix and tRNAs: 1x T7 RNA polymerase buffer; 10 mM DTT; 22.5 mM $MgCl_2$; 3.75 mM NTPs; 5 mM GMP; 30 ng/µL DNA template; 2 U/µL T7 RNA Polymerase, 0.024 M KOH

Flexizyme: 1x T7 RNA polymerase buffer; 10 mM DTT; 30 mM $MgCl_2$; 5 mM NTPs; 30 ng/µL DNA template; 2 U/µL T7 RNA Polymerase, 0.03 M KOH

tRNA Type		5' to 3' Sequence
RNA	$fMet_{ATG}$	GGCGGGGUGGAGCAGCCUGGUAGCUCGUCGGG**CUCAUAA**CCCGAAG AUCGUCGGUUCAAAUCCGGCCCCCGCAACCA
DNA	$fMet_{ATG}$	GGCGTAATACGACTCACTATAGGCGGGGTGGAGCAGCCTGGTAGCT CGTCGGG**CTCATAA**CCCGAAGATCGTCGGTTCAAATCCGGCCCCCG CAACCA
RNA	Glu_{ATG}	GUCCCCUUCGUCUAGAGGCCCAGGACACCGCC**CUCAUAA**GGCGGUA ACAGGGGUUCGAAUCCCCUAGGGGACGCC
DNA	Glu_{ATG}	GGCGTAATACGACTCACTATAGTCCCCTTCGTCTAGAGGCCCAGGA CACCGCC**CTCATAA**GGCGGTAACAGGGGTTCGAATCCCCTAGGGGA CGCCA

Table of the DNA and RNA sequences of the reprogrammable tRNAs by flexizyme acylation. The tRNA **anti-codon** is given in bold+underline and the adjacent **wobble base pairs** are given in bold. Note the CAU anti-codon reads the ATG codon.

10.14.1. Acylation of tRNA/Microhelix

Microhelix timecourses and tRNA acylation reactions were performed according to the protocol of Goto et al [228]. For acylation of 500 pmol worth of tRNA/microhelix: 2 µL of 0.5 M HEPES (pH 7.5) (0.22 µm filtered) was added to a 0.5 mL Eppendorf tube, followed by either tRNA or microhelix RNA (500 pmol worth in H_2O) and further H_2O was added to ensure the final mixture would reach 20 µL. The mixture was heated at 95 °C for 2 minutes, allowed to cool at room temperature for 5 minutes, then 4 µL of 3 M $MgCl_2$ was added and mixed. The

solution was incubated at room temperature for 5 minutes, then on ice for 5 minutes. Finally, 4 µL of activated amino acid (in DMSO) was added and the solution was mixed by pipette. The reaction was incubated on ice for the optimised acylation time. The reaction was quenched with 0.1 M NaOAc (pH 5) in 70% EtOH (280 µL) and centrifuged at 15k g for 15 mins. Supernatant was removed, then 100 µL of quench solution was further added to the pellet and the pellet was re-suspended and then centrifuged at 15k g for 5 mins. The supernatant was removed, and the wash step repeated. The supernatant was removed and 70% EtOH (100 µL) was added and then the sample centrifuged at 15k g for 3 mins. The supernatant was removed and the pellet was air dried (~25 mins). Each batch of acylated tRNA was verified by IVTT and MALDI-TOF-MS.

10.15. In-Vitro Transcription/Translation (IVTT) Reactions

Components for the IVTT reaction were mixed together in a low-binding tube on ice, then incubated at 37 °C for at least 2 hours. Typically, 500 pmol of each pre-acylated tRNA was sufficient for one reaction. IVTT reactions were carried out using a PURExpress kit (NEB) on a 5 µL scale as per the manufacturer's instructions, with 30 ng of DNA template, a customised amino acid mixture and pre-acylated tRNA (dissolved in 1 mM NaOAc, pH 5, immediately prior to use).

Crude IVTT reaction mixtures were purified by C_{18} (for Cyclic Peptides) or SCX (for Histone Peptides) ZipTip (Millipore) according to the manufacturer's instructions. The elute from this purification was spotted on a steel MALDI-plate and analysed by MALDI-TOF-MS.

10.16. Magnetic Bead Loading Optimisation for Selection

Loading Buffer: 20 mM phosphate, pH 7.2, 150 mM NaCl, 2 mM BME, 50 µM $ZnCl_2$, 0.1% BSA

0.6 mL low-binding tubes (StarLab) were used for all steps.

Dynabeads M-280 Streptavidin (Invitrogen) were pelleted in an Eppendorf tube by application of a magnet and washed by re-suspension in loading buffer. The beads were pelleted and the loading buffer was removed, and then this wash step was repeated a further time, before final re-suspension in loading buffer at the original volume of beads. Beads were then aliquoted to 0, 6, 12.5, 25, 50 µL portions. The target protein to be loaded onto the beads was diluted to 2.5 µM in the loading buffer and 20 µL (50 pmol) was added to each tube. Tubes were incubated with rotation at 4 °C for 30 minutes, then 1 µL of 500 µM biotin was added for 5 minutes. Beads were then pelleted by magnet and the supernatants recovered. The beads were then washed twice by pelleting and re-suspension in loading buffer and finally re-suspended in 20 µL loading buffer. 4xNuPAGE loading dye (6.66 µL) was added to each sample (either recovered supernatant or final bead re-suspension) and each sample was heated at 95 °C for 5 minutes. The bead samples were pelleted by magnet whilst hot and their supernatant recovered as the bead sample. 10 µL of each sample (original supernatant or recovered from bead) was analysed by SDS-PAGE (12.5% Bis-Tris NuPAGE gel, 160 V, 45 mins, instant blue). Lanes that corresponded to the minimum bead volume showing protein in both 'bead' and 'supernatant' samples indicated the minimum volume of beads required for saturation with the biotinylated protein. Result found for $KDM7B_{[PHD+JmjC]}$-biotin: 2 pmol protein per 1 µL beads.

10.17. Procedure for mRNA Display Selection on KDM7B$_{[PHD+JmjC]}$

10.17.1. mRNA Ligation

The final mRNA ligation reaction contained 1X T4 RNA ligase buffer, cyclic peptide library mRNA at 10 µM, 15 µM puromycin-linker, 0.2 U of RNase inhibitor, 4.7 U of T4 RNA ligase and MilliQ H_2O was added to achieve a final volume of 7.5 µL. The reaction was assembled in a PCR tube and allowed to react at 20 °C for 1 hour, then heat inactivated at 70 °C for 10 minutes.

10.17.2. In-vitro Translation

All translation components were kept on ice during preparation and assembled in a low-binding tube. Pre-acylated tRNA (1x 500 pmol ClAcTyr) was dissolved in 0.8 µL of 1 mM NaOAc (pH 5.2), and centrifuged at 15k g for 20 seconds immediately before adding to translation mix (5 µL final volume). The translation kit (Sol A, Sol B) was from a custom order version of a PURExpress kit with ΔAAs and ΔRF1 (New England Biolabs). Reaction was usually carried out in 5 µL final volume, with RNA inhibitor (0.8 U), amino acids (ΔMet, 0.3mM) and mRNA-puromycin ligated product (2.5 µM). The translation reaction was carried out at 37 °C for 1 h, then incubated at 60 °C for 10 minutes to dissociate the ribosomal machinery and aid cyclisation.

10.17.3. Reverse Transcription (RT)

Reverse transcription was carried out using M-MLV RTase, using Primer HS_P2. The reaction (10 µL) contained 53 U M-MLV RTase, 3X M-MLV buffer, 0.15 mM dNTPs and input was the translation reaction. The reaction was incubated at 42 °C for 1 hour. On completion, the RT mixture was diluted with 10 µL of 2X selection buffer and 2 µL H_2O (total vol = 20 µL). 0.25 µL from the diluted RT mixture was taken and further diluted to 1000 µL in H_2O (Sample A).

Selection buffer: PBS, 0.01% BSA, 0.05% Tween20, 2 mM beta-mercaptoethanol (BME).

10.17.4. Magnetic Bead Preparation

All steps used low-binding tubes to minimise protein/peptide binding.

Strep-Mag (50 µL): Beads from 50 µL Dynabeads-Streptavidin M280 were pelted by applying a magnet, then the supernatant was removed. The beads were further washed twice with 5 volumes of 1X selection buffer and further re-suspended with 50 µL of 1X selection buffer.

Biotin-Strep-Mag (12 µL): 1.2 µL of 1 mM biotin (f.c. 100 µM) was added to the washed beads (12 µL), mixed and incubated at 4 °C for 10 mins with rotation. Beads were pelleted (magnet), supernatant was removed, and the beads were washed twice with 5 volumes of 1X selection buffer. This was further re-suspended in 12 µL 1X selection buffer.

Equal volumes of Strep-Mag and Biotin-Strep-Mag beads (12 µL/12 µL, Total = 24 µL) were added, then split equally into three separate low-binding tubes, 8 µL each (labelled NS1, NS2, NS3).

10.17.5. Negative Selections (NS)

NS1: RT mix was added to pelleted NS1 beads, re-suspended, and incubated at 4 °C for 30 min with rotation. NS2: Beads were pelleted from NS1 and the supernatant was recovered which was used to re-suspend NS2 beads and incubated at 4 °C for 30 min with rotation. NS3: Beads were pelleted from NS2 and recovered supernatant was used to re-suspend NS3 beads, followed by incubation at 4 °C for 30 min with rotation. A small aliquot (0.25 µL) was taken from the supernatant and diluted to 1000 µL in MilliQ H_2O (Sample C). The supernatant from NS3 contained the ligated cyclic peptide library and was used in positive selection.

10.17.6. Loading of Magnetic Beads with the Target Protein

All steps were performed at 4 °C. 10 µL of KDM7B$_{[PHD+JmjC]}$-biotin (36 µM) (thawed on ice) was taken and diluted with an equal volume of 1X selection buffer (with added 10 µM $ZnCl_2$). 5 µL of this diluted solution was taken and further diluted with 1X selection buffer (no added $ZnCl_2$) to a final protein concentration of 1 µM. 10 µL of washed Strep-Mag beads* were added to a low-binding tube, the supernatant was removed and the beads were re-suspended in 20 µL of the 1 µM solution of KDM7B$_{[PHD+JmjC]}$-biotin. The beads were incubated at 4 °C for 30 min with rotation, then beads were pelleted and washed five times with 6 volumes of ice-cold 1X selection buffer.

*(Bead loading was previously determined to be 2 pmol protein per µL bead slurry. The final volume of the cyclic peptide library-containing solution was 20 µL. Therefore, 20 pmol of KDM7B$_{[PHD+JmjC]}$-biotin was added to achieve a final concentration of 1 µM target protein during selection).

10.17.7. Positive Selection and DNA Recovery

The beads loaded with the target KDM7B protein were pelleted by magnet, and the supernatant removed, immediately prior to re-suspension in the solution containing the cyclic peptide library (supernatant from NS3). The positive selection was then incubated at 4 °C for 30 min with rotation. The beads were then pelleted, and the supernatant was removed before re-suspension of the beads in 100 µL of 1X selection buffer. The re-suspended beads were transferred to a new low-binding tube and the supernatant was removed. The re-suspension and transfer processes were repeated twice more. The beads from the positive selection were then pelleted and the supernatant removed before final re-suspension of the beads in 100 µL of PCR mix. The final concentrations of components in the PCR mix were: 1X NH_4 buffer, 2.5

mM $MgCl_2$, 0.025 mM dNTPs, 1 µM HS_P1, 1 µM HS_P2, topped up with additional MilliQ H_2O.

The beads that were suspended in the PCR mix were then heated at 95 °C for 5 mins and then immediately pelleted by application of a magnet, in order to recover the supernatant containing the liberated DNA which encoded for the cyclic peptide library.

10.17.8. Quantitative PCR (qPCR)

Equipment: C1000 Thermal Cycler (BioRad). **Software:** Bio-Rad CFX Manager 3.1

qPCR was performed to quantitate the amount of DNA present in Sample A, Sample C and the recovered library DNA from the selection. A qPCR master mix was made containing all necessary reagents except for the unknown sample. The qPCR master mix contained: 1.05X SYBR GREEN, 0.26 µM of HS_P1 and 0.26 µM of HS_P2, diluted in MilliQ H_2O. 19 µL of qPCR master mix and 1 µL of recovered library or a known standard were added to each sample well in the qPCR plate. Standards were used in addition to unknown samples to enable quantitation of the DNA and were generated by reverse transcription of a known quantity of purified library RNA, assuming a 1:1 transcription ratio. All qPCR samples were run in triplicate. The program used for qPCR was: 95 °C (3 min); [95 °C (40 s), 61 °C (40 s), 72 °C (40 s), plate read] x40.

DNA was quantitation was calculated with respect to the standards, using them to calculate a mean starting quantity number of DNA molecules in each qPCR sample, which was then multiplied by its respective dilution factors and starting volume. The percentage DNA recovery for the positive selection was then calculated by taking the total number of DNA molecules in the positive selection sample and dividing it by the total number of DNA molecules present in Sample C, then multiplying that number by 100 to get a percentage.

10.17.9. PCR Library Amplification

The recovered library from the positive selection was amplified by PCR (1 µL of BioTaq polymerase was added per 100 µL of solution). The PCR product was purified by phenol-chloroform extraction and analysed on a 3% agarose gel (150 V for 30 mins). The concentration of the purified DNA was estimated by NanoDrop. The PCR amplification program used was: 94 °C 5 min; [94 °C 40s, 61 °C 40s, 72 °C 40s] x n cycles, where n is determined by Cq value (no greater than 28).

10.17.10. PCR and qPCR Primers

HS_P1: 5'-TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATG-3'

HS_P2: 5'- TTTCCGCCCCCCGTCCTAGCTGCCGCTGCCGCTGCCGCA-3'

Peptide Library: 5'-GCTGCCGCTGCCGCTGCCGCMNNCATATGTATATCTCCTTCTTAAAG-3' – note this is in the reverse direction to translation, MNN becomes NNK after PCR amplification with HS_P1 and HS_P2. The number of MNN (NNK) codons can be varied. Puromycin Linker: 5'-Puromycin-CC-Linker-CCTGCCCCCCGCCCTC-3'

10.17.11. DNA Transcription to mRNA

The amplified DNA from the positive selection was converted into mRNA for use as the input library in the next round of selection. The transcription reaction was set-up to contain: 1X T7 buffer, 10 mM DTT, 20 mM $MgCl_2$, 3.75 mM NTPs, 0.02 M KOH, 20 U of RNase inhibitor, 2 U of T7 RNA polymerase and DNA template (PCR product) at 150 ng/µL. The reaction was topped up with MilliQ H_2O to a final volume of 20 µL, the pH was checked to ensure 8 < pH < 10, and the reaction was incubated at 37 °C for 16 hours. After 16 hours, 2 µL of 100mM $MnCl_2$ and 0.2 µL of DNaseI were added to the reaction and it was incubated at 37 °C for 30 min. The reaction was then finally quenched with an equal volume of a solution containing 600 mM

NaCl, 10mM EDTA. The mRNA was purified by phenol chloroform extraction, before final dissolution of the pellet in MilliQ H_2O.

10.17.12. Next Generation Sequencing and Data Analysis

To construct an Illumina-compatible sequencing library from the DNA sequences recovered after selection, specific forward (NGS_F) and reverse (NGS_R) primers were designed that contained priming sequences complementary to the constant region of the recovered peptide library DNA at the 3' end of the NGS primer [164]. NGS adaptor sequences for sequencing were placed at the 5' end of the NGS primer. PCR with Q5 polymerase was used to install in the adaptors to a sample of library recovered from positive selection. Each amplified PCR product was quantitated using a high sensitivity DNA tape on the Tapestation 2000 (Agilent) prior to pooling, size-selected (AMPure XP, Beckman) and final quantification of the library was performed by quantitative PCR using pre-diluted standards (Kapa Biosystems). Paired-end sequencing was performed using a NextSeq 500 (Illumina) with 80 reads in each direction.

NGS_F:

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTA
GGGTTAACTTTAAGAAGGAGATATAC-3'

NGS_R:

5'-CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCG
ATCTGTCCTAGCTGCCGCTG-3', where XXXXXX represents the index, which was one of: GCGGAC, TTTCAC, GGCCAC or CGAAAC.

After sequencing, the variable region of each amplicon was extracted by removing the constant region and sequencing adaptors from reads using Cutadapt (version 1.8.1) [370]. The

forward and reverse reads for each variable region were compared and non-identical reads were discarded in order to reduce downstream error due to sequencing errors. The resulting validated variable region sequences were tabulated using R and the resultant peptide determined using the Biostrings package [371].

10.18. DNA Agarose Gel Analysis

For analysis of DNA fragments ≤ 500 bp, a 3% agarose gel was prepared: 1.5 g agarose (Sigma Aldrich) and 50 mL of 1X TBE buffer, added to conical flask and microwaved to boil and dissolve the agarose. 5 µL of SYBR-SAFE dye was added to the cooling solution before pouring into a gel mould. Samples were prepared by addition of the DNA sample to 6X loading dye (Thermo Fisher, R0611) and 10 µL per sample was loaded per well of the gel to be run for 30 minutes at 150 V in TBE buffer.

10.19. Phenol-Chloroform Purification of DNA/RNA

DNA/RNA products were transferred (e.g. 100 µL) into a standard 1.5 mL Eppendorf tube. An equivalent volume (e.g. 100 µL) of phenol/chloroform/isoamylalcohol (25:24:1, Sigma Aldrich) was added to the sample and it was mixed by vigorous shaking, then centrifuged at ≥15,000 g for 5 min. The top layer (aqueous layer) was recovered by pipette and added to a new pre-labelled Eppendorf tube. (Note: protein could precipitate between the two layers, so care was taken to ensure the pipette tip did not touch the interface). The same volume (e.g. 100 µL) of chloroform/isoamylalcohol (24:1, Sigma Aldrich) was added and mixed by vigorous shaking, then centrifuged at ≥15,000 g for 5 min. The top layer was recovered into a new tube. The volume of the extract was determined using a pipette. If no NaCl had been added previously *, then 3 M NaCl was added at approximately 1/9 of the volume of the sample (e.g. 11 µL). With the correct NaCl concentration, 2x the resultant volume (e.g. add

220 µL) of 100% ethanol was added and mixed well by vortex to precipitate the DNA/RNA, then centrifuged for 15 min at ≥15,000 g and the supernatant was removed. Half the initial volume (e.g. 50 µL) of 70% ethanol was added to wash the pellet by vortexing, then centrifuged for 3 min at ≥15,000 g. The supernatant was carefully removed to leave the pellet intact, which was then air dried for 5-10 min. The DNA/RNA pellet was dissolved in MilliQ H_2O (e.g. 20 µL) for downstream applications. *For small sample starting volumes (< 30 µL) an equal volume of 0.6 M NaCl was added prior to the first phenol/chloroform addition to make the volume more manageable for the phenol/chloroform extraction steps. The same steps were then followed, with the omission of adding further NaCl, as the solution already had the correct salt concentration for DNA/RNA precipitation.

10.20. RNA Urea Acrylamide Gel Analysis

For analysis of tRNA: 10 % TBE Urea PAGE Gel. **For analysis of flexizyme:** 15 % TBE Urea PAGE Gel. **For analysis of acylated microhelix RNA:** 20 % Acid PAGE Gel.

For 1 Gel	TBE Urea Acrylamide					For 1 Gel	Acid PAGE
Reagent	**10%**	**12%**	**15%**	**20%**		**Reagent**	**20%**
Urea (g)	2.1	2.1	2.1	2.1		**Urea (g)**	2.1
5X TBE (µL)	1200	1200	1200	600		**3M NaOAc pH 5.2 (µL)**	100
40% acrylamide (µL)	1500	1800	2250	3000		**40% acrylamide (µL)**	3000
MilliQ (Up to 6 mL)	1800	1500	1050	300		**MilliQ (Up to 6 mL)**	1000
TEMED (µL)	4	4	4	4		**TEMED (µL)**	4
10% APS (µL)	60	60	60	60		**10% APS (µL)**	60

Samples were prepared by addition of the RNA sample to 2X RNA loading dye (ThermoFisher, R0641), or Acid PAGE loading dye, and 10 to 15 µL per sample was loaded to be run for 80 minutes at 200 V in TBE buffer. Acid PAGE loading dye: 150 mM NaOAc pH 5.2, 10 mM EDTA, 93 % formamide, 0.016 % Bromophenol Blue.

10.21. Solid Phase Peptide Synthesis (SPPS): General Protocol

Equipment: Liberty Blue Automated Microwave Peptide Synthesiser with HT12 resin loader (CEM), Lyophilizer (CoolSafe, Scanvac). **Software:** Liberty Blue Application (CEM)

10.21.1. Synthesis

Peptides were synthesised from their C to N terminus on a Liberty Blue machine using a Rink Amide MBHA resin (100-200 mesh, AGTC Bioproducts) with fluorenylmethyloxycarbonyl (Fmoc) amino acids (Sigma Aldrich, Novabiochem, Fluorochem). Resin was weighed out according to the manufacturer loading recommendation for synthesis (0.1 or 0.05 mmol scale). Amino acid masses were calculated with the software and dissolved in DMF to 0.2 M. Amide couplings were performed under standard Liberty Blue microwave conditions for each amino acid, with diisopropylcarbodiimide (DIC, 0.5 M stock, Fluorochem) as activating agent and Oxyma (1 M stock, Novabiochem) as activating base with N,N-diisopropylethylamine additive (DIPEA, 0.1 M stock, Sigma Aldrich) in DMF (Ultra Pure, Cambridge Reagents). Fmoc-deprotection cycles were performed under standard Liberty Blue microwave conditions using 20 % piperidine in DMF (Ultra Pure, Cambridge Reagents). A final Fmoc-deprotection was performed as the last step to leave the free amino group at the N-terminus.

10.21.2. Modification and Deprotection & Cleavage

The resin-bound peptide was suspended in ~5 mL of DMF in a falcon tube. For cyclic peptides, it was treated with 10 equivalents of chloroacetic anhydride (Sigma Aldrich), pre-dissolved in 0.5 mL of DMF, for 2 hours at 20 °C with manual agitation every 30 minutes. A further 5 equivalents were then added similarly for 1 hour. For linear versions of the cyclic peptides, acetic anhydride was used as a capping agent instead.

The resin was subsequently washed in a phase separator (TELOS, 25 mL) under reduced pressure with 10 mL DMF, 3x 20 mL DCM and dried to a free-flowing texture under reduced pressure. Cleavage and removal of side-chain protection groups was performed with ~4 mL of a solution of trifluoroacetic acid (TFA, 92.5%), triisopropylsilane (TIPS, 2.5%), 1,3-dimethoxybenzene (2.5%) and H_2O (2.5%) in a glass vial for 3 hours. Peptide-containing solution was collected through a phase separator into a pre-weighed 50 mL falcon tube and precipitated with addition of ~45 mL cold diethyl ether (Et_2O) and vortexing. Precipitate was pelleted by centrifugation at 3000 g for 5 minutes and washed similarly a further three times with diethyl ether. A small sample of solid could be taken for MALDI-TOF-MS analysis to confirm synthesis and N-terminus modification. The peptide could also be re-dissolved in a small volume of an appropriate ratio of MeCN/H_2O, snap frozen in liquid nitrogen and lyophilized overnight to remove any remaining traces of TFA. The approximate yield of crude peptide could be estimated from falcon tube weight comparison to tare weight.

10.21.3. Cyclisation

Method 1: 10-20 mg of crude peptide was dissolved in ~1 mL MeCN/H_2O. 0.1 M NaOH (aq) was added in small increments (~10 µL) until pH~8-9 was achieved (pH paper). TCEP was added to 1 mM if oxidation was a concern. The solution was heated at 60 °C in a water bath for 30-60 minutes. MALDI-TOF-MS or LCMS was used to confirm cyclisation. The reaction was quenched with a volume of the same ratio of MeCN/H_2O but containing TFA to arrive at 0.1 % TFA final and it was ensured the quenched solution was acidic (pH < 2, by pH paper).

Method 2: 10-20 mg of crude peptide was dissolved in ~1 mL DMSO. Triethylamine (NEt_3, ~30 µL) was added to adjust the pH ~8-9 (by pH paper). TCEP was added to 1 mM if oxidation was a concern. The peptide was incubated with continuous agitation at 37 °C for 60 minutes.

MALDI-TOF-MS or LCMS was used to confirm cyclisation. The reaction was quenched with aqueous TFA to 0.1 % TFA final and it was ensured the quenched solution was acidic (pH < 2, by pH paper).

10.22. HPLC Peptide Purification: General Protocol

Equipment: JASCO HPLC System, Column Phenomenex Gemini-NX 5 µm C18 110 Å 250x30 mm, UV Detection at 220 nm. **Software:** JASCO ChromNAV v1.19.01

HPLC solvents: A = H_2O, 0.1 % (v/v) TFA and B = MeCN, 0.1 % (v/v) TFA.

Following cyclisation, a small sample volume of the quenched reaction was diluted 100-fold into H_2O/MeCN (0.1 % v/v TFA) to ensure the peptide remained soluble. The lowest tolerable ratio of MeCN for a given peptide determined the starting solvent ratio for the HPLC. DMSO reactions were diluted in the starting ratio of solvent such that no more than 25 % (v/v) DMSO was injected. Up to 2 mL of peptide solution was injected at once and purified by increasing MeCN gradient elution, customised for each peptide following a pilot run, over 60 minutes. Product containing fractions were identified by MALDI-TOF-MS, combined according to their chromatographic purity (estimated >90 % peak area), flash frozen in falcon tubes using liquid nitrogen and lyophilized to a white powder. Re-dissolution in MeCN/H_2O and lyophilization into a pre-weighed Eppendorf tube was performed for compact storage and estimation of mass yield. Final characterisation was performed by MALDI-TOF-MS & HR-ESI-MS, with NMR calibration of peptide solution concentration (in DMSO or H_2O-based solutions).

10.23. Histone Peptide Sourcing

All histone H3 [1-21] peptides, including methylated lysine derivatives, were purchased from GL Biochem (Shanghai), received as solid and dissolved to mM stocks in H_2O, or DMSO, with identity verified by MALDI-TOF-MS and concentration verified by NMR. Biotinylated histone peptides were purchased from AnaSpec, with a GGK-biotin unit at the C-terminus of the sequence.

Histone H3 [1-21] Sequence = ARTKQTARKSTGGKAPRKQLA

10.24. MALDI-TOF-MS Spectra Acquisition

Equipment: Bruker MALDI-TOF Microflex LRF, Bruker microScout target plate (ground or polished steel). **Software:** Bruker Compass for flexSeries 1.4: flexControl v3.4 and flexAnalysis v3.4

MALDI-TOF mass spectra were obtained using a Bruker MALDI-TOF Microflex LRF (positive, reflectron mode) by spotting 1 µL of sample onto a 1 µL droplet of α-cyano-4-hydroxycinnamic acid (CHCA) matrix (dissolved in 50 % acetonitrile/50 % H_2O/0.1 % TFA v/v). Plates were dried at 37 °C prior to shooting. For purified peptides, a solution of the purified peptide diluted in MilliQ H_2O was used. For assays, a spot of the crude reaction mixture was used. Spectra were processed by applying default baseline subtraction and smoothing, with automated or manual peak picking as appropriate. Spectra from Bruker flexAnalysis were exported using the built-in software report tool. Alternatively, MALDI-TOF mass spectra were obtained on a Micromass MALDI-TOF micro MX mass spectrometer (flight tube voltage 12000 V, reflectron voltage 5200 V, positive mode) and processed using MASSLYNX 4.0 (Waters). Spectra from Masslynx Waters were generated using the built-in image export tool. Figures with overlaid MALDI-TOF spectra were created using Adobe Illustrator or Inkscape v0.91.

10.25. Cyclic Peptide High-Resolution ESI-MS (HRMS)

High-resolution electrospray ionisation mass spectrometry was used for characterisation of synthesised small molecules and peptides. Samples were prepared by dilution in H_2O/MeCN to < 0.1 mg/mL and run by the University of Oxford Mass Spectrometry Department on a Waters Acquity Ultraperformance LC system connected to a Thermo Orbitrap Exactive MS for electrospray in the regime of direct infusion. ESI capillary voltage 3000 V; ion source 200 °C; transfer capillary 350 °C. Elevated temperatures were used during the electrospray for desolvation of the ions and prevention of non-volatile neutral molecules contaminating the ion source.

10.26. NMR Protocols

10.26.1. Calibration of Peptide Concentration

Equipment: Bruker 500 MHz Avance II with BBO Probe or Avance III with TBO probe. **Software:** Bruker TopSpin 3.5.

Standard NMR tube: (Borosilicate Glass, OD = 4.97 mm, ID = 4.2 mm, Length = 178 mm, Concentricity = 0.020mm, Camber = 0.070mm (Norrell, Fisher, # 12800056)). Smaller NMR tube: (Borosilicate Glass, OD = 3.0 mm, ID = 2.4 mm, Length = 100 mm (Hilgenberg, #2001724)).

TSP Standard: 3-(Trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt (CAS = 24493-21-8, Sigma Aldrich) with ^1H chemical shift = 0 ppm.

Sample Preparation: NMR samples were prepared with 2.5 µL of peptide (stock estimated by mass/volume at minimum 5 mM) and 2.5 µL of TSP standard (10 mg/mL in D_2O) added to 450 µL of NMR solvent in a standard NMR tube or 160 µL in a small NMR tube. If peptide stock

was in DMSO then NMR solvent = DMSO-d_6, if peptide stock was in H_2O then NMR solvent = D_2O.

D_2O Experiments: ^{1}H-NMR, Number of Scans (NS) = 16, Dummy Scans (DS) = 2, TD (Size of FID) = 65536, Spectral Width (SW) = 12 ppm, O1P = 4.7 ppm, D1 (Relaxation Delay) = 30 sec

DMSO-d_6 Experiments: ^{1}H-NMR, Number of Scans (NS) = 16, Dummy Scans (DS) = 2, TD (Size of FID) = 131072, Spectral Width (SW) = 24 ppm, O1P = 2 ppm, D1 (Relaxation Delay) = 30 sec

Processing: The extended relaxation delay was to ensure full T_1 relaxation for accurate integral quantitation. Spectra were processed in TopSpin 3.5, with phase-correction, baseline correction and integration applied manually. Cyclic peptide concentrations were typically calibrated according to the total number of expected aromatic protons (His, Tyr, Phe, Trp), using integration of the 6-8 ppm region. Histone peptides were calibrated according to a characteristic peak, typically the low-chemical shift (0.5-1 ppm) methyl peak for valine/leucine/isoleucine. The reference TSP peak at ~0 ppm was integrated and set to 9.

Concentration Calculation: (assuming equal volumes of sample and standard added to tube):

TSP original concentration = a = 58.05 mM. Expected number of protons = b = 9. Integration = c = 9.

Peptide original concentration = x (unknown). Expected number of protons in chosen chemical shift region = y (known). Integration = z (measured).

$$x = a \times \frac{(z/y)}{(c/b)}$$

10.26.2. KDM7B$_{[PHD]}$ Protein Solution NMR

Nuclear Magnetic Resonance (NMR) spectra were recorded using a Bruker AVIII 700 MHz NMR spectrometer equipped with a 5-mm inverse cryoprobe using 3 mm MATCH NMR tubes (Cortectnet). For ^{1}H-^{15}N HSQC experiments, the Bruker pulse sequence *hsqcetfpf3gpsi* was used. The size of the FID was 2048 points (^{1}H) and 512 points (^{15}N), respectively. The spectral width was 16 ppm (^{1}H) and 40 ppm (^{15}N), respectively. $^{1}J_{NH}$ was set to 90 Hz. ^{15}N decoupling was achieved using a 170 µs 90-degree GARP sequence and the high power ^{15}N 90-degree pulse was achieved using a 40 µs GARP sequence. The relaxation delay was 1 second. All spectra were acquired at 308 K. Buffering was with 25 mM phosphate buffer, 100 mM NaCl, in 90% H_2O, 10% D_2O, pH 6.3. Data were processed with Bruker 3.2 software.

10.26.3. OC9 Solution NMR

Nuclear Magnetic Resonance (NMR) spectra were recorded using a Bruker Avance III 700 MHz equipped with TCI Cryoprobe, or a Bruker Avance III HD 600 MHz equipped with BBFO N_2 prodigy Cryoprobe, using 3 mm NMR tubes (Hilgenberg, #2001724). All spectra were acquired at 308 K. Buffering was with 25 mM phosphate buffer, 100 mM NaCl, in 90% H_2O, 10% D_2O, pH 6.3. Data were processed with Bruker Topspin 3.5 software. Water suppression was by excitation sculpting for OC9, and WATERGATE for OC9 P6A and OC9 P8A. Experiment conditions are set out in the table below:

	^{1}H-2D TOCSY	**^{1}H-2D TOCSY**	**^{1}H-2D ROESY**	**^{1}H-2D TOCSY**	**^{1}H-2D TOCSY**
ACQUISITION	**OC9**	**OC9**	**OC9**	**OC9 P6A**	**OC9 P8A**
MHz	700	700	700	600	600
Bruker Pulse Program	*dipsi2esgpph*	*dipsi2esgpph*	*roesyesgpph*	*dipsi2gpph19*	*dipsi2gpph19*
Mixing Time (ms)	120	60	60	80	80
TD (F1)	1200	1224	1224	512	800
TD (F2)	4096	4096	4096	2048	2048
Dummy Scans	16	16	16	16	16
Number of Scans	8	8	16	48	8
Spectral Width (ppm) (F1)	10	10	10	14	14
Spectral Width (ppm) (F2)	10	10	10	14	14
Increment for Delay (µs)	142.86	142.86	140	118.8	102
Acquisition Time (sec) (F1)	0.0857151	0.0874294	0.08568	0.0304128	0.0408
Acquisition Time (sec) (F2)	0.2924544	0.2924544	0.28672	0.1216512	0.104448
O1P (ppm) ^{1}H	4.692	4.692	4.692	4.700	4.700
Relaxation Delay (sec)	3	3	2	2	2
PROCESSING					
Window Function	QSINE	QSINE	QSINE	QSINE	QSINE
LB (Hz) (F1)	0.3	0.3	0.3	0.3	0.3
LB (Hz) (F2)	1	1	1	1	1
Real Spectrum Points for Zero Filling (F1)	2k	2k	1k	1k	1k
Real Spectrum Points for Zero Filling (F2)	8k	8k	1k	1k	1k

10.27. BioLayer Interferometry (BLI): General Protocol

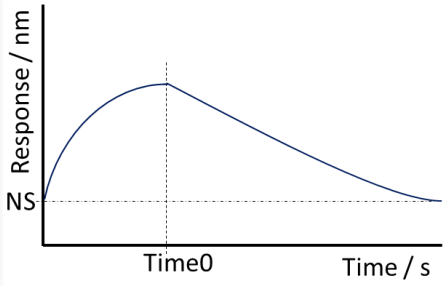
Equipment: OctetRed-384 System (Forte Bio), BLI plate (384-well tilted bottom, Forte Bio), BLI sensors (e.g. Ni-NTA, Forte Bio). **Software**: DataAcquisition v9 (Forte Bio), DataAnalysis v9 (Forte Bio), Prism v7 (GraphPad). **Buffer 1**: 1x PBS, 2 mM BME (β-mercaptoethanol)/TCEP (Tris(2-carboxyethyl)phosphine), 0.05 % Tween20, 1/0.1 % BSA. **Buffer 2**: Buffer 1 + x % DMSO

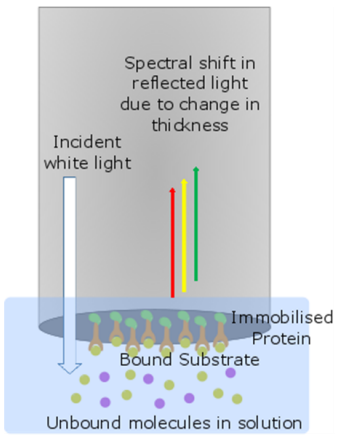
The protein of interest with an appropriate immobilisation tag (e.g. HisTag) was diluted to ~1 µM in Buffer 1 (minimum 560 µL) and kept at 4 °C until immediately before use. The compound of interest was first diluted into Buffer 1 from a concentrated stock solution down to the highest concentration to be tested (giving at least 160 µL). The resulting x % of DMSO was kept constant for the compound titration series (Buffer 2). Compound dissociation steps were performed in wells containing Buffer 2 only. Titration series were diluted 2-fold over 7 points total, (highest concentration marked on the plotted graphs), with the 8th position kept as Buffer 2 only to serve as a reference well for subtraction of Ni-NTA sensor drift in analysis (Data Analysis v9). Ni-NTA sensors were stripped between compounds by immersion in glycine, 0.5 M, pH 1 and regenerated with 10 mM $NiCl_2$. The BLI step-procedure for one compound in an Ni-NTA sensor experiment typically included: Baseline (Buffer 1, 15 s); Loading (Protein, threshold 1.5 nm / 180 s); Baseline (Buffer 1, 15 s); Baseline (Buffer 2, 15 s); Association (Compound Dilution Series, 120 s); Dissociation (Buffer 2, 300-600 s); 3x[Wash (glycine, 3 s), (Buffer 1, 3 s)]; Baseline (Buffer 1, 15s); Regeneration ($NiCl_2$, 60s); Baseline (Buffer 1, 15 s). One noteworthy practical observation emerged. The N-terminal thioredoxin/Z-Basic tag, located between the HisTag and KDM7 sequence, served as an effective and necessary spacer between the immobilisation surface and PHD finger. Immobilisation of a $KDM7_{[PHD]}$ via an N-terminal HisTag without a spacer resulted in no binding of the cognate H3 peptide.

10.27.1. BLI: Association then Dissociation Binding Model

Data were preliminarily analysed using DataAnalysis v9, with the no-peptide well (8th position) used as a reference for subtraction, to correct for signal drift due to slow protein dissociation from the Ni-NTA sensor. Alignment to baseline, inter-step correction to dissociation and Savitzky-Golay filtering were applied and data fitted with the global-fit 1:1 model, then results exported to a csv file to retrieve the processed background subtracted-raw data. These data were imported and fitted using Prism v7. Final reported kinetic values were derived from the 'Association then Dissociation' model.

The BLI plots show a different colour trace for each peptide concentration, with the top concentration noted on the plot and the rest a 2-fold dilution series.





$$K_D = {K_{off}}/{K_{on}} = \frac{[Ligand] \times [Receptor]}{[(Ligand.Receptor)]} = \frac{1}{K_{eq}} = \frac{1}{K_A}$$

$$\text{Ligand + Receptor} \underset{K_{off}}{\overset{K_{on}}{\rightleftharpoons}} \text{(Ligand.Receptor)}$$

For Fitting 'Association then Dissociation':

$$K_{obs} = (K_{on} \times [Compound]) + K_{off}$$

$$Y = \frac{B_{max} \times [Compound]}{([Compound] + K_D)}$$

$$Y_0 = Y \times (1 - e^{(-K_{obs} \times Time0)})$$

$$Association\text{: } Y_x = Y \times (1 - e^{(-K_{obs} \times x)})$$

$$Dissociation\text{: } Y_x = Y_0 \times e^{(-K_{off} \times (x - Time0))}$$

$$IF(X < Time0), Y = (Association) + NS$$
$$IF(X \geq Time0), Y = (Dissociation) + NS$$

K_D = affinity constant or equilibrium dissociation constant / M (calculated from K_{on} and K_{off})
Note that K_D gives [compound] for 50% occupancy of target
K_A = equilibrium association constant / M^{-1} = **K_{eq}** = equilibrium constant / M^{-1}
[Compound] = Concentration of Peptide / M (known parameter)
K_{obs} = observed rate constant / M^{-1} s^{-1} ([compound] dependent, determined from fit)
K_{on} = association rate constant / M^{-1} s^{-1} ([compound] dependent, calculated from K_{obs} and K_{off})
K_{off} = dissociation rate constant / s^{-1} ([compound] independent, determined from fit)
Y = binding response / nm
Y_x = binding response at time (x / sec)
Y_0 = binding response at Time0
Time0 = time dissociation is initiated
B_{max} = value of Y at [compound] and time ∞
NS = fixed non-specific binding term

10.28. Tryptophan Intrinsic Fluorescence Assay

Equipment: Pherastar FS (BMG Labtech), 280/350 nm excitation/emission filter.

Buffer = 1X PBS (1 % DMSO final)

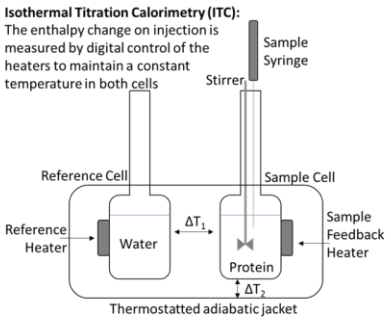
This procedure was adapted from the protocol reported by Dr Rebecca Hancock (DPhil Thesis). A titration (25, 10, 5 µM) of $KDM7C_{[PHD]}$ protein (expressed in the pNH-TrxT vector, with the purification tags cleaved) was first performed to optimise the concentration for the best tryptophan fluorescence signal, with each sample in triplicate at 50 µL sample per well.

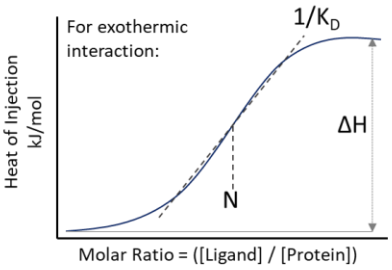
For testing the effect of the compounds, dilution series of compounds were made in DMSO and 1 µL of each 100X compound concentration was added to the bottom of a new Eppendorf tube, followed by 100 µL of protein and mixing by pipetting. Each mixture was split into 2x 50 µL for transfer to a flat clear-bottomed 384 well plate to give duplicate readings for each concentration point, ensuring all liquid arrived at the bottom of the well and no air bubbles were present. The focal distance was automatically optimised to 4.4 and the gain was manually adjusted to 600 for the best signal to noise ratio. The plate was shaken for 5 seconds prior to reading with the bottom optic of a Pherastar FS (BMG) plate reader, using single wavelength excitation (280 nm) and emission detection (350 nm). The final concentration of $KDM7C_{[PHD]}$ in each well was 25 µM. Control wells with protein in DMSO-containing buffer were used for the reference value of tryptophan fluorescence in free protein. The average of four control wells was calculated and subtracted from each sample fluorescence reading to give the change in fluorescence due to the compound. Results were fitted with Prism v7 using the log(agonist) vs response – Variable slope (four parameters) model.

10.29. Isothermal Titration Calorimetry (ITC): General Protocol

Equipment: Malvern Panalytical MicroCal PEAQ-ITC Automated (cell volume = 200 µL, cell material = Hastelloy, injection syringe volume = 40 µL). **Software**: Malvern Panalytical MicroCal PEAQ-ITC Analysis Software 1.1.0.1262. ITC Buffer = 50 mM HEPES, 100 mM NaCl, pH 7.4 (0.2 µm filtered).

Samples were prepared in ITC buffer, with the concentrated protein stock aliquot (stored at -80 °C) buffer exchanged using a PD-10 or PD-MiniTrap G-25 (GE Healthcare) prior to use. Protein concentrations were calibrated by taking an average of three Nanodrop readings, with extinction coefficients and mass predicted by ExPASy ProtParam based on the amino acid sequence. All KDM7 proteins had their purification tags cleaved prior to use. Peptides were dissolved from solid into ITC buffer. Peptide concentration was calibrated by ^{1}H-NMR. At least 370 µL of protein and 180 µL of peptide was prepared. The experiment was performed with the peptide titrated into the cell containing the protein. Titration consisted of 19 injections with 1x 0.4 µL (0.8 s), 18x 2 µL (4.0 s), spaced 150 s apart, stirring at 750 rpm and temperature held at 25 °C, with reference power set between 40-42 µW. Data were fitted using a one-set-of-sites binding model with the Malvern software.





For Sample, S, injection to Protein, P: $S + P \rightleftharpoons SP$

$$\Delta G = \Delta H - T\Delta S = -RTln(K_{eq}) = \mathrm{RTln}(K_D) \qquad K_{eq} = {}^{1}/_{K_D} = \frac{[SP]_{eq}}{[S]_{eq}\,[P]_{eq}}$$

K_D = affinity constant / M (determined from fit)
T = temperature / K (known)
ΔG = change in Gibbs free energy / kJ/mol^{-1} (calculated)
ΔS = change in entropy / kJ/mol^{-1} (calculated)
ΔH = change in enthalpy / kJ/mol^{-1} (measured)
R = universal gas constant

10.29.1. ITC: Experiment Conditions

The tabulated conditions below were used to conduct the ITC experiments.

Peptide	µM	KDM Protein	µM
OC9	113	7B PHD+JmjC	5.93
	113	7A PHD	10.0
	113	7B PHD	10.0
	113	7C PHD	10.0
	100	DIDO1 PHD	10.0
	50	Buffer	-
	50	5A PHD3	4.60
H3K4me3 [1-21]	100	7B PHD+JmjC	7.50
	100	7A PHD	10.0
	100	7B PHD	10.0
	100	7C PHD	10.0
	100	DIDO1 PHD	10.0
	50	5A PHD3	4
OC3	100	7B PHD+JmjC	7.50

10.30. HDxMS Protocol

Written by Dr Christopher Stubbs (AstraZeneca)

The hydrogen/deuterium exchange reactions were performed using a LEAP PAL RTC system (LEAP Technologies). Peptides were identified that covered 79% and 83% of the protein sequence for the PHD-finger only and PHD+JmjC (Dual Domain) constructs, respectively.

Protein ± ligand samples for HDX-MS were prepared in 20 mM HEPES-Na pH 7.5, 100 mM NaCl and 1 mM TCEP and kept at 4 °C, unless otherwise stated. The concentrations of the relevant species are listed in the table below, where DMSO was constant at 1% (v/v).

Ligand	Stock Solution		Labelling Reaction		Replicates
	Protein (µM)	Ligand (µM)	Protein (µM)	Ligand (µM)	
DMSO	10	-	0.5	-	3
H3K4me3 [1-21]	10	20	0.5	1	3
OC3	10	20	0.5	1	3
OC6	10	50	0.5	2.5	2
OC9	10	20	0.5	1	3

Protein = KDM7B Dual Domain OR PHD Finger

Deuterium exchange reactions were initiated by diluting the protein/ligand stock solutions 20-fold in 20 mM HEPES-Na pH 7.5, 100 mM NaCl, 1 mM TCEP prepared in D_2O (99.9% atom D, Sigma Aldrich; labelling buffer 96% (v/v) D_2O; labelling reaction 91.2% (v/v) D_2O), and incubating at 20 °C for 60 s, 300 s, 900 s, 1800 s and 3600 s. Undeuterated controls were prepared by performing the same dilution but in buffer prepared in H_2O. The labelling reactions were quenched by transferring 50 µL of the reaction mixture to 50 µL of pre-chilled quench solution (3.0 M urea, 1.6% (v/v) formic acid in water; 2 °C). Quenched samples were injected directly onto an Enzymate BEH immobilised pepsin column (2.1 x 300 mm, 3 µm; Waters) at 100 µL/min at 20 °C for 3 min at 10,000 psi. Peptic peptides were trapped and desalted on an Acquity BEH C18 VanGuard pre-column (130 Å, 2.1 x 5 mm, 1.7 µm; Waters) kept at 0.1 °C. The trapped peptides were eluted using a 6 min gradient of 5-35% acetonitrile in 0.1% (v/v) formic acid at 40 µL/min on an Acquity UPLC BEH C18 column (130 Å, 1.7 µm, 100 mm x 1 mm; Waters) at 0.1 °C. Peptides were detected on a SYNAPT G2-Si HDMS mass spectrometer (Waters) acquiring over an m/z range of 50–2000 with an electrospray source and lock mass calibration (Leucine Enkephalin, 200 pg/µL; Waters). The mass spectrometer was operated at a source temperature of 80 °C and a spray voltage of 2.5 kV. Spectra were collected in positive ion resolution mode.

Peptide identification was performed in Protein Lynx Global Server (Waters) using MS^E data collected for the undeuterated control samples. The resultant peptide lists were imported into DynamX (Waters) where peptides were filtered: minimum intensity of 5000, minimum of 0.2 products per amino acid, a maximum MH^+ error of 5 ppm, and found in all of the undeuterated datasets. The automatic peptide assignment in DynamX was performed using the standard parameters, but charge state assignment and retention times were verified manually for all peptides. Data were not corrected for back-exchange, so are relative rather

than absolute deuterium uptake values. The first two residues of each peptide are excluded from the analyses due to rapid back-exchange.

A Student's t-test was performed on the peptide-level data using the pooled standard deviation of all peptides across all samples and time-points (Pooled SD(DualDomain) = 0.08 Da; Pooled SD(PHD) = 0.05 Da). Peptides with a p-value <0.01 were considered significant and taken forward in the analyses. A custom MATLAB script was used to combine data from redundant peptides to generate a residue-level view of the differences in deuterium uptake for each time point; these results are shown in the Dual Domain Residues and PHD Residues Profiling figures.

The pooled variance was calculated by:

$$\sigma_p^2 = \frac{\sum_{i=1}^{k}(n_i - 1)\sigma_i^2}{\sum_{i=1}^{k}(n_i - 1)}$$

The two-tailed Student's t-test was calculated by:

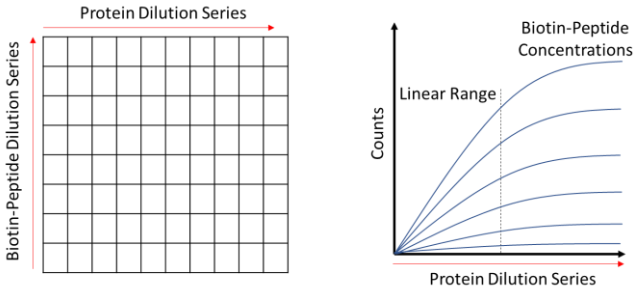
$$t = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{\sigma_p^2}{n_1} + \frac{\sigma_p^2}{n_2}}}$$

10.31. Displacement Assay (AlphaScreen): General Protocol

Equipment: ProxiPlate 384 Plus, White (PerkinElmer), BMG Pherastar FS with AlphaScreen detection module (680, 570), AlphaScreen Nickel Chelate Kit beads (6760619, Perkin Elmer). Note that in the following procedures: (#X refers to # times final desired concentration). Displacement AlphaScreen assays were carried out as described by Bhushan et al [137] with the following modifications.

10.31.1. Linear Range Determination AlphaScreen Displacement Assay

A dilution series of protein was tested against a dilution series of biotinylated peptide to find the optimal concentration of each required for a good signal to noise ratio (ideally >10), whilst remaining within the linear range of signal to be sensitive to any subsequent signal disruption by tested compounds. The DMSO % was matched to the planned compound assay, typically 0.5 % in a 20 µL assay.



General assay procedure in a ProxiPlate 384 Plus: 5 µL of 4X protein dilution series and 5 µL DMSO-buffer were added to each well and incubated at 20 °C for 5 minutes under foil seal. Then 5 µL of 4X biotinylated peptide (H3K4me3 [1-21] GGK-biotin, Anaspec) dilution series were added and incubated for 30 minutes at 20 °C. Then AlphaScreen nickel chelate kit beads (5 µL of 4X) (6760619M, pre-mixed and incubated for 20 minutes) were added under reduced ambient lighting and incubated under foil cover for 1 hour at 20 °C. The plate was read with a BMG Pherastar FS with AlphaScreen detection module (680, 570) and data were analysed in Microsoft Excel and Prism v7.

10.31.2. Dose-Response AlphaScreen Displacement Assay

Buffer: 50 mM HEPES pH 7.5, 0.01% Tween, 0.1% BSA, 150 mM NaCl. N.B. Buffer used for compound dilutions contained 2 % DMSO, to ensure consistency and result in 0.5 % DMSO in the final volume 20 µL assay. General assay procedure in a ProxiPlate 384 Plus: Each compound was tested in technical triplicate. The compound dilution series was prepared in a 96 well PCR plate, usually at least 8 points. Multichannel pipettes were used for liquid transfers.

5 µL of 4X protein was added to each well, followed by 5 µL of compound dilution series and incubated at 20 °C for 5 minutes under foil seal. Then 5 µL of 4X biotinylated peptide (H3K4me3 [1-21] GGK-biotin, Anaspec) were added and incubated for 30 minutes at 20 °C. Then AlphaScreen nickel chelate kit beads (5 µL of 4X) (6760619M, pre-mixed and incubated for 20 minutes) were added under reduced ambient lighting and incubated under foil cover for 1 hour at 20 °C. The plate was read with a BMG Pherastar FS with AlphaScreen detection module (680, 570).

Control wells had either no compound (maximum signal) or no biotinylated peptide (minimum signal). Measured counts were normalised to the maximum signal by dividing the counts for a given data point by the average maximum control signal and converted to a %. This was converted (100 - % signal) to represent % displacement of the interaction where appropriate. The data were plotted and fitted using GraphPad Prism 7 with the model: log(inhibitor) vs. normalized response -- Variable slope:

$$Y = 100/(1 + 10^{((LogIC_{50} - x) \times HillSlope)})$$

Where:

Y = normalised signal (% displacement), **x** = log(concentration), **IC/EC$_{50}$** = compound concentration to achieve normalised signal = 50%, **HillSlope** = steepness of the curve

Data are reported with the fitted IC$_{50}$ value and its 95% confidence interval values (95% chance that the confidence interval calculated contains the true population mean).

10.31.3. Table of AlphaScreen Displacement Assay Conditions

Protein							Biotin-Peptide		Assay	Temp
Organism	AAs	Domains	Vector	Tags Present	Name	nM	I.D. #	nM	Buffer	°C
HS (Q6ZMT4-1)	38 - 480	1,2	pNH-TrxT	Nt-His$_6$+Trx	KDM7A$_{[PHD+JmjC]}$	25	1	25	1	20
HS (Q6ZMT4-1)	38 - 98	2	pNH-TrxT	Nt-His$_6$+Trx	KDM7A$_{[PHD]}$	100	1	50	1	20
HS (Q9UPP1-2)	1 - 447	1,2	pNH-TrxT	Nt-His$_6$+Trx	KDM7B$_{[PHD+JmjC]}$	10	1	10	1	20
HS (Q9UPP1-2)	1 - 66	2	pNH-TrxT	Nt-His$_6$+Trx	KDM7B$_{[PHD]}$	50	1	20	1	20
HS (O75151)	1 - 449	1,2	pNIC-ZB	Nt-His$_6$+ZB	KDM7C$_{[PHD+JmjC]}$	100	1	6	1	20
HS (O75151)	1 - 67	2	pNIC-ZB	Nt-His$_6$+ZB	KDM7C$_{[PHD]}$	25	1	6	1	20
HS (Q9BTC0)	260 - 325	2	pNH-TrxT	Nt-His$_6$+Trx	DIDO1$_{[PHD]}$	50	1	20	1	20
HS (Q5VWG9)	857 - 924	2	pGTVL2	Nt-His$_6$+GST	TAF3$_{[PHD]}$	12.5	1	3	1	20
HS (P29375)	1542 - 1660	2	pGTVL2	Nt-His$_6$+GST	KDM5A$_{[PHD]}$	50	1	25	1	20
HS (Q9H160)	171 - 226	2	pGTVL2	Nt-His$_6$+GST	ING2$_{[PHD]}$	25	1	3	1	20
HS (Q9Y657)	26 - 262	3	pNIC-CTHF	Ct-His$_6$	SPIN1$_{[TTD]}$	25	1	6	1	20

I.D. #	DOMAIN
1	JmjC
2	PHD Finger
3	Triple Tudor

TAGS
Nt = N-terminal
Ct = C-terminal
Trx = Thioredoxin
His$_6$ = hexa-histidine
ZB = Z-basic

I.D. #	Biotin-PEPTIDE	Source
1	H3K4me3 (1-21)-GGK-biotin	Anaspec

I.D. #	ASSAY BUFFER
1	50 mM HEPES pH 7.5, 0.01% Tween, 0.1% BSA, 150 mM NaCl

10.32. ECHO Dispensing of Compounds

Where ECHO dispensing is stated for a procedure, all compounds were made up in DMSO and dispensed directly into assay playes using a Labcyte ECHO 550. Wells were back-filled with DMSO where appropriate for dilutions to achieve a consistent target DMSO % in the final assay.

Plates (LABCYTE): Source Plate: 384-Well Low Dead Volume Microplate, Cyclic Olefin Copolymer, RNase/DNase Free. Intermediate Dilutions: Echo Qualified 384-Well Polypropylene Microplate, Clear, Flat Bottom.

10.33. General Compounds for Demethylase Activity Assays

The following abbreviations are used in the procedures describing the demethylase activity assays.

Fe(II) = ammonium Iron(II) sulfur hexahydrate (Sigma Aldrich).

NaAsc = Sodium L-ascorbate (Sigma Aldrich).

2OG = 2-oxoglutarate (Sigma Aldrich)

10.34. Activity Assay (AlphaScreen): General Protocol

Equipment: ProxiPlate 384 Plus White (PerkinElmer), BMG Pherastar FS with AlphaScreen detection module (680, 570), AlphaScreen General IgG (Protein A) Detection Kit (Perkin Elmer, 6760617). Note that in the following procedures: (#X refers to # times final desired concentration). KDM AlphaScreen assays were carried out as described by Hopkinson et al [246] and Rose et al [160] with the following modifications.

A timecourse was performed according to the procedure below, with an equivalent DMSO % in place of compounds. Each timepoint was run at least in duplicate and in different wells.

Compound titration series were ECHO dispensed in triplicate (100 nL in DMSO) to a Proxi-Plate for a 10 µL total volume assay (1 % DMSO final). The AlphaScreen IgG beads were incubated with antibody for at least 30 minutes at 20 °C prior to use. A dilution of 1.33X enzyme was prepared to add 7.5 µL to each well and then incubated for 10 minutes at 20 °C. A 4X substrate mix was prepared containing Fe(II), 2OG, Asc and histone H3 peptide substrate in buffer. Fe(II) was initially diluted in 20 mM HCl, then subsequently in MilliQ, and added to the master substrate mix after Asc. 2.5 µL of substrate mix was added to each well to initiate the reaction which proceeded at 20 °C under a foil seal. The reaction was quenched at an

endpoint within the previously determined linear range by addition of 5 µL of (30 mM EDTA, 1.5 M NaCl), then 5 µL of IgG-antibody bead mix was added to each well under low ambient light conditions and the plate foil sealed and incubated at 20 °C for 1 hour and read with a BMG Pherastar FS with AlphaScreen detection module (680, 570). Measured counts were normalised to the average of DMSO-only control wells (maximum observable count signal from catalytic activity) and analysed with Prism v7, fitted using the model: log(inhibitor) vs. normalized response -- Variable slope: $Y = 100/(1 + 10^{((LogIC_{50} - x) \times HillSlope)})$

Where: **Y** = normalised signal (% binding or % inhibition), **x** = log(concentration), **IC/EC$_{50}$** = compound concentration to achieve normalised signal = 50%, **HillSlope** = steepness of the curve. Data are reported with the fitted IC_{50} value and associated 95% confidence interval values (95% chance that the confidence interval calculated contains the true population mean).

10.34.1. Table of AlphaScreen Activity Assay Conditions

Enzyme		Biotin-Substrate		Co-Factor (µM)			Assay	EndPoint	Antibody	Antibody	Temp
Name	nM	I.D. #	nM	Fe(II)	Ascorbate	2OG	Buffer	(min)		Dilution	°C
KDM7B$_{[PHD+JmjC]}$	20	1	150	10	100	20	1	10	Anti-H3K9me1 (Ab8896)	10,000	20
KDM4A$_{[JmjC]}$	1	2	30	1	100	10	2	20	Anti-H3K9me2 (Ab1220)	20,000	20
KDM5A$_{[JmjC]}$	4	3	100	10	100	5	2	20	Anti-H3K4me2 (CST9725S)	3,000	20

I.D. #	SUBSTRATE
1	H3K4me3K9me2 (1-21)-GGK-biotin
2	H3K9me3 (1-21)-GGK-biotin
3	H3K4me3 (1-21)-GGK-biotin

I.D. #	ASSAY BUFFER
1	50 mM HEPES, pH 7.05, 50 mM NaCl 0.1% BSA, 0.01% Tween20
2	50 mM HEPES, pH 7.5, 0.1% BSA, 0.01% Tween20

10.35. Activity Assays (MALDI-TOF-MS): General Protocol

Equipment: Bruker MALDI-TOF Microflex LRF, ThermoMixer C (Eppendorf) and SmartBlock, PCR plates (4ti-0740, 4titude). **Software**: Bruker Compass for flexSeries 1.4: flexControl v3.4 and flexAnalysis v3.4. Note that in the following procedures: (#X refers to # times final desired concentration). MALDI-TOF KDM assays were carried out as described by Hopkinson et al [246] and Rose et al [160] with the following modifications.

10.35.1. Single Point Concentration MALDI-TOF-MS Activity Assay

A timecourse was first performed according to the below procedure, with an equivalent DMSO % in place of compounds. Each timepoint was run in at least in duplicate and each run in a different well. To a 96-well PCR plate was added 2 µL of 5X compound (typically 1 % DMSO in final 10 µL assay volume). To each well was added 4 µL of 2.5X protein, incubated for 10 minutes at 20 °C and then reaction initiated on a temperature-controlled ThermoMixer C (Eppendorf) by addition of 4 µL of 2.5X substrate mix (Asc, 2OG, Fe(II), Histone H3 peptide substrate) – Fe(II) was initially diluted in 20 mM HCl, then subsequently in MilliQ, and added to the master substrate mix after Asc. The reaction was quenched at an endpoint within the previously determined linear range by addition of 4 µL of 6 % (v/v) formic acid. Each assay well was analysed by MALDI-TOF-MS, spotting 1 µL of sample with 1 µL of CHCA matrix solution (saturated, >10 mg/mL, in 1:1 MeCN:H_2O with 0.1 % TFA (v/v)).

The % demethylation was determined by analysis (Bruker flexAnalysis v3.4) of the monoisotopic peak intensity for each observable state of demethylation. Data were analysed using Prism v7 with statistical significance of the inhibition assessed by an ordinary one-way ANOVA test, assuming a normal distribution and common variance, followed by Dunnett's multiple comparison test (comparing the mean value for each compound result with the mean value of the DMSO-control) [372]. The significance of the resulting multiplicity adjusted P-

values was used to label the graphs (the chosen threshold values for declaration of significance are noted in the figure legends). Multiplicity adjusted P-values are defined by GraphPad using the thresholds noted in the figure legends.

10.35.2. Dose-Response MALDI-TOF-MS Activity Assay

For the dose-response assay versus $KDM7A_{[PHD+JmjC]}$, $KDM7B_{[PHD+JmjC]}$ a timecourse was first performed according to the procedure below to determine the linear range according to the conditions, but with DMSO in place of compound. Each compound titration series was ECHO dispensed (100 nL DMSO) in triplicate to a 96-well PCR plate, for a final assay volume of 20 µL (0.5 % DMSO final). 10 µL of 2X enzyme solution was added per well and incubated at 20 °C for 5 minutes, then the reaction initiated on a temperature controlled on a ThermoMixer C (Eppendorf) by addition of 20 µL per well of 2X substrate mix (Asc, 2OG, Fe(II), Histone H3 peptide substrate) – Fe(II) was initially diluted in 20 mM HCl, then subsequently in MilliQ, and added to the master substrate mix after Asc. The reaction was quenched at a fixed endpoint within the previously determined linear range by addition of 5 µL of 5% (v/v) formic acid. Each assay well was analysed by MALDI-TOF-MS, spotting 1 µL of sample with 1 µL of CHCA matrix. The % demethylation was determined by analysis (Bruker flexAnalysis v3.4) of the monoisotopic peak intensity corresponding to the H3K9me2/1/0 states and converted to % inhibition, normalised against DMSO-only control wells. For the KDM7A assay, where activation/ enhancement of demethylase activity was seen against H3K4me3K9me2, the % demethylation of H3K4me3K9me2 was instead normalised to the % of demethylation seen versus the preferred H3K9me2 substrate over the same duration.

Data were analysed with Prism v7, fitted using the model: log(inhibitor) vs. normalized response -- Variable slope:

$$Y = 100/(1 + 10^{((LogIC_{50} - x) \times HillSlope)})$$

Where:

Y = normalised signal (% inhibition/enhancement of demethylation), **x** = log(concentration), **IC/EC$_{50}$** = compound concentration to achieve a normalised signal of 50%, **HillSlope** = steepness of the curve. Data are reported with the fitted IC_{50} / EC_{50} value and associated 95% confidence interval values (95% chance that the confidence interval calculated contains the true population mean).

10.35.3. Table of MALDI Assay Conditions

For Single Concentration Point:

Enzyme							Substrate		Co-Factor (µM)			Temp	Assay	Timecourse	EndPoint	Replicates
Organism	AAs	Domains	Vector	Tags	Name	µM	I.D. #	µM	Fe(II)	Ascorbate	2OG	°C	Buffer	Performed?	(min)	per Point
HS (Q9Y2K7)	1 - 1162	1,2,3,4	Commercial	Nt-FLAG	**KDM2A**	0.29	4	10	50	100	50	37	1	Yes	25	2
HS (Q8NHM5)	1 - 1336	1,2,3,4	Commercial	Nt-FLAG	**KDM2B**	0.21	4	10	50	100	50	37	1	Yes	30	3
HS (Q9Y4C1)	515 - 1317	1,3	pFB-CT10HF-LIC	Nt-His_{10}	**KDM3A**	0.10	2	10	50	100	50	20	1	Yes	7	3
HS (O75164)	1 - 359	1,5	pNIC28-Bsa4	Nt-His_6	**KDM4A**	1.0	5	10	50	100	50	37	1	Yes	10	2
HS (P41229)	1 - 765	1,2,3,5,6	pFB-CT10HF-LIC	Ct-His_{10}	**KDM5C**	0.25	6	10	50	100	50	37	1	Yes	10	3
HS (O15054)	1141 - 1641	1	pNH-TrxT	Nt-His_6+Trxt	**KDM6B**	0.50	3	10	10	100	50	25	2	Yes	7	3
HS (Q6ZMT4-1)	38 - 480	1,2	pNH-TrxT	Cleaved	**KDM7A**	2	1,2	5	100	1000	200	25	1	Yes	17	3
HS (Q9UPP1-2)	1 - 447	1,2	pNH-TrxT	Nt-His_6+Trx	**KDM7B**	2	1	5	50	100	50	30	1	Yes	6	3
HS (Q9UPP1-2)	1 - 447	1,2	pNH-TrxT	Cleaved	**KDM7B**	2	2	5	100	1000	200	25	1	Yes	4	3
HS (Q9GZT9)	181-426	7	pNIC-Bio3	Nt-His_6 Ct-AviTag	**PHD2**	1	7	50	50	100	20	25	2	Yes	18	3

For Dose-Response:

Enzyme							Substrate		Co-Factor (µM)			Temp	Assay	Timecourse	EndPoint	Replicates
Organism	AAs	Domains	Vector	Tags	Name	µM	I.D. #	µM	Fe(II)	Ascorbate	2OG	°C	Buffer	Performed?	(min)	per Point
HS (Q6ZMT4-1)	38 - 480	1,2	pNH-TrxT	Nt-His_6+Trx	**KDM7A**	0.35	1,2	2	100	200	100	30	1	Yes	10	3
HS (Q9UPP1-2)	1 - 447	1,2	pNH-TrxT	Nt-His_6+Trx	**KDM7B**	0.35	1	1	100	200	100	30	1	Yes	4	3

Key:

I.D. #	SUBSTRATE
1	H3K4me3K9me2 (1-21)
2	H3K9me2 (1-21)
3	H3K27me3 (21-44)
4	H3K36me2 (30-41)
5	H3K9me3 (1-21)
6	H3K4me3 (1-21)
7	CODD (DLDLEALAPYIPADDDFQL)

I.D. #	DOMAIN
1	JmjC
2	PHD Finger
3	Other Zinc Finger
4	F-box
5	JmjN
6	ARID
7	Fe,2OG,Dioxygenase

I.D. #	ASSAY BUFFER
1	50 mM HEPES, pH 7.05, 50 mM NaCl, 1 mM TCEP
2	50 mM HEPES, pH 7.50, 150 mM NaCl

TAGS
Nt = N-terminal
Ct = C-terminal
Trx = Thioredoxin
His = histidine

CO-FACTOR I.D.
Fe(II) = CAS 7783-85-9
Ascorbate = CAS 134-03-2
2OG = CAS 328-50-7

10.36. Activity Assay (RapidFire): General Protocol

Equipment: RapidFire High-Throughput Mass Spectrometry System (Agilent). **Software:** RapidFire Integrator (Agilent), Agilent MassHunter. Note that in the following procedures: (#X refers to # times final desired concentration).

A timecourse with DMSO in place of compounds was performed prior to the dose-response assay to establish an end-point within the linear range.

Assay conditions: 0.5 µM $KDM7B_{[PHD+JmjC]}$, 100 µM sodium ascorbate, 20 µM 2OG, 5 µM Fe(II), 5 µM H3K4me3K9me2 [1-15], 0.5% DMSO. 50 µL total assay volume per well. Buffer: 50 mM HEPES, pH 7.45, 50 mM NaCl, 2 mM TCEP. Temperature of all steps was at 20 °C. Compound titration series in triplicate were ECHO dispensed to a polypropylene 384-well microplate, V-Shape (Greiner, 781280), and 25 µL of 2X protein was added and incubated for 10 minutes. Reaction was initiated by addition of 25 µL of substrate mix to each well and proceeded for 60 minutes before quenching with 5 µL of 10 % (v/v) formic acid. Data were first analysed using RapidFire Integrator to quantitate the peak intensities for the +4 charge state and then % demethylation converted to % inhibition and analysed with Prism v7 using the model: log(inhibitor) vs. normalized response -- Variable slope:

$$Y = 100/(1 + 10^{((LogIC_{50} - x) \times HillSlope)})$$

Where: **Y** = normalised signal (% inhibition/enhancement of demethylation), **x** = log(concentration), **IC/EC_{50}** = compound concentration to achieve normalised signal = 50%, **HillSlope** = steepness of the curve. Data are reported with the fitted IC_{50} / EC_{50} value and associated 95% confidence interval values (95% chance that the confidence interval calculated contains the true population mean).

10.37. Histone Extract Demethylase Activity Assay: General Protocol

Equipment: ThermoMixer C and SmartBlocks (Eppendorf); 96-well PCR Plate Skirted (4ti-0740, 4titude). Note that in the following procedures: (#X refers to # times final desired concentration).

Calf Thymus Histone Extracts (Type IIA, H9250, Sigma Aldrich) were dissolved to 10 mg/mL in MilliQ H_2O and stored as aliquots at -20 °C. Buffer = 50 mM HEPES, 50 mM NaCl, 1 mM TCEP, pH 7.4.

A DMSO-only control timecourse (0, 15, 30, 60, 90, 120 minutes) was run according to the procedure below to determine the enzyme concentration (0.5, 1, 2.5, 5 µM) necessary to observe changes in demethylation levels over time at 37 °C.

The assay was set-up in a 96-well PCR plate on a temperature-controlled ThermoMixer C 96-block. To each well was added 4 µL of compound (20 µM in final assay volume) followed by 8 µL of 2.5X enzyme, incubated for 5 minutes at 37 °C, then reaction initiated by addition of 8 µL of 2.5X substrate mix (Asc, 2OG, Fe(II), calf histones) – Fe(II) was initially diluted in 20 mM HCl, then subsequently in MilliQ, and added to the master substrate mix after Asc. Final assay volume of 20 µL (0.4 % DMSO). Reactions were quenched at desired timepoints (0, 15, 30, 60 minutes) by addition of 2 µL of 10 % (v/v) formic acid.

Samples were prepared for SDS-PAGE gel analysis by: 12 µL assay mixture + 4 µL 4X NuPAGE LDS loading buffer with 100 mM DTT to a PCR tube, heated at 95 °C for 6 minutes, and the entirety of the sample loaded to the SDS-PAGE gel. Separated histones were subsequently analysed by Western Blot. Compound-Assay Conditions:

		Co-Factors (µM)			Calf Histones
Protein	**µM**	**Fe(II)**	**2OG**	**Ascorbate**	**µg/mL**
$KDM7A_{[PHD+JmjC]}$	0.5	100	1000	1000	50
$KDM7B_{[PHD+JmjC]}$	1	100	1000	1000	50

10.38. Affinity Pull-Down of KDM7B from HEK293T Nuclear Lysate

HEK293T cells were transiently transfected (6-well plate, 800k cells per well, 2 µg DNA per well) with either NLuc-KDM7B(WT) or NLuc-KDM7B(PHD_Mutant[Y18A, W29A]), incubated for 18 hours at 37 °C with 5% CO_2 in normoxia and harvested by centrifugation at 100 g. Cell counting was performed prior to centrifugation to estimate $2x10^6$ of cells per aliquot for pelleting. Low-binding tubes were used for all following steps. Cell pellets were lysed with a Nuclear Extraction kit (Ab219177, Abcam) according the manufacturer's instructions which enabled separation of cytoplasmic, soluble nuclear and insoluble nuclear fractions. The soluble nuclear lysate was used for the pull-down experiment and the total amount of protein present within the sample was quantified by bicinchoninic acid (BCA) assay (Pierce BCA Protein Assay Kit 23227 and Albumin Standards 23209, Thermo Scientific). The soluble nuclear lysates from each cell pellet were then diluted in PBS-T to a standard concentration (0.5 mg/mL of 500 µL). Magnetic streptavidin beads (Dynabeads M-280 Streptavidin 11206D, Invitrogen) were prepared by washing three times in 5 volumes of PBS-T before final re-suspension to their original volume. A pre-clear step was performed by incubating 0.25 lysate-volumes' worth of beads alone with the lysate with rotation for 15 minutes at 4 °C. Magnetic beads used for the pre-clear were removed by magnetic separation. The pre-cleared nuclear soluble lysate was aliquoted (3x 100 µL) for three pull-down experiments. The final concentration of a pull-down probe was targeted to be 1 µM in 100 µL of soluble nuclear lysate. Probes were added (< 5 µL additional solution) and incubated with rotation for 30 minutes at 4 °C. For competition experiments, the unlabelled competitor was also added at this stage from a concentrated DMSO stock to a final concentration of 100 µM (final DMSO < 2 % for all pull-downs). Probes were recovered by addition of magnetic streptavidin beads. The streptavidin beads have an estimated binding capacity of 4 pmol / µL. The theoretical

maximum pull-down yield was 100 pmol (1 µM x 100 µL), therefore at least 25 µL of beads were required. 30 µL of beads were added to each pull-down and incubated with rotation for 15 minutes at 4 °C. Beads were recovered by magnetic separation and washed (3x 30 µL PBS-T), transferring to a new low-binding tube between each wash, before final re-suspension in 30 µL of loading buffer (PBS-T and NuPAGE 4X LDS Sample Buffer NP007, ThermoFisher, with 1 mM DTT). The suspended beads were heated at 95 °C for 5 minutes and the supernatant recovered to a new low-binding tube as the final pull-down sample. 8 µL of each final pull-down sample was loaded to SDS-PAGE gel. The input lysate samples were prepared to 8 µL total with loading buffer and 0.6 µL of the diluted pre-cleared nuclear soluble lysate (0.5 mg/mL). This represented 0.6 % of the total available protein in a 100 µL sample. *Note: this was not subsequently easily observable by Western Blot, and a larger input-sample proportion is recommended.* Gels were run for 45 minutes at 180 V. Further analysis of the gel was performed by Western Blot and ECL detection.

10.39. Western Blot Procedure

Equipment: iBlot (Invitrogen); iBlot Transfer Stacks Nitrocellulose Mini (Invitrogen); Mini PROTEAN Tetra Cell (BioRad). **Software:** BioRad Image Lab v5.0

Histone Transfer: The SDS-PAGE gel was transferred directly after electrophoresis to a nitrocellulose membrane using an iBlot transfer stack with an iBlot at 20 V for 6 minutes.

Non-Histone Transfer: Transfer buffer: (100 mL MeOH, 100 mL 10X TG, 800 mL H_2O, chilled to 4 °C). SDS-PAGE gels were equilibrated in transfer buffer for 15 minutes. Nitrocellulose membrane (0.45 µm, 10600003, Amersham Protran) was pre-wet in MilliQ H_2O for 5 minutes. The membrane, filter paper (3 mm, Chr) and gel were assembled into the blot-stack (Mini Trans-Blot Electrophoretic Transfer Cell, BioRad) according to manufacturer's instructions.

Transfer was performed at 100 V for 60 minutes, at 4 °C. The quality of the transfer was checked by total protein staining (Ponceau) for 5 minutes, then de-stained in MilliQ H_2O.

Membrane Development: The membrane was blocked in PBS-T with 5 % milk (skim milk powder 70166, Sigma Aldrich) for 90 minutes, then washed (3x PBS-T for 5 minutes). Primary antibody (in PBS with 0.02% NaN_3) was added and incubated for 16 hours at 4 °C with gentle rocking. The primary antibody solution was recovered, and the membrane washed (3x PBS-T) before addition of secondary antibody (in PBS with 5 % milk) for 1 hour at 20 °C with gentle rocking. Membranes were then washed thoroughly (5x PBS-T for 5 minutes) and imaged by ECL.

Electrochemiluminescence (ECL) Detection:

Performed using ECL detection kit (SuperSignal West Pico PLUS Chemiluminescent Substrate, ThermoFisher), according to the manufacturer's instructions and incubating the membrane with the with reagent for 5 minutes before imaging (ChemiDoc MP, BioRad).

10.39.1. Western Blot Antibodies

Class	Target	Name	Source	Dilution	Species	Type
Primary	Anti-H3K9me2	Ab1220	Abcam	1000	Mouse	mAb
Primary	Anti-H3K27me2	CST-D18C8	CST	1000	Rabbit	mAb
Primary	Anti-H4	Ab177840	Abcam	1000	Rabbit	r-mAb
Primary	Anti-KDM7B (AAs 800-900)	Ab36068	Abcam	500	Rabbit	pAb
Secondary	HRP-Anti-Mouse IgG	PI-2000	Vector Laboratories	5000	Horse	
Secondary	HRP-Anti-Rabbit IgG	PI-1000	Vector Laboratories	5000	Goat	

10.40. Synthesis of OC9-Biotin

A 1000X solution (800 mM) of anhydrous Cu(II)SO_4 was prepared by dissolution of 12.8 mg in 100 µL H_2O, followed by dilution to a 10X solution (8 mM). A 10X solution (24 mM) of THPTA (aka: Tris(3-hydroxypropyltriazolylmethyl)amine) was prepared by dissolution of 1.04 mg in 100 µL of H_2O. A 100X solution (400 mM) of sodium ascorbate was prepared by dissolution of 7.92 mg in 100 µL H_2O, followed by dilution to a 10X solution. A 100X solution of Biotin-PEG(4)-alkyne was prepared by dissolution of 2.01 mg in 100 µL H_2O, followed by dilution to a 10X solution.

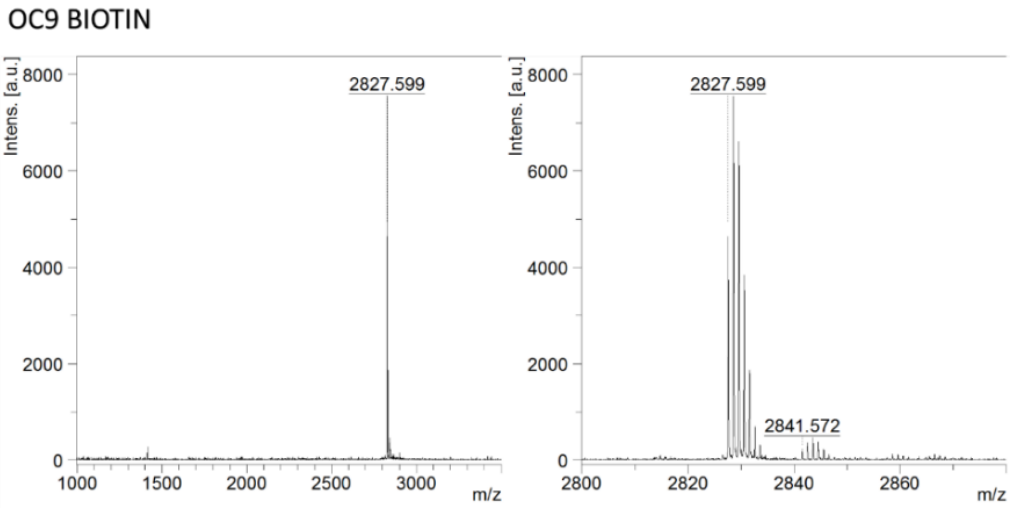
To a standard 1.5 mL Eppendorf tube was added 100 µL of 10X Cu(II)SO_4 solution (0.13 mg, 2 eqv) and 100 µL of 10X THPTA solution (1.04 mg, 6 eqv). The mixture was allowed to incubate at 20 °C for several minutes until a light blue colour emerged. Separately, to a standard 1.5 mL Eppendorf tube was added OC9-azide (0.95 mg, 1 eqv) in 100 µL of H_2O/MeCN (80/20), followed by 100 µL of 10X Biotin-PEG(4)-Alkyne solution (0.20 mg, 1.1 eqv) (aka: 15-[D(+)-Biotinylamino]-4,7,10,13-tetraoxapentadec-1-yne, CAS# 1262681-31-1) (Source: PEG4950.0250, Iris Biotech GmbH).

The Cu(II)SO_4/THPTA solution mixture was then added to the OC9-azide/alkyne solution, followed by 100 µL of 10X sodium ascorbate solution (0.79 mg, 10 eqv) and mixed gently with a pipette whereupon the solution turned clear. The reaction mixture was then buffered by addition of 500 µL of 20 mM sodium phosphate (pH 6.5), the lid sealed and the reaction incubated at 37 °C, 600 rpm on a Thermo Mixer C (Eppendorf). Concentrations in final reaction: OC9-azide (0.4 mM); Biotin-PEG(4)-Alkyne (0.44 mM); Cu(II)SO_4 (0.8 mM); THPTA (2.4 mM); Sodium Ascorbate (4 mM). The reaction was monitored by sampling of 1 µL aliquots into 50 µL of H_2O/MeCN (50/50) and spotting of 1 µL for MALDI-TOF-MS analysis. After 6

hours no starting material was detectable by MALDI-TOF-MS, the reaction was deemed complete and the reaction mixture was injected directly to HPLC for purification (see general HPLC protocol for peptides).

HR-ESI-MS: Calculated (C_{127} H_{187} O_{37} N_{33} $^{32}S_2$) [M+2H] = 1415.15980, Found = 1415.15894 (Δppm -0.61)

MALDI-TOF-MS assessment of peptide purity post-HPLC/lyophilisation [M+H] calculated: 2829.313.



F.10.40.i. MALDI-TOF-MS spectrum of OC9-biotin. The left-window shows a wide-mass range to give an indication of peptide purity. The right-window is a zoom in on the peak of interest.

10.41. NanoBRET Experiments

10.41.1. Cellular Assay

NanoBRET experiments were conducted using the NanoBRET™ Protein:Protein Interaction System, PPI Assay Flexi Starter System (N1821, Promega), according to the guidance in the manufacturer's protocol. Plasmid constructs purchased from Promega/Kazusa and verified by DNA Sanger Sequencing on receipt: i) FHC01149: KDM7B (variant 1)-N-terminal HaloTag in pFN21A. ii) FHC09287: H3F3A-N-terminal HaloTag in pFN21A. The Flexi® Cloning System (Promega) was used to sub-clone the KDM7B.v.1 sequence into a C-terminal NLuc tag plasmid (pFC32K) and to sub-clone the histone H3.3 sequence into a C-terminal HaloTag plasmid (pFC14K). The HaloTag Ligand structure is proprietary, but is believed to consist of a non-chloro-TOM dye (NCT) molecule (BRET acceptor) connected to a chloroalkane (to react with HaloTag protein).

2 mL of cultured HEK293T cells (800,000) in one well of a six-well plate were transiently transfected for 20 hours in DMEM media with 2 µg of DNA total alongside 8 µL of FuGENE HD (Promega), at 37 °C and 5 % CO_2. For signal optimisation of the assay, the transfection ratio of DNA between the HaloTag and NLuc containing plasmids was varied (1:10/100/1000 NLuc to HaloTag) and the N-terminal or C-terminal pairings of the plasmids were varied. Cells were washed with PBS and trpysinised for re-suspension in DMEM, then centrifugation at 100 g for 5 minutes and re-suspension in Opti-MEM I Reduced Serum Medium, no phenol red, 4 % FBS. Cell density was adjusted to 2.2 x10^5 cells per mL and divided into two pools. One pool received HaloTag NanoBRET 618 Ligand, the other DMSO vehicle as a control. Cells were plated to a 96-well plate: for signal optimisation 100 µL of cells per well was used and incubated overnight at 37 °C and 5 % CO_2; for compound testing 90 µL of cells per well was used, and 10 µL of 10X compound (with or without digitonin in the compound solution) was

added for incubation at 37 °C and 5 % CO_2 overnight (no digitonin) or 2 hours (with digitonin 50 ng/mL final). Each assay point was plated in at least duplicate. Plates were imaged by addition of 25 µL of 5X NanoBRET Nano-Glo Substrate (Promega) in Opti-MEM I Reduced Serum Medium, no phenol red, 4 % FBS under reduced ambient lighting. Plates were imaged within 15 minutes with a Pherastar FS (BMG Labtech) using a NanoBRET compatible filter (top optic, 460-480 nm donor emission, 610 nm acceptor emission detection). The corrected mean NanoBRET ratio for each sample was calculated by:

$$\frac{acceptor\ emission}{donor\ emission} = raw\ NanoBRET\ ratio = BU \qquad milliBRET\ (mBU) = BU\ \times 1000 = mBU$$

Determine the mean value for each assay point, then:

$$mean(mBU\ experimental) - mean(mBU\ no\ HaloLigand\ control) = mean\ corrected\ mBU$$

The mean corrected mBU for compounds could then be normalised to the DMSO compound control.

10.41.2. Site-Directed Mutagenesis

Site directed mutagenesis of full-length KDM7B was performed according to standard molecular biology PCR protocols (conditions are given below) on a KDM7B C-terminal NLuc template plasmid.

Primers for KDM7B PHD finger (Y14A, W29A):

Forwards Primer:

5'-CTGCCTgcCGATGTGACCCGCTTCATGATCGAGTGTGACATGTGCCAGGACgcGTTTCATGGCAGTTG-3'

Reverse Primer:

5'-GAAACgcGTCCTGGCACATGTCACACTCGATCATGAAGCGGGTCACATCGgcAGGCAGCCGGCAGAG-3'

PCR Reaction: The PCR reaction was assembled in a PCR tube and contained: 100 ng plasmid template (full-length KDM7B wild-type, sequence verified), 0.2 μM forwards primer, 0.2 μM reverse primer, 250 μM dNTPs, 5 μL 10X PfuUltra II reaction buffer, 1 μL PfuUltra II fusion HS DNA polymerase, MilliQ H_2O up to 50 μL.

PCR Parameters: 98°C 2 min; [98 °C 20 sec, 55 °C 20 sec, 72 °C 6 min (60 sec / kb)] x20; 72°C 3 min

Followed by digestion with 10 U DpnI at 37°C for 2 h. Then 2 μL of SDM reaction was transformed into 50 μL of NEB 5-alpha cells and overnight culture for mini-prep. Successful mutagenesis was confirmed by DNA sequencing.

10.42. Peptide Stability Testing

This section was provided by Dr William McCoull from AstraZeneca

Rat Hepatocyte Incubations

Hepatocyte incubation medium (L-15) was warmed to 37 °C in a water bath. Cryopreserved rat hepatocyte cells (BioreclamationIVT catalog. M00065, Lot PSO) were thawed and re-suspended in 50 mL of incubation medium. Cells were centrifuged for 3 min at 50 g and room temperature. The supernatant fraction was removed and viable cells were re-suspended to 1 x 10^6 /mL in incubation medium. Cell viability was determined using a Cellometer Vision and >80% viability required to proceed. Hepatocyte incubations were prepared in 1 mL 96-well plates to a final volume of 250 µL containing 0.25 x 10^6 cells/well. Hepatocyte incubations were shaken at 600 rpm at 50 °C to rise the temperature to 37 °C. the reaction was initiated through the addition of test compound (final concentration 1 µmol/L). The reaction was incubated at 900 rpm and 37 °C. The reaction was terminated at 0.5, 5, 15, 30, 45, 60, 80, 100 and 120 min by transferring 20 µL of incubation mixture into 100 µL of 100% acetonitrile containing internal standard (100, 200 and 100 nmol/L of alprazolam, caffeine and tolbutamide, respectively). Standards were prepared by diluting 100 µmol/L in 50:50 acetonitrile: water. All quenched samples were centrifuged for 20 min at 4000 rpm and 4 °C to pellet precipitated microsomal protein. 40 µL of the supernatant fraction was diluted with deionised water 1:4 and centrifuged for a further 15 min at 14000 rpm and 4 °C. 100 µL of the supernatant fraction was diluted 1:1 with deionised water and analysed by LC-MS/MS.

To determine the in vitro intrinsic clearance (CL_{int} (µL/min/x10^6 cells)) of parent compound, the slope value, k, is determined by linear regression of the natural logarithm of percent

parent remaining vs. incubation time plot. The CL_{int} is determined from the slope value using the following equation:

t1/2 = - (0.693 / k) (min)

Conversion of the t1/2 into the in vitro intrinsic clearance (CLint (µL/min/x10^6 cells)) was completed by the following equation:

CL_{int} = (0.693 / t1/2) * (V / N) ;V = incubation volume (0.25 mL); N = number of cells per well (0.25 x 10^6 cells).

Human Microsomal Incubations

Microsomal incubations were prepared in 1 mL 96-well plates to a final volume of 250 µL. Incubations contained 1 mg/mL human liver microsomes (Corning ultrapool 150 donors, catalog. 452117, Lot 38289), 1 mmol/L NADPH and 1 µM test compound in phosphate buffer pH 7.4. All incubations were completed in singlicate. Reactions were initiated through the addition of NADPH and incubated at 37 °C in a water bath. The reaction were terminated at 0.5, 5, 10, 15, 20 and 30 min by transferring 20 µL of incubation mixture into 100 µL of 100% acetonitrile containing internal standard (100, 200 and 100 nmol/L of alprazolam, caffeine and tolbutamide, respectively. Standards were prepared as detailed above but replacing phosphate buffer for NADPH. All quenched samples were centrifuged for 20 min at 4000 rpm and 4 °C to pellet precipitated microsomal protein. 40 µL of the supernatant fraction was diluted with deionised water 1:4 and centrifuged for a further 15 min at 14000 rpm and 4 °C. 100 µL of the supernatant fraction was diluted 1:1 with deionised water and analysed by LC-MS/MS. To determine the in vitro intrinsic clearance (CL_{int} (µL/min/x10^6 cells)) of parent compound, the slope value, k, is determined by linear regression of the natural

logarithm of percent parent remaining vs. incubation time plot. The CL_{int} is determined from the slope value using the following equation: The in vitro half-life (t1/2) was determined from the slope value: t1/2 = - (0.693 / k) (min)

Conversion of the t1/2 into the in vitro intrinsic clearance (CLint (µL/min/mg protein)) was completed by the following equation: CL_{int} = (0.693 / t1/2) * (V / N)

V = incubation volume (0.25 mL); N = amount of protein per well (1 mg/mL).

LC-MS/MS Methods

MS/MS instrument: Waters XEVO® TQ-S or Waters XEVO® TQ-D or API 4000 (AB Sciex). Ultra mass spectrometer was used to carry out the sample analysis in the MRM mode. (MS/MS). The LC conditions are shown in the Table. (Column: Waters XSelect HSS T3 C18, 2.5µm, 2.1 x 50mm, 40 °C, Solvent A: H_2O + 0.1 % formic acid, Solvent B: Acetonitrile + 0.1 % (v/v) formic acid). Flow = 0.5 mL/min.

Gradient	Time (min)	% A	% B
	0	99.8	0.2
		99.8	0.2
		5	95
		5	95
		99.8	0.2
	1.8	99.8	0.2

10.43. Physicochemical Property Testing

This section was provided by Dr William McCoull from AstraZeneca

$logD_{7.4}$: $logD_{7.4}$ measurements were made using a shake-flask method where the extent of partitioning between pH 7.4 buffer and octanol was measured. Compounds were dissolved in a known volume buffer, and following the addition of a known amount of octanol, the solutions were shaken for 30 min. Following centrifugation, analysis of the aqueous layer was performed by LC-UV to quantify the amount of compound in solution and then compared to analysis of the compound in solution before the addition of octanol to calculate the partitioning coefficient, $D_{7.4}$.

Solubility: Samples were incubated for 24 h in pH 7.4 phosphate buffer. After centrifugation, analysis of the supernatant liquid was performed by LC-UV to quantify the amount of compound in solution.

Permeability measurement: Caco-2 cells (American Type Culture Collection number HTB-37) were grown on Transwell inserts (24 well HTS Transwell microplate, Corning Inc.) for 14 d until a confluent monolayer of differentiated cells was achieved. Test compounds were dissolved in DMSO to a concentration of 2 mM and then further diluted in Hanks Balanced Salt Solution (HBSS) (containing 25 mM 4-morpholineethanesulfonic acid (MES), pH 6.5, 50 µM quinidine, 30 µM benzbromarone and 20 µM sulfasalazine) to give a final concentration of 5 µM. The culture medium was removed from the Transwell and the test compound solution (210 µL) was added to the Transwell insert (apical compartment) and 800 µL of HBSS receiver buffer (containing 25 mM *N*-(2-hydroxyethyl)piperazine-*N*'-(2-ethanesulfonic acid) (HEPES), pH 7.4, 50 µM quinidine, 30 µM benzbromarone, 20 µM sulfasalazine) added to the wells in the receiver plate (basolateral compartment). The plate was incubated at 37 °C with rotary

shaking at 480 rpm. Samples (10 µL) were taken from the apical (donor) side immediately after test compound solution addition, and following 45 and 120 min of incubation. These were diluted in HBSS receiver buffer (90 µL). Samples (100 µL) were taken from the basolateral (receiver) side following 45 and 120 min of incubation. Fresh HBSS receiver buffer (100 µL) was added to replace the volume taken. The samples were protein precipitated with acetonitrile (300 µL) containing internal standards (alprazolam 100 nM, caffeine 200 nM and tolbutamide 100 nM), vortexed for 5 min, and centrifuged (4000 rpm for 20 min). The supernatant was analysed for test compound concentration against a standard curve from 2.1 to 1500 nM using LC/MS/MS. The apparent permeability (P_{app}) in units of centimetres per second, was calculated using equation where C_R is the concentration of the receiver side, C_D is the concentration of the donor side, at the respective time points, *area* is the is the surface area of the Transwell insert membrane (0.33 cm^2), and *time* is the time between difference between the two samplings (4500 s).

$$P_{app} = \frac{C_R^{120} \times 0.8 - C_R^{45} \times 0.7}{(C_D^{45} + C_D^{120})/2 \times area \times time}$$

Chapter 11. Appendix

11.1. Human KDM7 Canonical Domain Sequences (UniProt)

11.1.1. KDM7A (Isoform 1, Q6ZMT4-1)

PHD Finger: AA Residues 37-88

PVYCVCRQPYDVNRFMIECDICKDWFHGSCVGVEEHHAVDIDLYHCPNCAVL

JmjC: AA Residues 230-386

FSDTKMSELVEVPDIAKKLSWVENYWPDDSVFPKPFVQKYCLMGVQDSYTDFHIDFGGTSVW

YHVLWGEKIFYLIKPTDENLARYESWSSSVTQSEVFFGDKVDKCYKCVVKQGHTLFVPTGWI

HAVLTSQDCMAFGGNFLHNLNIGMQLRCYEMEK

11.1.2. KDM7B (Isoform 2, Q9UPP1-2)

PHD Finger: AA Residues 4-56

PVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVL

JmjC: AA Residues 195-351

FSDTRLSNLVETPKIVRKLSWVENLWPEECVFERPNVQKYCLMSVRDSYTDFHIDFGGTSVW

YHVLKGEKIFYLIRPTNANLTLFECWSSSSNQNEMFFGDQVDKCYKCSVKQGQTLFIPTGWI

HAVLTPVDCLAFGGNFLHSLNIEMQLKAYEIEK

11.1.3. KDM7C (Isoform 1, O75151-1)

PHD Finger: AA Residues 5-56

PVYCVCRLPYDVTRFMIECDACKDWFHGSCVGVEEEEAPDIDIYHCPNCEKT

JmjC: AA Residues 197-353

FSDTRMSSFVEPPDIVKKLSWVENYWPDDALLAKPKVTKYCLICVKDSYTDFHIDSGGASAW
YHVLKGEKTFYLIRPASANISLYERWRSASNHSEMFFADQVDKCYKCIVKQGQTLFIPSGWI
YATLTPVDCLAFAGHFLHSLSVEMQMRAYEVER

11.1.4. KDM7B (Isoform 1, Q9UPP1-1)

Has the following sequence appended to the N-terminus of Isoform-2:

MNRSRAIVQRGRVLPPPAPLDTTNLAGRRTLQGRAK

11.2. Protein Construct Information

All plasmids used in expression by O.D.Coleman and Yijia Zhang were sequence-verified by DNA Sanger Sequencing prior to use. The following sequences detail the amino acid residues pertaining to the target gene, additional fusion tags were present (Expression Plasmid) unless cleaved.

The tag sequences (noted in the Expression Plasmid) that can be appended directly to a given protein sequence to generate the full sequence that was produced are as follows:

His_6+Thioredoxin: (N-terminal)

MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQG
KLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGTENL
YFQSM

His_6: (N-terminal)

MHHHHHHSSGVDLGTENLYFQSM

His_6: (C-terminal)

AENLYFQSHHHHHHDYKDDDDK

AviTag: (C-terminal)

GSKGGYGLNDIFEAQKIEWHE

His_6+ZBasic: (N-terminal)

MHHHHHHSSGVDNKFNKERRRARREIRHLPNLNREQRRAFIRSLRDDPSQSANLLAEAKKLN
DAQPKGTENLYFQSM

His_6+GST:

MHHHHHHSSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFP
NLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDF
ETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLV
CFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSSSGVDLGTENLYFQSM

Note: the TEV recognition sequence (ENLYFQS) results in cleavage at: ENLYFQ_S, leaving a Ser residue at the N-terminus, or alternatively ENLYFQ if tag placed at the C-terminus, after TEV cleavage.

11.2.1. KDM7B$_{[PHD+JmjC]}$-Biotin

HisTag TEV recognition sequence Thioredoxin KDM7B$_{[PHD+JmjC]}$ linker AviTag

UniProt Identifier: HS(Q9UPP1-2). **Expression Plasmid:** pNIC-Bio3 (N-terminal His$_6$ and C-terminal AviTag). **AA Residues:** 1-447

MHHHHHHSSGVDLGTENLYFQSMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIA
PILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKE
FLDANLAGTENLYFQSMASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADID
LYHCPNCEVLHGPSIMKKRRGSSKGHDTHKGKPVKTGSPTFVRELRSRTFDSSDEVILKPTG
NQLTVEFLEENSFSVPILVLKKDGLGMTLPSPSFTVRDVEHYVGSDKEIDVIDVTRQADCKM
KLGDFVKYYYSGKREKVLNVISLEFSDTRLSNLVETPKIVRKLSWVENLWPEECVFERPNVQ
KYCLMSVRDSYTDFHIDFGGTSVWYHVLKGEKIFYLIRPTNANLTLFECWSSSSNQNEMFFG
DQVDKCYKCSVKQGQTLFIPTGWIHAVLTPVDCLAFGGNFLHSLNIEMQLKAYEIEKRLSTA
DLFRFPNFETICWYVGKHILDIFRGLRENRRHPASYLVHGGKALNLAFRAWTRKEALPDHED
EIPETVRTVQLIKDLAREIRLVEDIFQQNSSKGGYGLNDIFEAQKIEWHE

*AvigTag is biotinylated during protein production

11.2.2. KDM7A$_{[PHD]}$

Gift from Yijia Zhang

UniProt Identifier: HS(Q6ZMT4-1). **Expression Plasmid:** pNH-TrxT (N-terminal His$_6$+Thioredoxin). **AA Residues:** 38-98:

VYCVCRQPYDVNRFMIECDICKDWFHGSCVGVEEHHAVDIDLYHCPNCAVLHGSSLMKKRR

11.2.3. KDM7A[PHD+JmjC]

UniProt Identifier: HS(Q6ZMT4-1). **Expression Plasmid:** pNH-TrxT (N-terminal His_6+Thioredoxin). **AA Residues:** 38-480:

VYCVCRQPYDVNRFMIECDICKDWFHGSCVGVEEHHAVDIDLYHCPNCAVLHGSSLMKKRRN
WHRHDYTEIDDGSKPVQAGTRTFIKELRSRVFPSADEIIIKMHGSQLTQRYLEKHGFDVPIM
VPKLDDLGLRLPSPTFSVMDVERYVGGDKVIDVIDVARQADSKMTLHNYVKYFMNPNRPKVL
NVISLEFSDTKMSELVEVPDIAKKLSWVENYWPDDSVFPKPFVQKYCLMGVQDSYTDFHIDF
GGTSVWYHVLWGEKIFYLIKPTDENLARYESWSSSVTQSEVFFGDKVDKCYKCVVKQGHTLF
VPTGWIHAVLTSQDCMAFGGNFLHNLNIGMQLRCYEMEKRLKTPDLFKFPFFEAICWFVAKN
LLETLKELREDGFQPQTYLVQGVKALHTALKLWMKKELVSEHAFEIPDNVRPGHLIKELSKV
IRAIEEENG

11.2.4. KDM7B[PHD]

UniProt Identifier: HS(Q9UPP1-2). **Expression Plasmid:** pNH-TrxT (N-terminal His_6+Thioredoxin). **AA Residues:** 1-66:

MASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVLHGPSIM
KKRR

11.2.5. KDM7B[JmjC]

Gift from Yijia Zhang

UniProt Identifier: HS(Q9UPP1-2). **Expression Plasmid:** pNH-TrxT (N-terminal His_6+Thioredoxin). **AA Residues:** 78-447:

PVKTGSPTFVRELRSRTFDSSDEVILKPTGNQLTVEFLEENSFSVPILVLKKDGLGMTLPSP
SFTVRDVEHYVGSDKEIDVIDVTRQADCKMKLGDFVKYYYSGKREKVLNVISLEFSDTRLSN
LVETPKIVRKLSWVENLWPEECVFERPNVQKYCLMSVRDSYTDFHIDFGGTSVWYHVLKGEK
IFYLIRPTNANLTLFECWSSSSNQNEMFFGDQVDKCYKCSVKQGQTLFIPTGWIHAVLTPVD
CLAFGGNFLHSLNIEMQLKAYEIEKRLSTADLFRFPNFETICWYVGKHILDIFRGLRENRRH
PASYLVHGGKALNLAFRAWTRKEALPDHEDEIPETVRTVQLIKDLAREIRLVEDIFQQN

11.2.6. KDM7B[PHD+JmjC]

UniProt Identifier: HS(Q9UPP1-2). **Expression Plasmid:** pNH-TrxT (N-terminal His_6+Thioredoxin). **AA Residues:** 1-447:

MASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVLHGPSIM
KKRRGSSKGHDTHKGKPVKTGSPTFVRELRSRTFDSSDEVILKPTGNQLTVEFLEENSFSVP
ILVLKKDGLGMTLPSPSFTVRDVEHYVGSDKEIDVIDVTRQADCKMKLGDFVKYYYSGKREK
VLNVISLEFSDTRLSNLVETPKIVRKLSWVENLWPEECVFERPNVQKYCLMSVRDSYTDFHI
DFGGTSVWYHVLKGEKIFYLIRPTNANLTLFECWSSSSNQNEMFFGDQVDKCYKCSVKQGQT
LFIPTGWIHAVLTPVDCLAFGGNFLHSLNIEMQLKAYEIEKRLSTADLFRFPNFETICWYVG
KHILDIFRGLRENRRHPASYLVHGGKALNLAFRAWTRKEALPDHEDEIPETVRTVQLIKDLA
REIRLVEDIFQQN

11.2.7. KDM7C[PHD]

Gift from Yijia Zhang

UniProt Identifier: HS(O75151). **Expression Plasmid:** pNIC-ZB (N-terminal His_6+ZBasic). **AA Residues:** 1-67:

MATVPVYCVCRLPYDVTRFMIECDACKDWFHGSCVGVEEEEAPDIDIYHCPNCEKTHGKSTL

KKKRT

11.2.8. KDM7C[PHD+JmjC]

Gift from Yijia Zhang

UniProt Identifier: HS(O75151). **Expression Plasmid:** pNIC-ZB (N-terminal His_6+ZBasic). **AA Residues:** 1-449:

MATVPVYCVCRLPYDVTRFMIECDACKDWFHGSCVGVEEEEAPDIDIYHCPNCEKTHGKSTL

KKKRTWHKHGPGQAPDVKPVQNGSQLFIKELRSRTFPSAEDVVARVPGSQLTLGYMEEHGFT

EPILVPKKDGLGLAVPAPTFYVSDVENYVGPERSVDVTDVTKQKDCKMKLKEFVDYYYSTNR

KRVLNVTNLEFSDTRMSSFVEPPDIVKKLSWVENYWPDDALLAKPKVTKYCLICVKDSYTDF

HIDSGGASAWYHVLKGEKTFYLIRPASANISLYERWRSASNHSEMFFADQVDKCYKCIVKQG

QTLFIPSGWIYATLTPVDCLAFAGHFLHSLSVEMQMRAYEVERRLKLGSLTQFPNFETACWY

MGKHLLEAFKGSHKSGKQLPPHLVQGAKILNGAFRSWTKKQALAEHEDELPEHFKPSQLIKD

LAKEIRLSENASKAV

11.2.9. DIDO1[PHD]

Cloned by the author. **UniProt Identifier:** HS(Q9BTC0-4). **Expression Plasmid:** pNH-TrxT (N-terminal His_6+Thioredoxin). **AA Residues:** 260-325:

ECEGYDPNALYCICRQPHNNRFMICCDRCEEWFHGDCVGISEARGRLLERNGEDYICPNCTI

LQVQQ

11.2.10. AIRE[PHD1]

Cloned by the author. **UniProt Identifier:** HS(O43918). **Expression Plasmid:** pNIC-Bio3 (N-terminal His_6 and C-terminal AviTag). **AA Residues:** 290-354:

QLHQKNEDECAVCRDGGELICCDGCPRAFHLACLSPPLREIPSGTWRCSSCLQATVQEVQPR
AEE

11.2.11. ING2[PHD]

Gift from the SGC. **UniProt Identifier:** HS(Q9H160). **Expression Plasmid:** pGTVL2 (N-terminal His_6+GST). **AA Residues:** 171-226.

11.2.12. TAF3 [PHD]

Gift from the SGC. **UniProt Identifier:** HS(Q5VWG9). **Expression Plasmid:** pGTVL2 (N-terminal His_6+GST). **AA Residues:** 857-924.

11.2.13. SPIN1[TTD]

Gift from the SGC. **UniProt Identifier:** HS(Q9Y657). **Expression Plasmid:** pNIC-CTHF (C-terminal His_6). **AA Residues:** 26-262.

11.2.14. KDM2A[Full-Length]

Purchased commercially (31485, Active Motif). **UniProt Identifier:** HS(Q9Y2K7) **AA Residues:** 1-1162.

11.2.15. KDM2B[Full-Length]

Purchased commercially (31455, Active Motif). **UniProt Identifier:** HS(Q8NHM5) **AA Residues:** 1-1336.

11.2.16. KDM3A[JmjC+ZincFinger]

Gift from Dr Rachel Schamo-Schiller (Kawamura Lab). **UniProt Identifier:** HS(Q9Y4C1). **Expression Plasmid:** pFB-CT10HF-LIC. **AA Residues:** 515-1317.

11.2.17. KDM4A[JmjC+JmjN]

Gift from Eidarus Salah (Kawamura Lab). **UniProt Identifier:** HS(O75164). **Expression Plasmid:** pNIC28-Bsa4. **AA Residues:** 1-359.

11.2.18. KDM5C[JmjC+PHD+ZincFinger+JmjN+ARID]

Gift from the SGC. **UniProt Identifier:** HS(P41229). **Expression Plasmid:** pFB-CT10HF-LIC. **AA Residues:** 1-765.

11.2.19. KDM6B[JmjC]

Gift from Eidarus Salah (Kawamura Lab). **UniProt Identifier:** HS(O15054). **Expression Plasmid:** pNH-TrxT. **AA Residues:** 1141-1641.

11.2.20. EGLN1[Fe-2OG-Dioxygenase]

Gift from Dr Tom McAllister (Kawamura Lab). **UniProt Identifier:** HS(Q9GZT9). **Expression Plasmid:** pNIC-Bio3. **AA Residues:** 181-426.

11.3. Protein Mass Spectrometry Data

Data is shown for recombinantly expressed proteins (E. coli) produced by O.D.Coleman (the author-designated as 'Made'). Gifted proteins were characterised by others.

	Molecular Mass	
Expressed Protein Construct	**Calculated**	**Observed**
His-Trx-KDM7A$_{[PHD+JmjC]}$	65533	65538
His-KDM7B$_{[PHD+JmjC]}$-AviTag-Biotinylated	54237	54240
His-Trx-KDM7B$_{[PHD]}$	21655	21656
KDM7B$_{[PHD]}$ (^{15}N ^{13}C labelled)	8153	8140
His-Trx-KDM7B$_{[PHD+JmjC]}$	65547	65551
His-Trx-DIDO1$_{[PHD]}$	21975	21975
His-AIRE$_{[PHD1]}$-AviTag-Biotinylated	12487	12486
KDM7 PHD Fingers Used in ITC:		
KDM7A$_{[PHD]}$	7347	7346
KDM7B$_{[PHD]}$	7728	7727
KDM7C$_{[PHD]}$	7786	7785

11.4. List of Protein Contributors

Tags?	Protein	Domain Construct	Sources
His+Trx	KDM7A	PHD	Yijia Zhang
Cleaved	KDM7A	PHD	Yijia Zhang
His+Trx	KDM7A	PHD+JmjC	Made
Cleaved	KDM7A	PHD+JmjC	Made
His+Trx	KDM7B	PHD	Made and Yijia Zhang
Cleaved	KDM7B	PHD	Made and Yijia Zhang
Cleaved	KDM7B	PHD (^{15}N labelled)	Finn Wolfreys
Cleaved	KDM7B	PHD (^{15}N, ^{13}C labelled)	Made
His+Trx	KDM7B	JmjC	Yijia Zhang
Cleaved	KDM7B	JmjC	Yijia Zhang
His+Trx	KDM7B	PHD+JmjC	Made and Yijia Zhang
Cleaved	KDM7B	PHD+JmjC	Made and Yijia Zhang
AviTagBiotin	KDM7B	PHD+JmjC	Made
His+ZB	KDM7C	PHD	Yijia Zhang
Cleaved	KDM7C	PHD	Yijia Zhang
His+ZB	KDM7C	PHD+JmjC	Yijia Zhang
Cleaved	KDM7C	PHD+JmjC	Yijia Zhang
His, AviTagBiotin	AIRE	PHD1	Made
His+Trx	DIDO1	PHD	Made
His+GST	ING2	PHD	SGC
His+GST	TAF3	PHD	SGC
His	SPIN1	TTD	SGC
	KDM2A	Full-Length	ActiveMotif (31485)
	KDM2B	Full-Length	ActiveMotif (31455)
	KDM3A	JmjC+ZincFinger	Rachel Schamo-Schiller
	KDM4A	JmjC+JmjN	Eidarus Salah
	KDM5C	JmjC+PHD+ZincFinger+JmjN+ARID	SGC
	KDM6B	JmjC	Eidarus Salah
	EGLN1	Fe-2OG-Dioxygenase	Tom McAllister

Table of proteins and their sources. Made = produced by the author (O.D.Coleman).

11.5. Cyclic Peptide High-Resolution Mass Spectrometry

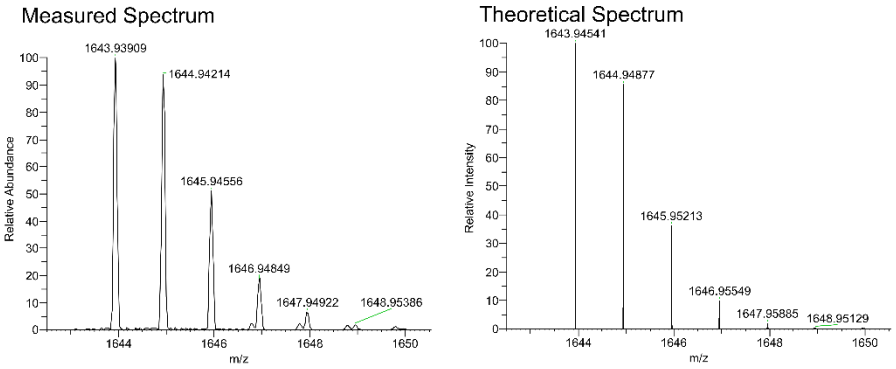
11.5.1. HRMS Summary Table

Peptide Name	Calculated Exact Mass [M+H]	Calculated Exact Mass [M+2H]	Observed m/z	Δ ppm	Molecular Formula	RDB
OC3	1643.94541		1643.93909	-3.85	C79 H127 O16 N20 S1	26.5
OC6	1906.01964		1906.0155	-2.17	C95 H137 O18 N22 S1	38.5
OC9	1785.84174		1785.83655	-2.91	C85 H117 O21 N20 S1	37.5
OC9_Biotin		1415.15980	1415.15894	-0.61	C127 H187 O37 N33 S2	51
OC9_Linear		894.43233	894.43201	-0.36	C85 H120 O21 N20 S1	36
OC9_Scrambled		893.42451	893.42383	-0.76	C85 H118 O21 N20 S1	37
OC9_Y1F		885.42705	885.42664	-0.47	C85 H118 O20 N20 S1	37
OC9_Y2F		885.42705	885.42566	-1.57	C85 H118 O20 N20 S1	37
OC9_Y7F		885.42705	885.42511	-2.19	C85 H118 O20 N20 S1	37
OC9_Y127F		869.43214	869.43176	-0.43	C85 H118 O18 N20 S1	37
OC9_T4 H5 P8 A		832.40050	832.40143	1.11	C79 H112 O20 N28 S1	33
OC9.2.0		821.41595	821.41614	0.22	C81 H114 O17 N18 S1	34
OC9_T4-dA		878.41922	878.41864	-0.67	C84 H116 O20 N20 S1	37
OC9_H5-dA		860.41361	860.41486	1.45	C82 H116 O21 N18 S1	34
OC9_P8-dA		880.41668	880.41528	-1.59	C83 H116 O21 N20 S1	36
OC9_trans-4F-P6		902.41980	902.41925	-0.60	C85 H117 O21 N20 F S1	37
OC9_cis-4F-P6		902.41980	902.41882	-1.08	C85 H117 O21 N20 F S1	37
OC9_Short	1035.50403		1035.50476	0.71	C44 H71 O13 N14 S1	16.5
OC9 Y2A	1693.81552		1693.81274	-1.64	C79 H113 O20 N20 S1	33.5
OC9 L3A	1743.79479		1743.79248	-1.32	C82 H111 O21 N20 S1	37.5
OC9 T4A	1755.83117		1755.82849	-1.53	C84 H115 O20 N20 S1	37.5
OC9 H5A	1719.81994		1719.81628	-2.13	C82 H115 O21 N18 S1	34.5
OC9 P6A	1759.82609		1759.82373	-1.34	C83 H115 O21 N20 S1	36.5
OC9 Y7A	1693.81552		1693.81213	-2.00	C79 H113 O20 N20 S1	33.5
OC9 P8A		880.41668	880.41522	-1.66	C83 H116 O21 N20 S1	36
OC9 S9A	1769.84682		1769.84595	-0.49	C85 H117 O20 N20 S1	37.5
OC9 V10A	1757.81044		1757.80896	-0.84	C83 H113 O21 N20 S1	37.5
OC9 F11A		855.40886	855.40839	-0.55	C79 H114 O21 N20 S1	33
OC9 R12A		850.89251	850.89185	-0.78	C82 H111 O21 N17 S1	36
OC9 T13A	1755.83117		1755.82825	-1.67	C84 H115 O20 N20 S1	37.5

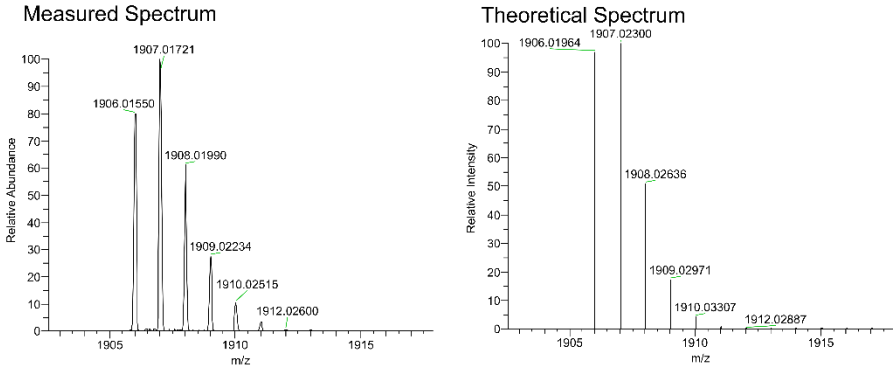
Table of data summarising the high-resolution ESI-MS used to characterise each synthesised peptide

11.5.2. Example HRMS Spectra

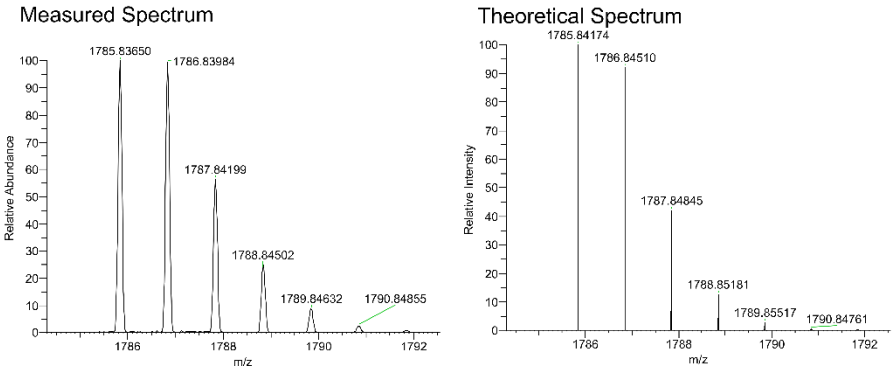
OC3
Calculated m/z: 1643.94541
Observed m/z: 1643.93909
Delta ppm: -3.85
Formula: C79 H127 O16 N20 ^{32}S
RDB: 26.5



OC6
Calculated m/z: 1906.01964
Observed m/z: 1906.01550
Delta ppm: -2.17
Formula: C95 H137 O18 N22 ^{32}S
RDB: 38.5



OC9
Calculated m/z: 1785.84174
Observed m/z: 1785.83650
Delta ppm: -2.91
Formula: C85 H117 O21 N20 ^{32}S
RDB: 37.5



Example spectra for high-resolution ESI-MS of the three cyclic peptides of primary interest

11.6. Cyclic Peptide MALDI-TOF-MS

11.6.1. MALDI-TOF-MS Summary Table

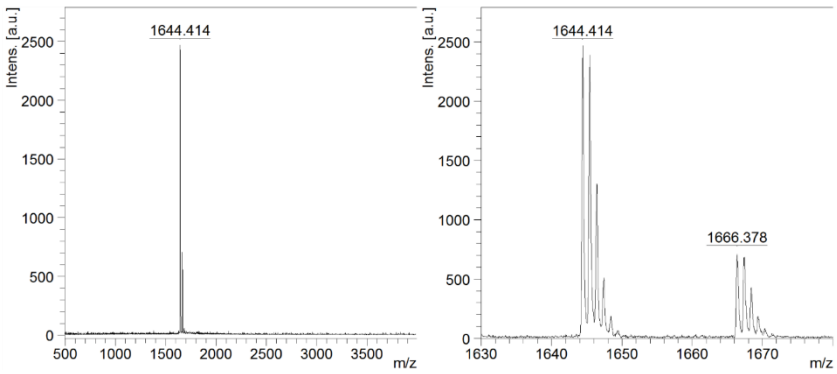
Peptide Name	Mw	Exact Mass	[M+H]			[M+Na]			[M+K]		
			Calc	Obs	Δ	Calc	Obs	Δ	Calc	Obs	Δ
OC3	1644.061	1642.938	1643.946	1644.414	-0.468	1665.928	1666.378	-0.450	1681.902	-	-
OC6	1906.329	1905.012	1906.020	1906.433	-0.413	1928.002	1928.437	-0.435	1943.976	1944.347	-0.371
OC9	1786.042	1784.834	1785.842	1786.151	-0.309	1807.824	1808.110	-0.286	1823.798	1824.101	-0.303
OC9.2.0	1641.560	1640.817	1641.825	1641.758	0.067	1663.807	1663.754	0.053	1679.781	1679.717	0.064
OC9 T4 H5 P8 A	1663.150	1662.786	1663.794	1664.002	-0.208	1685.776	1685.670	0.106	1701.750	-	-
OC9 Y127F	1738.045	1736.850	1737.858	1738.138	-0.280	1759.840	1760.055	-0.215	1775.814	1776.019	-0.205
OC9 Y1F	1770.043	1768.840	1769.848	1770.632	-0.784	1791.830	1792.314	-0.484	1807.804	-	-
OC9 Y2F	1770.043	1768.840	1769.848	1770.482	-0.634	1791.830	1792.461	-0.631	1807.804	-	-
OC9 Y7F	1770.043	1768.840	1769.848	1770.598	-0.750	1791.830	-	-	1807.804	-	-
OC9 T4a	1756.016	1754.824	1755.832	1755.898	-0.066	1777.814	1777.853	-0.039	1793.788	1793.817	-0.029
OC9 H5a	1719.979	1718.813	1719.821	1719.475	0.346	1741.803	1741.476	0.327	1757.777	-	-
OC9 P8a	1760.004	1758.819	1759.827	1759.947	-0.120	1781.809	1781.937	-0.128	1797.783	1797.890	-0.107
OC9 Linear	1788.058	1786.850	1787.858	1788.356	-0.498	1809.840	-	-	1825.814	-	-
OC9 Scrambled	1786.042	1784.834	1785.842	1786.344	-0.502	1807.824	-	-	1823.798	-	-
OC9-cis-4F-P6	1804.032	1802.825	1803.833	1804.199	-0.366	1825.815	1826.135	-0.320	1841.789	-	-
OC9-trans-4F-P6	1804.032	1802.825	1803.833	1804.226	-0.393	1825.815	1826.135	-0.320	1841.789	-	-
OC9-azide	2372.605	2371.080	2372.088	2372.595	-0.507	2394.070	-	-	2410.044	-	-
OC9-biotin	2830.191	2828.305	2829.313	2827.599	1.714	2851.295	-	-	2867.269	-	-
OC9 Short	1035.189	1034.497	1035.505	1035.389	0.116	1057.487	-	-	1073.461	1073.350	0.111
OC9 Y2A	1693.945	1692.808	1693.816	1693.743	0.073	1715.798	1715.733	0.065	1731.772	1731.728	0.044
OC9 L3A	1743.961	1742.788	1743.796	1744.012	-0.216	1765.778	1765.980	-0.202	1781.752	1781.937	-0.185
OC9 T4A	1756.016	1754.824	1755.832	1756.072	-0.240	1777.814	1778.038	-0.224	1793.788	1794.011	-0.223
OC9 H5A	1719.979	1718.813	1719.821	1720.095	-0.274	1741.803	1742.064	-0.261	1757.777	1758.028	-0.251
OC9 P6A	1760.004	1758.819	1759.827	1760.177	-0.350	1781.809	1782.169	-0.360	1797.783	1798.084	-0.301
OC9 Y7A	1693.945	1692.808	1693.816	1694.120	-0.304	1715.798	1716.112	-0.314	1731.772	1732.109	-0.337
OC9 P8A	1760.004	1758.819	1759.827	1760.177	-0.350	1781.809	1782.169	-0.360	1797.783	1798.122	-0.339
OC9 S9A	1770.043	1768.840	1769.848	1770.213	-0.365	1791.830	1792.190	-0.360	1807.804	1808.149	-0.345
OC9 V10A	1757.988	1756.803	1757.811	1758.105	-0.294	1779.793	1780.045	-0.252	1795.767	1796.066	-0.299
OC9 F11A	1709.944	1708.803	1709.811	1710.127	-0.316	1731.793	1732.109	-0.316	1747.767	1748.027	-0.260
OC9 R12A	1700.932	1699.770	1700.778	1701.095	-0.317	1722.760	1723.056	-0.296	1738.734	1739.009	-0.275
OC9 T13A	1756.016	1754.824	1755.832	1756.110	-0.278	1777.814	1778.038	-0.224	1793.788	-	-

Summarised table of data for recorded masses when analysing peptide purity by MALDI-TOF-MS

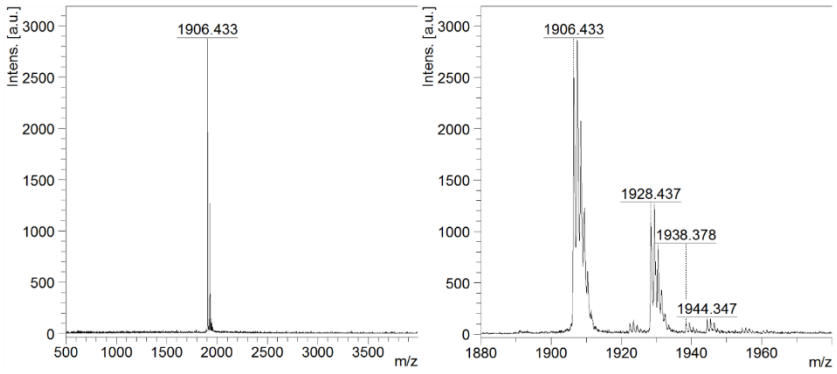
11.6.2. MALDI-TOF-MS Spectra

The following are spectra that were acquired to assess the purity of the purified peptide. On the left is a wide-mass-range spectrum, with the right hand-window a zoom view of the peak of interest.

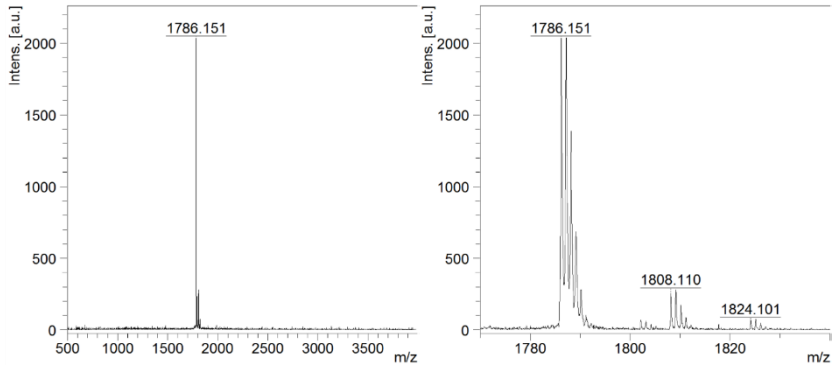
OC3



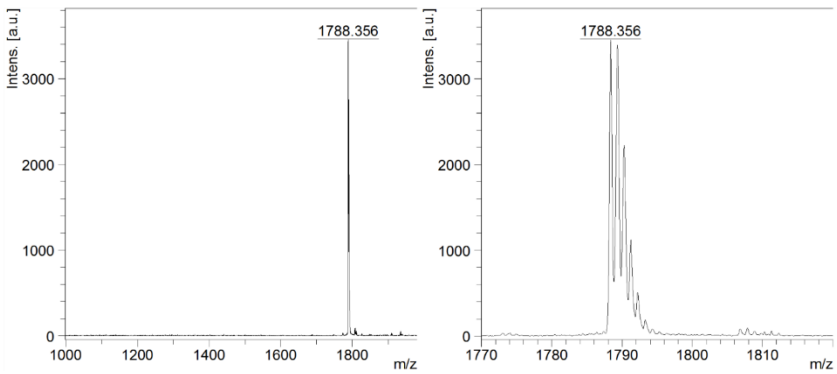
OC6



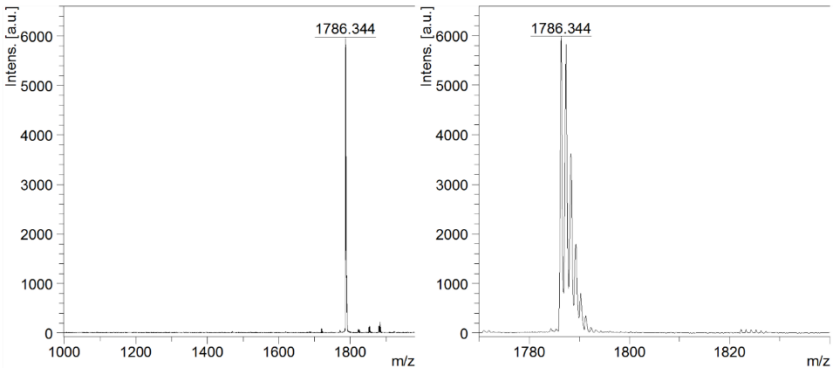
OC9



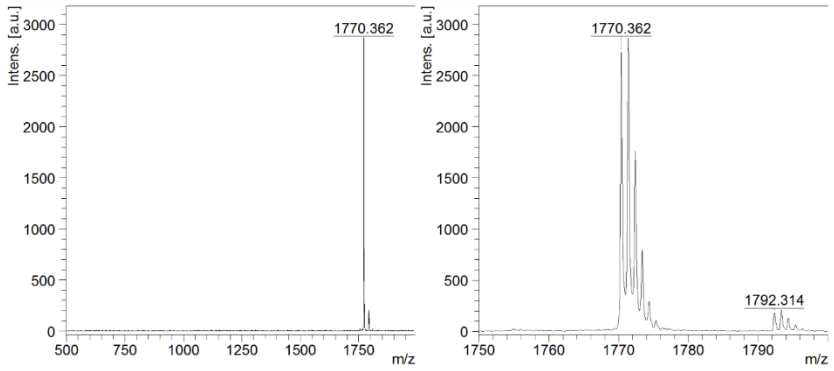
OC9 LINEAR



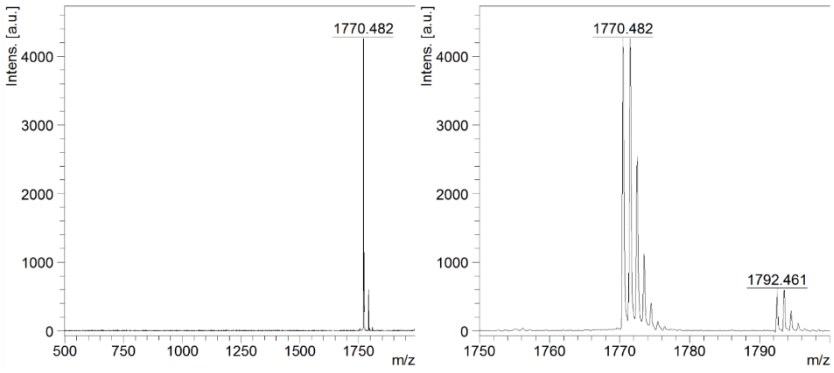
OC9 SCRAMBLED



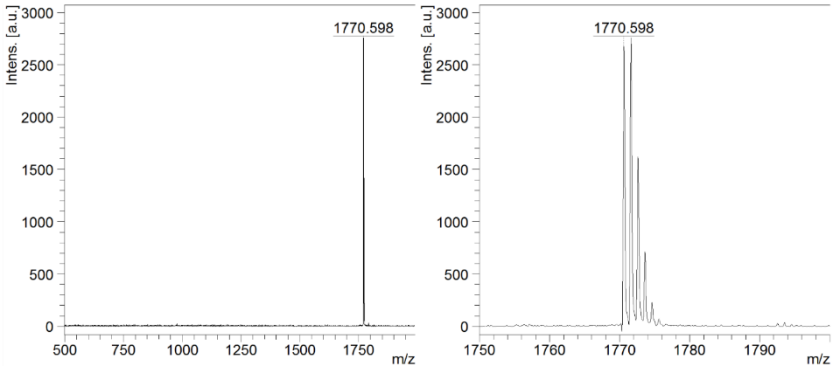
OC9 Y1F



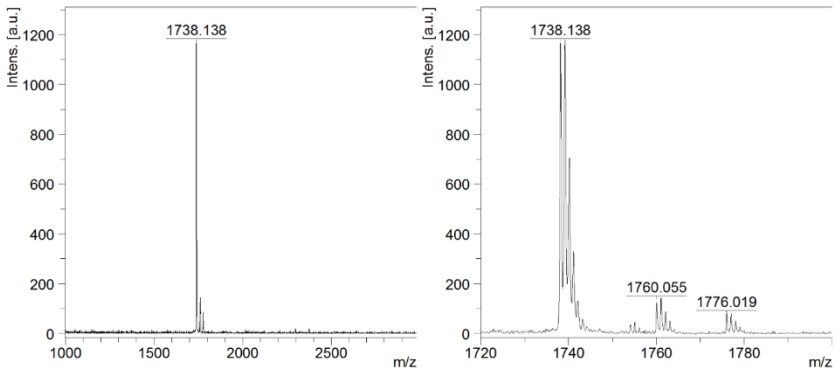
OC9 Y2F



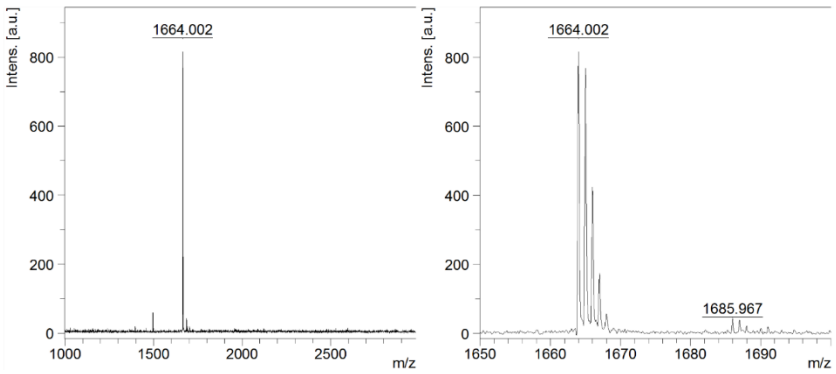
OC9 Y7F



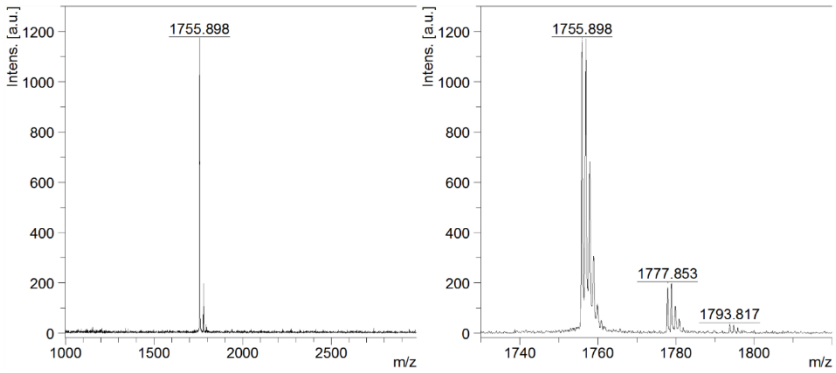
OC9 Y127F



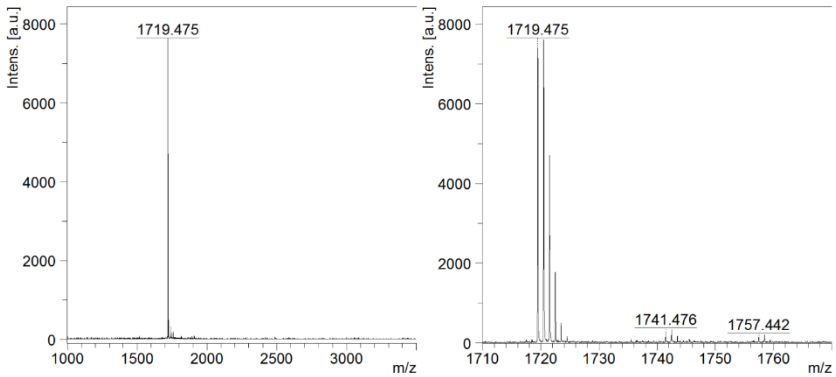
OC9 T4 H5 P8 A



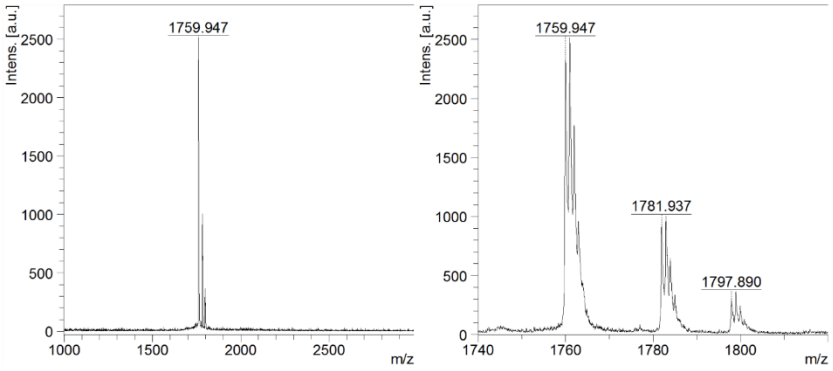
OC9 T4^{D}A



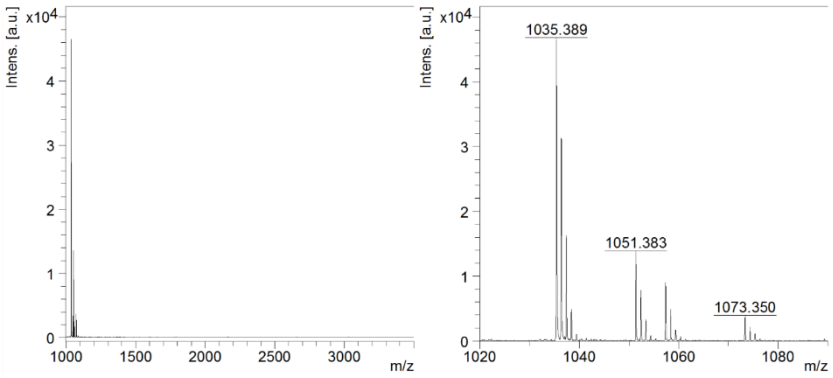
OC9 H5^{D}A



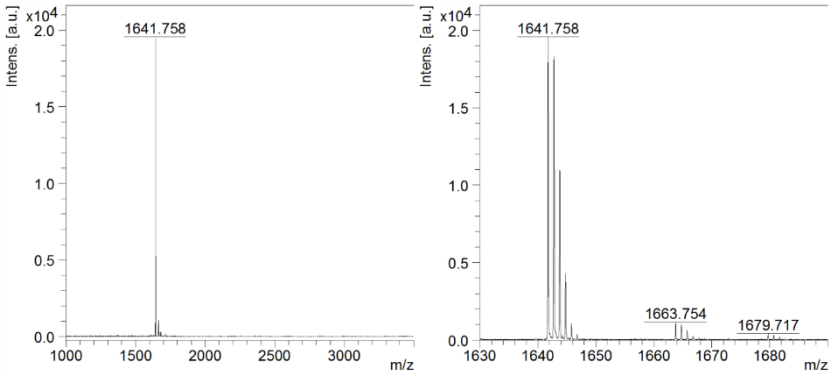
OC9 P8^{D}A



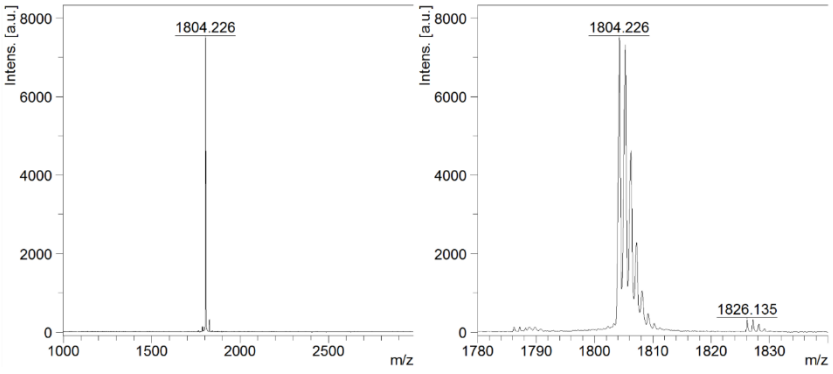
OC9 SHORT



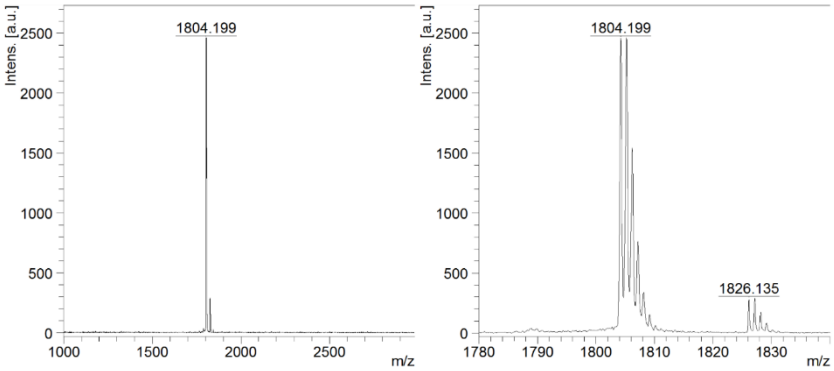
OC9.2.0



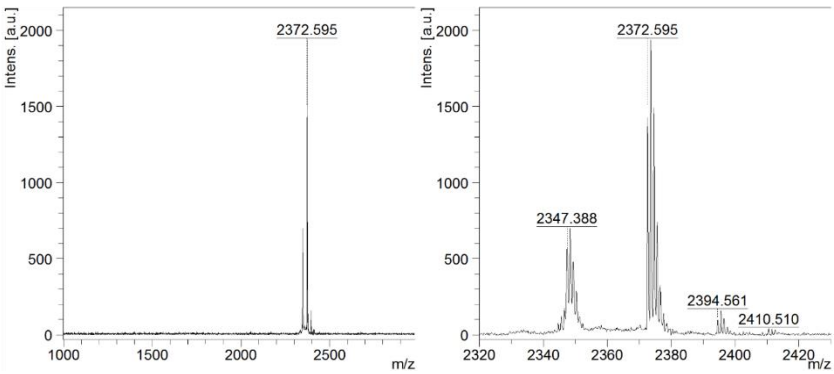
OC9 TRANS-4F-P6



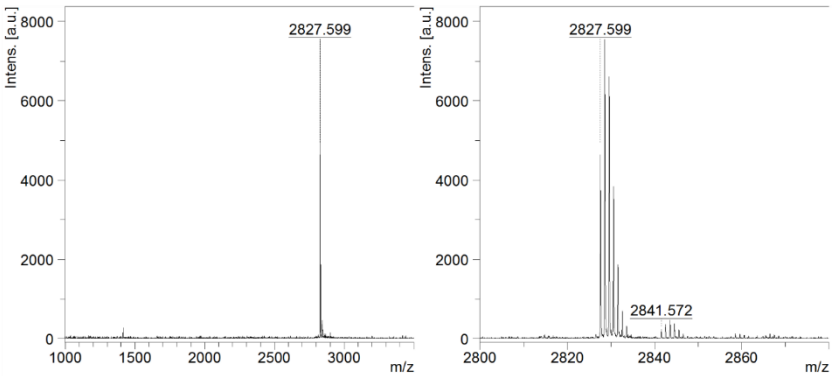
OC9 CIS-4F-P6



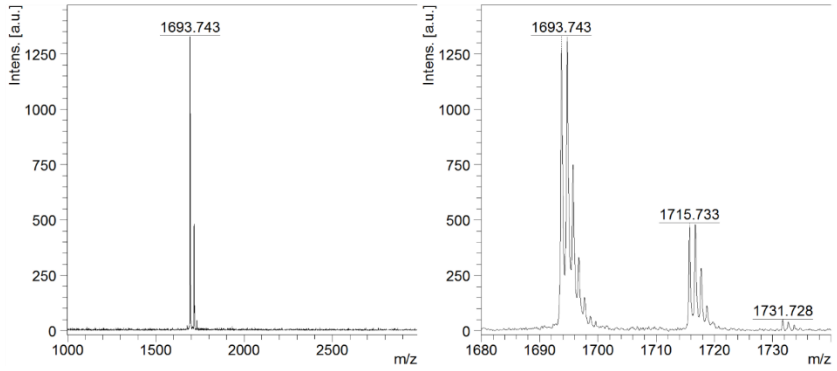
OC9 AZIDE



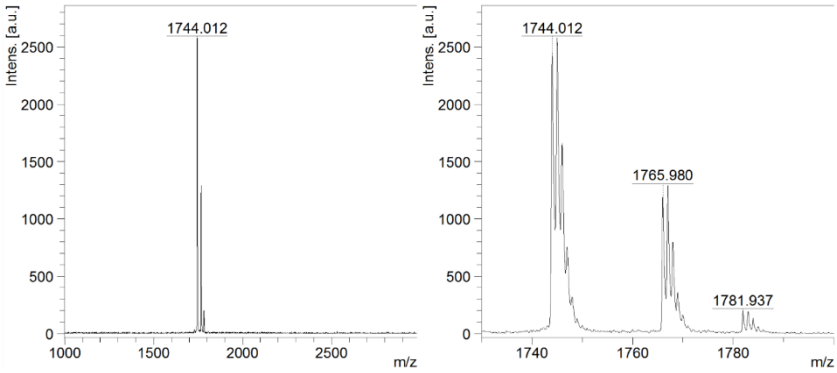
OC9 BIOTIN



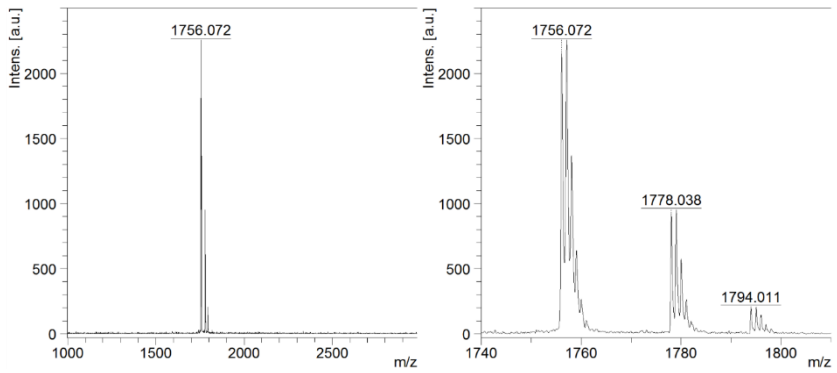
OC9 Y2A



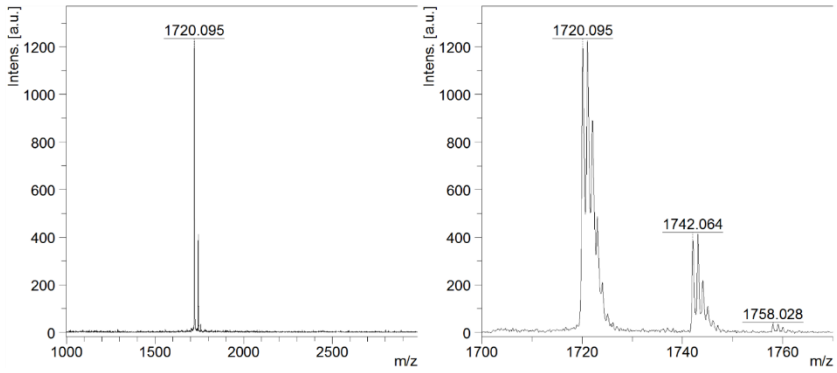
OC9 L3A



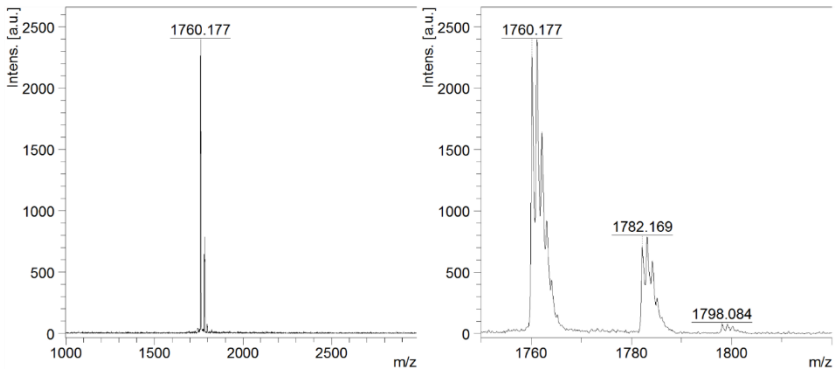
OC9 T4A



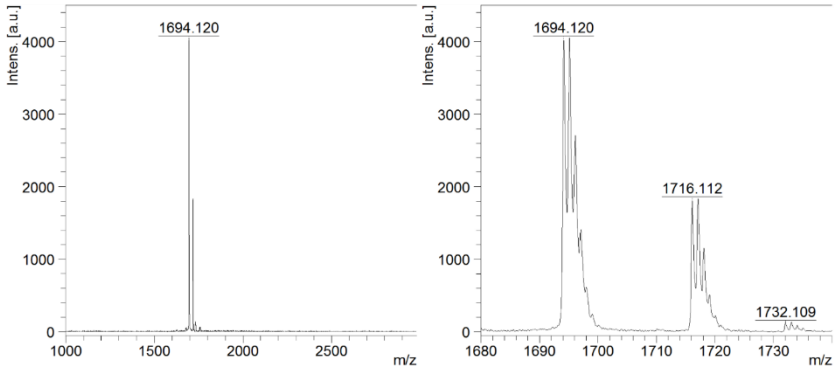
OC9 H5A



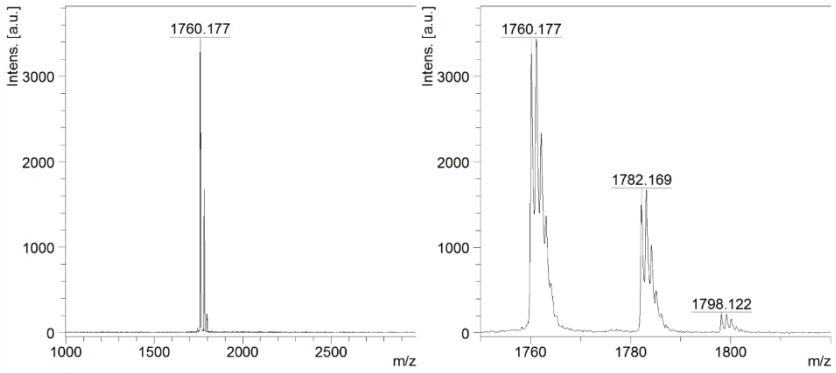
OC9 P6A



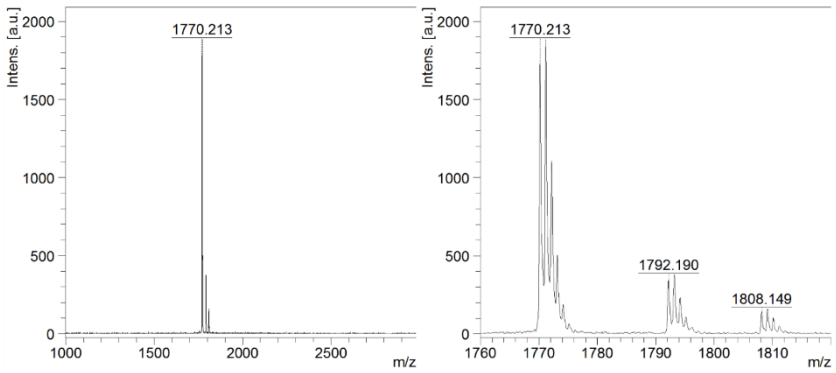
OC9 Y7A



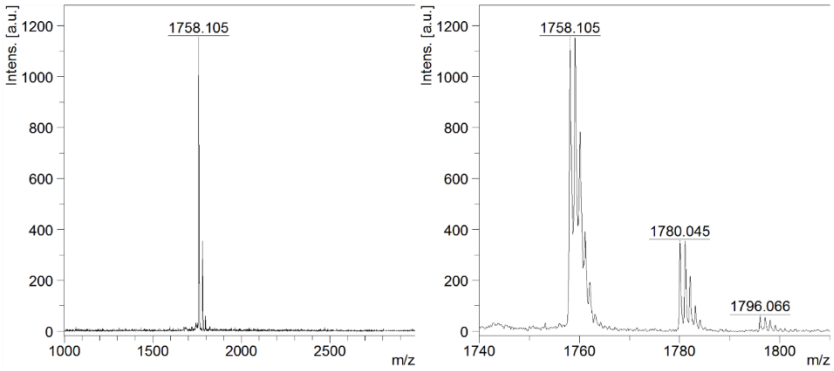
OC9 P8A



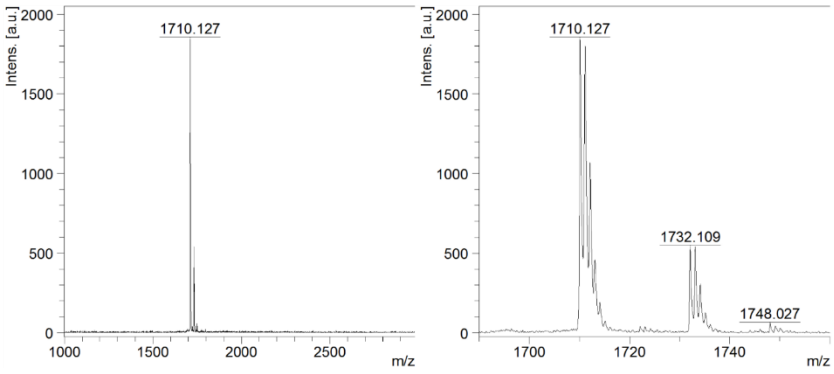
OC9 S9A



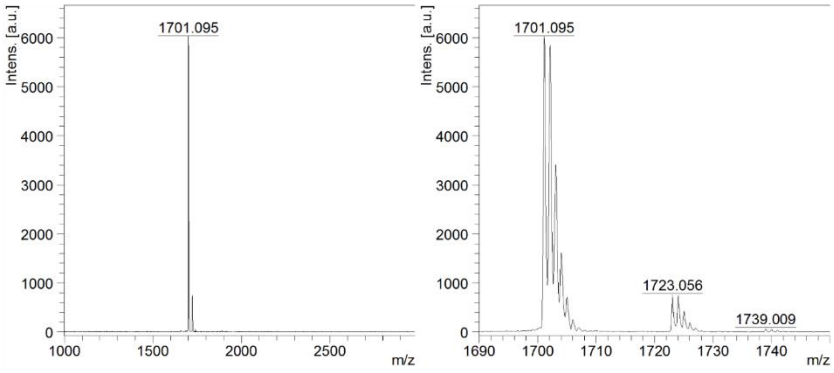
OC9 V10A



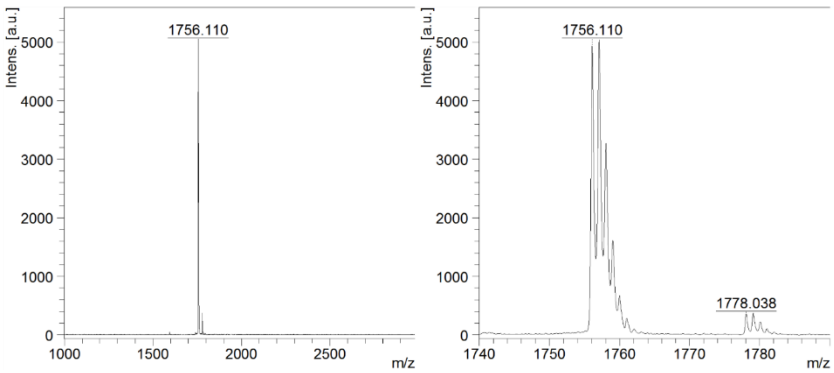
OC9 F11A



OC9 R12A



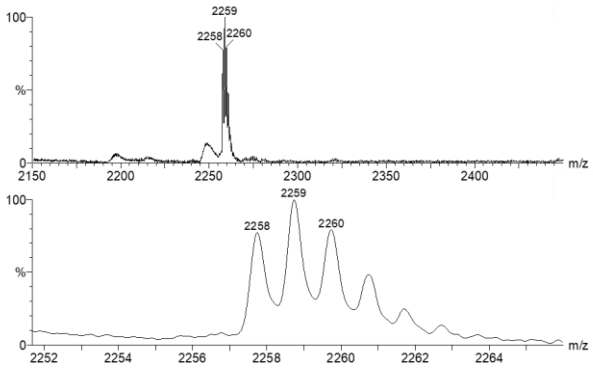
OC9 T13A



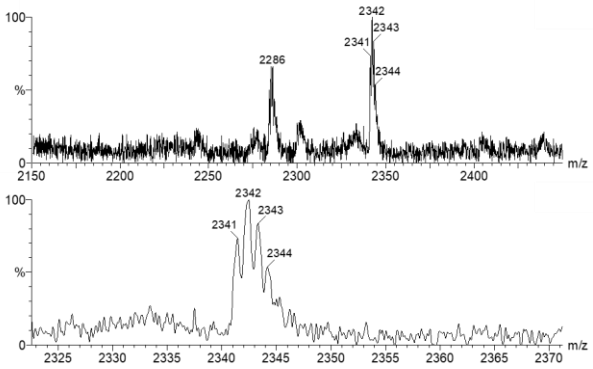
11.7. IVTT MALDI-TOF-MS Spectra

Spectra acquired to assess the synthesis and purity of peptides following ZipTip extraction from IVTT reactions incorporating artificial lysines (Chapter 1).

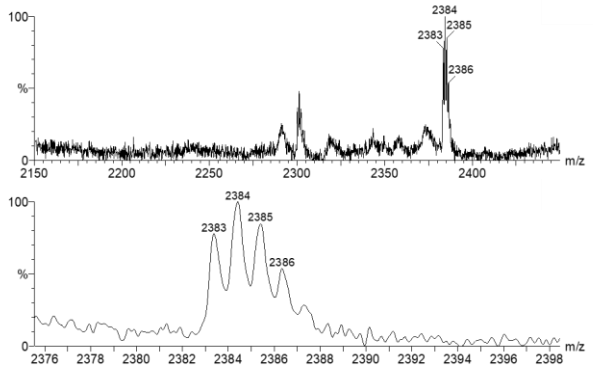
CP2 R6M Linear, Met Control



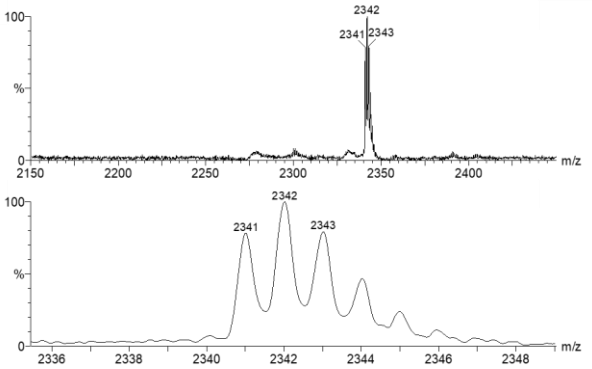
CP2 R6Kme3



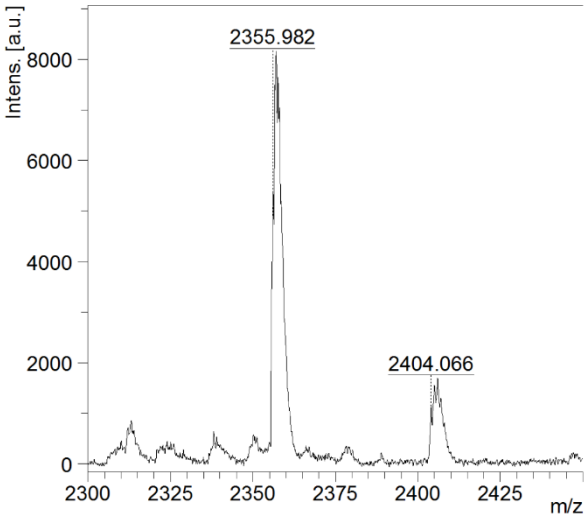
CP2 R6Ket3



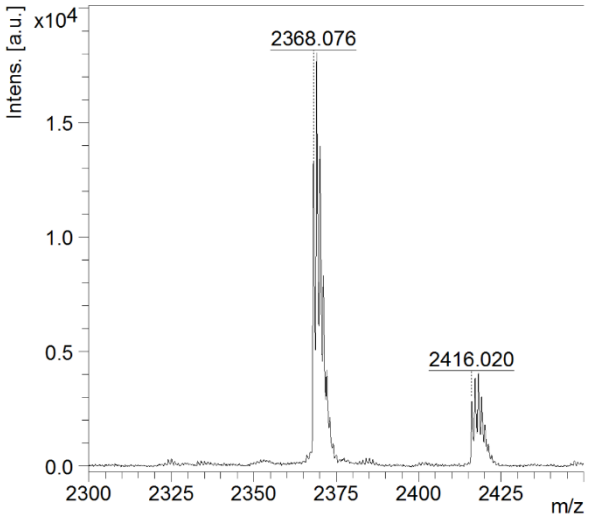
CP2 R6Kac



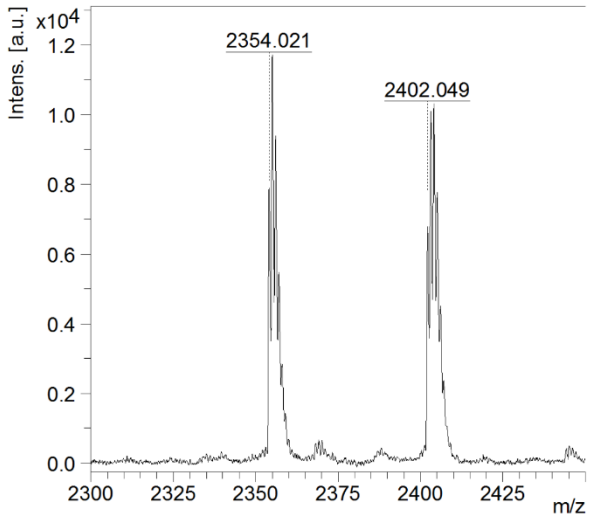
CP2 R6Kpr



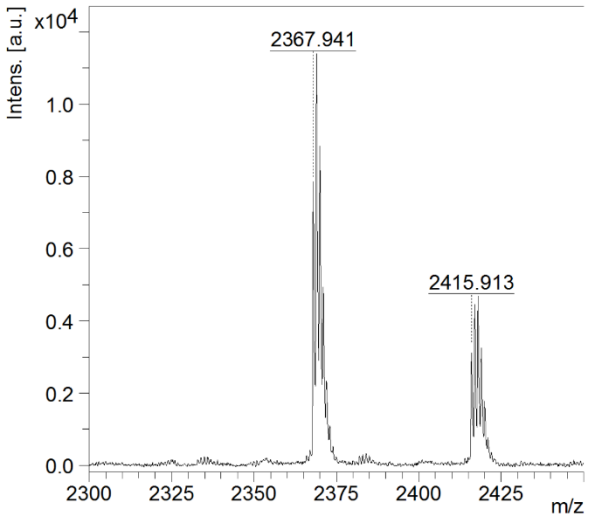
CP2 R6Kcr



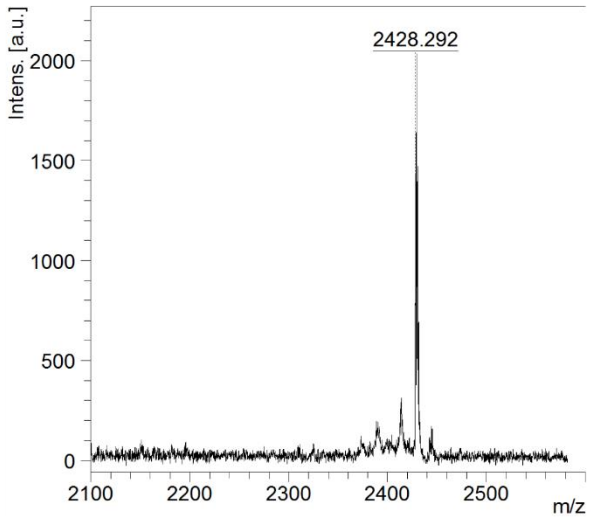
CP2 R6Kpyr



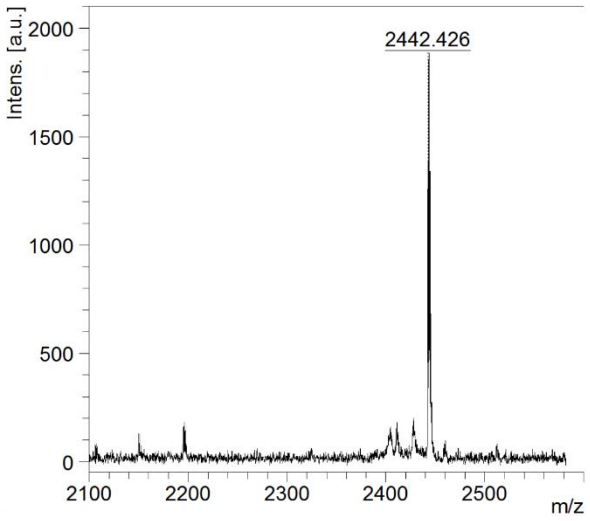
CP2 R6Kpip



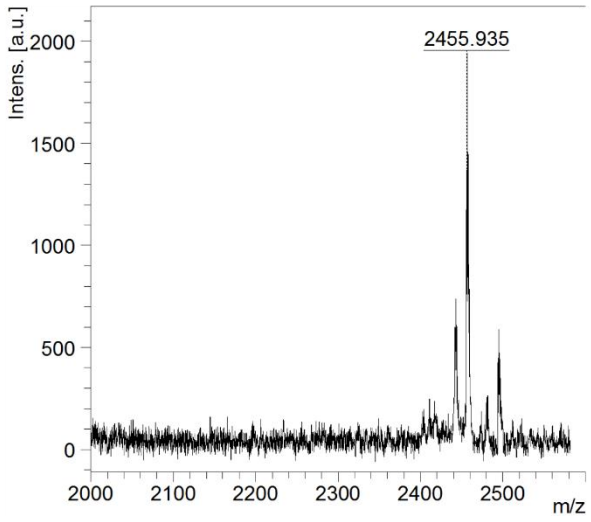
H3K14Kme1 [1-21]



H3K14Kme2 [1-21]

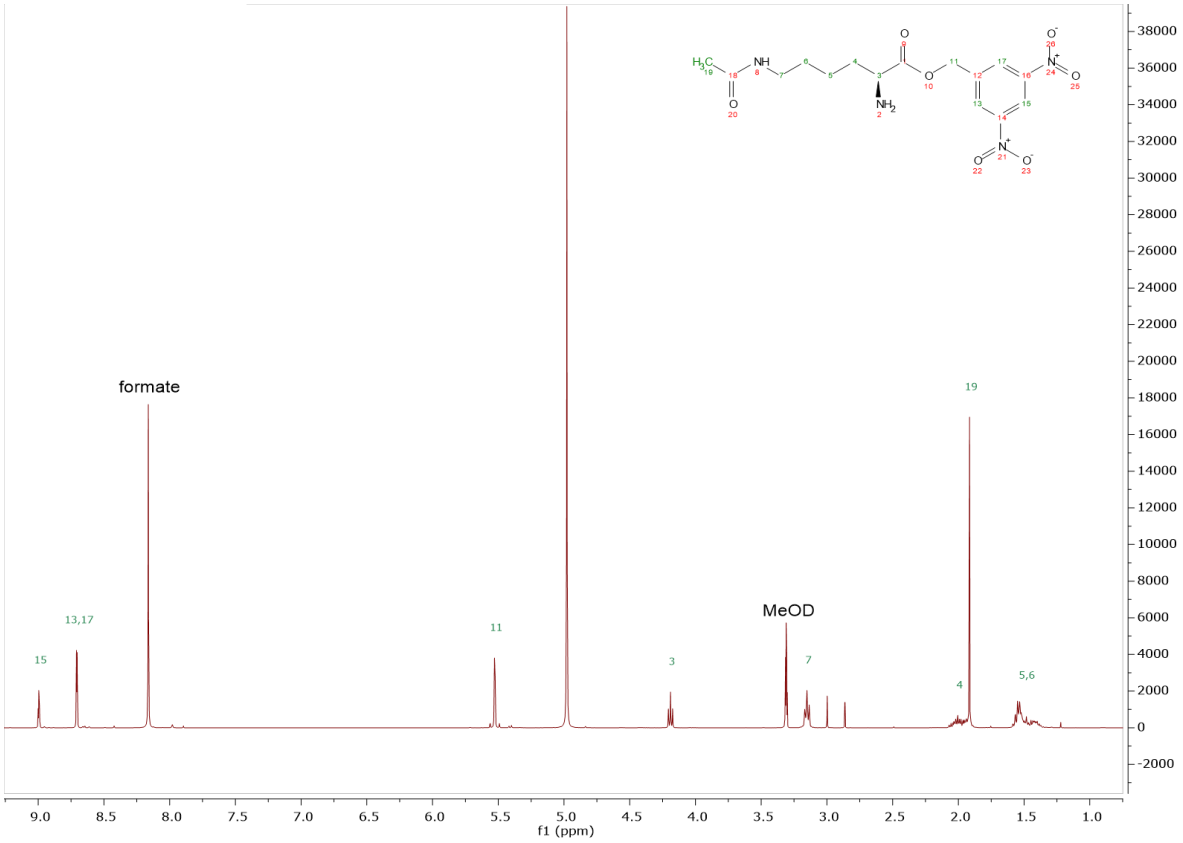


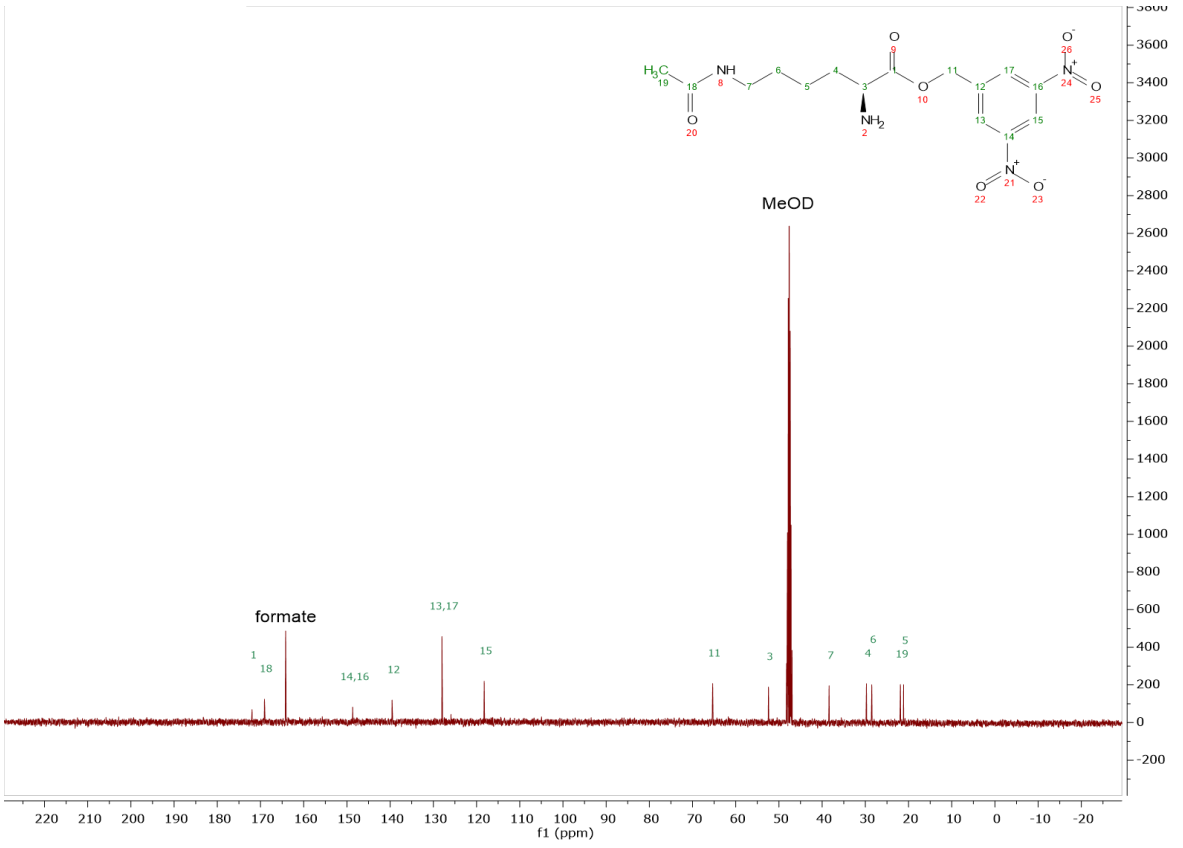
H3K14Kac [1-21]

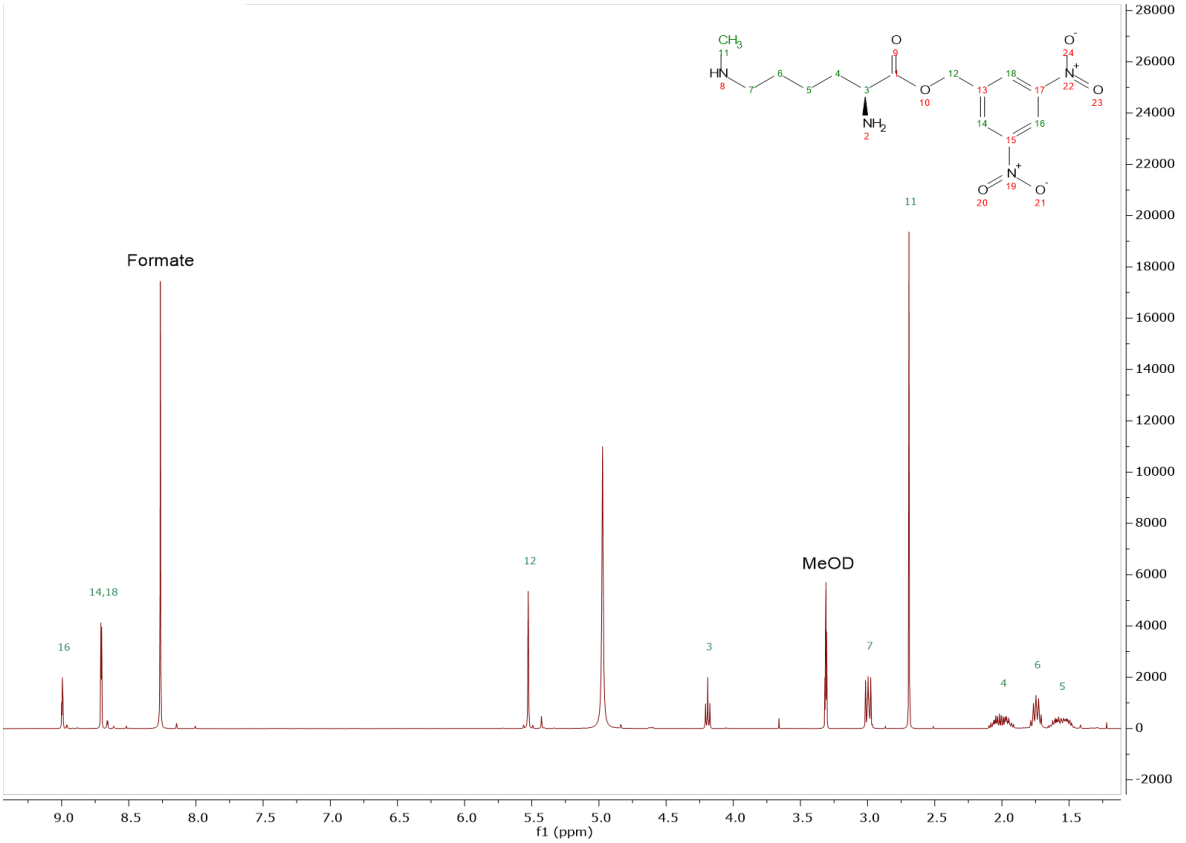


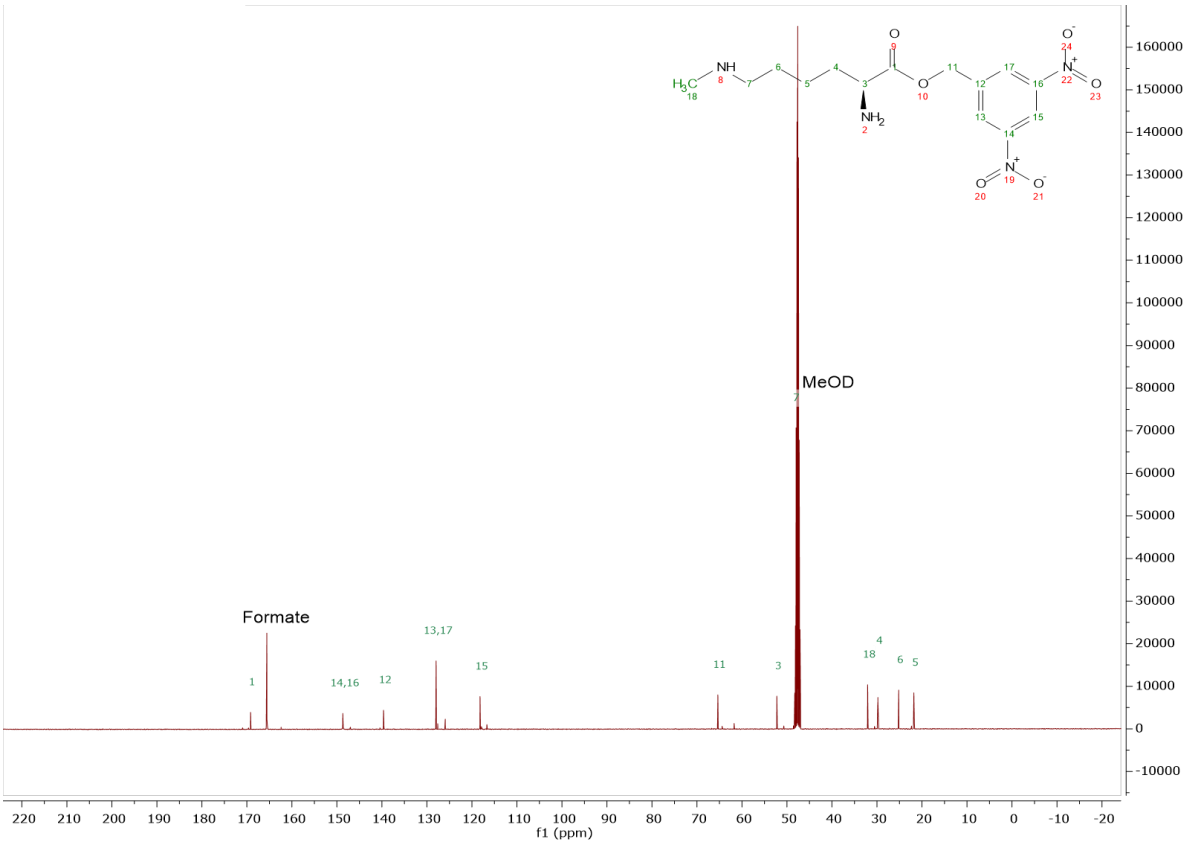
11.8. Small Molecule ^{1}H and ^{13}C 1D-NMR Spectra

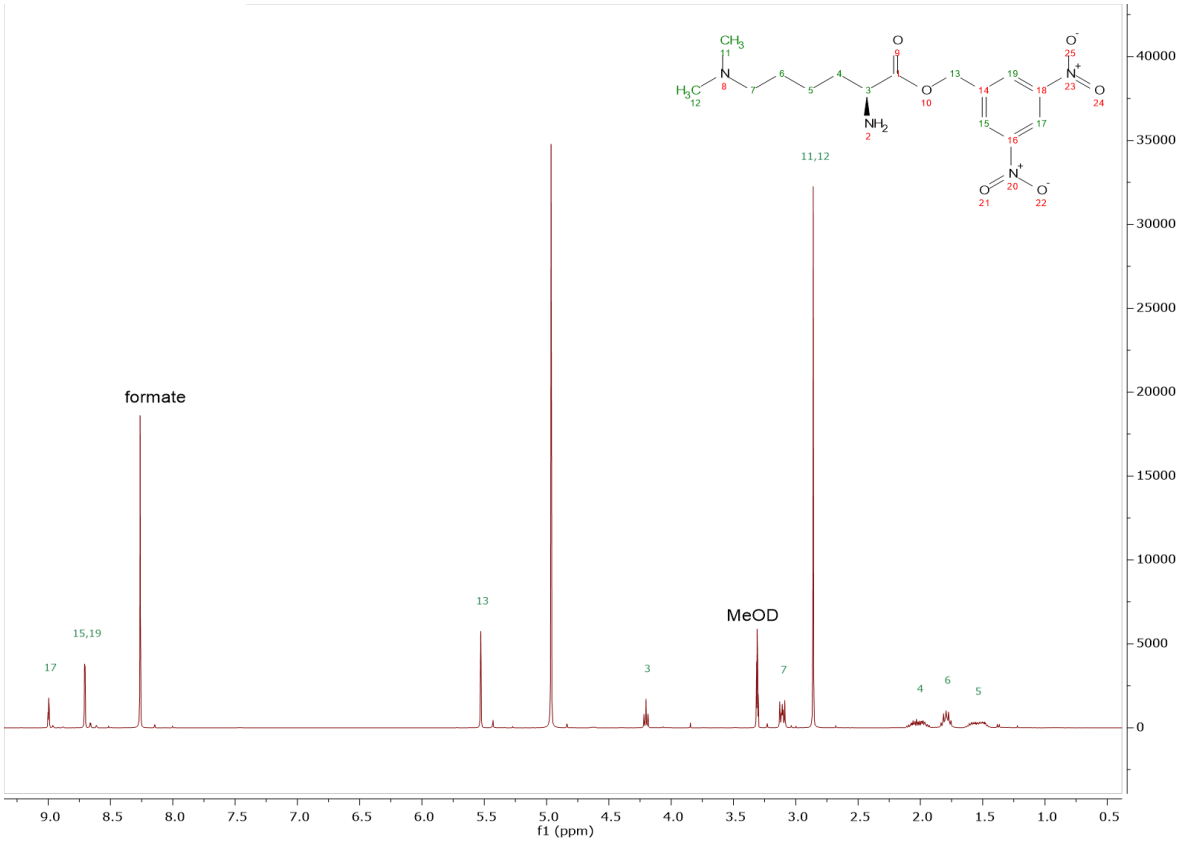
1D ^{1}H and ^{13}C NMR spectra for the final activated ester artificial lysine compounds (Chapter 1); syntheses reported in Materials and Methods.

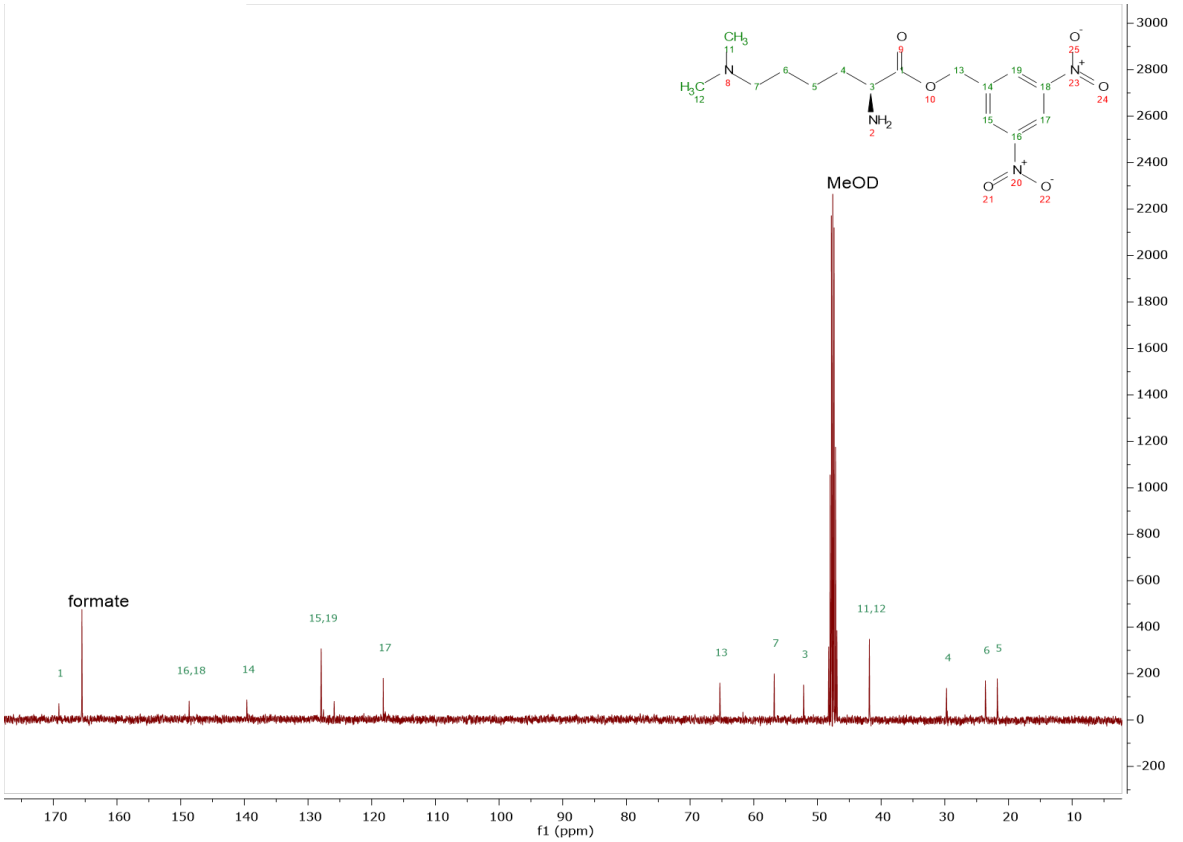


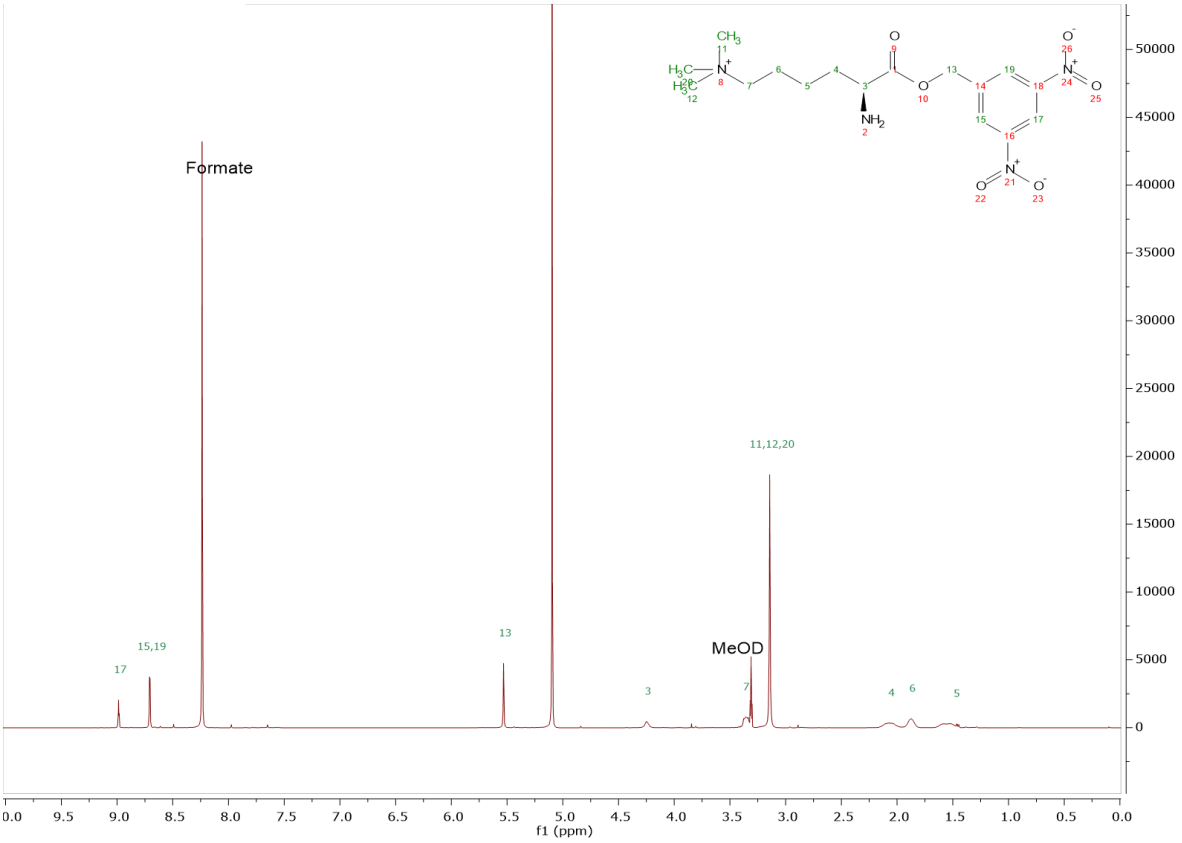


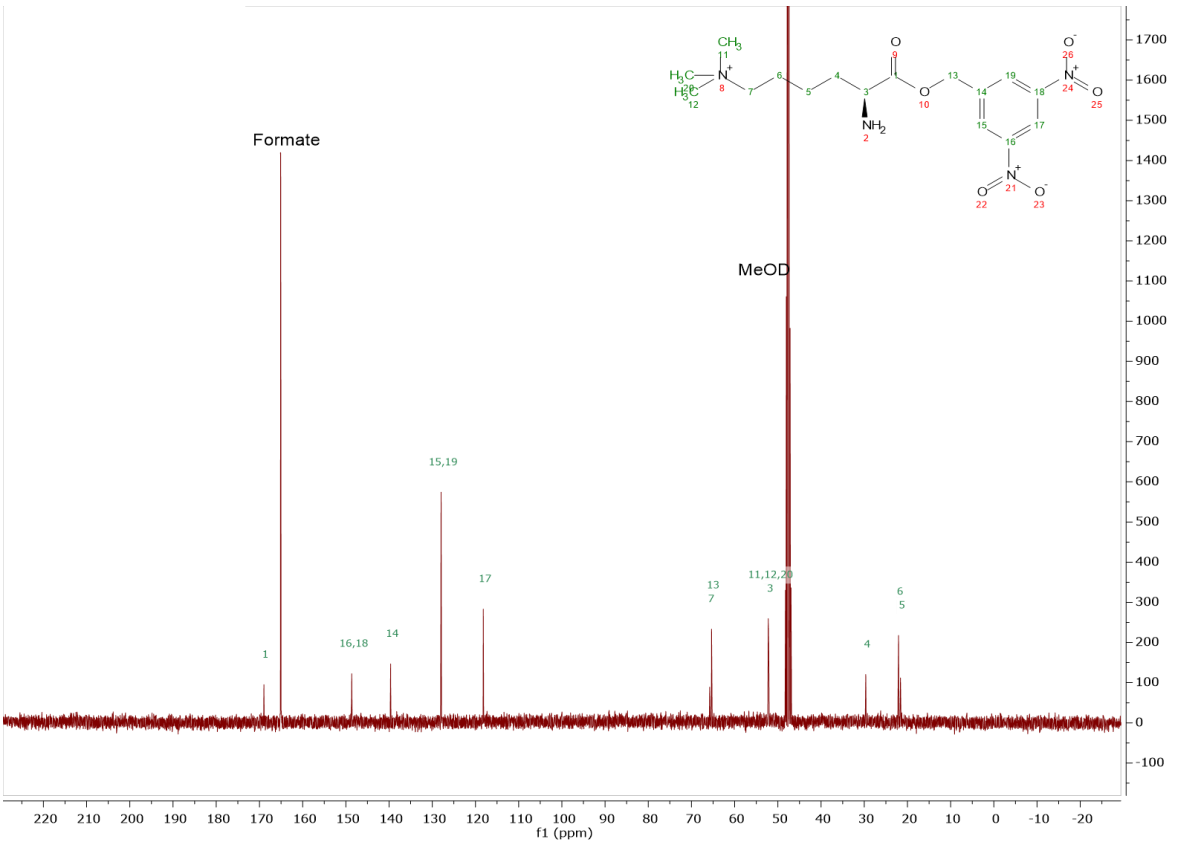

Formate
MeOD
16
14,18
12
3
7
11
4
6
5
f1 (ppm)

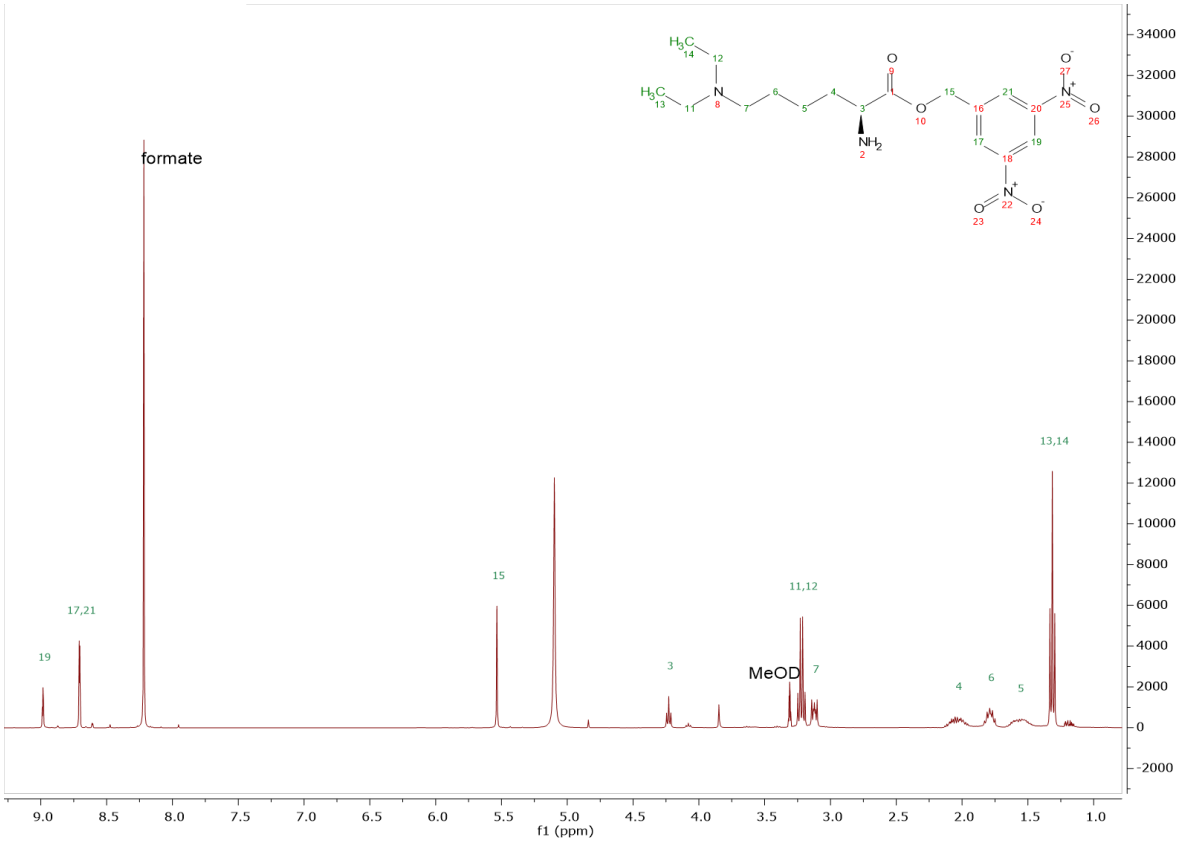

Formate
MeOD
1
14,16
12
13,17
15
11
3
18
4
6
5
f1 (ppm)

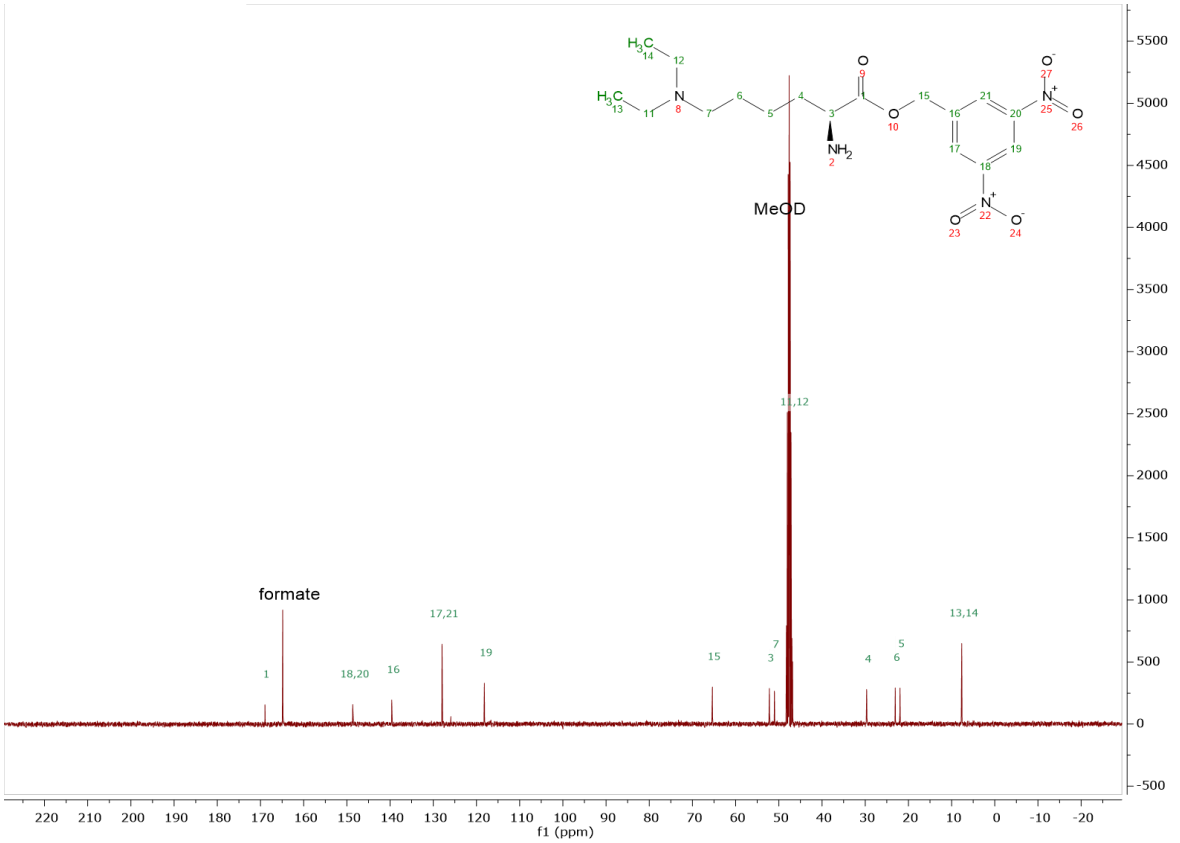

formate
MeOD
11,12
15,19
17
13
3
7
4
6
5
f1 (ppm)

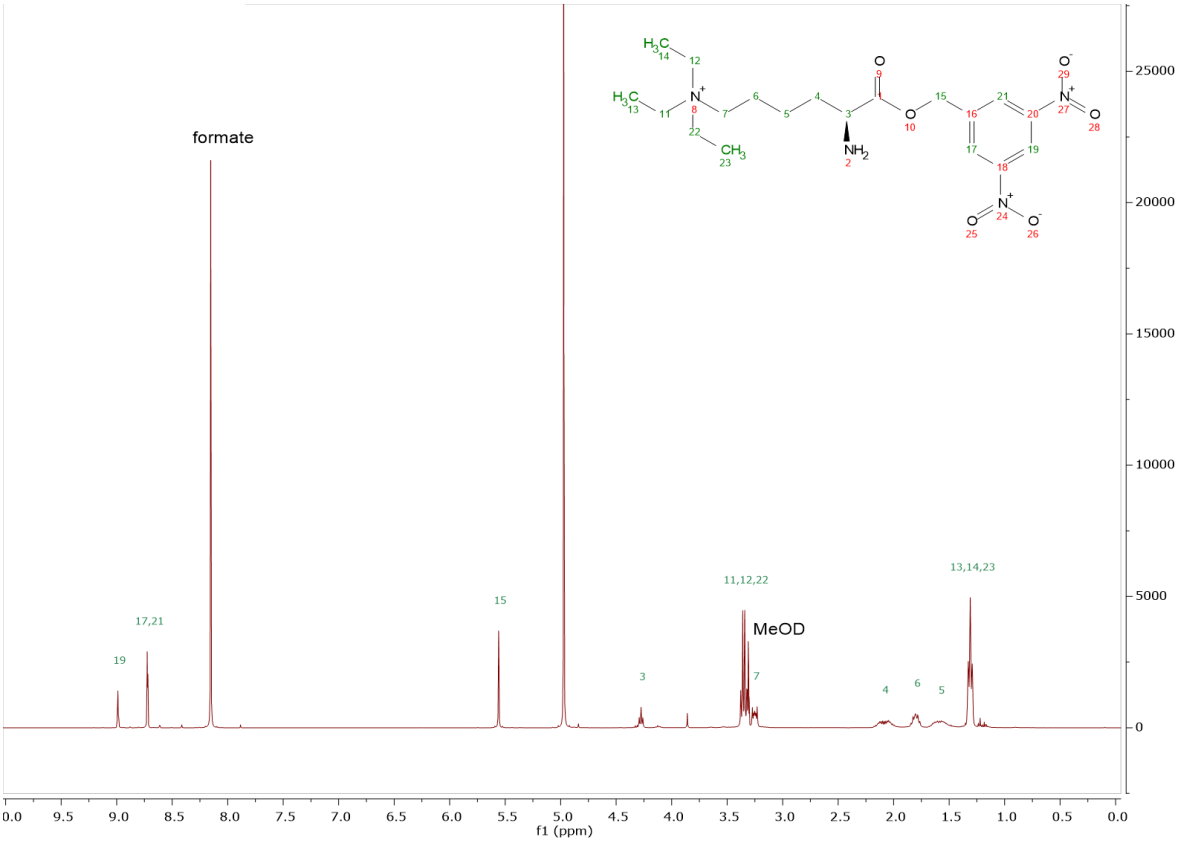

formate
MeOD
11,12
15,19
16,18
14
17
1
13
7
3
4
6
5
f1 (ppm)

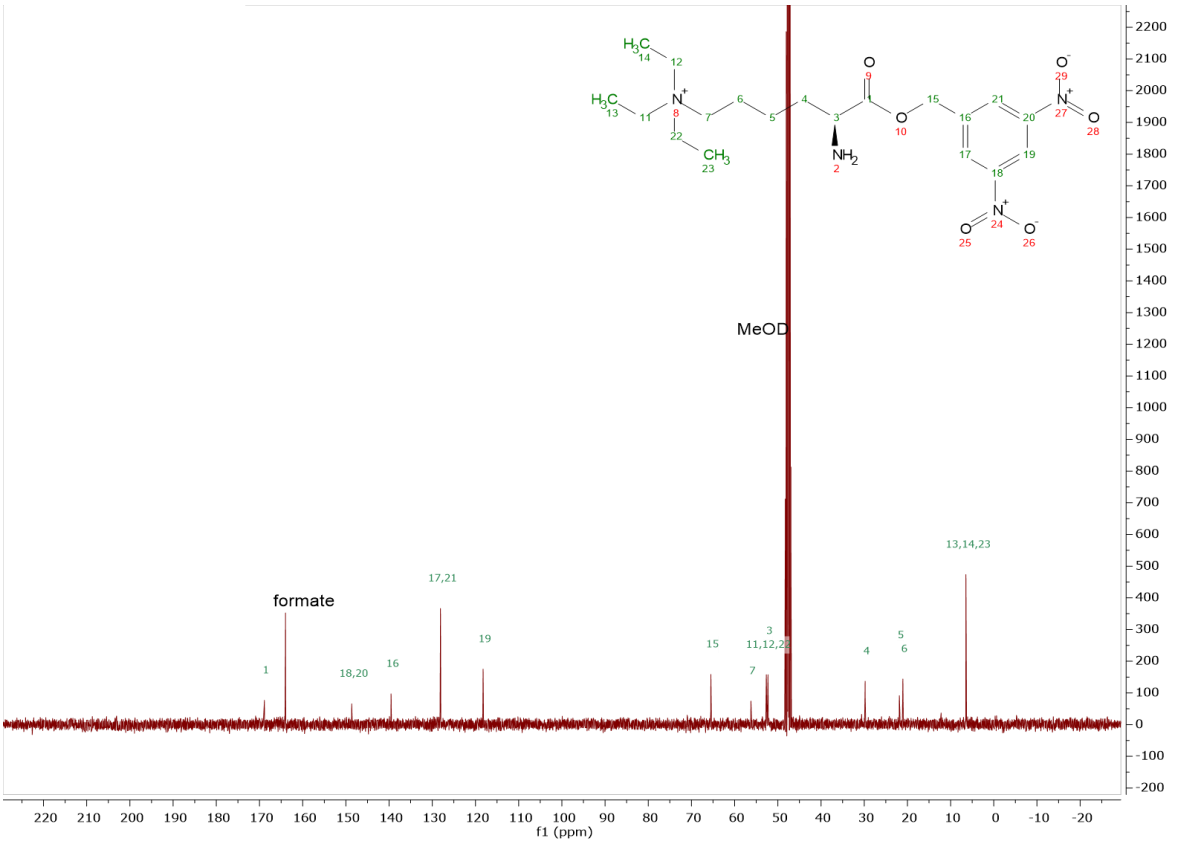

Formate
MeOD
17
15,19
13
3
7
11,12,20
4
6
5
f1 (ppm)


Formate
MeOD
1
16,18
14
15,19
17
13
7
11,12,20
3
4
6
5
f1 (ppm)


formate
MeOD
19
17,21
15
3
11,12
7
4
6
5
13,14
f1 (ppm)


MeOD
formate
1
18,20
16
17,21
19
15
3
7
4
5
6
13,14
11,12
f1 (ppm)


formate
MeOD
19
17,21
15
3
11,12,22
7
4
6
5
13,14,23
f1 (ppm)


MeOD
formate
1
18,20
16
19
17,21
15
3
11,12,22
7
4
5
6
13,14,23
f1 (ppm)

11.9. Crystallography Conditions Trialled for OC9 and KDM7s

KDM7+OC9 Crystallisation Attempts				
KDM7	**AAs**	**Domains**	**Co-Crystallisation**	**Buffer Screen**
B	1-447	PHD+JmjC	OC9	Pact Premier
B	1-447	PHD+JmjC	OC9, Mn(II), NOG	Pact Premier, Crystal
C	1-449	PHD+JmjC	OC9, Mn(II), NOG	JCSG+
B	1-66	PHD	OC9	Pact Premier

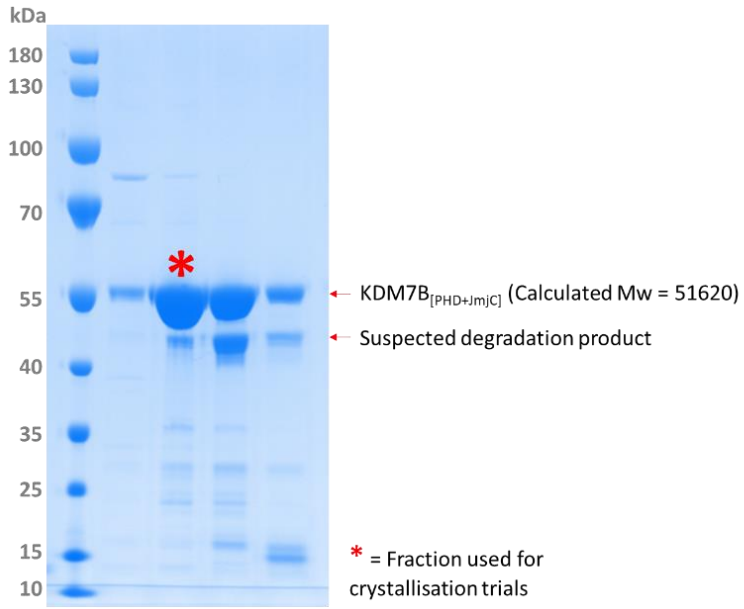
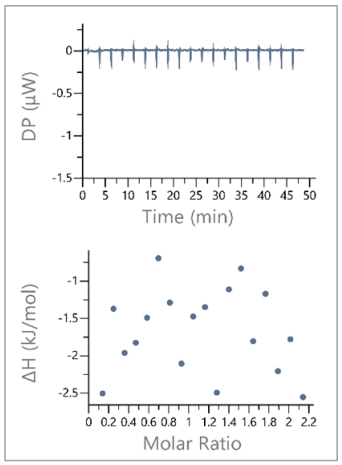


Table of conditions unsuccessfully trialled for attempted co-crystallisation of KDM7s with OC9, and SDS-PAGE gel of KDM7B$_{[PHD+JmjC]}$ protein that was used.

11.10. ITC Control Data



ITC Buffer vs OC9

K_D = ---

OC9 (50 µM stock) was titrated into ITC buffer as a control. OC9 did not give any notable exo/endotherm into the buffer alone which could result in possible interpretation as a binding response, compared to that observed for genuine binding to KDM7 targets (Chapter 2).

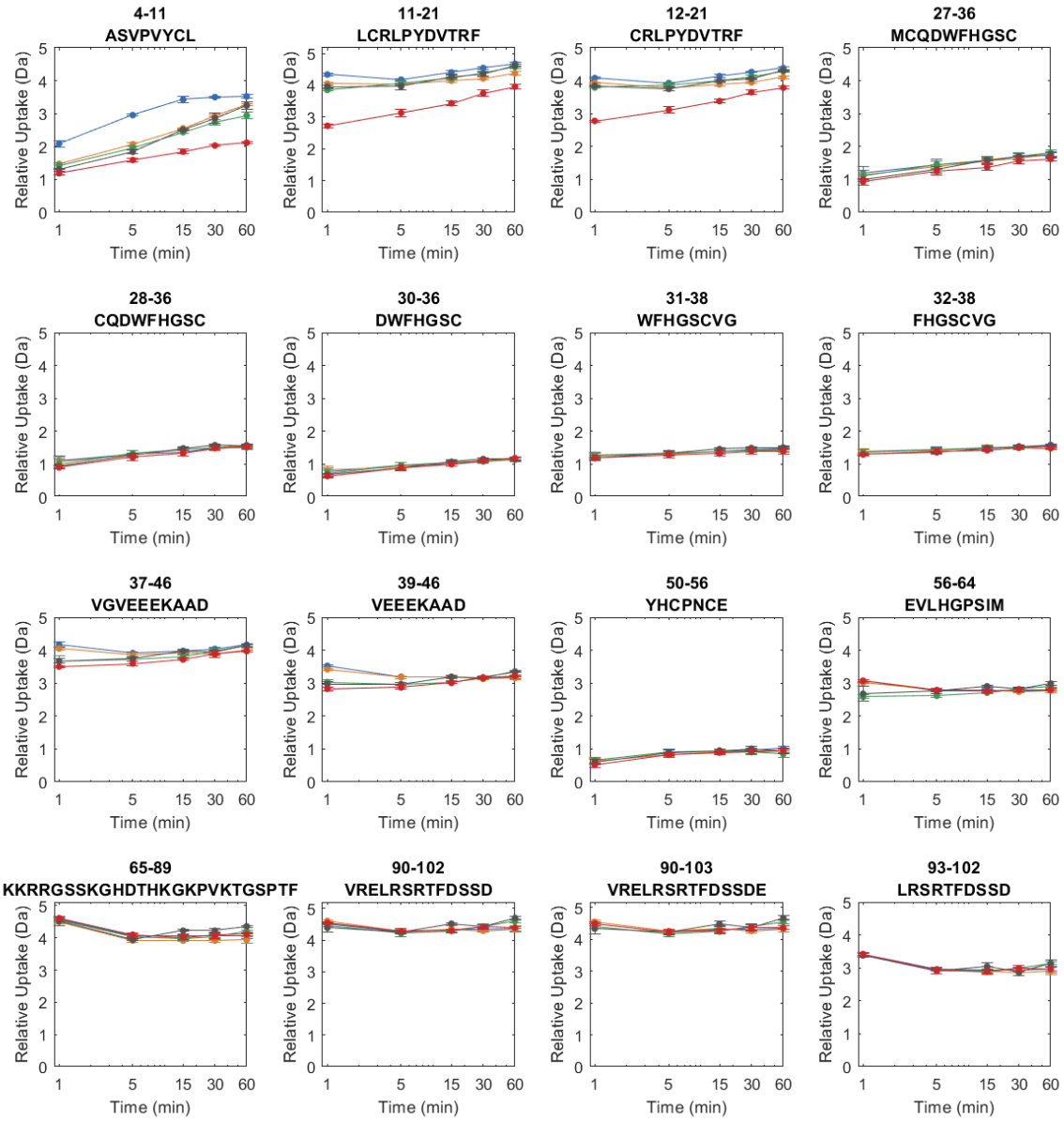
11.11. HDxMS Deuterium Uptake Plots

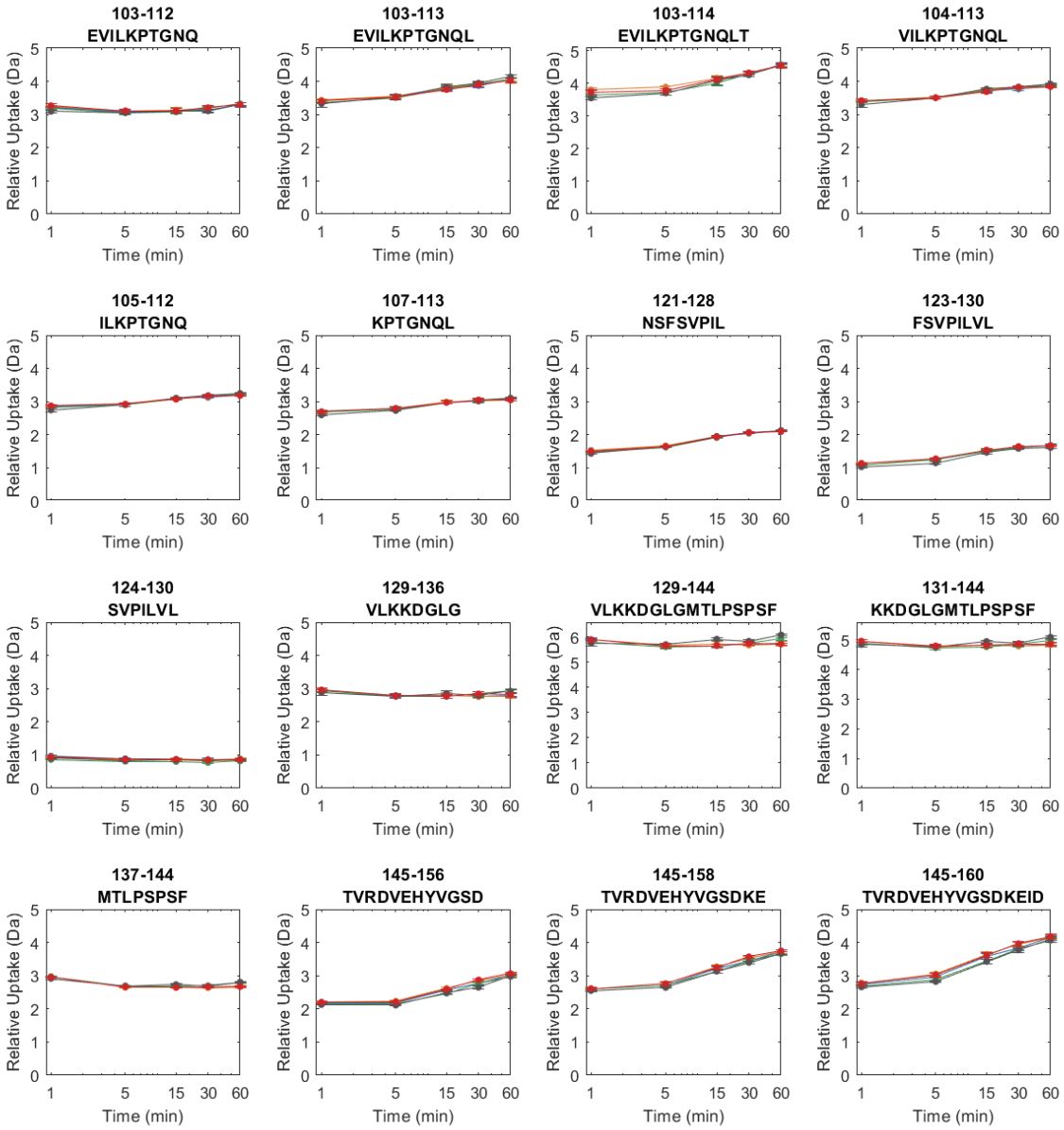
The time-dependent plots for deuterium uptake in the proteolytic fragments, generated during HDxMS analysis, in the presence of different compounds (Chapter 5). Produced by Dr Christopher Stubbs (AstraZeneca).

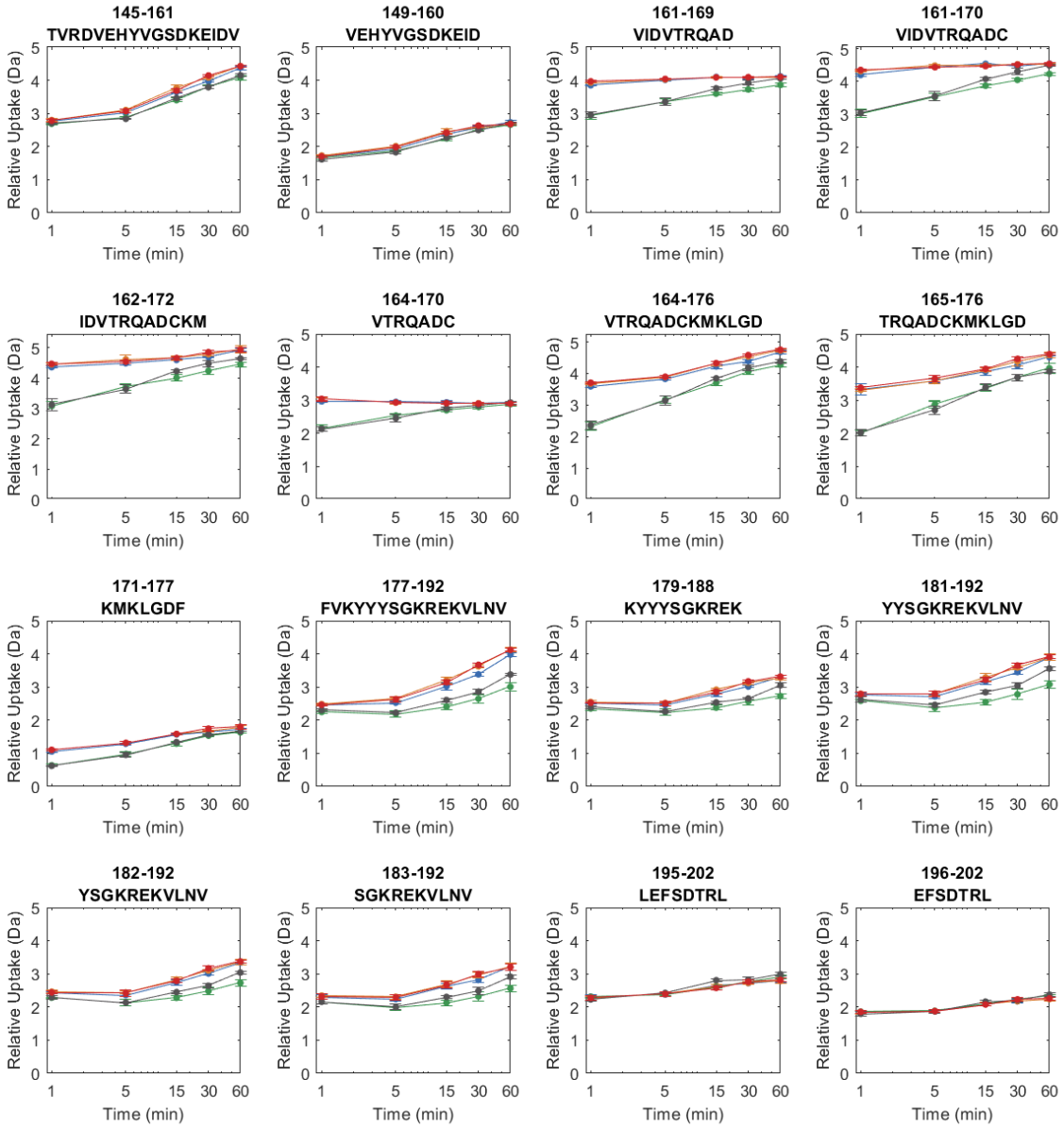
11.11.1. KDM7B[PHD+JmjC]

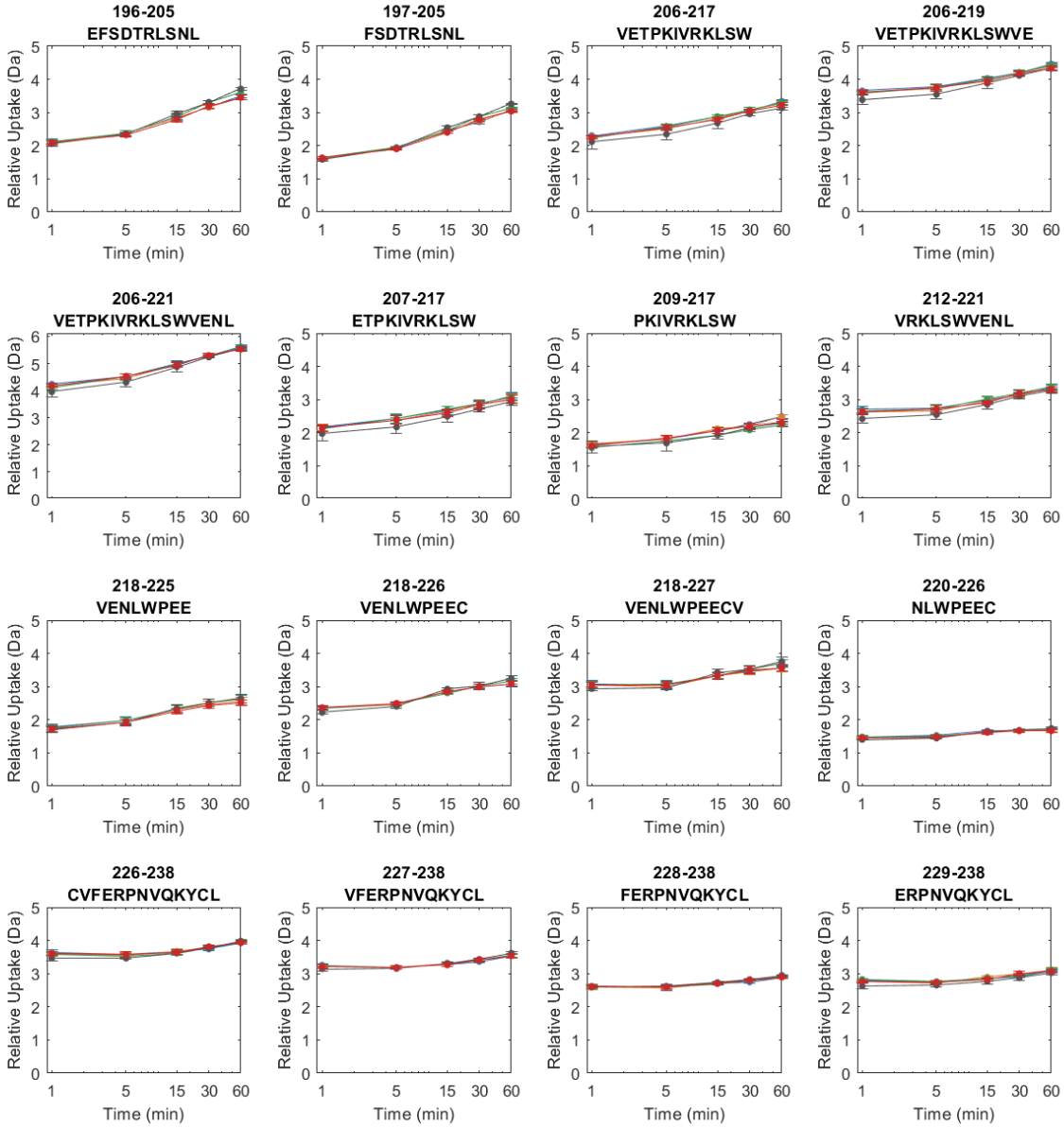
Colour Key:

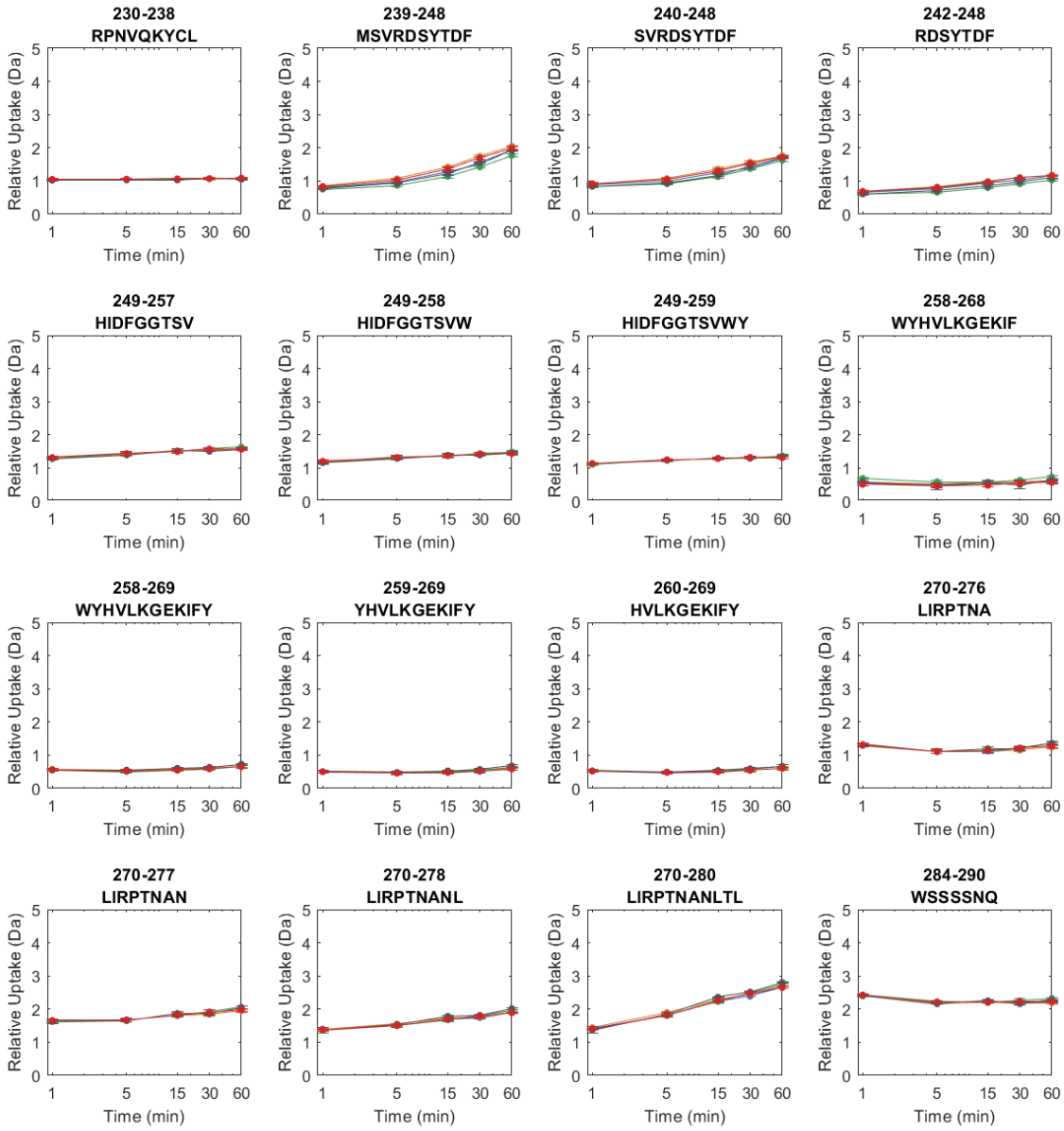
DMSO
H3K4Me3
OC3
OC6
OC9

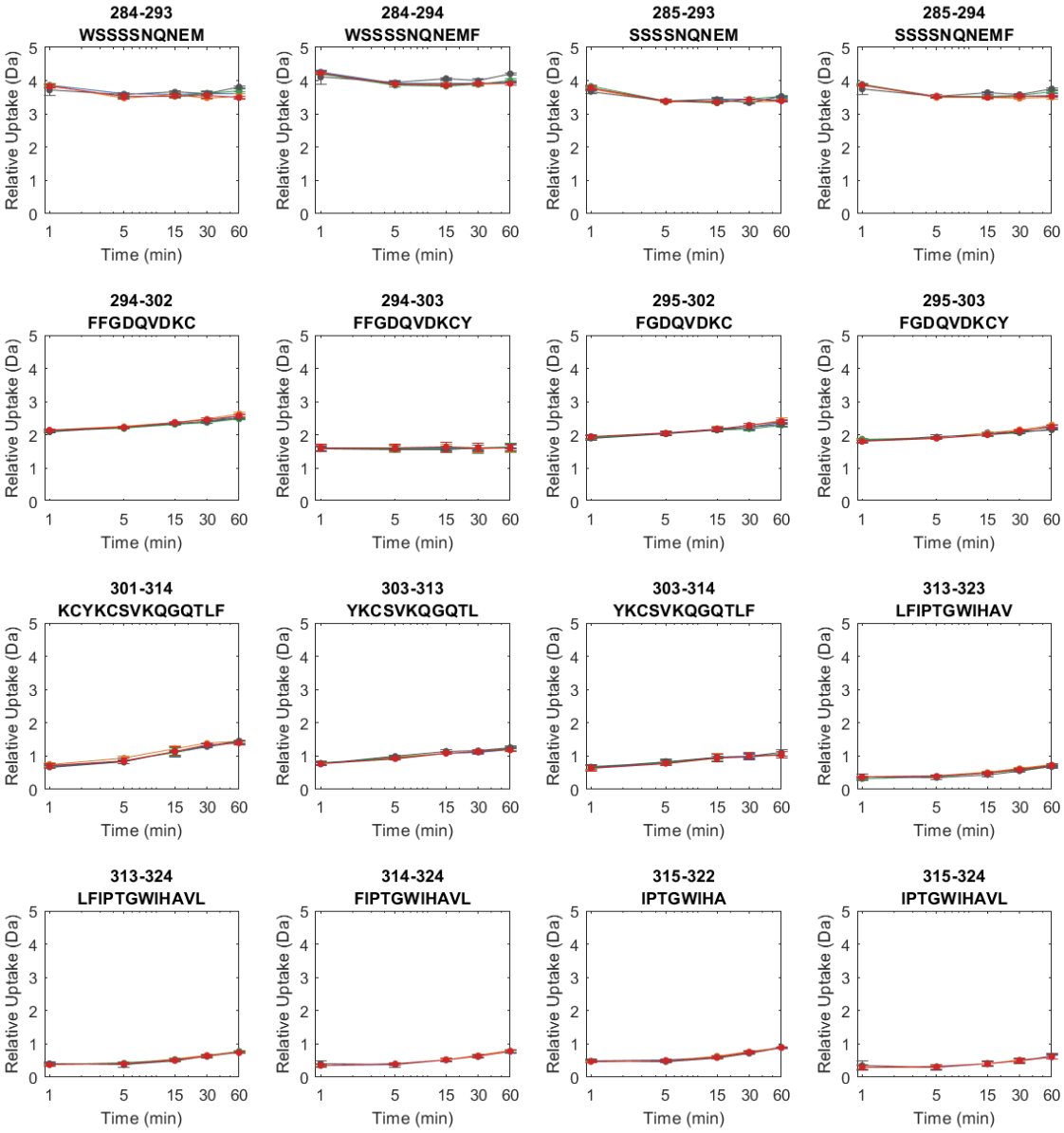


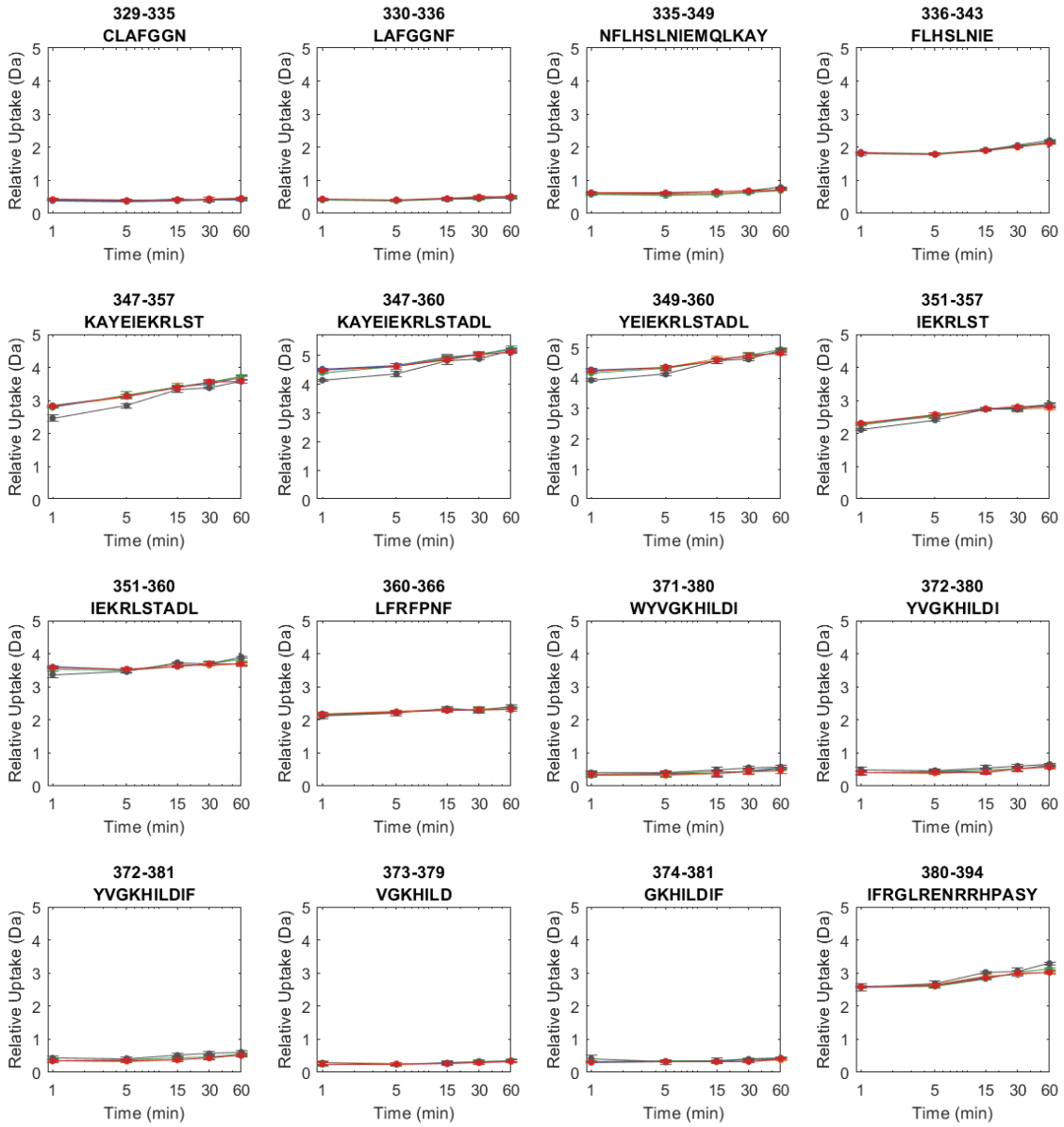

103-112
EVILKPTGNQ
103-113
EVILKPTGNQL
103-114
EVILKPTGNQLT
104-113
VILKPTGNQL
105-112
ILKPTGNQ
107-113
KPTGNQL
121-128
NSFSVPIL
123-130
FSVPILVL
124-130
SVPILVL
129-136
VLKKDGLG
129-144
VLKKDGLGMTLPSPSF
131-144
KKDGLGMTLPSPSF
137-144
MTLPSPSF
145-156
TVRDVEHYVGSD
145-158
TVRDVEHYVGSDKE
145-160
TVRDVEHYVGSDKEID
Relative Uptake (Da)
Time (min)

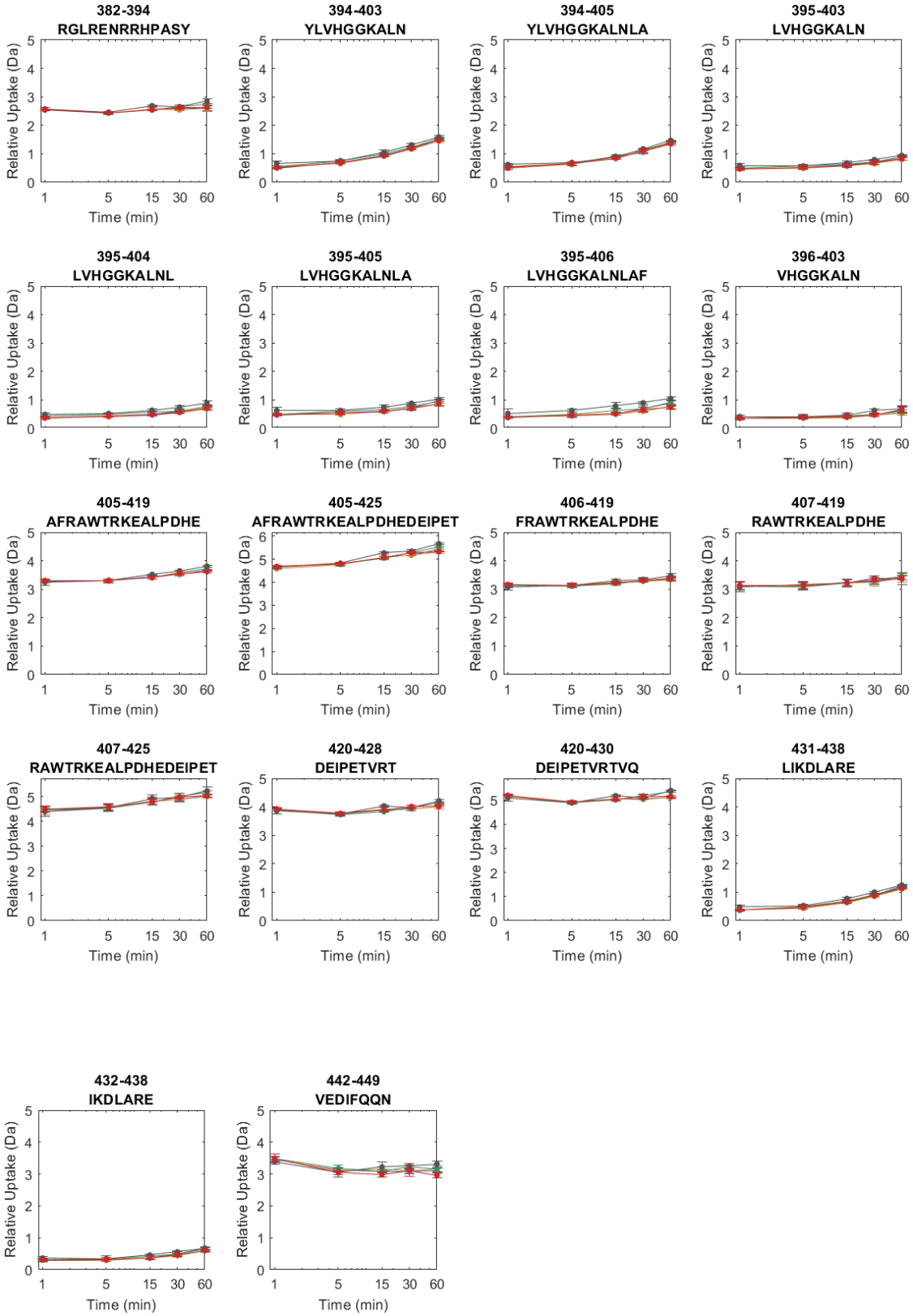

145-161
TVRDVEHYVGSDKEIDV
149-160
VEHYVGSDKEID
161-169
VIDVTRQAD
161-170
VIDVTRQADC
162-172
IDVTRQADCKM
164-170
VTRQADC
164-176
VTRQADCKMKLGD
165-176
TRQADCKMKLGD
171-177
KMKLGDF
177-192
FVKYYYSGKREKVLNV
179-188
KYYYSGKREK
181-192
YYSGKREKVLNV
182-192
YSGKREKVLNV
183-192
SGKREKVLNV
195-202
LEFSDTRL
196-202
EFSDTRL
Relative Uptake (Da)
Time (min)


196-205
EFSDTRLSNL
197-205
FSDTRLSNL
206-217
VETPKIVRKLSW
206-219
VETPKIVRKLSWVE
206-221
VETPKIVRKLSWVENL
207-217
ETPKIVRKLSW
209-217
PKIVRKLSW
212-221
VRKLSWVENL
218-225
VENLWPEE
218-226
VENLWPEEC
218-227
VENLWPEECV
220-226
NLWPEEC
226-238
CVFERPNVQKYCL
227-238
VFERPNVQKYCL
228-238
FERPNVQKYCL
229-238
ERPNVQKYCL
Relative Uptake (Da)
Time (min)
1
5
15
30
60


230-238
RPNVQKYCL
239-248
MSVRDSYTDF
240-248
SVRDSYTDF
242-248
RDSYTDF
249-257
HIDFGGTSV
249-258
HIDFGGTSVW
249-259
HIDFGGTSVWY
258-268
WYHVLKGEKIF
258-269
WYHVLKGEKIFY
259-269
YHVLKGEKIFY
260-269
HVLKGEKIFY
270-276
LIRPTNA
270-277
LIRPTNAN
270-278
LIRPTNANL
270-280
LIRPTNANLTL
284-290
WSSSSNQ
Relative Uptake (Da)
Time (min)


284-293
WSSSSNQNEM
284-294
WSSSSNQNEMF
285-293
SSSSNQNEM
285-294
SSSSNQNEMF
294-302
FFGDQVDKC
294-303
FFGDQVDKCY
295-302
FGDQVDKC
295-303
FGDQVDKCY
301-314
KCYKCSVKQGQTLF
303-313
YKCSVKQGQTL
303-314
YKCSVKQGQTLF
313-323
LFIPTGWIHAV
313-324
LFIPTGWIHAVL
314-324
FIPTGWIHAVL
315-322
IPTGWIHA
315-324
IPTGWIHAVL
Relative Uptake (Da)
Time (min)
0
1
2
3
4
5
1
5
15
30
60

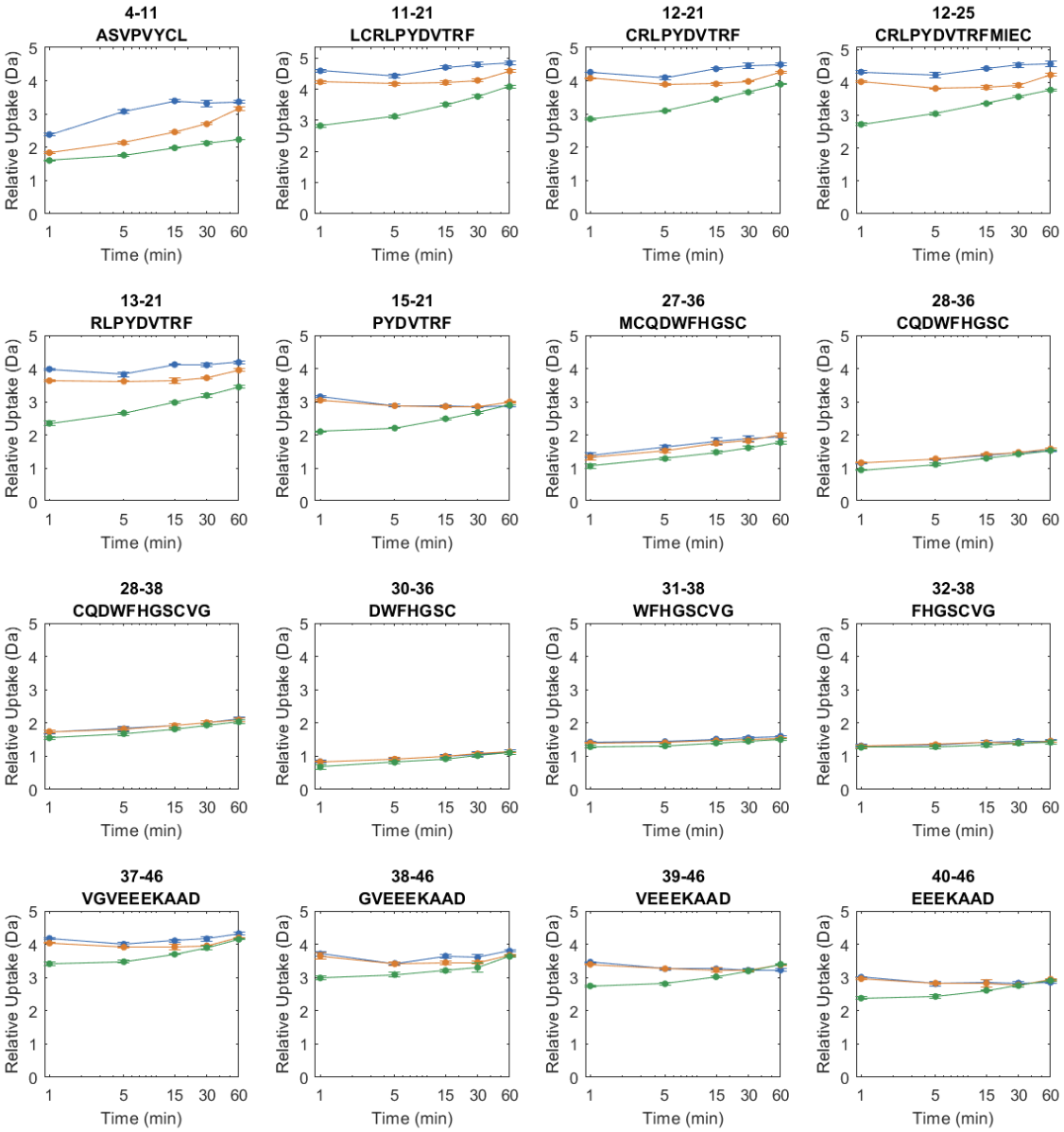

329-335
CLAFGGN
330-336
LAFGGNF
335-349
NFLHSLNIEMQLKAY
336-343
FLHSLNIE
347-357
KAYEIEKRLST
347-360
KAYEIEKRLSTADL
349-360
YEIEKRLSTADL
351-357
IEKRLST
351-360
IEKRLSTADL
360-366
LFRFPNF
371-380
WYVGKHILDI
372-380
YVGKHILDI
372-381
YVGKHILDIF
373-379
VGKHILD
374-381
GKHILDIF
380-394
IFRGLRENRRHPASY
Relative Uptake (Da)
Time (min)

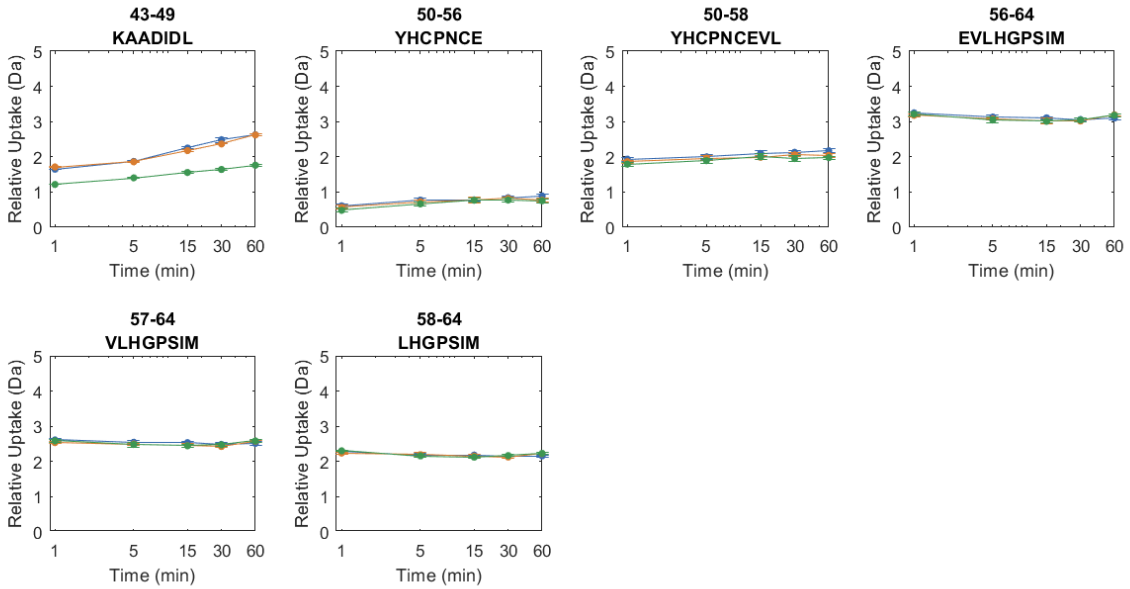

382-394
RGLRENRRHPASY
394-403
YLVHGGKALN
394-405
YLVHGGKALNLA
395-403
LVHGGKALN
395-404
LVHGGKALNL
395-405
LVHGGKALNLA
395-406
LVHGGKALNLAF
396-403
VHGGKALN
405-419
AFRAWTRKEALPDHE
405-425
AFRAWTRKEALPDHEDEIPET
406-419
FRAWTRKEALPDHE
407-419
RAWTRKEALPDHE
407-425
RAWTRKEALPDHEDEIPET
420-428
DEIPETVRT
420-430
DEIPETVRTVQ
431-438
LIKDLARE
432-438
IKDLARE
442-449
VEDIFQQN
Relative Uptake (Da)
Time (min)
1
5
15
30
60

11.11.2. KDM7B[PHD]

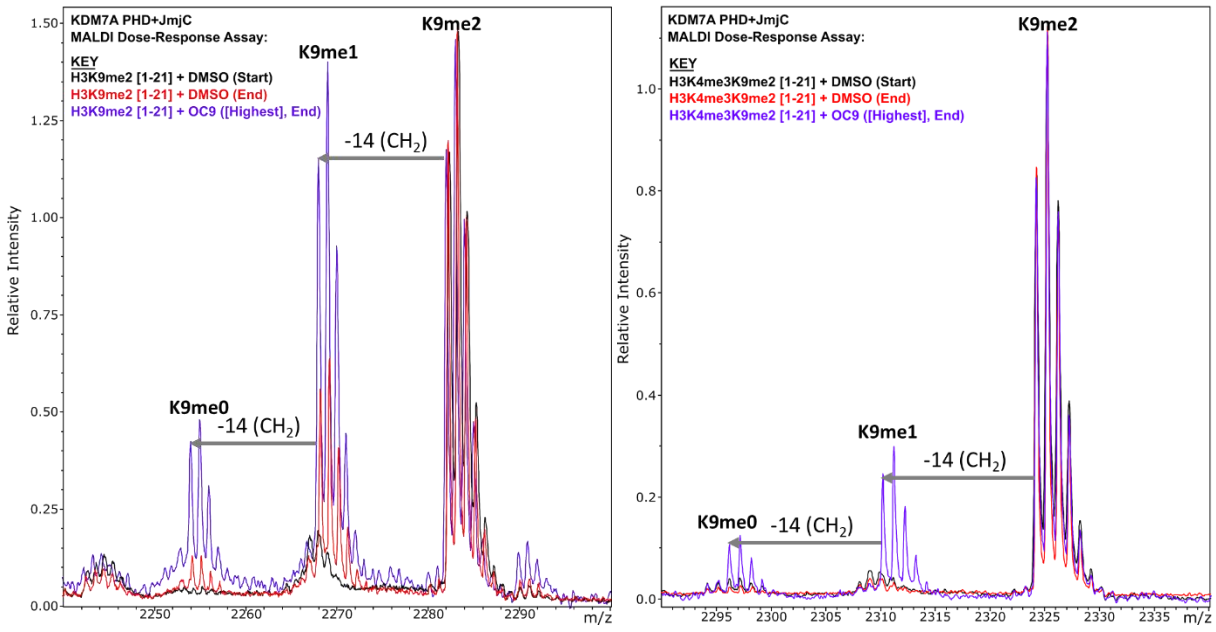
Colour Key:

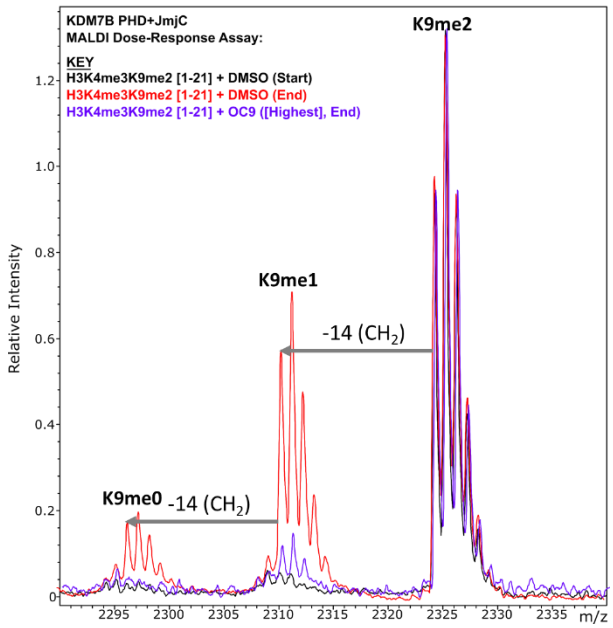
DMSO
H3K4Me3
OC9




43-49
KAADIDL
50-56
YHCPNCE
50-58
YHCPNCEVL
56-64
EVLHGPSIM
57-64
VLHGPSIM
58-64
LHGPSIM
Relative Uptake (Da)
Time (min)
0
1
2
3
4
5
1
5
15
30
60

11.12. Example MALDI-TOF Spectra for OC9 vs KDM7 Activity Assays



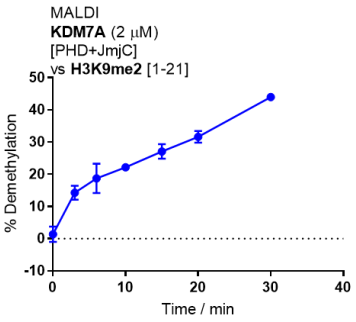


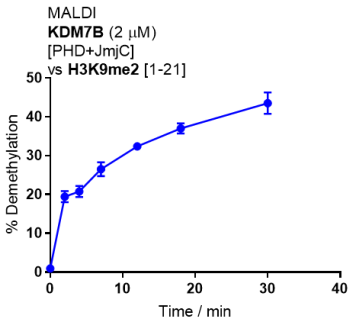
The relative intensities of each individual spectrum, within the overlaid spectra, are scaled to the most intense peak within the spectrum

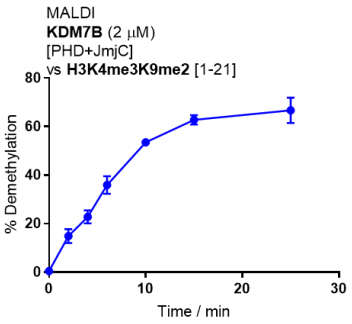
11.13. MALDI-TOF Activity Assay Timecourses

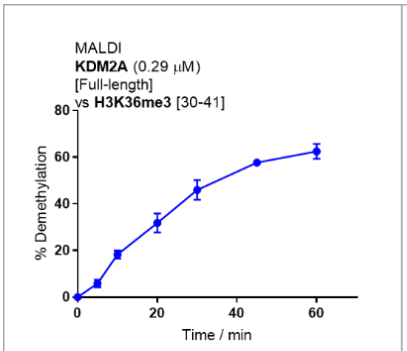
Timecourses were conducted to assess the linear range of activity for enzymes, in order to conduct assays to assess inhibitor/activator effects within a reasonable signal-window.

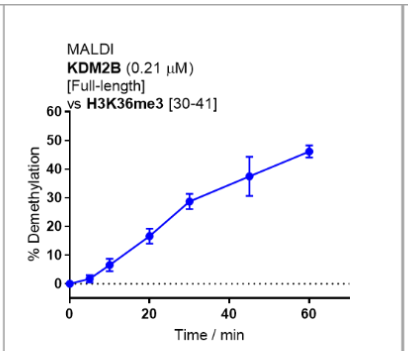
For Single Point Concentration MALDI-TOF-MS Assays:

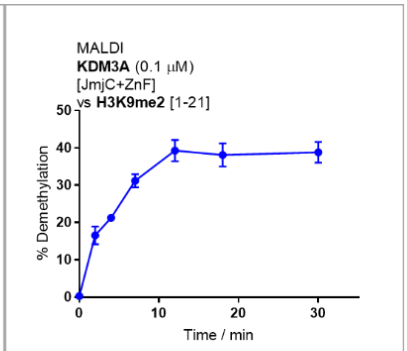


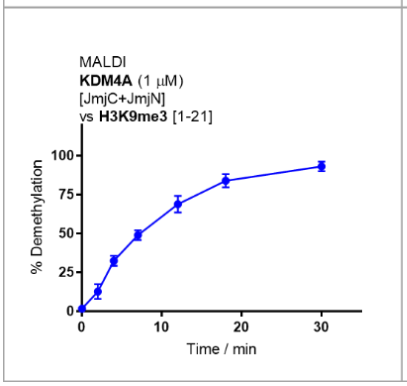


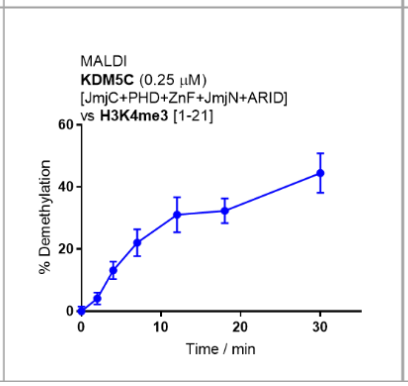


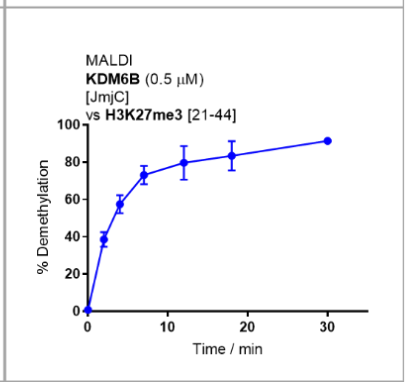


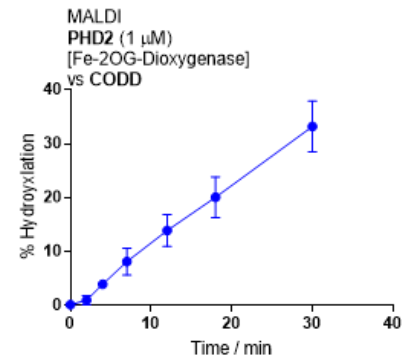




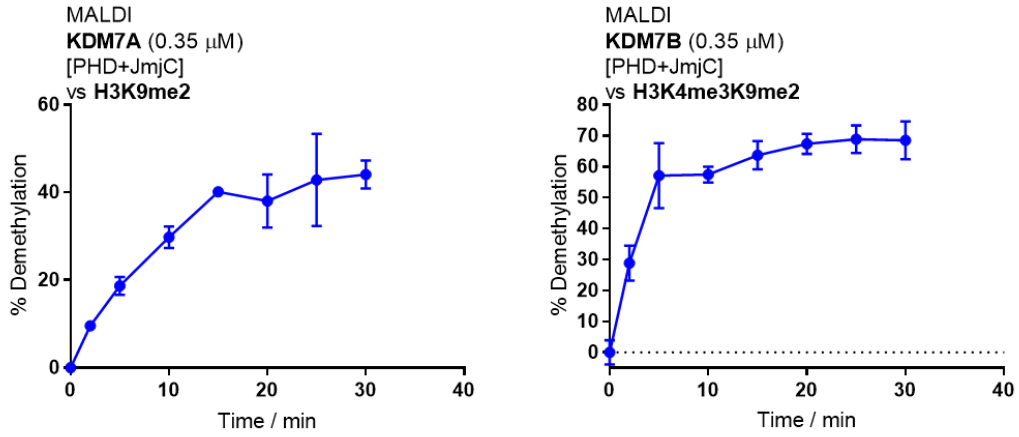




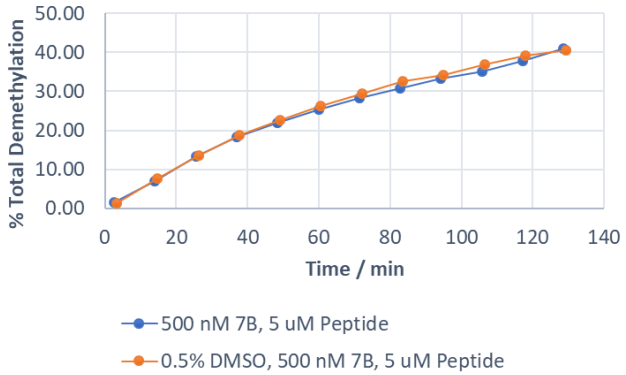




For Dose-Response MALDI-TOF-MS Assays:



11.14. RapidFire-MS Timecourse



A timecourse was conducted to assess the linear range of enzyme activity for KDM7B$_{[PHD+JmjC]}$ by RapidFire-MS (Chapter 3).

11.15. OC3 and OC6 Sequence WebLogo Plots from NGS Data



11.16. List of Publications and Prizes (Presentations, Conferences)

- Højrup, C. *, Coleman, O. D. *, Bukowski, John-Paul*, Clausen, R. P. & Kawamura, A. in *Epigenetic Drug Discovery* 263–296 (2018). *Authors contributed equally.
- Departmental commendation for Transfer of Status Report (end of research year 1)
- Invited Speaker, Runner-Up Prize for Best-Talk – RSC BMCS Post-Graduate Symposium, December 2017, Cambridge University
- Invited Speaker, Travel Prize Recipient – RSC Protein and Peptide Subject Group Early Stage Researcher Meeting, July 2018, Southampton University
- Poster Presentation – EMBO Chemical Biology Workshop, August 2018, EMBL Heidelberg

Chapter 12. References

1. Turner, B. M. Epigenetic responses to environmental change and their evolutionary implications. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 3403–3418 (2009).

2. Jiang, R., Jones, M. J., Sava, F., Kobor, M. S. & Carlsten, C. Short-term diesel exhaust inhalation in a controlled human crossover study is associated with changes in DNA methylation of circulating mononuclear cells in asthmatics. *Part. Fibre Toxicol.* **11**, 71 (2014).

3. Kaelin, W. G. & McKnight, S. L. Influence of Metabolism on Epigenetics and Disease. *Cell* **153**, 56–69 (2013).

4. Hanna, C. W., Demond, H. & Kelsey, G. Epigenetic regulation in development: is the mouse a good model for the human? *Hum. Reprod. Update* **24**, 556–576 (2018).

5. Sen, P., Shah, P. P., Nativio, R. & Berger, S. L. Epigenetic Mechanisms of Longevity and Aging. *Cell* **166**, 822–839 (2016).

6. Budhavarapu, V. N., Chavez, M. & Tyler, J. K. How is epigenetic information maintained through DNA replication? *Epigenetics Chromatin* **6**, 32 (2013).

7. Skvortsova, K., Iovino, N. & Bogdanović, O. Functions and mechanisms of epigenetic inheritance in animals. *Nat. Rev. Mol. Cell Biol.* (2018). doi:10.1038/s41580-018-0074-2

8. Heard, E. & Martienssen, R. A. Transgenerational Epigenetic Inheritance: Myths and Mechanisms. *Cell* **157**, 95–109 (2014).

9. Hamazaki, T., El Rouby, N., Fredette, N. C., Santostefano, K. E. & Terada, N. Concise Review: Induced Pluripotent Stem Cell Research in the Era of Precision Medicine. *Stem Cells* **35**, 545–550 (2017).

10. Edwards, J. R., Yarychkivska, O., Boulard, M. & Bestor, T. H. DNA methylation and DNA methyltransferases. *Epigenetics Chromatin* **10**, 23 (2017).

11. Wu, T. P. *et al.* DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333 (2016).

12. Dor, Y. & Cedar, H. Principles of DNA methylation and their implications for biology and medicine. *Lancet* **392**, 777–786 (2018).

13. Wu, H. & Zhang, Y. Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions. *Cell* **156**, 45–68 (2014).

14. Carell, T., Kurz, M. Q., Müller, M., Rossa, M. & Spada, F. Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA. *Angew. Chemie Int. Ed.* **57**, 4296–4312 (2018).

15. Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557 (2015).

16. Rothbart, S. B. & Strahl, B. D. Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839**, 627–643 (2014).

17. Noh, K.-M., Allis, C. D. & Li, H. Reading between the Lines: “ADD”-ing Histone and DNA Methylation Marks toward a New Epigenetic “Sum”. *ACS Chem. Biol.* **11**, 554–563 (2016).

18. Lawrence, M., Daujat, S. & Schneider, R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet.* **32**, 42–56 (2016).

19. Castillo, J., López-Rodas, G. & Franco, L. Histone Post-Translational Modifications and Nucleosome Organisation in Transcriptional Regulation: Some Open Questions. in *Protein Reviews: Volume 18* (ed. Atassi, M. Z.) 65–92 (Springer Singapore, 2017). doi:10.1007/5584_2017_58

20. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **169**, 1187–1200 (2017).

21. Helm, M. & Motorin, Y. Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* **18**, 275–291 (2017).

22. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).

23. Peschansky, V. J. & Wahlestedt, C. Non-coding RNAs as direct and indirect modulators of epigenetic regulation. *Epigenetics* **9**, 3–12 (2014).

24. Holoch, D. & Moazed, D. RNA-mediated epigenetic regulation of gene expression. *Nat. Rev. Genet.* **16**, 71–84 (2015).

25. Delcuve, G. P., Rastegar, M. & Davie, J. R. Epigenetic control. *J. Cell. Physiol.* **219**, 243–250 (2009).

26. Harvey, Z. H., Chen, Y. & Jarosz, D. F. Protein-Based Inheritance: Epigenetics beyond the Chromosome. *Mol. Cell* **69**, 195–202 (2018).

27. Houston, I. *et al.* Epigenetics in the Human Brain. *Neuropsychopharmacology* **38**, 183–197 (2013).

28. Kondilis-Mangum, H. D. & Wade, P. A. Epigenetics and the adaptive immune response. *Mol. Aspects Med.* **34**, 813–825 (2013).

29. Busslinger, M. & Tarakhovsky, A. Epigenetic Control of Immunity. *Cold Spring Harb. Perspect. Biol.* **6**, a019307–a019307 (2014).

30. Mehta, S. & Jeffrey, K. L. Beyond receptors and signaling: epigenetic factors in the regulation of innate immunity. *Immunol. Cell Biol.* **93**, 233–244 (2015).

31. Dunn, J. & Rao, S. Epigenetics and immunotherapy: The current state of play. *Mol. Immunol.* **87**, 227–239 (2017).

32. Anthony T. Annunziato. DNA Packaging: Nucleosomes and Chromatin. *Nature Education* (2008). Available at: https://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310. (Accessed: 30th April 2019)

33. Ou, H. D. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, eaag0025 (2017).

34. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244 (2017).

35. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500 (2016).

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

37. Raiber, E.-A. *et al.* 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nat. Chem.* (2018). doi:10.1038/s41557-018-0149-x

38. Zhao, Y. & Garcia, B. A. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb. Perspect. Biol.* **7**, a025064 (2015).

39. Iwasaki, W. *et al.* Contribution of histone N-terminal tails to the structure and stability of nucleosomes. *FEBS Open Bio* **3**, 363–369 (2013).

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

41. Huang, H., Lin, S., Garcia, B. A. & Zhao, Y. Quantitative Proteomic Analysis of Histone Modifications. *Chem. Rev.* **115**, 2376–2418 (2015).

42. Howe, F. S., Fischl, H., Murray, S. C. & Mellor, J. Is H3K4me3 instructive for transcription activation? *BioEssays* **39**, e201600095 (2017).

43. Becker, J. S., Nicetto, D. & Zaret, K. S. H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet.* **32**, 29–41 (2016).

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

45. Black, J. C., Van Rechem, C. & Whetstine, J. R. Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Mol. Cell* **48**, 491–507 (2012).

46. Martinez, S. & Hausinger, R. P. Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases. *J. Biol. Chem.* **290**, 20702–20711 (2015).

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

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

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

50. Clifton, I. J. *et al.* Structural studies on 2-oxoglutarate oxygenases and related double-stranded β-helix fold proteins. *J. Inorg. Biochem.* **100**, 644–669 (2006).

51. Hou, J. *et al.* Genomic amplification and a role in drug-resistance for the KDM5A histone demethylase in breast cancer. *Am. J. Transl. Res.* **4**, 247–256 (2012).

52. D 'oto, A., Tian, Q.-W., Davidoff, A. M. & Yang, J. Histone demethylases and their roles in cancer epigenetics. *J Med Oncl Ther J. Med. Oncol. J Ther. J Med Oncl Ther* **1**, 34–40 (2016).

53. Tong, D. *et al.* The HIF/PHF8/AR axis promotes prostate cancer progression. *Oncogenesis* **5**, e283 (2016).

54. Vallianatos, C. N. *et al.* Altered Gene-Regulatory Function of KDM5C by a Novel Mutation Associated With Autism and Intellectual Disability. *Front. Mol. Neurosci.* **11**, 1–12 (2018).

55. Li, Q. *et al.* Stage-Dependent and Locus-Specific Role of Histone Demethylase Jumonji D3 (JMJD3) in the Embryonic Stages of Lung Development. *PLoS Genet.* **10**, e1004524 (2014).

56. Huang, C. *et al.* Dual-specificity histone demethylase KIAA1718 (KDM7A) regulates neural differentiation through FGF4. *Cell Res.* **20**, 154–165 (2010).

57. Walsh, R. M. *et al.* Phf8 loss confers resistance to depression-like and anxiety-like behaviors in mice. *Nat. Commun.* **8**, 1–11 (2017).

58. Beyaz, S. *et al.* The histone demethylase UTX regulates the lineage-specific epigenetic program of invariant natural killer T cells. *Nat. Immunol.* **18**, 184–195 (2016).

59. Morera, L., Lübbert, M. & Jung, M. Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. *Clin. Epigenetics* **8**, 57 (2016).

60. Maes, T., Carceller, E., Salas, J., Ortega, A. & Buesa, C. Advances in the development of histone lysine demethylase inhibitors. *Curr. Opin. Pharmacol.* **23**, 52–60 (2015).

61. Højrup, C., Coleman, O. D., Bukowski, J.-P., Clausen, R. P. & Kawamura, A. JmjC-domain-Containing Histone Demethylases. in *Epigenetic Drug Discovery* 263–296 (2018).

62. Fortschegger, K. & Shiekhattar, R. Plant homeodomain fingers form a helping hand for transcription. *Epigenetics* **6**, 4–8 (2011).

63. Fortschegger, K. *et al.* PHF8 Targets Histone Methylation and RNA Polymerase II To Activate Transcription. *Mol. Cell. Biol.* **30**, 3286–3298 (2010).

64. Dephoure, N. *et al.* A quantitative atlas of mitotic phosphorylation. *Proc. Natl. Acad. Sci.* **105**, 10762–

10767 (2008).

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

66. Liu, W. *et al.* PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* **466**, 508–512 (2010).

67. Qi, H. H. *et al.* Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. *Nature* **466**, 503–507 (2010).

68. Wen, H. *et al.* Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. *J. Biol. Chem.* **285**, 9322–9326 (2010).

69. Baba, A. *et al.* PKA-dependent regulation of the histone lysine demethylase complex PHF2–ARID5B. *Nat. Cell Biol.* **13**, 669–676 (2011).

70. Stender, J. D. *et al.* Control of Proinflammatory Gene Programs by Regulated Trimethylation and Demethylation of Histone H4K20. *Mol. Cell* **48**, 28–38 (2012).

71. Horton, J. R., Upadhyay, A. K., Hashimoto, H., Zhang, X. & Cheng, X. Structural Basis for Human PHF2 Jumonji Domain Interaction with Metal Ions. *J. Mol. Biol.* **406**, 1–8 (2011).

72. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823–837 (2007).

73. Wang, J. *et al.* PHF8 and REST/NRSF co-occupy gene promoters to regulate proximal gene expression. *Sci. Rep.* **4**, 5008 (2015).

74. Feng, W., Yonezawa, M., Ye, J., Jenuwein, T. & Grummt, I. PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation. *Nat. Struct. Mol. Biol.* **17**, 445–450 (2010).

75. Lee, K. *et al.* Histone demethylase KDM7A controls androgen receptor activity and tumor growth in prostate cancer. *Int. J. Cancer* **143**, 2849–2861 (2018).

76. Choi, J.-Y. & Jo, S. A. KDM7A histone demethylase mediates TNF-α-induced ICAM1 protein upregulation by modulating lysosomal activity. *Biochem. Biophys. Res. Commun.* **478**, 1355–1362 (2016).

77. Chen, X. *et al.* Phf8 histone demethylase deficiency causes cognitive impairments through the mTOR pathway. *Nat. Commun.* **9**, 114 (2018).

78. Kleine-Kohlbrecher, D. *et al.* A Functional Link between the Histone Demethylase PHF8 and the Transcription Factor ZNF711 in X-Linked Mental Retardation. *Mol. Cell* **38**, 165–178 (2010).

79. Qiu, J. *et al.* The X-linked mental retardation gene PHF8 is a histone demethylase involved in neuronal differentiation. *Cell Res.* **20**, 908–918 (2010).

80. Shen, Y., Pan, X. & Zhao, H. The histone demethylase PHF8 is an oncogenic protein in human non-small cell lung cancer. *Biochem. Biophys. Res. Commun.* **451**, 119–125 (2014).

81. Miclaus, M. *et al.* Histone demethylase PHF8 promotes epithelial to mesenchymal transition and breast tumorigenesis. *Nucleic Acids Res.* **8**, 3256–3270 (2016).

82. Wang, Q. *et al.* Stabilization of histone demethylase PHF8 by USP7 promotes breast carcinogenesis. *J. Clin. Invest.* **126**, 2205–2220 (2016).

83. Fu, Y. *et al.* The histone demethylase PHF8 promotes adult acute lymphoblastic leukemia through interaction with the MEK/ERK signaling pathway. *Biochem. Biophys. Res. Commun.* 1–7 (2018). doi:10.1016/j.bbrc.2018.01.049

84. Arteaga, M. F. *et al.* The Histone Demethylase PHF8 Governs Retinoic Acid Response in Acute Promyelocytic Leukemia. *Cancer Cell* **23**, 376–389 (2013).

85. Maina, P. K. *et al.* c-MYC drives histone demethylase PHF8 during neuroendocrine differentiation and in castration-resistant prostate cancer. *Oncotarget* **7**, (2016).

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

87. Jin, W. *et al.* The histone demethylase PHF8 promotes prostate cancer cell growth by activating the oncomiR miR-125b. *Onco. Targets. Ther.* **8**, 1979 (2015).

88. Sun, X. *et al.* Oncogenic Features of PHF8 Histone Demethylase in Esophageal Squamous Cell Carcinoma. *PLoS One* **8**, e77353 (2013).

89. Li, S. *et al.* Histone demethylase PHF8 promotes progression and metastasis of gastric cancer. **7**, 448–461 (2017).

90. Zhou, W. *et al.* PHF8 upregulation contributes to autophagic degradation of E-cadherin, epithelial-mesenchymal transition and metastasis in hepatocellular carcinoma. *J. Exp. Clin. Cancer Res.* **37**, 215 (2018).

91. Gu, L. *et al.* The Histone Demethylase PHF8 Is Essential for Endothelial Cell Migration. *PLoS One* **11**, e0146645 (2016).

92. Liu, X., Wang, X., Bi, Y., Bu, P. & Zhang, M. The histone demethylase PHF8 represses cardiac hypertrophy upon pressure overload. *Exp. Cell Res.* **335**, 123–134 (2015).

93. Maina, P. K. *et al.* Histone demethylase PHF8 regulates hypoxia signaling through HIF1α and H3K4me3. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1860**, 1002–1012 (2017).

94. Erdoğan, Ö. *et al.* Proteomic dissection of LPS-inducible, PHF8-dependent secretome reveals novel roles of PHF8 in TLR4-induced acute inflammation and T cell proliferation. *Sci. Rep.* **6**, 24833 (2016).

95. Asensio-Juan, E. *et al.* The histone demethylase PHF8 is a molecular safeguard of the IFNγ response. *Nucleic Acids Res.* gkw1346 (2017). doi:10.1093/nar/gkw1346

96. Lim, H.-J. *et al.* The G2/M Regulator Histone Demethylase PHF8 Is Targeted for Degradation by the Anaphase-Promoting Complex Containing CDC20. *Mol. Cell. Biol.* **33**, 4166–4180 (2013).

97. Lee, K.-H. *et al.* PHF2 histone demethylase acts as a tumor suppressor in association with p53 in cancer. *Oncogene* **34**, 2897–2909 (2015).

98. Okuno, Y. *et al.* Epigenetic Regulation of Adipogenesis by PHF2 Histone Demethylase. *Diabetes* **62**, 1426–1434 (2013).

99. Lee, K.-H., Ju, U.-I., Song, J.-Y. & Chun, Y.-S. The Histone Demethylase PHF2 Promotes Fat Cell Differentiation as an Epigenetic Activator of Both C/EBPα and C/EBPδ. *Mol. Cells* **37**, 734–741 (2014).

100. Kim, H.-J. *et al.* Plant homeodomain finger protein 2 promotes bone formation by demethylating and activating Runx2 for osteoblast differentiation. *Cell Res.* **24**, 1231–1249 (2014).

101. Hata, K. *et al.* Arid5b facilitates chondrogenesis by recruiting the histone demethylase Phf2 to Sox9-regulated genes. *Nat. Commun.* **4**, 1–11 (2013).

102. 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).

103. Andrews, F. H., Strahl, B. D. & Kutateladze, T. G. Insights into newly discovered marks and readers of epigenetic information. *Nat. Chem. Biol.* **12**, 662–668 (2016).

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

105. Su, Z. & Denu, J. M. Reading the Combinatorial Histone Language. *ACS Chem. Biol.* **11**, 564–574 (2016).

106. Kwan, A. H. Y. *et al.* Engineering a Protein Scaffold from a PHD Finger. *Structure* **11**, 803–813 (2003).

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

108. Baker, L. A., Allis, C. D. & Wang, G. G. PHD fingers in human diseases: Disorders arising from misinterpreting epigenetic marks. *Mutat. Res. Mol. Mech. Mutagen.* **647**, 3–12 (2008).

109. Torres, I. O. *et al.* Histone demethylase KDM5A is regulated by its reader domain through a positive-feedback mechanism. *Nat. Commun.* **6**, 6204 (2015).

110. Gatchalian, J. & Kutateladze, T. G. *Chapter 2: PHD Fingers as Histone Readers*. (Springer International Publishing, 2015). doi:10.1007/978-3-319-18102-8

111. 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**, 670–679 (2009).

112. Chignola, F. *et al.* The solution structure of the first PHD finger of autoimmune regulator in complex with non-modified histone H3 tail reveals the antagonistic role of H3R2 methylation. *Nucleic Acids Res.* **37**, 2951–61 (2009).

113. Morrison, E. A. & Musselman, C. A. The Role of PHD Fingers in Chromatin Signaling. in *Chromatin Signaling and Diseases* 127–147 (Elsevier, 2016). doi:10.1016/B978-0-12-802389-1.00007-1

114. Yang, S., Bansal, K., Lopes, J., Benoist, C. & Mathis, D. Aire's plant homeodomain(PHD)-2 is critical for induction of immunological tolerance. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1833–8 (2013).

115. Musselman, C. A. & Kutateladze, T. G. PHD Fingers: Epigenetic Effectors and Potential Drug Targets. *Mol. Interv.* **9**, 314–323 (2009).

116. Gatchalian, J. *et al.* Chromatin condensation and recruitment of PHD finger proteins to histone H3K4me3 are mutually exclusive. *Nucleic Acids Res.* **44**, 6102–6112 (2016).

117. Musselman, C. A. *et al.* Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression. *Proc. Natl. Acad. Sci.* **109**, 787–792 (2012).

118. Houliston, R. S. *et al.* Conformational dynamics of the TTD–PHD histone reader module of the UHRF1 epigenetic regulator reveals multiple histone-binding states, allosteric regulation, and druggability. *J. Biol. Chem.* **292**, 20947–20959 (2017).

119. Herquel, B., Ouararhni, K. & Davidson, I. The TIF1α-related TRIM cofactors couple chromatin modifications to transcriptional regulation, signaling and tumor suppression. *Transcription* **2**, 231–236 (2011).

120. Shi, X. *et al.* Proteome-wide Analysis in Saccharomyces cerevisiae Identifies Several PHD Fingers as Novel Direct and Selective Binding Modules of Histone H3 Methylated at Either Lysine 4 or Lysine 36. *J. Biol. Chem.* **282**, 2450–2455 (2007).

121. Bua, D. J. *et al.* Epigenome Microarray Platform for Proteome-Wide Dissection of Chromatin-Signaling Networks. *PLoS One* **4**, e6789 (2009).

122. Wilkinson, A. W. & Gozani, O. Histone-binding domains: Strategies for discovery and characterization. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839**, 669–675 (2014).

123. Musco, G. & Peterson, P. PHD finger of autoimmune regulator: an epigenetic link between the histone modifications and tissue-specific antigen expression in thymus. *Epigenetics* **3**, 310–314 (2008).

124. Gaetani, M. *et al.* AIRE-PHD fingers are structural hubs to maintain the integrity of chromatin-associated interactome. *Nucleic Acids Res.* **40**, 11756–11768 (2012).

125. Org, T. *et al.* The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO Rep.* **9**, 370–376 (2008).

126. Wang, G. G. *et al.* Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD

finger. *Nature* **459**, 847–851 (2009).

127. Huang, Z. *et al.* NSD2 Is Recruited through Its PHD Domain to Oncogenic Gene Loci to Drive Multiple Myeloma. *Cancer Res.* **73**, 6277–6288 (2013).

128. Workman, P. & Collins, I. Probing the Probes: Fitness Factors For Small Molecule Tools. *Chem. Biol.* **17**, 561–577 (2010).

129. Blagg, J. & Workman, P. Choose and Use Your Chemical Probe Wisely to Explore Cancer Biology. *Cancer Cell* **32**, 9–25 (2017).

130. Plowright, A. T. *et al.* Joining Forces: The Chemical Biology–Medicinal Chemistry Continuum. *Cell Chem. Biol.* **24**, 1058–1065 (2017).

131. Schreiber, S. L. *et al.* Advancing Biological Understanding and Therapeutics Discovery with Small-Molecule Probes. *Cell* **161**, 1252–1265 (2015).

132. Drewes, G. & Knapp, S. Chemoproteomics and Chemical Probes for Target Discovery. *Trends Biotechnol.* **36**, 1275–1286 (2018).

133. Garbaccio, R. M. & Parmee, E. R. The Impact of Chemical Probes in Drug Discovery: A Pharmaceutical Industry Perspective. *Cell Chem. Biol.* **23**, 10–17 (2016).

134. Arrowsmith, C. H. *et al.* The promise and peril of chemical probes. *Nat. Chem. Biol.* **11**, 536–541 (2015).

135. McAllister, T. E. *et al.* Recent Progress in Histone Demethylase Inhibitors. *J. Med. Chem.* **59**, 1308–1329 (2016).

136. Wagner, E. K., Nath, N., Flemming, R., Feltenberger, J. B. & Denu, J. M. Identification and Characterization of Small Molecule Inhibitors of a Plant Homeodomain Finger. *Biochemistry* **51**, 8293–8306 (2012).

137. Bhushan, B. *et al.* Investigations on small molecule inhibitors targeting the histone H3K4 tri-methyllysine binding PHD-finger of JmjC histone demethylases. *Bioorg. Med. Chem.* **26**, 2984–2991 (2018).

138. Ali, F., Yamaguchi, K., Fukuoka, M., Elhelaly, A. E. & Kuwata, K. Logical design of an anti-cancer agent targeting the plant homeodomain in Pygopus2. *Cancer Sci.* **107**, 1321–1328 (2016).

139. Miller, T. C. R. *et al.* Competitive Binding of a Benzimidazole to the Histone-Binding Pocket of the Pygo PHD Finger. *ACS Chem. Biol.* **9**, 2864–2874 (2014).

140. Ali, M. *et al.* Molecular Insights into Inhibition of the Methylated Histone-Plant Homeodomain Complexes by Calixarenes. *J. Biol. Chem.* **290**, 22919–22930 (2015).

141. Wagner, T. *et al.* Identification of a small-molecule ligand of the epigenetic reader protein Spindlin1 via a versatile screening platform. *Nucleic Acids Res.* **44**, e88–e88 (2016).

142. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).

143. Doroshow, D. B., Eder, J. P. & LoRusso, P. M. BET inhibitors: a novel epigenetic approach. *Ann. Oncol.* **28**, 1776–1787 (2017).

144. James, L. I. *et al.* Discovery of a chemical probe for the L3MBTL3 methyllysine reader domain. *Nat. Chem. Biol.* **9**, 184–191 (2013).

145. He, Y. *et al.* The EED protein-protein interaction inhibitor A-395 inactivates the PRC2 complex. *Nat. Chem. Biol.* **13**, 389–395 (2017).

146. Qi, W. *et al.* An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED. *Nat. Chem. Biol.* **13**, 381–388 (2017).

147. Aho, E. R. *et al.* Displacement of WDR5 from Chromatin by a WIN Site Inhibitor with Picomolar Affinity. *Cell Rep.* **26**, 2916–2928.e13 (2019).

148. Milosevich, N. & Hof, F. Chemical Inhibitors of Epigenetic Methyllysine Reader Proteins. *Biochemistry* **55**, 1570–1583 (2016).

149. Milosevich, N., Warmerdam, Z. & Hof, F. Structural aspects of small-molecule inhibition of methyllysine reader proteins. *Future Med. Chem.* **8**, 1681–1702 (2016).

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

151. Thirstrup, K. *et al.* Endogenous 2-oxoglutarate levels impact potencies of competitive HIF prolyl hydroxylase inhibitors. *Pharmacol. Res.* **64**, 268–273 (2011).

152. Heinemann, B. *et al.* Inhibition of demethylases by GSK-J1/J4. *Nature* **514**, E1–E2 (2014).

153. Hatch, S. B. *et al.* Assessing histone demethylase inhibitors in cells: lessons learned. *Epigenetics Chromatin* **10**, 9 (2017).

154. Thinnes, C. C. *et al.* Betti reaction enables efficient synthesis of 8-hydroxyquinoline inhibitors of 2-oxoglutarate oxygenases. *Chem. Commun.* **51**, 15458–15461 (2015).

155. Kawamura, A. *et al.* Highly selective inhibition of histone demethylases by de novo macrocyclic peptides. *Nat. Commun.* **8**, 14773 (2017).

156. Vinogradova, M. *et al.* An inhibitor of KDM5 demethylases reduces survival of drug-tolerant cancer cells. *Nat. Chem. Biol.* **12**, 531–538 (2016).

157. Liang, J. *et al.* Lead optimization of a pyrazolo[1,5-a]pyrimidin-7(4H)-one scaffold to identify potent, selective and orally bioavailable KDM5 inhibitors suitable for in vivo biological studies. *Bioorganic Med. Chem. Lett.* **26**, 4036–4041 (2016).

158. Tumber, A. *et al.* Potent and Selective KDM5 Inhibitor Stops Cellular Demethylation of H3K4me3 at Transcription Start Sites and Proliferation of MM1S Myeloma Cells. *Cell Chem. Biol.* 1–10 (2017). doi:10.1016/j.chembiol.2017.02.006

159. England, K. S. *et al.* Optimisation of a triazolopyridine based histone demethylase inhibitor yields a potent and selective KDM2A (FBXL11) inhibitor. *Med. Chem. Commun.* **5**, 1879–1886 (2014).

160. Rose, N. R. *et al.* Plant Growth Regulator Daminozide Is a Selective Inhibitor of Human KDM2/7 Histone Demethylases. *J. Med. Chem.* **55**, 6639–6643 (2012).

161. Suzuki, T. *et al.* Identification of the KDM2/7 Histone Lysine Demethylase Subfamily Inhibitor and its Antiproliferative Activity. *J. Med. Chem.* **56**, 7222–7231 (2013).

162. Gerken, P. A. *et al.* Discovery of a Highly Selective Cell-Active Inhibitor of the Histone Lysine Demethylases KDM2/7. *Angew. Chemie Int. Ed.* **56**, 15555–15559 (2017).

163. Dorosz, J. *et al.* Structure-Based Design of a New Scaffold for Cell-Penetrating Peptidic Inhibitors of the Histone Demethylase PHF8. *ChemBioChem* **18**, 1369–1375 (2017).

164. McAllister, T. E. *et al.* Non-competitive cyclic peptides for targeting enzyme–substrate complexes. *Chem. Sci.* **9**, 4569–4578 (2018).

165. Cardote, T. A. F. & Ciulli, A. Cyclic and Macrocyclic Peptides as Chemical Tools To Recognise Protein Surfaces and Probe Protein-Protein Interactions. *ChemMedChem* **11**, 787–794 (2016).

166. Qian, Z., Dougherty, P. G. & Pei, D. Targeting intracellular protein–protein interactions with cell-permeable cyclic peptides. *Curr. Opin. Chem. Biol.* **38**, 80–86 (2017).

167. Boehm, M. *et al.* Discovery of Potent and Orally Bioavailable Macrocyclic Peptide–Peptoid Hybrid CXCR7 Modulators. *J. Med. Chem.* **60**, 9653–9663 (2017).

168. Empting, M. *Cyclic Peptides*. (Royal Society of Chemistry, 2017). doi:10.1039/9781788010153

169. White, C. J. & Yudin, A. K. Contemporary strategies for peptide macrocyclization. *Nat. Chem.* **3**, 509–

524 (2011).

170. Rogers, J. M. & Suga, H. Discovering functional, non-proteinogenic amino acid containing, peptides using genetic code reprogramming. *Org. Biomol. Chem.* **13**, 9353–9363 (2015).

171. Malins, L. R. *et al.* Peptide Macrocyclization Inspired by Non-Ribosomal Imine Natural Products. *J. Am. Chem. Soc.* **139**, 5233–5241 (2017).

172. Hacker, D. E., Hoinka, J., Iqbal, E. S., Przytycka, T. M. & Hartman, M. C. T. Highly Constrained Bicyclic Scaffolds for the Discovery of Protease-Stable Peptides via mRNA Display. *ACS Chem. Biol.* **12**, 795–804 (2017).

173. Kale, S. S. *et al.* Cyclization of peptides with two chemical bridges affords large scaffold diversities. *Nat. Chem.* **10**, 715–723 (2018).

174. Tan, Y. S., Lane, D. P. & Verma, C. S. Stapled peptide design: principles and roles of computation. *Drug Discov. Today* **21**, 1642–1653 (2016).

175. Sawyer, T. K. *et al.* Macrocyclic α helical peptide therapeutic modality: A perspective of learnings and challenges. *Bioorg. Med. Chem.* **26**, 2807–2815 (2018).

176. Verdine, G. L. & Hilinski, G. J. Stapled Peptides for Intracellular Drug Targets. in *Methods in Enzymology* **503**, 3–33 (Elsevier Inc., 2012).

177. Lau, J. L. & Dunn, M. K. Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorganic Med. Chem.* **26**, 2700–2707 (2018).

178. Morrison, C. Constrained peptides' time to shine? *Nat. Rev. Drug Discov.* **17**, 531–533 (2018).

179. Di, L. Strategic Approaches to Optimizing Peptide ADME Properties. *AAPS J.* **17**, 134–143 (2015).

180. Henninot, A., Collins, J. C. & Nuss, J. M. The Current State of Peptide Drug Discovery: Back to the Future? *J. Med. Chem.* **61**, 1382–1414 (2018).

181. Tucker, A. T. *et al.* Discovery of Next-Generation Antimicrobials through Bacterial Self-Screening of Surface-Displayed Peptide Libraries. *Cell* 1–11 (2018). doi:10.1016/j.cell.2017.12.009

182. Valdez, B. C. *et al.* Romidepsin targets multiple survival signaling pathways in malignant T cells. *Blood Cancer J.* **5**, e357–e357 (2015).

183. Marsault, E. & Peterson, M. L. Macrocycles Are Great Cycles: Applications, Opportunities, and Challenges of Synthetic Macrocycles in Drug Discovery. *J. Med. Chem.* **54**, 1961–2004 (2011).

184. Taylor, R. D., Rey-Carrizo, M., Passioura, T. & Suga, H. Identification of nonstandard macrocyclic peptide ligands through display screening. *Drug Discov. Today Technol.* **26**, 17–23 (2017).

185. Galán, A. *et al.* Library-based display technologies: where do we stand? *Mol. Biosyst.* **12**, 2342–2358 (2016).

186. Foster, A. D. *et al.* Methods for the Creation of Cyclic Peptide Libraries for Use in Lead Discovery. *J. Biomol. Screen.* **20**, 563–576 (2015).

187. Passioura, T., Katoh, T., Goto, Y. & Suga, H. Selection-based discovery of druglike macrocyclic peptides. *Annu. Rev. Biochem.* **83**, 727–52 (2014).

188. Huang, Y., Wiedmann, M. M. & Suga, H. RNA Display Methods for the Discovery of Bioactive Macrocycles. *Chem. Rev.* acs.chemrev.8b00430 (2018). doi:10.1021/acs.chemrev.8b00430

189. Franzini, R. M., Neri, D. & Scheuermann, J. DNA-Encoded Chemical Libraries: Advancing beyond Conventional Small-Molecule Libraries. *Acc. Chem. Res.* **47**, 1247–1255 (2014).

190. Belyanskaya, S. L., Ding, Y., Callahan, J. F., Lazaar, A. L. & Israel, D. I. Discovering Drugs with DNA-Encoded Library Technology: From Concept to Clinic with an Inhibitor of Soluble Epoxide Hydrolase. *ChemBioChem* (2017). doi:10.1002/cbic.201700014

191. Pande, J., Szewczyk, M. M. & Grover, A. K. Phage display: Concept, innovations, applications and future. *Biotechnol. Adv.* **28**, 849–858 (2010).

192. Hamzeh-Mivehroud, M., Alizadeh, A. A., Morris, M. B., Bret Church, W. & Dastmalchi, S. Phage display as a technology delivering on the promise of peptide drug discovery. *Drug Discov. Today* **18**, 1144–1157 (2013).

193. Sheehan, J. & Marasco, W. A. Phage and Yeast Display. *Microbiol. Spectr.* **3**, 1–12 (2015).

194. Deyle, K., Kong, X.-D. & Heinis, C. Phage Selection of Cyclic Peptides for Application in Research and Drug Development. *Acc. Chem. Res.* **50**, 1866–1874 (2017).

195. Molek, P., Strukelj, B. & Bratkovic, T. Peptide Phage Display as a Tool for Drug Discovery: Targeting Membrane Receptors. *Molecules* **16**, 857–887 (2011).

196. Krumpe, L. R. H. & Mori, T. The Use of Phage-Displayed Peptide Libraries to Develop Tumor-Targeting Drugs. *Int. J. Pept. Res. Ther.* **12**, 79–91 (2006).

197. Szardenings, M. Phage Display of Random Peptide Libraries: Applications, Limits, and Potential. *J. Recept. Signal Transduct.* **23**, 307–349 (2003).

198. Tavassoli, A. SICLOPPS cyclic peptide libraries in drug discovery. *Curr. Opin. Chem. Biol.* **38**, 30–35 (2017).

199. Horswill, A. R. & Benkovic, S. J. Cyclic peptides, a chemical genetics tool for biologists. *Cell Cycle* **4**, 552–555 (2005).

200. Tavassoli, A. & Benkovic, S. J. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in E. coli. *Nat. Protoc.* **2**, 1126–1133 (2007).

201. Birts, C. N. *et al.* A cyclic peptide inhibitor of C-terminal binding protein dimerization links metabolism with mitotic fidelity in breast cancer cells. *Chem. Sci.* **4**, 3046 (2013).

202. Spurr, I. B. *et al.* Targeting Tumour Proliferation with a Small-Molecule Inhibitor of AICAR Transformylase Homodimerization. *ChemBioChem* **13**, 1628–1634 (2012).

203. Hartman, M. C. T., Josephson, K., Lin, C.-W. & Szostak, J. W. An Expanded Set of Amino Acid Analogs for the Ribosomal Translation of Unnatural Peptides. *PLoS One* **2**, e972 (2007).

204. Josephson, K., Ricardo, A. & Szostak, J. W. mRNA display: from basic principles to macrocycle drug discovery. *Drug Discov. Today* **19**, 388–399 (2014).

205. Passioura, T. & Suga, H. A RaPID way to discover nonstandard macrocyclic peptide modulators of drug targets. *Chem. Commun.* **53**, 1931–1940 (2017).

206. Goto, Y., Katoh, T. & Suga, H. Flexizymes for genetic code reprogramming. *Nat. Protoc.* **6**, 779–790 (2011).

207. Hipolito, C. J. & Suga, H. Ribosomal production and in vitro selection of natural product-like peptidomimetics: The FIT and RaPID systems. *Curr. Opin. Chem. Biol.* **16**, 196–203 (2012).

208. Hayashi, Y., Morimoto, J. & Suga, H. In vitro selection of anti-Akt2 thioether-macrocyclic peptides leading to isoform-selective inhibitors. *ACS Chem. Biol.* **7**, 607–613 (2012).

209. Kawakami, T. *et al.* In Vitro Selection of Multiple Libraries Created by Genetic Code Reprogramming To Discover Macrocyclic Peptides That Antagonize VEGFR2 Activity in Living Cells. *ACS Chem. Biol.* **8**, 1205–1214 (2013).

210. Jongkees, S. A. K. *et al.* Rapid Discovery of Potent and Selective Glycosidase-Inhibiting De Novo Peptides. *Cell Chem. Biol.* 1–10 (2017). doi:10.1016/j.chembiol.2017.02.001

211. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354**, 1465–1465 (2016).

212. Simon, M. D. *et al.* The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell* **128**, 1003–1012 (2007).

213. Kulkarni, S. S., Sayers, J., Premdjee, B. & Payne, R. J. Rapid and efficient protein synthesis through expansion of the native chemical ligation concept. *Nat. Rev. Chem.* **2**, 122 (2018).

214. Hancock, S. M., Uprety, R., Deiters, A. & Chin, J. W. Expanding the Genetic Code of Yeast for Incorporation of Diverse Unnatural Amino Acids via a Pyrrolysyl-tRNA Synthetase/tRNA Pair. *J. Am. Chem. Soc.* **132**, 14819–14824 (2010).

215. Morimoto, J., Hayashi, Y., Iwasaki, K. & Suga, H. Flexizymes: Their Evolutionary History and the Origin of Catalytic Function. *Acc. Chem. Res.* **44**, 1359–1368 (2011).

216. Iwane, Y. *et al.* Expanding the amino acid repertoire of ribosomal polypeptide synthesis via the artificial division of codon boxes. *Nat. Chem.* **8**, 317–325 (2016).

217. Kawakami, T. & Murakami, H. Genetically Encoded Libraries of Nonstandard Peptides. *J. Nucleic Acids* **2012**, 1–15 (2012).

218. Morimoto, J., Hayashi, Y. & Suga, H. Discovery of Macrocyclic Peptides Armed with a Mechanism-Based Warhead: Isoform-Selective Inhibition of Human Deacetylase SIRT2. *Angew. Chemie Int. Ed.* **51**, 3423–3427 (2012).

219. Huhn, A. J., Guerra, R. M., Harvey, E. P., Bird, G. H. & Walensky, L. D. Selective Covalent Targeting of Anti-Apoptotic BFL-1 by Cysteine-Reactive Stapled Peptide Inhibitors. *Cell Chem. Biol.* **23**, 1123–1134 (2016).

220. Kang, T. J., Yuzawa, S. & Suga, H. Expression of Histone H3 Tails with Combinatorial Lysine Modifications under the Reprogrammed Genetic Code for the Investigation on Epigenetic Markers. *Chem. Biol.* **15**, 1166–1174 (2008).

221. Iwasaki, K., Goto, Y., Katoh, T. & Suga, H. Selective thioether macrocyclization of peptides having the N-terminal 2-chloroacetyl group and competing two or three cysteine residues in translation. *Org. Biomol. Chem.* **10**, 5783 (2012).

222. Effraim, P. R. *et al.* Natural amino acids do not require their native tRNAs for efficient selection by the ribosome. *Nat. Chem. Biol.* **5**, 947–953 (2009).

223. Goto, Y. & Suga, H. Translation Initiation with Initiator tRNA Charged with Exotic Peptides. *J. Am. Chem. Soc.* **131**, 5040–5041 (2009).

224. Goto, Y., Goto, Y., Katoh, T. & Suga, H. Preparation of materials for flexizyme reactions and genetic code reprogramming. *Protoc. Exch.* **100**, 1–14 (2011).

225. Müller, M. M. & Muir, T. W. Histones: At the Crossroads of Peptide and Protein Chemistry. *Chem. Rev.* **115**, 2296–2349 (2015).

226. Kaniskan, H. Ü., Martini, M. L. & Jin, J. Inhibitors of Protein Methyltransferases and Demethylases. *Chem. Rev.* acs.chemrev.6b00801 (2017). doi:10.1021/acs.chemrev.6b00801

227. James, L. I. *et al.* Small-Molecule Ligands of Methyl-Lysine Binding Proteins: Optimization of Selectivity for L3MBTL3. *J. Med. Chem.* **56**, 7358–7371 (2013).

228. Goto, Y., Goto, Y., Katoh, T. & Suga, H. Preparation of materials for flexizyme reactions and genetic code reprogramming. *Protoc. Exch.* (2011). doi:10.1038/protex.2011.209

229. Boucher, L. E. & Bosch, J. Development of a multifunctional tool for drug screening against plasmodial protein-protein interactions via surface plasmon resonance. *J. Mol. Recognit.* **26**, 496–500 (2013).

230. Subirós-Funosas, R., Prohens, R., Barbas, R., El-Faham, A. & Albericio, F. Oxyma: An Efficient Additive for Peptide Synthesis to Replace the Benzotriazole-Based HOBt and HOAt with a Lower Risk of Explosion. *Chem. - A Eur. J.* **15**, 9394–9403 (2009).

231. Subirós-Funosas, R., El-Faham, A. & Albericio, F. Aspartimide formation in peptide chemistry: occurrence, prevention strategies and the role of N-hydroxylamines. *Tetrahedron* **67**, 8595–8606 (2011).

232. Kamat, V. & Rafique, A. Designing binding kinetic assay on the bio-layer interferometry (BLI) biosensor to characterize antibody-antigen interactions. *Anal. Biochem.* **536**, 16–31 (2017).

233. Schlosshauer, M. & Baker, D. Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. *Protein Sci.* **13**, 1660–1669 (2004).

234. Pan, A. C., Borhani, D. W., Dror, R. O. & Shaw, D. E. Molecular determinants of drug–receptor binding kinetics. *Drug Discov. Today* **18**, 667–673 (2013).

235. Sahu, A. *et al.* Binding Kinetics, Structure-Activity Relationship, and Biotransformation of the Complement Inhibitor Compstatin. *J. Immunol.* **165**, 2491–2499 (2000).

236. Copeland, R. A., Pompliano, D. L. & Meek, T. D. Drug–target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.* **5**, 730–739 (2006).

237. Lu, H. & Tonge, P. J. Drug–target residence time: critical information for lead optimization. *Curr. Opin. Chem. Biol.* **14**, 467–474 (2010).

238. Tautermann, C. S. *et al.* Molecular basis for the long duration of action and kinetic selectivity of tiotropium for the muscarinic M3 receptor. *J. Med. Chem.* **56**, 8746–8756 (2013).

239. Andersson, K. & Hämäläinen, M. D. Replacing affinity with binding kinetics in QSAR studies resolves otherwise confounded effects. *J. Chemom.* **20**, 370–375 (2006).

240. Shuman, C. F., Vrang, L. & Danielson, U. H. Improved Structure−Activity Relationship Analysis of HIV-1 Protease Inhibitors Using Interaction Kinetic Data. *J. Med. Chem.* **47**, 5953–5961 (2004).

241. McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41**, W597–W600 (2013).

242. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).

243. Luo, X. *et al.* A Selective Inhibitor and Probe of the Cellular Functions of Jumonji C Domain-Containing Histone Demethylases. *J. Am. Chem. Soc.* **133**, 9451–9456 (2011).

244. Xu, W. *et al.* Quantitative Analysis of Histone Demethylase Probes Using Fluorescence Polarization. *J. Med. Chem.* **56**, 5198–5202 (2013).

245. Copeland, R. A. *Evaluation of Enzyme Inhibitors in Drug Discovery*. (John Wiley & Sons, Ltd, 2005).

246. 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).

247. Liu, L., Zhen, X. T., Denton, E., Marsden, B. D. & Schapira, M. ChromoHub: a data hub for navigators of chromatin-mediated signalling. *Bioinformatics* **28**, 2205–2206 (2012).

248. Finn, R. D. *et al.* InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190–D199 (2017).

249. Park, I. *et al.* Estimation of Hydrogen-Exchange Protection Factors from MD Simulation Based on Amide Hydrogen Bonding Analysis. *J. Chem. Inf. Model.* **55**, 1914–1925 (2015).

250. Radou, G., Dreyer, F. N., Tuma, R. & Paci, E. Functional Dynamics of Hexameric Helicase Probed by Hydrogen Exchange and Simulation. *Biophys. J.* **107**, 983–990 (2014).

251. Kaltashov, I. A., Bobst, C. E. & Abzalimov, R. R. Mass spectrometry-based methods to study protein architecture and dynamics. *Protein Sci.* **22**, 530–544 (2013).

252. Skinner, J. J., Lim, W. K., Bédard, S., Black, B. E. & Englander, S. W. Protein dynamics viewed by hydrogen exchange. *Protein Sci.* **21**, 996–1005 (2012).

253. Konermann, L., Pan, J. & Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**, 1224–1234 (2011).

254. Engen, J. R. Analysis of Protein Conformation and Dynamics by Hydrogen/Deuterium Exchange MS. *Anal. Chem.* **81**, 7870–7875 (2009).

255. Chalmers, M. J., Busby, S. A., Pascal, B. D., West, G. M. & Griffin, P. R. Differential hydrogen/deuterium exchange mass spectrometry analysis of protein–ligand interactions. *Expert Rev. Proteomics* **8**, 43–59 (2011).

256. Wang, H., Rempel, D. L., Giblin, D., Frieden, C. & Gross, M. L. Peptide-Level Interactions between Proteins and Small-Molecule Drug Candidates by Two Hydrogen−Deuterium Exchange MS-Based Methods: The Example of Apolipoprotein E3. *Anal. Chem.* **89**, 10687–10695 (2017).

257. Leurs, U. *et al.* Dissecting the Binding Mode of Low Affinity Phage Display Peptide Ligands to Protein Targets by Hydrogen/Deuterium Exchange Coupled to Mass Spectrometry. *Anal. Chem.* **86**, 11734–11741 (2014).

258. Gallagher, E. S. & Hudgens, J. W. Mapping Protein–Ligand Interactions with Proteolytic Fragmentation, Hydrogen/Deuterium Exchange-Mass Spectrometry. in *Methods in Enzymology* 357–404 (2016). doi:10.1016/bs.mie.2015.08.010

259. Chalmers, M. J. *et al.* Probing Protein Ligand Interactions by Automated Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* **78**, 1005–1014 (2006).

260. Abzalimov, R. R., Bobst, C. E. & Kaltashov, I. A. A New Approach to Measuring Protein Backbone Protection with High Spatial Resolution Using H/D Exchange and Electron Capture Dissociation. *Anal. Chem.* **85**, 9173–9180 (2013).

261. Maurer, T. NMR Studies of Protein–Ligand Interactions. in *Protein-Ligand Interactions* 197–214 (Humana Press, 2005). doi:10.1385/1-59259-912-5:197

262. Dias, D. M. & Ciulli, A. NMR approaches in structure-based lead discovery: Recent developments and new frontiers for targeting multi-protein complexes. *Prog. Biophys. Mol. Biol.* **116**, 101–112 (2014).

263. Gossert, A. D. & Jahnke, W. NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **97**, 82–125 (2016).

264. Furukawa, A., Konuma, T., Yanaka, S. & Sugase, K. Quantitative analysis of protein–ligand interactions by NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **96**, 47–57 (2016).

265. Pellecchia, M. *et al.* Perspectives on NMR in drug discovery: a technique comes of age. *Nat. Rev. Drug Discov.* **7**, 738–745 (2008).

266. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta - Proteins Proteomics* **1814**, 942–968 (2011).

267. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).

268. Aguirre, C., Cala, O. & Krimm, I. Overview of probing protein-ligand interactions using NMR. *Curr. Protoc. Protein Sci.* **2015**, 17.18.1-17.18.24 (2015).

269. Prestegard, J. H. *et al.* Chemical shift prediction for denatured proteins. *J. Biomol. NMR* **55**, 201–209 (2013).

270. Ulrich, E. L. *et al.* BioMagResBank. *Nucleic Acids Res.* **36**, D402–D408 (2007).

271. Albani, J. R. Origin of Tryptophan Fluorescence. in *Reviews in Fluorescence 2009* (ed. Geddes, C. D.)

225–280 (Springer New York, 2011). doi:10.1007/978-1-4419-9672-5_9

272. Chen, Y. & Barkley, M. D. Toward Understanding Tryptophan Fluorescence in Proteins †. *Biochemistry* **37**, 9976–9982 (1998).

273. Vivian, J. T. & Callis, P. R. Mechanisms of Tryptophan Fluorescence Shifts in Proteins. *Biophys. J.* **80**, 2093–2109 (2001).

274. Juszczak, L. J. & Eisenberg, A. S. The Color of Cation-π Interactions: Subtleties of Amine-Tryptophan Interaction Energetics Allow for Radical-like Visible Absorbance and Fluorescence. *J. Am. Chem. Soc.* **139**, 8302–8311 (2017).

275. Hamm, H. E., Meier, S. M., Liao, G. & Preininger, A. M. Trp fluorescence reveals an activation-dependent cation-π interaction in the Switch II region of Gα i proteins. *Protein Sci.* **18**, 2326–2335 (2009).

276. Luchansky, M. S. *et al.* Amino acid transport in thermophiles: characterization of an arginine-binding protein in Thermotoga maritima. *Mol. BioSyst.* **6**, 142–151 (2009).

277. Chatterjee, J., Laufer, B. & Kessler, H. Synthesis of N-methylated cyclic peptides. *Nat. Protoc.* **7**, 432–444 (2012).

278. White, T. R. *et al.* On-resin N-methylation of cyclic peptides for discovery of orally bioavailable scaffolds. *Nat. Chem. Biol.* **7**, 810–817 (2011).

279. Cheng, R. P., Gellman, S. H. & DeGrado, W. F. β-Peptides: From Structure to Function. *Chem. Rev.* **101**, 3219–3232 (2001).

280. Stawikowski, M. & Cudic, P. Depsipeptide Synthesis. in *Peptide Characterization and Application Protocols* (ed. Fields, G. B.) 321–339 (Humana Press, 2007). doi:10.1007/978-1-59745-430-8_13

281. Choudhary, A. & Raines, R. T. An Evaluation of Peptide-Bond Isosteres. *ChemBioChem* **12**, 1801–1807 (2011).

282. Meanwell, N. A. Synopsis of Some Recent Tactical Application of Bioisosteres in Drug Design. *J. Med. Chem.* **54**, 2529–2591 (2011).

283. Passioura, T., Bhushan, B., Tumber, A., Kawamura, A. & Suga, H. Structure-activity studies of a macrocyclic peptide inhibitor of histone lysine demethylase 4A. *Bioorg. Med. Chem.* **26**, 1225–1231 (2018).

284. Walport, L. J., Obexer, R. & Suga, H. Strategies for transitioning macrocyclic peptides to cell-permeable drug leads. *Curr. Opin. Biotechnol.* **48**, 242–250 (2017).

285. Crooks, G. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188–1190 (2004).

286. Nath Jha, A., Vishveshwara, S. & Banavar, J. R. Amino acid interaction preferences in proteins. *Protein Sci.* **19**, 603–616 (2010).

287. Salinas, R. Y., Baker, S. A., Gospe, S. M. & Arshavsky, V. Y. A Single Valine Residue Plays an Essential Role in Peripherin/rds Targeting to Photoreceptor Outer Segments. *PLoS One* **8**, e54292 (2013).

288. Razavi, A. M., Wuest, W. M. & Voelz, V. A. Computational Screening and Selection of Cyclic Peptide Hairpin Mimetics by Molecular Simulation and Kinetic Network Models. *J. Chem. Inf. Model.* **54**, 1425–1432 (2014).

289. Geng, H., Jiang, F. & Wu, Y.-D. Accurate Structure Prediction and Conformational Analysis of Cyclic Peptides with Residue-Specific Force Fields. *J. Phys. Chem. Lett.* **7**, 1805–1810 (2016).

290. Thompson, S. J., Hattotuwagama, C. K., Holliday, J. D. & Flower, D. R. On the hydrophobicity of peptides: Comparing empirical predictions of peptide log P values. *Bioinformation* **1**, 237–241 (2006).

291. Rossi Sebastiano, M. *et al.* Impact of Dynamically Exposed Polarity on Permeability and Solubility of Chameleonic Drugs Beyond the Rule of 5. *J. Med. Chem.* **61**, 4189–4202 (2018).

292. Wang, C. K., Swedberg, J. E., Harvey, P. J., Kaas, Q. & Craik, D. J. Conformational Flexibility Is a Determinant of Permeability for Cyclosporin. *J. Phys. Chem. B* **122**, 2261–2276 (2018).

293. Whitty, A. *et al.* Quantifying the chameleonic properties of macrocycles and other high-molecular-weight drugs. *Drug Discov. Today* **21**, 712–717 (2016).

294. Di, L. & Kerns, E. H. Biological assay challenges from compound solubility: strategies for bioassay optimization. *Drug Discov. Today* **11**, 446–451 (2006).

295. Cyprotex. ADME Guide. (2006). Available at: https://www.cyprotex.com/guides. (Accessed: 16th November 2018)

296. Rand, A. C. *et al.* Optimizing PK properties of cyclic peptides: the effect of side chain substitutions on permeability and clearance. *Medchemcomm* **3**, 1282 (2012).

297. Santos, G. B., Ganesan, A. & Emery, F. S. Oral Administration of Peptide-Based Drugs: Beyond Lipinski's Rule. *ChemMedChem* **11**, 2245–2251 (2016).

298. Nielsen, D. S. *et al.* Orally Absorbed Cyclic Peptides. *Chem. Rev.* **117**, 8094–8128 (2017).

299. Furukawa, A. *et al.* Passive Membrane Permeability in Cyclic Peptomer Scaffolds Is Robust to Extensive Variation in Side Chain Functionality and Backbone Geometry. *J. Med. Chem.* **59**, 9503–9512 (2016).

300. Villar, E. A. *et al.* How proteins bind macrocycles. *Nat. Chem. Biol.* **10**, 723–731 (2014).

301. Rezai, T. *et al.* Conformational flexibility, internal hydrogen bonding, and passive membrane permeability: Successful in silico prediction of the relative permeabilities of cyclic peptides. *J. Am. Chem. Soc.* **128**, 14073–14080 (2006).

302. Werner, H. M., Cabalteja, C. C. & Horne, W. S. Peptide Backbone Composition and Protease Susceptibility: Impact of Modification Type, Position, and Tandem Substitution. *ChemBioChem* **17**, 712–718 (2016).

303. Johansson, C. *et al.* Structural analysis of human KDM5B guides histone demethylase inhibitor development. *Nat. Chem. Biol.* **12**, 78–80 (2016).

304. Wang, C. K. *et al.* Rational design and synthesis of an orally bioavailable peptide guided by NMR amide temperature coefficients. *Proc. Natl. Acad. Sci.* **111**, 17504–17509 (2014).

305. Räder, A. F. B., Reichart, F., Weinmüller, M. & Kessler, H. Improving oral bioavailability of cyclic peptides by N-methylation. *Bioorganic Med. Chem.* **26**, 2766–2773 (2018).

306. Appavoo, S. D., Kaji, T., Frost, J. R., Scully, C. C. G. & Yudin, A. K. Development of Endocyclic Control Elements for Peptide Macrocycles. *J. Am. Chem. Soc.* **140**, 8763–8770 (2018).

307. Schmidpeter, P. A. M. & Schmid, F. X. Prolyl isomerization and its catalysis in protein folding and protein function. *J. Mol. Biol.* **427**, 1609–1631 (2015).

308. Pal, D. & Chakrabarti, P. Cis peptide bonds in proteins: residues involved, their conformations, interactions and locations. *J. Mol. Biol.* **294**, 271–288 (1999).

309. Stewart, D. E., Sarkar, A. & Wampler, J. E. Occurrence and role of cis peptide bonds in protein structures. *J. Mol. Biol.* **214**, 253–260 (1990).

310. 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).

311. Shi, T., Spain, S. M. & Rabenstein, D. L. Unexpectedly Fast Cis/Trans Isomerization of Xaa-Pro Peptide Bonds in Disulfide-Constrained Cyclic Peptides. *J. Am. Chem. Soc.* **126**, 790–796 (2004).

312. Koehnke, J., Naismith, J. & van der Donk, W. A. *Cyclic Peptides: From Bioorganic Synthesis to Applications*. (Royal Society of Chemistry, 2017).

313. Chung, B. K. W., White, C. J., Scully, C. C. G. & Yudin, A. K. The reactivity and conformational control of

cyclic tetrapeptides derived from aziridine-containing amino acids. *Chem. Sci.* **7**, 6662–6668 (2016).

314. Holzberger, B., Obeid, S., Welte, W., Diederichs, K. & Marx, A. Structural insights into the potential of 4-fluoroproline to modulate biophysical properties of proteins. *Chem. Sci.* **3**, 2924 (2012).

315. Pandey, A. K., Naduthambi, D., Thomas, K. M. & Zondlo, N. J. Proline Editing: A General and Practical Approach to the Synthesis of Functionally and Structurally Diverse Peptides. Analysis of Steric versus Stereoelectronic Effects of 4-Substituted Prolines on Conformation within Peptides. *J. Am. Chem. Soc.* **135**, 4333–4363 (2013).

316. Eberhardt, E. S., Panasik, N. & Raines, R. T. Inductive effects on the energetics of prolyl peptide bond isomerization: Implications for collagen folding and stability. *J. Am. Chem. Soc.* **118**, 12261–12266 (1996).

317. Shoulders, M. D. & Raines, R. T. Collagen Structure and Stability. *Annu. Rev. Biochem.* **78**, 929–958 (2009).

318. DeRider, M. L. *et al.* Collagen Stability: Insights from NMR Spectroscopic and Hybrid Density Functional Computational Investigations of the Effect of Electronegative Substituents on Prolyl Ring Conformations. *J. Am. Chem. Soc.* **124**, 2497–2505 (2002).

319. O'Hagan, D., Bilton, C., Howard, J. a. K., Knight, L. & Tozer, D. J. The preferred conformation of N-β-fluoroethylamides. Observation of the fluorine amide gauche effect. *J. Chem. Soc. Perkin Trans. 2* 605–607 (2000). doi:10.1039/b000205o

320. Hodges, J. A. & Raines, R. T. Energetics of an n → π * Interaction that Impacts Protein Structure. *Org. Lett.* **8**, 4695–4697 (2006).

321. Gutowsky, H. S. & Holm, C. H. Rate Processes and Nuclear Magnetic Resonance Spectra. II. Hindered Internal Rotation of Amides. *J. Chem. Phys.* **25**, 1228–1234 (1956).

322. Blixt, O. *et al.* A High-Throughput O -Glycopeptide Discovery Platform for Seromic Profiling. *J. Proteome Res.* **9**, 5250–5261 (2010).

323. Hilpert, K., Winkler, D. F. & Hancock, R. E. Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion. *Nat. Protoc.* **2**, 1333–1349 (2007).

324. Abegg, D. *et al.* Strained Cyclic Disulfides Enable Cellular Uptake by Reacting with the Transferrin Receptor. *J. Am. Chem. Soc.* **139**, 231–238 (2017).

325. Young Kim, H., Young Yum, S., Jang, G. & Ahn, D. Discovery of a non-cationic cell penetrating peptide derived from membrane-interacting human proteins and its potential as a protein delivery carrier. *Sci. Rep.* **5**, 11719 (2015).

326. Montrose, K., Yang, Y., Sun, X., Wiles, S. & Krissansen, G. W. Xentry, a new class of cell-penetrating peptide uniquely equipped for delivery of drugs. *Sci. Rep.* **3**, 1661 (2013).

327. Wang, F. *et al.* Recent progress of cell-penetrating peptides as new carriers for intracellular cargo delivery. *J. Control. Release* **174**, 126–136 (2014).

328. Guidotti, G., Brambilla, L. & Rossi, D. Cell-Penetrating Peptides : From Basic Research to Clinics. *Trends Pharmacol. Sci.* **38**, 406–424 (2017).

329. Collado Camps, E. & Brock, R. An opportunistic route to success: Towards a change of paradigm to fully exploit the potential of cell-penetrating peptides. *Bioorganic Med. Chem.* **26**, 2780–2787 (2018).

330. Palm-Apergi, C., Lönn, P. & Dowdy, S. F. Do Cell-Penetrating Peptides Actually "Penetrate" Cellular Membranes? *Mol. Ther.* **20**, 695–697 (2012).

331. Ramos-Molina, B. *et al.* Cationic Cell-Penetrating Peptides Are Potent Furin Inhibitors. *PLoS One* **10**, e0130417 (2015).

332. Fotin-Mleczek, M. Cationic cell-penetrating peptides interfere with TNF signalling by induction of TNF receptor internalization. *J. Cell Sci.* **118**, 3339–3351 (2005).

333. Lee, D. *et al.* Effective delivery of a rationally designed intracellular peptide drug with gold nanoparticle–peptide hybrids. *Nanoscale* **7**, 12356–12360 (2015).

334. Torchilin, V. Intracellular delivery of protein and peptide therapeutics. *Drug Discov. Today Technol.* **5**, e95–e103 (2008).

335. Fiacco, S. V. *et al.* Directed Evolution of Scanning Unnatural-Protease-Resistant (SUPR) Peptides for in Vivo Applications. *ChemBioChem* **17**, 1643–1651 (2016).

336. Howell, S. M. *et al.* Serum Stable Natural Peptides Designed by mRNA Display. *Sci. Rep.* **4**, 6008 (2015).

337. Penchala, S. C. *et al.* A biomimetic approach for enhancing the in vivo half-life of peptides. *Nat. Chem. Biol.* **11**, 793–798 (2015).

338. Rogers, J. M., Passioura, T. & Suga, H. Nonproteinogenic deep mutational scanning of linear and cyclic peptides. *Proc. Natl. Acad. Sci.* **115**, 10959–10964 (2018).

339. Ben-David, U. *et al.* Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* **560**, 325–330 (2018).

340. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative Proteomic Analysis of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins. *Mol. Cell. Proteomics* **11**, M111.014050 (2012).

341. Simon, G. M., Niphakis, M. J. & Cravatt, B. F. Determining target engagement in living systems. *Nat. Chem. Biol.* **9**, 200–205 (2013).

342. Schürmann, M., Janning, P., Ziegler, S. & Waldmann, H. Small-Molecule Target Engagement in Cells. *Cell Chem. Biol.* **23**, 435–441 (2016).

343. Dubach, J. M. *et al.* Quantitating drug-target engagement in single cells in vitro and in vivo. *Nat. Chem. Biol.* **13**, 168–173 (2017).

344. Machleidt, T. *et al.* NanoBRET-A Novel BRET Platform for the Analysis of Protein-Protein Interactions. *ACS Chem. Biol.* **10**, 1797–1804 (2015).

345. Hall, M. P. *et al.* Engineered Luciferase Reporter from a Deep Sea Shrimp Utilizing a Novel Imidazopyrazinone Substrate. *ACS Chem. Biol.* **7**, 1848–1857 (2012).

346. Lin, C.-J., Conti, M. & Ramalho-Santos, M. Histone variant H3.3 maintains a decondensed chromatin state essential for mouse preimplantation development. *Development* **140**, 3624–3634 (2013).

347. Richon, V. M. Cancer biology: mechanism of antitumour action of vorinostat (suberoylanilide hydroxamic acid), a novel histone deacetylase inhibitor. *Br. J. Cancer* **95**, S2–S6 (2006).

348. Robers, M. B. *et al.* Target engagement and drug residence time can be observed in living cells with BRET. *Nat. Commun.* **6**, 10091 (2015).

349. Stoddart, L. A., Kilpatrick, L. E. & Hill, S. J. NanoBRET Approaches to Study Ligand Binding to GPCRs and RTKs. *Trends Pharmacol. Sci.* **39**, 136–147 (2018).

350. Kilpatrick, L. E. *et al.* Real-time analysis of the binding of fluorescent VEGF 165 a to VEGFR2 in living cells: Effect of receptor tyrosine kinase inhibitors and fate of internalized agonist-receptor complexes. *Biochem. Pharmacol.* **136**, 62–75 (2017).

351. Lau, Y. H., Wu, Y., de Andrade, P., Galloway, W. R. J. D. & Spring, D. R. A two-component 'double-click' approach to peptide stapling. *Nat. Protoc.* **10**, 585–594 (2015).

352. Bautista, S. J. *et al.* mTOR complex 1 controls the nuclear localization and function of glycogen synthase kinase 3β. *J. Biol. Chem.* **293**, 14723–14739 (2018).

353. Greer, E. L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **13**, 343–357 (2012).

354. Wagner, T., Robaa, D., Sippl, W. & Jung, M. Mind the Methyl: Methyllysine Binding Proteins in Epigenetic Regulation. *ChemMedChem* **9**, 466–483 (2014).

355. James, L. I. & Frye, S. V. Chemical probes for methyl lysine reader domains. *Curr. Opin. Chem. Biol.* **33**, 135–141 (2016).

356. Lallous, N. & Ramón-Maiques, S. Chromatin Recognition Protein Modules: The PHD Finger. in *Encyclopedia of Life Sciences* 1–14 (John Wiley & Sons, Ltd, 2011). doi:10.1002/9780470015902.a0023176

357. Santiago, C., Nguyen, K. & Schapira, M. Druggability of methyl-lysine binding sites. *J. Comput. Aided. Mol. Des.* **25**, 1171–1178 (2011).

358. Yeh, T.-L. *et al.* Molecular and cellular mechanisms of HIF prolyl hydroxylase inhibitors in clinical trials. *Chem. Sci.* **8**, 7651–7668 (2017).

359. Moustakim, M. *et al.* Discovery of a PCAF Bromodomain Chemical Probe. *Angew. Chemie Int. Ed.* **56**, 827–831 (2017).

360. Moustakim, M. *et al.* Discovery of an MLLT1/3 YEATS Domain Chemical Probe. *Angew. Chemie Int. Ed.* **57**, 16302–16307 (2018).

361. Bowkett, D. *et al.* Identifying Small-Molecule Binding Sites for Epigenetic Proteins at Domain-Domain Interfaces. *ChemMedChem* **13**, 1051–1057 (2018).

362. Baba, R., Hori, Y., Mizukami, S. & Kikuchi, K. Development of a Fluorogenic Probe with a Transesterification Switch for Detection of Histone Deacetylase Activity. *J. Am. Chem. Soc.* **134**, 14310–3 (2012).

363. Murakami, H., Ohta, A., Ashigai, H. & Suga, H. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat. Methods* **3**, 357–359 (2006).

364. Mathew, A. E., Mejillano, M. R., Nath, J. P., Himes, R. H. & Stella, V. J. Synthesis and evaluation of some water-soluble prodrugs and derivatives of taxol with antitumor activity. *J. Med. Chem.* **35**, 145–151 (1992).

365. Gupta, P. K. *et al.* Inhibitors selective for HDAC6 in enzymes and cells. *Bioorg. Med. Chem. Lett.* **20**, 7067–7070 (2010).

366. Anderson, Z. J. & Fox, D. J. Total synthesis of the azolemycins. *Org. Biomol. Chem.* **14**, 1450–1454 (2016).

367. Huang, Z.-P., Su, X.-Y., Du, J.-T., Zhao, Y.-F. & Li, Y.-M. Facile synthesis of Nε-(benzyl, methyl)-lysine as a building block for site-specifically lysine monomethylated peptides. *Tetrahedron Lett.* **47**, 5997–5999 (2006).

368. Ferreira, R. S. *et al.* Synthesis, Biological Evaluation, and Structure–Activity Relationships of Potent Noncovalent and Nonpeptidic Cruzain Inhibitors as Anti- Trypanosoma cruzi Agents. *J. Med. Chem.* **57**, 2380–2392 (2014).

369. Gileadi, O. *et al.* High Throughput Production of Recombinant Human Proteins for Crystallography. in *Structural Proteomics: High-Throughput Methods* (eds. Kobe, B., Guss, M. & Huber, T.) 221–246 (Humana Press, 2008). doi:10.1007/978-1-60327-058-8_14

370. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).

371. Pagès, H., Aboyoun, P., Gentleman, R. & DebRoy, S. Biostrings.

372. McHugh, M. L. Multiple comparison analysis testing in ANOVA. *Biochem. Medica* **21**, 203–209 (2011).