# Structural and Biochemical Studies on Transcriptional Regulation and Metallo-β-Lactamase Inhibition




**David Leo Zollman**

Chemistry Research Laboratory
Department of Chemical Biology
Magdalen College
University of Oxford

**Hilary Term 2018**

*A thesis submitted to the Board of the Faculty of Mathematical, Physical and Life Sciences of the University of Oxford in partial fulfilment of the requirements for the degree of Doctor of Philosophy*

# Abstract

**Part I** of this thesis explores the utility and limitations of employing constructs of varying complexity as simplified models for the cellular environment in the study of transcriptional regulation; the background is given in **Chapter 1**.

**Chapter 2** describes work contributing to the successful design of a small molecule dual inhibitor capable of inducing a hypoxic-like transcriptional response while also inhibiting bromodomain 4 (BRD4) function in cells. The ability of candidate dual inhibitors to interrupt the binding between BRD4 and a histone fragment binding partner was investigated by AlphaScreen. These studies helped to inform on the importance of several structural features on inhibitor–BRD4 binding. AlphaScreen was subsequently used (**Chapter 3**) to investigate the effect of a newly detected histone modification, lysine hydroxyacetylation (KAcOH), on BRD binding. KAcOH was found to interrupt binding to BRD4, indicating a potential role for KAcOH as a 'PTM silencer'.

Studies detailed in **Chapter 4** probe the interactions between a histone binding protein (tripartite motif containing 24; TRIM24) and a recombinant, homogeneously post-translationally modified mononucleosome containing an $N^{\varepsilon}$-acetylthialysine residue at the 18-position of histone 3 (H3 sK18Ac). Crystallography was used to analyse the H3 sK18Ac nucleosome, as well as TRIM24. The binding of TRIM24 to the H3 sK18Ac nucleosome was investigated using non-denaturing mass spectrometry, small angle X-ray scattering and isothermal titration calorimetry. Together, these results provided insight into TRIM24–H3 sK18Ac nucleosome binding, as well as structural insight into the complex formed.

The importance of substrate complexity on the activity of JmjC $N^{\varepsilon}$-methyl lysine histone demethylases (KDMs) was the focus of **Chapter 5**. An antibody-based platform capable of detecting demethylation of $N^{\varepsilon}$-trimethylthialysine ($sKMe_3$) marks was optimised and used to compare the activity of the demethylase KDM4A against H3 $sK9Me_3$ peptide, histone, and nucleosome substitutes. Initial attempts to discover a demethylase for $K79Me_3$ were also made using this platform.

**Part II** details structural studies on another family of metallo proteins; the metallo-β-lactamases (MBLs), which are clinically important due to their ability to hydrolyse almost all known β-lactam antibiotics. Structural analyses on complexes of a range of MBLs with stereoisomers of the angiotensin converting enzyme (ACE) inhibitor, captopril, reveal information about binding in **Chapter 6**. A novel binding mode for a *de novo* designed MBL inhibitor is revealed in **Chapter 7**.

# Acknowledgements

As with all modern scientific progress, the work presented in this thesis would not have been possible without the scientific foundations built up by countless researchers, the advice and collaboration of workmates, and the support of friends and family.

I would first like to thank my supervisor Professor Chris Schofield for providing the most nurturing research environment I could have asked for. I greatly appreciate the enthusiasm, encouragement and ideas you consistently provided, along with the freedom to direct my own projects. For their practical supervision I would also like to thank Dr Jürgen Brem for welcoming me as an honorary 'Jürgling', and Dr Mike McDonough, Master of Crystallography, for taking me under his wing.

I would like to thank the BBSRC for funding, and Professor Gail Preston, along with the support staff at the DTP, for coordinating such an engaging training program. I would also like to thank the rest of my DTP cohort for making my first year so enjoyable.

I was very fortunate to collaborate with many knowledgeable, hard-working scientists during these projects. Specifically, I would like to thank Dr Lukas Lercher for his advice regarding nucleosome production, Dr Angelina Sekirnik for her contribution to the TRIM24 project, and Dr Hidenori Someya for his collaboration on the dual inhibitor studies. All specific contributions from collaborators are acknowledged in the main text.

Most members of the CJS group have provided help and advice to me one stage or another, which I appreciate wholeheartedly. In particular, I would like to thank Sarah Wilkins, Akane Kawamura, Roman Belle, Ritu Raj, Rebecca Hancock, Yijia Zhang, Larissa See, Sander van Berkel, Anne Makena, James McCullagh, Richard Hopkins, Kate Dunne, Bhaskar Bhushan, Ed Lowe, Saiful Islam and Herminio Manso. On a more personal note, I would like to thank Joanna Bonnici, Leah Taylor Kearney, Martine Abboud, Chris Lohans, Anne Makena and Kerstin Lippl for making life in the lab a pleasure, even on the days when experiments misbehaved.

There were many good times and difficult times over the course of my doctorate, but by going through them all with my friends, the last few years were some of the best of my life. I would like to express my warmest appreciation to Reinier van Straten, Alexandra Constantinescu, Matthias Brinkmann, Niels Martens, Chris Vogel, Ed and Imogen Doddridge, Alison Feder, Marta Arnaldi and many others for sharing their time in Oxford with me. I would also like to thank the cast and crew of Canary for enabling a necessary distraction from my studies.

Finally, and most importantly, I thank my parents, and sister, whose love and support over the course of my doctorate, and all the years before, enabled any of this to happen. And to Katie I give my love and appreciation; you never let me doubt myself.

# Abbreviations

| | |
|---|---|
| 2OG | 2-Oxoglutarate |
| 2-TY | 2 × Tryptone/yeast extract |
| ACE | Angiotensin converting enzyme |
| APS | Ammonium persulfate |
| ARNT | Aryl hydrocarbon receptor nuclear translocator |
| ATP | Adenosine triphosphate |
| BCII | *Bacillus cereus* II |
| BET | Bromodomain and extra-terminal |
| bp | Base pairs |
| BME | β-Mercaptoethanol |
| BRD | Bromodomain |
| BRD4(1) | Bromodomain 4 (first bromodomain) |
| BSA | Bovine serum albumin |
| *C*-terminus | Carboxy-terminus |
| CA9 | Carbonic anhydrase 9 |
| CADD | Computer assisted drug design |
| CB | Crystallisation buffer |
| CCD | Charge coupled device |
| CpG | Cytosine-phosphate-guanine |
| CV | Column volumes |
| DAMT | DNA $N^6$-adenine methyltransferase |
| DCLRE | DNA cross-link repair enzyme |
| DIC | *N*,*N*-Diisopropylcarbodiimide |
| $d_{max}$ | Maximum diameter of an object |
| DMF | Dimethylformamide |
| DNA | Deoxyribonucleic acid |
| DNMT | DNA methyltransferases |
| dNTP | Deoxyribonucleotide |
| DOT1L | Disruptor of telomeric silencing 1-like |
| DTT | Dithiothreitol |

| | |
|---|---|
| $EC_{50}$ | Half maximal effective concentration |
| EDTA | Ethylenediaminetetraacetic acid |
| FAC | Final assay concentration |
| FAD | Flavin adenine dinucleotide |
| FBS | Fetal bovine serum |
| Fmoc | Fluorenylmethyloxycarbonyl chloride |
| FPLC | Fast protein liquid chromatography |
| H1 | Histone 1 |
| H2A | Histone 2A |
| H2B | Histone 2B |
| H3 | Histone 3 |
| H4 | Histone 4 |
| HAGH | Hydroxyacylglutathione hydrolase (glyoxalase II) |
| HAT | Histone acetyltransferase |
| HDAC | Histone deacetylase |
| HeLa | A human cell line derived from cells obtained from Henrietta Lacks |
| HEPES | 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid |
| HIF | Hypoxia Inducible Factor |
| HMT | Histone methyltransferase |
| HPLC | High performance liquid chromatography |
| HRP | Horseradish peroxidase |
| HSB | High salt buffer |
| HTS | High throughput screening |
| $IC_{50}$ | Half maximal inhibitory concentration |
| IMP | Imipenemase |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| ITC | Isothermal calorimetry |
| JmjC | Jumonji-*C* (domain) |
| JmjN | Jumonji-*N* (domain) |
| JMJD | Jumonji domain-containing protein |
| KAc | $N^{\varepsilon}$-Acetyllysine |
| KAcOH | $N^{\varepsilon}$-Hydroxyacetyllysine |

| | |
|---|---|
| $K_D$ | Dissociation constant |
| KDM | Lysine demethylase |
| $K_i$ | Inhibitor constant |
| $K_M$ | Substrate binding constant |
| $KMe_1$ | $N^{\varepsilon}$-Monomethyllysine |
| $KMe_2$ | $N^{\varepsilon}$-Dimethyllysine |
| $KMe_3$ | $N^{\varepsilon}$-Trimethyllysine |
| KMT | Lysine methyltransferase |
| LC-MS | Liquid chromatography-mass spectrometry |
| LSB | Low salt buffer |
| LSD1 | Lysine specific demethylase 1 |
| MALDI | Matrix assisted laser desorption ionisation |
| MBL | Metallo-β-lactamase |
| MS | Mass spectrometry |
| MW | Molecular weight |
| *N*-terminus | Amino terminus |
| NDM | New Delhi MBL |
| NMAD | $N^6$-Methyladenine demethylase |
| NMR | Nuclear magnetic resonance |
| NVA | *N*-Vinylacetamide |
| $OD_{600}$ | Optical density at λ = 600 nm |
| PAGE | Polyacrylamide gel electrophoresis |
| PBP | Penicillin binding protein |
| PDB | Protein Data Bank |
| PHD | Plant homeodomain |
| PHD2 | Prolyl-hydroxylase domain-containing protein 2 |
| PHF2 | PHD finger protein 2 |
| RMSD | Root mean square deviation |
| P(r) | Pair-distance distribution function |
| PRMT | Protein arginine methyltransferase |
| PTM | Post-translational modification |
| pVHL | Von Hippel-Lindau tumour suppressor protein |

| | |
|---|---|
| Q-TOF | Quadrupole-time of flight |
| RNA | Ribonucleic acid |
| rpm | Revolutions per minute |
| RT | Room temperature (~20 °C) |
| SAH | *S*-Adenosyl-homocysteine |
| SAM | *S*-Adenosyl methionine |
| SAXS | Small-angle X-ray scattering |
| SBL | Serine-β-lactamase |
| SDM | Site-directed mutagenesis |
| SDS | Sodium dodecylsulfate |
| SEC | Size-exclusion chromatography |
| Sfh-1 | *Serratia fonticola*-1 |
| sKAc | $N^{\varepsilon}$-Acetylthialysine |
| $sKMe_1$ | $N^{\varepsilon}$-Monomethylthialysine |
| $sKMe_2$ | $N^{\varepsilon}$-Dimethylthialysine |
| $sKMe_3$ | $N^{\varepsilon}$-Trimethylthialysine |
| SPPS | Solid phase peptide synthesis |
| TEMED | Tetramethylethylenediamine |
| TET | Ten-eleven translocation |
| TFA | Trifluoroacetic acid |
| TMSP-$d_4$ | 3-(Trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid |
| TOF | Time of flight |
| Tris base | 2-Amino-2-(hydroxymethyl)-1,3-propanediol |
| TRIM24 | Tripartate motif-containing 24 |
| tRNA | Transfer RNA |
| UPLC | Ultra-performance liquid chromatography |
| VA-044 | 2,2'-Azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride |
| VEGF-A | Vascular endothelial growth factor A |
| VIM | Verona integron-encoded MBL |
| v/w | Volume per weight |
| v/v | Volume per volume |
| w/v | Weight per volume |

# Experiments performed in collaboration

**Chapter 2**

All compounds synthesised in **Chapter 2** were prepared by *Dr Hidenori Someya* unless otherwise stated. PHD2 binding experiments were conducted by *Kerstin Lippl*, pVHL binding experiments by *Lucia Serran Aguilera* and all in-cell data was obtained by *Tzu-Lan Yeh*.

**Chapter 3**

The synthesis of $N^{\alpha}$-Fmoc-protected hydroxyacetylated lysine was conducted by *Roman Belle*.

**Chapter 4**

TRIM24 protein was supplied by *Dr Angelina Sekirnik*. Non-denaturing MS data were generated by *Dr Idlir Liko* and *Dr Nisha Patel*.

**Chapter 5**

2OG-dependent KDM proteins were supplied by *Dr Rebecca Hancock* (KDM4A), *Joanna Bonnici* (KDM4E), *Dr Sarah Wilkins* (JMJD4, JMJD5 and JMJD7), *Dr Saiful Islam* (JMJD6) and *Yijia Zhang* (PHF2).

**Chapter 6**

L-, D-, *epi*-L-, and *epi*-D-captopril were synthesised by *Dr Sander van Berkel*. All MBL proteins used were produced and purified by *Dr Jürgen Brem*. Kinetic and nucleotide assays were also conducted by *Dr Jürgen Brem*.

**Chapter 7**

CADD, *in silico* screening, and MBL inhibitor synthesis were conducted by *Dr Ricky Cain*. VIM-2 protein was produced and purified by *Dr Jürgen Brem*.

Experiments performed in collaboration with others are also specified in the 'objectives' section of each experimental chapter and are referenced again in the main text.

# Table of contents

*Part I: Investigating the Function and Selectivity of Transcription Factors to Peptide, Histone and Nucleosome Substrates* ..... 1

*Chapter 1: Introduction* ..... 2
**1.1 Epigenetics** ..... 2
**1.2 Gene regulation throughout Life** ..... 3
1.2.1 Prokaryotes ..... 3
1.2.2 Eukaryotes ..... 3
1.2.3 Archaea ..... 4
**1.3 DNA storage in eukaryotes** ..... 4
1.3.1 Chromatin ..... 4
1.3.2 Nucleosomes ..... 6
1.3.3 Histones ..... 6
**1.4 Mechanisms for controlling gene expression** ..... 7
1.4.1 DNA modification ..... 8
1.4.2 Post-translational modifications of histones ..... 11
1.4.3 Epigenetic drug targets ..... 21
**1.5 Aims of Part I** ..... 22
**1.6 References** ..... 23

*Chapter 2: Using AlphaScreen to Aid the Design of a Dual Inhibitor* ..... 32
**2.1 Introduction** ..... 31
2.1.1 The hypoxic response ..... 31
2.1.2 HIF pathway ..... 31
2.1.3 Oxygen-dependent regulation of HIF-1α ..... 32
2.1.4 Small molecule induction of the hypoxic response ..... 33
2.1.5 Bromodomain and extra-terminal proteins ..... 34
2.1.6 Bi-functional inhibitors ..... 35
**2.2 Objectives** ..... 36
**2.3 AlphaScreen assays** ..... 37
2.3.1 BRD4(1) production ..... 38
2.3.2 AlphaScreen dose-response curve ..... 41
**2.4 Designing and testing a bi-functional BRD–PHD inhibitor** ..... 43
2.4.1 Selecting inhibitor linking sites ..... 43
2.4.2 Investigating the effect of linker length on the binding strength of a selection of dual BRD–PHD inhibitors to BRD4(1) ..... 44
2.4.3 BRD–PHD dual inhibitor AlphaScreen assay ..... 46
2.4.4 BRD–PHD dual inhibitor activity against PHD2 and *in vivo* activity ..... 48

**2.5 Designing and testing a bi-functional BRD–pVHL inhibitor** ..... 48
2.5.1 Designing a BRD–pVHL dual inhibitor ..... 48
2.5.2 BRD–pVHL dual inhibitor AlphaScreen assay ..... 51
2.5.3 BRD–pVHL dual inhibitor activity against pVHL and *in vivo* activity ..... 52
**2.6 Discussion** ..... 53
**2.7 References** ..... 55

*Chapter 3: Investigation of a Putative New PTM by AlphaScreen* ..... 57
**3.1 Introduction** ..... 57
3.1.1 Post-translational modifications ..... 57
3.1.2 KDM3A ..... 57
**3.2 Objectives** ..... 59
**3.3 Synthesis of peptides containing hydroxyacetylated lysine marks** ..... 60
3.3.1 Design of hydroxyacetylated $H4_{1-20}$ peptides ..... 60
3.3.2 Peptide synthesis ..... 61
**3.4 AlphaScreen experiments** ..... 65
**3.5 Discussion** ..... 67
**3.6 References** ..... 69

*Chapter 4: Investigating the Interactions between TRIM24 and the Nucleosome* ..... 70
**4.1 Introduction** ..... 70
4.1.1 Peptides and nucleosomes in epigenetic research ..... 70
4.1.2 TRIM24 ..... 70
**4.2 Objectives** ..... 73
**4.3 H3 sK18Ac nucleosome production** ..... 74
4.3.1 Histone production ..... 75
4.3.2 Methods for incorporating post-translational modifications into proteins ..... 75
4.3.3 Chemical incorporation of an acetyl group to K18 ..... 78
4.3.4 Octamer folding ..... 79
4.3.5 DNA production ..... 80
4.3.6 Nucleosome reconstitution ..... 84
**4.4 Crystallography** ..... 85
4.4.1 Crystallography of a modified nucleosome ..... 86
4.4.2 Crystallography of TRIM24 ..... 89
4.4.3 H3 sK18Ac nucleosome crystal structure ..... 91
4.4.4 TRIM24 crystal structure ..... 98
4.4.5 TRIM24–peptide crystallography ..... 101
4.4.6 Nucleosome–TRIM24 complex crystallography ..... 101
**4.5 Non-denaturing mass spectrometry** ..... 102
**4.6 Small angle X-ray scattering** ..... 107
4.6.1 Background ..... 107
4.6.2 Results ..... 107

**4.7 Isothermal titration calorimetry** ..... 112
4.7.1 Background ..... 112
4.7.2 Results ..... 113
**4.8 Discussion** ..... 116
**4.9 References** ..... 119

*Chapter Chapter 5: Development of Assays for Monitoring the Demethylation of Recombinant Nucleosomes* ..... 122
**5.1 Introduction** ..... 122
5.1.1 K79 methylation mark ..... 122
5.1.2 Methods for monitoring demethylation reactions catalysed by 2OG-oxygenases ..... 123
**5.2 Objectives** ..... 124
**5.3 Production of peptide and nucleosome methylated lysine analogues** ..... 125
5.3.1 Synthesis of methylated lysine analogues ..... 126
5.3.2 Reconstitution of nucleosomes containing methylated lysine analogues ..... 128
**5.4 Monitoring demethylation by mass spectrometry** ..... 129
**5.5 Antibody-based detection methods** ..... 131
5.5.1 Western blots ..... 131
5.5.2 Investigating KDM4A-catalysed H3 sK9$Me_3$ nucleosome demethylation by western blot ..... 131
5.5.3 Dot blots ..... 134
5.5.4 Comparing KDM4A-catalysed demethylation of peptide, histone and nucleosome substrates by dot blot ..... 135
5.5.5 Application of dot blot demethylation in the hunt for an H3 K79 demethylase ..... 137
**5.6 Discussion** ..... 138
**5.7 Part I conclusions** ..... 140
**5.8 References** ..... 143

*Part II: Structural Studies on Metallo-β-Lactamase Inhibition* ..... 145

*Chapter 6: Structural Basis of Metallo-β-lactamase Inhibition by Captopril Stereoisomers* ..... 146
**6.1 Introduction** ..... 146
6.1.1 Antibiotics ..... 146
6.1.2 Antibiotic resistance ..... 149
6.1.3 β-Lactamases ..... 150
6.1.4 β-Lactamase inhibitors ..... 155
6.1.5 Captopril ..... 158

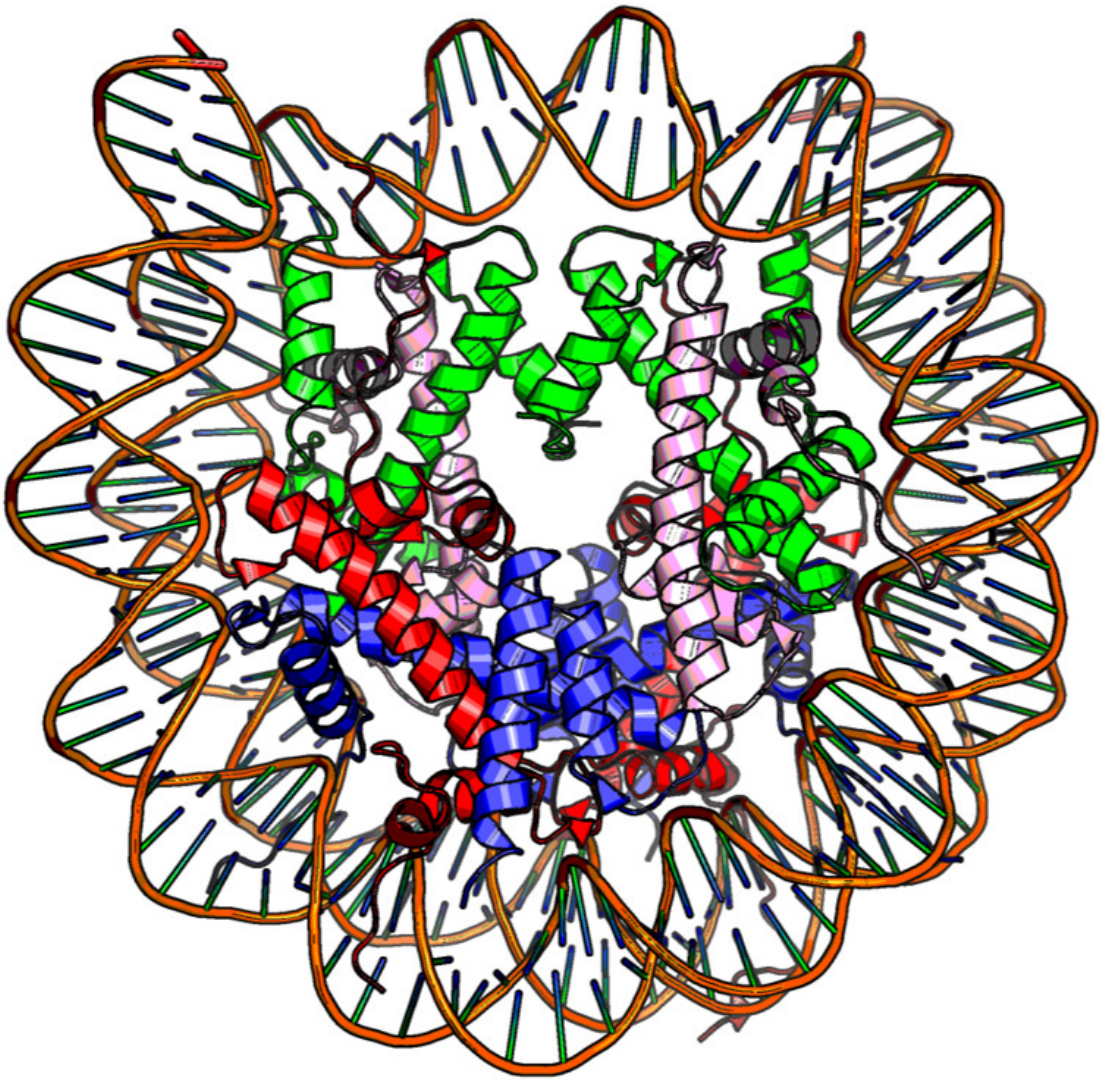
**6.2 Aims of Part II** ..... 160
**6.3 Objectives** ..... 161
**6.4 Target selection** ..... 162
6.4.1 Captopril ..... 162
6.4.2 Metallo-β-lactamases ..... 162
**6.5 Metallo-β-lactamase inhibition by captopril stereoisomers** ..... 163
6.5.1 $K_i$ determination using a fluorescent reporter assay ..... 163
6.5.2 Probing human metallo-β-lactamase-fold enzyme inhibition by captopril stereoisomers ..... 164
**6.6 Structural analysis of captopril binding to metallo-β-lactamases** ..... 165
6.6.1 Data collection ..... 165
6.6.2 Structural analysis ..... 171
**6.7 Discussion** ..... 181
**6.8 References** ..... 183

*Chapter 7: Structural Basis for the Inhibition of VIM-2 by a Novel,* In Silico-*Designed Small Molecule Inhibitor* ..... 189
**7.1 Introduction** ..... 189
7.1.1 Classical and reverse pharmacology ..... 189
7.1.2 Screening methods ..... 190
**7.2 Objectives** ..... 192
**7.3 *De novo* ligand design** ..... 193
**7.4 Metallo-β-lactamase inhibition by a *de novo*-designed ligand** ..... 194
**7.5 Structural analysis of a *de novo*-designed ligand binding to VIM-2** ..... 194
7.5.1 Data collection ..... 194
7.5.2 Structural analysis ..... 196
7.5.3 Structural optimisation ..... 197
**7.6 Discussion** ..... 199
**7.7 Part II conclusions** ..... 199
**7.8 References** ..... 201

*Chapter 8: Materials and Methods* ..... 203
**8.1 General methods** ..... 203
**8.2 Gel electrophoresis** ..... 205
8.2.1 SDS-PAGE ..... 205
8.2.2 Western blot procedure ..... 207
8.2.3 Dot blot ..... 208
8.2.4 Native PAGE ..... 208
**8.3 Peptide synthesis** ..... 208
8.3.1 Solid phase peptide synthesis ..... 208
8.3.2 Reverse phase high performance liquid chromatography ..... 209

**8.4 DNA techniques** ..... 210
8.4.1 DNA sequencing ..... 210
8.4.2 Plasmid purification ..... 210
8.4.3 Site directed mutagenesis ..... 210
8.4.4 145 bp Widom 601 DNA sequence ..... 211
8.4.5 Large scale DNA production buffers ..... 211
**8.5 Protein production** ..... 212
8.5.1 Materials ..... 212
8.5.2 Transformations ..... 213
8.5.3 Starter cultures ..... 214
8.5.4 Glycerol stocks ..... 214
8.5.5 Competent cell production ..... 214
8.5.6 Large-scale histone growths ..... 214
8.5.7 Large-scale BRD4 growth ..... 215
**8.6 Purification of recombinant proteins** ..... 215
8.6.1 Histone purification ..... 215
8.6.2 BRD4 purification ..... 217
**8.7 Chemical incorporation of post-translational modification mimics** ..... 218
8.7.1 Method for producing acetylated lysine analogues ..... 219
8.7.2 Method for producing dimethylated lysine analogues ..... 219
8.7.3 Method for producing trimethylated lysine analogues ..... 220
**8.8 AlphaScreen** ..... 221
**8.9 Nucleosome production** ..... 221
8.9.1 Buffers ..... 221
**8.10 Crystallography** ..... 222
8.10.1 General crystallography materials and methods ..... 222
8.10.2 Crystallography data analysis ..... 223
8.10.3 Protein Data Bank accession codes ..... 223
**8.11 Mass spectroscopy techniques** ..... 224
8.11.1 Whole protein mass spectrometry ..... 224
8.11.2 Quadrupole-time of flight mass spectrometry ..... 224
8.11.3 Non-denaturing Q-TOF MS ..... 225
8.11.4 MALDI-TOF MS ..... 225
**8.12 Small-angle X-ray scattering** ..... 225
8.12.1 Data collection ..... 225
8.12.2 Data analysis ..... 226
**8.13 Isothermal titration calorimetry** ..... 226
**8.14 $K_i$ calculations** ..... 227
**8.15 References** ..... 229

# Part I: Investigating the Function and Selectivity of Transcription Factors to Peptide, Histone and Nucleosome Substrates



*View from a novel H3 sKI8Ac nucleosome crystal structure described in* ***Chapter 4****.*

# Chapter 1: Introduction

## 1.1 Epigenetics

In 1958 Francis Crick published his "Central Dogma" of molecular biology which states that "the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible".[1] The flow of biological information has since been found to be more complex. Diploid cells containing identical deoxyribonucleic acid (DNA) have the ability to differentiate into a wide variety of different cell types with distinct phenotypic traits. The human genome comprises 3 billion base pairs (bp) of DNA, of which less than 3% is currently considered to be directly coding for proteins.[2,3] Only a small proportion of those genes are transcribed at any time, with gene expression patterns varying in response to numerous factors, including nutrient availability, time of day, cell type and the stage of the cell's life cycle.[4–6]

The term "epigenetics" was first coined in 1942 by Conrad Waddington to describe the study of "causal interactions between genes and their products".[7] This concept has developed over time, and in 2009 a consensus definition of the epigenetic trait was described as "a stably heritable phenotype resulting from changes in a chromosome without alterations in the ATGC DNA sequence".[8]

Epigenetic control of gene expression is required for cellular development and function, whereas disruption of these epigenetic processes has been linked to a range of illnesses including developmental deficiencies, cancer and dementia.[9–11]

A wide range of processes are involved in regulating gene expression, although the mechanisms utilised varies between different domains of life.

## 1.2 Gene regulation throughout Life

### 1.2.1 Prokaryotes

Prokaryotes tend to have a small genome, ranging from 0.13 Mbp to 14 Mbp in length.[12,13] Prokaryotes store their DNA in a supercoiled bundle known as a nucleoid which comprises a compacted central core surrounded by ~ 100 lobes of DNA.[14] Multiple related genes are arranged together into operons which are subsequently expressed together. Each operon contains a regulatory sequence of DNA that includes a promotor region which can recruit RNA polymerase, initiating the transcription of DNA to RNA. Transcription factors, proteins capable of binding specific DNA sequences, can also bind to these regulatory sequences which can, in turn, promote or inhibit the cell's transcriptional machinery.[15]

### 1.2.2 Eukaryotes

It is proposed that the small genome sizes of prokaryotes allows for gene expression to be controlled predominantly by transcription factors; however, with genomes that are often many Gbp ($10^9$ base pairs) in length, eukaryotes have developed more complex regulatory mechanisms.[2] Eukaryotes employ a highly condensed protein–DNA complex known as chromatin to regulate gene expression. Many different processes are exploited in order to control eukaryotic gene expression, including the use of transcription factors, non-coding ribonucleic acids (RNAs), chromatin remodelling, and chemical modifications to the DNA and the chromatin architectural proteins known as histones. Several of these processes are discussed in more detail below.

### 1.2.3 Archaea

Archaea are single-celled organisms that share many similarities with bacteria, such as their size, shape, and appearance. Like bacteria they lack organelles, including the nucleus, and have a similar sized genome, generally several Mbp in size.[16] They are, however, genetically distinct from bacteria with many archaeal species able to thrive under extreme environments including high temperature, high salinity and strongly acidic conditions.[17–20] One significant divergence between archaea and bacteria comes from the fact that prokaryotes are lacking in histones. In contrast, homologues of the eukaryotic architectural proteins can be found in many families of archaea.[21] While archaeal nucleosomes can be post-translationally modified they are much smaller than their eukaryotic counterparts, lacking the amino (*N*)-terminal tail so heavily modified in eukaryotes.[22] The use of histones to stabilise the DNA of archaea may also contribute to the ability of some species to withstand extreme conditions.[23]

## 1.3 DNA storage in eukaryotes

### 1.3.1 Chromatin

*Homo sapiens* have around 2 m of DNA in every diploid cell.[24] With ~$4 \times 10^{13}$ cells in the body each human contains a length of DNA sufficient to circle the $2.8 \times 10^{13}$ m orbit of Neptune.[25,26] In order to store 2 m of DNA in the cell's nucleus, an organelle 6 μm in diameter, the genome must be highly compacted. By utilising a range of structural and regulatory proteins, DNA is compacted into a nucleoprotein fibre known as chromatin. Chromatin not only provides an efficient packing of DNA but also strengthens the chain, protects against damage to the sequence and allows for the regulation of gene expression. Two distinct forms of chromatin are distinguished; heterochromatin and euchromatin. Tightly packaged heterochromatin is generally inactive to transcription due to the

impeded access of gene promotors to the DNA; examples include telomeres and centromeres. Euchromatin is comparatively open, leaving exposed DNA more accessible to transcription. A schematic representation of chromatin composition is given in **Figure 1.1**.

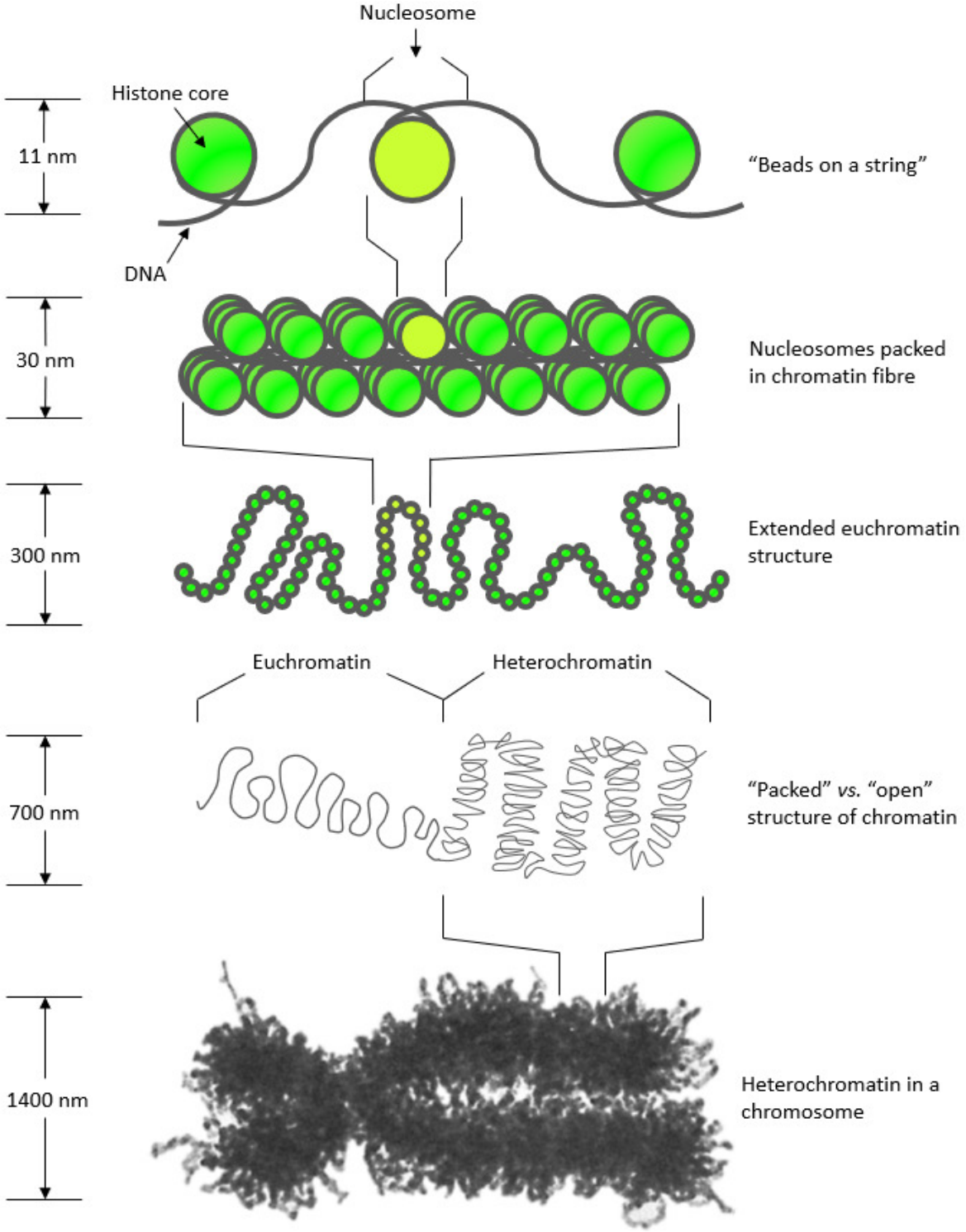



***Figure 1.1**: Schematic representation of the various levels of chromatin compaction.*

### 1.3.2 Nucleosomes

The basic repeating unit of chromatin is the nucleosome, which comprises a positively charged histone core particle wrapped in 145-147 bp of negatively charged DNA in ~1.65 left-handed superhelical turns.[27] Nucleosomes have a flat disk shape centred on a protein octamer normally assembled from two copies of the histones H2A, H2B, H3 and H4.[27] These nucleosomes repeat along the DNA, separated by a short section of linker DNA that is sometimes stabilised by the histone H1.[28] A crystal structure of a nucleosome viewed towards the histone face and towards the DNA is presented in **Figure 1.2**, revealing the flat disk shape of the nucleosome.

### 1.3.3 Histones

The histone core particle, around which DNA is spooled, is an octamer usually formed from an H3/H4 tetramer flanked by two H2A/H2B dimers.[27] Each histone contains a globular histone fold domain comprising three α-helices linked by two loops, and a long, flexible *N*-terminal tail.[27] Views from a crystal structure of a histone displaying the abovementioned structural features are given in **Figure 1.2**. The histone *N*-terminal tails protrude from the nucleosome core, beyond the DNA, where they are accessible to a range of transcription factors. Additionally, all four histone tails are required for efficient chromatin compaction.[29–31] The basic tails of H2A, H2B and H3 are able to shield some of the negative charge of the DNA through electrostatic interactions, resulting in tighter compaction.[30,31] Through its tail, H4 is able to bind to an acidic patch in the core of a neighbouring nucleosome stabilising the higher order structure of the chromatin.[27]

Humans also produce several homologues of H2A, H2B and H3, which contain a similar histone fold and can be incorporated into the nucleosome structure. These "non-canonical" histones can influence chromatin structure by altering the stability of the

nucleosome, restricting the post-translational modifications (PTMs) that can be placed on them, and altering the length of DNA that can wrap around them.[32] Unlike canonical histones, which are mainly expressed during synthesis (S)-phase, expression of these variants are not restricted to any particular stage of the cell's life cycle.[33]

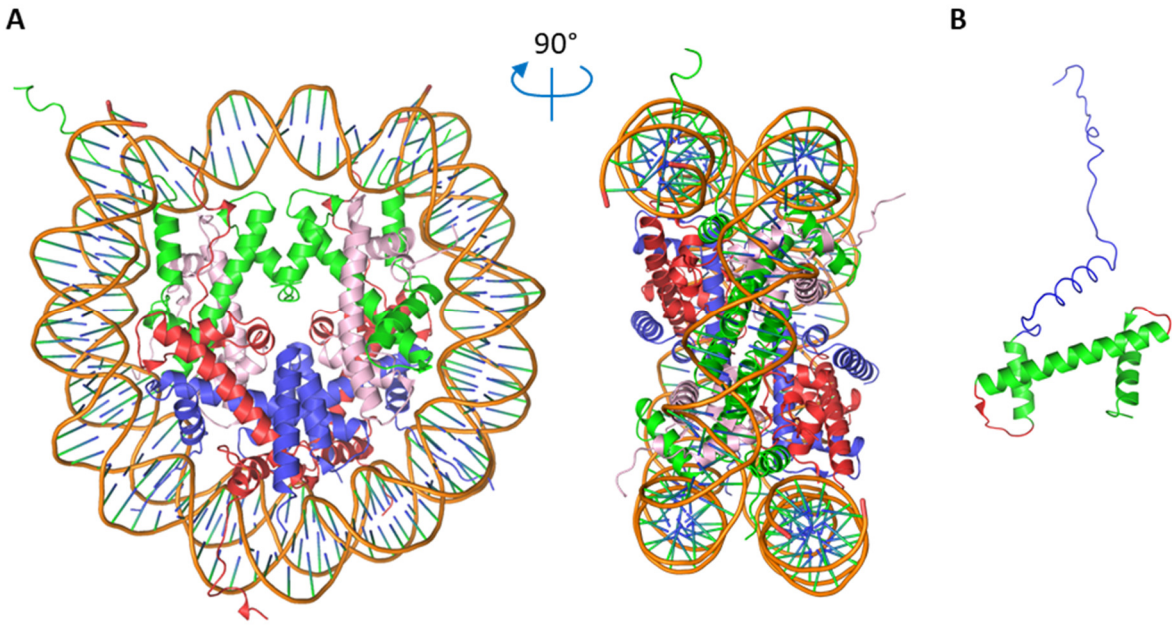



***Figure 1.2*** *Views from a crystal structure of a nucleosome (Protein Data Bank (PDB) ID: 1AOI).* ***A****- View facing the exposed histone core (****left****) and facing an edge of the DNA (****right****), highlighting the disk-like structure of the nucleosome.[27] DNA is shown as an orange helix encircling the core histones H2A (red), H2B (blue), H3 (green) and H4 (pink). Due to their flexibility the histone* N*-terminal tails are only partially observed in the crystal structure.* ***B****- Example histone (H3) extracted from 1AOI displaying the characteristic histone fold. α-Helices are displayed in green, loops in red and a portion of the* N*-terminal tail in blue.*

## 1.4 Mechanisms for controlling gene expression

In order to transcribe a gene, the cell's transcriptional machinery needs to be able to access the DNA. For this to occur, the packed chromatin structure needs to be disassembled, and subsequently reassembled after translation has occurred, thereby preventing excessive gene expression. The cell is able to regulate which regions of DNA

are exposed through use of histone variants, and enzymatic modification to the DNA and histones comprising chromatin.

### 1.4.1 DNA modification

Most eukaryotes are able to modify their DNA through methylation at the 5-position of cytosine.[34] The majority of vertebrate cell types incorporate this modification at sites where cytosine is followed directly by guanine in the 5′→3′ direction.[35] As these dinucleotides are separated by only a phosphate group, these positions are referred to as CpG sites. As CpG sites are symmetrically methylated in duplex DNA, methylation patterns can be passed onto the daughter strand during semiconservative replication. DNA methylation in eukaryotes is associated with long term gene silencing and has been linked to the formation of heterochromatic regions, such as on the inactive female X chromosome.[36–38]

DNA methylation can be carried out by DNA methyltransferases (DNMTs) through transfer of a methyl group from *S*-adenosyl methionine (SAM) to the 5-position of cytosine, producing *S*-adenosyl-homocysteine (SAH) as a by-product (see **Figure 1.3**). Methylation can also be caused by small molecule methylating agents such as methyl methanesulfonate.[39] While enzymatic methylation is directed and controlled, exogenous methylating molecules can act randomly with toxic effects. Unwanted DNA methylation marks can be removed by process including excision of the base, followed by base excision repair.[40]

In vertebrate DNA, CpG dinucleotides occur far less frequently than they statistically would be expected to. One explanation for this is the increased potential for methylated cytosine to mutate to thymine by spontaneous deamination. Thymine formed in this

manner is chemically identical to endogenous thymine and results in a base pair mismatch. During DNA mismatch repair the new thymine can either be replaced by a cytosine, or the pairing guanine may be replaced by an adenosine, causing a point mutation. While unmethylated cytosine can undergo a similar deamination, the product of the reaction is uracil, which can be easily identified by the cell and replaced (see **Figure 1.3**).

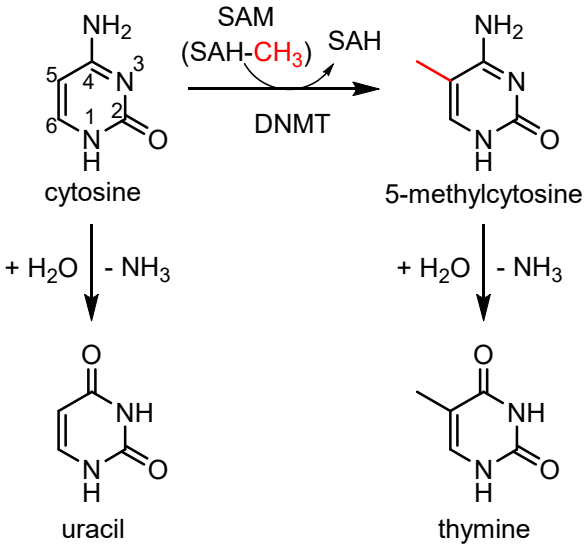



***Figure 1.3*** *Spontaneous deamination of cytosine and 5-methylcytosine produces uracil and thymine, respectively.*

As well as containing regions of low frequency, methylated CpG sites, vertebrate DNA also contains sections of unmethylated DNA referred to as CpG islands.[41] These islands are generally enriched in CG content, with higher incidences of CpG sites than observed in the rest of the genome. CpG islands are associated with gene activation and are often found in the promotors of mammalian genes.[42,43] The silencing of tumour suppressor genes through aberrant DNA methylation within the gene's promoters has been identified in a range of human cancers.[10]

CpG methylation is the most studied DNA modification in vertebrates, but other DNA modifications are possible. In addition to methylation at the 5-position on cytosine, hydroxymethylation, formylation and carboxylation at the same position have been detected in humans.[44–46] The conversion of 5-methylcytosine to hydroxymethyl-, formyl- and carboxycytosine is facilitated by the ten-eleven translocation (TET) enzymes.[46] TET enzymes have also been found to oxidise thymine bases at the 5-position, forming 5-hydroxymethyluracil.[47]

Methylation at the $N^6$-position of adenosine has been detected in both DNA and RNA, and is thought to be transcriptionally activating in prokaryotes.[48,49] The placement and removal of adenosine methylation marks are thought to be catalysed by DNA $N^6$-adenine methyltransferases (DAMTs) and $N^6$-methyladenine demethylases (NMADs), respectively.[50]

Structures of the abovementioned modifications to DNA bases in eukaryotes are shown in **Figure 1.4**.

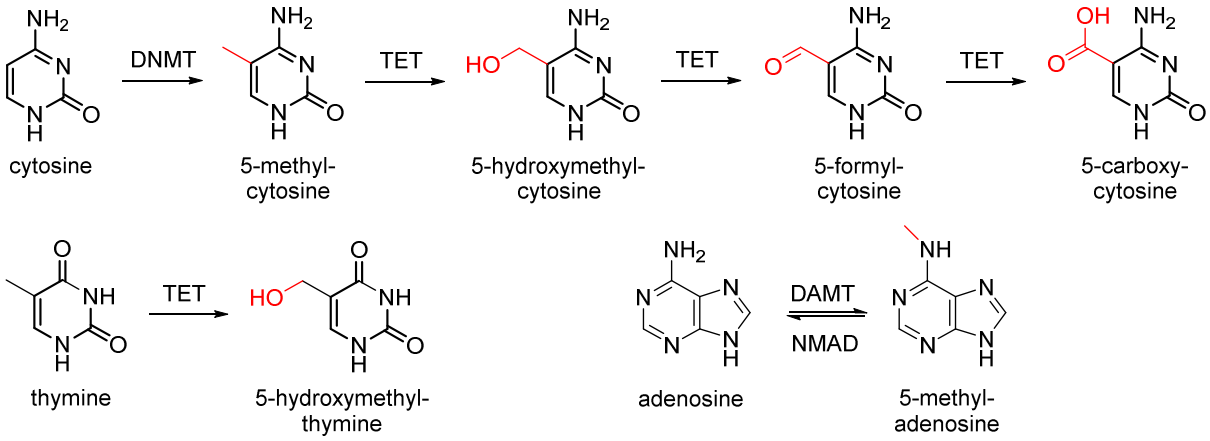



***Figure 1.4**: Scheme showing DNA modifications that have been identified as having epigenetic regulatory functions and some of the pathways that interconvert them.*[51]

### 1.4.2 Post-translational modifications of histones

Each of the canonical histones can undergo a range of chemical modifications at multiple different residues and positions, particularly on their *N*-terminal tails. A summary of identified modifications of histones and the types of residues on which they can occur can be found in **Table 1.1**.

***Table 1.1**: Showing known modifications of histone proteins and the residues on which they have been detected.*

| Residue | Post-translational modification |
|---|---|
| **Arginine** | Methylation (symmetric[52], asymmetric[53]), citrullination[54], ADP ribosylation[55] |
| **Glutamate** | ADP ribosylation[56] |
| **Glutamine** | Methylation[57] |
| **Histidine** | Phosphorylation[58] |
| **Lysine** | Methylation (mono-, di-, tri-)[59], acetylation[60], biotinylation[61], ubiquitination[62], formylation[63], crotonylation[64], sumoylation[65] |
| **Proline** | Isomerisation[66] |
| **Serine** | Phosphorylation[67], GlcNAcylation[68] |
| **Tyrosine** | Phosphorylation[69] |
| **Threonine** | Phosphorylation[67], GlcNAcylation[68] |

The 'histone code' hypothesis postulates that certain combinations of modifications at specific histone residues are able to facilitate activation or deactivation of the chromatin to transcription.[70] These modifications can invoke structural changes by recruiting transcription factors, enzymes capable modifying the structure of chromatin, that contain domains specialised to 'read' specific combinations of modifications.[71] Histone PTMs can also stabilise, or destabilise, the nucleosome directly such as through charge neutralisation or on addition of bulky groups to the side-chain.[72,73] Within a single nucleosome, combinations of marks can be placed on the same (*cis*), or different (*trans*) histones, with different combinations capable of producing different outcomes.[70] PTMs can influence the placement of subsequent marks by recruiting, or inhibiting

transcription factors, or by blocking the placement of an alternative modification by occupying the same amino acid residue. Enzymes capable of installing these marks are referred to as "writers", those that remove them, "erasers", and those that can detect the marks, "readers".

The work detailed in **Part I** of this thesis focusses on histone lysine methylation and acetylation marks, and the proteins that interact with them.

#### 1.4.2.1 Lysine methylation

Histone methylation is one of the best studied epigenetic modifications and has been found to occur on both lysine and arginine residues. Lysine residues can harbour one, two or three methyl groups, all of which are placed at the $N^{\varepsilon}$-position (see **Figure 1.5**). Lysine methylation is highly site specific, with the downstream effects dependent on the methylation state, and position. Methylation on lysine 4 on H3 (H3 K4) has been associated with facilitating centromere formation in its dimethylated state (H3 K4Me$_2$), whereas H3 K4Me$_3$ is thought to induce transcription *via* recruitment of chromatin modifying proteins.[74,75] Additionally, while H3 K4Me$_3$ is considered activating towards transcription, the same methylation state on H3 K9 is associated with gene repression through the recruitment of heterochromatic protein 1 (HP1) and the subsequent expansion of heterochromatic regions.[76]

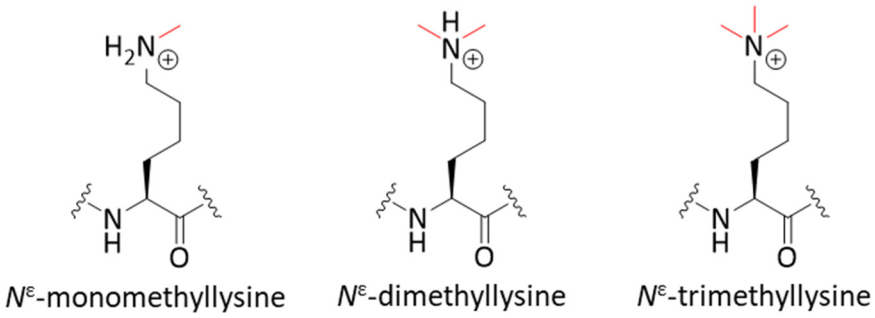



***Figure 1.5**: Displaying the possible methylation states of lysine residues.*

**Writers**

Lysine methylation is catalysed by the lysine methyltransferases (KMTs) *via* $S_N2$ nucleophilic substitution using SAM as the methyl donor.[77,78] Most KMTs facilitate lysine methylation through their su(var)3-9, enhancer-of-zeste, trithorax (SET) domain. The number of methyl groups transferred, the site of the lysine, and the required combination of PTMs already in proximity of the residue differ between KMTs. All four canonical histones have been found to undergo lysine methylation, predominantly on their *N*-terminal tails.[79] Additionally, lysine 79 on H3 (H3 K79) which is located in the histone core can be methylated by disruptor of telomeric silencing 1-like (DOT1L).[80] DOT1L holds the distinction of being the only identified KMT lacking a SET domain, and, unlike most other KMTs, cannot act on free histone, instead only methylating H3 when it is incorporated into a nucleosome.[81]

**Erasers**

As with writers, lysine methyl erasers are also site and methylation-state specific. Many lysine demethylases (KDMs) possess multiple domains capable of binding epigenetic modifications on histones or DNA. Through these interactions the different KDMs are able to regulate their activity. The removal of lysine methylation marks can be catalysed

by two families of KDM identified so far; the 2-oxoglutarate (2OG)-dependent dioxygenases and the flavin adenine dinucleotide (FAD)-dependent demethylases.

The 2OG demethylases are part of a large superfamily of Fe(II) and 2OG-dependent hydroxylases. These hydroxylases have been identified in almost all organisms and have been found to catalyse a wide range of oxidative processes including epimerisation, halogenation, desaturation, C-C bond cleavage, ring expansion and ring closure, as well as hydroxylation and demethylation.[82–87] Despite their varied functions, the 2OG oxygenases operate through a largely conserved mechanism. 2OG oxygenases contain a conserved double-stranded β-helix fold with which the enzyme can bind the required substrates and cofactors.[88,89] The active site contains an Fe(II) ion, generally coordinated with an acidic amino acid, two histidine residues, and three waters.[89] 2OG can displace two water molecules to bind the Fe(II) species. Fe(II) is converted into an Fe(IV)-oxo intermediate through the activation of molecular oxygen and the concomitant conversion of 2OG into succinate and $CO_2$.[90] This reactive Fe(IV)-oxo species is then poised to cleave the substrate C-H bond, forming a carbon radical and an Fe(III)-hydroxide. These high energy species subsequently collapse into the hydroxylated product and a succinate-bound Fe(II) species. Finally, succinate can be replaced by a water molecule to regenerate the starting Fe(II) species. **Figure 1.6** displays a scheme of the consensus mechanism of oxidation by the 2OG oxygenases.

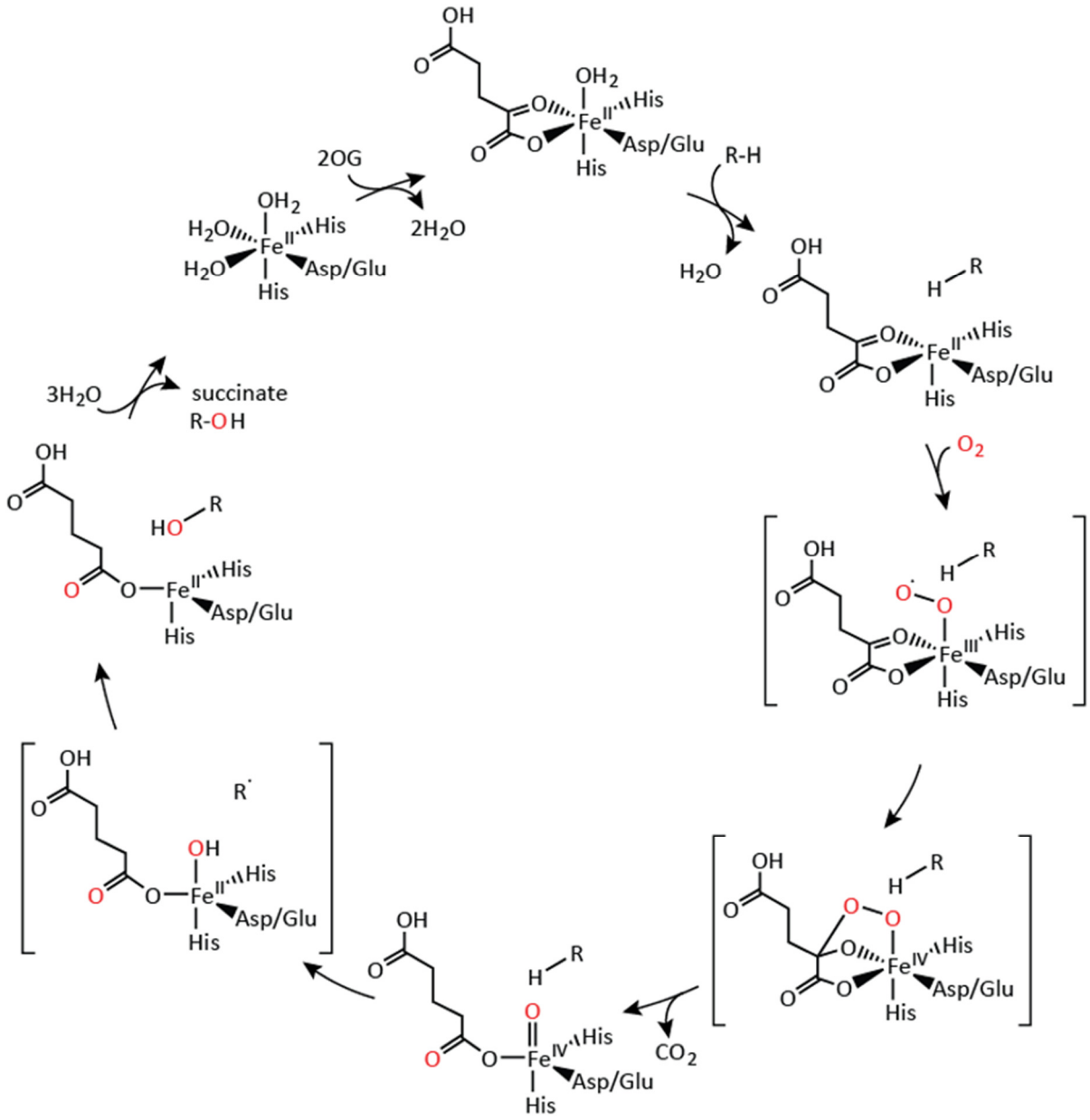



***Figure 1.6**: Scheme displaying the consensus mechanism of oxidation catalysed by 2OG oxygenases. R-H and R-OH represent the substrate and the oxidised product, respectively.*[91]

In the case of the Jumonji-*C* domain containing (JmjC) KDMs, lysine *N*$^{\varepsilon}$-methyl groups can be oxidised to a hemiaminal group by the previously described mechanism, which can spontaneously fragment into formaldehyde and the demethylated lysine.[92] Through this process 2OG KDMs are capable of catalysing demethylation for all three methylation states of lysine (see **Figure 1.7**).

The second family of KDM comprise only two known members, lysine-specific demethylase 1 (LSD1) and LSD2, and catalyse lysine demethylation *via* oxidation of FAD.[93,94] FAD KDMs demethylate lysine residues through conversion of the methyl group to an imine, which can react with water to form a hemiaminal. As with the mechanism of JmjC-mediated demethylation, these hemiaminals are able to fragment into the demethylated product and formaldehyde (see **Figure 1.7**).[95] Due to the requirement for the formation of an imine group during lysine demethylation, FAD-dependent KDMs are only able to demethylate mono- and di-, but not trimethylated, lysine residues.

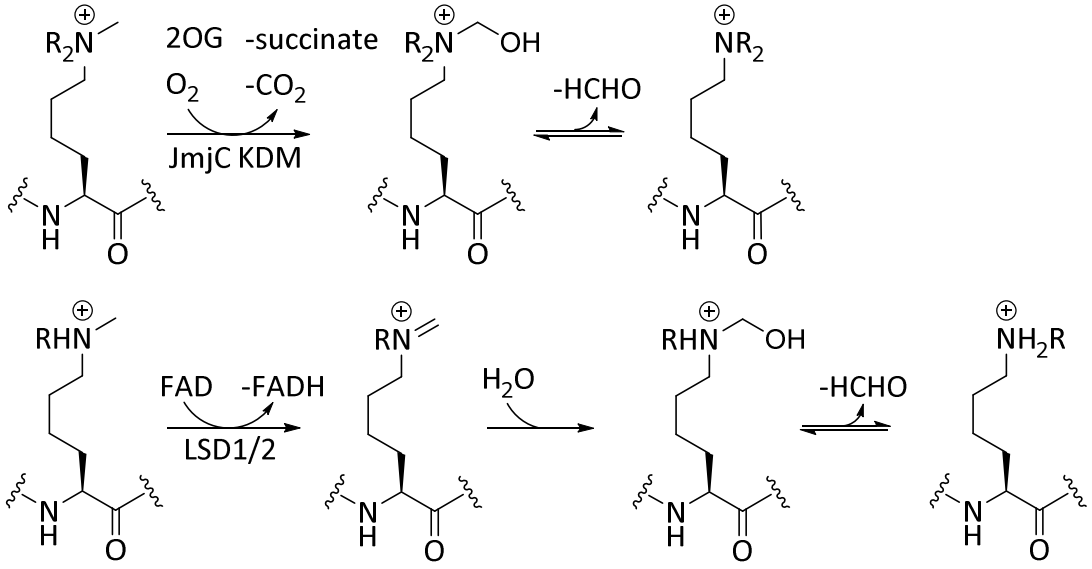



***Figure 1.7**: Scheme showing KDM catalysed demethylation of lysine residues by JmjC (**top**) and LSD1/2 (**bottom**). Due to the imine formed in the FAD-dependent demethylation pathway they are unable to demethylate quaternary amines.*

**Readers**

Lysine methylation marks can be 'read' by a wide range of different domains including plant homeodomains (PHDs), tudor domains, chromodomains, ankyrin repeats, Pro-Trp-Trp-Pro (PWWP) domains and Trp-Asp (WD) repeats.[96] Reader domains can selectively bind un-, mono-, di- and trimethylated lysine, making it useful to consider unmodified lysine as a legitimate methylation state. Proteins may contain multiple reader domains, allowing them to bind to multiple methylation marks simultaneously. Some proteins can even detect multiple methylation states with a single reader domain; tripartite motif-containing 33 (TRIM33) preferentially binds $K4Me_0$ and $K9Me_3$ through the same PHD domain.[97]

The binding pockets of most methylated lysine reader proteins tend to consist of a relatively rigid aromatic cage, the size of which depends on the methylation state of the binding partner. Binders for un-, mono-, and occasionally dimethylated lysine tend to accommodate the group in a small cavity, whereas tri- and sometimes dimethylated lysine residues are generally accommodated in less restrictive surface grooves. **Figure 1.8** displays an example binding site for each of the methylation states of lysine.

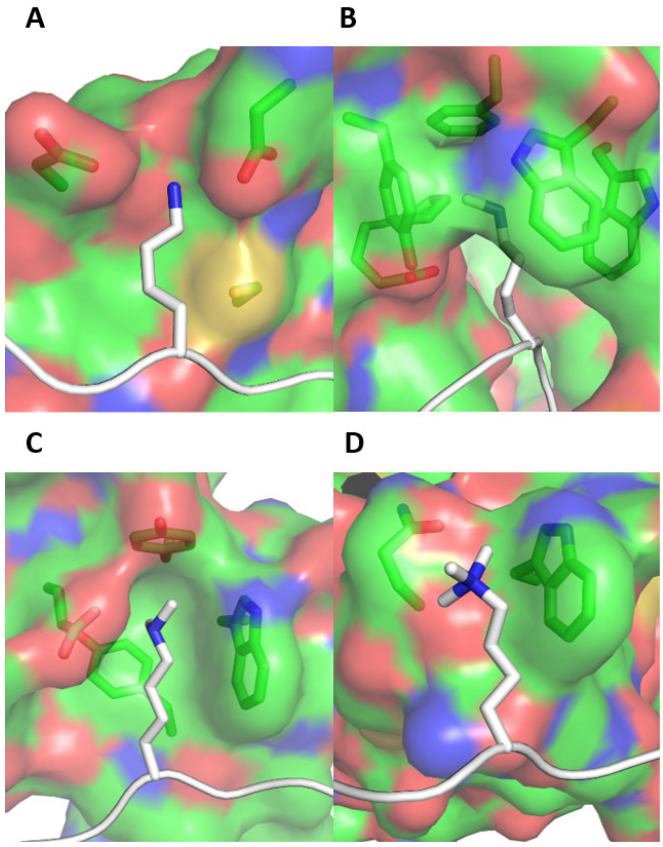



***Figure 1.8**: View from crystal structures of the lysine binding pockets of three reader proteins. Peptides are shown in white and reader proteins are shown in green. **A**- H3 K4[Me$_0$] bound in a negatively charged pocket in the PHD domain of TRIM33 (PDB ID: 3U5O).[97] **B**- H4 K20Me$_1$ surrounded by a full aromatic cage in the chromodomain of male specific lethal 3 (MSL3) (PDB ID: 3OA6).[98] **C**- H3 K4Me$_2$ embedded into an open aromatic cage in the PHD domain of bromodomain PHD finger transcription factor (BPTF) (PDB ID: 2RI7).[99] **D**- H3 K9Me$_3$ bound in an exposed surface groove in the PHD domain of TRIM33.[97] **A** and **D** are different views from the same TRIM33$_{881-1087}$–H3$_{1-22}$ K9Me$_3$, K14Ac, K18Ac peptide co-crystal structure.*

#### 1.4.2.2 Lysine acetylation

Acetylation, one of the most frequently occurring histone PTMs, occurs on lysine residues at the $N^{\varepsilon}$-position, and is associated with regions of euchromatin.[100,101] This may, in part, be due to the neutralisation of the lysine positive charge that arises on acetylation, reducing the histone tail's affinity to the negatively charged DNA, resulting in a less tightly bound histone–DNA complex. The formation of chromatin fibres can even be inhibited by the acetylation of a single residue (H4 K16Ac), which impacts on the histone's ability to form inter-nucleosome contacts.[102] As with methylation marks, lysine acetylation can also cause chromatin remodelling and gene transcription through the recruitment of appropriate transcription factors.[103] Along with the electronic effect of a change in charge and the recruitment of reader proteins, the presence alone of an acetylation mark can disrupt other epigenetic pathways. H3 K9 acetylation is considered transcriptionally activating, possibly because it prevents the placement of heterochromatin-inducing H3 K9 methylation marks.[104]

**Writers**

Histone acetyltransferases (HATs) are responsible for histone acetylation through facilitation of the transfer of an acetyl group from acetyl-CoA to the histone lysine $N^{\varepsilon}$-position (see **Figure 1.9**). In addition to their catalytic domains, HATs also contain reader domains capable of binding specific histone marks. Guided by these reader domains, HATs that target similar residues are able to perform different functions. For example Spt-Ada-Gcn5 acetyltransferase (SAGA) and Elongator complex both acetylate H3 K9 and H3 K14 residues, but act in promotor and coding regions, respectively.[105,106]

### Erasers

Acetylation marks on histones are generally highly dynamic, a feature considered important for temporal control of gene expression, although some long lived acetylation marks have been discovered.[107] The removal of acetylation marks is catalysed by histone deacetylases (HDACs) (see **Figure 1.9**). HDACs activate a water molecule for nucleophilic attack on the acetyl group carbonyl. The amino-diol then breaks down into acetic acid and lysine.[108] Regions of heterochromatin tend to be hypoacetylated, and many transcriptional co-repressor complexes contain HDAC subunits.[109]

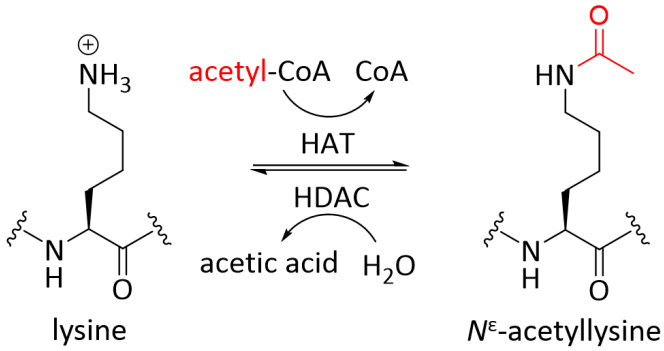



***Figure 1.9**: Scheme showing the interconversion of lysine and* N$^{\varepsilon}$*-acetyllysine (KAc).*

### Readers

Acetylated lysine marks are read by small (~110 amino acids) reader proteins known as bromodomains (BRDs). Despite low sequence homology, all BRDs contain a conserved BRD fold comprising a left-handed four-helix bundle joined by three variable linkers.[110] Two of these inter-helical loops form a hydrophobic pocket capable of binding acetylated lysine (see **Figure 1.10**). While the overall fold is conserved, the electrostatic surface properties of the area surrounding the KAc binding site differ dramatically, resulting in the sequence specificity of the BRDs.[110]

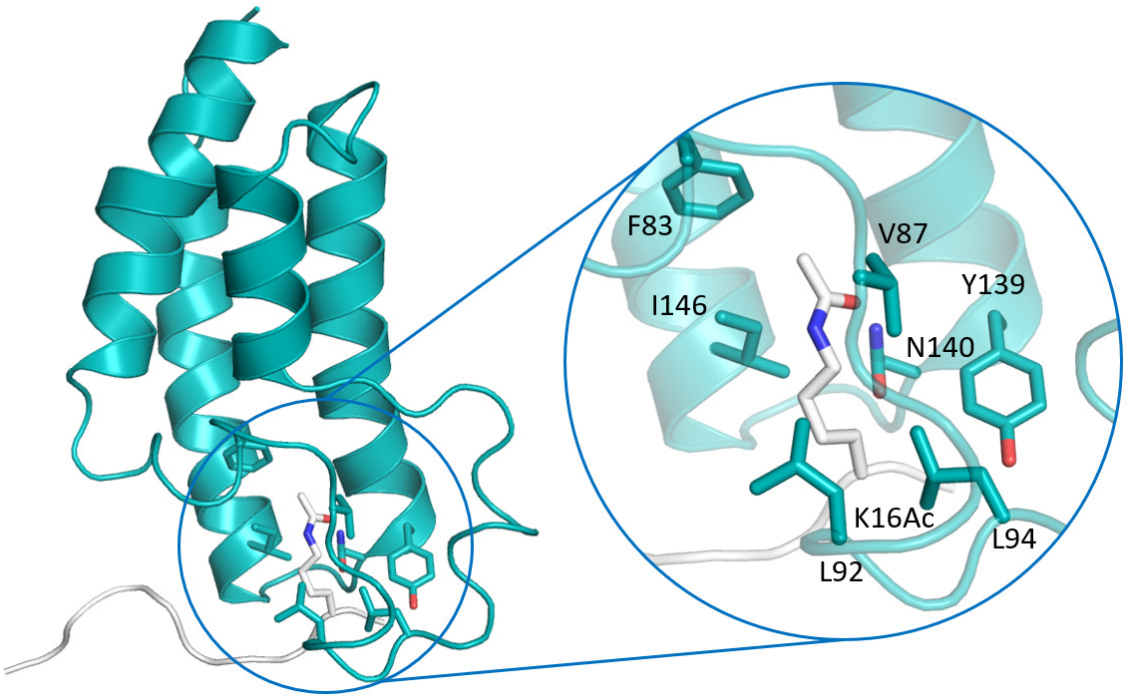



***Figure 1.10**: View from a crystal structure of bromodomain 4 (BRD4) bound to $H4_{16-26}$ K16Ac peptide (PDB ID: 3UVY).[110] Figure shows the first bromodomain of the reader protein BRD4 (teal), displaying the characteristic BRD fold comprising four α-helices connected by linker regions. The magnified insert shows the interactions between the H4 K16Ac residue (white) and the hydrophobic BRD binding site.*

BRDs are often contained in proteins alongside other reader domains, permitting a high degree of target specificity; for example TRIM33 contains both a PHD and BRD, through which it can target three PTM marks simultaneously (unmodified H3 K4, H3 $K9Me_3$, H3 K18Ac).[97] BRDs have been found in several different families of nuclear proteins such as methyltransferases, transcriptional coactivators, chromatin remodelling complexes and almost all HATs.[111–115]

### 1.4.3 Epigenetic drug targets

PTMs play a fundamental role in a wide range of cellular processes, and disruptions of these mechanisms are associated with numerous pathologies. Mutations to the KDM PHD finger protein 8 (PHF8) are associated with X-linked mental retardation, and knockout models of many PTM writers and erasers result in embryonic death.[116–119] Errant

levels of histone acetylation has been linked to certain forms of cancer, autoimmune diseases such as multiple sclerosis, and several neurological diseases.[120–123]

The histone acetylation pathway has therefore become an attractive therapeutic target, with a number of HDAC inhibitors currently approved for treatment of T-cell lymphoma.[124,125] The discovery of JQ1, a small molecule inhibitor of the BRD and extra terminal (BET) family highlighted reader proteins as viable targets for drug discovery, while providing a scaffold for the development of chemical probes.[126]

## 1.5 Aims of Part I

The work described in **Part I** of this thesis (**Chapters 1-5**) investigated the utility and limitations of histone fragment-based studies as simplified models for higher order proteins and complexes in the study of transcriptional regulation. Specific aims were to:

- Use a fragment-based affinity assay to aid in the design of a small molecule inhibitor capable of inducing a hypoxic-like response while simultaneously inhibiting BRD4 (**Chapter 2**).
- Investigate the effect of the putative PMT lysine hydroxyacetylation on BRD binding using the same fragment-based affinity assay (**Chapter 3**).
- Probe the interaction between TRIM24 and a recombinant, homogeneously post-translationally modified mononucleosome using a range of structural and biophysical techniques (**Chapter 4**).
- Develop an antibody-based system capable of monitoring nucleosome methylation states, and use said probe to investigate the importance of substrate complexity on the demethylase activity of KDM4A (**Chapter 5**).

## 1.6 References


1. Crick, F. Central dogma of molecular biology. *Nature* **227,** 561–3 (1970).
2. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–51 (2001).
3. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
4. Oehler, S., Eismann, E. R., Krämer, H. & Müller-Hill, B. The three operators of the lac operon cooperate in repression. *EMBO J.* **9,** 973–979 (1990).
5. Sahar, S. & Sassone-Corsi, P. The epigenetic language of circadian clocks. In *Circadian Clocks.* 29–44 (Springer Berlin Heidelberg, 2013).
6. Arbeitman, M. N. *et al.* Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297,** 2270–2275 (2002).
7. Waddington, C. H. The epigenotype. *Int. J. Epidemiol.* **41,** 10–13 (2012).
8. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **23,** 781–783 (2009).
9. Champagne, F. A. Epigenetics and developmental plasticity across species. *Dev. Psychobiol.* **55,** 33–41 (2013).
10. Eads, C. A. *et al.* Epigenetic patterns in the progression of esophageal adenocarcinoma. *Cancer Res.* **61,** 3410–3418 (2001).
11. Veerappan, C. S., Sleiman, S. & Coppola, G. Epigenetics of Alzheimer's disease and frontotemporal dementia. *Neurotherapeutics* **10,** 709–21 (2013).
12. McCutcheon, J. P. & Von Dohlen, C. D. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr. Biol.* **21,** 1366–1372 (2011).
13. Han, K. *et al.* Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. *Sci. Rep.* **3,** 2101 (2013).
14. Kavenoff, R. & Bowen, B. C. Electron microscopy of membrane-free folded chromosomes from *Escherichia coli*. *Chromosoma* **59,** 89–101 (1976).
15. Eisenberg, E. *et al.* Human housekeeping genes, revisited. *Trends Genet.* **29,** 569–74 (2013).
16. Koonin, E. V & Wolf, Y. I. Genomics of bacteria and archaea: The emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36,** 6688–6719 (2008).
17. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5088–5090 (1977).
18. Kashefi, K. & Lovley, D. R. Extending the upper temperature limit for life. *Science* **301,** 934 (2003).
19. Rodriguez-Valera, F., Ruiz-Berraquero, F. & Ramos-Cormenzana, A. Behaviour of mixed populations of halophilic bacteria in continuous cultures. *Can. J. Microbiol.*

**26,** 1259–63 (1980).

20. Macalady, J. L. *et al.* Tetraether-linked membrane monolayers in *Ferroplasma spp*: A key to survival in acid. *Extremophiles* **8,** 411–419 (2004).

21. Sandman, K. & Reeve, J. N. Archaeal histones and the origin of the histone fold. *Curr. Opin. Microbiol.* **9,** 520–525 (2006).

22. Fahrner, R. L., Cascio, D., Lake, J. A. & Slesarev, A. An ancestral nuclear protein assembly: Crystal structure of the *Methanopyrus kandleri* histone. *Protein Sci.* **10,** 2002–7 (2001).

23. Li, T., Sun, F., Ji, X., Feng, Y. & Rao, Z. Structure based hyperthermostability of archaeal histone HPhA from *Pyrococcus horikoshii*. *J. Mol. Biol.* **325,** 1031–7 (2003).

24. Alberts, B. *et al. Chromosomal DNA and its packaging in the chromatin fiber. Molecular Biology of the Cell. 4th edition* (Garland Science, 2002).

25. Bianconi, E. *et al.* An estimation of the number of cells in the human body. *Ann. Hum. Biol.* **40,** 463–471 (2013).

26. Williams, D. R. Neptune fact sheet. *NASA* 1–3 (2016). Available at: http://nssdc.gsfc.nasa.gov/planetary/factsheet/neptunefact.html. (Accessed: 11th August 2016).

27. 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,** 251–60 (1997).

28. Thoma, F. & Koller, T. Influence of histone H1 on chromatin structure. *Cell* **12,** 101–107 (1977).

29. Fletcher, T. M. & Hansen, J. C. Core histone tail domains mediate oligonucleosome folding and nucleosomal DNA organization through distinct molecular mechanisms. *J. Biol. Chem.* **270,** 25359–25362 (1995).

30. Tse, C. & Hansen, J. C. Hybrid trypsinized nucleosomal arrays: Identification of multiple functional roles of the H2A/H2B and H3/H4 *N*-termini in chromatin fiber compaction. *Biochemistry* **36,** 11381–11388 (1997).

31. Kan, P. Y., Lu, X., Hansen, J. C. & Hayes, J. J. The H3 tail domain participates in multiple interactions during folding and self-association of nucleosome arrays. *Mol Cell Biol* **27,** 2084–2091 (2007).

32. Talbert, P. B. & Henikoff, S. Histone variants — ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* **11,** 264–275 (2010).

33. Kamakaka, R. T. & Biggins, S. Histone variants: Deviants? *Genes and Development* **19,** 295–310 (2005).

34. Bird, A. The essentials of DNA methylation. *Cell* **70,** 5–8 (1992).

35. Jabbari, K. & Bernardi, G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* **333,** 143–149 (2004).

36. Bird, A. P. Gene expression: DNA methylation — how important in gene control? *Nature* **307,** 503–504 (1984).

37. Razin, A. & Cedar, H. DNA methylation and gene expression. *Microbiol. Rev.* **55,** 451–8 (1991).

38. Rose, N. R. & Klose, R. J. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim. Biophys. Acta* **1839,** 1362–1372 (2014).

39. Sega, G. A. & Owens, J. G. Methylation of DNA and protamine by methyl methanesulfonate in the germ cells of male mice. *Mutat. Res.* **111,** 227–244 (1983).

40. Bellacosa, A. & Drohat, A. C. Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. *DNA Repair (Amst).* **32,** 33–42 (2015).

41. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25,** 1010–1022 (2011).

42. Fatemi, M. *et al.* Footprinting of mammalian promoters: Use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. *Nucleic Acids Res.* **33,** e176 (2005).

43. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 1412–1417 (2006).

44. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324,** 929–930 (2009).

45. Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed.* **50,** 7008–7012 (2011).

46. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–3 (2011).

47. Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10,** 574–81 (2014).

48. Koziol, M. J. *et al.* Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* **23,** 24–30 (2016).

49. Fu, Y. *et al.* $N^6$-Methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell* **161,** 879–892 (2015).

50. Greer, E. L. *et al.* DNA methylation on $N^6$-adenine in *C. elegans*. *Cell* **161,** 868–878 (2015).

51. Breiling, A. & Lyko, F. Epigenetic regulatory functions of DNA modifications: 5-Methylcytosine and beyond. *Epigenetics Chromatin* **8,** 24 (2015).

52. Pal, S., Vishwanath, S. N., Erdjument-Bromage, H., Tempst, P. & Sif, S. Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes. *Mol. Cell. Biol.* **24,**

9630–45 (2004).

53. Lin, W. J., Gary, J. D., Yang, M. C., Clarke, S. & Herschman, H. R. The mammalian immediate-early TIS21 protein and the leukemia-associated BTG1 protein interact with a protein-arginine *N*-methyltransferase. *J. Biol. Chem.* **271,** 15034–15044 (1996).

54. Christophorou, M. A. *et al.* Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* **507,** 104–8 (2014).

55. Laing, S., Unger, M., Koch-Nolte, F. & Haag, F. ADP-ribosylation of arginine. *Amino Acids* **41,** 257–269 (2011).

56. Ogata, N., Ueda, K. & Hayaishi, O. ADP-ribosylation of histone H2B. *J. Biol. Chem* **255,** 7610–7615 (1980).

57. Tessarz, P. *et al.* Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* **505,** 564–8 (2014).

58. Chen, C. C. *et al.* Phosphorylation of nuclear proteins in rat regenerating liver. *Biochemistry* **16,** 4852–4855 (1977).

59. Guo, H. B. & Guo, H. Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 8797–8802 (2007).

60. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389,** 349–352 (1997).

61. Kuroishi, T., Rios-Avila, L., Pestinger, V., Wijeratne, S. S. K. & Zempleni, J. Biotinylation is a natural, albeit rare, modification of human histones. *Mol. Genet. Metab.* **104,** 537–545 (2011).

62. Weake, V. M. & Workman, J. L. Histone ubiquitination: Triggering gene activity. *Molecular Cell* **29,** 653–663 (2008).

63. Jiang, T., Zhou, X., Taghizadeh, K., Dong, M. & Dedon, P. C. *N*-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 60–65 (2007).

64. Tan, M. *et al.* Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146,** 1016–1028 (2011).

65. Nathan, D. *et al.* Histone sumoylation is a negative regulator in *Saccharomyces cerevisiae* and shows dynamic interplay with positive-acting histone modifications. *Genes Dev.* **20,** 966–976 (2006).

66. Nelson, C. J., Santos-Rosa, H. & Kouzarides, T. Proline isomerization of histone H3 regulates lysine methylation and gene expression. *Cell* **126,** 905–916 (2006).

67. Rossetto, D., Avvakumov, N. & Côté, J. Histone phosphorylation: A chromatin modification involved in diverse nuclear events. *Epigenetics* **7,** 1098–1108 (2012).

68. Sakabe, K., Wang, Z. & Hart, G. W. Beta-*N*-acetylglucosamine (O-GlcNAc) is part

of the histone code. *Proc. Natl. Acad. Sci.* **107,** 19915–19920 (2010).

69. Xiao, A. *et al.* WSTF regulates the H2A.X DNA damage response *via* a novel tyrosine kinase activity. *Nature* **457,** 57–62 (2009).

70. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403,** 41–45 (2000).

71. Turner, B. M. Histone acetylation and an epigenetic code. *BioEssays* **22,** 836–845 (2000).

72. Krajewski, W. A., Panin, V. M. & Razin, S. V. Acetylation of core histones causes the unfolding of 30 nm chromatin fiber: Analysis by agarose gel electrophoresis. *Biochem Biophys Res Commun* **196,** 455–460 (1993).

73. Lercher, L. *et al.* Generation of a synthetic GlcNAcylated nucleosome reveals regulation of stability by H2A-Thr101 GlcNAcylation. *Nat. Commun.* **6,** 7978 (2015).

74. Bergmann, J. H. *et al.* Epigenetic engineering shows H3K4Me$_2$ is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. *EMBO J.* **30,** 328–340 (2011).

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

76. Stunnenberg, R. *et al.* H3K9 methylation extends across natural boundaries of heterochromatin in the absence of an HP1 protein. *EMBO J.* **34,** 2789–2803 (2015).

77. Le, D. D. & Fujimori, D. G. Protein and nucleic acid methylating enzymes: Mechanisms and regulation. *Curr. Opin. Chem. Biol.* **16,** 507–515 (2012).

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

79. Huang, H., Sabari, B. R., Garcia, B. A., David Allis, C. & Zhao, Y. SnapShot: Histone modifications. *Cell* **159,** 458–458 (2014).

80. Van Leeuwen, F., Gafken, P. R. & Gottschling, D. E. Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* **109,** 745–756 (2002).

81. Feng, Q. *et al.* Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **12,** 1052–1058 (2002).

82. Clifton, I. J. *et al.* Crystal structure of carbapenem synthase (CarC). *J. Biol. Chem.* **278,** 20843–20850 (2003).

83. Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T. & Drennan, C. L. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **440,** 368–371 (2006).

84. Flashman, E. & Schofield, C. J. The most versatile of all reactive intermediates? *Nat. Chem. Biol.* **3,** 86–87 (2007).

85. Baldwin, J. E. *et al.* Purification and initial characterization of an enzyme with

deacetoxycephalosporin C synthetase and hydroxylase activities. *Biochem. J.* **245,** 831–41 (1987).

86. Zhang, Z. *et al.* Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat. Struct. Biol.* **7,** 127–133 (2000).

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

88. McDonough, M. A. *et al.* Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc. Natl. Acad. Sci. U. S. A.* **103,** 9814–9 (2006).

89. Valegård, K. *et al.* Structure of a cephalosporin synthase. *Nature* **394,** 805–9 (1998).

90. Hoffart, L. M., Barr, E. W., Guyer, R. B., Bollinger, J. M. & Krebs, C. Direct spectroscopic detection of a C-H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 14738–14743 (2006).

91. Bollinger, J. M., Price, J. C., Hoffart, L. M., Barr, E. W. & Krebs, C. Mechanism of taurine: α-Ketoglutarate dioxygenase (TauD) from *Escherichia coli*. *Eur. J. Inorg. Chem.* **2005,** 4245–4254 (2005).

92. Yi, C. *et al.* Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature* **468,** 330–3 (2010).

93. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119,** 941–953 (2004).

94. Fang, R. *et al.* Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4$Me_2$ methylation. *Mol. Cell* **39,** 222–233 (2010).

95. Scrutton, N. S. *et al.* Chemical aspects of amine oxidation by flavoprotein enzymes. *Nat. Prod. Rep.* **21,** 722 (2004).

96. Yun, M., Wu, J., Workman, J. L. & Li, B. Readers of histone modifications. *Cell Res.* **21,** 564–78 (2011).

97. Xi, Q. *et al.* A poised chromatin platform for TGF-β access to master regulators. *Cell* **147,** 1511–24 (2011).

98. Kim, D. *et al.* Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. *Nat. Struct. Mol. Biol.* **17,** 1027–9 (2010).

99. Li, H. *et al.* Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. *Mol. Cell* **28,** 677–691 (2007).

100. Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325,** 834–40 (2009).

101. Hebbes, T. R., Thorne, A. W. & Crane-Robinson, C. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.* **7,** 1395–1402

(1988).

102. Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311,** 844–7 (2006).

103. Struhl, K. Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* **12,** 599–606 (1998).

104. Kim, J. H. & Workman, J. L. Histone acetylation in heterochromatin assembly. *Genes Dev.* **24,** 738–740 (2010).

105. Koutelou, E., Hirsch, C. L. & Dent, S. Y. R. Multiple faces of the SAGA complex. *Curr. Opin. Cell Biol.* **22,** 374–382 (2010).

106. Wittschieben, B. *et al.* A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Mol. Cell* **4,** 123–128 (1999).

107. Zheng, Y., Thomas, P. M. & Kelleher, N. L. Measurement of acetylation turnover at distinct lysines in human histones identifies long-lived acetylation sites. *Nat. Commun.* **4,** 2203 (2013).

108. Lombardi, P. M., Cole, K. E., Dowling, D. P. & Christianson, D. W. Structure, mechanism, and inhibition of histone deacetylases and related metalloenzymes. *Curr. Opin. Struct. Biol.* **21,** 735–743 (2011).

109. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40,** 897–903 (2008).

110. Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149,** 214–231 (2012).

111. Gregory, G. D. *et al.* Mammalian ASH1L is a histone methyltransferase that occupies the transcribed region of active genes. *Mol. Cell. Biol.* **27,** 8466–8479 (2007).

112. Trotter, K. W. & Archer, T. K. The BRG1 transcriptional coregulator. *Nucl. Recept. Signal.* **6,** e004 (2008).

113. Brès, V., Yoh, S. M. & Jones, K. A. The multi-tasking P-TEFb complex. *Curr. Opin. Cell Biol.* **20,** 334–340 (2008).

114. Venturini, L. *et al.* TIF1gamma, a novel member of the transcriptional intermediary factor 1 family. *Oncogene* **18,** 1209–17 (1999).

115. Jeanmougin, F., Wurtz, J. M., Le Douarin, B., Chambon, P. & Losson, R. The bromodomain revisited. *Trends Biochem. Sci.* **22,** 151–153 (1997).

116. Laumonnier, F. *et al.* Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. *J. Med. Genet.* **42,** 780–6 (2005).

117. Loenarz, C. *et al.* PHF8, a gene associated with cleft lip/palate and mental retardation, encodes for an Nepsilon-dimethyl lysine demethylase. *Hum. Mol. Genet.* **19,** 217–22 (2010).

118. Kaneda, M. *et al.* Essential role for *de novo* DNA methyltransferase Dnmt3a in

paternal and maternal imprinting. *Nature* **429,** 900–903 (2004).

119. Welstead, G. G. *et al.* X-linked H3K27Me3 demethylase Utx is required for embryonic development in a sex-specific manner. *Proc. Natl. Acad. Sci.* **109,** 13004–13009 (2012).

120. Ropero, S. & Esteller, M. The role of histone deacetylases (HDACs) in human cancer. *Molecular Oncology* **1,** 19–25 (2007).

121. Pedre, X. *et al.* Changed histone acetylation patterns in normal-appearing white matter and early multiple sclerosis lesions. *J. Neurosci.* **31,** 3435–3445 (2011).

122. Bjarkam, C. R. *et al.* Further immunohistochemical characterization of BRD1 a new susceptibility gene for schizophrenia and bipolar affective disorder. *Brain Struct. Funct.* **214,** 37–47 (2009).

123. Muller, S., Filippakopoulos, P. & Knapp, S. Bromodomains as therapeutic targets. *Expert Rev. Mol. Med.* **13,** e29 (2011).

124. Mann, B. S., Johnson, J. R., Cohen, M. H., Justice, R. & Pazdur, R. FDA approval summary: Vorinostat for treatment of advanced primary cutaneous T-cell lymphoma. *Oncologist* **12,** 1247–1252 (2007).

125. Poole, R. M. Belinostat: First global approval. *Drugs* **74,** 1543–54 (2014).

126. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468,** 1067–73 (2010).

# Chapter 2: Using AlphaScreen to Aid the Design of a Dual Inhibitor

## 2.1 Introduction

### 2.1.1 The hypoxic response

In aerobically respiring cells a constant supply of oxygen is required for the transfer of chemical energy from organic substrates such as sugars into the high-energy terminal phosphate bond of adenosine triphosphate (ATP), the "molecular unit of energy currency" in the cell.[1] To maintain metabolic homeostasis eukaryotic organisms have developed numerous methods for adapting to changes in environmental $O_2$ supply, at both local, and organismal levels. Responses to a short-term decrease in $O_2$, for example during exercise, include faster breathing and increased heart rate. Longer term changes to oxygen demand or availability, such as in populations living at high altitude can be mitigated through modulation of gene expression.[2] However, this same low oxygen (hypoxic) response plays a significant role in tumour development, improving the viability of cells in the hypoxic environment caused by rapidly proliferating cells.[3]

### 2.1.2 HIF pathway

Hypoxia inducible factor (HIF) is a transcription factor responsible for the upregulation of over 100 genes (in humans) that are involved in the cellular response to hypoxia, including genes controlling angiogenesis (*e.g.* vascular endothelial growth factor A; VEGF-A), cell migration (*e.g.* lysine oxidase; LOX), pH regulation (*e.g.* carbonic anhydrase 9; CA9) and metabolism (*e.g.* lactate dehydrogenase A; LDHA).[4–7] The HIF complex consists of two subunits, HIF-α and HIF-β (identical in sequence to the aryl hydrocarbon receptor nuclear translocator; ARNT); both of which are expressed in all

metazoan species tested to date.[8] In humans there are 3 isoforms of HIF-α (HIF-1α, 2α and 3α), all of which can dimerise with HIF-β.[9]

As HIF-β is constitutively expressed in cells in an excess relative to HIF-α levels, HIF-α levels are considered the limiting factor in HIF-1 formation.[3,5] HIF-1α levels are associated with tumour angiogenesis, invasion, and resilience, and are highly regulated in an oxygen dependent manner.[10]

### 2.1.3 Oxygen-dependent regulation of HIF-1α

Under normoxic conditions HIF prolyl-hydroxylases, such as prolyl-hydroxylase domain-containing proteins (PHD1-3, in humans), catalyse the hydroxylation of HIF-α isoforms at residues P402 and P564 (using HIF-1α numbering).[11] While hydroxylated HIF-1α is still capable of forming a transcriptionally active (α/β-)HIF complex with HIF-β, hydroxylation at either P402 or P564 substantially promotes binding of HIF-α to von Hippel-Lindau tumour suppressor protein (pVHL).[12] pVHL acts as a recognition component for an E3 ubiquitin ligase complex, which in turn results in the ubiquitination, and subsequent proteosomal degradation of prolyl-hydroxylated HIF-α isoforms.[12–14] This mechanism of constant HIF-α expression and degradation under normoxic conditions ensures the maintenance of an appropriate level of translated HIF-1α protein, along with an associated mediation of activity, in cells.

The PHDs require molecular oxygen to function, and are consequently limited by oxygen availability.[15] Under hypoxia HIF-α prolyl-hydroxylation is reduced, and HIF-α is no longer recruited to pVHL, resulting in an increased level of HIF-α. Following translocation to the nucleus, HIF-α can form the HIF-1 complexes by dimerization with HIF-β, thereby initiating the hypoxic response (see **Figure 2.1**).

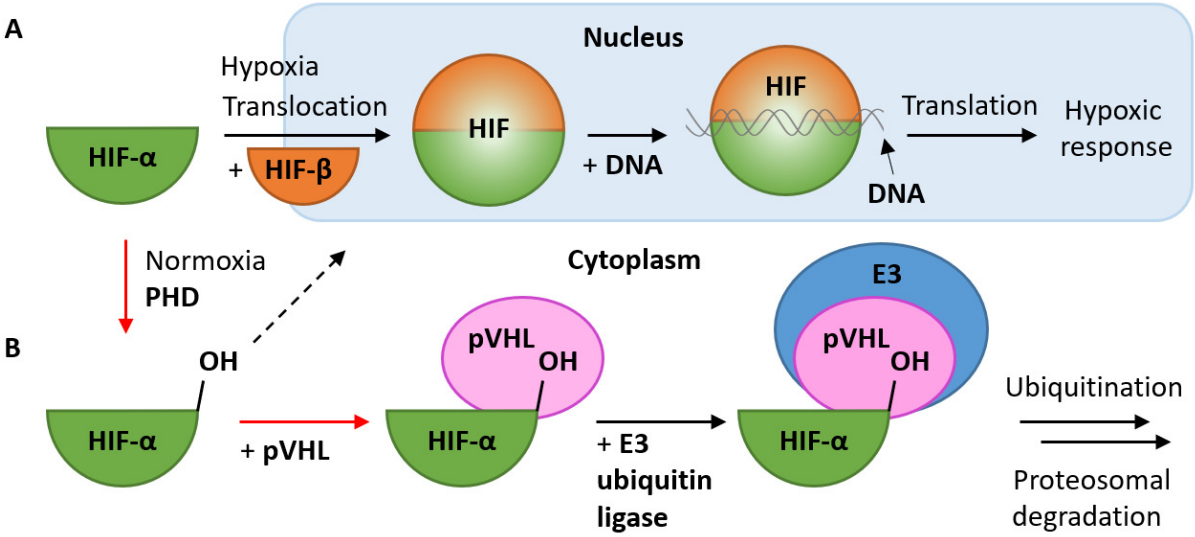



***Figure 2.1**: Schematic representation of the HIF pathway. Under hypoxic conditions (**A**) HIF-α is translocated from the cytoplasm to the nucleus where it dimerises with HIF-β to form the HIF complex, which in turn promotes the translation of genes involved in the hypoxic response. Under normoxia (**B**) HIF-α is hydroxylated at P402 or P564 by a PHD-containing protein. Hydroxylated HIF-α binds tightly to pVHL, marking HIF-α for ubiquitination by the E3 ubiquitin ligase complex, resulting in its proteosomal degradation. Note that hydroxylated HIF-α does not recruit the E3 ubiquitination complex in the absence of pVHL and is therefore still capable of forming the HIF complex with HIF-β (dotted arrow). Red arrows indicate promising intervention points that, if disrupted, result in a hypoxic-like response under normoxic conditions.*

### 2.1.4 Small molecule induction of the hypoxic response

Due to the importance of the hypoxic response to the survival of tumour cells, this pathway is of immense interest to the field of cancer research. However, investigating cultures under a hypoxic environment poses many challenges due to the equipment requirements and the impaired viability of the oxygen-starved cells under study. The ability to induce a hypoxia-like transcriptional response by dosing with a small molecule inhibitor is therefore a powerful tool for researchers.

Both the prolyl-hydroxylation of HIF-α, and the subsequent pVHL-mediated ubiquitylation of hydroxylated HIF-α are required for the restriction of HIF complex levels under normoxia. Disruption of the function of the HIF prolyl-hydroxylases or

pVHL results in an accumulation of HIF-α, or prolyl-hydroxylated HIF-α, respectively, triggering a hypoxic-like response (see **Figure 2.1**).

(4-Hydroxy-1-methyl-2-oxo-1,2-dihydroquinoline-3-carbonyl)glycine (**1**) is a selective PHD-catalysed HIF-1α prolyl-hydroxylation inhibitor, and (2*S*,4*R*)-1-((*S*)-2-acetamido-3,3-dimethylbutanoyl)-4-hydroxy-*N*-(4-(4-methylthiazol-5-yl)benzyl)pyrrolidine-2-carboxamide (**2**) inhibits the pVHL–HIF-1α interaction.[16,17] The chemical structures of compounds **1** and **2** are presented in **Figure 2.2**, **Top**. Through use of small molecule inhibitors such as **1** or **2** it is possible to artificially induce a hypoxic-like response under normal oxygen levels, and so elevate levels of HIF target genes of medical interest, such as erythropoietin and VEGF-A.[16,17]

### 2.1.5 Bromodomain and extra-terminal proteins

While HIF-1 is capable of promoting a large number of genes, each gene's expression profile can be further modified by a multitude of other transcription factors. One family of transcription factors able to influence the transcription of genes associated with the hypoxic response are the bromodomain and extra-terminal (BET) proteins.[18]

The BET proteins (BRD2-4, and bromodomain testis associated; BRDT) are a family of histone acetyl lysine 'reader' proteins that regulate transcription through the recruitment of transcription factors and epigenetic regulators.[19] Selective inhibition of the BET proteins with the inhibitor JQ1 (see **Figure 2.2**, **bottom**) has been found to alter the expression levels of ~45% of hypoxia-responsive genes, including the downregulation of oncogenes such as CA9 and VEGF-A.[18,19] JQ1 has been shown to have an effect on tumour survival and proliferation, highlighting BET inhibitors as an attractive target for anti-cancer drug discovery.[20]

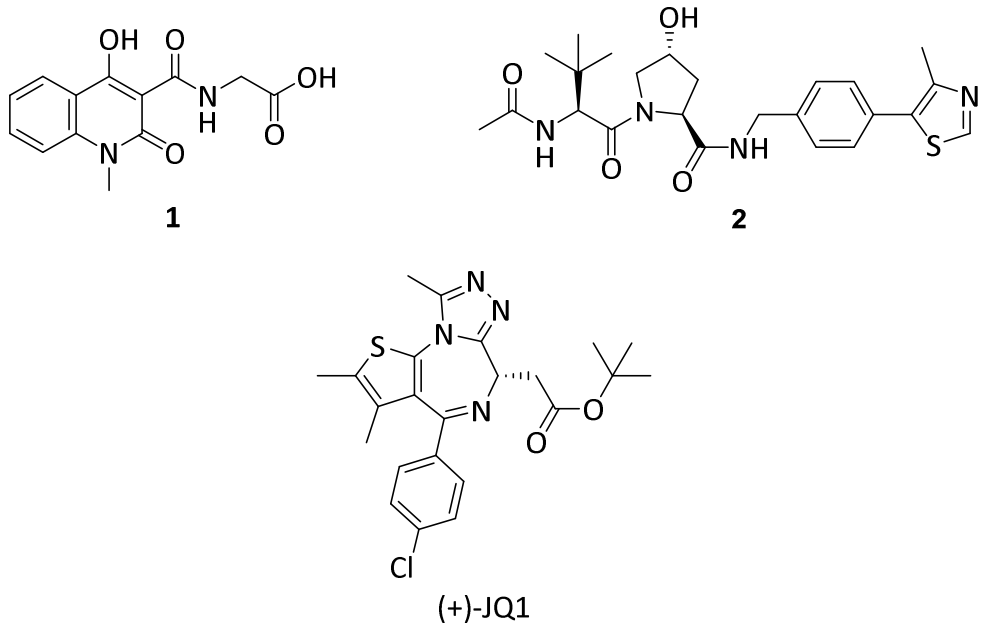



***Figure 2.2**: Chemical structures of a selection of epigenetically relevant inhibitors. **1**- Inhibits HIF prolyl hydroxylases (half maximal inhibitory concentration ($IC_{50}$) = 33 nM* vs. *PHD2).[16] **2**- Disrupts the binding of pVHL to prolyl-hydroxylated HIF-1α (dissociation constant ($K_D$)= 185 nM* vs. *pVHL).[17] **Bottom**- The active (+) enantiomer of the BET inhibitor JQ1 ($IC_{50}$ = 77 nM and 33 nM* vs. *the first and second bromodomains of BRD4, respectively).[21]*

### 2.1.6 Bi-functional inhibitors

Due to differences in drug bioavailability, metabolism and effective lifetime, co-dosing with multiple inhibitors can result in different local concentrations of the drugs. By designing a single small molecule with more than one target it should be possible to ensure identical local concentrations of the constituent activities. In the initial work described in this thesis two inhibitors are joined by a linker. While the local concentrations of the target proteins and the affinities of these target proteins to the dual inhibitor will vary, the ability to ensure an equimolar local concentration of the constituent inhibitor compounds represents an expansion of the toolbox of probes available to researchers investigating transcriptional regulation. In the longer-term, engineering more than one activity onto a single framework is desirable as the linker could potentially be cleaved in cells or have deleterious activities. This approach is of

particular interest in the field of epigenetic manipulation where phenotypes often appear to arise from more than one biochemical process.[22]

## 2.2 Objectives

Investigating transcriptional regulation is a highly complex task, so the development of new tools able to probe the combinatorial effects of multiple proteins may be key to the building up of our understanding of the larger system. This chapter details the process of designing a BET inhibitor covalently linked to either a pVHL or prolyl hydroxylase inhibitor.

The binding affinity of various dual inhibitors against the first bromodomain of BRD4 (BRD4(1)) was measured and used to inform subsequent rounds of inhibitor design. BRD4(1) was produced and purified, and the effect of linker style, length and the identity of the second inhibitor on the $IC_{50}$ value against BRD4(1) was investigated using AlphaScreen assays.

*All compounds synthesised in this chapter were prepared by Dr Hidenori Someya unless otherwise stated. PHD2 binding experiments were conducted by Kerstin Lippl, pVHL binding experiments by Lucia Serran Aguilera and all in-cell data was obtained by Tzu-Lan Yeh.*

## 2.3 AlphaScreen assays

AlphaScreen (Amplified Luminescence Proximity Homogeneous Assay Screen) is a fluorescence-based competition assay used to determine the relative binding affinity of a binding partner to a protein.[23] Assay conditions contain 'donor' and 'acceptor' beads. Donor beads contain phthalocyanine, a compound capable of converting ambient oxygen into an excited singlet form when illuminated by light at $\lambda$ = 680 nm.[24] The acceptor beads contain a thioxene-derivative that can react with singlet oxygen, resulting in the production of light with wavelengths between 520-620 nm.[24] Singlet oxygen is highly reactive, with a half-life of 4 μsec, during which time it can diffuse over around 200 nm.[25] Fluorescence is therefore only produced if the donor and acceptor beads are in close proximity. Both beads also contain a tag, which can be varied as required. In the experiments detailed below the acceptor was coupled to nickel-bound nitrilotriacetic acid, and the donor was coupled to streptavidin. In the presence of a his-tagged protein and a biotin-tagged binding partner the donor and acceptor beads are held together, and a fluorescent signal is observed on illumination at $\lambda$ = 680 nm. If an inhibitor is present and able to disrupt this protein–binding partner interaction, fewer donor and acceptor beads will be proximal to each other, resulting in a decrease in fluorescent signal. In these experiments the his-tagged protein was a $BRD4_{44\text{-}168}$ (*Homo sapiens*) construct, comprising the first bromodomain of BRD4 (BRD4(1)). The biotin-tagged binding partner was a $H4_{1\text{-}20}$ peptide tetra-acetylated at positions K5, K8, K12 and K16 (supplied by *Dr Oleg Fedorov*). This $H4_{1\text{-}20}$ (K5, 8, 12, 16)$Ac_4$ peptide tightly binds to BRD4(1) and has been established as a good binding partner for BRD4(1) in previously reported AlphaScreen studies.[24,26,27]

### 2.3.1 BRD4(1) production

#### Bacterial growth and harvesting

A his-tagged construct coding for BRD4(1) in a pNIC28-Bsa4 vector backbone was transformed into *Escherichia coli* (*E. coli*) BL21 (DE3) Gold cells. Flasks containing 600 mL 2TY media and 50 μL/mL kanamycin were inoculated with 1% of the bacterial starter culture. Cells were grown at 37 °C with shaking at 180 rotations per minute (rpm) until an $OD_{600}$ of 0.8-1.0 was reached. The temperature was lowered to 18 °C and protein expression was induced by the addition of IPTG at a final concentration of 0.25 mM. Cultures were incubated overnight (18 °C, 180 rpm) before harvesting by centrifugation. The resultant cell pellets were then stored at –80 °C. Full experimental details and buffer compositions are provided in **Materials and Methods 8.5.7**.

#### BRD4(1) purification

The cell pellet was dissolved in binding buffer (5-times volume to weight (v/w) of cell pellet) and lysed by sonication (10 s bursts at 10 s intervals for a total sonication time of 10 min). The lysate was clarified by centrifugation (50000*g*, 20 min, 4 °C). The supernatant was decanted off and filtered through a 0.45 μm filter. A 5 mL HisTrap FF chromatography column (GE Healthcare) was charged with 4 column volumes (CV) of 100 mM $NiSO_4$ and equilibrated with 5 CV of binding buffer before addition of the filtered lysate at 1 mL/min. The column was then washed with wash buffer (20 CV) and the column-bound protein was batch eluted with of elution buffer (4 CV). Fractions containing the highest concentration of protein were concentrated to 1-2 mL, then loaded onto a 120 mL Superdex 75 size exclusion chromatography (SEC) column (GE Healthcare) pre-equilibrated with gel filtration buffer. The protein was eluted with gel filtration buffer, and fractions identified by UV traces as containing protein were

analysed by SDS-PAGE. The identity of the purified protein was validated by mass spectrometry. Fractions containing the purest samples of BRD4(1) by SDS-PAGE were concentrated to 200 μM and flash cooled in liquid $N_2$, then stored at -80 °C until required. The UV trace following passage through the final Superdex 75 SEC column, as well as the results of the SDS-PAGE gel of the protein-containing fractions and the deconvoluted MS spectrum are provided in **Figure 2.3**. Buffer compositions are provided in **Materials and Methods 8.6.2**.

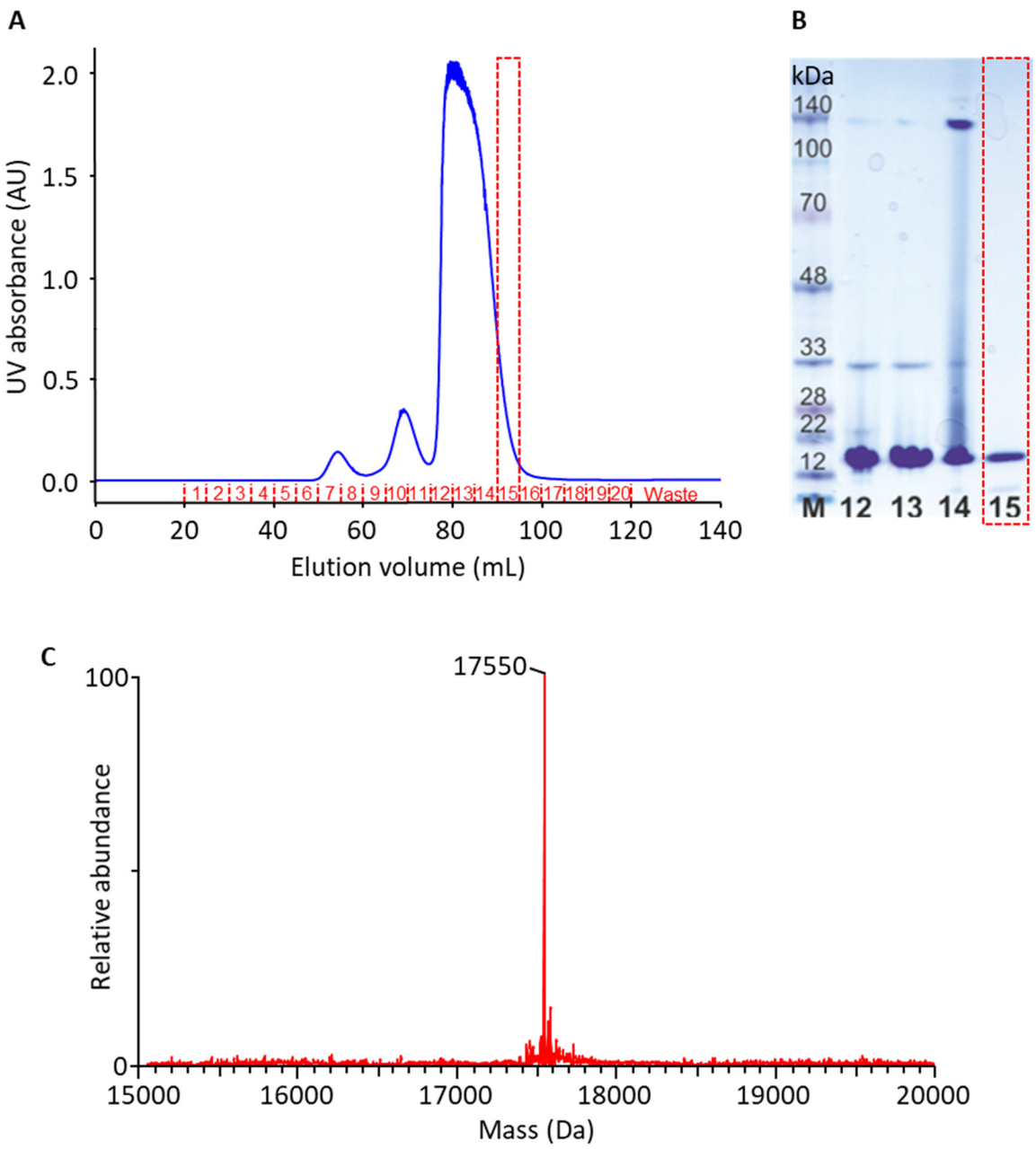



***Figure 2.3**: **A**- Elution profile of purified BRD4(1) from a 150 mL S75 gel filtration column. **B**- SDS-page gel displaying fractions containing high quantities of BRD4(1). 1.2 mg was concentrated at >95% purity (fraction 15), and 18 mg at ~90% purity (fractions 12-13), and flash frozen in liquid $N_2$. Aliquots from fraction 15 were used for all experiments described in this thesis. The marker (M) used was a Precision Plus Dual Colour™ protein ladder (Bio-Rad). **C**- Deconvoluted MS spectra of BRD4(1) (expected mass: 17549 Da) from fraction 15.*

### 2.3.2 AlphaScreen dose-response curve

The sensitivity of the AlphaScreen assay is dependent on the concentration of protein, binding partner and AlphaScreen beads in the assay mixture.[23] The optimal concentrations can be determined using a dose-response curve. This process can also provide validation that the sample of BRD4(1) produced in **Section 2.4** comprises properly folded protein. The AlphaScreen bead concentrations were kept constant at 10 μg/mL final assay concentration (FAC) for each experiment.

A 1000 nM sample of biotin-tagged $H4_{1\text{-}20}$ (K5, 8, 12, 16)$Ac_4$ peptide in AlphaScreen buffer was diluted into freshly prepared AlphaScreen buffer in a 12-step 2-fold dilution series resulting in a 1000-0.488 nM (250-0.122 nM FAC) range of peptide concentrations. BRD4(1) protein was tested at four concentrations (2, 5, 10 and 20 nM, FAC, in AlphaScreen buffer). To a 5 μL sample of each of the four concentrations of BRD4(1) was added a 5 μL sample of every peptide concentration in a 384-well plate. The plate was sealed using Aluminium StarSeal (STARLAB) and incubated at room temperature (RT) for 30 minutes. A mixture containing nickel chelate donor and streptavidin-bound acceptor beads (20 μg/mL, each) was subsequently added to each well (10 μL, 10 μg/mL FAC) under low light conditions. The plate was then sealed and incubated in the dark (1 hr, RT). Each well was then illuminated with 680 nm light and the absorbance at $\lambda$ = 520-620 nm was recorded using an EnVision Multilabel Plate Reader (PerkinElmer). The results for each concentration of BRD4(1) were plotted on a line graph (see **Figure 2.4**) and the optimal $H4_{1\text{-}20}$ (K5, 8, 12, 16)$Ac_4$ peptide and BRD4(1) concentrations were identified. Buffer compositions are provided in **Materials and Methods 8.8**.

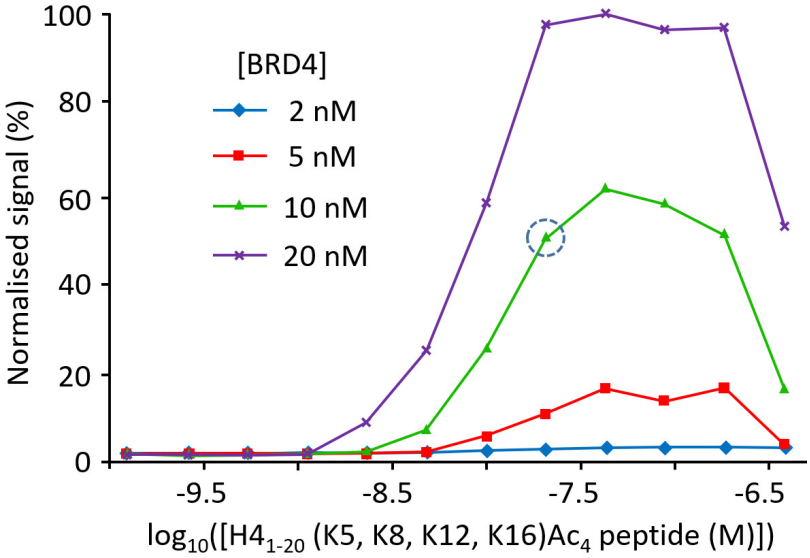



***Figure 2.4**: AlphaScreen dose-response curves for a range of biotin-tagged $H4_{1-20}$ (K5, K8, K12, K16)$Ac_4$ peptide, and his-tagged BRD4(1) concentrations. The normalised signal, presented as a percentage of the maximum signal detected is plotted against the $log_{10}$([$H4_{120}$ (K5, K8, K12, K16)$Ac_4$ peptide (M)]). The concentrations selected for use in future experiments are circled (10 nM BRD4(1) and 15.6 nM peptide).*

As the condition containing 10 nM BRD4(1) and 15.6 nM peptide displayed a strong signal while indicating a high sensitivity to a decrease in peptide concentration, these concentrations were selected for the main AlphaScreen experiments. While a stronger signal could be detected when using a higher concentration of peptide and protein, the minimum sensitivity of the assay is determined by the peptide concentration used, so a lower concentration was preferable. The fact that the signal intensity approached a peak when the protein and peptide concentrations were similar indicates the majority of BRD4(1) was folded appropriately for binding to the $H4_{1-20}$ (K5, K8, K12, K16)$Ac_4$ peptide.

Note that at high concentrations of peptide all binding sites on the donor beads become saturated. Beyond this point, increasing the peptide concentration results in a decrease in fluorescent signal observed as the unbound excess peptide begins to outcompete the donor bead-bound peptide. This phenomenon is known as the "hook effect" and can be seen at the highest concentration of peptide used in this experiment.

## 2.4 Designing and testing a bi-functional BRD–PHD inhibitor

The first approach to producing a small molecule capable of inducing a hypoxic-like response while also inhibiting BET proteins attempted was to design a compound comprising the BRD inhibitor JQ1 and the PHD inhibitor **1** connected by a linker (see **Figure 2.6**).

### 2.4.1 Selecting inhibitor linking sites

During the process of devising a way to join two biologically active molecules, care must be taken when siting the linker to avoid interrupting the original inhibitory function of the compounds. Using previously reported crystal structures of BRD4(1) complexed with JQ1, and PHD2 complexed with compound **1**, it was possible to identify promising linker locations for both inhibitors (see **Figure 2.5**, **top**).[16,21] The *t*-butyl group of JQ1 and the proton at the 7-position of the quinoline ring of compound **1** were selected for substitution with the linker (see **Figure 2.5**, **bottom**).

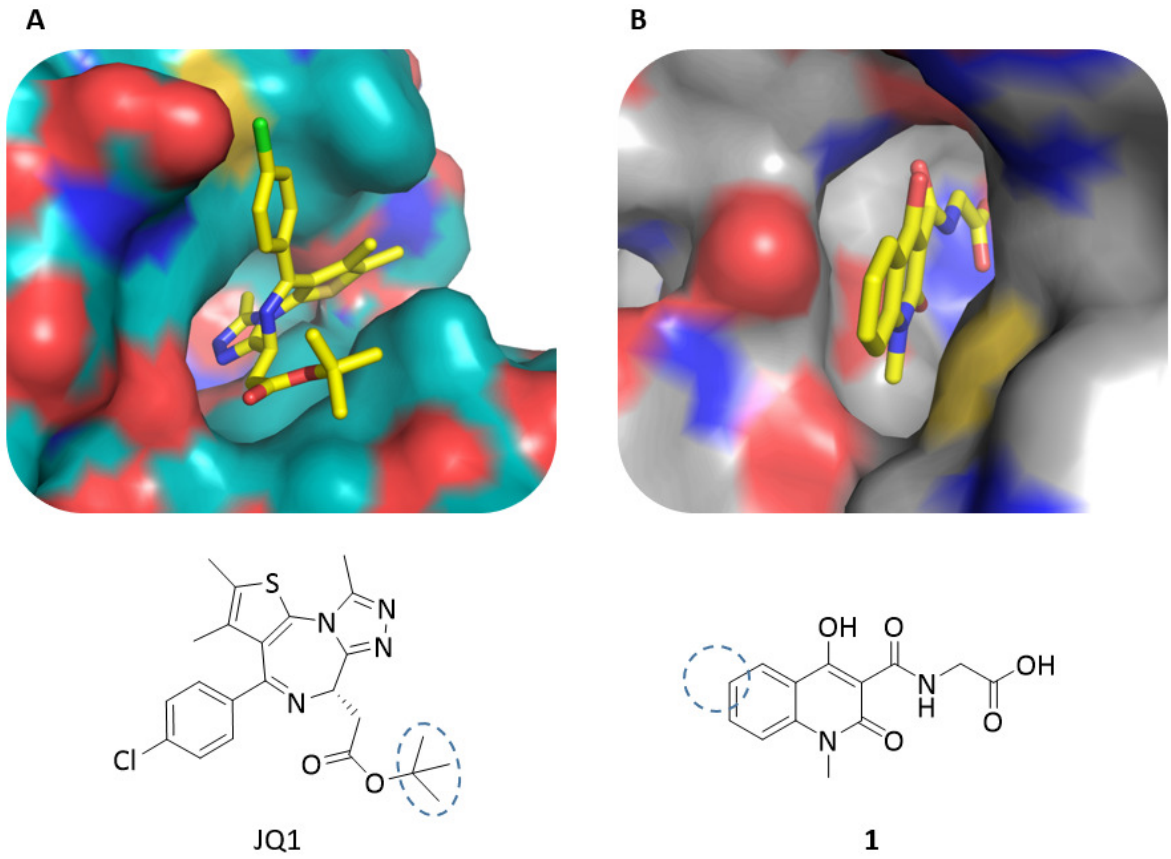



***Figure 2.5**: Views from crystal structures of; **A**- BRD4(1) complexed with JQ1 (PDB ID: 3MXF), and **B**- PHD2 complexed with compound **1** (PDB ID: 4BQW).[16,21] Chemical structures of JQ1 and compound **1** are also provided. Promising potential linker locations are circled.*

### 2.4.2 Investigating the effect of linker length on the binding strength of a selection of dual BRD–PHD inhibitors to BRD4(1)

The nature of the linker bridging the two inhibitors has the potential to dramatically affect the inhibition strength of the compound, as well as altering the biophysical properties of the molecule, such as solubility and stability. In order to investigate the effect of linker length on BRD4(1) inhibition, three inhibitors were designed containing linkers 9, 15 and 23 atoms long (compounds **4**-**6**, respectively). In addition, the carboxylate analogue of JQ1 (compound **3**) was prepared to probe the importance of the bulky *t*-butyl group in BRD4(1) binding. Compounds **3**-**6** were synthesised by *Dr Hidenori Someya* and the structures are provided in **Figure 2.6**.

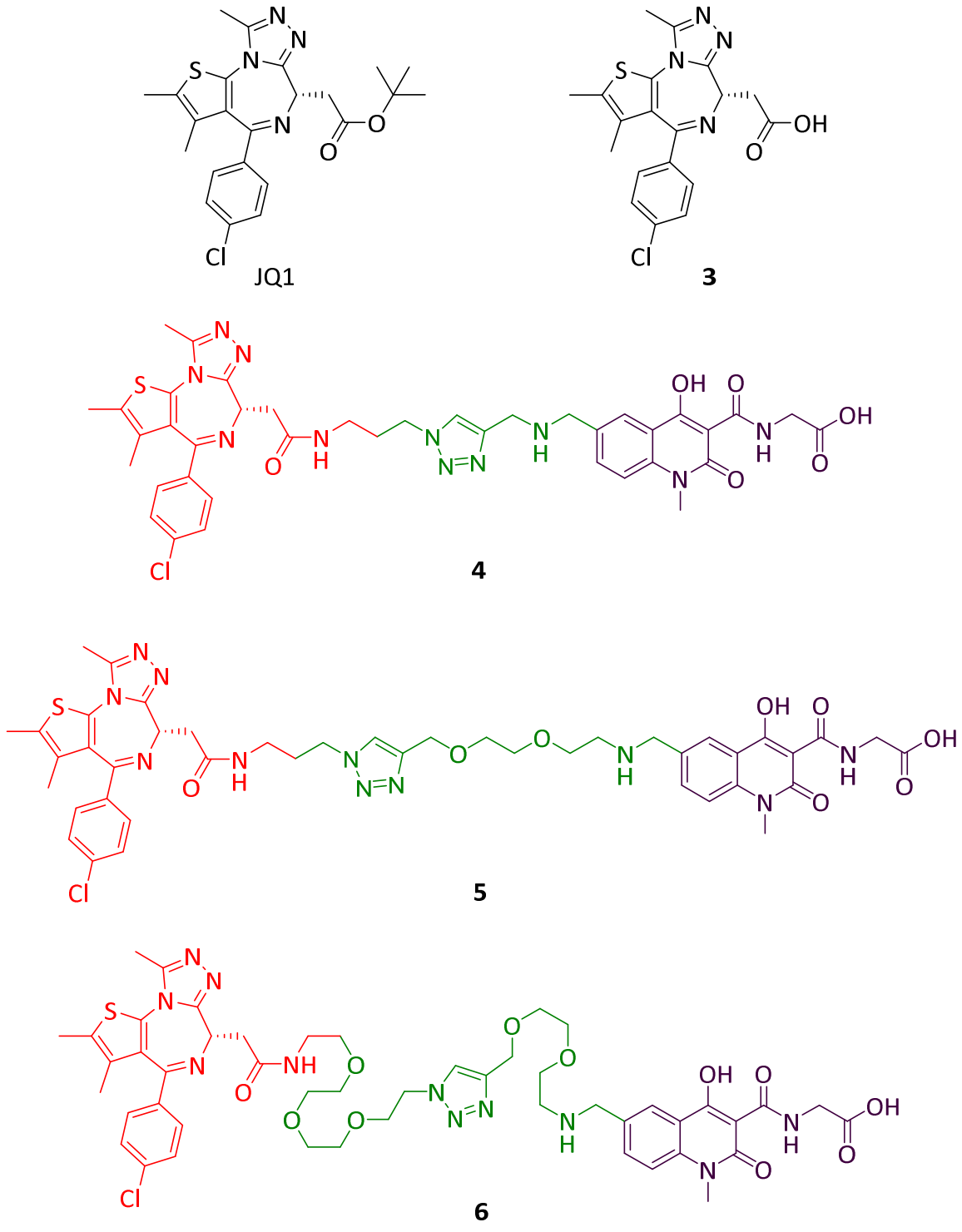



***Figure 2.6**: Chemical structures of the BRD inhibitor JQ1, an acid based on JQ1 (**3**), and the dual BRD–PHD inhibitors **4**-**6**. The BRD-inhibiting components of the dual inhibitors are shown in red, the PHD-inhibiting components in purple, and the linker in green.*

### 2.4.3 BRD–PHD dual inhibitor AlphaScreen assay

A sample of BRD4(1) was diluted in freshly prepared AlphaScreen buffer to 40 μM (10 μM FAC) and dispensed (5 μL) into a 384 well plate containing 20 μL reaction wells (Perkin Elmer). A 5 mM solution of each of the compounds **3-6** and JQ1 were prepared in DMSO before dilution to 200 μM by addition of AlphaScreen buffer. A 14-step 2-fold dilution series into AlphaScreen buffer supplemented with 4% DMSO was then prepared for each of the compounds, resulting in a concentration range of 200-0.012 μM (50-0.003 μM FAC). Three repeats of each compound at each concentration (5 μL) were added to the reaction wells. The wells were then sealed and incubated at RT for 30 minutes. $H4_{1\text{-}20}$ (K5, K8, K12, K16)$Ac_4$ peptide (10 μL, 31.2 μM; 15.6 μM FAC) in AlphaScreen buffer was added to each well. The plate was re-sealed and stored at RT for 30 minutes. A 5 μL sample containing nickel chelate donor and streptavidin-bound acceptor beads at 40 μg/mL (10 μg/mL FAC) in AlphaScreen buffer was prepared and added to each reaction well under low light conditions. The plate was sealed for a third time and incubated for 1 hour at RT in the dark. Each well was illuminated by light with a wavelength of 680 nm, and the fluorescence was recorded at λ = 520-620 nm. Results were plotted as $\log_{10}$([inhibitor (M)]) *vs.* % inhibition, and a variable slope (four parameter) curve was fitted using GraphPad prism (see **Figure 2.7**).[28]

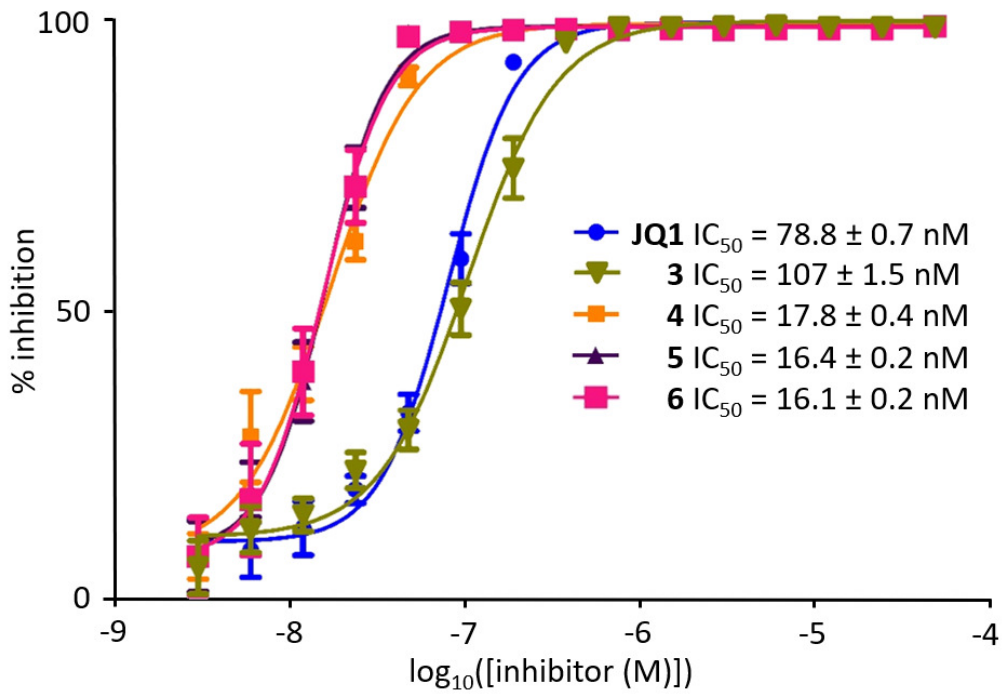



***Figure 2.7**: Line graph displaying AlphaScreen data for a range of JQ1-based dual-inhibitors containing a short (**4**), medium (**5**) and long (**6**) linker, as well as the JQ1-based acid (**3**). JQ1 is included as a reference compound. $IC_{50}$ values are presented for each compound ± standard error from a minimum of three separate experiments.*

The experimentally derived $IC_{50}$ value for JQ1 was 78.8 ± 0.7 nM, comparing well to the previously reported literature value of 77 nM against BRD4(1).[21] A small reduction in the apparent inhibitory effect was observed for compound **3** (107 ± 1.5 nM), indicating that the presence of the bulky *t*-butyl group on JQ1 was beneficial, but not essential for binding. Each of the three dual inhibitors showed a lower $IC_{50}$ against BRD4(1) than JQ1 ($IC_{50}$ = 17.8 ± 0.4; 16.4 ± 0.2; and 16.1 ± 0.2 nM for compounds **4**-**6**, respectively), indicating a 4-5-fold tighter binding than the parent compound. In each of the compounds **4**-**6**, the linker is identical for the first two $(CH_2)_2$ groups in the chain after the amide bond. These results indicate that an ethyl group may be a better substitute than the bulky *t*-butyl group on JQ1. Additionally, the linker length appears to have little effect on the ability of the dual inhibitor to bind BRD4(1).

**DMSO concentrations**

It should be noted that DMSO can act as an inhibitor of bromodomains and has been shown to interrupt binding between BRD4(1) and an $H4_{1-20}$ (K5, K8, K12, K16)$Ac_4$ peptide with an $IC_{50}$ of 281 mM in AlphaScreen experiments.[29] Each AlphaScreen condition tested contained a FAC of 1% DMSO (141 mM). Repeating the abovementioned AlphaScreen experiment using a dilution series of JQ1 diluted into AlphaScreen buffer (with no additional DMSO) allowed for the calculation of an $IC_{50}$ of 80 ± 3 nM. As JQ1 was found to have a similar $IC_{50}$ in both the 1% DMSO AlphaScreen experiments (78.8 ± 0.7 nM) and the low DMSO control (80 ± 3 nM) it was concluded that the presence of 1% DMSO did not significantly interfere with BRD4(1)–inhibitor binding.

### 2.4.4 BRD–PHD dual inhibitor activity against PHD2 and *in vivo* activity

Each of the BRD–PHD inhibitors (**4**-**6**) were tested against PHD2 by AlphaScreen assay (by *Kerstin Lippl,* data not shown) and found to have $IC_{50}$ values in the single μM range, representing a ~100-fold decrease in binding affinity relative to the PHD inhibitor alone (**1**) against PHD2 ($IC_{50}$ = 33 nM).[16] Additionally, *in vitro* studies investigating Henrietta Lacks (HeLa) cells treated with compounds **4**-**6** at 50 and 500 μM indicated no significant increase in HIF-1α levels compared to untreated cells (carried out by *Tzu-Lan Yeh,* data not shown) indicating a lack of activity in cells.

## 2.5 Designing and testing a bi-functional BRD–pVHL inhibitor

### 2.5.1 Designing a BRD–pVHL dual inhibitor

Larger molecules often have more difficulty passing into cells than smaller molecules. As the three linker lengths tested in compounds **4**-**6** did not seem to significantly alter the ability of the dual inhibitor to bind either of the target proteins *in vitro*, it was decided to produce a BRD–pVHL dual inhibitor with as short a linker as possible.

Regions of similarity between the BRD4(1) inhibitor JQ1, and the pVHL inhibitor compound **2**, provided the possibility of designing a dual inhibitor with the linker region incorporated into the original structure of the inhibitors (compound **7**) (see **Figure 2.8**). Compound **7** was synthesised by *Dr Hidenori Someya*.

A previously reported crystal structure of pVHL co-crystallised with the inhibitor **2** showed the amide-bound methyl group to be in a reasonably exposed location (see **Figure 2.8**, **top**).[17] Continuation of the dual inhibitor into the BRD region at that location appeared likely to minimise disruption to pVHL binding.

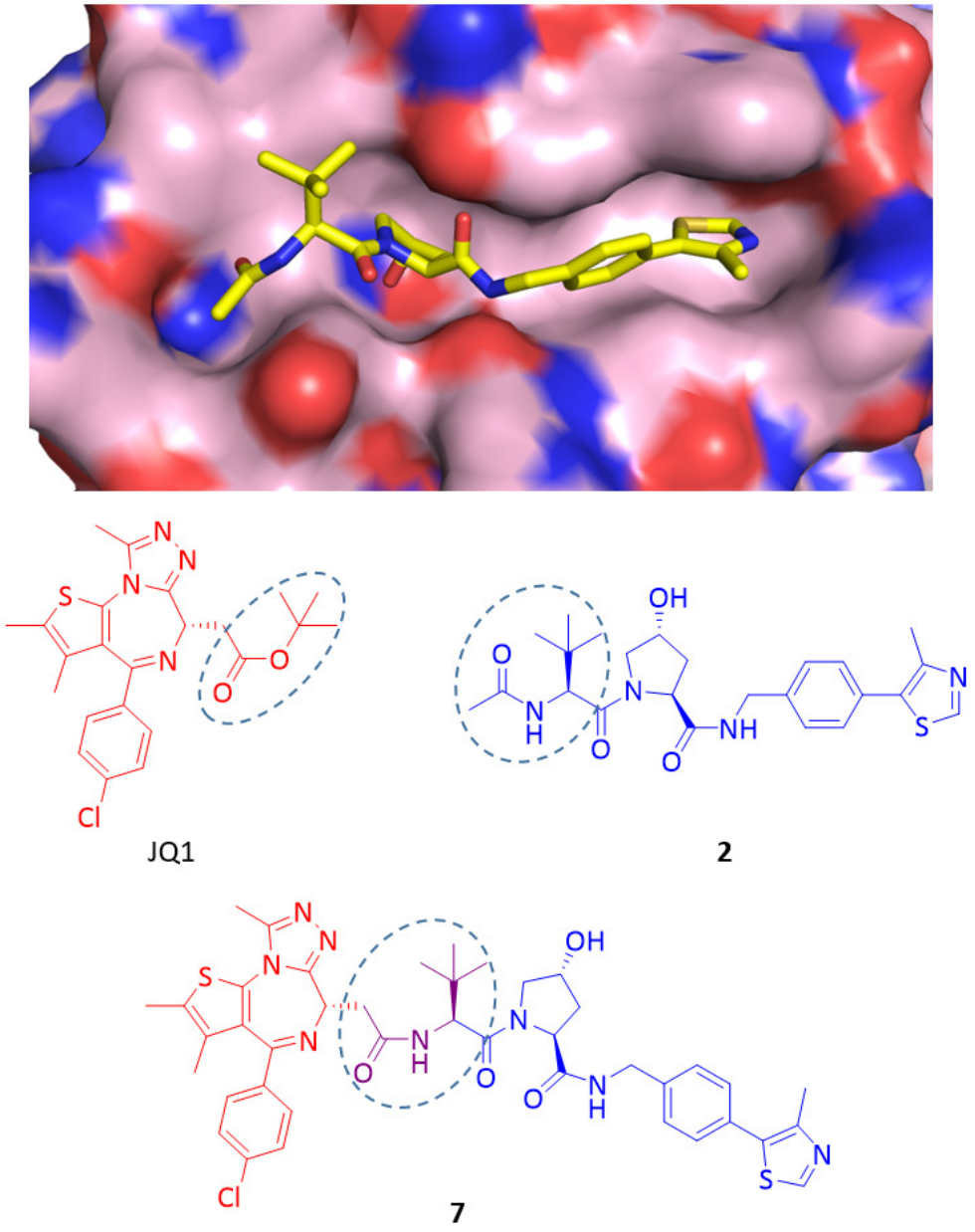



***Figure 2.8**: **Top**- View from a crystal structure of pVHL co-crystallised with an inhibitor (compound **2**) (PDB ID: 4W9H).[17] Chemical structures of the BRD4(1) inhibitor JQ1 and the pVHL inhibitor (**2**) are also shown. Areas of similarity between the two inhibitors are circled, and the structure of the dual BRD–pVHL **7** inhibitor is also provided. The BRD-inhibiting components and substructures are shown in red, the pVHL-inhibiting regions in blue, and the areas of overlap in the dual inhibitor in purple.*

### 2.5.2 BRD–pVHL dual inhibitor AlphaScreen assay

The potency of compound **7** against BRD4(1) was tested by AlphaScreen following the protocol described in **Section 2.4.3**. As before, JQ1 was used as a reference compound.

Results were plotted as $\log_{10}$([inhibitor (M)]) *vs.* % inhibition, and a variable slope (four parameter) curve was fitted using GraphPad prism (see **Figure 2.9**).[28]

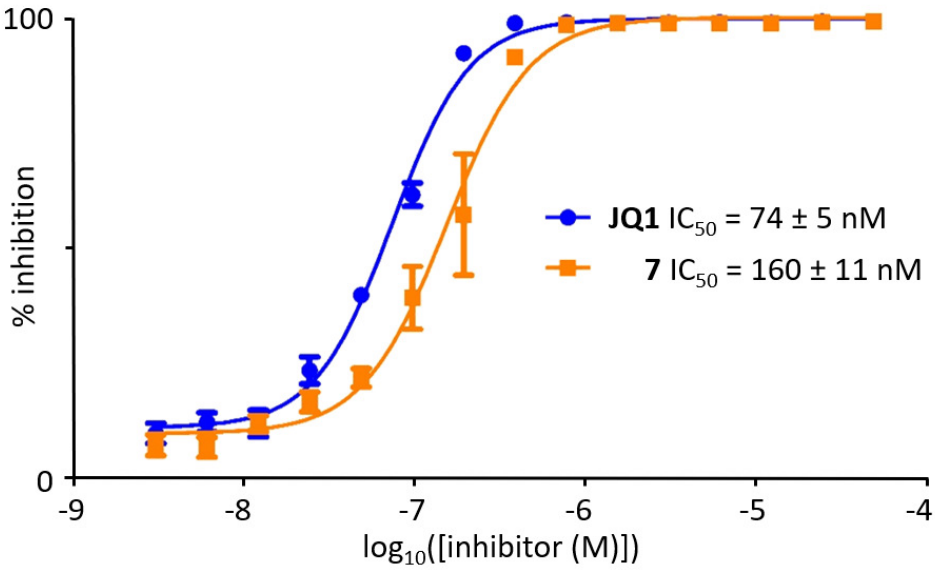



***Figure 2.9**: Line graph displaying AlphaScreen data for the BRD–pVHL dual inhibitor compound **7**. JQ1 is included as a reference compound. $IC_{50}$ values are presented for both compounds ± standard error from a minimum of three separate experiments.*

The AlphaScreen experiments indicated compound **7** to be active as an inhibitor of BRD4(1) *in vitro* with an $IC_{50}$ of 160 ± 11 nM. While this represents a 2-fold decrease in activity against BRD4(1) relative to JQ1 ($IC_{50}$ = 74 ± 5 nM), it was still a tight enough binder to test as a probe compound in cells.

The reference compound, JQ1, was found to have an $IC_{50}$ within error of the literature value ($IC_{50}$ = 77 nM), validating the fidelity of the assay.[21]

### 2.5.3 BRD–pVHL dual inhibitor activity against pVHL and *in vivo* activity

The binding strength of compound **7** to pVHL was then determined using isothermal calorimetry (ITC) (by *Lucia Serran Aguilera*, data not shown), resulting in a measured $K_D$ of 31.5 μM. A $K_D$ of 1.54 μM for the parent pVHL inhibitor (**2**) was also measured against the same batch of pVHL by ITC (carried out by *Lucia Serran Aguilera*, data not shown), indicating weaker binding than the reported literature value (literature $K_D$ = 0.185 μM).[17]

Treatment of HeLa cells with 100 μM compound **7** for 4 hours resulted in increased levels of hydroxylated HIF-1α being observed by Western blot analysis, presumably due to VHL inhibition (experiment performed by *Tzu-Lan Yeh*, data not shown).

Using quantitative polymerase chain reaction (qPCR), expression levels for several genes believed to be involved in the hypoxic response were analysed in HeLa cells treated with the BRD–pVHL dual inhibitor compound **7**, the pVHL inhibitor compound **2**, the BRD inhibitor JQ1, and no inhibitor (by *Tzu-Lan Yeh*, data not shown).[18,19] Similar levels of upregulation of the hypoxically-activated genes *PHD3* and *BNIP3* were observed in cells treated with compounds **2** and **7**. While the oncogene *CA9* was upregulated in cells treated with the pVHL inhibitor **2**, this was not the case with cells treated with the dual BRD–pVHL inhibitor **7**, consistent with the established role of JQ1 as a *CA9* suppressor.[21] Interestingly, levels of the oncogene *VEGF-A* appeared to be upregulated by compound **7**, even when compared to cells treated with compound **2**; this observation is particularly surprising considering JQ1 is currently reported to downregulate VEGF-A expression.[19]

## 2.6 Discussion

Deciphering the role of transcriptional regulation in epigenetics is a formidable task that will require many different approaches. The design of new tools capable of teasing apart the effects of different components of such a complex system is crucial. The development of dual inhibitors such as those discussed in this chapter could, for example, make it possible to use a small molecule to help probe the differences between transcriptional regulation controlled by bromodomains, and the HIFs, as well as the role of bromodomains in HIF-mediated expression.

This chapter detailed the design of a range of dual inhibitors capable of both inducing a hypoxic-like response, and inhibiting BET-domain proteins. Combining a BRD inhibitor with a PHD inhibitor by means of a linker of variable length resulted in a series of compounds with binding strengths against BRD4(1) greater even than the BRD inhibitor they were based on (JQ1).[21]

While these dual BRD–PHD inhibitors were also capable of binding to PHD2 with a μM affinity, they were unable to increase HIF-1α levels *in vitro*. This may have been due to an inability of the bulky compounds to permeate the cell membrane, or possibly due to the action of other PHD-containing proteins not inhibited by the dual inhibitor. Between each of the three linker lengths tested, no significant difference to BRD4(1) $IC_{50}$, PHD2 $IC_{50}$, or HIF-1α stabilisation was observed. Potential modifications to a BRD–PHD dual inhibitor design that might improve activity in cells include; incorporation of a shorter linker, changing the position of the linker on the PHD inhibitor component (*e.g.* to the 8-position of the quinoline ring), or selecting a different PHD inhibitor to base the dual inhibitor on.

In an attempt to design a dual BRD–pVHL inhibitor with as small a size as possible, the two component inhibitors were joined *via* a region of 'homology', resulting in a dual inhibitor smaller than the two 'combined' inhibitors that comprise it. This dual BRD–pVHL inhibitor was able to tightly bind both BRD4(1) and pVHL *in vitro*, albeit to a lesser degree than the single inhibitors.

The dual BRD–pVHL inhibitor **7** was active in cells, increasing levels of hydroxylated HIF-1α in HeLa cells. Compound **7** was capable of upregulating some proteins associated with the hypoxic response, while simultaneously downregulating HIF-promoted expression of the oncogene *CA9*, indicating successful inhibition of BRD4. Both CA9 and VEGF-A were expected to be downregulated through inhibition of BRD4, so the observation of an increase in VEGF-A levels in cells treated with dual BRD–pVHL inhibitor, when compared to cells treated with BRD inhibitor (JQ1), pVHL inhibitor (**3**) and no inhibitor, was surprising. This unexpected response highlights the complexity of transcriptional regulation, as well as the importance of the development of new tools to identify and probe gaps in our understanding.

## 2.7 References


1. Knowles, J. R. Enzyme-catalyzed phosphoryl transfer reactions. *Annu. Rev. Biochem.* **49,** 877–919 (1980).
2. Höpfl, G., Ogunshola, O. & Gassmann, M. Hypoxia and high altitude. The molecular response. *Adv. Exp. Med. Biol.* **543,** 89–115 (2003).
3. Semenza, G. L. Targeting HIF-1 for cancer therapy. *Nat. Rev. Cancer* **3,** 721–732 (2003).
4. Semenza, G. L. HIF-1 and mechanisms of hypoxia sensing. *Curr. Opin. Cell Biol.* **13,** 167–171 (2001).
5. Semenza, G. L. Hypoxia-inducible factors in physiology and medicine. *Cell* **148,** 399–408 (2012).
6. Ward, C. *et al.* New strategies for targeting the hypoxic tumour microenvironment in breast cancer. *Cancer Treat. Rev.* **39,** 171–179 (2013).
7. McIntyre, A. & Harris, A. L. Metabolic and hypoxic adaptation to anti-angiogenic therapy: A target for induced essentiality. *EMBO Mol. Med.* **7,** 368–79 (2015).
8. Loenarz, C. *et al.* The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, *Trichoplax adhaerens*. *EMBO Rep.* **12,** 63–70 (2011).
9. Ravenna, L., Salvatori, L. & Russo, M. A. HIF3α: The little we know. *FEBS J.* **283,** 993–1003 (2016).
10. Zagzag, D. *et al.* Expression of hypoxia-inducible factor 1α in brain tumors: Association with angiogenesis, invasion, and progression. *Cancer* **88,** 2606–2618 (2000).
11. Snell, C. E. *et al.* Proline-hydroxylated hypoxia-inducible factor 1α (HIF-1α) upregulation in human tumours. *PLoS One* **9,** e88955 (2014).
12. Jaakkola, P. *et al.* Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science* **292,** 468–472 (2001).
13. Maxwell, P. H. *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399,** 271–275 (1999).
14. Iwai, K. *et al.* Identification of the von Hippel–Lindau tumor-suppressor protein as part of an active E3 ubiquitin ligase complex. *Proc. Natl. Acad. Sci.* **96,** 12436–12441 (1999).
15. Kaelin, W. G. & Ratcliffe, P. J. Oxygen sensing by metazoans: The central role of the HIF hydroxylase pathway. *Mol. Cell* **30,** 393–402 (2008).
16. Chowdhury, R. *et al.* Selective small molecule probes for the hypoxia inducible factor (HIF) prolyl hydroxylases. *ACS Chem. Biol.* **8,** 1488–1496 (2013).
17. Galdeano, C. *et al.* Structure-guided design and optimization of small molecules

targeting the protein–protein interaction between the von hippel-lindau (VHL) E3 ubiquitin ligase and the hypoxia inducible factor (HIF) alpha subunit with *in vitro* nanomolar affinities. *J. Med. Chem.* **57,** 8657–8663 (2014).

18. da Motta, L. L. *et al.* The BET inhibitor JQ1 selectively impairs tumour response to hypoxia and downregulates CA9 and angiogenesis in triple negative breast cancer. *Oncogene* **36,** 122–132 (2017).

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

20. Rosner, M. & Hengstschläger, M. Targeting epigenetic readers in cancer. *N. Engl. J. Med.* **367,** 1764–1765 (2012).

21. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468,** 1067–73 (2010).

22. Willmore, K. E., Young, N. M. & Richtsmeier, J. T. Phenotypic variability: Its components, measurement and underlying developmental processes. *Evol. Biol.* **34,** 99–120 (2007).

23. Quinn, A. M. *et al.* A homogeneous method for investigation of methylation-dependent protein–protein interactions in epigenetics. *Nucleic Acids Res.* **38,** e11 (2009).

24. Roberts, J. M. & Bradner, J. E. A bead-based proximity assay for BRD4 ligand discovery. *Curr. Protoc. Chem. Biol.* **7,** 263–78 (2015).

25. Wilkinson, F., Helman, W. P. & Ross, A. B. Rate constants for the decay and reactions of the lowest electronically excited singlet state of molecular oxygen in solution. An expanded and revised compilation. *J. Phys. Chem. Ref. Data* **24,** 663–677 (1995).

26. Dey, A., Chitsaz, F., Abbasi, A., Misteli, T. & Ozato, K. The double bromodomain protein BRD4 binds to acetylated chromatin during interphase and mitosis. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 8758–8763 (2003).

27. Jung, M. *et al.* Affinity map of bromodomain protein 4 (BRD4) interactions with the histone H4 tail and the small molecule inhibitor JQ1. *J. Biol. Chem.* **289,** 9304–9319 (2014).

28. GraphPad Software. GraphPad Prism version 5.04 for Windows (2010).

29. Philpott, M. *et al.* Bromodomain-peptide displacement assays for interactome mapping and inhibitor discovery. *Mol. Biosyst.* **7,** 2899 (2011).

# Chapter 3: Investigation of a Putative New PTM by AlphaScreen

## 3.1 Introduction

### 3.1.1 Post-translational modifications

Over 20 types of PTMs have been established as occurring in histones, a number that continues to rise with the application of advanced mass spectrometric methods and the development of specific antibodies.[1–3] Lysine is considered the most modified histone residue, both in terms of the range of different PTMs occurring at lysine residues, and the number of locations modified.[4] Multiple different PTMs can occur at the same position, resulting in similar, or contrasting effects, depending on the location. Trimethylation and acetylation marks at H3 K9, for example, are thought to be transcriptionally suppressing and activating, respectively, whereas the presence of either mark at the K4 position is considered to be transcriptionally activating.[5–7] The placement of a PTM can therefore modulate gene expression not just through the recruitment of transcription factors, or the physiochemical change to the residue, but by preventing the placement of other PTMs at that same position.

### 3.1.2 KDM3A

KDM3A is a JmjC domain-containing demethylase acting at H3 K9Me$_{1/2}$, the demethylation activity of which appears to improve in the presence of H3 K4Me$_3$.[8,9] Recent peptide studies with histone fragment peptides (conducted by *Dr Rachel Shamo-Schiller*) have indicated KDM3A may have further functionality.[9] Incubation of KDM3A with a H3$_{1\text{-}15}$ peptide containing K4Me$_3$ and K9Ac substitutions under 'demethylation' reaction conditions (100 μM 2OG, 100 μM ascorbate, 60 μM Fe(II), 1 hr, 37 °C) resulted in the unexpected observation of a +16 Da peak by matrix assisted laser desorption

ionisation-time of flight mass spectrometry (MALDI-TOF MS) (see **Figure 3.1 A**).[9] The +16 Da mass shift indicated addition of a hydroxyl group to the peptide (see **Figure 3.1 B**), which is consistent with the established hydroxylation-based mechanism of JmjC-catalysed demethylation, as discussed in **Section 1.4.2**.[10] Previous work had revealed the potential for JmjC KDMs to catalyse the formation of stable alcohol products.[11] Similarly conducted KDM3A 'demethylation' experiments using $H3_{1-15}$ peptides containing either $K4Me_3$ or K9Ac substitutions did not result in an observed mass shift by MALDI-TOF MS.[9] Absence of a -42 Da mass shift indicated no demethylation had occurred, suggesting hydroxylation took place at the K9Ac position and not at $K4Me_3$.[9] Hydroxyacetylation was further verified by repeating KDM3A 'demethylation' experiments using a $H3_{1-15}$ peptide containing $K4Me_3$ and deuterated K9Ac. In this instance a +15 mass shift was observed, consistent with the replacement of a deuterium atom with a hydroxide group.

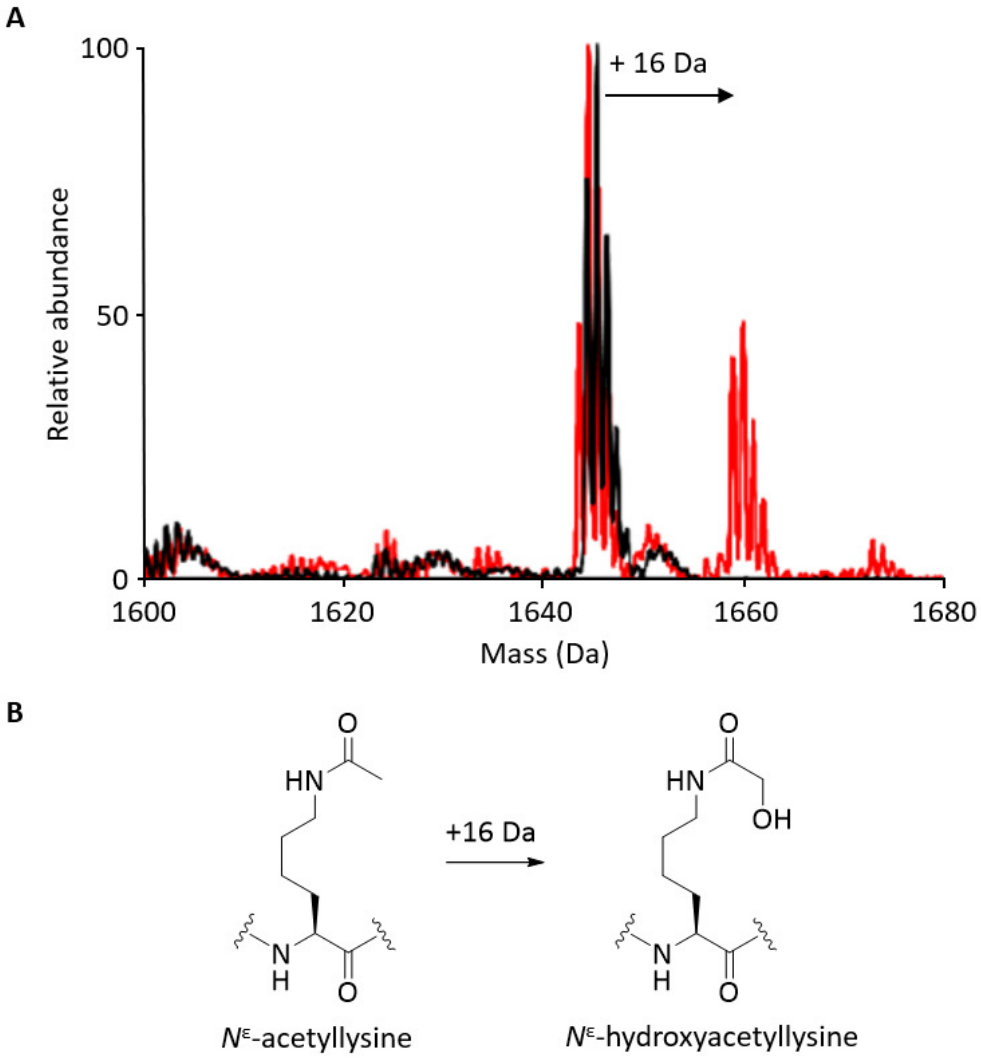



***Figure 3.1**: **A**- MALDI-TOF MS spectrum of a $H3_{1-15}$ K4Me3, K9Ac peptide (10 μM) following incubation with KDM3A (2 μM) for 1 hour at 37 °C under 'demethylation' reaction conditions. Reaction containing KDM3A is shown in red, and the no enzyme control in black. **B**- Chemical structures of* $N^{\varepsilon}$*-acetyllysine (KAc) and the proposed +16 Da* $N^{\varepsilon}$*-hydroxyacetyllysine (KAcOH) product. Figure reproduced with permission.*[9]

## 3.2 Objectives

This chapter describes the production of a set of peptides containing KAc, and KAcOH residues. AlphaScreen competition assays were then used to compare the binding affinity of these peptides to BRD4(1) as a model for acetylated lysine binding proteins.

*The synthesis of* $N^{\alpha}$*-Fmoc-protected hydroxyacetylated lysine was conducted by Roman Belle.*

## 3.3 Synthesis of peptides containing hydroxyacetylated lysine marks

In order to begin to investigate the role putative PTM KAcOH might play in gene expression, the effect of KAc hydroxylation on binding to a transcription factor was investigated. The 'model' reader protein BRD4(1) binds tightly to $H4_{1\text{-}20}$ (K5, 8, 12, 16)$Ac_4$ peptide, with a $K_D$ of 3.7 μM.[12] Comparing the binding affinity of BRD4(1) to variants of the $H4_{1\text{-}20}$ (K5, 8, 12, 16)$Ac_4$ peptide containing different combinations of KAcOH substitutions for KAc residues could therefore provide insight into the potential role of KAcOH in epigenetic regulation.

### 3.3.1 Design of hydroxyacetylated $H4_{1\text{-}20}$ peptides

Previously conducted time-resolved fluorescence energy transfer (TR-FRET) experiments on BRD4(1) binding to variously acetylated $H4_{1\text{-}20}$ peptides had led to the conclusion that both the location and number of KAc marks influences binding.[13] Singly acetylated peptides bound tighter than unacetylated peptide, multiply acetylated peptides bound tighter than singly acetylated peptides, and peptides containing four KAc substitutions bound strongest of all.[13] $H4_{1\text{-}20}$ K5Ac was the tightest binding singly acetylated peptide, and $H4_{1\text{-}20}$ (K5, 8)$Ac_2$ was the tightest binding doubly acetylated peptide to BRD4(1).[13]

Three peptide constructs were therefore selected for substitution with KAcOH, containing one, two or four KAcOH substitutions: $H4_{1\text{-}20}$ K5AcOH, (K8, 12, 16)$Ac_3$; $H4_{1\text{-}20}$ (K5, 8, 12, 16)$AcOH_2$, (K12, 16)$Ac_2$; and $H4_{1\text{-}20}$ (K5, 8, 12, 16)$AcOH_4$.

### 3.3.2 Peptide synthesis

The three abovementioned KAcOH-substituted peptides, as well as a $H4_{1\text{-}20}$ (K5, 8, 12, 16)$Ac_4$ reference peptide were prepared *via* fluorenylmethyloxycarbonyl chloride (Fmoc)-mediated solid phase peptide synthesis (SPPS) using a Liberty Blue Automated Microwave Peptide Synthesiser. $N^{\alpha}$-Fmoc-protected KAcOH was synthesised by *Roman Belle*.

SPPS occurs through the sequential addition of amino acids to a growing peptide attached to an insoluble resin. For this experiment the carboxy (*C*)-terminal K20 residue was attached to the (Rink Amide-4-methylbenzhydrylamine) resin by an acid-cleavable linker, before systematic addition of activated $N^{\alpha}$-Fmoc-protected *N*-terminal residues. By protecting the *N*-termini of each new amino acid with base-labile Fmoc it was possible to avoid unwanted over-addition of residues.

Each round of amino acid addition began with the cleaving of the $N^{\alpha}$-Fmoc of the growing peptide by addition of a DMF:piperidine 80:20 solution. The resin was washed and a new Fmoc-protected amino acid was added. Peptide coupling was facilitated by addition of a solution containing *N*,*N*-diisopropylcarbodiimide (DIC) and ethyl (hydroxyimino)cyanoacetate (OxymaPure®) in dimethylformamide (DMF), resulting in the activation of the Fmoc-protected amino acid carbonyl. Coupling was further accelerated through irradiation of the reaction mixture with microwaves. The mechanism for amino acid activation and peptide coupling is presented in **Figure 3.2**. Reaction conditions are provided in **Materials and Methods 8.3.1**.

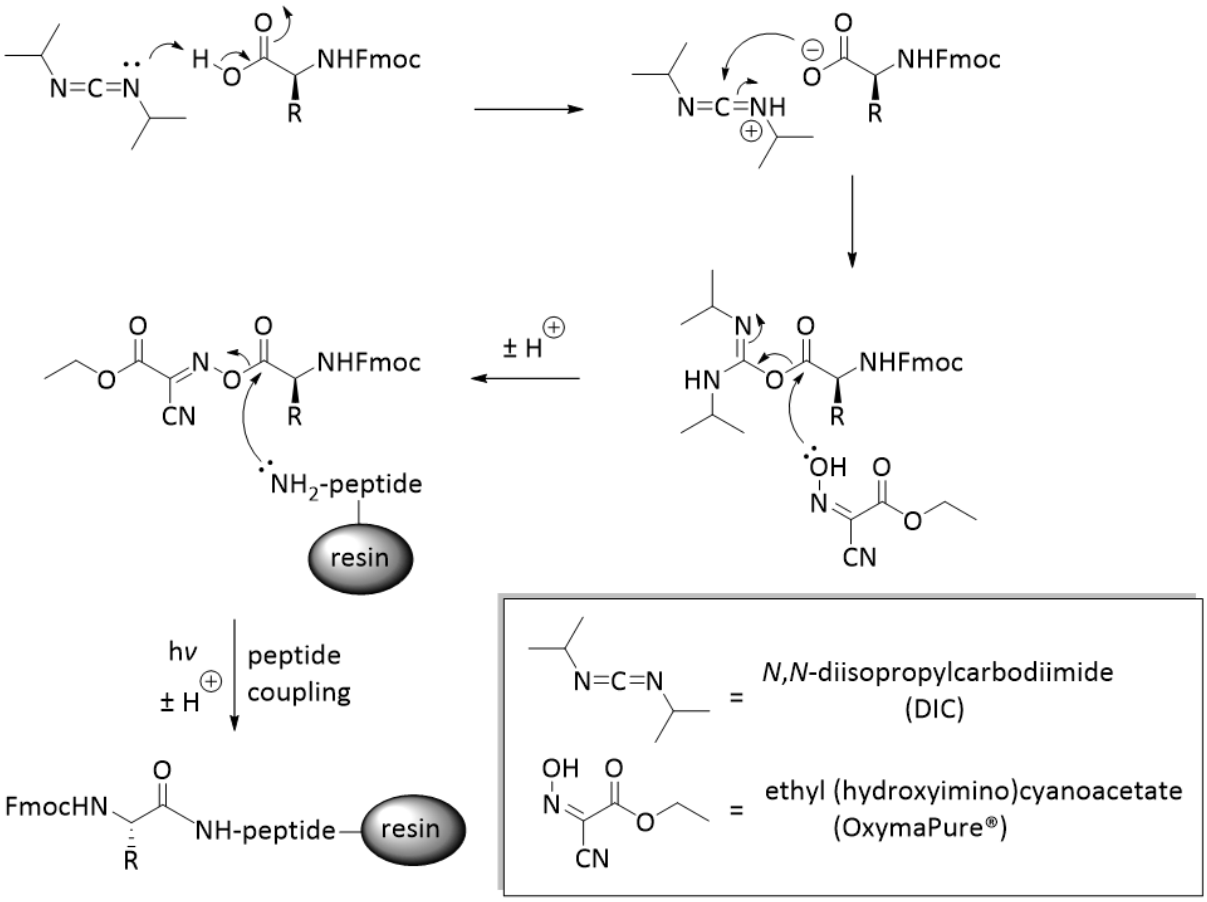



***Figure 3.2**: Mechanism for the activation of* N$^{\alpha}$*-Fmoc-protected amino acids by DIC and OxymaPure®, and the subsequent coupling to a peptide* N*-terminus. Amino acid carboxylic acid activation begins with acylation of DIC and the formation of an* O*-acylurea intermediate. Attack by the hydroxyl group of OxymaPure® produces the activated amino acid. Nucleophilic attack by the deprotected* N*-terminal peptide amine results in peptide coupling and the reformation of OxymaPure®.*

As the peptide is bound to a solid resin, it is possible to wash excess reagents such as unreacted amino acids from the reaction vessel between each step without losing product. The selectivity of the coupling reaction was ensured by the incorporation of acid-labile protecting groups on the side-chains of all nucleophilic amino acids. Once all required residues had been added to the peptide a final Fmoc deprotection step was carried out. DMF, and all insoluble reagents were removed from the resin by washing with diethyl ether, before drying under vacuum.

The peptide was cleaved from the resin by treatment with a solution containing 92.5% (v/v) trifluoroacetic acid (TFA), 2.5% (v/v) triisopropylsilane, and 2.5% $H_2O$ for 3 hours (RT). As all nucleophilic residues were protected using acid-labile protecting groups this TFA cleavage step also served to remove the remaining protecting groups. Cleavage of peptides from RINK-amide resin using TFA resulted in *C*-terminal amides.

While each round of peptide coupling was highly efficient (>95% conversion), lower MW peptides can still be produced during synthesis. The full length 20-mer peptides were separated from the lower MW products by high performance liquid chromatography (HPLC). MALDI-TOF MS verified the identity of the purified $H4_{1\text{-}20}$ peptides and confirmed that no shorter peptides contaminated the sample. MALDI traces of the purified peptides are shown in **Figure 3.3**. Experimental details for the purification of peptides are provided in **Materials and Methods 8.3.2**.

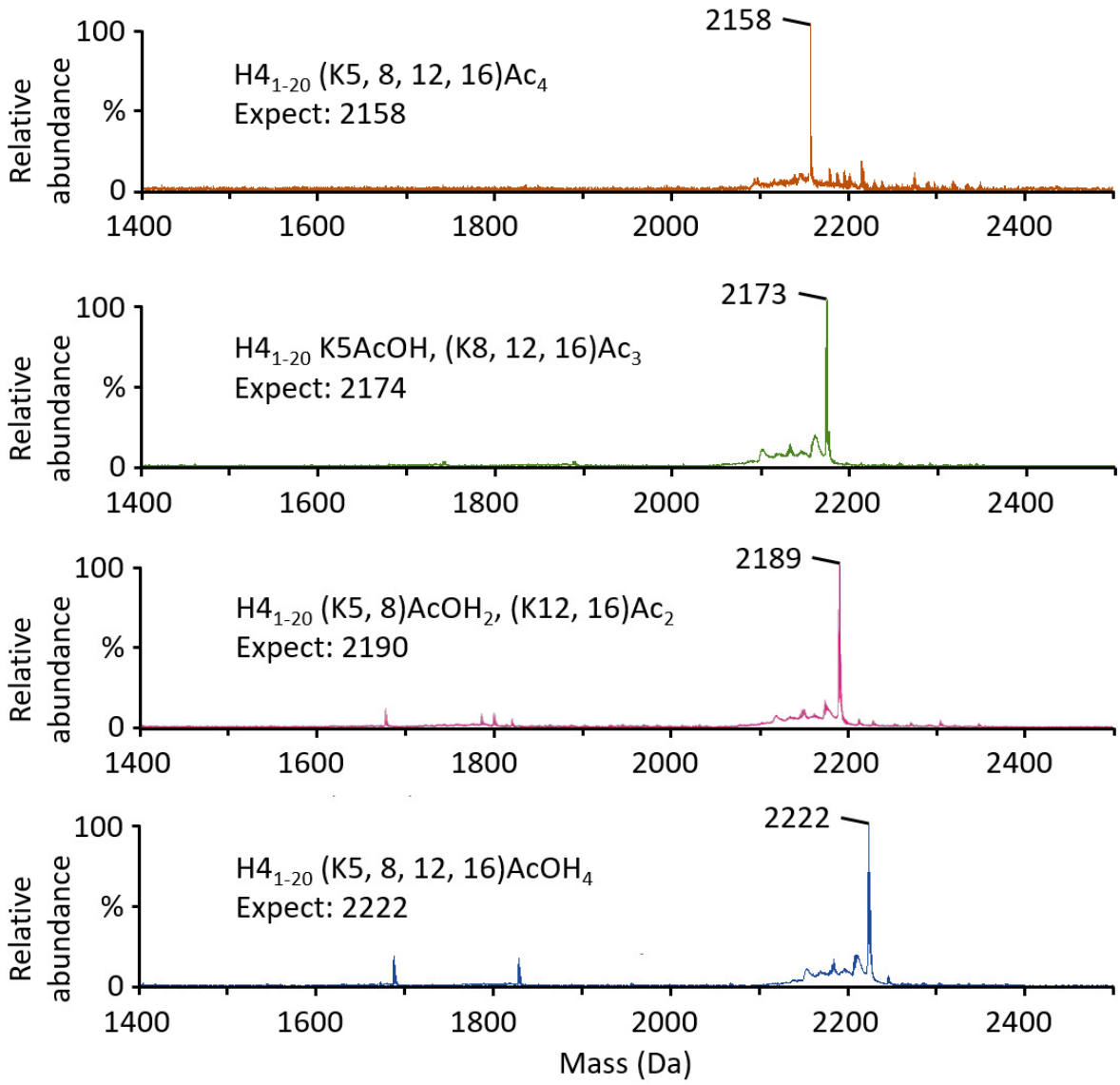



***Figure 3.3**: MALDI-TOF MS spectra of four purified H4$_{1-20}$ peptides containing KAc or KAcOH substitutions.*

Peptide concentrations were verified by $^1$H nuclear magnetic resonance (NMR) by comparing the concentration of each peptide against an internal standard (3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid; TMSP-$d_4$). The strengths of the integrated nine proton TMSP-$d_4$ $(CH_3)_3$ peak was compared to the distinct signal from the KAc $CH_3$ groups, or KAcOH $CH_2$ groups of the peptides.

## 3.4 AlphaScreen experiments

Using the AlphaScreen displacement assay optimised in **Section 2.4.3** the binding affinity of BRD4(1) to a range of $H4_{1-20}$ peptides containing different substitution patterns of KAc and KAcOH marks was tested. As a different batch of BRD4(1) was used in these experiments, a new dose-response curve was produced, as described in **Section 2.3.2**. 10 nM BRD4(1), and 15.6 nM peptide with 10 µg/mL AlphaScreen beads were again established as optimal FACs (see **Figure 3.4**).

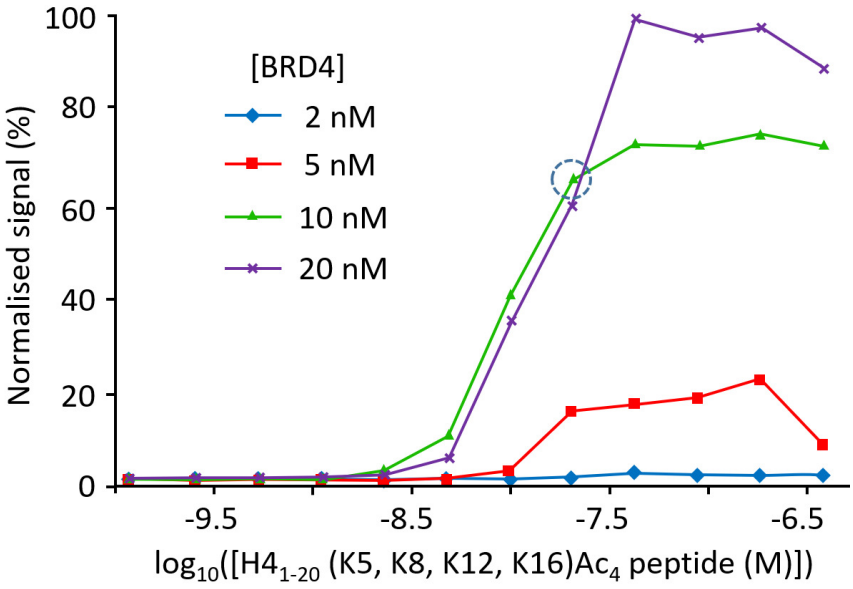



***Figure 3.4**: AlphaScreen dose-response curves for a range of biotin-tagged $H4_{1-20}$ (K5, 8, 12, 16)$Ac_4$ peptide, and his-tagged BRD4(1) concentrations. The normalised signal, presented as a percentage of the maximum signal detected is plotted against the $log_{10}$([$H4_{1-20}$ (K5, 8, 12, 16)$Ac_4$ peptide (M)]). The concentrations selected for use in future experiments are circled (10 nM BRD4(1) and 15.6 nM peptide).*

Following the procedure described in **Section 2.5.2** an AlphaScreen assay was conducted using a 14-step twofold dilution series of $H4_{1-20}$ KAc/KAcOH modified peptides (1250-0.076 µM FAC) diluted into AlphaScreen buffer. Results were plotted as $log_{10}$([inhibitor (M)]) *vs.* % inhibition, and a variable slope (four parameter) curve was fitted using GraphPad prism (see **Figure 3.5**).[14]

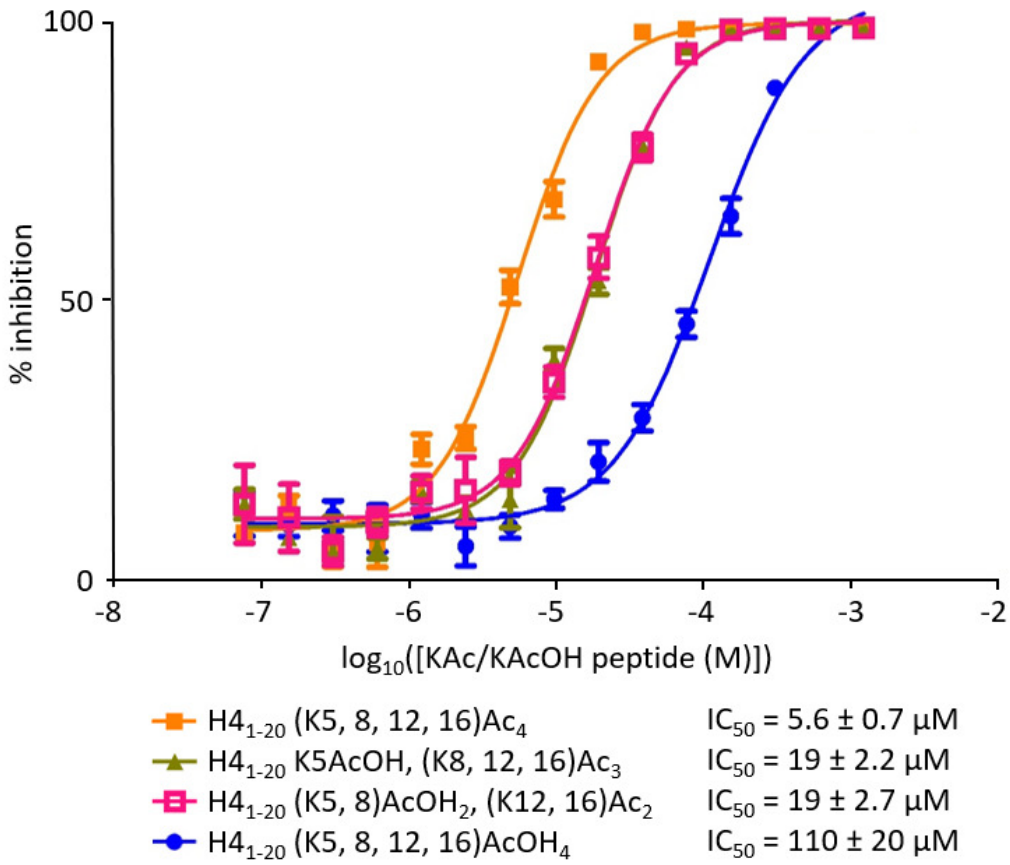



***Figure 3.5**: Line graph displaying AlphaScreen data for a range of H4$_{1-20}$ peptides containing different combinations of KAc and KAcOH marks. IC$_{50}$ values are provided for each compound ± standard error from a minimum of three separate experiments.*

The unsubstituted H4$_{1-20}$ (K5, 8, 12, 16)Ac$_4$ peptide displaced its biotinylated counterpart with an IC$_{50}$ value of 5.6 ± 0.7 µM. A previously reported AlphaScreen-derived IC$_{50}$ value for a Y-H4$_{1-20}$ (K5, 8, 12, 16)Ac$_4$ peptide against BRD4(1) was in close agreement with this figure (5.4 µM).[15]

Peptides containing one or two KAcOH groups replacing KAc groups resulted in a >3-fold increase in IC$_{50}$ relative to the H4$_{1-20}$ (K5, 8, 12, 16)Ac$_4$. Replacement of all four KAc residues with KAcOH residues resulted in a 20-fold increase in IC$_{50}$ from 5.6 ± 0.7 µM to 110 ± 20 µM.

## 3.5 Discussion

Comparison of the AlphaScreen-derived $IC_{50}$ binding results for $H4_{1\text{-}20}$ (K5, 8, 12, 16)$AcOH_4$ (110 ± 20 μM) *vs.* $H4_{1\text{-}20}$ (K5, 8, 12, 16)$Ac_4$ (5.6 ± 0.7 μM) clearly shows that this putative PTM is not well tolerated by the bromodomain BRD4(1). Interestingly, the di-KAcOH-substituted peptide exhibited a similar $IC_{50}$ against BRD4(1) to the mono-KAcOH-substituted peptide (18.6 μM, each). BRD4(1) has previously been shown to be able to bind up to two acetylation marks, separated by two or three residues, simultaneously (*e.g.* K5Ac + K8Ac, and K12Ac + K16Ac).[12] It therefore appears likely that replacement of the K5Ac residue with K5AcOH is sufficient to interrupt binding of BRD4(1) to the (K5, 8)[$Ac_2$] binding site, with the residual binding strength arising from the (K12, 16)$Ac_2$ residues. The weakest binding of the peptides tested to BRD4(1) was for the $H4_{1\text{-}20}$ (K5, 8, 12, 16)$AcOH_4$ construct, indicating BRD4(1) binding to the second (K12, 16)[$Ac_2$] binding site was also disrupted.

The results of these experiments provide a proof of principle that hydroxylation of KAc residues could influence bromodomain binding to histones. Furthermore, these results indicate a potential use for proteins capable of hydroxylating lysine acetyl groups, such a KDM3A, as a 'PTM silencer', *i.e.* reducing the ability of acetylation marks to recruit bromodomains such as BRD4. Whether or not this silencing effect applies to other reader proteins is yet to be determined. Crotonylation, for example, another recently discovered PTM, can be recognised by some KAc 'reader' domains (*e.g.* YEATS domains), but not others (*e.g.* most bromodomains), and could be further modified.[16,17] Further experiments using methods such as those detailed above could be used to investigate the

affinities of other 'reader' proteins to KAcOH marks, thereby providing a better understanding of the utility of KAcOH in the domain of transcriptional regulation.

## 3.6 References


1. Martin, C. & Zhang, Y. Mechanisms of epigenetic inheritance. *Curr. Opin. Cell Biol.* **19,** 266–272 (2007).
2. Tan, M. *et al.* Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146,** 1016–1028 (2011).
3. Huang, H., Lin, S., Garcia, B. A. & Zhao, Y. Quantitative proteomic analysis of histone modifications. *Chem. Rev.* **115,** 2376–2418 (2015).
4. Azevedo, C. & Saiardi, A. Why always lysine? The ongoing tale of one of the most modified amino acids. *Adv. Biol. Regul.* **60,** 144–150 (2016).
5. Stunnenberg, R. *et al.* H3K9 methylation extends across natural boundaries of heterochromatin in the absence of an HP1 protein. *EMBO J.* **34,** 2789–2803 (2015).
6. Kim, J. H. & Workman, J. L. Histone acetylation in heterochromatin assembly. *Genes Dev.* **24,** 738–740 (2010).
7. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442,** 86–90 (2006).
8. Yamane, K. *et al.* JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125,** 483–495 (2006).
9. Rachel Shamo-Schiller. Investigating the inhibitor and substrate diversity of the JmjC histone demethylases. (University of Oxford, 2016).
10. Bollinger, J. M., Price, J. C., Hoffart, L. M., Barr, E. W. & Krebs, C. Mechanism of taurine: α-Ketoglutarate dioxygenase (TauD) from *Escherichia coli*. *Eur. J. Inorg. Chem.* **2005,** 4245–4254 (2005).
11. Hopkinson, R. J. *et al.* Is JmjC oxygenase catalysis limited to demethylation? *Angew. Chem. Int. Ed.* **52,** 7709–7713 (2013).
12. Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149,** 214–231 (2012).
13. Jung, M. *et al.* Affinity map of bromodomain protein 4 (BRD4) interactions with the histone H4 tail and the small molecule inhibitor JQ1. *J. Biol. Chem.* **289,** 9304–9319 (2014).
14. GraphPad Software. GraphPad Prism version 5.04 for Windows (2010).
15. Philpott, M. *et al.* Bromodomain-peptide displacement assays for interactome mapping and inhibitor discovery. *Mol. Biosyst.* **7,** 2899 (2011).
16. Andrews, F. H. *et al.* The Taf14 YEATS domain is a reader of histone crotonylation. *Nat. Chem. Biol.* **12,** 396–8 (2016).
17. Flynn, E. M. *et al.* A subset of human bromodomains recognizes butyryllysine and crotonyllysine histone peptide modifications. *Structure* **23,** 1801–1814 (2015).

# Chapter 4: Investigating the Interactions between TRIM24 and the Nucleosome

## 4.1 Introduction

### 4.1.1 Peptides and nucleosomes in epigenetic research

As described in previous chapters, peptide fragments have commonly been used as models for much larger proteins and complexes. As peptides can be synthetically produced in large quantities with, in principle, any sequence and combination of PTMs, they are often favoured over the more difficult to produce full-length histones or nucleosomes for *in vitro* studies. Peptides have often been considered as a good model for the flexible histone tail, and many crystal structures of epigenetic proteins in complex with histone peptides have been obtained. Peptides are, however, unable to mimic the global histone environment, which may be influenced by its proximity to the ordered regions of the histone core, neighbouring histones, nucleosomal DNA, and other proteins.

Crystallography is a powerful tool for visualising protein–protein interactions at atomic resolution. Nucleosomes are considerably larger, more complex and harder to produce than peptides, but crystal structures have been obtained of nucleosomes in complex with a number of transcription factors.[1–5] To date, no crystal structures are available for proteins that interact with the H3 tail in the context of an intact nucleosome, despite the central role of such proteins in epigenetic regulation.

### 4.1.2 TRIM24

The epigenetic 'reader' protein TRIpartate Motif-containing 24 (TRIM24), is a chromatin regulator and has been found to be involved in the activation of oestrogen-dependent

genes associated with cellular proliferation through binding to chromatin and oestrogen receptor α.[6] Subsequently, TRIM24 was implicated in tumour development, and high levels of TRIM24 expression correlate with a low rate of survival in breast cancer patients.[6] TRIM24 is a 117 kDa multidomain protein consisting of a RING domain, two B-box domains, a coiled-coil region, a PHD, and a BRD.[7] Peptide-based AlphaScreen binding studies on constructs of TRIM24 containing only the PHD and BRD indicate that TRIM24 is able to simultaneously bind to a peptide with three H3 marks (unmethylated K4, K9Me$_3$ and K18Ac) with a 2 nM $K_D$.[8,9] ITC experiments indicate that the TRIM24 BRD alone is able to bind an H3$_{1-27}$ peptide containing only one modified residue, K18Ac, with a 21 nM $K_D$, the strongest interaction currently identified between a BRD and a singly modified binding target.[8,10]

There are currently no crystal structures available for TRIM24 binding to a peptide *via* its PHD and BRD, however, the closely related TRIM33 is predicted to bind to modified H3 in a similar manner.[8] **Figure 4.1** displays the binding mode of TRIM33 to H3$_{1-27}$ K4, K9Me$_3$ and K18Ac peptide.

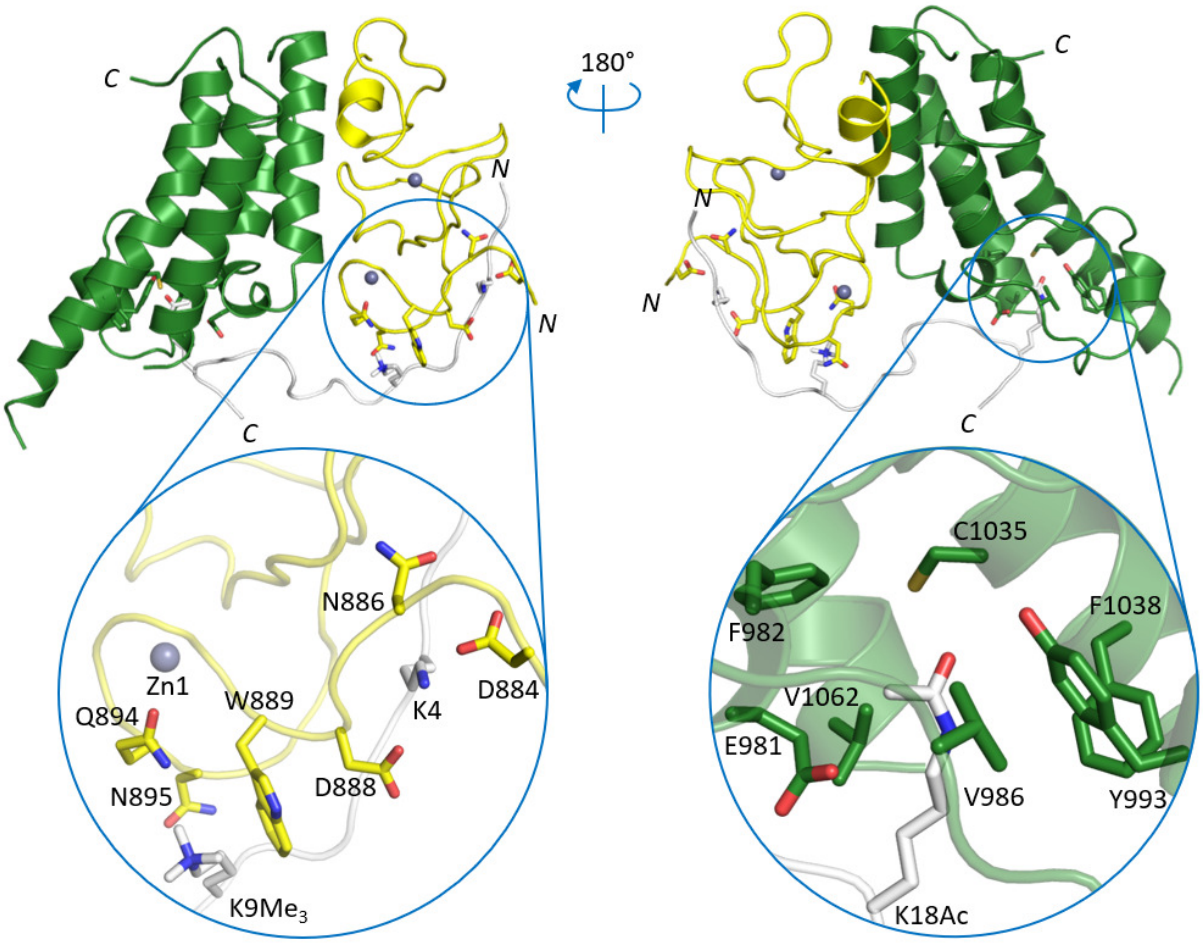



***Figure 4.1**: Views from a crystal structure of a TRIM33–H3$_{1-27}$ K9Me$_3$, K18Ac complex (PDB ID: 3U5O). The PHD domain is displayed in yellow, the BRD in dark green, the H3 peptide in white and the zinc ions in grey. H3 K4, K9Me$_3$ and K18Ac are presented as sticks, along with TRIM33 residues that interact with them. Areas of interest have been magnified. **Left**- PHD domain of TRIM33 accommodating K4 and K9Me$_3$. **Right**- K18Ac binding in a polar, hydrophobic pocket located in the TRIM33 BRD.*

## 4.2 Objectives

The work described in this chapter aimed to investigate the feasibility of crystallising TRIM24 in complex with an intact nucleosome by characterising the interactions between them. Initial peptide studies were employed to identify modifications to H3 that enhance TRIM24 binding and are easily incorporated into nucleosomes. This chapter details the production of a modified nucleosome and analysis of its binding to TRIM24 using non-denaturing electrospray ionisation mass spectrometry, small-angle X-ray scattering (SAXS) and ITC.

*TRIM24 protein was supplied by Dr Angelina Sekirnik. Non-denaturing MS data were generated by Dr Idlir Liko and Dr Nisha Patel.*

## 4.3 H3 sK18Ac nucleosome production

The observed tight binding of TRIM24 to $H3_{1-27}$ K18Ac peptide identified it as a feasible starting point for nucleosome complex crystallography. A flowchart surmising the processes used to produce homogeneous post-translationally modified nucleosomes can be found in **Figure 4.2**.

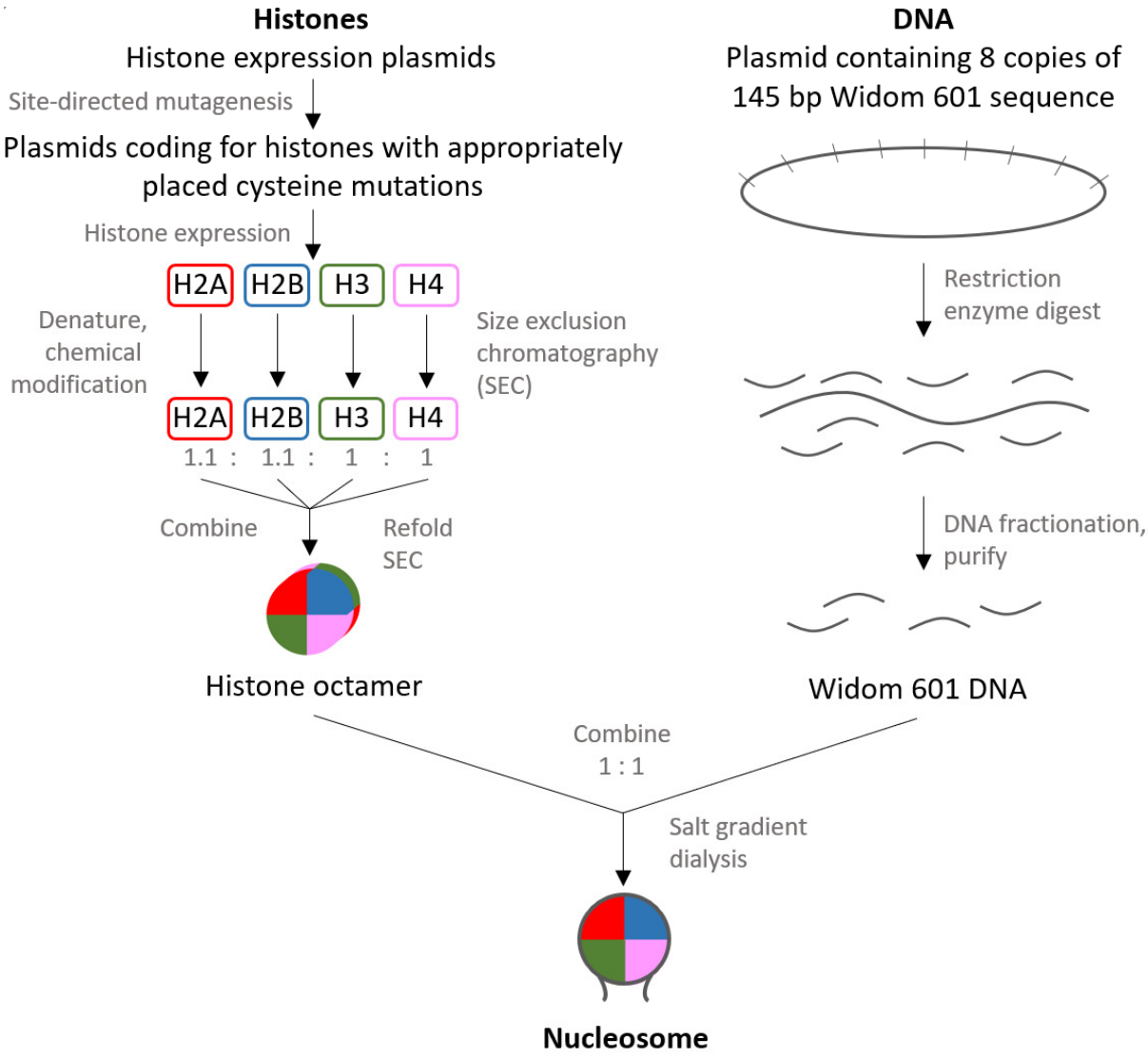



***Figure 4.2**: Flowchart of methods used to produce, modify, purify and assemble components of the nucleosome.*

### 4.3.1 Histone production

Each of the canonical histones; H2A, H2B, H3 and H4 (*Xenopus laevis*) were individually overexpressed in *E. coli* and extracted from exclusion bodies. Histones were purified by SEC, dialysed into deionised water, lyophilised, and stored at -80 °C. Purity was assessed by 16% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and liquid chromatography-mass spectrometry (LC-MS). Details of histone expressions and purifications are provided in **Materials and Methods 8.5.7** and **8.6.1**, respectively.

### 4.3.2 Methods for incorporating post-translational modifications into proteins

A number of different processes can be utilised in order to place a PTM, such as an acetyl group, onto a protein, each with their own benefits and caveats.

• Enzymatic incorporation of the PTM. *In vitro* use of HATs to place acetylation groups on lysine residues could produce modified histones and nucleosomes directly. Issues associated with this method include; the requirement for producing large amounts of HAT, the promiscuity of most HATs making acetylation at one specific desired site troublesome, problems driving the reaction to completion, and difficulties separating the desired product from unreacted substrate and alternatively acetylated product.[11]

• Translational incorporation of PTMs. The availability of *E. coli* strains that read through the UAG stop codon allows for codon reassignment.[12] Due to the availability of an orthogonal UAG KAc-tRNA synthetase/tRNA pair it is now possible to code at the genomic level for homogeneous incorporation of a desired PTM at any site in a protein.[13] Disadvantages of this technique include the high cost associated with producing large quantities of protein due to the large amounts of KAc that need to be supplemented to the bacterial growth media.

• Chemical ligation of a peptide to a protein. Multiple modified amino acids can be incorporated into a synthetic peptide, which could, in turn, be attached to the terminus of a protein *via* chemical ligation.[14] As with translational incorporation of PTMs, this method can be costly to use on a large scale due to the price of reagents such as the requisite resin and modified amino acids.

• Synthetic incorporation of PTMs to proteins at cysteine residues. While it is not currently possible to chemically modify a specific lysine residue in a protein containing more than one lysine, analogues of modified lysine residues can be introduced by reaction with cysteine. The thiol group of cysteine is a soft nucleophile with unique reactivity compared to other amino acids, allowing selective reaction.[15] Using site-directed mutagenesis (SDM) (**Materials and Methods 8.4.3**) cysteine residues can be placed at any position in a protein's sequence, then chemically reacted to form analogues of natural PTMs of lysine residues. Examples of PTM analogues that can be produced by this method include acetylated (see **Figure 4.3**) and methylated lysine residues, the latter of which is described in more detail in **Section 5.3.1**.[15,16]

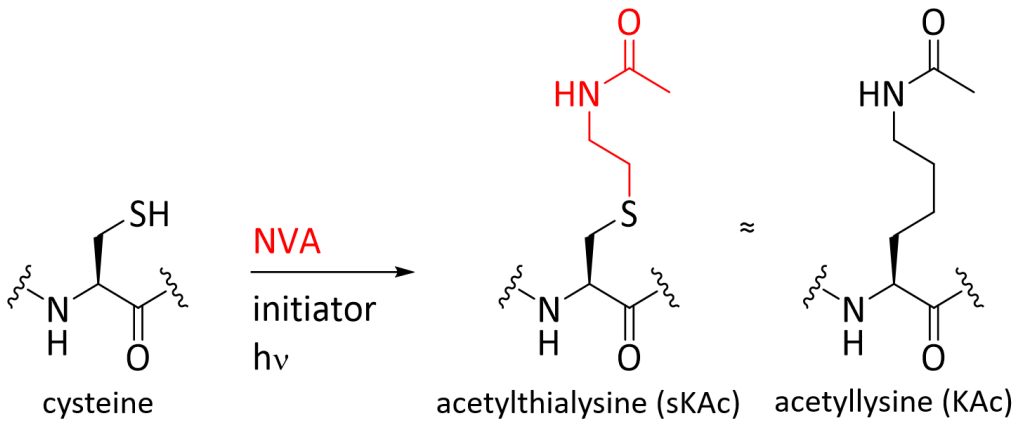



***Figure 4.3**: Scheme showing the conversion of cysteine to sKAc, a homologue of KAc. Reaction conditions: 500 μM cysteine sites, 75 mM* N*-vinylacetamide, radical initiator, hν, 25 °C, 2-3 hours.*[16] *The reaction occurs* via *a radical initiated thiol-ene coupling.*[16]

While these efficient reactions can be used to produce homogeneously modified proteins, this approach does have a number of downsides. First, in order for complete conversion of cysteine residues, the protein often must be unfolded. Second, this method results in the modification of all accessible cysteine residues present, and therefore might unfavourably alter proteins that contain cysteine residues important for the protein's structure (*i.e.* disulfide) or function (*i.e.* active sites). Finally, the products of these reactions contain a thioether at the γ-position instead of the native methylene bridge. However, as this substitution is somewhat isosteric and separated from the $N^{\varepsilon}$ atom by two carbon atoms, the lysine analogues are predicted to have similar biochemical properties to their natural analogues. The lysine analogues have been shown to be recognised by antibodies raised against natural PTM analogues, and can be utilised as substrates by appropriate enzymes such as demethylases or acetylases.[15,16]

In the case of lysine acetylation on H3, several disadvantages of chemical modification at cysteine don't apply. As the standard method for producing nucleosomes from recombinant histones involves the denaturation and subsequent refolding of histones, reacting H3 in its unfolded state is not an issue. Histones have a low abundance of cysteine residues, with H3 (*Xenopus laevis*) containing only a single cysteine at position 110, which can be removed through SDM to alanine or serine to avoid conflict with this method's incompatibility with native cysteine residues. Crystal structures of nucleosome constructs containing H3 C110A substitutions have been obtained, showing this substitution has no significant effect on nucleosome structure.[17,18]

Due to the high yield, high homogeneity of the product, and the ability to modify histones on a >10 mg scale, chemical modification of histones by reaction at a cysteine residue was selected as the method for incorporating an acetylated lysine analogue.

### 4.3.3 Chemical incorporation of an acetyl group to K18

An H3 C110A construct containing a K18C substitution was obtained by SDM (**Materials and Methods 8.4.3**), followed by expression and purification as described in **Materials and Methods 8.5.7** and **8.6.1**, respectively. The purified H3 variant was dissolved in sKAc reaction buffer, then reacted with a 150-fold excess of *N*-vinylacetamide (NVA) in the presence of UV light and a radical initiator (VA-044), (3 hr, RT) with occasional mixing, forming the $N^{\varepsilon}$-acetylthialysine (sKAc) product (see **Figure 4.3**).

The reaction was monitored by LC-MS until completion and validated by Western blot using a non-specific polyclonal rabbit α-KAc 1° antibody (ab61257) and a polyclonal goat α-rabbit 2° antibody conjugated to horseradish peroxidase (ab6721) (see **Figure 4.4**). A signal was observed only in lanes containing reacted H3 K18C proteins, and not in lanes that only contained starting material, indicating the formation of the desired H3 sK18Ac product. The sK18Ac modified H3 protein was separated from the small molecule contaminants by buffer transfer *via* PD-10 desalting column (GE Healthcare Life Sciences) into MQ water, then lyophilised. Typical yields for protein recovery were 75%.

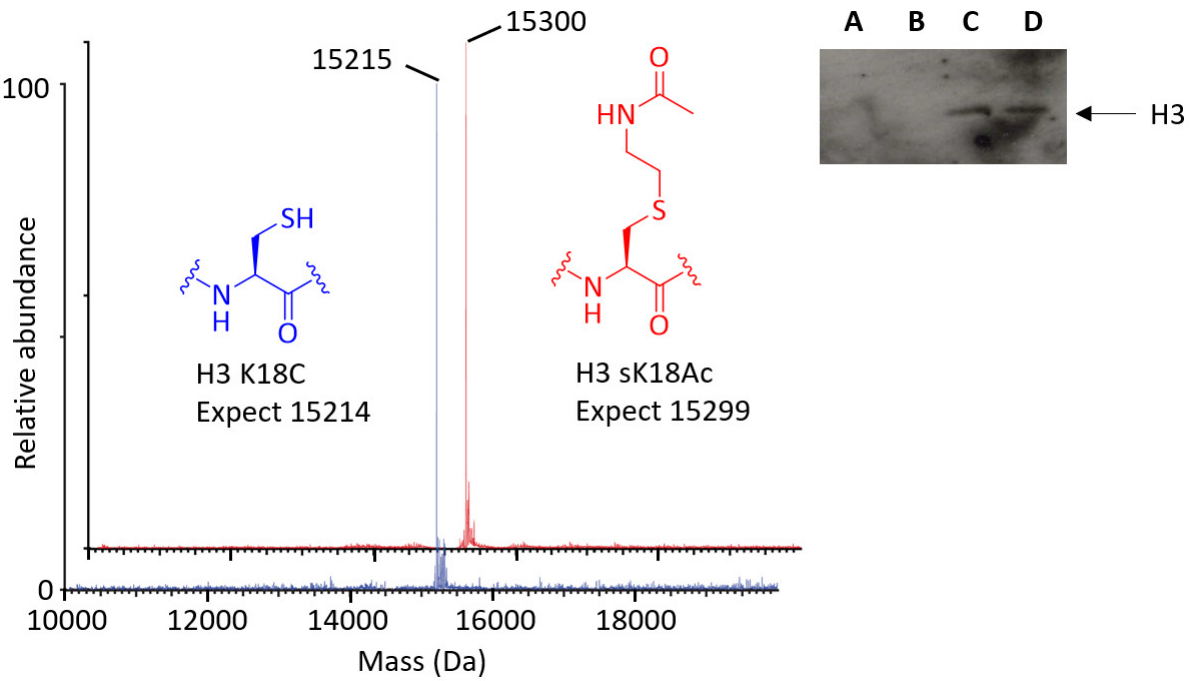



***Figure 4.4**: Analyses of H3 sK18Ac formation. **Left**-Deconvoluted spectra of H3 K18C before (blue) and after (red) reaction with NVA. A mass increase of 85 Da is observed, consistent with addition of NVA to H3. **Right**- Western blot analysis of H3 before and after reaction with NVA using polyclonal rabbit α-KAc 1° antibody (abcam-ab61257) and polyclonal goat α-rabbit horseradish peroxidase 2° antibody (abcam-ab6721). **A** and **B**- Before reaction; unmodified H3 K18C loaded at 1 µg and 5 µg, respectively. **C** and **D**- After reaction; H3 sK18Ac loaded at 1 µg and 5 µg, respectively.*

### 4.3.4 Octamer folding

H2A, H2B, H4 and modified H3 were combined in a 1.1:1.1:1:1 ratio under denaturing conditions (unfolding buffer), then dialysed into refolding buffer over the course of 48 hours with stirring. During the dialysis, histones fold together into H2A–H2B dimers and $(H3–H4)_2$ tetramers that subsequently combine in a 2:1 ratio to form the histone octamer.[18] Octamer was separated from contaminants including excess histones and H2A–H2B dimer by SEC using a 120 mL S200 gel-filtration column (GE Healthcare). As the $(H3–H4)_2$ tetramer co-elutes with the histone octamer, excess H2A and H2B were used in this experiment to minimise the presence of any $(H3–H4)_2$ tetramer in the post-dialysis solution. An example fast protein liquid chromatography (FPLC) UV trace

displaying a typical elution profile for the octamer refolding mixture is shown in **Figure 4.5**. Buffer details are provided in **Materials and Methods 8.9.1**.

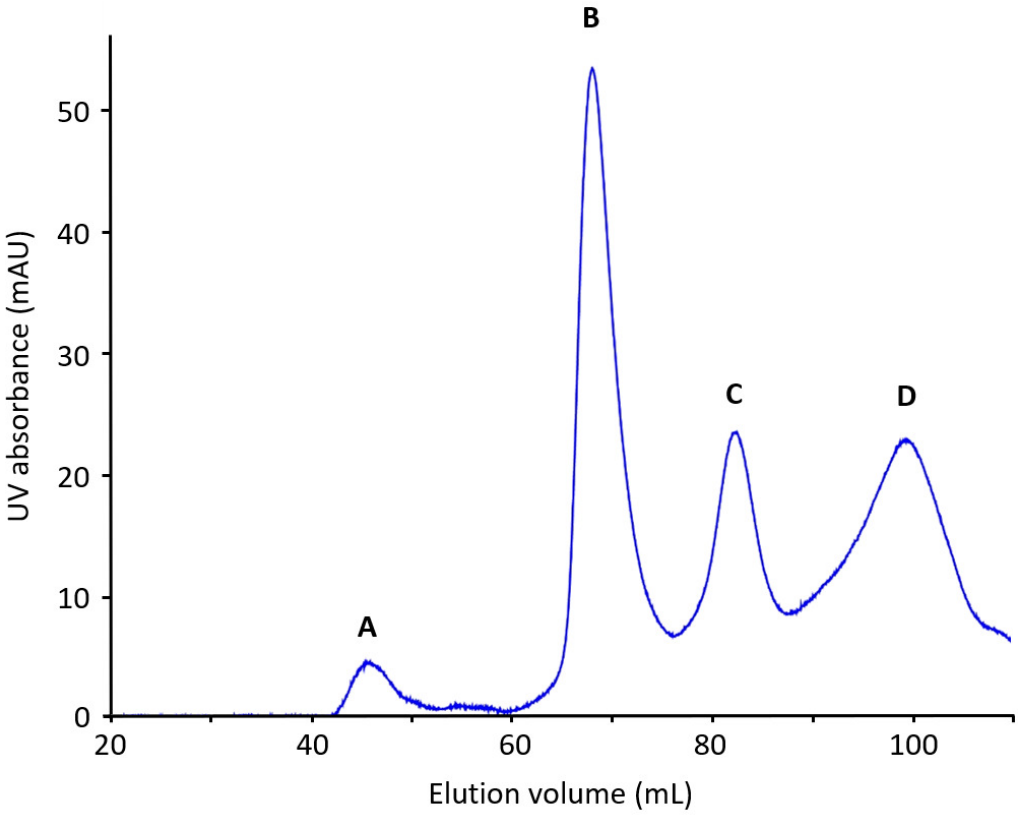



***Figure 4.5**: Elution profile of histone octamer dialysis mixture from a 120 mL S200 gel filtration column. **A**- High molecular weight aggregates elute between 40-50 mL; **B**- histone octamer at 68 mL; **C**- H2A–H2B dimer at 82 mL; **D**- monomeric histones between 90-105 mL. $(H3–H4)_2$ tetramer elutes at a similar volume to histone octamer (72 mL), so its presence is minimised through addition of an excess of H2A and H2B.*[19]

### 4.3.5 DNA production

#### Selecting a DNA sequence

When designing a nucleosome thought must be given to the DNA sequence selected for inclusion in the structure. Each nucleosome comprises a histone octamer wrapped with 145-147 bp of DNA. With four nucleotide bases to choose from (and not counting modified DNA bases) there are a total of $4^{147}$ possible DNA sequences that could potentially be incorporated into a nucleosome. Previous research has shown that the ability to crystallise nucleosomes, and the diffraction quality of those crystals depends

largely on the sequence of DNA used.[18] A number of highly positioning, tightly binding sequences have been identified, such as the naturally occurring alpha-satellite sequence, and the Widom 601 sequence, identified from a pool of randomly synthesised oligonucleotides.[20,21] All nucleosome complex structures determined prior to the commencement of this project incorporated the Widom 601 sequence in their nucleosomes. Therefore the 145 bp Widom 601 sequence was selected for the following described studies.[1–3] The nucleic acid sequence of the Widom 601 DNA used in these experiments is presented in **Materials and Methods 8.4.4**.

#### Large-scale DNA production

All buffer components used in the large-scale production of Widom 601 DNA are described in **Materials and Methods 8.4.5**.

A plasmid containing eight copies of the 145 bp Widom 601 sequence flanked by EcoRV restriction-sites was transformed into in HB 101 cells, a strain of *E. coli* deficient in the RecA DNA repair enzyme. A 7.2 L growth of transformed cells were incubated at 37 °C (24 hr) before harvesting by centrifugation (10000*g*, 8 min, 4 °C).

Cell pellets containing Widom 601 plasmids were resuspended in 400 mL alkaline lysis solution I (ALSI), then lysed by addition of 720 mL of ALSII with shaking, followed by incubation on ice (20 min). The resulting mixture was acidified by addition of 1.2 L ice cold ALSIII with swirling, then incubated on ice (20 min). The precipitant was removed by centrifugation (10000*g*, 30 min, 4 °C) followed by vacuum filtration through a 25 µm filter. The plasmid was precipitated with 0.52 volumes of isopropanol, then pelleted by centrifugation (10000*g*, 30 min, 20 °C). The supernatant was removed, and the pellet left to dry at RT overnight.

The plasmid was dissolved in 40 mL TE 10/50 and 2.0 mg RNase A was added. The mixture was then incubated overnight at 37 °C. The plasmid was precipitated again on addition of sodium chloride solution and PEG 6000 solution at a final concentration of 0.8 M and 16% (w/v), respectively, followed by incubation on ice (30 min). The plasmid was pelleted by centrifugation (3000*g*, 20 min, 4 °C) and the supernatant discarded.

The purified plasmid was dissolved in 30 mL TE 10/0.1 buffer by rocking at 37 °C overnight. The plasmid concentration was determined using a NanoDrop ND-1000 UV/Vis spectrophotometer (A = 260 nm, $\varepsilon = 2.31 \times 10^6$), and the purity of the sample was examined by agarose gel electrophoresis (see **Figure 4.6**). Stock solutions were added until the plasmid buffer contained 150 mM sodium chloride, 10 mM tris-HCl pH 8.0, 6 mM magnesium chloride, 1 mM DTT and 0.1 mM EDTA-NaOH pH 8.0. The Widom 601 inserts were excised from the plasmid on addition of 10 000 units of restriction enzyme (EcoRV) followed by incubation at 37 °C (20 hr). The progression of the digestion was monitored by agarose gel electrophoresis (see **Figure 4.6**), and the reaction was continued until no laddering was observed in the gel.

In order to precipitate the remaining empty plasmid vector, 0.192 volumes of a 4 M sodium chloride solution and 0.346 volumes of a 50% (w/v) PEG 6000 solution were added to the digestion mixture and the solution was incubated on ice (1 hr). The vector was precipitated by centrifugation (27000*g*, 20 min, 4 °C) and separated from the supernatant. The supernatant containing the Widom 601 sequence DNA fragments was poured into 4 volumes of ice cold ethanol and incubated on ice (20 min) resulting in the precipitation of the Widom 601 DNA fragments. The Widom 601 DNA was pelleted by

centrifugation (27000*g*, 20 min, 4 °C), the supernatant decanted, and the pellet allowed to dry at RT (~10 min).

The Widom 601 DNA was dissolved in 5 mL TE 10/0.1, then 2 mL phenol and 1 mL trichloromethane were added and the mixture was shaken vigorously. The aqueous layer was removed, then extracted from trichloromethane (1:1), twice. To the aqueous layer was added 0.1 volume of 3 M acetate buffer (sodium acetate-acetic acid) pH 5.2 and 3 volumes of ethanol. The solution was mixed until homogeneous then incubated at -20 °C (3 days). The precipitated Widom 601 sequence was pelleted by centrifugation (27000*g*, 30 min, 4 °C) and the supernatant was discarded. The pellet was dried at RT (20 min) then dissolved in nuclease-free $H_2O$. The purity of the final, purified DNA sequence was validated by agarose gel electrophoresis (see **Figure 4.6**) and stored at -20 °C.

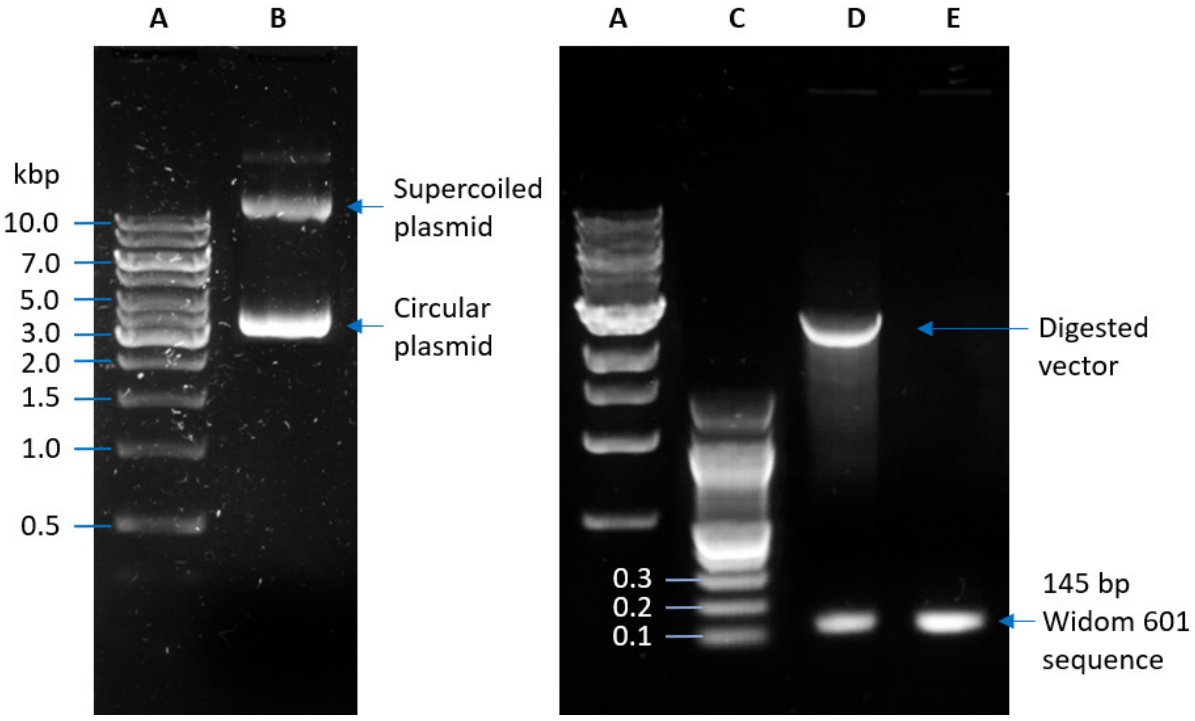



***Figure 4.6**: 1% (w/v) agarose gel displaying; **A**- 1 kbp DNA ladder (NEB), **B**- purified plasmid (3870 bp), **C**- 100 bp DNA ladder (NEB), **D**- plasmid following EcoRV digestion (2710 bp vector, 145 bp insert), **E**- purified Widom 601 sequence (145 bp).*

### 4.3.6 Nucleosome reconstitution

To a solution of 145 bp DNA in high salt refolding buffer (HSRB) was added one equivalent of the histone octamer. The sample was diluted to 5 μM and incubated at 4 °C for 30 minutes, then transferred to dialysis tubing (5 kDa molecular weight cut-off) and placed floating in 400 mL HSRB. The HSRB was slowly replaced with low salt refolding buffer (LSRB) at a rate of 1.5 mL/min with stirring over the course of 24 hours at 4 °C to ensure a smooth dialysis gradient. The remaining dialysis buffer was replaced with 400 mL LSRB and stirred for a final 3 hours before the mixture was extracted from the tubing and concentrated to 50 μM. Following the assumption that all DNA in the mixture was bound in nucleosomes, and that the DNA signal would dominate any contribution from the proteins, the concentration of the reconstituted nucleosome was determined by NanoDrop (A = 260 nm, $\varepsilon = 2.31 \times 10^6$). Nucleosome samples were stored at 4 °C.

It may be possible for DNA associating to the histone core to adopt multiple different conformations. These alternative binding modes result in the nucleosome migrating at a different rate through a non-denaturing gel, therefore, each batch of nucleosome was analysed by 6% TBE NativePAGE. To improve homogeneity, samples found to contain multiple nucleosome bands were incubated at 37 °C for 30 minutes in an attempt to heat shift the DNA into the more thermodynamically stable conformation (see **Figure 4.7**). In all nucleosome samples a small quantity of non-nucleosomal Widom 601 DNA was visible, however, this band was observable even when octamer was added in excess and may, therefore, have been an artefact of the NativePAGE. 6% TBE NativePAGE showed no detectable decrease in nucleosome folding over a period of up to 3 months (gel not shown).

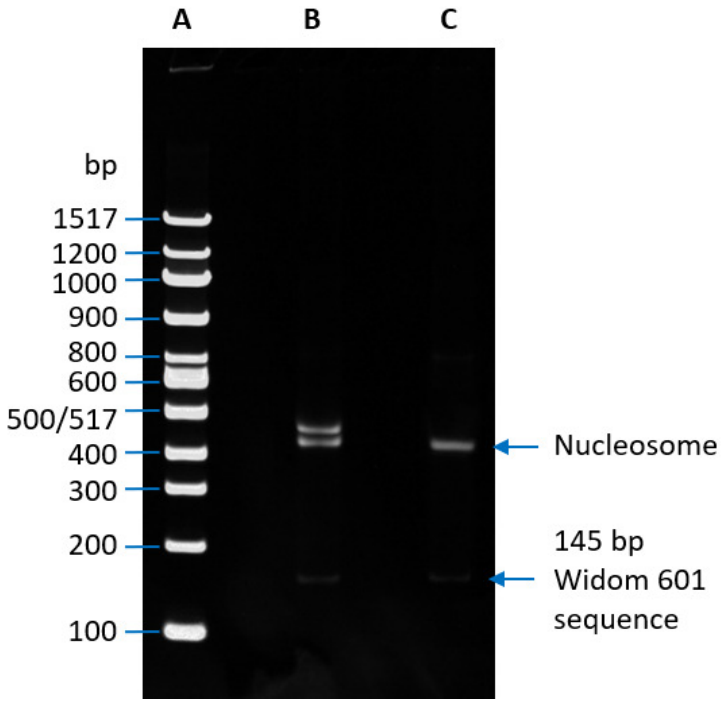



***Figure 4.7**: 6% TBE NativePAGE (Thermo Fisher Scientific) of partially misfolded, and properly reconstituted nucleosome. **A**- 100 bp ladder (NEB), **B**- Nucleosome sample displaying multiple DNA conformations, **C**- Homogeneous nucleosome sample. A small quantity of free 145 bp Widom 601 DNA was visible in both **B** and **C** (>5% intensity relative to the nucleosome band), but this was observed even in experiments where octamer was added in excess, so may have been an artefact of the gel.*

## 4.4 Crystallography

In the available Widom 601–nucleosome crystal structures the majority of the H3 tail is unmodeled due to its disordered nature in solution.[1] However, an H3 tail binding partner should stabilise it enough to be observed, as is seen in many structure of H3 binding partners when complexed with H3 tail fragments.[6] The strong interaction observed between a PHD–BRD construct of TRIM24 and an H3 K18Ac peptide (21 nM $K_D$) suggests the possibility of obtaining a nucleosome–TRIM24 crystal structure. As an additional check on purity and crystallisability, independent crystallisation trials of TRIM24 and the H3 sK18Ac-modified nucleosome were first conducted.

For instrumentation details see **Materials and Methods 8.10.1**.

### 4.4.1 Crystallography of a modified nucleosome

Sitting drop crystallisation trials of H3 sK18Ac nucleosome (10 mg/mL) (prepared as described in **Section 4.3**) in LSRB were set up at 4 °C and RT using a range of broad, sparse matrix and protein–DNA complex optimised crystal screens; JCSG+, PACT premier, WIZARD Classic 1 & 2 (Molecular Dimensions), Protein–Nucleic Acid Complex Crystal Screen (KeraFAST), Natrix, Natrix 2 and PEG ion, PEG ion 2 (Hampton Research). The protein solution was added to the crystallisation buffer in a 1:1, 1:2 and 1:3 ratio in a CrystalMation Intelli-Plate™ 96-3 well low-profile crystallisation plate (Hampton Research); the plates were then sealed. Crystals grew in a number of conditions at 4 °C, however, these same conditions produced larger crystals at RT. Details of the conditions in which crystals were observed are provided in **Table 4.1**, and photographs of the crystals that grew at RT are presented in **Figure 4.8**.

***Table 4.1**: Crystallisation hits of H3 sK18Ac nucleosome. P:B is the ratio of protein to crystallisation buffer in the condition photographed. 4 °C indicates whether the condition also produced crystals at 4 °C. Crystal photographs are presented in **Figure 4.8**.*

| Screen | Well ID | P:B | 4 °C | Crystal form | Photograph |
|---|---|---|---|---|---|
| JCSG+ | A7 | 1:3 | | Rhombohedral clusters | A |
| JCSG+ | A12 | 1:2 | Yes | Microcrystals | B |
| JCSG+ | D2 | 1:1 | Yes | Microcrystal clusters | C |
| JCSG+ | D7 | 1:2 | | Fibrous clusters | D |
| JCSG+ | G6 | 1:1 | Yes | Rhombohedral | E |
| KeraFAST | 23 | 1:1 | | Rhombohedral | F |
| Natrix | 9 | 1:3 | Yes | Hexagonal plate | G |
| Natrix 2 | 23 | 1:3 | Yes | Hexagonal plate | H |
| Natrix 2 | 24 | 1:1 | | Lenticular | I |
| Natrix 2 | 31 | 1:2 | | Rods | J |
| Natrix 2 | 47 | 1:2 | Yes | Rods | K |
| PACT premier | F9 | 1:1 | | Needle clusters | L |
| PEG ion | 47 | 1:1 | | Rhombohedral | M |
| PEG ion 2 | 33 | 1:1 | | Needle clusters | N |
| WIZARD | 35 | 1:3 | | Needle clusters | O |
| WIZARD | F8 | 1:2 | | Rod clusters | P |

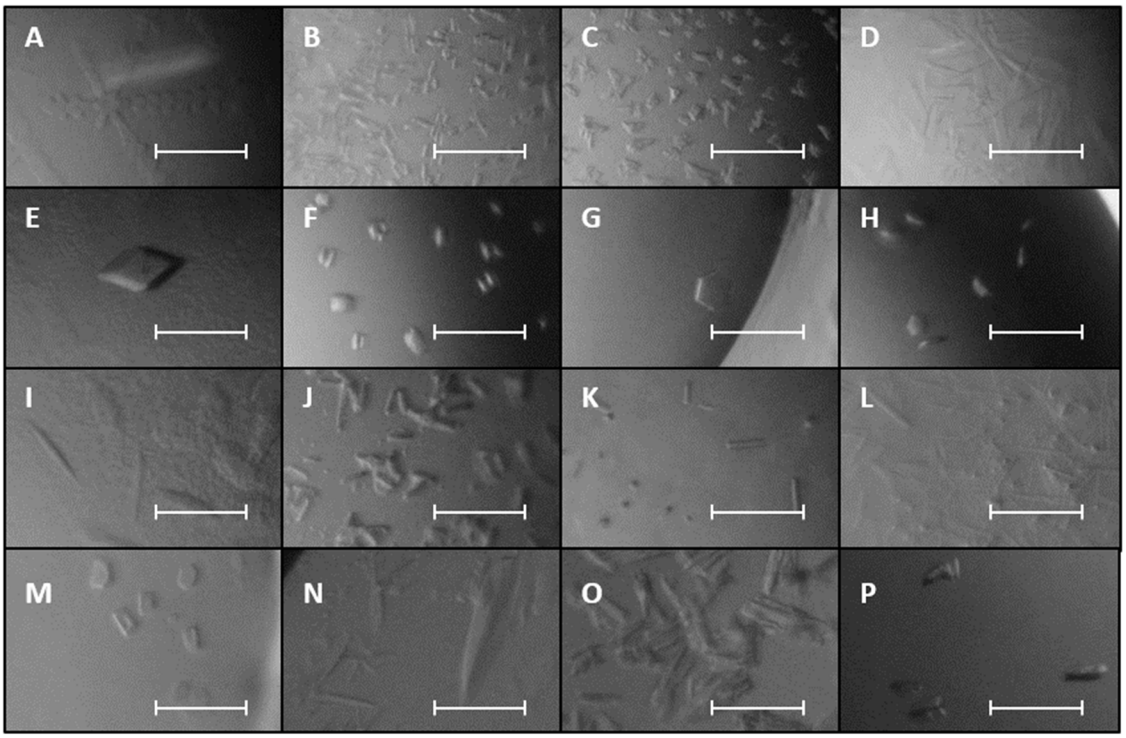



***Figure 4.8**: Photographs of H3 sK18Ac nucleosome crystallisation hits. Crystals were photographed after growing for 15 days at RT. Scale bar width is 10 μm. Crystals used in diffraction experiments were grown in condition **H** (Natrix 2 condition 23; 18 mM magnesium chloride, 50 mM sodium cacodylate-HCl pH 6.5, 10% (v/v) 2-propanol, 3 mM spermine).*

Attempts to optimise the crystallisation buffer based on the conditions in which crystals grew were unsuccessful in producing larger protein crystals. A single H3 sK18Ac nucleosome crystal (pictured in **Figure 4.8 H**) was harvested from Natrix 2 condition 23 (18 mM magnesium chloride, 50 mM sodium cacodylate-HCl pH 6.5, 10% (v/v) 2-propanol, 3 mM spermine). The crystal diffracted to 3.0 Å resolution, and the refinement statistics are displayed in **Table 4.2**. The H3 sK18Ac nucleosome crystallised in similar conditions (10 mg/ml nucleosome, 55 mM magnesium chloride, 30 mM potassium chloride, and 20 mM potassium-cacodylate-HCl pH 6.0), and in the same space group ($P\ 2_1\ 2_1\ 2_1$), as in a previously reported unmodified Widom 601 nucleosome structure (PDB ID: 3LZ0).[22]

### 4.4.2 Crystallography of TRIM24

A purified 21.5 kDa TRIM24$_{824\text{-}1006}$ construct containing the PHD and BRD of TRIM24 was supplied by *Dr Angelina Sekirnik*. It should be noted that future references to TRIM24 in this chapter refer to this TRIM24$_{824\text{-}1006}$ construct unless otherwise stated.

Initial crystallisation trials of TRIM24 (18 mg/mL) were set up using the commercially available JCSG+ (Molecular Dimensions) sparse matrix screen. Crystals were observed in two conditions (0.2 M magnesium chloride hexahydrate, 0.1 M 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)-NaOH pH 7.5, 30% (w/v) PEG 400; and 0.8 M sodium phosphate monobasic monohydrate, 0.1 M HEPES-NaOH pH 7.5), and hits were further optimised. Diffraction-quality crystals were grown by sitting drop, vapour diffusion at 20 °C, reaching their maximum size in 9 days. The crystals used in X-ray diffraction studies were grown in conditions containing 1.2 M sodium phosphate monobasic monohydrate and 0.1 M sodium HEPES-NaOH pH 8.0 and crystallised in space group *C* 1 2 1. The TRIM24 crystals diffracted to 2.0 Å resolution and refinement statistics are displayed in **Table 4.2**.

In the case of both the H3 sK18Ac nucleosome and TRIM24 crystallisation trials, diffraction data was indexed, integrated and scaled using HKL-3000.[23] The structures were solved by molecular replacement using Phaser, followed by refinement using PHENIX and Coot until the $R_{work}$ and $R_{free}$ values no longer decreased.[24–26]

***Table 4.2**: Processing and refinement statistics for TRIM24 and H3 sK18Ac nucleosome crystal structures. FRE+ is an in-house X-ray generator equipped with a Saturn 944+ charge coupled device (CCD) detector. †ASU = Asymmetric unit. ‡RMSD = root mean square deviation.*

| Dataset | TRIM24 | H3 sK18Ac nucleosome |
|---|---|---|
| **Processing** | | |
| X-ray source | FRE+ | FRE+ |
| Resolution (outer shell) (Å) | 27.9-1.96 (2.03-1.96) | 60.6-3.01 (3.50-3.01) |
| Unit cell dimensions a, b, c (Å) | 89.4, 36.8, 129.7 | 191.9, 110.2, 106.0 |
| Unit cell dimensions α, β, γ (degrees) | 90.0, 110.1, 90.0 | 90.0, 90.0, 90.0 |
| Space group | *C* 1 2 1 | *P* $2_1$ $2_1$ $2_1$ |
| Protein molecules per ASU† | 2 | 1 |
| Completeness (outer shell) (%) | 99.1 (96.6) | 98.0 (90.7) |
| $R_{merge}$ (outer shell) | 9.1 (65.5) | 6.0 (51.9) |
| No. of unique reflections | 28739 | 42791 |
| I/σ mean (outer shell) | 12.3 (2.1) | 9.4 (2.2) |
| No. of non-H protein atoms per ASU† | 3194 | 5998 |
| **Refinement** | | |
| RMSD from ideal bond length (Å) ‡ | 0.020 | 0.011 |
| RMSD from ideal angles (degrees) | 0.667 | 1.006 |
| $R_{work}$ (%) | 16.8 | 21.9 |
| $R_{free}$ (%) | 19.7 | 32.5 |

### 4.4.3 H3 sK18Ac nucleosome crystal structure

Although the nucleosome structure has been described in detail, this section discusses key features of the nucleosome fold and aspects relating to this investigation.

The H3 sK18Ac nucleosome crystal structure comprises a $(H3–H4)_2$ tetramer flanked by two (H2A–H2B) dimers encircled twice by 145 bp Widom 601 double helix DNA (see **Figure 4.11**). The crystal structure of the H3 sK18Ac nucleosome core aligns well with a previously published nucleosome containing the Widom 601 DNA sequence (PDB ID: 3LZ0), with an RMSD of 0.280 Å for all histone $C^{\alpha}$ atoms.[22] The model contains all 145 bp of the Widom 601 sequence, as before, as well as 78% of the core histone residues. While the $R_{free}$ value for the H3 sK18Ac nucleosome structure was high (32.5%), this is comparable to other reported nucleosome structures ($R_{free}$ = 30-32%).[18, 22]

The H3 sK18Ac nucleosome has a pseudo-twofold rotational axis passing through the midpoint of the DNA sequence. While the Widom 601 sequence is non-palindromic, and the two halves of the DNA share only a 28% similarity on the nucleotide level, the histone–DNA register is identical for both halves of the DNA. DNA bases are therefore numbered with respect to the DNA midpoint; *C/G0*, located in 'turn *0*' (see **Figures 4.9 B** and **4.10**).

#### Structural analysis of the histone core

Each histone displays the characteristic histone fold; an 11 residue α-helix (α1), followed by a 7-9 residue loop and short 2 residue β-strand (L1), a longer 27-28 residue α-helix (α2), a second 6-7 residue loop and short 2-3 residue β-strand(L2), and a third 11-12 residue α-helix (α3). In addition to this fold each histone also contains *N*- and *C*-terminal largely disordered regions.

In the case of H3 and H4 the *C*-termini are short sequences of 3 or 10 residues, respectively, located inside the histone core of the nucleosome. H2A has a much longer, 36 residue, *C*-terminal loop that extends along the surface of the histone core. The final 11 *C*-terminal residues of H2A are apparently disordered. The H2B *C*-terminal tail contains 24 residues arranged in an α-helix along the surface of the histone core. Unlike H3 and H4, the *C*-terminal tails of H2A and H2B have been found to be important for chromatin compaction and remodelling.[27,28]

The *N*-terminal 27 and 37 residues of the H2B and H3 tails, respectively, including the modified H3 sK18Ac, were not modelled due to disorder, which infers flexibility. Likewise, the *N*-terminal tails of H2A and H4, situated on the exposed histone face of the nucleosome, appear extended beyond the histone core and are disordered up to residues 16 and 20, respectively. Previously reported nucleosome crystal structures containing Widom 601 DNA also lack the same residues from their models.[1–3,18,22]

Histone dimers contain two histones, either H2A and H2B, or H3 and H4, crossing each other about half way through α2. The two histones are related to each other by pseudo-twofold rotational symmetry. The α1 and α3 helices bind to the α2 helix of both histones. The α1 helices form a positively charged surface constituting the 'α1α1' DNA binding site. The α3 helices in the dimer are not close enough to make contact with each other and do not bind to the DNA. The short L1 β-strand of one histone forms a two-residue two-stranded parallel β-sheet with the short L2 β-strand of the other histone in the dimer. These 'L1L2' loop regions are rich in positively charged lysine and arginine residues and provide an additional DNA binding site (see **Figure 4.9 A**).

The interface between the two (H3–H4) dimeric components of the $(H3–H4)_2$ tetramer occur as a four-helix bundle between the two H3 histones. The bundle comprises both H3 α3 helices and the *C*-terminal region of both α2 helices with the α2 helix of one histone running parallel to the α3 helix of the other (see **Figure 4.9 B**). The two (H2A–H2B) dimers associate to the $(H3–H4)_2$ tetramer in a similar way; the α2 and α3 helices of H2B form a four-helical bundle with the equivalent helices of H4.

#### Histone–DNA interactions

12 of the 14 DNA double helix turns come into contact with an αlα1 or L1L2 binding site. Each (H2A–H2B) dimer as well as both (H3–H4) dimers bind to three turns of DNA through their L1L2, αlα1, and second L1L2 binding sites, sequentially (see **Figure 4.9 B**). Each (H2A–H2B) dimer binds to either DNA turns *-6* to *-3* or *+3* to *+6*, and the $(H3–H4)_2$ tetramer binds to DNA turns *-3* to *+3*.

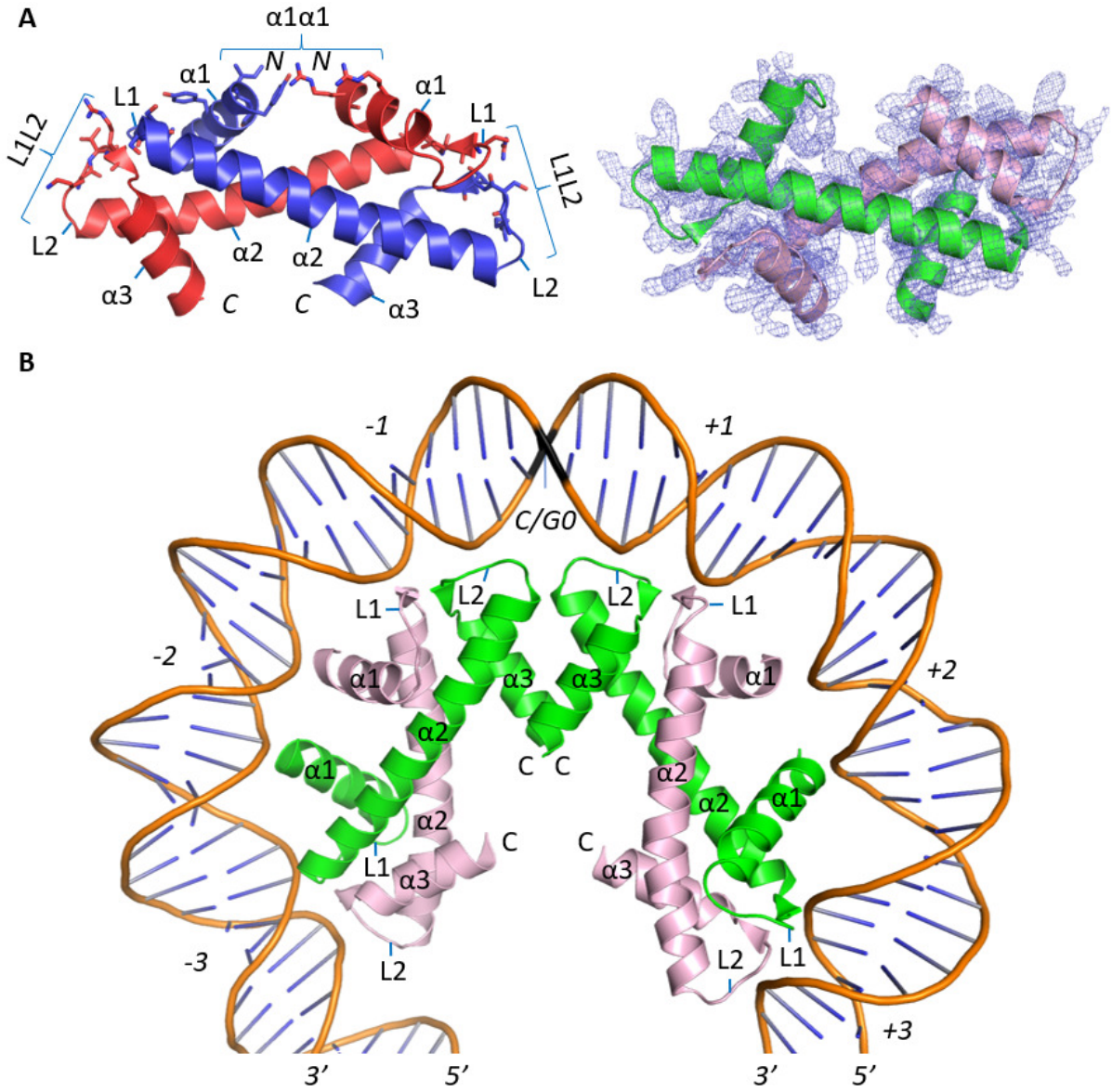



***Figure 4.9**: Views of histones and DNA extracted from a crystal structure of the H3 sK18Ac nucleosome. H2A is displayed in red, H2B in blue, H3 in green, and H4 in light pink. The DNA is displayed in grey with an orange phosphate backbone and the central* C/G0 *base pair is coloured black. The histone tails are not shown. **A**- An (H2A–H2B) (**left**) and (H3–H4) (**right**) dimer showing the characteristic histone dimer fold. The αIαI and LIL2 binding sites are also indicated. Histone dimers have pseudo-twofold rotational symmetry about the vertical axis as printed. Residues close enough to interact with the DNA (<3.5 Å) are shown as sticks (**left**). An example 2|Fo-Fc| electron density map for the H3–H4 dimer is shown (light blue mesh), contoured to 1 σ (**right**). **B**- The $(H3–H4)_2$ tetramer interacting with 6 turns of DNA. The labelled α2 and α3 helices comprise the four-helix bundle typical of histone dimer–dimer binding sites. The DNA turn number is displayed in italics in the major groove. Turns ±1 and ±3 interact with the LIL2 DNA binding site and turns ±2 interact with the αIαI DNA binding site of the tetramer.*

The *N*-terminal tails of H2B and H3 cross between two strands of double helical DNA *via* channels formed from the minor grooves of the DNA. The *N*-terminal tails of H2B cross between turns *+3* and *-5*, or *-3* and *+5*, and the *N*-terminal tails of H3 pass between turns *+1* and *-7*, or *-1* and *+7* (see **Figure 4.10**). Histone tails, therefore, pass between every second turn of DNA in the H3 sK18Ac nucleosome. The minor grooves of the terminal, ±*7* DNA turns do not come into contact with one of the histone core binding sites, and only interact with the histones through the H3 tail.

Histones interact with DNA using both backbone and side-chain atoms. The crystal structure indicates that extensive hydrogen bonds and salt bridges may form between the DNA phosphate oxygen atoms and amino acid side-chains. While the majority of these polar interactions are observed between arginine side-chains and phosphate groups, lysine, threonine and serine side-chains also form hydrogen bonds or salt bridges to the DNA phosphate groups. Additionally, the side-chain of an arginine residue binds into, or near the minor groove of each turn of DNA. One of these minor groove-binding arginine residues are located in each of the four *N*-terminal histone tails that cross between the DNA gyres. The other 10 arginine side-chains are located in 10 of the 12 α1α1 or L1L2 DNA binding sites in the histone core. The α1α1 DNA binding sites of the (H2A–H2B) dimers do not contain arginine residues that bind inside the nearby DNA turn ±*5* minor grooves; these DNA minor groove-binding arginine residues are instead supplied by the H2B *N*-terminal tails (see **Figure 4.10**). These buried arginine side-chains are positioned to form multiple hydrogen bonds with the deoxyribose backbone of the DNA.

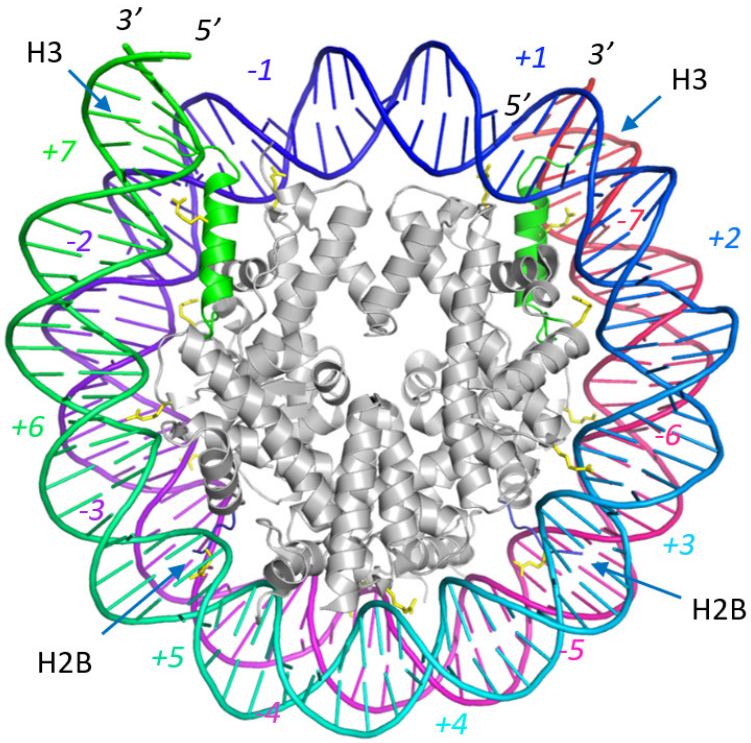



***Figure 4.10**: View from a crystal structure of the H3 sK18Ac nucleosome. H3 and H2B* N-*terminal tails are shown in green and blue, respectively, DNA minor groove-binding arginine residues in yellow, the remainder of the histone core in grey and the DNA in rainbow. H3 tails cross between the minor grooves of DNA turns* -7 *and* +1*, and turns* -1 *and* +7*. H2B tails cross between the minor grooves of DNA turns* -5 *and* +3*, and turns* -3 *and* +5*.*

DNA phosphate groups and histone backbone nitrogen atoms at the *N*-termini of the α2 helices of all histones, and the α1 helices of H2B and H3, are positioned to form hydrogen bonds that are potentiated by the dipole of the helix. Hydrogen bonds are also suggested between DNA phosphate oxygen atoms and histone backbone nitrogen atoms on each of the L1 and L2 loops.

An overall view of the crystal structure of the H3 sK18Ac nucleosome is shown in **Figure 4.11** along with zoomed in views depicting examples of histone core– and histone tail–DNA interactions.

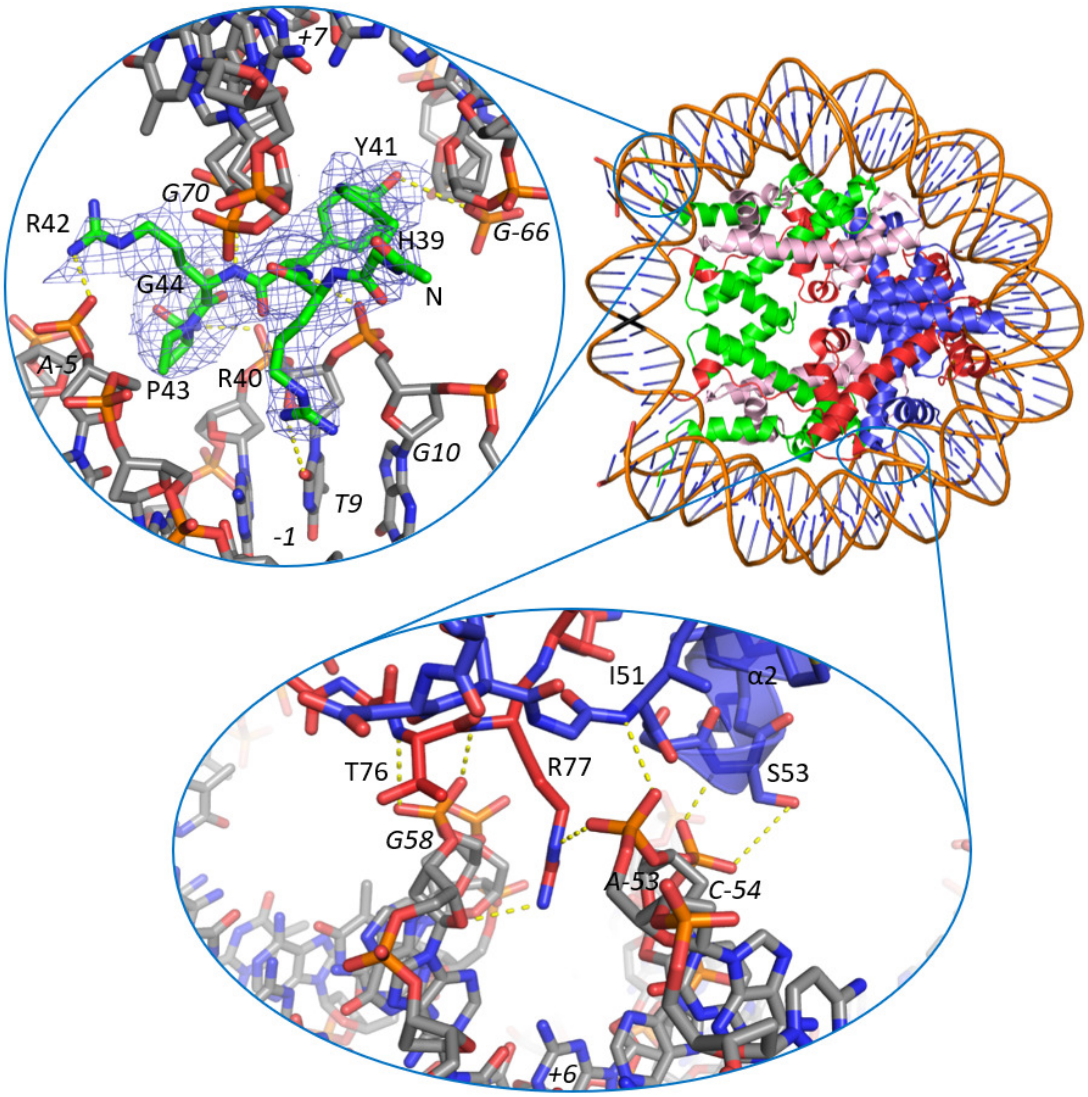



***Figure 4.11****: Views from a crystal structure of the H3 sKI8Ac nucleosome. H2A is displayed in red, H2B in blue, H3 in green, H4 in light pink, DNA in grey with an orange phosphate backbone and the DNA midpoint (C/G0) in black. DNA turn numbers and nucleotide base identifiers are provided in italics. Close contacts (within 3.5 Å) between histone residues and DNA are indicated with a yellow dotted line in the magnified regions.* ***Left inset****- The segment of the H3 N-terminal tail which crosses between the two DNA gyres. The amide backbone nitrogens of Y41, R42 and G44 are positioned to form hydrogen bonds with the phosphate oxygens of G10, G70 and T9, respectively. Hydrogen bonds between DNA phosphate groups and H3 side-chains appear likely between G-66 and the hydroxyl of Y41, and A-5 and an R41 $N^{\omega}$ nitrogen. An R39 $N^{\omega}$ nitrogen and the ortho-carbonyl oxygen of T9 are also positioned to form a hydrogen bond. The conformations of the side-chains depicted as sticks were all clearly indicated by the electron density, and an example 2|Fo-Fc| electron density map for the $H3_{39-44}$ sequence is shown (light blue mesh), contoured to 1 σ.* ***Bottom inset****- (H2B)L1(H2A)L2 DNA binding site showing multiple possible backbone nitrogen–DNA phosphate, and side-chain amide or hydroxide–DNA phosphate hydrogen bonds. Note that a hydrogen bond between the S53 amide and the C-54 phosphate oxygen would be enhanced by the α2 helix dipole. A hydrogen bond between an R77 $N^{\omega}$ nitrogen and the deoxyribose ester oxygen of G58 is also positioned to occur.*

### 4.4.4 TRIM24 crystal structure

The crystal structure of a TRIM24 183 residue construct encompassing the PHD (residues 824-883), linker (residues 884-898) and BRD (residues 899-1006) (see **Figure 4.12**) was crystallised in the *C* 1 2 1 space group. A crystal structure of a similar TRIM24 construct in the unbound state, crystallised under different conditions (0.2 M ammonium acetate, 0.1 M HEPES-NaOH pH 7.4, 25% (v/v) PEG 3350) in the *P* 1 space group has previously been reported (PDB ID 3O33).[6] The *C* 1 2 1 crystal form structure obtained in this investigation aligned well with the *P* 1 crystal form structure, having an RMSD of 0.263 Å for all $C^{\alpha}$ atoms. All residues observed in the *P* 1 structure were visible in the *C* 1 2 1 structure. Electron density for several additional residues in the inter-domain loop could be observed and were modelled.

The two TRIM24 PHD zinc ions are tetrahedrally coordinated through either four cysteine residues, or three cysteines and one histidine residue. Each zinc-binding cysteine or histidine residue is separated from a second zinc-coordinating cysteine residue by two amino acids forming a C/HxxC motif. Zn1 is coordinated by the first (C829 and C832) and third (H848 and C852) C/HxxC motifs, and Zn2 is coordinated by the second (C841 and C844) and fourth (C867 and C870) CxxC motifs (see **Figure 4.12 A** and **B**). This alternation of coordination pairs describes the 'cross-brace' topology characteristic of all structurally characterised PHD fingers.[29] It is through this PHD that TRIM24 reads unmethylated H3 K4 and H3 K9$Me_3$.

The BRD of TRIM24 contains a left-handed four helical bundle characteristic of the BRD fold. The helices are sequentially labelled α1-α4 from the *N*-*C* termini. Helix α1 is connected to α2 by one long 27-residue linker. Helices α2/α3, and α3/4 are connected by

shorter, 5 and 8 residue loops, respectively. The BRD KAc binding site comprises residues located on the α3 and α4 helices and on the α1/α2 linker (see **Figure 4.12 C**).

The TRIM24 KAc binding site residues are all hydrophobic, with the exceptions of C976 and N980. Previously reported crystal structures of TRIM24 complexed with acetylated H3 tail fragments show N980 is able to form a hydrogen bond between the asparagine side-chain nitrogen and the acetyl group carbonyl of a KAc residue.[6] A modified KAc side-chain is able to more efficiently fill the hydrophobic BRD binding site than an equivalent unmodified lysine residue. This efficient filling of the binding site, combined with the hydrogen bond to N980, may explain the increased affinity of TRIM24 to peptides containing KAc residues *vs.* those containing unmodified lysine residues.

The *C* 1 2 1 crystal structure solved in this chapter was aligned with a crystal structure of TRIM24 bound to an $H3_{13-32}$ peptide containing K23Ac (PDB ID: 3O34). No structure of TRIM24 bound to an H3 K18Ac tail is currently available.[6] The positioning of the BRD K23Ac-binding residues are identical between the peptide-bound TRIM24 structure and the structure of TRIM24 in the unbound state, indicating no conformational change to the BRD binding site is required for KAc binding. An H3 K23Ac residue extracted from the aligned 3O34 structure is shown in **Figure 4.12 C** to illustrate the KAc binding position in the TRIM24 BRD.

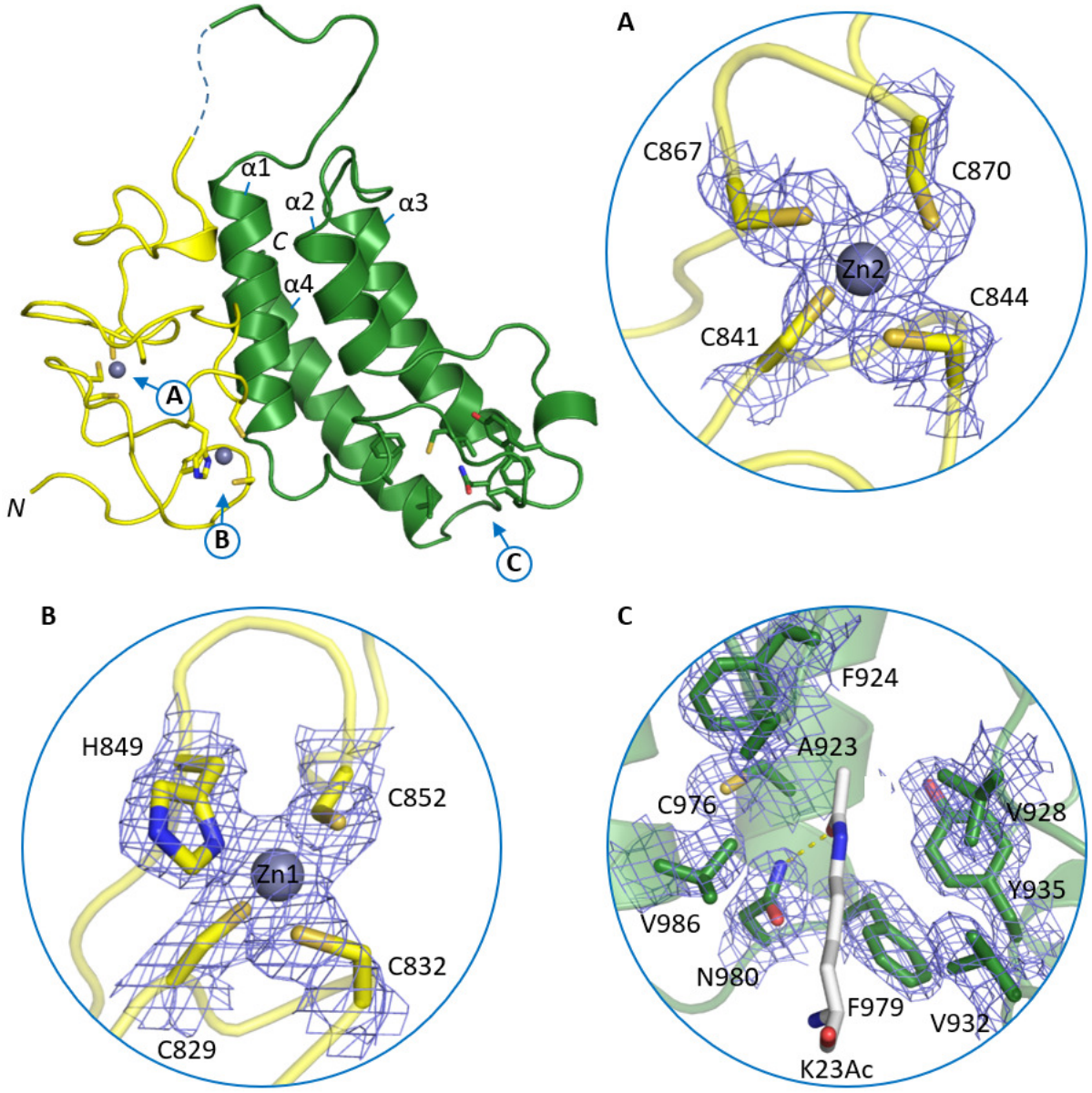



***Figure 4.12**: Cartoon representation of the C 1 2 1 crystal form structure of TRIM24 in the unbound state. The PHD is shown in yellow and the BRD in green. Arrows indicate the perspectives of the magnified regions. Magnifications of the PHD coordination environments around Zn2 and Zn1 are provided in **A** and **B**, respectively. **C**- The KAc binding site in the BRD of TRIM24. 2|Fo-Fc| electron density maps around metal-, or KAc-binding residues are displayed (light blue mesh) contoured to 1 σ. An H3 K23Ac residue superimposed from PDB ID: 3O34 onto the C 1 2 1 crystal form structure is shown in white to illustrate the interactions between the BRD of TRIM24 and a KAc residue.[6] Acetylated lysine can form a strong hydrogen bond with N980 (yellow dotted line) and hydrophobic interactions with F924 and V928.*

The success of the H3 sKI8Ac, and TRIM24 crystal trials showed that both species were purified to an appropriate level for crystallography.

### 4.4.5 TRIM24–peptide crystallography

While co-crystal structures of TRIM24 have been reported with peptides binding to either the PHD or BRD (PDB IDs: 3O37 and 3O34, respectively), no structure containing a peptide spanning both domains is currently available.[6] The tightly binding $H3_{1\text{-}27}$ K4, K9Me$_3$, K18Ac peptide used in the AlphaScreen experiments mentioned in **4.1.2** were selected for co-crystallisation trials.

Soaking of peptide into the new crystal form of TRIM24 was attempted, however, all soaking attempts resulted in the dissolution of the crystals. Broad screen co-crystallisation trials were set up for TRIM24 with peptide at molar ratios of 1:1.5 and 1:3 and hits were optimised. However, crystals that grew were small and diffracted to >10 Å resolution. Seeding attempts into optimisation trials using TRIM24 as seed stock were also unsuccessful.

### 4.4.6 Nucleosome–TRIM24 complex crystallography

Each nucleosome contains two H3 proteins, so it was predicted two TRIM24 molecules would bind to each nucleosome. Preparations of nucleosome and TRIM24 were combined in a 1:2 and 1:10 molar ratio and incubated together at 4 °C for 1, or 12 hours, respectively. Attempts were made to separate the nucleosome–$(TRIM24)_2$ complex from the constituent species by S200 SEC. Fractions containing the highest molecular weight species were analysed by 16% SDS-PAGE and found not to contain TRIM24, consisting instead of nucleosome alone (figure not shown). Therefore, it was decided to attempt crystallisation trials of the mixture without the SEC step.

Samples of nucleosome and TRIM24 in LSRB buffer were combined at a concentration of 1 mg/mL in a 1:2 molar ratio and incubated for 1 hour at 4 °C. The sample was concentrated to 9.1 mg/mL, then used to set up the same array of crystal screens used in

the earlier nucleosome crystallography experiments, at both RT and 4 °C. After 6 months no protein crystals were detected. Seeding attempts into optimisation trials using H3 sK18Ac nucleosome as seed stock were also unsuccessful.

Following the failure to observe the formation of a complex between H3 sK18Ac nucleosome and TRIM24 by SEC or co-crystallisation, attempts were made to better characterise their interactions using non-denaturing MS.

## 4.5 Non-denaturing mass spectrometry

Conventional methods of electrospray ionisation mass spectrometry involve measuring the mass of a protein under denaturing conditions (such as 50% (v/v) acetonitrile, 0.1% (v/v) formic acid).[30] These conditions interrupt the non-covalent inter- and intramolecular bonds of the protein, causing it to unfold and expose all residues to the solution. The unfolded protein has a high surface area to volume ratio and is able to pick up a high number of charges, leading to the majority of molecules having a low mass to charge ratio ($m/z$).[30]

Non-denaturing, or 'native' electrospray ionisation mass spectrometry (non-denaturing MS) uses buffer conditions in which non-covalent protein interactions are preserved.[30] Folded proteins have a lower surface area to volume ratio, and fewer charged residues exposed to the solution than their unfolded counterparts, resulting in a higher $m/z$ for folded proteins. As inter- and intramolecular protein bonds can be conserved while conducting this technique, non-denaturing MS can be useful in the investigation of folded, macromolecular complexes in the near-native state.[31,32]

The absence of TRIM24 in high molecular weight SEC fractions suggested that the affinity between the TRIM24 and the H3 sK18Ac nucleosome was much lower than

predicted by peptide studies. To investigate this further, non-denaturing MS was employed in order to ascertain whether TRIM24 could be observed binding to the H3 sK18Ac nucleosome at all, the number of TRIM24 molecules that could be detected associating to each H3 sK18Ac nucleosome, and what effect increasing TRIM24:H3 sK18Ac nucleosome ratio would have on complex formation.

A key requirement for non-denaturing MS is for the samples to remain ionised and volatile through use of an appropriate buffer such as ammonium acetate. Non-volatile buffer components including most salts are not compatible. The H3 sK18Ac nucleosome and TRIM24 were diluted to 1 mg/mL in LSRB, then combined in a 1:2 ratio and incubated together at 4 °C (1 hr). Samples were buffer exchanged into 250 mM ammonium acetate (ammonium hydroxide-acetic acid) buffer pH 7.4, then loaded to a gold-coated glass capillary. Nano electrospray ionisation measurements were performed using a Q Exactive™ hybrid quadrupole-Orbitrap™ mass spectrometer (Thermo Fischer Scientific) modified for the transmission and detection of high *m/z* ions in positive ion mode (carried out by *Dr Idlir Liko* or *Dr Nisha Patel*). Samples containing H3 sK18Ac nucleosome alone were also tested.

The ion distribution profile produced by samples containing nucleosome in the absence of TRIM24 was used to calculate the mass of the complex as 198.0 kDa. Due to the large mass of the sample, a large error was expected. However, the measured mass was within 1% of the predicted mass of the DNA-bound nucleosome (197.3 kDa). By comparison, addition of one, or two TRIM24 molecules to a nucleosome would cause a 10, or 18% increase in mass, respectively. In the sample containing the H3 sK18Ac nucleosome and TRIM24 mixture the same peaks were observed as for the independent nucleosome. In

addition to a recurrence of the nucleosome peak, less intense additional peaks were observed corresponding to a mass of 219.9 kDa, again, within 1% of the expected 218.8 kDa for a complex of one H3 sK18Ac nucleosome and one TRIM24 molecule. To confirm the identity of this new signal, MS/MS analysis of these ions was conducted (by *Dr Idlir Liko*), using a quadrupole time of flight II mass spectrometer (Waters), resulting in the detection of H2A, H4 and TRIM24 in the complex (see **Figure 4.13**).

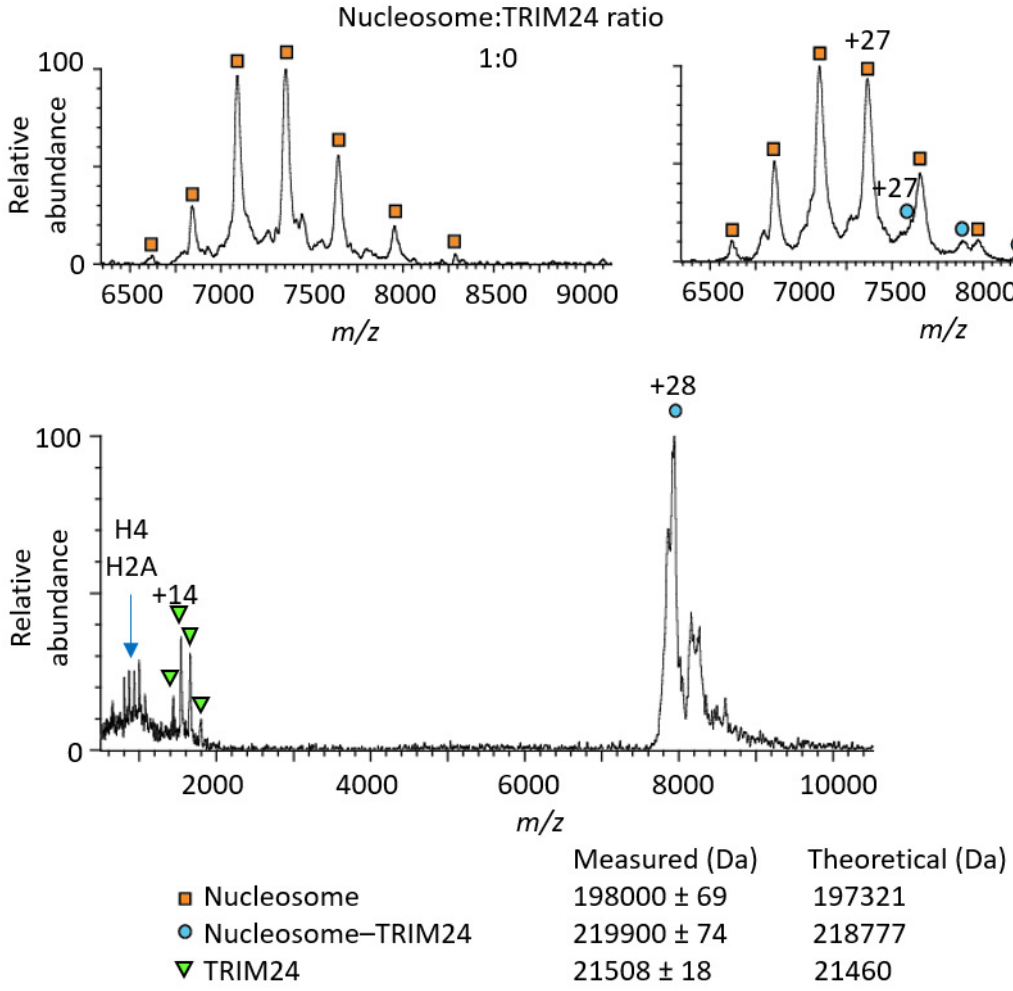



***Figure 4.13**: **Top**- Non-denaturing MS data for H3 sK18Ac nucleosome (**left**) and H3 sK18Ac nucleosome incubated with 2 equivalents of TRIM24 (**right**). **Bottom**- MS/MS spectra of H3 sK18Ac nucleosome–TRIM24 complex showing emission of TRIM24 and histones H4 and H2A.*

The non-denaturing MS experiments appeared to indicate only a small proportion of the H3 sK18Ac nucleosome and TRIM24 were complexed, and that the complex contained one H3 sK18Ac nucleosome and one TRIM24 molecule. To investigate the effect of relative TRIM24 concentration on nucleosome complex formation modified nucleosome was incubated with TRIM24 at a range of concentrations, then submitted to non-denaturing MS analysis as previously described. The proportion of signal corresponding to the nucleosome–TRIM24 complex increased for samples containing higher ratios of TRIM24 to nucleosome. At 3:1 ratios of TRIM24 to nucleosome and above, peaks corresponding to complexes of nucleosomes bound to two TRIM24 proteins began to become detectable. The H3 sK18Ac nucleosome–$(TRIM24)_2$ complex had an apparent mass of 241.5 kDa [predicted mass: 240.2 kDa] (see **Figure 4.14**). At TRIM24:nucleosome ratios greater than 15:1 the signal quality rapidly began to decrease.

While the intact nucleosome complex was observed by non-denaturing MS, H3 sK18Ac nucleosome–TRIM24 complexes containing one or two copies of TRIM24 were still in the minority, even when TRIM24 was added in a 15-fold excess.

The incompatibility of non-denaturing MS with non-volatile salts and the requirement for replacing the buffer with ammonium acetate solution may have adversely affected complex formation. TRIM24 tends to precipitate when in a low salt environment and, while no precipitation was visibly observed at 250 mM ammonium acetate, it is likely that the buffer may have altered the ability of TRIM24 to bind to the H3 sK18Ac nucleosome.

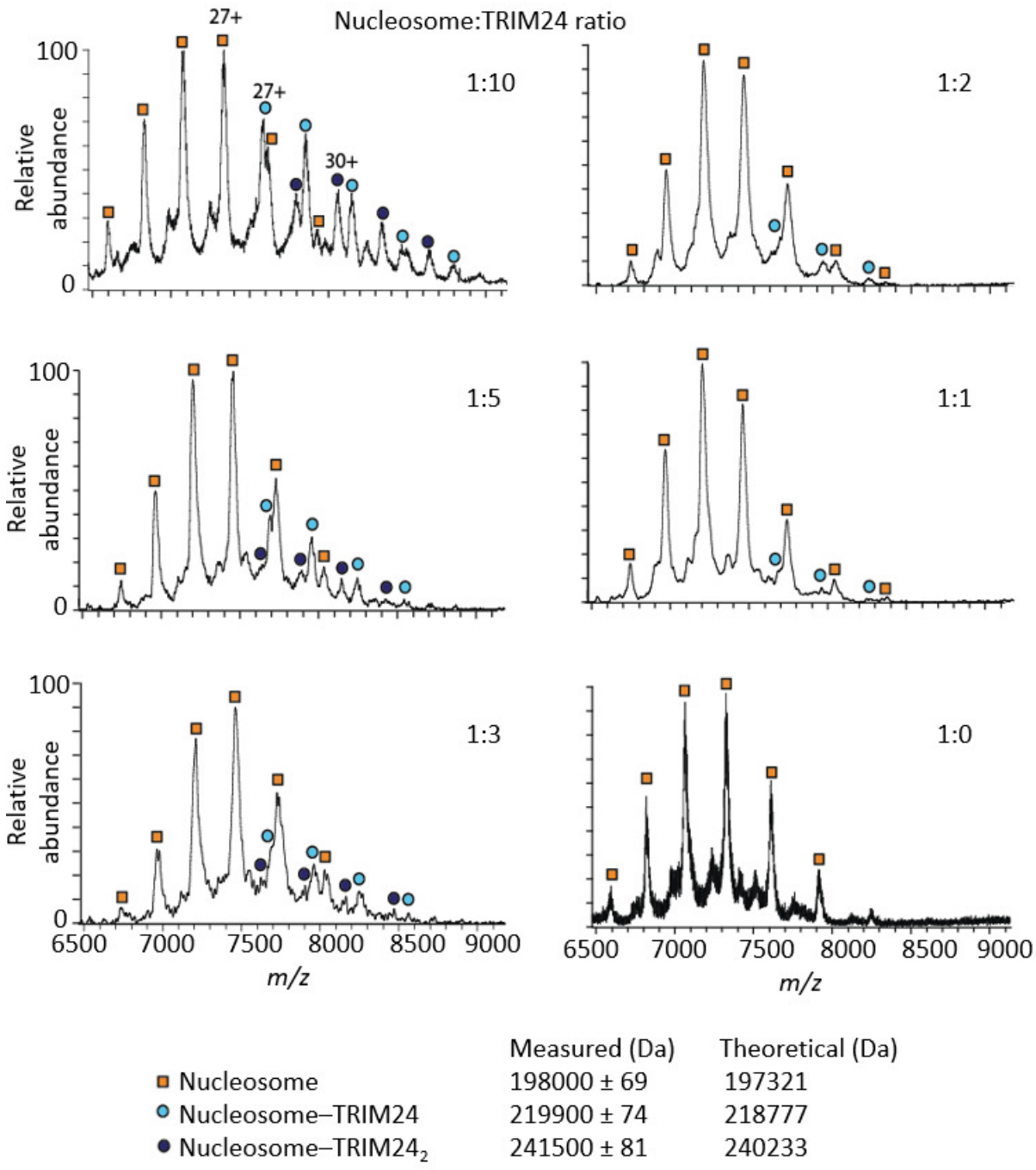



***Figure 4.14**: Non-denaturing MS data for H3 sK18Ac nucleosomes bound to 0, 1 and 2 molecules of TRIM24 after incubation at different ratios of binding partners.*

## 4.6 Small angle X-ray scattering

### 4.6.1 Background

SAXS is a biophysical method that is able to provide information on the low resolution (10-20 Å) "envelope" of solution-phase macromolecular complexes. SAXS can be conducted at a range of temperatures and buffers and is therefore a good technique for probing protein behaviour under near physiological conditions.

Electrons in a solution illuminated by X-ray light can interact with the incident beam resulting in a scattering pattern. Inhomogeneous areas of electron density, such as within a protein, scatter light differently to electrons in the bulk solvent. In a SAXS experiment a sample is illuminated with X-ray light, and the low-angle elastic scattering of X-ray photons by the sample is recorded.[33] The intensity of X-rays detected at different scattering angles can then be used to derive a pair-distance distribution (P(r)) function.[34] This P(r) function describes the distribution of all distances between electrons within an object, and can therefore be used to gain information about its shape. The P(r) function approaches 0 at the maximum diameter of the object ($d_{max}$).

All species present in the sample contribute to the X-ray scattering data dependent on their respective concentrations. Additionally, the degree of scattering caused by a particle is proportional to the number of electrons it contains, therefore larger complexes have a greater influence on a sample's scattering profile than smaller ones.[35]

### 4.6.2 Results

A 2-fold dilution series containing 35-0.3 μM H3 sK18Ac nucleosome with 2 equivalents of TRIM24 were subjected to SAXS analysis. For comparison, scattering data for H3 sK18Ac nucleosome in the absence of TRIM24 was also collected at the same

concentrations. The highest quality data was obtained from samples at 17.5 µM for the mixture, and 12.5 µM for the isolated H3 sK18Ac nucleosome, so these datasets were selected for analysis. P(r) plots were constructed from these datasets, indicating very different particle conformations (see **Figure 4.15**). SAXS data collection, and data analysis details are provided in **Materials and Methods 8.12.1** and **8.12.2**, respectively.

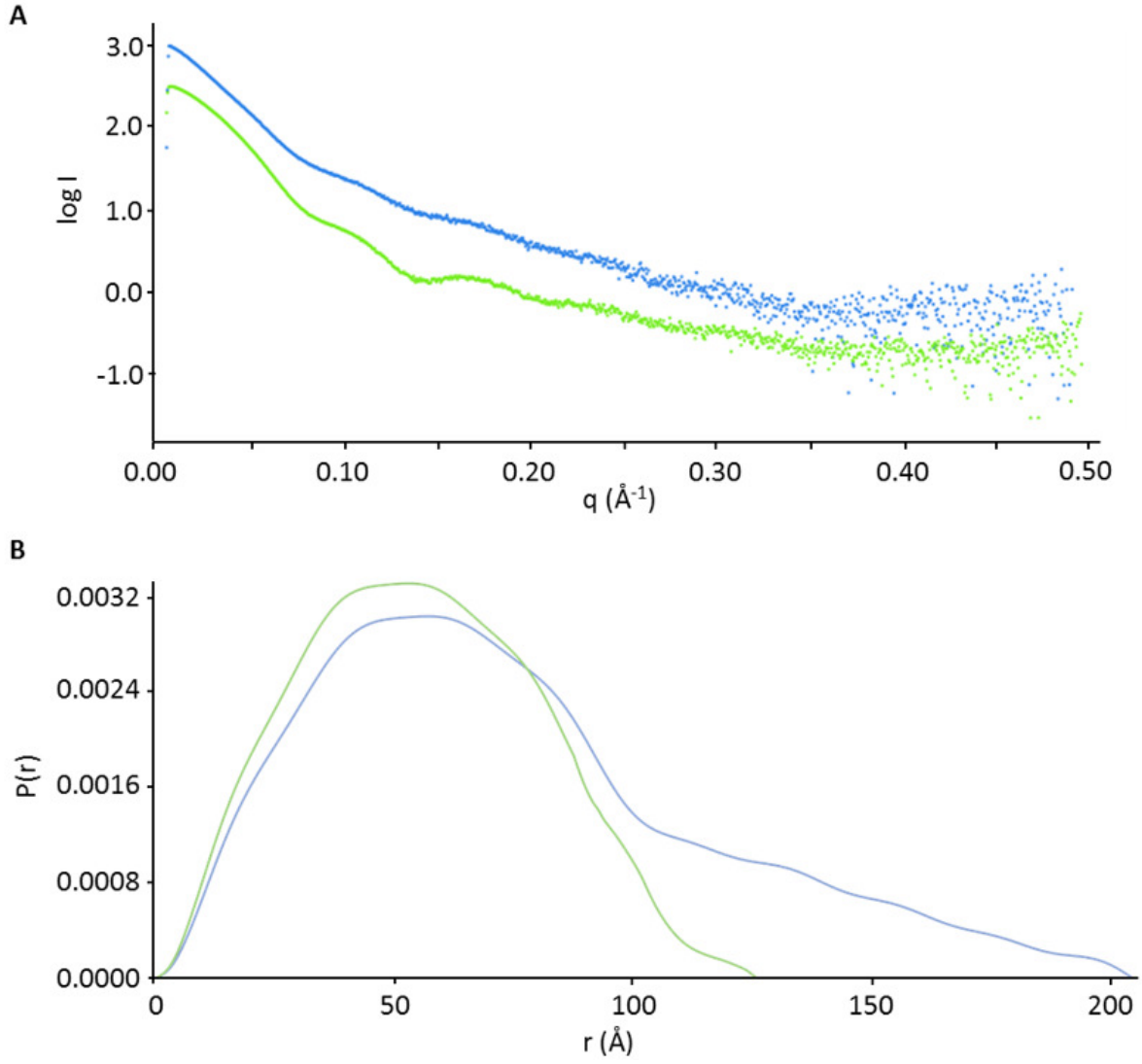



***Figure 4.15**: **A**- Scattering patterns recorded from samples containing independent H3 sK18Ac nucleosome (green) at 12.5 µM and H3 sK18Ac nucleosome + 2×TRIM24 (blue) at 17.5 µM H3 sK18Ac nucleosome. Log intensity (I) is plotted against scattering angle (q). **B**- P(r) histograms plotted using the above scattering patterns. P(r) is plotted against density separation (r).*

Guinier analysis was used to confirm that no aggregation was observed in the samples.[36] DAMIF, followed by DAMAVER, programs from the ATSAS suite, were used to predict 3D molecular envelopes from the P(r) data. The results are presented in **Figure 4.16**.

The P(r) function of the nucleosome is in the form of a bell-shaped curve, characteristic of globular structures, and indicates a $d_{max}$ of 120 Å. The P(r) function, and disk-shaped protein envelope derived from this data appear to have a similar form to previously published nucleosome SAXS, and crystallography structures (see **Figure 4.16 B**).[18,22,37,38] This indicates that the H3 sK18Ac nucleosome is amenable to SAXS studies.

The sample containing both nucleosome and TRIM24, however, appears to have a very different shape. A much larger $d_{max}$ of 205 Å indicates a wider complex, and the long tail of the P(r) plot suggest a less compact shape (see **Figure 4.15 B**). The molecular envelope calculated from the P(r) function indicates the presence of a complex containing a round, nucleosome-like region with a single extended tail. A potential explanation for this shape is that TRIM24 binds the nucleosome through the acetylated H3 tail at a point normally in close proximity to the DNA (see **Figure 4.16 C**). On TRIM24 binding the DNA may become displaced, unwinding from the histone octamer and resulting in the observed extended region of the envelope. Due to the large number of relatively electron rich phosphorus atoms in the DNA backbone, nucleic acids scatter X-rays more strongly than amino acids, perhaps partially explaining the large volume predicted for the tail. A schematic diagram of the suggested H3 sK18Ac nucleosome–TRIM24 complex is presented in **Figure 4.16 C**.

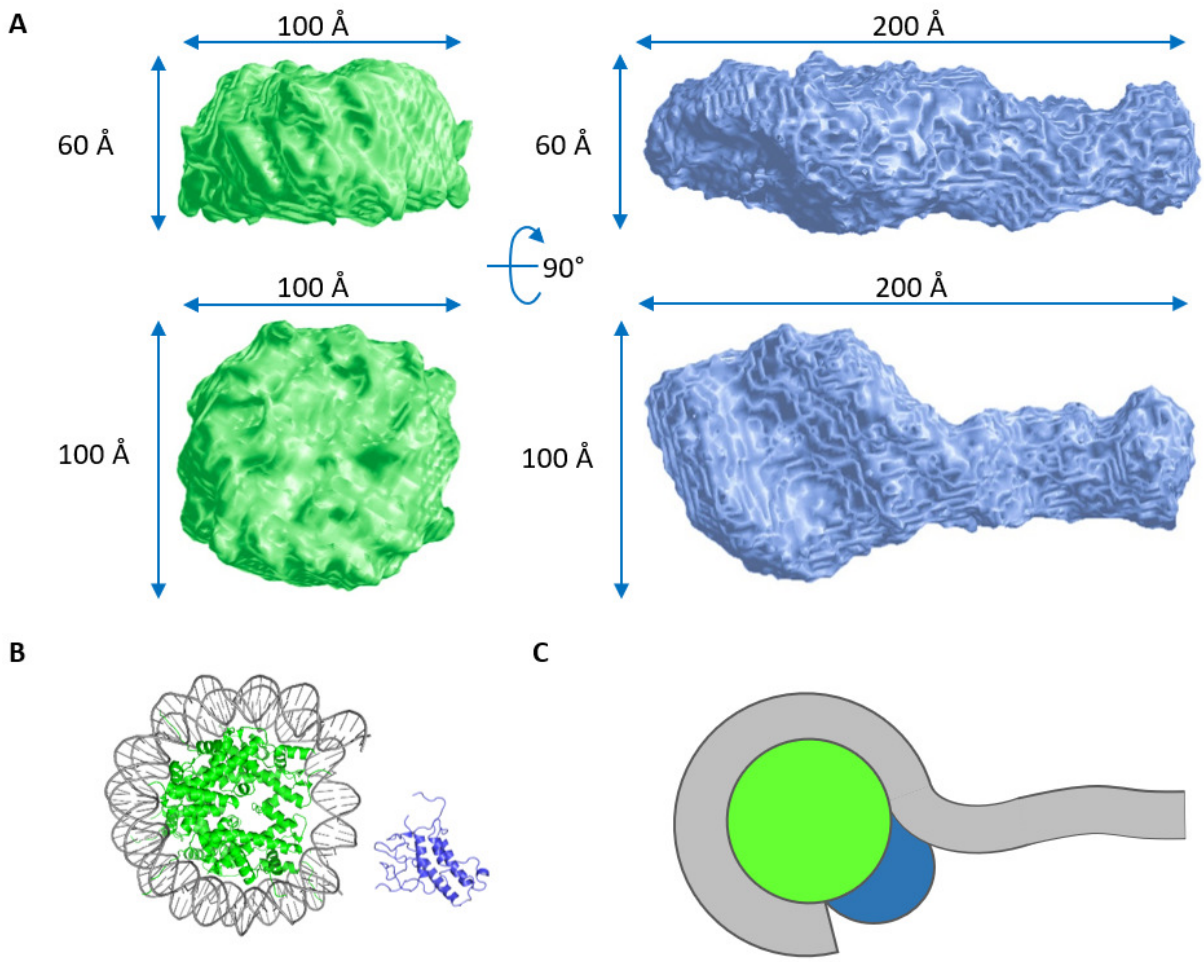



***Figure 4.16**: **A**- Molecular envelopes derived from SAXS data for H3 sK18Ac nucleosome (green) and H3 sKAc nucleosome + 2×TRIM24 (blue). **B**- Crystal structures solved in **Section 4.4.1** for reference. DNA is shown in grey, histones in green and TRIM24 in blue. Crystal structures are to scale with SAXS-derived molecular envelopes. **C**- Schematic diagram displaying a potential complex of modified nucleosome and TRIM24.*

Following on from the non-denaturing MS results and the poor binding indicated by SEC it would be expected that the nucleosome–TRIM24 solution would not be a homogeneous sample of nucleosome complex. Rather, the solution would likely contain a mixture of complexes containing nucleosome and one or two molecules of TRIM24 as well as an excess of the unbound constituent species.

Previous non-denaturing MS/SEC experiments were conducted after incubating the components together for 1 or 12 hours at 4 °C. For transport reasons SAXS experiments

were conducted after incubating at 4 °C for one week, and the scattering data was collected at 20 °C. The long incubation followed by a temperature change may have resulted in a higher proportion of H3 sK18Ac nucleosome–TRIM24 complexes than observed by non-denaturing MS.

The contribution to the scattering profile from smaller molecules (*i.e.* TRIM24) is less than that of the larger molecules present, due to the smaller molecules having a lower number of electrons. By comparison, 90% of an H3 sK18Ac nucleosome–TRIM24 complex's mass is contributed by the 197.6 kDa H3 sK18Ac nucleosome, therefore unbound nucleosome could have a large impact on the mixture's scattering profile. This influence would likely manifest as an increase in density in the globular region of the mixture's P(r) plot. For this reason, it cannot be confirmed from the SAXS data that the shape of the round, nucleosome-like region observed in the molecular envelope was not influenced by a contaminant. However, the region of the P(r) plot above the $d_{max}$ of the nucleosome could not have been influenced by the folded nucleosome or TRIM24, alone indicating a significant conformational change in the modified nucleosome due to the presence of TRIM24.

## 4.7 Isothermal titration calorimetry

The observed high binding affinity for TRIM24 to H3 K18Ac peptide did not appear to translate, as predicted, to a stable H3 sK18Ac nucleosome–TRIM24 complex. ITC experiments wherein TRIM24 was titrated into H3 sK18Ac protein and nucleosome were conducted in order to investigate the effects of higher order structure on binding.

### 4.7.1 Background

In ITC experiments the calorimeter contains two identical cells, the sample cell and the reference cell. The sample cell is carefully maintained at the same temperature as the reference cell by means of an electric heater and constant mixing in both cells.[39] The protein of interest (in this case H3 sK18Ac histone or nucleosome in LSRB) is loaded into the sample cell, and the same buffer is added to the reference. Throughout the experiment a small current constantly heats the reference cell, resulting in a similar heating of the sample cell. A known quantity of binding partner (such as TRIM24) is injected into the sample cell. As the solutions are mixed minute temperature changes in the sample cell occur due to the binding event. The heater increases or decreases its power to maintain the same temperature between the cells. Additional quantities of binding partner are added in subsequent injections, and the change in energy consumption of the heater over time is recorded. As binding approaches saturation, the heat change occurring from additional ligand injections decreases, allowing the deduction of the stoichiometric ratio of the binder–ligand complex. The energy produced during each injection can be deduced by integrating the change in energy consumption during the injection and mixing. These energies of injection can be plotted against molar ratio, and the shape of the graph can be used to derive the stoichiometry

of binding and the $K_D$, as well as the enthalpy, entropy and Gibbs free energy change of the reaction.[39]

An advantage of ITC over other methods for determining binding constants such as AlphaScreen or fluorescence polarisation is that ITC does not require either binding partner to contain a tag. The disadvantage is that it can require several hundred microliters of protein at a high concentration to obtain good data.

### 4.7.2 Results

To prevent a contribution to the temperature change of mixing being caused by the combination of different buffers all proteins were dialysed into the same batch of buffer (LSRB) overnight (TRIM24 and H3 sK18Ac nucleosome) or dissolved directly into LSRB (H3 sK18Ac histone) prior to the commencement of this experiment. A heat of dilution experiment was conducted by titrating buffer into buffer and subtracted from subsequent spectra. Experiments measuring TRIM24 binding to sK18Ac histone or nucleosome were conducted by placing either the histone or the nucleosome into the calorimeter cell at 20 μM and injecting 1.3 μL aliquots of TRIM24 (400 μM) (data displayed in **Figure 4.17**).

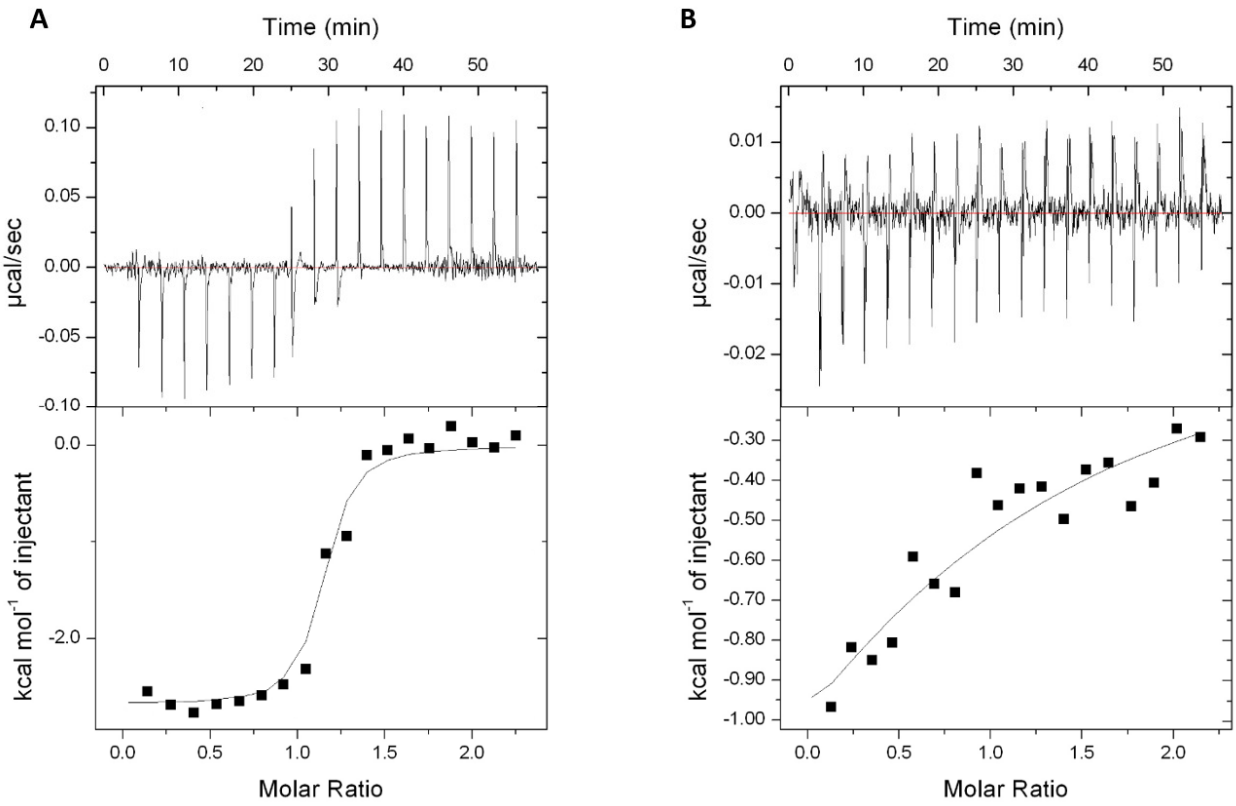



***Figure 4.17**: ITC experiments titrating TRIM24 (400 µM) into **A**- H3 sK18Ac histone (20 µM) and **B**- H3 sK18Ac nucleosome (20 µM).*

The batch of TRIM24 used in this experiment was previously used to measure a $K_D$ of 0.026 µM against $H3_{1-27}$ K18Ac peptide, similar to the reported ITC value of 0.021 µM.[8,9] In contrast, TRIM24 was found to have a $K_D$ of 1.56 µM and 52.9 µM against sK18Ac histone and nucleosome, respectively.

It should be noted that the histone and nucleosome differ from the peptide by more than just complexity; the peptides tested contained a native K18Ac mark whereas the histone and nucleosome had chemically modified sK18Ac marks (see **Figure 4.4**). As the sK18Ac histones can be recognised by α-KAc antibodies (see **Figure 4.5**), and the only difference between them and naturally acetylated lysine is a γ-thioether two carbons from the $N^{\varepsilon}$ atom, they were predicted to interact with TRIM24 with similar properties to the natural substrate. However, peptides containing sKAc groups have been found to be

inferior substrates for the histone deacetylase sirtuin-2 (SIRT2), becoming deacetylated at approximately 1/3 the rate of equivalently modified KAc peptides.[16] A 60-fold decrease in $K_D$ was observed for TRIM24 binding H3 sK18Ac histone *vs.* $H3_{1-27}$ K18Ac peptide, possibly due to the BRD having a high sensitivity to the change in side-chain length brought on by the thioether substitution. The increased complexity of the histone relative to the peptide may also have contributed to the differences in binding strength. For complete comparison, future ITC experiments should be used to establish the binding affinity of TRIM24 to $H3_{1-27}$ sK18Ac peptide.

A further 33-fold decrease in binding strength was observed for TRIM24 binding to modified nucleosome compared to the modified histone. In this case, the histone was identically modified to the nucleosome, as they both contained H3 sK18Ac from the same batch. The difference in $K_D$ must therefore have been due to the context in which the histone was situated. In a nucleosome the H3 tail is proximal to a range of different environments; the positively charged tails of neighbouring histones, the ordered histone core, and negatively charged DNA. Some features may offer opportunities for additional binding, while others provide barriers to accessing the binding site. SAXS data indicate that formation of a nucleosome complex may result in the displacement of DNA, potentially explaining the reduction in $K_D$.

## 4.8 Discussion

Peptide studies, while providing a versatile tool that is easy to work with, risk biasing our understanding of relevant transcription factor interactions away from the higher order structures present in a cellular setting. As our knowledge of epigenetic processes becomes more comprehensive, increasingly complex systems will need to be understood in order to further scientific knowledge in the field and identify future drug targets.

The work detailed in this chapter describes the process of identifying a potential target for nucleosome complex formation, the design and production of a modified nucleosome, and investigations into the binding of reader protein TRIM24 and its modified nucleosome binding partner. Crystal structures of the component proteins of the nucleosome complex examined in this report have been determined and the ability of an H3 sK18Ac nucleosome to bind up to two molecules of TRIM24 has been observed by non-denaturing MS. SAXS data of a modified nucleosome in the presence of TRIM24 has been obtained and suggests some potential conformational changes of the nucleosome complex occurs. ITC results reveal TRIM24's preference for binding to an H3 sK18Ac histone protein over a similarly modified nucleosome.

The observed decrease in the TRIM24 binding affinity to increasingly physiologically representative species appears counterintuitive; however, it is worth considering how TRIM24 was selected for this study in the first place. Peptide studies indicated TRIM24 contained a BRD with the highest observed binding affinity to a single H3 acetylation mark. However, the affinity did not appear to translate to tight binding to an intact nucleosome *in vitro*. As with peptide studies, working with truncated proteins is often easier than working with the full-length construct. The TRIM24 construct used in the

experiments detailed in this chapter contained only the PHD and BRD identified as important for peptide binding. These domains represent less than 1/5 of the full-length protein and lack several domains that likely play a role in nucleosome binding, including DNA binding domains. Larger, full-length transcription factors may be promising future targets for nucleosome crystallography, particularly if they contain multiple regions, such as DNA binding domains, capable of binding different sites in a nucleosome simultaneously. Larger transcription factor constructs may also be better models for the endogenous environment of the nucleus, allowing for the probing of more biologically relevant nucleosome–transcription factor interactions.

The presence of an acetylated lysine analogue may have also impinged on the reader protein's ability to bind as tightly as was observed for native KAc. Other less disruptive methods could be investigated in future, perhaps in combination, such as translational incorporation of KAc at position 18, followed by chemical modification at a C9 residue to permit the incorporation of the tightly binding K4, sK9Me$_3$, K18Ac combination into a single nucleosome.

The Widom 601 DNA sequence selected for these studies binds in a nucleosome more tightly than any identified naturally occurring DNA sequence.[21] The stability that affords the nucleosome core particle makes it an attractive feature for any experiment that requires homogeneity. It is possible that the unnaturally tight binding of the Widom 601 DNA sequence to the histone core interferes with the ability of TRIM24 to access and bind the nearby H3 tail. Perhaps a less tightly binding DNA sequence could be used to form a more stable nucleosome complex in future, particularly in the case of transcription factors that interact with histone tails. Nucleosomes containing less tightly

binding DNA may also provide a more biologically relevant model when probing interactions between transcription factors and nucleosomes.

In conclusion, a structure for a nucleosome complex with a modified H3 histone tail interacting with a transcription factor containing an epigenetic reader domain remains elusive. TRIM24 showed much lower affinity to a modified nucleosome than to an analogously modified histone. This indicates that future attempts to trap such a complex should focus on transcription factors that are shown to interact strongly with higher order structures.

Epigenetic reader domains may have a functionally important role in unwinding DNA from the nucleosome in chromatin remodelling processes. The observations detailed in this chapter may suggest such a role for TRIM24. Further studies will need to be performed to better understand the observed phenomena.

## 4.9 References


1. Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467,** 562–6 (2010).
2. McGinty, R. K., Henrici, R. C. & Tan, S. Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature* **514,** 591–596 (2014).
3. Armache, K. J., Garlick, J. D., Canzio, D., Narlikar, G. J. & Kingston, R. E. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science* **334,** 977–82 (2011).
4. Morgan, M. T. *et al.* Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science* **351,** 725–728 (2016).
5. Wilson, M. D. *et al.* The structural basis of modified nucleosome recognition by 53BP1. *Nature* **536,** 100–103 (2016).
6. Tsai, W. W. *et al.* TRIM24 links a non-canonical histone signature to breast cancer. *Nature* **468,** 927–932 (2010).
7. Thénot, S., Henriquet, C., Rochefort, H. & Cavaillès, V. Differential interaction of nuclear receptors with the putative human transcriptional coactivator hTIF1. *J. Biol. Chem.* **272,** 12062–12068 (1997).
8. Xi, Q. *et al.* A poised chromatin platform for TGF-β access to master regulators. *Cell* **147,** 1511–24 (2011).
9. Sekirnik, A. Probing tandem epigenetic reader domains. (University of Oxford, 2015).
10. Philpott, M. *et al.* Assessing cellular efficacy of bromodomain inhibitors using fluorescence recovery after photobleaching. *Epigenetics Chromatin* **7,** 14 (2014).
11. Robinson, P. J. J. *et al.* 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *J. Mol. Biol.* **381,** 816–25 (2008).
12. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342,** 357–360 (2013).
13. Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding $N^{\varepsilon}$-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **4,** 232–234 (2008).
14. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266,** 776–779 (1994).
15. Simon, M. D. *et al.* The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128,** 1003–12 (2007).
16. Li, F. *et al.* A direct method for site-specific protein acetylation. *Angew. Chem. Int. Ed. Engl.* **50,** 9611–4 (2011).
17. Frouws, T. D., Duda, S. C. & Richmond, T. J. X-ray structure of the MMTV-A nucleosome core. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 1214–1219 (2016).

18. 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,** 251–60 (1997).

19. Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375,** 23–44 (2004).

20. Satchwell, S. C., Drew, H. R. & Travers, A. A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191,** 659–675 (1986).

21. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276,** 19–42 (1998).

22. Vasudevan, D., Chua, E. Y. D. & Davey, C. A. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J. Mol. Biol.* **403,** 1–10 (2010).

23. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).

24. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

25. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68,** 352–367 (2012).

26. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

27. Vogler, C. *et al.* Histone H2A *C*-terminus regulates chromatin dynamics, remodeling, and histone H1 binding. *PLoS Genet.* **6,** 1–12 (2010).

28. Wang, C. Y. *et al.* The *C*-terminus of histone H2B is involved in chromatin compaction specifically at telomeres, independently of its monoubiquitylation at lysine 123. *PLoS One* **6,** e22209 (2011).

29. Sanchez, R. & Zhou, M. M. The PHD finger: A versatile epigenome reader. *Trends. Biochem. Sci.* **36,** 364–72 (2011).

30. Lorenzen, K. & van Duijn, E. Native mass spectrometry as a tool in structural biology. *Curr. Protoc. Protein Sci.* **12,** 62–121 (2010).

31. Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74,** 1402–1407 (2002).

32. Lössl, P. *et al.* The diverse and expanding role of mass spectrometry in structural and molecular biology. *EMBO J.* **16,** 155–166 (2016).

33. Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589,** 2570–2577 (2015).

34. Debye, P. & Bueche, A. M. Scattering by an inhomogeneous solid. *J. Appl. Phys.* **20,** 518–525 (1949).

35. Tuukkanen, A. T. & Svergun, D. I. Weak protein–ligand interactions studied by small-angle X-ray scattering. *FEBS J.* **281,** 1974–1987 (2014).

36. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: Defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40,** 191–285 (2007).

37. Yang, C., Van Der Woerd, M. J., Muthurajan, U. M., Hansen, J. C. & Luger, K. Biophysical analysis and small-angle X-ray scattering-derived structures of MeCP2-nucleosome complexes. *Nucleic Acids Res.* **39,** 4122–4135 (2011).

38. Arimura, Y. *et al.* Structural basis of a nucleosome containing histone H2A.B/H2A.Bbd that transiently associates with reorganized chromatin. *Sci. Rep.* **3,** 3510 (2013).

39. Du, X. *et al.* Insights into protein–ligand interactions: Mechanisms, models, and methods. *Int. J. Mol. Sci.* **17,** 144 (2016).

# Chapter 5: Development of Assays for Monitoring the Demethylation of Recombinant Nucleosomes

## 5.1 Introduction

### 5.1.1 K79 methylation mark

As described in **Chapter 4**, studies employing peptide fragments do not always reflect enzyme-substrate selectivity with proteins. An illustration of this can be found in the action of disruptor of telomeric silencing 1-like (DOT1L), the only identified methyltransferase capable of methylating the K79 position of H3.[1,2] Unlike modification marks on the transiently structured histone tails, H3 K79 is positioned in a comparatively structured environment. H3 K79 is located in the histone core on the L1 loop of H3, in close proximity to the DNA and the L2 loop of H4 (see **Figure 5.1**). It appears likely that this structured environment around K79 contributes to the recognition of DOT1L to the modified position. Unlike most other lysine methyltransferases, DOT1L has only been found to methylate K79 in the context of a nucleosome, and does not methylate free histones or peptides.[1]

Methylated H3 K79 is unique amongst lysine methylation marks in that is has no currently identified demethylase. The ability to investigate the effect of a purified protein on a target *in vitro* is a useful method for deconvoluting the complicated interplay of epigenetic regulators in cells. Studies attempting to identify novel demethylases often employ peptide studies, however, in this instance a homogeneously modified nucleosome may be a better search model.

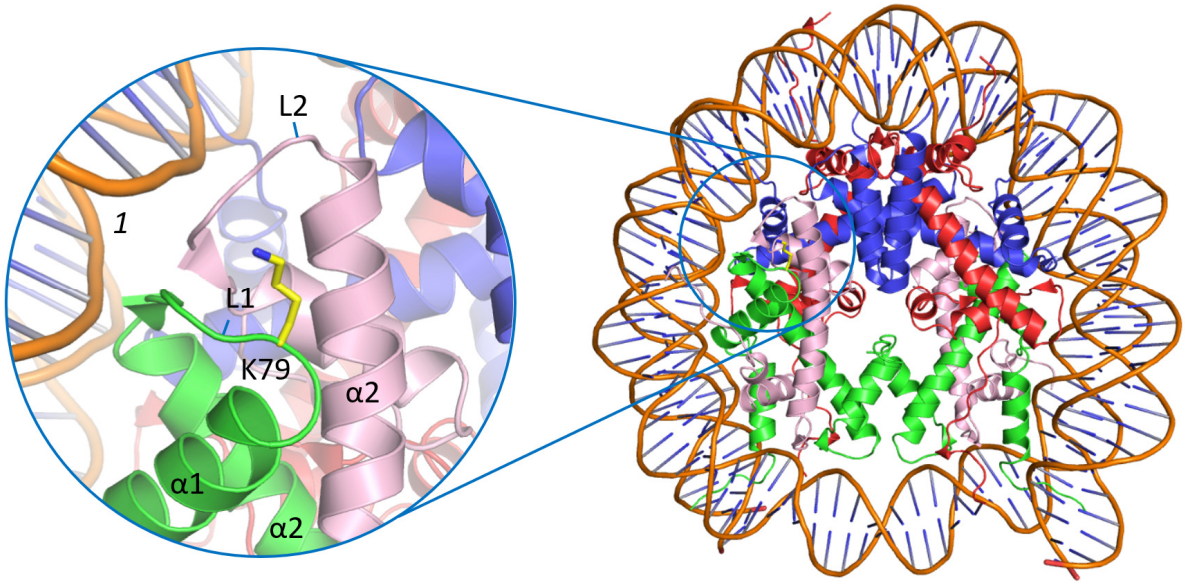



***Figure 5.1**: Views from a crystal structure of the H3 sK18Ac nucleosome described in **Chapter 4** showing the location of H3 K79. H2A is displayed in red, H2B in blue, H3 in green, H4 in light pink and DNA in grey with an orange phosphate backbone. The side-chain of H3 K79 is shown in yellow. The magnified region shows the quaternary structure of the nucleosome proximal to the L1 loop on which the H3 K79 site is located, including the α1 and α2 helices of H3, the L2 loop and α2 helix of H2B, the and the minor groove of the* (±)1 *turn of DNA.*

### 5.1.2 Methods for monitoring demethylation reactions catalysed by 2OG-oxygenases

Many different methods can be used to follow the progression of 2OG-oxygenase-catalysed demethylation reactions, including monitoring the loss of starting material, production of product, or following 2OG turnover/succinate production.[3,4] The production of the product can be measured directly by methods such as by NMR, or MS, or indirectly when the action of the enzyme changes the structure of the product sufficiently to alter the binding of the product to antibodies. Antibody systems can be coupled to light-inducing compounds/proteins, which can be subsequently detected.[5]

## 5.2 Objectives

The work in this chapter describes the development and application of a sensitive assay capable of monitoring demethylation experiments on recombinant nucleosome substrates. Mass spectrometry and antibody-based techniques were investigated as potential methods for following demethylation reactions. 'Dot blot' assays were identified as a preferred approach due to their high sensitivity and ease of use.[6]

Dot blot assays were first used to investigate the difference in reactivity between a histone demethylase; KDM4A, and a target H3 (s)K9$Me_3$ substrate in the context of a peptide, histone and nucleosome. A small panel of 2OG-dependent KDMs were then subjected to the dot blot-based demethylation assay using a homogeneously modified H3 sK79$Me_3$ nucleosome as a substrate in an initial step towards identifying an H3 K79 demethylase.

*2OG-dependent KDM proteins were supplied by Dr Rebecca Hancock (KDM4A), Joanna Bonnici (KDM4E), Dr Sarah Wilkins (JMJD4, JMJD5 and JMJD7), Dr Saiful Islam (JMJD6) and Yijia Zhang (PHF2).*

## 5.3 Production of peptide and nucleosome methylated lysine analogues

In a method analogous to the production of acetylated lysine mimics described in **Section 4.3.3**, methylated lysine mimics can be produced through alkylation at cysteine residues (see **Figure 5.2**).[7] This method results in the formation of a lysine analogue with a thioether at the γ-position, but as previously discussed, based on available evidence, this change from the native methylene bridge has not been considered to be detrimental to function.[7]

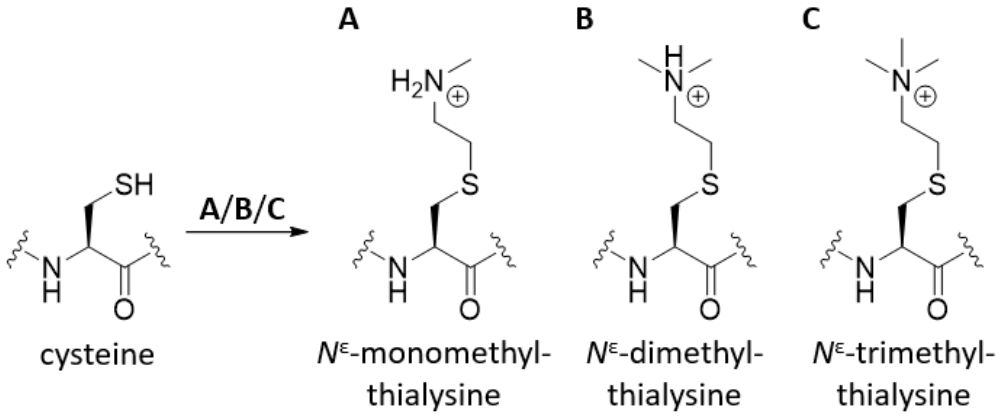



***Figure 5.2**: Scheme showing synthetic methods for producing methylated lysine analogues by chemical reaction with cysteine. **A**- 500 μM cysteine sites, 300 mM (2-chloroethyl)-methylammonium chloride, 20 mM DTT, 25 °C, 5 hours.[8] **B**- 500 μM cysteine sites, 50 mM (2-chloroethyl)-dimethylammonium chloride, 20 mM DTT, 25 °C, 4 hours.[8] **C**- 500 μM cysteine sites, 400 mM (2-bromoethyl) trimethylammonium bromide, 20 mM DTT, 50 °C, 5 hours.[8]*

In order to investigate if an assay could differentiate between nucleosomes containing H3 sK79Me$_3$ and the demethylated product, both H3 sK79Me$_3$ and H3 sK79Me$_2$ were produced. As a positive control, any demethylation assay developed would be used to monitor an established demethylation reaction. KDM4A has been found to demethylate H3$_{1\text{-}15}$ K9Me$_3$ peptide *in vitro*, therefore peptide and nucleosome constructs containing H3 sK9Me$_3$ substitutions were produced.[9,10] A recombinant construct of KDM4A

containing the Jumonji-*N* (JmjN) and JmjC domains required for catalytic activity was supplied by *Dr Rebecca Hancock*.

### 5.3.1 Synthesis of methylated lysine analogues

H3 K9C, and H3 K79C variants were obtained by conducting SDM on an *H3 (C11A)* construct (**Materials and Methods 8.4.3)**. Histones were expressed and purified as described in **Materials and Methods 8.5.7** and **8.6.1**, respectively. The $H3_{1-15}$ K9C peptide was synthesised and purified as per the method described in **Section 3.3.2**. Prior to alkylation, 0.5 μmol H3 histones or $H3_{1-15}$ peptides were dissolved in alkylating buffer (1 mL) before addition of DTT (to 20 mM). The reducing solution was then incubated for 1 hour at 37 °C.

#### Method for producing dimethylated lysine analogues

To produce the desired dimethylated lysine analogues, the reduced histones were reacted with 2-dimethylaminoethyl chloride HCl for 4 hours at RT with stirring, before the reaction was quenched with addition of β-mercaptoethanol (BME), following the procedure described by Simon *et al*.[7] The progression of the reaction was monitored by LC-MS, and no unreacted histone was detected after 4 hours of reaction (see **Figure 5.3**). The histone products were purified by buffer exchange into 2 mM BME solution using a PD-10 gel filtration column before lyophilisation and storage at -80 °C. Typical yields for protein recovery were 80%.

#### Method for producing trimethylated lysine analogues

To a solution containing reduced H3 histone/peptide was added (2-bromoethyl) trimethylammonium bromide, followed by shaking for 5 hours at 50 °C and quenching with BME. The progression of the reaction was monitored by LC-MS (histones) or

MALDI-TOF MS (peptides), and no unreacted histone/peptide was detected after 5 hours of reaction (see **Figure 5.3**). Full experimental details of histone/peptide alkylation reactions are provided in **Materials and Methods 8.7.2** and **8.7.3**. Typical yields for protein/peptide recovery were 80% for histones and 70% for peptides.

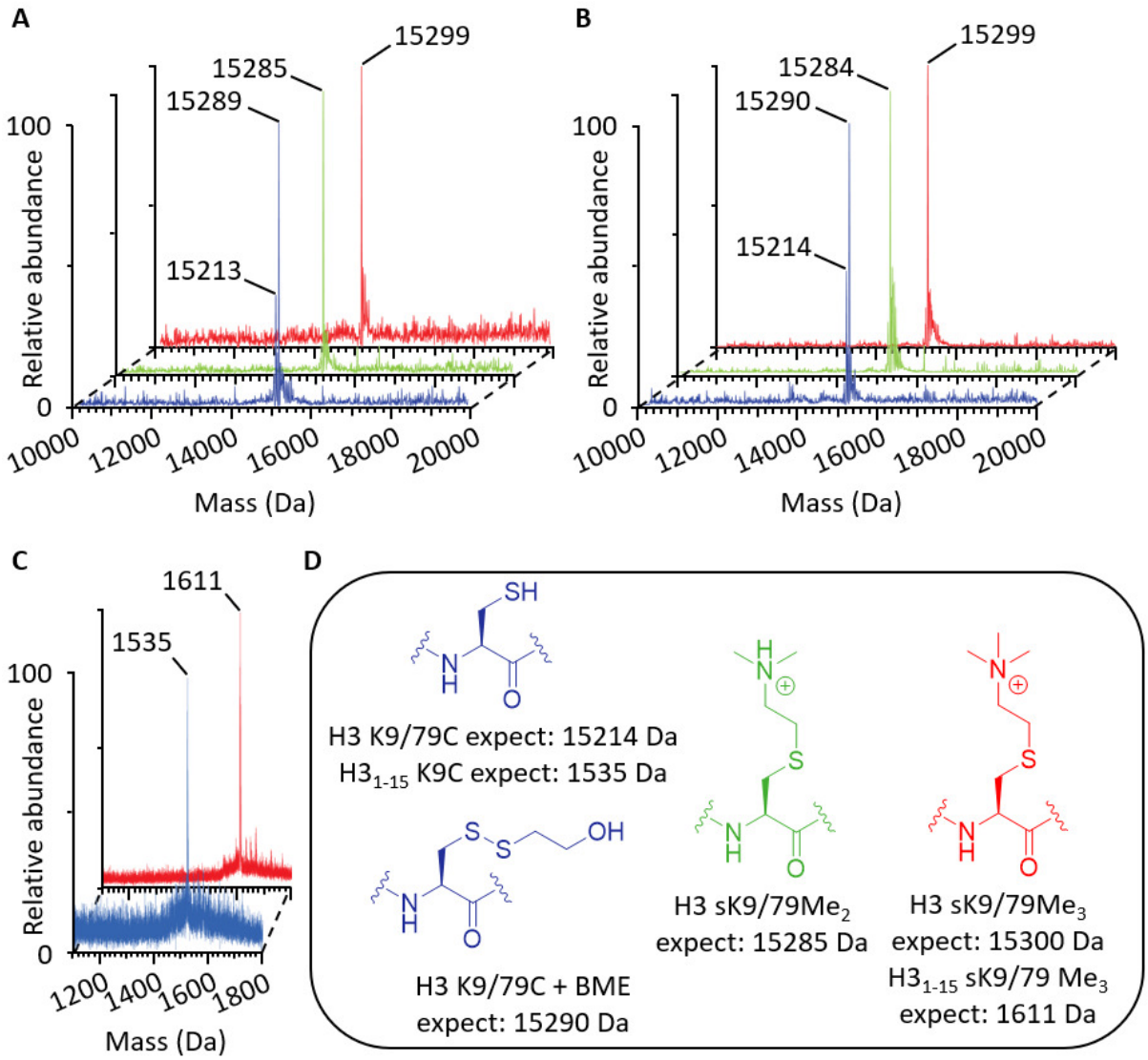



***Figure 5.3**: MS analyses of histone and peptide alkylation. Deconvoluted spectra of unreacted, and reacted H3 K9C histone are provided in **A**, H3 K79C histone in **B**, and $H3_{1-15}$ K9C peptide in **C**. Unreacted histone/peptide is displayed in blue, $sK9/79Me_2$ product in green, and $sK9/79Me_3$ product in red. **D**- Expected masses of products corresponding to the major peaks are provided, along with structures of the respective modifications.*

Note that in the presence of BME the major peak observed in the MS spectra of histones containing unreacted cysteines was the BME adduct. As the alkylation of cysteine was proceeded by a reduction step, this did not appear to adversely affect alkylation. Differences of ±1 Da between the expected and observed mass likely arose from mass spectrometer calibration errors.

### 5.3.2 Reconstitution of nucleosomes containing methylated lysine analogues

Following the processes detailed in **Sections 4.3.3-4.3.6** the modified H3 proteins were combined with H2A, H2B and H4, then refolded into an octameric histone core before reconstitution into a nucleosome with addition of 145 bp Widom 601 DNA. The homogeneity of the nucleosome product was assessed by 6% TBE non-denaturing PAGE analysis (see **Figure 5.4**).

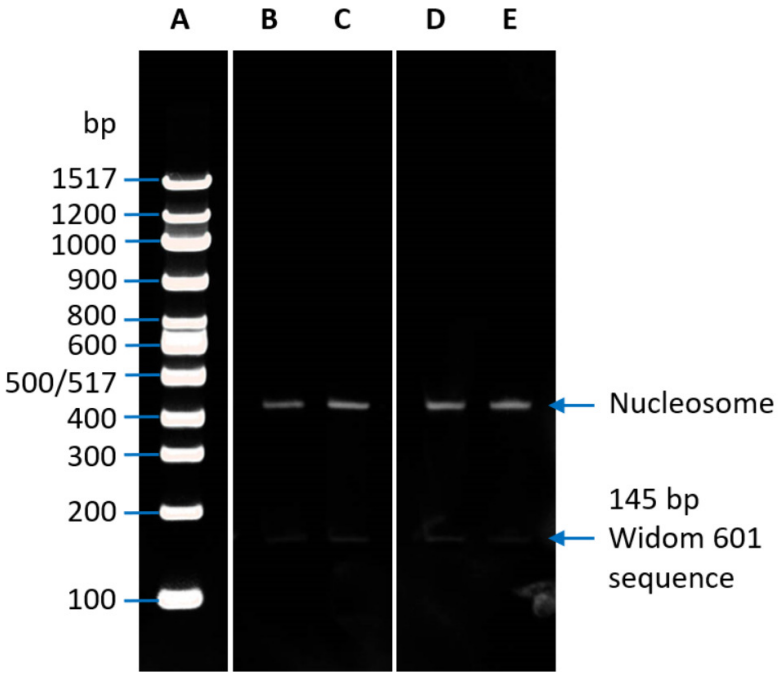



***Figure 5.4**: 6% TBE non-denaturing PAGE of homogeneous reconstituted nucleosomes. **A**- 100 bp ladder (NEB), **B**- H3 sK9Me$_2$ nucleosome, **C**- H3 sK9Me$_3$ nucleosome, **D**- H3 sK79Me$_2$ nucleosome, **E**- H3 sK79Me$_3$ nucleosome. As in previous nucleosome preparations (see **Figure 4.8**) a small quantity of free 145 bp Widom 601 DNA was visible in each of the wells containing nucleosome (>5% intensity relative to the nucleosome band).*

## 5.4 Monitoring demethylation by mass spectrometry

Quadrupole-time of flight mass spectrometry (Q-TOF MS) was selected as an initial method for monitoring reaction progression directly. Prior to MS analysis, solutions containing nucleosome were mixed with unfolding buffer (6 M guanidinium chloride, 20 mM tris-HCl pH 7.5, 1 mM DTT) in a 1:3 ratio. The unfolded proteins were then analysed by ultra-performance liquid chromatography (UPLC) Q-TOF MS (see **Figure 5.5**). Experimental details are provided in **Materials and Methods 8.11.2**.

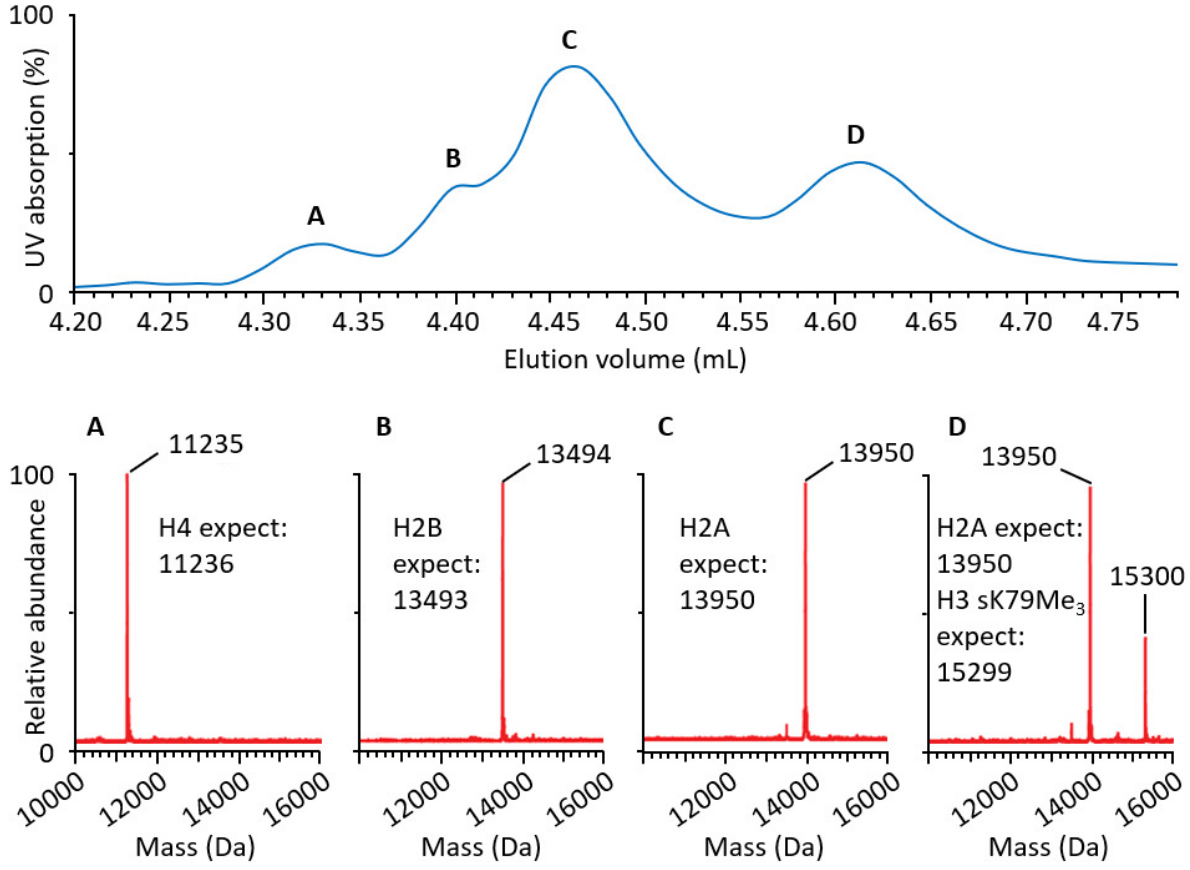



***Figure 5.5**: **Top**- Elution profile of histones from an unfolded 1 µM sample of H3 sK9Me3 nucleosome separated by an RP-18 2 mm × 50 mm reverse phase chromatography column. **Bottom**- Deconvoluted MS spectra of the major peaks **A**-**D**, corresponding to H4, H2B, H2A and a co-elution of H2A and H3, respectively.*

Separation using an RP-18 2 mm × 50 mm reverse phase chromatography column resulted in the elution of histones in four peaks. Three peaks contained a single histone

species, and the fourth contained a co-elution of histones H2A and H3 (see **Figure 5.5**). Attempts to separate H2A and H3 were unsuccessful, however both proteins could be clearly resolved in the deconvoluted spectra. Samples containing nucleosome concentrations lower than 3 μM, and samples in which the nucleosome was not unfolded prior to loading onto the column, resulted in poorly resolved or absent H3 peaks.

To assess the utility of this method for following demethylation experiments, equimolar samples of H3 sK79Me$_2$ nucleosome and H3 sK79Me$_3$ nucleosome were combined and subjected to UPLC Q-TOF MS analysis (see **Figure 5.6**).

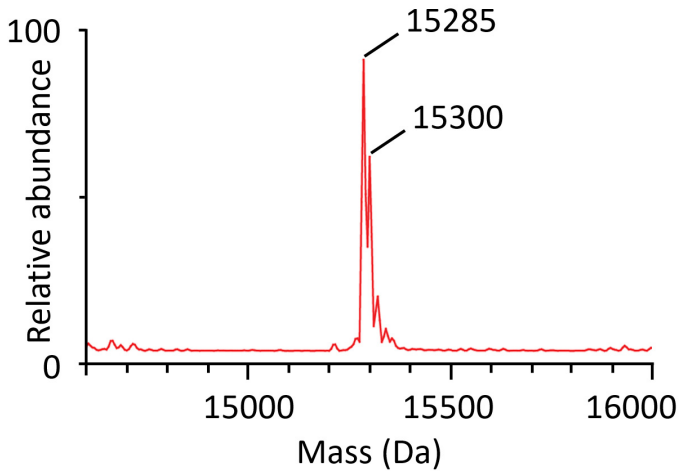



***Figure 5.6**: Deconvoluted MS spectra of H3 elution peak from a sample containing a 50:50 mixture of H3 sK9Me$_2$, and H3 sK9Me$_3$ nucleosome. The expected mass of H3 sK79Me$_2$ histone is 15285, and the expected mass of H3 sK79Me$_3$ histone is 15300.*

UPLC Q-TOF MS was able to resolve between the two H3 sK79 methylation states; however, for strong peaks to be observed, a high concentration of nucleosome was required (>20 μM nucleosome prior to dilution into unfolding buffer). It was therefore decided to investigate a more sensitive method for monitoring nucleosome demethylation.

## 5.5 Antibody-based detection methods

### 5.5.1 Western blots

A western blot is an antibody-based method for detecting proteins or protein modifications.[11] In western blotting, proteins are initially separated on a gel, *e.g.* by SDS-PAGE. Proteins are then transferred from the gel to a blotting membrane through application of an electric potential perpendicular to the plane of the gel. Once transferred, the remainder of the membrane is then saturated for binding through incubation in a protein solution such as 5% (v/v) fetal bovine serum (FBS) solution. This ensures that the membrane is saturated with proteins that won't interact with the primary (1°) or secondary (2°) antibodies used.[11] The membrane is washed and incubated in a solution containing a 1° antibody raised against a feature of interest, *e.g.* H3, or H3 K9Me$_3$. A 2° antibody raised against the organism that the 1° antibody was produced by is then incubated with the membrane. This 2° antibody is often conjugated to a species capable of producing a detectable readout such as a fluorophore (used in this chapter) or bound to an enzyme capable of catalysing a fluorogenic reaction such as horseradish peroxidase (used in **Section 4.3.3**).[12–14] Where fluorophores are used, different colours can be associated with the 2° antibodies, thereby allowing for the potential to detect multiple different marks on the same sample.[15]

### 5.5.2 Investigating KDM4A-catalysed H3 sK9Me$_3$ nucleosome demethylation by western blot

Reported demethylation experiments exploring demethylation of H3 sK9Me$_3$ nucleosomes used a large excess of demethylase (100-1000-fold) relative to the substrate.[7,10] A series of demethylation experiments using different concentrations of KDM4A was carried out in order to establish a more precise estimate of the required

enzyme concentration. To a solution containing H3 sK9Me$_3$ nucleosome (10 μM) in LSRB was added 2OG (300 μM), ascorbate (300 μM) and Fe(II) (30 μM), followed by a 4-step 5-fold dilution series of KDM4A (25-0.008 μM). The reaction was incubated for 1 hour at 37 °C before quenching with addition of a 3-fold excess of unfolding buffer. 20 μL samples were separated by SDS-PAGE using a 16% polyacrylamide gel, before transferring onto a blotting membrane. The blotting membrane was charged with mouse α-H3 1° antibody (ab10799) and rabbit α-H3 K9Me$_3$ 1° antibody (ab8898), followed by goat α-mouse 2° antibody conjugated to Alexa Fluor®790 (green) (ab186695) and goat α-rabbit 2° antibody conjugated to Alexa Fluor®680 (red) (ab186696). The fluorescent dye-labelled 2° antibodies enabled visualisation of protein bands containing H3 and H3 K9Me$_3$, through illumination at λ = 790 nm and 680 nm, respectively (see **Figure 5.7**). Experimental details are provided in **Materials and Methods 8.2.2**, and buffer compositions are provided in **Materials and Methods 8.9.1**.

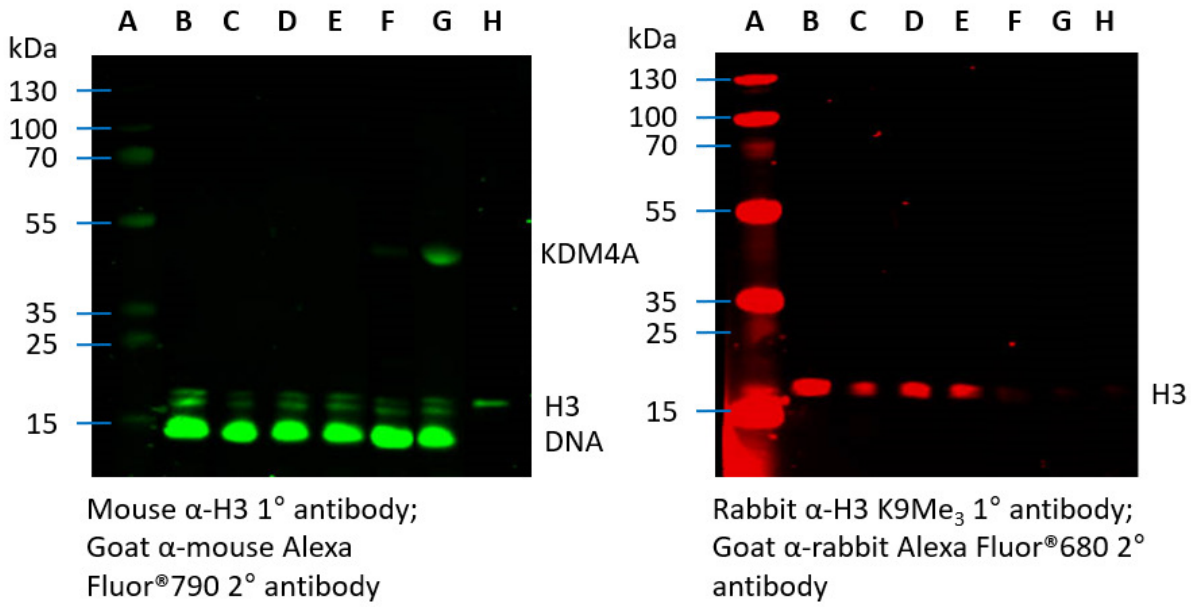



***Figure 5.7**: Western blot analyses presenting the products of a 1-hour demethylation reaction between H3 sK9Me3 nucleosome (10 μM) and varying concentrations of KDM4A. **Left**- 790 nm channel, revealing H3 protein. **Right**- 680 nm channel, revealing H3 K9Me3. **A**- PageRuler Plus™ Pre-stained Protein Ladder (Thermo Fisher Scientific). **B**-**G**- Reactions containing 0, 0.04, 0.2, 1, 5 and 25 μM KDM4A, respectively. **H**- H3 K9Me2 octamer (10 μM). Wells **B**-**H** also contained 2OG (300 μM), ascorbate (300 μM) and Fe(II) (30 μM).*

In each of the nucleosome-based experiments the H3 band appears to be split into two bands of comparative intensity in the 790 nm channel, possibly due to incomplete unfolding. The double H3 band was not observed in the samples prepared from octameric histone. The resolution was insufficient to resolve whether a single or double H3 band was illuminated in the 680 nm channel.

Non-specific binding of either the mouse α-H3 1° antibody, or goat α-mouse 2° antibody resulted in the visualisation of multiple bands under λ = 790 nm illumination (see **Figure 5.7**, **left**). These same bands were observed when a different (mouse α-H3 K9Me2; ab1220) 1° antibody was used (figure not shown), indicating the non-specific binding was due to the goat α-mouse 2° antibody. The intense band at ~13 kDa, absent in the octamer

condition is likely to be DNA. KDM4A (MW 41.7 kDa) also appears to interact with the goat α-mouse 2° antibody.

Only the H3 band (and the ladder) appears to be illuminated in the λ = 680 nm channel associated with the α-H3 K9Me$_3$ 1° antibody. A clear decline in intensity for the H3 band at increasing KDM4A concentrations was observed in the 680 nm channel. At the highest KDM4A:H3 sK9Me$_3$ nucleosome ratio tested (2.5:1) fluorescence of the H3 band in the 680 nm channel decreased to 5% intensity relative to the nucleosome demethylation condition containing no KDM4A (see **Figure 5.7**, condition **B**). It should be noted that the octameric H3 sK9Me$_2$ condition also produced an H3 band visible to low intensity (5% of condition **B**). These results indicate that; a) the rabbit α-H3 K9Me$_3$ 1°–goat α-rabbit 2° antibody system was selective for H3 sK9Me$_3$ over H3 sK9Me$_2$, and that b) KDM4A was able to demethylate H3 sK9Me$_3$ nucleosome.

### 5.5.3 Dot blots

Dot blots function similarly to western blots, but do not require a gel separation step, permitting the direct application of a sample to the membrane.[6] Subsequent blocking, antibody incubation and imaging steps proceed in the same way as with western blots. This can have advantages over western blotting as it is much faster, cheaper, and requires less sample, so is a desirable alternative to western blotting when the antibody system used is selective enough. The western blot experiments described in **Section 5.5.2** showed that by using a rabbit α-H3 K9Me$_3$ 1°–goat α-rabbit 2° antibody system it was possible to detect demethylation of sK9Me$_3$ residues. An absence of significant signal being observed for the other components of the reaction indicated that the rabbit α-H3 K9Me$_3$ 1°–goat α-rabbit 2° antibody system was indeed selective enough to use in dot blot experiments.

### 5.5.4 Comparing KDM4A-catalysed demethylation of peptide, histone and nucleosome substrates by dot blot

Samples of H3 sK9Me$_3$ peptide (20 μM), histone (20 μM) and nucleosome (10 μM), as well as H3$_{1\text{-}15}$ peptide containing the natural K9Me$_3$ methylation mark (20 μM) (supplied by *Dr Rebecca Hancock*) in LSRB were prepared. Double concentrations of peptide and histone were used relative to the concentration of nucleosome as each nucleosome contains two H3 sK9Me$_3$ proteins. To each of these solutions was added 2OG (300 μM), ascorbate (300 μM), Fe(II) (30 μM) and a 9-step 2-fold dilution series of KDM4A (50-0.1 μM). A 0 M KDM4A condition for each of the H3 species was also prepared. The mixtures were incubated for 1 hour at 37 °C before quenching with addition of a 3-fold excess of unfolding buffer. The unfolded sample (2 μL) was spotted onto a blotting membrane before blocking with 5% (v/v) FBS solution. The membrane was charged with α-H3 K9Me$_3$ 1° antibody followed by goat α-rabbit–Alexa Fluor®680 2° antibody. The intensity of the spots in the 680 nm channel was measured and quantified, normalising against the background fluorescence. The results were plotted as $\log_{10}$([KDM4A (M)]) *vs.* % fluorescence observed relative to the 0 M KDM4A control. A variable slope (four parameter) curve was fitted using GraphPad Prism (see **Figure 5.8**) and used to calculate the concentration of KDM4A that was effective at reducing the trimethylated lysine signal by half (EC$_{50}$) for each experiment.[16]

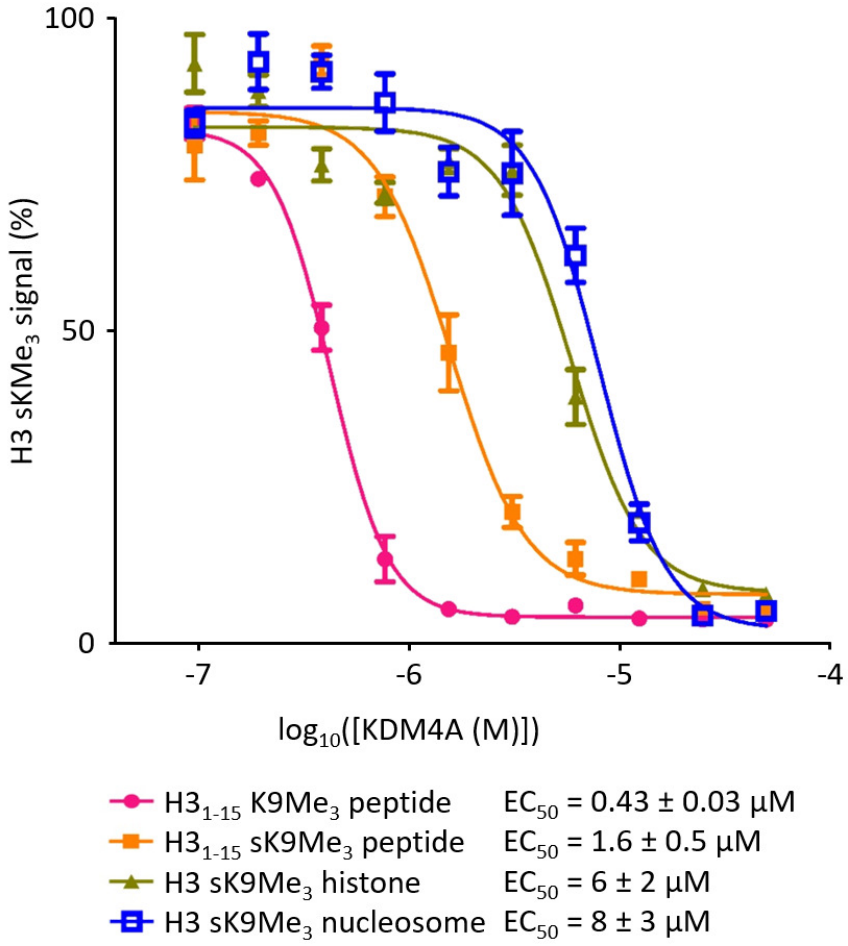



***Figure 5.8**: Line graph displaying dot blot data for KDM4A-catalysed demethylation reactions. The concentration of KDM4A effective in reducing H3 sKMe3 signal to 50% ($EC_{50}$) are presented for each species of H3 tested ± standard error from three separate experiments.*

The results of the demethylation experiments indicate KDM4A is more active against peptides than histones or nucleosomes, at least over a 1-hour reaction time. $H3_{1-15}$ peptides containing the natural $K9Me_3$ modification appear to be better substrates than $H3_{1-15}$ $sK9Me_3$ peptides, by a factor of ~4 (0.43 ± 0.03 µM *vs.* 1.6 ± 0.5 µM $EC_{50}$, respectively). This observation is consistent with previously published histone demethylase reactions on $sKMe_3$ peptides.[7,10] A further decline of ~4-fold activity was seen between the $H3_{1-15}$ $sK9Me_3$ peptide and H3 $sK9Me_3$ histones (1.6 ± 0.5 µM *vs.* 6 ± 2

μM $EC_{50}$, respectively). KDM4A appeared to display similar activity against H3 $sK9Me_3$ histone and nucleosome substrates (6 ± 2 μM *vs.* 8 ± 3 μM $EC_{50}$, respectively).

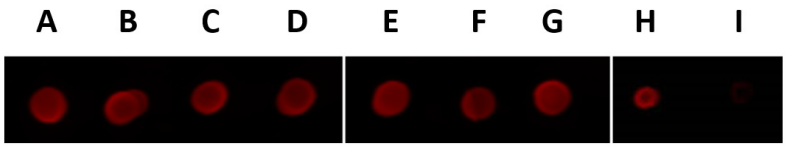
### 5.5.5 Application of dot blot demethylation in the hunt for an H3 K79 demethylase

The dot blot method detailed above provided an efficient, quantitative platform for monitoring nucleosome demethylation. It was therefore decided to employ this dot blot method in the search for an H3 $K79Me_3$ demethylase. The rabbit α-H3 $K9Me_3$ 1°–goat α-rabbit 2° antibody system was found to be appropriate for monitoring H3 $sK79Me_3$ demethylation, displaying preferential binding to spots containing H3 $sK79Me_3$ nucleosome over spots containing H3 $sK79Me_2$ nucleosome (see **Figure 5.9**).

#### H3 $sK79Me_3$ nucleosome demethylation experiments

A small panel of 2OG-oxygenases with potential, or demonstrated, KDM activity (KDM4A, KDM4E, Jumonji domain-containing protein (JMJD)4-7 and PHD finger protein 2 (PHF2)) were selected for initial demethylation experiments against the H3 $sK79Me_3$ nucleosome. Samples of purified KDM4A was supplied by *Dr Rebecca Hancock*, KDM4E by *Joanna Bonnici*, JMJD4, JMJD5 and JMJD7 by *Dr Sarah Wilkins*, JMJD6 by *Dr Saiful Islam* and PHF2 by *Yijia Zhang*.

To a solution of H3 $sK79Me_3$ nucleosome (10 μM) in LSRB was added 2OG (300 μM), ascorbate (300 μM) and Fe(II) (30 μM), followed by a 2OG-oxygenase (25 μM). Reactions were incubated for 1 hour at 37 °C before quenching with addition of a 3-fold excess of unfolding buffer. The reaction mixtures were then analysed by dot blot (see **Figure 5.9**).




Rabbit α- 'H3 K9Me$_3$' 1° antibody;
Goat α-rabbit Alexa Fluor®680 2° antibody

***Figure 5.9**: Dot blot analyses following a 1-hour incubation of H3 sK79Me$_3$ nucleosome (10 μM) under demethylating conditions in the presence of **A**- KDM4A; **B**- KDM4E; **C**- JMJD4; **D**- JMJD5; **E**- JMJD6; **F**- JMJD7; **G**- PHF2 (all 25 μM). **H**- is an H3 sK79Me$_3$ nucleosome no enzyme control, and **I**- is an H3 sK79Me$_2$ nucleosome no enzyme control. The spots were illuminated under light with λ = 680 nm.*

H3 sK79Me$_3$ nucleosomes did not appear to be significantly demethylated by any of the 2OG-oxygenases tested. No fluorescence was observed in 2OG-oxygenase controls (figure not shown), indicating no non-specific antibody binding to any of the enzymes occurred.

## 5.6 Discussion

A sensitive assay capable of monitoring lysine demethylation on substrates of varying complexity has been developed. Analysis through dot blots means large numbers of experiments can be probed quickly, cheaply and in parallel, with only 20 pmol of nucleosome required per experiment. Following initial hits identified by the dot blot method, additional validation could be provided, *e.g.* by LC-MS.

$H3_{1\text{-}15}$ sK9$Me_3$ peptide was shown to be recognised and demethylated by KDM4A, albeit with a 4-fold lower activity than an $H3_{1\text{-}15}$ K9$Me_3$ peptide. This verifies the utility of synthetically-incorporated thioether methylated lysine marks as a useful tool for experiments requiring homogeneously modified histones or nucleosomes.

KDM4A has been shown to be less effective at demethylating histones and nucleosomes by a factor of ~4 *vs*. peptides, highlighting the importance of substrate complexity on enzyme activity. In **Chapter 4**, TRIM24 was shown to bind to H3 sK18Ac histone with a 33-fold lower $K_D$ than to an H3 sK18Ac nucleosome. It is therefore interesting to note that KDM4A appears able to demethylate H3 sK9$Me_3$ histone, and nucleosome substrates with similar efficiency. One important factor that may have contributed to this apparent discrepancy is that the H3 K9$Me_3$ site is located further from the histone core and the surrounding DNA than the H3 K18Ac target of TRIM24, potentially making it more accessible to attack by KDM4A. The KDM4A construct used in this study only included the JmjN and JmjC domains, therefore, full length KDM4A may have additional binding regions that would improve its efficiency demethylating the more biologically relevant constructs.

No demethylation of H3 sK79Me$_3$ by the 2OG-oxygenases KDM4A, KDM4E, JMJD4-7 or PHF2 was detected, however these represents only a subset of human 2OG-dependent KDMs. Using the developed dot blot assay other KDMs could be easily tested for demethylation activity against H3 sK79Me$_3$, or any other methylation mark (with the right antibody system) in a nucleosome context. Additionally, only 2OG-dependent KDMs were tested; H3 K79Me$_3$ may be demethylated by another family of proteins such as the FAD-dependent KDMs, or a presently unidentified family.

While chemically-produced thioether lysine methylation marks are similar to the natural lysine methylation marks, some proteins could potentially be too selective to recognise H3 sK79Me$_3$. Other methods for implementing PTMs, such as through codon reassignment, could be used to produce even more physiologically relevant mononucleosomes. Furthermore, KDMs may require combinations of marks to act on H3 K79Me$_3$, such as in the case of DOT1L, which requires ubiquitination at H2B K123 for di- and tri-methylation to occur.[1] While the chemical modification of cysteine only allows for one type of PTM to be added per protein by this method, other modifications could be added to other histones. Additionally, different methods for implementing PTMs could be used on the same protein, such as by incorporating thioether-enabled modifications on histones pre-modified through codon reassignment. The assays developed in this chapter provide a platform on which combinatorial effects of different PTMs can be teased apart at the nucleosome level.

## 5.7 Part I conclusions

Transcriptional regulation involves the complex interplay of a vast range of regulatory processes. Deconvolution of these pathways will be enabled by a wide variety of probes and methods for investigating the action of these proteins and markers, both in isolation and in combination. One such method for simplifying an interaction under study is to use fragments of proteins instead of the harder to produce full length proteins or complexes.

Fragment-based AlphaScreen studies were used to assist in the design of a dual inhibitor capable of inducing a hypoxic-like response, while simultaneously inhibiting BRD4 function. The effect of linker length and constituent inhibitor identity on BRD4 binding was investigated, resulting in the design of a potent BRD–pVHL inhibitor. The inhibitor was found to be active in cells, inducing an expression profile that differed from what was expected based on studies conducted using BRD4 inhibitors and pVHL inhibitors alone.

Inclusion of a newly discovered putative histone post translational modification (KAcOH) into $H4_{1\text{-}20}$ peptides allowed for the probing of the affinity of BRD4 to KAcOH marks. Despite the tight binding of BRD4 to KAc marks, KAcOH was found by AlphaScreen to be poorly tolerated by BRD4, indicating a potential role for KAcOH as a 'PTM silencer'.

Peptide studies predicted a strong interaction between TRIM24 and an H3 K18Ac nucleosome, suggesting the pair to be promising candidates for the first co-crystal complex structure of a nucleosome bound to a transcription factor *via* a histone tail. Crystal structures were obtained for an H3 sK18Ac nucleosome as well as TRIM24, however co-crystal trials were unsuccessful. ITC experiments exposed significant

discrepancies between the binding strength of TRIM24 to peptides (0.026 μM), histones (1.56 μM), and nucleosomes (52.9 μM), highlighting the limitations of fragment-based studies. H3 sK18Ac nucleosome–TRIM24 complex formation was detected by non-denaturing MS and SAXS, revealing structural insights into the complex formed.

The difference in demethylase activity of KDM4A to H3 sK9Me$_3$ peptides, histones and nucleosomes were quantified using an antibody-based assay. A ~16-fold increase in demethylation activity was observed for KDM4A acting on H3$_{1\text{-}15}$ K9Me$_3$ peptide *vs.* H3 sK9Me$_3$ nucleosome, much lower than the 2000-fold tighter binding seen between TRIM24 binding to H3$_{1\text{-}27}$ K18Ac peptide *vs.* H3 sK18Ac nucleosome. This suggests that the biological relevance of results obtained from fragment-based studies is dependent on the interaction under investigation. The combination of synthetically produced methylated lysine analogues, and a sensitive antibody-based assay capable of tracking nucleosome methylation states provides an attractive platform for probing demethylase activity in a nucleosome context. Initial attempts have been made to apply this assay in the search for an H3 K79 demethylase.

These studies emphasise some of the uses and limitations of using simplified models such as protein fragments to investigate transcriptional regulation. Movement towards systems more representative of the 'natural' conditions within cells are a necessary step, and these studies aid towards making those studies more feasible.

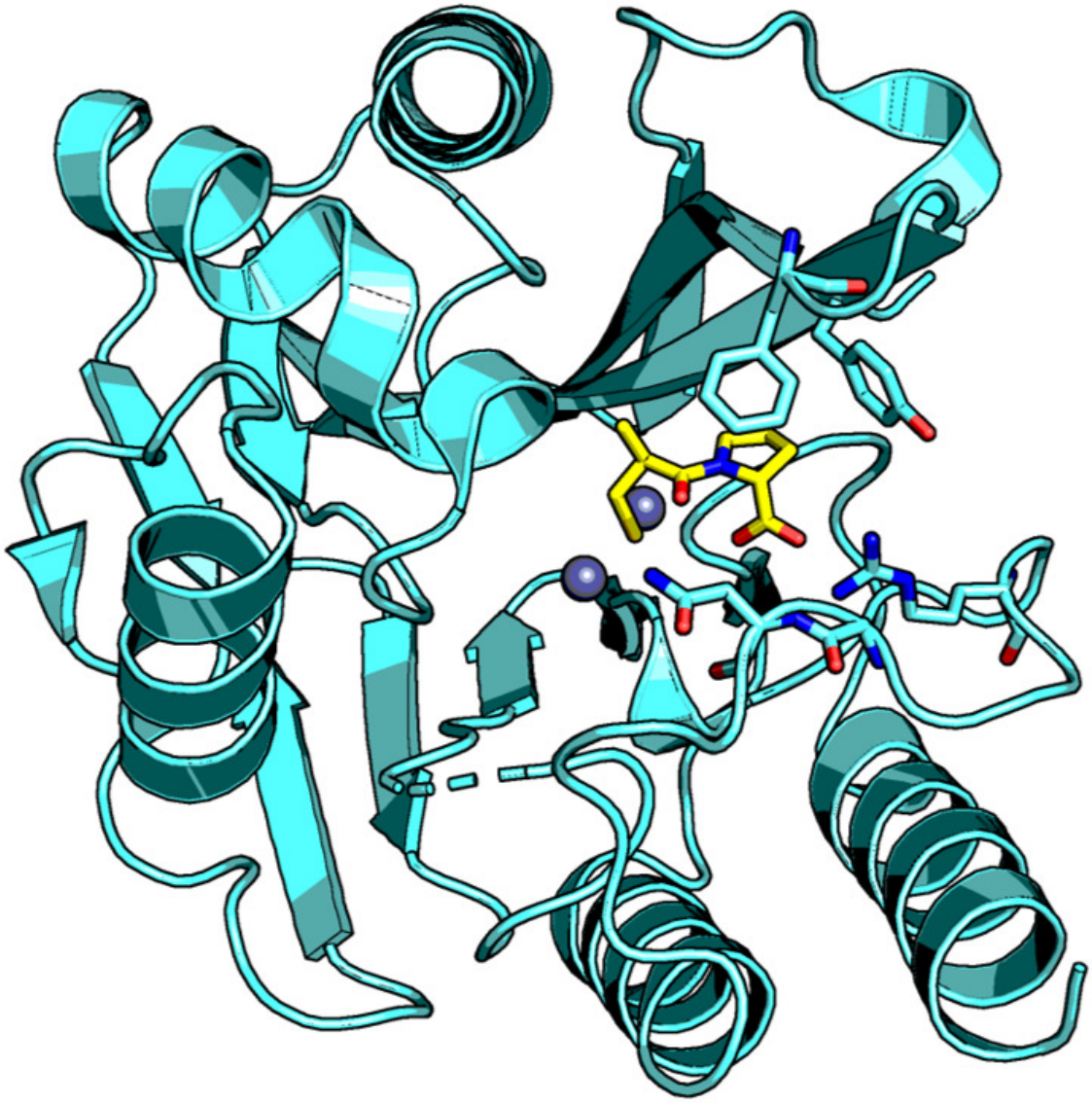
## 5.8 References


1. Van Leeuwen, F., Gafken, P. R. & Gottschling, D. E. Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* **109,** 745–756 (2002).
2. Feng, Q. *et al.* Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **12,** 1052–1058 (2002).
3. Hewitson, K. S. *et al.* Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. *J. Biol. Chem.* **277,** 26351–26355 (2002).
4. Welford, R. W. D. *et al.* Incorporation of oxygen into the succinate co-product of iron(II) and 2-oxoglutarate dependent oxygenases from bacteria, plants and humans. *FEBS Lett.* **579,** 5170–5174 (2005).
5. Kawamura, A. *et al.* Development of homogeneous luminescence assays for histone demethylase catalysis and binding. *Anal. Biochem.* **404,** 86–93 (2010).
6. Oprandy, J. J., Olson, J. G. & Scott, T. W. A rapid dot immunoassay for the detection of serum antibodies to eastern equine encephalomyelitis and St. Louis encephalitis viruses in sentinel chickens. *Am. J. Trop. Med. Hyg.* **38,** 181–6 (1988).
7. Simon, M. D. *et al.* The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128,** 1003–12 (2007).
8. Simon, M. D. Installation of site-specific methylation into histones using methyl lysine analogs. In *Current protocols in molecular biology* Unit 21.18.1-10 (John Wiley & Sons, Inc., 2010).
9. Whetstine, J. R. *et al.* Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125,** 467–481 (2006).
10. Shiau, C. *et al.* Reconstitution of nucleosome demethylation and catalytic properties of a jumonji histone demethylase. *Chem. Biol.* **20,** 494–499 (2013).
11. Mahmood, T. & Yang, P. C. Western blot: Technique, theory, and trouble shooting. *N. Am. J. Med. Sci.* **4,** 429–434 (2012).
12. Gallagher, S., Winston, S. E., Fuller, S. A. & Hurrell, J. G. R. Immunoblotting and immunodetection. In *Current protocols in molecular biology* Unit 10.8.1-10.8.28 (John Wiley & Sons, Inc., 2008).
13. Chau, Y. P. & Lu, K. S. Investigation of the blood-ganglion barrier properties in rat sympathetic ganglia by using lanthanum ion and horseradish peroxidase as tracers. *Cells Tissues Organs* **153,** 135–144 (1995).
14. Blake, M. S., Johnston, K. H., Russell-Jones, G. J. & Gotschlich, E. C. A rapid, sensitive method for detection of alkaline phosphatase-conjugated anti-antibody on western blots. *Anal. Biochem.* **136,** 175–179 (1984).
15. Mishra, M., Tiwari, S. & Gomes, A. V. Protein purification and analysis: Next generation western blotting techniques. *Expert Rev. Proteomics* **14,** 1037–1053

(2017).

16. GraphPad Software. GraphPad Prism version 5.04 for Windows (2010).

# Part II: Structural Studies on Metallo-β-Lactamase Inhibition



*View from a novel crystal structure of VIM-2 complexed with a D-captopril (PDB ID: 4C1E) described in* ***Chapter 6****.*

# Chapter 6: Structural Basis of Metallo-β-lactamase Inhibition by Captopril Stereoisomers

## 6.1 Introduction

### 6.1.1 Antibiotics

Since Alexander Fleming's pioneering observations in 1928, penicillin antibiotics have come to represent one of mankind's greatest weapons against pathogenic microbial attack.[1] During the 1940s an antibiotic drug discovery platform was developed in which organisms were screened for antimicrobial activity by detecting zones of inhibited microbial growth on bacterial growth plates.[2] Over the following two decades this highly successful method resulted in the discovery of the major classes of antibiotics currently in use.[3] As antibiotics entered into widespread clinical use, increasing instances of target microorganisms developing resistance to antibiotics began to emerge. In order to counter these antibiotic resistant bacteria, modified analogues of existing antibiotics were developed, revitalising their potency. Resistance to these new antibiotics inevitably arose, and while the rate of drug discovery initially appeared to outpace the rate of development of antibiotic resistance, by the 1970s the drug discovery pipeline apparently began to dry up. Since the discovery of the streptogramins in 1963 only one clinically useful antibiotic from an entirely new class has been discovered; the lipopeptide daptomycin reported in 1986.[4,5] Teixobactin, a new lipid II-binding compound with antibiotic properties, was discovered in 2015, however this potential drug has not yet entered phase I drug trials.[6] The proliferation of multidrug resistant bacteria, coupled with the lack of development of new antibiotics has resulted in the warning by the chief medical officer for England of an impending "antibiotic apocalypse".[7]

Antibiotics that target proteins specific to bacteria and are termed antibacterials. The majority of antibacterials disrupt one of four major pathways or targets, as listed below.

- The bacterial cell membrane can be compromised either through inhibition of proteins associated with cell wall synthesis (β-lactams and glycopeptides) or by direct destabilisation of the cell membrane (lipopeptides).[6,8–10]
- DNA replication can be disturbed by inhibition of DNA topoisomerases (quinolones) or through inhibition of RNA synthesis (ansamycins).[11,12]
- Binding to the 30S (tetracyclines) or 50S (chloramphenicols, macrolides, streptogramins and oxazolidinones) ribosomal subunits can result in perturbation of protein synthesis.[4,13–16]
- Inhibition of bacterial metabolic enzymes such as dihydropteroate synthetase, an enzyme involved in folate synthesis can also limit microbial growth (sulfonamides). Sulfonamides are considered bacteriostatic as they restrict the availability of folic acid to the bacteria, preventing bacterial proliferation, but do not kill the organism directly.[17]

**β-Lactam antibiotics**

Due to their high efficacy, low toxicity and affordability the β-lactams are one of the most frequently prescribed classes of antibiotics.[18,19] Over the latter half of the 20$^{th}$ century an arsenal of β-lactam variants were developed, including semi-synthetic analogues of penicillin, carbapenems, cephalosporins and monobactams, amongst others, which enabled improved drug activity, stability and absorption properties.[20] A conserved motif consistent within these compounds is the four-membered β-lactam ring (see **Figure 6.1**), crucial to activity.

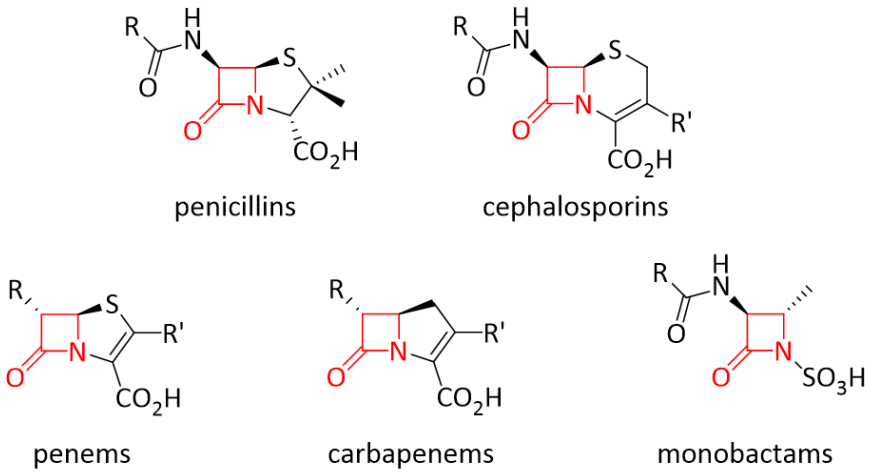



***Figure 6.1**: General structures of β-lactam antibiotics. The β-lactam rings are displayed in red. The 'R' groups vary between members of the β-lactam families.*

Throughout the life cycle of bacteria, the cell wall is in a permanent state of flux with a constant breaking down and building up of its peptidoglycan layer. The peptidoglycan layer determines cell shape, protects against osmotic rupture, and is vital to cell division and growth.[21] The β-lactams manifest their bactericidal nature by inhibiting the transpeptidase penicillin binding proteins (PBP) that are responsible for the formation of new peptidoglycan cross-links.[8,22] The structural similarity of β-lactam antibiotics to the D-alanine-D-alanine terminus (possibly in a conformationally restrained form) of the forming peptidoglycan strand is reported to result in efficient binding to the PBP active site. The hydroxide of a PBP active site serine subsequently attacks the β-lactam carbonyl, opening the torsionally constrained four-membered ring and forming a stable acyl-enzyme complex (see **Figure 6.2**).[23] This covalent inhibition of the PBPs prevents the formation of new peptidoglycan cross-links, compromising the integrity of the cell wall and preventing further bacterial cell division.

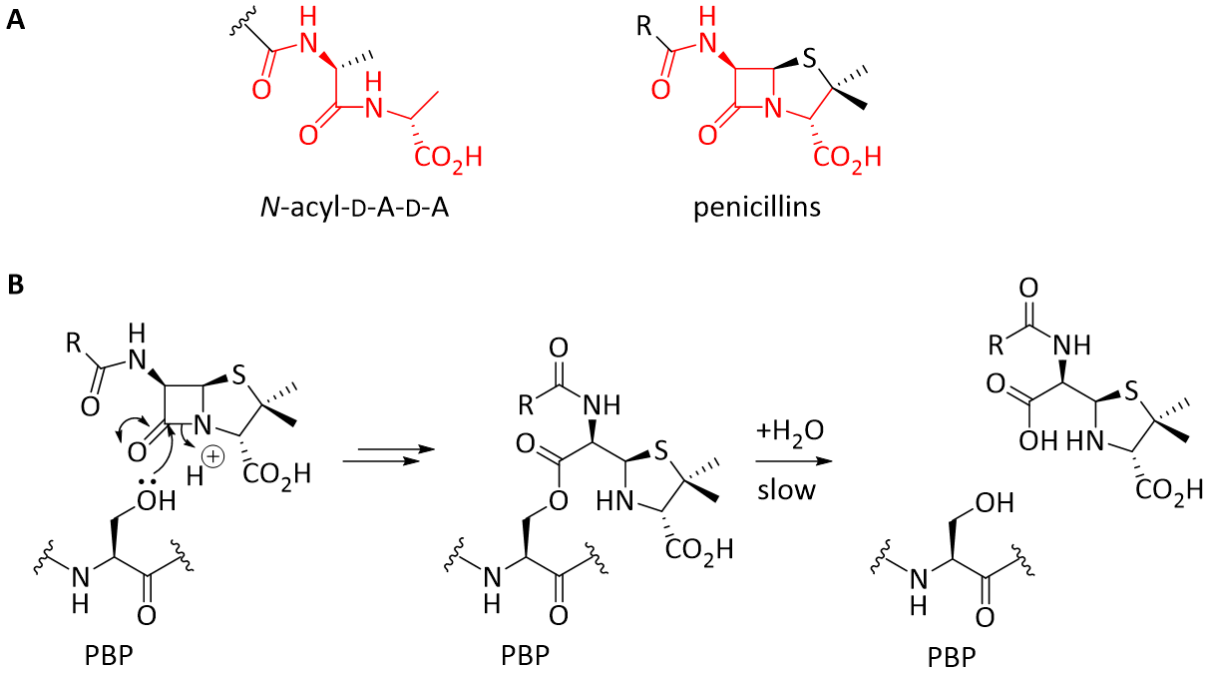



***Figure 6.2**: **A**- Structures of the D-alanine-D-alanine substrate of PBPs (**left**) and an example class of β-lactam antibiotic (penicillins) (**right**). The identity of the 'R' group varies between penicillins. Regions of structural similarity are displayed in red. **B**- Reaction scheme for the covalent inhibition of a PBP active site serine residue by penicillin. The driving force for the reaction is provided by the opening of the highly constrained β-lactam ring. The stable acyl-enzyme complex formed deactivates the PBP and leaves it unable to function until the antibiotic is removed* via *a slow hydrolysis step.*

### 6.1.2 Antibiotic resistance

The occurrence of antibiotic resistance is not a new phenomenon. The development of toxins and resistances are a commonly repeated pattern in nature where organisms vie for the same ecological niche. Samples of bacteria collected after lying below the Yukon permafrost for thirty thousand years were shown to contain a diverse range of genes coding for β-lactam antibiotics.[24] Similarly, a subterranean cave, separated from the outside world for over four million years, was shown to contain bacteria resistant to 14 commercially available antibiotics including daptomycin, an antibiotic of last resort in the treatment of gram-positive, drug resistant pathogens.[25] While resistance may be an ancient development the proportion of bacterial infections caused by antibiotic resistant cells has been steadily increasing since the drugs entered common usage.[26,27]

Specific mechanisms by which bacteria can become resistant to the β-lactams include the following:

- Alteration to the active site of the PBP can reduce the affinity between the protein and antibiotic resulting in an increase of resistance to the drug.[28]
- In order for β-lactams to reach the target proteins on the inner plasma membrane they first need to cross the membrane either by diffusion or through outer membrane proteins (OMPs). Some *Enterobacteriaceae* are able to reduce their susceptibility to certain antibiotics by decreasing expression of the specific OMP able to facilitate that drug's passage into the cell.[29,30]
- Conversely, efflux pumps have the ability to export a range of substrates out of the cell, including β-lactams.[31] The combination of down-regulated OMPs with efflux pumps represent an important mechanism used by many multidrug resistant gram-negative pathogens.[32]
- The most common mode of resistance is the production of β-lactamases which catalyses the hydrolysis of the β-lactam ring, rendering the molecule inactive.[33] It is the study of these enzymes that this investigation will centre on.

### 6.1.3 β-Lactamases

Through analysis of their sequence similarity the β-lactamases can be divided into four classes (A-D). Three of these classes (A, C and D) are thought to share a common ancestor with the PBPs and catalyse the degradation of β-lactams through nucleophilic attack of an activated serine residue to the carbonyl of the β-lactam, opening the ring.[34] Unlike PBPs these serine-β-lactamases (SBLs) are also able to efficiently catalyse the subsequent hydrolysis of the inhibitor from the active site, rapidly regenerating the enzyme whilst leaving the inhibitor deactivated (see **Figure 6.3**).[35]

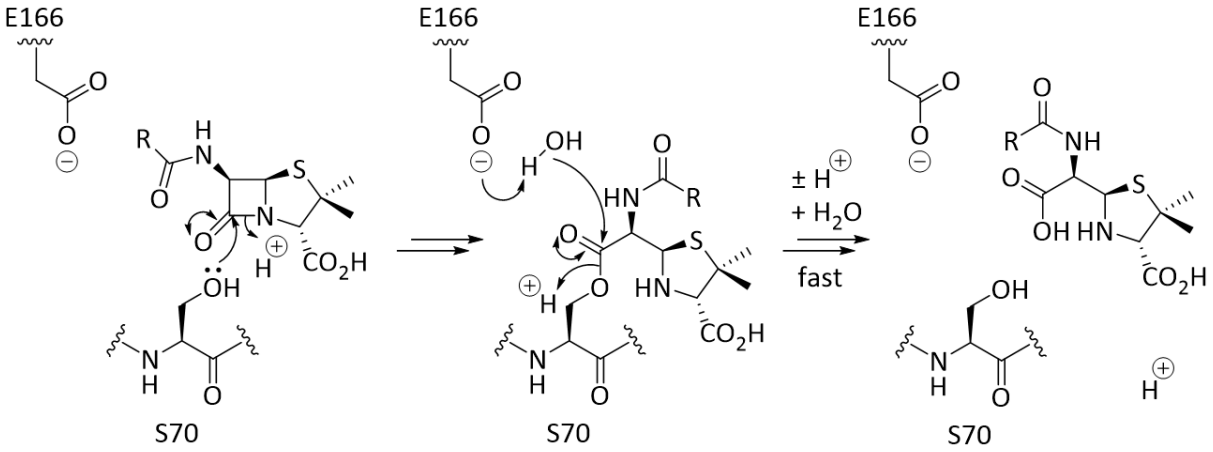



***Figure 6.3**: Scheme showing the hydrolysis of penicillin as catalysed by an SBL enzyme. The protons are provided by nearby activated serene and glutamic acid residues.*

Class B β-lactamases, or metallo-β-lactamases (MBLs), use either one or two Zn(II) ions to activate a water molecule in its active site (see **Figure 6.4**). The MBLs can be further divided by structural and sequence alignment into three subclasses (B1-B3).[36] In their active forms the B1 MBLs contain two Zn(II) ions and an activated bridging water or hydroxide ion in their active site. The first zinc is coordinated in a trigonal bipyramidal coordination environment by cysteine, aspartic acid and histidine residues and two waters, one of which bridges to the second zinc. The second zinc has a tetrahedral coordination sphere made up of three histidine residues and the bridging water.[33] B2 MBLs bind only one Zn(II) ion in their active sites in their catalytically active form, which is tetrahedrally coordinated by aspartic acid, cysteine and histidine residues as well as one water.[33] B3 MBLs generally contain an active site with two Zn(II) ions coordinated similarly to B1 MBLs but with a histidine replacing the cysteine in the coordination sphere of the second zinc.[33] The B1 subclass contains the most clinically relevant MBLs including New Delhi MBL (NDM), Verona integron-encoding MBL (VIM) and imipenemase (IMP) type MBLs, capable of catalysing the hydrolysis of almost all known β-lactam antibiotics, and is the focus of the work described in this chapter.[37]

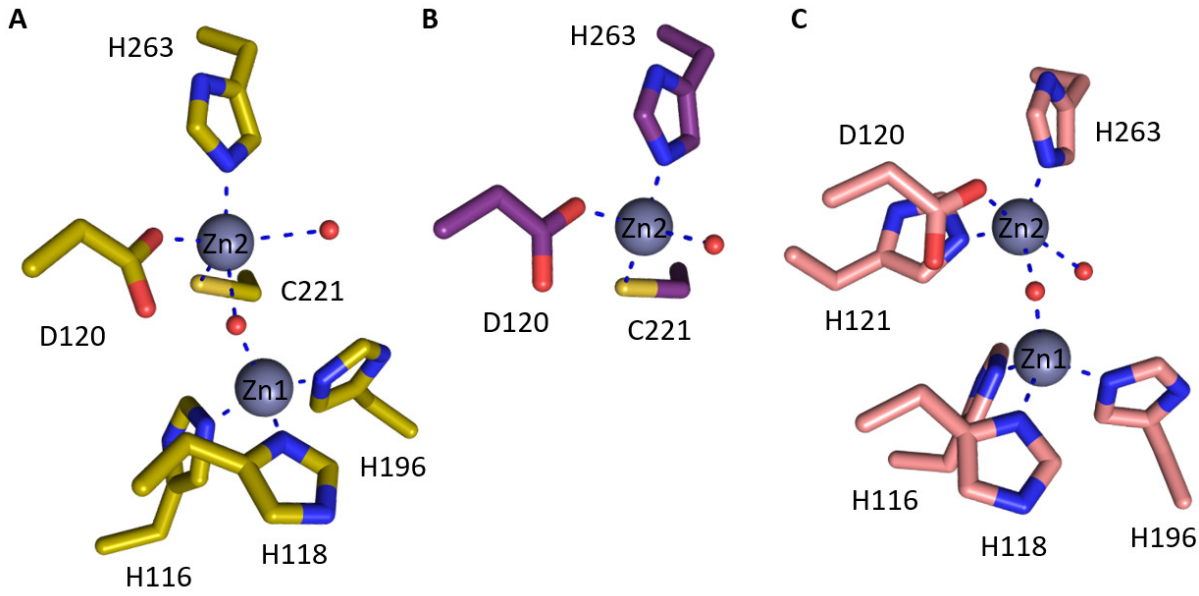



***Figure 6.4**: Views from crystal structures of a B1 (NDM-1), B2 (*Serratia fonticola*-1; Sfh-1), and B3 (L1) type MBL (PDB IDs: 3SPU, 3SD9, and 1SML, respectively) showing zinc coordination environments.*[38–40]

A number of potential mechanistic roles for the MBL Zn(II) ions have been suggested which may vary depending on the substrate, enzyme and experimental conditions such as pH.[41] B1 and B3 MBLs are proposed to coordinate to Zn2 through displacement of the non-bridging water molecule by a penicillin carboxylic acid oxygen atom.[42] Nucleophilic attack of the bridging water/hydroxide on the β-lactam carbonyl ultimately results in the breaking of the penicillin amide bond, however the precise mechanism by which this occurs is contentious.[42] One possibility is that the rate-limiting step of amide bond cleavage occurs concomitantly with protonation of the amide nitrogen. A second suggested possibility is that the amide bond breaks first, resulting in the formation of an anionic nitrogen intermediate, followed by the protonation of the nitrogen as the rate-limiting step.[43,44] Both mechanisms have been supported by experimental results. The rate-limiting step may be influenced by the ability of the enzyme and substrate to stabilise the anionic intermediate.[44–47] The general consensus is that a 'new', non-

bridging water coordinated to Zn2 is likely to be the proton donor for the amine nitrogen.[48,49] The hydrolysed product is subsequently released and the active enzyme is regenerated. The general reaction scheme for the hydrolysis of a β-lactam antibiotic as catalysed by a di-Zn(II) MBL is provided in **Figure 6.5**.

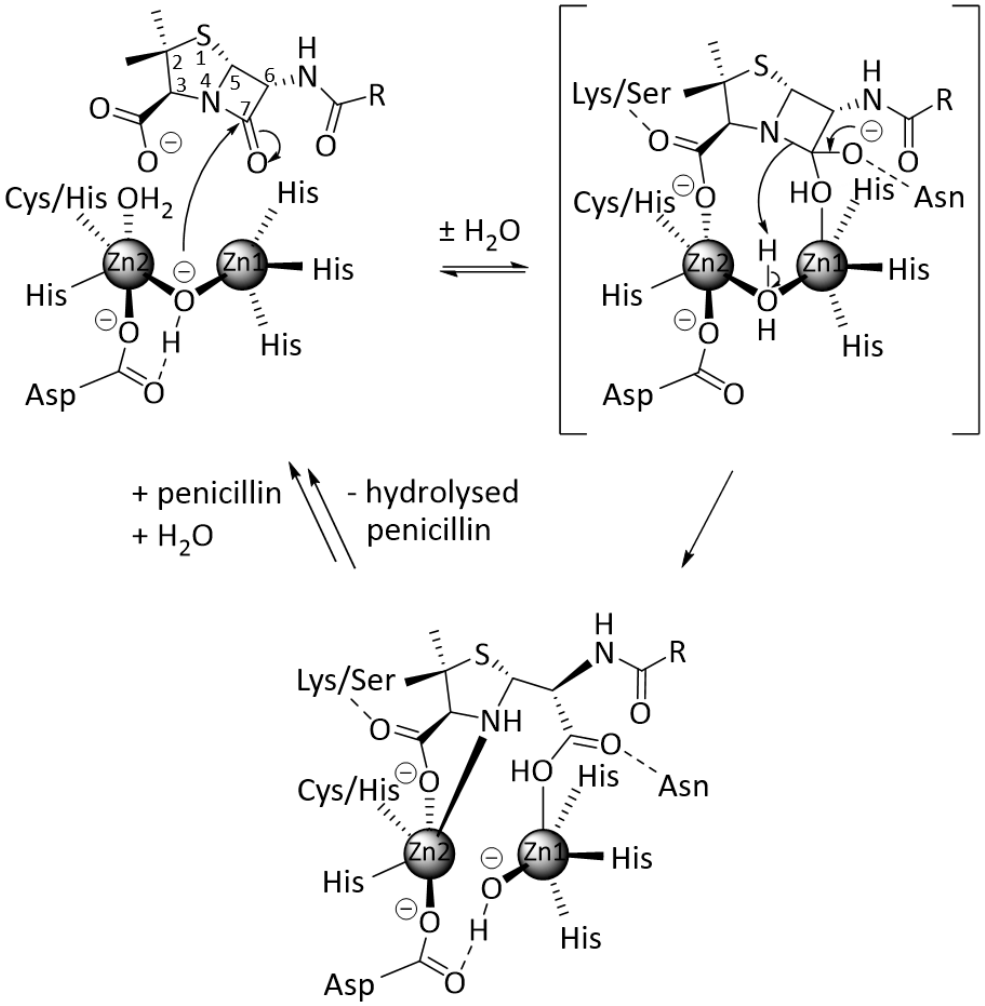



***Figure 6.5**: General mechanisms for the hydrolysis of penicillin by B1/B3 MBLs. The coordination geometry of the MBL–hydrolysed penicillin complex (**bottom**) is based on a crystal structure of NDM-1 complexed with hydrolysed penicillin (PDB ID: 4EYF).[50] The precise mechanism of hydrolysis and Zn(II) ion-substrate/product coordination environment depends on the MBL and β-lactam compound in question.[42]*

Co-crystal structures have been reported for the B1 MBL NDM-1 in complex with a range of hydrolysed β-lactam antibiotics, including penicillins, cephalosporins and

carbapenems.[42] These co-crystal structures indicate that an oxygen from the carboxylic acid formed during the β-lactam hydrolysis is positioned to coordinate to the MBL Zn1.

In the majority of reported NDM-1-hydrolysed β-lactam co-crystal structures, Zn2 appears able to coordinate the $N^4$ nitrogen previously located in the β-lactam ring as well as an oxygen in the carboxylic acid group attached to the $C^3$-position (see **Figure 6.5**, **bottom**). In each of the NDM-1-hydrolysed β-lactam co-crystal structures reported, the hydrolysed product is positioned to form hydrogen bonds through the $C^3$ carboxylic acid oxygen not involved in Zn2 binding to a nearby K211 $N^{\varepsilon}$. Similarly, each hydrolysed compound is positioned to hydrogen bond to the N233 side-chain nitrogen through a $C^7$ carboxylic acid oxygen.

A crystal structure of the B3 MBL L1 complexed with the hydrolysed cephalosporin moxolactam (PDB ID: 2AIO) indicates a similar Zn(II) ion coordination environment to the NDM-1-hydrolysed penicillin/cephalosporin crystal structures with the exception of the substitution of C208 for H89, characteristic of B3 MBLs.[51] Equivalently located lysine and asparagine residues to the NDM-1 K211 and N233 residues are absent from L1. A $C^3$ carboxylic acid oxygen of hydrolysed moxolactam is, however, positioned to hydrogen bond to the sidechain hydroxyl group of serine residues S221 and S223 in L1 (see **Figure 6.5**, **bottom**).

There are two main approaches that have been employed to combat the rise of the β-lactamases. The first is the development of new β-lactam antibiotics more resistant to β-lactam hydrolysis, normally through the incorporation of groups designed to conflict with the β-lactamase active site.[52] However, as previously mentioned, the slow rate of new antibiotic discovery, combined with the proliferation of strains of bacteria capable

of resisting even drugs of last resort indicate this approach cannot succeed alone.[53] The second approach to neutralising the threat posed by β-lactamases is through the discovery of β-lactamase inhibitors.

### 6.1.4 β-Lactamase inhibitors

As the spread of MBL-containing pathogens continues at an alarming pace, rendering some or all antibiotics impotent, it is hoped that the development of inhibitors of these enzymes could allow for an increase in the lifespan of β-lactams.[54] Co-dosing a patient with both a β-lactam antibiotic and a relevant β-lactamase inhibitor has been found to be a clinically successful method for regaining antibiotic potency against certain multi-drug resistant infections.[55]

To date there are four β-lactamase inhibitors that are clinically available, all of which have been found to be effective at inhibiting certain SBLs (see **Figure 6.6 A**). Sulbactam and tazobactam are effective against class A and C β-lactamases, whereas clavulanic acid only inhibits class A β-lactamases.[55] Sulbactam, tazobactam and clavulanic acid contain β-lactam rings and have been found to inhibit SBLs by a mechanism related to the β-lactam inhibition of PBPs.[55–57] As with other β-lactams, these three β-lactamase inhibitors form an acyl-enzyme intermediate with an activated SBL serine. Unlike in the case of β-lactam–PBP interactions, the SBL inhibitor-derived acyl-enzyme intermediates are relatively stable to hydrolysis.[55] The enzyme-bound inhibitors are then able to undergo rearrangement followed by either cross-linking to a second active site serine, irreversibly inactivating the enzyme, or slow hydrolysis of the deactivated inhibitor (see **Figure 6.6 B**).[58]

A fourth β-lactamase inhibitor, avibactam, is capable of inhibiting class A, C and some class D SBLs and lacks a β-lactam ring, containing instead an imidazolidinone ring.[59] Like the β-lactam-based β-lactamase inhibitors, avibactam is able to form a stable acyl-enzyme complex through attack of an activated SBL serine, in this case to the avibactam imidazolidinone carbonyl, opening the 5-membered ring.[60] The most distinct difference between the mode of inhibition of avibactam and the class A β-lactamase inhibitors is the fact that avibactam can be re-cyclised on cleavage with the SBL.[59] The positioning of the avibactam sulfate group in the SBL active site is thought to displace the activated water molecule normally involved in inhibitor hydrolysis, favouring intramolecular recyclization of avibactam over hydrolysis.[60]

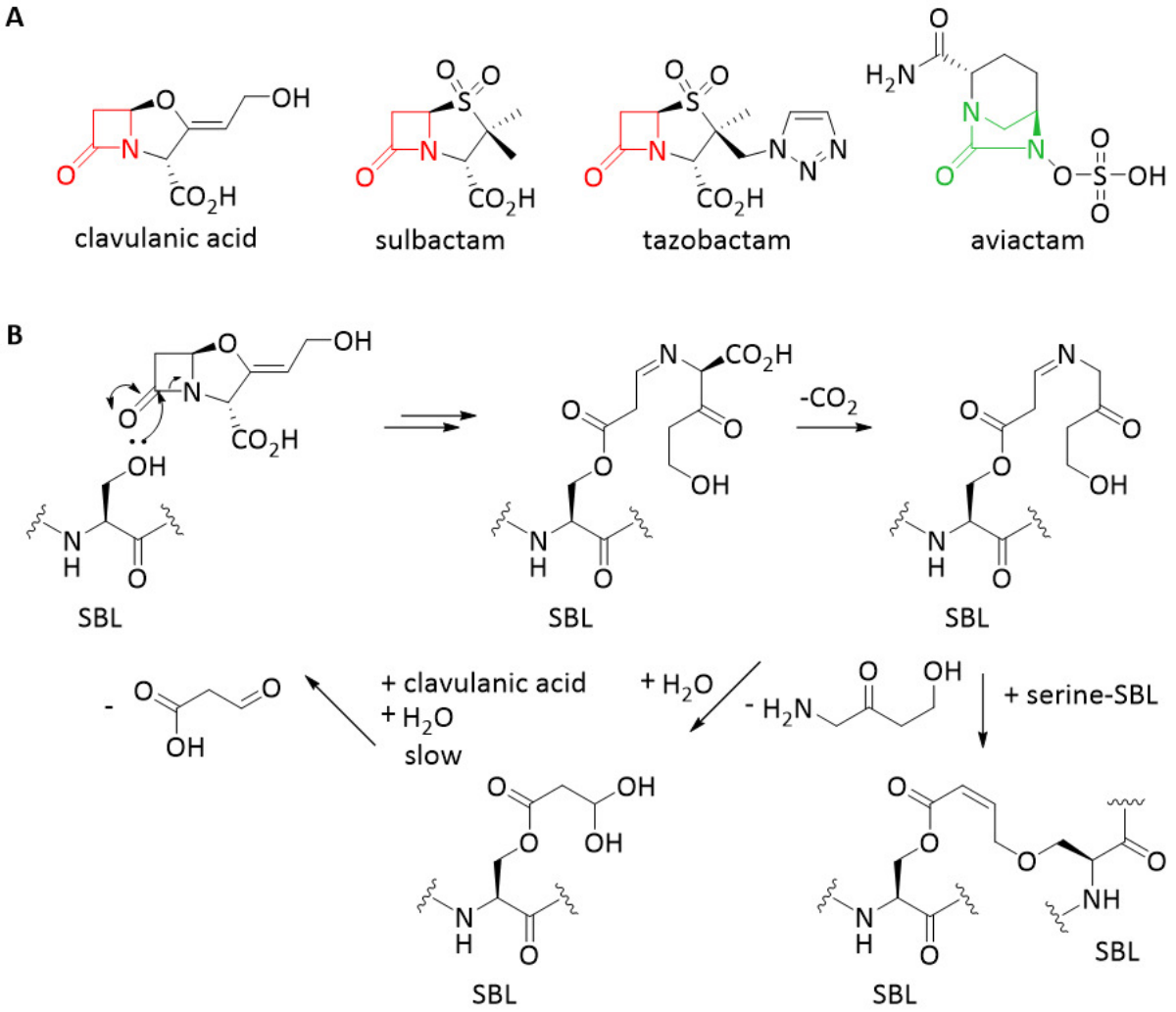



***Figure 6.6**: **A**- Structures of β-lactamase inhibitors currently in clinical use. β-Lactam rings are shown in red and the avibactam imidazolidinone ring is shown in green. **B**- Scheme typical of class A β-lactamase inhibition showing key intermediates formed in the reaction of an SBL with clavulanic acid.[58] The reaction mechanism is more complex than presented and variations in the mechanisms of inhibition also occur between the inhibitors.*

A selection of inhibitors targeting class A and class C SBLs are currently clinically available with others in pipeline. Due to differences in their mode of action, MBLs are not susceptible to the available β-lactamase inhibitors. As a result there are no drugs in clinical use designed to inhibit any MBL, highlighting it as a vital target for drug design.[33]

### 6.1.5 Captopril

Several classes of known metalloenzyme inhibitors inhibit MBLs, including thiols, carboxylic acids, trifluoromethyl ketones, hydroxamic acids, and rhodanines.[57,61,62] For an MBL inhibitor to be clinically useful it should ideally achieve a wide breadth of inhibition across a range of MBL subtypes without interfering with the action of structurally related human MBL-fold enzymes. In order for the inhibitor to be a viable drug candidate the molecule should also be "druglike"; *i.e.* have a low molecular weight, high potency, high stability, and an appropriate solubility in water and fat. (2*S*)-1-[(2*S*)-2-Methyl-3-sulfanylpropanoyl] pyrrolidine-2-carboxylic acid (captopril) is an angiotensin-converting enzyme (ACE) inhibitor that has been successfully used in clinics to control high blood pressure since the early 1980s.[63] The clinically used (2*S*,2*S*)-stereoisomer (L-captopril) has been found to inhibit a range of MBLs from each of the subclasses.[64–67] Additionally, an L-captopril stereoisomer (2*S*)-1-[(2*R*)-2-methyl-3-sulfanylpropanoyl] pyrrolidine-2-carboxylic acid (D-captopril) has been reported to be more active than L-captopril against some MBLs such as the clinically relevant NDM-1 and the model MBL *Bacillus cereus* II (BcII).[66,68]

Captopril has structural similarities to the hydrolysed product of MBL-catalysed degradation of penicillin (see **Figure 6.7 A**). Both compounds contain a carboxylic acid-substituted five-membered ring separated from a thiol/second carboxylic acid group by five atoms. The structural similarity of captopril to hydrolysed penicillin, combined with its potent inhibition, suggest that the two molecules may interact with the β-lactamases in a similar way.

Crystal structures of L- or D-captopril in complex with MBLs have been reported for each of the MBL subclasses. B1 MBLs NDM-1 and BlaB from *Chryseobacterium*

*meningosepticum* have been crystallised with L- and D-captopril, respectively.[50,69] The B2 MBL CphA from *Aeromonas hydrophila* has been crystallised with D-captopril.[67] B3 MBLs from *Fluoribacter gormanii* (FEZ-1) and *Stenotrophomonas maltophilia* (L1) have both been complexed with D-captopril.[64,70] The majority of these MBL–captopril structures support the hypothesis that captopril binds to the MBLs through both active site Zn(II) ions, displacing the bridging water.[50,67,69]

Captopril contains two stereogenic centres at the $C^2$ and $C^4$ positions in the molecule. L- and D-captopril are defined with respect to the $C^4$ stereogenic centre, and have both been found to have relatively high, and differing potencies against multiple MBLs. Two further stereoisomers of L- and D-captopril are possible, containing reversed stereochemistry at the $C^2$ chiral centre; (2*R*)-1-[(2*S*)-2-methyl-3-sulfanylpropanoyl] pyrrolidine-2-carboxylic acid and (2*R*)-1-[(2*R*)-2-methyl-3-sulfanylpropanoyl] pyrrolidine-2-carboxylic acid (*epi*-L-captopril and *epi*-D-captopril, respectively) (see **Figure 6.7 B**).

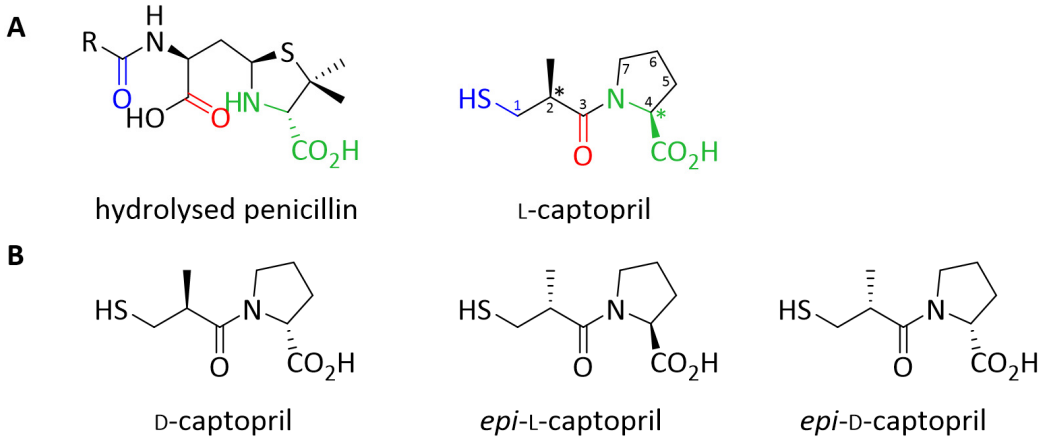



***Figure 6.7**: **Top**- Structures of hydrolysed penicillin and the structurally related ACE inhibitor L-captopril. Corresponding colours indicate regions of structural similarity between the two compounds. **Bottom**- Structures of the three stereoisomers of L-captopril.*

## 6.2 Aims of Part II

The aim of the work described in **Part II** of this thesis (**Chapters 6** and **7**) was to provide structural insights into the binding modes of several potent MBL inhibitors when complexed with a range of clinically relevant, or model MBL targets. Aims of the specific chapters were to:

- Solve the crystal structures of captopril stereoisomers in complex with BcII, IMP-1 and VIM-2, thereby enabling rationalisation of the captopril stereoisomer–MBL structure-activity relationship (**Chapter 6**).
- Investigate the binding mode of a potent, novel, *de novo*-designed MBL inhibitor to VIM-2 through crystallographic studies (**Chapter 7**).

## 6.3 Objectives

The rising wave of antibiotic resistance threatens the utility of humanity's most important weapon against microbial attack. One key mechanism of antibiotic resistance is through the hydrolysis of β-lactam antibiotics by MBLs. By inhibiting the action of the MBLs it may be possible to revitalise the antibiotic potency of the β-lactam class of antibiotic. To date no such inhibitor is clinically available.

The work described in this chapter aimed to investigate the inhibitory action of the four stereoisomers of the FDA approved and widely prescribed ACE inhibitor captopril against four clinically relevant MBLs (IMP-1, NDM-1, SPM-1 and VIM-2) and the model MBL BcII. The inhibitory effects of these inhibitors on four structurally related human MBL-fold metalloenzymes; DNA cross-link repair enzymes 1A and 1B (DCLRE1A and DCLRE1B), human glyoxalase II (hHAGH), and human angiotensin-converting enzyme-2 (hACE-2) are also tested.

The bulk of the work in this chapter concerns the crystallisation and structural analysis of BcII, IMP-1 and VIM-2 in the presence and absence of captopril inhibitors with a goal of rationalising the captopril stereoisomer-MBL structure-activity relationship.

*The work in this chapter has been published (Brem, van Berkel, Zollman* et al. *2016).* L-, D-, epi-L-, *and* epi-D-*captopril were synthesised by Dr Sander van Berkel. All proteins used in this chapter were produced and purified by Dr Jürgen Brem. Kinetic and nucleotide assays were also conducted by Dr Jürgen Brem.*

## 6.4 Target selection

### 6.4.1 Captopril

No reports on MBL inhibition by either *epi*-L-captopril or *epi*-D-captopril had been published prior to the commencement of this project. Additionally, no more than one MBL–captopril stereoisomer crystal structure was available for any MBL, preventing direct comparisons of binding modes between the stereoisomers. Systematic studies on the inhibition of a selection of MBLs by the four captopril stereoisomers were therefore conducted. L-Captopril, D-captopril, *epi*-L-captopril and *epi*-D-captopril were synthesised (by *Dr Sander van Berkel*) using reported procedures.[66,71,72]

### 6.4.2 Metallo-β-lactamases

Four clinically relevant MBLs (IMP-1, NDM-1, SPM-1 and VIM-2) and the model MBL BcII were chosen for this study. While reports of drug-resistant infections containing each of the selected clinically relevant MBLs continue to rise, the plasmidic origin of IMP-1, NDM-1 and VIM-2 further increases their risk to human health.[54,73–75] MBLs encoded on plasmids are capable of being readily passed between organisms, conveying β-lactam resistance by horizontal gene transfer. The selected MBLs are all members of the B1 subclass and displayed low sequence homology (26-38%). Recombinant forms of the five MBLs used in these experiments were produced in *E. coli* and purified (by *Dr Jürgen Brem*) following previously described procedures.[76,77]

## 6.5 Metallo-β-lactamase inhibition by captopril stereoisomers

### 6.5.1 $K_i$ determination using a fluorescent reporter assay

Inhibitor constant ($K_i$) values for the captopril stereoisomers against the selected MBLs were determined using an absorbance-based assay (carried out by *Dr Jürgen Brem*).[61,76] Each MBL was incubated in the presence of an inhibitor and the β-lactam-containing compound nitrocefin.[78] Hydrolysis of the nitrocefin β-lactam ring results in a colour change from yellow (λ ≈ 380 nm) to red (λ ≈ 500 nm).[79] Monitoring the colour change of reactions containing different concentration of inhibitor allows for the inference of the potency of the MBL inhibitor. $K_i$ data are presented in **Table 6.1**.

***Table 6.1**: $K_i$ (μM) values for the four captopril stereoisomers against different MBLs. All experiments were performed a minimum of three times. $K_i$ values and their 95% confidence intervals were calculated as described in **Materials and Methods 8.14**. The mean value of each trial is presented ± standard deviation. † Literature $K_i$ value.*

| | L-captopril | D-captopril | *epi*-L-captopril | *epi*-D-captopril |
|---|---|---|---|---|
| **BCII** | 19.7 ± 0.3 (65†)[66] | 2.6 ± 0.3 (45†)[66] | 103.7 ± 0.4 | > 125 |
| **IMP-1** | 16.1 ± 0.9 (12.5†)[80] | 5.0 ± 0.8 | > 200 | 119.5 ± 0.8 |
| **NDM-1** | 41.0 ± 0.3 (52.6†)[81] | 5.2 ± 0.4 (2.1†)[81] | > 125 | 16.8 ± 0.4 |
| **SPM-1** | > 200 | 102.2 ± 0.5 | > 200 | > 200 |
| **VIM-2** | 1.0 ± 0.2 | 0.016 ± 0.002 | 1.2 ± 0.2 | 1.2 ± 0.3 |

The $K_i$ data calculated were generally close to the literature values, with any differences likely arising from the different assay conditions used. D-Captopril was the most potent inhibitor in all cases showing an increase in potency compared to the second most potent captopril stereoisomer of ~8-fold for BcII, ~3-fold for IMP-1, ~3-fold for NDM-1, ~2-fold for SPM-1 and ~60-fold for VIM-2. L-Captopril was the second most potent captopril inhibitor of the MBLs tested except for in the case of NDM-1 where *epi*-D-captopril was more potent, and with SPM-1, where only D-captopril showed significant levels of inhibition. *Epi*-L-captopril showed little inhibition against IMP-1 or NDM-1, and neither

*epi*-L-captopril nor *epi*-D-captopril showed significant inhibition against BCII or SPM-1. *Epi*-D-captopril showed some activity against IMP-1 and NDM-1, and both *epi*-L-captopril and *epi*-D-captopril exhibited relatively potent inhibition against VIM-2.

### 6.5.2 Probing human metallo-β-lactamase-fold enzyme inhibition by captopril stereoisomers

One feature conserved between the MBLs is an αβ/βα sandwich fold (discussed in more detail in **Section 6.6.2**), which has also been identified in at least 18 human proteins to date.[82] These human MBL-fold proteins are involved in a diverse range of biological processes and are divided into three subgroups based on phylogenetic analyses.[82]

• Group 1 contains seven members which are related to glyoxalase II family enzymes. Members of group 1 with presently characterised functions (*e.g.* hydroxyacylglutathione hydrolase; HAGH) play significant roles in cell detoxification.[83]

• Group 2 comprises nine members that are involved in nucleic acid modification pathways such as DNA repair (*e.g.* DNA cross-link repair enzymes 1A and 1B; DCLRE1A and DCLRE1B, respectively) and RNA processing.[84]

• Group 3 consists of two members (*N*-acyl-phosphatidylethanolamine-hydrolysing phospholipase D; NAPE-PLD, and cytidine monophospho-*N*-acetylneuraminic acid hydroxylase; CMAH) which are thought to have pleiotropic roles in metabolism.[82]

As previously mentioned, a key feature of any drug candidate for the inhibition of MBLs would be the selective inhibition of MBLs over any human enzymes, thereby minimising unwanted side-effects. Although L-captopril is a well-studied ACE inhibitor, there are no reports of its selectivity *vs.* human MBL-fold enzymes. L- and D-captopril were tested against the group 1 human-MBL-fold enzyme HAGH, and the group 2 human-MBL-fold

enzymes DCLRE1A and DCLRE1B as well as the human angiotensin-converting enzyme-2 (hACE-2) (by *Dr Jürgen Brem*). hACE-2 and hHAGH were tested using commercially available recombinant hACE-2 and recombinant human glyoxalase II protein assays, respectively (R&D Systems). The DCLRE1A and DCLRE1B nuclease assays were performed following a previously described method (by *Dr Jürgen Brem*), using a 21-residue DNA oligonucleotide with a fluorescein label at its 3′ end as the substrate.[85] No inhibition of any of the human MBL-fold enzymes tested was observed by either L- or D-captopril at 100 μM.

## 6.6 Structural analysis of captopril binding to metallo-β-lactamases

### 6.6.1 Data collection

In order to investigate the mode of binding of the different captopril stereoisomers to MBLs, co-crystallisation trials of BcII, IMP-1 and VIM-2 with L- and D-captopril were set up. In the case of VIM-2 co-crystallisation trials with both *epi*-L- and *epi*-D-captopril were also conducted. Crystallisation trials of each of the three MBLs in the absence of inhibitor were set up in the same conditions to allow direct comparison of the MBLs in the inhibitor bound, and unbound state. Preliminary structural analysis indicated that partial oxidation of the metal-binding cysteine (C221) had occurred in both the crystallized BcII and VIM-2 proteins, in a manner similar to that observed in previous crystallographic studies on these B1 MBLs. No cysteine oxidation was observed for IMP-1. Addition of tris(2-carboxyethyl)phosphine (TCEP) was found to prevent cysteine oxidation in BcII and VIM-2.

Crystals were observed growing in 10 of the 11 trials attempted, with only apo-di-Zn(II) IMP-1 failing to produce viable hits. High-resolution crystal structures for L- and

D-captopril in complex with BcII (1.10 and 1.18 Å resolution, respectively), IMP-1 (2.01 and 1.71 Å resolution, respectively), and VIM-2 (1.20 and 1.40 Å resolution, respectively) as well as di-Zn(II) BcII and VIM-2 in the absence of inhibitor (1.30 and 1.20 Å resolution, respectively) were determined. Structures for crystals grown in the *epi*-L- and *epi*-D-captopril–VIM-2 co-crystallisation trials were also solved (1.30 and 1.40 Å resolution, respectively). The crystals reached a maximum size after 1, 10 and 3 days for BcII, IMP-1 and VIM-2 experiments, respectively, in the conditions described in **Table 6.2**.

Crystals were cryoprotected using well solution diluted with 25% (v/v) glycerol before being flash cooled in liquid nitrogen. All data sets were collected at 100 K. All data were indexed, integrated, and scaled using HKL-3000.[86] The structures were solved by molecular replacement using Phaser.[87] The structures were then refined using PHENIX and Coot until the $R_{work}$ and $R_{free}$ values no longer decreased.[88,89] Data collection and refinement statistics for BcII, IMP-1 and VIM-2 complexes are provided in **Tables 6.3**-**6.5**, respectively.

***Table 6.2**: Crystallisation conditions for various MBLs in the apo-di-Zn(II) state or co-crystallised with a stereoisomer of captopril. Trials were conducted at RT.*

| | Sample composition | Crystallisation buffer (CB) | Vapor diffusion conditions |
|---|---|---|---|
| **Apo-di-Zn(II) BcII** | BcII (25 mg/mL) in CB + 1 mM TCEP | 0.1 M Bis-tris-HCl pH 5.5, 19% (w/v) PEG 3350, 0.05 M ammonium sulfate | Sitting drop, protein-to-CB ratio, 1:2, 293 K |
| **L-Captopril BcII** | BcII (25 mg/mL) in CB + 1 mM TCEP + 7.5 mM L-captopril | 0.1 M Bis-tris-HCl pH 5.5, 17% (w/v) PEG 3350, 0.15 M ammonium sulfate | Sitting drop, protein-to-CB ratio, 1:2, 293 K |
| **D-Captopril BcII** | BcII (25 mg/mL) in CB + 1 mM TCEP + 7.5 mM D-captopril | 0.1 M Bis-tris-HCl pH 5.5, 23% (w/v) PEG 3350, 0.1 M ammonium sulfate | Sitting drop, protein-to-CB ratio, 1:2, 293 K |
| **L-Captopril IMP-1** | IMP-1 (9 mg/mL) in CB + 1 mM TCEP + 7.5 mM L-captopril | 0.1 M sodium acetate pH 4.5, 23% (w/v) PEG 8000, 0.2 M $LiSO_4$ | Sitting drop, protein-to-CB ratio, 1:2, 293 K |
| **D-Captopril IMP-1** | IMP-1 (9 mg/mL) in CB + 1 mM TCEP + 7.5 mM D-captopril | 0.1 M sodium acetate pH 4.5, 23% (w/v) PEG 8000, 0.05 M $LiSO_4$ | Sitting drop, protein-to-CB ratio, 1:1, 293 K |
| **Apo-di-Zn(II) VIM-2** | VIM-2 (8 mg/mL) in CB + 1 mM TCEP | 0.25 M magnesium formate, 22% (w/v) PEG 3350 | Sitting drop, protein-to-CB ratio, 1:1, 293 K |
| **L-Captopril VIM-2** | VIM-2 (8 mg/mL) in CB + 1 mM TCEP + 7.5 mM L-captopril | 0.1 M magnesium formate, 23% (w/v) PEG 3350 | Sitting drop, protein-to-CB ratio, 1:2, 293 K |
| **D-Captopril VIM-2** | VIM-2 (8 mg/mL) in CB + 1 mM TCEP + 7.5 mM D-captopril | 0.1 M magnesium formate, 25% (w/v) PEG 3350 | Sitting drop, protein-to-CB ratio, 1:2, 293 K |
| ***Epi*-L-captopril VIM-2** | VIM-2 (8 mg/mL) in CB + 1 mM TCEP + 7.5 mM *epi*-L-captopril | 0.2 M magnesium formate, 22% (w/v) PEG 3350 | Sitting drop, protein-to-CB ratio, 1:2, 293 K |
| ***Epi*-D-captopril VIM-2** | VIM-2 (8 mg/mL) in CB + 1 mM TCEP + 7.5 mM *epi*-D-captopril | 0.1 M magnesium formate, 21% (w/v) PEG 3350 | Sitting drop, protein-to-well ratio, 1:1, 293 K |

***Table 6.2**: Processing and refinement statistics for BcII in the apo-di-Zn(II) state or co-crystallised with L- or D-captopril. I03 is a beamline at Diamond Light Source Oxford. †ASU = asymmetric unit. ‡RMSD = root mean square deviation.*

| BcII dataset | Apo-di-Zn(II) | L-Captopril | D-Captopril |
|---|---|---|---|
| **PDB ID** | 4C09 | 4C1H | 4C1C |
| **X-ray source** | I03 | I03 | I03 |
| **Resolution (outer shell) (Å)** | 34.7-1.20 (1.23-1.20) | 50.0-1.10 (1.14-1.10) | 40.2-1.18 (1.21-1.18) |
| **Unit cell dimensions a, b, c (Å)** | 53.3, 62.0, 69.5 | 53.0, 61.3, 69.6 | 53.1, 61.5, 69.5 |
| **Unit cell dimensions α, β, γ (degrees)** | 90.0, 93.0, 90.0 | 90.0, 93.0, 90.0 | 90.0, 93.0, 90.0 |
| **Space group** | *C* 1 2 1 | *C* 1 2 1 | *C* 1 2 1 |
| **Protein molecules per ASU†** | 1 | 1 | 1 |
| **Completeness (outer shell) (%)** | 98.3 (95.6) | 99.0 (92.5) | 95.0 (90.4) |
| **$R_{merge}$ (outer shell)** | 5.7 (51.1) | 8.7 (18.7) | 5.1 (59.5) |
| **No. of unique reflections** | 69108 | 89049 | 69108 |
| **I/σ mean (outer shell)** | 9.1 (2.0) | 18.1 (5.0) | 11.5 (2.3) |
| **No. of non-H protein atoms per ASU†** | 1698 | 1678 | 1662 |
| **RMSD from ideal bond length (Å) ‡** | 0.015 | 0.009 | 0.010 |
| **RMSD from ideal angles (degrees)** | 1.520 | 1.360 | 1.430 |
| **$R_{work}$ (%)** | 11.9 | 13.2 | 12.0 |
| **$R_{free}$ (%)** | 13.9 | 13.9 | 13.6 |

***Table 6.3**: Processing and refinement statistics for IMP-1 co-crystallised with L- or D-captopril. I03 is a beamline at Diamond Light Source Oxford. †ASU = asymmetric unit. ‡RMSD = root mean square deviation.*

| IMP-1 dataset | L-Captopril | D-Captopril |
|---|---|---|
| **PDB ID** | 4C1F | 4C1G |
| **X-ray source** | I03 | I03 |
| **Resolution (outer shell) (Å)** | 52.5-2.01 (2.08-2.01) | 64.86-1.71 (1.76-1.71) |
| **Unit cell dimensions a, b, c (Å)** | 193.6, 49.9, 54.6 | 194.6, 50.2, 54.6 |
| **Unit cell dimensions α, β, γ (degrees)** | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 |
| **Space group** | $P\ 2_1\ 2_1\ 2_1$ | $P\ 2_1\ 2_1\ 2_1$ |
| **Protein molecules per ASU†** | 2 | 2 |
| **Completeness (outer shell) (%)** | 99.9 (99.1) | 100 (100) |
| **$R_{merge}$ (outer shell)** | 9.3 (35.3) | 8.6 (63.1) |
| **No. of unique reflections** | 36158 | 58579 |
| **I/σ mean (outer shell)** | 9.7 (3.8) | 13.7 (3.2) |
| **No. of non-H protein atoms per ASU†** | 3290 | 3252 |
| **RMSD from ideal bond length (Å) ‡** | 0.004 | 0.020 |
| **RMSD from ideal angles (degrees)** | 0.766 | 1.730 |
| **$R_{work}$ (%)** | 18.4 | 14.5 |
| **$R_{free}$ (%)** | 22.4 | 17.6 |

***Table 6.5**: Processing and refinement statistics for VIM-2 in the apo-di-Zn(II) state or co-crystallised with a stereoisomer of captopril. I04-1 and I03 are beamlines at Diamond Light Source Oxford. FRE+ is an in-house X-ray generator equipped with a Saturn 944+ CCD detector. †ASU = asymmetric unit. ‡RMSD = root mean square deviation.*

| VIM-2 dataset | Apo-di-Zn(II) | L-Captopril | D-Captopril | *Epi*-L-captopril | *Epi*-D-captopril |
|---|---|---|---|---|---|
| **PDB ID** | 4BZ3 | 4C1D | 4C1E | - | - |
| **X-ray source** | I04-1 | I03 | FRE+ | FRE+ | FRE+ |
| **Resolution (outer shell) (Å)** | 50.0-1.30 (1.35-1.30) | 50.0-1.20 (1.24-1.20) | 50.0-1.40 (1.45-1.40) | 50.0-1.30 (1.35-1.30) | 50.0-1.40 (1.45-1.40) |
| **Unit cell dimensions a, b, c (Å)** | 103.0, 79.7, 68.1 | 102.2, 79.4, 67.8 | 103.9, 79.0, 67.2 | 102.9, 79.3, 67.6 | 103.1, 79.5, 67.8 |
| **Unit cell dimensions α, β, γ (degrees)** | 90, 130.6, 90 | 90.0, 130.3, 90.0 | 90.0, 131.6, 90.0 | 90.0, 130.9, 90.0 | 90.0, 130.5, 90.0 |
| **Space group** | *C* 1 2 1 | *C* 1 2 1 | *C* 1 2 1 | *C* 1 2 1 | *C* 1 2 1 |
| **Protein molecules per ASU†** | 2 | 2 | 2 | 2 | 2 |
| **Completeness (outer shell) (%)** | 98.2 (97.0) | 99.1 (91.0) | 98.7 (95.3) | 98.6 (95.1) | 98.9 (95.5) |
| **$R_{merge}$ (outer shell)** | 11.0 (25.3) | 17.1 (61.2) | 10.0 (34.7) | 11.5 (37.7) | 14.0 (30.1) |
| **No. of unique reflections** | 101604 | 126706 | 78943 | 75921 | 79873 |
| **I/σ mean (outer shell)** | 10.9 (3.5) | 10.6 (1.9) | 10.7 (2.0) | 10.7 (2.2) | 10.9 (2.7) |
| **No. of non-H protein atoms per ASU†** | 3477 | 3464 | 3484 | 3457 | 3464 |
| **RMSD from ideal bond length (Å) ‡** | 0.025 | 0.025 | 0.024 | 0.026 | 0.025 |
| **RMSD from ideal angles (degrees)** | 1.239 | 1.403 | 1.420 | 1.228 | 1.391 |
| **$R_{work}$ (%)** | 11.3 | 14.3 | 13.2 | 15.4 | 14.8 |
| **$R_{free}$ (%)** | 13.8 | 17.2 | 16.7 | 16.6 | 17.6 |

### Protein Data Bank accession numbers

Coordinates and structure factors can be found in the Protein Data Bank (PDB) under the following accession codes: apo-di-Zn(II) BcII, PDB entry 4C09; BcII + L-captopril, PDB entry 4C1H; BcII + D-captopril, PDB entry 4C1C; IMP-1 + L-captopril, PDB entry 4C1F; IMP-1 + D-captopril, PDB entry 4C1G; apo-di-Zn(II) VIM-2, PDB entry 4BZ3; VIM-2 + L-captopril, PDB entry 4C1D; VIM-2 + D-captopril, PDB entry 4C1E. Structures for VIM-2 bound to *epi*-L- or *epi*-D-captopril were not deposited due to the poorly defined electron density observed in the active site of each structure.

### 6.6.2 Structural analysis

As well as the conserved Zn(II) ion binding sites described in **Section 6.1.3** members of the MBL family typically display an αβ/βα sandwich fold consisting of a core of β-sheets flanked by α-helices.[90] The Zn(II) binding site is located in a groove at one end of the two β-sheets and is surrounded by several loops. Two of the active-site loops, termed L3 and L10, are crucial for substrate binding and recognition (see **Figure 6.8**).[91] The L3 loop (residues 61-66 in the class B β-lactamase (BBL) numbering) connects the β3 and β4 strands and is thought to be most important for substrate specificity, as mutation of only two residues is required to dictate what substrates will bind.[92] The L10 loop (residues 223-241), located between strand β11 and helix α4, also contains residues crucial for substrate and inhibitor binding including K224/R228 and N233. Crystal structures of apo-di-Zn(II) BcII and VIM-2, and IMP-1 bound to D-captopril are shown in **Figure 6.8** displaying the abovementioned features. Note that the standard BBL numbering system is used throughout this chapter.[36]

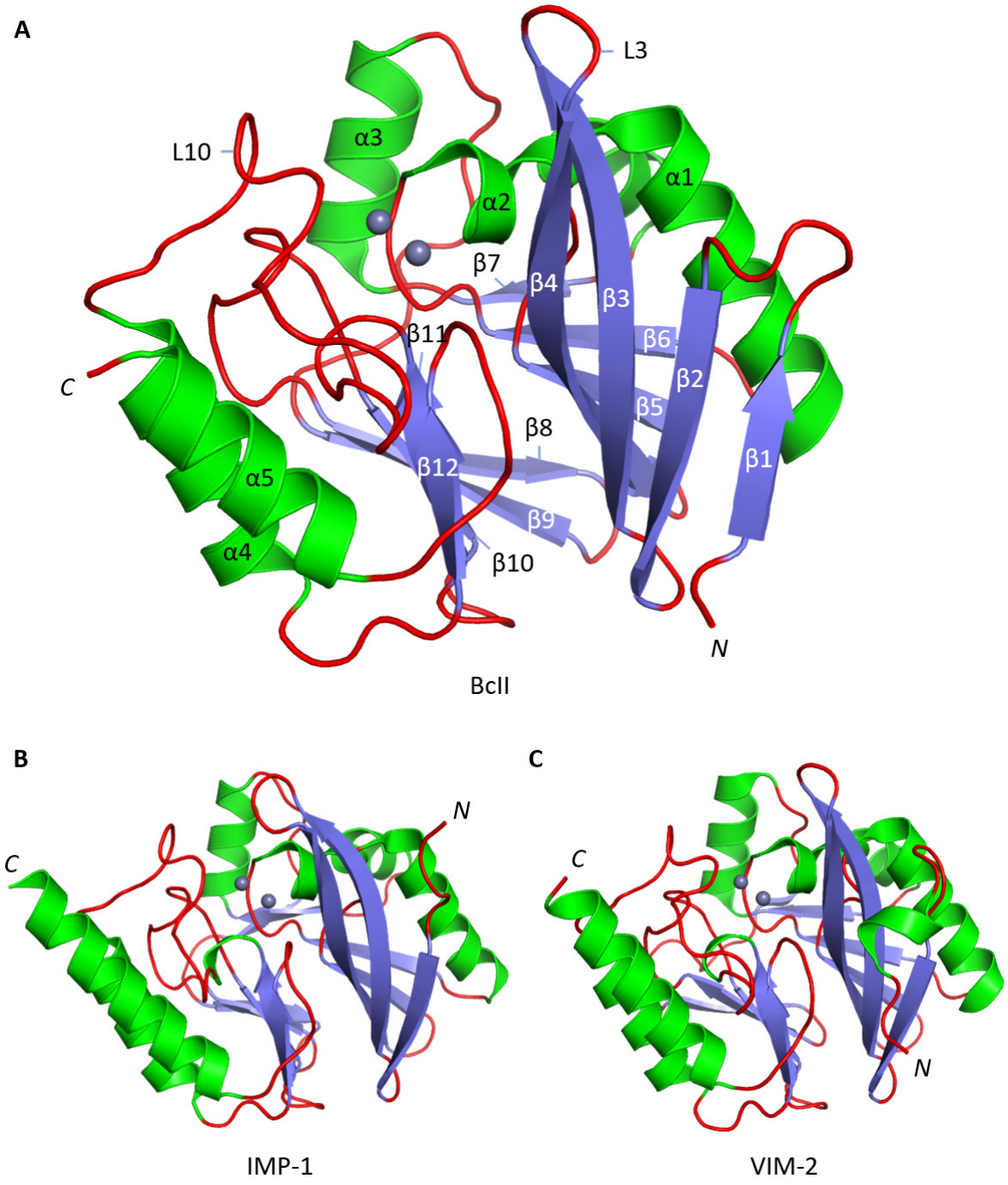



***Figure 6.8**: Views from crystal structures of B1 MBLs. α-Helices are shown in green, β sheets in blue, loops in red. Zn(II) ions are shown as grey spheres. **A**- BcII showing the characteristic MBL fold. The labelled L3 and L10 loops flank the zinc active site and are involved in substrate binding. The BcII αβ/βα sandwich comprises α1-α2, β1-β7, β8-β12 and α4-α5, sequentially. **B** and **C**- D-captopril-bound IMP-1 and apo-di-Zn(II) VIM-2, respectively. For the sake of clarity D-captopril is not shown in the IMP-1 structure. Despite both proteins sharing only 38% sequence homology with BcII, and 33% sequence homology with each other they all adopt a very similar tertiary structure. The β1 strand in BcII is replaced by a loop (IMP-1) or a loop and α-helix (VIM-2) in the other MBLs presented.*

**BcII**

The BcII structures comprise 222 amino acids and only one molecule per asymmetric unit. The L3 loop is highly flexible, and residues 64-67 were not modelled in the D-captopril structure. Comparisons between the active site of BcII in the apo-di-Zn(II) and the L- and D-captopril-bound state are provided in **Figure 6.9**. All structures appear to have multiple conformations of the active site cysteine (C221) with 80:20 occupancies. As the zinc it coordinates (Zn2) and both inhibitors also appear to be at ~80% occupancy it seems likely that the effects are due to partial zinc occupancy. The apo-di-Zn(II) structure contains a bridging water/hydroxide molecule between the active site zinc atoms and a second water bound to Zn1. Both of these Zn(II)-bound water molecules are displaced on binding by L- or D-captopril. Additionally, binding of the captopril molecule results in a ~0.7 Å shift in the position of Zn2.

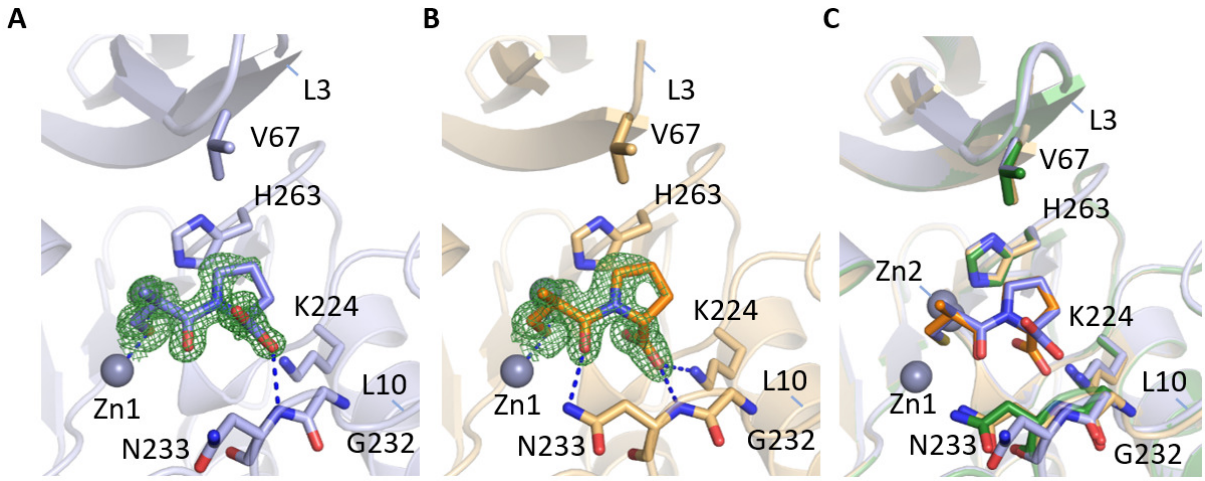



***Figure 6.9**: Views of the zinc binding site from crystal structures of BcII complexed with L- and D-captopril (**A** and **B**, respectively). L-Captopril-bound BCII is shown in grey, L-captopril in slate, D-captopril-bound BcII in light orange and D-captopril in orange. Close contacts (within 3.5 Å) are indicated with a blue dotted line. |Fo-Fc| positive electron density is displayed (green mesh) contoured to 3 σ. Density was calculated from the final model, following omission of the ligand, against experimental diffraction data. **C**- Overlay of the apo-di-Zn(II) (dark green), L- and D-captopril–BcII structures.*

The most potent inhibitor of BcII is D-captopril. D-Captopril binds to the active site zinc ions through its thiol leaving the carbonyl group in a position to hydrogen bond to the side-chain of N233 (3.0 Å). The carboxylate group is additionally positioned able to hydrogen bond to the side-chain of K224 (3.0 Å) and the backbone nitrogen of N233 (2.9 Å) (see **Figure 6.9 B**).

The conformations of L- and D-captopril in the BcII active site are almost identical, differing only at the carboxylic acid. This is somewhat surprising as L-captopril is only positioned to hydrogen bond to one of the three groups that D-captopril can access; the N233 backbone nitrogen through its carboxyl group (2.7 Å). The L-captopril carboxyl group pushes the L10 loop away from the active site by up to 2 Å, prohibiting hydrogen bonding with the N233 side-chain.

**IMP-1**

The IMP-1 structures are made up of 219 residues in each of the two molecules in the asymmetric unit. One of these molecules is highly disordered making deductions about the contents of its active site difficult. As no high-resolution structure of native IMP-1 crystallised under the conditions used in these experiments has been reported comparisons are made between the L- and D-captopril structures only (see **Figure 6.10 C**). In both L- and D-captopril structures of IMP-1, the ligands were modelled and refined with 100% occupancy in chain A, but the residual density present in chain B was interpreted as too weak (<50% occupancy) to include in the model. The following analyses concern the IMP-1 chain A molecules only.

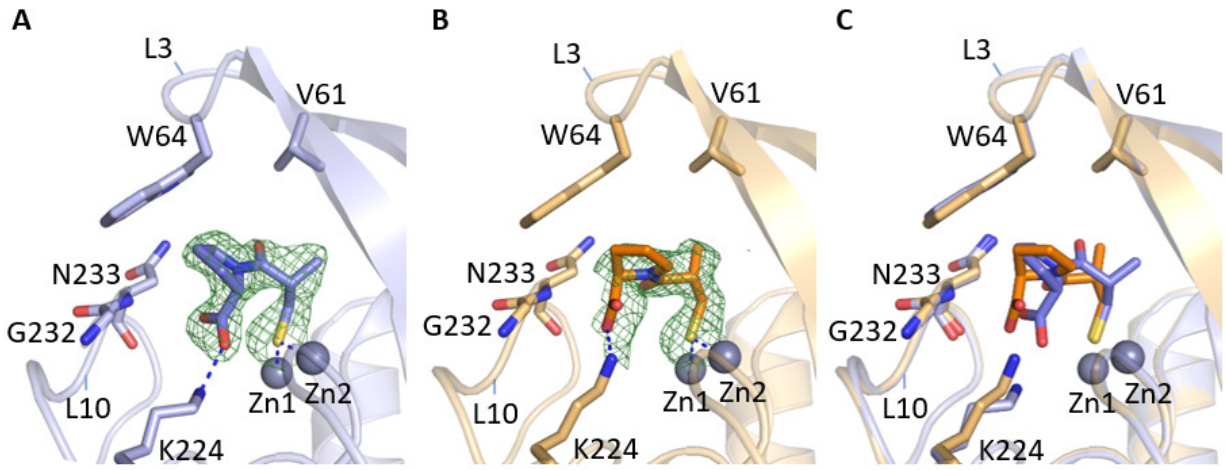



***Figure 6.10**: Views of the zinc binding site from crystal structures of IMP-1 complexed with L- and D-captopril (**A** and **B**, respectively). L-Captopril-bound IMP-1 is shown in grey, L-captopril in slate, D-captopril-bound IMP-1 in light orange and D-captopril in orange. Close contacts (within 3.5 Å) are indicated with a blue dotted line. |Fo-Fc| positive electron density is displayed (green mesh) contoured to 3 σ. Density was calculated from the final model, following omission of the ligand, against experimental diffraction data. **C**- Overlay of the L- and D-captopril–IMP-1 structures.*

As with BcII, both captopril molecules bind to the two Zn(II) ions of IMP-1 through their thiol group (see **Figure 6.10**). A second major interaction observed is the salt bridge between the captopril carboxyl group and the K224 side-chain amine located on the L10 loop. The inter-atomic distances between the closest D-captopril carboxylate group oxygen and the side-chain $N^{\varepsilon}$ of N233 are too great to allow for a strong hydrogen bond (3.6 Å) (see **Figure 6.10 B**). While in the BcII structures only the captopril carboxylate differed in position between captopril structures, significant conformational adjustments are required for the stereoisomers to bind the same locations in IMP-1 with their thiol and carboxylate groups. D-Captopril appears able to form a salt bridge with to the side-chain of K224 through both oxygens of its carboxylate group (2.3 Å, each), whereas L-captopril appears to be able to interact only through one oxygen (2.3 Å). In all structures reported in this chapter bar one, the captopril carbonyl points towards a conserved N233 residue, often close enough to form a hydrogen bond with the asparagine

side-chain. The exception is the IMP-1–L-captopril co-crystal structure, wherein the carbonyl of L-captopril is apparently forced away from N233 by the torsional requirements of reaching K224 with its carboxylic acid (see **Figure 6.10 A**). While L-captopril may form a strong interaction with K224, the effect of pushing its carbonyl towards the hydrophobic L3 loop may explain why L-captopril is still not as strong an inhibitor of IMP-1 as D-captopril.

No other major binding interactions are observed and, aside from the side-chain of K224, there is little movement in either the L3 or L10 loop between the IMP-1 crystal structures. Lacking an apo-di-Zn(II) IMP-1 structure crystallised under equivalent conditions, it is not currently possible to determine whether both, or neither captopril stereoisomers induce a movement in the loop.

#### VIM-2

All the structures of VIM-2 described here are comprised of 231 residues, and two molecules are present in the asymmetric unit. Both molecules in the asymmetric unit display the same inhibitor binding conformations. L-Captopril was modelled at 80% occupancy and D-captopril was modelled at 100% occupancy. The largest change to the crystal structure seen on substrate binding is in the L3 loop, which contains residues F61 and Y67 (see **Figure 6.11**). Poor density for F61 indicates disorder in the residue, suggesting a lack of strong interactions with the phenyl ring in the native structure (see **Figure 6.11 C**). Views from the binding site of the L- and D-captopril–VIM-2 co-crystal structures, as well as the apo-di-Zn(II) VIM-2 structure, are shown in **Figure 6.11**.

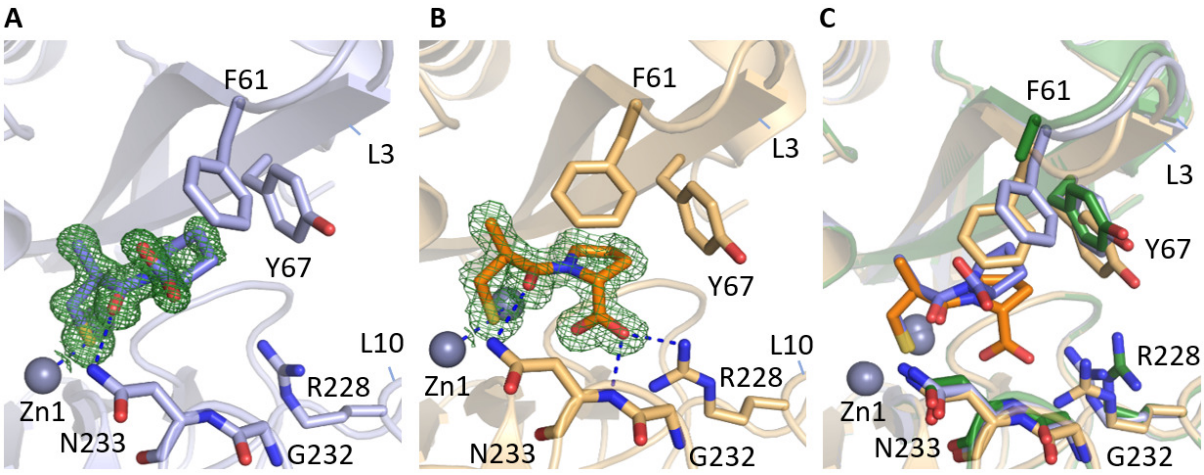



***Figure 6.11**: Views of the zinc binding site from crystal structures of VIM-2 complexed with L- and D-captopril (**A** and **B**, respectively). L-Captopril-bound VIM-2 is shown in grey, L-captopril in slate, D-captopril-bound VIM-2 in light orange and D-captopril in orange. Close contacts (within 3.5 Å) are indicated with a blue dotted line. |Fo-Fc| positive electron density is displayed (green mesh) contoured to 3 σ. Density was calculated from the final model, following omission of the ligand, against experimental diffraction data. **C**- Overlay of the apo-di-Zn(II)- (dark green), L- and D-captopril–VIM-2 structures.*

The most potent MBL–inhibitor combination tested in in this chapter was the inhibition of VIM-2 by D-captopril. As in the previously described MBL–captopril binding modes, D-captopril binds to both Zn(II) ions in the VIM-2 binding site *via* displacement of the bridging water/hydroxide ion by the D-captopril thiol group. As in the case of BcII, Zn2 is also displaced slightly (~0.4 Å) on binding by the captopril stereoisomers. D-Captopril is also positioned to form hydrogen bonds analogous to those described between D-captopril and BcII. The D-captopril carbonyl is positioned to form a hydrogen bond with the side-chain nitrogen of N233 (3.0 Å). A D-captopril carboxylate oxygen is also close enough to form a salt bridge with both the side-chain guanidinium group of R228 (3.0 Å), and the N233 backbone amide (2.9 Å). Of all VIM-2 structures discussed in this report only the D-captopril co-crystal structure shows R228 bent towards the active site. Inhibitor binding causes a lowering of the L3 loop into the active site by up to 2.0 Å, likely due to cation-π interactions between F61 and the D-captopril nitrogen (4.1-4.6 Å

separation). This movement of the L3 loop is consistent with an induced fit mechanism during inhibitor binding. Clearer electron density of F85 in this structure compared to the apo-di-Zn(II) structure indicates an increase in order likely brought on by favourable inhibitor interactions. The ionic form of the captopril amide bond is likely stabilised by both cation-π interactions with F85 and hydrogen bonding with N233.

L-Captopril is shown to bind to the Zn(II) ions in an identical way to D-captopril and it is only after the $C^2$-position that the binding modes of the compounds diverge. The L-captopril carbonyl oxygen is still close enough to pick up hydrogen bonding interactions with the N233 side-chain nitrogen (2.9 Å), but due to the inversion of chirality at the carboxylate position, that group is no longer able to hydrogen bond to either N233 or R288. Additionally, the ring *S*-stereocenter points the carboxylate towards F85 prohibiting cation-π interactions occurring and instead forces the residue, and so the L3 loop, upwards, out of the active site by up to 0.7 Å. These observations correlate well with the relative inhibition levels observed against VIM-2.

#### *Epi*-captopril co-crystal structures

The refined maps were used to attempt to fit *epi*-L-captopril into the active site. While it was possible to fit the majority of the inhibitor into a 2|Fo-Fc| electron density cloud, much of the molecule was still not observed. *Epi*-D- and *epi*-L-captopril are worse inhibitors of VIM-2 than the other stereoisomers of captopril, so a less clear density map was expected. Omit maps calculated for the *epi*-L-captopril–VIM-2 dataset that excluded the inhibitor indicated electron density was present in the binding site where the inhibitor was expected. However, as the shape of the unbiased density was indistinguishable from that of a glycerol molecule, it was decided to abandon the dataset.

Omit maps calculated from the *epi*-D-captopril–VIM-2 dataset suggested the low occupancy presence of D-captopril instead of the expected *epi*-D-captopril enantiomer. NMR analysis of the *epi*-D-captopril compound showed it to be contaminated with ~5% D-captopril (data not shown). This revelation explained the observed density and may also suggest that the highly active D-captopril may have been responsible for the relatively strong inhibition previously attributed to *epi*-D-captopril. The *epi*-D-captopril dataset was therefore also discarded.

The conserved binding mode of L- and D-captopril up until the $C^2$-position indicates that it may be a reasonably stable interaction, which may suggest why epimerisation at the $C^2$-position results in a loss of activity. For L- and D-captopril the $C^2$ geometry allows for the carbonyl and the carboxylate groups to point out of the active site towards the bulk of the water and positively charged residues. By inverting the stereochemistry an equivalent binding mode would force the bulk of the molecule into the less flexible core of the binding site where fewer positively charged groups are accessible. This may explain the decrease in activity of the *epi*-captoprils.

**Comparisons of captopril binding between metallo-β-lactamases**

Due to the stereochemistry on the proline ring, captopril can be considered to have two binding faces. The carboxylic acid defines the hydrophilic binding face and is used, in every D-captopril-MBL structure presented in this chapter, to bind to the conserved K224/R228 residue on the L10 loop, perhaps explaining the relatively high levels of inhibition by D-captopril. This conserved positively charged residue is predicted to bind to the carboxylate group present in most β-lactam antibiotics.[93] It should be noted that the *R*-stereochemistry at the D-captopril carboxylic acid is more similar to that of the

hydrolysed penicillin carboxylic acid group than the equivalent position in L-captopril (see **Figure 6.7**). While MBLs have been found to be susceptible to product inhibition, and D-captopril is more structurally similar to hydrolysed β-lactams, D-captopril appears to adopt a differing binding mode.[65] Crystal structures that have previously been solved of MBLs in complex with hydrolysed β-lactams show the antibiotics binding to Zn1 through an oxygen atom of the carboxylic acid formed on β-lactam hydrolysis, and to Zn2 through the imidazole nitrogen and an adjacent carboxylic acid oxygen (see **Figure 6.5**).[42] D-Captopril, by comparison, binds to the two Zn(II) ions only once, simultaneously through the captopril thiol, binding instead K224/R228 through its carboxylic acid (see **Figures 6.9-6.11 B**).

The obverse, hydrophobic face of the proline ring points in the same direction as the hydrophobic captopril methyl group in D-captopril structures, which can interact with the hydrophobic residues on the L3 loop (V67; BcII, W64 and V61; IMP-1, F61 and Y67; VIM-2) (see **Figures 6.9-6.11 B**). The relative ease with which D-captopril can take advantage of these interaction may go some way to explaining its potency of inhibition.

A feature common to all L- and D-captopril structures solved in this chapter is the binding of the captopril thiol group to the two binding-site Zn(II) ions, displacing the assigned catalytic bridging water/hydroxide (assumed for IMP-1). This is in agreement with the captopril binding modes previously observed for B1 (NDM-1–L-captopril and BlaB–D-captopril) and B3 (L1–L-captopril) MBLs.[50,64,69] The binding mode of D-captopril to the mono-Zn(II), B2 MBL, CphA differs distinctly from the conformation adopted on bonding to di-Zn(II) MBLs.[66] Rather than interacting with the Zn(II) ion through the captopril thiol group, zinc binding is instead observed through the captopril carbonyl.[66]

A markedly different binding mode for the di-Zn(II) B3 enzyme FEZ-1 to D-captopril has also been reported, placing the inhibitor at a different location in the binding site, not interacting with the Zn(II) ions at all.[70] It should be noted, however, that half of the atoms in the D-captopril structure are not observed in the FEZ-1 model, and those that are can be equally well modelled as solvent.

## 6.7 Discussion

Antibiotic resistance in MBL-expressing organisms is an increasingly urgent public health concern.[94] VIM-2 in particular has become hugely problematic in parts of Asia, being present in up to 99% of all MBL-positive multidrug-resistant strains.[95] As the arsenal of antibiotics the medical community has at its disposal becomes increasingly ineffective it has become imperative to develop new ways of fighting drug resistant infections. The development of inhibitors of the enzymes that convey antibiotic resistance to pathogenic organisms is one such possibility.

The effect of stereochemistry on the potency of an inhibitor against a range of MBL targets was investigated. D-Captopril was identified as the most potent inhibitor of the stereoisomers of captopril against every MBLs tested. Crystal structures were solved of L- and D-captopril in complex with two clinically relevant MBLs, IMP-1 and VIM-2, as well as the model MBL BcII. While several crystal structures of MBLs in combination with either L- or D-captopril have been previously reported, this investigation provides the first examples of MBLs being co-crystallised with multiple different captopril stereoisomers (separately). This provided the first opportunity to directly probe the effect of captopril stereochemistry on MBL binding.

Correlations can be identified between the binding modes of the captopril stereoisomers and their relative potency of inhibition. Each MBL crystallised contained a positively charged K224/R228 residue and an aptly placed N233 residue on the L10 loop, as well as a mobile, hydrophobic L3 loop hanging over the binding site. D-Captopril had the appropriate geometry to make the most of all of these interactions, and at the same time displays the highest potency of inhibition of all captopril stereoisomers. L-Captopril was consistently poised to engage in fewer hydrogen bonds to the MBL than D-captopril, and subsequently induced lower inhibition. The smallest measured gap between L- and D-captopril $K_i$ values was observed for IMP-1. IMP-1 was the only enzyme that L-captopril was able to bind to in such a way as to form a salt bridge with K224, potentially explaining the similarity between the L- and D-captopril binding strength.

D-Captopril has been shown to inhibit a mechanism of antibiotic resistance for which there is no current medical treatment. D-Captopril is able to inhibit a wide range of MBLs across all the MBL subclasses while showing no inhibition at 100 μM against the four human MBL-fold enzymes tested. In addition, D-captopril is "druglike", as evidenced by the clinical use of L-captopril as an ACE inhibitor. The insights provided by the binding-mode analyses detailed in this investigation will hopefully aid in the design of potent, broad-spectrum, or targeted, enzyme specific, inhibitors of the MBLs in the future.

## 6.8 References


1. Fleming, A. G. Responsibilities and opportunities of the private practitioner in preventive medicine. *Can. Med. Assoc. J.* **20,** 11–3 (1929).
2. Schatz, A., Bugie, E. & Waksman, S. A. Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. 1944. *Clin. Orthop. Relat. Res.* **55,** 3–6 (2005).
3. Lewis, K. Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.* **12,** 371–87 (2013).
4. Vazquez, D. The streptogramin family of antibiotics. *Antibiotics* **1,** 387–403 (1967).
5. Eliopoulos, G. M. *et al. In vitro* and *in vivo* activity of LY 146032, a new cyclic lipopeptide antibiotic. *Antimicrob. Agents Chemother.* **30,** 532–535 (1986).
6. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517,** 455–459 (2015).
7. Gallagher, J. BBC News - Antibiotic 'apocalypse' warning. *19 November 2015* (2013). Available at: http://www.bbc.co.uk/news/health-21702647. (Accessed: 20th January 2018).
8. Strominger, J. L., Park, J. T. & Thompson, R. E. Composition of the cell wall of *Staphylococcus aureus*: Its relation to the mechanism of action of penicillin. *J. Biol. Chem.* **234,** 3263–3268 (1959).
9. Van Bambeke, F. Glycopeptides and glycodepsipeptides in clinical development: A comparative review of their antibacterial spectrum, pharmacokinetics and clinical efficacy. *Curr. Opin. Investig. Drugs* **7,** 740–749 (2006).
10. Hutchinson, J. A., Burholt, S. & Hamley, I. W. Peptide hormones and lipopeptides: From self-assembly to therapeutic applications. *J. Pept. Sci.* **23,** 82–94 (2017).
11. Suto, M. J., Domagala, J. M., Roland, G. E., Mailloux, G. B. & Cohen, M. A. Fluoroquinolones: Relationships between structural variations, mammalian cell cytotoxicity, and antimicrobial activity. *J. Med. Chem.* **35,** 4745–4750 (1992).
12. Calvori, C., Frontali, L., Leoni, L. & Tecce, G. Effect of rifamycin on protein synthesis. *Nature* **207,** 417–8 (1965).
13. Chopra, I. & Roberts, M. Tetracycline antibiotics: Mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev.* **65,** 232–260 (2001).
14. Jardetzky, O. Studies on the mechanism of action of chloramphenicol. *J. Biol. Chem.* **238,** 2498–2508 (1963).
15. Gaynor, M. & Mankin, A. S. Macrolide antibiotics: Binding site, mechanism of action, resistance. *Curr. Top. Med. Chem.* **3,** 1–12 (2003).
16. Bozdogan, B. & Appelbaum, P. C. Oxazolidinones: Activity, mode of action, and mechanism of resistance. *Int. J. Antimicrob. Agents.* **23,** 113–119 (2004).

17. Yun, M. K. *et al.* Catalysis and sulfa drug resistance in dihydropteroate synthase. *Science* **335,** 1110–4 (2012).

18. Demain, A. L. & Elander, R. P. The β-lactam antibiotics: Past, present, and future. *Antonie Van Leeuwenhoek* **75,** 5–19 (1999).

19. Elander, R. P. Industrial production of β-lactam antibiotics. *Appl. Microbiol. Biotechnol.* **61,** 385–392 (2003).

20. Leonard, D. A., Bonomo, R. A. & Powers, R. A. Class D β-lactamases: A reappraisal after five decades. *Acc. Chem. Res.* **46,** 2407–2415 (2013).

21. Livermore, D. & Williams, J. Mode of action and mechanisms of bacterial resistance. In *Antibiotics in Laboratory Medicine* 502–577 (Williams and Wilkins, Baltimore, 1996).

22. Wise, E. M. & Park, J. T. Penicillin: Its basic site of action as an inhibitor of a peptide cross-linking reaction in cell wall mucopeptide synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **54,** 75–81 (1965).

23. Waxman, D. J., Yocum, R. R. & Strominger, J. L. Penicillins and cephalosporins are active site-directed acylating agents: Evidence in support of the substrate analogue hypothesis. *Philos. Trans. R. Soc.* **289,** 257–71 (1980).

24. D'Costa, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477,** 457–461 (2011).

25. Bhullar, K. *et al.* Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS One* **7,** e34953 (2012).

26. Hughes, V. M. & Datta, N. Conjugative plasmids in bacteria of the 'pre-antibiotic' era. *Nature* **302,** 725–726 (1983).

27. Knapp, C. W., Dolfing, J., Ehlert, P. A. I. & Graham, D. W. Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940. *Environ. Sci. Technol.* **44,** 580–587 (2010).

28. Laible, G., Spratt, B. G. & Hakenbeck, R. Interspecies recombinational events during the evolution of altered PBP 2*x* genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Mol. Microbiol.* **5,** 1993–2002 (1991).

29. Hopkins, J. M. & Towner, K. J. Enhanced resistance to cefotaxime and imipenem associated with outer membrane protein alterations in *Enterobacter aerogenes*. *J. Antimicrob. Chemother.* **25,** 49–55 (1990).

30. Jacoby, G. A., Mills, D. M. & Chow, N. Role of β-lactamases and porins in resistance to ertapenem and other β-lactams in *Klebsiella pneumoniae*. *Antimicrob. Agents Chemother.* **48,** 3203–3206 (2004).

31. Poole, K. Efflux-mediated multiresistance in gram-negative bacteria. *Clin. Microbiol. Infect.* **10,** 12–26 (2004).

32. Li, X. Z., Ma, D., Livermore, D. M. & Nikaido, H. Role of efflux pump(s) in intrinsic resistance of *Pseudomonas aeruginosa*: Active efflux as a contributing factor to β-lactam resistance. *Antimicrob. Agents Chemother.* **38,** 1742–1752 (1994).

33. Bebrone, C. Metallo-β-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. *Biochem. Pharmacol.* **74,** 1686–1701 (2007).

34. Pratt, R. F. Functional evolution of the serine β-lactamase active site. *J. Chem. Soc. Perkin Trans.* **20,** 851–861 (2002).

35. Lamotte-Brasseur, J., Dive, G., Dehareng, D. & Ghuysen, J. M. Electrostatic potential maps at the quantum chemistry level of the active sites of the serine peptidases, α-chymotrypsin and subtilisin. *J. Theor. Biol.* **145,** 183–98 (1990).

36. Garau, G. *et al.* Update of the standard numbering scheme for class B β-lactamases. *Antimicrob. Agents Chemother.* **48,** 2347–2349 (2004).

37. Galleni, M. *et al.* Standard numbering scheme for class B β-lactamases. *Antimicrob. Agents Chemother.* **45,** 660–663 (2001).

38. King, D. & Strynadka, N. Crystal structure of New Delhi metallo-β-lactamase reveals molecular basis for antibiotic resistance. *Protein Sci.* **20,** 1484–1491 (2011).

39. Fonseca, F., Bromley, E. H. C., Saavedra, M. J., Correia, A. & Spencer, J. Crystal structure of *Serratia fonticola* Sfh-I: Activation of the nucleophile in mono-zinc metallo-β-lactamases. *J. Mol. Biol.* **411,** 951–959 (2011).

40. Ullah, J. H. *et al.* The crystal structure of the L1 metallo-β-lactamase from *Stenotrophomonas maltophilia* at 1.7 Å resolution. *J. Mol. Biol.* **284,** 125–136 (1998).

41. Weston, J. Mode of action of bi- and trinuclear zinc hydrolases and their synthetic analogues. *Chemical Reviews* **105,** 2151–2174 (2005).

42. Meini, M. R., Llarrull, L. I. & Vila, A. J. Overcoming differences: The catalytic mechanism of metallo-β-lactamases. *FEBS Lett.* **589,** 3419–3432 (2015).

43. Palzkill, T. Metallo-β-lactamase structure and function. *Ann. N. Y. Acad. Sci.* **1277,** 91–104 (2013).

44. Spencer, J., Clarke, A. R. & Walsh, T. R. Novel mechanism of hydrolysis of therapeutic β-lactams by *Stenotrophomonas maltophilia* L1 metallo-β-lactamase. *J. Biol. Chem.* **276,** 33638–33644 (2001).

45. Oelschlaeger, P. *et al.* Meropenem and chromacef intermediates observed in IMP-25 metallo-β-lactamase-catalyzed hydrolysis. *Antimicrob. Agents Chemother.* **59,** 4326–4330 (2015).

46. Yang, H. *et al.* Mechanistic and spectroscopic studies of metallo-β-lactamase NDM-1. *Biochemistry* **51,** 3839–3847 (2012).

47. Wang, Z., Fast, W. & Benkovic, S. J. On the mechanism of the metallo-β-lactamase from *Bacteroides fragilis*. *Biochemistry* **38,** 10013–10023 (1999).

48. Concha, N. O., Rasmussen, B. A., Bush, K. & Herzberg, O. Crystal structure of the wide-spectrum binuclear zinc β-lactamase from *Bacteroides fragilis*. *Structure* **4,** 823–836 (1996).

49. Page, M. I. & Badarau, A. The mechanisms of catalysis by metallo β-lactamases. *Bioinorg. Chem. Appl.* **2008,** 1–14 (2008).

50. King, D. T., Worrall, L. J., Gruninger, R. & Strynadka, N. C. J. New Delhi metallo-β-lactamase: Structural insights into β-lactam recognition and inhibition. *J. Am. Chem. Soc.* **134,** 11362–11365 (2012).

51. Spencer, J. *et al.* Antibiotic recognition by binuclear metallo-β-lactamases revealed by X-ray crystallography. *J. Am. Chem. Soc.* **127,** 14439–14444 (2005).

52. Essack, S. Y. The development of beta-lactam antibiotics in response to the evolution of beta-lactamases. *Pharm. Res.* **18,** 1391–1399 (2001).

53. Meletis, G. Carbapenem resistance: Overview of the problem and future perspectives. *Ther. Adv. Infect. Dis.* **3,** 15–21 (2016).

54. Kumarasamy, K. K. *et al.* Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: A molecular, biological, and epidemiological study. *Lancet Infect. Dis.* **10,** 597–602 (2010).

55. Drawz, S. M. & Bonomo, R. A. Three decades of β-lactamase inhibitors. *Clin. Microbiol. Rev.* **23,** 160–201 (2010).

56. Brown, A. G. Clavulanic acid, a novel beta-lactamase inhibitor: A case study in drug discovery and development. *Drug Des. Deliv.* **1,** 1–21 (1986).

57. Fast, W. & Sutton, L. D. Metallo-β-lactamase: Inhibitors and reporter substrates. *Biochim. Biophys. Acta* **1834,** 1648–1659 (2013).

58. Brown, R. P. A., Aplin, R. T. & Schofield, C. J. Inhibition of TEM-2 β-lactamase from *Escherichia coli* by clavulanic acid: Observation of intermediates by electrospray ionization mass spectrometry. *Biochemistry* **35,** 12421–12432 (1996).

59. Ehmann, D. E. *et al.* Avibactam is a covalent, reversible, non–β-lactam β-lactamase inhibitor. *Proc. Natl. Acad. Sci.* **109,** 11663–11668 (2012).

60. Lahiri, S. D. *et al.* Avibactam and class C β-lactamases: Mechanism of inhibition, conservation of the binding pocket, and implications for resistance. *Antimicrob. Agents Chemother.* **58,** 5704–5713 (2014).

61. Brem, J. *et al.* Rhodanine hydrolysis leads to potent thioenolate mediated metallo-β-lactamase inhibition. *Nat. Chem.* **6,** 1084–1090 (2014).

62. Buynak, J. D. β-Lactamase inhibitors: A review of the patent literature (2010-2013). *Expert Opin. Ther. Pat.* **23,** 1469–81 (2013).

63. Li, N. *et al.* Simplified captopril analogues as NDM-1 inhibitors. *Bioorganic Med. Chem. Lett.* **24,** 386–389 (2014).

64. Nauton, L., Kahn, R., Garau, G., Hernandez, J. F. & Dideberg, O. Structural insights into the design of inhibitors for the L1 metallo-β-lactamase from *Stenotrophomonas maltophilia*. *J. Mol. Biol.* **375,** 257–269 (2008).

65. Badarau, A. & Page, M. I. The variation of catalytic efficiency of *Bacillus cereus*

metallo-β-lactamase with different active site metal ions. *Biochemistry* **45,** 10654–10666 (2006).

66. Heinz, U. *et al.* Coordination geometries of metal ions in D- or L-captopril-inhibited metallo-β-lactamases. *J. Biol. Chem.* **278,** 20659–20666 (2003).

67. Liénard, B. M. R. *et al.* Structural basis for the broad-spectrum inhibition of metallo-β-lactamases by thiols. *Org. Biomol. Chem.* **6,** 2282–94 (2008).

68. Rydzik, A. M. *et al.* Monitoring conformational changes in the NDM-1 metallo-β-lactamase by $^{19}$F NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **53,** 3129–3133 (2014).

69. García-Sáez, I. *et al.* The 1.5-Å structure of *Chryseobacterium meningosepticum* zinc β-lactamase in complex with the inhibitor, D-captopril. *J. Biol. Chem.* **278,** 23868–23873 (2003).

70. García-Sáez, I. *et al.* Three-dimensional structure of FEZ-1, a monomeric subclass B3 metallo-β-lactamase from *Fluoribacter gormanii*, in native form and in complex with D-captopril. *J. Mol. Biol.* **325,** 651–660 (2003).

71. Suh, J. T. *et al.* Angiotensin-converting enzyme inhibitors. New orally active antihypertensive (mercaptoalkanoyl)- and [(acylthio)alkanoyl]glycine derivatives. *J. Med. Chem.* **28,** 57–66 (1985).

72. Kim, B. H., Lee, H. B., Hwang, J. K. & Kim, Y. G. Asymmetric induction in the conjugate addition of thioacetic acid to methacrylamides with chiral auxiliaries. *Tetrahedron-Asymmetry* **16,** 1215–1220 (2005).

73. Watanabe, M., Iyobe, S., Inoue, M. & Mitsuhashi, S. Transferable imipenem resistance in *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **35,** 147–151 (1991).

74. Poirel, L. *et al.* Characterization of VIM-2, a carbapenem-hydrolyzing metallo-β-lactamase and its plasmid- and integron-borne gene from a *Pseudomonas aeruginosa* clinical isolate in France. *Antimicrob. Agents Chemother.* **44,** 891–897 (2000).

75. Berendonk, T. U. *et al.* Tackling antibiotic resistance: The environmental framework. *Nat. Rev. Microbiol.* **13,** 310–317 (2015).

76. Van Berkel, S. S. *et al.* Assay platform for clinically relevant metallo-β-lactamases. *J. Med. Chem.* **56,** 6945–6953 (2013).

77. Lassaux, P. *et al.* Biochemical and structural characterization of the subclass B1 metallo-β-lactamase VIM-4. *Antimicrob. Agents Chemother.* **55,** 1248–1255 (2011).

78. Xie, H. *et al.* Rapid point-of-care detection of the tuberculosis pathogen using a BlaC-specific fluorogenic probe. *Nat. Chem.* **4,** 802–9 (2012).

79. O'Callaghan, C. H., Morris, A., Kirby, S. M. & Shingler, A. H. Novel method for detection of β-lactamases by using a chromogenic cephalosporin substrate. *Antimicrob. Agents Chemother.* **1,** 283–288 (1972).

80. Vella, P. *et al.* The identification of new metallo-β-lactamase inhibitor leads from fragment-based screening. *Bioorganic Med. Chem. Lett.* **21,** 3282–3285 (2011).

81. Guo, Y. *et al.* A structural view of the antibiotic degradation enzyme NDM-1 from a superbug. *Protein Cell* **2,** 384–394 (2011).

82. Pettinati, I., Brem, J., Lee, S. Y., McHugh, P. J. & Schofield, C. J. The chemical biology of human metallo-β-lactamase fold proteins. *Trends Biochem. Sci.* **41,** 338–355 (2016).

83. Xue, M., Rabbani, N. & Thornalley, P. J. Glyoxalase in ageing. *Semin. Cell Dev. Biol.* **22,** 293–301 (2011).

84. Dominski, Z. Nucleases of the metallo-β-lactamase family and their role in DNA and RNA metabolism. *Crit. Rev. Biochem. Mol. Biol.* **42,** 67–93 (2007).

85. Sengerová, B. *et al.* Characterization of the human SNM1A and SNM1B/Apollo DNA repair exonucleases. *J. Biol. Chem.* **287,** 26254–26267 (2012).

86. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).

87. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

88. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68,** 352–367 (2012).

89. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

90. Carfi, A. *et al.* The 3-D structure of a zinc metallo-β-lactamase from *Bacillus cereus* reveals a new type of protein fold. *EMBO J.* **14,** 4914–21 (1995).

91. Brem, J. *et al.* Studying the active-site loop movement of the São Paolo metallo-β-lactamase-1. *Chem. Sci.* **6,** 956–963 (2015).

92. Merino, M. *et al.* Role of changes in the L3 loop of the active site in the evolution of enzymatic activity of VIM-type metallo-β-lactamases. *J. Antimicrob. Chemother.* **65,** 1950–1954 (2010).

93. Mojica, M. F. *et al.* Exploring the role of residue 228 in substrate and inhibitor recognition by VIM metallo-β-lactamases. *Biochemistry* **54,** 3183–3196 (2015).

94. Cornaglia, G., Giamarellou, H. & Rossolini, G. M. Metallo-β-lactamases: A last frontier for β-lactams? *Lancet Infect. Dis.* **11,** 381–393 (2011).

95. Edelstein, M. V *et al.* Spread of extensively resistant VIM-2-positive ST235 *Pseudomonas aeruginosa* in Belarus, Kazakhstan, and Russia: A longitudinal epidemiological and clinical study. *Lancet Infect. Dis.* **13,** 867–876 (2013).

# Chapter 7: Structural Basis for the Inhibition of VIM-2 by a Novel, *In Silico*-Designed Small Molecule Inhibitor

## 7.1 Introduction

### 7.1.1 Classical and reverse pharmacology

The processes involved in favoured paths to drug discovery have changed substantially over the years. Traditional medicines have been used for millennia, with records dating from as early as 2600 B.C. detailing products such as poppy juice, cedar oil, and myrrh, which have been used to treat a wide range of ailments.[1] Initial successes in the field of focussed drug discovery began in the mid-nineteenth century, progressing through the identification of bioactive compounds in natural remedies that were known to treat specific illnesses.[2] This process developed into the 'classical pharmacology' method, wherein libraries of natural products and compounds were screened against animal disease models in search of a desirable phenotypic change.[3] Knowledge of a specific target pathway on which the prospective drug might act is not required in classical pharmacology. Many widely used, clinically important drugs were identified using the classical pharmacology method, such as aspirin, morphine, and digoxin.[4–6]

As the field of organic, and biological chemistry matured, scientists began to develop a better understanding of the chemical structures of these bioactive molecules, along with the proteins and pathways on which they acted. A new approach to drug discovery began to come into favour from the 1950s onward, in which a target protein involved in a pathological process was first identified. Potential drugs were then screened against the purified protein in an attempt to modulate its activity in a process known as 'reverse

pharmacology'.[3] Identified compounds of interest would subsequently be tested in animal models in order to establish whether they induce a desired phenotypic response. This new approach allowed for the efficient screening of focussed libraries of compounds based on known bioactive molecules, and quickly resulted in the discoveries of blockbuster drugs such as beta blockers, statins and $H_2$ antagonists.[7–9]

### 7.1.2 Screening methods

#### High throughput screening

Currently the most widely used screening method employed in the pharmaceutical industry for early stage drug discovery is high throughput screening (HTS).[10] HTS is a brute force process in which large libraries of compounds are screened against cell lines (classical pharmacology), or a purified biological target (reverse pharmacology). Many different types of libraries can be used including small molecule, genomic, peptide, and protein libraries. Developments in automation mean hundreds of thousands of compounds can be screened in a single day.[10] HTS has the potential to identify novel bioactive molecules without prior knowledge of the target, however, the technique comes with an exceptionally low hit rate. Vast libraries of compounds are therefore required to conduct HTS, greatly restricting the use of this technique.[11]

#### Computer assisted drug design

Over the past decade computer assisted drug design (CADD) has emerged as a useful tool for improving the efficiency of screening.[11] A 'virtual screen' can be conducted *in silico* in order to predict which compounds will bind tightly to a target protein. Through a process of docking the optimal binding mode of each compound to the protein, and the strength of the protein–compound interactions, are predicted and compared.[12] Compounds

predicted to interact strongly with the target can then be prioritised, while those expected to bind poorly can be ignored, greatly reducing the numbers of compounds needed to be screened. In one project aiming to identify novel inhibitors of the enzyme protein tyrosine phosphatase-1B (PTP1B), both conventional HTS, and CADD were used, enabling comparison of success rates.[13] HTS using a library of 400 000 compounds resulted in the identification of 85 hits ($IC_{50}$ < 100 μM), therefore a hit rate of 0.021%.[13] When CADD was first employed against a virtual library of 235 000 compounds, and the top 365 scoring compounds were screened, 127 hits were identified, resulting in a hit rate of ~35% for compounds tested following CADD pre-screening.[13] CADD therefore has the potential to substantially reduce the material cost, instrumentation requirements and time needed to identify new potential drugs, without decreasing the hit rate.

An important feature of CADD that can assist in identifying potential lead compounds is *de novo* drug design (*i.e.* drug design prior to screening).[14] Using a model of a target protein, features of the active site such as cavity shape, locations of H-bonding sites, metal binding sites, electrostatic interaction sites and hydrophobic regions can be mapped.[15] Moieties and functional groups predicted to interact favourably within this site can then be placed. By linking various combinations of these fragments together in different ways large libraries can quickly be generated tailored to a specific target. A noteworthy early success of *de novo* drug design was the ACE inhibitor (L-)captopril.[16,17]

As CADD does not directly involve experimentation, the success of *in silico* screening and *de novo* drug design is highly dependent on the quality of the models used. High-resolution crystal structures of the protein of interest are desirable, even more so when that protein is in complex with a binding partner. While co-crystal structures of MBLs

complexed with either hydrolysed β-lactam antibiotics, or inhibitors were available (see **Chapter 6**), at the commencement of this study, no trials employing CADD to assist in identifying potential MBL inhibitors had been reported.[18,19]

## 7.2 Objectives

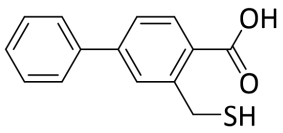
The ability to design and screen new inhibitors *in silico* is an attractive approach that had not previously been utilised for the identification of MBL inhibitors. The work detailed in this chapter contributes towards a collaborative effort to identify a new class of MBL inhibitor through CADD, involving quantifying the binding strength of potential hits to a selection of MBLs experimentally, and investigating the binding mode crystallographically.

The specific focus of the work conducted in this chapter was towards the co-crystallisation, and structural analysis of the B1 MBL VIM-2 with the novel, CADD-identified MBL inhibitor 4-phenyl-2-sulfonylmethylbenzoic acid (compound **8**). It was hoped that investigating such a complex would yield information on the structure-activity relationship of this new class of compound as well as inform future drug discovery and optimisation attempts.

*The findings in this chapter have been published (Cain, Brem, Zollman* et al. *2017). CADD,* in silico *screening, and MBL inhibitor synthesis were conducted by Dr Ricky Cain. VIM-2 protein was produced and purified by Dr Jürgen Brem.*

## 7.3 *De novo* ligand design

The *de novo* molecular design program SPROUT was used for compound library design and *in silico* screening (by *Dr Ricky Cain*).[15,20] A 1.3 Å X-ray crystal structure of NDM-1 in complex with hydrolysed ampicillin (PDB ID: 3Q6X) was selected as the initial model.[21] A 10 Å diameter sphere centred on the hydrolysed ampicillin was defined as the receptor site. Within the receptor site, four regions for targeting were selected: the water/hydroxide 'bridging' between the two zinc ions; the region within bonding distance of the K224 side-chain; the hydrophobic region proximal to W87; and the area close to the side-chain of Q119. Fragments capable of interacting with these regions for targeting were subsequently docked into the receptor site. Complete ligands formed by linking combinations of docked fragments were generated, resulting in the design of around 34 000 molecules. Binding affinities of each compound to NDM-1 were estimated and ranked. The 100 ligands predicted to bind tightest to NDM-1 were analysed for complexity by the ALLIGATOR module of SPROUT.[20] The top 10 ligands predicted to bind tightest to NDM-1, and deemed synthetically accessible by ALLIGATOR, were docked into NDM-1 using AutoDock.[22] Visual inspection of the docking poses resulted in the identification of 4-phenyl-2-sulfonylmethylbenzoic acid (compound **8**) as an attractive potential MBL inhibitor (see **Figure 7.1**).




**8**

***Figure 7.1****: Chemical structure of 4-phenyl-2-sulfonylmethylbenzoic acid (**8**).*

## 7.4 Metallo-β-lactamase inhibition by a *de novo*-designed ligand

Compound **8** was synthesised (by *Dr Ricky Cain*) and tested in a fluorescent reporter assay as described in **Section 6.4.1** against IMP-1, NDM-1 and VIM-2 (by *Dr Jürgen Brem*). $K_i$ values of 0.22 ± 0.04, 4.35 ± 0.04 and 0.17 ± 0.03 μM were recorded for **8** against IMP-1, NDM-1 and VIM-2, respectively, comparing favourably with other thiol-MBL inhibitors such as the captopril stereoisomers (**Table 7.1**).[23]

***Table 7.1****:* $K_i$ *(μM) values for* ***8*** *against three different MBLs. Equivalent measurements for the two tightest binding captopril stereoisomers are provided for context.*[23] $K_i$ *values and their 95% confidence intervals were calculated as described in* ***Materials and Methods 8.14*** *using nonlinear regression (GraphPad Prism).*[24] *The mean value of each of the trials is presented ± standard deviation from three separate experiments.*

| | IMP-1 | NDM-1 | VIM-2 |
|---|---|---|---|
| **8** | 0.22 ± 0.04 | 4.35 ± 0.04 | 0.17 ± 0.03 |
| **L-captopril** | 16.1 ± 0.9 | 41.0 ± 0.3 | 1.0 ± 0.2 |
| **D-captopril** | 5.0 ± 0.8 | 5.2 ± 0.4 | 0.016 ± 0.002 |

## 7.5 Structural analysis of a *de novo*-designed ligand binding to VIM-2

### 7.5.1 Data collection

As a clinically relevant, model B1 MBL, VIM-2 was chosen to be a protein binding partner for co-crystallisation trials with **8**. As in the previous VIM-2–captopril co-crystallisation experiments, VIM-2–**8** crystals were observed growing in conditions containing 0.2 M magnesium formate, 20% (w/v) PEG 3350, 1 mM TCEP, 213 μM VIM-2, 12.5 mM **8**, reaching a maximum size after incubation at RT for three days.[23] Crystals were cryoprotected using well solution diluted with 25% (v/v) glycerol before being flash cooled in liquid nitrogen. The VIM-2–**8** crystal complex X-ray data set was collected at 100 K, diffracting to 1.7 Å. The data were indexed, integrated, and scaled using HKL-3000.[25] The structures were solved by molecular replacement using Phaser with

PDB ID: 4BZ3 as a search model.[23,26] The structures were then refined using PHENIX and Coot until the $R_{work}$ and $R_{free}$ values no longer decreased.[27,28] Data collection and refinement statistics are provided in **Table 7.2**.

***Table 7.2****: Processing and refinement statistics for VIM-2 co-crystallised with compound* ***8****. FRE+ is an in-house X-ray generator equipped with a Saturn 944+ CCD detector.* [†]*ASU = asymmetric unit.* [‡]*RMSD = root mean square deviation.*

| | VIM-2–compound 8 complex |
|---|---|
| **X-ray source** | FRE+ |
| **Resolution (outer shell) (Å)** | 50.0-1.74 (1.81-1.74) |
| **Unit cell dimensions a, b, c (Å)** | 102.89, 79.54, 67.89 |
| **Unit cell dimensions α, β, γ (degrees)** | 90.0, 130.6, 90.0 |
| **Space group** | *C* 1 2 1 |
| **Protein molecules per ASU[†]** | 2 |
| **Completeness (outer shell) (%)** | 95.0 (85.1) |
| **$R_{merge}$ (outer shell)** | 7.9 (38.9) |
| **No. of unique reflections** | 40087 |
| **I/σ mean (outer shell)** | 14.2 (1.9) |
| **No. of non-H protein atoms per ASU[†]** | 3457 |
| **Average B factor** | 27.65 |
| **RMSD from ideal bond length (Å) [‡]** | 0.003 |
| **RMSD from ideal angles (degrees)** | 0.600 |
| **$R_{work}$ (%)** | 16.6 |
| **$R_{free}$ (%)** | 19.4 |

Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession code 5K48.

### 7.5.2 Structural analysis

As in the case of most, but not all thiol carboxy-MBL inhibitors co-crystallised with a B1 subfamily MBL, **8** bound to the Zn(II) ion binding site through displacement of the bridging water/hydroxide with its thiol. While previously described tightly binding thiol-MBL inhibitors containing carboxyl groups have appeared to favour hydrogen bonding to R228, **8** instead is positioned to form a hydrogen bond to D119 *via* an active site water (see **Figure 7.2 A**). With its carboxylate positioned pointing away from R228, **8** instead appears able to form π-stacking interactions with both W87 and Y67. Comparisons of the VIM-2–**8** co-crystal structure with the apo-di-Zn(II) VIM-2 structure described in **Chapter 6** (PDB ID: 4BZ3) indicates that binding to **8** results in the hydrophobic L3 loop being pulled into the active site by around 1.1 Å (see **Figure 7.2 B**). A weak hydrogen bond (3.6 Å) between a carboxylate oxygen and the side-chain amide of N233 also appears likely.

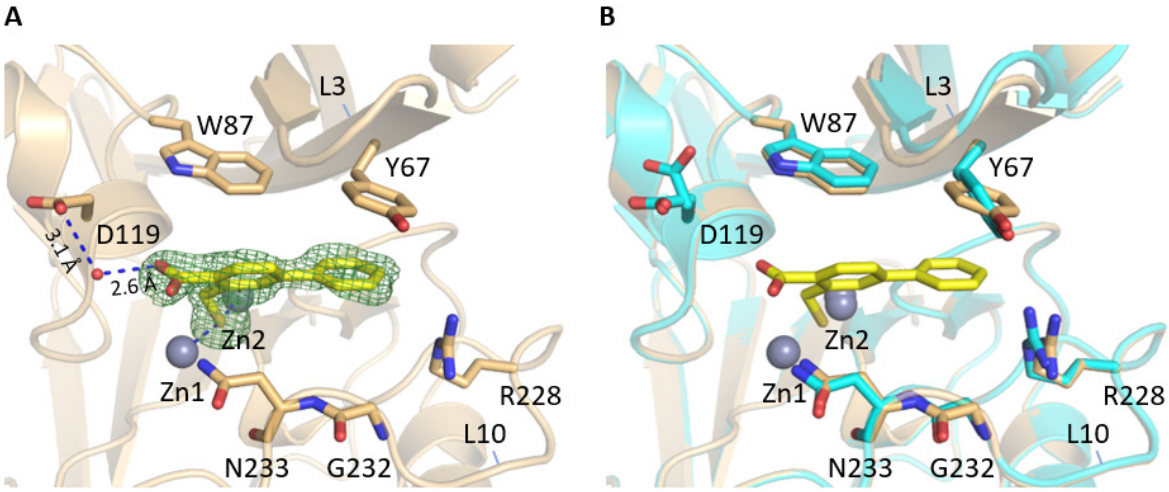



***Figure 7.2**: **A**- View of the zinc binding site from a crystal structure of VIM-2 complexed with compound **8**. Compound **8**-bound VIM-2 is shown in sand and compound **8** in yellow. Close contacts (within 3.5 Å) are indicated with a blue dotted line. |Fo-Fc| positive electron density is displayed (green mesh) contoured to 3 σ. Density was calculated from the final model, following omission of the ligand, against experimental diffraction data. **B**- Overlay of the compound **8**- and apo-Zn(II)-bound VIM-2 structures (sand and light blue, respectively).*

### 7.5.3 Structural optimisation

MBLs contain a conserved basic residue at the R228 position, thought to be important for binding the carboxylic acid/sulfate group common to β-lactam antibiotics.[29] As discussed in **Chapter 6** the ability of D-captopril to form an apparent hydrogen bond to the sidechain of K224/R228 appears to be a vital contribution to the higher binding strength of D-captopril relative to L-captopril against the MBLs tested.[23] An absence of a strong interaction between **8** and the conserved basic residue at the R228 position highlights a promising location for compound optimisation.[23] A group capable of forming a hydrogen bond, such as an alcohol, carbonyl, carboxylic acid or sulphonyl in the *meta*- or *para*-position of the monosubstituted benzene ring ($R^3$ and $R^4$, respectively) would likely be well placed to form one with the K224/R228 side-chain.

In the solved structure, the carboxylic group of **8** appears able to form a hydrogen bond with D119 *via* an active site water (see **Figure 7.2 A**). Addition of a one carbon spacer between the carboxylic acid of **8** and the rest of the molecule may result in the displacement of this bridging water, allowing for direct hydrogen bonding between the inhibitor and D119.

The carboxylic acid of **8** is only positioned to form a weak hydrogen bond (3.6 Å) to the side-chain of N233, a key binding partner for the tightest binding captopril stereoisomers, located on the L10 loop. Placement of a hydrogen bond acceptor at any of the unsubstituted positions of the trisubstituted phenyl ring appear unlikely to be able to reach N233 without disrupting the favourable π-stacking interactions between **8** and W87. The *ortho* ($R^2$)-position of the monosubstituted phenyl ring, however, is a promising location for a hydrogen bonding group, either bound directly to the ring, or

separated by one carbon. A group capable of forming a hydrogen bond at this position may be able to bond to either the side-chain or main-chain of N233.

Suggestions for the optimisation of compound **8** for binding to MBLs, as derived from the above-described VIM-2–**8** complex analysis, are presented in **Figure 7.3**.

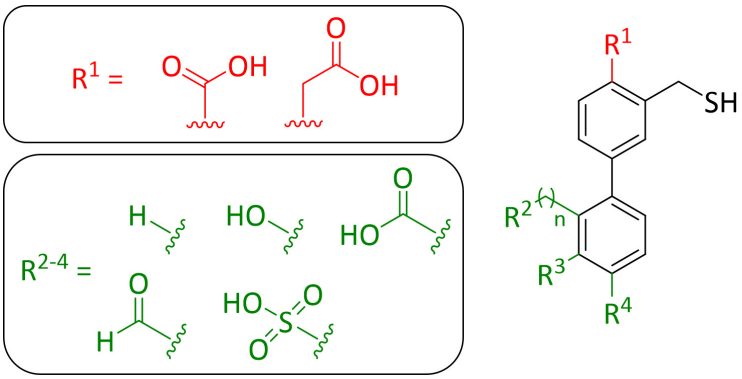



***Figure 7.3**: Potential adaptations to compound **8** ($R^1 = CO_2H$, $R^{2-4} = H$, $n = 0$) for increased binding to VIM-2 and other MBLs where n = 0 or 1.*

## 7.6 Discussion

For the first time a *de novo* method has been used to design a novel, potent inhibitor of the MBLs. Co-crystal structural analysis of compound **8** complexed with VIM-2 reveals a new binding mode for MBL inhibitors. As is typical with thiol-based MBL inhibitors, metal binding occurred through the thiol of **8**. Strong π-stacking interactions were identified as important contributors to the binding strength of the compound to the protein; however, several key binding opportunities were shown to be missed. Suggestions have been made for improving the potency of an inhibitor based on **8** against VIM-2, and likely other MBL enzymes.

The potency of binding to a target protein is not the only feature important in a drug candidate. The ability to inhibit a range of MBLs would be desirable, but selectivity over other proteins including human MBL-fold enzymes is also an important consideration. Additional functionalisation of the molecule may serve to preclude the compound from binding to non-target enzymes, improving the selectivity of the drug. A more active compound could also be prescribed at lower concentrations, reducing side-effects.

## 7.7 Part II conclusions

The rise of antibiotic resistance has been characterised as one of the greatest imminent threats posed to public health. An important factor in the progress of antibiotic resistance is the wide range of MBL-coding pathogens, capable of deactivating the most widely prescribed (β-lactam) class of antibiotic, that continue to proliferate.[30] At present there are no clinically prescribed drugs designed to inhibit any MBL. The work presented in **Part II** of this thesis has the potential to help towards this goal through the structural

analysis of a range of potent small molecule inhibitors of MBL enzymes in complex with several MBL targets.

CADD has been introduced as a tool with great promise for assisting in the search for an MBL inhibitor drug. Drug discovery methods such as CADD rely on high quality models, such as high-resolution protein–binding partner co-crystal structures, for success. Nine such complexes are presented in **Part II**, revealing a hierarchy of preferred binding interactions for captopril stereoisomers (thiol–di-Zn(II) > carboxylic acid–K224/R228 > ketone–N223), and a novel binding mode for compound **8**. Both the "drug-like" D-captopril and the potent inhibitor **8** represent promising starting points for drug optimisation. Any attempts to develop inhibitors based on captopril or compound **8** would be greatly assisted by these newly revealed structures.

Should a clinically viable MBL inhibitor be discovered it may be possible to co-dose it with β-lactam antibiotics, shielding the drug from hydrolysis by the MBL enzymes. By this method this vital class of antibiotic may be re-potentiated, helping to grant society a reprieve from the predicted "antibiotic apocalypse".[31]

## 7.8 References


1. Cragg, G. M. & Newman, D. J. Biodiversity: A continuing source of novel drug leads. *Pure Appl. Chem.* **77,** 7–24 (2005).
2. Dias, D. A., Urban, S. & Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* **2,** 303–336 (2012).
3. Takenaka, T. Classical *vs*. reverse pharmacology in drug discovery. *BJU Int.* **88,** 7–10 (2008).
4. Sneader, W. The discovery of aspirin: A reappraisal. *BMJ* **259,** 614–617 (1997).
5. Stefano, G. B., Ptácek, R., Kuzelová, H. & Kream, R. M. Endogenous morphine: Up-to-date review 2011. *Folia Biologica (Czech Republic)* **58,** 49–56 (2012).
6. Withering, W. *An account of the foxglove, and some of its medical uses with practical remarks on dropsy, and other diseases*. (G. G. J. and J. Robinson, Paternoster-Row, London, 1785).
7. Freemantle, N., Cleland, J., Young, P., Mason, J. & Harrison, J. β Blockade after myocardial infarction: Systematic review and meta regression analysis. *BMJ* **318,** 1730–1737 (1999).
8. Endo, A., Kuroda, M. & Tsujita, Y. ML-236A, ML-236B, and ML-236C, new inhibitors of cholesterogensis produced by *Penicillium citrinum*. *J. Antibiot. (Tokyo).* **29,** 1346–1348 (1976).
9. Brogden, R. N., Heel, R. C., Speight, T. M. & Avery, G. S. Cimetidine: A review of its pharmacological properties and therapeutic efficacy in peptic ulcer disease. *Drugs* **15,** 93–131 (1978).
10. Szymański, P., Markowicz, M. & Mikiciuk-Olasik, E. Adaptation of high-throughput screening in drug discovery-toxicological screening tests. *Int. J. Mol. Sci.* **13,** 427–452 (2012).
11. Sliwoski, G., Kothiwale, S., Meiler, J. & Lowe, E. W. Computational methods in drug discovery. *Pharmacol. Rev.* **66,** 334–395 (2013).
12. Lengauer, T. & Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* **6,** 402–406 (1996).
13. Doman, T. N. *et al.* Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.* **45,** 2213–2221 (2002).
14. Hartenfeller, M. & Schneider, G. *De novo* drug design. In *Chemoinformatics and computational chemical biology* **672,** 299–323 (Humana Press, Totowa, NJ, 2011).
15. Gillet, V., Johnson, A. P., Mata, P., Sike, S. & Williams, P. SPROUT: A program for structure generation. *J. Comput. Aided. Mol. Des.* **7,** 127–153 (1993).
16. Cushman, D. W., Cheung, H. S., Sabo, E. F. & Ondetti, A. Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and

mercaptoalkanoyl amino acids. *Biochemistry* **16,** 5484–5491 (1977).

17. Ondetti, M. A., Rubin, B. & Cushman, D. W. Design of specific inhibitors of angiotensin-converting enzyme: New class of orally active antihypertensive agents. *Science* **196,** 441–4 (1977).

18. King, D. T., Worrall, L. J., Gruninger, R. & Strynadka, N. C. J. New Delhi metallo-β-lactamase: Structural insights into β-lactam recognition and inhibition. *J. Am. Chem. Soc.* **134,** 11362–11365 (2012).

19. Cain, R. *et al. In silico* fragment-based design identifies subfamily B1 metallo-β-lactamase inhibitors. *J. Med. Chem.* **61,** 1255–1260 (2018).

20. Gillet, V. J. *et al.* SPROUT: Recent developments in the *de novo* design of molecules. *J. Chem. Inf. Comput. Sci.* **34,** 207–217 (1994).

21. Zhang, H. & Hao, Q. Crystal structure of NDM-1 reveals a common β-lactam hydrolysis mechanism. *FASEB J.* **25,** 2574–2582 (2011).

22. Morris, G. M. *et al.* Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30,** 2785–2791 (2009).

23. Brem, J. *et al.* Structural basis of metallo-β-lactamase inhibition by captopril stereoisomers. *Antimicrob. Agents Chemother.* **60,** 142–150 (2016).

24. GraphPad Software. GraphPad Prism version 5.04 for Windows (2010).

25. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).

26. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

27. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68,** 352–367 (2012).

28. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

29. Mojica, M. F. *et al.* Exploring the role of residue 228 in substrate and inhibitor recognition by VIM metallo-β-lactamases. *Biochemistry* **54,** 3183–3196 (2015).

30. Johnson, A. P. & Woodford, N. Global spread of antibiotic resistance: The example of New Delhi metallo-β-lactamase (NDM)-mediated carbapenem resistance. *J. Med. Microbiol.* **62,** 499–513 (2013).

31. Gallagher, J. BBC News - Antibiotic 'apocalypse' warning. *19 November 2015* (2013). Available at: http://www.bbc.co.uk/news/health-21702647. (Accessed: 20th January 2018).

# Chapter 8: Materials and Methods

## 8.1 General methods

Reagents were purchased from the Sigma-Aldrich Corporation unless otherwise stated.

### Sterile practice

All methods were conducted using standard sterile practices. Before use, equipment and media were sterilised by autoclaving at 121 °C for 20 minutes in a TouchClave® II autoclave (LTE Scientific). Milli-Q® (MQ) water was purified using a Millipore Elix® 10 system followed by filtration through a 0.22 μm filter. All buffers were filtered through a 0.22 μm filter prior to use in experiments involving protein or DNA.

### Water baths

Incubations at 37 °C were conducted in a Grant Y14 waterbath; incubations at 42 °C were conducted in a Grant W14 waterbath, unless otherwise stated.

### pH determination

pH measurements were made using a HI 9321 microprocessor pH meter (Hanna Instruments). Prior to use, the probe was calibrated against standard calibration solutions (Fisher Scientific). Adjustments to solution pH were made through addition of solutions of either NaOH or HCl, with stirring.

### Centrifuges

For low volumes of liquid (<1.5 mL) an AccuSpin™ micro benchtop centrifuge (Fisher Scientific) was used; for volumes between 1.5 and 50 mL an Allegra X-30R benchtop centrifuge (Beckman Coulter™) was used; for volumes of liquid >50 mL an Avanti J25 centrifuge (Beckman Coulter™) was used.

### Fast protein liquid chromatography systems

All FPLC purification steps were conducted using one of the following systems; an ÄKTA FPLC (GE Healthcare) with a Frac-920 fraction collector (GE Healthcare), an ÄKTApurifier FPLC (GE Healthcare) with a Frac-940 fraction collector (GE Healthcare), or an ÄKTA pure FPLC (GE Healthcare) with an F9-C fraction collector (GE Healthcare). FPLC systems were controlled using Unicorn version 5.1 (ÄKTA or ÄKTApurifier systems) or Unicorn version 6.4.1 (ÄKTA pure system). All columns, loops and column media were purchased from GE Healthcare. UV monitoring was conducted at $\lambda$ = 220 nm. Buffers for use in SEC were degassed before use.

### DNA concentration determination

DNA concentration was assessed using a NanoDrop ND-1000 UV/Vis spectrophotometer (NanoDrop Technologies Inc.) to measure the absorbance of a 2 µL sample at $\lambda$ = 260 nm. DNA purity was assessed by comparing the $A_{260}/A_{280}$ ratio.

### Protein concentration determination

Protein concentration was assessed using a NanoDrop ND-1000 UV/Vis spectrophotometer (NanoDrop Technologies Inc.) to measure the absorbance of a 2 µL sample at $\lambda$ = 280 nm. The absorbance spectra were normalised against the sample buffer, and the extinction coefficient of the protein was calculated using the ExPASy ProtParam tool (https://web.expasy.org/protparam/).

### Lyophilisation

Aqueous solutions were lyophilised by flash freezing in liquid $N_2$ followed by lyophilisation using a CoolSafe chiller unit (LaboGene) attached to a RC6 vacuum pump (VACUUBRAND).

#### Gel imaging

Gels were imaged using a Gel Doc™ XR+ (Bio-Rad) and bands were quantified using the Image Lab version 5.2.1 (Bio-Rad).[1]

## 8.2 Gel electrophoresis

### 8.2.1 SDS-PAGE

*Separating gel buffer*

| Composition | Concentration |
|---|---|
| **Tris-HCl† pH 8.8** | 1.5 M |
| **SDS** | 0.2% (w/v) |

†2-Amino-2-(hydroxymethyl)-1,3-propanediol.

*Stacking gel buffer*

| Composition | Concentration |
|---|---|
| **Tris-HCl pH 6.8** | 0.1 M |
| **SDS** | 0.2% (w/v) |

*SDS-PAGE separating gel*

| | Volume (for 2 gels) | |
|---|---|---|
| **Composition** | **12.5%** | **16%** |
| **Separating gel buffer** | 6 mL | 5 mL |
| **30% (w/v) Acrylamide** | 5 mL | 5 mL |
| **$H_2O$** | 890 µL | - |
| **10% (w/v) APS†** | 100 µL | 100 µL |
| **TEMED‡** | 10 µL | 10 µL |

†APS = ammonium persulfate; ‡TEMED = tetramethylethylenediamine.

*SDS-PAGE stacking gel*

| Composition | Volume (for 2 gels) |
|---|---|
| **Stacking gel buffer** | 2.5 mL |
| **30% (w/v) Acrylamide** | 500 µL |
| **$H_2O$** | 1.9 mL |
| **10% (w/v) APS** | 100 µL |
| **TEMED** | 15 µL |

*SDS-PAGE running buffer*

| Composition | Concentration |
|---|---|
| **Tris-HCl pH 8.3** | 25 mM |
| **Glycine** | 200 mM |
| **SDS** | 0.1% (w/v) |

*SDS-Page sample loading buffer*

| Composition | Concentration |
|---|---|
| **SDS** | 2% (w/v) |
| **Glycerol** | 1.6 M |
| **BME** | 700 mM |
| **Tris-HCl pH 6.8** | 50 mM |
| **Bromophenol blue** | 3 mM |

*SDS-PAGE stain/de-stain*

| Composition | |
|---|---|
| **Acetic acid** | 10% (w/v) |
| **Methanol** | 30% (w/v) |
| **Coomassie® brilliant blue R-250** | 0.25% (w/v, stain only) |

**SDS-PAGE gel casting**

SDS-PAGE gels were prepared by combining and thoroughly mixing the SDS-PAGE separating gel reagents (APS and TEMED last), before quick addition of the solution to the gel cast. Once the gel had solidified (~10 min) SDS-PAGE stacking gel was prepared in a similar manner and dispensed on top of the SDS-PAGE separating gel. Well combs were added, and the gel was left to solidify (~10 min). SDS-PAGE gel acrylamide % was selected based on the size of the proteins to be separated.

**SDS-PAGE method**

Samples were combined with SDS-PAGE sample loading buffer in a 1:1 ratio and heated for 5 min at 100 °C. Samples were loaded into the gel, with one lane reserved for PageRuler™ pre-stained protein ladder (Thermo Fisher Scientific). SDS-PAGE gels were

run (180 V, RT, 40-60 min) in a Bio-Rad Mini-PROTEAN II apparatus. Gels were visualised by incubation in SDS-PAGE stain solution (20 min), then in SDS-PAGE de-stain solution overnight.

### 8.2.2 Western blot procedure

Protein samples were separated by SDS-PAGE, then transferred onto blotting membrane using an iBlot western blot transfer system (Thermo Fisher Scientific) following the manufacturer's instructions. Protein transfer was confirmed with Ponceau S dye. The membrane was destined by washing with MQ $H_2O$, then incubated in blocking buffer containing 5% (v/v) FBS in tris-buffered saline-tween (TBS-T) solution (25 mM tris-HCl pH 8.8, 137 mM NaCl, 2.68 mM KCl, 0.1% (v/v) Tween 20) overnight at 4 °C. The blocked membranes were charged with 1° antibodies diluted as per manufacturer's instructions in blocking solution for 1 hour (RT). Membranes were washed in TBS-T solution (3 × 10 min), then probed by 2° antibodies at 1:15000 dilution in blocking buffer (1 hr, RT). The 2° antibody solution was removed with washing in TBS-T solution (3 × 10 min). Visualisation of the bands was carried out by one of two methods, depending on the antibody system used:

- Where a 2° antibody coupled to horseradish peroxidase had been used, bands were visualised using ECL™ reagents 1 and 2. ECL™ reagents 1 and 2 were added to the membrane (0.5 mL, each) and incubated at RT for 1 minute. A sheet of Amersham™ Hyperfilm ECL film (GE Healthcare) was exposed to the charged membranes, then the film was developed using a Compack X4 (Xograph) film developer.
- Where fluorescent 2° antibodies were used, bands were visualised directly using a ChemiDoc imaging system (Bio-Rad).

All antibodies used were purchased from AbCam.

### 8.2.3 Dot blot

Dot blot experiments were conducted by application of the protein reaction sample directly onto the blotting membrane (Hybond ECL membrane, 0.2 μM pore size, GE Healthcare). The membrane was subsequently left to dry, before being treated with antibodies and visualised as described for western blot experiments.

### 8.2.4 Native PAGE

Nucleosome folding was validated by native PAGE a using 6% TBE gel (Thermo Fisher Scientific). Nucleosome samples were loaded (250 ng) with a 100 bp DNA ladder for reference (NEB). Gels were run in TBE buffer (Thermo Fisher Scientific) in a XCell4 SureLock® Midi-Cell (Thermo Fisher Scientific) (150 V, 4 °C, 90 min). Bands containing DNA were visualised by incubation in TBA buffer containing 0.01% (v/v) SYBR™ Gold (Thermo Fisher Scientific) in the dark (RT, 20 min) before imaging using a ChemiDoc imaging system (Bio-Rad).

## 8.3 Peptide synthesis

Proteinogenic amino acids were purchased from Novabiochem, CSBio, Alfa Aesar, or AGTC bioproducts, and Fmoc-protected KAc/KAcOH residues were synthesised by *Roman Belle*.

### 8.3.1 Solid phase peptide synthesis

SPPS was carried out on a 50 μM scale using a Liberty Blue™ automated microwave peptide synthesiser (CEM). The peptide was synthesised from the *C*- to the *N*-terminus on a rink amide 4-methylbenzhydrylamine resin (100-200 mesh) through sequential reaction of the propagating peptide with $N^{\alpha}$-Fmoc-protected amino acids. Amino acid

sidechains were protected, where appropriate, with acid labile protecting groups. Coupling was achieved through incubation of the resin-bound peptide in the presence of an $N^{\alpha}$-Fmoc-protected amino acid (0.2 M) in DMF, DIC (TCI Europe), and OxymaPure® (Merck).

Fmoc deprotection was induced through incubation of the peptide in 20% (v/v) piperidine in DMF (AGTC Bioproducts). The coupling and deprotection steps were both microwave assisted using the Liberty Blue™ default settings. The resin was washed with DMF before repetition of the coupling and deprotection processes with new $N^{\alpha}$-Fmoc-protected amino acids as required. Following the final Fmoc deprotection step, DMF and all insoluble reagents were removed from the resin by washing with diethyl ether, before drying under vacuum. The peptide was cleaved from the resin by incubation in 92.5% (v/v) trifluoroacetic acid (TFA), 2.5% (v/v) triisopropylsilane solution (RT, 3 hr). This cleavage step also served to remove all the sidechain protecting groups from the peptide residues. The peptide was precipitated with addition of diethyl ether and incubation at -20 °C (24 hr). The peptide was pelleted by centrifugation, and the supernatant was discarded. The peptide was then dissolved in $H_2O$, flash cooled, and lyophilised, yielding a brown solid.

### 8.3.2 Reverse phase high performance liquid chromatography

Crude peptide was dissolved in 2 mL $H_2O$, with addition of small volumes of acetonitrile when the peptide solubility was low. The peptide solution was filtered through a 0.45 µM filter, then loaded onto a Grace Vydac 218TP 250 mm × 22 mm C18 chromatography column fitted to a Jasco HPLC system (PU-2086 Plus pumps, MX-2080-32 Dynamic Mixer and DG-2080-54 Degasser). Separation of the peptide from impurities was performed using a 5-30% (v/v) acetonitrile gradient over 15 minutes. The mobile phase

also contained 0.1% (v/v) TFA. Fractions containing the pure, desired peptide, as determined by MALDI-TOF MS, were pooled, lyophilised and stored at -20 °C.

## 8.4 DNA techniques

### 8.4.1 DNA sequencing

Plasmid sequencing was carried out by Source BioScience. The plasmids coding for all proteins produced in this thesis were sequenced using T7F and T7R primers.

### 8.4.2 Plasmid purification

Plasmid DNA was purified from a 5 mL overnight cell culture using a GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) following the manufacturer's protocol. DNA was eluted with MQ $H_2O$ (50 µL) and the DNA concentration was assessed by nanodrop. DNA was diluted to 100 ng/µL and stored at -20 °C.

### 8.4.3 Site directed mutagenesis

The oligonucleotides that were used in SDM reactions are provided in the table below. Differences from the native DNA sequence are presented in bold and underlined.

*Primers used for SDM*

| Construct | Direction | Sequence |
|---|---|---|
| **H3 K9C** | Forward | GACCGCCCGT**TGC**TCCACCGGAGG |
| **H3 K9C** | Reverse | CCTCCGGTGGA**GCA**ACGGGCGGTC |
| **H3 K18C** | Forward | GGGAAGGCTCCCCGC**TGC**CAGCTGGCCACCAAG |
| **H3 K18C** | Reverse | CTTGGTGGCCAGCTG**GCA**GCGGGGAGCCTTCCC |
| **H3 K79C** | Forward | GGGAGATCGCTCAGGACTTC**TGC**ACCGACCTGC |
| **H3 K79C** | Reverse | GCAGGTCGGT**GCA**GAAGTCCTGAGCGATCTCCC |

The SDM reaction mixture was prepared through sequential addition of the reagents listed in the following table. The solution was mixed by pipetting before, and after, addition of the Q5 enzyme. 5× Q5 buffer and Q5 polymerase were purchased from NEB.

*SDM reaction components*

| Composition | Concentration | Volume (µL) |
|---|---|---|
| **MQ $H_2O$** | | 33.5 |
| **Q5 buffer** | 5× | 10 |
| **Template DNA** | 20-100 ng/µL | 1 |
| **Reverse primer** | 10 µM | 1.25 |
| **Reverse primer** | 10 µM | 1.25 |
| **Deoxyribonucleotide (dNTP)** | 10 mM | 3 |
| **Q5 polymerase** | 2000 units/mL | 1 |

The SDM mixture was placed in a thermocycler programmed with the following settings:

| Step | | Number of cycles |
|---|---|---|
| **i** | 98 °C, 10 s | 1 |
| **ii** | 98 °C, 10 s | 25 |
| | 50-72 °C*, 10 s | |
| | 72 °C, 30 s/kb | |
| **iii** | 72 °C, 10 min | 1 |

*Temperature dependent on the melting point of the primer.

Parental DNA was digested with addition of Dpn1 enzyme to the mixture (1 µL), followed by incubation at 37 °C for 30 minutes. The mutated plasmid was then transformed into XL10 Gold competent cells for growth, miniprep and sequencing.

### 8.4.4 145 bp Widom 601 DNA sequence

```
ATCAGAATCC CGGTGCCGAG GCCGCTCAAT TGGTCGTAGA CAGCTCTAGC

ACCGCTTAAA CGCACGTACG CGCTGTCCCC CGCGTTTTAA CCGCCAAGGG

GATTACTCCC TAGTCTCCAG GCACGTGTCA GATATATACA TCGAT
```

### 8.4.5 Large scale DNA production buffers

*TE 10/50 buffer*

| Composition | Concentration |
|---|---|
| **Tris-HCl pH 8.0** | 10 mM |
| **EDTA-NaOH† pH 8.0** | 50 mM |

†Ethylenediaminetetraacetic acid-NaOH.

*TE 10/0.1 buffer*

| Composition | Concentration |
|---|---|
| **Tris-HCl pH 8.0** | 10 mM |
| **EDTA-NaOH pH 8.0** | 0.1 mM |

*ALS I buffer*

| Composition | Concentration |
|---|---|
| **Glucose** | 50 mM |
| **Tris-HCl pH 8.0** | 25 mM |
| **EDTA-NaOH pH 8.0** | 10 mM |

*ALS II buffer*

| Composition | Concentration |
|---|---|
| **Sodium hydroxide** | 200 mM |
| **SDS** | 1% (w/v) |

*ALS III buffer*

| Composition | Concentration |
|---|---|
| **Sodium acetate** | 4 M |
| **Acetic acid** | 2 M |

Experimental details for the large-scale production of Widom 601 DNA are provided in **Section 4.3.5**.

## 8.5 Protein production

### 8.5.1 Materials

#### Incubators

Small scale incubations (100 mL or less) were carried out in a Series 25 incubator (New Brunswick). Large scale incubations were carried out in a New Brunswick Scientific-Innova 44 incubator.

#### Growth media

*2 × Tryptone/yeast extract (2-TY) media*

| Composition | Amount (g/L) |
|---|---|
| **Bacto-tryptone** | 16 |
| **Yeast extract** | 10 |
| **NaCl** | 5 |

2-TY media was prepared and sterilised by *Herminio Manso*.

#### Antibiotics

Antibiotics were used at a final concentration of 100 µg/mL of ampicillin, 30 µg/mL of kanamycin, or 35 µg/mL of chloramphenicol.

#### Protein plasmid sources

Plasmids coding for H2A, H2B, H3 and H4 (*Xenopus laevis*) were provided in pET3d vectors by *Dr Lukas Lercher*. *H3 K4C*, *H3 K9C* and *H3 K18C* variants were obtained through SDM to an *H3 C110A* (*Xenopus laevis*) variant. A $BRD4_{44\text{-}168}$ (*Homo sapiens*) construct coding for BRD4(1) in a pNIC28-Bsa4 vector was provided by *Dr Angelina Serkis*.

#### Optical density measurements

Optical density $(OD)_{600}$ values were measured using a Novaspec® II spectrophotometer (Pharmacia) against a reference sample of sterile growth media.

### 8.5.2 Transformations

Competent cells were thawed on ice, before addition of plasmid DNA (1 µL, 100 ng/µL or equivalent). The plasmid and cells were incubated together on ice for 20 minutes before undergoing 'heat shock'. The sample was quickly heated to 42 °C (45 s), then cooled again on ice for 2 minutes. 2-TY media (250 µL) was added to the mixture before incubation at

37 °C with shaking (1 hr). The transformed cells were spread onto an agar plate infused with the appropriate antibiotic. Plates were incubated overnight at 37 °C resulting in the formation of colonies when plasmid transformation had been successful.

### 8.5.3 Starter cultures

Starter cultures were prepared by inoculating a 100 mL flask of 2-TY media with a single colony from an agar plate or glycerol stock. The media was incubated at 37 °C overnight with shaking at 250 rpm.

### 8.5.4 Glycerol stocks

Starter cultures were prepared and grown overnight with shaking at 37 °C. A 750 µL sample of bacterial culture was added to 250 µL sterile glycerol. The sample was mixed, then flash frozen in liquid $N_2$. The glycerol stock was then stored at -80 °C until required.

### 8.5.5 Competent cell production

Cells were streaked onto an antibiotic-free agar plate and grown overnight at 37 °C. A flask of 2-TY media (100 mL) was inoculated with a single colony of cells before incubating overnight with shaking (37 °C, 250 rpm). 4 × 200 µL of the starter culture was used to inoculate 4 flasks of sterile 2-TY media (100 mL), which were incubated at 37 °C with shaking (250 rpm) until they reached an $OD_{600}$ of 0.6-0.8. The cells were pelleted by centrifugation (750*g*, 4 °C, 5 min) before gentle resuspension in 100 mM $CaCl_2$, 20% (v/v) glycerol solution (8 × 25 mL). Cells were again pelleted by centrifugation (750*g*, 4 °C, 5 min), then gently resuspended in 100 mM $CaCl_2$, 20% (v/v) glycerol solution (8 × 1 mL). Cells were aliquoted, and flash cooled in liquid $N_2$, then stored at -80 °C.

### 8.5.6 Large-scale histone growths

Plasmids coding for the desired histone were transformed into BL21 (DE3) pLysS Rosetta 2 (H2B) cells, or BL21 (DE3) pLysS cells (all other histones). Starter cultures of

the transformed cells were used to inoculate 3 × 2 L flasks containing 600 mL 2-TY media and ampicillin (6 mL starter culture/flask). The flasks were incubated at 37 °C, 180 rpm, until the cells reached an $OD_{600}$ of 0.4-0.6. Histone expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) solution at a final concentration of 1 mM. Cells were incubated at 37 °C, 180 rpm for 3 hours, then harvested by centrifugation (10000*g*, 4 °C, 10 min). The cell pellet was resuspended in histone wash buffer (35 mL) (see **Section 8.6.1** for buffer composition) and stored until required at -80 °C.

### 8.5.7 Large-scale BRD4 growth

Plasmids coding for BRD4 were transformed into BL21 (DE3) pLysS cells. Starter cultures of the transformed cells were grown and used to inoculate 12 × 2 L flasks containing 600 mL 2-TY media and kanamycin (6 mL starter culture/flask). The flasks were incubated at 37 °C, 180 rpm, until the cells reached an $OD_{600}$ of 0.8-1.0, at which point the temperature was lowered to 18 °C and IPTG was added at a final concentration of 0.25 mM. Cells were incubated at 18 °C, 180 rpm overnight, then harvested by centrifugation (10000*g*, 4 °C, 10 min) and stored until required at -80 °C.

## 8.6 Purification of recombinant proteins

### 8.6.1 Histone purification

Histone purification was carried out following a modified version of the method described by Dyer *et al.*[2]

**Buffers**

*SAU-1000 buffer*

| Composition | Concentration |
|---|---|
| **Sodium acetate-acetic acid pH 5.2** | 20 mM |
| **Urea** | 7 M |
| **NaCl** | 1 M |
| **EDTA-NaOH pH 8.0** | 1 mM |
| **BME** | 5 mM |

*Histone wash buffer*

| Composition | Histone wash buffer | Histone wash + Triton buffer (HWTB) |
|---|---|---|
| **Tris-HCl pH 7.5** | 50 mM | 50 mM |
| **NaCl** | 100 mM | 100 mM |
| **EDTA-NaOH pH 8.0** | 1 mM | 1 mM |
| **Triton-x-100** | - | 1% (v/v) |

*Unfolding buffer*

| Unfolding Buffer | Concentration |
|---|---|
| **Guanidinium chloride** | 6 M |
| **Tris-HCl pH 7.5** | 20 mM |
| **DTT†** | 5 mM |

†Dithiothreitol.

Histones were purified from exclusion bodies. The resuspended cell pellet in histone wash buffer was thawed at RT and lysed by sonication (10 s on/off for a total sonication time of 2 min) with cooling on ice. The insoluble fraction was pelleted by centrifugation (50000*g*, 20 min, 4 °C), and the supernatant was discarded. The pellet was washed by resuspending in HWTB (40 mL) before pelleting by centrifugation (50000*g*, 20 min, 4 °C) and discarding the supernatant. This wash step was repeated once using HWTB, then a second time using histone wash buffer. The pellet was resuspended in DMSO (1 mL) and incubated at RT with shaking (1 hr). Unfolding buffer (7 mL) was added to the mixture followed by incubation (30 min, RT). The cell debris was pelleted by

centrifugation (50000*g*, 20 min, 4 °C). The supernatant was extracted and further clarified by centrifugation for a second time (50000*g*, 10 min, 4 °C) prior to being passed through a 0.45 µM filter. The histone solution was passed through a 300 mL Superdex 200 SEC column pre-equilibrated with SAU-1000 buffer, and fractions were collected. Fractions containing protein were analysed by SDS-PAGE, and fractions were combined based on purity. Histone solution (10-30 mL) was transferred into 6-8 kDa dialysis tubing, which was in turn placed into 2 mM BME solution (4 L) with stirring. Dialysis was allowed to proceed for at least 24 hours at 4 °C, with two changes of 2 mM BME solution spaced at least 4 hours apart. The dialysed histone solution was flash cooled in liquid $N_2$ and lyophilised, before storing at -80 °C.

### 8.6.2 BRD4 purification

BRD4 purification was carried out following the method described by Filippakopoulos *et al.*[3]

*BRD4 purification buffers*

| Composition | Charge | Binding | Wash | Elution | Gel filtration |
|---|---|---|---|---|---|
| **HEPES-NaOH pH 7.6** | - | 50 mM | 50 mM | 50 mM | 50 mM |
| **NaCl** | - | 0.5 M | 0.5 M | 0.5 M | 0.5 M |
| **Imidazole-HCl pH 8.0** | - | 5 mM | 30 mM | 500 mM | - |
| **$NiSO_4$** | 100 mM | - | - | - | - |

#### Method

The cell pellet was dissolved in a 5-fold excess (w/v) of binding buffer and lysed by sonication (10 s on/off for 10 min total sonication time). The insoluble lysate fraction was removed by centrifugation (50000*g*, 4 °C, 20 min) followed by passage of the lysate through a 0.45 µm filter. A 3-6 mL HisTrap™ chromatography column (GE Healthcare) was loaded with charge buffer (4 CV), then equilibrated with binding buffer (5 CV)

before addition of the clarified lysate. The column was washed with wash buffer (20 CV) and eluted with elution buffer (4 CV). Fractions were analysed by SDS-PAGE, and the fractions containing the highest concentration of pure BRD4 were combined. The combined fractions were concentrated and loaded onto a 150 mL Superdex 75 SEC column pre-equilibrated with gel filtration buffer. Following gel filtration, the fractions containing the highest concentration of pure BRD4 (determined by SDS-PAGE) were combined, concentrated, flash cooled in liquid $N_2$, and stored at -80 °C until required. Protein identity was validated by LC-MS.

## 8.7 Chemical incorporation of post-translational modification mimics

Synthetic incorporation of thioether acetyl lysine analogues was conducted essentially as described by Li *et al.* and methylated lysine analogues were prepared as described by Simon *et al.*[4,5]

### Buffers

*sKAc reaction buffer*

| Composition | Concentration |
|---|---|
| **Guanidinium chloride** | 6 M |
| **Sodium acetate-acetic acid pH 4.0** | 200 mM |
| **DL-Methionine** | 1 mM |

*Alkylation buffer*

| Composition | Concentration |
|---|---|
| **Guanidinium chloride** | 4 M |
| **HEPES-NaOH pH 7.8** | 200 mM |
| **DTT** | 5 mM |
| **DL-Methionine** | 1 mM |

### 8.7.1 Method for producing acetylated lysine analogues

Lyophilised H3 K18C was dissolved in sKAc reaction buffer at a final histone concentration of 0.5 mM. The mixture was incubated with shaking (RT, 30 min) before addition of NVA (50 mM), VA-044 (5 mM) and glutathione (15 mM). The solution was mixed by inversion several times, then incubated under UV irradiation (λ = 365 nm) (2 hr, RT). Progression of the reaction was monitored by LC-MS after 2 hours, and the reaction was extended if unreacted histone was observed. To extend the reaction, half the amount of NVA, VA-044 and glutathione that had been added at the start of the reaction was added again, and the mixture was incubated under UV irradiation (λ = 365 nm) for a further hour at RT before analysis by LC-MS. No unreacted histone was detected by LC-MS after a third hour of reaction. The protein was purified by buffer exchange into 2 mM BME solution using a PD-10 gel filtration column before lyophilisation and storage at -80 °C. Typical yields for protein recovery were 75%.

### 8.7.2 Method for producing dimethylated lysine analogues

Lyophilised histone was dissolved in alkylation buffer to a protein concentration of 0.5 mM before addition of DTT solution (1 M, 20 µL/mL). The mixture was incubated for 1 hour at 37 °C. To the reduced histone solution (950 µL) was added 2-dimethylaminoethyl chloride HCl (1 M, 50 µL). The mixture was incubated for 2 hours at RT with gentle shaking. Completion of the reaction was assessed by LC-MS. If the unreacted histone starting material was still detectable the reaction was extended through addition of DTT (1 M, 10 µL) followed by 2-dimethylaminoethyl chloride HCl (1 M, 50 µL) and a further 2-hour incubation at RT with shaking. No unreacted histone was detected by LC-MS after this second incubation step. The reaction was quenched with addition of 50 µL BME, and the histone products were purified by buffer exchange

into 2 mM BME solution using a PD-10 gel filtration column before lyophilisation and storage at -80 °C. Typical yields for protein recovery were 80%.

### 8.7.3 Method for producing trimethylated lysine analogues

Lyophilised histone or peptide was dissolved in alkylation buffer to a protein/peptide concentration of 0.5 mM before addition of DTT solution (1 M, 20 µL/mL). The mixture was incubated for 1 hour at 37 °C. To the reduced H3 protein/peptide solution in alkylation buffer (1 mL) was added 0.4 µmol (2-bromoethyl) trimethyl ammonium bromide. The mixture was incubated for 2.5 hours at 50 °C with shaking. Progression of the reaction was assessed by LC-MS or MALDI-TOF MS for the H3 histone or $H3_{1\text{-}15}$ peptide, respectively. If unreacted histone/peptide was detected DTT was added (1 M, 10 µL) and the reaction was continued for a further 2.5 hours at 50 °C with shaking. No histone/peptide starting material was detectable by MS after the second 2.5-hour incubation. The reaction was quenched with addition of 50 µL BME. The histone products were purified by buffer exchange into 2 mM BME solution using a PD-10 gel filtration column, and the peptide products were purified by HPLC before lyophilisation and storage at -80 °C. Typical yields of recovery following alkylation and purification was 80% for histones and 70% for peptides.

## 8.8 AlphaScreen

AlphaScreen assays were conducted following a method adapted from Roberts *et al.*[6]

*AlphaScreen buffer*

| Composition | Concentration |
|---|---|
| **HEPES-NaOH pH 7.6** | 50 mM |
| **NaCl** | 100 mM |
| **CHAPS** | 50 µM |
| **BSA**† | 0.1% (w/v) |

†Bovine serum albumin.

Experimental details are provided in **Section 2.3.2** (dose-response curve) and **Section 2.4.3** (AlphaScreen assay). Nickel chelate donor and streptavidin-bound acceptor beads were purchased from Perkin Elmer. AlphaScreen assays were conducted in 384-well white Proxiplates (Perkin Elmer), and were analysed using an EnVision™ 2104 Multilabel Reader (Perkin Elmer).

## 8.9 Nucleosome production

Nucleosomes were assembled from purified histones and 145 bp Widom 601 DNA following a method adapted from Luger *et al.*[7] Experimental details for octamer folding are provided in **Section 4.3.4**, and for nucleosome reconstitution in **Section 4.3.6**.

### 8.9.1 Buffers

*Unfolding buffer*

| Composition | Concentration |
|---|---|
| **Guanidinium chloride** | 6 M |
| **Tris-HCl pH 7.5** | 20 mM |
| **DTT** | 5 mM |

*Refolding buffer*

| Composition | Concentration |
|---|---|
| **NaCl** | 2 M |
| **Tris-HCl pH 7.5** | 10 mM |
| **BME** | 5 mM |

*High salt refolding buffer (HSRB)*

| Composition | Concentration |
|---|---|
| **KCl** | 2 M |
| **Tris-HCl pH 7.5** | 10 mM |
| **DTT** | 1 mM |

*Low salt refolding buffer (LSRB)*

| Composition | Concentration |
|---|---|
| **KCl** | 250 mM |
| **Tris-HCl pH 7.5** | 10 mM |
| **DTT** | 1 mM |

## 8.10 Crystallography

Experimental details for specific crystallisation experiments are provided in the main text of the respective chapters (**Sections 4.4.1**, **4.4.2**, **6.5.1** and **7.5.1**).

### 8.10.1 General crystallography materials and methods

Crystal screens were purchased from Molecular Dimensions, KeraFAST or Hampton Research. Broad screens were prepared and stored in sealed 96 well blocks at 4 °C prior to use. Crystallisation trials were dispensed into CrystalMation Intelli-Plate™ 96-3 well low-profile crystallisation plates (Hampton Research) using a Crystal Phoenix protein crystallography dispenser (Art Robbins Instruments) equipped with a 96-syringe head and a single nano dispenser. 80 µL of crystallisation condition was dispensed into each plate reservoir, and three sitting drop experiments per condition were prepared containing protein solution and well solution volumes of 0.1 and 0.2, 0.1 and 0.1, and 0.2 and 0.1 µL, respectively. The plates were sealed using StarSeal advanced polyolefin film

(STARLAB). Plates were incubated at either 4 °C or RT with regular monitoring. Prior to harvesting, crystals were soaked in a cryoprotectant solution comprising 80% (v/v) well solution and 20% (v/v) glycerol. Crystals were fished using pin-mounted nylon loops before flash cooling in liquid $N_2$. Crystals were stored in cryovials (Hampton Research) under liquid $N_2$ until required. Crystals were analysed at 100 K either using a Rigaku FR/E Superbright rotating copper anode diffractometer equipped with a Saturn 944+ CCD detector, or by synchrotron radiation at beamline IO3 or I04-1 (Diamond Light Source). I03 and I04-1 beamlines were equipped with Pilatus 6M-F detectors.

### 8.10.2 Crystallography data analysis

All data was indexed, integrated and scaled using HKL-3000.[8] The structures were solved by molecular replacement using Phaser, followed by refinement using PHENIX and Coot until the $R_{work}$ and $R_{free}$ values no longer decreased.[9–11]

### 8.10.3 Protein Data Bank accession codes

Coordinates and structure factors have been deposited in the PDB under the following accession codes.

| Protein/complex | PDB accession code |
|---|---|
| **Apo-di-Zn(II) BcII** | 4C09 |
| **BcII–L-captopril** | 4C1H |
| **BcII–D-captopril** | 4C1C |
| **IMP-1–L-captopril** | 4C1F |
| **IMP-1–D-captopril** | 4C1G |
| **Apo-di-Zn(II) VIM-2** | 4BZ3 |
| **VIM-2–L-captopril** | 4C1D |
| **VIM-2–D-captopril** | 4C1E |
| **VIM-2–compound 8** | 5K48 |

Structures solved in VIM-2–*epi*-L-captopril, and VIM-2–*epi*-D-captopril co-crystallisation trials were not submitted to the PDB due to ambiguous electron density

in the enzyme active site. The TRIM24 and H3 K18Ac nucleosome crystal structures described in **Chapter 4** have not been deposited due to their close similarities to reported structures.

## 8.11 Mass spectroscopy techniques

### 8.11.1 Whole protein mass spectrometry

Protein samples were analysed using an LCT Premier mass spectrometer (Waters) coupled to a NanoAquity UPLC system (Waters) equipped with a ProSwift™ RP-4H 1 mm × 50 mm reverse phase chromatography column (Thermo Fisher Scientific), unless otherwise stated. Protein samples were diluted to 10 µg/mL in $H_2O$ before analysis. Samples were loaded onto the column in solvent A (5% (v/v) acetonitrile, 0.1% (w/v) ammonium formate-formic acid pH 7.0 in $H_2O$) and eluted in a linear gradient to solvent B (95% acetonitrile (v/v), 0.1% (w/v) ammonium formate-formic acid pH 7.0 in $H_2O$) over 4 minutes before introduction into the electrospray ion source. Analyses of the data were performed using MassLynx version 4.0, with deconvolution using the MaxEnt1 algorithm.[12]

### 8.11.2 Quadrupole-time of flight mass spectrometry

MS analysis of H3 methylation states in nucleosome demethylation experiments was conducted using Q-TOF MS. Samples containing nucleosome (20 µM) were first unfolded by addition of a 3-fold excess of unfolding buffer, resulting in a 10 µM concentration of H3. The mixture was separated by UPLC using a Chromolith® RP-18 2 mm × 50 mm reverse phase chromatography column (Merck) using a 5-85% (v/v) gradient of acetonitrile in aqueous 0.1% (v/v) TFA solution as the mobile phase. The eluate was analysed directly using a Xevo G2-XS Q-TOF mass spectrometer (Waters).

Control of the spectrometer and analyses of the data were performed using MassLynx version 4.1, with deconvolution using the MaxEnt1 algorithm.[13]

### 8.11.3 Non-denaturing Q-TOF MS

Non-denaturing MS was conducted by *Idlir Liko* or *Nisha Patel* using the procedure described by Sobott *et al*.[14]

Prior to non-denaturing MS experiments, samples were buffer exchanged into 250 mM ammonium acetate buffer (pH 7.4), then loaded into a gold glass capillary. Nano electrospray ionisation measurements were performed using a Q Exactive™ hybrid quadrupole-Orbitrap™ mass spectrometer (Thermo Fischer Scientific) modified for the transmission and detection of high *m/z* ions in positive ion mode. MS/MS analysis was conducted using a quadrupole time of flight II mass spectrometer (Waters).

### 8.11.4 MALDI-TOF MS

*MALDI matrix*

| Composition | Concentration |
|---|---|
| **Acetonitrile** | 350 µL |
| **CHCA**† | 5 µg |
| **$H_2O$** | 150 µL |
| **TFA** | 0.5 µL |

†α-Cyano-4-hydroxycinnamic acid (LaserBioLabs).

MALDI matrix was prepared and spotted (1 µL) onto a MALDI target plate (Waters). 1 µL of sample was spotted onto the MALDI matrix, and the spot was left to dry. MALDI-TOF MS was carried out using a MALDI MX Micro (Waters) MALDI-TOF mass spectrometer in positive ion reflection mode. Spectra were processed using MassLynx™ version 4.0.[12]

## 8.12 Small-angle X-ray scattering

### 8.12.1 Data collection

SAXS data was acquired at the BioSAXS beamline BM29 at the European Synchrotron Radiation Facility (ESRF) using a beam energy of 12.5 keV. The distance between the sample and the (Pilatus 1M) detector was 2.81 m. A 2-fold, 7-step dilution series containing 35-0.3 μM H3 sK18Ac nucleosome with 2 equivalents of TRIM24 in LSRB, were irradiated in a 1 mm diameter quartz glass capillary. A similar dilution series of H3 sK18Ac nucleosome alone was also tested, along with a buffer only reference condition.

### 8.12.2 Data analysis

Scattering intensity curves were measured at all concentrations and corrected for buffer scattering using ScÅtter.[15] The highest quality data was obtained from samples at 17.5 μM for the mixture, and 12.5 μM for the isolated H3 sK18Ac nucleosome, so these datasets were selected for analysis. The radius of gyration was estimated by Guinier analysis in the PRIMUS suite using low-angle data.[16] P(r) plots, histograms of inter-atomic distances, were calculated using GNOM (PRIMUS).[16] 10 low-resolution molecular envelopes were calculated using the program DAMMIF (part of the ATSAS suite).[17] Molecular envelopes were compared using DAMSEL (part of the DAMAVER suite), and the 5-10 envelopes in the closest agreement were aligned using DAMSUP (DAMAVER).[18] Envelopes were averaged using DAMAVER.[18,19] DAMAVER works by assigning an effective occupancy for each voxel, and removing voxels with below half-maximal occupancy.[18] The resulting averaged molecular envelope was visualised and compared to the H3 sK18Ac nucleosome structure using sculptor (part of the Situs suite).[20]

## 8.13 Isothermal titration calorimetry

ITC experiments were conducted using an $iTC_{200}$ MicroCalorimeter™ (Malvern). Prior to analysis, TRIM24 and H3 sK18Ac nucleosome were dialysed into the same batch of LSRB, and H3 sK18Ac histone was dissolved in LSRB, to ensure all buffers were matched. A heat of dilution experiment was conducted by titrating buffer into buffer and subtracted from subsequent spectra. Experiments measuring TRIM24 binding to sK18Ac histone or nucleosome were conducted by placing either the histone or nucleosome into the calorimeter cell at 20 µM and injecting 1.3 µL aliquots of TRIM24 (400 µM) into the cell with stirring (1000 rpm). Data was analysed using ORIGIN version 7.0. The raw power-time curve was corrected for heats of dilution, and the binding constants were calculated by non-linear least squares fitting.

## 8.14 $K_i$ calculations

$IC_{50}$ values for each MBL inhibitor tested in **Chapter 6** and **Chapter 7** were obtained using a fluorescent reporter assay as described in **Section 6.4.1** (carried out by *Dr Jürgen Brem*). $IC_{50}$ values and their 95% confidence intervals were calculated using non-linear regression (GraphPad Prism).[21]

*MBL inhibitor $IC_{50}$ values (µM)*

| | L-captopril | D-captopril | *epi*-L-captopril | *epi*-D-captopril | 8 |
|---|---|---|---|---|---|
| **BCII** | 80 ± 1 | 11 ± 1 | 424 ± 1 | > 500 | - |
| **IMP-1** | 23 ± 1 | 7 ± 1 | 436 ± 1 | 173 ± 1 | 0.23 ± 0.04 |
| **NDM-1** | 157 ± 1 | 20 ± 2 | > 500 | 65 ± 1 | 5.59 ± 0.05 |
| **SPM-1** | > 500 | 262 ± 1 | > 500 | > 500 | - |
| **VIM-2** | 4.4 ± 0.8 | 0.07 ± 0.01 | 5.5 ± 0.7 | 5 ± 2 | 0.23 ± 0.04 |

Using these $IC_{50}$ values $K_i$ calculations were subsequently performed using the Cheng-Prusoff equation for competitive inhibitors:

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_M}}$$

Where [S] is the concentration of substrate used (nitrocefin; 25 μM for captopril stereoisomer experiments, 2.5 μM for compound **8** experiments) and $K_M$ is the substrate binding constant, in this instance for nitrocefin to each MBL (BCII = 8.1 μM; IMP-1 = 55.7 μM; NDM-1 = 8.8 μM; SPM-1 = 16.0 μM; VIM-2 = 7.2 μM).[22,23]

## 8.15 References


1. Bio-Rad. Image Lab version 5.2.1 for Window (2016).
2. Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375,** 23–44 (2004).
3. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468,** 1067–73 (2010).
4. Li, F. *et al.* A direct method for site-specific protein acetylation. *Angew. Chem. Int. Ed. Engl.* **50,** 9611–4 (2011).
5. Simon, M. D. *et al.* The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128,** 1003–12 (2007).
6. Roberts, J. M. & Bradner, J. E. A bead-based proximity assay for BRD4 ligand discovery. *Curr. Protoc. Chem. Biol.* **7,** 263–78 (2015).
7. 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,** 251–60 (1997).
8. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
9. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
10. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68,** 352–367 (2012).
11. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).
12. MassLynx version 4.0 for Windows (2002).
13. MassLynx version 4.1 for Windows (2005).
14. Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74,** 1402–1407 (2002).
15. Robert Rambo. ScÅtter version 3.1R for windows (2017).
16. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.* **36,** 1277–1282 (2003).
17. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76,** 2879–2886 (1999).
18. Volkov, V. V. & Svergun, D. I. Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Cryst.* **36,** 860–864 (2003).

19. The PyMOL Molecular Graphics System, Version 2.0 (2017).

20. Wahle, M. & Wriggers, W. Multi-scale visualization of molecular architecture using real-time ambient occlusion in Sculptor. *PLoS Comput. Biol.* **11,** e1004516 (2015).

21. GraphPad Software. GraphPad Prism version 5.04 for Windows (2010).

22. Cheng, Y.-C. & Prusoff, W. H. Relationship between the inhibition constant ($K_i$) and the concentration of inhibitor which causes 50 per cent inhibition ($I_{50}$) of an enzymatic reaction. *Biochem. Pharmacol.* **22,** 3099–3108 (1973).

23. Van Berkel, S. S. *et al.* Assay platform for clinically relevant metallo-β-lactamases. *J. Med. Chem.* **56,** 6945–6953 (2013).