BIOCHEMICAL AND BIOPHYSICAL STUDIES ON THE JUMONJI-C HYDROXYLASES JMJD7 AND JMJD4



Suzana Markolovic

Chemistry Research Laboratory
Department of Chemistry
Jesus College, University of Oxford

Hilary Term 2017

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

Suzana Markolovic
Jesus College

DPhil Thesis, Abstract
Hilary Term 2017

Biochemical and biophysical studies on the Jumonji-C hydroxylases JMJD7 and JMJD4

The human Jumonji-C (JmjC) oxygenases, which belong to the 2-oxoglutarate (2OG)-dependent oxygenase superfamily, are a group of ~60 enzymes that catalyse the hydroxylation and demethylation (via hydroxylation) of proteins, including histones, ribosomes, and release factors. These enzymes are of particular interest due to their roles in transcriptional regulation and in modifying components of translational machinery. This thesis details the biochemical assignment and structural studies of a previously uncharacterised JmjC oxygenase, Jumonji domain-containing 7 (JMJD7) and related JmjC lysyl hydroxylase, Jumonji domain-containing 4 (JMJD4). Biochemical and biophysical studies aiming to elucidate the effects of JMJD7-catalysed hydroxylation and of a previously unreported JMJD4-catalysed modification, γ-lactam formation, are described.

Biochemical studies reveal JMJD7 as a catalytically active 2OG oxygenase, catalysing a previously unreported type of post-translational modification (PTM), (*3S*)-lysyl hydroxylation. A cellular substrate trapping approach, combined with peptide screening, identified two highly related members of the Translation Factor family of GTPases, Developmentally Regulated GTP Binding Proteins 1 and 2 (DRG1-2), as JMJD7 substrates; DRG1-2 were then validated as JMJD7 substrates by cellular studies. Mass spectrometry (MS) and amino acid analyses revealed that JMJD7 catalyses specific (3*S*)-lysyl hydroxylation of highly conserved residues in DRG1-2. Complementary structural studies by x-ray crystallography validated JMJD7 as a JmjC domain-containing protein and indicated a unique dimerisation mode. Biophysical and mutational studies confirmed that JMJD7 dimerisation comprises disulphide bond formation and interactions between *N*- and *C*-terminal α-helices. Given the reported associations between JMJD7 and cancer/autism, the effect of clinically-relevant substitutions on JMJD7 activity, structure, and stability was investigated using biophysical techniques, which identified P272S as a loss-of-function substitution. Overall, the functional assignment of JMJD7 described in this thesis will enable future studies to define its role in normal physiology and disease.

Biophysical studies were also undertaken on the related JmjC lysyl hydroxylase, JMJD4, which regulates translation termination via C4 hydroxylation of K63 from the eukaryotic release factor, eRF1. While JMJD4-catalysed hydroxylation of K63 has been found to increase translation termination efficiency, the exact molecular mechanism by which hydroxylation influences termination was unclear at the start of this thesis, and thus the initial subject of investigation. However, these studies identified an unprecedented JMJD4-catalysed '+30 Da' modification, the chemical assignment of which became the primary subject of investigation. MS and NMR analyses of an isolated +30 Da-containing peptide enabled the characterisation of the modification as oxidative γ-lactam formation, formed via the K63 side chain and S64 backbone amide of eRF1. Protein NMR analyses indicated that this modification induced major chemical shift perturbations, indicative of structural/conformational changes that would certainly affect translation termination should the modification be identified in human cells.

Overall, the work described in this thesis expands the reported chemistry of human 2OG oxygenases by detailing the identification of two previously unreported PTMs, (*3S*)-lysyl hydroxylation and γ-lactam formation. The work describes the first functional assignment of JMJD7 and lays a foundation for all future studies on JMJD7 in normal physiology and disease. Finally, this thesis provides the basis for the future exploration of lactam formation as a biologically-relevant human PTM and the potential of JMJD4 as a biocatalyst.

ACKNOWLEDGEMENTS

I would like to express my sincerest thanks to Professor Chris Schofield for introducing me to the world of 2OG oxygenases and providing me with the incredible opportunity to learn under his guidance and supervision. Thank you especially for be being a positive role model and inspiring me with your passion, scientific curiosity, willingness to collaborate, and unwavering enthusiasm. A *huge* thank you must also go to Dr Sarah Wilkins, my "unofficial" supervisor, mentor, and friend. Thank you for introducing a former chemist to the world of biochemistry and enabling me to develop my skills as a scientific researcher.

The work described in this thesis would certainly not have been possible without collaboration and experimental assistance from colleagues. I would like to thank those who have directly assisted with the experimental work undertaken in this thesis, including: Dr Mathew Coleman and members of his lab, Dr Qinqin Zhuang and Charlotte Eaton, for cellular experiments and productive collaboration; Dr Rasheduzzaman Chowdhury for guidance with x-ray crystallography; Dr Weston Struwe for biophysical analyses; Dr Ivanhoe Leung for protein NMR spectroscopy; Dr Roman Fischer, for guidance with MS analysis of cellular data; and Owain Jones for undertaking a challenging Part II project under my supervision. Their contributions are greatly appreciated and have been acknowledged in the text.

Thank you to the Schofield Laboratory. It has been an absolute pleasure working alongside a dedicated and hardworking bunch of individuals from all corners of the world. It has been an interesting three and half years, but living through Brexit and the 2016 US Election have certainly been manageable thanks to you. Tea-times with you will certainly be missed. Thank you very much to Dr Adam Hardy, Wendy Sobey, Jenny Houlsby, Val DeNewtown, and Zahn Gower for keeping the lab running smoothly. A big thank you must also go to Herminio Manso Jubier for providing me with many litres of media and fond memories of the fermentation lab. I genuinely appreciate the help and encouragement I received from all past and present CJS members, especially (but in no particular order) from Dr Sarah Wilkins, Dr Martine Abboud, Dr Joan Gannon, Dr Christopher Lohans, Dr Richard Hopkinson, Gareth Langley, Thomas Leissing, Dr Louise Walport, Dr Ilaria Pettinati, Dr Jürgen Brem, Dr Adam Hardy, Dr Marios Markoulides, Dr Anne Makena, Rebecca Hancock, Joanna Bonnici, Leah Kearney, Saiful Islam, Robert Leśniak, Emily Howe and others. An extra thank you to those who kindly and constructively proofread this thesis, including Sarah, Chris L., Martine, Rashed, and Rich.

A special thanks must also go the Mass Spectrometry Facility Staff, particularly Dr James Wickens, Professor James McCullagh, Dr Colin Sparrow, Lingzhi Gong, and Elisabete Pires, for always being so generous with their time and so willing to provide help and guidance. Thank you also to Professor Timothy Claridge for providing me with technical NMR support and helping me finally crack the structure of that mysterious '+30' modification.

My studies in Oxford would certainly not have been possible without financial support. I am grateful to the Clarendon Fund and Jesus College for providing me with the opportunity to undertake my doctoral studies at the University of Oxford.

Thank you to my family and friends who have provided me with love and support, particularly during these last three and half years. To my sisters, Anna, Gaby, Dayna, and Sam for providing lots of needed laughter. To my Nani, whose love I will never forget. To the Redos family for your encouragement. To my dear friend, Sarah McCuaig, for being the best running buddy a girl could ask for. To my fiancé, John Redos, for always being my rock and for sticking around in Oxford these last three years for me.

Last, but certainly not least, I would like to thank my parents, Djuro and Kjara Markolovic. Thank you for all that you have sacrificed for me to be where I am today. I love you both very much. This thesis is dedicated to you.

COLLABORATOR CONTRIBUTIONS

All experiments described in this thesis were performed by myself, unless otherwise specified in the text. Collaborator contributions are also summarised below:

Dr Wei Ge, University of Oxford

- Cloning of the pET28a(+)-JMJD7 plasmid used throughout this study

Dr Ming Yang (under the supervision of *Dr Mathew Cockman* & *Prof Sir Peter Ratcliffe*), Target Discovery Institute, University of Oxford

- Preliminary JMJD7 co-immunoprecipitation experiments used for JMJD7 peptide substrate hunting

Dr Qinqin Zhuang and ***Charlotte Eaton*** (under supervision from *Dr Mathew Coleman*), University of Birmingham

- All cellular JMJD7 experiments, including co-immunoprecipitation and immunofluorescence experiments

Dr Roman Fischer, Target Discovery Institute, University of Oxford

- Analysis of cellular mass spectrometry data generated by the Coleman Laboratory

Dr Rasheduzzaman Chowdhury, Department of Chemistry, University of Oxford

- Collection, determination, and refinement of JMJD7 crystal structures

Dr Weston Struwe, Departments of Chemistry and Biochemistry, University of Oxford

- Non-denaturing mass spectrometry and size exclusion chromatography-multi-angle light scattering experiments of JMJD7 oligomerisation and clinically-relevant variants (performed in collaboration)

Dr Ivanhoe Leung, University of Auckland (formerly at the University of Oxford)

- Protein NMR studies and analyses of eRF1, including backbone assignment, solution dynamics, and rRNA binding

Owain Jones, Biochemistry Part II student under my supervision, University of Oxford

- Production of full-length eRF1, linear and cyclic eRF1 peptides, and the +30 Da modified $eRF1_{ND}$ peptide for NMR analyses; Intact protein MS, MS/MS, and elemental composition analyses on $eRF1_{FL}$

Professor Timothy D. W. Claridge, Director of NMR Spectroscopy Facility, Department of Chemistry, University of Oxford

- Assistance with the assignment of the +30 Da modified eRF1 peptide and determination of its chemical identity by NMR spectroscopy

TABLE OF CONTENTS

Abstract i
Acknowledgements iii
Collaborator contributions iv
Table of contents v
Abbreviations xiv
1 Introduction to 2OG oxygenase-catalysed protein hydroxylation 1
1.1 2-Oxoglutarate-dependent oxygenases 1
1.1.1 Structural features of 2OG oxygenases 2
1.1.2 Consensus mechanism 4
1.2 Hydroxylation catalysed by human 2OG oxygenases 6
1.2.1 Collagen biosynthesis 7
1.2.2 Epidermal growth factor domain hydroxylases 9
1.2.3 Oxygen sensing 10
1.3 The JmjC subfamily of human 2OG oxygenases 13
1.3.1 Identification and characteristic features of the JmjC oxygenases 13
1.3.2 JmjC histone demethylases 15
1.3.3 JmjC hydroxylases 19
1.3.3.1 Jumonji domain-containing 6 (JMJD6) 20
1.3.3.2 tRNA wybutosine-synthesising protein 5 (TYW5) 21
1.3.3.3 Ribosomal JmjC hydroxylases 21
1.3.3.4 Jumonji domain-containing 4 (JMJD4) 22
1.3.4 JmjC hydroxylases and protein biosynthesis 23
1.3.5 Uncharacterised JmjC oxygenases 24
1.4 Aims of this thesis 25
2 Biochemical characterisation of Jumonji domain-containing 7 (JMJD7) 27
2.1 Introduction 27
2.1.1 Identification of JMJD7 as a putative 2OG oxygenase 27
2.1.2 JMJD7 expression in human cells and tissues 29
2.1.3 JMJD7 and disease 31
2.1.4 Aims of this chapter 32
2.2 Identification of a JMJD7 substrate 33
2.2.1 Substrate hunting strategy 33
2.2.2 Bioinformatics analysis of JMJD7 36
2.2.3 Identification of proteins that interact with JMJD7 37

2.2.4 Recombinant expression and purification of JMJD7 38
2.2.5 Recombinant expression and purification of JMJD7-H178A 40
2.2.6 JMJD7 catalyses uncoupled conversion of 2OG to succinate 40
2.2.7 Identification of potential JMJD7 substrates via MALDI-TOF-MS-based peptide screening 42
2.3 Biochemical characterisation of JMJD7 as a (*3S*)-lysyl hydroxylase 43
2.3.1 JMJD7 catalyses hydroxylation of residue K22 of DRG1 43
2.3.2 DRG1 peptide substrate optimisation for use as a biochemical probe 46
2.3.3 JMJD7 is a bona fide 2OG oxygenase 50
2.3.3.1 JMJD7 catalysis relies on 2OG and Fe(II) *50*
2.3.3.2 Investigation of the effect of ascorbate on JMJD7 catalysis *52*
2.3.3.3 Investigation of the catalytic activity of JMJD7-H178A *52*
2.3.4 JMJD7 catalyses formation of *(3S)*-hydroxylysine 53
2.3.4.1 Amino acid analysis *53*
2.3.4.2 Optimisation of sample preparation for amino acid analysis *56*
2.3.4.3 Determining the regiochemistry of JMJD7-catalysed lysyl hydroxylation *58*
2.3.4.4 Determining the stereochemistry of JMJD7-catalysed lysyl hydroxylation *59*
2.3.5 JMJD7 catalyses hydroxylation of recombinant full-length DRG1 61
2.3.5.1 DRG1 expression trials and small scale purification *61*
2.3.5.2 Recombinant expression and purification of DRG1 ..... *62*
2.3.5.3 Investigation of full-length DRG1 as a JMJD7 substrate *63*
2.3.6 Investigation of DRG2 as a JMJD7 substrate 63
2.3.6.1 DRG2 is closely related to DRG1 ..... *63*
2.3.6.2 Investigation of DRG2 peptides as JMJD7 substrates *65*
2.3.6.3 Investigation of full-length DRG2 as a JMJD7 substrate *67*
2.3.7 Investigation of JMJD7 peptide substrate selectivity – DRG1 v. DRG2 68
2.3.7.1 Limitations of determining kinetic parameters by MALDI-TOF-MS *68*
2.3.7.2 Competition assays reveal that DRG2 hydroxylation is inhibited by the presence of DRG1 ..... *70*
2.3.8 Investigation of JMJD7 as a histone lysine demethylase 72
2.3.9 K22/K21 of DRG1-2 are hydroxylated by JMJD7 in cells 74
2.3.9.1 JMJD7 interacts with DRG complexes in an activity-dependent manner *74*
2.3.9.2 MS/MS analyses of exogenous and endogenous DRGs *76*
2.3.9.3 Subcellular localisation of JMJD7 and DRG1-2 ..... *78*
2.4 Discussion and future perspectives 80
2.4.1 Roles of the DRG Proteins 80
2.4.1.1 DRG proteins belong to the Translation Factor (TRAFAC) GTPases *80*
2.4.1.2 DRG proteins and disease *81*
2.4.2 2OG oxygenases that modify GTPases 82

2.4.3 JmjC oxygenases that modify translational machinery.....**Error! Bookmark not defined.**
2.4.4 JMJD7-catalysed hydroxylation expands known 2OG oxygenase-catalysed lysyl modifications 83
2.4.5 JMJD7 and disease.....85
2.4.6 Potential roles for JMJD7-catalysed hydroxylation.....86

3 Structural characterisation of Jumonji domain-containing 7 87
3.1 Introduction.....87
3.1.1 Aims of this chapter.....88
3.2 Crystal structures of human JMJD7 and JMJD7-R260C.....88
3.2.1 Crystallisation trials and hit identification.....88
3.2.1.1 Full-length His_6-JMJD7 and His_6-JMJD7-R260C 88
3.2.2 Crystallisation and data collection.....90
3.2.2.1 JMJD7 crystallisation data collection in H32 crystal form..... 90
3.2.2.2 JMJD7-R260C crystallisation data collection in $P2_12_12$ crystal form 91
3.2.3 Structural comparison of JMJD7 with JMJD7-R260C.....91
3.2.4 Overall structure of JMJD7.....94
3.2.5 The JMJD7 active site.....96
3.3 Investigation of the JMJD7 oligomerisation domain.....97
3.3.1 Structural analysis of the JMJD7 oligomerisation domain 97
3.3.2 Design of JMJD7 constructs for investigation of oligomerisation status 98
3.3.3 Cloning, expression, and purification of JMJD7 oligomerisation constructs 99
3.3.4 Investigation of the effect of truncations on JMJD7 structural fold by circular dichroism 102
3.3.5 Investigation of oligomeric state of JMJD7 in solution.....104
3.3.5.1 Non-denaturing MS analyses of JMJD7 oligomerisation constructs 104
3.3.5.2 SEC-MALS analyses..... 106
3.3.6 Investigation of the activity of JMJD7 oligomerisation constructs 108
3.4 Discussion and future perspectives.....110
3.4.1 Overall structural comparison of JMJD7 with JmjC demethylases and hydroxylases 110
3.4.1.1 The importance of the structural characterisation of JMJD7 110
3.4.1.2 Overall domain architectures..... 111
3.4.1.3 Oligomerisation..... 114
3.4.1.4 The JmjC active site and substrate binding..... 117
3.4.2 Summary and future perspectives.....120

4 Investigation of clinically-relevant JMJD7 substitutions.....123

4.1 Introduction 123
4.1.1 JMJD7 variants associated with disease 123
4.1.2 Aims of this chapter 125
4.2 Production of JMJD7 variants 126
4.2.1 Rationale for JMJD7 variants chosen for investigation 126
4.2.2 Mutagenesis, expression, and purification of JMJD7 variants 128
4.3 Comparison of hydroxylation activity between JMJD7 clinically-relevant variants 130
4.4 Biophysical analyses of clinically-relevant JMJD7 variants 132
4.4.1 Limitations of studying the P272S variant biophysically 132
4.4.2 Investigation of the effects of clinically-relevant JMJD7 substitutions on secondary structure 133
4.4.3 Investigation of clinically-relevant JMJD7 variant structural stability 133
4.4.4 Investigation of JMJD7 variant oligomerisation status 134
4.4.4.1 Non-denaturing MS 135
4.4.4.2 SEC-MALs 136
4.5 Structural analyses of the JMJD7 clinically-relevant variants 137
4.6 Discussion and future perspectives 140

5 Identification of a new PTM catalysed by Jumonji domain-containing 4 (JMJD4). 143
5.1 Introduction 143
5.1.1 JMJD4 is a 2OG oxygenase acting on eRF1 143
5.1.2 Translation termination and the role of eRF1 144
5.1.3 eRF1, the NIKS motif, and stop codon recognition 146
5.1.4 Aims of this chapter 147
5.2 Expression and purification of $eRF1_{ND}$ and JMJD4 reveal an unprecedented +30 Da modification 149
5.2.1 Strategies to isolate $eRF1_{ND}$-OH for study by NMR spectroscopy 149
5.2.2 Expression trials 151
5.2.2.1 Co-expression studies 151
5.2.2.2 Co-lysis studies 153
5.2.2.3 Attempts to express and purify ^{15}N-$eRF1_{ND}$-OH for NMR studies 154
5.2.3 The '+30 Da' modification is JMJD4-dependent 156
5.2.4 Validation of the +30 Da modification by high resolution protein MS 158
5.2.5 Modified aims 159
5.3 Isolation of $eRF1_{ND}$-OH and $eRF1_{ND}$-30 160
5.3.1 Isolation of $eRF1_{ND}$-OH 160
5.3.2 Isolation of $eRF1_{ND}$-30 162
5.4 The +30 Da modification is apparently localised to K63 164

5.4.1 LC-MS/MS analyses of $eRF1_{ND}$ 165
5.4.2 Amino acid analysis of $eRF1_{ND}$ 171
5.5 Investigation of JMJD4-catalysed +30 Da modification of full-length eRF1 172
5.5.1 Expression and purification of $eRF1_{FL}$ 173
5.5.2 LC-MS analyses of JMJD4-catalysed +30 Da modification of $eRF1_{FL}$ 173
5.5.3 LC-MS/MS analyses of JMJD4-catalysed +30 Da modification of $eRF1_{FL}$ 174
5.6 Investigation of the effects of hydroxylation and +30 Da modification on $eRF1_{ND}$ structure, stability, and protein dynamics 177
5.6.1 Investigation of the effects of hydroxylation and +30 Da modification on $eRF1_{ND}$ secondary structure 177
5.6.2 Investigation of the effect of hydroxylation and +30 Da modification on $eRF1_{ND}$ stability 178
5.6.3 Investigation of JMJD4-catalysed $eRF1_{ND}$ modifications by protein NMR spectroscopy 179
5.6.3.1 Protein NMR spectroscopy 179
5.6.3.2 Expression of ^{15}N and $^{15}N/^{13}C$-labelled $eRF1_{ND}$ 181
5.6.3.3 ^{1}H-^{15}N HSQC spectra suggest structural differences between wild type and JMJD4-catalysed $eRF1_{ND}$-30 182
5.6.3.4 Protein backbone assignment by triple resonance experiments 184
5.6.3.5 ^{15}N and $^{13}C\alpha$ chemical shift perturbations suggest $eRF1_{ND}$-30 is structurally/conformationally distinct 184
5.6.3.6 Investigation of protein dynamics 186
5.6.3.7 $eRF1_{ND}$ interactions with ribosomal RNA 188
5.7 Characterisation of the +30 Da modification 189
5.7.1 Elemental composition analyses of $eRF1_{FL}$ tryptic peptide fragments by LC-MS/MS 190
5.7.2 Peptide optimisation for study by NMR 191
5.7.2.1 Linear peptides 193
5.7.2.2 Cyclic peptides 194
5.7.2.3 Isolation of a +30 Da modified $eRF1_{ND}$ tryptic peptide fragment 197
5.7.3 Analysis of the isolated +30 Da modified peptide by NMR spectroscopy 199
5.7.4 Proposed reaction mechanism for γ-lactam formation 207
5.7.5 Mutation studies support oxidative γ-lactam structure 208
5.8 The +30 Da modification is efficiently formed in bacterial cells 210
5.8.1 Co-expression of $eRF1_{ND}$ and JMJD4 210
5.8.2 Co-expression of $eRF1_{FL}$ and JMJD4 211
5.8.3 Co-expression offers a better alternative for the isolation of $eRF1_{ND}$-30 212
5.9 Investigation of the +30 Da modification in human cells 212

5.10 Discussion and future perspectives 216
5.10.1 Effect of hydroxylation and +30 Da modification on $eRF1_{ND}$ 216
5.10.2 Isolation of a peptide for NMR characterisation 218
5.10.3 Localisation and identity of the +30 Da modification 219
5.10.4 2OG oxygenase catalysis and ring formation 221
5.10.5 Recommendations for future studies on JMJD4-catalysed γ-lactam formation 224

6 Summary and recommendations for future work 227
6.1 Summary 227
6.2 Recommendations for future work 230
6.2.1 Functional studies of JMJD7 230
6.2.2 Investigation of JMJD4-catalysed γ-lactam formation 231

7 Materials and methods 233
7.1 General considerations 233
7.1.1 pH measurements 233
7.2 Microbiological Techniques 233
7.2.1 Preparation of oligonucleotide primers 233
7.2.2 Plasmids used in this thesis 234
7.2.3 Ligation independent cloning (LIC) 236
7.2.3.1 DNA amplification 236
7.2.3.2 Vector preparation and cohesive end generation 237
7.2.4 Agarose gel electrophoresis 238
7.2.5 Site-directed mutagenesis 238
7.2.6 Plasmid purification 239
7.2.7 DNA concentration determination 240
7.2.8 DNA Sequencing 240
7.2.9 Competent cells 240
7.2.10 Bacterial transformations 241
7.2.10.1 Bacterial co-transformations 241
7.2.11 Luria Broth (LB) agar plates 241
7.2.12 Antibiotics 242
7.3 Protein expression 242
7.3.1 2TY media 242
7.3.2 M9 minimal growth media 243
7.3.3 Glycerol stock preparation 243
7.3.4 Starter cultures 244
7.3.5 Large scale growths 244

7.4 Protein purification 245
7.4.1 General considerations 245
7.4.2 Small scale purification with loose His resin 245
7.4.3 Nickel(II) affinity chromatography 246
7.4.4 Size exclusion chromatography 248
7.4.5 Protein concentration determination 249
7.4.6 SDS-PAGE gel electrophoresis 249
7.4.7 Western blot analyses for JMJD7 removal optimisation 250
7.5 Solid phase peptide synthesis (SPPS) 251
7.5.1 Intavis Multipep 251
7.5.2 Liberty Blue 252
7.5.3 Cyclic peptide synthesis 253
7.5.3.1 Peptide synthesis 253
7.5.3.2 Cyclisation 253
7.6 HPLC peptide purification 253
7.7 Hydroxylation assays 254
7.7.1 JMJD7 hydroxylation assays 254
7.7.1.1 Hydroxylation assays by MALDI-TOF MS or LC-MS 254
7.7.1.2 Co-factor/substrate dependence assays & time course conditions 255
7.7.1.3 Competition & activity assays 255
7.7.2 JMJD4 assays 255
7.7.2.1 $eRF1_{ND}$ hydroxylation trials 255
7.7.2.2 Hydroxylation assays 256
7.7.2.3 '+30 Da' assays 256
7.7.2.4 Optimised large scale purification of (un)labelled $eRF1_{ND}$-OH 257
7.7.2.5 Optimised large scale modification and purification of (un)labelled $eRF1_{ND}$-30 for NMR spectroscopy or tryptic digestion and isolation of +30 Da peptide 257
7.8 Mass spectrometry methods 257
7.8.1 MALDI-TOF-MS 257
7.8.2 Intact protein LC-MS 258
7.8.2.1 Open access LCT Premier 258
7.8.2.2 XEVO G2-S QTof 259
7.8.2.3 Exactive and Q-Exactive 259
7.8.3 LC-MS/MS 260
7.8.3.1 In-solution tryptic digest 260
7.8.3.2 Sep-Pak C18 Purification 261
7.8.3.3 Synapt G2-Si high definition mass spectrometry (HDMS) 261
7.8.3.4 XEVO G2-S QTof 262

7.8.4 Non-denaturing mass spectrometry 262
7.8.5 Amino acid analysis 263
7.8.5.1 Sample Preparation 263
7.8.5.2 LC-MS analyses 264
7.9 JMJD7 protein crystallisation 265
7.9.1 Crystallisation 265
7.9.2 Data collection, structure solution, and refinement 266
7.10 Circular dichroism 267
7.11 Differential scanning fluorimetry (DSF) 267
7.12 Size exclusion chromatography-Multi-angle light scattering 268
7.13 Nuclear magnetic resonance 268
7.13.1 Peptide concentration determination by ^{1}H-NMR 268
7.13.2 ^{1}H-NMR 2OG turnover time course 269
7.13.3 Peptide NMR experiments 270
7.13.4 Protein NMR experiments 270
7.14 Protein Basic Local Alignment Search Tool (BLAST) DRG1-2 analyses 271

8 References 273

A Appendix: Supplementary Information 291
A1. JMJD7 co-immunoprecipitation results 291
A2. ^{13}Cα chemical shift index (CSI) 292

B Appendix: Investigation of JMJD7 as a methionine oxidase 297
B.1. Introduction: Physiological relevance of methionine oxidation 297
B.1.1. Aims 299
B.2. JMJD7 catalyses methionine oxidation is an enzyme-dependent manner 299
B.2.1. Alanine scanning mutagenesis 299
B.2.2. MALDI-TOF-MS/MS analyses 301
B.2.3. Co-factor dependence assays 302
B.3. Investigation of the stereospecificity of JMJD7-catalysed methionine oxidation 303
B.3.1. MALDI-TOF-MS assays 303
B.3.2. Amino acid analysis 304
B.3.3. Reversed-phase chromatography analyses 306
B.4. Summary and future work 308
B.5. Materials and methods 310
B.5.1. JMJD7 oxidation assays 310
B.5.2. Reductase assays with MsrA and MsrB 310
B.5.2.1. MALDI-TOF-MS assays 310

B.5.2.2. Amino acid analysis 310
B.5.2.3. Reversed-phase chromatography assays 310
B.5.3. Solid phase peptide synthesis 311
B.5.4. MALDI-TOF-MS/MS 311
B.5.5. Amino acid analysis 311

C Appendix: Copyright permissions 313

ABBREVIATIONS

2OG	2-oxoglutarate
2OG oxygenase	Fe(II)- and 2OG-dependent oxygenases
ALKBH	AlkB Homologue
AQC	6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate
ARID	AT-rich interaction domains
ASPH	human aspartyl/asparaginyl β-hydroxylase
ASU	asymmetric unit
BBOX	γ-butyrobetaine hydroxylase
CAS	clavaminate synthase
CD	circular dichroism
CHCA	α-cyano-4-hydroxycinnamic acid
C-P3H	collagen prolyl-3-hydroxylases
C-P4H	collagen prolyl-4-hydroxylases
CSI	Chemical Shift Index
CSP	chemical shift perturbation
DFRP1-2	DRG family regulatory proteins 1 and 2
DMOG	dimethyloxalyl glycine
DRG1-2	developmentally-regulated GTP-binding proteins 1 and 2
DSBH	double-stranded β-helix
DSF	differential scanning fluorimetry
DTT	dithiothreitol
E. coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
EF-Tu	elongation factor-thermo unstable
EGF	epidermal growth factor
eRF1	eukaryotic release factor 1
eRF1$_{FL}$	full-length eRF1
eRF1$_{ND}$	N-domain of eukaryotic release factor 1
eRF1$_{ND}$-30	+30 Da modified N-domain of eukaryotic release factor 1
eRF1$_{ND}$-OH	hydroxylated N-domain of eukaryotic release factor 1
eRF3	eukaryotic release factor 3
ESI	electrospray ionisation
FIH	factor inhibiting HIF
FTO	fat mass and obesity-associated protein
GDP	guanosine diphosphate
GTP	guanosine triphosphate
H#K#me3/2/1	histone H# tri-/di-/mono-methylated lysine #
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HIF	hypoxia inducible factor
His$_6$	polyhistidine-tag
HMGB1-2	high-mobility group box proteins 1 and 2
HSPBAP1	heat shock protein beta-1 associated protein 1

IPNS	isopenicillin N synthase
IPTG	isopropyl-β-D-1-thiogalactopyranoside
JmjC	Jumonji-C
JMJD#	Jumonji domain-containing #
KDM(#)	histone lysine demethylase (#)
LIC	ligation independent cloning
LC-MS	liquid chromatography mass spectrometry
LC-MS/MS	liquid chromatography tandem mass spectrometry
MALDI-TOF-MS	matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry
MALDI-TOF-MS/MS	matrix-assisted laser desorption/ionisation time-of-flight tandem mass spectrometry
MINA53	Myc-induced nuclear antigen 53
mRNA	messenger RNA
MS	mass spectrometry
NMR	nuclear magnetic resonance
NO66	nucleolar protein 66
NOE	nuclear overhauser effect
NOG	*N*-oxalylglycine
PCR	polymerase chain reaction
PDB	Protein Data Bank
PHD	prolyl hydroxylase domain proteins
PLA2G4B	phospholipase A2 group IVB
PLODs	procollagen lysyl hydroxylases
PPHD	*Pseudomonas* prolyl-hydroxylase domain containing protein
PTM	post-translational modification
RBG1	ribosome-interacting GTPase 1
RiOX	ribosomal oxygenase
RPL#	ribosomal protein L#
RPS#	ribosomal protein S#
rRNA	ribosomal RNA
SDM	site-directed mutagenesis
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SEC-MALS	size exclusion chromatography-multi-angle light scattering
SPPS	solid phase peptide synthesis
TET	ten eleven translocation
TMLH	N^{ϵ}-trimethyllysine hydroxylase
TRAFAC	translation factor
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
tRNA	transfer RNA
TYW5	tRNA wybutosine-synthesising protein 5
U2AF65	U2 auxiliary factor 65 kDa subunit
UPLC	ultra-performance liquid chromatography
WH	winged helix

1 INTRODUCTION TO 2OG OXYGENASE-CATALYSED PROTEIN HYDROXYLATION

** Parts of **Chapter 1** have been published in a modified form in the following review: Markolovic, S., Wilkins, S. E., and Schofield, C.J. (2015) Protein hydroxylation catalyzed by 2-oxoglutarate-dependent oxygenases. Journal of Biological Chemistry, 290:20712-22.*

1.1 2-Oxoglutarate-dependent oxygenases

The ferrous iron- and 2-oxoglutarate-dependent oxygenases (2OG oxygenases) are an enzyme superfamily that catalyse the two electron oxidation of a wide range of substrates, including nucleic acids, amino acids, peptides, proteins, lipids, and natural products[1,2]. In addition to their extensive substrate range, 2OG oxygenases catalyse a wide range of oxidative reactions, including oxidative cyclisations[3,4], desaturations[3,5–7], epimerisations[5–7], halogenations[8,9], hydroxylations[10,11], and demethylations[12,13]. Despite their chemical versatility in other organisms, particularly in prokaryotes, only hydroxylation (i.e. the incorporation of an alcohol group) and mechanistically related demethylation reactions are reported for oxygenases in humans and animals[1,14]. There are ~60 2OG oxygenases that have been identified in humans[2] (**Figure 1.1**), the majority of which have been confirmed as catalytically active 2OG oxygenases. For many of these enzymes, further biochemical, cellular, and structural characterisations have been reported.

Human 2OG oxygenases are divided into two structurally distinct groups, the Jumonji-C (JmjC) and non-JmjC oxygenases[12,15]. As shown in **Figure 1.1**, the JmjC oxygenases encompass about half of the enzymes belonging to the 2OG oxygenase superfamily, specifically the JmjC hydroxylases and histone lysine demethylase (KDM) subfamilies 2-7. The remaining enzymes in the 2OG oxygenase superfamily include the small-molecule hydroxylases, prolyl hydroxylases, collagen prolyl hydroxylases, and nucleotide hydroxylases, which are non-JmjC oxygenases.

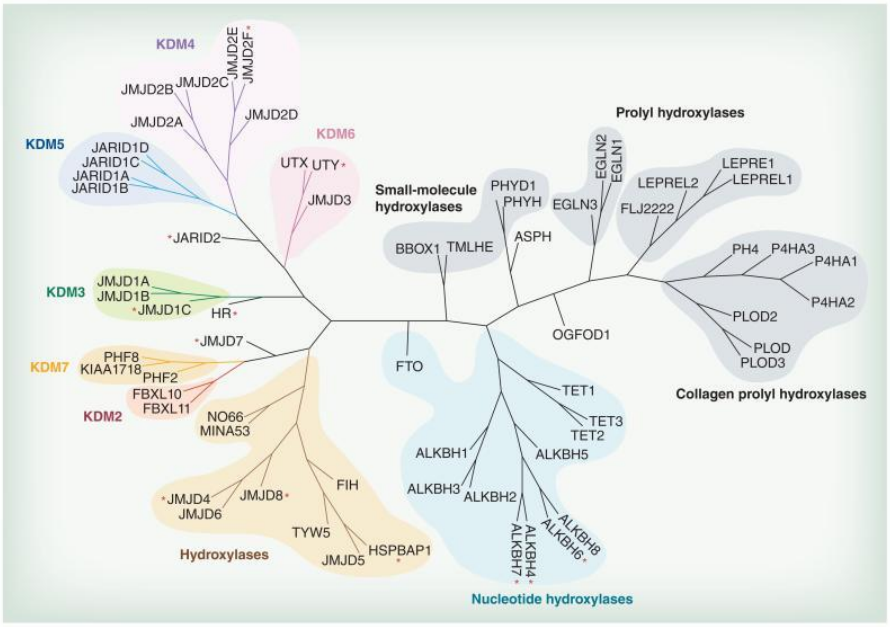


Figure 1.1. Phylogenetic tree of human 2OG oxygenases. Human 2OG oxygenases are structurally divided into two different groups – the Jumonji-C (JmjC) and non-JmjC oxygenases. The JmjC oxygenases encompass the JmjC hydroxylases and histone lysine demethylase (KDM) subfamilies 2-7 (left side of the phylogenetic tree), while the non-JmjC oxygenases are made up of the small-molecule hydroxylases, prolyl hydroxylases, collagen prolyl hydroxylases, and nucleotide hydroxylases (right side of phylogenetic tree). Figure reprinted with permission under the Creative Commons Attribution License 3.0 (https://creativecommons.org/licenses/by/3.0/): Johansson, C. *et al.* (2014)[16].

1.1.1 Structural features of 2OG oxygenases

2OG oxygenases are characterised by the universal presence of a distorted double-stranded β-helix (DSBH) core fold (also known as a jelly-roll, cupin, or Jumonji-C fold), which supports the active site[4,17,18]. The DSBH is made up of eight β-strands organised into two anti-parallel β-sheets. The major sheet (comprised of βI, βVIII, βIII, βVI) and minor sheet (βII, βVII, βIV, βV) are arranged in such a way as to form a barrel-like structure, which contains the catalytic active site (**Figure 1.2**). The open end of the DSBH active site supports the Fe(II)- and 2OG-binding residues. Fe(II) can be found octahedrally-complexed at the centre of the active site. The metal is typically coordinated by a characteristic Fe(II)-binding HX(D/E)…H motif[18,19], though variations of this motif have been observed [e.g. the human aspartyl/asparaginyl β-hydroxylase (ASPH) lacks the acidic Fe(II)-binding residue, as evidenced by crystallographic analyses (Protein Data Bank (PDB) ID: 5APA)]. Crystallographic studies have revealed that 2OG binds

the iron in a bidentate manner, occupying two ligation sites. 2OG is further stabilised by electrostatic interactions arising between its C5 carboxylate and the side chains of a specific lysine or arginine residue, in addition to hydrogen bonding/electrostatic interactions with another alcohol side chain. The final site of the octahedrally-complexed active site Fe(II) is occupied by water.

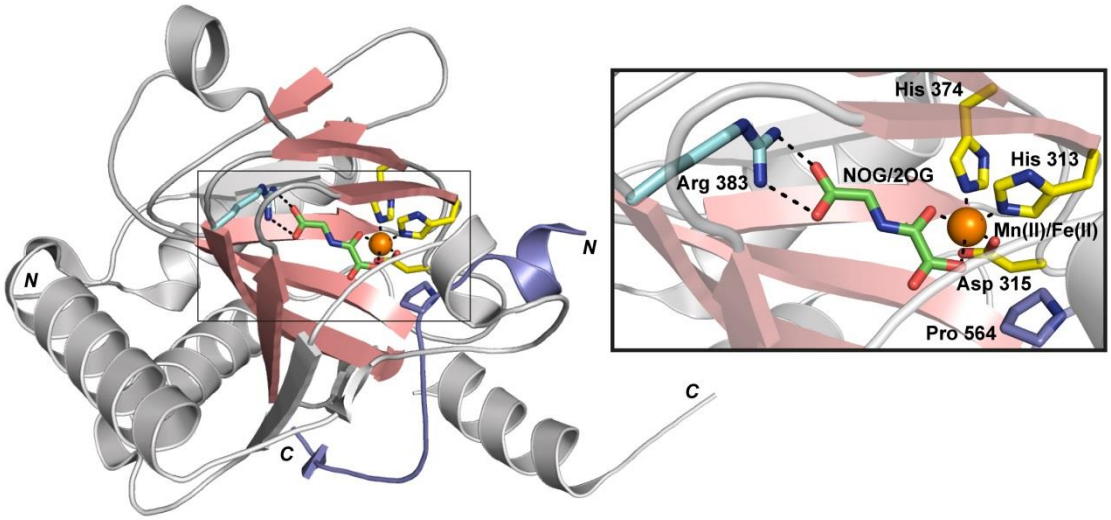


Figure 1.2. Characteristic fold of 2OG oxygenases. Views from a crystal structure of a representative 2OG oxygenase, prolyl hydroxylase domain 2 [PHD2; Protein Data Bank (PDB) ID: 3HQR[20])], reveals the characteristic core double-stranded β-helix (DSBH) fold, which is comprised of eight β-strands (salmon) that make two anti-parallel β-sheets. Mn(II) [orange sphere; an inert substitute for Fe(II) used for crystallography] is bound within the DSBH active site by the characteristic Fe(II)-binding residues (yellow sticks). *N*-Oxalylglycine (NOG; substitute for 2OG and pan-2OG oxygenase inhibitor) binds the metal in a bidentate manner. The C5 carboxylate group of NOG forms hydrogen bonding/electrostatic interactions with the 2OG binding residue (cyan sticks). The substrate peptide (purple) binds to orient the target Pro 564 residue toward the metal, facilitating hydroxylation. The box shows a zoomed view of the active site. Figure reprinted with permission from the American Society for Biochemistry and Molecular Biology (ASBMB; see Appendix C for permissions): Markolovic, S. *et al.* (2015)[11].

Notably, the position and identity of the basic 2OG C5 carboxylate-interacting residue differs between the JmjC and non-JmjC oxygenases. The JmjC oxygenases tend to bind the 2OG C5 carboxylate with a lysine residue, located on the βIV strand of the major sheet (though there are exceptions)[21,22]. By contrast, the non-JmjC oxygenases tend to have an arginine residue located on either the βIV or βVIII strand. Furthermore, the JmjC oxygenases tend to have β-strands immediately *N*-terminal to the DSBH, which non-JmjC oxygenases lack (**Figure 1.3**)[21,22].

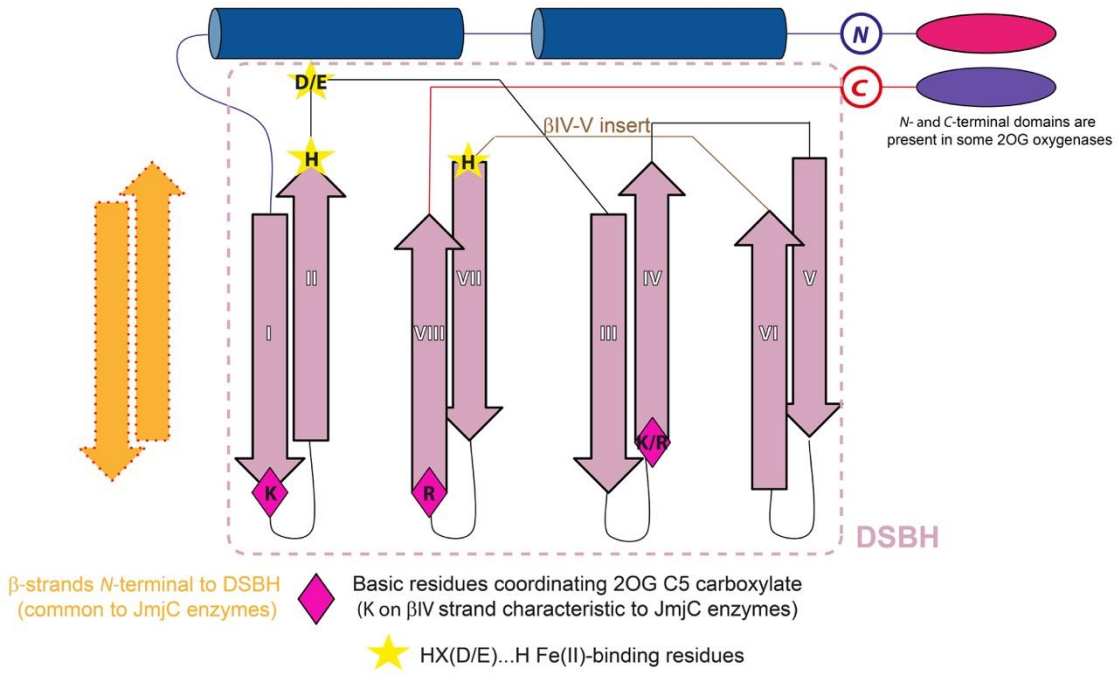


Figure 1.3. Domain architecture of 2OG oxygenases. The DSBH (salmon) is comprised of eight β-strands (labelled I-VIII). Characteristic Fe(II)-binding residues are shown as yellow stars. Basic residues that are involved in coordinating the 2OG C5 carboxylate are shown as magenta diamonds. Note that a lysine residue located on the βIV strand is characteristic of JmjC oxygenases. Figure adapted from: Markolovic, S. & Leissing, T. M. *et al.* (2016)[21].

1.1.2 Consensus mechanism

2OG oxygenases are characterised by their near complete dependence on 2OG and atmospheric oxygen as co-substrates; Fe(II) is also a requirement for catalysis[23,24]. Despite the wide range of oxidative modifications catalysed by 2OG oxygenases, these enzymes likely employ a conserved catalytic mechanism[23] (**Figure 1.4**). 2OG first binds the active site, displacing two water molecules and coordinating the active site-bound Fe(II) in a bidentate manner. Substrate binding follows, causing the displacement of a third water molecule, which frees an Fe(II) coordination site to allow for atmospheric oxygen to bind. Oxygen binding effects oxidation of Fe(II) to Fe(III), which is followed by the oxidative decarboxylation of 2OG to yield a reactive Fe(IV)=O ferryl-oxo intermediate species that effects two-electron oxidation (typically hydroxylation) on the prime substrate.

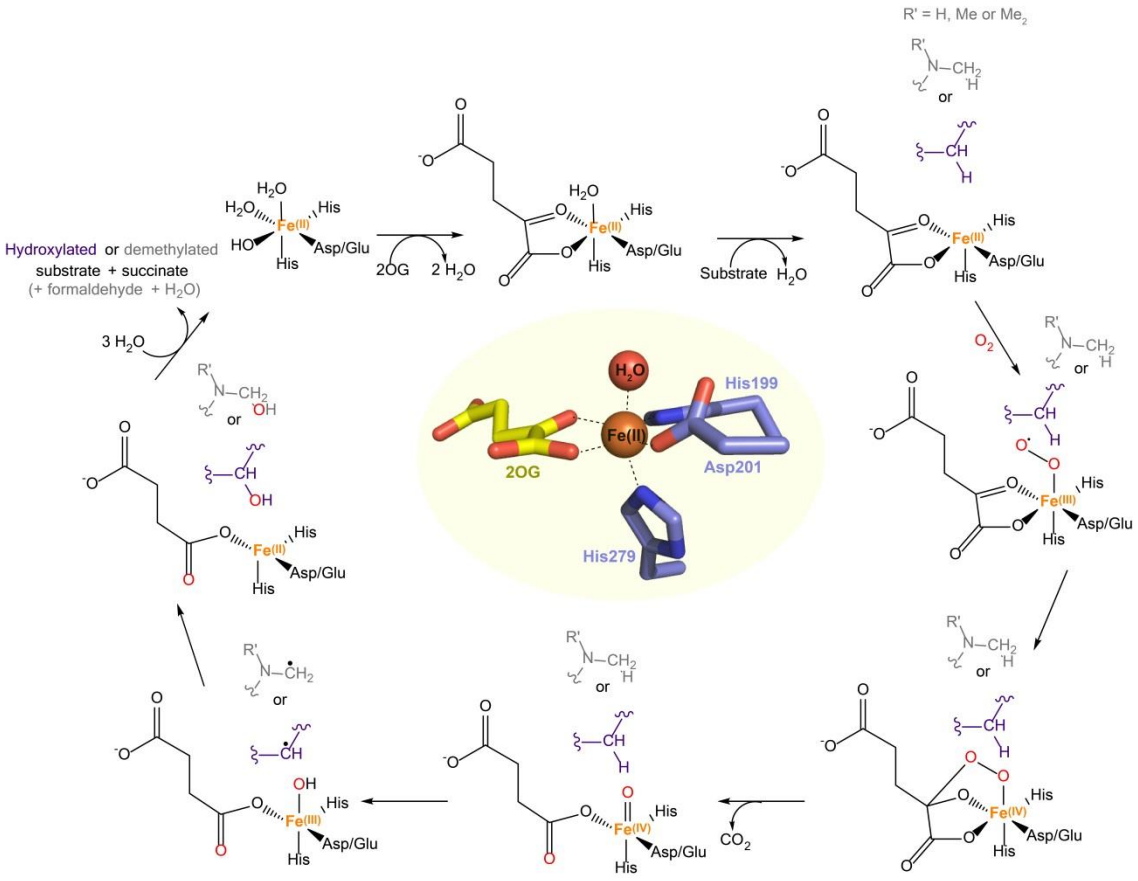


Figure 1.4. The consensus 2OG oxygenase mechanism. Fe(II) (orange) sits within the 2OG oxygenase active site and is octahedrally-bound to the HX(D/E)…H binding residues and three water molecules. 2OG first binds the active site, displacing two water molecules. Substrate binding follows, displacing the third water molecule to allow oxygen to then occupy the coordination site. Oxygen (red) binding generates an Fe(III)-superoxo intermediate, which promotes the distal oxygen to attack the C2 position of 2OG to yield a peroxohemiketal bicyclic intermediate. This species initiates oxidative decarboxylation of 2OG and release of CO_2 via an Fe(IV)-oxo intermediate, which remains bound to succinate (note that one of the oxygen atoms from O_2 is incorporated into succinate). The Fe(IV)-oxo species abstracts a hydrogen atom to form an Fe(III)-OH intermediate and activated C-H radical that is subsequently hydroxylated. Substrate hydroxylation (purple) yields release of the product and succinate. If carbon is attached to a heteroatom (e.g. nitrogen; grey substrate), hydroxylation will yield an unstable hemiaminal intermediate, which fragments to give the demethylated product and formaldehyde. Fe(II) (orange sphere) is shown in the centre of the catalytic mechanism octahedrally-bound to 2OG (yellow), water (red), and the Fe(II)-binding ligands (H199, D201, and H279) of a representative 2OG oxygenase, Factor inhibiting hypoxia inducible factor (FIH; PDB ID: 4Z2W[25]). Figure adapted from Martinez and Hausinger (2015)[23].

Evidence of the Fe(IV)=O intermediate has been accrued using methods that can detect transient intermediate species, including UV-visible[26], electron paramagnetic resonance[26], Mössbauer[27], and Raman[28] spectroscopies. Evidence for the abstraction of a hydrogen atom from a non-activated C-H bond of a ^{2}H-labelled prime substrate by the Fe(IV)=O species has also been shown by Mössbauer studies on two different 2OG oxygenases[27,29]. Coupled oxidative decarboxylation and prime substrate oxidation are followed by release of CO_2,

succinate, and the oxidised product. Demethylation of methylated lysine or arginine residues occurs via initial hydroxylation of the *N*-methyl group to form a hemiaminal intermediate[30,31]; this intermediate then fragments to release the demethylated product and formaldehyde (**Figure 1.4**).

It is noteworthy that, in some cases, 2OG oxygenases in isolated forms and in cells are reported to have enhanced catalytic activity in the presence of ascorbate[32–34]. However, the exact mechanism by which ascorbate stimulates catalytic activity remains unclear. In an *in vitro* or isolated protein context, ascorbate acts as a reducing agent to keep Fe(II) reduced in order to maintain 2OG oxygenase catalytic activity[32]. Yet, even in these contexts, ascorbate is not always required, and the exact mechanism is not known.

1.2 Hydroxylation catalysed by human 2OG oxygenases

As described above, human 2OG oxygenase catalysis is limited to hydroxylation and related demethylation reactions[11,14,35]. The first evidence for 2OG oxygenase-catalysed protein hydroxylation was reported in 1967 with the discovery of the collagen prolyl hydroxylases[10]. While protein hydroxylation was initially thought to be rare (though prevalent in collagen), work over the last decade has shown that hydroxylation is a more common post-translational modification (PTM) than at first perceived[11,14]. The addition of a hydroxyl group is a fairly small, chemically neutral modification; nevertheless, hydroxylation has been found to have major biological effects in processes including collagen biosynthesis, oxygen sensing, and protein biosynthesis[11]. The functional roles of 2OG oxygenases that catalyse protein hydroxylation are described in more detail below.

While the primary focus of this introduction is on 2OG oxygenase-catalysed protein hydroxylation, it is worth commenting on the 2OG oxygenases that catalyse hydroxylation of small molecules and nucleic acids, given the important roles that they play in fatty acid

metabolism and DNA/RNA repair, respectively. For example, two 2OG oxygenases have been shown to catalyse stereospecific hydroxylation of small molecules involved in the biosynthesis of carnitine (**Figure 1.5**)[36], an important energy metabolite responsible for transporting long-chain fatty acids into mitochondria for β-oxidation to take place[36]. In addition, there are several 2OG oxygenases [e.g. ten eleven translocation (TET) enzymes, AlkB Homolog (ALKBH) proteins, and fat mass and obesity-associated (FTO) protein] that have been identified as DNA/RNA hydroxylases and/or demethylases, with roles in DNA/RNA repair and gene activation/silencing[37,38]. 2OG oxygenase catalysis of small molecules and nucleic acids highlights their wide substrate specificity and more generally, their roles in important cellular processes.

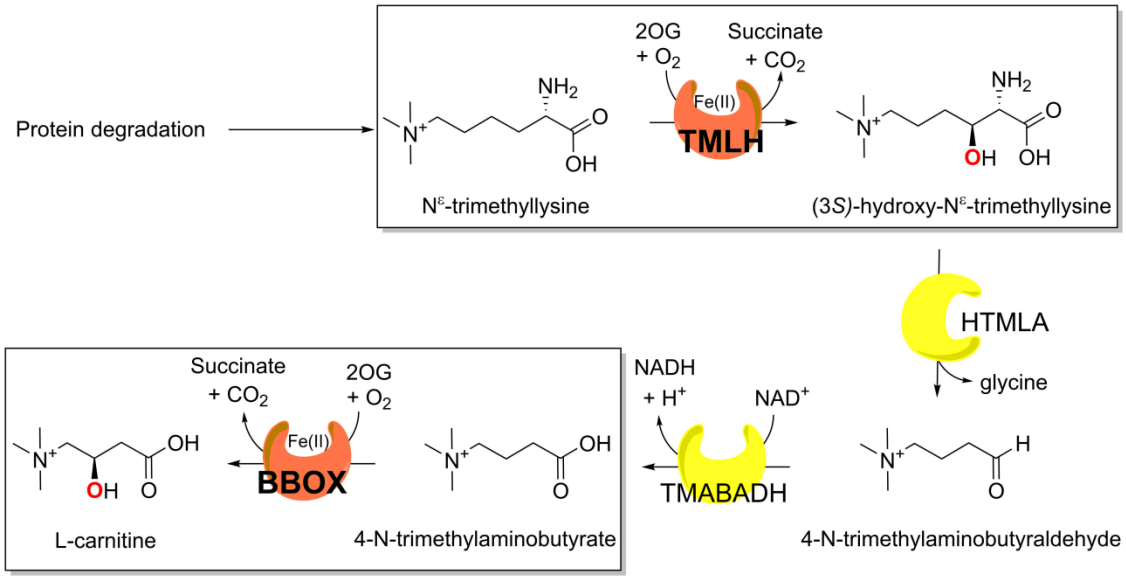


Figure 1.5. 2OG oxygenase-catalysed hydroxylations in carnitine biosynthesis. The carnitine biosynthesis pathway involves two 2OG oxygenases: N^{ε}-trimethyllysine hydroxylase (TMHL), which catalyses formation of (3*S*)-hydroxy-N^{ε}-trimethyllysine (first step), and γ-butyrobetaine hydroxylase (BBOX), which catalyses hydroxylation of 4-N-trimethylaminobutyrate to L-carnitine (final step). HTMLA, (3*S*)-hydroxy-N^{ε}-trimethyllysine aldolase; TMABADH, 4-N-trimethylaminobutyrate dehydrogenase.

1.2.1 Collagen biosynthesis

Collagen is the most abundant protein in all mammals, making up the main structural component of skin and almost all connective tissues[39]. It is synthesised from procollagen, which undergoes several PTMs prior to its assembly in triple helical bundles[40]. Among these

modifications are hydroxylations of proline residues at the C3 and C4 positions as well as C5 lysyl hydroxylation; the stereoselective hydroxylation of these residues is catalysed by three different groups of 2OG oxygenases – the collagen prolyl-3-hydroxylases (C-P3Hs), C-P4Hs, and procollagen lysyl hydroxylases (PLODs), respectively[40] (**Figure 1.6**).

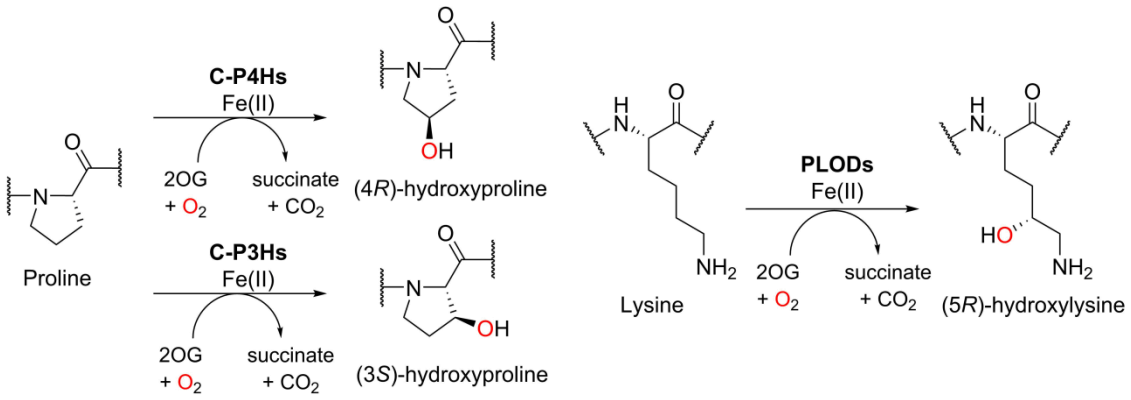


Figure 1.6. 2OG oxygenases involved in collagen biosynthesis. The collagen prolyl-3- and 4-hydroxylases (C-P3Hs and CP4Hs, respectively) catalyse formation of (3*S*)- and (4*R*)-hydroxyproline in procollagen, respectively. The procollagen lysyl hydroxylases (PLODs) catalyse (5*R*)-hydroxylation. Catalysis of C-P3Hs, C-P4Hs, and PLODs require Fe(II) as a co-factor and 2OG and oxygen as co-substrates for hydroxylase catalysis.

(4*R*)-Prolyl hydroxylation is the most abundant PTM in humans, in large part due to the abundance of collagen, which has several hydroxylation sites. Hydroxylation occurs on multiple proline residues at the Y position in the repeating *Gly-X-Y* motif found in the collagen triple helix domain[41]. The functional effect of (4*R*)-prolyl hydroxylation in stabilising the triple helix fold is well-established[42]. By contrast, (3*S*)-prolyl hydroxylation is less abundant[43,44]. While it has been proposed that collagen C3 hydroxylation destabilises the triple helix fold[45], its role in collagen biosynthesis is less well understood than for (4*R*)-prolyl hydroxylation. Likewise, the biological roles of procollagen (5*R*)-lysyl hydroxylation are similarly not well understood[39], though procollagen lysyl hydroxylation enables O-glycosylation (with galactose or glucosylgalactose)[46] and subsequent cross-linking via its side chain[39,47,48]. Given the important roles of the collagen hydroxylases in maintaining the structural integrity of human connective tissues, it is not surprising that loss-of-function mutations have reported links with connective tissue diseases (e.g. *C-P3H* mutations are associated with osteogenesis imperfecta,

or brittle bone disease[41,49]). Furthermore, the C-P4Hs rely on ascorbate to a greater extent than most 2OG oxygenases[32]; interestingly, a vitamin C (ascorbate) deficiency causes scurvy, a connective tissue disease[50]. Overall, the collagen prolyl hydroxylases exemplify the biological importance of 2OG oxygenase-catalysed protein hydroxylation.

1.2.2 Epidermal growth factor domain hydroxylases

About 15 years after the discovery of collagen prolyl hydroxylases, (3*R*)-hydroxylation of asparaginyl and aspartyl residues was identified in epidermal growth factor (EGF)-like domains (**Figure 1.7**)[51,52], including vitamin-K dependent proteins, coagulation factors, complementation factors, and ligands of the Notch signalling pathway[53]. These hydroxylations were later identified to be catalysed by aspartyl/asparaginyl β-hydroxylase (ASPH)[54,55], a 2OG oxygenase that contains a unique Fe(II)-binding motif, with only the two His residues of the typical HX(D/E)…H Fe(II)-binding triad present (PDB ID: 5APA). While little is known about the functional consequences of ASPH-catalysed hydroxylation, ASPH is of medical interest given that its overexpression is associated with cancer progression[53], and high levels of ASPH are correlated with highest grade glioblastoma multiforme[56]. Furthermore, ASPH mutations are linked with severe facial abnormalities, as exhibited by *ASPH* knockout mice[57]. The links between ASPH and disease highlight the importance of 2OG oxygenases from a medical perspective, indicating the study of hydroxylation may elucidate the molecular mechanisms underlying certain diseases.

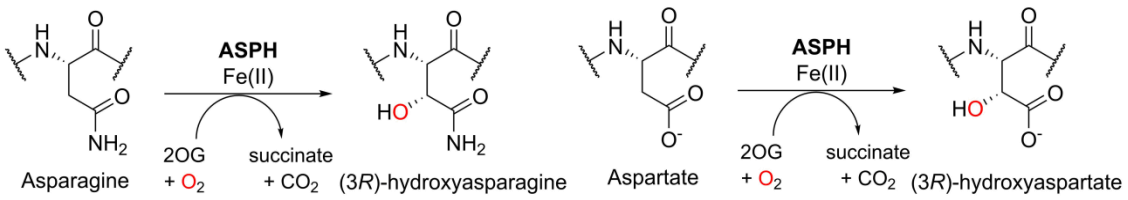


Figure 1.7. ASPH-catalysed hydroxylation of EGF-like domain proteins. ASPH catalyses (3*R*)-hydroxylation of asparaginyl and aspartyl residues in EGF-like domain proteins.

1.2.3 Oxygen sensing

The cell biology of all animals is regulated by the availability of oxygen. Limited oxygen availability (hypoxia) is linked to multiple diseases including cancer, cardiovascular disease, and anaemia[58–60]. Oxygen sensing on a chronic time scale is monitored by the post-translational hydroxylation of the hypoxia inducible factor (HIF)[61]. HIF is a heterodimeric (α/β) transcription factor that regulates the transcription of a large set of genes that enable the cell to adapt to low oxygen conditions, including genes encoding proteins that are involved in cell metabolism, angiogenesis, erythropoiesis, proliferation, and viability (e.g. *EPO*, *VEGF*)[60]. The HIF hydroxylases, which include the prolyl hydroxylase domain proteins (PHDs; PHD1-3) and Factor inhibiting HIF (FIH), are members of the 2OG oxygenase superfamily. The identification of roles for 2OG oxygenase-catalysed hydroxylation of the hypoxia inducible factor (HIF) by the HIF hydroxylases was a ground-breaking discovery in the field of oxygen sensing.

Under normal oxygen availability conditions (normoxia), PHD2 catalyses the (4*R*)-prolyl hydroxylation of P402 and P564[62–66], which are located in two conserved regions of HIF-α, the *N*-and *C*-terminal oxygen-dependent degradation domains (NODD and CODD), respectively (**Figure 1.8**; **Figure 1.9**). Hydroxylation enables HIF-α binding to the von Hippel-Lindau protein (pVHL)-Elongin B/C complex, which targets HIF-α for ubiquitination by an E3 ubiquitin ligase and subsequent proteasomal degradation[62,63]. However, under hypoxic conditions, PHD2-catalysed hydroxylation significantly decreases due to limited oxygen availability. Therefore, instead of being 'tagged' for pVHL binding and proteasomal degradation, HIF-α levels accumulate, "switching on" the hypoxic response and transcriptional expression of genes that enable the cell to adapt. The affinity for atmospheric oxygen is such that it responds to even minor changes in oxygen concentrations within physiological ranges[67]. Due to its oxygen sensitivity and 'switch-like' control, PHD2 is a major oxygen sensor regulating the hypoxic response[61].

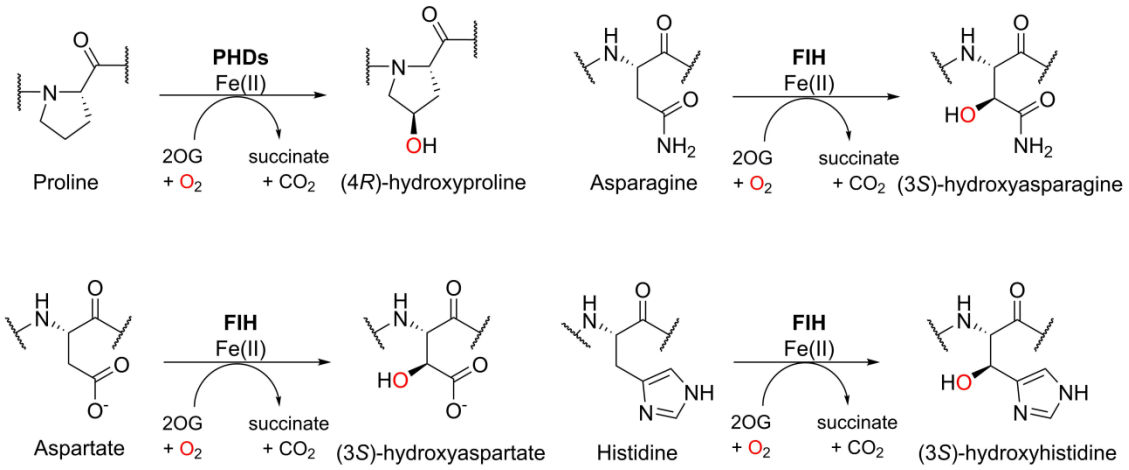


Figure 1.8. Catalysis of 2OG oxygenases involved in oxygen sensing. The PHDs, like the C-P4Hs, catalyse (4*R*)-hydroxylation of two proline residues in conserved regions of HIF-α. FIH is a more promiscuous enzyme, capable of catalysing (3*S*)-asparaginyl hydroxylation of HIF-α and (3*S*)-asparaginyl, aspartyl, and histidinyl hydroxylation of ankyrin repeat domain (ARD) proteins.

HIF transcriptional activity is also regulated by another 2OG oxygenase, FIH, which catalyses (3*S*)-asparaginyl hydroxylation of N803 in the *C*-terminal transcriptional activation domain (CAD) of HIF-1α[15,68] (**Figure 1.8**; **Figure 1.9**). It is worth noting that FIH and PHD2 belong to distinct structural subfamilies of 2OG oxygenases; FIH is a JmjC oxygenase[15], while PHD2 is more similar to the collagen prolyl hydroxylases. FIH-catalysed hydroxylation disrupts the normally tight interaction between the HIF CAD and the CREB-binding/p300 transcriptional coactivator proteins[69]; thus, under hypoxic conditions that reduce FIH activity, tight binding of HIF to coactivator proteins results in transcriptional activation of hypoxia response genes. Studies in cells and with recombinant proteins have indicated that FIH is less sensitive to oxygen than PHD2 (i.e. more active under hypoxic conditions), suggesting that PHD2 is more likely to function as a major oxygen sensor[70–72]. However, given that all 2OG oxygenases rely on molecular oxygen for catalysis, there are potentially other 2OG oxygenases that may act as oxygen sensors; these remain to be identified, if there are any.

FIH has been shown to be a promiscuous enzyme, catalysing the (3*S*)-asparaginyl (preferred), (3*S*)-aspartyl, and (3*S*)-histidinyl hydroxylation of multiple ankyrin repeat domain (ARD) proteins[73–75]. While the roles of ARD hydroxylation are unclear, hydroxylation has a stabilising

effect in some cases[76,77]. The available evidence to date suggests that most human 2OG oxygenases specifically and stereoselectively target a single amino acid residue within a protein. However, FIH is an example of a 2OG oxygenase that accepts multiple substrates, and thus suggests the possibility that other 2OG oxygenases may have wider substrate scopes than currently known.

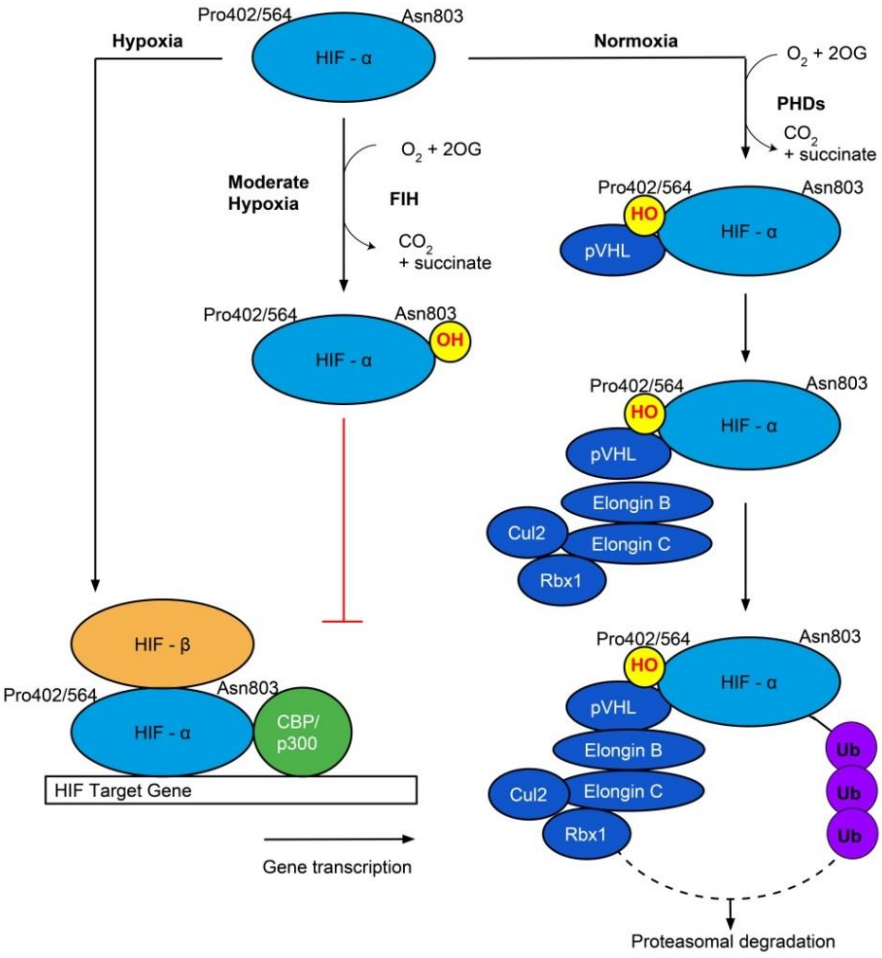


Figure 1.9. PHD- and FIH-catalysed hydroxylation in the transcriptional control of HIF-1α. In normoxia, the PHDs catalyse human HIF-1α prolyl hydroxylation of P402 and P564 [in the *N*/*C*-oxygen-dependent degradation domains (NODD/CODD), respectively]. Hydroxylation is a signal for pVHL binding, which causes subsequent ubiquitination (Ub) and proteasomal degradation of HIF. However, under limiting oxygen conditions (hypoxia), HIF-α cannot be hydroxylated by the PHDs, leading to the accumulation of HIF-α, which "switches on" the transcription of genes that enable the cell to adapt. FIH catalyses N803 hydroxylation within the *C*-terminal transactivation domain, causing disruption of HIF binding with CBP/p300. However, in hypoxia, FIH cannot hydroxylate HIF effectively, causing CBP/p300 to bind and activate the transcription of hypoxia response genes. Figure adapted from Wilkins, S.E. *et al.* (2015)[61].

The discovery of the roles of PHD2 and FIH in oxygen sensing has sparked intense medical and pharmaceutical interest in 2OG oxygenases. For example, PHD2 is a major target for the development of inhibitors, particularly for the treatment of anaemia and cardiovascular disease[78,79]. Inhibitors targeting PHD2, which act to upregulate the production of erythropoietin, are currently in clinical trials for the treatment of anaemia[78,79]. Given that most PHD inhibitors target the catalytic site, which is structurally similar to those in other 2OG oxygenases, the development of selective inhibitors is challenging.

1.3 The JmjC subfamily of human 2OG oxygenases

1.3.1 Identification and characteristic features of the JmjC oxygenases

As described in **Section 1.1**, the human 2OG oxygenases are made up of the JmjC and non-JmjC enzymes (**Figure 1.1**). Notably, FIH is a JmjC oxygenase[12], while PHD2 is structurally more similar to the collagen prolyl hydroxylases and other non-JmjC oxygenases[11]. The first JmjC domain-containing gene, *jmj*, was identified in 1995 by gene trap capture in mice and reported to encode the putative Jumonji protein (human JARID2)[80]. The JmjC domain was named 'Jumonji' (Japanese word for 'cruciform') after the abnormal 'cross-like' appearance of neural grooves observed in *jmj* knockout mice[80]. The protein product of the *jmj* gene was predicted to contain DNA-binding motifs, and two previously undescribed domains, which were designated the JmjC and JmjN domains because of their locations within the protein. Initially, the JmjC and JmjN domains were thought to co-occur[81]; however, subsequent bioinformatic studies have indicated that these domains exist as independent structural and functional units[82]. These bioinformatic studies identified JmjC domains in >100 eukaryotic and bacterial proteins, demonstrating that the JmjC domain was more widespread than initially believed. Because several of these proteins were predicted to have roles in transcriptional control and contain DNA/chromatin-binding motifs [e.g. AT-rich interaction domains (ARIDs), plant homeobox domains (PHDs), and Zn fingers], they were collectively called the Jumonji

family of transcription factors[82]. Significant sequence and active site similarities with the cupin superfamily led to the proposal that the Jumonji family of transcription factors was an unrecognised branch of metalloenzymes that adopt a characteristic cupin, or DSBH, core fold[82]. While these studies proposed that JmjC proteins were likely to be Zn(II)-dependent transcriptional regulators, given their similarity to a Zn(II)-utilising enzyme with a DSBH fold, structural and biochemical analyses of FIH, the first of any JmjC oxygenase, revealed that the JmjC enzymes are Fe(II)-dependent enzymes that adopt a DSBH core fold[15,68,83,84].

The biochemical assignment of FIH as a hydroxylase acting on HIF (described in detail in **Section 1.2.3**) indicated a role for JmjC oxygenases in transcriptional regulation[15,68]. This discovery was very shortly followed by biochemical studies on the first characterised JmjC histone lysine demethylases, belonging to the KDM2, 3, and 4 subfamilies[30,85,86], with structural studies of KDM4A providing the first insights into distinguishing features of KDM architecture[87,88]. These discoveries expanded the roles of JmjC oxygenase catalysis to the regulation of gene expression via histone modifications, and further identified the JmjC KDMs (also known as the JHDMs) as the largest class of demethylase enzymes[12]. The pioneering discoveries of FIH and the KDMs have since led to the identification and biochemical characterisation of several JmjC oxygenases involved in the regulation of gene expression and protein biosynthesis[11]. To date, more than 30 human JmjC oxygenases have been identified, the majority (~20) of which catalyse demethylation of N^{ϵ}-methyl lysine. The remainder (known as the JmjC hydroxylases) catalyse the formation of chemically stable alcohol products. The JmjC oxygenases are thought to follow the consensus 2OG oxygenase mechanism described in **Section 1.1.2**.

While there is no specific definition of a JmjC protein, there are structural features that distinguish the JmjC oxygenases from other members of the 2OG oxygenase superfamily (though they possess the DSBH fold common to 2OG oxygenases). As described briefly in **Section 1.1.1**, the JmjC oxygenases bind the C5 carboxylate group of 2OG using a conserved

lysine residue located near the *N*-terminus of the βIV strand of the DSBH[21,22]. However, there are variations on this binding mode in some family members, including the KDM6 subfamily[89,90] and KDM3B (PDB ID: 4LXL), where the lysine is located on βI and βVIII strands, respectively[21]. A notable difference between the JmjC hydroxylases and KDMs is that the JmjC hydroxylases tend to contain only one recognisable domain – the JmjC domain; thus, they are sometimes called the 'JmjC-only' oxygenases[12,21]. By contrast, the KDMs are larger, multi-domain proteins containing DNA- and histone-binding domains (**Table 1.1**). Other structural features that distinguish the JmjC hydroxylases from the JmjC KDMs, including domain architecture, protein oligomerisation, and active site substrate binding, are described in detail in **Chapter 3**.

1.3.2 JmjC histone demethylases

Each cell contains ~1.8 metres of DNA, which is compactly packed in the nucleus of a cell by wrapping itself around histone proteins[91]. This tightly packed DNA-histone complex, which is held together primarily via electrostatic interactions arising from the negatively charged DNA and positively charged regions of histone proteins, is called a nucleosome[91]. The nucleosome is one basic unit of chromatin and is formed from two of each histone proteins H2A, H2B, H3, and H4 to make an octamer[92].

Gene expression is in large part regulated by PTMs made to the tails of histone proteins[93]. Among these modifications, methylation of the N^{ϵ}-amino group of lysine residues in histone tails is amongst the most well-established as a regulatory mechanism of gene expression[93]. N^{ϵ}-methylation of lysine (and arginine) residues of histone proteins is important from an epigenetic perspective, acting to activate or silence gene expression in a context-dependent manner. For example, methylations of the histone H3 lysine 4 (H3K4), H3K36, and H3K79 sites have been linked to transcriptional activation, while H3K9 and H3K27 methylations are linked to transcriptional repression[94]. Lysine residues can exist in the unmodified, mono-, di-, or tri-

methylated states. For several years, lysine methylation was believed to be an irreversible PTM. However, the discovery of the KDMs in 2004 indicated that methyl marks could be removed from histone proteins[95], highlighting potential regulatory roles for these enzymes during transcription. The first identified KDMs were the flavin-dependent lysine-specific demethylases[95], which make up the smaller KDM1 subfamily of histone demethylases that are structurally and mechanistically distinct from the 2OG oxygenases. The JmjC KDMs were discovered shortly after and have since become the largest identified group of histone lysine demethylases[30,85,86], comprising the KDM2-7 subfamilies. The KDMs have been reported to catalyse the demethylation (via hydroxylation) of specific tri-, di-, and/or mono-methylated lysine residues within histone proteins (**Figure 1.10**) likely via the consensus 2OG oxygenase mechanism (**Section 1.1.2**; **Figure 1.4**).

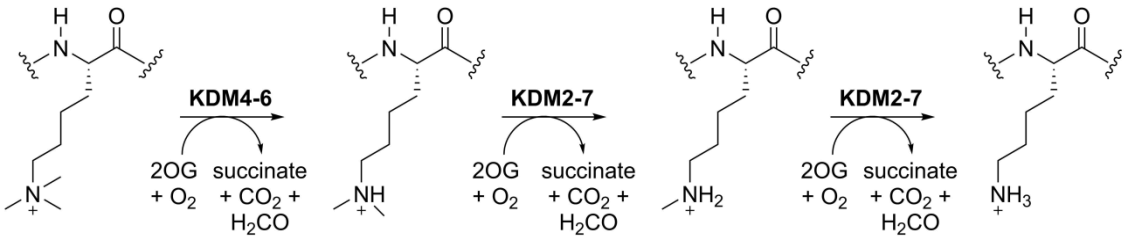


Figure 1.10. Histone lysine demethylation of tri-, di-, and mono-methylated lysine residues by indicated KDMs.

Beyond methylation, histone lysine residues can undergo several other PTMs, including acetylation, ubiquitination, crotonylation, and hydroxylation[93]. The extensive modifications that these residues can undergo highlights the difficulty in gaining a mechanistic understanding of how the dynamic regulation of methylation/demethylation affects transcription. Nevertheless, significant efforts have been made in the biochemical, structural, and functional characterisation of JmjC KDMs as clinical epigenetic targets[96,97], given that several of these enzymes show strong associations with various cancers and intellectual disabilities[16]. Structural studies of KDMs bound to their substrates have notably demonstrated key roles for DNA/chromatin-binding domains in directing histone lysine residue specificity in the active site. Structural studies of the JmjC active site have also provided insights into lysine N^{ε}-

methylation state specificity, an understanding of which is important for the design of selective inhibitors. Currently, there are ongoing efforts to identify selective KDM inhibitors; in the case of KDM1A, there are currently three inhibitors in phase I or I/II clinical trials[98]. Reported KDM catalytic activities and substrates are summarised in **Figure 1.10** and **Table 1.1**.

Like lysine methylation, arginine methylation of histone tails similarly influences gene expression and chromatin remodelling[99]. Recent biochemical and structural studies using purified proteins and histone peptides have shown that a subset of JmjC KDMs (e.g. KDM3A, KDM4E, KDM5C, and KDM6B) can catalyse demethylation of histone N^{ω}-methylated arginine residues (and non-histone substrates)[100]. While evidence remains to be found for their arginine demethylase activity in cells, their observed activity on peptides is, in a way, consistent with their KDM activity and the generally broad scope of 2OG oxygenase substrates.

Table 1.1. Catalytic functions and substrates of the JmjC oxygenases.

	Subfamily	Name	Other Names	NCBI/ UniProt ID	Catalytic Function	Reported Biochemical & Cellular Substrate(s)	Domain Architecture†
Histone Demethylases	KDM2 (FBXL)	KDM2A*	FBXL11; JHDM1A	22992/Q9Y2K7	Lys(me2/1) demethylase	H3K36, p65	1162
		KDM2B	FBXL10; JHDM1B	84678/Q8NHM5	Lys(me2/1) demethylase	H3K36, H3K4me3	
	KDM3 (JMJD1)	KDM3A*	JMJD1A; JHDM2A	55818/Q9Y4C1	Lys(me2/1) demethylase	H3K9	1321
		KDM3B	JMJD1B; JHDM2B	51780/Q7LBC6	Lys(me2/1) demethylase	H3K9	
		JMJD1C	-	221037/Q15652	Lys(me2/1) demethylase	H3K9, MDC1	
	KDM4 (JMJD2)	KDM4A*	JMJD2A; JHDM3A	9682/O75164	Lys(me3/2/1) demethylase	H3K9, H3K36, H1.4K26	1064
		KDM4B	JMJD2B; JHDM3B	23030/O94953	Lys(me3/2/1) demethylase	H3K9, H3K36, H1.4K26	
		KDM4C	JMJD2C; JHDM3C	23081/Q9H3R0	Lys(me3/2/1) demethylase	H3K9, H3K36, H1.4K26; Pc2	
		KDM4D*	JMJD2D; JHDM3D	55693/Q6B0I6	Lys(me3/2/1) demethylase	H3K9, H3K36, H1.4K26	523
		KDM4E	JMJD2E	390245/B2RXH2	Lys(me3/2/1) demethylase	H3K9	
	KDM5 (JARID)	KDM5A*	JARID1A; RBP2	5927 P29375	Lys(me3/2/1) demethylase	H3K4	1690
		KDM5B	JARID1B; PLU1	10765/Q9UGL1			
		KDM5C	JARID1C; SMCX	8242/P41229			
		KDM5D	JARID1D; SMCY	8284/Q9BY66			
	KDM6	KDM6A*	UTX	7403/O15550	Lys(me3/2/1) demethylase	H3K27	1401
		KDM6B‡	JMJD3	23135/O15054	Lys(me3/2/1) demethylase	H3K27	
		KDM6C	UTY	7404/O14607	Lys(me3/2/1) demethylase	H3K27	
	KDM7	KDM7A	KIAA1718; JHDM1D	80853/Q6ZMT4	Lys(me2/1) demethylase	H3K9, H3K27, H4K20	1060
		KDM7B*	PHF8; JHDM1F	23133/Q9UPP1	Lys(me2/1) demethylase	H3K9, H4K20	
		KDM7C	PHF2; JHDM1E	5253/O75151	Lys(me2/1) demethylase	H3K9, H3K27, H4K20me3, ARID5B	
Controversial Assignments	JmjC-only	JMJD6	-	23210/Q6NYC1	Lys (*5S*)-hydroxylase; Arg(me1) demethylase**	Splicing proteins (e.g. U2AF65, LUC7L2), p53 H3R2**; H4R3**	403
		MINA53	MINA	84864/Q8IUF8	His (*3S*)-hydroxylase; Lys(me3/2) demethylase**	RPL27; H3K9**	465
		NO66	-	79697/Q9H6W3	His *3S*-hydroxylase; Lys(me3/2) demethylase**	RPL8; H3K4**, H3K36**	
		JMJD5	KDM8	79831/Q8N371	Lys(me2) demethylase** Hydroxylase**	H3K36** NFATc1**	416
Protein/nucleic acid hydroxylases	JmjC-only	FIH	HIF1AN	55662/Q9NWT6	Asn (*3S*)-hydroxylase; Asp/His (*3S*)-hydroxylase	HIF-1α; Multiple ankyrin repeat domain proteins	349
		JMJD4	-	65094/Q9H9V9	Lys C-4 hydroxylase	eRF1	463
		TYW5	-	129450/ A2RUC4	Wybutosine β-hydroxylase	$tRNA^{Phe}$ (yW-72)	315
Uncharacterized JmjC oxygenases	JmjC-only	HSPBAP1	PASS1	6767/Q96EW2	n/d	n/d	488
		JMJD7	-	100137047/ P0C870	n/d	n/d	316
		JMJD8	-	339123/Q96S16	n/d	n/d	334

JmjC JmjN F-box LRR CXXC Zn Finger C6 Zn Finger PHD Zn Finger ARID TPR Tudor Dimerization WH

† Domain architecture based on reported domains as defined by PROSITE-ProRuler[101] or structures (MINA53[102], FIH[84], and TYW5[103]). * Domain architecture representation is based on this enzyme. ** Controversial functional assignments. ‡ KDM6B does not have reported TPR domains, though is roughly the same size as KDM6A and KDM6C. Table reprinted under the Creative Commons Attribution License 3.0 (https://creativecommons.org/licenses/by/3.0/): Markolovic, S., Leissing, T. M., *et al.* (2016)[21].

1.3.3 JmjC hydroxylases

Since the discovery and functional assignment of FIH in 2001/2002[15,104], several of the remaining putative JmjC hydroxylases (~10) have been the subject of investigations, including Jumonji domain-containing 6 (JMJD6), tRNA wybutosine-synthesising protein 5 (TYW5), Myc-induced nuclear antigen 53 (MINA53), nucleolar protein 66 (NO66), and Jumonji domain-containing 4 (JMJD4) (**Figure 1.11**; **Table 1.1**).

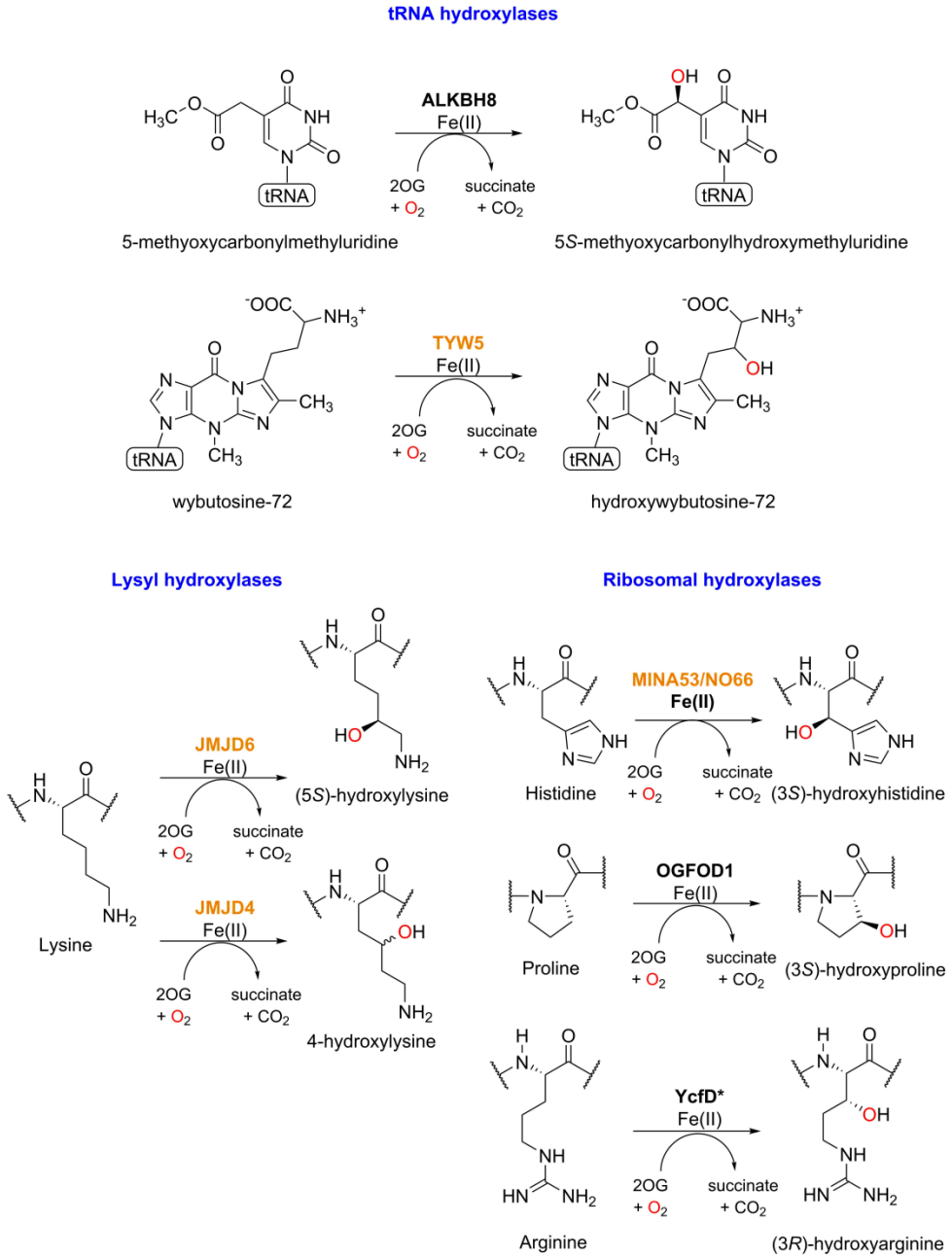


Figure 1.11. 2OG oxygenases with reported roles in human protein biosynthesis. The JmjC hydroxylases are shown in orange. ALKBH8 and OGFOD1 are non-JmjC oxygenases involved in protein biosynthesis. *YcfD is the only bacterial hydroxylase depicted.

1.3.3.1 Jumonji domain-containing 6 (JMJD6)

The first report on the catalytic activity of JMJD6, which used a combination of antibody and mass spectrometry (MS)-based detection methods, assigned JMJD6 as a histone arginine demethylase acting specifically on H3R2 and H4R3[105]. While demethylase activity is consistent with the larger KDM subfamily of JmjC proteins, efforts to replicate this finding have yielded conflicting results[106]. Subsequent biochemical studies revealed that JMJD6 can also catalyse lysyl hydroxylation of mRNA splicing regulatory proteins, including the splicing regulatory factor U2 auxiliary factor 65 kDa subunit (U2AF65) and other related proteins[107,108]. mRNA splicing is the process by which pre-mRNA is edited to remove introns and ligate the remaining exons to make up the mature mRNA. Extensive biochemical characterisation of JMJD6, including by NMR and MS analyses, has shown that JMJD6 catalyses stereospecific lysyl (5*S*)-hydroxylation[109] (**Figure 1.11**; the opposite stereochemistry than that catalysed by the PLODs, as described in **Section 1.2.1**). Recent work has confirmed a biological role for JMJD6 in mRNA splicing, demonstrating that JMJD6 and U2AF65 co-regulate mRNA splicing in a way that is dependent on JMJD6 hydroxylase activity on U2AF65[110].

In addition to acting as an arginyl demethylase and lysyl hydroxylase, JMJD6 has also been reported to act as a histone lysine hydroxylase[111]; whether this activity has a role in transcription remains to be determined. Evidence for its H4 arginyl demethylase activity has since been supported by work identifying a role for JMJD6 in transcriptional pause release[112]. Despite conflicting reports on JMJD6 catalytic activity, JMJD6 is currently accepted as a dual arginyl demethylase and lysyl hydroxylase acting on several substrates (**Table 1.1**). The understanding of its wide substrate scope and dual catalytic activity are currently of intense interest given emerging implications between JMJD6 and cancers in the lung[113], breast[114], and colon[115].

1.3.3.2 *tRNA wybutosine-synthesising protein 5 (TYW5)*

Though it has not been identified as a protein hydroxylase, TYW5 is notably the first (and to date, only) JmjC oxygenase identified to act on transfer RNA (tRNA) or any other type of nucleic acid. TYW5 catalyses hydroxylation of a precursor of wybutosine[116], a hypermodified guanosine of phenylalanine-tRNA that maintains the reading frame by stabilising anticodon-codon pairing[117] (**Figure 1.11**). While TYW5 is the only reported JmjC oxygenase acting on nucleic acids, there are several non-JmjC 2OG oxygenases that act on DNA/RNA (e.g. TETs, ALKBHs, and FTO). Like TYW5, there is one other 2OG oxygenase that has been shown to act on aminoacyl-tRNA. ALKBH8, which also exhibits methyltransferase activity via its methyltransferase domain[118], catalyses hydroxylation of 5-methoxycarbonylmethyluridine, a hypermodified base located within the anticodon loop and at the wobble position of glycine-tRNA[119,120]. Overall, the diverse substrate scope of the 2OG oxygenases is similarly reflected within the JmjC hydroxylases.

1.3.3.3 *Ribosomal JmjC hydroxylases*

Initially assigned as lysine demethylases[121], MINA53 and NO66 were subsequently identified as (3*S*)-histidinyl hydroxylases acting on the ribosomal proteins, RPL27a (His39) and RPL8 (His216), respectively (**Figure 1.11**)[122]. They are structurally distinguishable from other JmjC hydroxylases because they contain a winged-helix domain, a domain with ancient roles in transcription via DNA binding, RNA recognition, and protein-protein interactions[123]. As yet, the biological functions of MINA53- and NO66-catalysed hydroxylations are not yet clear, though their modification sites notably extend into the ribosome; in the case of RPL8, the side chain of the target H216 residue extends close to the peptidyl transferase centre[122], the ribosomal site where peptide bond formation and nascent peptide release occur[124]. MINA53 and NO66 are also of interest from a medical perspective, given that high expression of these proteins, particularly for MINA53, has been linked with poor prognosis in lung, gastric, and breast cancers[125–127]; for this reason, MINA53 has been proposed a diagnostic and prognostic marker in these cancers[125,128].

Notably, MINA53 and NO66 are part of the recently identified ribosomal oxygenases (RiOXs), a group of 2OG oxygenases that act on the ribosome. The third human member of this group is the non-JmjC 2OG oxygenase, OGFOD1, a (3*S*)-prolyl hydroxylase acting on Pro62 of the small ribosomal protein RPS23 (**Figure 1.11**). Knockdown and/or inactivation of OGFOD1 in human cells, yeast, and *Drosophila* causes translational arrest, growth impairment, and in yeast, translation inaccuracy[129–131]. A ribosomal oxygenase from *E. coli*, YcfD, has also been identified. YcfD catalyses (3*R*)-arginyl hydroxylation of ribosomal protein RPL16 (Arg81)[122], which like the target RPL8 residue of NO66, is adjacent to the peptidyl transferase centre, thus suggesting a role in protein translation. While the exact biological consequences of RiOX-catalysed hydroxylation remain to be understood, an effect on translation via ribosomal modifications is consistent with reported roles for the RiOXs in regulating cell proliferation and growth[122,126,129]. Furthermore, current interest in identifying antibiotics that target the ribosome to inhibit protein synthesis highlights the importance of gaining a better mechanistic understanding of YcfD-catalysed hydroxylation of RPL16[132], which may be a viable target for antibiotic development.

1.3.3.4 Jumonji domain-containing 4 (JMJD4)

In addition to modifying tRNA and ribosomal proteins, JmjC oxygenases modify other components of the translational machinery. Jumonji domain-containing 4 (JMJD4) catalyses the C4 hydroxylation of a highly conserved lysine residue in eukaryotic release factor 1 (eRF1; **Figure 1.11**)[133]. eRF1 is a release factor primarily responsible for recognising stop codons and hydrolysing the release of the nascent polypeptide chain. The target residue of JMJD4-catalysed lysyl hydroxylation is located within a highly conserved NIKS motif, which is functionally important for stop codon recognition[134,135]. Early or late protein termination can have huge effects on protein structure, and consequently, detrimental effects on protein function. The functional effect of JMJD4-catalysed hydroxylation is to increase translation

termination efficiency[133], though the exact molecular mechanism by which a hydroxyl group causes this observed increase has yet to be determined.

Though not related to JMJD4, it is worth mentioning that the non-JmjC 2OG oxygenase, *Pseudomonas* prolyl hydroxylase domain containing protein (PPHD), also modifies translational machinery via prolyl hydroxylation of the translation Elongation Factor-Tu (EF-Tu)[136]. While the exact biological function of this modification remains to be understood, the identification of 2OG oxygenases in bacteria indicates that 2OG oxygenase-catalysed hydroxylation has an ancient role in regulating protein biosynthesis[11].

1.3.4 JmjC hydroxylases and protein biosynthesis

A common theme emerging from the functional assignments of the JmjC hydroxylases is that most, if not all, of these enzymes play a role in protein biosynthesis, including (1) transcriptional regulation via FIH-catalysed HIF hydroxylation, (2) mRNA splicing via JMJD6-catalysed hydroxylation of splicing regulatory proteins, (3) the hydroxylation of ribosomal proteins by MINA53/NO66, and (4) JMJD4-catalysed hydroxylation of a release factor responsible for translation termination (**Figure 1.12**). Because protein biosynthesis is one of the most fundamental cellular processes, it is of no surprise that dysregulation of this process has pathological consequences[137]. An understanding of the mechanistic and functional roles of 2OG oxygenases in protein biosynthesis is thus likely to create a more complete picture of this complex process and highlight 2OG oxygenases as attractive therapeutic targets. Given all JmjC oxygenases rely on oxygen, it is also possible that the JmjC oxygenases play a role in adapting to chronic hypoxia via a mechanism involving global reduction in protein translation, which would allow the cell to conserve energy.

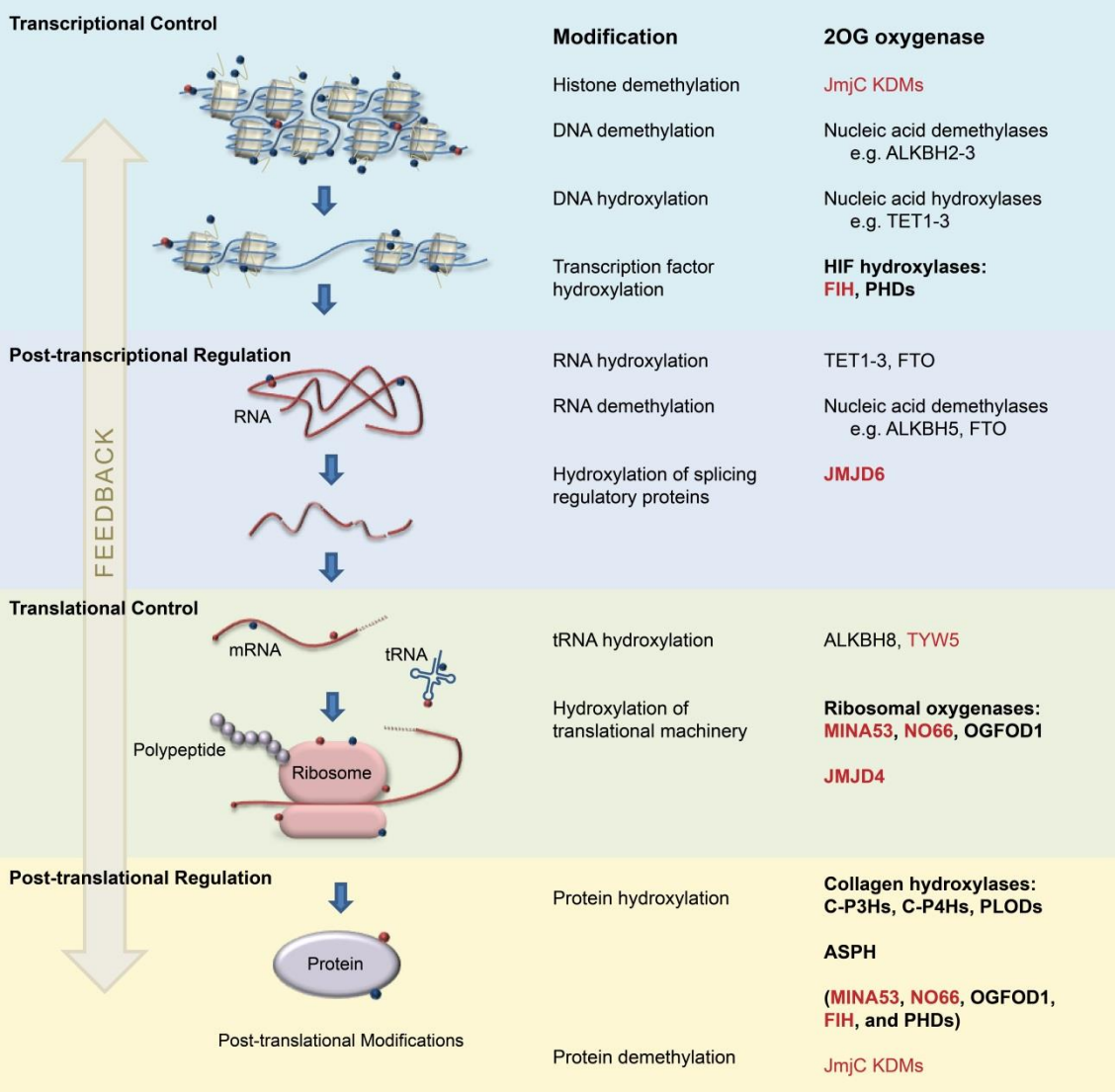


Figure 1.12. 2OG oxygenases involved in protein biosynthesis. 2OG oxygenases in bold catalyse protein hydroxylation. 2OG oxygenases in red are JmjC enzymes. Figure adapted from: Markolovic, S. *et al.* (2015)[11].

1.3.5 Uncharacterised JmjC oxygenases

To date, there are ~20 human JmjC KDMs that have been identified to act on histone lysine residues. While most of these enzymes have been biochemically assigned, several of these assignments have yet to be validated in cells (e.g. JMJD1C, KDM6C, KDM7A, KDM7C and others). The remaining ~10 JmjC oxygenases are hydroxylases or putative hydroxylases based on sequence and structural similarities (**Table 1.1**). At the start of this thesis, there were only three completely uncharacterised 'JmjC-only' oxygenases – HSPBAP1, JMJD7, and JMJD8. Of these, JMJD8 is predicted to have three His Fe(II)-binding residues (UniProt KB: Q96S16),

rather than the characteristic HX(D/E)…H motif, based on sequence alignments of the JmjC domains from the JmjC-only enzymes. Thus, it is possible that JMJD8 cannot bind Fe(II) efficiently and consequently may be a 'pseudo' enzyme, that is an enzyme that shares several features with the JmjC hydroxylases, but that is catalytically inactive. Other examples of potential 'pseudo' enzymes include the first identified JmjC domain-containing protein, JARID2 (predicted to have only one conserved His residue) and KDM7C (HXC…Y)[12,138]. Unlike JMJD8, the uncharacterised JmjC proteins, HSPBAP1 and JMJD7, both contain the typical Fe(II)-binding motif, and are therefore more likely to be catalytically active enzymes. Given that both HSPBAP1 and JMJD7 have only an identifiable JmjC domain (i.e. they do not contain DNA/chromatin-binding domains typically found in the JmjC KDMs), they are likely to have hydroxylase activity.

1.4 Aims of this thesis

At the onset of my work, there were only a few remaining JmjC proteins with completely uncharacterised enzymatic functions. Given the fundamental roles that many 2OG oxygenases, and particularly the JmjC oxygenases, play in oxygen sensing and protein biosynthesis, the biochemical, structural, and functional characterisation of these enzymes is of interest. JMJD7 is one such enzyme for which there are no reports on its biochemical function. Nevertheless, JMJD7 has been linked with major human diseases including cancer and autism[139–141] and is likely of biological (and potentially medicinal) importance. Thus, detailed biochemical and structural studies were undertaken to identify the biochemical and functional roles of human JMJD7.

The work described in **Chapters 2-4** comprises the biochemical and structural characterisation of JMJD7, which aims to lay the foundation for future studies of JMJD7 in normal physiology and disease. The specific aims of this work are to:

(1) Determine whether JMJD7 is a functional 2OG oxygenase;

(2) Assign the catalytic activity of JMJD7;

(3) Identify a cellular substrate;

(4) Solve the crystal structure of human JMJD7; and

(5) Investigate the effect(s) of clinically-relevant JMJD7 variants.

Chapter 5 describes a related, well-characterised JmjC hydroxylase, JMJD4. While JMJD4 has been reported to act on eRF1 and consequently improve translation termination efficiency, the molecular details underlying this observed functional effect were unknown at the start of this project. Thus, the work described in **Chapter 5** details the biophysical studies of eRF1 with the aim of gaining molecular insights into the effect of hydroxylation on eRF1 (1) structure; (2) dynamics, and (3) stability. This chapter further details the identification and characterisation of a previously unreported JMJD4-catalysed PTM.

2 BIOCHEMICAL CHARACTERISATION OF JUMONJI DOMAIN-CONTAINING 7 (JMJD7)

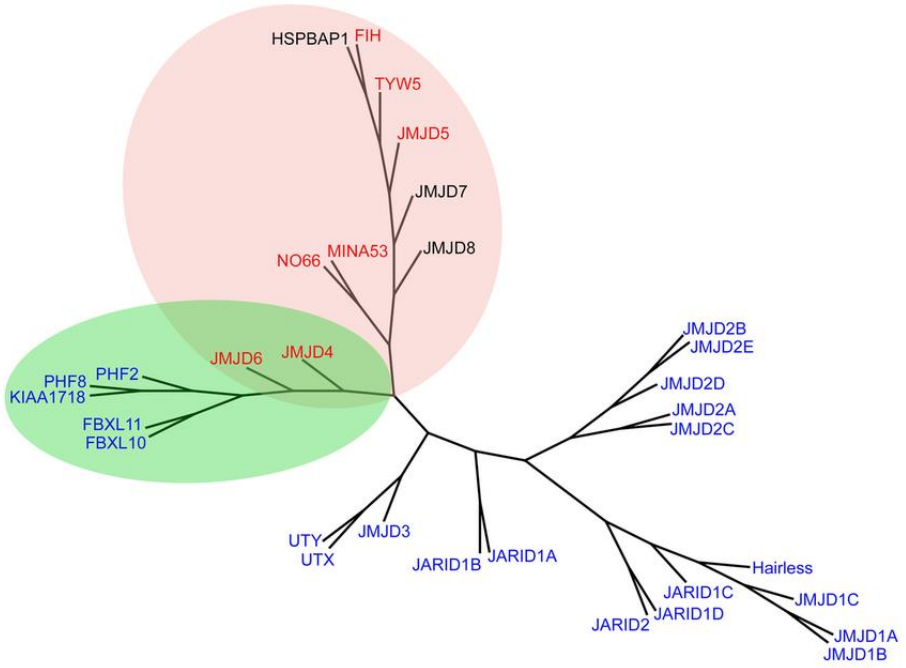
2.1 Introduction

** The majority of the biochemical studies on JMJD7 described in* ***Chapter 2*** *has been prepared in a modified form as a co-first author manuscript, which is currently under review for publication in* Nature Chemical Biology*: Markolovic, S. and Zhuang, Q. et al. (2017) (3S)-Lysyl hydroxylation of TRAFAC GTPases is catalyzed by the human Jumonji-C oxygenase JMJD7.*

2.1.1 Identification of JMJD7 as a putative 2OG oxygenase

Given the important roles that the Jumonji-C (JmjC) 2OG oxygenases play in fundamental cellular processes such as collagen biosynthesis, oxygen sensing, fatty acid metabolism, and protein biosynthesis[2,11] (described in **Chapter 1**), it is of intense interest to determine the functions of the few remaining uncharacterised human 2OG oxygenases, amongst which is JMJD7. The first identification of JMJD7 as a putative metalloprotein can be traced to the 2001 bioinformatic study by Clissold and Ponting, which identifies the cupin metalloenzyme-like JmjC domain-containing proteins in a range of eukaryotes[82]. In this study, JMJD7 is identified as a putative cytosolic phospholipase A2β protein containing a *C*-terminally truncated JmjC domain. Subsequent biochemical and structural studies by Hewitson *et al*. (2002) identified FIH as a JmjC enzyme belonging to the 2OG oxygenase superfamily, and by implication, other related members[15]. A 2006 bioinformatic study by Zhang and colleagues describing the evolutionary relationship of JmjC domain-containing proteins similarly identified JMJD7 as an alternatively spliced form of phospholipase A2β (phospholipase A2 group IVB, PLA2G4B)[12]. Given that complete sequencing of the human genome was not completed until 2003[142] and annotations of the human genome are on-going, it was later revealed that the *JMJD7* gene exists alone (NCBI Gene ID: 100137047[143]) and as a read-through transcript with *PLA2G4B* (NCBI Gene ID: 8681[143]). According to the National Center for Biotechnology Information

(NCBI)[143], it has been proposed that the *JMJD7-PLA2G4B* read-through transcript encodes a fusion protein combining the *N*-terminus of JMJD7 (including part of its JmjC domain) with a nearly complete coding region of PLA2G4B.



Gene	Protein name	UniProtKB accession number
FBXL11/JHDM1A/KDM2A	F-box and leucine-rich repeat protein 11	Q9Y2K7
FBXL10/JHDM1B/KDM2B	F-box and leucine-rich repeat protein 10	Q8NHM5
JMJD1A/KDM3A	Jumonji domain containing 1A	Q9Y4C1
JMJD1B/KDM3B	Jumonji domain containing 1B	Q7LBC6
JMJD1C/JHD2C	Jumonji domain containing 1C	Q15652
JMJD2A/JHDM3A/KDM4A	Jumonji domain containing 2A	O75164
JMJD2B/KDM4B	Jumonji domain containing 2B	O94953
JMJD2C/KDM4C	Jumonji domain containing 2C	Q9H3R0
JMJD2D/KDM4D	Jumonji domain containing 2D	Q6B0I6
JMJD2E/KDM4E	Jumonji domain containing 2E	B2RXH2
JARID1A/RBBP2/KDM5A	Jumonji, AT rich interactive domain 1A	P29375
JARID1B/KDM5B	Jumonji, AT rich interactive domain 1B	Q9UGL1
JARID1C/KDM5C	Jumonji, AT rich interactive domain 1C	P41229
JARID1D/KDM5D	Jumonji, AT rich interactive domain 1D	Q9BY66
UTX/KDM6A	Ubiquitously transcribed tetratricopeptide repeat protein, chromosome X	O15550
UTY	Ubiquitously transcribed tetratricopeptide repeat protein, chromosome Y	O14607
JMJD3/KDM6B	Jumonji domain containing 3	O15054
K1AA1718/ JHDM1D/KDM7	Lysine-specific demethylase 7	Q6ZMT4
PHF8/KIAA1111	PHD finger protein 8	Q9UPP1
PHF2/JHDM1E	PHD finger protein 2	O75151
JMJD4	Jumonji domain containing 4	Q9H9V9
JMJD5	Jumonji domain containing 5	Q8N371
JMJD6	Jumonji domain containing 6	Q6NYC1
JMJD7	Jumonji domain containing 7	P0C870
JMJD8	Jumonji domain containing 8	Q96S16
JARID2/JMJ	Jumonji/ARID domain-containing protein 2/ Protein Jumonji	Q92833
TYW5	tRNA wybutosine-synthesizing protein 5	A2RUC4
FIH	Factor inhibiting HIF	Q9NWT6
HSPBAP1/PASS1	HSPB1-associated protein 1	Q96EW2
Hairless	Protein hairless	O43593
MINA53	MYC induced nuclear antigen	Q8IUF8
NO66	Nucleolar protein 66	Q9H6W3

Figure 2.1. Parsimony tree analysis of the human JmjC subfamily of 2OG oxygenases showing divergent branches. JmjC hydroxylases (red), demethylases (blue), and hydroxylases/demethylases (green) are colour-coded. Genes encoding proteins with uncharacterised functions (including JMJD7) are in black. Figure reprinted with permission from Nature Publishing Group (See Appendix C for permissions): Chowdhury, R. *et al.* (2016)[102].

Even so, Clissold, Ponting, Zhang, and colleagues[12,82] identified that the putative JMJD7-PLA2G4B fusion protein contains only a recognisable JmjC protein domain within its *N*-terminal region, a feature shared by other JmjC-only 2OG oxygenases (e.g. FIH, MINA53, NO66, JMJD4, and HSPBAP1)[12,15]. More recent bioinformatic analyses of the JmjC oxygenases similarly identify JMJD7 to be most closely related to the JmjC-only oxygenases (**Figure 2.1**)[102]; notably, all biochemically characterised JmjC-only oxygenases (FIH[15,68], TYW5[116], MINA53[102], NO66[102], JMJD4[133], JMJD5[144], JMJD6[107]) have reported hydroxylase activity[21]. Beyond the identification and sequence-based assignment of JMJD7 as a putative 2OG oxygenase, there are no reports on its biological or chemical functions.

2.1.2 JMJD7 expression in human cells and tissues

While there are few reports focusing on JMJD7, publically available databases provide some data on JMJD7 gene and protein expression, which may offer insights into its physiological relevance. The gene annotation portal, BioGPS[145], provides information on *JMJD7* gene expression in a range of human cells. A comparison of *JMJD7* gene expression levels in different human cell types indicates that *JMJD7* is more highly expressed in immune cells as compared to other cells types, with expression levels elevated at least 3-fold above the median in CD4+ T-cells, CD56+ NK (natural killer) cells, and CD14+ monocytes (**Figure 2.2**). These observations implicate a potential role for JMJD7 in the immune response.

The Human Protein Atlas[146] is a publically-available database that provides information on gene expression in normal cells as well as protein expression profiles in a range of normal and cancer tissues. Protein expression is reported for JMJD7 in a variety of normal tissue types (**Figure 2.3A**). However, this data should be interpreted with caution, as the antibody (HPA005726) used to detect JMJD7 recognises an antigen that is present in both JMJD7 and the JMJD7-PLA2G4B fusion protein. Thus, it is unclear whether the reported expression levels correspond to the wildtype or fusion protein (or both). This is reflected in the observation that

staining for JMJD7 and JMJD7-PLA2G4B is nearly identical for a variety of cancer tissues, likely due to the use of the same antibody (**Figure 2.3B** and data not shown). While the data reported require external validation, they indicate that JMJD7 and JMJD7-PLA2G4B are likely ubiquitously expressed at the protein level across different cell types.

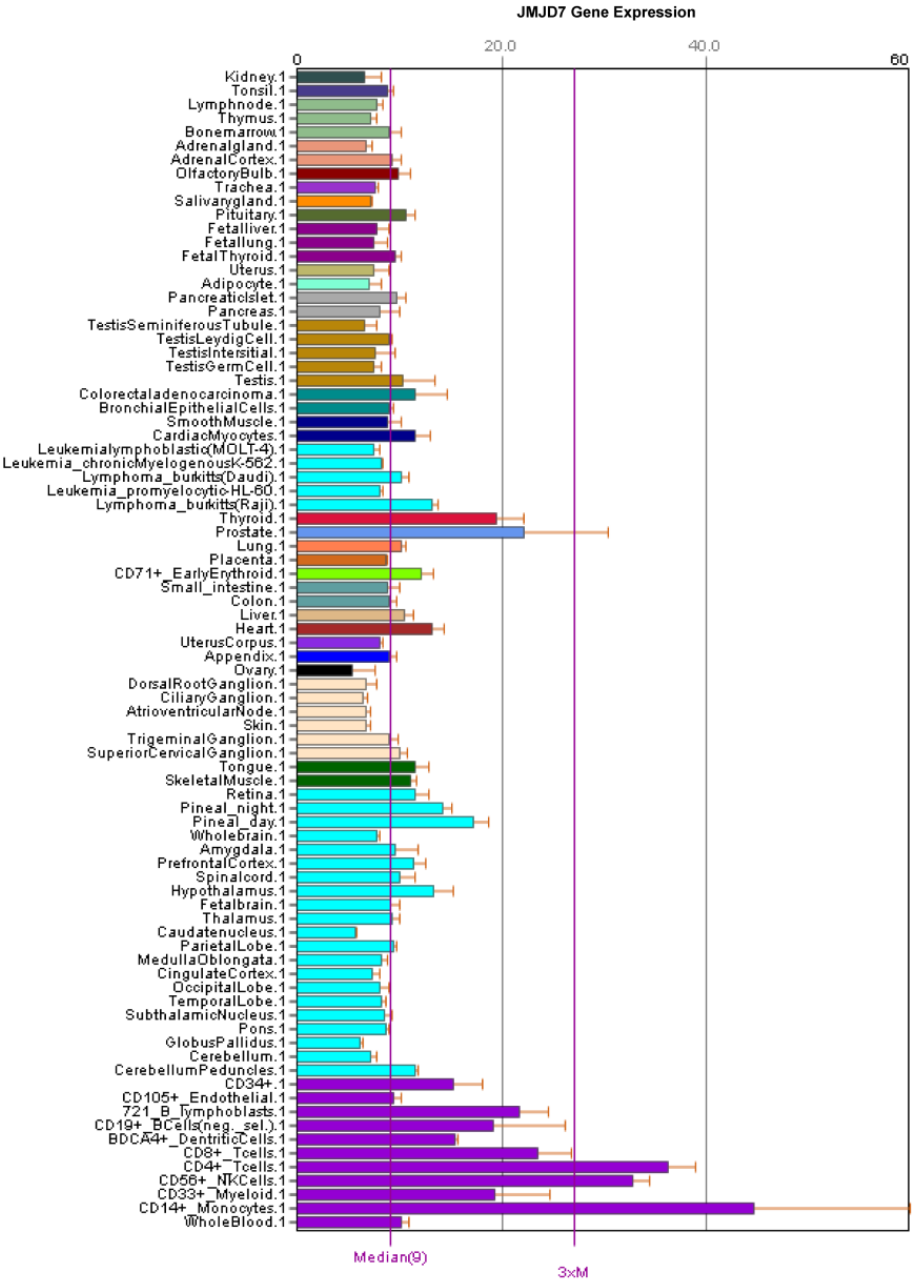
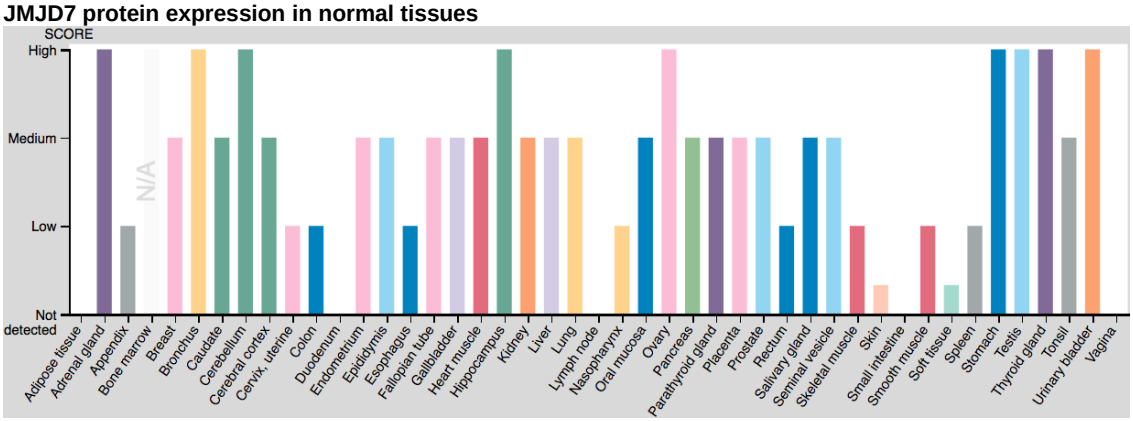


Figure 2.2. JMJD7 gene expression. JMJD7 gene expression data were obtained from and generated by the BioGPS database (www.biogps.org)[145]. The *JMJD7* gene is more highly expressed in immune cells (purple) as compared with other cell types.

JMJD7 protein expression in normal tissues



JMJD7 protein expression in cancer tissues

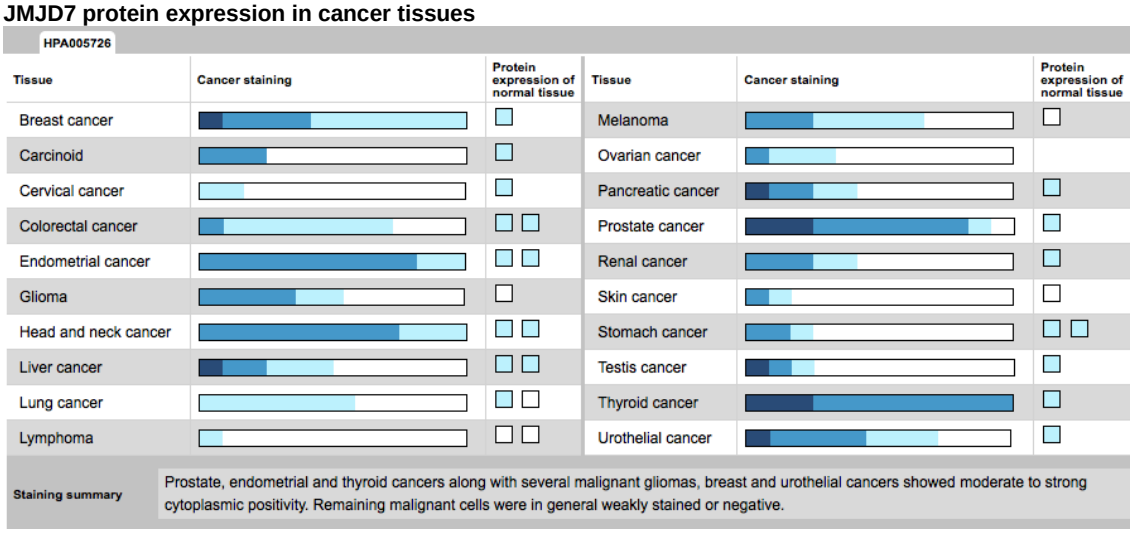


Figure 2.3. JMJD7 protein expression in (top) normal and (bottom) cancer tissues. Figure obtained from and generated by the Human Protein Atlas using the HPA005726 antibody[146].

2.1.3 JMJD7 and disease

Beyond general scientific interests to biochemically characterise JMJD7 and create a more comprehensive picture of human biochemical pathways, the motivation to study JMJD7 is also driven by its links to human disease, especially cancer. Interestingly, pan-cancer analysis using the cBioPortal for Cancer Genomics[147,148], an open-access database of cancer genomics data, indicates that the *JMJD7* gene is deleted in a variety of tumour types (**Figure 2.4A**). Furthermore, patients with tumours that have low *JMJD7* expression or *JMJD7* deletions are associated with poor prognosis in several cancers, including gastric, lung, and colorectal (**Figure 2.4B-E**). These observations are consistent with JMJD7 having apparent tumour suppressor activity. In addition to links with cancer, *JMJD7* mutations have also been reported

in genetic studies in high-risk autism families and individuals with severe intellectual disabilities[139,140].

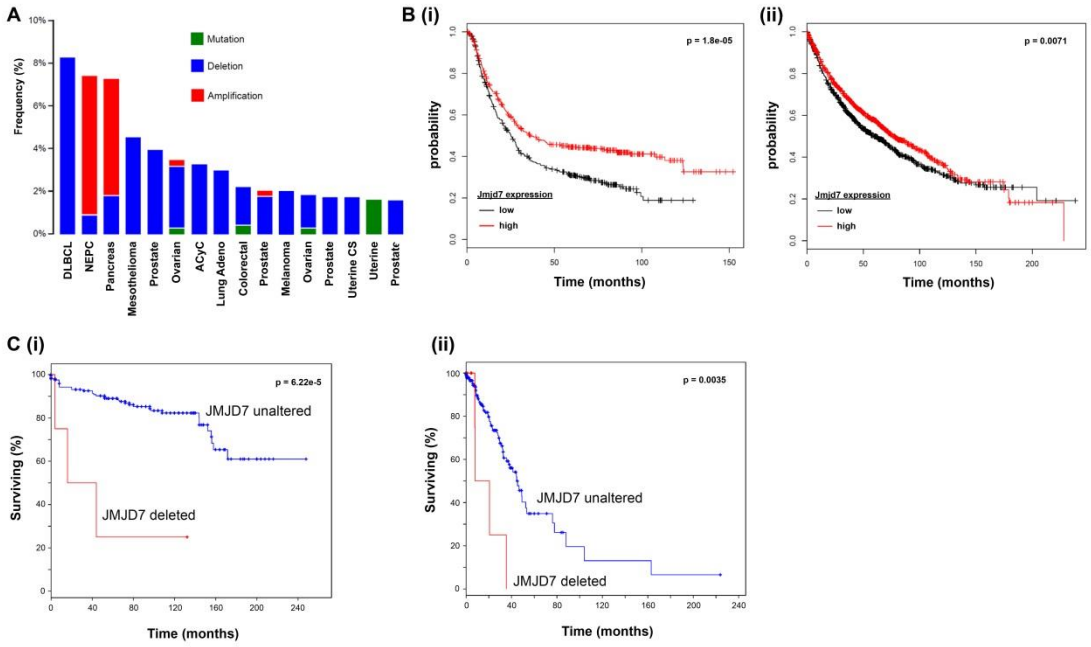


Figure 2.4. JMJD7 is a potential tumour suppressor protein. (A) JMJD7 copy number alteration and mutation across cancers as compiled by cbioportal.org[147]. DLBCL, Diffuse Large B-Cell Lymphoma; NEPC, Neuroendocrine prostate cancer; ACyC, Adenoid Cystic Carcinoma; CS, Carcinosarcoma. ***(B)*** Kaplan-Meier plots showing lower *JMJD7* mRNA expression in tumours from ***(i)*** gastric cancer and ***(ii)*** lung cancer in patients with poor prognosis, compiled using kmplotter (kmplot.com)[149]. ***(C)*** Overall patient survival analysed in ***(i)*** colorectal adenocarcinoma[150] and ***(ii)*** lung adenocarcinoma, using TCGA datasets from cBioPortal.org[147]. Figure kindly provided by Dr Mathew Coleman (University of Birmingham).

2.1.4 Aims of this chapter

Given that the cell biology of all animals is regulated by oxygen, JMJD7 is of interest because it belongs to a family of oxygenases. Furthermore, it is interesting from a medical perspective, as *JMJD7* mutations and deletions have been linked with cancer progression and intellectual disabilities[139,140]. These associations suggest that JMJD7 plays an important role in human disease and therefore may be of therapeutic interest. To understand JMJD7 in a disease and/or cancer setting, it is important to first characterise the chemical and biological roles of JMJD7 in normal physiology.

The work described in this chapter aims to provide a biochemical characterisation of JMJD7 via:

(1) the identification of physiologically-relevant JMJD7 substrate(s);

(2) the biochemical characterisation of JMJD7 catalysis; and

(3) the functional effect(s) of JMJD7 catalysis.

The results presented in this chapter aim to lay the foundation for future JMJD7 biochemical and functional studies in both normal physiology and disease.

2.2 Identification of a JMJD7 substrate

2.2.1 Substrate hunting strategy

To identify a JMJD7 substrate, a peptide screening strategy using matrix-assisted laser desorption/ionisation-time of flight-mass spectrometry (MALDI-TOF-MS) was designed based on analogous approaches that had been successful in identifying substrates for other JmjC hydroxylases, (e.g. MINA53, NO66, and the bacterial oxygenase YcfD[122]). As shown in (**Figure 2.5**), this strategy entailed a bioinformatic search to predict JMJD7 catalytic activity, followed by a co-immunoprecipitation (co-IP) experiment to identify JMJD7-interacting proteins in cells. The approach then required the production of active, recombinant JMJD7 and the synthesis of peptides derived from JMJD7 candidate-interacting proteins to screen for peptides that undergo JMJD7-catalysed oxidation. The synthesis and screening of a tiled array of 20-mer overlapping peptides corresponding to proteins of interest would theoretically minimise false negatives that might arise as an artefact of using linear peptides (i.e. short, linear peptides lack secondary structure elements that might be required for enzyme:substrate recognition).

It was envisaged that peptides would then be screened against JMJD7 for JMJD7-dependent oxidation using MALDI-TOF-MS. MALDI-TOF-MS employs soft ionisation to analyse

biomolecules[151,152]. Samples are prepared by uniform mixing of the peptide with a matrix (e.g. α-cyano-4-hydroxycinnamic acid; CHCA), which promotes vaporisation upon irradiation of the sample with a laser. Upon absorbing UV light, the solid peptide/matrix sample is vaporised and undergoes ionisation. The soft ionisation method is advantageous for analysing biopolymers, which are prone to fragmentation and multiple charge states; however, the results of MALDI-TOF-MS should be interpreted with caution as the chemistry of the sample/matrix ionisation is not well-understood and 'peak artefacts' are sometimes observed. Nevertheless, MALDI-TOF-MS is an extremely efficient method as compared with those required to analyse whole protein substrates, such as intact protein MS (5-10 min/sample) and tandem MS, the latter of which requires proteolytic digestion (16-20 h) prior to MS analysis (10-90 min/sample).

The JMJD7 substrate hunting approach shown in **Figure 2.5** was chosen because it offered two main advantages. The first is that solid-phase peptide synthesis (SPPS) using the Intavis Multipep (**Section 7.5.1**) provides a rapid and efficient way to synthesise up to 48 peptides (≤ 25 amino acids in length) per run; 20-mer peptides spanning an entire protein sequence could be synthesised and ready for screening within a week. Peptide synthesis is an easier and faster alternative to cloning, expressing, and purifying proteins for subsequent screening. The second advantage is that peptides between 1000-4000 Da can be efficiently analysed by MALDI-TOF-MS (~30-60 s/sample) in a high-throughput (96-well) format.

Nonetheless, the peptide screening strategy (**Figure 2.5**) also has potential limitations, including the potential for false positives and false negatives (i.e. instances where the peptide is a substrate, but the full-length protein is not a cellular substrate, and vice versa). False negatives can arise for several reasons, some of which include: (1) the lack of secondary structure found in linear peptides, which may be required for enzyme:substrate recognition; (2) the absence of a structural/sequence motif distal to the target residue of a protein substrate that is not covered in the linear peptide sequence, but is required for enzyme:substrate recognition; and/or (3) lack of other factors present in a cellular context.

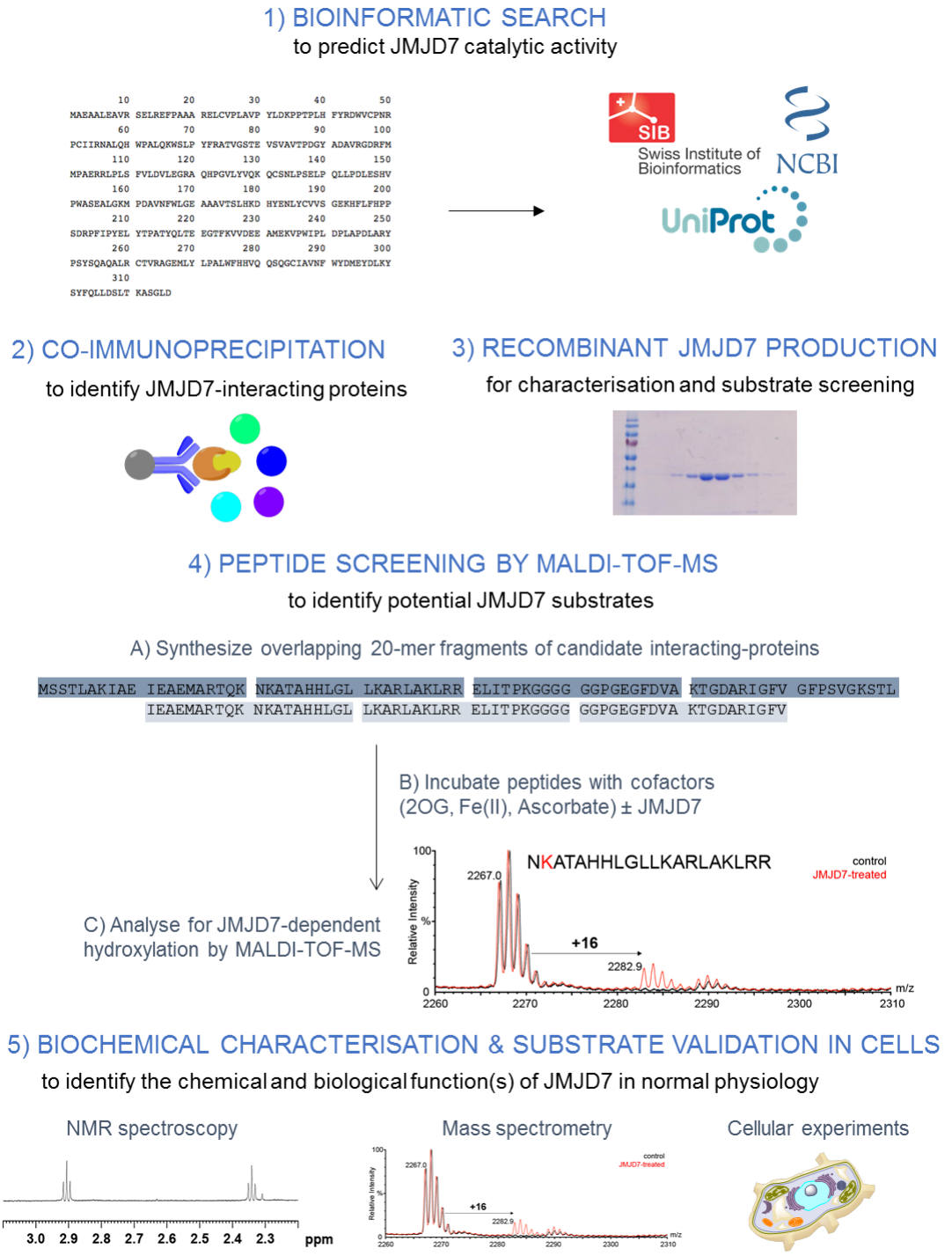


Figure 2.5. JMJD7 peptide screening approach using MALDI-TOF-MS for substrate identification.

Despite these limitations, a similar peptide screening strategy has been successfully used to identify substrates for several JmjC oxygenases, including the ribosomal oxygenases, MINA53, NO66, and the bacterial JmjC hydroxylase, YcfD[122]. The identification of a peptide substrate would then lead to further investigation, biochemical characterisation, and validation using a variety of biochemical and biophysical techniques.

2.2.2 Bioinformatics analysis of JMJD7

A structure-based bioinformatic analysis of JMJD7 using Swiss-Model[153,154] predicts that JMJD7 is a JmjC-domain containing 2OG oxygenase. Among the human JmjC oxygenases, JMJD7 is predicted to be most closely related to FIH[83,84,155], TYW5[103], JMJD5[156,157], and the ribosomal oxygenases (RiOXs)[102], MINA53 and NO66 (**Figure 2.5**). These enzymes all have reported hydroxylase functions (and in some cases, debated demethylase functions) and roles in gene regulation (as described in detail in **Chapter 1**). A sequence alignment of the JmjC domain of JMJD7 with those of the JmjC hydroxylases indicates that JMJD7 contains several conserved residues that are likely of functional importance, including the Fe(II)-binding HXD/E…H triad (H178, D180, and H277) and 2OG-binding lysine residue (K193, **Figure 2.6**), characteristic of most 2OG oxygenases[17,21]. The presence of these residues implies that JMJD7 is likely to be catalytically active, and its predicted structural similarity to the JmjC hydroxylases suggests that it is more likely to function as a hydroxylase, rather than a demethylase (though dual functionality is also a possibility).

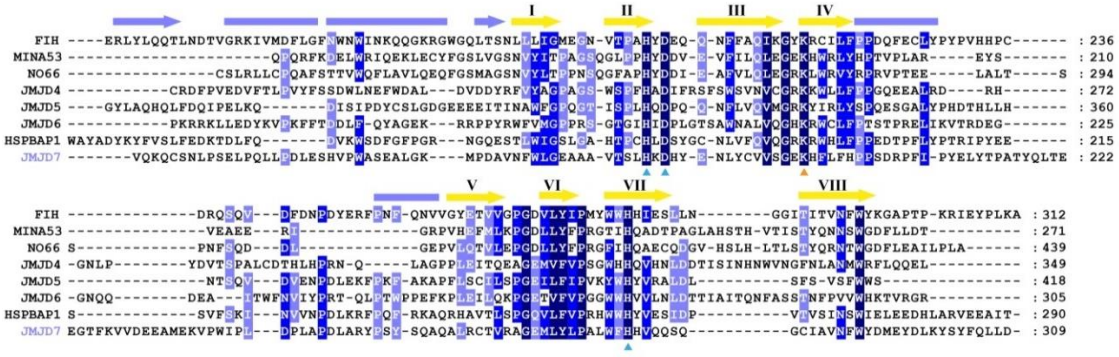


Figure 2.6. Sequence alignment of JMJD7 with the human JmjC-only 2OG oxygenases. Sequences of JmjC domains from the JmjC-only subfamily of human 2OG oxygenases (UniProtKB Entries: FIH, Q9NWT6; MINA53, Q8IUF8; NO66, Q9H6W3; JMJD4, Q9H9V9; JMJD5, Q8N371; JMJD6, Q6NYC1; HSPBAP1, Q96EW2; JMJD7, P0C870) were aligned using Clustal Omega[158] and shaded using GeneDoc. The secondary structure of human FIH (based on crystallographic evidence, PDB ID: 1H2K[84]) is shown above the sequence as rectangles (α-helices) and arrows (β-strands). The eight β-strands (I-VIII) comprising the double-stranded β-helix (DSBH) fold are shown in yellow (arrows). Conserved residues binding Fe(II) (blue) and 2OG (orange) are indicated by triangles.

2.2.3 Identification of proteins that interact with JMJD7

To identify proteins that interact with JMJD7 in cells, a co-immunoprecipitation (co-IP) experiment was performed in HEK293T cells using FLAG-tagged JMJD7 in the presence and absence of dimethyloxalyl glycine (DMOG; co-IP performed by Dr Ming Yang, University of Oxford). DMOG is a cell-permeable 2OG oxygenase inhibitor precursor that is metabolised to the 2OG analogue and oxygenase inhibitor, *N*-oxalylglycine (NOG; **Figure 2.7**)[159]. As described in **Section 1.1.2**, the consensus mechanism for 2OG oxygenase catalysis proceeds via 2OG binding, which is followed by substrate binding to the active site. Oxygen binding then causes formation of the ferryl-oxo [Fe(IV)=O] species, which oxidises the prime substrate and 2OG, causing subsequent release the product, succinate, and CO_2. Thus, the use of a 2OG mimetic allows for the 2OG analogue and substrate to bind the active site, but prevents subsequent coupled co-substrate oxidation and product release from occurring (i.e. without a 2OG-mimetic inhibitor, the co-IP method might not work because genuine substrates would be converted to product and released from JMJD7). Thus, proteins that are not substrates might be expected to non-specifically bind JMJD7 in the presence and absence of DMOG, whereas genuine substrates should be enriched or only bind in the presence of DMOG. Notably, DMOG has been successfully employed to "trap" proteins that interact with 2OG oxygenases, in particular for the JmjC hydroxylases, FIH and JMJD4[133,160].

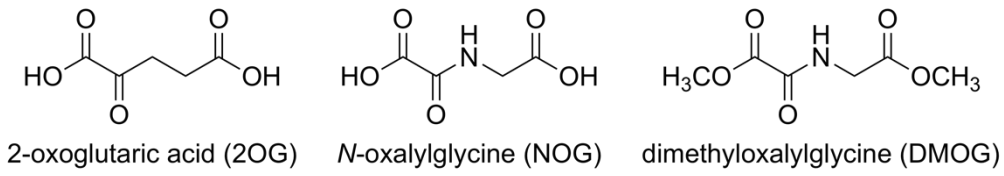


Figure 2.7. Structural comparison of 2OG with the pan-2OG oxygenase inhibitors, NOG and DMOG.

Following analysis of the co-IP by MS/MS peptide sequencing, a list of candidate JMJD7-interacting proteins was generated and ranked according their enrichment in the DMOG-treated cells versus the non-DMOG-treated cells (**Appendix A, Table A1**). Of these, Developmentally-regulated GTP-binding protein 1 (DRG1; UniProtKB Entry: Q9Y295) and

High mobility group protein B2 (HMGB2; UniProtKB Entry: P26583) were identified as top interactors. A protein related to HMGB2, High mobility group protein B1 (HMGB1; UniProtKB Entry: P09429), also showed enhanced binding to JMJD7 in the presence of DMOG, though to a lesser extent than HMGB2.

Overall, the bioinformatic predictions (**Section 2.2.1**), combined with the absence of histones identified to interact with JMJD7 in the co-IP, suggested that JMJD7 is more likely to function as a JmjC hydroxylase rather than a KDM and provided further validation for the substrate hunting approach described in **Section 2.2.1**.

2.2.4 Recombinant expression and purification of JMJD7

The gene encoding full-length human JMJD7 (UniProt ID: P0C870; residues 1-316) was inserted into a pET28a(+) vector for bacterial production with an *N*-terminal His_6-tag (MGSSHHHHHHSSGLVPR*GSH; plasmid cloned and kindly provided by Dr Wei Ge, University of Oxford). pET28a(+)-JMJD7 was transformed into *E. coli* Rosetta(DE3) pLysS cells and produced as described in **Section 7.3.5**. JMJD7 was isolated from bacterial cell lysate by nickel affinity chromatography (**Section 7.4.3**), whereby the electron donating imidazole rings of the His_6-tag bind to the charged Ni(II) resin through coordination bonds[161]. JMJD7 was eluted from the nickel affinity column by applying a step-wise gradient of increasing free imidazole to a final concentration of 250 mM (see **Table 2.1** for details). Protein-containing fractions, identified from a UV trace, were visualised by SDS-PAGE, and the fractions containing JMJD7 were pooled for further purification by size exclusion chromatography (SEC; **Section 7.4.4**). SEC, also known as gel filtration, is a purification technique that separates proteins based on molecular size[162]. Concentrating JMJD7 in a high imidazole concentration buffer promoted protein aggregation and precipitation; therefore, JMJD7-containing fractions post-nickel affinity chromatography were immediately injected onto the size exclusion column without further concentration. This two-step purification method yielded

crystallography-grade enzyme (>95% purity as assessed by SDS-PAGE; **Figure 2.8**). Unless otherwise specified, all purified JMJD7 proteins were produced with an *N*-terminal His_6-tag.

Table 2.1. Nickel affinity chromatography gradient elution method.

Column Volume	Binding Buffer (%)*	Elution Buffer (%)**
1	90	10
2	80	20
3	70	30
4	60	40
7	0	100
32		End Method.

* Binding Buffer = 50 mM HEPES, 500 mM NaCl, 5 mM imidazole, pH 7.5
** Elution Buffer = Binding Buffer + 250 mM imidazole

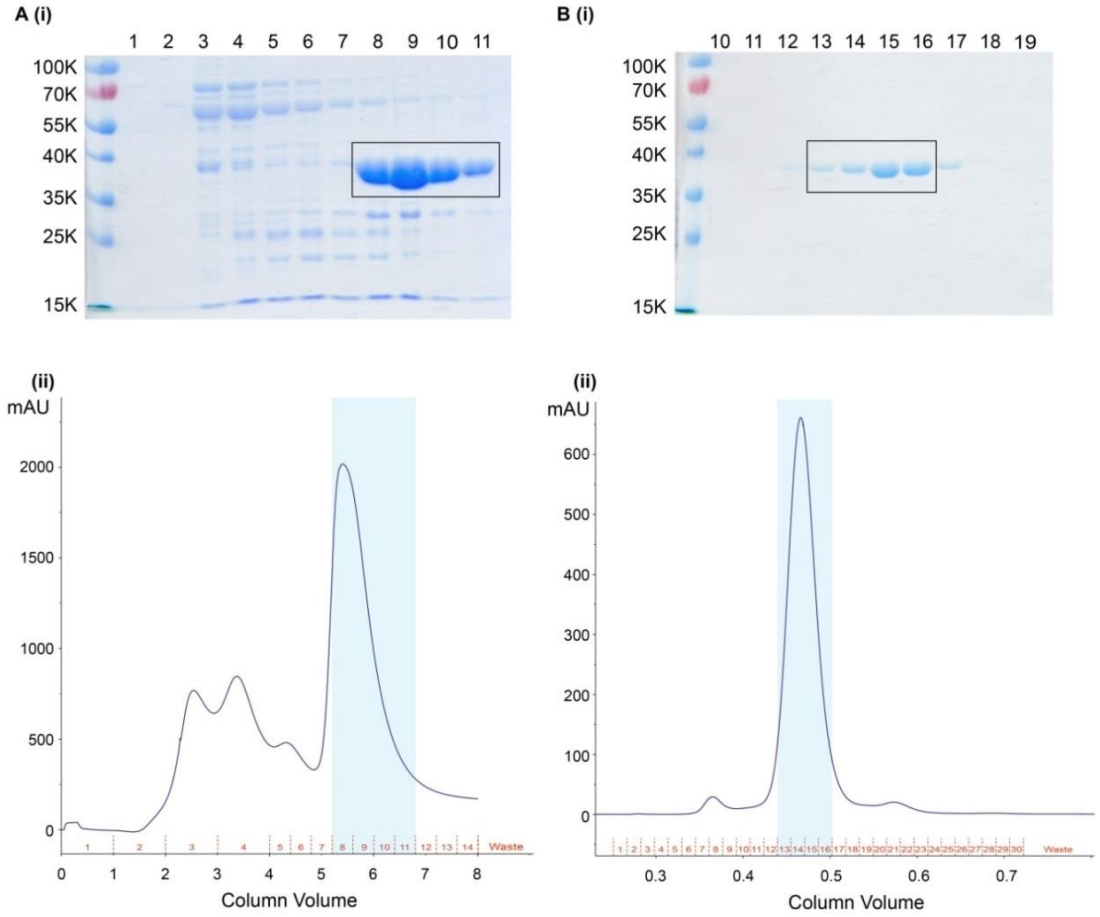


Figure 2.8. Representative two-step purification of JMJD7. Coomassie-stained 12% SDS-PAGE gels with respective elution UV traces are shown for JMJD7 purification by ***(A)*** nickel affinity chromatography and subsequent ***(B)*** SEC. Highlighted regions of ***(A)*** represent JMJD7-containing fractions that were loaded onto the SEC column for further purification. Highlighted regions of ***(B)*** represent purest JMJD7-containing fractions used for subsequent biochemical assays, crystallisation trials, and characterisation.

2.2.5 Recombinant expression and purification of JMJD7-H178A

One characteristic feature of 2OG oxygenases is the Fe(II)-binding triad in the active site (**Figure 1.2**). Typically, substitution of one of the Fe(II)-binding residues ablates or substantially reduces enzymatic activity by hindering Fe(II)-binding within the active site. To generate a catalytically inactive JMJD7 variant, one of the predicted Fe(II)-binding residues (His178, **Figure 2.6**), was substituted with an alanine residue. The pET28a(+)-JMJD7-H178A plasmid was generated via site-directed mutagenesis (**Section 7.2.5**) using custom primers (**Table 7.1**) and the wildtype pET28a(+)-JMJD7 plasmid as a template. JMJD7-H178A was produced and purified exactly as described for wildtype JMJD7 (**Section 2.2.4**).

2.2.6 JMJD7 catalyses uncoupled conversion of 2OG to succinate

In 2OG oxygenase catalysis, substrate oxidation is coupled to the oxidative decarboxylation of 2OG to succinate. In some cases (e.g. collagen prolyl hydroxylases, BBOX, KDMs), 2OG oxygenases can catalyse conversion of 2OG to succinate in the absence of a substrate (i.e. 'uncoupled' turnover)[31,163]. To investigate the potential catalytic activity of recombinant JMJD7, ^{1}H-NMR spectroscopy was employed to determine whether JMJD7 can catalyse uncoupled 2OG turnover (**Section 7.13.2**)[31]. Consistent with 2OG oxygenase catalysis, JMJD7 catalyses the Fe(II)-dependent conversion of 2OG to succinate in the absence of a substrate (**Figure 2.9**). Furthermore, the presence of NOG[159,164], a broad spectrum 2OG oxygenase inhibitor, apparently inhibits JMJD7 activity by reducing JMJD7-catalysed 2OG turnover/succinate formation, indicating the potential viability of JMJD7 as target for inhibition. Substitution of one of the predicted JMJD7 Fe(II)-binding residues (H178) to an alanine ablated 2OG turnover activity. It is worth noting that lack of uncoupled turnover does not necessarily mean that the enzyme is inactive. For example, the rate of uncoupled turnover for PHD2 is very slow[71,165]. However, uncoupled turnover certainly supports the likelihood that

JMJD7 is a functional 2OG oxygenase. Overall, the results show that JMJD7 is a catalytically active enzyme.

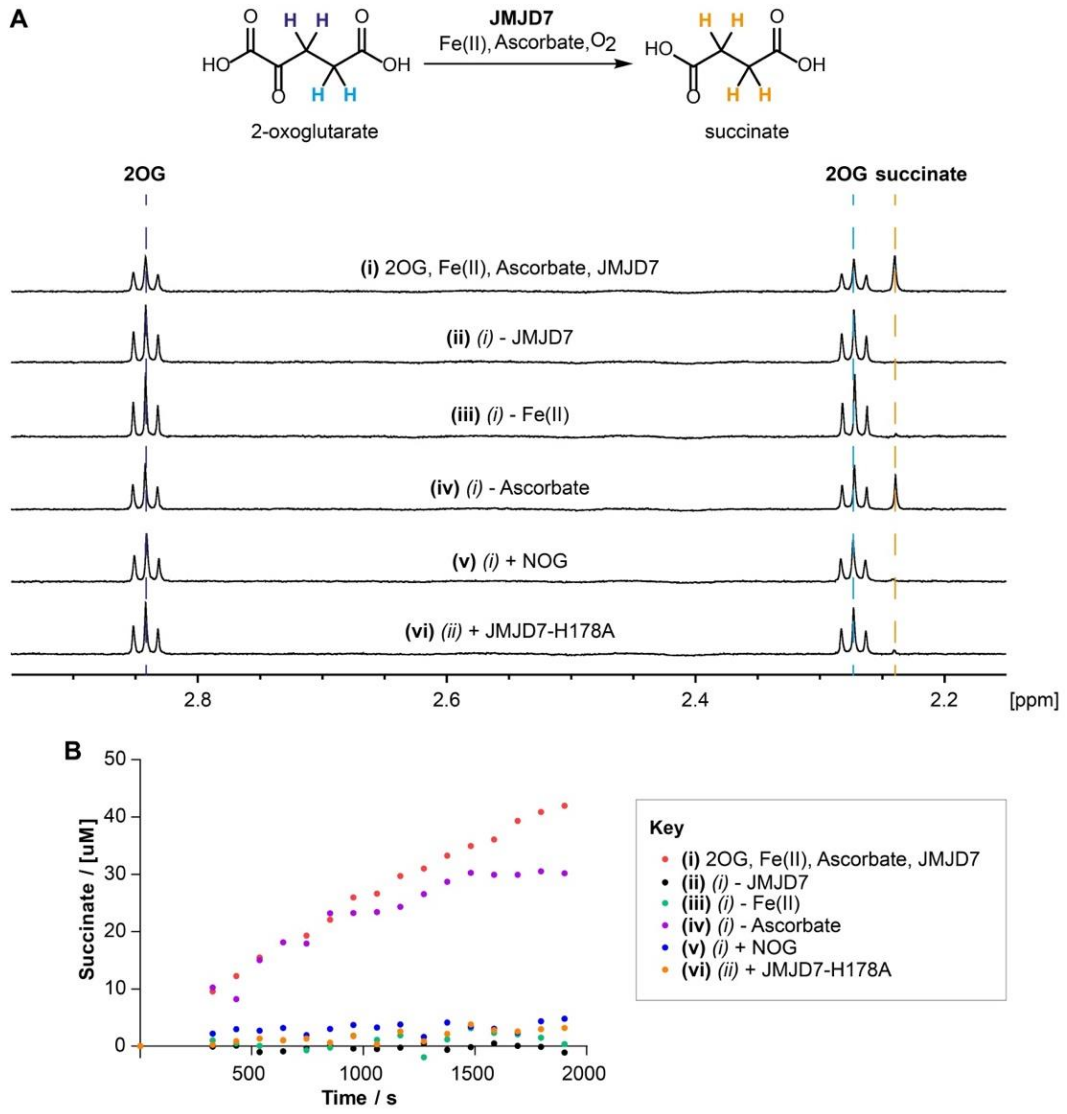


Figure 2.9. JMJD7 catalyses uncoupled conversion of 2OG to succinate. (A) (*Top*) Schematic of proposed uncoupled JMJD7-catalysed turnover of 2OG to succinate. (*Bottom*) Representative end point ^{1}H-NMR spectra of 2OG and succinate following incubations (25°C, 1900 s) in the presence or absence of indicated enzymes, co-factors, or inhibitors: 10 µM JMJD7/JMJD7-H178A, 100 µM Fe(II), 200 µM 2OG, 500 µM ascorbate, 1 mM *N*-oxalylglycine (NOG), 50 mM Tris-d11, pH 7.5. ***(B)*** Plot of succinate formation [µM] over time in the presence or absence of indicated co-factors/substrates or inhibitors, as monitored by ^{1}H-NMR spectroscopy.

2.2.7 Identification of potential JMJD7 substrates via MALDI-TOF-MS-based peptide screening

After validating that recombinantly produced JMJD7 was active, peptide screening by MALDI-TOF-MS (**Section 2.2.1**, **Figure 2.5**) ensued. The peptides shown in **Table 2.2** were synthesised by SPPS and screened for JMJD7-catalysed hydroxylation in the presence of Fe(II), 2OG, and ascorbate using the reaction conditions described in **Section 7.7.1.1**.

Table 2.2. Peptides synthesised and screened against JMJD7. Peptides were synthesised by SPPS with *C*-terminal amides. Red font indicates peptides for which a +16 Da shift was observed.

Peptide	Gene Name	UniProtKB Entry	Peptide Sequence	Modification Observed?
1	HMGB1$_{1-20}$	P09429	MGKGDPKKPRGKMSSYAFFV	**(+)16**
2	HMGB1$_{21-40}$	P09429	QTCREEHKKKHPDASVNFSE	None
3	HMGB1$_{41-60}$	P09429	FSKKCSERWKTMSAKEKGKF	None
4	HMGB1$_{61-80}$	P09429	EDMAKADKARYEREMKTYIP	None
5	HMGB1$_{81-100}$	P09429	PKGETKKKFKDPNAPKRPPS	None
6	HMGB1$_{101-120}$	P09429	AFFLFCSEYRPKIKGEHPGL	None
7	HMGB1$_{121-140}$	P09429	SIGDVAKKLGEMWNNTAADD	None
8	HMGB1$_{141-160}$	P09429	KQPYEKKAAKLKEKYEKDIA	None
9	HMGB1$_{161-180}$	P09429	AYRAKGKPDAAKKGVVKAEK	None
10	HMGB2$_{1-20}$	P26583	MAKGDPKKPKGKMSAYAFFV	**(+)16**
11	DRG1$_{1-20}$	Q9Y295	MSSTLAKIAEIEAEMARTQK	None
12	DRG1$_{21-40}$	Q9Y295	NKATAHHLGLLKARLAKLRR	**(+)16**
13	DRG1$_{41-60}$	Q9Y295	ELITPKGGGGGGPGEGFDVA	None
14	DRG1$_{61-80}$	Q9Y295	KTGDARIGFVGFPSVGKSTL	None
15	DRG1$_{81-100}$	Q9Y295	LSNLAGVYSEVAAYEFTTLT	Not tested
16	DRG1$_{101-120}$	Q9Y295	TVPGVIRYKGAKIQLLDLPG	None
17	DRG1$_{121-140}$	Q9Y295	IIEGAKDGKGRGRQVIAVAR	None
18	DRG1$_{141-160}$	Q9Y295	TCNLILIVLDVLKPLGHKKI	None
19	DRG1$_{161-180}$	Q9Y295	IENELEGFGIRLNSKPPNIG	None
20	DRG1$_{181-200}$	Q9Y295	FKKKDKGGINLTATCPQSEL	None
21	DRG1$_{201-220}$	Q9Y295	DAETVKSILAEYKIHNADVT	None
22	DRG1$_{221-240}$	Q9Y295	LRSDATADDLIDVVEGNRVY	Not tested
23	DRG1$_{241-260}$	Q9Y295	IPCIYVLNKIDQISIEELDI	Not tested
24	DRG1$_{261-280}$	Q9Y295	IYKVPHCVPISAHHRWNFDD	None
25	DRG1$_{281-300}$	Q9Y295	LLEKIWDYLKLVRIYTKPKG	None
26	DRG1$_{301-320}$	Q9Y295	QLPDYTSPVVLPYSRTTVED	None
27	DRG1$_{321-340}$	Q9Y295	FCMKIHKNLIKEFKYALVWG	None
28	DRG1$_{341-360}$	Q9Y295	LSVKHNPQKVGKDHTLEDED	None

As indicated in **Table 2.2**, a 20-mer peptide (12) derived from DRG1 (DRG1$_{21-40}$; NKATAHHLGLLKARLAKLRR-NH$_2$) was found to undergo a +16 Da shift in the presence of JMJD7, consistent with the incorporation of one oxygen atom, i.e. a single hydroxylation (**Figure 2.10**). Notably, DRG1 was among the top JMJD7-interacting proteins identified in the

co-IP (**Section 2.2.3**; **Appendix A, Table A1**). Note that two other 20-mer peptides, corresponding to HMGB1 peptide (1) and HMGB2 peptide (10), were also observed to undergo a +16 Da shift in the presence of JMJD7. The +16 Da shift observed for HMGB1 and HMGB2 was identified to be methionine oxidation, a common artefact observed in 2OG oxygenase-catalysed hydroxylation assays due to the oxidising environment. However, preliminary experiments suggest that methionine oxidation is JMJD7-dependent in this case, as evidenced by the work on JMJD7-catalysed methionine oxidation summarised in **Appendix B**.

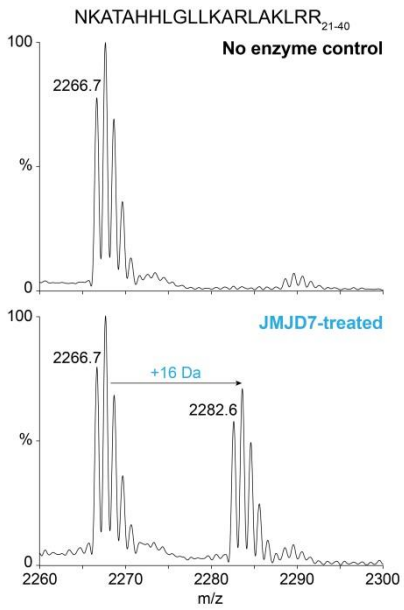


Figure 2.10. A peptide derived from DRG1 undergoes JMJD7-catalysed modification. A 20-mer peptide derived from $DRG1_{21-40}$ (NKATAHHLGLLKARLAKLRR-NH_2) undergoes a JMJD7-dependent +16 Da mass shift. Incubation conditions: 37°C, 1 h, 50 μM peptide, ±10 μM JMJD7, 100 μM Fe(II), 200 μM 2OG, 500 μM ascorbate, 50 mM HEPES, pH 7.5.

2.3 Biochemical characterisation of JMJD7 as a (*3S*)-lysyl hydroxylase

2.3.1 JMJD7 catalyses hydroxylation of residue K22 of DRG1

To identify the target residue, overlapping 20-mer peptides ($DRG1_{11-30}$ and $DRG1_{31-50}$) were synthesised by SPPS and assayed for JMJD7-catalysed hydroxylation (**Figure 2.11**). However, neither of these peptides were substrates for JMJD7, indicating that residues 21-40 likely play an important role in enzyme:substrate recognition.

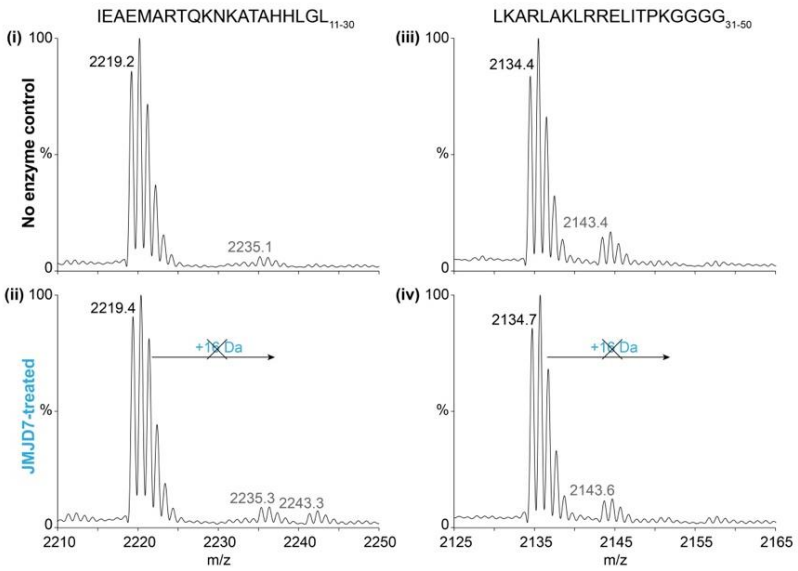


Figure 2.11. MALDI-TOF mass spectra of DRG1$_{11-30}$ and DRG1$_{31-50}$, incubated in the absence (top) and presence (bottom) of JMJD7, indicate that these peptides are not JMJD7 substrates. See **Figure 2.10** for incubation conditions.

Alanine scanning was therefore employed on the DRG1$_{21-40}$ peptide substrate to identify the target residue. Alanine scanning is a technique that can be used to identify residues that are important or required for enzymatic activity by systematically substituting each residue to an alanine. 20-mer peptides with an alanine substitution at each position of DRG1$_{21-40}$ were synthesised and screened for JMJD7-catalysed hydroxylation (**Table 2.3**).

The results revealed that only the K22A peptide (corresponding to alanine substitution of the lysine at peptide position 2) was not a substrate for hydroxylation, suggesting that K22 is specifically hydroxylated by JMJD7 (**Figure 2.12**). However, it is worth noting that alanine scanning typically cannot provide absolute evidence for the identification of the target residue. As shown in **Figure 2.12B**, substitution of H26, L28, G29, and K32 to an alanine residue results in a substantial reduction in observed hydroxylation. While it is possible that JMJD7 can hydroxylate more than one lysine residue within the peptide sequence (which has been reported for the JmjC hydroxylase JMJD6 *in vitro*[105]), this is unlikely for two reasons: (1) only a +16 Da modification (and no +32 Da modification) was observed in the JMJD7-treated wildtype peptide mass spectrum; and (2) the extent of hydroxylation observed for the variant peptides

was confidently above the limits of detection for MALDI-TOF-MS. By contrast, no hydroxylation was observed for the K22A peptide within the limits of detection (**Figure 2.12A**). Therefore, the reduction of hydroxylation observed for the H26A, L28A, G29A, and K32A variants is likely due to the importance of these residues for enzyme:substrate recognition and/or binding. Note that because these peptides were not purified prior to screening, alanine scanning serves only as a qualitative measure of hydroxylation (i.e. hydroxylation either is or is not observed).

Table 2.3. DRG1$_{21-40}$ alanine variant peptides.

Peptide	Sequence (-NH_2)*	Calculated Mass (Da)**
DRG1$_{21-40}$	NKATAHHLGLLKARLAKLRR	2266.75
N1A	AKATAHHLGLLKARLAKLRR	2223.73
K2A	NAATAHHLGLLKARLAKLRR	2209.66
T4A	NKAAAHHLGLLKARLAKLRR	2236.73
H6A	NKATAAHLGLLKARLAKLRR	2200.69
H7A	NKATAHALGLLKARLAKLRR	2200.69
L8A	NKATAHHAGLLKARLAKLRR	2224.67
G9A	NKATAHHLALLKARLAKLRR	2280.78
L10A	NKATAHHLGALKARLAKLRR	2224.67
L11A	NKATAHHLGLAKARLAKLRR	2224.67
K12A	NKATAHHLGLLAARLAKLRR	2209.66
R14A	NKATAHHLGLLKAALAKLRR	2181.64
L15A	NKATAHHLGLLKARAAKLRR	2224.67
K17A	NKATAHHLGLLKARLAALRR	2209.66
L18A	NKATAHHLGLLKARLAKARR	2224.67
R19A	NKATAHHLGLLKARLAKLAR	2181.64
R20A	NKATAHHLGLLKARLAKLRA	2181.64

* Alanine variant peptides were synthesised (as *C*-terminal amides) and assayed for JMJD7-catalysed hydroxylation.
** Average masses calculated using http://www.peptidesynthetics.co.uk/tools/

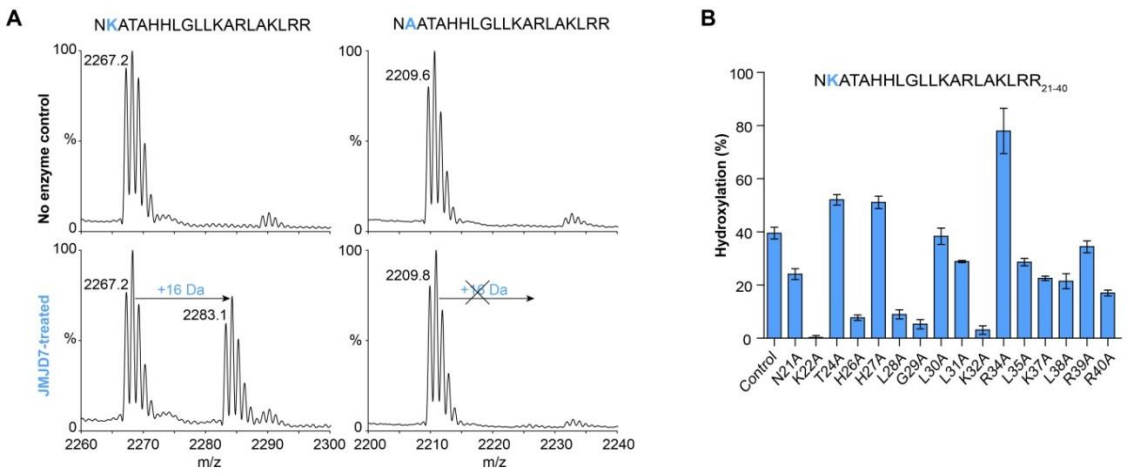


Figure 2.12. JMJD7 catalyses lysyl hydroxylation at position 22 of DRG1$_{21-40}$. **(A)** MALDI-TOF mass spectra corresponding to (left) the wildtype DRG1 peptide (residues 21-40) and (right) the analogous K22A DRG1 variant peptide, assayed in the absence (black) or presence (blue) of JMJD7. Substitution of K22 for an alanine residue results in complete loss of DRG1 peptide hydroxylation in the presence of JMJD7. **(B)** The substituted peptide resulting in ablation of hydroxylation is the K22A variant, suggesting K22 is hydroxylated. Data represent mean ± s.d. (n = 3). See **Table 2.3** for peptide sequences. Note that because of variations in peptide purity, the extent of hydroxylation between variants is a qualitative measurement. Incubation conditions described in **Figure 2.10**.

2.3.2 DRG1 peptide substrate optimisation for use as a biochemical probe

Upon identifying the target hydroxylated residue, the importance of the position of the K22 residue within the peptide sequence was investigated. As described in **Section 2.3.1**, placing the target residue in the centre of the peptide (position 12 of 20; DRG1$_{16-35}$) resulted in no JMJD7-catalysed hydroxylation (**Figure 2.11**). Truncations to the *C*-terminal end of the DRG1$_{21-40}$ peptide (**Table 2.4**) similarly ablated JMJD7-catalysed hydroxylation, suggesting that these residues were likely involved in enzyme:substrate recognition (i.e. no hydroxylation was observed for the peptides listed in **Table 2.4** within the limits of MALDI-TOF-MS detection).

Though no structure is available for human DRG1, a crystal structure of the DRG1 yeast homologue from *Saccharomyces cerevisiae* (Ribosome-interacting GTPase 1; RBG1) has been reported[166] (**Figure 2.13**). A sequence alignment of RBG1 (UniProtKB Entry: P39729) with human DRG1 using Clustal Omega[158] indicates that the proteins share 66% sequence identity, including the target DRG1 K22 residue. In fact, a sequence alignment of human DRG1 with

earlier eukaryotic organisms [identified using a Protein Basic Local Alignment Search Tool (BLAST)[143] search; **Section 7.14**] reveals that the target K22 residue and surrounding amino acids, particularly those immediately *N*-terminal to K22, are highly conserved (**Figure 2.14A**). A crystal structure of RBG1 reveals that the K22 side chain protrudes from a loop on a highly conserved region of a helix-turn-helix (HTH) motif. Strikingly, the position of K22 within the HTH resembles that of the target lysyl hydroxylation site of the eukaryotic release factor eRF1, the substrate for a related JmjC lysyl hydroxylase, JMJD4[133] (**Figure 2.14B-C**). The distinct position of K22 protruding from a turn between two α-helices suggests that secondary structure (e.g. a tight turn or α-helical recognition motif) may be important in substrate recognition by JMJD7, as has been observed for JMJD4/eRF1[133] and FIH with its ARD protein substrates[167].

Table 2.4. DRG1 peptides screened for hydroxylation optimisation.

Peptide	Sequence (-NH_2)*	Calculated Mass (Da)**	Substrate?
$DRG1_{21-40}$ †	NKATAHHLGLLKARLAKLRR†	2266.75	Yes
$DRG1_{21-36}$	NKATAHHLGLLKARLA	1713.04	No
$DRG1_{21-34}$	NKATAHHLGLLKAR	1528.81	No
$DRG1_{21-32}$	NKATAHHLGLLK	1301.54	No
$DRG1_{20-35}$	KNKATAHHLGLLKARL	1770.14	No
$DRG1_{19-34}$	QKNKATAHHLGLLKAR	1785.11	No

* Peptides were synthesised (with *C*-terminal amides) and screened for JMJD7-catalysed hydroxylation. ** Average masses calculated using http://www.peptidesynthetics.co.uk/tools/
† ~40% hydroxylation observed for wildtype DRG1 20-mer; no hydroxylation was observed for other peptides listed within the limits of MALDI-TOF-MS detection.

```
DRG1       MSSTLAKIAEIEAEMARTQKNKATAHHLGLLKARLAKLRRELITPK-GGGGGGPGEGFDV
RBG1       MSTTVEKIKAIEDEMARTQKNKATSFHLGQLKAKLAKLRRELLTSASSGSGGGAGIGFDV
           **:*: **  ** *********:.*** ***:********:*   .*.*** * ****

DRG1       AKTGDARIGFVGFPSVGKSTLLSNLAGVYSEVAAYEFTTLTTVPGVIRYKGAKIQLLDLP
RBG1       ARTGVASVGFVGFPSVGKSTLLSKLTGTESEAAEYEFTTLVTVPGVIRYKGAKIQMLDLP
           *:** * :****************:*:*. **.* ******.**************:****

DRG1       GIIEGAKDGKGRGRQVIAVARTCNLILIVLDVLKPLGHKKIIENELEGFGIRLNSKPPNI
RBG1       GIIDGAKDGRGRGKQVIAVARTCNLLFIILDVNKPLHHKQIIEKELEGVGIRLNKTPPDI
           ***:*****:***:***********::*:*** *** **:***:****.*****..**:*

DRG1       GFKKKDKGGINLTATCPQSELDAETVKSILAEYKIHNADVTLRSDATADDLIDVVEG-NR
RBG1       LIKKKEKGGISITNTVPLTHLGNDEIRAVMSEYRINSAEIAFRCDATVDDLIDVLEASSR
            :***:****.:* * * :.*. : ::::::**:*:.*::::*.***.******:*. .*

DRG1       VYIPCIYVLNKIDQISIEELDIIYKVPHCVPISAHHRWNFDDLLEKIWDYLKLVRIYTKP
RBG1       RYMPAIYVLNKIDSLSIEELELLYRIPNAVPISSGQDWNLDELLQVMWDRLNLVRIYTKP
            *:*.********.:*****:::*::*:.****: : **:*:**: :** *:********

DRG1       KGQLPDYTSPVVLPYSRTTVEDFCMKIHKNLIKEFKYALVWGLSVKHNPQKVGKDHTLED
RBG1       KGQIPDFTDPVVLRSDRCSVKDFCNQIHKSLVDDFRNALVYGSSVKHQPQYVGLSHILED
           ***:**:*.****  .* :*:*** :***.*:.:*: ***:* ****:** ** .* ***

DRG1       EDVIQIVKK
RBG1       EDVVTILKK
           ***: *:**
```

Figure 2.13. Sequence alignment of H. sapiens DRG1 with S. cerevisiae RBG1. A sequence alignment of DRG1 (UniProtKB Entry: Q9Y295) with RBG1 (UniProtKB Entry: P39729) using Clustal Omega[158] indicates that the proteins share 66% sequence identity, and 75% sequence identity within the HTH domain (residues M1 – T44) of RBG1 (blue). The GTP-binding (orange), S5D2L (light blue), and TGS (pink) domains of RBG1 are highlighted. The target K22 residue of DRG1 is conserved in RBG1 (red highlight).

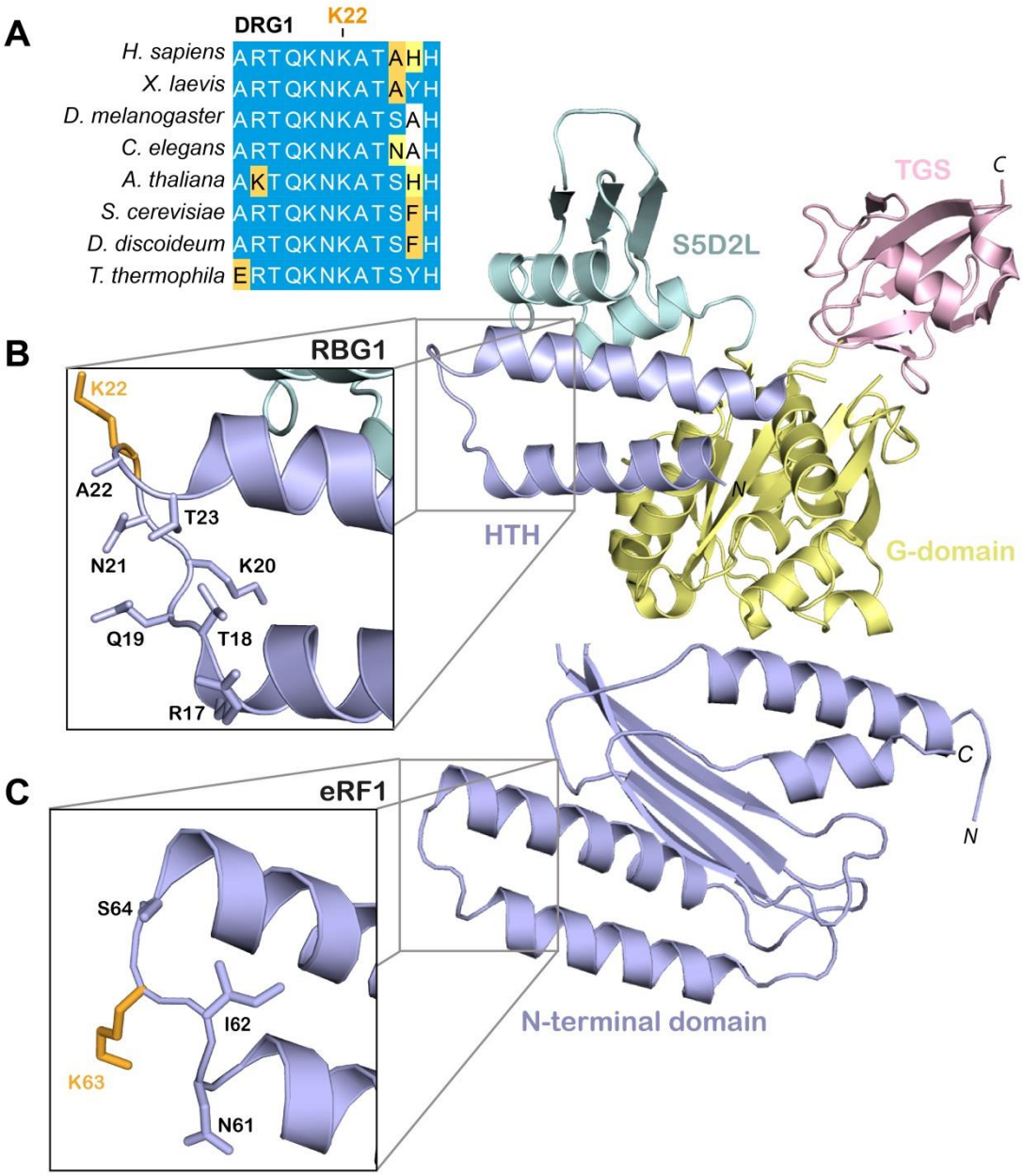


Figure 2.14. K22 of DRG1 is a highly conserved residue likely located on the apex of a helix-turn-helix (HTH) motif. (A) Sequence alignment of the HTH region of DRG1 reveals complete conservation of K22 and the RTQKNKAT$_{17\text{-}24}$ sequence (human numbering) across diverse species. Protein BLAST[143] searches identified DRG1 homologs from indicated organisms using the human protein sequence as the entry query (NCBI Reference IDs – *H. sapiens*: NP_004138.1; *X. laevis*: 001084013.1; *D. melanogaster*: 536733.1; *C. elegans*: 001255126.1; *A. thaliana*: 195662.1; *S. cerevisiae*: 009364.1; *D. discoideum*: 636294.1; *T. thermophila*: 001025967.2). ***(B)*** View from a crystal structure of the *S. cerevisiae* DRG1 homologue, RBG1 (PDB ID: 4A9A[166]), indicates that the target lysine residue is located at the apex of the HTH motif (light blue). A zoomed view of the HTH motif (black outline) reveals that K22 (orange stick) protrudes from a tight turn between two α-helices. ***(C)*** View from a crystal structure of the *N*-terminal domain of human eRF1 (PDB ID: 1DT9[168]) similarly shows the target lysine residues on a loop situated between two α-helices. A zoomed view (black outline) highlights the target K63 residue (orange stick).

The striking amino acid conservation across eukaryotic species, together with the structural insights gained from the RBG1 crystal structure, suggest that lengthening the peptide at the *N*-terminal end might create a better DRG1 peptide substrate. Indeed, it was found that increasing the peptide length from the *N*-terminus to include additional amino acids within the HTH (ARTQKNKATAHHLGLLKARLAKLRR-NH_2; $DRG1_{16\text{-}40}$) results in substantially more efficient hydroxylation (~80% after 1 h) as compared with the 20-mer $DRG1_{21\text{-}40}$ peptide substrate (~40% hydroxylation after 1 h) (**Figure 2.15**).

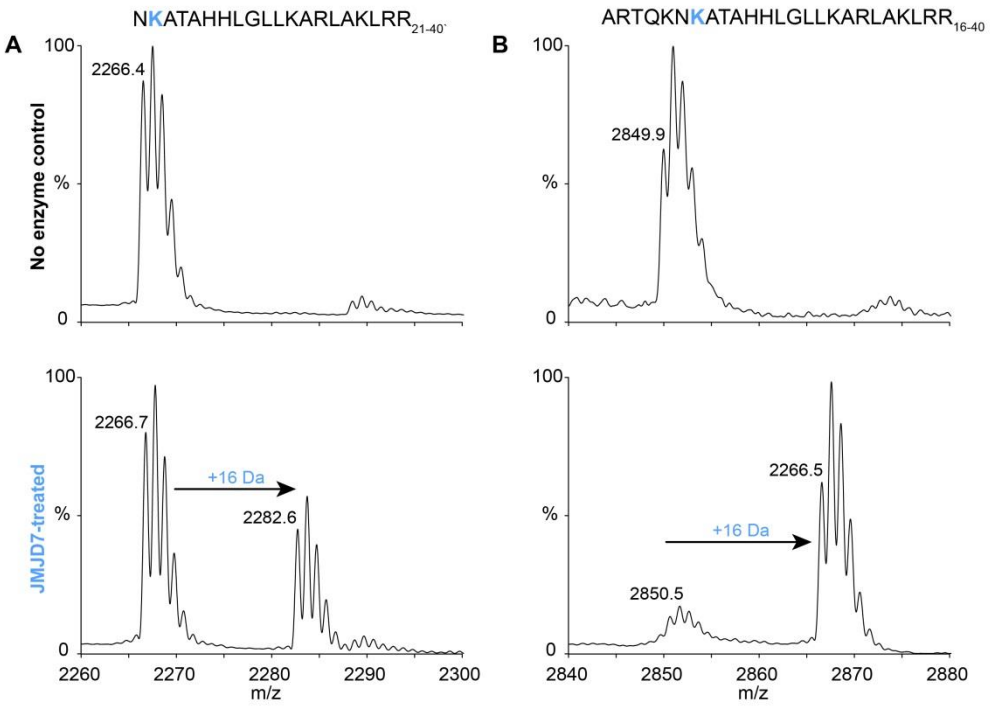


Figure 2.15. The 25-mer $DRG1_{16\text{-}40}$ peptide is a better JMJD7 substrate than the 20-mer $DRG1_{21\text{-}40}$ peptide. MALDI-TOF mass spectra corresponding to the ***(A)*** $DRG1_{21\text{-}40}$ peptide and ***(B)*** $DRG1_{16\text{-}40}$ peptide after incubation in the (top) absence and (bottom) presence of JMJD7 under the conditions described in **Figure 2.10**. The $DRG1_{16\text{-}40}$ peptide undergoes more JMJD7-catalysed hydroxylation under these conditions, indicating possible 'long-range' interactions between JMJD7 and DRG1 residues 16-21.

2.3.3 JMJD7 is a bona fide 2OG oxygenase

2.3.3.1 JMJD7 catalysis relies on 2OG and Fe(II)

As described in **Section 1.1.2**, 2OG oxygenase catalysis relies on 2OG and oxygen as co-substrates and Fe(II) as a co-factor; ascorbate is a requirement for some 2OG oxygenases (e.g. the pro-collagen prolyl hydroxylases), but in most cases, it is not required for 2OG oxygenase

activity[34]. To investigate the co-factor/substrate dependence of JMJD7 hydroxylase activity, assays were performed in the presence and/or absence of indicated co-factors/substrates (**Section 7.7.1.2**); $DRG1_{16-40}$ hydroxylation was monitored using MALDI-TOF-MS.

Consistent with standard 2OG oxygenase requirements, JMJD7 hydroxylation was found to be impaired by the absence of 2OG and Fe(II) as well as by the presence of the broad spectrum 2OG oxygenase inhibitor, NOG, under non-limiting conditions. The addition of ascorbate had little effect on JMJD7 activity in the linear range of enzyme catalysis. As predicted, substitution of one of the predicted Fe(II)-binding residues, H178, to an alanine, resulted in significantly less hydroxylation than wildtype JMJD7 (**Figure 2.16**). Overall, these results demonstrate that JMJD7 is a bona fide 2OG oxygenase.

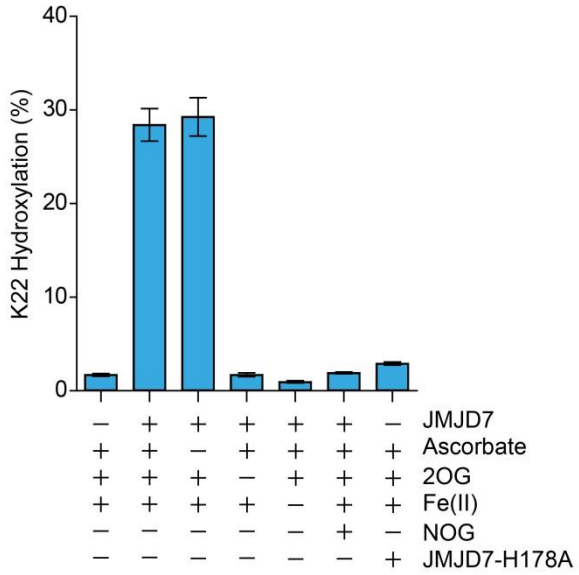


Figure 2.16. JMJD7 catalysis relies on Fe(II) and 2OG. Assays in the presence (+) and absence (-) of indicated co-factors/substrates and inhibitors reveal JMJD7 is a ferrous iron and 2OG-dependent oxygenase; JMJD7 is inhibited by the generic 2OG oxygenase inhibitor, NOG. JMJD7 activity is not dependent on ascorbate. Substitution of the iron-binding H178 residue to an alanine residue results in reduction of JMJD7 activity. $DRG1_{16-40}$ (50 μM) was incubated for 5 min at 37°C in the absence and/or presence of indicated co-factors under the following conditions: 100 μM Fe(II), 500 μM 2OG, 1 mM ascorbate, 1 mM NOG, 1 μM JMJD7 or JMJD7-H178A, 50 mM HEPES, 37°C. Data represent mean ± s.d. (n = 3) and analysis by MALDI-TOF-MS.

2.3.3.2 Investigation of the effect of ascorbate on JMJD7 catalysis

While ascorbate is not a requirement for JMJD7 catalysis, it was found that during longer incubations (1 hour), its absence caused a reduction in $DRG1_{16-40}$ hydroxylation from 72 ± 5% (**Figure 2.16**) to 38 ± 7% (**Figure 2.17A**). This phenomenon is likely due to the reducing nature of ascorbate, which helps to maintain the reduced state iron [Fe(II)] under the oxidising incubation conditions. While ascorbate is likely a relatively strict requirement for some 2OG oxygenases, (e.g. the collagen prolyl hydroxylases[32,34]), it is not essential for catalysis by several 2OG oxygenases, including most JmjC hydroxylases (e.g. JMJD4[133], JMJD6[107], and MINA53[122]). While the precise mechanistic and functional roles of ascorbate in 2OG oxygenase catalysis remain to be understood, ascorbate has been proposed to promote hydroxylase activity *in vitro* by reducing enzyme-oxidised Fe(III) species back to the active Fe(II) state, thus enabling 2OG oxygenases to continue the catalytic cycle[33,34]. This proposal is certainly in agreement with the observation that ascorbate promotes JMJD7 activity during longer incubations.

2.3.3.3 Investigation of the catalytic activity of JMJD7-H178A

Although the JMJD7-H178A variant does not catalyse substrate-uncoupled 2OG turnover and manifests reduced hydroxylation of $DRG1_{16-40}$ in the linear range, the JMJD7-H178A variant was observed to catalyse 2OG turnover in the presence of $DRG1_{16-40}$ (**Section 7.13.2**), though less efficiently than wildtype JMJD7 (**Figure 2.16**). Furthermore, JMJD7-H178A was observed to hydroxylate $DRG1_{16-40}$ to levels of ~37% after 1 hour (**Figure 2.17A**). Interestingly, while JMJD7-H178A cannot catalyse uncoupled 2OG turnover (**Figure 2.9**), it can catalyse 2OG conversion to succinate in the presence of a DRG1 peptide substrate, albeit less efficiently, consistent with observations made by MALDI-TOF-MS hydroxylation assays (**Figure 2.17**).

Overall, these data indicate that H178A is not a catalytically inactive JMJD7 variant in isolated protein assays, and therefore mutation of two Fe(II) binding residues is likely required for full

inactivity. Related observations have been made with FIH, where mutational studies indicated that the presence of two histidine residues was sufficient for Fe(II)-binding and catalysis[169]. Interestingly, there are 2OG oxygenases from lower organisms that catalyse halogenation (e.g. SyrB2[8] and WelO5[9]), which crystallographic studies have revealed are missing the carboxylate residue of the Fe(II)-binding triad to accommodate binding of the halogen ion in the active site[8,9]. The implications of these observations are important for producing catalytically inactive 2OG oxygenases for use in biochemical and cellular assays as negative controls. They may also be an important consideration in the design of selective 2OG oxygenase inhibitors.

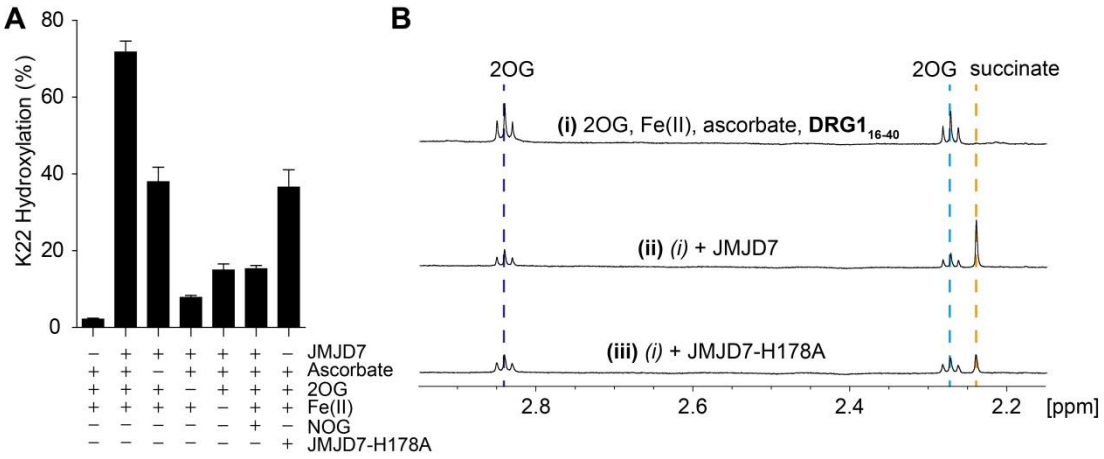


Figure 2.17. The presence of ascorbate and mutation of the H178 Fe(II)-binding residue have effects on JMJD7 catalysis during longer incubations. (A) Graphical representation of $DRG1_{16-40}$ hydroxylation corresponding to co-factor/substrate dependence assays quenched after 1 h incubations at 37°C in the presence (+) or absence (-) of indicated co-factors/substrates and inhibitors as described in **Figure 2.9**. Note that while the presence of ascorbate has little effect on $DRG1_{16-40}$ hydroxylation after 5 min (**Figure 2.16**), its presence has an effect after 1 h. Furthermore, the H178A Fe(II)-binding variant shows substantial activity (~40%) after 1 h. ***(B)*** ^{1}H-NMR spectra monitoring 2OG turnover following incubation (25°C, 1900 s) of (i) 50 µM $DRG1_{16-40}$ in the (ii) presence of JMJD7 or (iii) JMJD7-H178A under the conditions described in **Figure 2.9**. Albeit not as efficient as wildtype JMJD7, JMJD7-H178A catalyses coupled 2OG conversion to succinate in the presence of DRG1.

2.3.4 JMJD7 catalyses formation of *(3S)*-hydroxylysine

2.3.4.1 Amino acid analysis

Amino acid analysis (**Section 7.8.5**) is a technique that is used to determine the amino acid content of samples containing amino acids, peptides, and proteins[170]. The first step of amino acid analysis of peptides/proteins entails the hydrolysis of all amide bonds, which link the

constituent amino acids in a peptide or protein. Hydrolysis for amino acid analysis is typically achieved by two methods: (1) acid hydrolysis or (2) protease hydrolysis. Following sample hydrolysis, the amino acids are derivatised using 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate (AQC)[171], which produces more hydrophobic and fluorescent amino acids that can be suitably separated by a C18 column. The constituent amino acids are then analysed using ultra-performance liquid chromatography (UPLC) coupled to mass spectrometry[170,171]. The derivatised amino acids can be visualised by monitoring absorbance at 250 nm and identified by their respective retention times from extracted ion chromatograms, which isolates ion intensities corresponding to the molecular weight of the AQC-derivatised amino acid. The described workflow is outlined in **Figure 2.18**. Coupling the UPLC system to a mass spectrometer offers two main advantages: (1) the ability to generate extracted ion chromatograms and filter for the mass of a specific derivatised amino acid, enabling confident characterisation; and (2) increased sensitivity, allowing for the study of nano- and microgram amounts of material, which is important particularly when working with limited amounts of peptide and protein samples. The amino acids are ultimately identified and characterised by comparison of retention times and mass spectra with synthetic standards. The described workflow, especially the derivatisation, allows for diastereomeric mixtures to be separated.

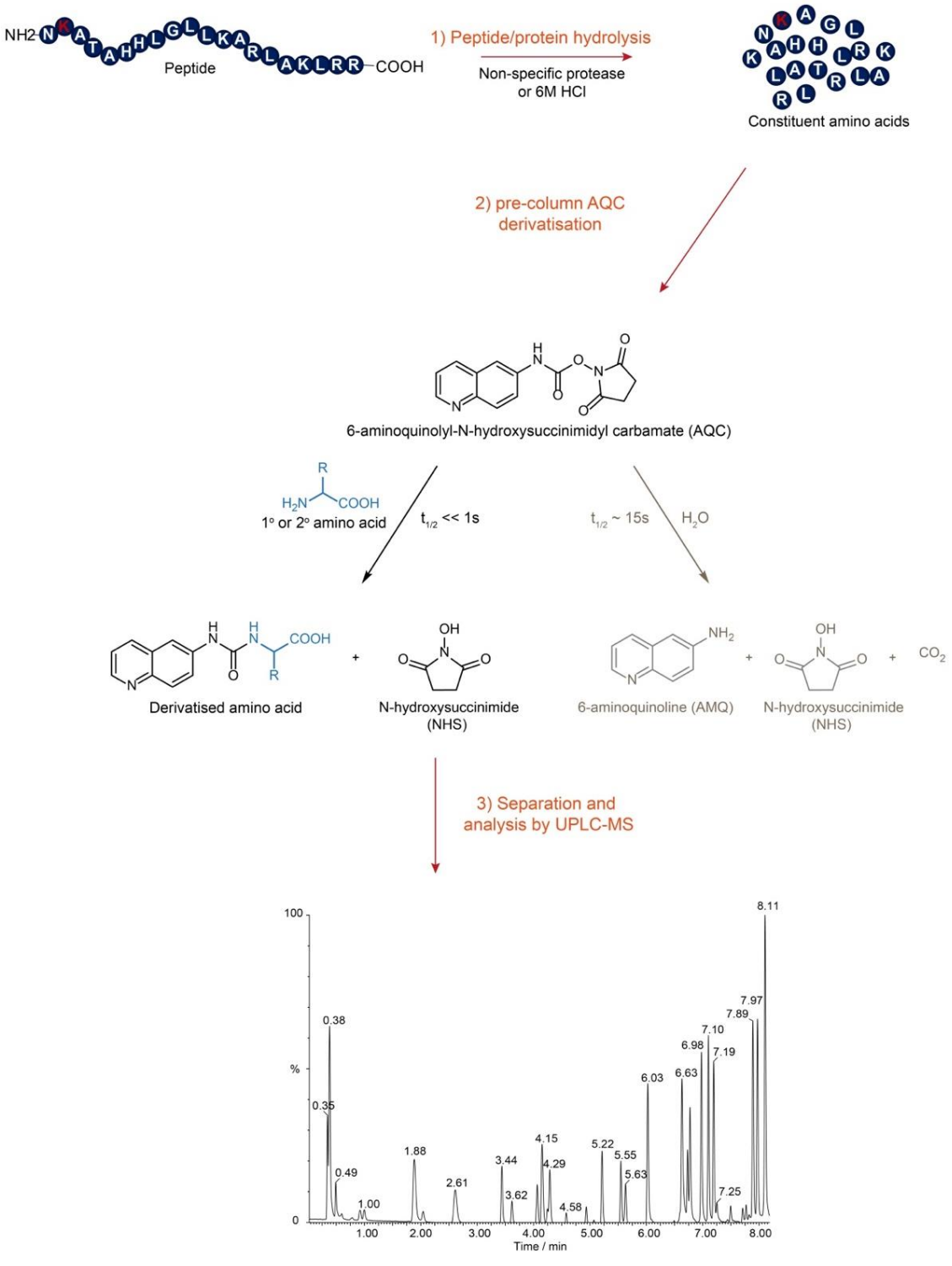


Figure 2.18. Amino acid analysis workflow. Peptides/proteins of interest are first hydrolysed to their constituent amino acids using either a non-specific protease and/or 6M HCl. The constituent amino acids are then derivatised with AQC and analysed by UPLC coupled to MS by comparison with synthetic standards.

2.3.4.2 Optimisation of sample preparation for amino acid analysis

Preliminary amino acid analysis experiments indicated that the presence of JMJD7 in the sample complicated the analyses (i.e. hydroxylysine was unable to be resolved). Therefore, several strategies were employed to remove JMJD7 following hydroxylation of the $DRG1_{21-40}$ peptide. The JMJD7 removal methods included: (1) enzyme precipitation by the addition of acetonitrile or methanol; (2) enzyme precipitation by heat; (3) removal of enzyme using a molecular weight cut-off (MWCO) spin column; (4) separation of peptide from enzyme using charged Ni(II) resin; and (5) isolation of the peptide from JMJD7 using a C18 Sep-Pak cartridge (**Figure 2.19**).

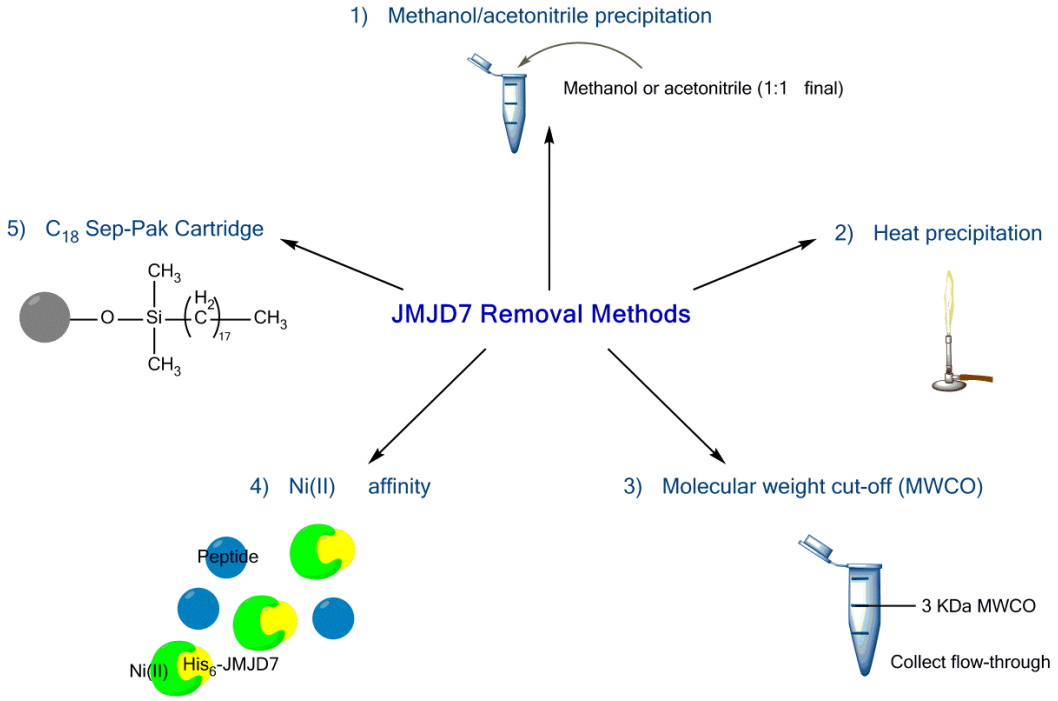


Figure 2.19. JMJD7 removal methods. Strategies to isolate the peptide from JMJD7 post-incubation under hydroxylation conditions are illustrated.

Methods 1-4 were not efficient for various reasons:

- Method (1): Acetonitrile and methanol did not yield JMJD7 precipitation.
- Method (2): Given that the peptide was not observed in the MALDI-TOF mass spectrum following JMJD7 precipitation with heat (60°C for 10 min followed by 70°C for 5 min), it is likely that tight substrate binding led to co-precipitation.

- Method (3): As for method (2), no peptide was observed in the MALDI-TOF mass spectrum following collection of flow-through from a 3K MWCO spin column, suggesting that the peptide stuck to MWCO membrane.
- Method (4): As for methods (2) and (3), no peptide was observed in the MALDI-TOF mass spectrum following binding of His_6-JMJD7 to the charged Ni(II) resin, possibly due to a tight substrate interaction with JMJD7 and/or non-specific peptide binding to the Ni(II) resin.

Method (5), the use of a C18 Sep-Pak cartridge for separation of the peptide from enzyme (**Section 7.8.3.2**), was found to be a successful method for removing JMJD7 from the sample, though it required optimisation. Rather than eluting in one-step with high percentage acetonitrile solution, a stepwise gradient elution (300 μL of 12%, 25%, 40%, 50%, and 65% acetonitrile, 0.1% (v/v) formic acid) was employed to determine the optimal concentration of acetonitrile necessary to elute the peptide from the column without eluting JMJD7. Following peptide elution, the samples were dried in a speed concentrator and re-suspended in a 0.1% (v/v) formic acid solution. A concentration of > 25% acetonitrile was found to be required to efficiently remove the $DRG1_{21\text{-}40}$ peptide from the column, as determined by MALDI-TOF-MS. The absence of His_6-JMJD7 was confirmed by western blot analyses (**Figure 2.20**; **Section 7.4.7**). Though a band corresponding to His_6-JMJD7 was not detected in any sample with an acetonitrile concentration < 65%, a one-step elution in 25% (v/v) acetonitrile, 0.1% (v/v) formic acid was chosen for peptide elution and enzyme removal. This concentration was chosen because it was found to be sufficient for peptide elution, and a lower concentration of acetonitrile was desirable to avoid JMJD7 elution. The C18 Sep-Pak method with a one-step elution in 25% acetonitrile, 0.1% formic acid was therefore used to remove JMJD7 from all subsequent samples of $DRG1_{21\text{-}40/16\text{-}40}$ for amino acid analyses. Following the removal of solvent using a speed concentrator, samples were re-suspended in 20 mM HEPES, pH 7.5 and hydrolysed overnight at 37°C using a non-specific protease from *Streptomyces griseus* in a 1:5 mass ratio of protease:sample as described in **Section 7.8.5**.

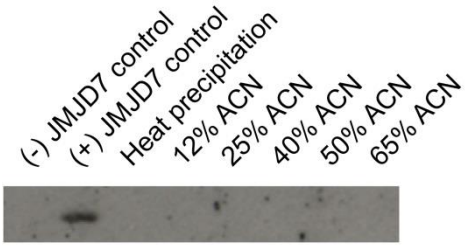


Figure 2.20. Western blot reveals efficient JMJD7 removal using a C18 Sep-Pak column. The efficiency of the C18 Sep-Pak column method for JMJD7 removal is evidenced by the lack of a band present for His_6-JMJD7. His_6-JMJD7 was visualised using 1° α-His and 2° α-mouse HRP (horseradish peroxidase) antibodies.

2.3.4.3 Determining the regiochemistry of JMJD7-catalysed lysyl hydroxylation

Following the preparation of samples for amino acid analysis using the optimised method described in **Section 2.3.4.2**, hydrolysed samples and amino acid standards were suspended in borate buffer and derivatised with AQC for analysis by UPLC-MS. Analysis of amino acids from the JMJD7-treated $DRG1_{21-40}$ peptide revealed the appearance of a new peak at 5.86 minutes, which was not present in the no enzyme control (**Figure 2.21A**). The mass of the new peak (m/z = 503.2162) was consistent with that of derivatised hydroxylysine (theoretical m/z = 503.2037). Note that this peak was not observed in the JMJD7-only and protease-only controls, indicating that it is likely a product of JMJD7 catalysis on DRG1.

Comparison of the new peak with derivatised C3, C4, and C5 hydroxylysine standards (diastereomeric mixtures containing the *(2S,XR)/(2R,XS)/(2S,XS)/(2R,XR)* isomers, where X is the hydroxylation site) revealed that the JMJD7 product co-elutes at the same retention time (5.86 minutes) as one of the C3 hydroxylysine enantiomeric pairs (**Figure 2.21A-B**). Note that because each standard is a mixture of diastereomers, two peaks are observed, each corresponding to a pair of enantiomers. To further validate the assignment of the additional peak, the JMJD7-treated sample was spiked with each of the C3, C4, and C5 hydroxylysine standards. The spiked samples clearly indicate that the product co-elutes with the second set of C3 hydroxylysine enantiomeric pairs, corresponding to either (2*R*,3*S*)/(2*S*,3*R*) or (2*R*,3*R*),(2*S*,3*S*) (**Figure 2.21C**). By contrast, the product appears as a separate peak when

spiked with the C4 and C5 hydroxylysine standards, confirming its assignment as C3 hydroxylysine (**Figure 2.21C**). These data indicate that JMJD7 catalyses C3 lysyl hydroxylation.

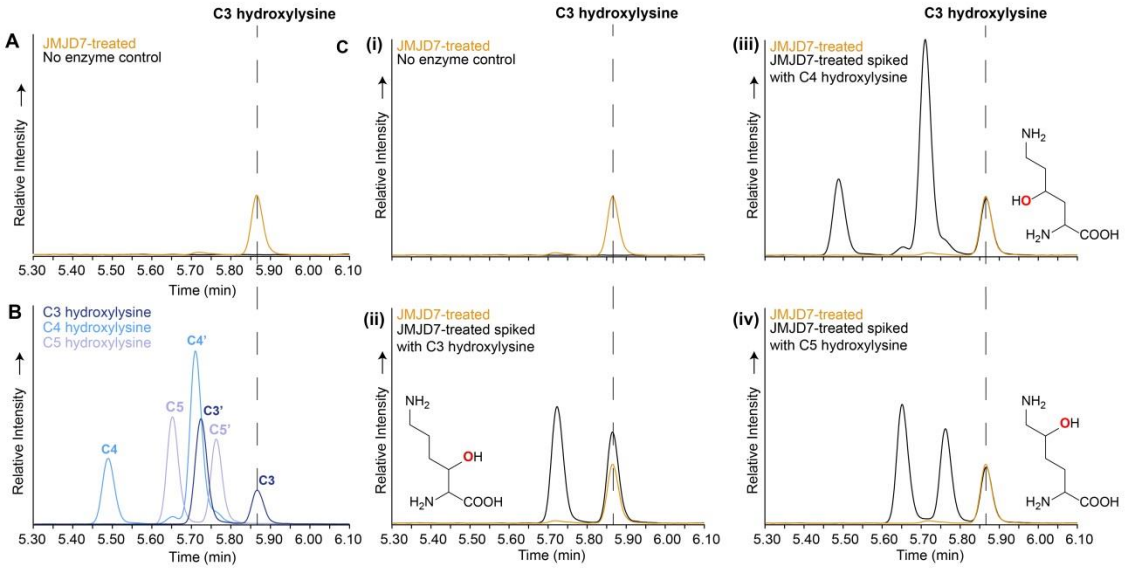


Figure 2.21. JMJD7 is a C3 lysyl hydroxylase. Extracted ion chromatograms corresponding to AQC-derivatised hydroxylysine (m/z = 503.2). ***(A)*** Chromatogram for the $DRG1_{21-40}$ peptide post-enzymatic hydrolysis indicates the presence of a new product (5.86 min) following treatment with JMJD7 (orange), which is not present in the no enzyme control (black). ***(B)*** Comparison of the JMJD7-treated sample with the C3, C4, and C5 hydroxylysine standards indicates that the new peak corresponds with one of the C3 hydroxylysine enantiomeric pairs at 5.86 min. ***(C)*** Spiking the JMJD7-treated sample with the C3, C4, and C5 hydroxylysine standards indicates that the new product only co-elutes with one set of the C3 hydroxylysine enantiomeric pairs at 5.86 min. Note that each standard is a mixture of pairs of stereoisomers.

2.3.4.4 Determining the stereochemistry of JMJD7-catalysed lysyl hydroxylation

Following the identification of JMJD7 as a C3 lysyl hydroxylase, further analyses were undertaken to determine the stereospecificity of the modification. Comparison of the derivatised C3 hydroxylysine standard (diastereomeric mixture) with a synthetic (2*S*,3*R*)/(2*R*,3*S*)-3-hydroxylysine standard (synthesised by Robert K. Leśniak, University of Oxford[172]) allowed for the diastereomeric pairs to be distinguished (**Figure 2.22A-B**). Comparison of the synthetic standard with the JMJD7-treated sample indicated the presence of two distinct peaks, implying that the product is not the (2*S*,3*R*) stereoisomer, but rather the (2*S*,3*S*) stereoisomer (assuming retention of the 2*S*-stereochemistry of the *3S* stereoisomer,

Figure 2.22A-B). Furthermore, no evidence for products of other isomers were observed within the limits of detection, thus revealing that JMJD7 stereospecifically catalyses *S*-hydroxylation at the C3 position.

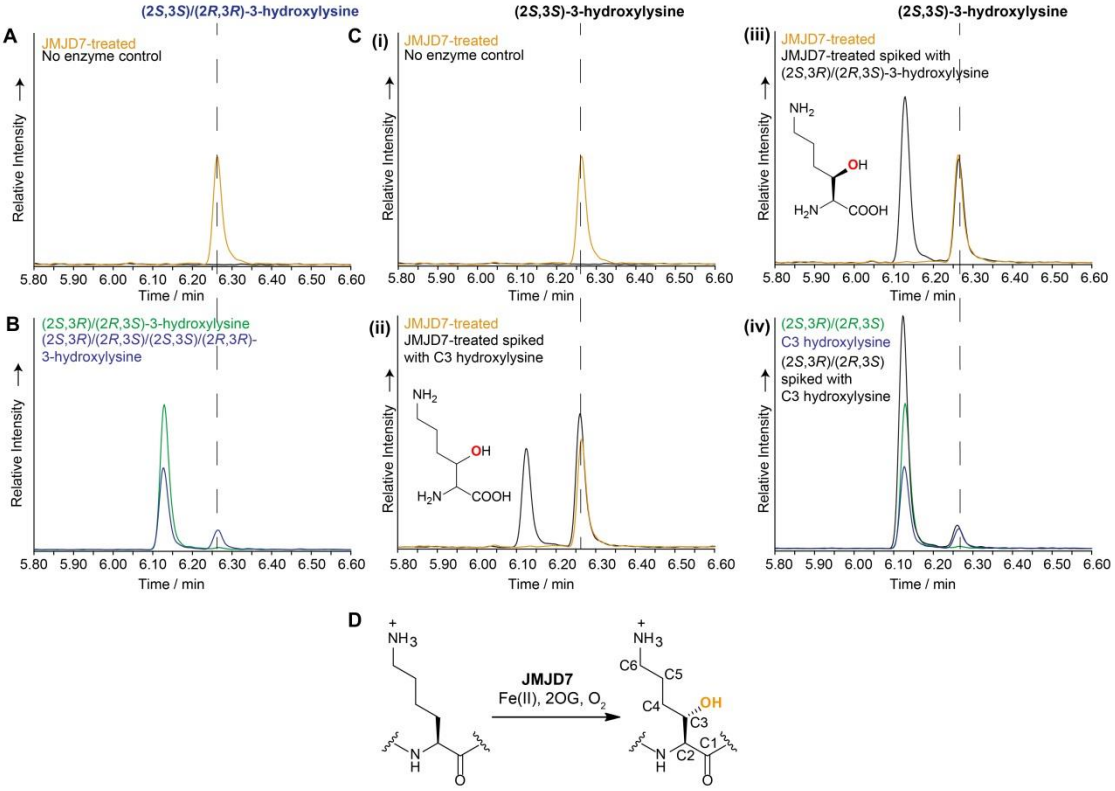


Figure 2.22. JMJD7 is a (3S)-lysyl hydroxylase. Extracted ion chromatograms corresponding to AQC-derivatised hydroxylysine (m/z = 503.2). ***(A)*** Following treatment with JMJD7 (orange), the $DRG1_{21-40}$ peptide sample contains a new amino acid product at 6.26 min, which is not present in the no enzyme control (black). ***(B)*** Comparison of the JMJD7-treated sample with the C3 hydroxylysine (blue) and (2*S*,3*R*)/(2*R*,3*S*)-3-hydroxylysine (green) standards indicates that the JMJD7 product corresponds with the opposite set of enantiomers, (2*R*,3*R*)/(2*S*,3*S*). ***(C)*** Spiking the JMJD7-treated sample (orange) with the C3 hydroxylysine standard (black) indicates that the new product co-elutes with one set of the (ii) C3 hydroxylysine pairs of enantiomers - (2*R*,3*S*)/(2*S*,3*R*) or (2*R*,3*R*)/(2*S*,3*S*) at 6.26 min. (iii) Spiking the JMJD7-treated sample (orange) with a (2*S*,3*R*)/(2*R*,3*S*)-3-hydroxylysine standard (black) results in two distinct peaks, implying the JMJD7-catalysed product is not the (2*S*,3*R*) stereoisomer, but rather the (2*S*,3*S*) stereoisomer (assuming retention of the 2*S*-stereochemistry). (iv) Overlaid extracted ion chromatograms of the C3 hydroxylysine standard (blue), the (2*S*,3*R*)/(2*R*,3*S*)-3-hydroxylysine standard (green), and two standards spiked (black) confirm that the first peak of the C3 hydroxylysine standard (6.13 min) corresponds to the (2*S*,3*R*) stereoisomer, thereby identifying the second peak at 6.26 min as the (2*S*,3*S*) stereoisomer. ***(D)*** Schematic outline of JMJD7-catalysed (3*S*)-lysyl hydroxylation.

The JMJD7-treated sample was spiked with C3 hydroxylysine and (2*S*,3*R*)/(2*R*,3*S*)-3-hydroxylysine standards to confirm that the JMJD7-catalysed product co-elutes with one set of

the C3 hydroxylysine sets of enantiomers, but not with the (2*S*,3*R*)/(2*R*,3*S*)-3-hydroxylysine synthetic standard (**Figure 2.22C**). Ultimately, these findings identify JMJD7 as the first human 2OG oxygenase, or indeed enzyme, to catalyse (3*S*)-lysyl hydroxylation (**Figure 2.22D**).

2.3.5 JMJD7 catalyses hydroxylation of recombinant full-length DRG1

2.3.5.1 DRG1 expression trials and small scale purification

To investigate whether JMJD7 catalyses hydroxylation of full-length DRG1 protein (rather than a peptide fragment), full-length DRG1 was produced in *E. coli* for subsequent testing as a JMJD7 substrate.

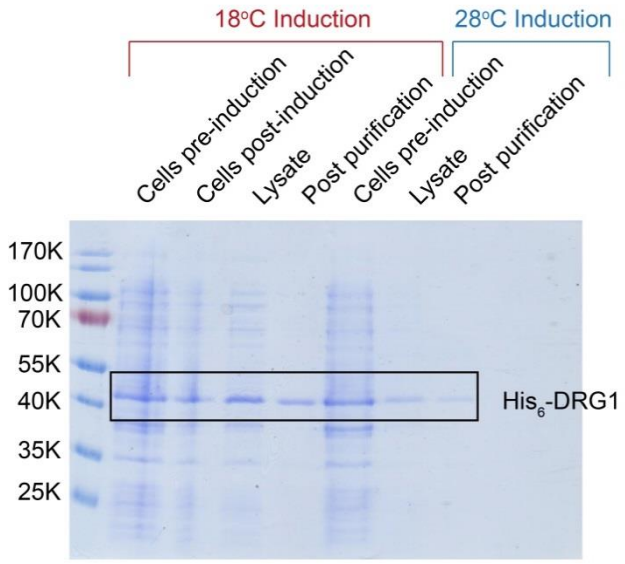


Figure 2.23. Coomassie-stained 12% SDS-PAGE gel analysis of His_6-DRG1 expression trials. Production of His_6-DRG1 was induced with 0.5 mM IPTG and grown overnight at 18°C (red) or 28°C (blue). Soluble protein expression was assessed in cells pre-induction, cells post-induction, and cell lysate. Protein following small-scale affinity purification using Ni(II) Sepharose resin was also analysed. Bands corresponding to DRG1 (theoretical molecular weight = 43095 Da) are outlined (black).

The gene encoding full-length DRG1 (UniProt: Q9Y295; residues 1-367) was inserted into a pNIC28-Bsa4 bacterial expression vector (kindly provided by the Structural Genomics Consortium, SGC) with an *N*-terminal His_6-tag (MHHHHHHSSGVDLGTENLYFQSM) using

custom primers (**Table 7.1**) as described in **Section 7.2.3**. pNIC28-Bsa4-DRG1 was transformed in Rosetta(DE3)2 pLysS cells. Protein overexpression was induced with the addition of 0.5 mM IPTG and overnight incubation at either 18°C or 28°C to determine optimal expression conditions. As shown in **Figure 2.23**, overnight incubation at 18°C post-IPTG induction was found to be more favourable. Protein was purified on a small scale using loose Ni(II)-affinity resin as described in **Section 7.4.2**.

2.3.5.2 Recombinant expression and purification of DRG1

His_6-DRG1 was produced using the optimised expression conditions described in **Section 2.3.5.1**. Overexpressed His_6-DRG1 was isolated from cell lysate using Ni(II) affinity chromatography followed by SEC, as described in **Section 7.4.4**, to > 90% purity as assessed by SDS-PAGE (**Figure 2.24**). Note that DRG1 contains an *N*-terminal His_6-tag unless otherwise stated.

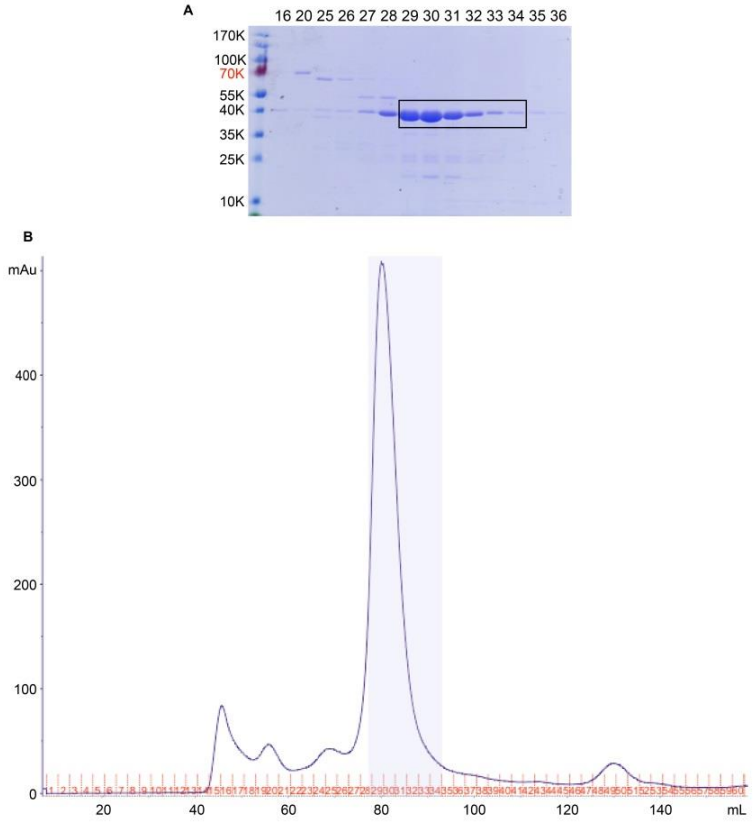


Figure 2.24. Representative SDS-PAGE gel following purification of DRG1 by SEC. (***A***) Representative 12% SDS-PAGE gel highlighting DRG1-containing fractions (black outline) with ***(B)*** the respective UV trace corresponding to DRG1 purification using a HiLoad 16/600 Superdex 200 120 mL SEC column.

2.3.5.3 Investigation of full-length DRG1 as a JMJD7 substrate

To investigate whether full-length DRG1 is a JMJD7 substrate, recombinant DRG1 was incubated with JMJD7 at 4°C or 37°C for 1 h using standard hydroxylation conditions described in **Section 7.7.1.1**. DRG1 protein was analysed for hydroxylation by intact protein LC-MS, revealing a complete shift in the mass of DRG1 by +16 Da in the presence of JMJD7 in both the 4°C and 37°C samples. Isolated protein assays thus revealed that full-length DRG1 is a substrate for JMJD7, supporting the likelihood that it might also be a cellular substrate.

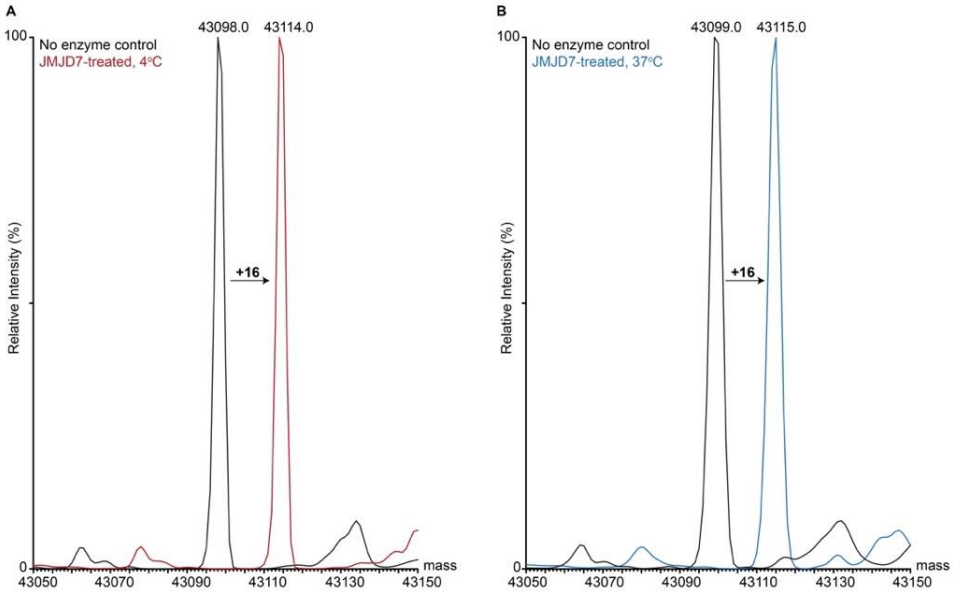


Figure 2.25. Deconvoluted intact protein mass spectra of full-length DRG1 post-incubation with JMJD7 at (A) 4 °C and (B) 37 °C for 1 h. Incubation conditions described in **Figure 2.10**.

2.3.6 Investigation of DRG2 as a JMJD7 substrate

2.3.6.1 DRG2 is closely related to DRG1

Two closely related *drg* genes are found in humans – *drg1* and *drg2* – which correspond to the DRG1 and Developmentally regulated GTP-binding protein 2 (DRG2) proteins, respectively[173]. DRG1 and DRG2 are both members of the Translation Factor (TRAFAC) family of GTPases[174]. While DRG2 was not identified in the initial co-IP (**Appendix A, Table A1**), it was hypothesised that DRG2 might also be a JMJD7 substrate due to high sequence similarity with DRG1 (57% sequence identity at the protein level), particularly in the region surrounding

the target K22 residue of DRG1. Interestingly, DRG2 is also highly conserved throughout various eukaryotic species (**Figure 2.26**).

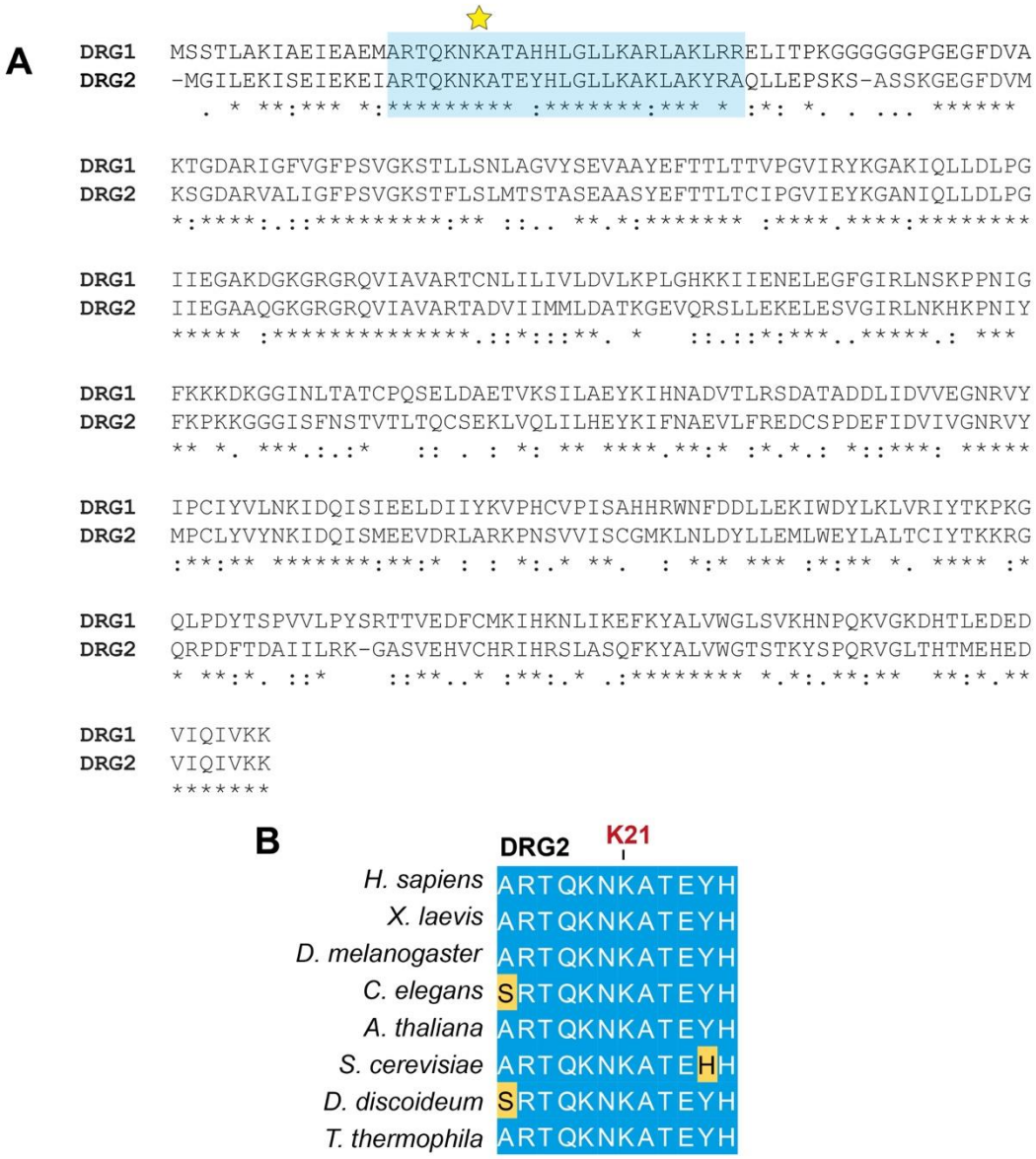


Figure 2.26. Sequence alignment of DRG1 and DRG2. (A) Amino acid sequences of DRG1 (UniProtKB Entry: Q9Y295) and DRG2 (UniProtKB Entry: P55039) were aligned using Clustal Omega[158]. The target K22 residue of DRG1, and corresponding K21 residue of DRG2, is indicated by the yellow star. Residues highlighted in light blue, corresponding to the $DRG1_{16-40}$ and $DRG2_{15-39}$ peptides, share 80% sequence identity. ***(B)*** Sequence alignment of the HTH region of DRG2 reveals complete conservation of K21 and strong conservation of the $ARTQKNKATEYH_{15-26}$ sequence (human numbering). Protein BLAST[143] searches identified DRG2 homologs from indicated organisms using the human protein sequence as the entry query across diverse species identified by BLAST (NCBI Reference IDs – *H. sapiens:* 001379.1; *X. laevis*: 001079983.1; *D. melanogaster*: 650822.1; *C. elegans:* 498808.2; *A. thaliana:* 173190.1; *S. cerevisiae:* 011689.1; *D. discoideum:* 641633.1; *T. thermophila*: 001021328.1). Sequences were aligned using Clustal Omega[158] and shaded using GeneDoc.

2.3.6.2 Investigation of DRG2 peptides as JMJD7 substrates

2.3.6.2.1 DRG2 peptide synthesis and screening

To investigate whether DRG2 might also be a JMJD7 substrate, DRG2 20-mer ($DRG2_{20-39}$) and 25-mer ($DRG2_{15-39}$) peptides, equivalent to the DRG1 20-mer ($DRG1_{21-40}$) and 25-mer ($DRG1_{16-40}$) peptides, respectively, were synthesised and screened for JMJD7-catalysed hydroxylation. Analyses by MALDI-TOF-MS indicated that while the shorter $DRG2_{20-39}$ peptide was not a JMJD7 substrate, the 25-mer $DRG2_{15-39}$ peptide was efficiently hydroxylated by JMJD7 (**Figure 2.27**). These results are consistent with the observation that the $DRG1_{21-40}$ peptide is a less efficient JMJD7 substrate as compared to the $DRG1_{16-40}$ peptide. However, this observation could not be explained using crystallographic data because no crystal structure is available for DRG2 from any species. Nevertheless, the high amino acid sequence similarity between DRG1 and DRG2 in this region suggests that these regions are also likely to be structurally similar.

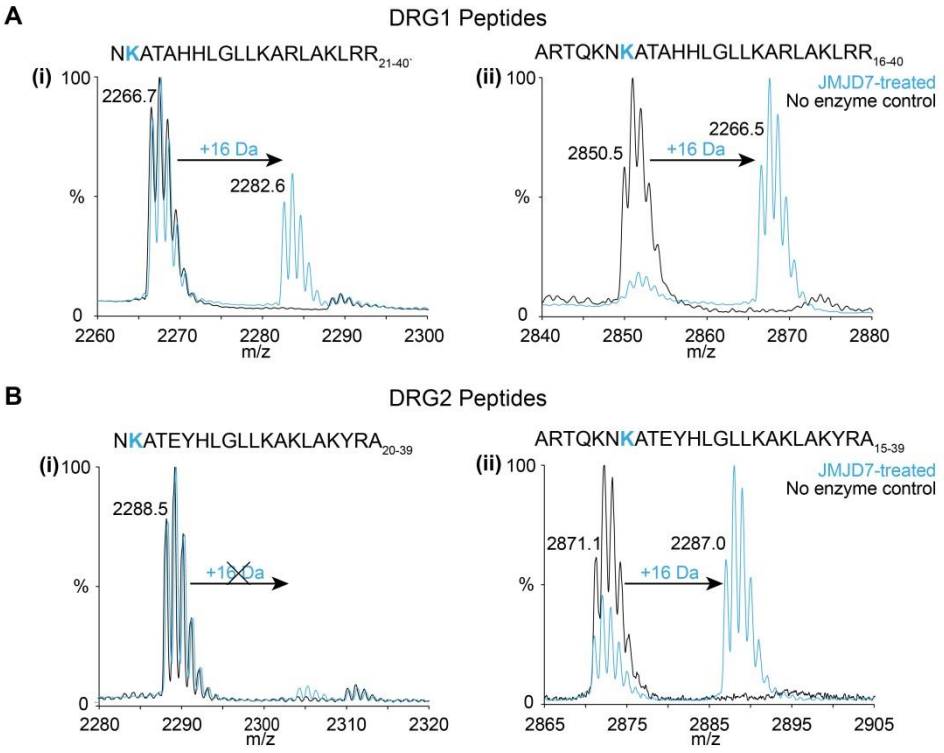


Figure 2.27. $DRG2_{15-39}$ is a more efficient JMJD7 peptide substrate than $DRG2_{20-39}$. MALDI-TOF mass spectra corresponding to ***(A)*** *(i)* $DRG1_{21-40}$ and *(ii)* $DRG1_{16-40}$ peptides and ***(B)*** *(i)* $DRG2_{20-39}$ and *(ii)* $DRG2_{15-39}$ peptides after incubation with (blue) or without (black) JMJD7 under the conditions described in **Figure 2.10**.

2.3.6.2.2 Amino acid analysis

Amino acid analysis (**Section 7.8.5**) was performed on the $DRG2_{15-39}$ peptide to validate the chemistry of the JMJD7-catalysed modification as described in **Section 2.3.4**. Consistent with the results for the DRG1 peptides, a new peak corresponding to the observed mass (m/z = 503.2162) and retention time (6.26 min) of AQC-derivatised (2*S*,3*S*)-3-hydroxylysine was present in the chromatogram for the JMJD7-treated $DRG2_{15-39}$ peptide sample (**Figure 2.28**). These results indicated that DRG2 is also likely to be a genuine JMJD7 substrate.

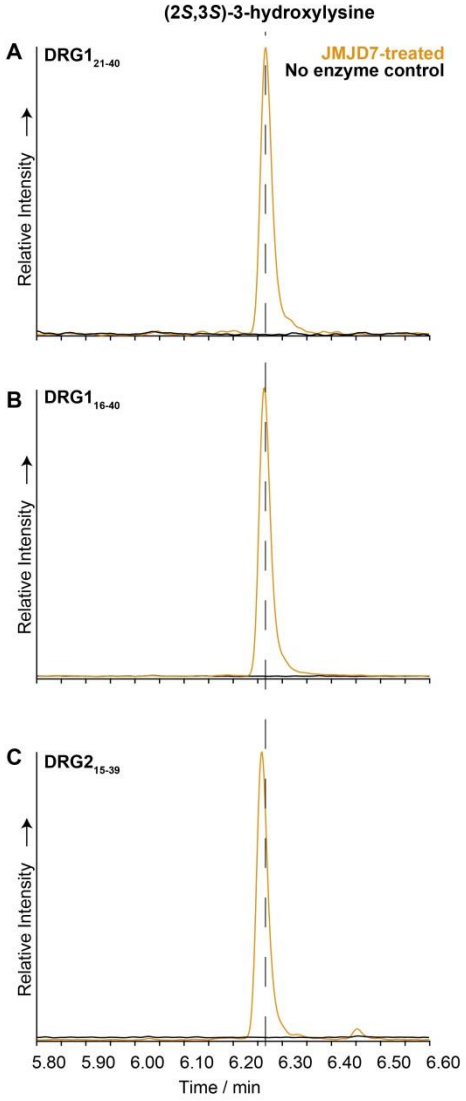


Figure 2.28. JMJD7 catalyses (3S)-lysyl hydroxylation of DRG2$_{15-39}$. Extracted ion chromatograms (m/z = 503.2) for ***(A)*** $DRG1_{21-40}$, ***(B)*** $DRG1_{16-40}$, and ***(C)*** $DRG2_{15-39}$ peptides post-enzymatic hydrolysis indicate that the JMJD7-catalysed (orange) product elutes at the same retention time as (2*S*,3*S*)-3-hydroxylysine.

2.3.6.3 Investigation of full-length DRG2 as a JMJD7 substrate

The gene encoding full-length DRG2 (UniProt: P55039; residues 1-364) was successfully amplified and inserted into a pNIC28-Bsa4 bacterial expression vector with an *N*-terminal His_6-tag (MHHHHHHSSGVDLGTENLYFQSM) using custom primers (**Table 7.1**) as described in **Section 7.2.3**. pNIC28-Bsa4-DRG2 was transformed in Rosetta(DE3)2 pLysS cells and overexpressed as described for full-length DRG1 in **Section 2.3.5.2**. However, isolation and purification of DRG2 using nickel(II) affinity and size exclusion chromatography as described in **Section 2.3.5.2** proved to be unsuccessful likely due to DRG2 instability. This was evidenced by the likely aggregation/degradation products observed by SDS-PAGE analyses (**Figure 2.29**).

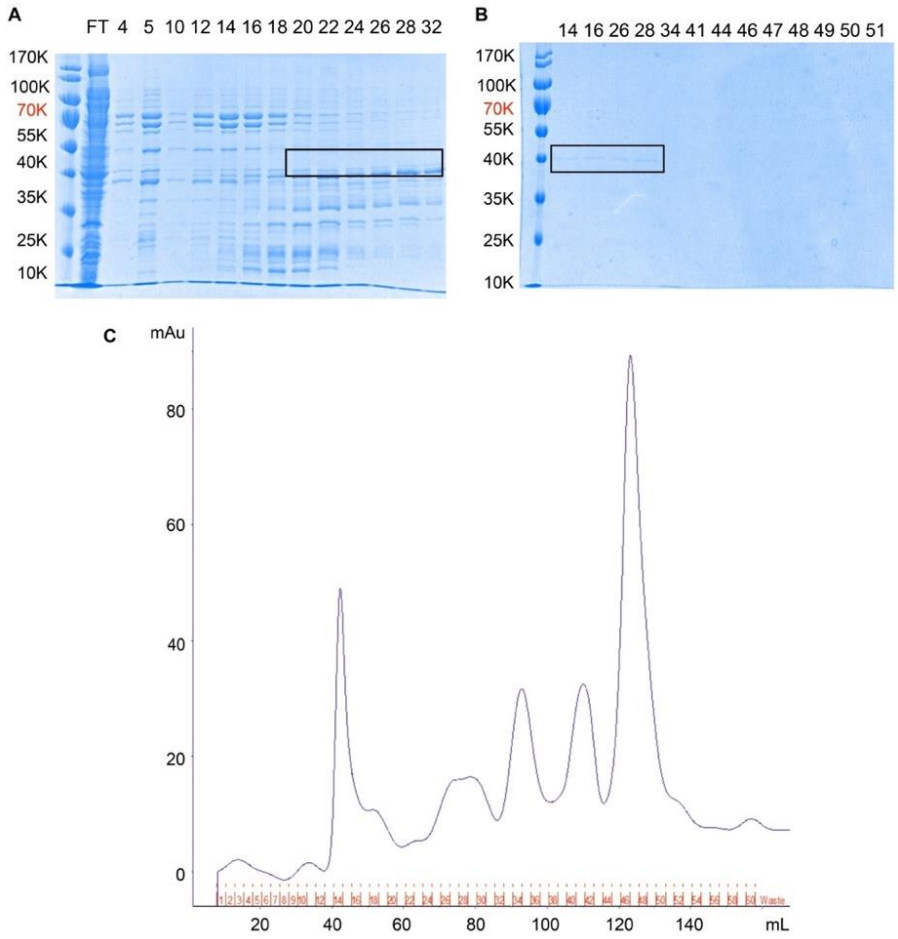


Figure 2.29. SDS-PAGE analyses following Ni(II)-affinity chromatography and SEC purification attempts of DRG2. Representative Coomassie-stained 12% SDS-PAGE gels highlighting DRG1-containing fractions (black outline) following ***(A)*** Ni(II) affinity purification and ***(B)*** SEC. ***(C)*** The respective UV trace corresponding to DRG1 purification using a HiLoad 16/600 Superdex 200 120 mL SEC column.

Attempts to stabilise the protein in subsequent purifications included the addition of fresh 2 mM dithiothreitol (DTT) to all purification solutions. While this changed the behaviour of the protein during purification, isolation of the protein was not successful due to poor expression and likely aggregation (note that protein precipitation was observed upon concentrating).

2.3.7 Investigation of JMJD7 peptide substrate selectivity – DRG1 v. DRG2

2.3.7.1 Limitations of determining kinetic parameters by MALDI-TOF-MS

Given the identification of two JMJD7 peptide substrates, it was of interest to determine which of these was the preferred JMJD7 substrate. It was thought that peptide substrate preferentiality might offer insights into JMJD7-DRG1/DRG2 interactions and catalysis in a cellular context. Kinetic parameters, such as k_{cat} and k_m, are measures of the catalytic properties of enzymes that follow Michaelis-Menten kinetics[175]. Specifically, k_{cat} (s^{-1}) is a measure of product production under saturating enzyme conditions, given as the time required for one molecule of enzyme to catalyse production of one molecule of product. The Michaelis constant, k_m, can correlate with enzyme:substrate affinity, as it represents the concentration of substrate required for the enzyme to reach half of its reaction maximum velocity (V_{max}). Substrate specificity can be measured by determining the specificity constant (k_{cat}/k_m), which takes into how well the substrate binds and the rate of product formation upon binding.

Determining kinetic parameters requires the use of a sensitive detection method to monitor enzyme catalysis. Earlier methods of monitoring 2OG oxygenase catalysis have involved the detection of enzyme-catalysed $^{14}CO_2$ formation from radioactively ^{14}C-labelled 2OG and fluorescence detection of 2OG depletion following post-incubation derivatisation with o-phenylenediamine[176]. Because 2OG oxygenases couple prime substrate oxidation with oxidative decarboxylation of 2OG to succinate and CO_2, these methods offer an indirect measure of 2OG oxygenase catalysis, rather than direct detection of hydroxylated or

demethylated product formation. Reliably measuring kinetic parameters using these assays is complicated by the fact that some 2OG oxygenases, like JMJD7, can catalyse uncoupled 2OG turnover to succinate as previously shown by ^{1}H-NMR (**Section 2.2.6**; **Figure 2.9**). MALDI-TOF-MS has therefore emerged as a commonly used direct detection method to monitor 2OG oxygenase-catalysed peptide substrate hydroxylation and demethylation[20,177,178]. MALDI-TOF-MS is a favourable technique because it is amenable to high-throughput analyses and requires low quantities of enzyme and substrate. However, kinetic parameter measurement by MALDI-TOF-MS is restricted by the instrument limits of detection and how well the peptide travels from the matrix to the MS detector.

Attempts to measure kinetic parameters for JMJD7-catalysed hydroxylation of the $DRG1_{16-40}$ and $DRG2_{15-39}$ peptide substrates were made using MALDI-TOF-MS. However, MALDI-TOF-MS proved to be an unfavourable method of determining the k_m values for the $DRG1_{16-40}$ and $DRG2_{15-39}$ peptide substrates due to the efficient turnover of substrate even at low enzyme concentrations (i.e. hydroxylation at very low peptide concentrations could not be reliably measured due to the limits of detection). Furthermore, the $DRG1_{16-40}$ and $DRG2_{15-39}$ peptide substrates apparently did not easily 'travel' to the MS detector and thus, could not be reliably detected at concentrations less than ~20 μM. Representative time courses monitoring hydroxylation are shown for $DRG1_{16-40}$ and $DRG2_{15-39}$ (**Figure 2.30**). Because several unsuccessful attempts were made to obtain kinetic parameters to compare $DRG1_{16-40}$ and $DRG2_{15-39}$ substrate selectively, competition assays were instead optimised and performed to determine the preferred JMJD7 peptide substrate (**Section 2.3.7.2**).

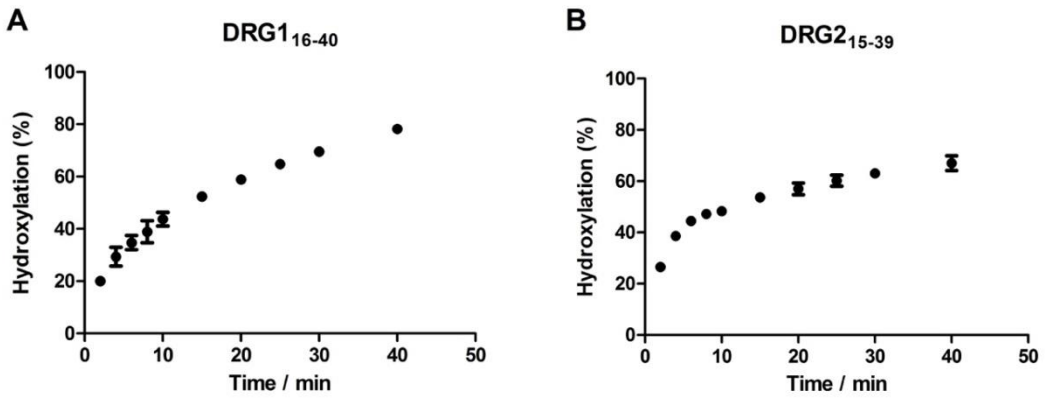


Figure 2.30. Representative time courses monitoring formation of hydroxylated product over time by MALDI-TOF-MS for the (A) $DRG1_{16-40}$ and (2) $DRG2_{15-39}$ peptides. Co-factor/substrate and buffer concentrations: 100 μM Fe(II), 500 μM 2OG, 1 mM ascorbate, 50 mM HEPES, pH 7.5. Incubation conditions for (A): 25°C, 50 μM $DRG1_{16-40}$, 1 μM JMJD7, and (B) 37°C, 52 μM $DRG2_{15-39}$, 0.65 μM JMJD7. Data represent mean ± s.d. (n = 3).

2.3.7.2 Competition assays reveal that DRG2 hydroxylation is inhibited by the presence of DRG1

To investigate whether $DRG1_{16-40}$ or $DRG2_{15-39}$ is the preferred JMJD7 peptide substrate, competition assays were employed. A competition assay was designed to monitor hydroxylation of equimolar amounts of $DRG1_{16-40}$ and $DRG2_{15-39}$ by MALDI-TOF-MS over time. To ensure that equimolar amounts of peptide were added to each reaction, absolute concentrations of peptide stocks were determined using ^{1}H-NMR spectroscopy. Peptide concentration determination was achieved by spiking a known amount of trimethylsilylpropanoic acid (TSP) in each peptide sample as described in **Section 7.13.1**. Using TSP as an internal standard and known proton amino acid chemical shifts/couplings[179], the concentrations of the peptide stocks were determined (with kind assistance from Dr Richard Hopkinson, Department of Chemistry).

Prior to the execution of the competition assays, initial velocity (or linear range) conditions of $DRG1_{16-40}$ and $DRG2_{15-39}$ hydroxylation by JMJD7 were determined by monitoring peptide hydroxylation over time under non-limiting co-factor/substrate concentrations by MALDI-TOF-MS, as described in **Section 7.7.1.2**. It is important to investigate substrate selectivity in the linear range because under these conditions, the rate is most likely to be unaffected by

factors such as reverse reaction rates, product inhibition (by either hydroxylated peptide or succinate), or enzyme inactivation due to instability or other factors. Five minutes was determined to be an appropriate time for the comparison of hydroxylation under linear, non-limiting conditions, and 40 minutes was chosen for hydroxylation comparison in the plateau or non-linear region (**Figure 2.30**).

Using stocks of $DRG1_{16-40}$ and $DRG2_{15-39}$ peptides of known concentrations, competition assays were performed using 50 µM $DRG1_{16-40}$ and/or 50 µM $DRG2_{15-39}$ under the conditions described in **Section 7.7.1.3** to investigate the peptide substrate selectivity of JMJD7. As shown in **Figure 2.31**, $DRG1_{16-40}$ hydroxylation is unaffected by the presence of $DRG2_{15-39}$ (i.e. the extent of $DRG1_{16-40}$ hydroxylation is the same in the absence and presence of $DRG2_{15-39}$). However, $DRG2_{15-39}$ hydroxylation is completely inhibited by the presence of $DRG1_{16-40}$. These results indicate that $DRG1_{16-40}$ is the preferred JMJD7 peptide substrate and apparently acts as an inhibitor of JMJD7-catalysed hydroxylation of $DRG2_{15-39}$ or as an inhibitor of JMJD7:DRG2 interactions required for catalysis. These results imply that full-length DRG1 may act as an inhibitor of JMJD7-catalysed DRG2 hydroxylation or that DRG1 more tightly binds to JMJD7. Interestingly, DRG2 was not identified as binding to JMJD7 in the co-IP experiment, further suggesting that DRG1 may preferentially bind to JMJD7. Analyses at the protein and cellular levels are required to validate these observations and better understand JMJD7 substrate specificity.

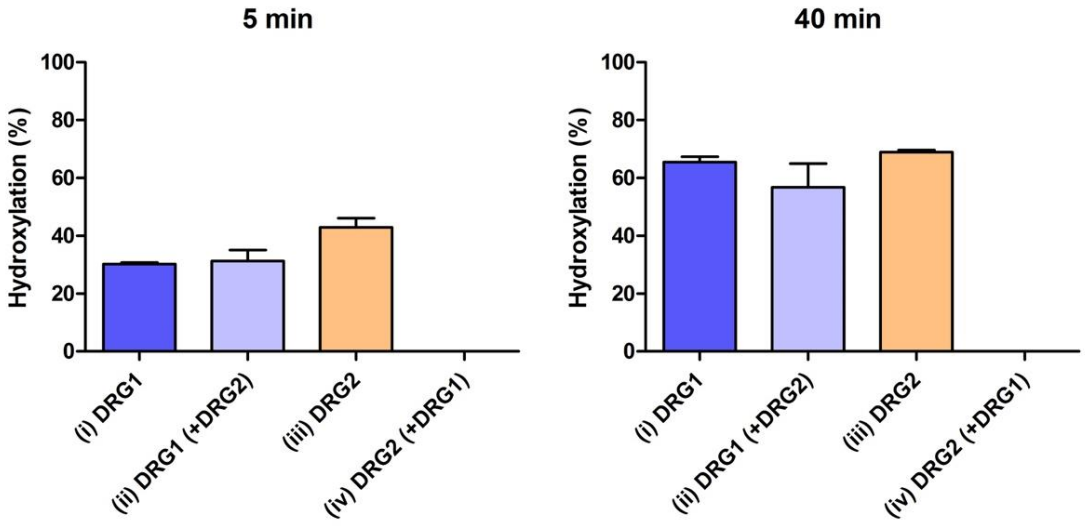


Figure 2.31. Competition assays performed with equimolar amounts of the DRG1$_{16-40}$ and DRG2$_{15-39}$ peptides quenched after 5 and 40 minutes. As compared with JMJD7-catalysed (i) DRG1$_{16-40}$ hydroxylation in the absence of DRG2$_{15-39}$, (ii) DRG1$_{16-40}$ hydroxylation is unaffected by the presence of an equimolar amount of DRG2$_{15-39}$. By contrast, in comparison to JMJD7-catalysed (iii) DRG2$_{15-39}$ hydroxylation in the absence of DRG1$_{16-40}$, (iv) DRG2$_{15-39}$ hydroxylation is completely ablated in the presence of an equimolar amount of DRG1$_{16-40}$. Data represent mean ± s.d. (n = 3).

2.3.8 Investigation of JMJD7 as a histone lysine demethylase

Histones were not identified to interact with JMJD7 in the co-IP studies (**Section 2.2.7**; **Appendix A, Table A1**); furthermore, bioinformatic analyses predicted closer structural similarity of JMJD7 to the JmjC hydroxylases (e.g. FIH and TYW5) than KDMS (**Figure 2.6**). However, most JmjC oxygenases are assigned as KDMs, and some have been controversially assigned dual hydroxylase and demethylase functionality (e.g. JMJD6[105,107,109,112,111], MINA53[121,122], NO66[122,180], and JMJD5[144,181]). Furthermore, a sequence comparison of the DRG1$_{16-40}$ and DRG2$_{15-39}$ peptide substrates with certain histone H3.1 peptide fragments known to be JmjC KDM substrates indicated strikingly sequence similarity in the region *N*-terminal to the target lysine residue (**Table 2.5**). The observed similarity between the characterised DRG substrates with certain histone H3.1 sequences suggested that KDM functionality or even JMJD7-catalysed hydroxylation of histones, though unlikely, was a possibility.

Table 2.5. Comparison of DRG1 and DRG2 peptide sequences with histone H3.1 peptide sequences.

Histone Peptide Library (Histone H3.1)*		Calculated Mass (Da)
DRG1$_{16-40}$	ARTQKN**K**ATAHHLGLLKARLAKLRR	2851.43
DRG2$_{15-39}$	ARTQKN**K**ATEYHLGLLKAKLAKYRA	2872.39
K4/K9	ART**K**QTARKSTGGKAPRKQLA	2253.62
K4me1	ART**K1**QTARKSTGGKAPRKQLA	2267.64
K4me2	ART**K2**QTARKSTGGKAPRKQLA	2281.64
K4me3	ART**K3**QTARKSTGGKAPRKQLA	2295.64
K9me1	ARTKQTAR**K1**STGGKAPRKQLA	2267.64
K9me2	ARTKQTAR**K2**STGGKAPRKQLA	2281.64
K9me3	ARTKQTAR**K3**STGGKAPRKQLA	2295.64
K27	KAPRKQLATKAAR**K**SAPATGG	2107.49
K27me1	KAPRKQLATKAAR**K1**SAPATGG	2121.49
K27me2	KAPRKQLATKAAR**K2**SAPATGG	2135.49
K27me3	KAPRKQLATKAAR**K3**SAPATGG	2149.49
K36	SAPATGGV**K**KPHRYRPGTVAL	2162.53
K36me1	SAPATGGV**K1**KPHRYRPGTVAL	2176.53
K36me2	SAPATGGV**K2**KPHRYRPGTVAL	2190.53
K36me3	SAPATGGV**K3**KPHRYRPGTVAL	2204.53

* Peptides were synthesised (with *C*-terminal amides) by Dr Sarah E. Wilkins (Schofield Laboratory, University of Oxford) and assayed for JMJD7-catalysed hydroxylation and/or demethylation. DRG1$_{16-40}$ and DRG2$_{15-39}$ are listed for sequence comparison. The target lysine is bolded.

To remain open to the possibility that JMJD7 has dual KDM/hydroxylase functionality, JMJD7 histone N^{ϵ}-methyl lysine demethylase functionality was investigated. JMJD7 was assayed against a set of histone H3.1 peptide fragments known to be JmjC KDM substrates (peptides synthesised and provided by Dr Sarah Wilkins, University of Oxford). Peptides are mimics of the unstructured histone tails, which are major sites of methylation, acetylation, and other PTMs[182]. Under conditions where JMJD7-catalysed DRG1$_{16-40}$ and DRG2$_{15-39}$ peptide hydroxylations are observed, no demethylation was observed, suggesting that JMJD7 is unlikely to function as an N^{ϵ}-methyl lysine demethylase (**Figure 2.32**).

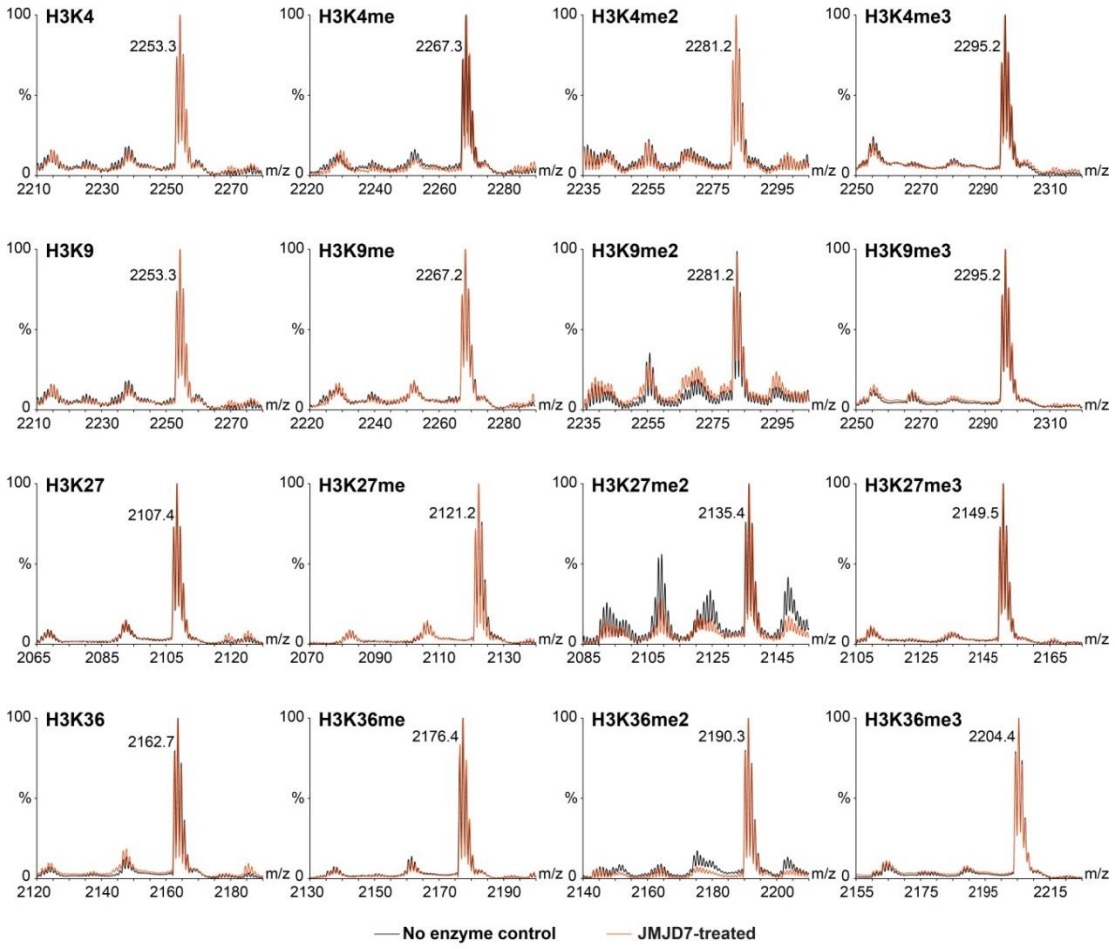


Figure 2.32. JMJD7 does not have lysine demethylase activity on histone H3.1 peptides. MALDI-TOF mass spectra of histone H3.1 peptides incubated in the absence (black) and presence (red) of JMJD7 under conditions where hydroxylation of $DRG1_{16-40}$ and $DRG2_{15-39}$ is observed. Incubation conditions are described in ***Figure 2.10***.

2.3.9 K22/K21 of DRG1-2 are hydroxylated by JMJD7 in cells

All cellular experiments described in **Section 2.3.9 were performed by Dr Qinqin Zhuang and Charlotte Eaton (immunofluorescence experiments only) under the supervision of an external collaborator, Dr Mathew Coleman (University of Birmingham). MS/MS analyses for the corresponding cellular experiments were performed by Dr Roman Fischer, Dr Simon Davis, and Rebecca Konietzny [Target Discovery Institute (TDI), University of Oxford].*

2.3.9.1 JMJD7 interacts with DRG complexes in an activity-dependent manner

To validate JMJD7 activity-dependent interactions with DRG1 and DRG2 (DRG1-2) in human cells, 3X-FLAG JMJD7 or JMJD7-H178A was expressed using doxycycline-inducible HEK293T stable cells prior to anti-FLAG immunopurification. Independent MS and western

blotting analyses identified DRG1-2 (DRG1-2) and their respective binding partners, DRG family regulatory proteins 1 and 2 (DFRP1-2)[183], as significantly enriched in wildtype JMJD7 versus control (empty vector or H178A variant) pulldowns (i.e. indicating a functionally-relevant interaction; **Figure 2.33**). The relative abundances of DRG1-2 compared to DFRP1-2 by MS were consistent with DRG1-2 being direct JMJD7 interactors. Interestingly, despite evidence for catalytic activity of JMJD7-H178A in isolated protein assays (**Figure 2.17**), the variant appeared to be catalytically inactive in cells, as evidenced by the inability of JMJD7-H178A to pulldown DRG1-2. This observation is consistent with the consensus mechanism for 2OG oxygenases, where Fe(II)-binding precedes 2OG and prime substrate binding in the active site (**Section 1.1.2**; **Figure 1.4**).

A

Protein	Accession	Peptides	Anova (p)	Abundance (emPAI)		
				EV	Jmjd7	Jmjd7 H178A
Drg1	Q9Y295	23	0.004	-	12.5	-
Drg2	P55039	25	0.004	-	8.9	-
Dfrp1	Q8WU90	24	0.004	-	7.8	-
Dfrp2	Q9H446	11	0.003	-	5.4	-

B

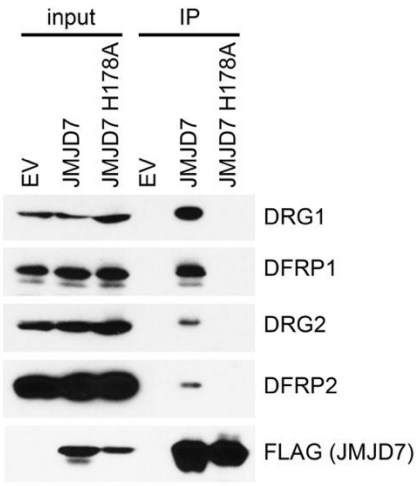


Figure 2.33. MS and western blotting analyses of 3X-FLAG JMJD7 immunopurification indicate functionally-relevant JMJD7 interactions with DRG1-2. (A) MS analyses of immunopurified complexes identified proteins significantly enriched in wildtype JMJD7 versus control empty vector or H178A mutant pulldowns. ***(B)*** Following anti-FLAG immunoprecipitation, western blotting was performed with the indicated antibodies: DRG1 (Proteintech, 13190-1-AP), DRG2 (Proteintech, 14743-1-AP), DFRP1 (SIGMA, HPA031099), DFRP2 (GeneTex, GTX120331), and FLAG (SIGMA, A8592). Figure was kindly provided by Dr Mathew Coleman (University of Birmingham).

2.3.9.2 MS/MS analyses of exogenous and endogenous DRGs

To further investigate DRG1-2 as cellular JMJD7 substrates, MS/MS analyses were undertaken on overexpressed HA-DRG/V5-DFRP complexes purified from HEK293T cells overexpressing control empty vector or FLAG-JMJD7 prior to digestion with trypsin. Consistent with the biochemical assignment of the DRG1-2 hydroxylation sites (**Figure 2.14**), single lysyl hydroxylation sites at K22 and K21 of DRG1 and DRG2, respectively, were identified. Interestingly, K22/K21 hydroxylation was assigned to peptides containing one 'missed' tryptic cleavage at the hydroxylation site: DRG1 (NK(OH)ATAHHLGLLK) and DRG2 (NK(OH)ATEYHLGLLK). The proportion of K22/K21-hydroxylated peptides showing a missed tryptic cleavage was significantly increased in the JMJD7-overexpressing cells versus the control (**Figure 2.34A**). These data are consistent with the K22/K21 residues being JMJD7 targets within DRG/DFRP complexes. Furthermore, these observations suggest that JMJD7-mediated lysyl hydroxylation impairs the activity of trypsin.

JMJD7 gene knockdown by siRNA was subsequently performed to determine whether JMJD7 is necessary for hydroxylation of endogenous DRG1-2. MS/MS analyses of endogenous DRG1-2, immunopurified from HEK293T, identified abundant K22 and K21 hydroxylation, respectively (**Figure 2.34B-F**). However, the ratio of hydroxylated versus unmodified peptides were significantly reduced upon *JMJD7* knockdown, indicating that K22/21 hydroxylation of DRG1-2 is dependent on JMJD7 expression. Consistent with the overexpression analyses, tryptic cleavage at K22/21 of endogenous DRG1-2 was apparently reduced when these sites were hydroxylated. Interestingly, impaired trypsinolysis has not been reported for the hydroxylation of lysine at the C4 or C5 position by other hydroxylases[109,133,184]. This observation suggests that JMJD7 catalyses a novel form of lysyl hydroxylation, which further supports the assignment of JMJD7 as the first identified human (3*S*)-lysyl hydroxylase.

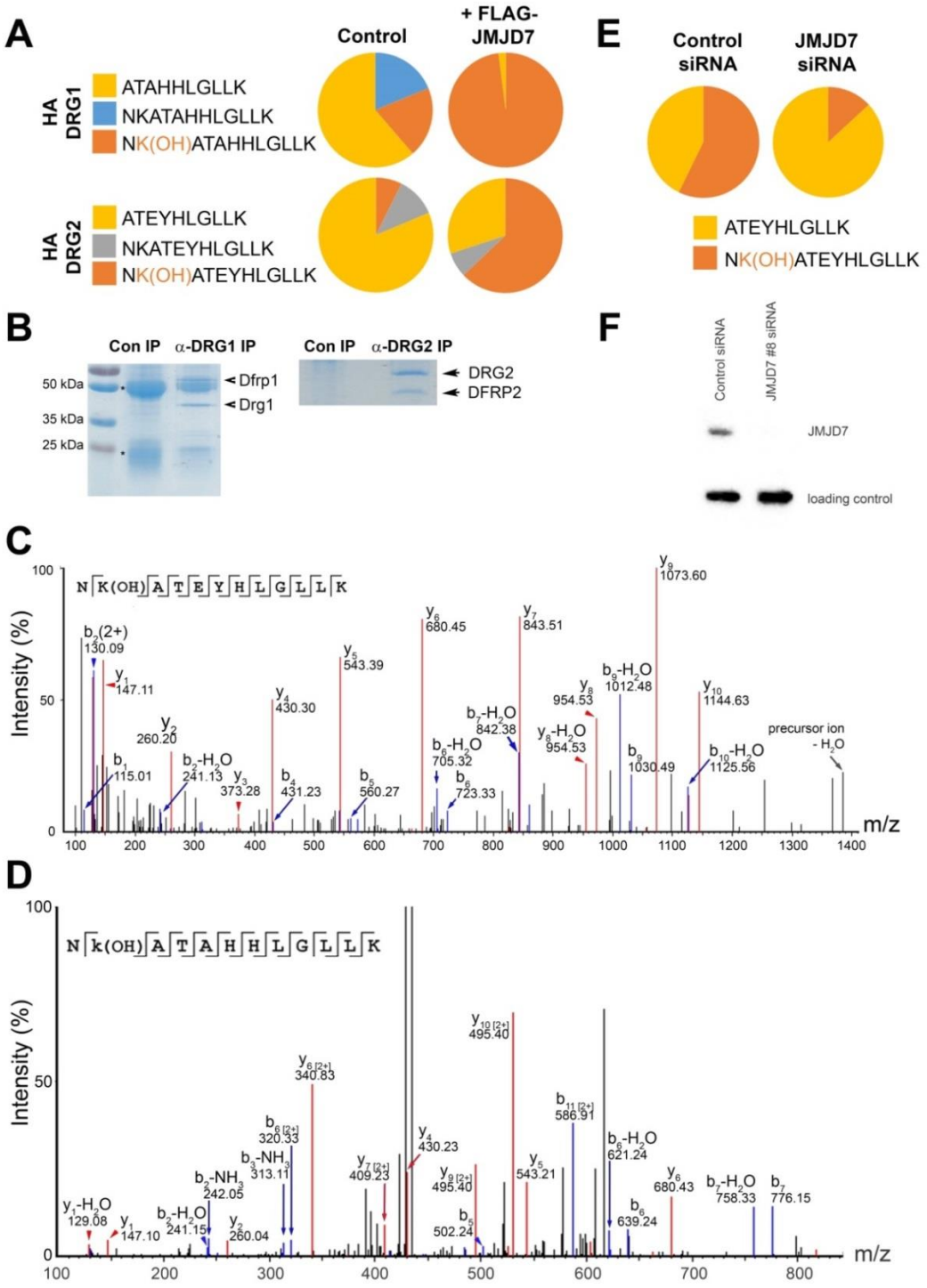


Figure 2.34. MS/MS analyses of exogenous and endogenous DRGs. (A) (Top) DRG1 K22 and (bottom) DRG2 K21 hydroxylation was analysed semi-quantitatively following HA-DRG1 or HA-DRG2 overexpression in the (left) absence or (right) presence of FLAG-JMJD7. Note that because K22/K21 hydroxylation impeded trypsinolysis, precursor intensity-based quantification was complicated. Thus, the abundance of indicated tryptic peptides was identified by their respective extracted precursor ion chromatogram, and ratios of peptides were compared. ***(B)*** Coomassie-stained SDS-PAGE gel containing eluted samples following immunopurification of endogenous (left) DRG1 and (right) DRG2 from HEK293T cells. MS/MS spectra of hydroxylated ***(C)*** DRG2 and ***(D)*** DRG1 tryptic peptide fragments (see ***Table 2.6***). ***(E)*** Quantification of endogenous DRG2 K21 hydroxylation following control or JMJD7 siRNA. ***(F)*** Western blot validation of JMJD7 knockdown in HEK293T cells transfected with control or JMJD7 siRNA oligonucleotide (JMJD7 antibody: Novus, NBP1-91110). Figure was kindly provided by Dr Mathew Coleman (University of Birmingham).

Table 2.6. MS/MS fragments corresponding to indicated endogenous DRG2 and DRG1 peptides as shown in respective MS/MS spectra in Figure 2.34C-D.*

Endogenous DRG2 N[K(OH)][A][T][E][Y][H][L][G][L][L][K]

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	115.05	97.04	98.02	58.03	N					12
2	259.14	241.13	242.11	130.09	K(+15.99)	1288.73	1270.72	1271.70	644.86	11
3	330.18	312.16	313.15	165.59	A	1144.63	1126.63	1127.61	572.82	10
4	431.23	413.21	414.20	216.11	T	1073.60	1055.57	1056.57	537.30	9
5	560.27	542.26	543.24	280.63	E	972.55	954.53	955.55	486.78	8
6	723.33	705.32	706.31	362.17	Y	843.51	825.50	826.44	422.25	7
7	860.38	842.38	843.37	430.70	H	680.45	662.44	663.42	340.72	6
8	973.47	955.46	956.45	487.24	L	543.39	525.38	526.36	272.19	5
9	1030.49	1012.48	1013.47	515.75	G	430.30	412.29	413.28	215.65	4
10	1143.58	1125.56	1126.55	572.29	L	373.28	355.27	356.25	187.14	3
11	1256.66	1238.65	1239.64	628.83	L	260.20	242.19	243.17	130.60	2
12					K	147.11	129.10	130.09	74.06	1

Endogenous DRG1 N[k(OH)][A][T][A][H][H][L][G][L][L][K]

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	115.05	97.04	98.02	58.03	N					12
2	259.14	241.15	242.04	130.10	K(+15.99)	1204.72	1186.71	1187.69	603.34	11
3	330.18	312.09	313.11	165.59	A	1060.63	1042.62	1043.60	530.87	10
4	431.23	413.21	414.14	216.11	T	989.55	971.50	972.56	495.40	9
5	502.24	484.38	485.24	251.25	A	888.45	870.53	871.28	444.97	8
6	639.24	621.24	621.99	320.33	H	817.41	799.49	800.48	409.23	7
7	776.15	758.33	759.35	388.69	H	680.43	662.35	663.42	340.83	6
8	889.10	871.28	872.44	444.97	L	543.21	525.25	526.23	272.19	5
9	946.49	928.48	929.46	473.74	G	430.23	412.29	413.28	215.65	4
10	1059.57	1041.64	1042.54	529.89	L	373.48	355.27	356.25	187.14	3
11	1172.65	1154.64	1155.63	586.91	L	260.04	242.04	243.17	130.10	2
12					K	147.10	129.08	130.10	74.06	1

***** Tables were generated using PEAKS Studio 7.5 (Bioinformatics Solutions) by Dr Roman Fischer (TDI, University of Oxford). Peaks highlighted in blue or red were observed in respective MS/MS spectra. Tables were kindly provided by Dr Mathew Coleman (University of Birmingham).

2.3.9.3 Subcellular localisation of JMJD7 and DRG1-2

Identifying the subcellular localisation of a protein is important for gaining insights into protein function and networks. Immunofluorescence is an imaging technique commonly used to investigate protein subcellular localisation by using fluorescently-conjugated antibodies on a specified protein within a cell. Immunofluorescence staining of endogenous DRG1-2 was performed in the presence or absence of exogenous HA-tagged JMJD7 in HeLa cells to investigate JMJD7/DRG1-2 co-localisation. Consistent with previous reports[183], DRG1 was predominantly found in the cytoplasm, where it co-localises with JMJD7, which is observed both within the cytoplasm and nucleus (**Figure 2.35A**). Consistent with a JMJD7 interaction, expression of HA-JMJD7 relocated a fraction of DRG1 to the nucleus, where it overlapped

with JMJD7 staining. DRG2 was similarly found to co-localise with JMJD7 in the cytoplasm, but expression of HA-JMJD7 did not relocate DRG2 to the nucleus as profoundly as for DRG1 (**Figure 2.35B**).

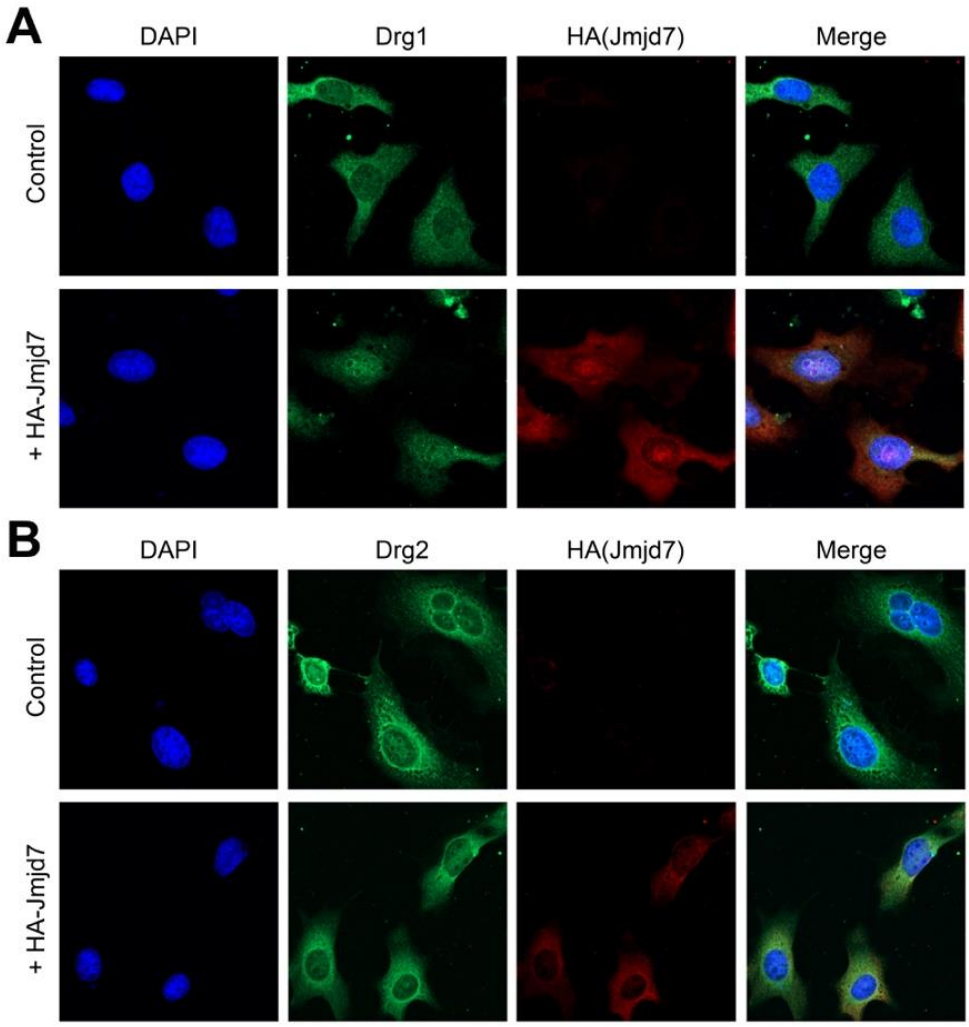


Figure 2.35. Immunofluorescence staining for (A) DRG1 and (B) DRG2 in the absence/presence of HA-JMJD7. Endogenous DRG1 is predominantly localised to the cytoplasm. Upon overexpression of HA-tagged JMJD7, DRG1 co-localises with JMJD7 in the nucleus and cytoplasm, consistent with a JMJD7 interaction. Like DRG1, endogenous DRG2 co-localises with JMJD7 in the cytoplasm; however, the extent of DRG2 re-location to the nucleus is not as profound as for DRG1 upon JMJD7 overexpression. Note that DAPI (4',6-diamidino-2-phenylindole) is a fluorescent stain commonly used in fluorescent microscopy[185]. DAPI binds to AT rich regions in DNA, and thus is used to stain for the nucleus to enable visualisation of nuclei for cellular localisation analyses.

2.4 Discussion and future perspectives

2.4.1 Roles of the DRG Proteins

2.4.1.1 DRG proteins belong to the Translation Factor (TRAFAC) GTPases

The DRG proteins, comprising DRG1 and DRG2, are highly conserved proteins belonging to the Obg subfamily of the Translation Factor (TRAFAC) class of GTPases[173,174]. GTPases are enzymes characterised by their ability to bind and hydrolyse guanosine-5'-triphosphate (GTP) to guanosine diphosphate (GDP). They have important roles in fundamental cellular processes including signal transduction, cell division, and protein biosynthesis[186]. GTPases are structurally characterised by their GTP-binding domain, or G-domain, which has five characteristic motifs – G1-G5[186]. Several GTPases are components of translational machinery; the GTPases that are involved in protein biosynthesis have been classified as the Translation Factor (TRAFAC) class of GTPases[174]. The DRGs belong to the TRAFAC GTPases and are highly conserved in archaea, fungi, yeast, plants, and humans[173,187–189]. Their evolutionary conservation suggests that the DRGs play essential roles in fundamental cellular processes, though the physiological roles of these enzymes remain poorly understood to date.

In eukaryotic organisms including mice and humans, there are two DRG homologs, DRG1 and DRG2[173]. The *drg1* gene encoding DRG1 was first identified in mice; *DRG1* mRNA expression was found to be constant during embryonic murine development but decreased in post-natal mice[187,190]. The observation, which led to the proposal that the DRGs play a role in the developing and differentiating mouse brain[190], was similarly made in *Xenopus laevis* embryos[188,191]. The expression of DRG1 and DRG2 is also linked to cell growth and/or proliferation in other species including zebrafish[192], plants (e.g. pea and *Arabidopsis*)[189,193], and humans[194,195]. Thus, their expression patterns suggest an important role in cell growth and proliferation. DRG1-2 have also been proposed to be associated with translation due to their presence in ribosome fractions in yeast, plants, and humans[196–198] and similarity to other OBG family GTPases, which have roles in translation and ribosome biogenesis[166,174].

2.4.1.2 DRG proteins and disease

While there are links between the DRG proteins and cell growth/proliferation across several species, the molecular mechanisms underlying their regulation remain poorly understood. The understanding of these mechanisms is of interest given that abnormal cellular proliferation is a hallmark of cancer cells. Recent studies have indicated links between DRG expression and tumourigenesis. For example, DRG1 levels are elevated in lung cancer and melanoma[194,199], and DRG1 knockdown inhibits the growth of lung tumour cells by inducing mitotic arrest[194]. Interestingly, the latter study reported DRG1 to be localised to the mitotic spindles[194], though others have reported DRG1 to be largely cytoplasmic[183]. The immunofluorescence experiments described in **Section 2.3.9.3** also reveal that DRG1 is predominantly found in the cytoplasm, though overexpression of JMJD7 relocated a fraction of cellular DRG1 to the nucleus (**Figure 2.35A**). While JMJD7 and DRG1 mitotic co-localisation has yet to be studied, these observations imply a potential role for DRG1 and JMJD7 in mitosis, which warrants future investigation.

The role of the DRGs is almost certainly context-dependent, and its regulation by other factors likely differs. DRG1 overexpression inhibits the expression of cancer cell matrix metalloproteinases in breast cancer cells, which are reported to stimulate cancer progression[200]. DRG2 overexpression was similarly reported to suppress the growth of Jurkat cells[201], an immortal T lymphocyte cell line. The complex roles of the DRGs in cancer and cell growth/proliferation likely reflect the same complexity for JMJD7 (**Section 2.1.3**). The DRGs may also have roles in intellectual development and disabilities. DRG1 was thought to play a role in neurogenesis given its initial identification in the developing mouse brain[190]. In addition, *DRG2* has been identified as a potential candidate gene implicated in Smith-Magenis syndrome, a developmental disorder affecting the body and brain, given its location in a critical interval on chromosome 17[202].

2.4.2 2OG oxygenases that modify GTPases

Interestingly, the DRGs are not the first GTPases identified as 2OG oxygenase substrates. Prokaryotic Elongation Factor-Thermo unstable (EF-Tu), which has previously been identified as a substrate for *Pseudomonas* prolyl hydroxylase domain containing protein (PPHD), is also a TRAFAC GTPase. As part of a ternary EF-Tu•GTP•aminoacyl-tRNA complex, EF-Tu is responsible for delivering appropriate amino acids to the ribosome during prokaryotic translation[203]. PPHD is a bacterial 2OG oxygenase with striking similarity to human PHD2, a well-studied oxygen sensor and HIF prolyl hydroxylase[136]. While the effect of PPHD-catalysed prolyl hydroxylation on the GTPase activity of EF-Tu was investigated, no effect was observed[136]. Given that the target K22 lysyl residue in DRG1 points away from the G-domain (assuming DRG1 structural similarity with RBG1, **Figure 2.14**), JMJD7-catalysed lysyl hydroxylation of DRG1-2 is also unlikely to have any effect on GTPase activity. However, the investigation of this effect could be the subject of future studies by monitoring DRG-catalysed GTP-GDP turnover by NMR spectroscopy.

2.4.3 JmjC oxygenases that modify translational machinery

The association of the DRGs with components of translational machinery is certainly consistent with reported substrates for other JmjC hydroxylases. JmjC hydroxylases have ancient roles in protein synthesis via the modification of components of translational machinery (**Section 1.3.4**; **Figure 1.12**). The human RiOXs, MINA53 and NO66, target ribosomal proteins RPL27a and RPL8, respectively, via hydroxylation of histidinyl residues[102,122]. YcfD, a bacterial homolog of the human RiOXs, similarly catalyses arginyl hydroxylation of the ribosomal protein RPL16[102,122]. The functional importance of hydroxylation of wybutosine to hydroxywybutosine in $tRNA^{Phe}$ by TYW5 has yet to be determined, but notably extends the known roles of JmjC proteins to include RNA hydroxylation[116]. JMJD4 catalyses C4 lysyl hydroxylation of the translation release factor eRF1, which increases the efficiency of translation termination[133]. Other characterised JmjC hydroxylases have reported roles in other aspects of protein synthesis.

For example, FIH regulates the transcriptional activity of HIF via asparaginyl hydroxylation[15,68,69], and JMJD6 catalyses (5*S*)-lysyl hydroxylation of splicing regulatory proteins[107,110]. Though not a JmjC oxygenase, OGFOD1 is another 2OG oxygenase involved in hydroxylating ribosomal proteins[129].

2.4.4 JMJD7-catalysed hydroxylation expands known 2OG oxygenase-catalysed lysyl modifications

The biochemical and cellular results presented in this chapter define the previously uncharacterised enzyme, JMJD7, as a lysyl hydroxylase that targets highly conserved residues within two closely related TRAFAC GTPases, DRG1 and DRG2. The dependence of JMJD7 on Fe(II) and 2OG is consistent with the assignment of JMJD7 as a 2OG oxygenase. Still, structural validation would provide direct evidence for the presence of a core DSBH fold that defines the 2OG/JmjC oxygenase catalytic domain. The finding that JMJD7 catalyses *(3S)*-lysyl hydroxylation is of interest, as it is the first time that this PTM has been observed in humans. Furthermore, it expands the repertoire of 2OG oxygenase-catalysed modifications of basic residues. To date, 2OG oxygenases have been reported to catalyse lysyl C4/C5 hydroxylation[109,133,204], N^{ε}-methyl lysyl demethylation[13], histidinyl hydroxylation[75,122], arginyl C3 hydroxylation[122], and *N*-methyl arginyl demethylation[105] (including by some recombinant KDMs[100]).

Among 2OG oxygenase-catalysed modifications of basic residues, lysyl hydroxylation is among the most prominent. The discovery of JMJD7-catalysed C3 lysyl hydroxylation further expands known lysyl hydroxylation PTMs, which currently include the C3, C4, and C5 (5*S* and 5*R*) lysyl side chain positions (**Figure 2.36**). Seven different lysine residue oxidations catalysed by 2OG oxygenases have now been identified (as the JmjC KDMs catalyse demethylation of all three N^{ε}-methylation states). The remaining unidentified C3 and C4 stereoisomers remain to be identified and characterised, though they are also likely catalysed by 2OG oxygenases given

that all presently identified lysyl hydroxylases are 2OG oxygenases. Lysyl oxidation at the C2 and C6 positions has not yet been identified, though it is likely that these hydroxylated products are unstable. Beyond expanding reported 2OG oxygenase catalysis, this is the first time (3*S*)-lysyl hydroxylation has been reported in humans to date.

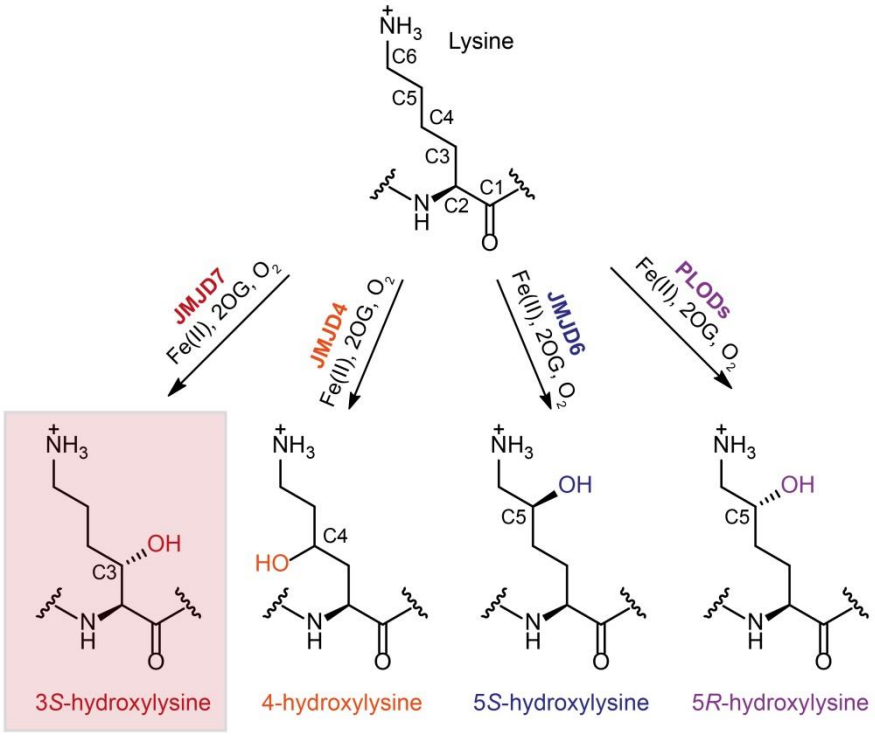


Figure 2.36. Schematic of known 2OG oxygenase-catalysed lysyl hydroxylations.

Beyond being a major target for 2OG oxygenase-catalysed hydroxylation and demethylation, lysine is one of the most modified amino acids. Lysine side chains have been reported to undergo several enzyme-catalysed PTMs, including hydroxylation, acetylation, methylation, ubiquitination, sumoylation, and phosphorylation (among others)[205]. A recent review by Azevedo and Saiardi (2016)[205] provides an overview of enzyme-catalysed lysine modifications focusing on the variability of these modifications with a key speculative question in mind: "Why always lysine?" The authors highlight the roles of the long, flexible lysine side chain and the intrinsic chemical reactivity of its positively charged ε-amino group (under physiological conditions). The long side chain and hydrophilic ε-amino group is likely why lysine tends to be solvent-exposed and consequently accessible to modifications. The reactive lysine side chain is

essential for maintaining protein stability and mediating interactions with proteins (e.g. histone-KDM interactions[21]), nucleic acids (e.g. histone-DNA interactions[206]) and other small molecules (e.g. 2OG with the characteristic lysine 2OG-binding residue found in the JmjC 2OG oxygenase active site[21])[205]. Arginine is the most chemically similar amino acid to lysine. Though arginine and lysine share a relatively long side chain and roles in maintaining protein structure, the positively charged guanidinium group (under physiological conditions), which contains three asymmetric nitrogen atoms, is likely to be less flexible and confer stricter structural rigidity given its potential to create up to three hydrogen bonding/electrostatic interactions in three different directions[205]. Nevertheless, it is likely that as for lysine, there are many PTMs that remain to be discovered for arginine. This may be true given that some recombinant JmjC KDMs have exhibited arginyl demethylase activity on peptides[100], and recombinant JMJD5 exhibits arginyl hydroxylation on peptide substrates (Dr Sarah Wilkins, University of Oxford, personal communication).

2.4.5 JMJD7 and disease

In addition to understanding JMJD7 and JMJD7-catalysed lysyl hydroxylation in normal physiology, it is of interest for future studies to investigate the role of JMJD7 in disease. Further to the identified links between *JMJD7* gene mutations and cancer described in **Section 2.1.3**, genetic studies have identified two missense mutations in the *JMJD7* gene, resulting in M160V[139] and R260C[140] amino acid substitutions. The M160V substitution was identified in a genetic sequencing study of patients with severe intellectual disabilities[139], while the R260C substitution was identified as a rare variant found in a high-risk autism family[140]. Overall, the links between JMJD7/DRGs with protein synthesis, intellectual disability/autism, and cancer provide ample reason for future studies to undertake JMJD7 structural studies and investigate the functional role of JMJD7 and the JMJD7-DRG pathways in normal physiology and disease.

2.4.6 Potential roles for JMJD7-catalysed hydroxylation

As described in **Chapter 1**, hydroxylation is a small, chemically neutral modification that has important biological consequences in collagen biosynthesis, oxygen sensing, and protein biosynthesis. Hydroxylation of prolyl residues in procollagen increases the stability of the collagen triple helix, enabling collagen to maintain its structural integrity[42]. Hydroxylation can also act as a "switch" by signalling for the transcriptional regulation of genes that enable the cell to adapt to hypoxia[61] (as exemplified by PHD2-catalysed hydroxylation of the HIF pathway). Furthermore, lysine PTMs beyond hydroxylation have been identified to play essential roles in epigenetics by mediating histone-DNA interactions; for example, the acetylation of lysine residues, which weakens histone-DNA interactions by neutralising the positive lysine side chain and consequently "relaxing" chromatin structure, is thought to enable transcriptional activitation[206].

In the case of JMJD7, an understanding of the functional effect of DRG1-2 hydroxylation and the molecular mechanism by which hydroxylation acts remain to be explored. Current understanding is in part limited by the incomplete functional characterisation of DRG1-2. While DRG1-2 have been identified to interact with components of translational machinery (e.g. ribosomes), their functional roles remain largely unknown. Thus, a more complete characterisation of these JMJD7 substrates would offer additional insights into the role of JMJD7 itself. Future studies that wish to further understand the physiological role of JMJD7 should focus on characterising DRG1-2. In addition, future structural studies of JMJD7 (bound with substrate) and the investigation of the effect of JMJD7 clinically-relevant substitutions may also provide insights into the functional role of JMJD7.

3 STRUCTURAL CHARACTERISATION OF JUMONJI DOMAIN-CONTAINING 7

3.1 Introduction

** The structural and biophysical studies on JMJD7 described in **Chapter 3** have been prepared in a modified form as a co-first author manuscript, which is currently under review for publication in* Nature Chemical Biology*: Markolovic, S. and Zhuang, Q. et al. (2017) (3S)-Lysyl hydroxylation of TRAFAC GTPases is catalyzed by the human Jumonji-C oxygenase JMJD7.*

The biochemical studies described in **Chapter 2** define human JMJD7 as a 2OG oxygenase based on its sequence similarity to members of the JmjC subfamily and dependence on Fe(II) and 2OG for catalytic activity. However, 2OG oxygenases are *structurally* defined by a conserved double-stranded β-helix (DSBH; or jelly-roll/cupin/JmjC) fold, which comprises the Fe(II)-binding triad (HXD/E…H motif) and subfamily conserved 2OG-binding residues[17]. Furthermore, the JmjC subfamily is structurally distinct from other members of the 2OG oxygenase superfamily; for example, the JmjC oxygenases tend to bind the 2OG C5 carboxylate via a conserved lysine residue located on the βIV strand of the DSBH. As described in **Chapter 1**, the JmjC oxygenases can be divided into two functionally distinct groups – the hydroxylases and KDMs. Structural differences between the JmjC hydroxylases and KDMs, including in domain architecture, oligomerisation, and active site substrate binding features, can offer insights into enzyme catalytic function and substrate binding[21]. Thus, to verify the biochemical assignment of JMJD7 as a JmjC hydroxylase and further identify structural features that facilitate (3*S*)-lysyl hydroxylase catalysis, structural studies were pursued.

3.1.1 Aims of this chapter

Given the structural definition of 2OG oxygenases and the controversial assignments of some JmjC hydroxylases as KDMs, it was considered important to structurally characterise JMJD7. Because no structures of human JMJD7 are currently available, a structural understanding of JMJD7 can provide insights into JMJD7:substrate binding, clinically-relevant substitutions, and the future development of selective JMJD7 therapeutics.

The work described in this chapter aims to provide a structural characterisation of JMJD7 via:

(1) the determination of a JMJD7 and/or JMJD7:DRG crystal structure;

(2) the functional effects of JMJD7 oligomerisation; and

(3) the structure-function relationship of JMJD7, as compared with other JmjC oxygenases.

The results presented in this chapter aim to lay the foundation for future JMJD7 structural studies and the development of selective JMJD7 therapeutics and/or inhibitors.

3.2 Crystal structures of human JMJD7 and JMJD7-R260C

3.2.1 Crystallisation trials and hit identification

**Crystallisation trials were done in collaboration with Dr Rasheduzzaman Chowdhury (Schofield Laboratory, University of Oxford).*

3.2.1.1 Full-length His_6-JMJD7 and His_6-JMJD7-R260C

Crystallisation trials with full-length His_6-JMJD7 (residues 1-316) were undertaken with the addition of $MnCl_2$ and 2OG or the 2OG-mimetic inhibitor, NOG. $MnCl_2$ was chosen because it is an inert Fe(II) substitute commonly used to prevent uncoupled 2OG turnover and enable crystallographically stable complexes of 2OG oxygenases, including for several JmjC hydroxylases (e.g. MINA53 [PDB ID: 4BU2], NO66 [4CCK] and bacterial oxygenase YcfD [4CCL][102]) and other 2OG oxygenases (e.g. PHD2 [3HQR][20], *P. putida* PPHD [4J25][136], *S.*

cerevisiae Tpa1p [4NHL][178], and several KDMs). Crystallisation trials were concurrently undertaken with a full-length His_6-JMJD7-R260C construct, which was chosen due to its associations with high-risk autism families[140]. Note that the JMJD7 and JMJD7-R260C proteins used for crystallisation were purified with an *N*-terminal His_6-tag, which was not removed for crystallography, as described in **Section 2.2.4.**

Crystallisation screening experiments began with several commercial screens at 20ºC, including the PEG/Ion, Crystal Screen, Index HT, and SaltRx (Hampton Research), JCSG-plus (Molecular Dimensions), and Wizard (Emerald, now Rigaku) crystal screens. Several JMJD7 crystals were observed in SaltRx D7 (1.5 M sodium nitrate and 0.1 M sodium acetate trihydrate pH 4.6); one crystal was observed in Wizard C11 [20% (v/v) 1,4-butanediol and 0.1 M sodium acetate/acetic acid pH 4.5] (**Figure 3.1**). Attempts to co-crystallise JMJD7 with $DRG1_{16\text{-}40}$, $DRG2_{15\text{-}39}$, and DRG_{KNKAT} peptides using the commercial screens described above did not yield any suitable crystals. Crystals from SaltRx D7 were soaked with the $DRG1_{16\text{-}40}$ or $DRG2_{15\text{-}39}$ peptide in an attempt to solve an enzyme:substrate complex structure; however, peptide soaking did not reveal any electron density for the substrate residues. As for JMJD7, JMJD7-R260C crystals were also observed in SaltRx well D7, but with different morphology (orthorhombic).

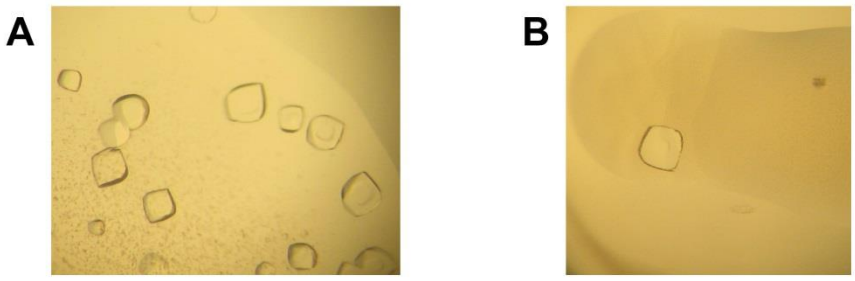


Figure 3.1. JMJD7.Mn(II).2OG initial crystal hits from (A) SaltRx D7 and (B) Wizard C11. Images are not to scale.

Further optimisation of the initial crystal hits for JMJD7 and JMJD7-R260C with $MnCl_2$ and/or 2OG or NOG from SaltRx D7 involved the use of the commercial crystallisation reagent (SaltRx 1 #43: 1.5 M sodium nitrate, 0.1 M sodium acetate, pH 4.6 solution) and commercial optimisation screens, including Additive, Detergent, and Silver Bullets (Hampton Research)

and an in-house custom ligand screen (made by Dr Rasheduzzaman Chowdhury). Although these attempts were not successful in obtaining crystals for JMJD7, several JMJD7-R260C crystals were observed in Silver Bullets Bio (HR2-088) well F1, which contained the following additives: 0.25% β-Nicotinamide adenine dinucleotide phosphate tetrasodium salt, 0.25% Adenosine 5'-triphosphate disodium salt hydrate, 0.25% *N*-Acetyl-*D*-galactosamine, 0.25% Gentamicin sulphate salt hydrate, 0.02 M HEPES sodium pH 6.8. These crystals had the same orthorhombic-like morphology, differing from the initial JMJD7 hits (**Figure 3.1**). Optimised conditions for JMJD7 crystallisation are summarised in **Section 7.9.1**.

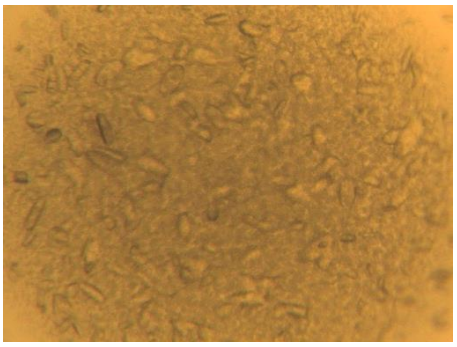


Figure 3.2. JMJD7-R260C.Mn(II).2OG crystal optimisation from Silver Bullets Bio F1. Image brightened and not to scale.

3.2.2 Crystallisation and data collection

**Data set collection, structure determination, and structure refinement were performed by Dr Rasheduzzaman Chowdhury (Schofield Laboratory, University of Oxford).*

JMJD7 crystals were harvested, cryoprotected with 25% glycerol in the mother liquor, and flash frozen in liquid nitrogen for diffraction studies. Data were collected using single crystals at 100 K at the Diamond Light Source (DLS; Didcot, UK) beam lines and processed using HKL2000[207].

3.2.2.1 JMJD7 crystallisation data collection in H32 crystal form

A highly redundant dataset to 3.0 Å resolution from the JMJD7 crystals shown in **Figure 3.1A** was collected in the rhombohedral *H*32 space group with four JMJD7 molecules in the asymmetric unit (ASU) (**Table 3.1**). The data were strongly anisotropic, with resolution limits

of 3.5 Å, 3.5 Å, and 3.0 Å along the a*, b* and c* axes (at an F/σ cut-off of 3). The wildtype JMJD7 structure was solved by molecular replacement using PHASER[208] and a search model generated from an ensemble of models using the crystal structures of the related enzymes, FIH (PDB ID: 1H2K)[84], TYW5 (PDB ID: 3AL5)[103], JMJD5 (PDB ID: 4GJZ)[156], and JMJD6 (PDB ID: 3K2O)[209] as templates. Only JMJD7 residues that are highly conserved in these JmjC hydroxylases were included in the homology models to maximise the degree of accuracy. Details for initial and final rounds of model building and refinement can be found in **Section 7.9.2**. Mn(II), 2OG, and water molecules were modelled in the final refinement stages based on the electron density map (F_{obs}-F_{calc}) as described in **Section 7.9.2**

3.2.2.2 JMJD7-R260C crystallisation data collection in $P2_12_12$ crystal form

A complete dataset to 2.2 Å resolution from the JMJD7-R260C crystals shown in **Figure 3.2** was collected in the orthorhombic $P2_12_12$ space group with two JMJD7 molecules in the ASU (**Table 3.1**). The structure was solved by molecular replacement using PHASER[208] and wildtype JMJD7 structure described in **Section 3.2.2.1** as a search model. Initial model building and refinement were performed using CNS[210]/PHENIX[211] until R_{factor}/R_{free} converged with iterative cycles of model building using COOT[212]. Additional details for model building, structure refinement, and modelling of Mn(II), 2OG, and water molecules are described in **Section 7.9.2**.

3.2.3 Structural comparison of JMJD7 with JMJD7-R260C

A comparison of the structures of JMJD7 and JMJD7-R260C reveals no observable differences between their overall structural folds. JMJD7, which crystallised with four molecules per ASU, is an apparent crystallographic dimer of dimers, while JMJD7-260C is a dimer with two molecules per ASU. As shown in **Figure 3.3**, JMJD7-R260C overlays in a nearly identical manner with one set of the JMJD7 dimers. In addition to having no observable effect on the overall JMJD7 structural fold, the R260C substitution has a negligible effect on the structure of

the monomer/dimer (**Figure 3.4**). Interestingly, however, the R260C substitution enabled different crystal packing (orthorhombic compared with rhombohedral) is likely not possible with wildtype JMJD7 due to an apparent steric clash between R260 and two surface-exposed residues (E192 and R264; **Figure 3.5**). Because no substantial differences were identified between the two structures [as evidenced by low Cα root-mean-square deviation (RMSD) values, **Figure 3.6**] and given that the R260C variant yielded a higher resolution structure, subsequent analyses were done using the JMJD7-R260C structure (i.e. all subsequent figures were made using the JMJD7-R260C structure).

Table 3.1. Data collection and refinement statistics for JMJD7 and JMJD7-R260C.

Datasets	JMJD7.Mn.2OG	JMJD7-R260C.Mn.2OG
PDB acquisition codes	5NFN	5NFO
Data collection		
Beamline (Wavelength, Å)	DLS I04 (0.9795)	DLS I03 (0.9795)
Detector	Pilatus 6M-F	Pilatus3 6M
Data processing	HKL2000	HKL2000
Space group	*H*32	$P2_12_12$
Cell dimensions		
***a*, *b*, *c* (Å)**	207.24, 207.24, 211.81	136.99, 60.93, 85.86
***α*, *β*, *γ* (°)**	90, 90, 120	90, 90, 90
No. of molecules/ ASU	4	2
No. reflections	35036 (3435)*	36904 (3611)*
Resolution (Å)	48.46 – 3.00 (3.11 - 3.00)*	49.69 – 2.21 (2.29 - 2.21)*
R_{sym} **or** R_{merge}**	0.17 (1.2)*	0.13 (1.0)*
I*/σ*I	14.5 (1.5)*	14.8 (2.0)*
CC(1/2)	0.941 (0.793)*	0.997 (0.723)*
Completeness (%)	100 (100)*	100 (100)*
Redundancy	15.0 (15.0)*	9.0 (8.9)*
Refinement		
Resolution (Å)	48.46 – 3.00 (3.11 - 3.00)*	49.69 – 2.21 (2.29 - 2.21)*
R_{work}/ R_{free}‡	0.245/ 0.263	0.179 / 0.217
R.m.s deviations		
-Bond lengths (Å)	0.006	0.005
-Bond angles (°)	1.176	0.805

*Highest resolution shell shown in parenthesis.

**R_{sym} = $\sum|I-\langle I\rangle|/\sum I$, where I is the intensity of an individual measurement and $\langle I\rangle$ is the average intensity from multiple observations.

‡R_{factor} = $\sum_{hkl}||F_{obs}(hkl)| - k\,|F_{calc}(hkl)||/\sum_{hkl}|F_{obs}(hkl)|$ for the working set of reflections; R_{free} is the R_{factor} for ~5% of the reflections excluded from refinement.

Table kindly prepared and provided by Dr Rasheduzzaman Chowdhury (University of Oxford).

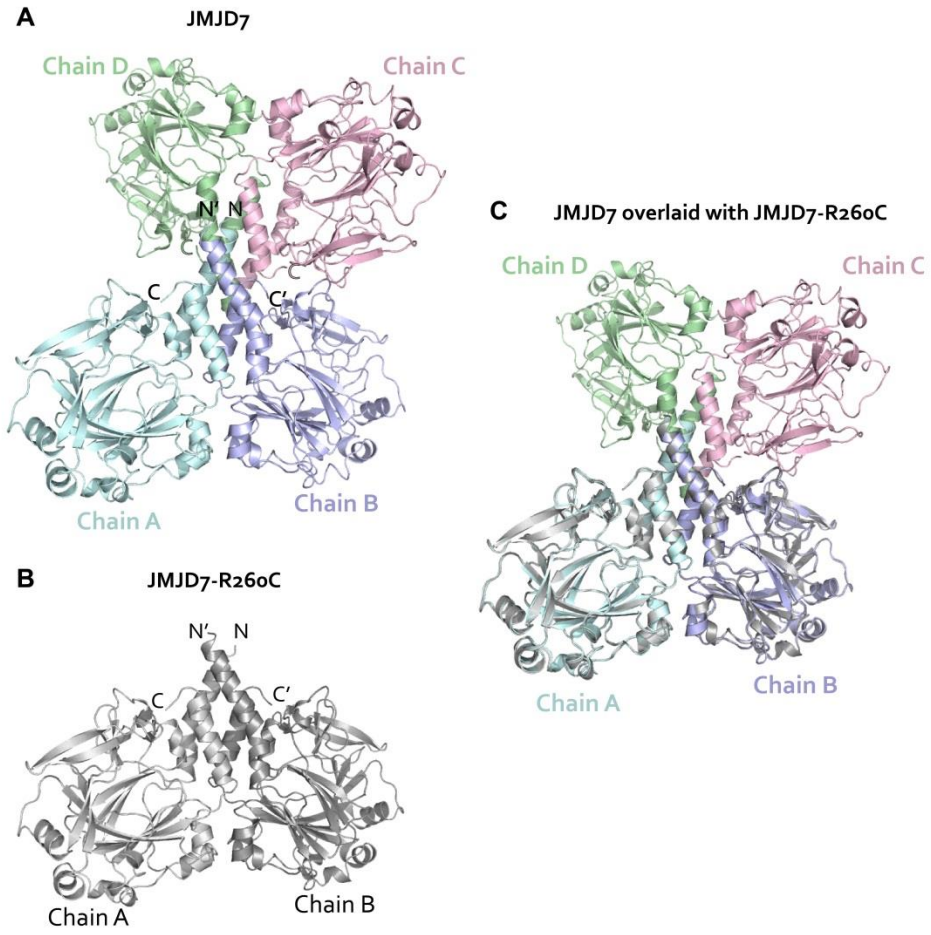


Figure 3.3. Views from crystal structures of JMJD7 complexes. ***(A)*** JMJD7, which crystallised in the *H*32 space group (3.0 Å), contains four molecules per ASU, an apparent dimer of dimers. ***(B)*** JMJD7-R260C crystallised in the $P2_12_12$ space group (2.2 Å) with two molecules per ASU, a crystallographic dimer. ***(C)*** An overlay of wildtype JMJD7 and JMJD7-R260C structures reveals no substantial differences in their overall structures.

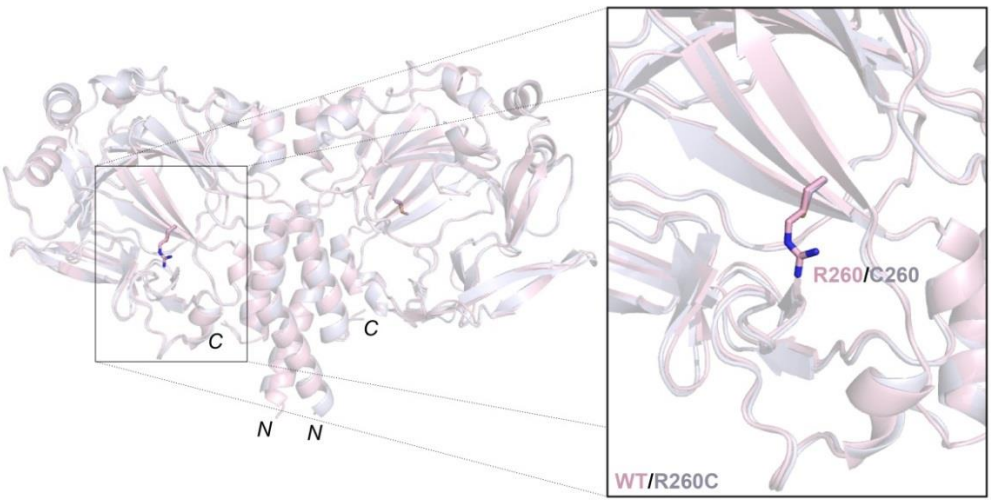


Figure 3.4. Comparison of JMJD7 (pink) and JMJD7-R260C (light purple). The R260C substitution has no substantial effect on neither the overall JMJD7 overall nor in the local vicinity of the R260 position (zoomed region, black outline).

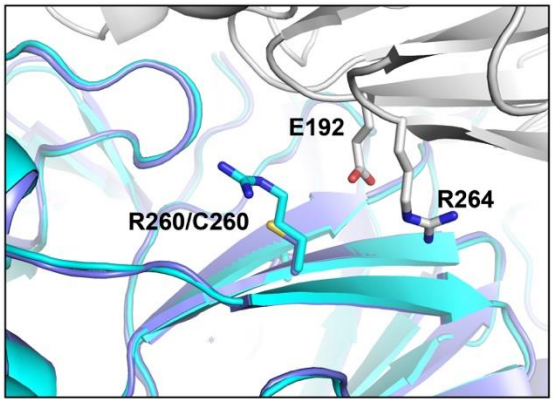


Figure 3.5. Wildtype JMJD7 (cyan) and JMJD7-R260C (light purple) crystal packing differs. Symmetry mates were generated for JMJD7-R260C within 4 Å (grey ribbon). Wildtype JMJD7 crystals cannot pack in the same manner as JMJD7-R260C because of a steric clash, which likely arises between the flexible R260 residue and E192/R264 residues from a symmetry mate (grey). Note that while this figure does not show steric clash, it is a very likely possibility given that R260 can have multiple side chain conformations. However, steric clash is neither observed nor anticipated with the R260C substitution.

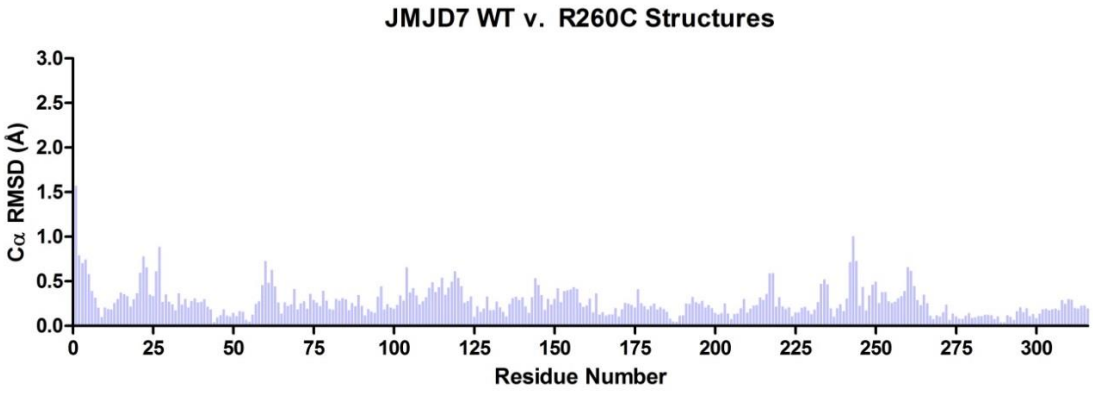


Figure 3.6. Plot of Ca RMSD (Å) values between wildtype JMJD7 and JMJD7-R260C structures indicates little differences between the two structures, with the exception of the flexible N-terminal M1 residue.

3.2.4 Overall structure of JMJD7

Analysis of the crystal structure of JMJD7 reveals that each JMJD7 molecule is comprised of nine α-helices, 18 β-strands, and two 3_{10} helices (**Figure 3.7**). JMJD7 forms an apparent crystallographic dimer mediated by interactions between *N*- and *C*-terminal α-helices (α1, α2, and α9) and disulphide bond formation between residues C47 of each monomer. The JMJD7 oligomerisation domain, which is unique among JmjC hydroxylases, is discussed in greater detail in **Section 3.3**.

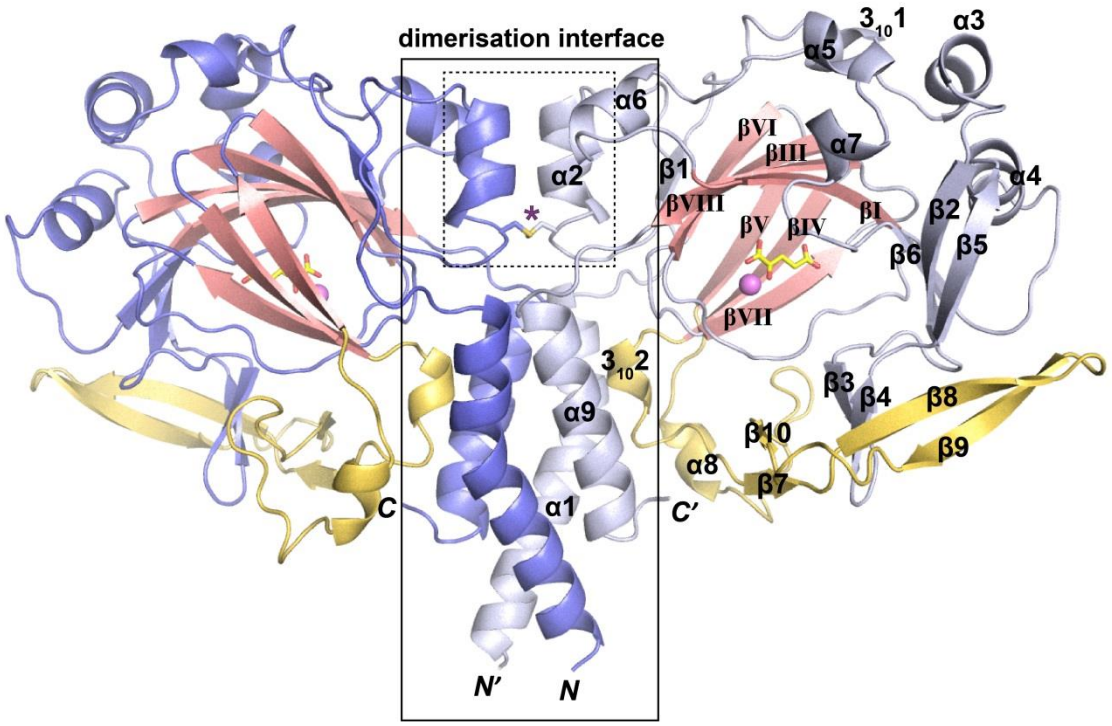


Figure 3.7. View from a crystal structure of human JMJD7 in complex with Mn(II) and 2OG. JMJD7 contains a 2OG oxygenase-characteristic double-stranded β-helix (DSBH) core fold, highlighted in salmon (βI-βVIII). Mn(II) (pink sphere), an inert substitute for Fe(II), and 2OG (yellow sticks) are bound in the active site. The apparent dimerisation interface is highlighted (black outline) and the disulphide bond formed between residues C47 of each monomer is indicated by a violet star. The βIV-V strand, which is distinctly longer for JMJD7 than most JmjC hydroxylases, is highlighted in light orange. Secondary structure features are labelled in bold.

As for all structurally characterised 2OG oxygenases, JMJD7 contains the characteristic distorted DSBH core fold[17,22]. The DSBH fold is comprised of eight β-strands arranged as two antiparallel β-sheets – the major sheet (βI, βVIII, βIII, βVI) and minor sheet (βII, βVII, βIV, βV). The major and minor sheets pack to form a barrel-like structure that encloses the catalytic active site, the open end of which supports the Fe(II)- and 2OG-binding sites. As predicted from sequence alignments of JMJD7 with other JmjC hydroxylases (**Figure 2.6**), the βIV-V insert, which is comprised of residues P199-A258 (3_{10}2-β7-β8-β9-β10-α8), is much longer than for most hydroxylases (**Figure 3.7**). Because there are examples of structurally characterised 2OG oxygenases with 'longer' βIV-V inserts that play important roles in substrate binding (e.g. FIH and clavaminate synthase, CAS)[22], it is possible that a similar role might exist for JMJD7; as yet, the role of the JMJD7 βIV-V insert is unknown.

3.2.5 The JMJD7 active site

Consistent with other structurally characterised 2OG oxygenases, the JMJD7 active site contains the Fe(II)- and 2OG-binding residues[17,22]. As predicted from sequence alignments and structural predictions (**Figure 2.6**), JMJD7 contains the conserved HXD/E…H triad of metal-binding residues (His178, Asp180, and His277); together with the oxalyl group of 2OG and a water molecule, the Fe(II)-binding residues coordinate the metal in an octahedral geometry (**Figure 3.8**).

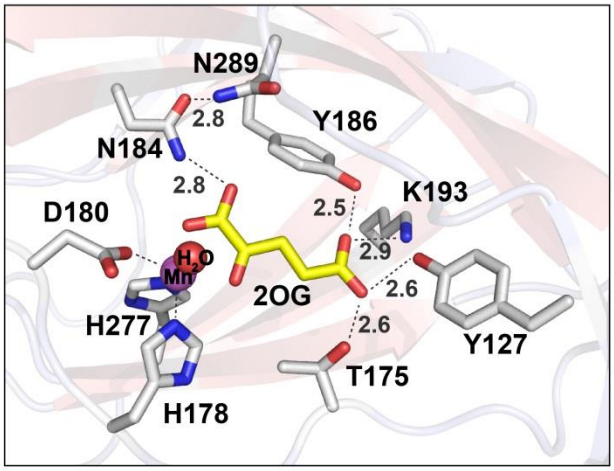


Figure 3.8. Zoomed view of the active site from a crystal structure of human JMJD7 complexed with Mn(II) and 2OG. Mn(II) [violet sphere, inert substitute for Fe(II)] is octahedrally coordinated by conserved Fe(II)-binding residues (H178, D180, and H277), the 2OG (yellow sticks) oxalyl group, and a water (red sphere). Apparently tight 2OG-binding is stabilised by hydrogen bonding/electrostatic interactions between the 2OG C5 carboxylate and Y127, T175, Y186, and K193 side chains. The 2OG C1 carboxylate is further stabilised by a hydrogen bonding network with N184 and N289. Dotted lines represent H-bond distances in Å.

2OG-binding residues are located at subfamily conserved sites within the DSBH fold[22]. JmjC oxygenases are characterised by a conserved lysine residue near the *N*-terminus of the βIV strand[22] [with the exception of some JmjC KDMs; e.g. UTX[89] and KDM3B (PDB ID: 4LXL) have 2OG-binding lysine residues positioned near the *C*-terminus of the βI strand]. By contrast, non-JmjC 2OG oxygenases have a 2OG-binding lysine or arginine residue located on the βVIII strand or an arginine residue on the βIV strand[22] (**Figure 1.3**). The JMJD7 C5 carboxylate group of 2OG interacts via electrostatic and hydrogen bonding interactions with the apparently conserved JmjC-subfamily-specific lysine residue (Lys193) situated on βIV (**Figure 3.8**), consistent with the bioinformatic assignment of JMJD7 as a JmjC 2OG oxygenase[12] (**Figure**

2.6). While most 2OG oxygenases typically have two or three polar side chains positioned to interact with the 2OG C5 carboxylate group, JMJD7 has four such side chains; the 2OG C5 carboxylate group is positioned to form additional hydrogen bonding interactions with the hydroxyl groups of Y127, T175 and Y186. The C1 2OG carboxylate group is further supported in the active site by a hydrogen bonding network comprised of interactions with the amide groups of N184 and N289 (**Figure 3.8**). The apparently tight binding of 2OG is consistent with the low level of substrate uncoupled turnover of 2OG observed by ^{1}H-NMR (**Section 2.2.6**).

3.3 Investigation of the JMJD7 oligomerisation domain

3.3.1 Structural analysis of the JMJD7 oligomerisation domain

Like other structurally characterised JmjC hydroxylases, including FIH[213], TYW5[103], and the RiOXs[102,214,215], JMJD7 is a crystallographic oligomer, specifically a dimer. The oligomerisation interface of JMJD7 is formed via apparent hydrophobic interactions involving α-helices from both the *N*- and *C*-terminal regions (α1, α2, and α9). Hydrophobic contacts mediate interactions between α1-α1' (Ala2, Leu6, and Val9), α2-α2' (Leu39 and Tyr42), α1-α9' (Leu13, Phe16, Tyr302, Leu305, and Leu309) and α9-α9' (Tyr302, Leu305, Le306, and Leu309). Furthermore, a disulphide bond is formed between residues C47 of each monomer on a two-fold symmetry axis (**Figure 3.9**).

The JMJD7 oligomerisation interface is unique among JmjC hydroxylases, which tend to oligomerise via interactions between *C*-terminal α-helices[21] (a more detailed comparison of JMJD7 with JmjC oxygenases is discussed in **Section 3.4.1**). Moreover, there have been no reports of disulphide bonds involved in mediating 2OG oxygenase oligomerisaton to date. Given the crystallographic evidence revealing the unique dimerisation interface of JMJD7, efforts were made to validate JMJD7 as a dimer in its near-native state and investigate the apparent contribution of each region to JMJD7 dimerisation.

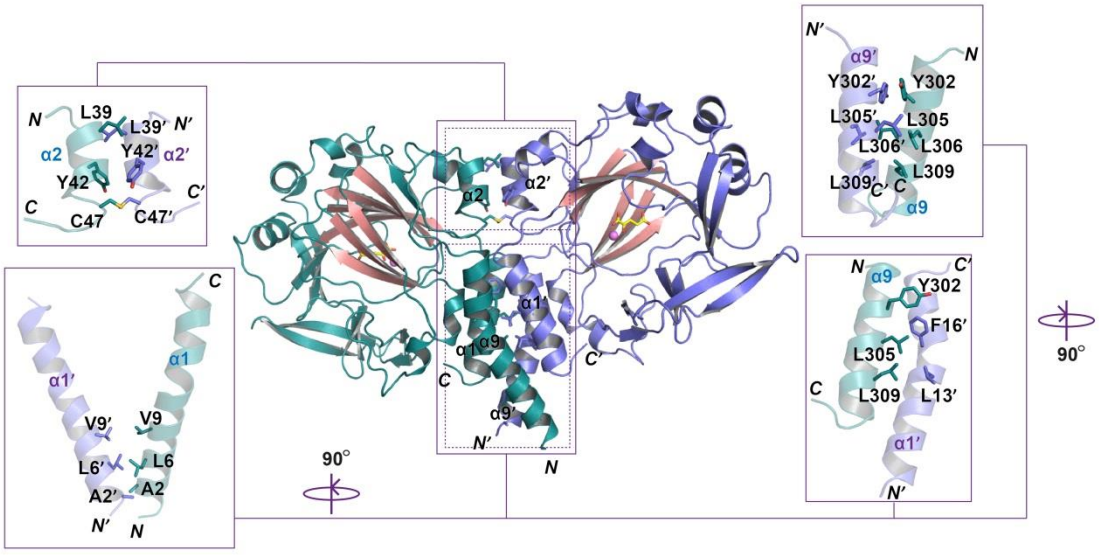


Figure 3.9. Detailed view of the JMJD7 dimerisation interface. Crystallographic analyses indicate that JMJD7 dimerisation is apparently mediated by hydrophobic interactions between highlighted residues in α_1-α_1′, α_2-α_2′, α_1-α_9′, and α_9-α_9′ and disulphide bond formation between residue C47 of each monomer (on α_2).

3.3.2 Design of JMJD7 constructs for investigation of JMJD7 oligomerisation status

Based on crystallographic analyses of the JMJD7 dimerisation interface (**Figure 3.9**), JMJD7 constructs were designed and produced with *N*- and *C*-terminal truncations with and without alanine substitution of C47 for investigation (**Table 3.2**). JMJD7 truncations representing the JMJD7 oligomerisation constructs are outlined in **Table 3.2** and illustrated in **Figure 3.10**.

Table 3.2. JMJD7 oligomerisation constructs investigated in this study.

Construct Name	Description
JMJD7$_{1\text{-}316}$	Full-length, wildtype JMJD7; residues M1-D316
JMJD7-C47A$_{1\text{-}316}$	As above, with Ala substitution of Cys47
JMJD7$_{27\text{-}316}$	*N*-terminally truncated JMJD7 with deletion of α1; residues L27-D316
JMJD7-C47A$_{27\text{-}316}$	As above, with Ala substitution of Cys47
JMJD7$_{1\text{-}295}$	*C*-terminally truncated JMJD7 with deletion of α9; residues M1-E295
JMJD7-C47A$_{1\text{-}295}$	As above, with Ala substitution of Cys47
JMJD7$_{27\text{-}295}$	*N*- and *C*-terminally truncated JMJD7 with deletions of α1 and α9; residues L27-E295
JMJD7-C47A$_{27\text{-}295}$	As above, with Ala substitution of Cys47

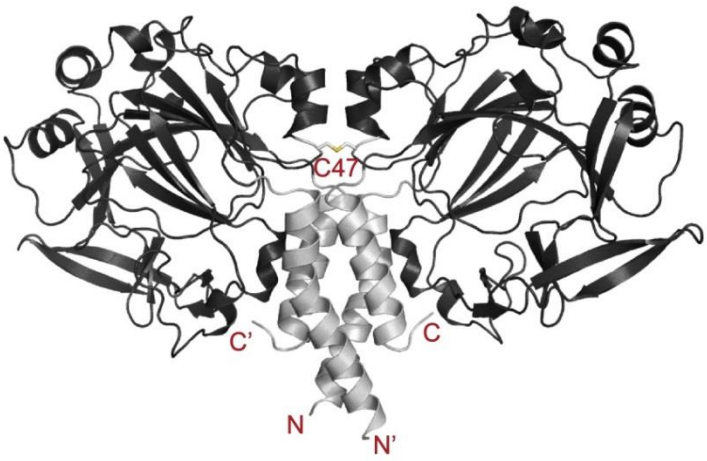


Figure 3.10. View from a JMJD7 crystal structure highlighting the JMJD7 oligomerisation constructs investigated this in this study. The highlighted (grey) *N*- and *C*-terminal α-helices were systematically truncated with or without C47A substitution for oligomerisation studies as described in **Table 3.2**: $JMJD7_{1\text{-}316}$, JMJD7-C47$A_{1\text{-}316}$, $JMJD7_{27\text{-}316}$, JMJD7-C47$A_{27\text{-}316}$, $JMJD7_{1\text{-}295}$, JMJD7-C47$A_{1\text{-}295}$, $JMJD7_{27\text{-}295}$, and JMJD7-C47$A_{27\text{-}295}$.

3.3.3 Cloning, expression, and purification of JMJD7 oligomerisation constructs

Plasmids encoding the JMJD7 constructs listed in **Table 3.2** were cloned by ligation independent cloning (LIC; **Section 7.2.3**) using the pET28a(+)-JMJD7 plasmid as a template and the custom primers in **Table 7.1**. Site-directed mutagenesis (SDM; **Section 7.2.5**) was performed to generate the C47A variant of each construct using the primers in **Table 7.1**. All proteins were produced in *E. coli* and purified as described for JMJD7 (**Section 2.2.4**). Notably, purification of the $JMJD7_{1\text{-}295}$ and $JMJD7_{27\text{-}295}$ constructs by SEC resulted in two major peaks, (A) and (B), likely corresponding to two different oligomeric states (**Figure 3.11**); one major peak, corresponding to either Peak (A) or (B), was observed in the SEC UV trace for all other constructs. Note that (A) and (B) thus correspond to the distinct peaks following SEC purification of $JMJD7_{1\text{-}295}$ and $JMJD7_{27\text{-}295}$ (**Figure 3.11**); this nomenclature is used throughout the text and figures.

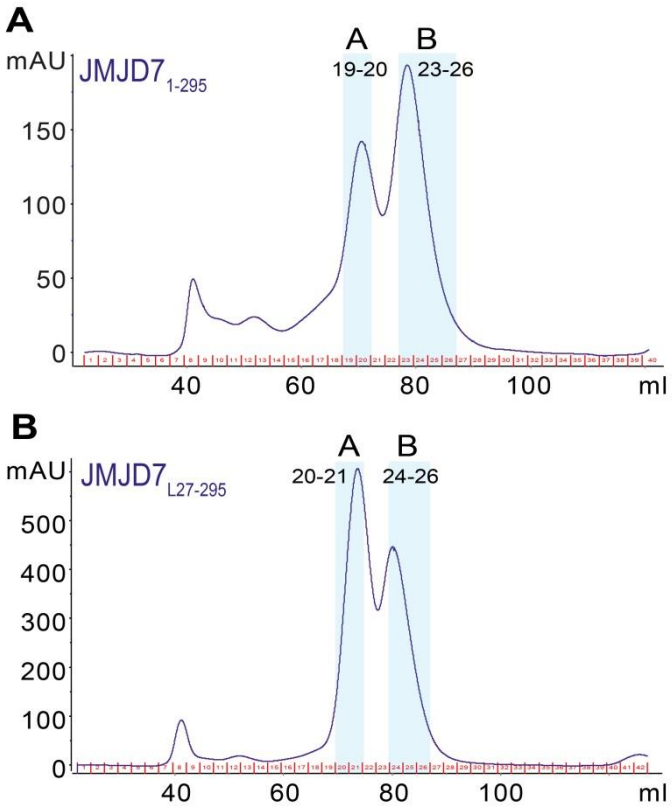


Figure 3.11. SEC UV traces of the (A) JMJD7$_{1-295}$ and (B) JMJD7$_{27-295}$ constructs. The JMJD7$_{1-295}$ and JMJD7$_{27-295}$ constructs were the only constructs that eluted as two peaks, labelled (A) and (B), during protein purification, likely representing a mixture of dimeric and monomer states in different ratios.

The purified JMJD7 oligomerisation constructs were further characterised by SDS-PAGE analyses (**Figure 3.12**) and intact protein LC-MS (**Figure 3.13**).

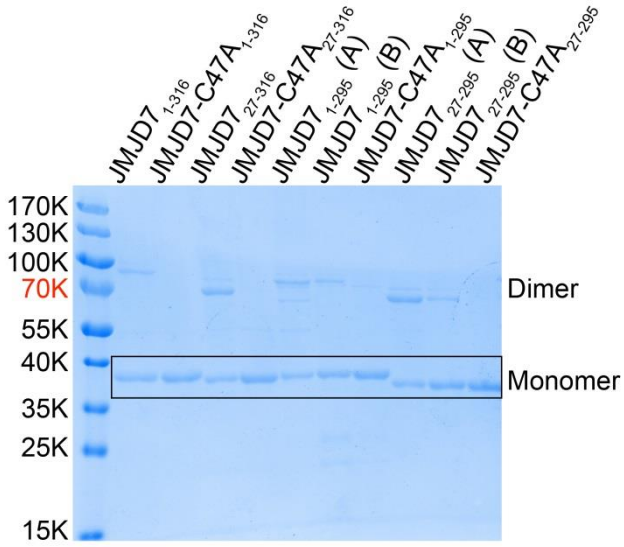


Figure 3.12. Coomassie-stained 12% SDS-PAGE gel corresponding to purified JMJD7 oligomerisation constructs. Bands outlined in the black box are consistent with the molecular weights of the monomeric protein forms (~33-38 kDa), while the higher molecular weight bands are consistent with dimeric JMJD7 (~66-76 kDa).

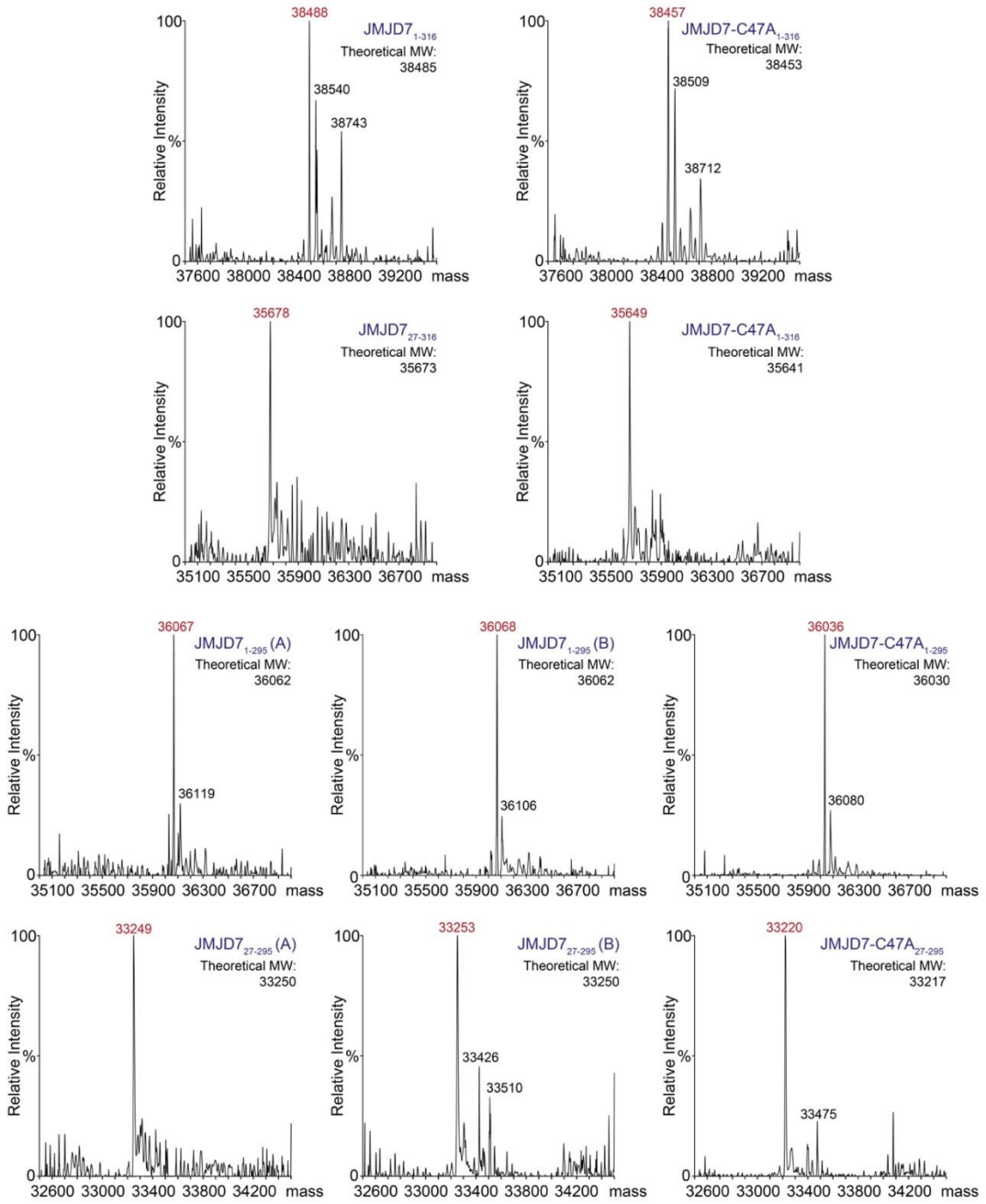


Figure 3.13. Deconvoluted intact protein mass spectra of indicated JMJD7 oligomerisation constructs under denaturing conditions. Red peaks correspond to the highest intensity ion. Theoretical molecular weights are noted in the upper right hand side of each spectrum. Note that in some cases (e.g. $JMJD7_{1\text{-}316}$, $JMJD7\text{-}C47A_{1\text{-}316}$, $JMJD7_{27\text{-}295}$ (B), and $JMJD7\text{-}C47A_{27\text{-}295}$), a peak is observed approximately +258 Da from the major peak, likely corresponding to proposed α-*N*-6-phosphogluconoylation, a common PTM reported to occur in proteins expressed in *E. coli* with an *N*-terminal (G-S-S-$[H]_6$-) His_6-tag[216] (see **Figure 3.14**).

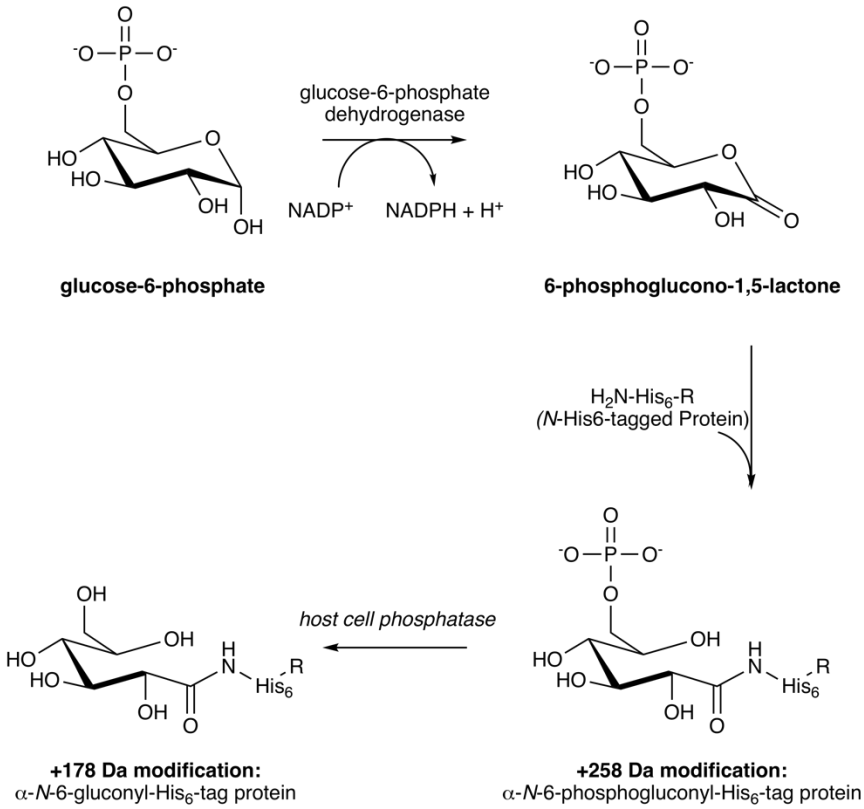


Figure 3.14. Schematic of α-N-6-phosphogluconoylation (+258 Da mass adduct) and α-N-6-gluconoylation (+178 Da mass adduct) formation. α-*N*-6-(Phospho)gluconoylation is a commonly observed PTM reported to occur in proteins expressed in *E. coli* with an *N*-terminal (G-S-S-$[H]_6$-) His_6-tag[216].

3.3.4 Investigation of the effect of truncations on JMJD7 structural fold by circular dichroism

Given the fairly large truncations made at the *N*(Δ26)- and/or *C*(Δ21)-terminal ends of JMJD7 (± alanine substitution of C47), it was important to validate that all JMJD7 constructs maintained a similar fold so that any observed differences in oligomerisation could be attributed to the regions/residues in question (i.e. changes in oligomerisation state were not a result of protein misfolding caused by the truncations). Circular dichroism (CD) studies were therefore undertaken to investigate JMJD7 secondary structure/folding in solution (as described in **Section 7.10**).

CD is a spectroscopic technique that measures the differential absorption of plane polarised light by chiral molecules. Plane polarised light is comprised of two circularly polarised

components (counter-clockwise [left-handed, L] and clockwise [right-handed, R]) of equal magnitude. CD signals are observed when a chromophore is optically active (i.e. chiral or asymmetric) and when the L and R components are absorbed at either equal or unequal extents, the latter of which results in elliptical polarisation[217–219]. Because proteins are comprised of peptide bonds (absorbing below 240 nm), measurements in the far UV CD spectra can offer information on protein secondary structure composition from the peptide bond region. Folded proteins have common types of secondary structure (e.g. α-helix, anti-parallel β-sheet, and random coil) that are highly asymmetric and give rise to characteristic CD spectra in the far UV region (180-260 nm)[217–219].

Optimal protein concentration was first determined to maximise signal-to-noise by measuring CD spectra of JMJD7 in the far UV region at the following concentrations – 0.5, 0.4, 0.3, 0.25. and 0.2 mg/mL; 0.2 mg/mL was determined to yield the best quality CD spectrum. CD analyses of the JMJD7 oligomerisation constructs revealed that all constructs maintain a stable, similar fold in solution (**Figure 3.15**).

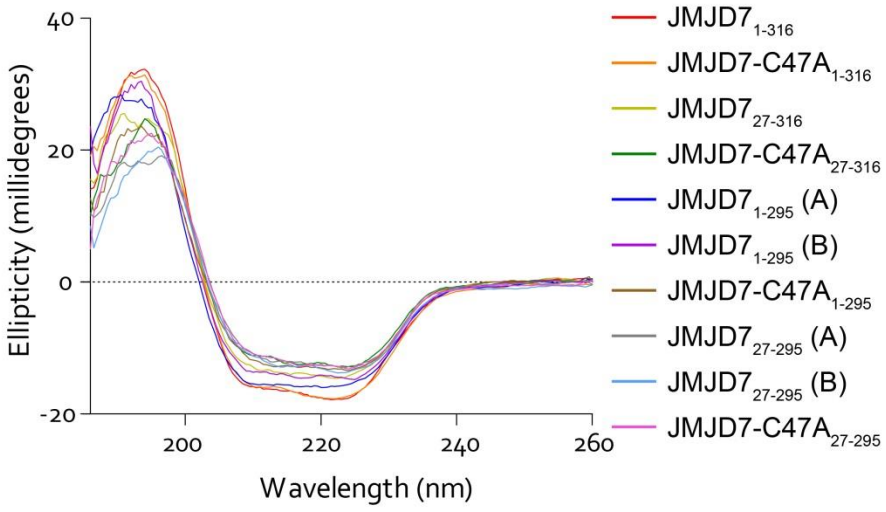


Figure 3.15. Raw CD spectra corresponding to indicated JMJD7 constructs indicate that, despite N- and/or C-terminal truncations, the constructs maintain a stable fold in solution. Spectral data were averaged (n=3) and smoothed (smooth = 8).

3.3.5 Investigation of oligomeric state of JMJD7 in solution

3.3.5.1 Non-denaturing MS analyses of JMJD7 oligomerisation constructs

**Non-denaturing MS experiments were performed with guidance from and in collaboration with Dr Weston B. Struwe (University of Oxford).*

Non-denaturing electrospray ionisation (ESI)-MS is a technique that is used for the study of intact protein complexes in their near-native state in the gas phase. Non-denaturing MS typically employs nano-ESI, which requires less solvent evaporation to bear a native protein ion. Consequently, gentler pressure and temperature conditions are required for native protein ionisation[220]. Furthermore, to promote ionisation in the near-native state, protein samples analysed by non-denaturing MS are typically in a volatile buffer, in contrast to the solutions used for denaturing MS (e.g. acetonitrile, formic acid, and methanol), which promotes protein denaturation[220].

Non-denaturing ESI-MS experiments were performed to investigate differences in oligomerisation between the JMJD7 constructs in their near-native state as described in **Section 7.8.4**. As shown in the denaturing mass spectra of wildtype JMJD7 and the oligomerisation constructs in **Figure 3.13**, all have monomeric molecular weights ranging from (~33-38 kDa). Non-denaturing ESI-MS analysis of wildtype JMJD7 in the positive ion mode indicates that JMJD7 is a dimer (~76 kDa) in its near-native state, with no evidence for higher oligomeric states observed (**Figure 3.16**). Notably, this data is consistent with the crystallographically observed dimer of dimers in the wildtype JMJD7 structure (and dimer in the JMJD7-R260C structure; **Figure 3.3**), indicating that JMJD7 is likely to be a dimer in solution.

The effects of *N*- and *C*-terminal JMJD7 truncations with or without C47A substitution on JMJD7 oligomerisation were also investigated by non-denaturing MS. As shown in **Figure 3.16**, substitution of C47A alone does not disrupt the dimeric state of JMJD7. Truncation of the *N*-terminus ($JMJD7_{27\text{-}316}$), alone or in combination with C47A substitution, disrupts JMJD7

dimerisation to a small extent, as evidenced by the appearance of charge states corresponding to the monomeric protein. Consistent with the SEC UV traces observed during purification (**Figure 3.11**), the two peaks (A) and (B), arising from the *C*- ($JMJD7_{1-295}$) and *N*-/*C*- ($JMJD7_{27-295}$) terminally truncated constructs, do indeed correspond to dimeric and monomeric forms, respectively (**Figure 3.16**). While the mixture of dimeric and monomeric states suggests that interactions involving the *C*-termini play a large role in JMJD7 dimerisation, interactions arising from the *N*- and *C*-terminal regions (α1 and α9, respectively) alone are not the only mediators of JMJD7 dimerisation. The only constructs that completely disrupted JMJD7 dimerisation were JMJD7-$C47A_{1-295}$ and JMJD7-$C47A_{27-295}$. Taken together, the non-denaturing MS data suggest that the *N*- and *C*-terminal regions, in combination with C47, all contribute to JMJD7 oligomerisation.

Based on the JMJD7 crystallographic analyses presented in **Figure 3.10**, a greater contribution from the *N*-terminal region was expected. However, it is possible that the hydrophobic interactions arising from α2-α2', which was not deleted from the *N*-terminally truncated construct, make greater contributions to the dimerisation domain. Nevertheless, the overall non-denaturing MS studies are in agreement with the crystallographic data.

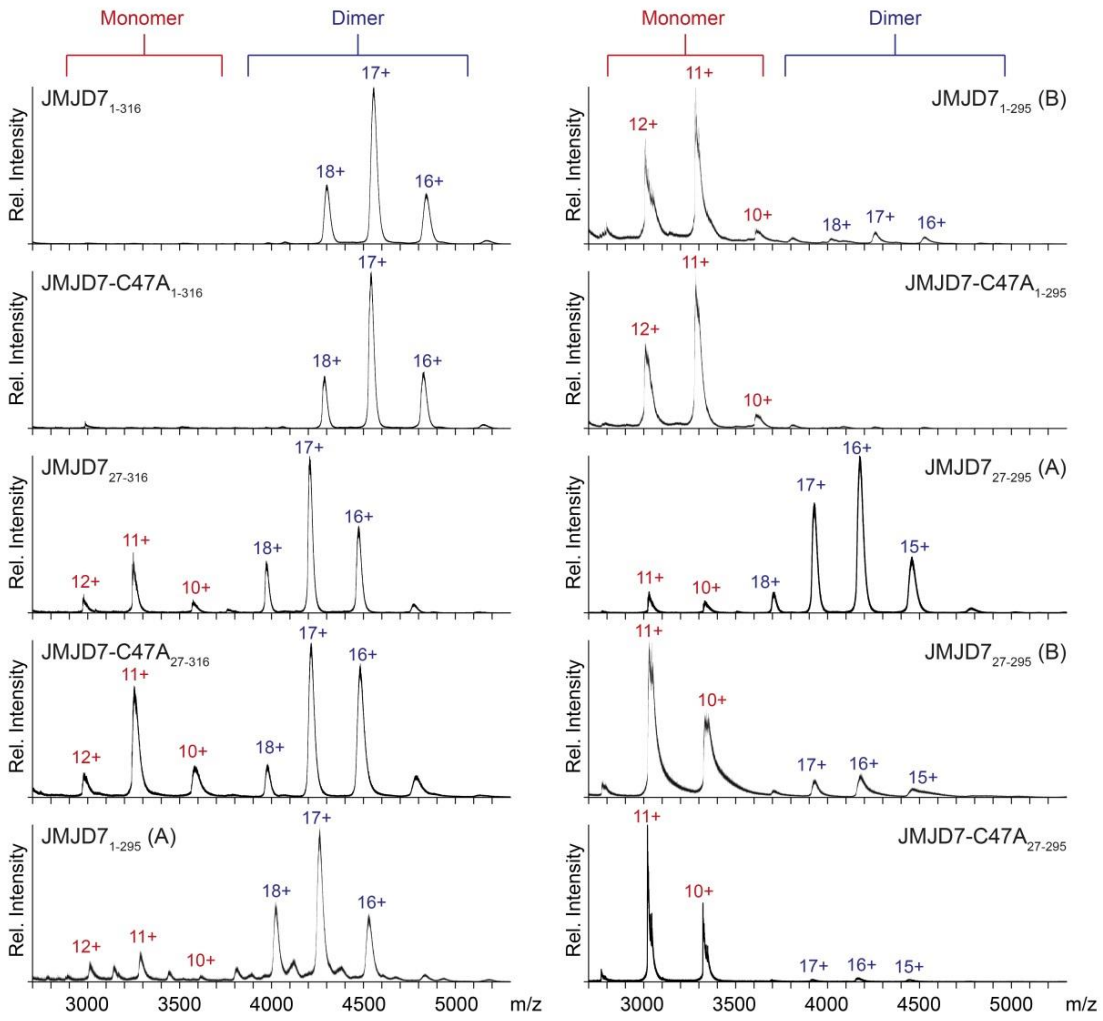


Figure 3.16. Non-denaturing ESI-mass spectra of wildtype JMJD7 and oligomerisation constructs in the positive ion mode. Protein molecular weights were calculated by assigning the correct charge state for each ion in the electrospray spectrum using Mass and Charge State Evaluation and Determination v 0.3 (http://benesch.chem.ox.ac.uk/resources.html). Given the theoretical monomeric molecular weights of wildtype JMJD7 and oligomerisation constructs (~33-38 kDa), JMJD7 constructs were assigned as dimers (~66-76 kDa) and/or monomers (~33-38 kDa). Dimeric and monomeric protein charge states are labelled in blue and red, respectively.

3.3.5.2 SEC-MALS analyses

**SEC-MALS experiments were performed with guidance from and in collaboration with Dr Weston B. Struwe (University of Oxford).*

SEC-MALS is a biophysical technique that couples size exclusion chromatography (SEC) with multi-angle light scattering (MALS) to investigate protein oligomerisation. Because protein samples are analysed in solution, SEC-MALS provides a method to study protein oligomerisation in solution and in an arguably more representative native conformation (as compared with non-denaturing MS which ionises samples to the gas phase). For SEC-MALS,

an analytical size exclusion column is used to separate proteins by molecular size, using a porous stationary phase that retains smaller [i.e. lower molecular weight (MW)] species[221]. Theoretically, the MW of a sample can be determined based on retention time, though factors like geometric compactness and interactions between the protein and stationary phase can affect retention time. However, unlike with traditional SEC, which requires running a standard calibration curve for MW determination, the coupling of SEC with MALS enables MW determination by exploiting multi-angle polarised light interactions with the protein[221].

SEC-MALS experiments were performed to corroborate the oligomeric status of the JMJD7 constructs observed by non-denaturing MS (**Section 7.12**). As shown in **Figure 3.17**, wildtype JMJD7 is a dimer in solution. Consistent with the non-denaturing MS data (**Figure 3.16**), substitution of C47A alone does not alter the dimeric state of JMJD7. However, there is less evidence by SEC-MALS than for non-denaturing MS that truncation of the *N*-terminus alone ($JMJD7_{27\text{-}316}$), and in combination with C47A substitution, disrupts dimerisation, as there is no observable peak corresponding to the monomeric form. Thus, it is important to consider the ionisation effects and gas-phase environment of non-denaturing MS.

Consistent with non-denaturing MS studies, SEC-MALS analyses also revealed that peaks (A) and (B), arising from the *C*- ($JMJD7_{1\text{-}295}$) and *N*-/*C*- ($JMJD7_{27\text{-}295}$) terminally truncated constructs, correspond to the dimeric and monomeric forms (as major forms), respectively. SEC-MALS also showed that JMJD7 dimerisation was completely disrupted by the JMJD7-$C47A_{1\text{-}295}$ and JMJD7-$C47A_{27\text{-}295}$ truncated variants. Overall, the SEC-MALS analyses, together with non-denaturing MS studies, indicate contributions from the *N*- and *C*-terminal regions, in combination with C47, to JMJD7 oligomerisation. The results indicate that the crystallographically observed dimer likely reflects JMJD7 in its native state in solution.

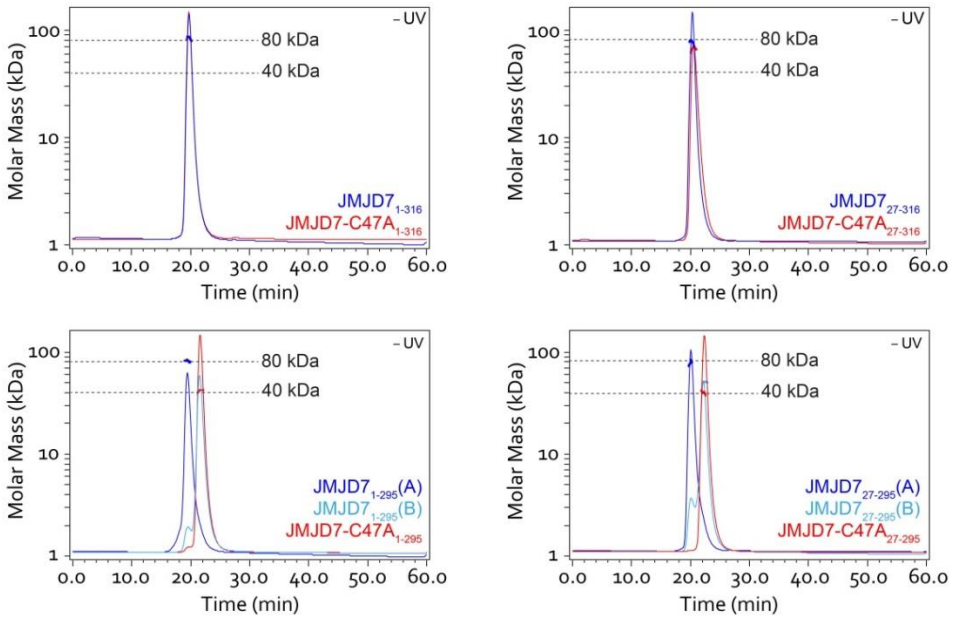


Figure 3.17. SEC-MALS UV traces of indicated JMJD7 oligomerisation constructs. JMJD7 oligomerisation constructs were analysed by SEC (~30 μg protein in 50 mM HEPES, 200 mM NaCl, 2% glycerol, pH 7.5). The solid coloured lines represent the UV traces arising from SEC analyses; the respectively coloured, bolded dashes reflect the molar masses (kDa) of each indicated construct as calculated by MALS. The dotted black lines at 80 and 40 kDa, which respectively correspond to JMJD7 dimeric and monomeric forms, are shown as reference markers for the MALS traces. As compared with ***(A)*** wildtype JMJD7 ($JMJD7_{1-316}$), no changes in retention time were observed for neither $JMJD7_{1-316}$-C47A nor the ***(B)*** *N*-terminally truncated JMJD7 constructs with or without C47A substitution ($JMJD7_{27-316}$ and $JMJD7_{27-316}$-C47A). Dimeric oligomerisation is supported by MALS traces (~80 kDa). As observed during their purifications (**Figure 3.11**), the ***(C)*** *C*-terminally truncated JMJD7 constructs [$JMJD7_{1-295}$ (A) and (B)] and the ***(D)*** *N*-/C-terminally truncated constructs [$JMJD7_{27-295}$ (A) and (B)] represent a mixture of dimeric and monomeric forms as evidenced by differences in retention elution times and MALS traces (~40 kDa). However, C47A substitution, in combination with these truncations ($JMJD7_{1-295}$ and $JMJD7_{27-295}$), results in only the monomeric form.

3.3.6 Investigation of the activity of JMJD7 oligomerisation constructs

To investigate the effects of oligomerisation on JMJD7 hydroxylase activity, MALDI-TOF-MS hydroxylation assays were performed with the $DRG1_{16-40}$ peptide substrate and JMJD7 oligomerisation constructs. As shown in **Figure 3.18**, alanine substitution of C47 had little effect on JMJD7 hydroxylase activity. The *N*-terminally truncated JMJD7 construct with and without alanine substitution of C47 showed a significantly reduced effect on JMJD7 hydroxylase activity, while all other constructs with *C*- and *N*-/*C*-terminal truncations were catalytically inactive (within the limits of MALDI-TOF-MS detection).

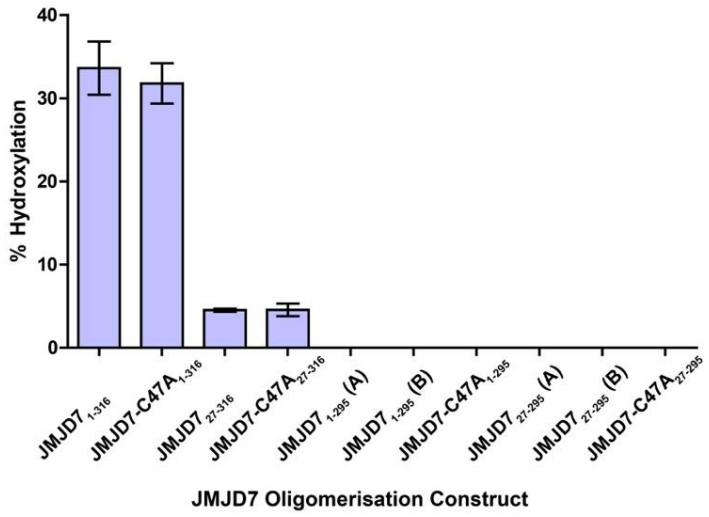


Figure 3.18. MALDI-TOF-MS hydroxylation assays with DRG1$_{16-40}$ and JMJD7 oligomerisation constructs. Assays were performed under the conditions described in **Section 7.7.1.3**, with 50 μM DRG1$_{16-40}$, 1 μM enzyme, 100 μM Fe(II), 500 μM 2OG, and 1 mM ascorbate in 50 mM HEPES, pH 7.5. Reactions were incubated at 25°C and quenched after 5 minutes. Data represent mean ± s.d. (n = 2).

While it is difficult to define a definitive relationship between JMJD7 dimerisation and activity given the combined oligomerisation data and activity assays (summarised in **Table 3.3**), it can be concluded that dimerisation is likely required. All catalytically active JMJD7 constructs are dimers according to non-denaturing MS and SEC-MALs analyses; all JMJD7 monomers are inactive. However, certain truncations that maintain JMJD7 dimerisation, e.g. JMJD7$_{27-316}$, JMJD7$_{1-295}$ (A), and JMJD7$_{27-295}$ (A), are not catalytically active. These data suggest that the *N*- and *C*-terminal regions are either (i) directly involved in substrate binding, or (ii) involved in forming an appropriate dimerisation interface that enables efficient substrate binding (as the dimerisation interface formed between the indicated dimeric JMJD7 truncated constructs has not been crystallographically resolved). Furthermore, the combined data provide evidence to show that while C47 plays a role in JMJD7 dimerisation, it is not the only contributing factor. Overall, the solution data are consistent with the crystallographically observed JMJD7 dimerisation interface (**Figure 3.9**), indicating that JMJD7 dimerisation is not a crystallographic artefact and that the observed dimerisation interface reflects JMJD7 dimerisation in solution.

Table 3.3. Summary of effect on JMJD7 N- and C-terminal truncations on oligomeric state and hydroxylase activity.

Construct Name	Non-denaturing MS	SEC-MALS	Hydroxylase Activity
JMJD7$_{1-316}$	Dimer	Dimer	Control
JMJD7-C47A$_{1-316}$	Dimer	Dimer	No observable effect
JMJD7$_{27-316}$	Dimer; trace monomer	Dimer	Significantly reduced
JMJD7-C47A$_{27-316}$	Dimer; trace monomer	Dimer	Significantly reduced
JMJD7$_{1-295}$ (A)	Dimer; trace monomer	Dimer	Inactive
JMJD7$_{1-295}$ (B)	Monomer; trace dimer	Monomer; trace dimer	Inactive
JMJD7-C47A$_{1-295}$	Monomer	Monomer	Inactive
JMJD7$_{27-295}$ (A)	Dimer; trace monomer	Dimer	Inactive
JMJD7$_{27-295}$ (B)	Monomer; trace dimer	Monomer; trace dimer	Inactive
JMJD7-C47A$_{27-295}$	Monomer	Monomer	Inactive

3.4 Discussion and future perspectives

** Parts of this section, in particular those focused on detailed comparisons on the structure-function relationships of human JmjC hydroxylases and KDMs (**Section 3.4.1**), have been published in a modified form as a co-first author review: Markolovic S. et al. (2016). Structure-function relationships of JmjC oxygenases-demethylases versus hydroxylases. Current Opinion in Structural Biology, 41:62-72*[21].

3.4.1 Overall structural comparison of JMJD7 with JmjC demethylases and hydroxylases

3.4.1.1 The importance of the structural characterisation of JMJD7

While the biochemical assignment of JMJD7 as a (3*S*)-lysyl hydroxylase described in **Chapter 2** is largely conclusive, structural studies were also undertaken to validate its assignment as a hydroxylase. Because the human JmjC oxygenases are reported to catalyse (i) N^{ϵ}-methyl lysine demethylation (via initial hydroxylation of the N^{ϵ}-methyl group) or (ii) oxidation of specific amino acid side chains in proteins to form a stable hydroxylated product, their dual functionality has led to conflicting assignments of some JmjC oxygenases. For example, the RiOXs, MINA53 and NO66, have been characterised as both ribosomal protein hydroxylases[122] and KDMs[121,180]. JMJD6 has been reported to catalyse both lysyl hydroxylation[107,111] and *N*-

methyl arginyl demethylation[105,112]. Finally, JMJD5 has been independently reported to function as either a hydroxylase[144] or KDM[181]. Due to the controversial assignments of several JmjC hydroxylases as KDMs, structural studies were important to support the biochemical assignment of JMJD7 as a hydroxylase.

3.4.1.2 Overall domain architectures

Despite the conserved catalytic JmjC domain shared among all JmjC oxygenases, one striking difference between the hydroxylases and KDMs is the variation in their domain architectures[21] (**Figure 3.19**). The KDMs are multi-domain proteins that tend to be 2-3x larger than the JmjC hydroxylases. The additional, non-catalytic domains present in KDMs include DNA- and histone-binding domains (e.g. PHD, zinc fingers, and ARID), consistent with their established roles in transcriptional control and targeted histone lysine demethylation. Several KDMs bind their histone substrates in part via pi-cation-pi/hydrophobic interactions arising from aromatic residues within a non-catalytic histone-binding domain and the methylated ammonium group of a non-substrate lysine residue. For example, KDM4A (an H3K9me3/2, H3K36me3/2, and H1.4K26me3/2 demethylase) forms stabilising interactions between an aromatic cage (F932, W967, and Y973) in its Tudor domain and the H3K4me3 site[222–224] (**Figure 3.20A**). Similarly, KDM7B (PHF8) forms pi-cation-pi/hydrophobic interactions between an aromatic cage (Y7, Y14 and W29) from within its PHD domain and H3K4me3 residue, enabling histone binding and promoting targeted demethylation of its H3K9me2/1 substrate[225] (**Figure 3.20B**). These interactions offer insights into the mechanisms by which KDMs can recognise specific methylation sites on histone amino acid sequences. However, unlike the KDMs, JMJD7 clearly lacks any discernible histone/DNA-recognition domains; rather the only recognisable domain present in JMJD7, as previously noted by Klose and colleagues using sequence-based bioinformatic analyses[12], is the JmjC catalytic domain (**Figure 3.7**).

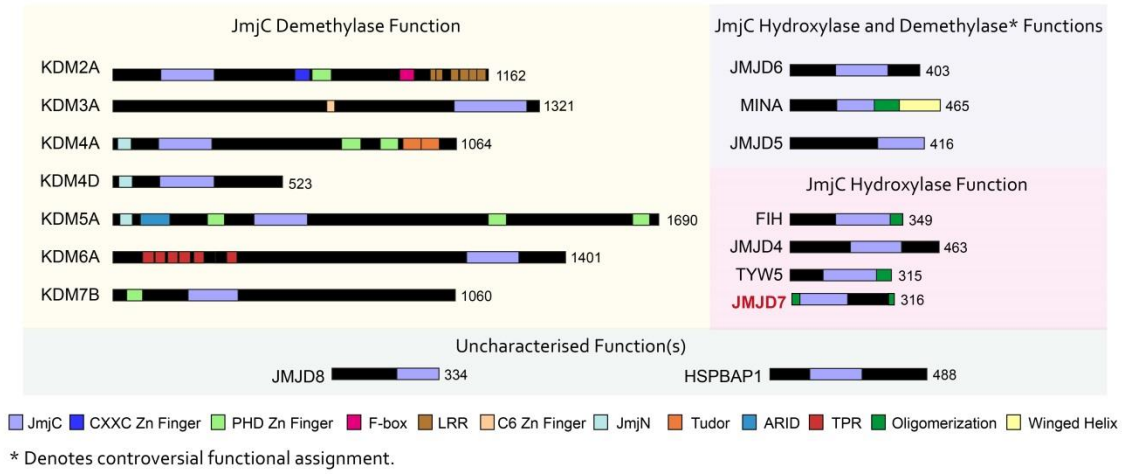


Figure 3.19. Overall domain architectures of representative JmjC oxygenases.

In contrast to the KDMs, the JmjC hydroxylases tend to be smaller proteins lacking DNA/chromatin-associated domains (e.g. PHD, Tudor, and Zn fingers) (**Figure 3.19**). However, they can have other domains, such as the winged helix (WH) and dimerisation domains, which are involved in oligomerisation and/or substrate binding. For example, the human RiOXs, MINA53 and NO66, contain a *C*-terminal WH domain that enables ribosomal substrate binding[102]. In the case of FIH, the *C*-terminal α-helical dimerisation domain promotes HIF-1α binding via hydrophobic interactions formed between FIH residues I318, I322, M325, and L326 and residue L795 of the HIF-1α substrate (**Figure 3.21**)[84]. The tRNA hydroxylase, TYW5, has been proposed to engage substrate binding via a positively charged patch formed in part by its *C*-terminal α-helices, which are also involved in protein dimerisation[103].

Consistent with sequence-based predictions[12], crystallographic analyses revealed that JMJD7 contains a JmjC catalytic domain. Like other JmjC hydroxylases, JMJD7 has a JmjC hydroxylase-characteristic oligomerisation domain, though a unique one involving both *N*- and *C*-termini and disulphide bond formation (**Section 3.3**; **Figure 3.9**); however, no additional domains, particularly KDM-characteristic DNA-binding and/or chromatin-associated domains, are present. Overall structure and domain architecture comparisons indicate that JMJD7 is

indeed more similar to the JmjC hydroxylases, in agreement with its biochemical assignment as a lysyl hydroxylase (**Chapter 2**).

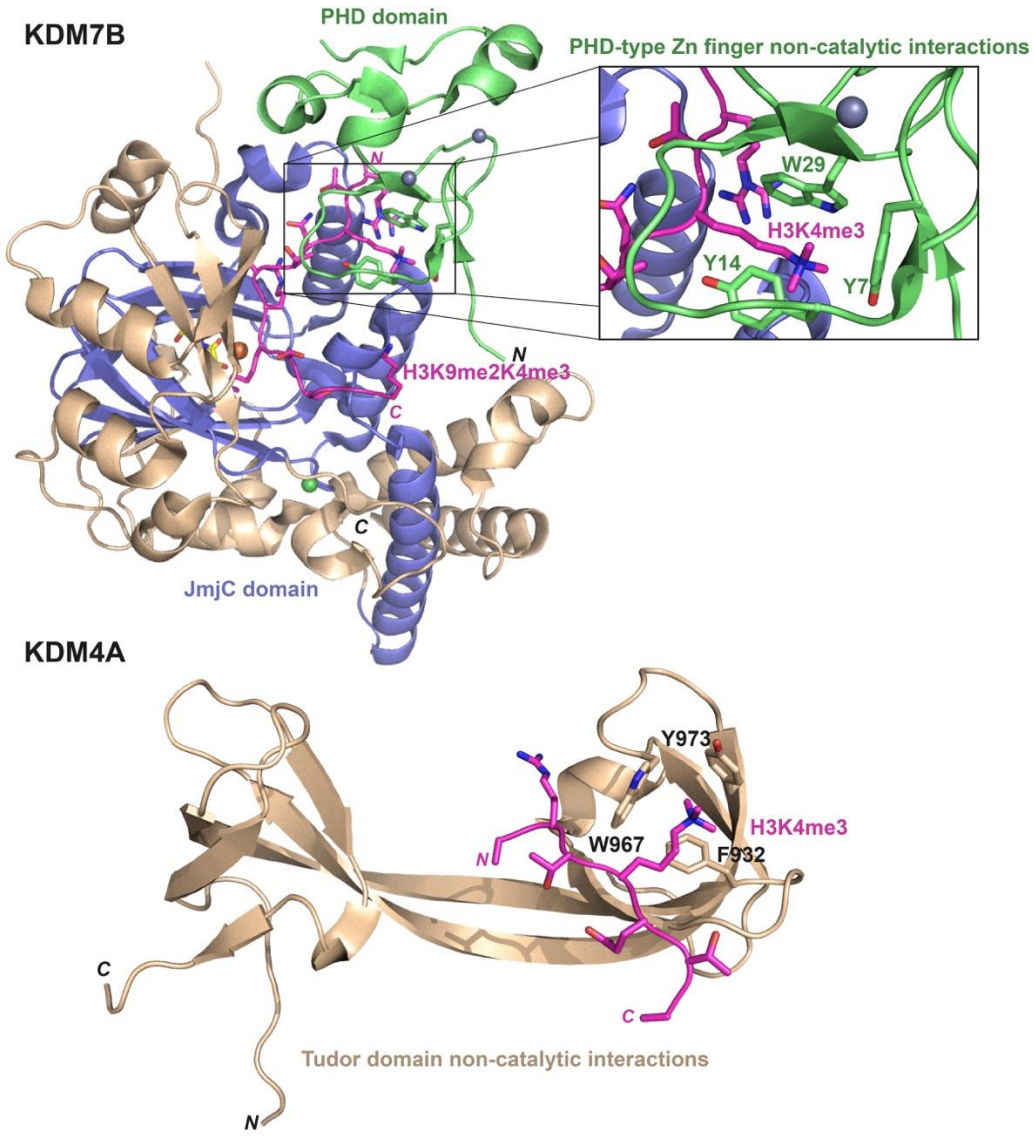


Figure 3.20. KDM7B and KDM4A non-catalytic domain histone substrate binding. *(Top)* View from a crystal structure of KDM7B in complex with an H3K9me2K4me3 peptide fragment (magenta) (PDB ID: 3KV4[225]) reveals the role of the PHD domain (green) in histone binding and targeted substrate (H3K9me2) binding within the JmjC catalytic domain (purple). Zoomed view of the PHD domain (black outline) indicates the pi-cation-pi/hydrophobic interactions between an aromatic cage (Y7, Y14 and W29) and the H3K4me3 residue. *(Bottom)* View from the structure of the KDM4A Tudor domain in complex with an H3K4me3 peptide fragment (magenta) (PDB ID: 2GFA[222]) reveals the binding interactions between the aromatic cage (F932, W967, and Y973) and the non-substrate H3K4me3 residue.

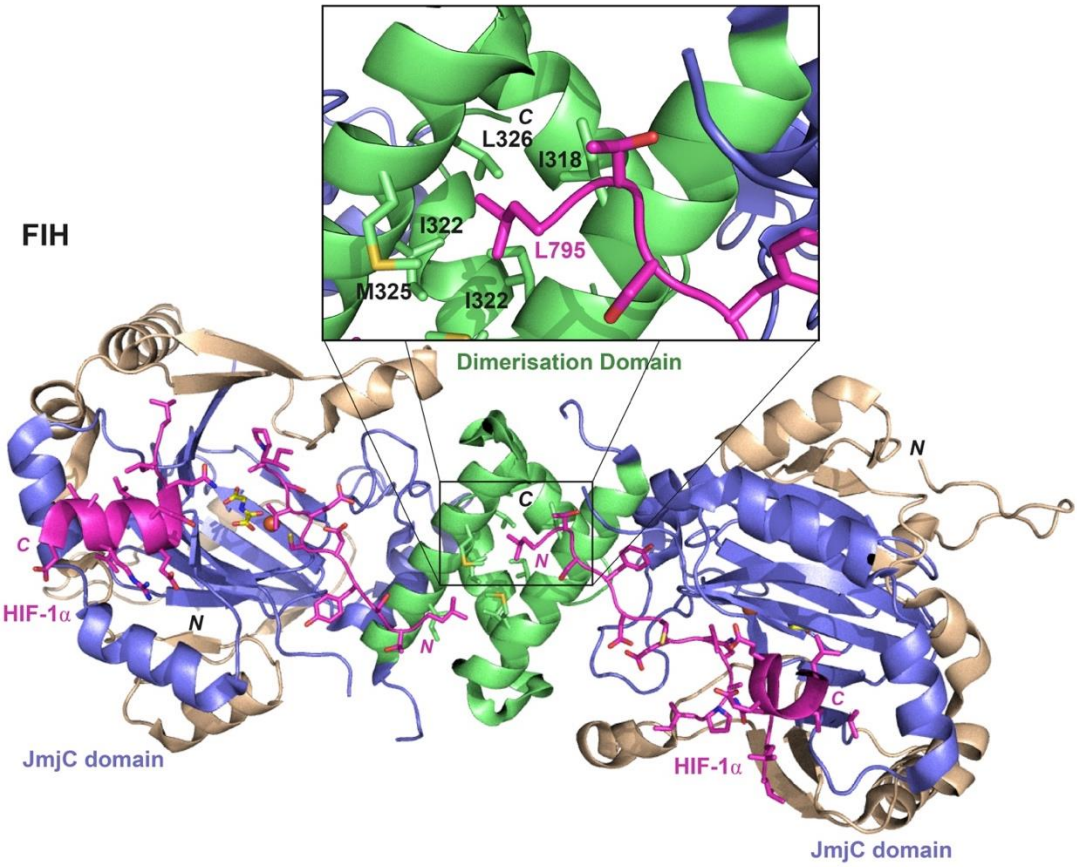


Figure 3.21. View of a crystal structure of FIH complexed with a HIF-1α peptide fragment (PDB ID: 1H2K[84]). The FIH dimerisation domain (green) is involved in binding the HIF-1α substrate (magenta) away from the JmjC active site (purple). Zoomed view of the dimerisation domain highlights the I318, I322, M325, and L326 residues (green sticks) involved in forming stabilising hydrophobic interactions with L795 of HIF-1α.

3.4.1.3 Oligomerisation

Likely in part due to their large sizes and multi-domain architectures, oligomerisation of full-length KDMs is not well characterised (i.e. domains tend to be studied in isolation, rather than in the context of the full-length protein due to their size). Only crystal structures of truncated KDMs have been solved to date, none of which reveal the presence of an oligomerisation domain. One exception is the di-/mono-methyl KDM, KDM3A, which solution studies have indicated forms a homodimer via interactions between its JmjC domain and Zn finger motif[226]. Another exception is the KDM4 subfamily, which cell-based immunoprecipitation experiments suggest is capable of homo- (KDM4A, C, and D) and hetero- (KDM4A and C)

oligomerisation[227]. However, these results have yet to be validated by detailed structural analyses.

By contrast, the JmjC hydroxylases tend to oligomerise in a catalytically important manner[21]. The dimeric interface of FIH is formed via its *C*-terminal α-helices, which interlock primarily via hydrophobic interactions (**Figure 3.21**; **Figure 3.22**)[213]. Consistent with crystallographic evidence, mutational studies, and biophysical analyses supporting FIH homodimerisation in solution, the dimerisation interface is involved in substrate binding and is required for hydroxylase activity[84,213]. The JmjC nucleic acid hydroxylase, TYW5, also forms a homodimer via its *C*-terminal α-helical bundle, but in an anti-parallel orientation that distinguishes its dimerisation from FIH (**Figure 3.22**)[103]. As for FIH, biochemical studies support TYW5 as a dimer in solution and indicate that disruption of the dimerisation domain results in a loss of activity[103]. The RiOXs similarly oligomerise via α-helical bundles (**Figure 3.22**), which are found within a structurally conserved dimerisation domain located between the JmjC and WH domains[102,214,215]. Biochemical studies have shown that mutations that disrupt the *C*-terminal dimerisation domain of the RiOXs cause a loss in activity, further indicating the importance of oligomerisation in hydroxylase catalysis [102,215].

Like other hydroxylases, JMJD7 oligomerises via α-helical bundles, though uniquely the interactions involve both *N*- and *C*-terminal α-helices, as evidenced by the crystallographic and solution studies described earlier in this chapter (**Section 3.3**). Interestingly, the contributions arising from the *C*-terminal α-helix (α9 of JMJD7) to its dimerisation interface are apparently greater than those from the *N*-terminal α-helix (α1). This is consistent with the dimerisation modes of all structurally characterised JmjC hydroxylases, which oligomerise via their *C*-terminal α-helices. As described in **Section 3.3.6**, dimerisation is likely required for JMJD7 catalytic activity, though the precise mechanism is unknown. Therefore, the determination of an enzyme:substrate complex structure would provide valuable insights into the potential role of

the dimerisation domain in substrate binding and JMJD7 catalysis. Overall, the crystallographic evidence for JMJD7 dimerisation via α-helical bundles further supports the biochemical assignment of JMJD7 as a hydroxylase.

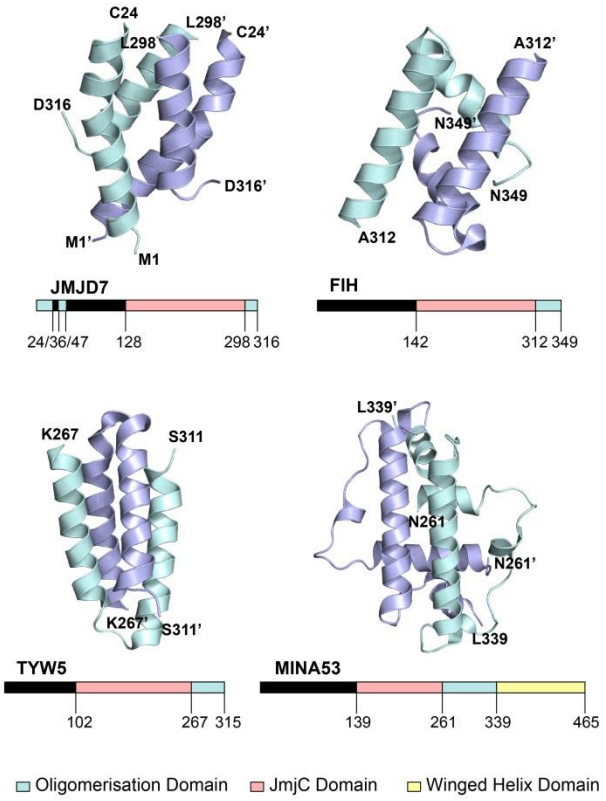


Figure 3.22. Comparison of the oligomerisation domains of representative JmjC hydroxylases. Structural analyses of the oligomerisation domains of JMJD7 with those of other JmjC hydroxylases, including FIH (PDB ID: 1H2K)[84], TWY5 (PBD ID: 3AL6)[103], and MINA53 (PDB ID: 4BXF)[102], reveal a conserved role for α-helices in oligomerisation between monomers, despite differences in the nature of these interactions. Overall domain architectures are shown below the dimerisation modes[102,103,213], highlighting the presence of the dimerisation domain (light blue) with respect to *N*- and *C*-termini, the JmjC domain (salmon), and for MINA53, the winged helix domain (yellow). JMJD7 dimerises via a unique mode involving both *N*- (α_1 and α_2) and *C*- (α_9) terminal α-helical regions.

As described in **Section 3.4.1.1**, there are controversial assignments of some JmjC hydroxylases as demethylases, including for JMJD6[106]. It is noteworthy that in the case of JMJD6, no crystal structures of full-length JMJD6 or JMJD6 bound to substrate have yet been reported. While crystal structures of truncated JMJD6 constructs have either one or two molecules per ASU[209,228], these likely do not completely reflect JMJD6 oligomerisation in solution (which include monomeric, trimeric, pentameric, and larger oligomeric states[229–231]). JMJD6 also has a positively charged surface, which has been suggested to enable RNA

binding[228]. While it is possible that JMJD6 has dual demethylase and hydroxylase functionality, its demethylase function is controversial though currently accepted[106]. Thus, the case of JMJD6 exemplifies the importance of combining biochemical assignments with structural analyses of enzymes, as has been performed for JMJD7.

3.4.1.4 The JmjC active site and substrate binding

In addition to differences in overall domain architecture, the JmjC KDMs and hydroxylases also differ in substrate binding within their active sites[21]. Compared with the hydroxylases, the KDMs tend to have a deeper and oftentimes narrower substrate binding pocket in the active site to accommodate target N^{ϵ}-methylated lysine residues. The KDM active site also tends to have a hydrophobic region that enables interactions between the hydrophobic N^{ϵ}-methylated lysine side chain and N^{ϵ}-methyl group. For example, a hydrophobic region of the KDM7B active site interacts with the side chain of the target H3K9me2 residue[225] (**Figure 3.23**). Notably, JMJD7 does not have a hydrophobic region within the DSBH core fold, suggesting that it unlikely to accommodate methylated lysine residues in its active site in a manner characteristic to the KDMs.

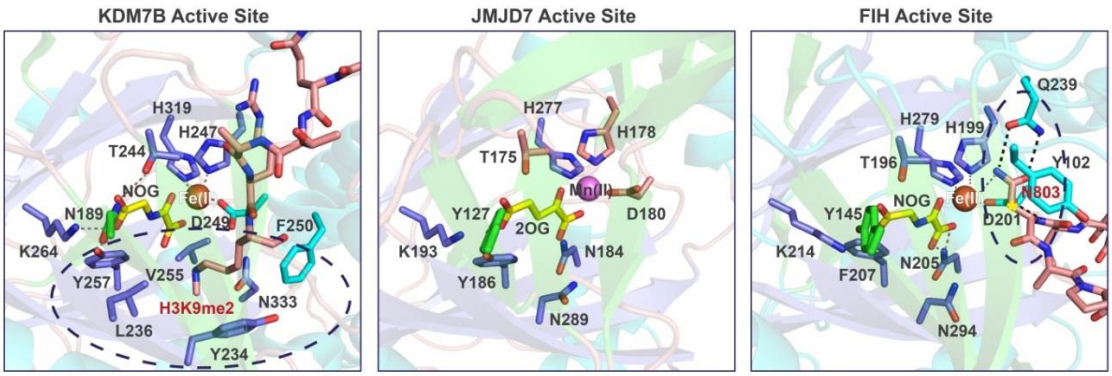


Figure 3.23. Comparison of the JMJD7 active site with representative JmjC oxygenases. KDM active site substrate binding is characterised by a hydrophobic region (dotted circle) that is involved in binding the target N^{ϵ}-methylated lysine residue, as exemplified by *(left)* KDM7B complexed with an H3K9me2 peptide fragment (salmon) (PDB ID: 3KV4[225]). By contrast, the JmjC hydroxylases tend to bind their substrates via main and side chain hydrogen bonding interactions (dotted circle), which enable stereoselective oxidation of the C-H bond (yellow star) of the target residue within the active site, as exemplified by *(right)* a structure of FIH complexed with a HIF-1α peptide substrate (salmon) (PDB ID: 1H2K[84]). *(Middle)* The active site of JMJD7 shares closer structural similarity to the JmjC hydroxylases, as exemplified by the more open active site and lack of hydrophobic residues in the active site.

In contrast to the JmjC KDMs, most structurally characterised JmjC hydroxylases generally have more open binding sites within the DSBH catalytic domain. Rather than the primarily hydrophobic interactions used by KDMs, the hydroxylases tend to bind their substrates principally through main and side chain hydrogen bonds, which is proposed to enable stereoselective hydroxylation[21]. This is exemplified by active site substrate binding in FIH, where hydrogen bonding interactions formed between Q239, D201, and the substrate HIF-1α N803 residue specifically orient the target C-H bond to Fe(II) in the active site[84] (**Figure 3.23**). Stereoselectively is likely less important for the KDMs for the following reasons: (1) oxidation of the methyl group leads to the formation of achiral formaldehyde, and (2) sequential KDM demethylation is likely promoted by the flexibility of the hydrophobic interactions involved in KDM substrate residue binding. Substrate flexibility is less desirable for the hydroxylases, which catalyse stereoselective hydroxylation. Given its more open binding active site and biochemical assignment as a stereoselective lysyl hydroxylase, it is likely that substrate binding of JMJD7 will reflect that of the JmjC hydroxylases. The determination of a JMJD7:substrate complex structure would elucidate active site binding interactions and unique features that can be used to specifically target JMJD7 over other hydroxylases.

The JmjC KDMs can also be distinguished from the hydroxylases by the presence of an extended flexible loop region, located immediately *N*-terminal to β1 of the DSBH[21]. This extended loop is typically involved in binding the target N^{ϵ}-methylated lysine residue. In the case of KDM4A, this loop contains E169, which together with Y175 on the β1 strand form electrostatic and hydrogen bonding interactions with R8 to stabilise the substrate and orient the target H3K9me3 residue deeper into the active site[232] (**Figure 3.24**). The involvement of this loop in substrate binding is also observed in the substrate-bound crystal structures of KDM2A[233], KDM6A[89], and KMD7B[225]. However, this extended loop is not observed in the structurally characterised JmjC hydroxylases, as illustrated by JMJD7 and FIH[84] (**Figure 3.24**).

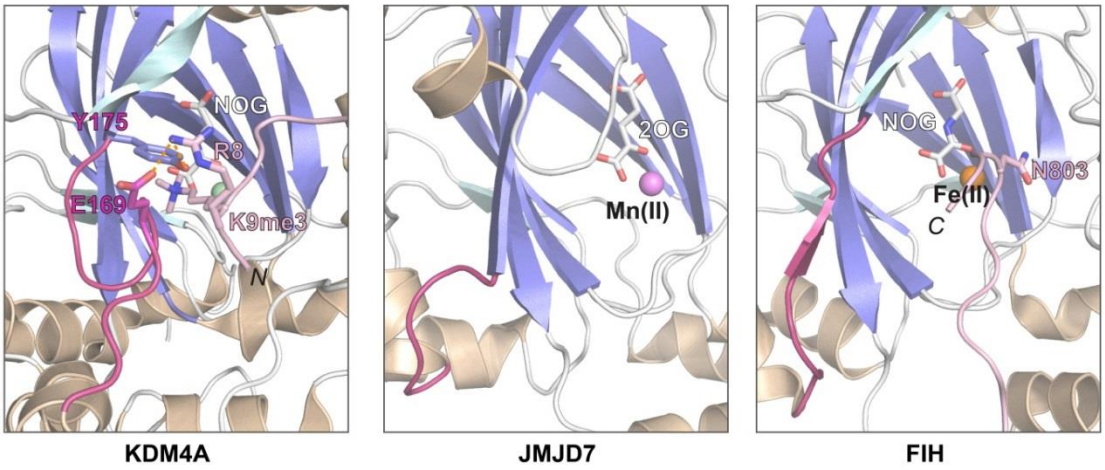


Figure 3.24. Comparison of KDM-characteristic extended loop in substrate binding. An extended loop located immediately *N*-terminal to the DSBH βI strand (dark pink) is characteristic of most JmjC KDMs. This is exemplified by *(left)* KDM4A (PDB ID: 2OQ6[232]), where the loop protrudes into the active site and forms stabilising electrostatic/hydrogen bonding interactions between residues E169 and Y175 with R8, which neighbours the target H3K9me3 residue within the H3K9me3K14Ac peptide substrate (pink). These interactions orient the target residue into the active site. By contrast, *(right)* FIH (PDB ID: 1H2K[84]) and *(middle)* JMJD7 have shorter loops *N*-terminal to the DSBH βI strand. As shown for FIH, this strand is not involved in binding the HIF-1α fragment peptide substrate (residues 795-806; pink). Metal colouring: Ni(II), green; Mn(II), violet; Fe(II), orange. DSBH β-strands are coloured in purple.

Given the described active site differences between JmjC KDMs and hydroxylases, JMJD7 is structurally more similar to the JmjC hydroxylases for the following reasons: (1) it has a more open binding site than that observed for the KDMs; (2) it lacks hydrophobic residues typically involved in KDM substrate binding; and (3) it lacks a loop *N*-terminal to DSBH β1 involved in KDM substrate binding (**Figure 3.23** and **Figure 3.24**). Because the JMJD7 structure reported in this chapter does not have a substrate bound in the active site, it is not possible to determine which residues are involved in substrate binding. However, it is likely that as for other hydroxylases, main and side chain hydrogen bonds will direct JMJD7:DRG binding and enable stereoselective (3*S*)-lysyl hydroxylation. Overall, the structural analyses of the JMJD7 active site are consistent with the biochemical assignment of JMJD7 as a JmjC hydroxylase.

3.4.2 Summary and future perspectives

The results presented in this chapter structurally define JMJD7 as a JmjC-subfamily 2OG oxygenase, supporting the biochemical and cellular characterisation of JMJD7 described in **Chapter 2**. Consistent with all structurally characterised 2OG oxygenases, JMJD7 contains a DSBH fold, which supports the conserved Fe(II)-binding residues (H178, D180, and H277) and a JmjC-subfamily-specific 2OG-binding lysine residue on the βVI strand. As described in this chapter, JMJD7 shares close structural similarity to the JmjC hydroxylases as evidenced by its overall domain architecture (i.e. lack of DNA-/chromatin-associated domains typically found in KDMs), tendency to oligomerise, and active site binding features (e.g. more open active site and lack of KDM-characteristic extended substrate-binding loop *N*-terminal to βI strand)[21]. Though JMJD7 is structurally more similar to the JmjC hydroxylases, its dimerisation interface is distinct. While most JmjC hydroxylases tend to oligomerise via their *C*-terminal α-helices[102,103,213], JMJD7 dimerisation forms via interactions involving *N*- and *C*-terminal α-helices and disulphide bond formation between residues C47 from each monomer, as described by the detailed biophysical and biochemical studies of truncated JMJD7 constructs presented in this chapter.

While attempts have been made to solve a JMJD7:DRG complex structure using full-length JMJD7 and DRG peptides ($DRG1_{16-40}$, $DRG2_{15-39}$, DRG_{KNKAT}), they have not been successful. To date, there are no reported enzyme:substrate complex structures of any lysyl hydroxylases. Future structural studies on JMJD7 should direct efforts towards solving a JMJD7:DRG complex structure to investigate JMJD7:DRG substrate binding at a molecular level. This understanding would likely offer insights into the roles of clinically-relevant JMJD7 variants and identify variants for future investigation. Furthermore, a detailed characterisation of the JMJD7 and DRG1-2 residues involved in enzyme:substrate binding would enable a better understanding of regio- and stereo-selective lysyl hydroxylation. Future JMJD7:substrate structural analyses would guide the development of selective JMJD7 therapeutics, which is of

interest given that the links between JMJD7 and diseases, including cancer (**Section 2.1.3**) and autism[139,140], remain to be understood. It is likely that distinct structural features elucidated by a JMJD7:substrate complex structure would highlight unique features (e.g. in the active site or dimerisation domain) that distinguish JMJD7 from other hydroxylases and enable *selective* inhibition and/or activation, contingent on its role(s) in cancer and/or autism.

4 INVESTIGATION OF CLINICALLY-RELEVANT JMJD7 SUBSTITUTIONS

4.1 Introduction

4.1.1 JMJD7 variants associated with disease

The original motivation to study JMJD7 was in large part due to its association with diseases, including cancer and autism[139,140]. As described in **Section 2.1.3**, pan-cancer analysis using the cBioPortal for Cancer Genomics[147,148] indicates that the *JMJD7* gene is not only deleted in a variety of tumour types, but also that low levels of *JMJD7* expression found in patients with tumours from gastric, lung, and colorectal cancer are associated with a poorer prognosis (**Figure 2.4**). As a potential tumour suppressor gene, it is likely that the catalytic activity of JMJD7 is required to prevent cancer development and/or progression. Therefore, gene mutations that result in a loss or reduction in JMJD7 function are likely to contribute to cancer progression. The Catalogue of Somatic Mutations in Cancer (COSMIC)[235] is a publically available database that compiles somatic mutations identified in human cancer from scientific literature and experimental screens from the Cancer Genome Project at the Sanger Institute (Cambridgeshire, UK). cBioPortal[147,148] similarly provides analysis of large scale cancer genomic datasets. A search for JMJD7 in COSMIC and cBioPortal identified several *JMJD7* mutations in various cancer genetic studies, summarised in **Figure 4.1.**

In addition to associations with cancer, certain *JMJD7* gene mutations have reported associations to intellectual disabilities, such as cognitive impairment and autism[139,140]. A mutation in the *JMJD7-PLA2G4B* gene resulting in an M160V amino acid substitution was identified as a *de novo* point mutation in patients with unexplained, severe intellectual disabilities (i.e. cognitive impairment or mental retardation); mutations were identified via family-based exome sequencing of 100 patients with unexplained intellectual disability and

their unaffected parents[139]. A separate study identified a *JMJD7* gene mutation resulting in R260C amino substitution as a rare causal autism variant found in several members (nine) of an extended, high-risk autism spectrum disorder family[140].

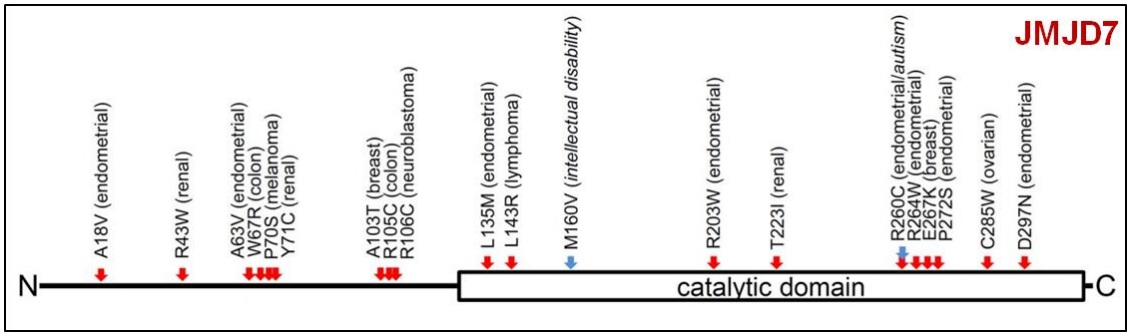


Figure 4.1. JMJD7 disease hot spots. Figure generated using publically available data from the COSMIC database[235] by Charlotte Eaton (under the supervision of Dr Mathew Coleman, University of Birmingham) and kindly provided by Dr Mathew Coleman.

Notably, mutations in genes of other JmjC oxygenases also have reported links to intellectual disabilities. For example, truncating mutations in *KDM7B* (*PHF8*) have been identified in males affected with X-linked mental retardation and (in some cases) associated cleft lip/cleft palate from two unrelated families (3-4 affected males per family)[236]. A missense *KDM7B* mutation resulting in F279S substitution has also been reported in a genetic study in multiple males similarly affected with mild mental retardation and cleft lip/palate from a single Finnish family[237]. Biochemical analyses of the F279S variant revealed that this substitution caused loss of *KDM7B* histone lysine demethylase function, suggesting a potential link between its catalytic activity and role in mental retardation[238]. There are also reported links between several mutations in *KDM5C* (*JARID1C*), X-linked intellectual disability, and autism spectrum disorders[234,239,240].

Because both cancer and intellectual disabilities such as mental retardation are both polygenic diseases, it is difficult to elucidate the precise role that missense mutations play in the disease. Nevertheless, the investigation of missense mutations resulting in amino acid substitutions in proteins may offer a better understanding of the functional role of a protein within the context

of the disease. While the precise functional role of JMJD7 in normal physiology remains to be understood, the investigation of cancer- and intellectual disability-associated amino acid substitutions on JMJD7 structure, stability, and activity may provide insights into the role of JMJD7 in disease.

4.1.2 Aims of this chapter

The study of clinically-relevant JMJD7 variants was of interest given the potential tumour suppressor activity of JMJD7 and the identification of several *JMJD7* mutations found in a variety of cancer types. The biochemical and structural characterisation of JMJD7, described in **Chapters 2 and 3**, provided a platform for the investigation of clinically-relevant JMJD7 variants. Thus, biochemical and biophysical studies were undertaken to identify variants that affect JMJD7 function, stability, and/or structure to provide insights into the potential role(s) that JMJD7 play(s) in cancer and/or intellectual disabilities.

The work described in this chapter aims to provide biochemical and structural characterisations of clinically-relevant JMJD7 variants by investigating the effect of these substitutions on:

(1) hydroxylase activity;

(2) secondary structure;

(3) structural stability; and

(4) oligomerisation status.

The results presented in this chapter aim to provide insights into how these genetic mutations/ amino acid substitutions might impact the activity of recombinant JMJD7 and consequently provide direction for future cellular studies of JMJD7 in a disease context. Additionally, the work described aims to provide a foundation for the future development of JMJD7 agonists and/or inhibitors in light of the effects of these mutations on JMJD7 activity, structure, and stability.

4.2 Production of JMJD7 variants

4.2.1 Rationale for JMJD7 variants chosen for investigation

Reported genetic mutations in *JMJD7* resulting in amino acid substitutions from a variety of cancer types and intellectual disabilities are shown in **Figure 4.1**. The following JMJD7 clinically-relevant variants were chosen for investigation for the reasons described below:

- R260C – The R260C variant was not only identified in endometrial cancer, but also in high-risk autism families[140]. Furthermore, the mutation lies within a cancer mutation 'hotspot', or region of DNA where mutations are likely to occur. High mutation prevalence in a cancer mutation hotspot strongly implies functional consequence.

- P272S – The P272S variant was also identified in endometrial cancer and observed to be in the same cancer mutation 'hotspot' as the R260C variant. Furthermore, the wildtype proline residue is conserved among all JmjC hydroxylases, suggesting an important structural and/or functional role (**Figure 4.2**).

- M160V – Though not present in a mutation 'hotspot,' the M160V variant, like R260C, was identified in persons with extreme intellectual disability[139].

- T37E – T37 has been widely identified as a target for phosphorylation in proteomic screens (Dr Mathew Coleman, University of Birmingham, personal communication). Tumour suppressor genes are often inactivated by multiple mechanisms, both genetic (e.g. via mutations) and biochemical (e.g. via PTMs). Phosphorylation, which is catalysed by protein kinases, is one of the most well-studied PTMs, with roles in cell growth, differentiation, and apoptosis[241]. Phosphorylation is essential for altering and mediating protein structure, function, and regulation in signalling pathways and the cell cycle; thus, disruption of phosphorylation, resulting from kinase mutations and/or overexpression, has been identified as a major contributor to cancer[241]. Thus, it was of

interest to investigate the effect of the phosphomimetic T37E variant. A phosphomimetic is an amino acid substitution that mimics a phosphorylated amino acid, which can cause protein activation or inactivation. Glutamic acid, which contains a negatively charged carboxylate side chain under physiological conditions, is chemically similar to phosphorylated threonine (**Figure 4.3**).

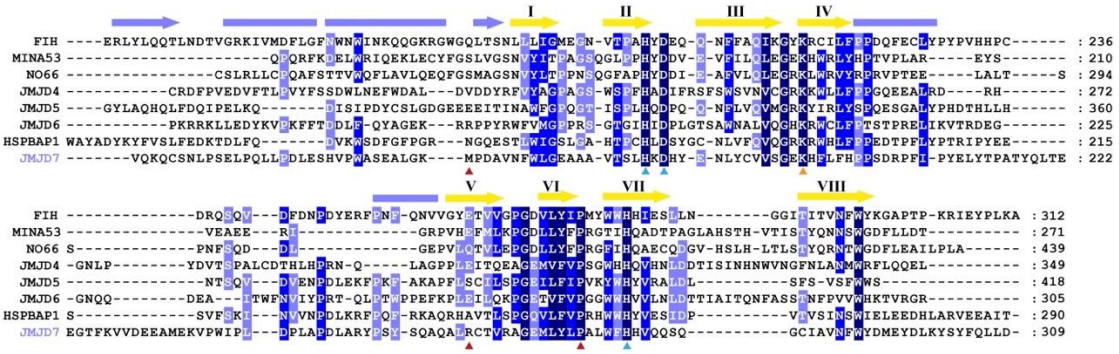


Figure 4.2. A sequence alignment of the JmjC domains of JmjC hydroxylases shows that P272 is conserved among all JmjC hydroxylases. Sequences of JmjC domains from the JmjC-only subfamily of human 2OG oxygenases (UniProtKB Entries: FIH, Q9NWT6; MINA53, Q8IUF8; NO66, Q9H6W3; JMJD4, Q9H9V9; JMJD5, Q8N371; JMJD6, Q6NYC1; HSPBAP1, Q96EW2; JMJD7, P0C870) were aligned using Clustal Omega[158] and shaded using GeneDoc. Secondary structure assignment is based on crystallographic analyses of human FIH (PDB ID: 1H2K[84]) and indicated as α-helices (rectangles) and β-strands (arrows). The eight β-strands (I-VIII) comprising the conserved DSBH fold are in yellow (arrows). Conserved Fe(II)-binding (blue) and 2OG-binding (orange) residues are indicated by triangles. The clinically-relevant variants chosen for study are indicated by red triangles.

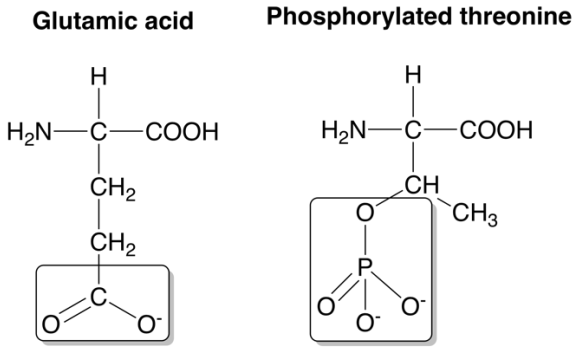


Figure 4.3. Comparison of glutamic acid and phosphorylated threonine structures. T37E is a phosphomimetic substitution.

4.2.2 Mutagenesis, expression, and purification of JMJD7 variants

Plasmids for the bacterial expression/production of the JMJD7 variants were generated via site-directed mutagenesis (**Section 7.2.5**) using the custom primers outlined in **Table 7.1.** All variants were produced with an *N*-terminal His_6-tag as described for JMJD7 (**Section 2.2.4**). Following purification, the variants were assessed for purity by SDS-PAGE analyses (**Figure 4.4**). Amino acid substitutions were validated by intact protein MS (**Figure 4.4** and **Table 4.1**). Notably, the UV traces corresponding to the Ni(II)-affinity and SEC purifications for all JMJD7 variants were nearly identical to wildtype JMJD7 (JMJD7), except for JMJD7-P272S, which eluted in two peaks, labelled in **Figure 4.5** as P272S Fractions (A) and (B). As shown in **Figure 4.5**, JMJD7-P272S fractions (A) and (B) likely correspond to degraded and aggregated protein, respectively.

Table 4.1. Summary the theoretical and observed molecular weights of JMJD7 clinical variants by intact protein MS (Figure 4.4).

Variant	Theoretical MW*	Obs. Mass (Major Peak)	Obs. Mass (Minor Peak)	Obs. Major ΔMW†	Theoretical Variant ΔMW‡	Obs. Variant ΔMW#	Obs. Major/Minor ΔMW◊
WT	38095	37962	38140	133	0	0	178
R260	38042	37909	38086	133	53	53	177
M160V	38063	37930	38107	133	32	32	177
T37E	38123	37990	38167	133	28	28	177
P272S (A)	38085	37951	38128	134	10	11	177
P272S (B)	38085	37950	38129	135	10	12	179

* Theoretical molecular weights (MW) were calculated using the ExPASy ProtParam tool (http://web.expasy.org/protparam/).
† Obs. Major ΔMW = |Theoretical MW – Observed Mass of Major Peak|
‡ Theoretical Variant ΔMW = Theoretical Mass of | WT– Variant|
# Obs. Variant ΔMW = Observed Mass of | WT Major Peak – Variant Major Peak|
◊ Obs. Major/Minor ΔMW = Observed Mass of |Major Peak – Minor Peak|

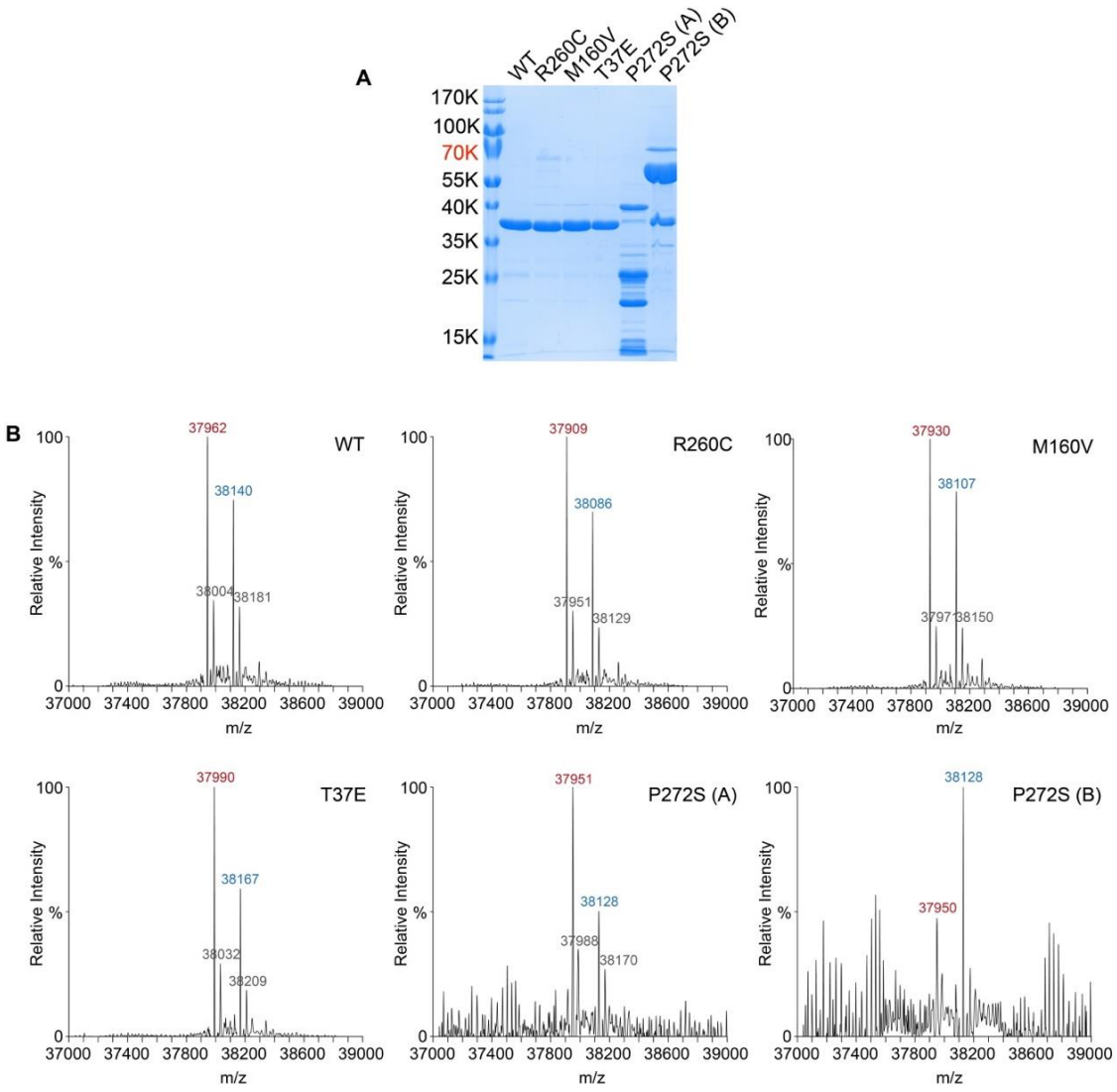


Figure 4.4. (A) Coomassie-stained 12% SDS-PAGE gel of JMJD7 variants post-purification and corresponding (B) deconvoluted intact protein mass spectra. Red and blue peaks correspond to "major" and "minor" peaks as described in **Table 4.1**. Note that the observed difference between major (red) and minor (blue) peaks is 178 Da, the latter peak of which corresponds to α-*N*-6-gluconoylation (via α-*N*-6-phosphogluconoylation), a common PTM reported to occur in proteins expressed in *E. coli* with an *N*-terminal (G-S-S-$[H]_6$-) His_6-tag[216] (see **Figure 3.14**). The mass of each major (red) peak differs from the respective theoretical mass by 133 Da, corresponding to the commonly reported loss of the *N*-terminal initiator methionine by *E. coli* aminopeptidases[242,243]. Note that P272S eluted as two distinct peaks, (A) and (B), during SEC purification (**Figure 4.5**).

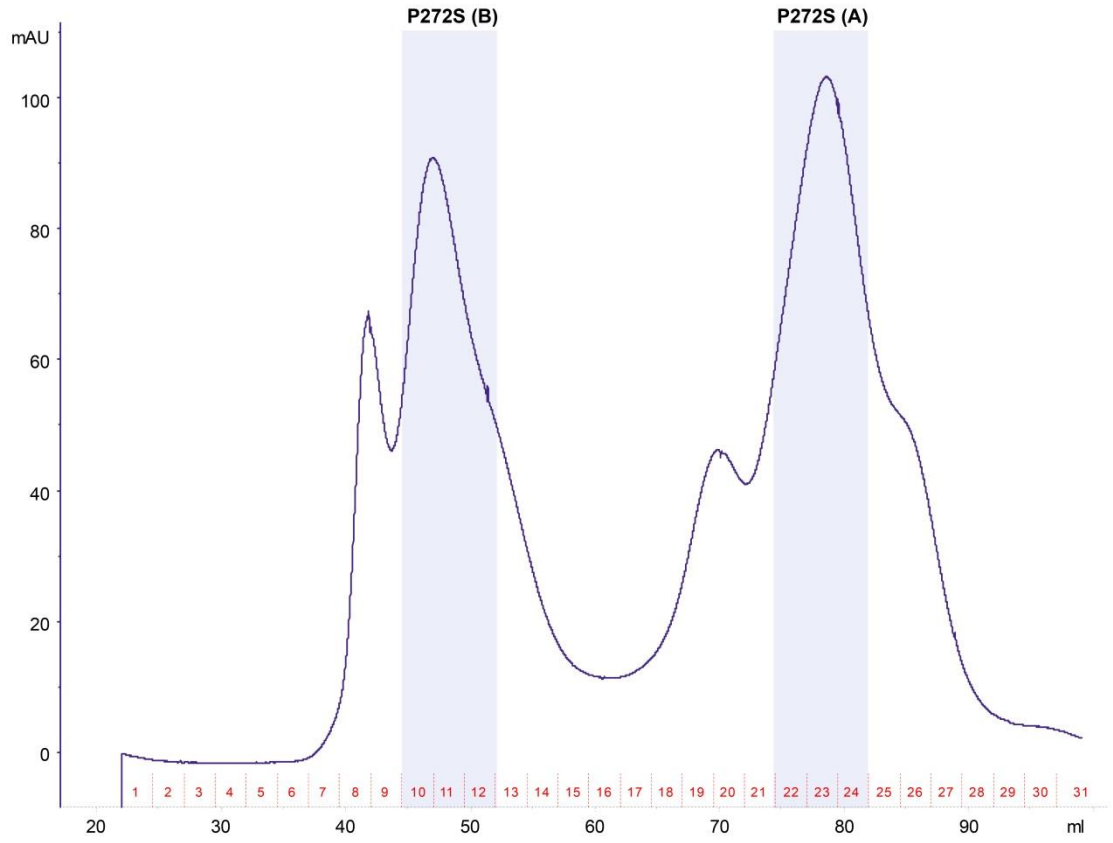


Figure 4.5. UV trace corresponding to JMJD7-P272S purification using a 16/600 Superdex 200 120 mL SEC column. Indicated JMJD7-P272S fractions were separately combined and concentrated, representing P272S (A) and P272 (B).

4.3 Comparison of hydroxylation activity between JMJD7 clinically-relevant variants

Because JMJD7 is a potential tumour suppressor gene, normal catalytic function may be required to prevent cancer progression. Similarly, investigating the effects of R260C and M160V on JMJD7 catalytic activity could provide insights into the role of these substitutions and JMJD7 in autism. For example, if these variants affect JMJD7 catalytic activity, it might support the idea that JMJD7 activity is important for normal brain function. The effect of these amino acid substitutions on JMJD7 activity was investigated by performing MALDI-TOF-MS hydroxylation assays with $DRG1_{16\text{-}40}$ and $DRG2_{15\text{-}39}$ peptides. Short (5 min; catalytic, linear range) and long (40 min) incubations with each peptide revealed that the R260C, M160V, and T37E variants have minor effects on JMJD7 hydroxylase activity; however, neither of the P272S fractions displayed any catalytic activity (**Figure 4.6**). Because enzyme kinetic

parameters were unable to be determined using the available methods (due to the limitations described in **Section 2.3.7.1**), SDS-PAGE analysis was performed for each enzyme input (10 μM enzyme) and one representative reaction (1 μM enzyme) to ensure that differences in enzyme concentration were not responsible for observed differences in activity.

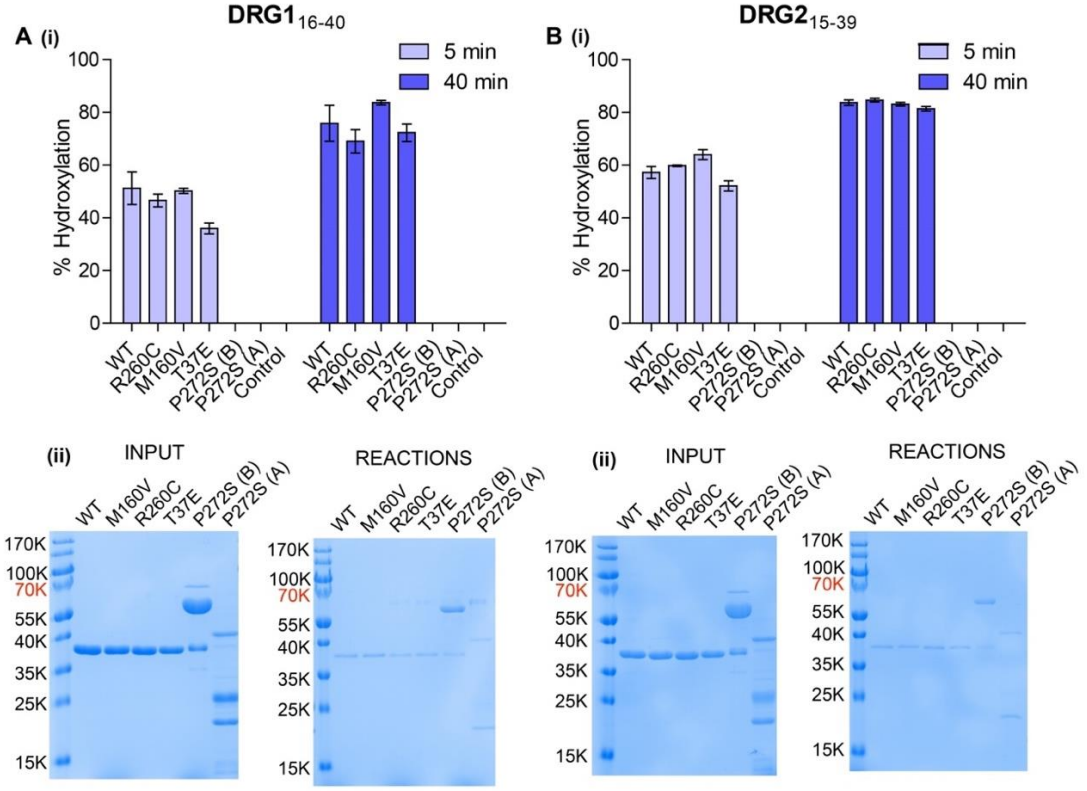


Figure 4.6. Graphical representations of MALDI-TOF-MS hydroxylation assays and respective Coomassie-stained 12% SDS-PAGE gels with clinically-relevant JMJD7 variants and (A) DRG1$_{16-40}$ or (B) DRG2$_{15-39}$. Activity assays following incubation at 25°C for 5 min (light blue bars) and 40 min (dark blue bars) reveal that the R260C, M160V, and T37E variants have no effect on activity as compared with wildtype JMJD7 (WT), while the P272S variant is inactive. Data represent mean ± s.d. (n = 3). The respective SDS-PAGE gels illustrate that relatively equal amounts of protein (except for P272S) are present in the 10 μM enzyme stocks (input) and 1 μM enzyme assay (reactions). Assay conditions: 5 or 40 minute incubation at 25° of 50 μM DRG1$_{16-40}$ or DRG1$_{15-39}$, 1 μM enzyme, 100 μM Fe(II), 500 μM 2OG, 1 mM ascorbate in 50 mM HEPES, pH 7.5.

Competition assays with DRG1$_{16-40}$ and DRG2$_{15-39}$ were also performed with the catalytically-active JMJD7 variants to determine whether the substitutions affect substrate selectivity. However, the R260C, M160V, and T37E variants exhibited the same substrate specificity as wildtype JMJD7. As reported in **Section 2.3.7.2**, DRG1$_{16-40}$ hydroxylation is unaffected by the

presence of $DRG2_{15-39}$, while $DRG2_{15-39}$ hydroxylation is apparently completely ablated (i.e. completely inhibited) by $DRG1_{16-40}$ (**Figure 4.7**).

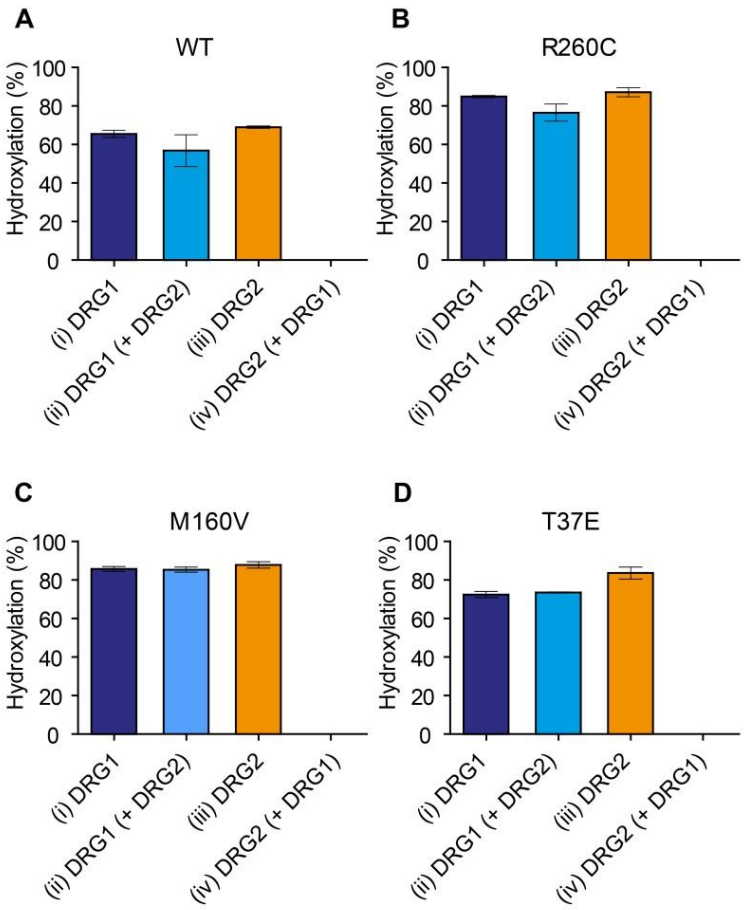


Figure 4.7. Competition assays performed with indicated JMJD7 variants and equimolar amounts of the $DRG1_{16-40}$ and $DRG2_{15-39}$ peptides following incubation at 25°C for 40 min. The same trend is observed for ***(A)*** wildtype JMJD7 and the tested variants, ***(B)*** R260C, ***(C)*** M160V, and ***(D)*** T37E: (i) the extent of $DRG1_{16-40}$ hydroxylation is not affected by (ii) the presence of $DRG2_{15-39}$; however, (iii) $DRG2_{15-39}$ hydroxylation is completely ablated in (iv) the presence of $DRG1_{16-40}$. Data represent mean ± s.d. (n = 3).

4.4 Biophysical analyses of clinically-relevant JMJD7 variants

4.4.1 Limitations of studying the P272S variant biophysically

As shown in **Figure 4.4** and **Figure 4.5**, the P272S variant was not successfully purified from *E. coli*. Consequently, the P272S variant could not be appropriately characterised and biophysical analyses were not undertaken.

4.4.2 Investigation of the effects of clinically-relevant JMJD7 substitutions on secondary structure

CD analyses (described in **Section 3.3.4**) were undertaken to study the effect of the R260C, M160V, and T37E substitutions on JMJD7 secondary structure. A comparison of the CD spectra of JMJD7 and the clinically-relevant variants at 0.2 mg/mL revealed no major differences (**Figure 4.8**), suggesting that these substitutions have a negligible effect on JMJD7 secondary structure. These results are consistent with the minor differences observed between the hydroxylase activities of these variants compared with that of wildtype JMJD7.

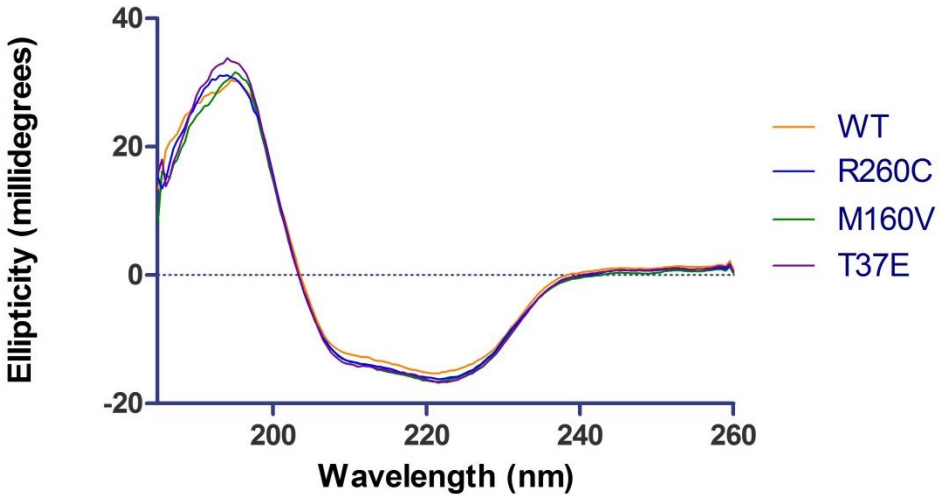


Figure 4.8. Raw CD spectra of wildtype JMJD7 (WT, orange) and clinically-relevant JMJD7 variants (R260C, blue; M160V, green; and T37E, purple). Spectra were averaged (n=3) and smoothed (smooth = 8).

4.4.3 Investigation of clinically-relevant JMJD7 variant structural stability

CD analyses were also undertaken to investigate the effect of amino acid substitutions on JMJD7 thermal stability. However, because JMJD7 did not re-fold after thermal denaturation, reliable thermodynamic properties (i.e. melting temperature) could not be determined by this method[218,219]. Therefore, differential scanning fluorimetry (DSF) with SYPRO Orange dye was used to study differences in thermal stability between JMJD7 variants as described in **Section 7.11**. DSF is a biophysical technique that indirectly measures apparent melting temperature

(T_m) by the use of a fluorescent probe to detect changes in protein hydrophobicity[244]. SYPRO Orange is a fluorescent stain ($\lambda_{Excitation}/\lambda_{Emission}$ = 492/610 nm) that binds the hydrophobic regions of proteins and is quenched by water in aqueous buffer solutions. Protein denaturation occurs by stepwise heating of the sample; as the protein is heated, unfolding exposes hydrophobic patches, and thus the dye is able to bind, causing an increase in fluorescence signal and quenching by water[244]. Apparent T_m values for clinically-relevant JMJD7 variants were subsequently determined by DSF with SYPRO Orange. As shown in **Figure 4.9**, clinically-relevant substitutions have very slightly destabilising effects on JMJD7 stability.

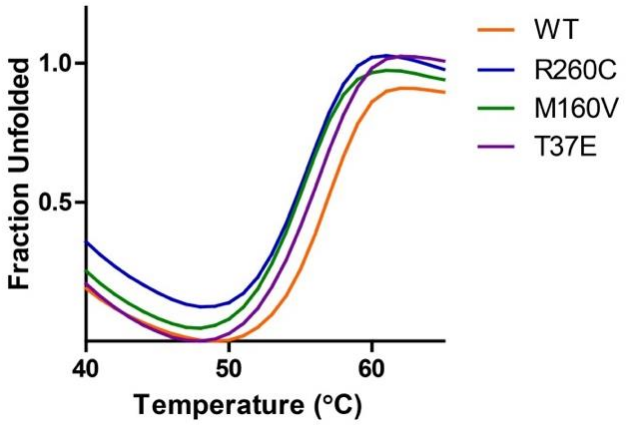


Figure 4.9. DSF analyses for wildtype JMJD7 and clinically-relevant variants (R260C, M160V, and T37E). Apparent T_m values for clinically-relevant JMJD7 variants (6 μM in 50 mM HEPES, pH 7.5) were determined by DSF using SYPRO Orange dye as follows: *WT*, 56.71 ± 0.13; *R260C*, 55.25 ± 0.21; *M160V*, 54.98 ± 0.21; and *T37E*, 55.87 ± 0.17. Data represent mean ± s.d. (n = 4).

4.4.4 Investigation of JMJD7 variant oligomerisation status

The oligomeric state of a protein can significantly affect its catalytic activity, stability, allosteric regulation, and subcellular localisation[245]. One striking similarity between the structurally characterised JmjC hydroxylases is their tendency to oligomerise. For example, dimerisation is required for FIH substrate binding and hydroxylase activity[84,213]. Similarly, biochemical studies have shown that disrupting the dimerisation of TYW5 and the RiOXs results in a loss of hydroxylase activity[102,103,215]. To determine whether clinically-relevant substitutions disrupt oligomerisation, native ESI-MS and SEC-MALS analyses were undertaken.

4.4.4.1 Non-denaturing MS

**Non-denaturing MS experiments were performed with guidance from and in collaboration with Dr Weston B. Struwe (University of Oxford).*

Table 4.1 lists the theoretical monomeric molecular weights (~38 kDa) for the JMJD7 constructs used in this study. As described in **Section 3.3.5.1**, non-denaturing ESI-MS analysis of JMJD7-WT in the positive ion mode indicates that the enzyme is a dimer (~76 kDa) in its near-native state, with no evidence for higher oligomeric states observed (**Figure 4.10**). The clinically-relevant variants are also dimers (~76 kDa), including the R260C variant, which had the greatest likelihood of forming other oligomeric states via disulphide bond formation.

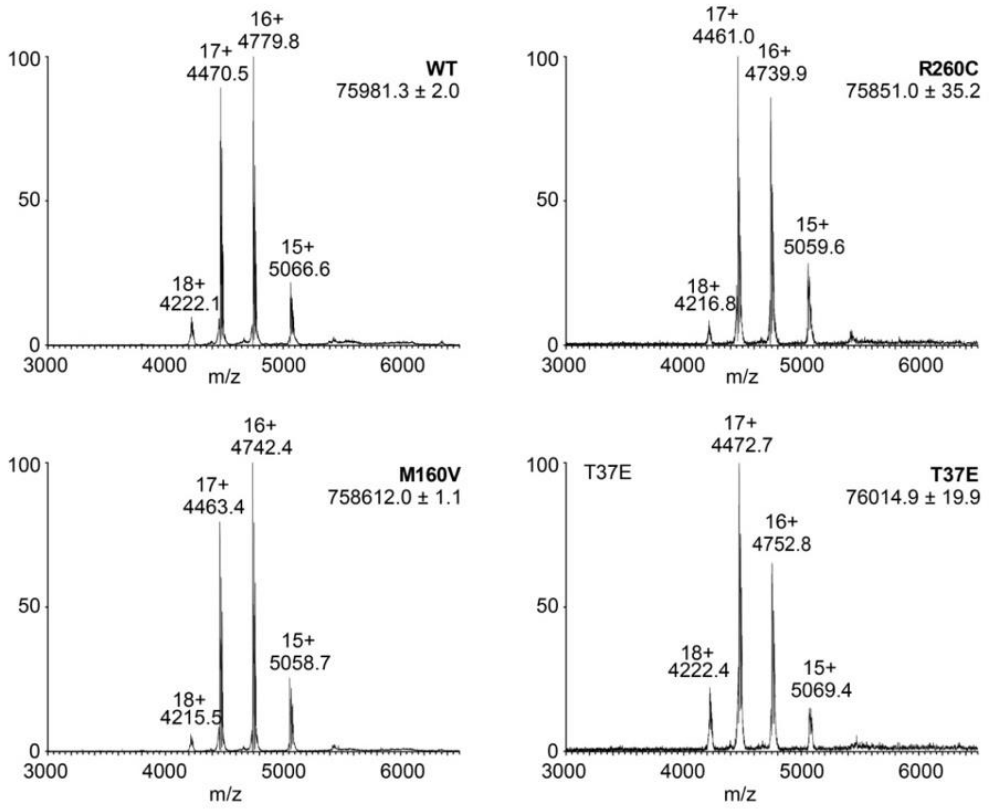


Figure 4.10. Non-denaturing ESI-mass spectra of wildtype JMJD7 and clinically-relevant variants in the positive ion mode indicate that all proteins are dimeric in their near-native state. Protein molecular weights (~76 kDa) were calculated by determining the correct charge state assignment for each ion in the electrospray spectrum using Mass and Charge State Evaluation and Determination v 0.3 (http://benesch.chem.ox.ac.uk/resources.html). Given that the theoretical monomeric molecular weights of wildtype JMJD7 and the clinically relevant variants is ~38 kDa, all JMJD7 constructs were identified as dimers, with no evidence for higher or lower oligomers.

4.4.4.2 SEC-MALs

**SEC-MALs experiments were performed with guidance from and in collaboration with Dr Weston B. Struwe (University of Oxford).*

As described in **Section 3.3.5.2**, SEC-MALS analysis revealed that wildtype JMJD7 was also a dimer in its near-native state. Consistent with non-denaturing MS analyses, the clinically-relevant variants were also determined to be dimers in solution by SEC-MALS (**Figure 4.11**), further supporting that the variants have no effect on JMJD7 oligomerisation.

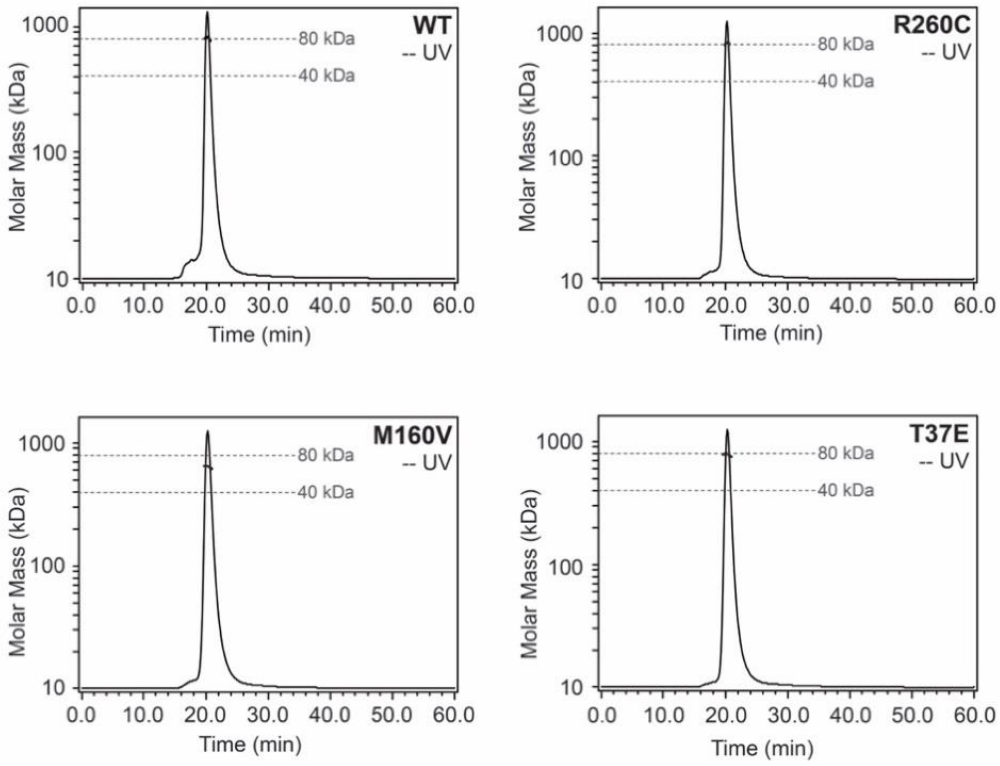


Figure 4.11. SEC-MALS traces of JMJD7 clinical variants reveal all proteins are dimers in solution. JMJD7 clinically-relevant variants were analysed by SEC (~30 μg protein in 50 mM HEPES, 200 mM NaCl, 2% glycerol, pH 7.5) and found to elute at the same retention time as wildtype JMJD7, suggesting that all constructs share the same the same oligomeric state. Molecular weight determination by MALS reveals that all JMJD7 constructs are dimers (~75-80 kDa), given that the theoretical monomeric molecular weight of wildtype JMJD7 (and the clinically-relevant variants) is ~38 kDa.

4.5 Structural analyses of the JMJD7 clinically-relevant variants

Given the structural characterisation of JMJD7 described in **Chapter 3**, the clinically-relevant JMJD7 variants were also analysed from a structural perspective to gain potential insights into the structure-function effects of these amino acid substitutions. For example, a surface-exposed residue indicates accessibility to PTMs (e.g. phosphorylation), a role in potential (substrate or allosteric) binding interactions, and/or a potential role in protein oligomerisation. Alternatively, a substitution of a residue in the DSBH catalytic active site may disrupt active site folding and consequently have deleterious effects on JMJD7 structure and activity. A similar effect could possibly be observed because of active site substitutions to residues involved in substrate and/or 2OG binding.

Figure 4.12 illustrates the positions of T37, M160, R260C, and P272 within the JMJD7 fold. All the residues are surface-exposed, except for P272, which is located immediately *C*-terminal to βVI and buried within the DSBH fold. T37 is situated immediately *N*-terminal to α2, which comprises part of the JMJD7 dimerisation domain. M160 is located on a flexible loop *N*-terminal to βI, and R260 is situated on βV pointing away from the active site.

The highly conserved, active site-buried P272 residue is apparently important in maintaining the structural integrity of the DSBH core fold by forming a tight turn between βVI and βVII in all structurally characterised hydroxylases[15,102,103,156,209], including JMJD7 (**Figure 4.12**). The structural importance of P272 supports the apparent instability of the P272S variant during protein expression and purification. Taken together, the crystallographic and biochemical analyses indicate that P272S substitution affects protein production, folding, and stability. Thus, it is likely that P272S cannot be produced in cells (as observed by Dr Mathew Coleman, personal communication) or is degraded as a result of misfolding, suggesting an aberrant functional role (i.e. loss-of-function) for this clinically-relevant variant.

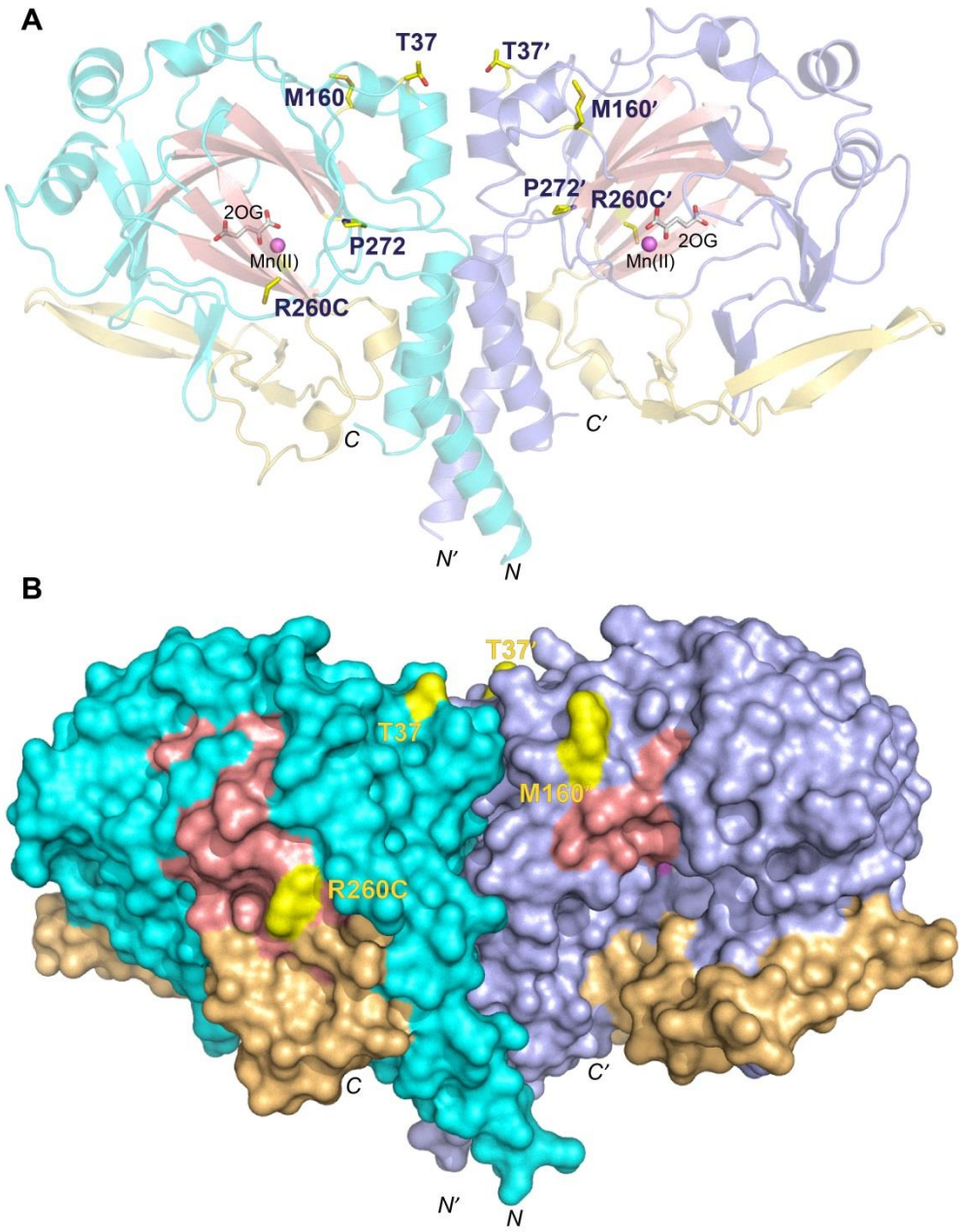


Figure 4.12. View of a JMJD7-R260C structure highlighting the clinically-relevant amino acids as (A) cartoon and (B) surface representations. Colouring is as follows: JMJD7 monomer A, cyan; JMJD7 monomer B, purple; DSBH β-strands, salmon; βVI-V insert, light orange; Mn(II), violet sphere; 2OG, white sticks; and clinically-relevant variants; yellow sticks.

Given the structural data, it is not surprising that T37E substitution does not influence the biophysical or catalytic properties of JMJD7. Although it is situated within the dimerisation domain, T37 is not involved in forming interactions that promote dimerisation. Rather, its surface-exposed position is consistent with the identification of T37 as a phosphorylation site in proteomic screens (Dr Mathew Coleman, personal communication). Given its orientation away from the active site (**Figure 4.12**), together with the finding that T37E substitution had no

observable effect on JMJD7 activity in isolated protein studies, it is possible that phosphorylation of T37 has no effect on JMJD7 catalytic activity in cells, but rather may play an alternative cellular role (e.g. signalling or allosteric binding).

M160 is a surface-exposed residue, situated *N*-terminal to βI and pointing away from the JmjC active site (**Figure 4.12**). Thus, crystallographic analyses are consistent with the minimal effect observed on JMJD7 hydroxylase activity for this variant. JMJD7 crystallographic analyses reveal no apparent structural or catalytic roles for this residue. Although the M160V substitution has been associated with autism, it is noteworthy that this substitution was made within the context of the JMJD7-PLA2G4B protein, which is likely structurally distinct from JMJD7. Overall, the structural and biochemical analyses suggest it is unlikely that a disruption in JMJD7 catalytic activity is a cause of autism/cancer.

Given the surface-exposed nature of the R260C variant, it was initially somewhat surprising that this substitution did not promote disulphide bond formation and/or affect protein aggregation/oligomerisation. However, as shown and described in **Section 3.3**, there are multiple interactions involved in forming a strong JMJD7 dimerisation interface (arising from the hydrophobic interactions between *N*- and *C*-terminal α-helices and disulphide bond formation between C47 of each monomer). Thus, it is possible that the interactions arising from the JMJD7 dimerisation interface override any effect that might otherwise arise from the R260C substitution. While the R260C substitution could theoretically be forming a different dimerisation interface, which would not be detected by the biophysical methods (i.e. non-denaturing MS and SEC-MALS) used in this study, it is unlikely given that JMJD7 activity is likely dependent on its native dimerisation mode (**Section 3.3.6**) and that R260C substitution was found to have no effect on hydroxylase activity (**Section 4.3**).

Overall, crystallographic analyses are consistent with the biophysical and biochemical analyses of the JMJD7 clinically-relevant variants. However, it is possible that these variants affect full-length substrate binding and/or allosteric binding with binding partners that may have not yet been identified.

4.6 Discussion and future perspectives

The results presented in this chapter demonstrate that the cancer (T37E, M160V)- and autism (M160V, R260C)-associated JMJD7 substitutions[139,140] do not have any observable effects on isolated JMJD7 catalytic activity, structural fold, protein stability, and oligomeric state. It is important to consider these activity assays in light of the fact that they were performed with recombinant JMJD7 and DRG1-2 peptide substrates. Thus, the possibility remains that these substitutions may play a different role within the context of a complex cellular environment. Given that no effect on JMJD7 hydroxylase activity was observed with the DRG1-2 peptide substrates, it is unlikely that these peptides exhibit differences in JMJD7 binding. However, it is possible that the JMJD7 R260, M160, and/or T37 residues are involved in binding residues in full-length DRG1-2 that are not present in the peptide substrates. Thus, it would be interesting to investigate whether these substitutions have any effect on JMJD7 activity using full-length, recombinant DRG1-2 and analysis by intact protein LC-MS. Isothermal calorimetry experiments can also be performed to investigate whether any differences in hydroxylase activity can be attributed to changes in enzyme:substrate binding affinity arising from these substitutions. Beyond differences in binding, the investigation of JMJD7 phosphorylation sites and the identification of kinases acting on JMJD7 may provide insights into functionally relevant modified residues in normal physiology and disease.

While the R260C, M160V, and T37E JMJD7 variants were adequately produced in *E. coli*, the endometrial cancer-associated P272S variant was not. Given the inability to produce/purify the P272S variant (**Figure 4.5**), together with bioinformatic and structural analyses that indicate an

important and conserved structural role for this residue within the DSBH (**Figure 4.2** and **Figure 4.12**), it is likely that this point mutation cannot maintain the structural integrity of the DSBH catalytic core fold and consequently has deleterious effects on JMJD7 stability and catalytic activity. Given the highly conserved nature of the JMJD7 P272 residue within the DSBH (**Figure 4.2**) and the striking effects on JMJD7 stability, further study of the P272S variant is warranted. However, given the difficulty in expressing the P272S variant in HEK293T cells (Dr Mathew Coleman, University of Birmingham, personal communication) and in purifying the variant from bacterial cells, further biophysical analyses pose a challenge.

Though the T37E, M160V, and R260C variants have no substantial effect on activity, stability, and oligomeric state, it is possible that these variants play other roles. For example, these substitutions may affect DRG1-2 substrate binding (as described above), the binding of presently unidentified allosteric binding partners, JMJD7 subcellular localisation, and/or play a role independent of hydroxylase activity. While great efforts were made to biochemically characterise JMJD7, the limited reports on the known physiological functions of the DRGs made it a challenge to define role for JMJD7 in DRG regulation. Thus, future studies should focus their efforts on identifying the physiological role of JMJD7 (as discussed in **Section 2.4.6**) prior to returning to the study of JMJD7 clinically-relevant variants.

Targeting 2OG oxygenases for disease is not a novel concept given their roles in epigenetic regulation and hypoxia. Currently, there are compounds in clinical trials for the treatment of cancer (e.g. inhibitors for the JmjC KDM5 subfamily)[246] and anaemia (e.g. inhibitors for PHD)[79] targeting the human 2OG oxygenases. Thus, JMJD7 may also prove to be a viable target for the treatment of disease. Following the identification of a physiological role for JMJD7, further biochemical investigation of the clinically-relevant variants is encouraged to direct future cell-based studies with the goal of developing targeted JMJD7 therapeutics. JMJD7 inhibition by the broad spectrum 2OG oxygenase inhibitor, NOG, has demonstrated that JMJD7 is a druggable target (**Sections 2.2.6 and 2.3.3.1**). Given the current evidence for

JMJD7 as a tumour suppressor gene, the future development of *JMJD7*-targeted gene therapy or small molecule agonists may be undertaken to restore JMJD7 tumour suppressor function. Given that JMJD7 is the only human C3 lysyl hydroxylase discovered to date, it might be possible to exploit (3*S*)-hydroxylysine as a biomarker and develop anti-(3*S*)-hydroxylysine antibodies as a diagnostic tool in cancer and/or intellectual disabilities.

As yet, the development of any targeted JMJD7 therapeutic is likely far away, given that the determination of the functional role(s) of JMJD7 in normal physiology and more detailed structural characterisation (i.e. JMJD7:DRG1-2 complex structure) are still required. However, the results presented in **Chapters 2-4** provide the first reported biochemical and structural characterisation of the previously uncharacterised 2OG oxygenase, JMJD7, laying the foundation and direction for future studies of JMJD7 in normal physiology and disease.

5 IDENTIFICATION OF A NEW PTM CATALYSED BY JUMONJI DOMAIN-CONTAINING 4 (JMJD4)

5.1 Introduction

5.1.1 JMJD4 is a 2OG oxygenase acting on eRF1

As described in **Chapter 1**, there are several recent reports linking human 2OG oxygenases with the regulation of protein biosynthesis at the translational level (**Figure 1.12**)[11]. These enzymes include TYW5[116], which acts in the biosynthesis of hydroxywybutosine in tRNA, and the RiOXs, MINA53[122], NO66[122], and OGFOD1[129], which catalyse the hydroxylation of ribosomal proteins. As described in **Chapter 2**, JMJD7 is another 2OG oxygenase acting on translational machinery; its substrates, DRG1 and DRG2, have been associated with polysomal and ribosomal binding[166,196–198], though the exact functional roles of these proteins and JMJD7-catalysed hydroxylation have yet to be determined.

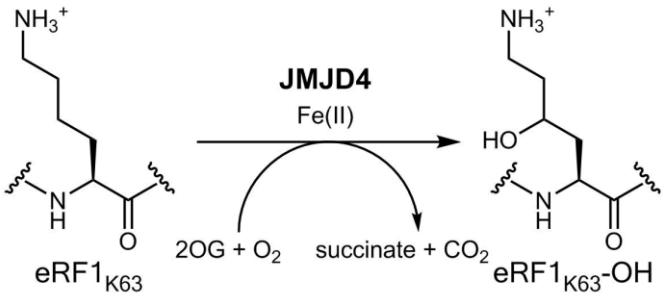


Figure 5.1. Schematic outline of $eRF1_{K63}$ C4 lysyl hydroxylation catalysed by JMJD4.

Like the enzymes described above (with the exception of OGFOD1), JMJD4 is a translation-associated 2OG oxygenase belonging to the JmjC subfamily. Like JMJD7, JMJD4 was shown to interact with and modify a piece of ribosome-associated translational machinery – eukaryotic release factor 1 (eRF1)[133]. eRF1 is a key release factor responsible for the accurate and efficient termination of protein biosynthesis[247]. Using combined cellular, mass spectrometric, and functional studies, JMJD4 was characterised as a C4 lysyl hydroxylase acting specifically on

K63 of eRF1 (**Figure 5.1**)[133]. These studies further revealed that JMJD4-catalysed hydroxylation of K63 increased translation efficiency in cells. Established stop codon read-through reporter assays (i.e. *Renilla*:stop:*firefly* luciferase assay) indicated that siRNA-mediated knockdown of *JMJD4* or *eRF1* resulted in read-through of all three stop codons. Further evidence was provided by an *in vitro* reconstituted translation system, which measures the release of translated peptides from stalled pretermination complexes; the assay similarly demonstrated that partial (> 60%) K63 hydroxylation increased translation termination efficiency as compared with wildtype eRF1 and a K63R eRF1 variant[133].

5.1.2 Translation termination and the role of eRF1

Translation, or ribosome-mediated protein biosynthesis, is the process by which messenger RNA (mRNA; transcribed from DNA) is "translated" to a sequence of amino acids that make up a protein. Translation, which occurs in three phases (initiation, elongation, and termination), takes place in the cytoplasm by translational machinery within the ribosome. Following translation initiation, polypeptide chain elongation occurs via the transfer of the appropriate aminoacyl-transfer RNA (tRNA), which covalently links the amino acid corresponding to the mRNA codon[247]. Termination of polypeptide elongation, the final step of protein biosynthesis, occurs when a stop codon enters the ribosomal A site following mRNA translocation, signifying the end of a coding region[247] (**Figure 5.2**). Class 1 release factors are then responsible for recognising the stop codons in the decoding site of the small ribosomal subunit. In eukaryotes, the class 1 release factor, eRF1, is capable of recognising all three stop codons (UAA, UGA, or UAG)[248,249]. Interestingly, though the three stop codons are conserved in humans and bacteria, prokaryotic class 1 release factors, RF1 and RF2, do not exhibit the same omnipotent (i.e. ability to recognise all three stop codons) specificity as eRF1; RF1 exclusively recognises UAA and UAG stop codons, while RF2 recognises only UAA and UGA[250].

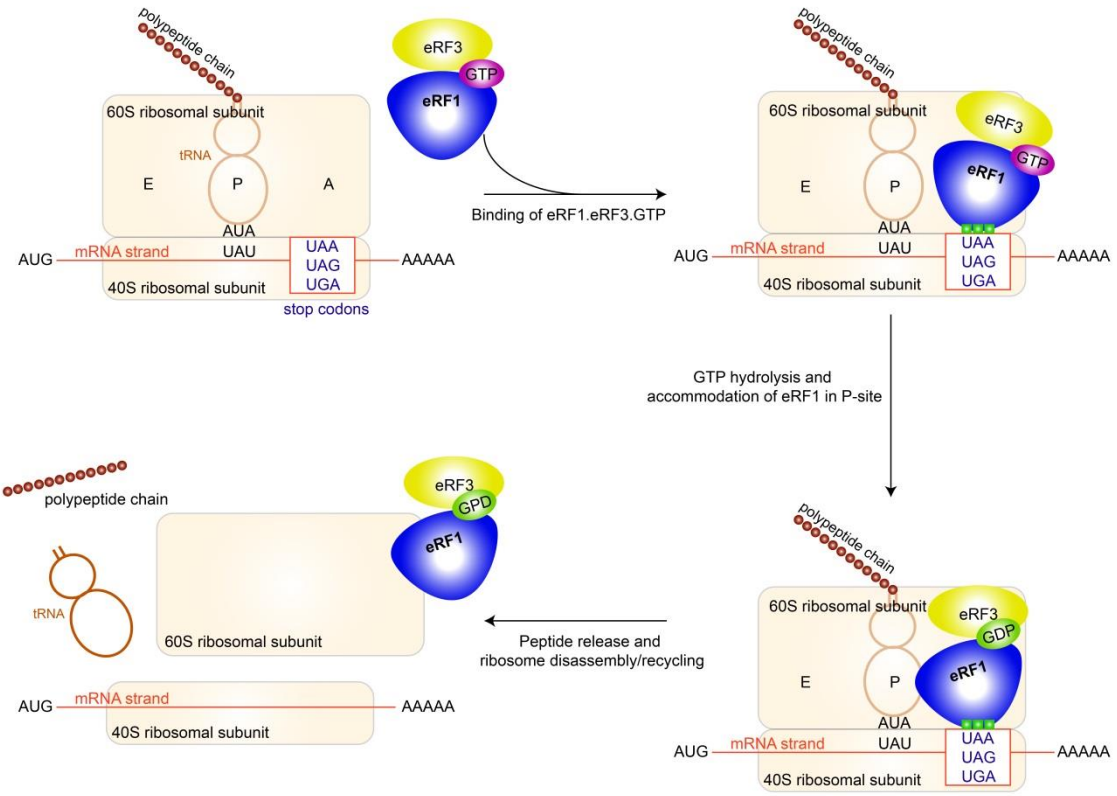


Figure 5.2. Schematic illustration of eRF1-mediated stop codon recognition and polypeptide chain release in eukaryotic translation termination. When a stop codon enters the ribosomal aminoacyl (A) site, eRF1 (blue) is responsible for recognising the stop codon at the decoding site of the 40S ribosomal subunit[248]. eRF1 exists in a ternary complex with eukaryotic release factor 3 (eRF3; yellow) and guanosine triphosphate (GTP; purple)[251]; eRF1 structurally resembles a transfer RNA (tRNA)[168,252]. eRF1 binding and accommodation in the peptidyl site (P-site) is promoted by GTP hydrolysis (GDP; green)[253,254]. eRF1 then interacts with the messenger RNA (mRNA) transcript and ribosome[254,255]. eRF1 triggers hydrolysis of the peptidyl-tRNA ester bond, releasing the nascent polypeptide chain[248,249]. Ribosome disassembly follows.[247,256] Figure adapted from Shoemaker, CJ & Green, R (2011)[256].

eRF1 exists in a ternary complex with the class II release factor, eukaryotic release factor 3 (eRF3), and guanosine triphosphate (GTP)[251,257]. Interestingly, eRF3 is a translation factor-associated GTPase that is related to bacterial elongation factor-Tu (EF-Tu)[174], as both belong to the TRAFAC superfamily of GTPases[174]. EF-Tu is a substrate for PPHD[136], a bacterial 2OG oxygenase related to human PHD2 that catalyses prolyl hydroxylation. The ternary eRF1·eRF3·GTP complex engages the ribosome via GTP hydrolysis by eRF3[251], which enables eRF1 to interact with the ribosome and mRNA, allowing recognition of stop codons in the A-site[253,254]. Stop codon recognition is then followed by eRF1-mediated hydrolysis of the peptidyl-tRNA ester bond in the ribosomal peptidyl transferase centre (P-site), releasing the

nascent polypeptide chain[255]. Ribosome disassembly and recycling follow, enabling future translation cycles[247,256] (**Figure 5.2**).

5.1.3 eRF1, the NIKS motif, and stop codon recognition

eRF1 is a multi-domain protein comprised of *N*-terminal (*N*-), middle (M-), and *C*-terminal (*C*-) domains[168]. The *N*-domain is primarily responsible for decoding stop codons[134,135,258], while the M-domain of eRF1 promotes nascent polypeptide chain release[168,259,260], and the *C*-domain facilitates interactions with eRF3[252,255,261]. A crystal structure of eRF1 highlights its three distinct domains[168] (**Figure 5.3**).

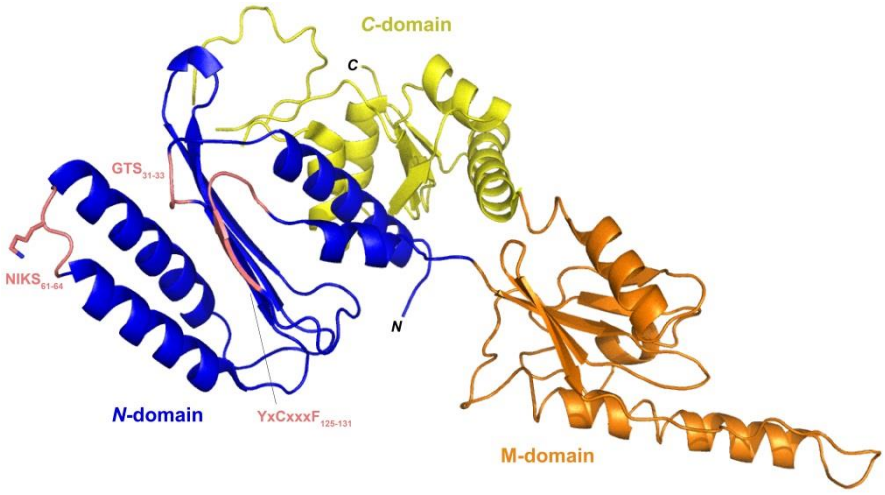


Figure 5.3. View from a crystal structure of eRF1 (PDB ID: 1DT9[168]). eRF1 is comprised of three domains – the *N*- (blue), M- (orange) and *C*-domains (yellow). JMJD4 catalyses the C4 hydroxylation of K63 (salmon, stick representation), which protrudes from the $NIKS_{61-64}$ motif (salmon). The NIKS motif is located within a tight turn between two α-helices in the *N*-domain. The NIKS, GTS_{31-33}, and $YxCxxxF_{125-131}$ motifs (salmon) have been identified as having important roles in stop codon recognition.

The exact molecular mechanism of stop codon recognition by eRF1 remains unknown, though several functional studies have identified regions within the *N*-domain that are important or essential for stop codon recognition and ribosome binding. For example, crosslinking and mutation studies have shown that the highly conserved $NIKS_{61-64}$ loop is functionally essential in translation[134,135]. Notably, K63, the residue hydroxylated by JMJD4, was identified to cross-link with the invariant uracil in the first position of all three stop codons in studies using photo-

activatable stop codons[134]. In addition, mutation of I62, K63, and S64 were shown to reduce ribosome binding and release factor activity[135]. Both studies highlighted the structural mimicry between the tRNA anticodon loop and the distinct position of the NIKS loop on the apex of a helix-turn-helix[134,135,168]. Other regions within the *N*-domain, including the conserved GTS_{31-33}[252,262,263] and $YxCxxxF_{125-131}$ motifs[252,262,264–266], have been identified as having roles in stop codon recognition through a combination of indirect (e.g. genetic and computational) and direct (e.g. structural and mutation) studies. The motifs reported to be involved in stop codon recognition are highlighted in **Figure 5.3**.

Recent structural studies of eRF1 using cryo-electron microscopy (EM) have provided insights into the mechanism of stop codon recognition[254,267]. For example, a cryo-EM-derived structure of an eRF1-eRF3-guanosine 5′-[β,γ-imido] triphosphate (GMPPNP) complex bound to a human ribosomal pre-termination complex (i.e. pre-GTP hydrolysis) provided direct structural evidence for interactions between the decoding centre of the 40S ribosomal subunit and the *N*-terminal domain of eRF1[254]. Protein NMR studies of the omnipotent *N*-domain of eRF1 ($eRF1_{ND}$) and a unipotent (UGA-specific) $eRF1_{ND}$ variant ($Q^{122}FM(Y)F^{126}$) have shown that the spatially distant GTS loop adopts alternate conformations in solution, indicating that the GTS loop likely plays a role in the omnipotent nature of eRF1 and has direct access to the stop codons[263]. This study also provided direct evidence for interactions between $eRF1_{ND}$ and ribosomal RNA (rRNA) fragments in solution[263]. Despite these insights and intense efforts to understand stop codon recognition, its precise molecular mechanisms remain largely unknown.

5.1.4 Aims of this chapter

While hydroxylation of K63 by JMJD4 has been found to increase translational termination efficiency,[133] the exact molecular mechanism by which hydroxylation influences termination remains unclear. Given the functional and structural studies highlighting the role of the NIKS motif in stop codon recognition, a more detailed understanding of the molecular mechanism by

which JMJD4-catalysed hydroxylation increases translation termination efficiency is required. The understanding of the role of hydroxylation in stop codon recognition is also important from a medical perspective, given that nonsense mutations leading to premature translation termination account for ~11% of gene lesions that cause inherited human diseases, such as cystic fibrosis and Duchenne muscular dystrophy[268,269].

A truncated construct corresponding to $eRF1_{ND}$ (residues 1-142) has been extensively studied by protein NMR spectroscopy[263,270]. Importantly, the *N*-domain was found to fold in a similar manner as within the full-length protein[270,271]. Given that reported optimised conditions for protein NMR spectroscopy and NMR resonance assignments of $eRF1_{ND}$ were available, protein NMR studies investigating the effect of hydroxylation on $eRF1_{ND}$ structure and dynamics was considered feasible and likely to provide relevant insights into the structural effects of hydroxylation. A PTM like hydroxylation may cause changes in the structure or dynamics of the NIKS motif, which may facilitate intermolecular interactions (e.g. protein/protein or protein/RNA interactions) that cause translation termination. Alternatively, hydroxylation may cause a more global effect by affecting protein stability, as exemplified by the stabilising effect of (4*R*)-hydroxylation of proline residues in collagen by the C-P4Hs[42]. There is also the possibility that JMJD4-catalysed hydroxylation of eRF1 triggers protein turnover and/or acts as a "switch" in a similar manner to PHD2-catalysed hydroxylation of HIF, where hydroxylation (in normoxia) or HIF accumulation (in hypoxia) mediates the transcriptional regulation of genes that enable the cell to adapt to low oxygen conditions[61]. Finally, hydroxylation of eRF1 might disrupt protein-protein/rRNA interactions in a similar way that FIH-catalysed hydroxylation of HIF disrupts CREB/p300 binding[69].

To date, there are no studies on how hydroxylation of K63 by JMJD4 affects eRF1 structure, dynamics, or stability. The work described in this chapter aims to understand the effect of JMJD4-catalysed hydroxylation on $eRF1_{ND}$ via structural, dynamics, and biophysical studies of hydroxylated $eRF1_{ND}$ ($eRF1_{ND}$-OH) using protein NMR spectroscopy and other biophysical

techniques. The results presented in this chapter aim to help elucidate the mechanism by which JMJD4-catalysed hydroxylation of eRF1 improves translation termination efficiency.

5.2 Expression and purification of $eRF1_{ND}$ and JMJD4 reveal an unprecedented +30 Da modification

5.2.1 Strategies to isolate $eRF1_{ND}$-OH for study by NMR spectroscopy

Solution or protein NMR spectroscopy is a biophysical technique that can be used to study proteins and determine protein structures in solution[179]. Because proteins are large biological macromolecules, proteins must be uniformly labelled with ^{15}N- and/or ^{13}C-isotopes to minimise resonance overlaps in 2D and 3D experiments. ^{15}N- and ^{13}C-isotopic labelling is commonly achieved by producing the protein of interest in bacteria grown in minimal media. Minimal media contain a mixture of salts and elements required for bacteria to grow, and allow for the substitution of nitrogen and/or carbon sources (**Section 7.3.2**), which can be introduced by supplementing the media with $^{15}NH_4Cl$ and/or wildtype or ^{13}C-labelled glucose to achieve ^{15}N- and/or ^{13}C-isotopic labelling, respectively[272,273]. The study of proteins by NMR spectroscopy requires high protein yields, high concentrations, and high purity due to the long acquisition time associated with the 3D NMR experiments required for protein backbone assignment[272,273]. Given the strict requirements for the study of proteins by NMR, expression and small-scale trials were initially performed to determine feasibility and optimal conditions.

While conditions have been reported for isotopic labelling and purification of $eRF1_{ND}$ for solution NMR studies[270], the study of hydroxylated $eRF1_{ND}$ ($eRF1_{ND}$-OH) by NMR had not yet been undertaken and thus required optimisation. In particular, high sample purity (> 90% hydroxylation) was amongst the highest priorities given that NMR spectroscopy is a highly sensitive technique; a mixture of unmodified and hydroxylated $eRF1_{ND}$ would certainly result in resonance overlap and ambiguity, consequently complicating NMR analyses. The isolation

of eRF1$_{ND}$-OH would not only rely on efficient separation from JMJD4 post-hydroxylation, but also requires complete hydroxylation (> 90%) due to an inability to separate unmodified eRF1$_{ND}$ from eRF1$_{ND}$-OH.

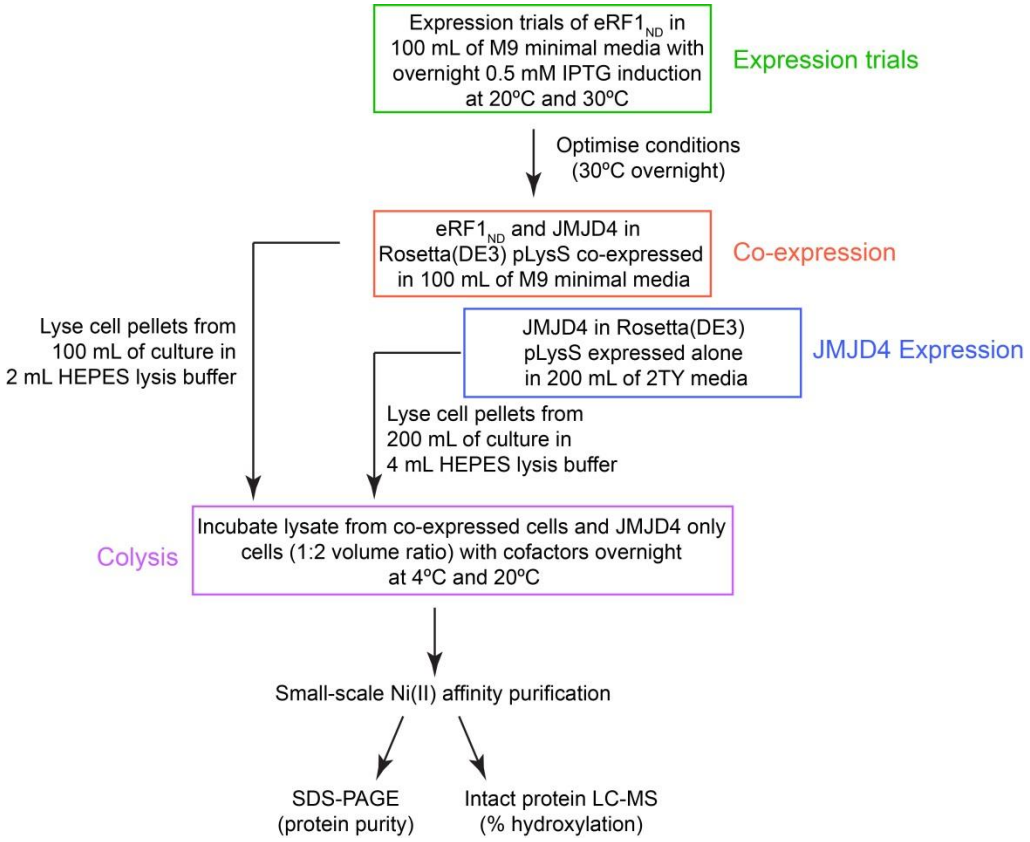


Figure 5.4. Strategies to identify optimal production of JMJD4-catalysed eRF1$_{ND}$-OH. Expression trials (green) were to be performed to identify optimal conditions for eRF1$_{ND}$ production in M9 minimal media. eRF1 and JMJD4 co-expression trials (orange) in M9 minimal media were performed to determine optimal conditions for JMJD4-catalysed hydroxylation of eRF1$_{ND}$. If co-expression alone did not yield near complete hydroxylation of eRF1$_{ND}$, co-expression would be combined with co-lysis (purple) of eRF1$_{ND}$/JMJD4-expressing bacterial cells with JMJD4-expressing bacterial cells (blue). Samples were to be purified using loose Ni(II) Sepharose resin and assessed by SDS-PAGE analyses. The extent of hydroxylation would be monitored by intact protein LC-MS analyses.

The strategies used to identify optimal conditions for the production and isolation of pure eRF1$_{ND}$-OH (> 90%) in M9 minimal media are illustrated in **Figure 5.4** and described below:

(1) Expression trials using the reported eRF1$_{ND}$ construct to determine optimal conditions for the production of *C*-terminally His$_6$-tagged protein in *E. coli* in minimal media.

(2) Co-production of recombinant eRF1$_{ND}$ and JMJD4 in minimal media;

(3) (2) Followed by co-lysis with additional JMJD4-expessing bacterial cells and incubation with added co-factors overnight.

5.2.2 Expression trials

5.2.2.1 Co-expression studies

Protein co-production can be achieved by co-transforming two plasmids with different selectable markers. Thus, sole transformation of pET23b(+)-eRF1$_{ND}$ (Amp^{R}) and co-transformation of pET23b(+)-eRF1$_{ND}$ (Amp^{R}) and pET28a(+)-JMJD4 (Kan^{R}) were performed in Rosetta(DE3) pLysS bacterial cells as described (**Section 7.2.10.1**). Initially, co-expression trials were performed in M9 minimal media (without isotopic labelling), supplemented with the appropriate antibiotics; eRF1$_{ND}$ was expressed either alone (control) or co-expressed with JMJD4 as described in **Table 7.3.5**. Protein production was induced at either 20°C or 30°C with 0.5 mM IPTG to determine optimal conditions for eRF1$_{ND}$ production and JMJD4-catalysed hydroxylation. Proteins were subsequently purified on a small scale using loose Ni(II) Sepharose resin and analysed by SDS-PAGE for differences in eRF1$_{ND}$ production and purity. The extent of JMJD4-catalysed hydroxylation resulting from co-expression was analysed by intact protein LC-MS. As shown in **Figure 5.5**, overnight growth at 30°C following IPTG induction resulted in greater eRF1$_{ND}$ production. However, the extent of hydroxylation under both co-expression conditions was not optimal for subsequent NMR analyses, yielding only ~40-50% of the hydroxylated product.

Note that in the mass spectra shown in **Figure 5.5** and nearly all subsequent mass spectra corresponding to eRF1$_{ND}$ in this chapter, +42 Da mass adduct peaks are observed. The adduct corresponds to the mass of acetonitrile $[M+H+41.05]^{+}$ and is caused by acetonitrile in the solvent elution gradient[274]. Furthermore, note that the resolving power of the open access Waters LCT Premier XE mass spectrometer used for data acquisition and the Waters MaxEnt1

software (MassLynx V4.1) used for deconvolution of raw protein mass spectra account for the observed mass inaccuracies (2-3 Da).

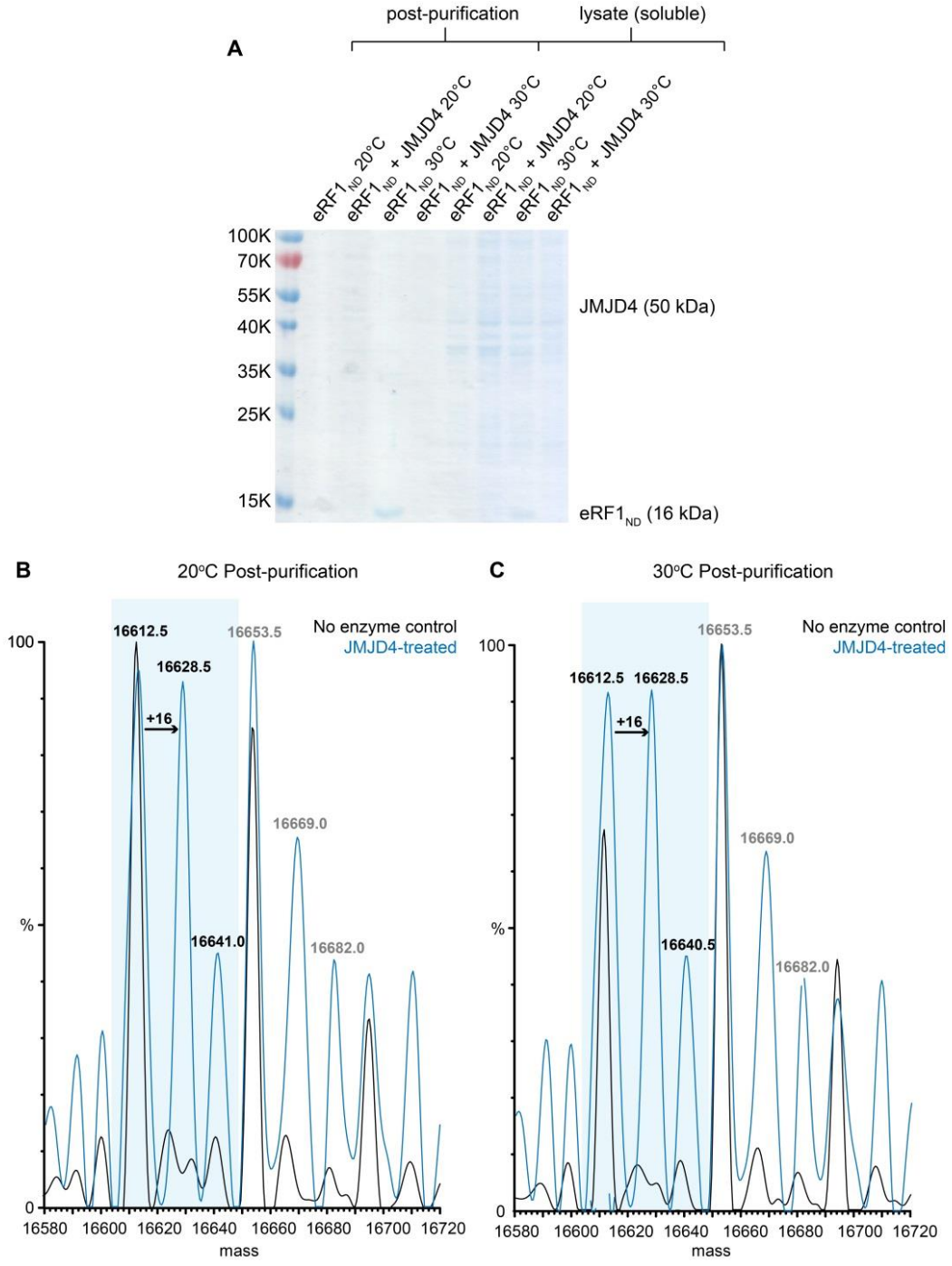


Figure 5.5. Coomassie-stained 12% SDS-PAGE gel and deconvoluted intact protein mass spectra of $eRF1_{ND}$/JMJD4 co-expression trials in M9 minimal media. (A) SDS-PAGE analyses indicate that overnight induction at 30°C yields greater $eRF1_{ND}$ production in M9 minimal media than 20°C. Deconvoluted intact protein MS analyses of $eRF1_{ND}$ expressed alone (black) and co-expressed with JMJD4 (blue) following induction at ***(B)*** 20°C and ***(C)*** 30°C indicates that the extent of hydroxylation (~40-50%) is nearly the same under both conditions and similarly insufficient for NMR analyses. Peaks outside of the blue band (grey) are mass adducts arising from acetonitrile in the solvent gradient[274].

5.2.2.2 Co-lysis studies

Attempts to promote further hydroxylation of $eRF1_{ND}$ (> 90%) were made by performing the co-lysis of co-expressing $eRF1_{ND}$/JMJD4 bacterial cells with JMJD4-expressing cells. $eRF1_{ND}$/JMJD4 co-expression was carried out in 100 mL of M9 minimal media under the optimal conditions described in **Section 5.2.2.1**. Note that recombinant JMJD4 was produced using reported optimised conditions[133] (induction with 0.5 mM IPTG followed by overnight incubation at 20°C; Dr Sarah Wilkins, personal communication). Following co-lysis with JMJD4-expressing cells (from 200 mL of 2TY media), additional co-factors [100 μM Fe(II), 200 μM 2OG, and 500 μM ascorbate in 50 mM HEPES, pH 7.5; final concentrations] were added to the lysate prior to incubation overnight at either 4°C or 20°C to promote hydroxylation of $eRF1_{ND}$.

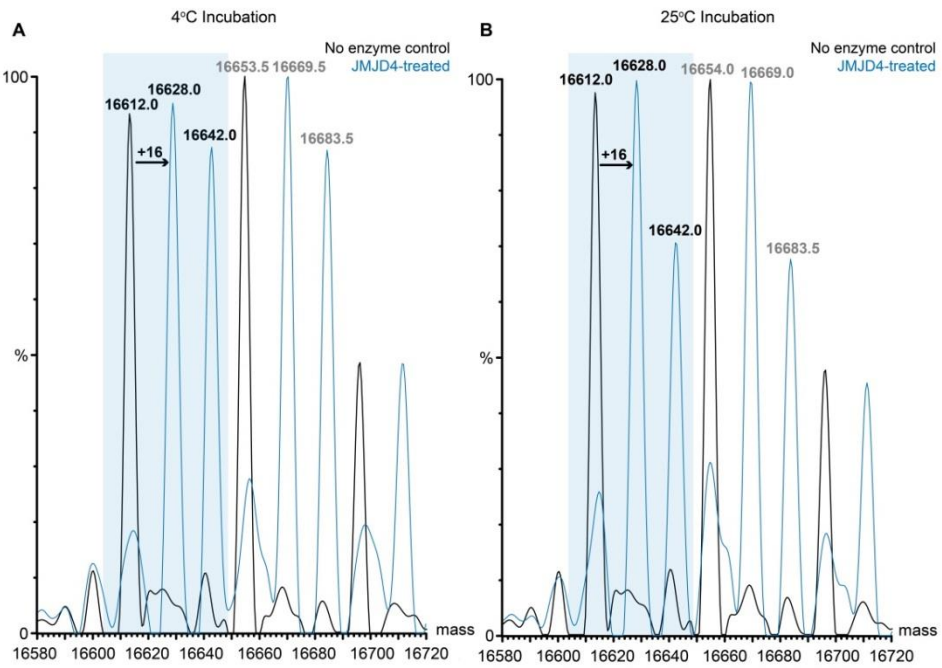


Figure 5.6. Deconvoluted intact protein mass spectra of $eRF1_{ND}$/JMJD4 co-expression/co-lysis trials following overnight incubations with JMJD4 lysate at (A) 4 °C and (B) 25 °C in the absence (black) and presence (blue) of JMJD4 co-expression/co-lysis. Note that both conditions yield sufficient hydroxylation (85-90%) for subsequent isolation and analysis by protein NMR spectroscopy. Peaks outside of the blue band (grey) are mass adducts arising from acetonitrile in the solvent gradient[274].

Assays were analysed by intact protein LC-MS following small-scale purification with loose Ni(II) Sepharose resin. As shown in **Figure 5.6**, the extent of hydroxylation was > 85%. The appearance of a dominant third species (roughly +30 Da from the unmodified $eRF1_{ND}$ peak) was observed in JMJD4-treated samples, but was not present in the no enzyme controls. Because the apparent +30 Da species was not consistent with any known JMJD4-catalysed modifications and due to the poor signal-to-noise, the species was initially regarded as a likely MS artefact/adduct.

5.2.2.3 Attempts to express and purify ^{15}N-$eRF1_{ND}$-OH for NMR studies

Given that the described co-expression/co-lysis conditions described in **Section 5.2.2.2** yielded 85-90% apparent hydroxylation of $eRF1_{ND}$, these conditions were selected to produce ^{15}N-labelled $eRF1_{ND}$-OH on a larger scale. $eRF1_{ND}$ was expressed alone (1 L of M9 minimal media supplemented with $^{15}NH_4Cl$) or co-expressed with JMJD4 (2 L of M9 minimal media supplemented with $^{15}NH_4Cl$) as described above; JMJD4 was produced separately (3.6 L of 2TY media). Pellets were re-suspended in HEPES lysis buffer (**Table 7.14**) and co-lysis was performed as outlined in **Figure 5.7A-B**. $eRF1_{ND}$ was then purified away from JMJD4 using an S75 300 mL gel filtration column. Overall, the described expression and purification of ^{15}N-$eRF1_{ND}$ indicated that (1) $eRF1_{ND}$ could be successfully ^{15}N-labelled using the optimised expression conditions in M9 minimal media, as evidenced by intact protein LC-MS (**Figure 5.7C**), and (2) $eRF1_{ND}$ (16.6 kDa) could be successfully purified from JMJD4 (49.5 kDa) using SEC (**Figure 5.7A-B**).

Strikingly, however, analysis of purified ^{15}N-$eRF1_{ND}$ by intact protein LC-MS following co-expression/co-lysis with JMJD4 did not yield any evidence for the hydroxylated product, which is consistent with a +16 Da mass shift from wildtype $eRF1_{ND}$; rather an unprecedented and frankly unexpected +27 Da shift was observed (**Figure 5.7C**). The near complete conversion of $eRF1_{ND}$ to this unexpected modified state strongly suggested that the (roughly +30 Da) third

dominant species previously observed in the small-scale co-expression/co-lysis trial was not a MS adduct as initially anticipated (**Figure 5.6**). Furthermore, the appearance of this modification complicated the isolation of hydroxylated $eRF1_{ND}$ given near complete conversion to this '+30 Da' state. Thus, the characterisation of this modification was required prior to subsequent biophysical studies of $eRF1_{ND}$ and $eRF1_{ND}$-OH. The determination of whether this modification is enzyme-dependent or merely an artefact of the described assay conditions also required further investigation.

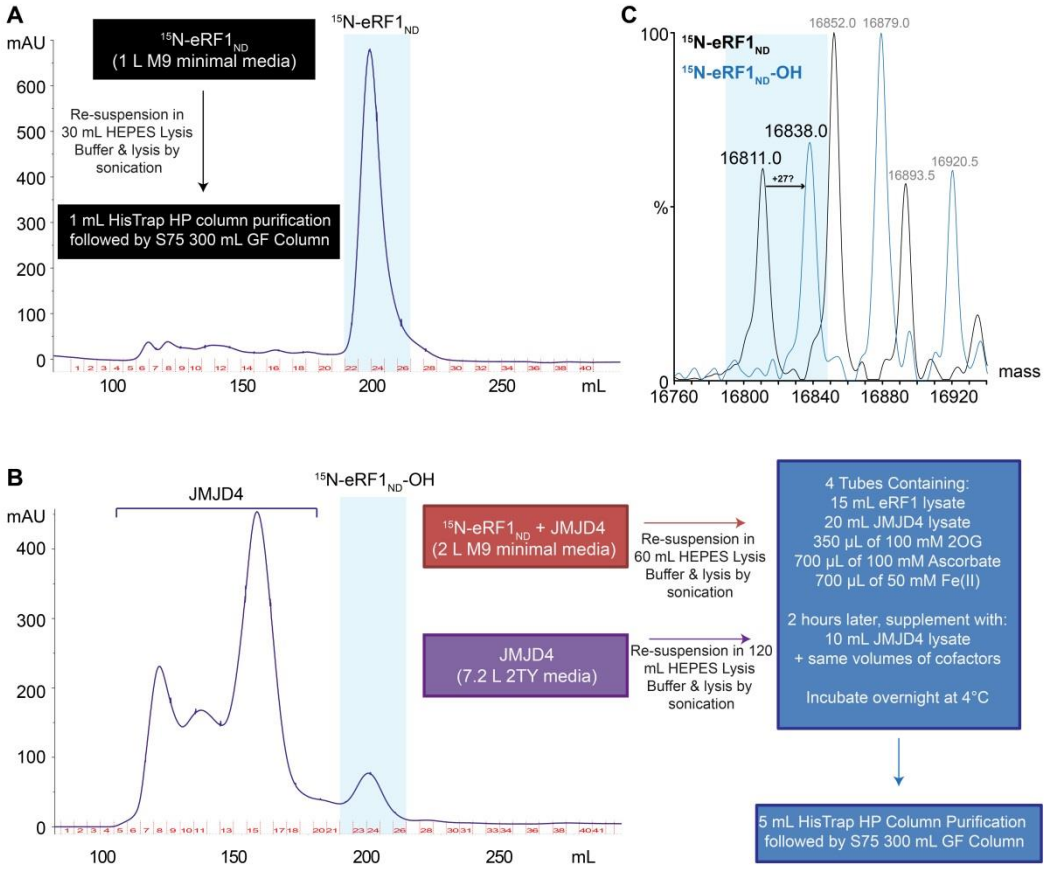


Figure 5.7. Expression, hydroxylation, and purification of ^{15}N-eRF1$_{ND}$ yields an unprecedented modification. SEC UV traces (S75 300 mL GF column) and purification outline of ***(A)*** ^{15}N-eRF1$_{ND}$ and ***(B)*** ^{15}N-eRF1$_{ND}$-OH following co-expression and co-lysis with JMJD4. Note that efficient separation of ^{15}N-eRF1$_{ND}$ is possible by SEC. ***(C)*** Deconvoluted intact protein LC-MS analyses of ^{15}N-eRF1$_{ND}$ (black) indicates successful isotopic ^{15}N-labelling (theoretical monoisotopic/average molecular weights: 16808.1 / 16817.6 Da, as calculated by PROTEIN CALCULATOR v3.4 - http://protcalc.sourceforge.net/). However, LC-MS analyses of ^{15}N-eRF1$_{ND}$-OH indicates near complete conversion (~90%) to a species +27 Da ('+30 Da') from unmodified ^{15}N-eRF1$_{ND}$, which is inconsistent with hydroxylation (+16 Da mass shift).

5.2.3 The '+30 Da' modification is JMJD4-dependent

Given the near complete conversion of $eRF1_{ND}$ to the '+30 Da' modified state, it was important to determine whether this unprecedented modification was JMJD4-dependent. Using reported optimised conditions described in **Section 7.3.5**, $eRF1_{ND}$-His_6 and His_6-JMJD4 were separately expressed in Rosetta(DE3) pLysS cells in 2TY media and purified using Ni(II) affinity chromatography followed by SEC (**Figure 5.8**).

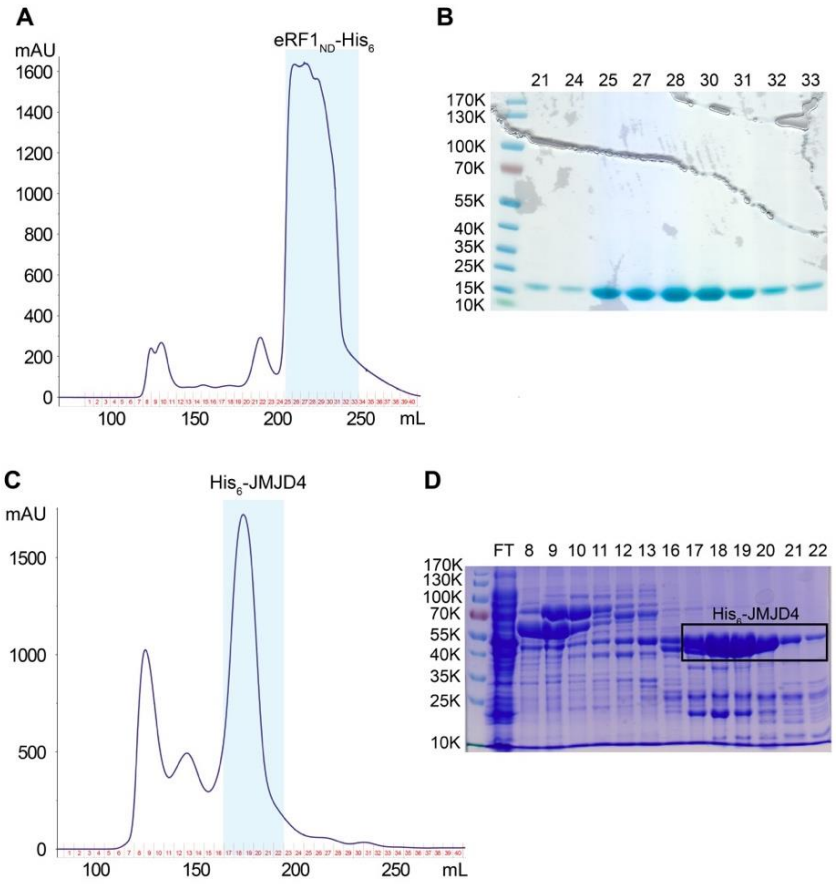


Figure 5.8. Representative SEC UV traces and corresponding Coomassie-stained 12% SDS-PAGE gels following purification by S75 300 mL GF column for (A-B) $eRF1_{ND}$ and (C-D) JMJD4.

Isolated protein assays were then performed using recombinant $eRF1_{ND}$ in the absence (no enzyme control) or presence of JMJD4 with the indicated co-factors [2OG, Fe(II), and/or ascorbate]; assays were analysed by intact protein LC-MS. As shown in **Figure 5.9A**, the hydroxylated and '+30 Da' modified protein states were observed to a small extent following incubation of $eRF1_{ND}$ with JMJD4 in a 5:1 substrate:enzyme ratio at 37°C for 2 hours. Formation of the +30 Da modified state was promoted by adding an excess of JMJD4 (1:3

substrate:enzyme ratio) (**Figure 5.9B**). Notably, neither the hydroxylated nor +30 Da modified $eRF1_{ND}$ states are observed in the no enzyme control [$eRF1_{ND}$ plus Fe(II), 2OG, and ascorbate]. Consistent with 2OG oxygenase catalysis, JMJD4 relies on 2OG, but not ascorbate (**Figure 5.9C and E**; note that ascorbate was consequently not used in subsequent assays).

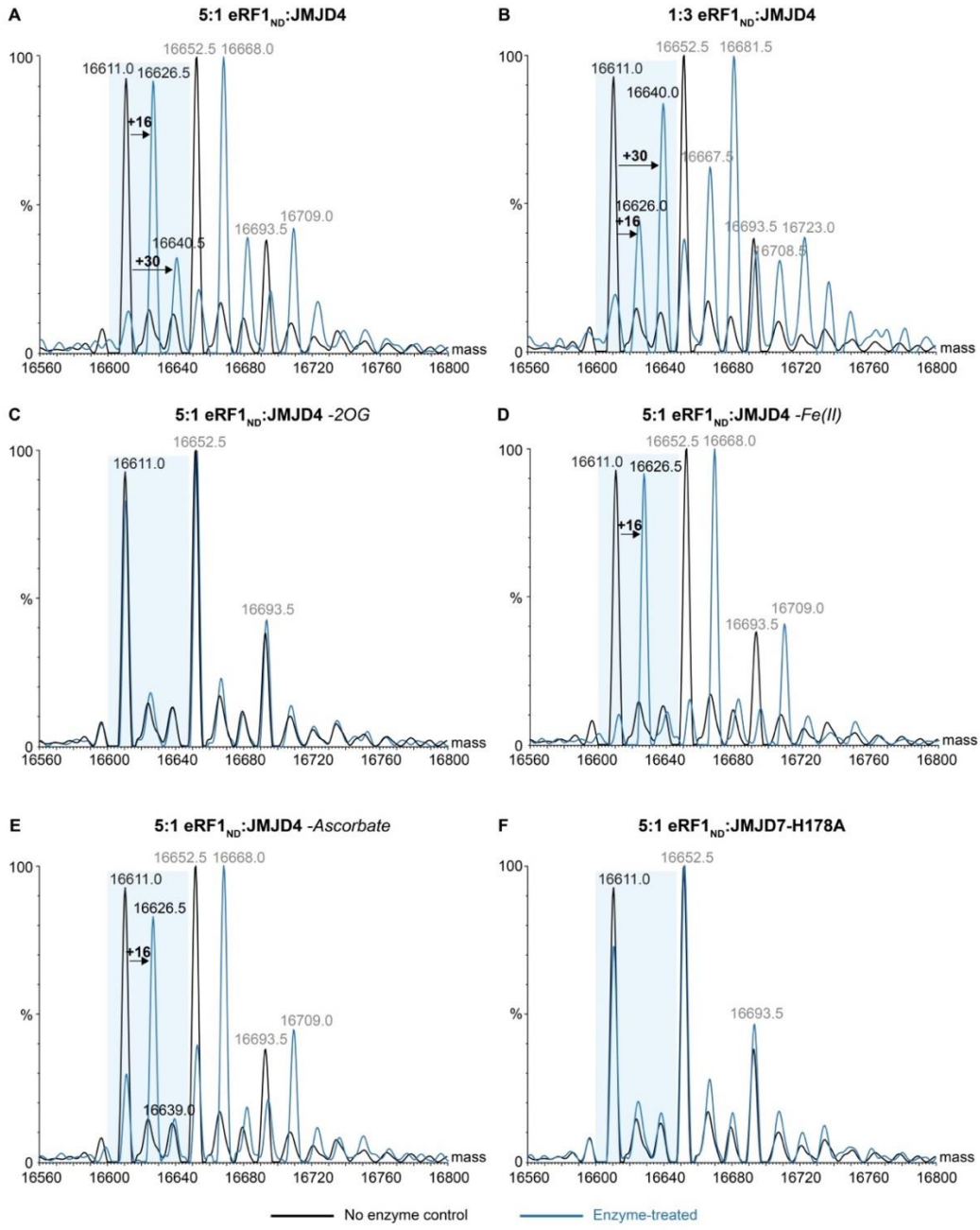


Figure 5.9. Deconvoluted intact protein mass spectra corresponding to JMJD4-dependence assays performed with recombinant $eRF1_{ND}$ in the absence (no enzyme control, black) or presence (enzyme-treated, blue) of JMJD4 as indicated: (A) 5:1 molar ratio of $eRF1_{ND}$:JMJD4; ***(B)*** 1:3 molar ratio of $eRF1_{ND}$:JMJD4; ***(C)*** 5:1 molar ratio of $eRF1_{ND}$:JMJD4 in the absence of 2OG; ***(D)*** 5:1 molar ratio of $eRF1_{ND}$:JMJD4 in the absence of Fe(II); ***(E)*** 5:1 molar ratio of $eRF1_{ND}$:JMJD4 in the absence of ascorbate; and ***(F)*** 5:1 molar ratio of $eRF1_{ND}$:JMJD7-H178A. Reaction conditions: 8 µM $eRF1_{ND}$, 1.6 or 24 µM JMJD4, 200 µM Fe(II), 200 µM 2OG, 500 µM ascorbate, 50 mM Tris, pH 7.5, at 37°C, 2 h. Samples were analysed using a Waters LCT Premier XE instrument (**Section 7.8.2.1**).

Interestingly, $eRF1_{ND}$ hydroxylation is observed in the absence of Fe(II), a requirement for 2OG oxygenase catalysis (**Figure 5.9D**). However, this result should be interpreted with caution given that it is likely for JMJD4 to have co-purified with Fe(II) in the active site during bacterial expression and purification (note that JMJD4 was not produced in the *apo*-form). Furthermore, $eRF1_{ND}$ was incubated with JMJD7-H178A to determine whether the modification was specific to JMJD4 oxygenase catalysis; however, neither hydroxylation nor the +30 Da modification was observed (**Figure 5.9F**). Overall, the isolated protein JMJD4 dependence assays indicated that the +30 Da modification is JMJD4-dependent and not an artefact of bacterial co-expression.

5.2.4 Validation of the +30 Da modification by high resolution protein MS

Analysis of this previously unreported JMJD4-catalysed modification by intact protein LC-MS yielded a fairly large mass range (+27-30 Da difference). Due to the observed peak broadness, it was initially thought that this +30 Da peak might be indicative of a second hydroxylation, which would be consistent with reported JMJD4 hydroxylase activity and a +32 Da mass shift. The observed peak broadness and mass accuracy is a result of the instrument used for intact protein analysis (open access Waters LCT Premier/XE) and deconvolution of raw protein mass spectra (Waters Mass Lynx V4.1, MaxEnt1). Therefore, more accurate and higher resolution MS analyses were performed using a Thermo Scientific™ Q Exactive™ Hybrid Quadrupole-Orbitrap™ (Waters) mass spectrometer[275] (**Section 7.8.2.3**), which confirmed a mass difference of 30 Da (**Figure 5.10**).

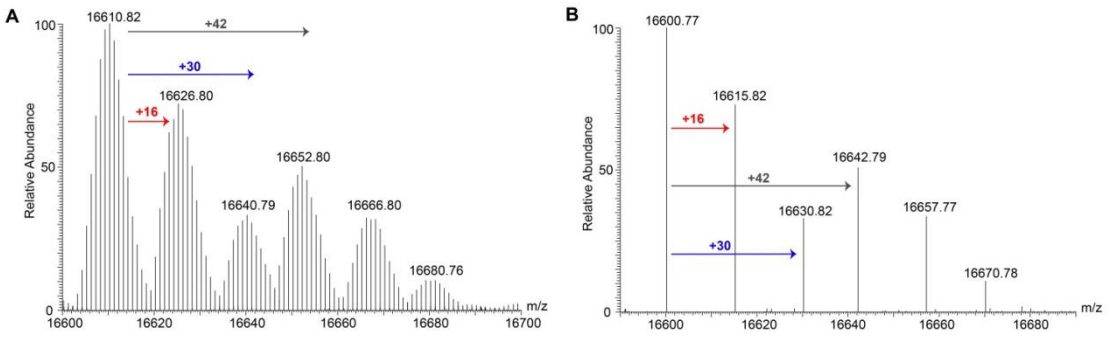


Figure 5.10. High resolution MS confirms a mass difference of 30 Da. A JMJD4-treated $eRF1_{ND}$ sample (1.5 h, 37°C) was spiked with unmodified $eRF1_{ND}$. A mass difference of 30 Da was confirmed using a high resolution Thermo Scientific Q-Exactive Hybrid Quadrupole-Orbitrap™ mass spectrometer, as shown in the deconvoluted ***(A)*** average isotopic and ***(B)*** monoisotopic mass spectra. This excluded the possibility of two hydroxylations, consistent with a 32 Da mass difference. Note that the +42 Da mass shift is consistent with a mass adduct caused by acetonitrile, which is present in the solvent gradient[274].

5.2.5 Modified aims

To date, JMJD4 has only been reported to catalyse a single hydroxylation at K63 of eRF1[133]. Thus, the appearance of the dominant +30 Da species under certain conditions was surprising and required further investigation. Given the unanticipated appearance of the +30 Da modification and the important role of K63 in stop codon recognition, the aims of this project/chapter were adjusted to include the characterisation of the JMJD4-catalysed +30 Da modification and understand its biological relevance.

The work described in this chapter aims to:

(1) identify the target residue and characterise the JMJD4-catalysed +30 Da modification;

(2) investigate the effects of hydroxylation and the +30 Da modification on JMJD4 structure, dynamics, stability, and rRNA binding using protein NMR spectroscopy and other biophysical techniques; and

(3) determine the biological relevance of the '+30 Da modification.'

Overall, the results presented in this chapter aim to characterise an unreported JMJD4-catalysed PTM and understand the effects of hydroxylation and the +30 modification on eRF1 and protein translation.

5.3 Isolation of $eRF1_{ND}$-OH and $eRF1_{ND}$-30

To study the effect of hydroxylation and the +30 Da modification on $eRF1_{ND}$ by NMR spectroscopy and other biophysical techniques, the determination of optimal conditions for the isolation and purification of the hydroxylated and +30 Da protein ($eRF1_{ND}$-30) states was still required. The enzyme-dependence assays described in **Section 5.2.3** revealed that the hydroxylated product appeared to be kinetically favoured, while formation of $eRF1_{ND}$-30 relied on initial hydroxylation. The sequential product formation was therefore exploited to isolate $eRF1_{ND}$-OH and $eRF1_{ND}$-30 for subsequent biophysical analyses.

5.3.1 Isolation of $eRF1_{ND}$-OH

Isolated protein assays revealed that the hydroxylated product was kinetically favoured; formation of $eRF1_{ND}$-OH was observed after 5 minutes of incubation at 37°C. After 30 minutes of incubation at 37°C (200 μM $eRF1_{ND}$, 20 μM JMJD4, 100 μM Fe(II) and 200 μM 2OG in 20 mM Tris, pH 7.5), > 90% of the hydroxylated product was observed, with no evidence of the +30 Da modified protein (**Figure 5.11**). Thus, these conditions were determined to be optimal for the isolation of the hydroxylated product for NMR studies (rather than the combined co-expression/co-lysis strategy described in **Section 5.2.2.2**). To prevent formation of $eRF1_{ND}$-30, the reaction was quenched with 5 – 7 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0 (final concentration) and incubated on ice.

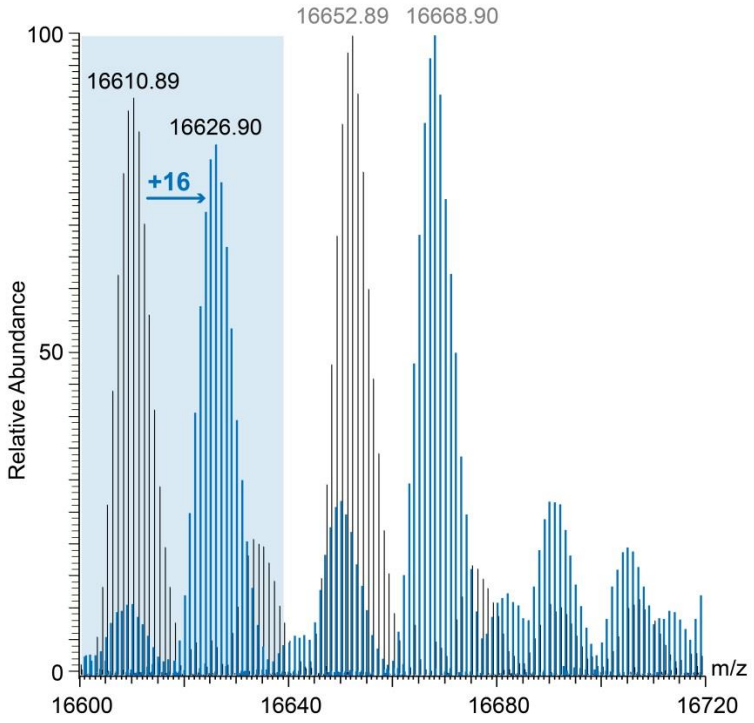


Figure 5.11. Representative deconvoluted mass spectra showing optimised conditions for eRF1$_{ND}$-OH formation. Optimised conditions for eRF1$_{ND}$-OH are as follows: Incubation of 200 µM eRF1$_{ND}$, 100 µM Fe(II), and 200 µM 2OG (black) with 20 µM JMJD4 (blue) for 30 minutes at 37°C. Reaction is quenched with 5-7 mM EDTA prior to purification by SEC. Proteins were analysed using a Thermo Scientific Exactive mass spectrometer (**Section 7.8.2.3**).

eRF1$_{ND}$-OH was then successfully purified from JMJD4 and the reaction mixture by injecting the sample onto a Superdex 75 300 mL GF column (pre-equilibrated with 20 mM Tris, 200 mM NaCl, pH 6.8), which enabled efficient separation of the proteins as previously described (**Figure 5.7** and **Figure 5.12**). Note that prior to injecting the sample onto the GF column, residual protein precipitate was removed from the reaction mixture by carefully collecting the supernatant following centrifugation (see **Section 7.7.2.4** for experimental details). Ultimately, this optimised method allowed for successful isolation and purification of eRF1$_{ND}$-OH.

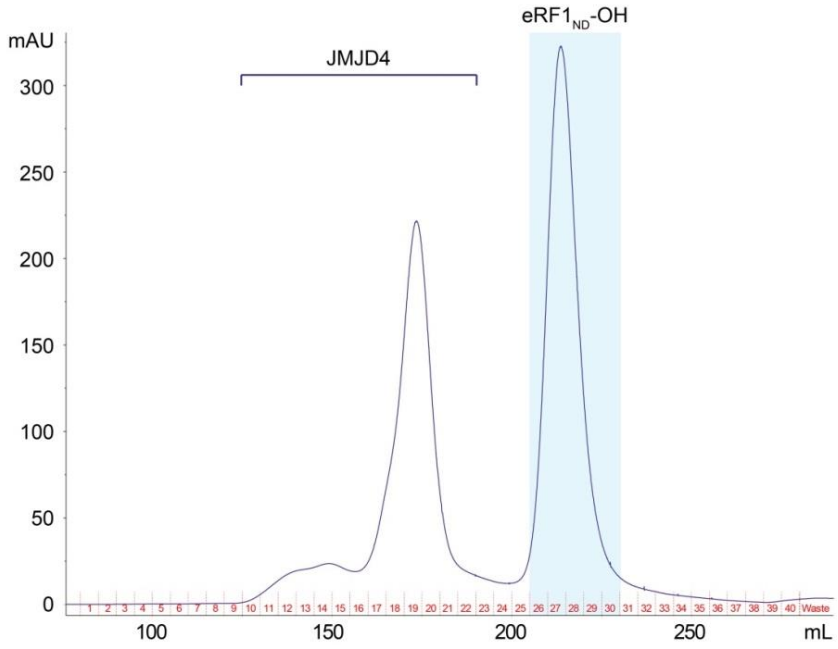


Figure 5.12. Representative UV trace following separation of eRF1$_{ND}$-OH from JMJD4 using an S75 300 mL GF column. The representative trace arises from the purification of $^{15}N/^{13}C$-labelled eRF1$_{ND}$-OH from JMJD4.

5.3.2 Isolation of eRF1$_{ND}$-30

While the hydroxylated product appeared to be kinetically favoured, formation of eRF1$_{ND}$-30 appeared to be dependent on initial hydroxylation under non-limiting conditions. Near complete formation of eRF1$_{ND}$-30 was observed after overnight (16-20 h) incubation at 4°C with shaking (80 μM eRF1$_{ND}$, 80 μM JMJD4, 1 mM Fe(II) and 2 mM 2OG in 20 mM Tris, pH 7.5; **Figure 5.13**). Note that likely due to the high protein content in the reaction mixture, some precipitation was observed. Following overnight incubation, the reaction was quenched with 5 – 7 mM EDTA, and eRF1$_{ND}$-30 was purified away from JMJD4 and the reaction mixture as described for eRF1$_{ND}$-OH (**Section 5.3.1**; **Figure 5.14**). These described modification and purification conditions enabled for the successful isolation of eRF1$_{ND}$-30 (experimental conditions are described in **Section 7.7.2.5**).

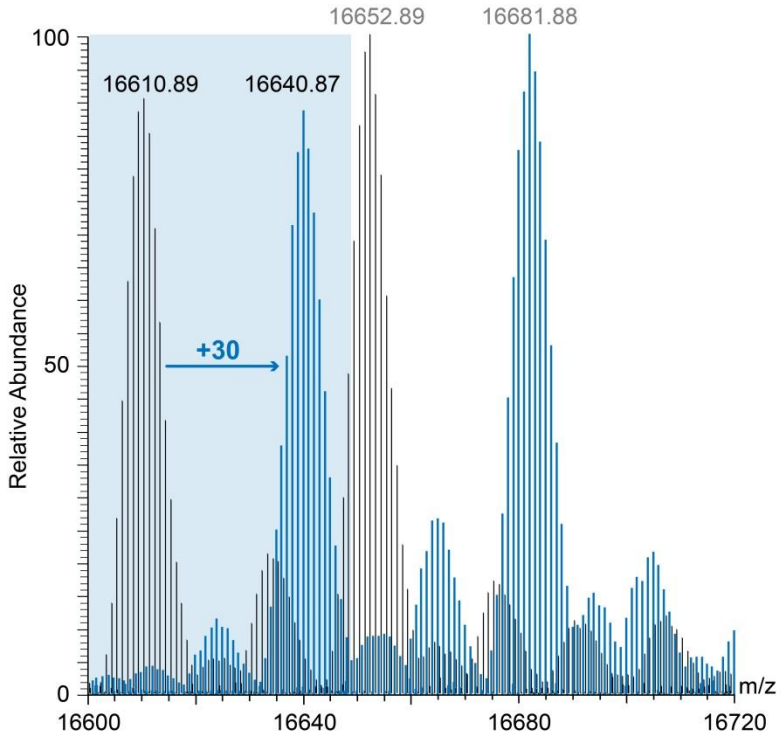


Figure 5.13. Representative deconvoluted mass spectra showing optimised conditions for eRF1$_{ND}$-30 formation. The optimised conditions for eRF1$_{ND}$-30 formation are as follows: Incubation of 80 µM eRF1$_{ND}$, 1 mM Fe(II), and 2 mM 2OG (black) with 80 µM JMJD4 (blue) overnight (16-20 h) at 4°C. Reaction was quenched with 5-7 mM EDTA prior to purification by SEC. Proteins were analysed using a Thermo Scientific Exactive mass spectrometer (**Section 7.8.2.3**).

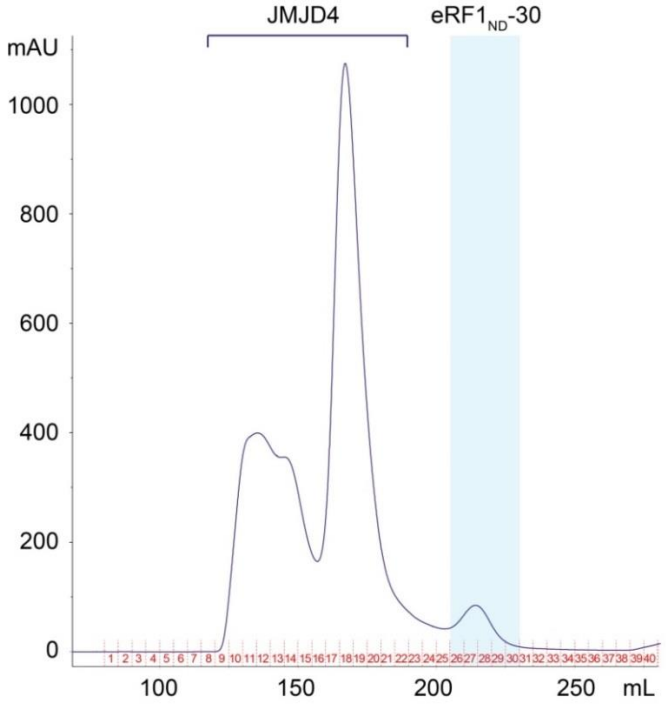


Figure 5.14. Representative UV trace following separation of eRF1$_{ND}$-30 from JMJD4 using an S75 300 mL GF column. The representative trace arises from purification of $^{15}N/^{13}C$-labelled eRF1$_{ND}$-30 from JMJD4.

Overall, the results presented in **Section 5.3** reveal optimised conditions for forming and isolating hydroxylated and +30 Da modified $eRF1_{ND}$ proteins with high purity (> 90%, **Figure 5.15**) and suitable for protein NMR and other biophysical studies.

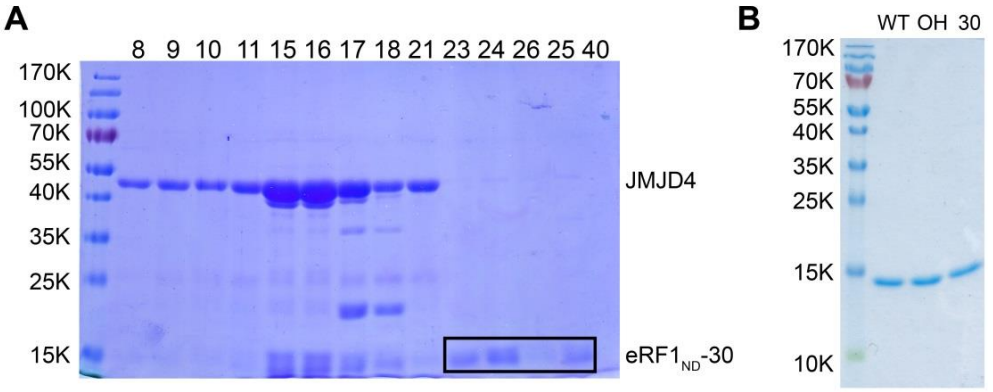


Figure 5.15. Representative Coomassie-stained 12% SDS-PAGE gels showing (A) modified $eRF1_{ND}$ separation from JMJD4 using an S75 300 mL GF column and (B) purified $eRF1_{ND}$ wildtype (WT), hydroxylated (OH) and +30 Da modified (30).

5.4 The +30 Da modification is apparently localised to K63

Previous studies have shown that JMJD4 acts a C4 lysyl hydroxylase modifying K63 of eRF1[133]. However, there are no reports of the +30 Da modification, which is identified in this study. While previous MS analyses have shown that hydroxylation is localised to K63[133], the exact site of the +30 Da modification was unknown and required investigation. As described in **Sections 5.2.3** and **5.3**, isolated protein assays indicated that $eRF1_{ND}$-OH was the initially formed product; however, after hours of incubation, the hydroxylated product completely disappeared and formation of $eRF1_{ND}$-30 was observed. These observations suggested that the hydroxylated product was likely an 'intermediate' for $eRF1_{ND}$-30 formation. Therefore, it was initially hypothesised that the +30 Da modification was also localised to K63, though crosslinking with local or even distant amino acid residues was considered a possibility. To investigate the site of the +30 Da modification, LC-MS/MS and amino acid analyses were undertaken.

5.4.1 LC-MS/MS analyses of eRF1$_{ND}$

LC-MS/MS analyses of tryptic eRF1$_{ND}$, eRF1$_{ND}$-OH, and eRF1$_{ND}$-30 fragments for peptide sequencing[276] were performed using their respective purified proteins. eRF1$_{ND}$ proteins were digested with trypsin and purified using a Sep-Pak C18 cartridge as described in **Section 7.8.3**. LC-MS/MS analyses of the tryptic eRF1$_{ND}$ digests were performed using a Synapt G2-Si high definition mass spectrometer on a service basis by Mr Lingzhi Gong (Mass Spectrometry, Department of Chemistry, University of Oxford). Data were manually analysed by employing a targeted search for peptides of interest using MassLynx v4.0 (Waters). An online fragmentation ion tool (http://db.systemsbiology.net:8080/proteomicsToolkit/FragIonServlet.html; Fragment Ion Calculator) was used to generate theoretical peptide and b/y ion fragmentation patterns (**Table 5.1**).

Extracted ion chromatograms were generated for tryptic fragments corresponding to wildtype eRF1$_{ND}$ [MLADEFGTASNI**K**, $(M+2H)^{2+}$ = 698.84], eRF1$_{ND}$-OH [MLADEFGTASNI$\mathbf{K^{OH}}$, $(M+2H)^{2+}$ = 706.84] and eRF1$_{ND}$-30 [MLADEFGTASNI$\mathbf{K^{+30}}$, $(M+2H)^{2+}$ = 713.84; and MLADEFGTASNI$\mathbf{K^{+30}}$SR; $(M+2H)^{2+}$ = 835.41]. Consistent with previous reports, the MS/MS analyses revealed that the JMJD4-catalysed eRF1 hydroxylation site is K63, as evidenced by complete y ion coverage of the MLADEFGTASNI$\mathbf{K^{OH}}$ peptide fragment (**Figure 5.16**; **Table 5.1**). Interestingly, only a peptide corresponding to the mass of one missed tryptic cleavage site (MLADEFGTASNI$\mathbf{K^{+30}}$SR) could be identified in the extracted ion chromatogram for the eRF1$_{ND}$-30 sample. The fragmentation pattern of the eRF1$_{ND}$-30 tryptic peptide revealed that the +30 Da modification is localised to K63, as evidenced by the near complete y ion coverage (**Figure 5.17**; **Table 5.1**). Peptide sequencing indicated that the y1 and y2 fragment ions ($[R]^+$ and $[SR]^+$, respectively) did not contain the +30 Da modification. However, y3 ($[\mathbf{K^{+30}}SR]^+$) and all subsequent y ions contained the +30 Da modification, indicating that, like hydroxylation, this modification is localised to K63.

Notably, trypsin hydrolysis was performed on an equal amount of purified $eRF1_{ND}$, $eRF1_{ND}$-OH, and $eRF1_{ND}$-30 at the same time and under the same conditions. Therefore, the +30 Da modification apparently inhibits tryptic hydrolysis at K63 as only a +30 Da modified K63-containing peptide with one missed cleavage site could be identified in the sample. This observation further supports the assignment of K63 as the target residue. Overall, the MS/MS data are consistent with the observed formation of $eRF1_{ND}$-30 following initial hydroxylation at K63.

Table 5.1. Theoretical and observed b and y ions corresponding to wildtype, hydroxylated, and +30 Da modified $eRF1_{ND}$ K63-containing tryptic peptide fragments as generated by the Fragment Ion Calculator*.

Wildtype Sequence: MLADEFGTASNIK

Charge	Mass	
	Monoisotopic	**Average**
(M)	1395.67046	1396.57812
$(M+H)^{+}$	1396.67773	1397.58540
$(M+2H)^{2+}$	698.84253	699.29636
$(M+3H)^{3+}$	466.23079	466.533354
$(M+4H)^{4+}$	349.92492	350.15184

Seq	#	b	y	# (+1)
M	1	132.04781	1396.67773	13
L	2	245.13187	1265.63725	12
A	3	316.16898	1152.55319	11
D	4	431.19593	1081.51607	10
E	5	560.23852	966.48913	9
F	6	707.30693	837.44654	8
G	7	764.32840	690.37812	7
T	8	865.37608	633.35666	6
A	9	936.41319	532.30898	5
S	10	1023.44522	461.27187	4
N	11	1137.48814	374.23984	3
I	12	1250.57221	260.19691	2
K	13	1378.66717	147.11285	1

Hydroxylated Sequence: MLADEFGTASNIK^{OH}

Charge	Mass	
	Monoisotopic	**Average**
(M)	1411.66046	1412.56812
$(M+H)^{+}$	1412.66773	1413.57540
$(M+2H)^{2+}$	706.83753	707.29136
$(M+3H)^{3+}$	471.56079	471.86335
$(M+4H)^{4+}$	353.92242	354.14934

Seq	#	b	y	# (+1)
M	1	132.04781	1412.66773	13
L	2	245.13187	1281.62725	12
A	3	316.16898	1168.54319	11
D	4	431.19593	1097.50607	10
E	5	560.23852	982.47913	9
F	6	707.30693	853.43654	8
G	7	764.32840	706.36812	7
T	8	865.37608	649.34666	6
A	9	936.41319	548.29898	5
S	10	1023.44522	477.26187	4
N	11	1137.48814	390.22984	3
I	12	1250.57221	276.18691	2
K	13	1394.65717	163.10285	1

+30 Da Sequence: MLADEFGTASNIK^{+30}**

Charge	Mass	
	Monoisotopic	Average
(M)	1425.67046	1426.57812
$(M+H)^{+}$	1426.67773	1427.58540
$(M+2H)^{2+}$	713.84253	714.29636
$(M+3H)^{3+}$	476.23079	476.53335
$(M+4H)^{4+}$	357.42492	357.65184

Seq	#	b	y	# (+1)
M	1	132.04781	1426.67773	13
L	2	245.13187	1295.63725	12
A	3	316.16898	1182.55319	11
D	4	431.19593	1111.51607	10
E	5	560.23852	996.48913	9
F	6	707.30693	867.44654	8
G	7	764.32840	720.37812	7
T	8	865.37608	663.35666	6
A	9	936.41319	562.30898	5
S	10	1023.44522	491.27187	4
N	11	1137.48814	404.23984	3
I	12	1250.57221	290.19691	2
K	13	1408.66717	177.11285	1

+30 Da Sequence: MLADEFGTASNIK^{+30}SR (1 missed cleavage)

Charge	Mass	
	Monoisotopic	Average
(M)	1668.80360	1669.84380
$(M+H)^{+}$	1669.81087	1670.85108
$(M+2H)^{2+}$	835.40910	835.92920
$(M+3H)^{3+}$	557.27517	557.62191
$(M+4H)^{4+}$	418.20821	418.46826

Seq	#	b	y	# (+1)
M	1	132.04781	1669.81087	15
L	2	245.13187	1538.77039	14
A	3	316.16898	1425.68632	13
D	4	431.19593	1354.64921	12
E	5	560.23852	1239.62227	11
F	6	707.30693	1110.57968	10
G	7	764.32840	963.51126	9
T	8	865.37608	906.48980	8
A	9	936.41319	805.44212	7
S	10	1023.44522	734.40501	6
N	11	1137.48814	647.37298	5
I	12	1250.57221	533.33005	4
K	13	1408.66717	420.24599	3
S	14	1495.69920	262.15102	2
R	15	1651.80031	175.11900	1

Blue/red highlighted ions were observed at > 10% of maximum ion intensity. Highlighted ions were observed between 5-10% of maximum peak intensity.

* Theoretical fragmentation patterns generated using: http://db.systemsbiology.net:8080/proteomicsToolkit/FragIonServlet.html (Fragment Ion Calculator)

** No peak was observed in the extracted ion chromatogram corresponding to the mass of this peptide.

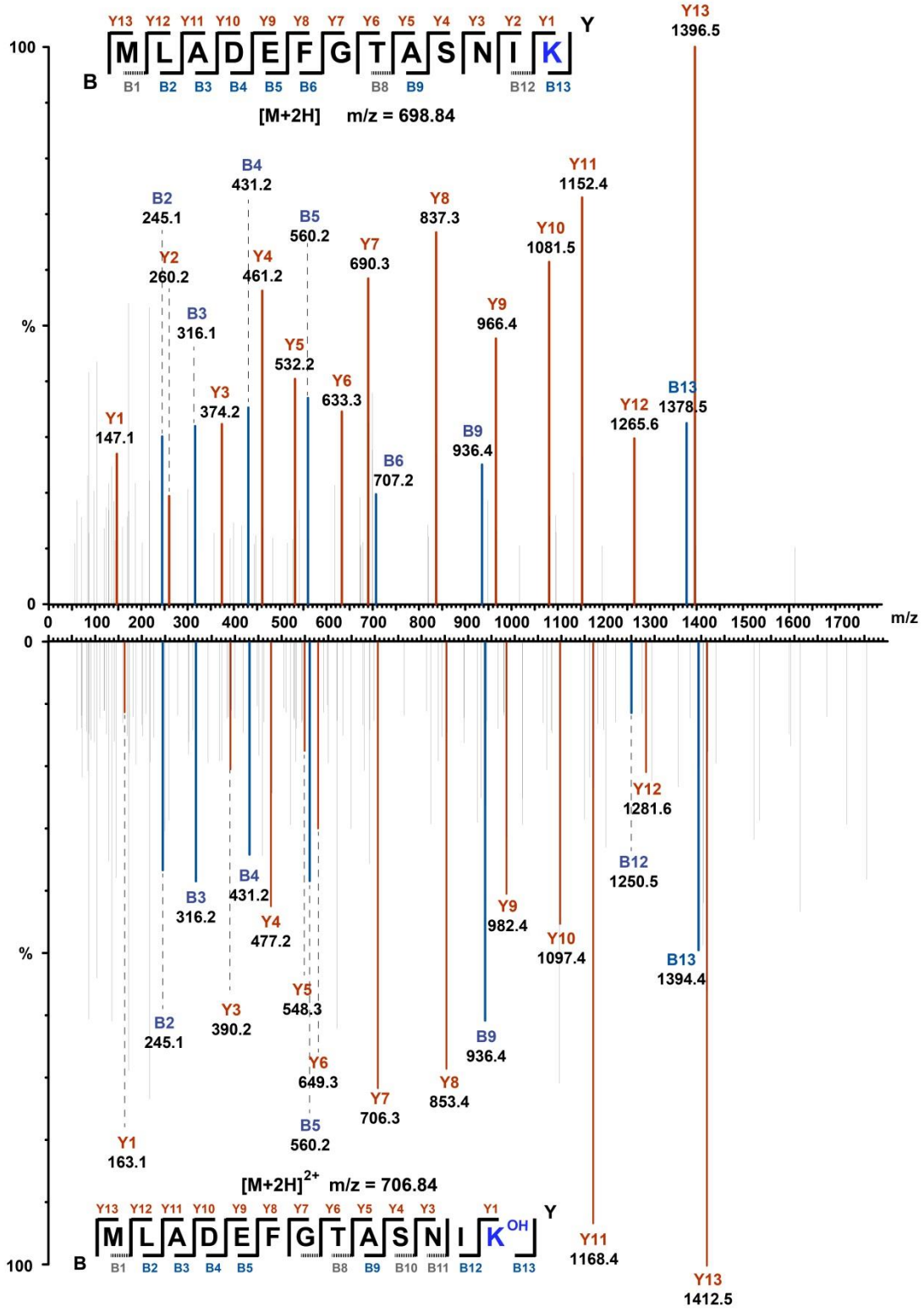


Figure 5.16. Deconvoluted MS/MS spectra of K63-containing tryptic fragments arising from (top) $eRF1_{ND}$ [MLADEFGTASNIK; m/z = 698.84] and (bottom) $eRF1_{ND}$-OH [$MLADEFGTASNIK^{OH}$; m/z = 706.84]. Note that ions below 10% were removed from the spectra for clarity. Ions shown in grey were observed between 5-10% of the maximum ion intensity. Spectra were deconvoluted using MaxEnt3 Mass Lynx V4.1 (Waters) deconvolution software.

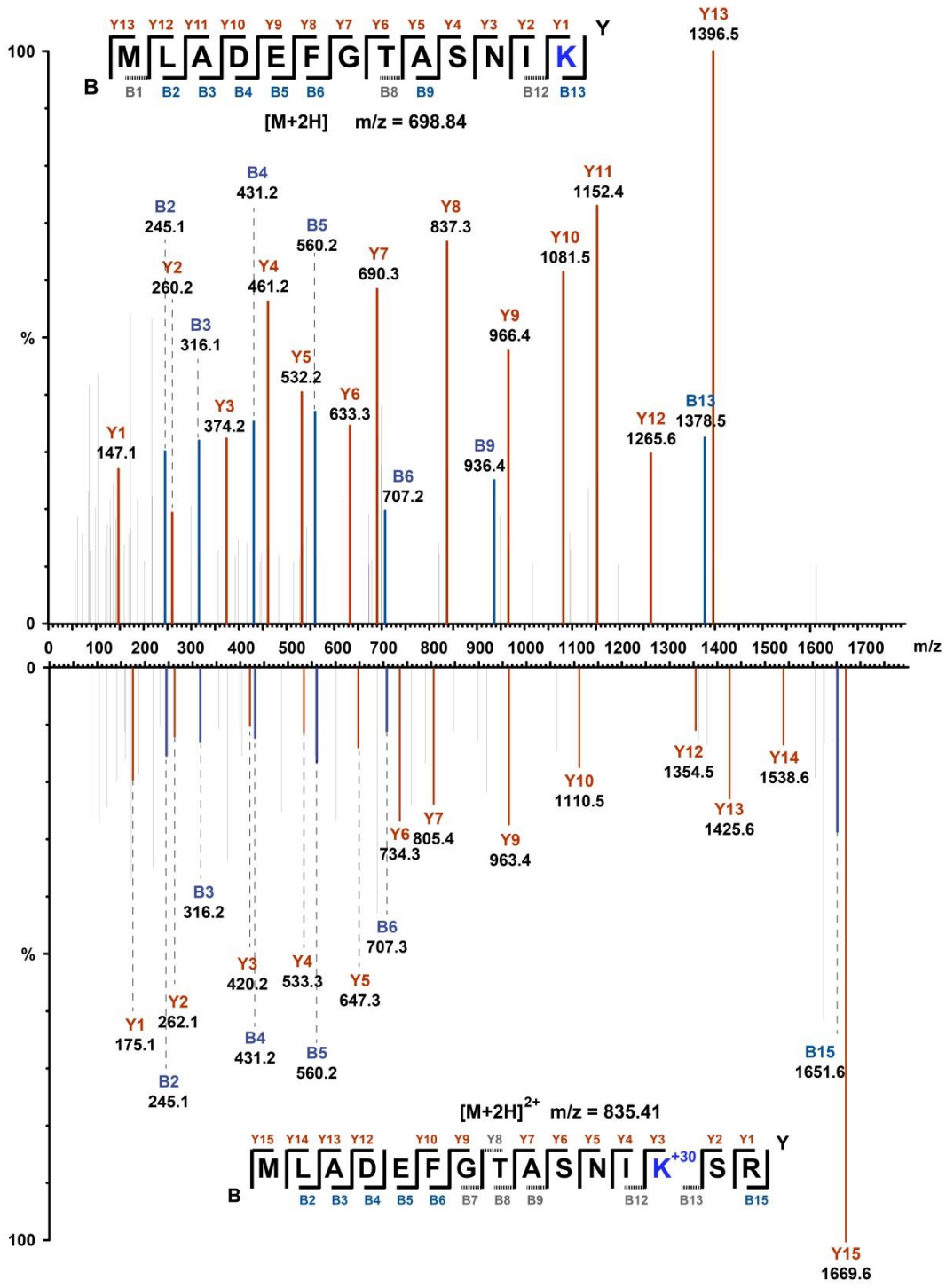


Figure 5.17. Deconvoluted MS/MS spectra of K63-containing tryptic fragments arising from (top) $eRF1_{ND}$ [MLADEFGTASNIK; m/z = 698.84] and (bottom) $eRF1_{ND}$-30 [$MLADEFGTASNIK^{+30}SR$; m/z = 835.41]. Note that ions below 10% were removed from the spectra for clarity. Ions shown in grey were observed between 5-10% of the maximum ion intensity. No K63-containing tryptic fragment arising from $eRF1_{ND}$-30 with no missed cleavage [$MLADEFGTASNIK^{+30}$; m/z = 713.84] was observed in the extracted ion chromatogram. Spectra were deconvoluted using MaxEnt3 Mass Lynx V4.1 (Waters) deconvolution software.

5.4.2 Amino acid analysis of eRF1$_{ND}$

Amino acid analyses (described in **Section 2.3.4.1**) of purified eRF1$_{ND}$, eRF1$_{ND}$-OH, and eRF1$_{ND}$-30 were undertaken to validate the localisation of the +30 Da modification and gain potential insights into the identity of the modification. Purified eRF1$_{ND}$, eRF1$_{ND}$-OH, and eRF1$_{ND}$-30 were hydrolysed using a non-specific protease from *Streptomyces griseus* and analysed by UPLC-MS following derivatisation with 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate as described in **Section 7.8.5**[277]. A comparison of the extracted ion chromatograms (m/z = 503.2, corresponding to the mass of AQC-derivatised hydroxylysine) of a C4 hydroxylysine standard[133] [a mixture of diastereomers – (2*S*,4*R*)/(2*R*,4*S*)/(2*S*,4*S*)/(2*R*,4*R*)] with eRF1$_{ND}$ and eRF1$_{ND}$-OH indicates the presence of C4 hydroxylysine in the eRF1$_{ND}$-OH sample (**Figure 5.18**). No evidence for C4 hydroxylysine is observed in the eRF1$_{ND}$ (control) sample. Interestingly, there is also no evidence for C4 hydroxylysine in the eRF1$_{ND}$-30 sample, consistent with LC-MS/MS analyses, which indicated that the +30 Da modification is localised to the K63 residue (**Section 5.4.1**; **Figure 5.17**). Furthermore, these data are consistent with isolated protein assays that suggest that +30 Da formation relies on initial hydroxylation, as evidenced by the disappearance of the hydroxylated peak and near complete conversion to the +30 Da state (**Sections 5.2.3 and 5.3**).

Notably, a new peak corresponding to the mass of +30 Da modified AQC-derivatised lysine (m/z = 517.5) was not observed in the eRF1$_{ND}$-30 sample (nor any other sample). It is possible that, as observed for trypsin, the +30 Da modification inhibited cleavage by the nonspecific protease from *S. griseus* at the K63 site. It was not possible to generate the extracted ion chromatogram for the mass corresponding to AQC-derivatised $K^{+30}S_{63\text{-}64}$ (theoretical m/z = 604.2) peptide fragment as the theoretical m/z value was greater than the acquisition window (100-600 Da).

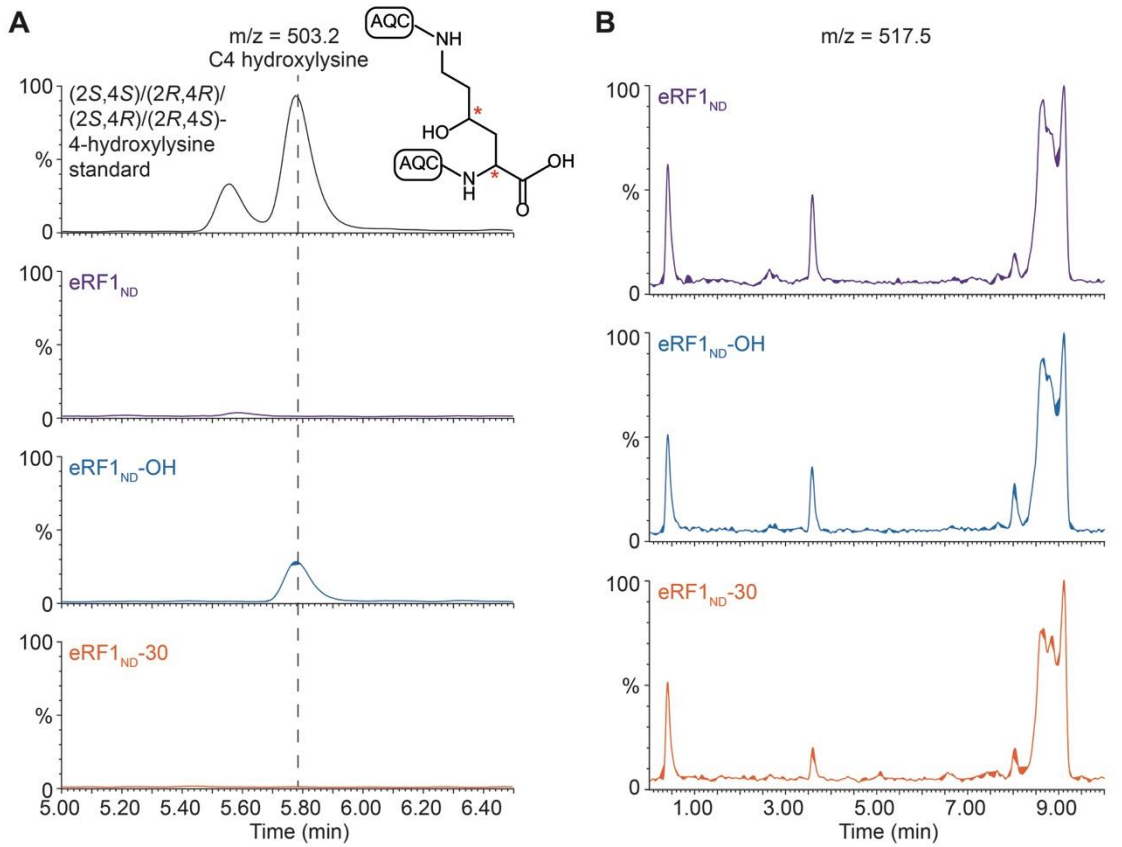


Figure 5.18. Amino acid analyses of JMJD4-catalysed eRF1$_{ND}$ modifications. (A) Extracted ion chromatograms corresponding to the mass of 4-hydroxylysine (m/z =503.2) for a C4 hydroxylysine standard (black), eRF1$_{ND}$ (purple), eRF1$_{ND}$-OH (blue), and eRF1$_{ND}$-30 (orange). 4-hydroxylysine is present only in eRF1$_{ND}$-OH, but notably not present in eRF1$_{ND}$-30. Note that the two peaks observed in the C4 hydroxylysine standard represent pairs of enantiomers. The structure of the AQC-derivatised C4 hydroxylysine standard is shown, with stereocentres highlighted with a red star. ***(B)*** Extracted ion chromatograms corresponding to the mass of +30 Da modified lysine (m/z =517.5) for eRF1$_{ND}$ (purple), eRF1$_{ND}$-OH (blue), and eRF1$_{ND}$-30 (orange). Note that the appearance of a new peak corresponding to +30 Da modified lysine was not observed in the eRF1$_{ND}$-30 sample.

5.5 Investigation of JMJD4-catalysed +30 Da modification of full-length eRF1

The experiments described in **Section 5.5 were performed by Owain Jones, under my supervision and guidance, and appear in a modified form in his Biochemistry Part II thesis*[278].

To investigate whether the +30 Da modification is an artefact arising from the eRF1 *N*-terminal domain construct, full-length eRF1 (eRF1$_{FL}$), a more physiologically relevant construct, was recombinantly produced and tested for JMJD4-catalysed +30 Da modification.

5.5.1 Expression and purification of eRF1$_{FL}$

eRF1$_{FL}$-His$_6$ (eRF1$_{FL}$) was produced in *E. coli* BL21(DE3) and subsequently purified by Ni(II) affinity chromatography and SEC as described in **Sections 7.3 and 7.4** (**Figure 5.19**).

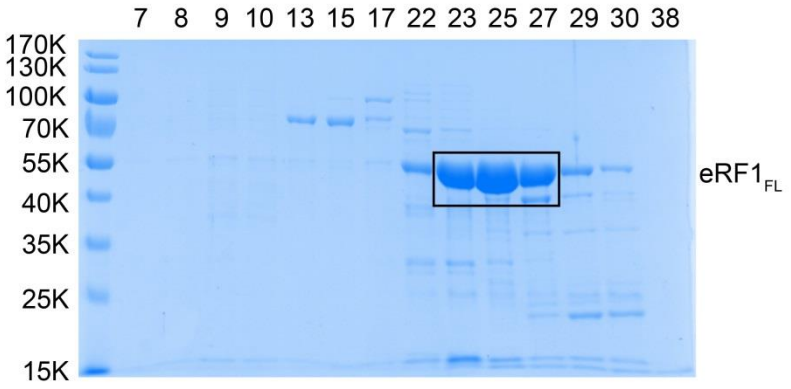


Figure 5.19. Representative Coomassie-stained 12% SDS-PAGE gel showing eRF1$_{FL}$ purification following SEC using an S200 300 mL GF column. Figure adapted from Jones, O. (2016)[278].

5.5.2 LC-MS analyses of JMJD4-catalysed +30 Da modification of eRF1$_{FL}$

Recombinant eRF1$_{FL}$ was incubated with JMJD4 under the conditions optimised for +30 Da formation (**Section 5.3.2**). Analysis by intact protein LC-MS (**Section 7.8.2.1**) suggested that eRF1$_{FL}$ likely undergoes JMJD4-catalysed +30 Da modification. As shown in **Figure 5.20**, the JMJD4-treated sample underwent a +25 Da shift as compared to the no enzyme control. This observed +25 Da species likely represents a mixture of hydroxylated (+16 Da) and +30 Da modified eRF1$_{FL}$ states and likely arises from the combined inability of the mass spectrometer (Waters Xevo G2-S QToF) and MaxEnt1 deconvolution software (Mass Lynx V4.1) to resolve 14 Da when analysing a 50 kDa protein (eRF1$_{FL}$-OH, 49,588.6 Da; eRF1$_{FL}$-30, 49,618.6 Da). While it is likely that this +25 Da species represented a combination of eRF1$_{FL}$-OH and eRF1$_{FL}$-30, LC-MS/MS analyses were required to confidently confirm the presence of the +30 Da modification in the JMJD4-treated eRF1$_{FL}$ sample.

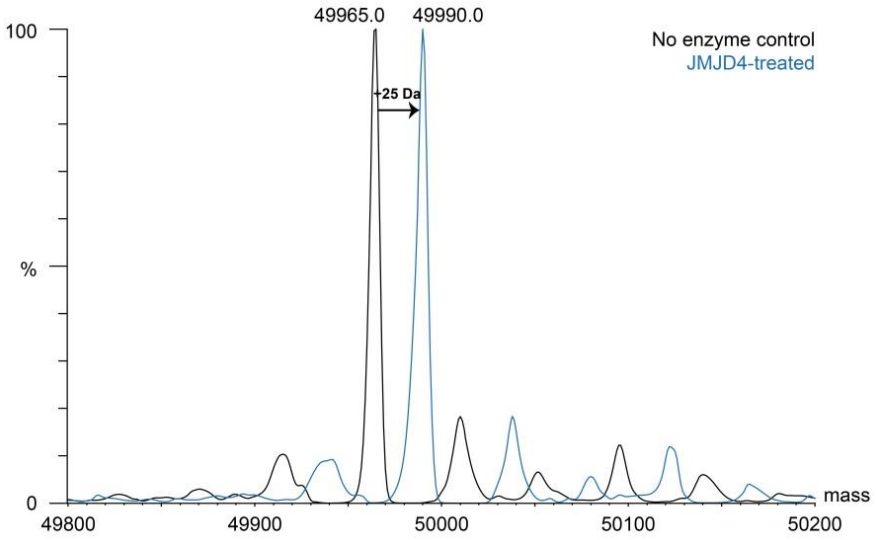


Figure 5.20. Deconvoluted intact protein mass spectra of eRF1$_{FL}$ following incubation in the absence (black) or presence (blue) of JMJD4 under the optimised +30 Da conditions. Proteins were analysed using a Xevo GS-2 QTof mass spectrometer (**Section 7.8.2.2**). Figure adapted from Jones, O. (2016)[278].

5.5.3 LC-MS/MS analyses of JMJD4-catalysed +30 Da modification of eRF1$_{FL}$

To confirm the presence of the +30 Da modification on eRF1$_{FL}$, the JMJD4-treated eRF1$_{FL}$ samples described in **Section 5.5.2** were digested with trypsin and purified using Sep-Pak C18 cartridges (**Section 7.8.3**) for subsequent LC-MS/MS analyses (**Section 7.8.3.4**). A targeted MS/MS search was manually performed using MassLynx V4.1 as described in **Section 5.4.1**. Using the theoretical m/z values for expected eRF1$_{FL}$ tryptic peptide b and y ion fragments (**Table 5.1**), extracted ion chromatograms were generated for tryptic fragments for wildtype eRF1$_{FL}$ (MLADEFGTASNI**K**, $(M+2H)^{2+}$ = 698.8), eRF1$_{FL}$-OH (MLADEFGTASNI**K^{OH}**, $(M+2H)^{2+}$ = 706.8) and eRF1$_{FL}$-30 (MLADEFGTASNI**K^{+30}**, $(M+2H)^{2+}$ = 713.8; and MLADEFGTASNI**K^{+30}**SR; $(M+2H)^{2+}$ = 835.4) in the JMJD4-treated sample.

Consistent with LC-MS/MS studies on purified eRF1$_{ND}$-30 (**Figure 5.17**), the same tryptic fragment containing the +30 Da modification (MLADEFGTASNI**K$^{(+30)}$**SR) was observed to have one missed tryptic cleavage site in the JMJD4-treated eRF1$_{FL}$ sample. The presence and identity of the unmodified, hydroxylated, and +30 Da modified tryptic eRF1$_{FL}$ peptide

fragments were further validated by the presence of theoretical b and y ions, which were observed in the product ion MS/MS spectra arising from respective extracted ion chromatograms (**Figure 5.21**). Note that no peptide corresponding to the +30 Da modified peptide was observed in the control sample.

Overall, the results indicated that $eRF1_{FL}$ can undergo JMJD4-catalysed +30 Da modification, suggesting that the +30 Da modification is not an artefact arising from the use of a truncated eRF1 domain. Furthermore, the evidence that $eRF1_{FL}$ can be +30 Da modified suggests that the modification is more likely to be a physiologically-relevant PTM.

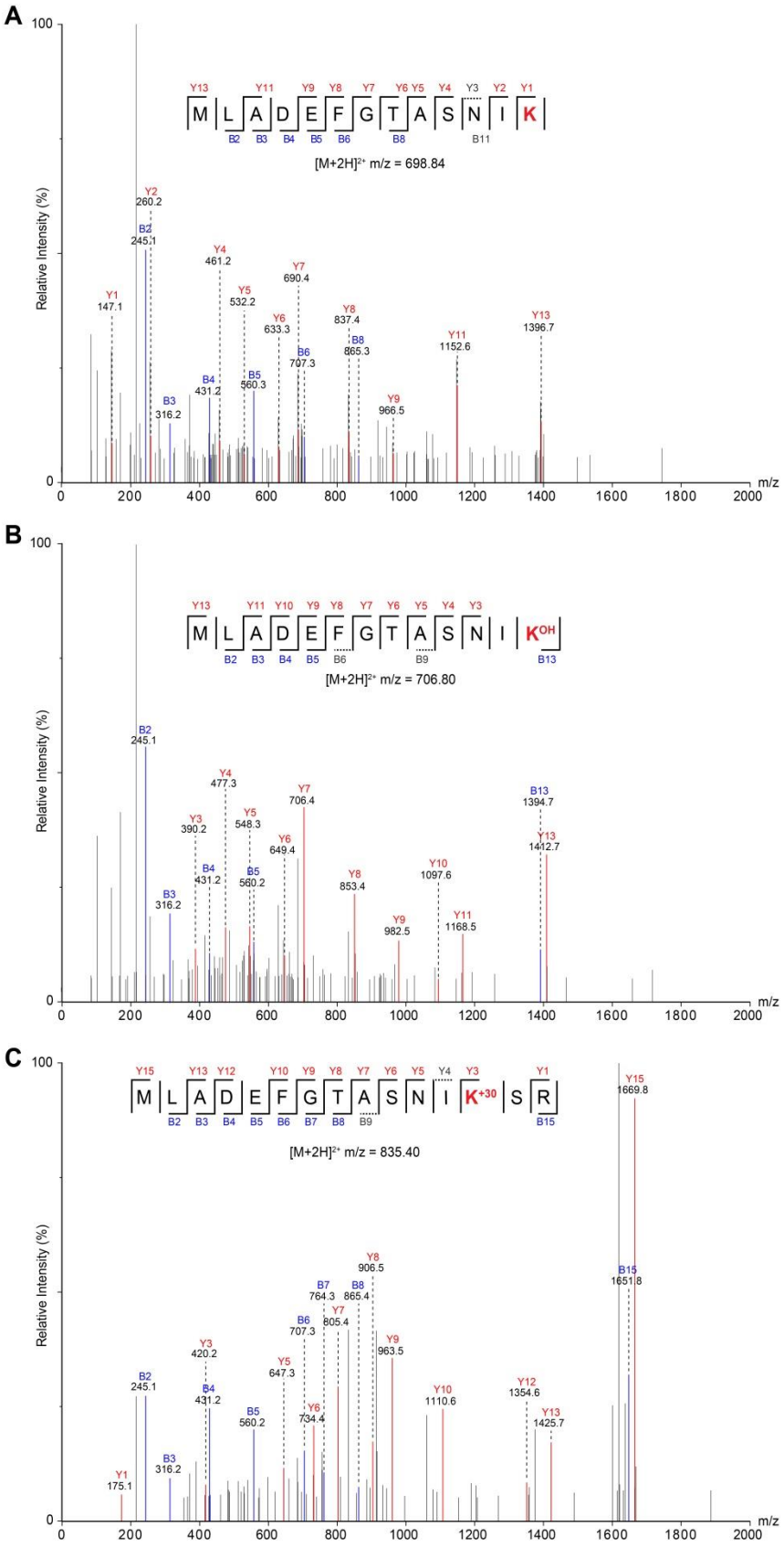


Figure 5.21. Deconvoluted MS/MS spectra corresponding to K63-containing tryptic fragments of (A) unmodified, (B) hydroxylated, and (C) +30 Da modified eRF1$_{FL}$ forms present in the JMJD4-treated sample (Figure 5.20) following trypsin digestion. Digested proteins were analysed using a Waters Xevo GS-2 QTof mass spectrometer (**Section 7.8.3.4**). Figure adapted from Jones, O. (2016)[278].

5.6 Investigation of the effects of hydroxylation and +30 Da modification on eRF1$_{ND}$ structure, stability, and protein dynamics

Given the likely biological relevance of the previously unreported +30 Da modification, biophysical analyses comparing the effect of hydroxylation and the +30 Da modification on eRF1$_{ND}$ structure, stability, and protein dynamics were undertaken to gain insights into the potential effect of these modifications. As yet, hydroxylation has been reported to increase translation termination efficiency[133]; however, the exact mechanism by which hydroxylation affects termination remains to be understood.

5.6.1 Investigation of the effects of hydroxylation and +30 Da modification on eRF1$_{ND}$ secondary structure

To investigate the effect of hydroxylation and +30 Da modification on eRF1$_{ND}$ secondary structure, CD analyses[217,218] (described in **Section 3.3.4**) were employed. Optimal protein concentration was first determined to minimise signal-to-noise by measuring CD spectra of eRF1$_{ND}$ in the far UV region at the following concentrations – 0.5, 0.4, 0.3, and 0.25 mg/mL; 0.25 mg/mL was determined to yield the best quality CD spectrum.

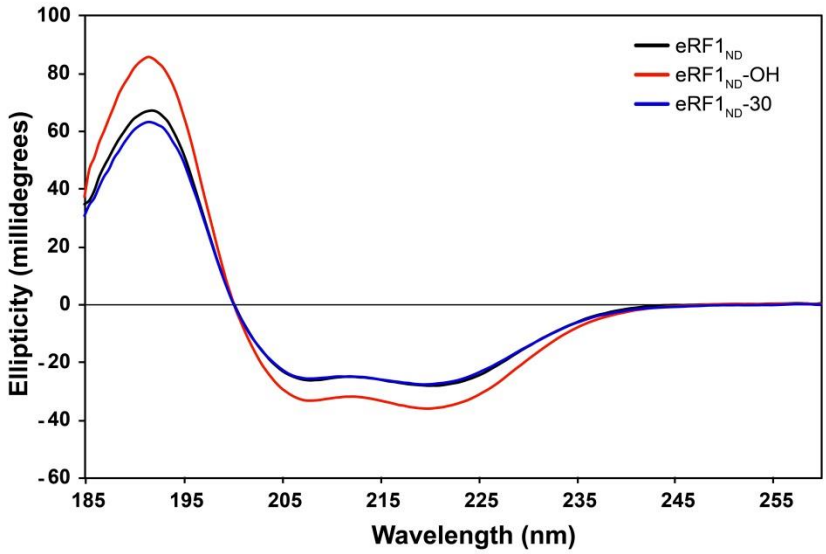


Figure 5.22. Raw CD spectra of eRF1$_{ND}$ (black), eRF1$_{ND}$-OH (red), and eRF1$_{ND}$-30 (blue). Spectra were averaged (n = 3) and smoothed (n = 8). Note that intensity differences (for eRF1$_{ND}$-OH) likely arise from small differences in protein concentration.

A comparison of the CD spectra of $eRF1_{ND}$, $eRF1_{ND}$-OH, and $eRF1_{ND}$-30 at 0.25 mg/mL revealed minimal differences (**Figure 5.22**), suggesting that these modifications have no substantial effect on $eRF1_{ND}$ secondary structure elements. Note that the differences observed for $eRF1_{ND}$-OH likely arise from differences in protein concentration.

5.6.2 Investigation of the effect of hydroxylation and +30 Da modification on $eRF1_{ND}$ stability

DSF[244,279] (described in **Section 4.4.3**) was used to study the effect of hydroxylation and the +30 Da modification on $eRF1_{ND}$ stability. Optimal protein concentration and buffer conditions for $eRF1_{ND}$ DSF studies were first determined; DSF analyses of 2, 4, and 6 μM $eRF1_{ND}$ in 50 mM HEPES, pH 7.5 buffer or 50 mM Tris, pH 7.5 revealed that 6 μM protein concentration in 50 mM HEPES, pH 7.5 gave optimal assay performance. Apparent T_m values for $eRF1_{ND}$, $eRF1_{ND}$-OH, and $eRF1_{ND}$-30 were subsequently determined with 6 μM protein concentration in 50 mM HEPES, pH 7.5 by DSF with SYPRO Orange. As shown in **Figure 5.23**, hydroxylation has little to no effect on $eRF1_{ND}$ stability, as evidenced by the minor differences between apparent T_m values for $eRF1_{ND}$ (59.94 ± 0.47°C) and $eRF1_{ND}$-OH (58.14 ± 1.47°C). However, the +30 Da modification has an apparent destabilising effect (apparent T_m = 51.84 ± 0.95°C).

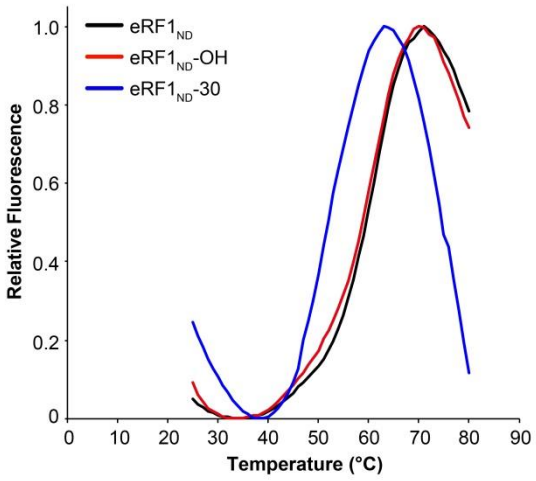


Figure 5.23. DSF analyses of $eRF1_{ND}$ (black), $eRF1_{ND}$-OH (red), and $eRF1_{ND}$-30 (blue) indicate that the +30 Da modification has a destabilising effect on $eRF1_{ND}$. Apparent T_m values: $eRF1_{ND}$, 59.94 ± 0.47°C; $eRF1_{ND}$-OH, 58.14 ± 1.47°C; and $eRF1_{ND}$-30, 51.84 ± 0.95°C. Data represent mean ± s.d. (n = 3).

5.6.3 Investigation of JMJD4-catalysed eRF1$_{ND}$ modifications by protein NMR spectroscopy

** The protein NMR experiments (HSQC, triple resonance, dynamics, and rRNA binding), backbone assignments, and related analyses described in **Section 5.6.3** were performed by Dr Ivanhoe Leung (University of Oxford, currently at the University of Auckland).*

5.6.3.1 Protein NMR spectroscopy

Protein NMR spectroscopy is a biophysical technique that is used to study protein structure and dynamics in solution[179]. As described in **Section 5.2.1**, proteins require ^{15}N and/or ^{13}C isotopic labelling to minimise resonance overlap and enable chemical shift assignments of each peak to the corresponding atom giving rise to the observed resonance [note that for very large proteins (>250 residues) ^{2}H-labelling may also be required[272,273]]. Without the assignment of each atom in the protein backbone, the in-solution protein structure cannot be determined. Thus, protein backbone assignment is typically achieved using a combination of heteronuclear NMR experiments.

Standard 2D ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) experiments, which monitor the one-bond coupling correlation between H and N, are amongst the most common ways to examine protein structure by NMR spectroscopy[272,280]. Because each amino acid residue (except proline) has an amide proton attached to a nitrogen in the peptide bond, an ^{1}H-^{15}N HSQC spectrum shows a signal for almost every amide bond in the protein backbone. Thus, in theory, each peak observed in a protein ^{1}H-^{15}N HSQC spectrum represents one amino acid residue. Exceptions to this include proline, where no amide bond is present, and the *N*-terminal residue, which contains a free amino group that undergoes exchange with the deuterated solvent. Other exceptions include side chains that have N-H bonds (e.g. Trp, Asn, Gln, and Arg), where more than one peak is typically observed per amino acid. Because backbone amide chemical shifts are sensitive to the local chemical environment and overall secondary/tertiary structure, they enable the study of protein structure and dynamics.

To study larger proteins (10-25 kDa) and conduct sequential protein backbone assignment, standard 3D ^{1}H,^{13}C,^{15}N triple resonance experiments are typically performed[272,280]. 3D triple resonance experiments employ magnetisation transfers through multiple, one-bond magnetisation steps, involving the nuclei (typically) named in the experiment (note that chemical shifts are not observed for nuclei shown in brackets). The most common 3D experiments include [HN(CO)CA, HNCA], [HNCO, HN(CA)CO], and [HN(CO)CACB, HNCACB][280]; these experiments are typically performed and analysed in pairs (brackets) with one experiment revealing inter-residue (*i* – *i-1*) correlations and the other showing both inter- (*i* – *i-1*) and intra- (*i* – *i*) residue correlations (**Figure 5.24**). A sequential assignment of the observed ^{1}H, ^{13}C, and ^{15}N resonances with their respective amino acids can thus be conducted using these experiments. Note that the $^{13}C_{\beta}$ chemical shifts (obtained from HN(CO)CACB and HNCACB experiments) can further enable backbone assignment by providing diagnostic chemical shifts for amino acids like serine, threonine, and alanine.

Because the solution NMR structure of the *N*-domain of eRF1 has been previously described and well-studied[263,270], the backbone resonance assignment of $eRF1_{ND}$ could theoretically be simplified by using ^{15}N-labelled $eRF1_{ND}$, measuring ^{1}H-^{15}N HSQC spectra, and employing a 'nearest-neighbour' resonance assignment approach. This method provides a simpler, more time efficient, and cheaper alternative than using $^{15}N/^{15}C$-labelled protein (cost of D-glucose-$^{13}C_6$ is ~$200 per gram), acquiring triple resonance experiments (3-6 days of acquisition time depending on the protein sample concentration), and employing sequential backbone resonance assignment analyses as described above. Thus, preliminary protein NMR studies were first performed using ^{15}N-labelled protein.

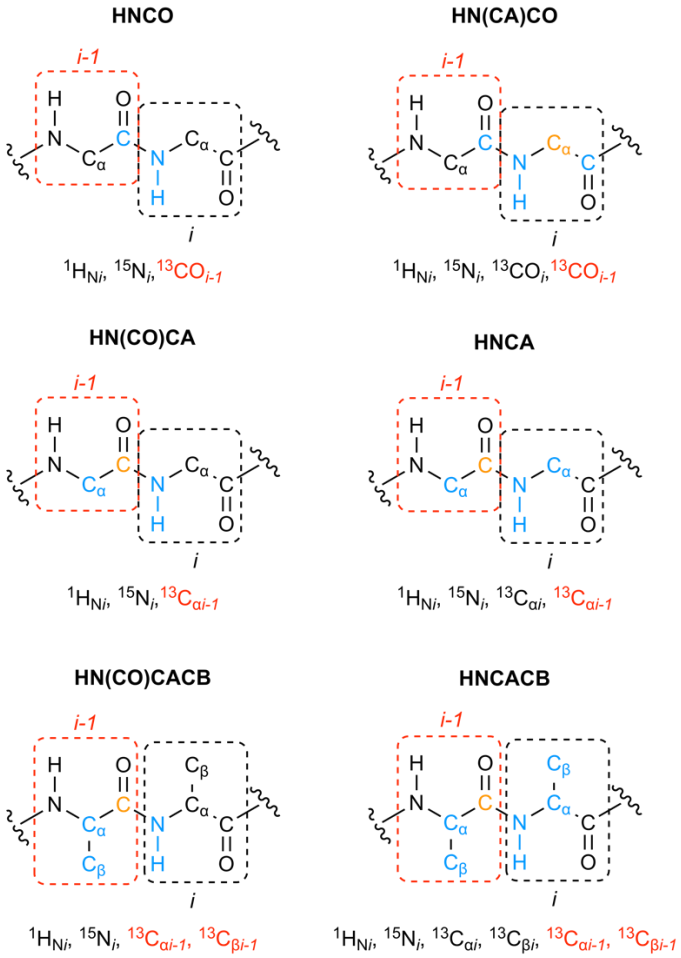


Figure 5.24. Triple resonance experiments used for sequential assignment of protein backbone resonances. Magnetisation transfer pathways are shown in blue. Chemical shifts for atoms shown in orange are not observed in resultant 3D spectra. Figure adapted from http://www.protein-nmr.org.uk/.

5.6.3.2 Expression of ^{15}N and $^{15}N/^{13}C$-labelled $eRF1_{ND}$

^{15}N and $^{15}N/^{13}C$-labelled $eRF1_{ND}$ were produced in Rosetta(DE3) pLysS cells grown in M9 minimal media using the expression conditions described in **Section 7.3.5**. ^{15}N and $^{15}N/^{13}C$-labelled $eRF1_{ND}$-OH and $eRF1_{ND}$-30 were produced and purified using the optimised assay and purification conditions described in **Sections 5.3.1** and **5.3.2**, respectively. Expression and purification conditions were modified from Polshakov, V. I. *et al.* (2012)[270]. Adequate yields and successful purification of each construct enabled the study of the proteins by 2D and 3D NMR spectroscopy.

5.6.3.3 ^{1}H-^{15}N HSQC spectra suggest structural differences between wild type and JMJD4-catalysed $eRF1_{ND}$-30

^{1}H-^{15}N HSQC experiments are usually the first heteronuclear experiments performed on protein samples to determine the quality of the protein sample and whether additional labelling (i.e. ^{13}C) and/or triple resonance NMR experiments are required. The hallmark of a high-quality protein HSQC spectrum is dispersity, which indicates a well-structured protein and non-overlapping peaks, given that the assignment of each peak to its respective amino acid is crucial for analysis and/or structure determination[281]. Preliminary ^{1}H-^{15}N HSQC experiments were first performed on ^{15}N-labelled-$eRF1_{ND}$ to confirm that the protein was folded in solution and that the spectrum was of a suitable quality for subsequent analyses of $eRF1_{ND}$-OH and $eRF1_{ND}$-30. A ^{1}H-^{15}N HSQC spectrum was acquired for $eRF1_{ND}$ using previously reported conditions[270]. Each peak and its corresponding ^{1}H and ^{15}N chemical shifts were assigned to their respective amino acid in the $eRF1_{ND}$ backbone by comparison with published chemical shift data (BMRB Entry 18092[270]) using a 'nearest-neighbour' approach (**Figure 5.25**).

^{1}H-^{15}N HSQC experiments were then performed on ^{15}N-labelled $eRF1_{ND}$-OH and $eRF1_{ND}$-30 to investigate whether hydroxylation and/or the +30 Da modification caused substantial chemical shifts changes. ^{1}H and ^{15}N chemical shifts are reflective of the chemical environment, and thus differences can indicate local chemical, structural, and/or conformational changes. As shown in **Figure 5.25A**, hydroxylation had little-to-no effect on ^{1}H and ^{15}N chemical shifts in the HSQC spectrum, except on a few residues which likely included K63 and nearby residues, given the expected change in local chemical environment caused by the introduction of a hydroxyl group on the K63 side chain. The minimal chemical shift changes thus enabled near-complete peak assignment by comparison with $eRF1_{ND}$ and the 'nearest-neighbour' approach. Overall, the HSQC experiment suggested that hydroxylation likely did not induce changes on the structure and/or conformation of $eRF1_{ND}$.

By contrast, comparison of $eRF1_{ND}$-30 with $eRF1_{ND}$ revealed that the +30 Da modification induced several large chemical shift changes in the protein backbone. The data suggested that this previously unidentified modification likely caused a major structural and/or conformational effect on $eRF1_{ND}$ (**Figure 5.25B**). However, due to the large chemical shift changes observed, the peaks in the $eRF1_{ND}$-30 HSQC spectrum could not be assigned to the protein backbone using the 'nearest-neighbour' approach. Thus, triple resonance protein NMR experiments were required to fully assign the $eRF1_{ND}$-OH and $eRF1_{ND}$-30 backbones for subsequent structural determination.

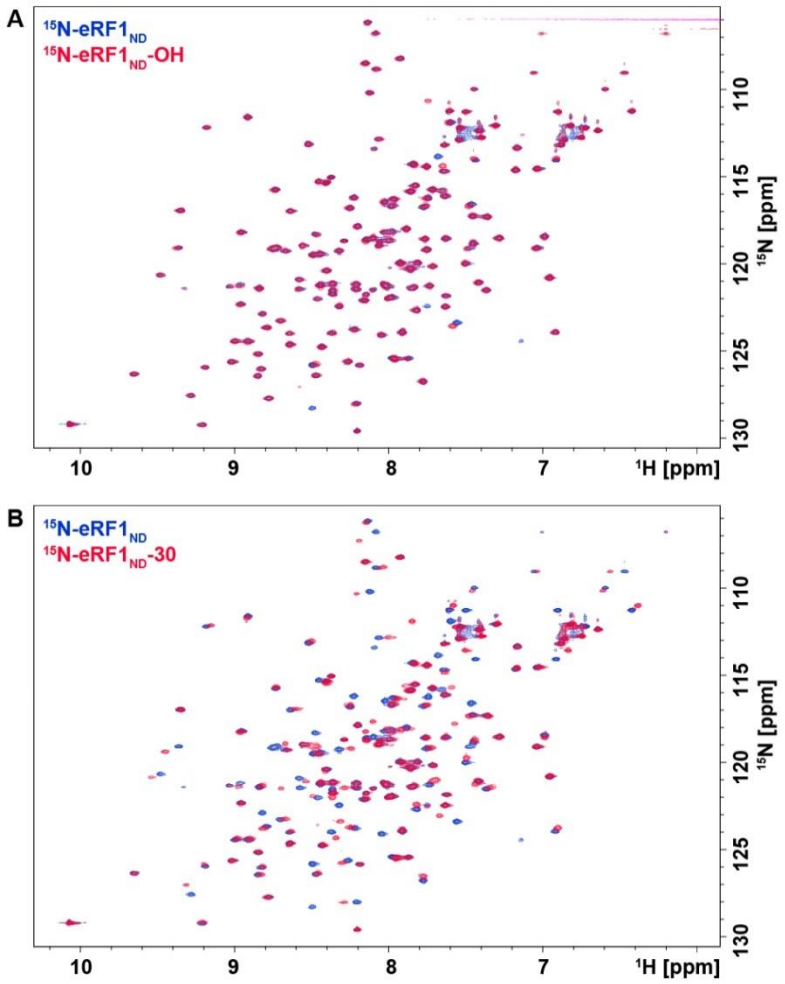


Figure 5.25. Overlaid ^{1}H-^{15}N HSQC spectra of (A) ^{15}N-$eRF1_{ND}$ (blue) and ^{15}N-$eRF1_{ND}$-OH (red) and (B) ^{15}N-$eRF1_{ND}$ (blue) and ^{15}N-$eRF1_{ND}$-30 (red). The +30 Da modification induces more global chemical shift perturbations, while only minor chemical shift changes are observed for $eRF1_{ND}$-OH, likely limited to K63-OH and neighbouring residues.

5.6.3.4 Protein backbone assignment by triple resonance experiments

Standard triple resonance experiments [HN(CO)CA, HNCA, HNCO, HN(CA)CO, HN(CO)CACB, and HNCACB (**Figure 5.24**)] were performed on $^{15}N/^{13}C$-labelled $eRF1_{ND}$-OH and $eRF1_{ND}$-30 for protein backbone assignment. While complete assignment of $eRF1_{ND}$-OH could be determined from the combined 3D NMR experiments, complete assignment of $eRF1_{ND}$-30 was not possible, particularly in the region of interest (residues surrounding K63) due to overlapping and 'missing' peaks in the spectra (**Figure 5.26; Appendix A, Table A2**).

5.6.3.5 ^{15}N and $^{13}C\alpha$ chemical shift perturbations suggest $eRF1_{ND}$-30 is structurally/conformationally distinct

Despite the incomplete assignment of $eRF1_{ND}$-30, differences in ^{15}N and $^{13}C\alpha$ chemical shift perturbations (CSPs) were analysed using reported ^{15}N and $^{13}C\alpha$ chemical shifts for wildtype $eRF1_{ND}$ (BMRB Entry 18092[270]) and assigned using chemical shifts for $eRF1_{ND}$-OH and $eRF1_{ND}$-30 from acquired triple resonance experiments. As shown in **Figure 5.26A**, significant ^{15}N CSPs in the ^{1}H-^{15}N HSQC spectra were observed between $eRF1_{ND}$ and $eRF1_{ND}$-30 (**Figure 5.25B**), particularly on the two α-helices, α2 and α3, located on either side of the modified NIKS motif. These large CSPs likely indicate differences in the structure and/or conformation of the two $eRF1_{ND}$ forms. However, ^{15}N CSPs between $eRF1_{ND}$ and $eRF1_{ND}$-OH were narrowly localised to K63 and neighbouring residues (**Figure 5.25A and Figure 5.26A**); the narrow ^{15}N CSP range suggests that these shifts likely reflect changes in the local chemical environment (i.e. introduction of a hydroxyl group on K63) and that hydroxylation induces no major structural changes.

Because $^{13}C\alpha$ chemical shifts are less sensitive to the local chemical environment, $^{13}C\alpha$ CSPs are more likely to reflect changes in overall protein structure, secondary structure, and/or conformation[282]. Consistent with the ^{15}N CSPs between $eRF1_{ND}$ and $eRF1_{ND}$-OH, the largest $^{13}C\alpha$ CSPs were observed at the K63 residue, the site of hydroxylation (**Figure 5.26B**). The

localised ^{13}Cα CSPs provided further evidence indicating that JMJD4-catalysed hydroxylation of eRF1$_{ND}$ has no effect on eRF1$_{ND}$ structure. Consistent with ^{15}N chemical shift perturbations between eRF1$_{ND}$ and eRF1$_{ND}$-30, large differences in ^{13}Cα chemical shifts were observed in a more expansive region surrounding K63 (residues 60-72), though to a lesser extent than for ^{15}N (residues 30-32, 52-79, 126-129). Despite the missing assignment of key residues, the large and expansive effect on ^{13}Cα chemical shifts was clear, indicating that the +30 Da modification likely had a large effect on eRF1$_{ND}$ structure and/or conformation.

Given the large effects observed on the residues surrounding K63, the likely target of the +30 Da modification, it was hypothesised that the +30 Da modification induced unravelling of the α-helices (α2 and α3), which are immediately *N*- and *C*-terminal to the loop from which K63 protrudes. To investigate whether these changes in ^{13}Cα chemical shifts reflected changes in secondary structure (i.e. α-helical unravelling), ^{13}Cα chemical shifts were indexed[282]. The ^{13}Cα chemical shift index (CSI) is a statistical approach that compares observed ^{13}Cα chemical shifts with a set of optimised backbone 'coil' chemical shifts for each amino acid as a quick (indirect, non-experimental) method to predict protein secondary structure[282]. While it was hypothesised that ^{13}Cα CSPs were caused by the unravelling of α2 and α3 in the eRF1$_{ND}$-30 structure, the ^{13}Cα CSI predicted that these helices maintain their α-helical nature in eRF1$_{ND}$-30 (**Appendix A2**). While this prediction is consistent with the CD studies, which revealed no observable differences in secondary structure between eRF1$_{ND}$, eRF1$_{ND}$-OH, and eRF1$_{ND}$-30 (**Section 5.6.1**; **Figure 5.22**), this prediction requires validation by direct experimental evidence (e.g. structure determination by protein NMR spectroscopy and/or crystallography).

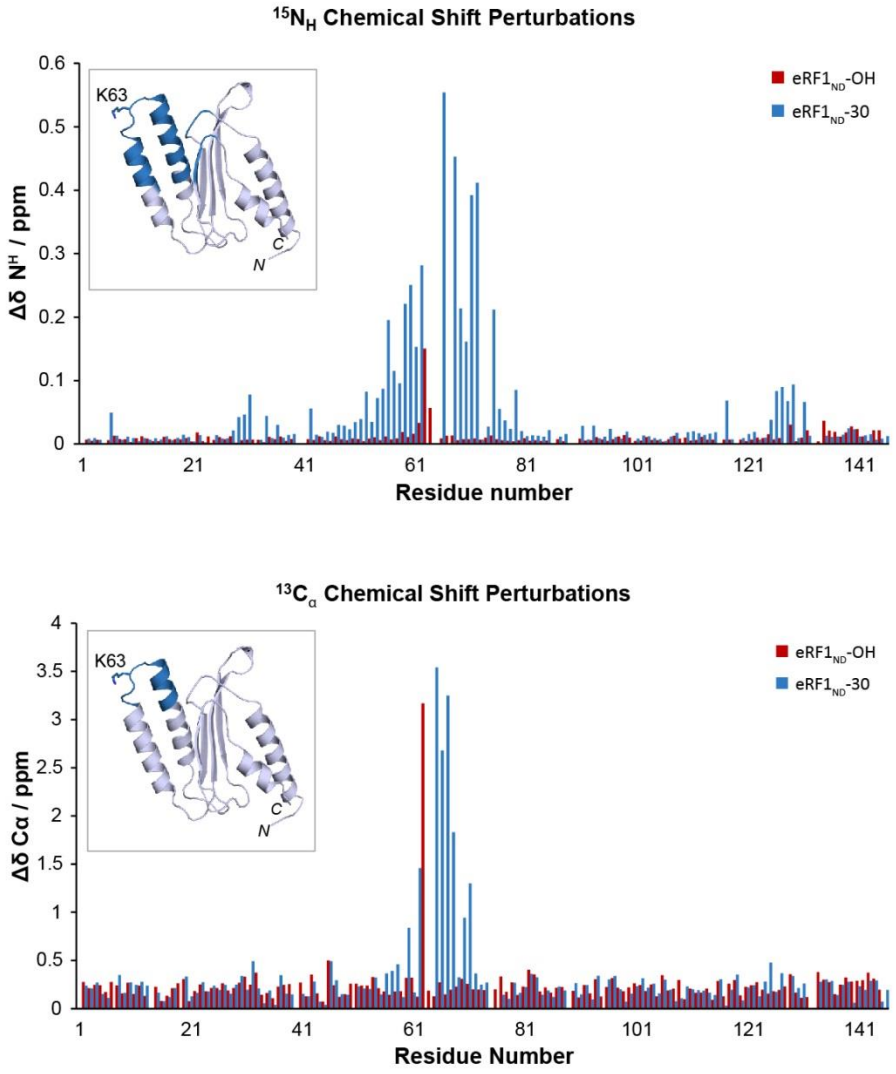


Figure 5.26. (A) ^{15}N and (B) $^{13}C\alpha$ chemical shift perturbations (CSPs) for eRF1$_{ND}$-OH (red) and eRF1$_{ND}$-30 (blue). eRF1$_{ND}$-30 undergoes significant ^{15}N and $^{13}C\alpha$ CSPs. While the ^{15}N chemical shift changes for eRF1$_{ND}$-OH (red) versus wildtype eRF1$_{ND}$ (BMRB 18092[270]) are narrowly localised to K63, eRF1$_{ND}$-30 experiences significant ^{15}N chemical changes in a much broader region. Similarly, eRF1$_{ND}$-30 experiences larger changes in $^{13}C\alpha$ chemical shifts around K63 in comparison to the $^{13}C\alpha$ chemical shifts of wildtype eRF1$_{ND}$ (BMRB 18092[270]), suggesting structural changes rather than sole changes in the local environment. Note that backbone assignment of K63 and several residues surrounding K63 are missing due to incomplete backbone assignment of eRF1$_{ND}$-30. Regions of eRF1$_{ND}$-OH and eRF1$_{ND}$-30 experiencing the greatest ^{15}N and $^{13}C\alpha$ chemical shift changes are highlighted (blue) in the eRF1$_{ND}$ crystal structure (purple, PDB ID: 1DT9[168]). $^{15}N_H$ CSPs were calculated using the equation: N_H CSP = $[(\Delta\delta_H)^2 + (\Delta\delta_N/10)^2]^{1/2}$. $^{13}C\alpha$ CSPs were calculated using the equation: $^{13}C\alpha$ CSP = $|\Delta\delta_{C\alpha}|$.

5.6.3.6 *Investigation of protein dynamics*

Enzymes are dynamic macromolecules that tend to undergo conformational changes during catalysis and/or substrate, ligand, or allosteric binding. Measuring protein dynamics can therefore offer insights into the location and energetics of these conformational changes[283]. NMR is a particularly suitable technique to investigate protein dynamics because it can monitor

fast time-scale dynamics (ps-µs range) by monitoring the signal relaxation times in the x, y, and z directions. T_1 (longitudinal relaxation time), T_2 (transverse relaxation time), and NOE (Nuclear Overhauser Effect) signals were measured to study the µs and ns dynamics of the wildtype and modified $eRF1_{ND}$ forms. However, as shown in **Figure 5.27**, no significant differences were observed between the three forms, suggesting that both $eRF1_{ND}$-OH and $eRF1_{ND}$-30 behave like the wildtype protein in solution.

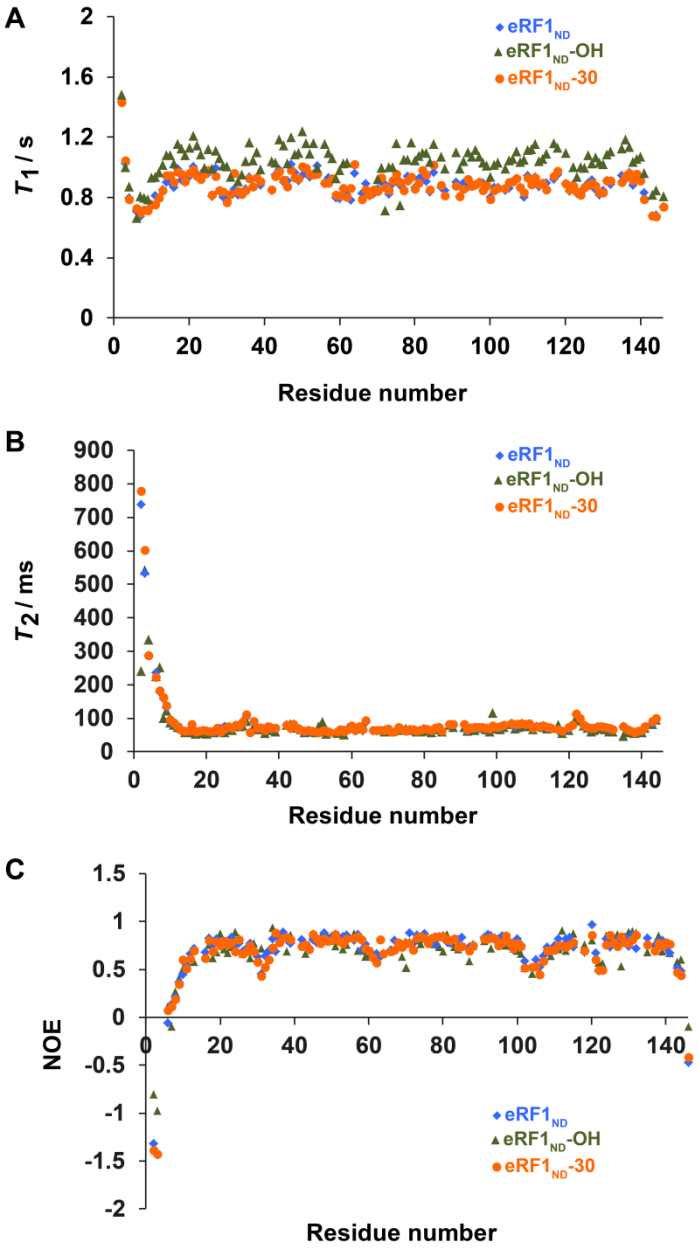


Figure 5.27. (A) T1, (B) T2, and (C) NOE NMR experiments for $^{15}N/^{13}C$-labelled $eRF1_{ND}$, $eRF1_{ND}$-OH, and $eRF1_{ND}$-30. Experiments revealed no significant differences in backbone dynamics between wildtype (blue), hydroxylated (green), and +30 Da modified (orange) $eRF1_{ND}$, indicating that modifications to $eRF1_{ND}$ do not cause changes in protein backbone dynamics.

5.6.3.7 $eRF1_{ND}$ interactions with ribosomal RNA

Previous NMR studies have shown that $eRF1_{ND}$ interacts with ribosomal RNA (rRNA)[263]. Several model RNA oligonucleotides mimicking Helix 44 of 18S rRNA were shown to bind $eRF1_{ND}$[263]. A 15-mer duplex RNA mimicking the decoding region of Helix 44 in 18S rRNA was synthesised (by Dr Afaf El-Sagheer, Brown Group, University of Oxford) for NMR binding studies with $eRF1_{ND}$-OH and $eRF1_{ND}$-30 (**Figure 5.28A**).

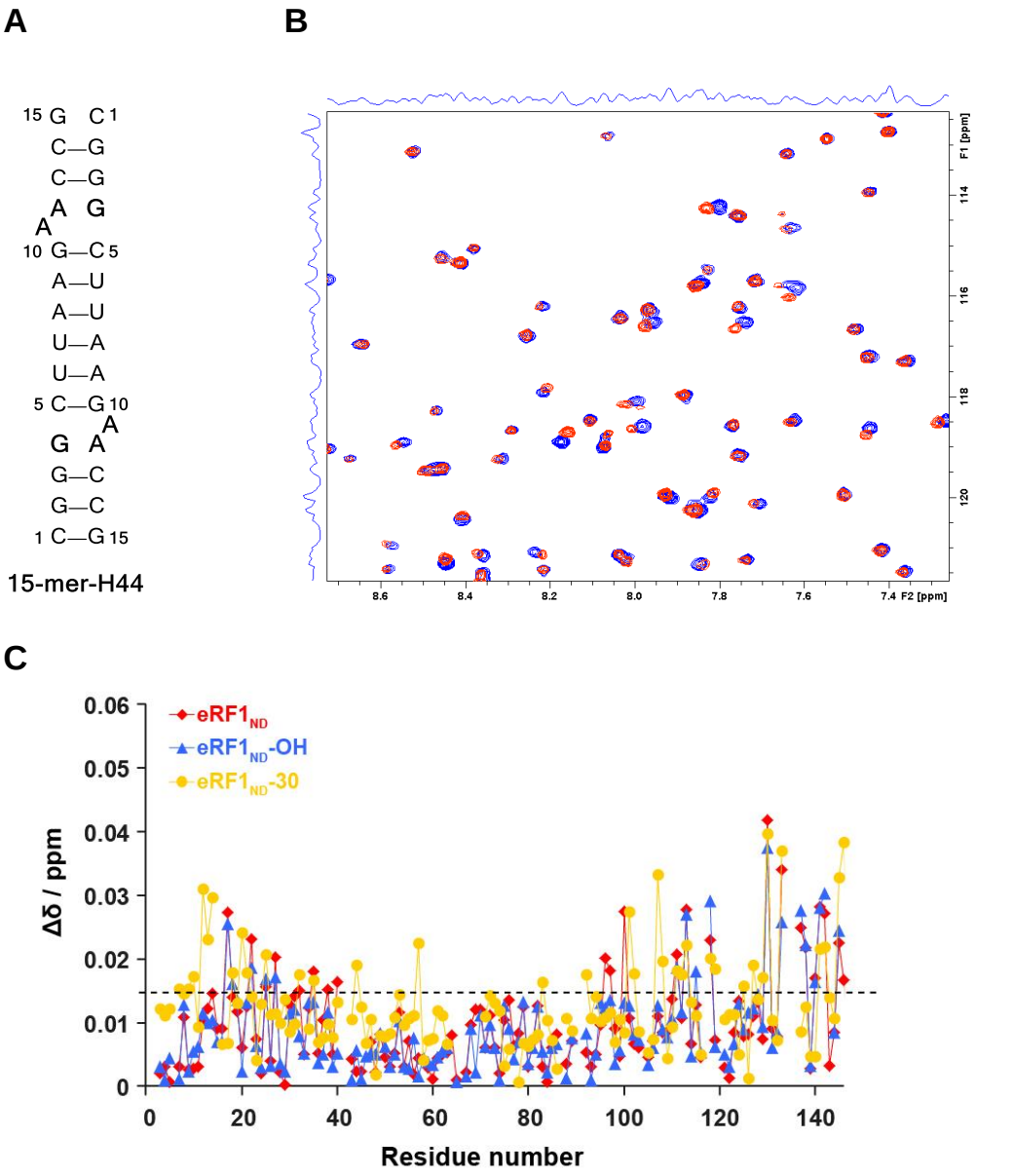


Figure 5.28. ^{1}H-^{15}N HSQC reveals no significant differences in rRNA binding between $eRF1_{ND}$, $eRF1_{ND}$-OH, and $eRF1_{ND}$-30. (A) 15-mer duplex RNA was used to mimic the decoding region of Helix 44 in 18S rRNA[263]. ***(B)*** Representative ^{1}H-^{15}N HSQC spectra corresponding to ^{15}N-$eRF1_{ND}$-OH in the absence (red) and presence (blue) of the 15-mer duplex rRNA reveals weak interactions. ***(C)*** Differences between ^{15}N chemical shifts upon rRNA binding for $eRF1_{ND}$ (red), $eRF1_{ND}$-OH (blue), and $eRF1_{ND}$-30 (yellow) reveal no clear differences.

In agreement with reported data[263], weak interactions were observed between α1, β2, β3, β4 and α4 of $eRF1_{ND}$ and the 15-mer duplex RNA via ^{1}H-^{15}N HSQC (**Figure 5.28**). However, no clear differences in rRNA binding were observed between the three $eRF1_{ND}$ forms, indicating that ribosomal binding may not be affected by JMJD4-catalysed modifications.

5.7 Characterisation of the +30 Da modification

To date, human 2OG oxygenases have only been reported to catalyse hydroxylation and related demethylation reactions. However, 2OG oxygenases from other species can catalyse desaturations, ring closures, halogenations, and epimerisations[24]. Thus, it was of interest to characterise the +30 Da modification given the potential for novel chemistry catalysed by a human 2OG oxygenase. Furthermore, the substantial CSPs observed by protein NMR spectroscopy and the apparent destabilising effect of the +30 Da modification suggested that this previously unreported modification has a significant effect on $eRF1_{ND}$ structure and stability, and thus could play an important role in translation termination. Intense efforts were therefore made to structurally characterise the +30 Da modification.

Characterisation of the +30 Da modification was not trivial. No information on the identity of the modification could be gathered from the protein NMR data due to the incomplete assignment of the protein backbone, including of the target K63 and neighbouring residues. Additionally, because no new peak was observed by amino acid analysis, MS characterisation of the +30 Da modified amino acid was not possible. Therefore, efforts were directed towards characterising the +30 Da modification using a +30 Da modified eRF1 peptide (as catalysed by JMJD4) as a chemical probe for analysis by MS and NMR spectroscopy.

5.7.1 Elemental composition analyses of eRF1$_{FL}$ tryptic peptide fragments by LC-MS/MS

** Elemental composition analyses described in this section were performed by Owain Jones, under my supervision and guidance, and appear in a modified form in his Biochemistry Part II thesis*[278].

The elemental composition (i.e. molecular formula determination) for molecular or fragment ions can be determined using high resolution MS analyses and an elemental composition calculator using MassLynx V4.1 software (Waters). To gain insights into the identity of the +30 Da modified species, elemental composition analysis was performed using the y3 fragment ion (420.2194 Da) arising from the LC-MS/MS data of the eRF1$_{FL}$-30 tryptic peptide (MLADEFGTASNI**K**$^{+30}$SR) described in **Section 5.5.3** (**Figure 5.20**). Analysis was similarly performed on the y2 fragment ions arising from the unmodified (MLADEFGTASNI**K**) and hydroxylated (MLADEFGTASNI**K**^{OH}) eRF1$_{FL}$ tryptic peptides as positive controls to assess the quality of the analyses, quantified by the degree of error (**Figure 5.20**).

As shown in **Table 5.2**, the elemental composition analyses of the y3 ion, corresponding to the K^{+30}-S-R tripeptide arising from the eRF1$_{FL}$-30 tryptic peptide, predicts a molecular formula of $C_{15}H_{30}N_7O_7^+$ (-3.1 ppm error). Notably, the error is consistent with those generated for the molecular formulas of the unmodified y2 ($C_{12}H_{26}N_3O_3$, -3.1 ppm) and hydroxylated y3 ($C_{16}H_{32}N_5O_6^+$, 4.1 ppm) positive ions (**Table 5.2**). The theoretical chemical formula of an unmodified K-S-R tripeptide y ion fragment is $C_{15}H_{32}N_7O_5^+$. A comparison with the predicted chemical formula for the analogous +30 Da modified y3 fragment ion indicates that the +30 Da modification theoretically corresponds to the loss of two hydrogen atoms and incorporation of two oxygen atoms ($C_{15}H_{30}N_7O_7^+$). A proposed structure that fits the chemical formula for the +30 Da modification corresponds to an α-hydroxyketone species, formed by three oxidative events – two hydroxylations at the C4 position, resulting in ketone formation via condensation, and a proposed third hydroxylation at a different position on the K63 side chain (**Figure 5.29**).

Table 5.2. Elemental composition analyses of eRF1$_{FL}$ tryptic fragment y ions.*

eRF1$_{FL}$ y2 ion (I-K) *Theoretical formula:* $C_{12}H_{26}N_3O_3^+$ *Observed mass:* 260.1966 Da			**eRF1$_{FL}$-OH y3 ion (N-I-K^{OH})** *Theoretical formula:* $C_{16}H_{32}N_5O_6^+$ *Observed mass:* 390.2369			**eRF1$_{FL}$-30 y3 ion (K^{+30}-S-R)** *Observed mass:* 420.2194		
Formula	**m/z**	**Error (ppm)**	**Formula**	**m/z**	**Error (ppm)**	**Formula**	**m/z**	**Error (ppm)**
$C_{12}H_{26}N_3O_3$	260.1974	-3.1	$C_{16}H_{32}N_5O_6$	390.2353	4.1	$C_{15}H_{30}N_7O_7$	420.2207	-3.1
$C_{17}H_{26}NO$	260.2014	-18.4	$C_{15}H_{32}N_7O_5$	390.2465	-24.6	$C18H_{26}N_7O_5$	420.1995	47.4
$C_{16}H_{24}N_2O$	260.1889	29.6	$C_{15}H_{30}N_6O_6$	390.2227	36.4	$C_{15}H_{32}N_8O_6$	420.2445	-59.7
$C_{13}H_{26}NO_4$	260.1862	40.0	$C_{16}H_{34}N_6O_5$	390.2591	-56.9	$C_{17}H_{24}N_8O_5$	420.1870	77.1
$C_{13}H_{28}N_2O_3$	260.2100	-51.5	$C_{17}H_{36}N_5O_5$	390.2716	-88.9	$C_{16}H_{34}N_7O_6$	420.2571	-89.7
$C_{15}H_{22}N_3O$	260.1763	78.0	$C_{15}H_{28}N_5O_7$	390.1989	97.4			
$C_{12}H_{24}N_2O_4$	260.1736	88.4						
$C_{12}H_{28}N_4O_2$	260.2212	-94.5						
$C_{14}H_{30}NO_3$	260.2226	-99.9						
Limits: $C_{12\text{-}50}H_{10\text{-}50}N_{1\text{-}50}O_{1\text{-}50}$			**Limits:** $C_{15\text{-}50}H_{15\text{-}50}N_{5\text{-}50}O_{5\text{-}50}$			**Limits:** $C_{15\text{-}50}H_{15\text{-}50}N_{7\text{-}50}O_{5\text{-}50}$		

** Chemical formulas were generated using the Elemental Analysis tool on Mass Lynx V4.1 (Waters).*

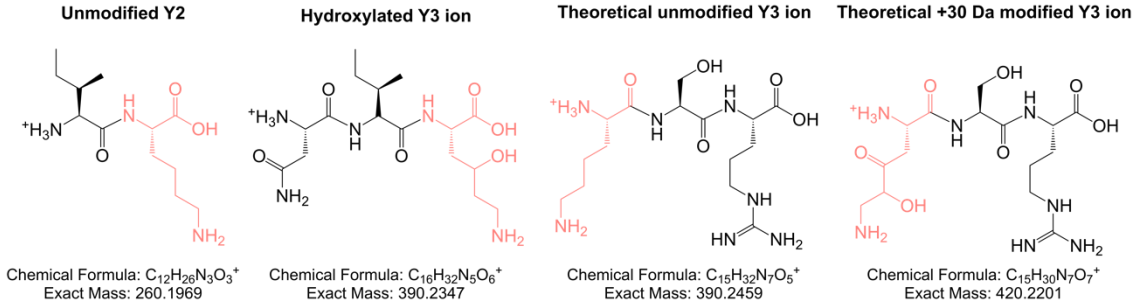


Figure 5.29. Theoretical y ions with respective chemical formulas and exact masses corresponding to the observed y ions arising from eRF1$_{FL}$ unmodified and modified tryptic fragments. Theoretical y ion chemical formulas and exact masses were generated using ChemDraw Professional 16.0.

5.7.2 Peptide optimisation for study by NMR

** The experiments described in **Section 5.7.2** were performed by Owain Jones, under my supervision and guidance, and appear in a modified form in his Biochemistry Part II thesis*[278].

Because the study of the +30 Da modification by protein NMR was limited by the incomplete backbone assignment of the $^{15}N/^{13}C$-eRF1$_{ND}$-30 dataset, attempts were made to identify an eRF1 peptide substrate that could undergo the +30 Da modification for study by NMR spectroscopy. This approach was motivated by several reasons, including the following advantages of using peptide substrates as biochemical probes:

(1) Peptide synthesis by SPPS, as compared with *E. coli* expression and purification of protein substrates, is more time efficient and produces greater yields. This is an advantage given that NMR studies require high protein yields and purity.

(2) NMR spectroscopy offers the most detailed and reliable information about the functional groups present in the +30 Da modification, as compared with other techniques, such as MS and chemical tests, which provide limited structural information.

(3) Peptides can be more easily studied by NMR spectroscopy due to their smaller size. Therefore, 2D NMR experiments can be typically performed on peptides using natural ^{15}N and ^{13}C abundances (i.e. peptides do not require expensive ^{15}N- and/or ^{13}C-isotopic labelling). In addition, 2D experiments typically require significantly less acquisition time than 3D experiments on protein samples.

(4) Peptide substrates can be efficiently separated from JMJD4 post-modification by precipitation with methanol[133], rather than requiring the use of SEC, which is used to separate $eRF1_{ND}$ from JMJD4.

Though there are several advantages associated with using a peptide substrate, there were also limitations envisaged, the biggest of which was the identification of a peptide substrate that can undergo +30 Da modification. Previous studies on JMJD4-catalysed eRF1 hydroxylation detailed that a linear eRF1 peptide substrate could not be identified for JMJD4[133]. These reports highlighted the likely importance of secondary structure in enzyme:substrate recognition; the target K63 residue of eRF1 is located on the apex of a sharp turn (loop) situated between two α-helices (**Figure 5.3**). However, a cyclic eRF1 peptide substrate was identified as a substrate for JMJD4-catalysed hydroxylation in these studies[133]. Thus, efforts were made to identify a fairly short linear or cyclic peptide substrate that could undergo JMJD4-catalysed +30 Da modification for subsequent characterisation by NMR spectroscopy.

5.7.2.1 Linear peptides

Previous studies showed that a 24-mer eRF1 peptide (LADEFGTASNIKSRVNRLSVLGAI-NH_2) was a poor substrate for JMJD4-catalysed hydroxylation[133]. Notably, the target K63 residue was situated in the centre of the peptide. As described in **Section 2.3.2**, the position of the target K63 residue of eRF1 mimics that of the target K22 residue of human DRG1 (assumed to be structurally similar to its yeast homologue, RBG1, due to high sequence identity in this region), a substrate for a related JmjC hydroxylase, JMJD7. Both target residues protrude from a loop situated between two α-helices in a helix-turn-helix motif. Like the 24-mer eRF1 linear peptide, a 20-mer DRG1 linear peptide containing the target K22 residue in the centre of the sequence was similarly not found to be a JMJD7 substrate (**Figure 2.11**). However, linear DRG1 peptides containing the target K22 residue near the *N*-terminus were found to be JMJD7 substrates. Furthermore, lengthening the linear peptide from the *N*-terminus to include highly conserved residues yielded efficient JMJD7 substrates (**Figure 2.15**).

Based on the structural similarity observed between eRF1 and DRG1 as JmjC hydroxylase substrates, eRF1 peptides ($eRF1_{61-80}$, NIKSRVNLSVLGAITSVQQ-NH_2; $eRF1_{57-81}$, GTASNIKSRVNRLSVLGAITSVQQR-NH_2) were designed with the target K63 residue situated closer to the *N*-terminus and synthesised by SPPS. The peptides were initially tested for JMJD4-catalysed hydroxylation. As shown in **Figure 5.30**, $eRF1_{61-80}$ was not a JMJD4 substrate; however, $eRF1_{57-81}$ was identified as a linear peptide substrate, though only limited (~35%) JMJD4-catalysed hydroxylation was observed. Following the identification of $eRF1_{57-81}$ as a linear peptide substrate, $eRF1_{57-81}$ was incubated with JMJD4 under the conditions optimised for +30 Da modification; however, $eRF1_{57-81}$ was not a substrate for JMJD4-catalysed +30 Da modification, though near complete hydroxylation was observed under these conditions.

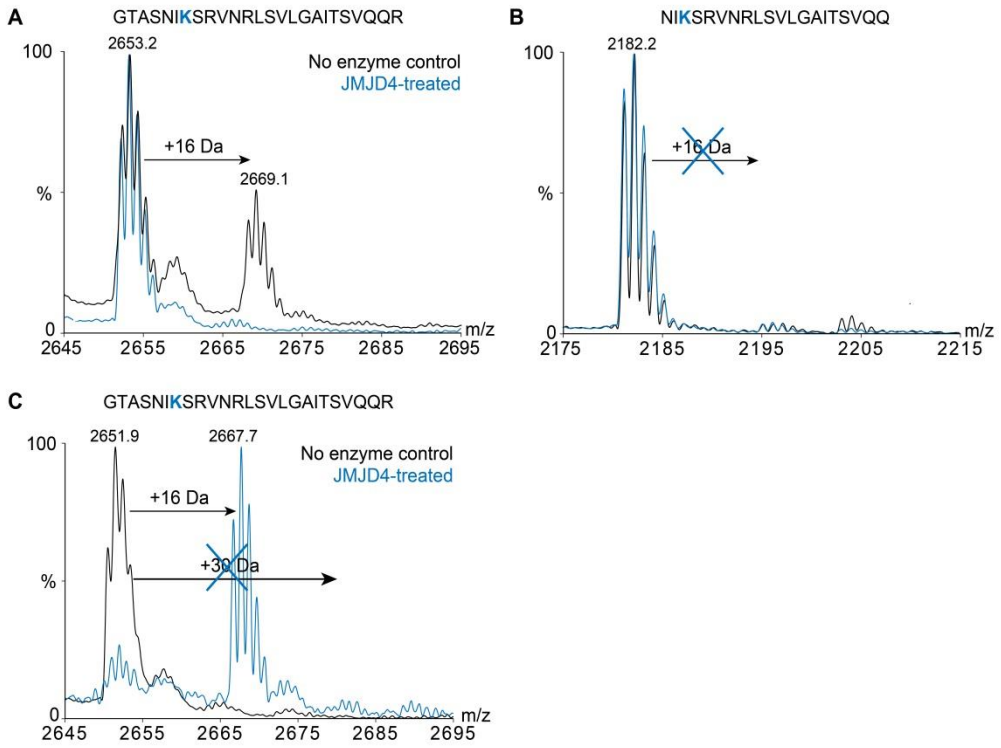


Figure 5.30. MALDI-TOF mass spectra of eRF1 linear peptides incubated in the absence (black) and presence (blue) of JMJD4. (A) $eRF1_{61\text{-}80,}$ but not ***(B)*** $eRF1_{57\text{-}81}$ can undergo JMJD4-catalysed hydroxylation under the optimised hydroxylation conditions described in **Section 5.3.1**. ***(C)*** While $eRF1_{61\text{-}80}$ undergoes near complete hydroxylation under the optimised +30 Da conditions (**Section 5.3.2**), there is no evidence for the formation of the +30 Da modified peptide product.

5.7.2.2 Cyclic peptides

As described in previous studies, a cyclic eRF1 peptide was proposed to more closely mimic the tight turn of the NIKS loop region than a linear peptide, thus enabling enzyme:substrate recognition[133]. This study reports a 'dimeric' eRF1 cyclic peptide as a substrate for JMJD4-catalysed hydroxylation. However, the dimerisation of the eRF1 cyclic peptide was likely an artefact resulting from harsh cyclisation conditions (80°C, 10 minutes), and the resultant peptide did not fly well by MALDI-TOF-MS, thus yielding relatively poor spectra (Dr Sarah Wilkins, personal communication). Consequently, efforts were made to synthesise a similar 'monomeric' cyclic peptide that would yield a higher quality MALDI-TOF mass spectrum and serve as a good substrate for the +30 Da modification.

A linear eRF1 precursor peptide (linear $eRF1_{57\text{-}70}$; GTASNIKSRVNRLS[D-Cys]) was synthesised by SPPS and subsequently cyclised as described[133] (**Figure 5.31**). A chloroacetyl group was added to the *N*-terminal Gly residue by reacting the peptide with chloroacetic anhydride, as monitored by MALDI-TOF-MS. Despite incomplete chloroacetylation following a 3 h incubation at room temperature, the unmodified and chloroacetylated peptides were successfully separated and purified by high-performance liquid chromatography (HPLC). The pure chloroacetylated eRF1 peptide was cyclised as described in **Section 7.5.3** [room temperature, 2 h – note gentler cyclisation conditions than those described in Feng, T. *et al (2014)*[133]]. Following HPLC purification of cyclic $eRF1_{57\text{-}70}$, both linear and cyclic $eRF1_{57\text{-}70}$ peptides were incubated with JMJD4 under the optimised +30 Da conditions and analysed by MALDI-TOF-MS. While both peptides underwent near complete JMJD4-catalysed hydroxylation, no evidence for the +30 Da modified product was observed (**Figure 5.32**).

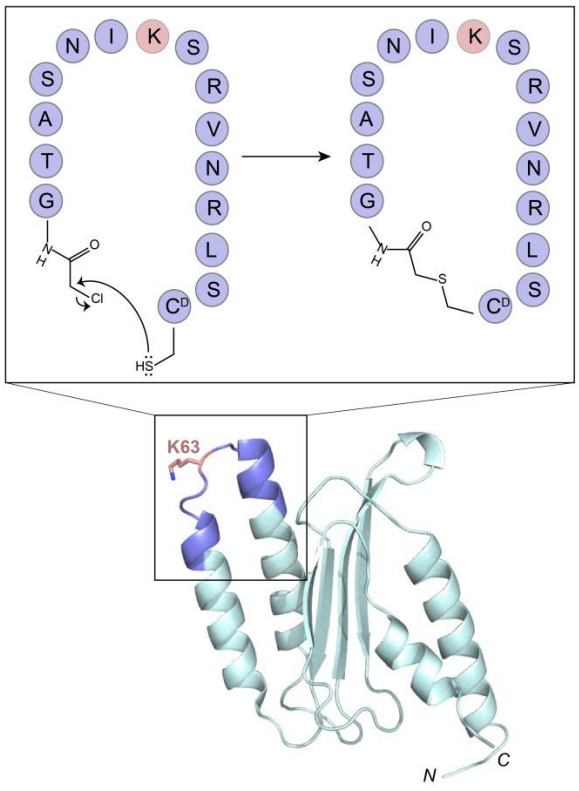


Figure 5.31. Design of cyclic eRF1$_{57\text{-}70}$ peptide. The residues encompassing cyclic $eRF1_{57\text{-}70}$ are highlighted in purple within a crystal structure of $eRF1_{ND}$ (PDB ID: 1DT9[168]). Cyclisation occurs under slightly basic conditions via nucleophilic attack (S_N2) of the *N*-terminal chloroacetyl group by the *C*-terminal thiol to a form a thioether linkage.

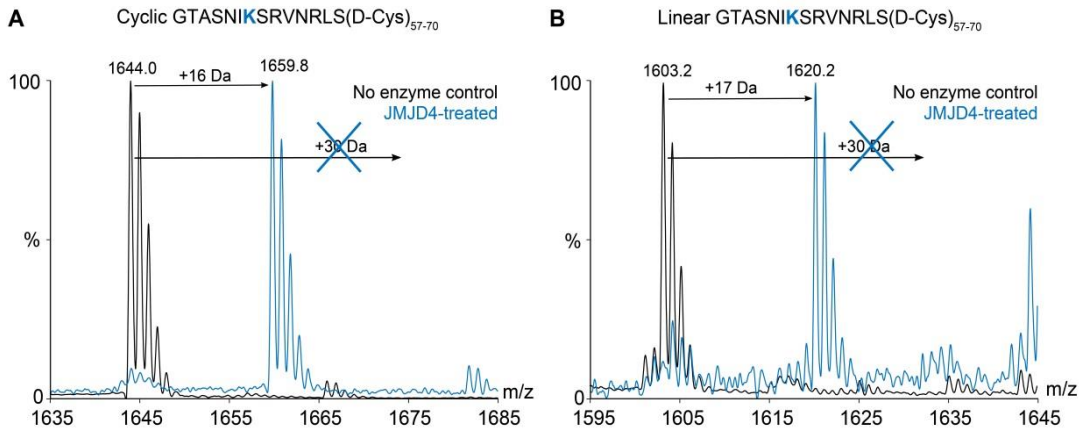


Figure 5.32. MALDI-TOF mass spectra of (A) cyclic and (B) linear eRF1$_{57-70}$ following incubation in the absence (black) or presence (blue) of JMJD4 under the optimised +30 Da conditions. While hydroxylation of both the cyclic and linear eRF1$_{57-70}$ peptides is observed, there is no evidence for the formation of the +30 Da modified peptide products.

Note that longer peptides (45-46 amino acids), designed to encompass the NIKS loop and flanking α-helices, were also tested as substrates to produce the JMJD4-catalysed +30 Da modification (**Figure 5.33**; **Table 5.3**). Because these peptides were too long to conveniently synthesise by SPPS, they were recombinantly produced in *E. coli* with *N*-terminal thiorodoxin (Trx)- or glutathione S-transferase (GST)-tags. Details for the cloning, expression, and purification of these eRF1 constructs can be found in Jones, O. (2016)[278]. Despite the design of longer eRF1 peptides that likely more closely mimic the α-helical secondary structure of eRF1, these peptides were not substrates for JMJD4-catalysed +30 Da modification (data not shown).

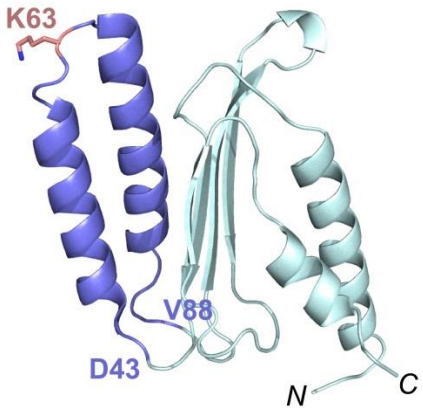


Figure 5.33. View from a crystal structure of eRF1$_{ND}$ (PDB ID: IDT9) highlighting the region encompassed by the GST- and Trx-tagged eRF1$_{43-87/88}$ peptides (Table 5.3).

Table 5.3. eRF1 constructs for E. coli expression.

eRF1 Construct	Amino acid sequence
GST-46mer	DQISRVAKMLADEFGTASNI**K**SRVNRLSVLGAITSVQQRLKLYNKV
Trx-45mer	DQISRVAKMLADEFGTASNI**K**SRVNRLSVLGAITSVQQRLKLYNK

It was striking that none of the tested eRF1 peptides underwent JMJD4-catalysed +30 Da modification under such co-factor/substrate and enzyme saturating conditions. Despite efforts to identify a peptide substrate for JMJD4-catalysed +30 Da modification, the *N*-terminal construct ($eRF1_{ND}$) remains the smallest +30 Da substrate identified to date. This observation suggests that formation of the +30 Da modification likely relies on JMJD4:eRF1 interactions distant from the eRF1 K63 hydroxylation/+30 Da modification site. Alternatively, it is possible that the intact *N*-terminal domain contains secondary structure elements, not found in the short eRF1 linear or cyclic peptides, that promote +30 Da formation. Given these possibilities, a new strategy was devised to isolate a +30 Da modified eRF1 peptide for study by NMR spectroscopy.

5.7.2.3 Isolation of a +30 Da modified $eRF1_{ND}$ tryptic peptide fragment

Because the $eRF1_{ND}$ protein was the smallest identified substrate for JMJD4-catalysed +30 Da modification, a strategy entailing (1) the large-scale +30 Da modification of $eRF1_{ND}$, (2) tryptic digestion of $eRF1_{ND}$-30, and (3) isolation of a +30 Da modified tryptic peptide fragment by HPLC for subsequent analysis by NMR spectroscopy was designed (**Figure 5.34**). This technique provided a viable, though challenging and labour-intensive method to isolate a +30 Da modified eRF1 15-mer peptide (MLADEFGTASNI$\mathbf{K}^{+30}$SR) for study by NMR. Importantly, prior MS/MS analyses of tryptic peptides indicated that formation of this 15-mer $eRF1_{ND}$-30 tryptic peptide is reproducible and of a suitable size for NMR studies.

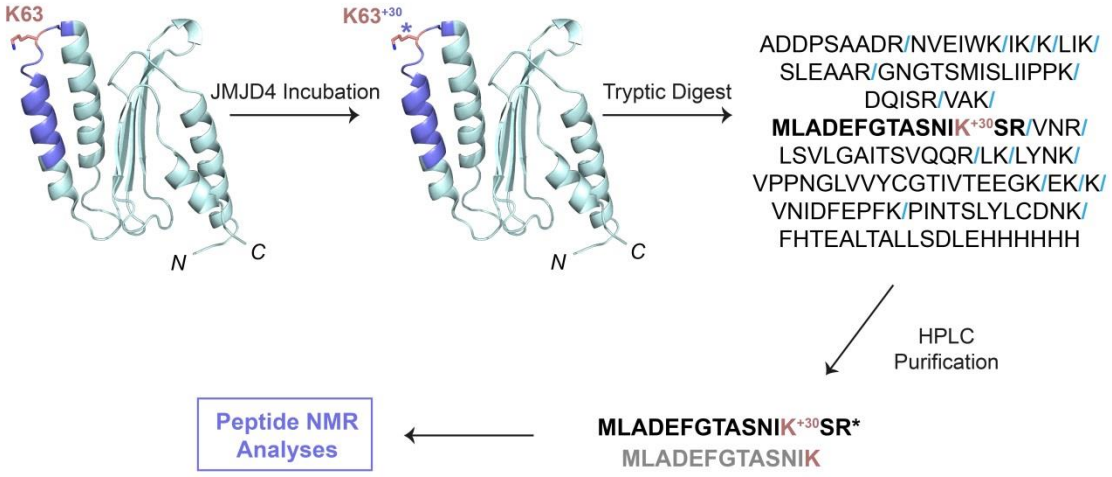


Figure 5.34. Outline of the strategy used to isolate a +30 Da modified eRF1$_{ND}$ tryptic peptide for study by NMR spectroscopy. eRF1 PDB ID: 1DT9[168].

Despite the viability of this strategy, the largest limitation was the large amounts of JMJD4 and eRF1$_{ND}$ proteins required in the initial large-scale modification to ensure that an adequate yield of the +30 Da modified peptide resulted for studies by NMR. Loss of the protein/peptide of interest was anticipated at all steps, including from (1) incomplete formation of eRF1$_{ND}$-30 during large scale modification, (2) loss of eRF1$_{ND}$ during purification from JMJD4 by SEC, (3) potential incomplete tryptic digestion, and (4) loss of peptide during HPLC purification. Furthermore, this strategy was limited by the large amount of JMJD4 required, especially given the relatively low yields of JMJD4 isolated following *E. coli* expression and purification (~2.5 mg/L).

In addition to yield limitations, the strategy also relied on high peptide purity (> 85%) following the final HPLC purification. To test the viability of this method, a tryptic digest was first performed using wildtype eRF1$_{ND}$, and HPLC conditions were optimised for isolation of the peptide containing the target K63 residue (data not shown)[278]. However, it is noteworthy that the eRF1$_{ND}$ tryptic digest generates a different K63-containing 13-mer peptide (MLADEFGTASNI**K**) than that for the eRF1$_{ND}$-30 15-mer (MLADEFGTASNI**K**$^{+30}$SR), due to the missed tryptic cleavage. Given the predicted incorporation of two oxygen atoms by accurate MS elemental composition analyses (**Section 5.7.1**) and the addition of serine and arginine to

the *C*-terminal end of the peptide, the +30 Da modified peptide was predicted to be more hydrophilic than the unmodified 13-mer. These observations and predictions were considered when designing an HPLC gradient for the ultimate purification of the $eRF1_{ND}$-30 15-mer peptide fragment.

5.7.2.3.1 Large scale +30 Da modification of $eRF1_{ND}$ and purification of $eRF1_{ND}$-30

Large scale assays (3 x 6 mL volumes), containing ~8 mg $eRF1_{ND}$ and ~24 mg JMJD4 per reaction, were performed using the optimised +30 Da conditions (**Section 5.3.2**). The large amounts of $eRF1_{ND}$ and JMJD4 were required to ultimately isolate enough $eRF1_{ND}$-30 15-mer peptide for NMR analysis. $eRF1_{ND}$-30 was purified from JMJD4 as described in **Section 5.3.2**, yielding ~8 mg.

5.7.2.3.2 Tryptic digestion and HPLC purification of $eRF1_{ND}$-30 15-mer

Purified $eRF1_{ND}$-30 (~8 mg) was digested with trypsin in a 1:25 ratio as described in **Section 7.8.3.1** (note that the peptides were not purified by Sep-Pak C18 cartridges following digestion to avoid loss of material). The resultant peptides were then purified by HPLC as described in **Section 7.6** to isolate the $eRF1_{ND}$-30 15-mer. MALDI-TOF-MS was used to identify $eRF1_{ND}$-30 15-mer-containing fractions and assess target peptide purity. The purest HPLC fractions were combined and lyophilised for subsequent NMR analyses.

5.7.3 Analysis of the isolated +30 Da modified peptide by NMR spectroscopy

** The NMR experiments described in **Section 5.7.3** were performed in collaboration with Professor Timothy D. W. Claridge (NMR Facility, Department of Chemistry, University of Oxford).*

The purified $eRF1_{ND}$-30 15-mer peptide (MLADEFGTASNI**K**$^{+30}$SR; ~0.5 mg) was re-suspended in 320 µL of D_2O. Note that despite efforts to weigh the mass of the peptide, no

accurate measurement could be achieved. Therefore, the peptide was re-suspended in the smallest volume possible to attain the highest peptide concentration for NMR experiments. D_2O was chosen to optimise detailed COrrelated SpectroscopY (COSY) and TOtal Correlated SpectroscopY (TOCSY) spectra of the K63 side chain to enable resonance assignments. However, given the proposed α-hydroxyketone +30 Da structure (**Figure 5.29**), the use of D_2O was anticipated to limit the detection of amide hydrogens and a potential enol, likely due to hydrogen-deuterium exchange.

1D ^{1}H and 2D ^{1}H-^{1}H COSY, TOCSY, and Nuclear Overhauser Effect SpectroscopY (NOESY) experiments were performed to assign the peptide backbone resonances (**Section 7.13.3**). In COSY experiments, magnetisation transfer occurs via scalar or through-bond coupling, giving rise to correlated ^{1}H-^{1}H signals that are two or three bonds apart[179]. By contrast, the magnetisation transfer in TOCSY experiments is dispersed over a complete spin system by successive scalar couplings, giving rise to ^{1}H-^{1}H correlations between all protons in the same spin system, even those connected by more than three chemical bonds[179]. Note that different amino acids belong to different spin systems due to lack of scalar coupling across the amide bond. Thus, TOCSY spectra include the signals that appear in COSY spectra, plus those of other protons in the same spin system. Amino acids have characteristic signal patterns that enable their identification, though some amino acids have similar spin systems and consequently identical patterns, which complicates resonance peak assignment. However, ^{1}H-^{1}H NOESY experiments support resonance assignments by providing 'through-space' ^{1}H-^{1}H correlations for protons within ~5 Å via dipolar coupling[179]. Peptide resonance peak assignment was carried out using the COSY, TOCSY, and NOESY spectra in combination with 2D ^{1}H-^{13}C HSQC and Heteronuclear Multiple Bond Correlation (HMBC) (described in more detail below; **Table 5.4**). While complete peptide assignment was not possible due to peak overlap and missing resonances, most amino acid residues were assigned, including the resonances arising from the proposed target K63 and neighbouring S64 residues.

As shown in the COSY and TOCSY spectra (**Figure 5.36**), ^{1}H-^{1}H correlations were observed for the K(αH) and two K(βH) protons. However, no correlation between K(αH) and K(γH), K(δH), and K(εH) were observed in the TOCSY spectrum. Interestingly, correlations between K(δH) and the two K(εH) protons were observed in the COSY and TOCSY spectra, though in a different spin state than that of K(αH) and the two K(βH) protons. The COSY/TOCSY spectra therefore indicated that there was likely no K(γH) present, consistent with the presence of a ketone at the C4 position in the previously proposed α-hydroxyketone species (**Figure 5.29**). Note that the characteristic ^{1}H-^{1}H correlations are highlighted for S64 in **Figure 5.36**. Though there are two serine residues in the peptide sequence (S60 and S64), the NOESY spectrum provided through-space correlations that enabled S64 to be distinguished, primarily via the unexpected correlation observed between S(αH) and K(δH).

Table 5.4. eRF1$_{ND}$-30 15-mer peptide chemical shift assignments.

		Hα	Hβ	Hγ	Hδ	Hε	Cα	Cβ	Cγ	Cδ	Cε
51	M				-	1.90				-	13.8
52	L					-					-
53	A	4.16	1.23	-	-	-	49.84	16.2	-	-	-
54	D	4.54	2.63, 2.58	-	-	-	50.3	35.9	-	-	-
55	E				-	-				-	-
56	F	4.44	3.04, 2.82	-			54.9	36.5	-		
57	G	3.80, 3.74	-	-	-	-	42.4	-	-	-	-
58	T	4.14	4.06	1.02	-	-	59.0	66.9	18.7	-	-
59	A	4.05	1.19	-	-	-	49.8	16.2	-	-	-
60	S	4.11	4.05, 3.82	-	-	-	55.5	59.3	-	-	-
61	N	4.40	2.60	-	-	-	50.3	35.9	-	-	-
62	I					-					-
63	K^{+30}	4.32	2.76, 1.84	-	3.79	3.19, 2.99	47.0	34.8	90.2	68.6	36.5
64	S	4.20	3.67, 3.63	-	-	-	55.6	60.7	-	-	-
65	R	4.04	1.65, 1.54	1.35	2.97	-	53.3	28.3	24.2	40.7	

Dash (-) denotes that no chemical shift is expected for this atom. A53/A59 and D54/N61 were not distinguishable. Target K63 and S64 residues are highlighted in red. *Aromatic side chain: ^{1}H chemical shifts: 7.08, 7.15; ^{13}C chemical shifts: 126.6, 127.1, 128.9, 129.0, and 136.4.

2D ^{1}H-^{13}C HSQC and HMBC experiments were performed to support the peptide resonance assignment and characterisation of the +30 Da modification. While the HSQC provides one-bond ^{1}H-^{13}C correlations, HMBC provides ^{1}H-^{13}C correlations arising between two, three, and sometimes four bonds[284]. Notably, there were no peaks observed in the ^{13}C chemical shift range

for a ketone carbonyl group (200-220 ppm) in the HMBC spectrum, suggesting that the proposed α-hydroxyketone species was unlikely (**Figure 5.35**).

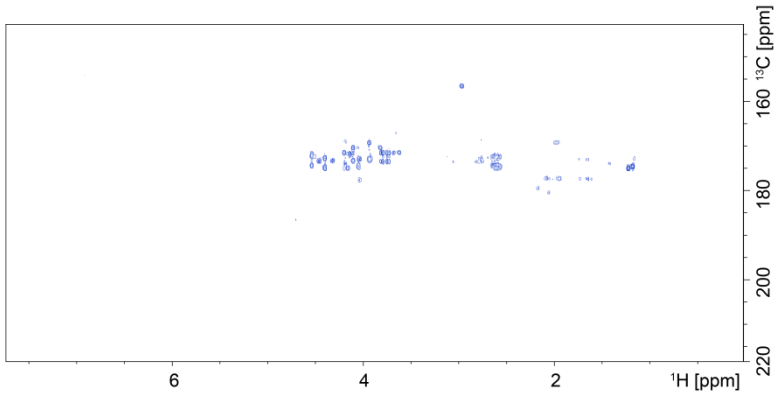


Figure 5.35. ^{1}H-^{13}C HSQC spectrum of the eRF1$_{ND}$-30 15-mer peptide reveals no correlation peaks corresponding to the presence of a ketone (200-220 ppm).

Strikingly, a non-protein-characteristic ^{13}C chemical shift was observed at ~90 ppm in the ^{1}H-^{13}C HMBC spectrum, which was not observed in the ^{1}H-^{13}C HSQC spectrum (**Figure 5.37**). The absence of this peak in the HSQC spectrum indicated that this carbon atom did not have a hydrogen attached to it. However, key correlations between the carbon at ~90 ppm were observed with K(βH), K(εH), and notably, S(αH) in the HSQC spectrum. Because ^{1}H-^{13}C in HMBC spectra arise from scalar, or through-bond couplings[284], these correlations with K(βH) and K(εH) indicated that the chemical shift at ~90 ppm was arising from the K(γC). Furthermore, the observed K(γC)-S(αH) correlation also had to arise from a two or three bond coupling. The only way this correlation could be possible is through a chemical bond formed via the S64 amide with the K63(γC), a structure consistent with γ-lactam formation (**Figure 5.37**).

Given that ^{13}C chemical shifts of alcohols are 50-65 ppm and those of carbonyl groups tend to appear between 170-220 ppm, the observed 90 ppm chemical shift for the K(γC) likely accounts for a single carbon atom attached to an oxygen atom (i.e. hydroxyl group) *and* nitrogen atom (i.e. S64 amide). In addition, the lack of correlations observed between K(γH)

with K(αH), K(βH), K(δH), and K(εH) in the COSY, TOCSY, and NOESY spectra are consistent with the absence of γ-hydrogens at K63. Finally, the ^{13}C chemical shift corresponding to K(δC), which appears at ~65 ppm, is consistent with that of an alcohol, indicating the presence of a hydroxyl group at the C5 or δC position of the lysine side chain. Overall, the only structure corresponding to a +30 Da modification that is consistent with the presented NMR data is oxidation to give a γ-lactam/hemiaminal ring represented in **Figure 5.36** and **Figure 5.37**. The characterisation of this modification as γ-lactam formation is further supported by theoretical ^{1}H and ^{13}C chemical shift predictions (MestreNova) that closely match the observed ^{1}H and ^{13}C chemical shifts, particularly for the strikingly non-protein-characteristic ^{13}C chemical shift assignment of K(γC) at ~90 ppm (**Figure 5.38**).

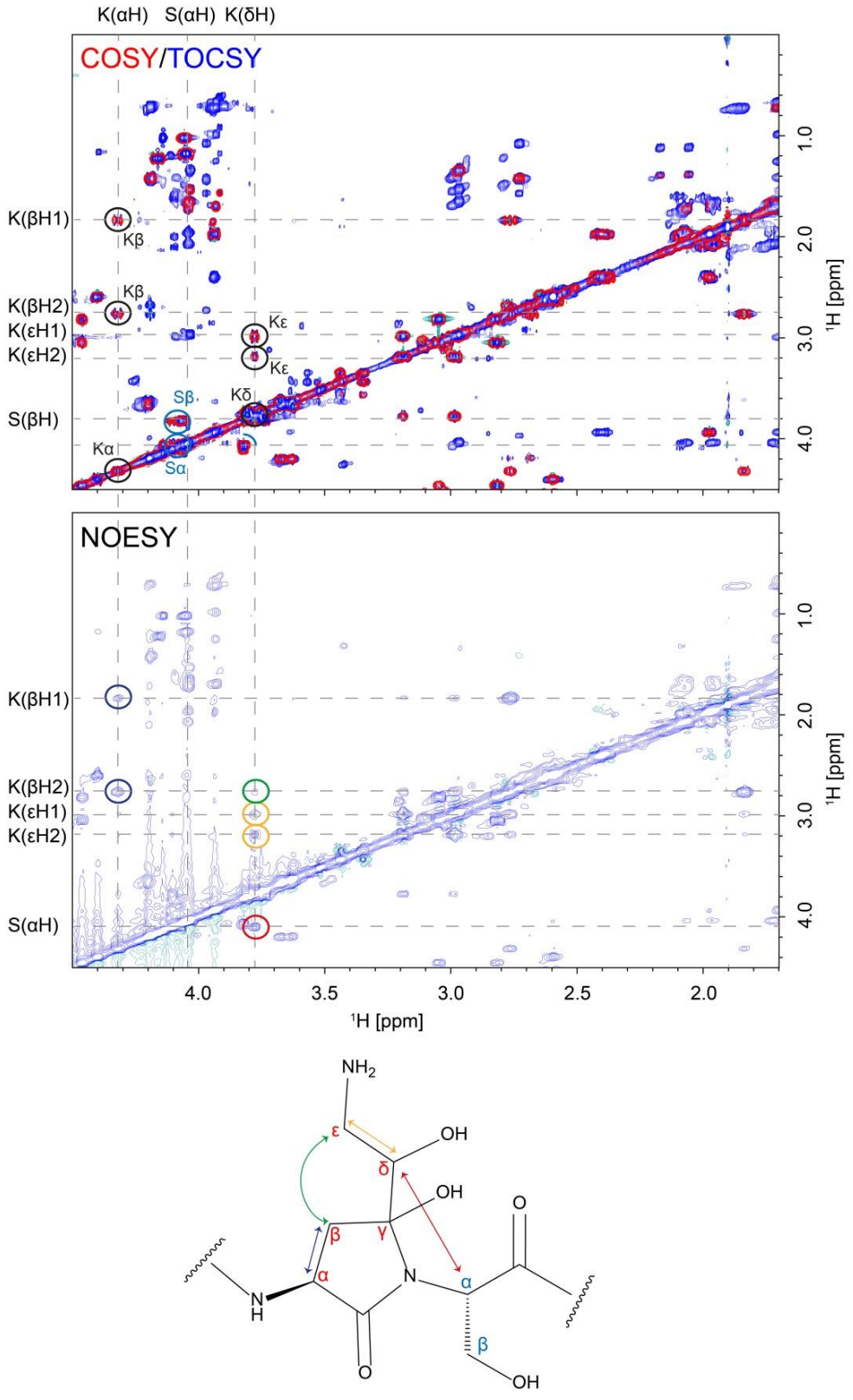


Figure 5.36. ^{1}H-^{1}H COSY, TOCSY, and NOESY NMR spectra used for the assignment of the eRF1$_{ND}$-30 15-mer peptide (MLADEFGTASNIK^{+30}SR). Peaks corresponding to hydrogen atoms in K63 and S64 are indicated. NOESY correlations are indicated in the spectrum and highlighted in the proposed γ-lactam structure corresponding to the +30 Da modification (see ***Figure 5.37***).

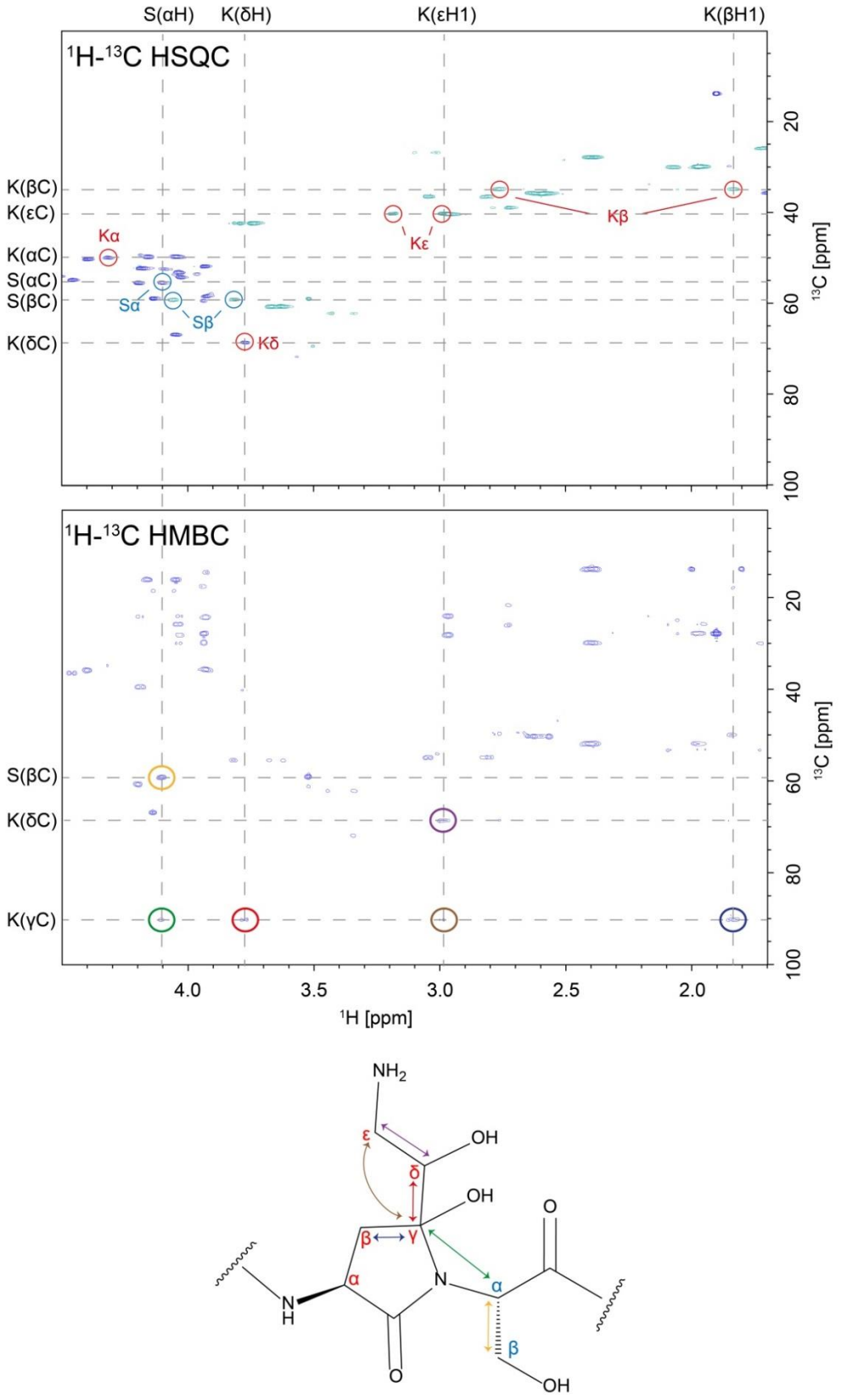


Figure 5.37. ^{1}H-^{13}C HSQC and HMBC NMR spectra used for the structure assignment of the +30 Da modification as oxidised γ-lactam formation. Peaks corresponding to K63 and S64 hydrogen and carbon atoms are indicated. HMBC correlations are indicated in the spectrum and highlighted in the proposed oxidised γ-lactam structure corresponding to the +30 Da modification (bottom).

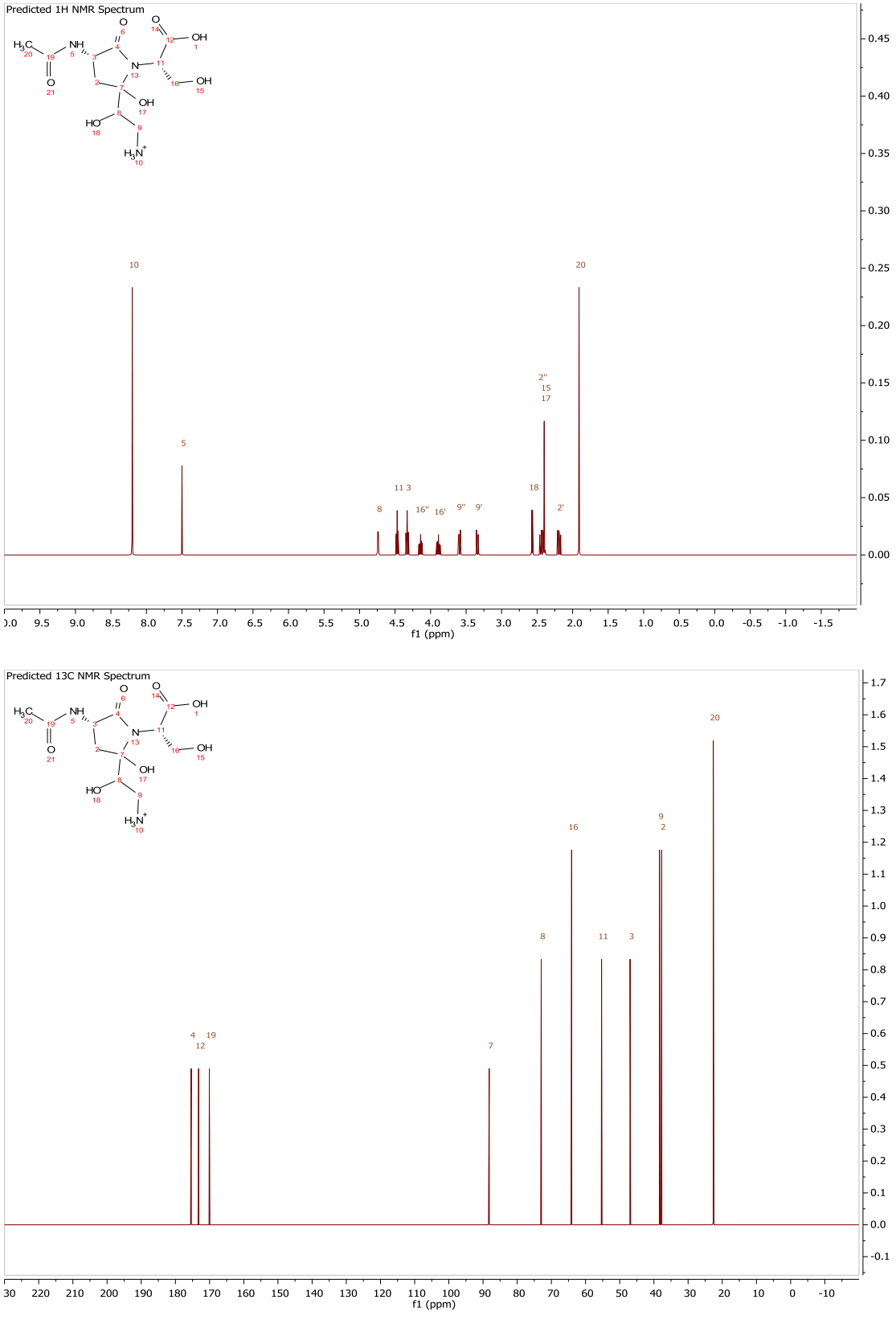


Figure 5.38. Predicted ^{1}H and ^{13}C chemical shifts for the proposed oxidative γ-lactam structure corresponding to the +30 Da modification. Chemical shift predictions were made using MestreNova 10.0 (MestreLab Research). Note that the peak arising at ~90 ppm in the predicted ^{13}C NMR spectrum (bottom) corresponds to the K(γC) of the proposed γ-lactam structure.

5.7.4 Proposed reaction mechanism for γ-lactam formation

While the characterisation of the +30 Da modification as containing a γ-lactam was initially speculative, the proposed reaction mechanism is, in a sense, consistent with the reported hydroxylase activity of JMJD4. As outlined in **Figure 5.39**, there are two related, potential reaction mechanisms by which JMJD4-catalysed γ-lactam formation might proceed.

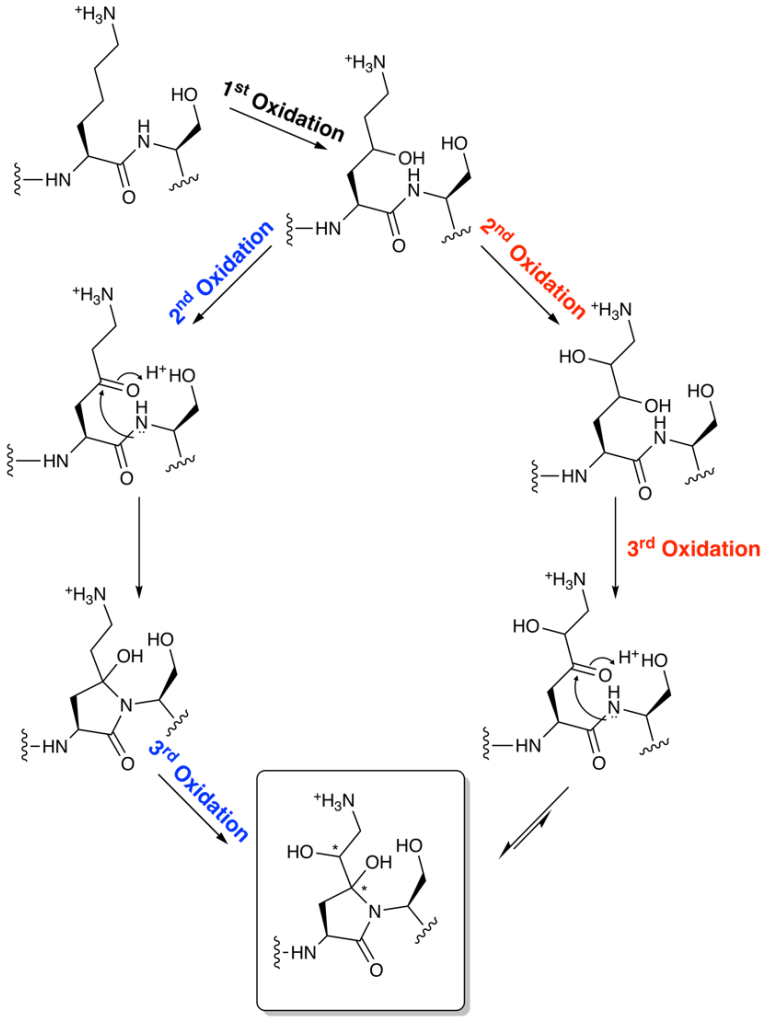


Figure 5.39. Proposed possible reaction mechanisms of JMJD4-catalysed oxidised γ-lactam formation of K63. Figure adapted from Jones, O. (2016)[278].

In both proposed mechanisms, the first step involves C4 lysyl hydroxylation of the K63 side chain, which has been previously reported[133] and validated by the work described in this chapter. Following initial hydroxylation, it is then possible that JMJD4 catalyses a second hydroxylation (blue pathway) at the C4 position, forming a possibly unstable geminal diol,

which hydrolyses to form a ketone. Nucleophilic attack of the carbonyl at the C4 position is then possible by the amide of the neighbouring S64 residue, which forms a γ-lactam and hemiaminal group (common 2OG oxygenase intermediate for the JmjC KDMs; see **Figure 1.4**) at the C4 carbon. A third hydroxylation is then proposed to be catalysed by JMJD4 at the C5 position of the lysine side chain. The second related pathway is proposed to occur in the opposite manner, where C5 lysyl hydroxylation follows initial C4 hydroxylation, and a subsequent third oxidative event at the C4 position occurs to yield the geminal diol/ketone species, which can undergo nucleic attack by the S64 amide to form the γ-lactam species. Note that while only evidence for the γ-lactam species was observed by NMR spectroscopy, it is likely that the γ-lactam species exists in equilibrium with the α-hydroxyketone species, as indicated in **Figure 5.39.**

Currently, no kinetic analyses have been undertaken to provide direct evidence for this reaction mechanism, though it would be of interest to monitor JMJD4-catalysed modification of eRF1 using techniques such as mass spectrometry and NMR to identify reaction intermediates and validate the proposed reaction mechanism.

5.7.5 Mutation studies support oxidative γ-lactam structure

A likely key step in forming the γ-lactam is the nucleophilic attack of a proposed ketone species at the C4 position of the K63 side chain by the backbone amide of the neighbouring S64 residue (**Figure 5.39**). To further investigate the NMR characterisation of the +30 Da modification and proposed reaction mechanism, an $eRF1_{ND}$-S64P variant was tested as a substrate for the +30 Da modification. It was hypothesised that the S64P variant would be unable to undergo γ-lactam formation due to the presence of a tertiary amide in the proline backbone. Thus, $eRF1_{ND}$-S64P was produced and purified on a small scale (100 mL 2TY media growth cultures), as described in **Sections 7.3.5 and 7.4.2**, and tested as a substrate for JMJD4.

Under the optimised +30 Da conditions (**Section 5.3.2**), eRF1$_{ND}$-S64P was found to undergo K63 hydroxylation, though it was not further modified to the +30 Da oxidative γ-lactam species (**Figure 5.40**). The results suggested the following: (1) potential structural changes induced by substituting a serine with a proline residue did not affect enzyme:substrate recognition for hydroxylation (this proposal is further supported by the fact that the K63 and S64 residues are naturally situated in a tight turn between two α-helices; because proline residues are typically involved in forming tight turns, it is possible that proline substitution does not greatly affect the structure); (2) assuming structural changes are not a factor in preventing JMJD4-catalysed γ-lactam formation, the absence of a +30 Da modification in the JMJD4-treated eRF1$_{ND}$-S64P mass spectrum is likely attributable to the inability of the tertiary proline amide to form a γ-lactam, and (3) there are likely different substrate selectivities for hydroxylation and oxidative γ-lactam formation. Given that a second hydroxylation (i.e. +32 Da modification) was not observed by MS, it is likely that JMJD4-catalysed C5 hydroxylation follows γ-lactam formation (blue pathway, **Figure 5.39**) and the consequent conformational change induced by γ-lactam formation allows for the (intermediate) product to be recognised as a JMJD4 substrate.

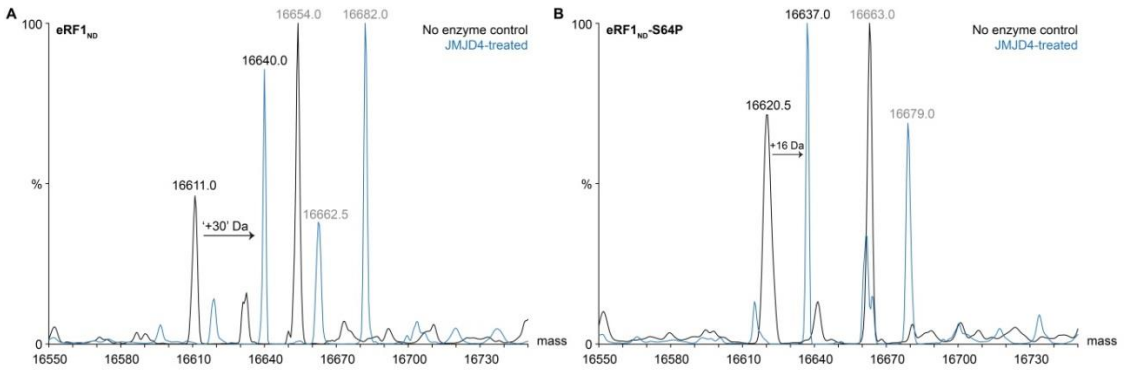


Figure 5.40. Deconvoluted intact protein mass spectra corresponding to (A) eRF1$_{ND}$ and (B) eRF1$_{ND}$-S64P in the absence (black) and presence (blue) of JMJD4 under the optimised +30 Da conditions. S64P substitution of eRF1$_{ND}$ prevents formation of the +30 Da modification, consistent with the proposed characterisation of the +30 Da modification as a γ-lactam formed via the K63 side chain and S64 amide. Proteins were analysed using a Xevo GS-2 QTof mass spectrometer (**Section 7.8.2.2**).

5.8 The +30 Da modification is efficiently formed in bacterial cells

To gain insights into the biological relevance of JMJD4-catalysed +30 Da modification, co-expression experiments were re-visited to determine whether the +30 Da modification could be efficiently produced in bacterial cells or whether the modification was an artefact of isolated protein assays.

5.8.1 Co-expression of eRF1$_{ND}$ and JMJD4

eRF1$_{ND}$-His$_6$ was expressed in the absence or presence of His$_6$-JMJD4 in Rosetta(DE3) pLysS cells in 2TY media following (co-)transformation (**Section 7.2.10**). Proteins were purified on a small scale (**Section 7.4.2**) using loose Ni(II) Sepharose resin, and JMJD4-catalysed +30 Da modification was assessed by intact protein LC-MS (**Section 7.8.2.2**). As shown in **Figure 5.41**, eRF1$_{ND}$ undergoes near-complete +30 Da modification when co-produced with JMJD4 Rosetta(DE3) pLysS cells in 2TY media. Notably, the +30 Da modification was not observed to such a large extent in the preliminary eRF1$_{ND}$/JMJD4 co-expression trials in M9 minimal media. In fact, it is difficult to confirm that eRF1$_{ND}$-30 is present in the preliminary co-expression experiments (if at all) due to the high signal-to-noise in the deconvoluted mass spectra (**Figure 5.6**). However, given the large enzyme:substrate ratio required in the isolated protein assays and generally poor expression of JMJD4 compared with eRF1, it is likely that JMJD4 does not express well in M9 minimal media. Poor JMJD4 expression in M9 minimal media would account for the limited eRF1$_{ND}$-30 formed upon co-expression. Overall, co-expression of eRF1$_{ND}$ with JMJD4 in 2TY media provides evidence for the possibility that JMJD4-catalysed γ-lactam formation can occur in bacterial cells.

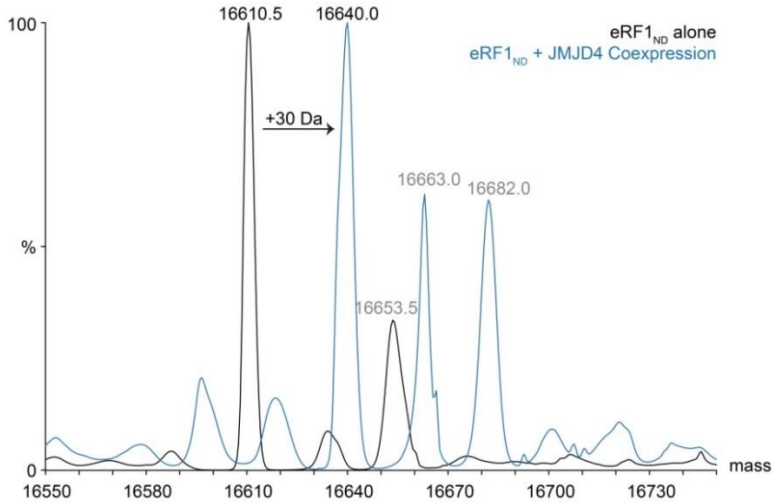


Figure 5.41. Deconvoluted intact protein mass spectra of eRF1$_{ND}$ following expression alone (black) or co-expression with JMJD4 (blue) in bacterial cells. Formation of the eRF1$_{ND}$-30 is possible in bacterial cells, as evidenced by the near complete conversion to the +30 Da modified state. Proteins were analysed using a Xevo GS-2 QTof mass spectrometer (**Section 7.8.2.2**).

5.8.2 Co-expression of eRF1$_{FL}$ and JMJD4

To further investigate the possibility of γ-lactam formation in bacterial cells, co-expression experiments were repeated as described in **Section 5.8.1** with eRF1$_{FL}$ to create a more "physiologically-relevant" environment, despite the overall overexpression conditions. LC-MS analyses of eRF1$_{FL}$ following co-expression with JMJD4 provide further evidence for JMJD4-catalysed γ-lactam formation in bacterial cells (**Figure 5.42**). Overall, the co-expression experiments suggest that the +30 Da modification may well also be present in human cells.

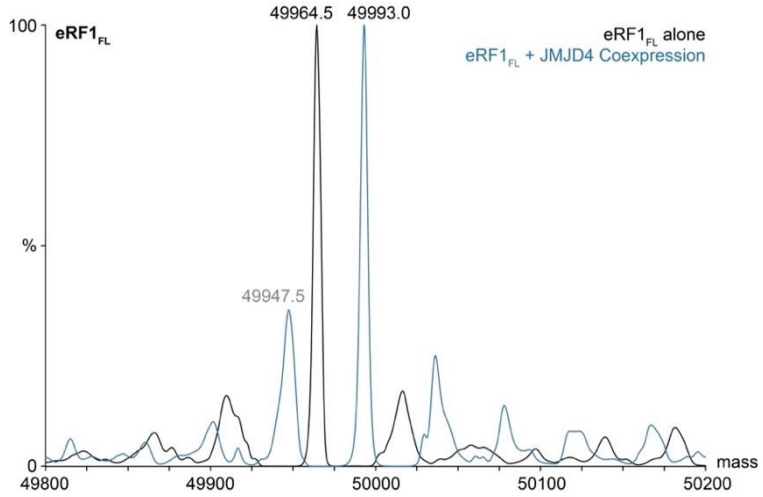


Figure 5.42. Deconvoluted intact protein mass spectra of eRF1$_{FL}$ following expression alone (black) or co-expression with JMJD4 (blue) in bacterial cells. Formation of the eRF1$_{FL}$-30 is possible in bacterial cells, as evidenced by the near complete conversion to the +30 Da modified state. Proteins were analysed using a Xevo GS-2 QTof mass spectrometer (**Section 7.8.2.2**).

5.8.3 Co-expression offers a better alternative for the isolation of eRF1$_{ND}$-30

In addition to revealing that JMJD4-catalysed γ-lactam formation can occur in bacterial cells, the co-expression experiments also offer insights into a better alternative for the isolation of unlabelled eRF1$_{ND}$-30. Rather than performing large scale isolated protein assays (**Sections 5.3.2 and 5.7.2.3.1**), eRF1$_{ND}$ can be co-expressed with JMJD4 to yield near complete conversion to eRF1$_{ND}$-30. The protein can then be purified using standard Ni(II)-affinity chromatography and SEC procedures to isolate eRF1$_{ND}$-30 from JMJD4 (**Section 5.3.2**). Co-expression provides a much more time efficient and cost effective strategy than isolated protein assays, which require time-consuming, independent expression and purification of large volumes of overexpressing JMJD4 bacterial cells due to poor yields. Because the extent of hydroxylation cannot be controlled by co-expression, isolated protein assays remain the most effective method for eRF1$_{ND}$-OH isolation and purification.

5.9 Investigation of the +30 Da modification in human cells

Given that JMJD4-catalysed γ-lactam formation is possible in bacterial cells, it was considered important to identify whether this unprecedented modification occurs in human cells. The presence of the +30 Da modification in human cells would verify γ-lactam formation as a novel PTM found in humans. Furthermore, given the large chemical shift changes induced by the +30 Da modification, γ-lactam formation could play an important role in translation termination.

MS/MS datasets arising from protease digests of previous siRNA knockdown and overexpression JMJD4 experiments in various human cell lines were manually interrogated to search for a +30 Da-containing eRF1 peptide fragment[133]. **Table 5.5** outlines the MS experiments and respective masses that were manually interrogated. Despite efforts to identify a +30 Da-containing eRF1 peptide fragment, no such fragment was found in these analyses. These datasets were also interrogated using PEAKS software (in collaboration with Dr Roman

Fischer, Target Discovery Institute, University of Oxford) by creating a new PTM corresponding to the mass of the oxidised γ-lactam structure (**Figure 5.43**). However, the PEAKS software was also unable to identify the +30 Da-containing eRF1 peptide fragment.

Table 5.5. Summary of the targeted search of unmodified, hydroxylated, and +30 Da modified peptides in human cells.

Experiment	**Description**	
PF139	Protease: ArgC Cell line: HEK 293T Antibody: α-eRF1 (*endogenous* eRF1)	
Fragments Searched	**Mass**	**Observed?**
VAKMLADEFGTASNIKSR	$(M+2H)^{2+}$ = 969.51	Relatively poor coverage of the *wild type* peptide.
VAKMLADEFGTASNIK^{OH}SR	$(M+2H)^{2+}$ = 977.51	Very good coverage of the *hydroxylated* peptide.
VAKMLADEFGTASNIK^{+30}SR	$(M+2H)^{2+}$ = 984.51	No fragmentation pattern identified for *+30 Da* peptide with no missed cleavage.
VAKMLADEFGTASNIK^{+30}SRVNR	$(M+2H)^{2+}$ = 1169.12 $(M+3H)^{3+}$ = 779.75	No fragmentation pattern identified for *+30 Da* peptide with 1 missed cleavage.
Experiment	**Description**	
MLC46	Protease: Trypsin Cell line: HEK 293T Transfection: V5 eRF1 + HA eRF3 + HA-JMJD4 Antibody: α-V5 (exogenous eRF1 and exogenous JMJD4)	
Fragments Searched	**Mass**	**Observed?**
MLADEFGTASNIK	$(M+2H)^{2+}$ = 698.84	Good coverage of the *wildtype* peptide.
MLADEFGTASNIK^{OH}	$(M+2H)^{2+}$ = 706.84	Very good coverage of the *hydroxylated* peptide.
MLADEFGTASNIK^{+30}	$(M+2H)^{2+}$ = 713.84	No fragmentation pattern identified for *+30 Da* peptide with no missed cleavage.
MLADEFGTASNIK^{+30}SR	$(M+2H)^{2+}$ = 828.41	No fragmentation pattern identified for *+30 Da* peptide with 1 missed cleavage.
MLADEFGTASNIK $^{+30}$SRVNR	$(M+2H)^{2+}$ = 1169.12 $(M+3H)^{3+}$ = 779.75	No fragmentation pattern identified for *+30 Da* peptide with 2 missed cleavages.
Experiment	**Description**	
MLC45	Protease: Trypsin Cell line: HEK 293T Transfection: V5-eRF1 + HA-eRF3 Antibody: α-V5 (*exogenous* eRF1, no JMJD4 overexpression)	
Fragments Searched	**Mass**	**Observed?**
MLADEFGTASNIK	$(M+2H)^{2+}$ = 698.84	Good coverage of the *wildtype* peptide.
MLADEFGTASNIK^{OH}	$(M+2H)^{2+}$ = 706.84	No fragmentation pattern identified for *hydroxylated* peptide.
MLADEFGTASNIK^{+30}	$(M+2H)^{2+}$ = 713.84	No fragmentation pattern identified for *+30 Da* protein with no missed cleavage.
MLADEFGTASNIK^{+30}SR	$(M+2H)^{2+}$ = 828.41	No fragmentation pattern identified for *+30 Da* peptide with 1 missed cleavage.
MLADEFGTASNIK^{+30}SRVNR	$(M+2H)^{2+}$ = 1169.12 $(M+3H)^{3+}$ = 779.75	No fragmentation pattern identified for *+30 Da* peptide with 2 missed cleavages.

Experiment	Description	
PF282	Protease: Trypsin Cell line: HeLa Description: siControl (no JMJD4 knockdown) + HA-JMJD4 rescue Antibody: α-eRF1 (*endogenous* eRF1, endogenous + exogenous JMJD4) SILAC labelling*	
Fragments Searched	**Mass**	**Observed?**
MLADEFGTASNIK	$(M+2H)^{2+}$ = 701.85	No fragmentation pattern identified for *wildtype* peptide.
MLADEFGTASNIK^{OH}	$(M+2H)^{2+}$ = 709.85	Good coverage of the *hydroxylated* peptide.
MLADEFGTASNIK^{+30}	$(M+2H)^{2+}$ = 716.89	No fragmentation pattern identified for *+30 Da* protein with no missed cleavage.
MLADEFGTASNIK^{+30}SR	$(M+2H)^{2+}$ = 838.46	No fragmentation pattern identified for *+30 Da* peptide with 1 missed cleavage.
Experiment	**Description**	
PF279	Protease: Trypsin Cell line: HeLa Description: siJMJD4 (JMJD4 knockdown) + HA-JMJD4 rescue Antibody: α-eRF1 (*endogenous* eRF1, exogenous JMJD4) SILAC labelling*	
Fragments Searched	**Mass**	**Observed?**
MLADEFGTASNIK	$(M+2H)^{2+}$ = 701.85	Good coverage of the *wildtype* peptide.
MLADEFGTASNIK^{OH}	$(M+2H)^{2+}$ = 709.85	Good coverage of the *hydroxylated* peptide.
MLADEFGTASNIK^{+30}	$(M+2H)^{2+}$ = 716.89	No fragmentation pattern identified for *+30 Da* protein with no missed cleavage.
MLADEFGTASNIK^{+30}SR	$(M+2H)^{2+}$ = 838.46	No fragmentation pattern identified for *+30 Da* peptide with 1 missed cleavage.
Experiment	**Description**	
PF280	Protease: Trypsin Cell line: HeLa Description: siControl + empty vector rescue Antibody: α-eRF1 (*endogenous* eRF1, endogenous JMJD4) SILAC labelling*	
Fragments Searched	**Mass**	**Observed?**
MLADEFGTASNIK	$(M+2H)^{2+}$ = 701.85	Good coverage of the *wildtype* peptide.
MLADEFGTASNIK^{OH}	$(M+2H)^{2+}$ = 709.85	Good coverage of the *hydroxylated* peptide.
MLADEFGTASNIK^{+30}	$(M+2H)^{2+}$ = 716.89	No fragmentation pattern identified for *+30 Da* protein with no missed cleavage.
MLADEFGTASNIK^{+30}SR	$(M+2H)^{2+}$ = 838.46	No fragmentation pattern identified for *+30 Da* peptide with 1 missed cleavage.

***** *SILAC label* - $^{13}C_6$-Lysine (+6 Da to every lysine in the sequence; in the case of the peptide of interest, there is only one lysine, so add 6 Da to the theoretical peptide mass)**.**
Experiments were performed by Dr Tianshu (Penny) Feng, published in Feng *et al.* (2016)[133].
Theoretical MS and MS/MS fragmentations masses were generated using the *Fragment Ion Calculator* (http://db.systemsbiology.net:8080/proteomicsToolkit/FragIonServlet.html).

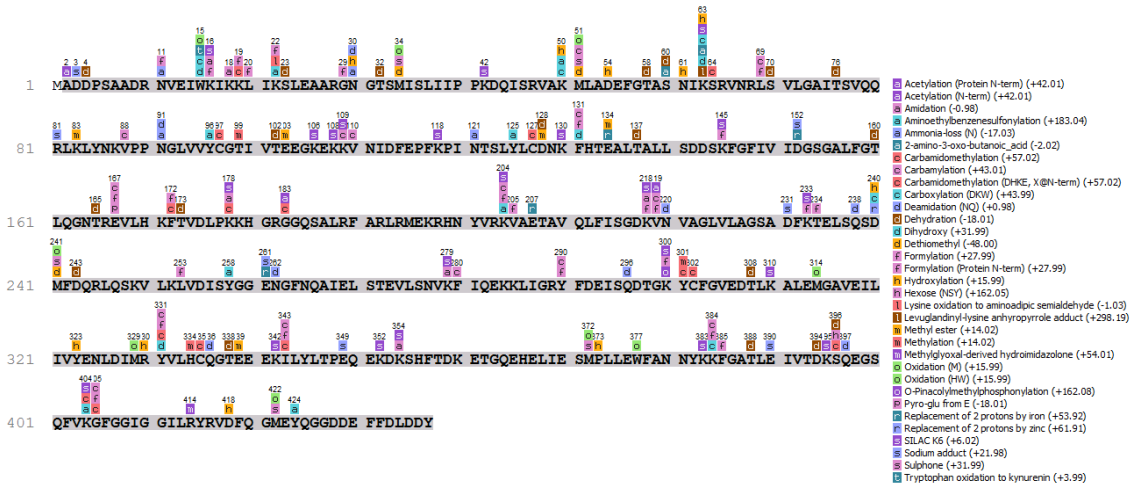


Figure 5.43. PEAKS analyses of the experiments outlined in Table 5.5 identify no +30 Da peptide-containing fragments. Figure kindly provided by Dr Roman Fischer (TDI, University of Oxford).

There are several possible reasons as to why the oxidised γ-lactam modification could not be identified in the above cellular experiments including:

(1) The modification is present, but at a very low abundance that cannot be detected;

(2) There was technical error during the execution of these experiments;

(3) Further modification occurs; and

(4) Given the destabilising effect of the modification on $eRF1_{ND}$ (**Section 5.6.2**), it is possible that γ-lactam formation is a signal for degradation, and therefore cannot be detected because it has been degraded.

For these reasons, JMJD4-catalysed γ-lactam formation of eRF1 cannot be ruled out as a physiologically-relevant PTM.

5.10 Discussion and future perspectives

The biochemical and biophysical results presented in this chapter reveal that JMJD4, a C4 lysyl hydroxylase, catalyses a previously unreported PTM, i.e. oxidative γ-lactam formation occurring between residues K63 and S64 of eRF1. Recombinant protein assays were optimised to isolate and purify $eRF1_{ND}$-OH and $eRF1_{ND}$-30, which enabled the study of these proteins and the ultimate characterisation of the +30 Da modification by NMR spectroscopy, tandem MS, and other biophysical techniques.

5.10.1 Effect of hydroxylation and +30 Da modification on $eRF1_{ND}$

Previous work has reported that JMJD4-catalysed hydroxylation of eRF1 improves translation termination efficiency[133]. One of the major aims of this chapter was to gain insights into the mechanism by which hydroxylation mediates improved termination. Previous structure determination of $eRF1_{ND}$ by protein NMR spectroscopy encouraged structural, dynamics, and rRNA binding studies of $eRF1_{ND}$, $eRF1_{ND}$-OH, and $eRF1_{ND}$-30[263,270]. Overall, NMR studies indicated that hydroxylation had little-to-no effect on $eRF1_{ND}$ structure, dynamics, and rRNA binding, as evidenced by the limited chemical shift perturbations induced by hydroxylation in the 2D and 3D spectra. Therefore, it is unlikely that JMJD4-catalysed hydroxylation mediates translation termination by inducing a global structural change or affecting the dynamics of the NIKS motif.

Recently, higher resolution cryo-EM studies of the human 80S ribosome bound in complex with eRF1 decoding a UAA(A) stop codon (3.8 Å)[285] and mammalian ribosomal complexes in complex with eRF1 and respectively bound to all three stop codons (3.5- 3.8 Å)[286] have revealed key interactions between the eRF1 NIKS, GTS, and YxCxxxF motifs with mRNA and ribosomal RNA (rRNA). Notably, both studies indicated that only the uridine in the first position of the stop codon on the mRNA transcript can form hydrogen bonding interactions

with K63 of the NIKS motif[285,286], offering major insights into the role of the NIKS motif and K63 in translation termination (**Figure 5.44**). While hydroxylation was not observed at 3.8 Å resolution, it was proposed that, given JMJD4-catalysed hydroxylation of K63 increases translational termination efficiency, additional hydrogen bonding interactions introduced by hydroxylation likely improve stop codon discrimination at position 1[285]. Overall, the structural studies reveal specificity for the invariant uridine in the first stop codon position due to interactions with the eRF1 NIKS region and provide the first direct structural evidence for the role of the NIKS region in stop codon recognition[285,286]. Thus, the proposal that C4 hydroxylation can improve base discrimination likely accounts for the observed improved translation termination efficiency. These structures also suggest that stop codon mRNA binding studies by NMR would likely provide greater insights than rRNA binding studies.

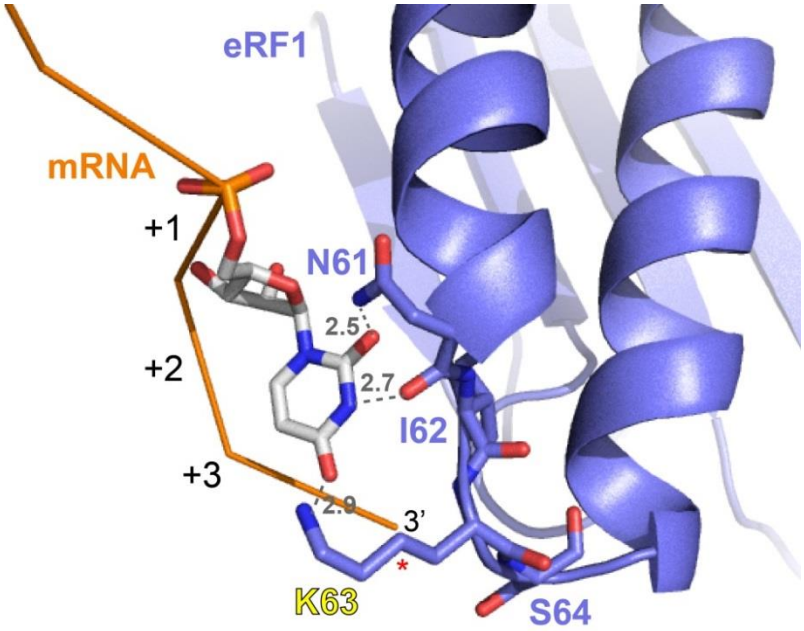


Figure 5.44. View from a cryo-EM structure of a mammalian ribosomal structure in complex with eRF1 (purple cartoon) and bound to an mRNA transcript (orange ribbon) containing the UAG stop codon (PBD ID: 3JAH)[286]. The invariant uridine, located in position 1 of all stop codons, is shown in sticks. Hydrogen bonding interactions between uridine and the amine side chain of K63, carbonyl main chain of I63, and amide side chain of N61 form the key interactions that facilitate eRF1:uridine base stop codon binding specificity. Red star denotes the JMJD4-catalysed hydroxylation site of K63 (yellow label). Dotted, grey lines represent H-bond distances in Å.

While hydroxylation had little effect on $eRF1_{ND}$ structure, oxidative γ-lactam formation induced fairly large, global chemical shift perturbations as observed by protein NMR, indicative of structural and/or conformational changes (**Section 5.6.3.5**). While γ-lactam formation had no significant effects on protein dynamics or rRNA binding in preliminary studies (**Sections 5.6.3.6** and **5.6.3.7**), the structural effect will almost certainly effect translation termination, given that K63 and S64 have been previously shown to be important in stop codon recognition[134,135]. Furthermore, the recent cryo-EM structures of the stalled mammalian ribosome termination complex with various stop codons show that the amine side chain of K63 is within hydrogen bonding distance of the uracil in the first stop codon position[285,286] (**Figure 5.44**). Certainly, γ-lactam formation occurring between K63 and S64 would induce large conformational constraints on the K63 side chain and likely cause more extensive structural changes as indicated by the 2D and 3D protein NMR studies. The pursuit of high-resolution cryo-EM structures (or use of other biophysical techniques) with eRF1 containing the oxidised γ-lactam would provide great insights into the effect of this modification on stop codon recognition and mRNA binding.

5.10.2 Isolation of a peptide for NMR characterisation

Ultimately, the characterisation of the +30 Da modification was identified by peptide NMR studies. Due to the incomplete backbone assignment of $eRF1_{ND}$-30 by protein NMR (including of the target K63 and S64 residues), significant efforts were made (by Owain Jones, Biochemistry Part II under my supervision)[278] to identify a peptide substrate for simplified NMR analyses. While the identification of a linear eRF1 peptide as a hydroxylation substrate for JMJD4 had previously been unsuccessful (Dr Sarah E Wilkins, University of Oxford, personal communication)[133], a linear peptide substrate was identified in this study. However, it was only able to undergo hydroxylation, even under the saturating, optimised +30 Da conditions. The pursuit of a cyclic peptide substrate, which was expected and reported to more closely mimic the helix-loop-helix architecture than linear peptides[133], was also undertaken,

though the cyclic peptide similarly underwent only a single hydroxylation. The inability of JMJD4 to catalyse γ-lactam formation on the peptide substrates designed in this study suggests that additional enzyme:substrate contacts distant from the NIKS region were likely required.

Because $eRF1_{ND}$ was the only efficient and reliable substrate for JMJD4-catalysed γ-lactam formation, a strategy was designed to isolate an oxidised γ-lactam-containing $eRF1_{ND}$ tryptic peptide by HPLC for NMR characterisation. While this method successfully provided suitable peptide purity and yield for NMR analyses, the method was extremely time consuming and required a substantial amount of starting material (~75 mg JMJD4; ~24 mg $eRF1_{ND}$) due to protein loss at each step of the digestion and purification process (e.g. gel filtration, HPLC purification, and incomplete tryptic digestion). While $eRF1_{ND}$ expresses well in *E. coli*, only ~2.5 mg of JMJD4 can be successfully purified from 1 litre of 2TY media (thus, ~30 L of bacterial growth cultures were required to isolate enough JMJD4 starting material for one experiment). While this proved to be a successful strategy, later studies indicated that co-expression of $eRF1_{ND}$ with JMJD4 in 2TY media yields near complete conversion to $eRF1_{ND}$-30 in bacterial cells. This strategy was initially not employed due to the incomplete hydroxylation observed during the co-expression trials in M9 minimal media. Given the already limited yields of JMJD4 obtained from bacterial cultures in 2TY media, the differences in the extent of γ-lactam formation can likely be accounted for by even poorer expression of JMJD4 in M9 minimal media than 2TY. Overall, future studies requiring an $eRF1_{ND}$-30 tryptic fragment should employ the more time efficient and material-friendly co-expression method in 2TY media prior to tryptic digestion and HPLC peptide purification.

5.10.3 Localisation and identity of the +30 Da modification

High resolution MS/MS analyses were initially performed to localise the +30 Da modification and gain insights into the identity of the modification using elemental composition analyses. Importantly, the NMR characterisation of the +30 Da modification as γ-lactam formation was

consistent with the MS data presented in this chapter. MS/MS analyses indicated that the +30 Da modification was localised solely to K63, as evidenced by the presence of the unmodified y2 (SR) and +30 Da modified y3 (K^{+30}SR) ions (**Section 5.4.1**). However, it is likely that the γ-lactam species exists as the α-hydroxyketone species under the collision-induced dissociation ionisation conditions, which likely causes opening of the lactam by hydrolysis of the bond formed between the K63 γ-carbon and S64 amide nitrogen[278]. The complementary NMR data indicate that γ-lactam formation between the oxidised K63 side chain and S64 amide nitrogen is apparently favoured in solution (given the absence of a ketone ^{13}C peak between 200-220 ppm in the ^{1}H-^{13}C HMBC spectrum). Finally, the α-hydroxyketone and oxidative γ-lactam species are both consistent with elemental composition analyses, which indicate that the +30 Da modification involves the incorporation of two oxygen atoms and loss of two hydrogen atoms (**Section 5.7.1**).

Overall, γ-lactam formation between K63 and S64 is the only structure that exclusively fits the combined NMR and MS data (**Figure 5.36** and **Figure 5.37**). As yet, exclusive characterisation of the modification would require comparison with a synthetic, small molecule standard, such as the one shown in **Figure 5.45**. While preparation of this standard would certainly be a synthetic challenge, comparison between the standard and modified peptide by NMR would validate the identity of the modification. Notably, JMJD4-catalysed γ-lactam formation introduces new stereocentres at the C4 and C5 positions of the K63 side chain, which remain to be characterised (potentially by comparison with synthetic standards using NMR spectroscopy and/or amino acid analysis). Finally, the identification of the +30 Da modification may provide guidance on the backbone assignment of the original protein NMR spectra of $eRF1_{ND}$-30, as a missing amide proton resonance for S64 would be anticipated due to γ-lactam formation.

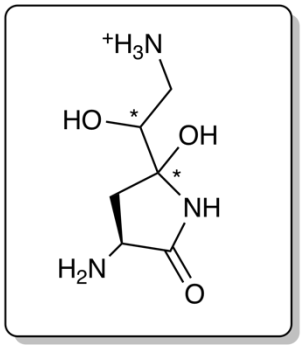


Figure 5.45. Structure of proposed synthetic standard required for NMR validation of JMJD4-catalysed +30 Da product. *Denotes that the stereochemistry at these carbon atoms is currently unknown. The (2*S*)-stereochemistry is assumed given that all naturally occurring proteins are made up of L-amino acids.

5.10.4 2OG oxygenase catalysis and ring formation

To date, human 2OG oxygenases have been reported to catalyse hydroxylation and related demethylation reactions. Therefore, the characterisation of JMJD4 as a γ-lactam forming oxygenase substantially expands the known chemistry catalysed by human 2OG oxygenases. However, 2OG oxygenases from other organisms have been reported to catalyse a wide range of oxidative reactions (e.g. halogenations, epimerisations, desaturations, and ring closures)[2,24]. In fact, oxidative ring formation is precedent by two enzymes involved in the biosynthesis of β-lactam drugs/antibiotics - isopenicillin N synthase (IPNS) and clavaminate synthase (CAS). IPNS is a 2OG oxygenase from fungus that catalyses oxidative double ring closure of a linear tripeptide (δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine; ACV) to yield isopenicillin N (**Figure 5.46A**)[287,288], a major step in the biosynthesis of a variety of penicillin and cephalosporin antibiotics. Penicillin and cephalosporin antibiotics are part of the β-lactam class of antibiotics, the most widely used antibiotics to treat a variety of bacterial infections[289]. IPNS catalyses both β-lactam and thiazolidine closure, thus yielding two new hetero-bonds (i.e. C-N and C-S bonds) in an early step in the biosynthetic pathway. IPNS contains a 2OG oxygenase-characteristic DSBH core fold and is thus structurally a member of the 2OG oxygenase superfamily[4]; however, it is a rare exception of a 2OG oxygenase that does not use 2OG as a co-substrate.

Oxidative cyclisation catalysed by a prokaryotic 2OG oxygenase CAS has also been reported in the clavulanic acid biosynthetic pathway from deoxyguanidino-proclavaminic acid (**Figure 5.46B**)[290]. In addition to desaturation, CAS, which is involved in three steps of the clavulanic acid biosynthetic pathway, catalyses hydroxylation and oxidative ring formation[290]. Clavulanic acid is a β-lactam drug that acts as an inhibitor of bacterial β-lactamase enzymes, which are major contributors to antibiotic resistance by hydrolysing the β-lactam of β-lactam antibiotics (e.g. penicillin), consequently inactivating them[289]. Therefore, clavulanic acid is oftentimes combined with penicillin-like antibiotics to overcome antibacterial resistance. Oxidative ring formation has also been observed for the promiscuous JmjC hydroxylase, FIH, under LC-MS conditions following hydroxylation of a serine residue, though the physiological relevance of this modification has yet to be determined[291].

As described, 2OG oxygenases play a major role in the biosynthesis of β-lactam drugs/antibiotics, including penicillins and cephalosporins. According to the World Health Organisation, antibiotic resistance is rising to dangerous levels throughout the world[292]. Given the rise in antibiotic resistance, major efforts are simultaneously being made to understand antibiotic resistance mechanisms and develop new antibiotics (ideally non-β-lactam antibiotics given several current modes of resistance including by the bacterial β-lactamase enzymes[289]). γ-Lactams are of interest due to their potential use as antibacterial compounds, and thus they have been studied from chemical synthetic and biosynthetic perspectives in large part for making biologically active molecules[293,294]. Notably, IPNS has been reported to catalyse γ-lactam formation (and desaturation) of an ACV tripeptide analogue containing a homocysteine residue in place of cysteine[295] (**Figure 5.46A**). Similarly, CAS has also been reported to accept a γ-lactam proclavaminic acid analogue as a substrate[296] (**Figure 5.46B**). IPNS and CAS exemplify 2OG oxygenases that catalyse a related type of chemistry to JMJD4. Thus, it would be of interest from a biosynthetic perspective to also investigate JMJD4 γ-lactam catalysis for the

synthesis of biologically active molecules and JMJD4 substrate specificity by testing known IPNS and CAS substrates and/or small molecule intermediates.

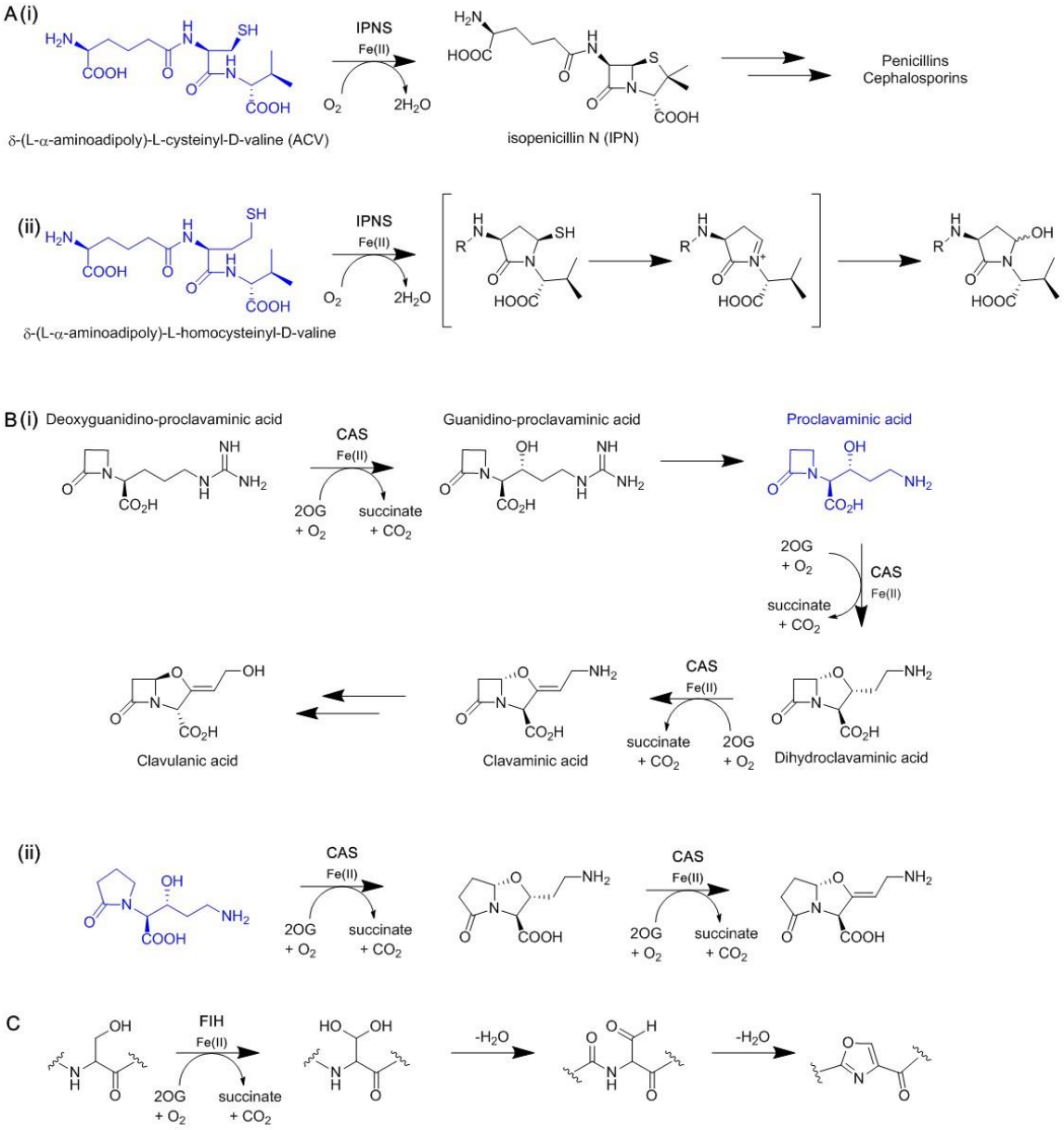


Figure 5.46. Oxidative ring formations as catalysed by 2OG oxygenases: (A) isopenicillin N synthase (IPNS) (B) clavaminate synthase (CAS) and (C) factor-inhibiting hypoxia inducible factor (FIH). (A) (i) IPNS catalyses oxidative double ring closure of a linear tripeptide (δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine, ACV; blue) to yield isopenicillin N, an early step in the biosynthesis of penicillin and cephalosporin antibiotics. (ii) IPNS has also been show to catalyse related oxidative double ring closure of an AVC analogue containing a homocysteine residue (blue), showing the potential of 2OG oxygenase catalysed γ-lactam formation. ***(B)*** (i) CAS catalyses hydroxylation and oxidative ring formation in the clavulanic acid biosynthetic pathway. (ii) CAS has also been reported to accept a γ-lactam analogue of proclavaminic acid as a substrate. ***(C)*** Reported oxidative ring formation following FIH-catalysed hydroxylation of serine.

5.10.5 Recommendations for future studies on JMJD4-catalysed γ-lactam formation

To date, γ-lactam formation of $eRF1_{ND}$ at K63/S64 has not been reported. Despite efforts to identify the modification in MS data from previous experiments performed in human cells (**Table 5.5**), no such +30 Da containing peptide was identified, as described in **Section 5.9**. Given that γ-lactam formation was found to have a destabilising effect on $eRF1_{ND}$ (**Section 5.6.2**), it is possible that γ-lactam formation is a signal for degradation, and thus cannot be detected. To rule out this possibility, it would be worthwhile for future cellular studies to be performed in the presence of a proteasomal inhibitor. It is also noteworthy that given the likely large structural changes induced by γ-lactam formation, the use of a specific anti-eRF1 antibody for the immunoprecipitation of endogenous eRF1 should be carried out with caution, since it is possible that the antibody may not interact with full-length eRF1-30 as efficiently or at all.

While there is evidence that γ-lactam formation is catalysed by JMJD4, future studies should focus on characterising the mechanism by which γ-lactam formation occurs. Proposed mechanisms for γ-lactam formation, described in **Section 5.7.4** and illustrated in **Figure 5.39**, require validation. The use of RapidFire MS with recombinant $eRF1_{ND}$, $eRF1_{ND}$-OH, and JMJD4 is recommended as a future strategy to monitor the kinetics of hydroxylation and γ-lactam formation, respectively, as it provides an appropriate, time efficient method of reaction monitoring of intact proteins (versus LC-MS, which requires 5-10 minutes/sample, or MS/MS analyses). Kinetic analyses and the identification of reaction intermediates would provide important insights into JMJD4 catalysis and understanding of the reaction mechanism.

In addition to studying the kinetics and mechanism of JMJD4-catalysed γ-lactam formation, it is of interest to investigate the effect of γ-lactam formation on translation termination. Previous studies investigating the effect of eRF1 hydroxylation on translation termination efficiency

were performed using a reconstituted *in vitro* translation system, which monitors peptide release from stalled pre-termination complexes[133]. It would thus be of interest to repeat these assays with $eRF1_{FL}$-OH and $eRF1_{FL}$-30, though the JMJD4-catalysed production and subsequent purification of $eRF1_{FL}$-OH and $eRF1_{FL}$-30 would certainly require optimisation.

As described in **Section 5.10.4**, there is biosynthetic potential for JMJD4 in the development of new antimicrobial, pharmaceutically active agents, given the precedent set by IPNS for 2OG oxygenases in the biosynthesis of β-lactam antibiotics (i.e. penicillins and cephalosporins) and γ-lactam formation. Given the potential for multifaceted roles of JMJD4-catalysed γ-lactam formation, future studies should primarily focus on (1) identifying the +30 Da modification in human cells, (2) performing biochemical assays aimed at determining the physiological consequences of previously unreported JMJD4-catalysed γ-lactam formation and (3) investigating the biosynthetic potential of JMJD4 by testing small molecules.

6 SUMMARY AND RECOMMENDATIONS FOR FUTURE WORK

6.1 Summary

The human 2OG oxygenases are a family of ~60 enzymes that catalyse hydroxylation and related demethylation reactions in fundamental cellular processes, including collagen biosynthesis, oxygen sensing, and protein biosynthesis. The JmjC subfamily of 2OG oxygenases is of particular interest due to their roles in transcriptional regulation, for example via the demethylation of methylated lysine residues in histone tails or the hydroxylation of HIF, a major regulator of the hypoxic response. Given these roles, it not only of interest to functionally assign the few remaining uncharacterised JmjC oxygenases, but also to further understand the molecular mechanisms underlying their biochemical and biological links. A major portion of the work described in this thesis has aimed to biochemically, structurally, and functionally assign a previously uncharacterised JmjC oxygenase, JMJD7. Given the associations between *JMJD7* gene mutations and major diseases, including cancer and autism, this work has further investigated the biochemical consequences of some reported mutations. In addition to the biochemical and structural studies of JMJD7, a large portion of this thesis has focused on understanding the effect(s) of JMJD4-catalysed modifications [hydroxylation and (previously unreported) γ-lactam formation] of eRF1, a release factor essential for translation termination.

The biochemical studies described in **Chapter 2** reveal that JMJD7 is indeed a catalytically active 2OG oxygenase. Using a substrate hunting approach that combined the identification of JMJD7-interacting proteins in human cells with peptide screening, two members of the TRAFAC family of GTPases, DRG1 and DRG2, were identified as JMJD7 substrates. Using

peptides as biochemical probes, NMR and mass spectrometry analyses revealed that JMJD7 catalyses the Fe(II)- and 2OG-dependent hydroxylation of highly conserved lysine residues in DRG1-2 (K22 and K21, respectively). Amino acid analyses of JMJD7 peptide substrates enabled the investigation of the regio- and stereo-chemistries of JMJD7-catalysed hydroxylation, demonstrating that JMJD7 catalyses a previously unreported type of PTM, (*3S*)-lysyl hydroxylation. Complementary cellular studies have validated DRG1 and DRG2 as genuine JMJD7 substrates. Overall, the work described in **Chapter 2** defines JMJD7 as a bona fide 2OG oxygenase catalysing (*3S*)-lysyl hydroxylation of specific residues in DRG1-2, laying the foundation for future studies to identify roles for JMJD7 in normal physiology and disease.

The biochemical assignment of JMJD7 defined in **Chapter 2** was supported by structural studies described in **Chapter 3**. Crystallographic analyses confirm JMJD7 as a JmjC domain-containing protein, as evidenced by characteristic features within its DSBH core fold, which is present in all 2OG oxygenases. Structural studies, combined with mutational and biophysical analyses, revealed that JMJD7 has a unique dimerisation mode, comprised of interactions involving both *N*- and *C*-terminal α-helical regions and disulphide bond formation between monomers. Overall, the structural studies in **Chapter 3** further corroborate the biochemical assignment of JMJD7 as a JmjC hydroxylase due to its closer structural similarities to the JmjC hydroxylases (rather than KDMs).

Given reported links between *JMJD7* mutations and diseases, such as cancer and autism, **Chapter 4** describes the investigation of the effects of clinically-relevant amino acid substitutions (R260C, M160V, T37E, and P272S) on JMJD7 (1) hydroxylase activity, (2) secondary structure, (3) structural stability, and (4) oligomerisation status. Of these JMJD7 variants, only the P272S amino acid substitution was identified as having a major effect on JMJD7 expression/structure, and consequently hydroxylase activity on DRG1 and DRG2. Bioinformatic and structural analyses indicate that the P272 residue, which is highly conserved in all JmjC hydroxylases, is important in maintaining the structural integrity of the DSBH

catalytic site by facilitating a tight turn between the DSBH βVI and βVII strands in all structurally characterised hydroxylases. Overall, the implications of *JMJD7* as a tumour suppressor gene, together with the apparent loss-of-function effect caused by the P272S substitution, suggest that that the catalytic activity of JMJD7 is likely important in normal physiology.

In many cases, including for JMJD7, the links between the biochemistry and biology remain to be elucidated. However, for the related JmjC hydroxylase, JMJD4, enzyme-catalysed hydroxylation of eRF1, a release factor responsible for accurate translation termination, has been shown to improve translation termination efficiency. **Chapter 5** describes the investigation of the effect(s) of JMJD4-catalysed hydroxylation on eRF1 structure, dynamics, and stability using a variety of biophysical techniques. Notably, this investigation led to the identification of a previously unreported JMJD4-catalysed ('+30 Da') modification, the characterisation of which became the primary focus of **Chapter 5**. Efforts to elucidate the chemical identity of this modification included the identification and isolation of a +30 Da-containing eRF1 peptide fragment. MS and peptide NMR studies identified the chemical identity of this modification as a γ-lactam, which the data suggests is formed via the K63 side chain and neighbouring S64 amide backbone. Protein NMR studies revealed that while hydroxylation had no apparent effect on eRF1 structure and dynamics, γ-lactam formation induced widespread chemical shift changes beyond the K63/S64 residues, suggestive of structural and/or conformational changes. While this modification has yet to be identified in human cells, bacterial JMJD4/eRF1 co-expression studies demonstrated that this previously unreported modification could be formed in bacterial cells. Given the large structural/conformational changes that this modification induces on recombinant eRF1, it is almost certain that γ-lactam formation would have major consequences on translation termination if identified as a bona fide human PTM. Beyond a potential role in translation, JMJD4-catalysed γ-lactam formation further expands the repertoire of modifications catalysed

by human 2OG oxygenases. Given that there is precedent for oxidative ring formation, the identification of an enzyme that can catalyse γ-lactam formation highlights a potential role for JMJD4 in the biocatalysis of biologically active γ-lactam molecules. The future investigation of JMJD4-catalysed γ-lactam formation in both human translation termination and biocatalysis is therefore of interest.

Overall, the work presented in this thesis has identified two previously unreported 2OG oxygenase-catalysed modifications – (3*S*)-lysyl hydroxylation and oxidative γ-lactam formation. The identification of the previously uncharacterised JmjC oxygenase, JMJD7, as specifically catalysing the hydroxylation of conserved lysine residues in two translation-associated GTPases is consistent with previously defined roles of JmjC oxygenases in protein biosynthesis. While JMJD4-catalysed γ-lactam formation requires validation in human cells, its identification highlights the potential catalytic diversity of the human 2OG oxygenases, as reflected in their lower organism counterparts. The biochemical and structural studies undertaken in this thesis provide a strong foundation for future functional studies of both JMJD7- and JMJD4-catalysed modifications.

6.2 Recommendations for future work

6.2.1 Functional studies of JMJD7

Recommendations for future experiments and studies on JMJD7 include:

(1) The investigation of the effect(s) of JMJD7-catalysed hydroxylation on DRG1-2 GTPase activity by ^{1}H NMR spectroscopy;

(2) The investigation of differences between full-length DRG1 and DRG2 substrate binding with JMJD7 (and JMJD7 clinically-relevant variants) using isothermal calorimetry and/or non-denaturing mass spectrometry;

(3) The pursuit of JMJD7:DRG1-2 complex structures to study differences in substrate binding, gain insights into structural features that enable regio- and stereo-specific (3*S*)-lysyl hydroxylation, and enable future therapeutic development (e.g. agonists).

(4) The biochemical characterisation of the JMJD7-PLA2G4B fusion protein, given recent reports on its oncogenic role in head and neck squamous cell carcinoma cell proliferation and survival[141].

(5) The identification and characterisation of other JMJD7 substrates using the described peptide screening approach (with caution, given likely structural requirements for enzyme:substrate recognition), so as to gain additional insights into the functional role of JMJD7.

6.2.2 Investigation of JMJD4-catalysed γ-lactam formation

Recommendations for future experiments and studies on JMJD4-catalysed γ-lactam formation include:

(1) The identification of the eRF1 γ-lactam product in human cells by overexpressing JMJD4 and immunopurifying endogenous eRF1 (initially to determine whether γ-lactam formation is possible in human cells).

(2) The analysis of the $eRF1_{ND}$-30 protein NMR data to re-assign the resonances corresponding to the protein backbone, considering the characterisation of the +30 Da modification as oxidative γ-lactam formation;

(3) The investigation of the effect of γ-lactam formation on translation termination using a previously described *in vitro* translation assay[133] and purified $eRF1_{FL}$, $eRF1_{FL}$-OH, and $eRF1_{FL}$-30 proteins;

(4) The investigation of JMJD4-catalysed hydroxylation and γ-lactam formation via enzyme kinetics to gain insights into the reaction mechanism using RapidFire MS and $eRF1_{ND}$/$eRF1_{FL}$ protein constructs;

(5) The synthesis of a small molecule standard resembling the K63-S64 γ-lactam product for definitive characterisation by NMR; and

(6) The investigation of JMJD4 biocatalytic activity via the screening of small molecules (including the IPNS and CAS substrates) as potential JMJD4 substrates and for the synthesis of biologically-active γ-lactam small molecules/antibiotics.

7 MATERIALS AND METHODS

7.1 General considerations

All microbiological techniques were carried out under standard sterile conditions. Pipette tips and solutions were autoclaved in a TouchClave® II autoclave (LTE Scientific) at 121°C for 20 minutes prior to use. All chemicals, reagents, and oligonucleotide primers were purchased from Sigma-Aldrich unless otherwise specified. Water was purified using an Elix® system (Millipore) and then by passing through a 0.22 µm filter by a Milli-Q® system (Millipore).

7.1.1 pH measurements

The pH of buffer solutions was measured using a Jenway pH meter (Jenway 3305) coupled to a glass/calomel electrode (Aldrich). The pH meter was calibrated immediately before each use in the 4.0-7.0 or 7.0-10.0 pH range (as required) with phthalate pH 4.0, phosphate pH 7.0, and borate pH 10.0 (Fischer Scientific). HCl, NaOH, acetic acid, or phosphoric acid were added to adjust pH as necessary. DCl or NaOD were used to adjust pH for Tris-D11 (Cambridge Isotope Laboratories) buffer used for NMR spectroscopy. The probe was stored in a saturated 4M KCl solution.

7.2 Microbiological Techniques

7.2.1 Preparation of oligonucleotide primers

Custom primers were prepared by making a 100 mM storage stock with Milli-Q purified water (MQ-H_2O). A 10 mM working stock in MQ-H_2O was subsequently prepared for ligation independent cloning (LIC) and/or site-directed mutagenesis (SDM). Primer sequences are outlined in **Table 7.1**.

Table 7.1. Primers used for ligation independent cloning (LIC) and site-directed mutagenesis (SDM).

Protein	Primer	Sequence (5'-to-3')
Primers for LIC		
$DRG1_{1-367}$	Forward	TACTTCCAATCCATGAGCAGCACCTTAGCTAAGAT
	Reverse	TATCCACCTTTACTGTCACTTCTTCACAATTTGAATGACA
$DRG2_{1-364}$	Forward	TACTTCCAATCCATGGGGATCTTAGAGAAGATCTC
	Reverse	TATCCACCTTTACTGTTACTTCTTCACGATCTGGATG
$JMJD7_{1-316}$	Forward	TACTTCCAATCCATGGCGGAGGCGGCTT
	Reverse	TATCCACCTTTACTGTCAGTCAAGGCCTGAAGCC
$JMJD7_{27-316}$	Forward	TACTTCCAATCCCTTGCTGTGCCCTACCTG
	Reverse	TATCCACCTTTACTGTCAGTCAAGGCCTGAAGCC
$JMJD7_{1-295}$	Forward	TACTTCCAATCCATGGCGGAGGCGGCTT
	Reverse	TATCCACCTTTACTGTCATTCCATGTCATACCAGAAATTCA
$JMJD7_{27-295}$	Forward	TACTTCCAATCCCTTGCTGTGCCCTACCTG
	Reverse	TATCCACCTTTACTGTCATTCCATGTCATACCAGAAATTCA
Primers for SDM		
JMJD7-H178A	Upper	GCAGTGACTTCTTTG**GCC**AAGGACCACTATGAG
	Lower	CTCATAGTGGTCCTT**GGC**CAAAGAAGTCACTGC
JMJD7-T37E	Upper	GGACAAACCCCCA**GAA**CCGCTCCACTTC
	Lower	GAAGTGGAGCGG**TTC**TGGGGGTTTGTCC
JMJD7-M160V	Upper*	GCCCTGGGAAAG**GTG**CCCGATGCTGTG
	Lower*	CACAGCATCGGG**CAC**CTTTCCCAGGGC
JMJD7-R260C	Upper	CCCAGGCCCTT**TGC**TGCACGGTG
	Lower	CACCGTGCA**GCA**AAGGGCCTGGG
JMJD7-P272S	Upper	GAGATGCTCTATCTG**TCG**GCTCTGTGGTTCCA
	Lower	TGGAACCACAGAGC**CGA**CAGATAGAGCATCTC
JMJD7-C47A	Upper	CGGGACTGGGTC**GCC**CCCAACAGGC
	Lower	GCCTGTTGGG**GGC**GACCCAGTCCCG
eRF1-S64P	Upper	GAACTGCATCTAACATTAAG**CCA**CGAGTAAACCGCCTTTC
	Lower	GAAAGGCGGTTTACTCG**TGG**CTTAATGTTAGATGCAGTTC

LIC overhang sequences are shown in blue. Start and stop codons are underlined. Alternations from native DNA sequences are **bold**.

7.2.2 Plasmids used in this thesis

The plasmids used in this thesis were either prepared by LIC, SDM, or provided by a collaborator as summarised in **Table 7.2**. pET28a(+) plasmids contained an *N*-terminal His_6-tag and thrombin protease cleavage site (MGSSHHHHHHSSGLVPR*GSH). pNIC28-Bsa4

plasmids contained *N*-terminal His_6-tag and tobacco etch virus (TEV) protease cleavage site (MHHHHHHSSGVDLGTENLYFQ*SM). pET23b(+) plasmids contained a *C*-terminal His_6-tag.

Table 7.2. Plasmids used in this study.

Plasmid	NCBI Gene ID	Residues	Antibiotic Resistance	Source
JMJD7 Plasmids				
pET28a(+)-JMJD7	100137047	1-316	Kanamycin	Dr W Ge
pNIC28-Bsa4-JMJD7$_{1-316}$	100137047	1-316	Kanamycin	LIC
pNIC28-Bsa4-JMJD7-C47A$_{1-316}$	100137047	1-316, C47A	Kanamycin	LIC
pNIC28-Bsa4-JMJD7$_{27-316}$	100137047	27-316	Kanamycin	LIC
pNIC28-Bsa4-JMJD7-C47A$_{27-316}$	100137047	27-316, C47A	Kanamycin	LIC
pNIC28-Bsa4-JMJD7$_{1-295}$	100137047	1-295	Kanamycin	LIC
pNIC28-Bsa4-JMJD7-C47A$_{1-295}$	100137047	1-295, C47A	Kanamycin	LIC
pNIC28-Bsa4-JMJD7$_{27-295}$	100137047	27-295	Kanamycin	LIC
pNIC28-Bsa4-JMJD7-C47A$_{27-295}$	100137047	27-295, C47A	Kanamycin	LIC
pET28a(+)-JMJD7-H178A	100137047	1-316, H178A	Kanamycin	SDM
pET28a(+)-JMJD7-T37E	100137047	1-316, T37E	Kanamycin	SDM
pET28a(+)-JMJD7-M160V	100137047	1-316, M160V	Kanamycin	SDM
pET28a(+)-JMJD7-R260C	100137047	1-316, R260C	Kanamycin	SDM
pET28a(+)-JMJD7-P272S	100137047	1-316, P272S	Kanamycin	SDM
DRG Plasmids				
pNIC28-Bsa4-DRG1	4733	1-367	Kanamycin	LIC
pNIC28-Bsa4-DRG2	1819	1-364	Kanamycin	LIC
eRF1 Plasmids				
pET23b(+)-eRF1$_{ND}$	2107	1-142	Ampicillin	Dr M Coleman
pET23b(+)-eRF1$_{FL}$	2107	1-437	Ampicillin	Dr M Coleman
pET23b(+)-eRF1$_{ND}$-S64P	2107	1-142, S64P	Ampicillin	SDM
JMJD4 Plasmids				
pET28a(+)-JMJD4	65094	47-447*	Kanamycin	Dr W Ge

*It is uncertain whether M1 or M47 is the site of translation initiation according to the UniProt curator (UniProtKB - Q9H9V9)[297].

7.2.3 Ligation independent cloning (LIC)

7.2.3.1 DNA amplification

DNA was amplified for LIC cloning by polymerase chain reaction (PCR) using the Q5® High-fidelity DNA polymerase (Catalogue number: M0491; New England BioLabs®, NEB) as per the manufacturer's protocol and with the modifications outlined in the reaction (**Table 7.3**) and thermocycler (**Table 7.4**) conditions. PCR reactions were run in a MJ Mini Personal Thermal Cycler (Bio-Rad). Amplified DNA was verified by agarose gel electrophoresis and subsequently purified (**Section 7.2.4**).

Table 7.3. Q5 High-fidelity DNA polymerase reaction setup.

Reagent	Stock	Volume (µL)
Q5 Reaction buffer	5x	5
dNTPs	10 mM	1
Forward primer (Sigma-Aldrich)	10 µM	1.25
Reverse primer (Sigma-Aldrich)	10 µM	1.25
Template DNA	Plasmid or cDNA*	1 - 2
Q5 GC Enhancer	5x	5
MQ-H_2O	-	9 - 10
Q5 High-fidelity DNA polymerase	2 U/mL	0.5
Total	-	25

* cDNA from U2-OS cells (kindly provided by Dr Louise Walport, University of Oxford) was used as a template for DRG1 and DRG2 gene amplification.

Table 7.4. Thermocycler conditions for DNA amplification.

Segment	Cycles	Temperature (°C)	Time (s)
Initial denaturation	1	98	30
Denaturation	40	98	10
Annealing*		55-58	30
Extension**		72	60
Final extension	1	72	5 min
Final hold	1	4°C	∞

* 55°C for JMJD7 and DRG2 constructs, 58°C for DRG1;
General rule: Annealing temperatures should be -5°C below primer T_m
** 20-30 s/kb

7.2.3.2 Vector preparation and cohesive end generation

The empty pNIC28-Bsa4 and pNH-TrxT vectors (kindly provided by the Structural Genomics Consortium, SGC) were digested with Bsa1 in NEB Buffer 3 for 2-3 hours at 37ºC according to the SGC protocol[162] outlined in **Table 7.5**. Note that Bsa1 digest releases the *sacB* gene, which is sensitive to sucrose. DNA was purified using a GeneJET Gel Extraction Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. The DNA was eluted in 50 µL of MQ-H_2O.

Table 7.5. Vector digestions.

Reagent	Stock	Volume (µL)
NEB Buffer 3	10x	10
BSA (NEB, 1 mg/ml)	10x	10
Plasmid	Concentration dependent	3 µg equivalent
Bsa1	90 units	3
MQ-H_2O	-	Up to 100 µL
Total	-	100

Table 7.6. Cohesive end generation of cut vectors and DNA inserts.

Reagent	Vector Reaction (µL)	Insert Reaction (µL)
Bsa1-digested plasmid	50	-
Insert amplified DNA	-	5
10x T4 DNA polymerase buffer (Novagen or NEB buffer 2)	10	1
dGTP (25 mM)	10	-
dCTP (25 mM)	-	1
BSA (NEB, 1 mg/ml)	1	1
DTT (100 mM)	5	0.5
10x T4 DNA polymerase (Novagen, LIC-qualified)	2.5	0.5
MQ-H_2O	21.5	1
Total	100	100

Cohesive ends of the Bsa1-treated vectors and amplified DNA (insert) were generated according to the modified SGC protocol[162] outlined in **Table 7.6**. Reactions were incubated for

30 minutes at 22ºC to generate sticky ends, then 20 minutes at 75ºC to inactive the enzyme. Thermocycler lid was preheated to 90ºC. 2 µL of the T4 DNA polymerase-treated inserts were incubated with 1 µL of the treated vector at room temperature for 10-30 minutes. The mixtures were transformed into XL10-Gold competent cells (**Section 7.2.10**) and plated onto agar plates containing 5% sucrose (selecting against uncut vectors).

7.2.4 Agarose gel electrophoresis

Amplified DNA PCR products were verified by agarose gel electrophoresis. 1% (w/v) agarose gels were prepared by dissolving 0.5 g of agarose in 50 mL of 1x Tris base, acetic acid, EDTA (TAE) buffer (**Table 7.7**) in a long-neck conical flask by microwaving carefully. 2.5 µL of SYBR SAFE (Invitrogen) was then swirled in, and the solution was poured into a cast, covered with aluminium foil, and allowed to set for 1 h. Samples were prepared by adding 5 µL of 6x DNA loading dye (NEB) to the PCR products (25 µL). Each 30 µL sample and the appropriate 100 bp or 1 kb DNA ladder (NEB) was pipetted into respective lanes. The gel was run in a Bio-Rad® tank filled with 1x TAE buffer at a constant 100V for 30-45 minutes. The band corresponding to the correct size of the amplification product was excised with a sterile razor, and the DNA was extracted using a GeneJET Gel Extraction Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. The DNA was eluted in 30 µL of MQ-H_2O. DNA was stored at -20°C.

Table 7.7. 50x Tris base, acetic acid, EDTA (TAE) buffer recipe.

Reagent	Per 0.5 Litre
Tris Base	121.14 g
Glacial acetic acid	28.55 mL
0.5 M EDTA pH 8.0	50 mL

7.2.5 Site-directed mutagenesis

Reactions for SDM were performed using the Q5 High-fidelity DNA polymerase (NEB) according to the manufacturer's protocol for the reaction setup (**Table 7.3**) and thermocycler

(**Table 7.8**) conditions kindly provided by Dr Azer Düzgün. PCR reactions were run in a MJ Mini Personal Thermal Cycler (Bio-Rad).

Table 7.8. **PCR thermocycler conditions for SDM.**

Segment	Cycles	Temperature (°C)	Time (min)
Initial denaturation	1	94	2
Denaturation	18	94	45 s
Annealing		55	1
Extension*		68	12
Final extension	1	68	20 min
Final hold	1	4	∞

* 1 min/kb of plasmid length

After the thermocycler had completed, 1 µL of *Dpn* I restriction enzyme (20 U/µL, NEB) was added to the reaction mixture and incubated for 3 hours at 37°C to digest the parental (non-mutated) supercoiled double-stranded DNA. Following incubation, 4 µL of the reaction mixture was immediately transformed into XL10-Gold competent cells (**Section 7.2.10**). Plasmids were purified using a GeneJET Plasmid Miniprep Kit (Thermo Scientific) as described in **Section 7.2.6** and sequenced by Source BioScience (Nottingham, UK) for verification.

7.2.6 Plasmid purification

Following transformation (**Section 7.2.10**) of LIC reaction, SDM reaction, or plasmid into XL10-Gold competent cells, a freshly transformed colony was picked and grown overnight in 100 mL of 2TY media supplemented with the appropriate antibiotic as described in **Section 7.3.4**. Plasmid DNA was purified from 5 mL of bacterial cell culture using a GeneJET Plasmid MiniPrep Kit (Thermo Fisher Scientific) as per the manufacturer's protocol. The purified plasmid DNA was eluted in 30 µL of sterile MQ-H_2O and stored at -20°C.

7.2.7 DNA concentration determination

Plasmid DNA concentrations were determined by measuring the absorbance at 260 nm (A_{260}) using a ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies). DNA purity was assessed by measuring the A_{260}/A_{280} ratio (1.7 – 2.0).

7.2.8 DNA Sequencing

All plasmid sequences were verified by Sanger DNA Sequencing carried out by Source BioScience (Nottingham, UK).

7.2.9 Competent cells

XL10-Gold® (Stratagene®) and Rosetta™(DE3)2 pLysS (Novagen) competent cells were prepared from commercial stocks. *E. coli* cells were streaked from commercial stocks onto an antibiotic-free LB-agar plate and grown overnight at 37°C. A single colony was used to inoculate 100 mL of 2TY media (antibiotic-free) and grown overnight at 37°C in a shaking incubator (180 rpm). 100 mL of sterile 2TY (antibiotic-free) media was inoculated with 200 µL of the starter culture and grown at 37°C in a shaking incubator (180 rpm) to an optical density at 600 nm (OD_{600}) of 0.6 – 0.8. Sterile 50 mL Falcon (centrifuge) tubes were pre-chilled, and 50 mL of cells were pelleted by centrifugation (4°C, 750 x g, 5 minutes). Cells were re-suspended in 25 mL of sterile, ice-cold 100 mM $CaCl_2$/20% glycerol solution (autoclaved prior to use) and incubated on ice for 20 minutes. Cell were again pelleted by centrifugation (4°C, 750 x g, 5 minutes) and re-suspended in 1 mL of sterile, ice-cold 100 mM $CaCl_2$/20% glycerol. Cells were aliquoted (100 µL) into sterile Eppendorf tubes, flash frozen in liquid nitrogen, and immediately stored at -80°C.

7.2.10 Bacterial transformations

Plasmid DNA was transformed into *E. coli* XL10-Gold (for plasmid purification following LIC or SDM) or Rosetta(DE3) pLysS / BL21(DE3) (for protein production) competent cells according to the following protocol. Bacterial cells (40 µL) were thawed and added to 1.0-1.5 µL of bacterial plasmid DNA (20-100 ng) or 4 µL of SDM reaction mixture. After resting for 5-10 minutes on ice, the bacteria were heat shocked at 42°C for 45 s (XL10-Gold) or 30 s (Rosetta), and then left to rest on ice for 10 minutes. 2TY growth media (100 µL, **Section 7.3.1**) was added to the bacteria, and solution was incubated at 37°C for 10-20 minutes (plasmid transformation) or 1 hour (LIC and SDM transformations). Following incubation, 100 µL of the bacteria solution was spread onto LB-agar plates with relevant antibiotic (final concentrations: 30 µg mL^{-1} kanamycin or 100 µg mL^{-1} ampicillin) and allowed to grow at 37°C overnight. Note that plates used for LIC cloning were made with 5% sucrose as described in **Section 7.2.11**.

7.2.10.1 Bacterial co-transformations

1.0 µL of each of the two plasmids with different antibiotic resistances were mixed with 40 µL of competent bacterial cells (e.g. pET23b(+)-eRF1$_{ND/FL}$ and pET28a(+)-JMJD4 in Rosetta(DE3) pLysS cells). Co-transformations were then carried out as described for single plasmid transformations (**Section 7.2.10**) with the transformed competent cells streaked onto agar plates containing both antibiotics (e.g. 30 µg mL^{-1} kanamycin and 100 µg mL^{-1} ampicillin; final concentrations).

7.2.11 Luria Broth (LB) agar plates

An LB agar solution was made in a glass Schott bottle using the recipe in **Table 7.9**. The solution was then autoclaved to dissolve the agar and sterilise the solutions. Upon cooling, yet prior to solidifying, appropriate antibiotics were added (final concentrations: kanamycin, 30 µg/mL; ampicillin, 100 µg/mL; chloramphenicol, 35 µg/mL). Note that for LIC cloning, agar

plates containing 5% (w/v) sucrose were made by preparing a 2x LB agar solution and adding to an equal volume of 10% (w/v) sucrose solution [sterile-filtered with a Filtropur 0.20 µm syringe filter (Sarstedt) due to heat sensitivity] prior to the addition of antibiotics. Plates were poured into petri dishes, left to solidify, then inverted and stored at 4ºC.

Table 7.9. LB agar recipe.

Reagent	Per 200 mL (1x)
Agar (1.5% final)	3 g
Bactotryptone	2 g
Yeast extract	1 g
NaCl	2 g

7.2.12 Antibiotics

Antibiotics used in this study, chloramphenicol, ampicillin sodium, and kanamycin sulphate (1:1), were obtained from Apollo Scientific Limited. 100x antibiotic stocks (defined as kanamycin, 30 mg/mL in MQ-H_2O; ampicillin, 100 mg/mL in MQ-H_2O; chloramphenicol, 35 mg/mL in ethanol) were prepared and filter sterilised with a Filtropur 0.45 µm syringe filter (Sarstedt). Antibiotic stocks were stored at -20ºC. Kanamycin and ampicillin stocks were allowed only one freeze-thaw cycle prior to use.

7.3 Protein expression

7.3.1 2TY media

2TY bacterial growth medium (**Table 7.10**) was either prepared individually or kindly prepared and autoclaved by Mr Herminio Manso Jubier.

Table 7.10. 2TY media recipe.

Reagent	Per 1 L
Bactotryptone	16 g
Yeast extract	10 g
NaCl	5 g

7.3.2 M9 minimal growth media

M9 minimal media, supplemented with ^{15}N- and/or ^{13}C isotopic labelling, was used for protein production as required for NMR experiments. A 10x M9 minimal salt solution (**Table 7.11**) was made up to 500 mL MQ-H_2O before diluting to 1x stocks (1 L final volume) and autoclaving. Ammonium chloride (^{15}N, 99%, Cambridge Isotope Laboratories or Cortecnet) was used for ^{15}N-labelling.

Table 7.11. 10x M9 minimal salts.

Reagent	Per 500 mL
Na_2HPO_4	30.0 g
KH_2PO_4	15.0 g
NH_4Cl (or $^{15}NH_4Cl$)	5.0 g
NaCl	2.5 g

The solutions listed in **Table 7.12** were sterile-filtered (due to heat sensitivity) using a Minisart 0.20 µm syringe filter (Sartorius Stedim Biotech) and then added to each 1x stock of 1 L M9 salt solution after autoclaving. D-glucose-$^{13}C6$ (99% ^{13}C, Cortecnet) was used for ^{13}C-labelling.

Table 7.12. Supplements to M9 minimal media

Reagent	Per 1 L
1M $MgSO_4 \cdot 7H_2O$	100 mL
20% (w/v) D-$C_6H_{12}O_6$ (or D-$C_6H_{12}O_6$-$^{13}C6$)	10 mL
0.5% (w/v) Thiamine HCl (Vitamin B_1)	10 µL
1 M $CaCl_2 \cdot H2O$	5 µL

7.3.3 Glycerol stock preparation

Glycerol stocks were made by gently mixing 1 mL of starter culture with 0.5 mL of sterile glycerol in a sterile cryotube. Glycerol stocks were stored at -80°C.

7.3.4 Starter cultures

Starter cultures containing 100 mL of 2TY growth media, supplemented with appropriate antibiotics (30 μg/mL kanamycin, 100 μg/mL ampicillin, and/or 35 μg/mL chloramphenicol) were inoculated with either a bacterial colony from a freshly transformed plate (preferred method), or small amount of a glycerol stock containing the plasmid of interest, and grown overnight at 37°C in a shaking incubator (180 rpm).

7.3.5 Large scale growths

Large scale cultures containing either 2TY (2 L or 7.2 – 12 L) were inoculated with a 1 to 100 volume of starter culture to total growth media. Note that large scale cultures containing M9 minimal growth media (5 L) were inoculated with a 2:100 volume of starter culture to total growth media with typical antibiotic concentrations halved to expedite growth (**Table 7.13**). The media was supplemented with appropriate sterile-filtered antibiotics and grown at 37°C, 180 rpm until an OD_{600} of 0.6 – 0.8 was reached. Induction was carried out at an optimised temperature (**Table 7.13**) overnight after the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG), prepared as a sterile-filtered 1 M stock (note IPTG solutions were either freshly prepared or stored at -20°C and allowed one freeze-thaw cycle). Cells were harvested by centrifugation (11,325 x g, 4°C, 10 min) and stored at -80°C. Note that small scale cultures (100 mL 2TY media) for expression trials were grown in the same manner with overnight induction temperatures at 18, 20 and/or 30°C. Optimised recombinant protein expression conditions are summarised in **Table 7.13**.

Table 7.13.Optimised recombinant protein expression conditions.

Protein*	Media	Antibiotics (μg/mL) **	Induction Temp. (°C)
$eRF1_{ND/FL}$-His_6	M9 minimal 2TY	Kan (15) / Chlor (17.5) Kan (30) / Chlor (35)	30
$eRF1_{ND/FL}$-His_6 His_6-JMJD4 Co-expression	M9 minimal 2TY	Kan (15) / Amp (50) / Chlor (17.5) Kan (15) / Amp (50) / Chlor (17.5)	20
His_6-JMJD4	2TY	Kan (15) / Chlor (17.5)	18
His_6-JMJD7†	2TY	Kan (30) / Chlor (35)	18
His_6-DRG1 and His_6-DRG2	2TY	Kan (30) / Chlor (35)	18

* All proteins were expressed in Rosetta(DE3) pLysS cells, except $eRF1_{ND/FL}$-His_6, which was sometimes expressed in BL21(DE3).
** Kan = kanamycin, Amp = ampicillin; Chlor = chloramphenicol.
† Represents conditions for all JMJD7 constructs and variants.

7.4 Protein purification

7.4.1 General considerations

All solutions for protein purification were prepared as a 5x or 10x concentrated stocks (with respect to their final concentrations) using MQ-H_2O and sterile-filtered with a Filtropur 0.20 μm or 0.45 μm syringe filter (Sarstedt). Solutions were diluted to a 1x stock with MQ-H_2O prior to use.

7.4.2 Small scale purification with loose His resin

Cell pellets from 100 mL of bacterial growth culture were re-suspended in 2 mL ice-cold HEPES lysis buffer (**Table 7.14**) by pipetting. Re-suspended cells were transferred to a clean, disposable VWR 5 mL freestanding mailing tubes and lysed on ice by sonication (3x 5 s with 20 s rest; Sonics Vibracell Soniprep VCX 500/750 attached to a microtip). Samples were transferred to 1.5 mL Eppendorf tubes and cell debris pelleted by centrifugation (18,000 x g, 30 min, 4°C). 10 μL of the supernatant was transferred to a fresh tube for SDS-PAGE analysis.

Loose Ni(II) Sepharose™ high-performance resin (GE Healthcare; 400 µL in a 50% slurry in ethanol) was transferred to a 15 mL Falcon tube and pre-charged with a 50 mM Ni(II) sulphate·7H_2O charge buffer. The resin was then pre-equilibrated with HEPES binding buffer (**Table 7.14**) to make a 400 µL 50% slurry in binding buffer. Protein samples were added to the 50% charged His resin slurry and incubated with rocking at 4°C for 1 hour. Loose resin was then pelleted by centrifugation (3,000 x g, 1 min, 4°C). The resin was washed once with 5-10 mL HEPES binding buffer then twice with 5-10 mL HEPES wash buffer (**Table 7.14**) with centrifugation steps in between to remove supernatant. The protein was eluted from the resin by incubating with 0.5 mL HEPES elution buffer for 10-15 minutes on ice with occasional gentle mixing. The loose resin was again pelleted by centrifugation and the supernatant transferred to a 1.5 mL Eppendorf tube. 10 µL of the purified protein was transferred to a fresh tube to assess purity by SDS-PAGE. Protein concentrations were determined as described in **Section 7.4.5**. When appropriate, samples were concentrated using an Amicon Ultra 0.5 mL centrifugal filter (Millipore) with an appropriately selected molecular weight cut off (MWCO) according to the manufacturer's protocol, and samples were flash frozen in liquid nitrogen in 10-50 µL aliquots and stored at -80°C.

7.4.3 Nickel(II) affinity chromatography

Cell pellets (~30g) containing the protein of interest were re-suspended 120 mL of ice-cold HEPES lysis buffer (**Table 7.14**), and gently stirred for 10 minutes at room temperature followed by additional stirring for 10-20 minutes at 4°C until the cells were fully re-suspended. Cells were lysed on ice by sonication (Sonics Vibracell VC 500/750 Ultrasonic Cell Disruptor) at 60% amplitude with a pulse of 3 s followed by a rest of 7 s, for a total pulse time of 3.5 min. Re-suspension volume and sonication time were adjusted according to cell pellet mass.

The lysate supernatant was separated from the cell debris via centrifugation (58,545 x g, 20-25 min, 4°C), and filtered through a Filtropur 0.45 µm syringe filter (Sarstedt) before loading onto

a 5 mL HisTrap HP column (GE Healthcare) at flow rate of 1 mL/min. The 5 mL HisTrap HP column had been charged with a 50 mM Ni(II) sulphate·7H_2O charge buffer and pre-equilibrated with HEPES binding buffer (**Table 7.14**). All following protein purification steps took place at 4°C and at a flow rate of 1 mL/min using an AKTA FPLC system. After the lysate was loaded onto the column, the column was washed with 25 column volumes (CVs) of HEPES wash buffer (**Table 7.14**) to remove non-specifically bound proteins. The protein of interest was typically eluted from the column in one-step with HEPES elution buffer (**Table 7.14**) and the eluent was collected in 1 mL fractions. In the case of His_6-JMJD7 (including all constructs and variants), the proteins were eluted using a gradient of HEPES elution buffer in HEPES binding buffer to increase protein purity (**Table 7.15**). Following use, the HisTrap HP column was washed with MQ H_2O, stripped with 100 mM EDTA strip buffer, pH 8.0 and stored in 20% ethanol according to manufacturer's protocol. SDS-PAGE was used to identify proteins of interest via their molecular weight.

Note that a 1 mL HisTrap HP column (GE Healthcare) was used for cell pellets (5-10 g) arising from bacterial expression in 1-2 L of 2TY media. The protocol was scaled accordingly, except the wash step, which consisted of 120 CV of HEPES wash buffer, and eluent collection in 0.5 mL fractions.

Table 7.14.Protein affinity chromatography purification buffers

Buffer*	Components
HEPES lysis buffer, pH 7.5	50 mM HEPES, 500 mM NaCl, 5 mM imidazole, 1 mM PMSF, ~0.02 mg/mL DNase I (Roche), ~0.2 mg/mL lysozyme (Sigma)
HEPES binding buffer, pH 7.5	50 mM HEPES, 500 mM NaCl, 5 mM imidazole
HEPES wash buffer, pH 7.5	50 mM HEPES, 500 mM NaCl, 20 mM imidazole
HEPES elution buffer, pH 7.5	50 mM HEPES, 500 mM NaCl, 250 mM imidazole

*Note for His6-DRG2 purification, 2 mM DTT was sometimes added to each solution.

Table 7.15. Elution method for JMJD7 proteins.

Column Volume	Binding Buffer (%)*	Elution Buffer (%)**
1	90	10
2	80	20
3	70	30
4	60	40
7	0	100
32		End Method.

7.4.4 Size exclusion chromatography

Proteins were further purified via size exclusion chromatography (SEC) using one of the following 16/600 prep grade 120 mL or 300 mL columns – Superdex 75 or 200 (S75 or S200; GE Healthcare). Size exclusion resin was from Amersham Biosciences. The column was washed with 1 CV of de-gassed MQ H_2O and pre-equilibrated with 1-1.5 CVs of the appropriate de-gassed gel filtration buffer prior to use (**Table 7.16**). The protein was loaded onto the column using a 10 mL (for 300 mL columns) or 2 mL injection loop (for 120 mL columns) at a flow rate of 0.5 mL/min. Note that if the combined protein-containing fractions totalled > 7 mL (for 10 mL loop) or > 2 mL (2 mL loop), the protein was concentrated to < 6 mL or < 2 mL, respectively, using an Amicon Ultra 4 or 15 mL centrifugal filter (Millipore) with an appropriately selected MWCO prior to loading onto the gel filtration column, according to the manufacturer's protocol. However, note that concentration of JMJD7 proteins in elution buffer was avoided due to protein instability/precipitation in high imidazole concentrations. The protein was eluted in the appropriate gel filtration buffer at 1 mL/min and collected in 5 mL fractions (**Table 7.16**). SDS-PAGE was used to identify fractions containing the protein of interest, which were then combined and concentrated using an appropriately selected MWCO Amicon Ultra 15 mL centrifugal filter (Millipore). Once the proteins had reached a concentration of 3-26 mg/mL (depending on subsequent use), the sample was centrifuged (18,000 x g, 4°C, 10 min) to remove any precipitate. Protein purity was assessed by SDS-PAGE. The supernatant was aliquoted, flash frozen in liquid nitrogen and stored at -80°C.

Table 7.16. Gel filtration elution buffers for purified proteins.

Protein	Gel Filtration Column	Gel filtration buffer
(^{15}N and/or ^{13}C) $eRF1_{ND}$-His_6 (wildtype, hydroxylated, and +30 Da modified)	S75 300 mL	20 mM Tris, 200 mM NaCl, pH 6.8
$eRF1_{FL}$-His_6		50 mM HEPES, 200 mM NaCl, 2% glycerol, pH 7.5
His_6-DRG1	S200 120 mL	50 mM HEPES, 200 mM NaCl, 2% glycerol, pH 7.5
His_6-DRG2	S200 120 mL	50 mM HEPES, 200 mM NaCl, 2% glycerol, pH 7.5 ± 2 mM DTT
His_6-JMJD4	S75 300 mL	20 mM Tris, 200 mM NaCl, 5% glycerol pH 7.5
His_6-JMJD7* (large scale)	S75 300 mL	50 mM HEPES, 200 mM NaCl, 2% glycerol, pH 7.5
His_6-JMJD7* (1-2 L scale)	S200 120 mL	

* Represents conditions for all JMJD7 constructs and variants.

7.4.5 Protein concentration determination

Protein concentration was quantified using the absorbance at 280 nm measured by a NanoDrop ND-1000 spectrophotometer and an extinction coefficient calculated by ExPASy ProtParam (http://web.expasy.org/protparam/).

7.4.6 SDS-PAGE gel electrophoresis

SDS-PAGE was used to analyse proteins by their molecular weights. 12% SDS-PAGE gels were made as outlined in **Table 7.17**. Protein samples were diluted with 2x SDS-PAGE sample loading buffer (**Table 7.19**). Samples were boiled at 100°C for 5-10 minutes to denature the protein samples before loading onto the gel. The gel was run at a constant current of 30 mA in GTS running buffer (200 mM glycine, 25 mM Tris, 0.1% SDS) in a Bio-Rad Mini-PROTEAN® Tetra Cell. The gel was then rinsed with water, stained for 20 minutes with

shaking in stain buffer (40% methanol, 10% acetic acid, 0.05% Coomassie Blue R), rinsed again with water and subsequently destained (40% methanol, 10% acetic acid) overnight.

Table 7.17. 12% SDS-PAGE gel composition.

Reagent	Separating gel (2 gels)	Stacking gel (2 gels)
2x separating buffer, pH 8.8*	5 mL	-
2x stacking buffer, pH 6.8*	-	2.5 mL
30% acrylamide	4 mL	830 µL
Milli-Q water	1 mL	1.67 mL
10% ammonium persulfate (APS)	100 µL	50 µL
N,N,N'N'-Tetramethylethylenediamine (TEMED)	10 µL	5 µL

*See **Table 7.18** below.

Table 7.18. 2x separating and stacking gel buffer recipes for SDS-PAGE gel.

Reagent	2x Separating gel buffer, pH 8.8		2x Stacking gel buffer, pH 6.8	
	Per 500 mL	Concentration	Per 500mL	Concentration
Trizma® Base	90.9 g	0.75 M	30.3 g	0.25 M
10% SDS	10 mL	0.2%	10 mL	0.2%

Table 7.19. SDS-PAGE sample loading buffer recipe.

Reagent	2x stock*	Final Concentration (1x stock)
1M Tris-HCl, pH 6.8	1 mL	50 mM
0.5% (w/v) bromophenol blue	1 mL	0.025%
Glycerol	2 mL	10%
SDS	0.4 g	2%
Milli-Q water	4 mL	-

*Stored in 400 µL aliquots at -20°C. Added 100 µL of 1M DTT to a freshly thawed aliquot before use.

7.4.7 Western blot analyses for JMJD7 removal optimisation

Samples were run on a 12% SDS-PAGE gel as described in **Section 7.4.6**, including a lane with PageRuler™ pre-stained protein ladder. The un-stained gel was then transferred to a Hybond ECL membrane (0.2 µm pore size, GE Healthcare), inserted between two layers of 3M paper in

transfer buffer (25 mM Tris, 192 mM glycine, 10% methanol), by running a constant current (200 mA) for 90 minutes on ice using a Mini Trans-Blot Electrophore (Bio-Rad). The membrane was then blocked with 5% skim milk in PBS (137 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KCl, pH 7.4) for 1 hour at room temperature with shaking. The blocked membrane was treated with a primary monoclonal mouse α-His_6 antibody (Abcam, ab18184) at a 1:500 dilution in 1.4 mL PBS buffer overnight at 4°C with rocking. The membrane was then washed with PBS-T [PBS + 0.1% (v/v) Tween-20] 3 times for 5 minutes with shaking. A horseradish peroxidase (HRP)-conjugated goat α-mouse secondary antibody (Thermo Scientific, 32430) was added to the membrane at a 1:20,000 dilution in 40 mL PBS-T for 1 hour at room temperature followed by washing with PBS-T as described. The blot was then treated for 5 minutes at room temperature with a 1.5 mL solution of a 1:1 ratio of Solution A (Luminol Enhancer Solution):Solution B (Peroxide Solution) from the Amersham™ ECL™ Prime Western Blotting Detection Reagent kit (GE Healthcare, Lot No. 9732255) for chemiluminescence detection. The membrane was then exposed to Amersham™ Hyperfilm ECL film (GE Healthcare) for 1 minute and the film was developed using a Compact X4 (Xograph) film developer.

7.5 Solid phase peptide synthesis (SPPS)

The 25-mer $DRG1_{16\text{-}40}$ (ARTQKNKATAHHLGLLKARLAKLRR-NH_2, >95%) and $DRG2_{15\text{-}39}$ (ARTQKNKATEYHLGLLKAKLAKYRA-NH_2, >95%) peptides were commercially synthesised by Severn Biotech (United Kingdom). Note that all peptides used in this study were synthesised with *C*-terminal amides.

7.5.1 Intavis Multipep

Peptides (16- to 20-mers) were prepared with *C*-terminal amides by SPPS on a Multipep RS (Intavis) peptide synthesiser on a plate module (1-5 mg scale) and using 0.2 M FMOC-

protected amino acids in dimethylformamide (DMF) and N,N'-diisopropylcarbodiimide (DIC) as a coupling reagent. Peptides were cleaved from the Tantagel S RAM (RAPP Polymere) resin and protective side chain groups were removed by treatment with 600 µL of 97.5% trifluoroacetic acid (TFA), 2.5% triisopropylsilane (TIPS) (3x 200 µL over 3 hours) directly into each well. The peptides were collected in the deprotection solution in small vials below the plate. The peptides were then precipitated by the addition of 4 mL of ice-cold ether. The mixture was centrifuged (3000 x g, 10 minutes, 4ºC) and the supernatant discarded; peptides were dried for 1-2 hours or overnight. Dried peptides were re-suspended in 1 mL of water (with the addition of acetonitrile if the peptides were insoluble in water) and lyophilised to remove residual traces of TFA. Peptides were made up to a 1 mM solution in MQ-H_2O and verified by MALDI-TOF-MS as described in **Section 7.8.1**. Note that peptides used for screening were not purified.

7.5.2 Liberty Blue

Peptides were prepared with *C*-terminal amides by SPPS on a Liberty Blue microwave peptide synthesiser (CEM) using 0.2 M FMOC-protected amino acids in DMF and DIC as a coupling reagent. Following synthesis, DMF was removed from the resin-linked peptide using a phase separator, and the resin was washed with 3 x 15 mL dichloromethane (DCM). After the resin was left to dry, peptides were deprotected and cleaved from the resin (Rink amide – MBHA resin; AGTC Bioproducts) with 1-5 mL of deprotection solution (92.5% TFA, 2.5% TIPS, 2.5% 1,3-dimethoxybenzene, and 2.5% MQ-H_2O). The peptide was collected in a 50 mL Falcon tube and precipitated by the addition of up to 40 mL ice-cold ether. The precipitated peptide was then pelleted by centrifugation as described (**Section 7.5.1**). The ether was removed, and the peptides were left to dry overnight. Peptides were re-suspended in 1-3 mL water (and acetonitrile, if necessary), flash frozen in liquid nitrogen, and lyophilised to remove trace TFA. Peptides were made up to a 1 mM solution in water and verified by MALDI-TOF-MS as described in **Section 7.8.1**.

7.5.3 Cyclic peptide synthesis

7.5.3.1 Peptide synthesis

The eRF1 linear cyclic peptide precursor (GTASNIKSRVNRLS[D-Cys]-NH_2) was synthesised by SPPS on a Liberty Blue microwave peptide synthesiser (CEM) as described in **Section 7.5.2**. Prior to cleavage and deprotection of the peptide with the TFA deprotection solution, the peptide was treated with 170 mg of chloroacetic anhydride for 3 hours at room temperature with intermittent, vigorous shaking. The extent of chloroacetylation was monitored by MALDI-TOF-MS (**Section 7.8.1**). The peptide was then washed with DCM, cleaved from the resin, and treated exactly as described in **Section 7.5.2**. The peptide was then purified by HPLC (**Section 7.6**) prior to cyclisation (**Section 7.5.3.2**).

7.5.3.2 Cyclisation

The purified, chloroacetylated eRF1 peptide (GTASNIKSRVNRLS[D-Cys]-NH_2) was re-suspended in 0.5 mL of 1M tetraethylammonium acetate buffer, pH 9.5. The pH was adjusted to 8.5 with triethylamine. The peptide was left to cyclise for 2 hours at room temperature. Conversion of the cyclised peptide was assessed by MALDI-TOF-MS (**Section 7.8.1**) prior to purification by HPLC purification (**Section 7.6**).

7.6 HPLC peptide purification

Crude peptides (re-suspended in MQ-H_2O plus MeCN, if required for solubility) and the dried tryptic eRF1 digests (re-suspended in 97.9% MQ-H_2O, 2% MeCN, 0.1% TFA) were filtered with a PFTE 4-4 0.45 µm syringe filter (Sulpeco) prior to loading onto a Grace Vydac 218TP C18 column controlled by a JASCO HPLC system. Peptides were eluted using the optimised gradients outlined in **Table 7.20** [Solvent A = 0.1% (v/v) TFA in water; Solvent B = 0.1% (v/v) TFA in acetonitrile]. Note that Solvent A was degassed for 30-60 minutes prior to use.

Table 7.20. Optimised HPLC elution gradients.

Peptide Sample	Time	Solvent B (%)
Chloroacetylated linear and cyclised eRF1 peptides	0-40 40-45 45-50	0-50 50-98 98-2
$eRF1_{ND}$ tryptic digest for *MLADEFGTASNIK* peptide isolation	0-40 40-45 45-50	0-40 40-98 98-2
$eRF1_{ND}$-30 tryptic digest for *MLADEFGTASNIK(+30)SR* peptide isolation	0-5 5-40 40-45 45-50	0-18 18-35 35-98 98-2

7.7 Hydroxylation assays

Isolated protein assays were performed using recombinant JMJD4 or JMJD7 as detailed in the sections below. Co-factors were prepared in 100 mM stocks and further diluted to 10x stocks in the appropriate assay buffer, except Fe(II), which was prepared in a 100 mM stock in 20 mM HCl and further diluted to a 10x stock with Milli-Q water (**Table 7.21**).

Table 7.21. Co-factors for isolated protein assays.

Co-factors	Abbreviation	Molecular Weight (g/mol)
Ammonium iron(II) sulphate	Fe(II)	392.14
Alpha-ketoglutaric acid disodium salt	2OG	190.1
(+)-sodium L-ascorbate	ascorbate	198.1

7.7.1 JMJD7 hydroxylation assays

7.7.1.1 Hydroxylation assays by MALDI-TOF MS or LC-MS

JMJD7 assay mixtures (final volume 20 µL in 50 mM HEPES, pH 7.5) containing 50 µM peptide or protein substrate and 10 µM His_6-JMJD7 were incubated for 60 minutes at 37°C (or as indicated) with 100 µM Fe(II), 200 µM 2OG, and 500 µM ascorbate in the presence or absence of 1 mM *N*-oxalylglycine, NOG (synthesised by Dr Marina Demetriades)[159], unless

otherwise indicated. Reactions containing peptide were quenched by immediately spotting on a MALDI plate in a 1:1 sample:matrix ratio. Protein substrates were quenched with an equal volume of formic acid (2% final concentration) and diluted accordingly with water. Samples were analysed by MALDI-TOF-MS (peptides; **Section 7.8.1**) or LC-MS (proteins; **Section 7.8.2**).

7.7.1.2 Co-factor/substrate dependence assays & time course conditions

Co-factor/substrate dependence assays were performed as described in **Section 7.7.1.1** with or without the following modifications: 50 µM $DRG1_{16-40}$, ± 1 µM JMJD7 or JMJD7-H178A, ± 100 µM Fe(II), ± 500 µM 2OG, ± 1 mM ascorbate, ± 1 mM NOG, and quenching after 5 minutes of incubation at 25° or 37°C (as indicated) with an equal volume formic acid (2% final concentration) prior to spotting on a MALDI plate.

7.7.1.3 Competition & activity assays

DRG competition assays and enzyme activity assays were performed under the modified conditions described in **Section 7.7.1.2** with the following variations: ± 50 µM $DRG1_{16-40}$, ± 50 µM $DRG2_{15-}$, ± 1 µM JMJD7, and quenching after 5 minutes and 40 minutes of incubation at 25°C with an equal volume of formic acid (2% final concentration) prior to spotting on a MALDI plate.

7.7.2 JMJD4 assays

7.7.2.1 $eRF1_{ND}$ hydroxylation trials

***7.7.2.1.1** Co-expression trials*

Co-production of $eRF1_{ND}$-His_6 and His_6-JMJD4 was carried out as described in **Section 7.3.5** in 100 mL 2TY media at the optimal induction temperature (30°C). Bacterial pellets were lysed

and purified (as described in **Section 7.4.2**), and proteins were analysed by intact protein LC-MS (**Section 7.8.2.1**), revealing only ~40% hydroxylation.

7.7.2.1.2 Co-lysis trials

To promote hydroxylation, pellets from eRF1$_{ND}$-His$_6$/His$_6$-JMJD4 co-producing Rosetta(DE3) pLysS cells (in 100 mL M9 minimal media; **Section 7.3.5**) were lysed and combined with lysate from Rosetta(DE3) pLysS cells expressing His$_6$-JMJD4 alone (in 200 mL 2TY media) in a 2:1 volume ratio. Cell lysis was carried out as described in **Section 7.4.2**. Lysate was supplemented with 100 µM Fe(II), 200 µM 2OG, and 500 µM ascorbate in 50 mM HEPES, pH 7.5 (prepared as described in **Section 7.7**; final concentrations) and incubated overnight at 4°C. Proteins were purified as described in **Section 7.4.2** and analysed by intact protein LC-MS (**Section 7.8.2.1**).

7.7.2.2 Hydroxylation assays

JMJD4 hydroxylation assay mixtures (final volume 20 µL in 20 mM Tris-HCl, pH 7.5) containing 50 µM peptide or protein substrate and 10 µM His$_6$-JMJD4 were incubated for 60 minutes at 37°C with 100 µM Fe(II) and 200 µM 2OG (or as indicated). Reactions containing peptide were quenched by immediately spotting on a MALDI plate in a 1:1 sample:matrix ratio. Protein substrates were quenched with an equal volume of formic acid (2% final concentration) and diluted accordingly with MQ-H_2O. Samples were analysed by MALDI-TOF-MS (peptides; **Section 7.8.1**) or LC-MS (proteins; **Section 7.8.2**).

7.7.2.3 '+30 Da' assays

JMJD4 '+30 Da' assay mixtures (final volume 20 µL in 20 mM Tris-HCl, pH 7.5) containing 50 µM peptide or protein substrate and 50 µM His$_6$-JMJD4 were incubated overnight (16-20 h) at 4°C with 1 mM Fe(II) and 2 mM 2OG. Samples were quenched and analysed as described in **Section 7.7.2.2**.

7.7.2.4 Optimised large scale purification of (un)labelled $eRF1_{ND}$-OH

A reaction mixture containing 200 μM of purified $eRF1_{ND}$-His_6 (in 20 mM Tris, 200 mM NaCl, pH 6.8) was incubated with 20 μM of purified His_6-JMJD4 (in 20 mM Tris, 200 mM NaCl, 5% glycerol, pH 7.5) in the presence of 200 μM Fe(II) and 400 μM 2OG in 20 mM Tris, pH 7.5 at 37°C for 30 minutes. The reaction was quenched with 5.12 mM EDTA and centrifuged (18,000 x g, 4°C, 10 min) to remove precipitated enzyme. $eRF1_{ND}$-OH was purified away from JMJD4 and the reaction mixture by SEC (**Section 7.4.4**) on a Superdex 75 300 mL column pre-equilibrated with 20 mM Tris, 200 mM NaCl, pH 6.8 (for NMR) or 7.5. Fractions containing $eRF1_{ND}$-OH alone were combined, concentrated, and flash frozen as described in **Section 7.4.4**.

7.7.2.5 Optimised large scale modification and purification of (un)labelled $eRF1_{ND}$-30 for NMR spectroscopy or tryptic digestion and isolation of +30 Da peptide

A reaction mixture [6-7 mL] containing a 1:1 molar ratio of $eRF1_{ND}$-His_6 and His_6-JMJD4 (50-80 μM) was incubated in the presence of 1 mM Fe(II) and 2 mM 2OG in 20 mM Tris, pH 7.5 at 4°C for 18-20 h. Note that precipitate was observed. The reaction was quenched with 5.14 - 7.14 mM EDTA, and the samples centrifuged (18,000 x g, 4°C, 10 min) to remove precipitated enzyme. $eRF1_{ND}$-30 was purified from JMJD4 and the reaction as described in **Section 7.7.2.5**. Note that $eRF1_{ND}$-30 prepared for the isolation of a +30 Da-containing peptide fragment was digested with trypsin as described in **Section 7.8.3.1**, though was not subsequently purified using a Sep-Pak C18 cartridge to avoid material loss. Rather, the peptide was purified by HPLC as described in **Section 7.6** and using the conditions outlined in **Table 7.20**.

7.8 Mass spectrometry methods

7.8.1 MALDI-TOF-MS

The extent of enzyme-catalysed peptide modifications were analysed using a Waters Micromass™ MALDI micro MX™ mass spectrometer in the positive ion reflectron mode with

a laser energy of 220-240, pulse of 1950, detector of 2353 and suppression of 500-1000 as described[177]. Samples were prepared by spotting 1 μL of sample and 1 μL of α-cyano-4-hydroxycinnamic acid (CHCA) matrix solution (saturated α-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 0.1% (v/v) trifluoroacetic acid) directly onto a Waters 96-spot MALDI plate. Instrument control and data processing were performed using MassLynx V4.1; data processing by smoothing and/or baseline subtraction are outlined in **Table 7.22**.

Table 7.22. MALDI-TOF-MS data processing parameters.

	Parameter	Value
Subtract	Polynomial order	15 / 1
	Below curve (%)	10 / 40
	Tolerance	0.01
Smooth	Smooth window (channels)	10
	Number of smooths	2
	Smooth method	Savitzky Golay

7.8.2 Intact protein LC-MS

LC-MS analyses were performed in collaboration with Dr James Wickens and Dr James McCullagh (Mass Spectrometry, Department of Chemistry, University of Oxford) according to the following protocols kindly provided by Dr Wickens.

7.8.2.1 Open access LCT Premier

Protein samples were subjected to LC-MS analysis on an open access service by injecting 5 μL of protein sample (0.01-0.02 mg/mL) onto a Merck Chromolith C18 2x5 mm guard column coupled to a Waters LCT Premier XE, equipped with an electrospray interface, and using a gradient of 5-100% acetonitrile in a 0.1% formic acid solution (v/v) in Milli-Q water over the course of 8 minutes. Data processing was performed using Mass Lynx V4.1. Electrospray mass spectra of protein samples were deconvoluted using MaxEnt1 software (Waters, **Table 7.23**) and spectra were baseline subtracted using the parameters outlined in **Table 7.22**.

Table 7.23. MaxEnt1 deconvolution parameters.

Parameter	Value
Resolution	1 Da / channel
Damage model	Uniform Gaussian
Width at half height	0.75 Da
Minimum intensity ratios (left and right)	33%

7.8.2.2 XEVO G2-S QTof

Intact protein LC-MS analyses were performed in collaboration with Dr James Wickens. Instrument control, data processing, and data acquisition were performed using MassLynx V4.1 with the mass spectrometer in the positive ion mode. Source conditions were adjusted to maximum sensitivity, and Lockspray™ was used to maintain mass accuracy during analysis. An ACQUITY UPLC system and Chromolith PR18 2x50 mm column (Merck) was used to perform liquid chromatography. 5-10 µL of each sample was injected onto the system. Elution solvents included Solvent A (MQ-H_2O with 0.1% (v/v) analytical-grade formic acid) and Solvent B (HPLC-grade acetonitrile). The elution program for intact protein LC-MS is outlined in **Table 7.24**.

Table 7.24. 10 minute UPLC gradient for intact protein LC-MS.

Time*	Solvent B (%)
0	5
1	5
7	99
8	99
8	5

*Rate = 0.2 ml/min

7.8.2.3 Exactive and Q-Exactive

High resolution LC-MS analyses were carried out in collaboration with Dr James McCullagh and Dr James Wickens using either a Thermo Q-Exactive S instrument equipped with a Thermo U3000 chromatography system or Thermo Exactive mass spectrometer equipped with an Aligent 1100 chromatography system. The mass spectrometers were operated using heated electrospray (HESI-II) probes, with all analysis performed in the positive ion mode. For both

instruments, 1 µL of sample (0.2 – 0.3 mg/mL) was injected onto a Chromolith Fastgradient RP18e 50x2 mm HPLC column at a flow rate of 0.4 mL/min using a gradient of 5-99% HPLC grade acetonitrile in a 0.1% formic acid solution (v/v) in Milli-Q water over the course of 7 minutes. Instrument control and data processing were performed using Thermo Xcalibur Software. Systems were calibrated on the day of analysis using an external myoglobin calibration standard.

7.8.3 LC-MS/MS

7.8.3.1 In-solution tryptic digest

Purified protein samples were digested with Sequencing Grade Modified Trypsin (Promega) according to an adapted protocol from the Kessler Lab Proteomics Group (Target Discovery Institute, University of Oxford, https://www.tdi.ox.ac.uk/_asset/file/solution.pdf). Proteins were re-suspended or diluted to 50-100 µL with 20 mM Tris, 200 mM NaCl, pH 6.8. 100 µL of 6M urea was added, and the sample was vortexed and subsequently sonicated in a Grant XB3 ultrasonic bath for 2 minutes. 5 µL of freshly-prepared 200 mM dithiothreitol (DTT) in 0.1 M Tris, pH 7.8 was then added, vortexed, and incubated for 30-60 min at room temperature. 20 µL of 200 mM iodoacetamide in 0.1 M Tris, pH 7.8 was added, vortexed, and incubated for 30-60 min at room temperature. 20 µL of 200 mM dithiothreitol (DTT) in 0.1 M Tris, pH 7.8 was again added, vortexed, and incubated for an additional 30-60 min at room temperature. The mixture was diluted to 1 mL with Milli-Q water and vortexed. A 2 µg/mL sequencing-grade modified trypsin (Promega) reagent was prepared by dissolving 20 µg of trypsin in 10 mL of cold 0.1 M Tris, pH 7.8. Trypsin was added in a 1:20 or 1:25 ratio to the total protein content and carefully mixed by pipetting. The reaction mixtures were incubated at 37°C overnight and purified by Sep-Pak C18 cartridges (**Section 7.8.3.2**).

7.8.3.2 Sep-Pak C18 Purification

Tryptic digest reactions were quenched with glacial acetic acid to pH < 6, and subsequently de-salted and purified using a Sep-Pak® Light C18 cartridge (Waters) according to a protocol from the Kessler Lab Proteomics Group (Target Discovery Institute, University of Oxford, http://www.ccmp.ox.ac.uk/_asset/file/sep-pak-C18.pdf). The Sep-Pak column was equilibrated with 5 mL of solution B (65% acetonitrile, 35% Milli-Q water, 0.1% formic acid; HPLC grade) followed by 10 mL of solution A (98% Milli-Q water, 2% HPLC-grade acetonitrile, 0.1% formic acid; HPLC grade). The peptide digest sample was then slowly let into the Sep-Pak column using a syringe. The column was then washed with 10 mL of solution A. The sample was eluted twice with 1 mL of solution B and collected in two separate 1.5 mL tubes. When the samples had been dried down to ~500 mL by vacuum centrifugation, they were combined and allowed to dry completely. Samples were re-suspended in 20 mL of solution A for analysis via LC-MS/MS.

7.8.3.3 Synapt G2-Si high definition mass spectrometry (HDMS)

LC-MS/MS analysis was done on eRF1 (wildtype, hydroxylated, and +30 Da modified) tryptic peptides on a service basis by Mr Lingzhi Gong (Mass Spectrometry, Department of Chemistry, University of Oxford) according to the following protocol kindly provided by Lingzhi Gong. The protein digests were analysed using a nanoACQUITY UPLC coupled to SYNAPT G2-Si HDMS interfaced with a nanoelectrospray source (Waters). Protein digests were injected onto a 5 mm symmetry C18 trap column (180 mm × 20 mm) and washed for 1 min at 15 mL/min with 0.1% (v/v) formic acid. Peptides were then separated and eluted for MS analysis using a gradient of acetonitrile containing 0.1% (v/v) formic acid on a nanoACQUITY UPLC column (BEH130 C18 1.7 µm particle size, 75 mm × 250mm). The column temperature was set at 35^{o}C. The eluted peptides were analysed in the positive ionisation mode over a mass range of 50–1990 m/z with a scan time of 0.6 s. The eluted peptides were analysed using an MS^{E} method collecting MS/MS data using collision energy ramping from 15 to 35 eV. LC-

MS/MS data were analysed using MassLynx V4.1 (Waters). Mass spectra were deconvoluted using MaxEnt3 software (Waters) from 0-2000 Da with a max number of charges = 4 (peptide isotope distribution).

7.8.3.4 XEVO G2-S QTof

LC-MS/MS analyses were performed in collaboration with Dr James Wickens as described in **Section 7.8.2.2**. The elution gradient for targeted LC-MS/MS is outlined in **Table 7.25**; MS/MS was performed in the collision-induced dissociation (CID) mode. Elemental composition was calculated from high resolution LC-MS/MS mass spectra using MassLynx V4.1 (Waters).

Table 7.25. 10 minute UPLC gradient for peptide LC-MS/MS.

Time*	Solvent B (%)
0	5
1	5
6	50
7	95
8	95
8.1	5

*Rate = 0.4 ml/min

7.8.4 Non-denaturing mass spectrometry

The following protocol was kindly provided by Dr Weston Struwe (Department of Chemistry, University of Oxford):

Stoichiometry measurements were performed using a hybrid quadrupole-Orbitrap (Thermo QExactive) or quadrupole-time-of-flight (Waters Synapt G2-Si) mass spectrometer modified for detection of high mass ions as described[298,299]. Samples were buffer exchanged into 200 mM ammonium acetate using Bio-Spin 6 (Bio-Rad) centrifuge columns and diluted to ~10 µM immediately prior to analysis. Ions were generated by static nanoelectrospray using gold-coated capillaries prepared in-house. Parameters for data acquisition for each instrument are outlined in **Table 7.26**. Data were obtained with a wide acquisition window (2,000-10,000 m/z) and

sufficient desolvation was achieved with moderate HCD voltage applied (10-50 V for Thermo QExactive) or moderate CID (4V for Waters Synapt G2-Si). Spectra were obtained with 10 microscans, averaged over 100 scans. Data were processed using Xcalibur 2.1 software (Thermo Scientific) or MassLynx V4.1 (Waters). Masses were assigned using in-house software (Mass/Charge State Error Determination v 0.3; http://benesch.chem.ox.ac.uk/resources.html).

Table 7.26. Instrument parameters for non-denaturing MS analyses.

Parameters	Thermo QExactive	Waters Synapt G2-Si
Capillary voltage	1.1 kV	1.4 kV
Source temperature	50°C	50°C
Max injection time	50	-
S-lens RF	100	-
C-trap entrance lens	5.8	-
Sample cone	-	10 V
DC bias	-	10 V

7.8.5 Amino acid analysis

7.8.5.1 Sample Preparation

7.8.5.1.1 JMJD7 peptide samples

Peptides (50 µM; $DRG1_{21-40}$, NKATAHHLGLLKARLAKLRR-NH_2, $DRG1_{16-40}$, ARTQKNKATAHHLGLLKARLAKLRR-NH2; $DRG2_{15-39}$, ARTQKNKATEYHLGLL-KAKLAKYRA-NH_2) were incubated with 100 µM Fe(II), 200 µM 2OG, and 500 µM ascorbate in the presence or absence of 10 µM recombinant His_6-JMJD7 for 1 hour at 37°C in 50 mM HEPES, pH 7.5 (final volume 500 µL). The reaction was quenched with 2 drops of neat formic acid, and hydroxylation (~30-80%) of the peptide was confirmed by MALDI-TOF MS. Recombinant JMJD7 was removed using a Sep-Pak C18 column (**Section 7.8.3.2**), eluting in 25% acetonitrile, 0.1% formic acid (v/v) solution in ddH_2O. Samples were dried by vacuum centrifugation and subsequently re-suspended in 60 µL of 20 mM HEPES, pH 7.5.

7.8.5.1.2 eRF1 protein samples

Purified $eRF1_{ND}$ wildtype, hydroxylated, and +30 Da modified protein samples (133 µg) were made up to a 60 µL final volume 20 mM HEPES, pH 7.5.

7.8.5.1.3 Enzymatic hydrolysis

Enzymatic hydrolysis was carried out at 37°C with a non-specific protease from *Streptomyces griseus* (1 mg/mL in 20 mM HEPES, Sigma-Aldrich) in a 1:10 mass ratio of enzyme:sample overnight as described[277]. The hydrolysed samples were then dried by vacuum centrifugation.

7.8.5.2 LC-MS analyses

Hydrolysed samples were dried by vacuum centrifugation followed by re-suspension in 80 µL borate buffer and derivatisation with 20 µL 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate (AQC, 1 mg/mL in acetonitrile) according to the AccQ-Tag Ultra Derivatization Protocol (Waters). LC-MS analyses were performed as described[109,133] using a Waters Acquity ultra-performance liquid chromatography system coupled to a Xevo G2-S QTof (Waters) equipped with an electrospray ionisation source (for JMJD7 samples) or LCT Premier XE orthogonal acceleration time-of-flight mass spectrometer (Waters) equipped with an electrospray ionisation source (for eRF1 samples) as described in **Table 7.27**[133]. Gradient conditions for amino acid separation were as described in the AccQ-Tag Ultra Amino Acid Analysis Derivatization Protocol (Waters). Concentrations of synthetic standards for spiking experiments were optimised based on extracted ion intensities of samples.

MS data were acquired and analysed using MassLynx V4.1 (Waters). Extracted ion chromatograms were produced for m/z values of either 503.2, corresponding to the theoretical mass of AQC-derivatised hydroxylysine (503.2037). Extracted ion chromatograms were smoothed (number of smooths, 2; smooth window channels, 3; Savitzky-Golay) and total ion

current chromatograms were further baseline subtracted (polynomial order, 1; below curve %, 40; tolerance, 0.01).

C3, C4, and C5 hydroxylysine standards were synthesised or commercially obtained as described[133]. The synthesis of the (2*S*,3*R*)/(2*R*,3*S*)-3-hydroxylysine standard has been reported elsewhere[172].

Table 7.27. MS parameters for amino acid analysis.

Parameters	Xevo G2-S QTof	LCT Premier XE
Mode	Electrospray positive-ion	Electrospray positive-ion
Analyser	Resolution mode	V mode
Desolvation temperature	600°C	250°C
Source temperature	100°C	120°C
Capillary voltage	3000 V	3000 V
Sample cone	20 V	100 V
Cone gas flow	30 L/min	30 L/min
Desolvation gas flow	1000 L/min	550 L/min

7.9 JMJD7 protein crystallisation

7.9.1 Crystallisation

Proteins used for crystallography were prepared by incubating 25 mg/mL protein with 2 mM $MnCl_2$-$4H_2O$ (added from a 100 mM stock in ddH_2O water) on ice for 15 minutes. This was followed by a 1 hour incubation with 4 mM 2OG [disodium salt, added from a 100 mM stock in gel filtration buffer (50 mM HEPES, 200 mM NaCl, 2% glycerol, pH 7.5)]. Precipitate was removed by centrifugation (18,000 x g, 5 minutes, 4°C) prior to laying protein crystallisation trays. Crystals of JMJD7 and JMJD7-R260C in complex with Mn(II) and 2OG were grown in sitting drops at 25°C using the vapour diffusion method (drop size: 200-300 nL) in 96-well Intelliplates (Art Robbins).

For wildtype JMJD7, drops were formed in a 1:1 ratio of a 25 mg/mL His_6-JMJD7, 2 mM $MnCl_2$, 4 mM 2OG solution with a 1.5 M sodium nitrate, 0.1 M sodium acetate, pH 4.6 solution (SaltRx 1 #43/SaltRx HT well D7, Hampton Research). For the R260C variant, drops were formed in a 1:1 ratio of a 26 mg/mL His_6-JMJD7-R260C, 2 mM $MnCl_2$, 4 mM 2OG solution with the commercial crystallisation reagent, 1.5 M sodium nitrate, 0.1 M sodium acetate, pH 4.6 solution (SaltRx 1 #43), containing the following additives: 0.25% β-Nicotinamide adenine dinucleotide phosphate tetrasodium salt, 0.25% Adenosine 5'-triphosphate disodium salt hydrate, 0.25% *N*-Acetyl-*D*-galactosamine, 0.25% Gentamicin sulphate salt hydrate, 0.02 M HEPES sodium pH 6.8 (Silver Bullets Bio, Hampton Research, HR2-088). Crystals were cryoprotected with 25% glycerol prior to harvesting in nylon loops (Hampton Research) and flash freezing in liquid nitrogen.

7.9.2 Data collection, structure solution, and refinement

The following protocol was kindly provided by Dr Rasheduzzaman Chowdhury (Schofield Laboratory, University of Oxford):

Data were collected using single crystals at 100 K at the Diamond Light Source (DLS) beam lines and processed using HKL2000[207] (**Table 3.1**). His_6-JMJD7.Mn(II).2OG data were strongly anisotropic (the spread in values of the three principal components of scale factors is 62.37 $Å^2$), with resolution limits of 3.5 Å, 3.5 Å, and 3.0 Å along the a*, b* and c* axes (at an F/σ cut-off of 3). An ensemble of models was generated by the ENSEMBLER built in PHENIX[211] using the crystal structures of related enzymes, FIH (PDB ID: 1H2K)[84], TYW5 (PDB ID: 3AL5)[103], JMJD5 (PDB ID: 4GJZ)[156], and JMJD6 (PDB ID: 3K2O)[209] as templates. To ensure maximum degree of accuracy, only JMJD7 residues that are highly conserved in these four JmjC hydroxylases were included in the homology models. The wildtype JMJD7 structure was then solved by molecular replacement using PHASER[208]. An initial model was generated by PHENIX AUTOBUILD[211], which included 82% of JMJD7 residues making up the four molecules in the asymmetric unit. The data were ellipsoidally truncated and

anisotropically scaled by using the 'UCLA Diffraction Anisotropy Server'[300] for initial rounds of refinements. Iterative cycles of model building and refinement using Buster[301] were performed until converging R_{factor}/R_{free} no longer decreased. Final rounds of intensity-based refinements were performed using CNS[210] and PHENIX[211]. JMJD7-R260C structure was solved by molecular replacement using PHASER[208] with two molecules per asymmetric unit using the wildtype JMJD7 structure as the initial model and refined by CNS[210]/ PHENIX[211] until R_{factor}/R_{free} converged with iterative cycles model building using COOT[212]. MOLPROBITY[302] was used to monitor the geometric quality of the models between refinement cycles. Water molecules were added to peaks (> 1.5 σ $2F_oF_c$) in electron density maps that were within H-bond distance to protein with reasonable geometry to form H-bonds.

7.10 Circular dichroism

CD spectra were acquired using a Chirascan CD spectrometer (Applied Photophysics model) equipped with a Peltier temperature-control cell holder according to the manufacturer's protocol. Experiments were performed in a 0.1 cm pathlength cuvette at 20°C using an optimised protein concentration (JMJD7 proteins, 0.2 mg/mL; eRF1 proteins, 0.25 mg/mL) in 10 mM sodium phosphate buffer (NaH_2PO_4, Na_2HPO_4), pH 8.0. Samples were buffer exchanged into phosphate buffer using Bio-Spin 6 (Bio-Rad) centrifuge columns and diluted to 0.2 mg/mL (JMJD7 proteins) or 0.25 mg/mL (eRF1 proteins) immediately prior to analysis. Data were recorded in triplicate from 185-260 nm with a 1 nm bandwidth in 0.5 nm steps. Spectral data were averaged (n = 3) and smoothed (smooth = 8).

7.11 Differential scanning fluorimetry (DSF)

DSF was performed using SYPRO® Orange Protein Gel Stain (Life Technologies Corporation) and a MiniOpticon™ Real-Time PCR Detection System (Bio-Rad). Fluorescence readings were taken every 1°C on 50 µL of 6 µM protein samples (JMJD7 proteins, n = 4; eRF1

proteins, n = 3) between 25-90°C with a 1°C/min linear increase in temperature. Raw data were fitted to a Boltzmann curve in GraphPad Prism 5 to determine apparent melting temperature values.

7.12 Size exclusion chromatography-Multi-angle light scattering

The following protocol was kindly provided by Dr Weston Struwe (Department of Chemistry, University of Oxford):

Analyses of JMJD7 oligomerisation constructs and clinically-relevant variants were performed using an Agilent 1260 HPLC with a Superdex 200 Increase 3.2/300 SEC column (GE Healthcare) coupled to a DAWN HELEOS II scattering detector (Wyatt Technologies) and Optilab T-rEX (Wyatt Technologies) differential refractive index detector. To ensure minimal background light scatter, the MALS detector was equilibrated overnight in running buffer (50 mM HEPES, 200 mM NaCl, 2% glycerol, pH 7.5). Roughly 30 µg of protein was loaded on-column at a flow rate of 75 µL/min in running buffer. The column compartment was set to 20°C. Data acquisition and analysis were performed with Astra 6.1 software (Wyatt Technologies).

7.13 Nuclear magnetic resonance

7.13.1 Peptide concentration determination by ^{1}H-NMR

Solid peptides (i.e. DRG1$_{16\text{-}40}$ and DRG2$_{15\text{-}39}$ peptides described in **Section 7.5**) were dissolved in MQ-H_2O to a 10 mM concentration. 5 µL of the 10 mM peptide stock was combined with 5 µL of a 1 mg/mL solution of trimethylsilylpropanoic acid (TSP) and 65 µL of D_2O in a 0.5 mL Eppendorf tube and transferred to a 2 mm Bruker MATCH^{TM} NMR tube. ^{1}H NMR spectra were acquired on a Bruker Avance III 700 MHz spectrometer equipped with a 5 mm inverse TCI cryoprobe and controlled by TopSpin 3.5 software. Spectra were acquired with water

suppression using the ^{1}H excitation sculpting suppression with perfect-echo NMR with 32 scans, 2 dummy scans, and a relaxation delay of 1 s. A 2 ms Sinc pulse was used for water suppression. The solvent lock deuterium was used as a lock signal. Prior to Fourier transformation, data were multiplied with an exponential function with 0.1 Hz line broadening. TopSpin 3.2 software was used to process and integrate the peaks for concentration determination using automated routines.

7.13.2 ^{1}H-NMR 2OG turnover time course

NMR assays were carried out as described[31]. Reactions (160 µL final volume) were prepared in a 1.5 mL Eppendorf tube with 16 µL D_2O (Sigma Aldrich), 200 µM 2OG, ±100 µM Fe(II), ±500 µM ascorbate, ±1 mM NOG, ±50 µM $DRG1_{16-40}$, ±10 µM His_6-JMJD7, and ±10 µM His_6-JMJD7-H178A in 20 mM Tris-D11 in MQ-H_2O (Cambridge Isotope Laboratories), pH 7.5. Note that all enzymes were buffer exchanged into 20 mM Tris-D11 in MQ-H_2O using Bio-spin 6 (Bio-Rad) centrifuge columns on the morning of analysis.

Each reaction was initiated by the addition of enzyme and immediately transferred to a 3 mm Shigemi NMR tube. The reaction was monitored at 298K in real-time by ^{1}H NMR on a Bruker Avance III 700 MHz spectrometer equipped with a 5 mm inverse TCI cryoprobe and controlled by TopSpin 3.5 software. The first proton spectrum was acquired with water suppression using the ^{1}H excitation sculpting suppression with perfect-echo NMR experiment 325 s after mixing (start of data acquisition was 220s). Spectra were acquired every 105 s for 1900 s and obtained using 16 scans, 4 dummy scans, and a relaxation delay of 1 s. A 2 ms Sinc pulse was used for water suppression. The solvent lock deuterium was used as a lock signal. Prior to Fourier transformation, data were multiplied with an exponential function with 0.1 Hz line broadening. TopSpin 3.2 software was used to process data using automated routines and integrate the peaks corresponding to 2OG and succinate for analysis of 2OG consumption/succinate formation using Microsoft Excel 2003.

7.13.3 Peptide NMR experiments

Peptide NMR experiments and analyses were performed with guidance from and in collaboration with Prof Timothy Claridge (NMR Facility, University of Oxford). The peptide sample was prepared by re-suspending the HPLC purified $eRF1_{ND}$-30 tryptic 15-mer peptide (MLADEFGTASNIK(+30)SR) in 320 µL of D_2O. The sample was transferred into a 5mm Shigemi NMR tube matched with D2O, bottom L 8 mm. The experiments were conducted on a Bruker Avance III 700 MHz spectrometer equipped with a 5 mm inverse TCI cryoprobe (AV700). The instrument was controlled by TopSpin 3.5 software. All experiments were conducted at 300 K and spectra were acquired with 2 dummy scans, a relaxation delay of 1 s. The solvent lock deuterium was used as a lock signal. Pulse sequences and experiment-specific parameters are summarised in **Table 7.28**.

Table 7.28. NMR pulse sequences and parameters for 1D and 2D experiments.

Experiment	Bruker sequence	Size of FID (^{1}H) / (^{1}H) or (^{13}C)	Spectral width / Centre of spectrum (^{1}H)	Spectral width / Centre of spectrum (^{13}C)	Number of scans
^{1}H	zg60	65536	15.9 / 5.0	-	16
^{1}H-^{1}H COSY	cosygpprqf	2048 / 256	10.0 / 4.70	-	8
^{1}H-^{1}H TOCSY	dipsi2phprf2	2048 / 256	6.0 / 3.0 and 4.7	-	24
^{1}H-^{1}H NOESY	noesygpphpr	2048 / 256	9.99 / 4.7	-	64
^{1}H-^{13}C HSQC	hsqcedetgpsp.3	1024 / 256	6.00 / 3.0	120.0 / 60.0	96
^{1}H-^{13}C HMBC	hmbcetgpl2nd	2048 / 320	8.0 / 4.0	220.0 / 110.0	128

7.13.4 Protein NMR experiments

All protein NMR experiments (^{1}H-^{15}N HSQC, T_1, T_2, NOE, triple resonance, and ^{1}H-^{15}N HSQC rRNA binding experiments) were performed by Dr Ivanhoe Leung (CJS/TDWC Group, University of Oxford, currently at the University of Auckland). ^{15}N- and/or ^{13}C-labelled $eRF1_{ND}$, $eRF1_{ND}$-OH and $eRF1_{ND}$-30 protein samples were prepared in 20 mM Tris, 200 mM NaCl, pH 6.8 to final concentrations of 1 mM, 750 µM and 350 µM, respectively and supplemented with 2 mM 2-mercaptoethanol and 0.02% sodium azide as described in Polshakov *et al.* (2013)[270]. Samples were transferred into a 5 mm Shigemi NMR tube matched with D2O, bottom L 5 mm and topped with a drop of Nujol mineral oil to prevent evaporation.

Experiments were conducted on either a Bruker Avance III 700 MHz spectrometer equipped with a 5 mm inverse TCI cryoprobe (AV700) or Bruker Avance III 600 MHz spectrometer equipped with a Prodigy broadband cryoprobe. The instruments were controlled by TopSpin 3.5 software. All experiments were conducted at 298 K, and the pulse sequence used for each experiment is listed in **Table 7.29**. Data obtained from triple resonance experiments were analysed by Dr Ivanhoe Leung using the CCPN Analysis software (The Collaborative Computing Project for NMR)[303].

Table 7.29. Pulse sequences for NMR experiments.

Experiment	Pulse Sequence
^{1}H-^{15}N HSQC	hsqcetfpf3gpsi
HNCO	hncogp3d
HN(CA)CO	hncacogp3d
HN(CO)CACB	hncocacbgp3d
HNCACB	hncacbgp3d

7.14 Protein Basic Local Alignment Search Tool (BLAST) DRG1 and DRG2 analyses

Protein BLAST[143] searches identified DRG1-2 homologs from indicated organisms using the human protein sequence as the entry query (NCBI Reference IDs – *H. sapiens:* NP_004138.1/001379.1; *X. laevis:* 001084013.1/001079983.1; *D. melanogaster:* 536733.1/650822.1; *C. elegans:* 001255126.1/498808.2; *A. thaliana:* 195662.1/173190.1; *S. cerevisiae:* 009364.1/011689.1; *D. discoideum:* 636294.1/641633.1; *T. thermophila:* 001025967.2/001021328.1).

8 REFERENCES

(1) Hausinger, R. P., and Schofield, C. J. (Eds.). (2015) 2-Oxoglutarate-dependent oxygenases. *RSC Met Ser No 3*. Royal Society of Chemistry, Cambridge.
(2) Loenarz, C., and Schofield, C. J. (2008) Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat Chem Biol 4*, 152–156.
(3) Zhang, Z., Ren, J., Stammers, D. K., Baldwin, J. E., Harlos, K., and Schofield, C. J. (2000) Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat Struct Biol 7*, 127–133.
(4) Roach, P. L., Clifton, I. J., Fülöp, V., Harlos, K., Barton, G. J., Hajdu, J., Andersson, I., Schofield, C. J., and Baldwin, J. E. (1995) Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. *Nature 375*, 700–704.
(5) Clifton, I. J., Doan, L. X., Sleeman, M. C., Topf, M., Suzuki, H., Wilmouth, R. C., and Schofield, C. J. (2003) Crystal structure of carbapenem synthase (CarC). *J Biol Chem 278*, 20843–20850.
(6) Topf, M., Sandala, G. M., Smith, D. M., Schofield, C. J., Easton, C. J., and Radom, L. (2004) The unusual bifunctional catalysis of epimerization and desaturation by carbapenem synthase. *J Am Chem Soc 126*, 9932–9933.
(7) Borowski, T., Broclawik, E., Schofield, C. J., and Siegbahn, P. E. M. (2006) Epimerization and desaturation by carbapenem synthase (CarC). A hybrid DFT study. *J Comput Chem 27*, 740–748.
(8) Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T., and Drennan, C. L. (2006) Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature 440*, 368–371.
(9) Mitchell, A. J., Zhu, Q., Maggiolo, A. O., Ananth, N. R., Hillwig, M. L., Liu, X., and Boal, A. K. (2016) Structural basis for halogenation by iron- and 2-oxo-glutarate-dependent enzyme WelO5. *Nat Chem Biol 12*, 636–640.
(10) Hutton, Jr., J. J., Tappel, A. L., and Udenfriend, S. (1967) Cofactor and substrate requirements of collagen proline hydroxylase. *Arch Biochem Biophys 118*, 231–240.
(11) Markolovic, S., Wilkins, S. E., and Schofield, C. J. (2015) Protein hydroxylation catalyzed by 2-oxoglutarate-dependent oxygenases. *J Biol Chem 290*, 20712–20722.
(12) Klose, R. J., Kallin, E. M., and Zhang, Y. (2006) JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet 7*, 715–727.
(13) Dong, C., Zhang, H., Xu, C., Arrowsmith, C. H., and Min, J. (2014) Structure and function of dioxygenases in histone demethylation and DNA/RNA demethylation. *IUCrJ 1*, 540–549.
(14) Loenarz, C., and Schofield, C. J. (2011) Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem Sci 36*, 7–18.
(15) Hewitson, K. S., McNeill, L. A., Riordan, M. V., Tian, Y.-M., Bullock, A. N., Welford, R. W., Elkins, J. M., Oldham, N. J., Bhattacharya, S., Gleadle, J. M., Ratcliffe, P. J., Pugh, C. W., and Schofield, C. J. (2002) 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.
(16) Johansson, C., Tumber, A., Che, K., Cain, P., Nowak, R., Gileadi, C., and Oppermann, U. (2014) The roles of Jumonji-type oxygenases in human disease. *Epigenomics 6*, 89–120.
(17) McDonough, M. A., Loenarz, C., Chowdhury, R., Clifton, I. J., and Schofield, C. J. (2010) Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr Opin Struct Biol 20*, 659–672.
(18) Aik, W. S., Chowdhury, R., Clifton, I. J., Hopkinson, R. J., Leissing, T., McDonough, M. A., Nowak, R., Schofield, C. J., and Walport, L. J. (2015) Introduction to structural studies on 2-oxoglutarate-dependent oxygenases and related enzymes, in *2-Oxoglutarate-dependent oxygenases* (Hausinger, R. P., and Schofield, C. J., Eds.) RSC Metall., pp 59–94. Royal Society of Chemistry, Cambridge.

(19) Hegg, E. L., and Que, Jr., L. (1997) The 2-His-1-carboxylate facial triad-An emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur J Biochem 250*, 625–629.
(20) Chowdhury, R., McDonough, M. A., Mecinović, J., Loenarz, C., Flashman, E., Hewitson, K. S., Domene, C., and Schofield, C. J. (2009) Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure 17*, 981–989.
(21) Markolovic, S., Leissing, T. M., Chowdhury, R., Wilkins, S. E., Lu, X., and Schofield, C. J. (2016) Structure–function relationships of human JmjC oxygenases—demethylases versus hydroxylases. *Curr Opin Struct Biol 41*, 62–72.
(22) Aik, W., McDonough, M. A., Thalhammer, A., Chowdhury, R., and Schofield, C. J. (2012) Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr Opin Struct Biol 22*, 691–700.
(23) Martinez, S., and Hausinger, R. P. (2015) Catalytic mechanisms of Fe(II)- and 2-oxoglutarate-dependent oxygenases. *J Biol Chem 290*, 20702–20711.
(24) Hausinger, R. P. (2015) Biochemical diversity of 2-oxoglutarate-dependent oxygenases, in *2-Oxoglutarate-dependent oxygenases* (Hausinger, R. P., and Schofield, C. J., Eds.) RSC Metall., pp 1–58. Royal Society of Chemistry, Cambridge.
(25) Taabazuing, C. Y., Fermann, J., Garman, S., and Knapp, M. J. (2016) Substrate promotes productive gas binding in the α-ketoglutarate-dependent oxygenase FIH. *Biochemistry 55*, 277–286.
(26) Price, J. C., Barr, E. W., Tirupati, B., Bollinger, Jr., J. M., and Krebs, C. (2003) The first direct characterization of a high-valent iron untermediate in the reaction of an α-ketoglutarate-dependent dioxygenase: a high-spin Fe(IV) complex in Taurine/α-ketoglutarate Dioxygenase (TauD) from Escherichia coli. *Biochemistry 42*, 7497–7508.
(27) Hoffart, L. M., Barr, E. W., Guyer, R. B., Bollinger, Jr., J. M., and Krebs, C. (2006) Direct spectroscopic detection of a C-H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. *Proc Natl Acad Sci 103*, 14738–14743.
(28) Proshlyakov, D. A., Henshaw, T. F., Monterosso, G. R., Ryle, M. J., and Hausinger, R. P. (2004) Direct detection of oxygen intermediates in the non-heme Fe enzyme Taurine/α-ketoglutarate dioxygenase. *J Am Chem Soc 126*, 1022–1023.
(29) Price, J. C., Barr, E. W., Glass, T. E., Krebs, C., and Bollinger, Jr., J. M. (2003) Evidence for hydrogen abstraction from C1 of Taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by Taurine:α-ketoglutarate Dioxygenase (TauD). *J Am Chem Soc 125*, 13008–13009.
(30) Yamane, K., Toumazou, C., Tsukada, Y., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006) JHDM2A, a JmjC-Containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell 125*, 483–495.
(31) Hopkinson, R. J., Hamed, R. B., Rose, N. R., Claridge, T. D. W., and Schofield, C. J. (2010) Monitoring the activity of 2-oxoglutarate dependent histone demethylases by NMR spectroscopy: direct observation of formaldehyde. *ChemBioChem 11*, 506–510.
(32) Myllylä, R., Kuuti-Savolainen, E.-R., and Kivirikko, K. I. (1978) The role of ascorbate in the prolyl hydroxylase reaction. *Biochem Biophys Res Commun 83*, 441–448.
(33) Flashman, E., Davies, S. L., Yeoh, K. K., and Schofield, C. J. (2010) Investigating the dependence of the hypoxia-inducible factor hydroxylases (factor inhibiting HIF and prolyl hydroxylase domain 2) on ascorbate and other reducing agents. *Biochem J 427*, 135–142.
(34) Kuiper, C., and Vissers, M. C. M. (2014) Ascorbate as a co-factor for Fe- and 2-oxoglutarate dependent dioxygenases: physiological activity in tumor growth and progression. *Front Oncol 4*, 1–11.
(35) Cheng, X., and Trievel, R. C. (2015) JmjC lysine demethylases, in *2-Oxoglutarate-dependent oxygenases* (Hausinger, R. P., and Schofield, C. J., Eds.) RSC Metall., pp 210–245. Royal Society of Chemistry, Cambridge.
(36) Vaz, F. M., and Wanders, R. J. A. (2002) Carnitine biosynthesis in mammals. *Biochem J 361*, 417–429.
(37) Zheng, G., and He, C. (2015) RNA demethylation by FTO and ALKBH5, in *2-Oxoglutarate-dependent oxygenases* (Hausinger, R. P., and Schofield, C. J., Eds.) RSC Metall., pp 263–274. Royal Society of Chemistry, Cambridge.

(38) Aravind, L., Zhang, D., and Iyer, L. M. (2015) The TET/JBP family of nucleic acid base-modifying 2-oxoglutarate and iron- dependent dioxygenases, in *2-Oxoglutarate-dependent oxygenases* (Hausinger, R. P., and Schofield, C. J., Eds.) RSC Metall., pp 289–308. Royal Society of Chemistry, Cambridge.
(39) Yamauchi, M., and Sricholpech, M. (2012) Lysine post-translational modifications of collagen. *Essays Biochem 52*, 113–133.
(40) Gjaltema, R. A. F., and Bank, R. A. (2017) Molecular insights into prolyl and lysyl hydroxylation of fibrillar collagens in health and disease. *Crit Rev Biochem Mol Biol 52*, 74–95.
(41) Hudson, D. M., and Eyre, D. R. (2013) Collagen prolyl 3-hydroxylation: a major role for a minor post-translational modification? *Connect Tissue Res 53*, 245–251.
(42) Berg, R. A., and Prockop, D. J. (1973) The thermal transition of a non-hydroxylated form of collagen. Evidence for a role of hydroxyproline in stabilizing the triple-helix of collagen. *Biochem Biophys Res Commun 52*, 115–120.
(43) Ogle, J. D., Arlinghaus, R. B., and Logan, M. A. (1962) 3-Hydroxyproline, a new amino acid of collagen. *J Biol Chem 237*, 3667–3673.
(44) Weis, M. A., Hudson, D. M., Kim, L., Scott, M., Wu, J.-J., and Eyre, D. R. (2010) Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly. *J Biol Chem 285*, 2580–2590.
(45) Jenkins, C. L., Bretscher, L. E., Guzei, I. A., and Raines, R. T. (2003) Effect of 3-hydroxyproline residues on collagen stability. *J Am Chem Soc 125*, 6422–6427.
(46) Butler, W. T., and Cunningham, L. W. (1966) Evidence for the linkage of a disaccharide to hydroxylysine in tropocollagen. *J Biol Chem 241*, 3882–3888.
(47) Pinnell, S. R., and Martin, G. R. (1968) The cross-linking of collagen and elastin: enzymatic conversion of lysine in peptide linkage to α-aminoadipic-δ-semialdehyde (allysine) by an extract from bone. *Biochemistry 61*, 708–716.
(48) Siegel, R. C., and Fu, J. C. C. (1976) Collagen cross-linking: purification and substrate specificity of lysyl oxidase. *J Biol Chem 251*, 5779–5785.
(49) Cabral, W. A., Chang, W., Barnes, A. M., Weis, M., Scott, M. A., Leikin, S., Makareeva, E., Kuznetsova, N. V, Rosenbaum, K. N., Tifft, C. J., Bulas, D. I., Kozma, C., Smith, P. A., Eyre, D. R., and Marini, J. C. (2007) Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. *Nat Genet 39*, 359–365.
(50) Grosso, G., Bei, R., Mistretta, A., Marventano, S., Calabrese, G., Masuelli, L., Giganti, M. G., Modesti, A., Galvano, F., and Gazzolo, D. (2013) Effects of vitamin C on health: a review of evidence. *Front Biosci 18*, 1017–1029.
(51) Drakenberg, T., Fernlund, P., Roepstorff, P., and Stenflo, J. (1983) β-Hydroxyaspartic acid in vitamin K-dependent protein C. *Proc Natl Acad Sci 80*, 1802–1806.
(52) McMullen, B. A., Fujikawa, K., Kisiel, W., Sasagawa, T., Howald, W. N., Kwa, E. Y., and Weinstein, B. (1983) Complete amino acid sequence of the light chain of human blood coagulation factor X: evidence for identification of residue 63 as β-hydroxyaspartic acid. *Biochemistry 22*, 2875–2884.
(53) Yang, H., Li, J., Tang, R., Li, J., Liu, Y., Ye, L., Shao, D., Jin, M., Huang, Q., and Shi, J. (2015) The aspartyl (asparaginyl) beta-hydroxylase in carcinomas. *Front Biosci 20*, 902–909.
(54) Stenflo, J., Holme, E., Lindstedt, S., Chandramouli, N., Tsai Huang, L. H., Tam, J. P., and Merrifield, R. B. (1989) Hydroxylation of aspartic acid in domains homologous to the epidermal growth factor precursor is catalyzed by a 2-oxoglutarate-dependent dioxygenase. *Proc Natl Acad Sci U S A 86*, 444–447.
(55) Gronke, R. S., Welsch, D. J., VanDusen, W. J., Garsky, V. M., Sardana, M. K., Stern, A. M., and Friedman, P. A. (1990) Partial purification and characterization of bovine liver aspartyl β-hydroxylase. *J Biol Chem 265*, 8558–8565.
(56) Sturla, L.-M., Tong, M., Hebda, N., Gao, J., Thomas, J.-M., Olsen, M., and de la Monte, S. M. (2016) Aspartate-β-hydroxylase (ASPH): a potential therapeutic target in human malignant gliomas. *Heliyon 2*, e00203.
(57) Patel, N., Khan, A. O., Mansour, A., Mohamed, J. Y., Al-Assiri, A., Haddad, R., Jia, X.,

Xiong, Y., Mégarbané, A., Traboulsi, E. I., and Alkuraya, F. S. (2014) Mutations in ASPH cause facial dysmorphism, lens dislocation, anterior-segment abnormalities, and spontaneous filtering blebs, or Traboulsi syndrome. *Am J Hum Genet 94*, 755–759.
(58) Masson, N., and Ratcliffe, P. J. (2014) Hypoxia signaling pathways in cancer metabolism: the importance of co-selecting interconnected physiological pathways. *Cancer Metab 2*, 3.
(59) Nakazawa, M. S., Keith, B., and Simon, M. C. (2016) Oxygen availability and metabolic adaptations. *Nat Rev Cancer 16*, 663–673.
(60) Meneses, A. M., and Wielockx, B. (2016) PHD2: from hypoxia regulation to disease progression. *Hypoxia 4*, 53–67.
(61) Wilkins, S. E., Flashman, E., Scotti, J. S., Hopkinson, R. J., Chowdhury, R., and Schofield, C. J. (2015) The role of 2-oxoglutarate-dependent oxygenases in hypoxia sensing, in *2-Oxoglutarate-dependent oxygenases* (Hausinger, R. P., and Schofield, C. J., Eds.) RSC Metall., pp 169–209. Royal Society of Chemistry, Cambridge.
(62) Jaakkola, P., Mole, D. R., Tian, Y.-M., Wilson, M. I., Gielbert, J., Gaskell, S. J., von Kriegsheim, A., Hebestreit, H. F., Mukherji, M., Schofield, C. J., Maxwell, P. H., Pugh, C. W., and Ratcliffe, P. J. (2001) Targeting of HIF-α to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science 292*, 468–472.
(63) Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J. M., Lane, W. S., and Kaelin Jr., W. G. (2001) HIFα targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science 292*, 464–468.
(64) Epstein, A. C. R., Gleadle, J. M., McNeill, L. A., Hewitson, K. S., O'Rourke, J., Mole, D. R., Mukherji, M., Metzen, E., Wilson, M. I., Dhanda, A., Tian, Y. M., Masson, N., Hamilton, D. L., Jaakkola, P., Barstead, R., Hodgkin, J., Maxwell, P. H., Pugh, C. W., Schofield, C. J., and Ratcliffe, P. J. (2001) C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell 107*, 43–54.
(65) Ivan, M., Haberberger, T., Gervasi, D. C., Michelson, K. S., Günzler, V., Kondo, K., Yang, H., Sorokina, I., Conaway, R. C., Conaway, J. W., and Kaelin, W. G. (2002) Biochemical purification and pharmacological inhibition of a mammalian prolyl hydroxylase acting on hypoxia-inducible factor. *Proc Natl Acad Sci U S A 99*, 13459–13464.
(66) Masson, N., Willam, C., Maxwell, P. H., Pugh, C. W., and Ratcliffe, P. J. (2001) Independent function of two destruction domains in hypoxia-inducible factor-α chains activated by prolyl hydroxylation. *EMBO J 20*, 5197–5206.
(67) Ehrismann, D., Flashman, E., Genn, D. N., Mathioudakis, N., Hewitson, K. S., Ratcliffe, P. J., and Schofield, C. J. (2007) Studies on the activity of the hypoxia-inducible-factor hydroxylases using an oxygen consumption assay. *Biochem J 401*, 227–34.
(68) Lando, D., Peet, D. J., Gorman, J. J., Whelan, D. A., Whitelaw, M. L., and Bruick, R. K. (2002) FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev 16*, 1466–1471.
(69) Lando, D., Peet, D. J., Whelan, D. A., Gorman, J. J., and Whitelaw, M. L. (2002) Asparagine hydroxylation of the HIF transactivation domain: a hypoxic switch. *Science 295*, 858–861.
(70) Koivunen, P., Hirsila, M., Günzler, V., Kivirikko, K. I., and Myllyharju, J. (2004) Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases. *J Biol Chem 279*, 9899–9904.
(71) Flashman, E., Hoffart, L. M., Hamed, R. B., Bollinger Jr., J. M., Krebs, C., and Schofield, C. J. (2010) Evidence for the slow reaction of hypoxia-inducible factor prolyl hydroxylase 2 with oxygen. *FEBS J 277*, 4089–4099.
(72) Tian, Y.-M., Yeoh, K. K., Lee, M. K., Eriksson, T., Kessler, B. M., Kramer, H. B., Edelmann, M. J., Willam, C., Pugh, C. W., Schofield, C. J., and Ratcliffe, P. J. (2011) Differential sensitivity of hypoxia inducible factor hydroxylation sites to hypoxia and hydroxylase inhibitors. *J Biol Chem 286*, 13041–13051.
(73) Cockman, M. E., Lancaster, D. E., Stolze, I. P., Hewitson, K. S., McDonough, M. A., Coleman, M. L., Coles, C. H., Yu, X., Hay, R. T., Ley, S. C., Pugh, C. W., Oldham, N. J., Masson, N., Schofield, C. J., and Ratcliffe, P. J. (2006) Posttranslational hydroxylation of ankyrin repeats in IκB proteins by the hypoxia-inducible factor (HIF) asparaginyl hydroxylase,

factor inhibiting HIF (FIH). *Proc Natl Acad Sci 103*, 14767–14772.
(74) Cockman, M. E., Webb, J. D., and Ratcliffe, P. J. (2009) FIH-dependent asparaginyl hydroxylation of ankyrin repeat domain-containing proteins. *Ann N Y Acad Sci 1177*, 9–18.
(75) Yang, M., Chowdhury, R., Ge, W., Hamed, R. B., McDonough, M. A., Claridge, T. D. W., Kessler, B. M., Cockman, M. E., Ratcliffe, P. J., and Schofield, C. J. (2011) Factor-inhibiting hypoxia-inducible factor (FIH) catalyses the post-translational hydroxylation of histidinyl residues within ankyrin repeat domains. *FEBS J 278*, 1086–1097.
(76) Hardy, A. P., Prokes, I., Kelly, L., Campbell, I. D., and Schofield, C. J. (2009) Asparaginyl β-hydroxylation of proteins containing ankyrin repeat domains influences their stability and function. *J Mol Biol 392*, 994–1006.
(77) Kelly, L., McDonough, M. A., Coleman, M. L., Ratcliffe, P. J., and Schofield, C. J. (2009) Asparagine β-hydroxylation stabilizes the ankyrin repeat domain fold. *Mol BioSyst 5*, 52–58.
(78) Maxwell, P. H., and Eckardt, K. (2015) HIF prolyl hydroxylase inhibitors for the treatment of renal anaemia and beyond. *Nat Rev Nephrol 12*, 157–168.
(79) Chan, M. C., Holt-Martyn, J. P., Schofield, C. J., and Ratcliffe, P. J. (2016) Pharmacological targeting of the HIF hydroxylases – A new field in medicine development. *Mol Aspects Med 47–48*, 54–75.
(80) Takeuchi, T., Yamazaki, Y., Katoh-Fukui, Y., Tsuchiya, R., Kondo, S., Motoyama, J., and Higashinakagawa, T. (1995) Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes Dev 9*, 1211–1222.
(81) Balciunas, D., and Ronne, H. (2000) Evidence of domain swapping within the jumonji family of transcription factors. *Trends Biochem Sci 25*, 274–276.
(82) Clissold, P. M., and Ponting, C. P. (2001) JmjC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A2β. *Trends Biochem Sci 26*, 7–9.
(83) Dann 3rd, C. E., Bruick, R. K., and Deisenhofer, J. (2002) Structure of factor-inhibiting hypoxia-inducible factor 1: an asparaginyl hydroxylase involved in the hypoxic response pathway. *Proc Natl Acad Sci 99*, 15351–15356.
(84) Elkins, J. M., Hewitson, K. S., McNeill, L. A., Seibel, J. F., Schlemminger, I., Pugh, C. W., Ratcliffe, P. J., and Schofield, C. J. (2003) Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1α. *J Biol Chem 278*, 1802–1806.
(85) Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., and Zhang, Y. (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature 439*, 811–816.
(86) Cloos, P. A. C., Christensen, J., Agger, K., Maiolica, A., Rappsilber, J., Antal, T., Hansen, K. H., and Helin, K. (2006) The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. *Nature 442*, 307–311.
(87) Whetstine, J. R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M., and Shi, Y. (2006) Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell 125*, 467–481.
(88) Chen, Z., Zang, J., Whetstine, J., Hong, X., Davrazou, F., Kutateladze, T. G., Simpson, M., Mao, Q., Pan, C.-H., Dai, S., Hagman, J., Hansen, K., Shi, Y., and Zhang, G. (2006) Structural insights into histone demethylation by JMJD2 family members. *Cell 125*, 691–702.
(89) Sengoku, T., and Yokoyama, S. (2011) Structural basis for histone H3 Lys 27 demethylation by UTX/KDM6A. *Genes Dev 25*, 2266–2277.
(90) Walport, L. J., Hopkinson, R. J., Vollmar, M., Madden, S. K., Gileadi, C., Oppermann, U., Schofield, C. J., and Johansson, C. (2014) Human UTY(KDM6C) is a male-specific Nε-methyl lysyl demethylase. *J Biol Chem 289*, 18302–18313.
(91) Annunziato, A. T. (2008) DNA packaging: nucleosome and chromatin. *Nat Educ 1*, 26.
(92) Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature 389*, 251–260.
(93) Rothbart, S. B., and Strahl, B. D. (2014) Interpreting the language of histone and DNA modifications. *Biochim Biophys Acta 1839*, 627–643.
(94) Shi, Y. (2007) Histone lysine demethylases: emerging roles in development, physiology and disease. *Nat Rev Genet 8*, 829–833.

(95) Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., and Shi, Y. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell 119*, 941–953.
(96) Maes, T., Carceller, E., Salas, J., Ortega, A., and Buesa, C. (2015) Advances in the development of histone lysine demethylase inhibitors. *Curr Opin Pharmacol 23*, 52–60.
(97) McAllister, T. E., England, K. S., Hopkinson, R. J., Brennan, P. E., Kawamura, A., and Schofield, C. J. (2016) Recent progress in histone demethylase inhibitors. *J Med Chem 59*, 1308–1329.
(98) Licheva, M., Neudolt, A., and Bennani-Baiti, I. M. (2016) Clinical trials of lysine demethylase inhibitors in cancer. *Cancer Epigenetics Drug Database - Clin data, v11 Cancer Epigenetics Soc*.
(99) Poulard, C., Corbo, L., and Le Romancer, M. (2016) Protein arginine methylation/demethylation and cancer. *Oncotarget 7*, 67532–67550.
(100) Walport, L. J., Hopkinson, R. J., Chowdhury, R., Schiller, R., Ge, W., Kawamura, A., and Schofield, C. J. (2016) Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat Commun 7*, 11974.
(101) De Castro, E., Sigrist, C., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P., Gasteiger, E., Bairoch, A., and Hulo, N. (2006) ScanProsite: detection of PROSITE signature matches and ProRule- associated functional and structural residues in proteins. *Nucleic Acids Res 24*, W362-365.
(102) Chowdhury, R., Sekirnik, R., Brissett, N. C., Krojer, T., Ho, C., Ng, S. S., Clifton, I. J., Ge, W., Kershaw, N. J., Fox, G. C., Muniz, J. R. C., Vollmar, M., Phillips, C., Pilka, E. S., Kavanagh, K. L., von Delft, F., Oppermann, U., McDonough, M. A., Doherty, A. J., and Schofield, C. J. (2014) Ribosomal oxygenases are structurally conserved from prokaryotes to humans. *Nature 510*, 422–426.
(103) Kato, M., Araiso, Y., Noma, A., Nagao, A., Suzuki, T., Ishitani, R., and Nureki, O. (2011) Crystal structure of a novel JmjC-domain-containing protein, TYW5, involved in tRNA modification. *Nucleic Acids Res 39*, 1576–1585.
(104) Mahon, P. C., Hirota, K., and Semenza, G. L. (2001) FIH-1: a novel protein that interacts with HIF-1α and VHL to mediate repression of HIF-1 transcriptional activity. *Genes Dev 15*, 2675–2686.
(105) Chang, B., Chen, Y., Zhao, Y., and Bruick, R. K. (2007) JMJD6 is a histone arginine demethylase. *Science 318*, 444–447.
(106) Böttger, A., Islam, M. S., Chowdhury, R., Schofield, C. J., and Wolf, A. (2015) The oxygenase Jmjd6 - a case study in conflicting assignments. *Biochem J 468*, 191–202.
(107) Webby, C. J., Wolf, A., Gromak, N., Dreger, M., Kramer, H., Kessler, B., Nielsen, M. L., Schmitz, C., Butler, D. S., Yates 3rd, J. R., Delahunty, C. M., Hahn, P., Lengeling, A., Mann, M., Proudfoot, N. J., Schofield, C. J., and Böttger, A. (2009) Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science 325*, 90–93.
(108) Heim, A., Grimm, C., Müller, U., Häußler, S., Mackeen, M. M., Merl, J., Hauck, S. M., Kessler, B. M., Schofield, C. J., Wolf, A., and Böttger, A. (2014) Jumonji domain containing protein 6 (Jmjd6) modulates splicing and specifically interacts with arginine–serine-rich (RS) domains of SR- and SR-like proteins. *Nucleic Acids Res 42*, 7833–7850.
(109) Mantri, M., Loik, N. D., Hamed, R. B., Claridge, T. D. W., McCullagh, J. S. O., and Schofield, C. J. (2011) The 2-oxoglutarate-dependent oxygenase JMJD6 catalyses oxidation of lysine residues to give 5S-hydroxylysine residues. *ChemBioChem 12*, 531–534.
(110) Yi, J., Shen, H.-F., Qiu, J.-S., Huang, M.-F., Zhang, W.-J., Ding, J.-C., Zhu, X.-Y., Zhou, Y., Fu, X.-D., and Lui, W. (2016) JMJD6 and U2AF65 co-regulate alternative splicing in both JMJD6 enzymatic activity dependent and independent manner. *Nucleic Acids Res [Epub ahead of print]*.
(111) Unoki, M., Masuda, A., Dohmae, N., Arita, K., Yoshimatsu, M., Iwai, Y., Fukui, Y., Ueda, K., Hamamoto, R., Shirakawa, M., Sasaki, H., and Nakamura, Y. (2013) Lysyl 5-hydroxylation, a novel histone modification, by jumonji domain containing 6 (JMJD6). *J Biol Chem 288*, 6053–6062.
(112) Liu, W., Ma, Q., Wong, K., Li, W., Ohgi, K., Zhang, J., Aggarwal, A. K., and Rosenfeld,

M. G. (2013) Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. *Cell 155*, 1581–1595.
(113) Zhang, J., Ni, S.-S., Zhao, W.-L., Dong, X.-C., and Wang, J.-L. (2013) High expression of JMJD6 predicts unfavorable survival in lung adenocarcinoma. *Tumor Biol 34*, 2397–2401.
(114) Poulard, C., Rambaud, J., Lavergne, E., Jacquemetton, J., Renoir, J.-M., Trédan, O., Chabaud, S., Treilleux, I., Corbo, L., and Le Romancer, M. (2015) Role of JMJD6 in breast tumourigenesis. *PLoS One* (Migliaccio, A., Ed.) *10*, e0126181.
(115) Wang, F., He, L., Huangyang, P., Liang, J., Si, W., Yan, R., Han, X., Liu, S., Gui, B., Li, W., Miao, D., Jing, C., Liu, Z., Pei, F., Sun, L., and Shang, Y. (2014) JMJD6 promotes colon carcinogenesis through negative regulation of p53 by hydroxylation. *PLoS Biol* (Manfredi, J., Ed.) *12*, e1001819.
(116) Noma, A., Ishitani, R., Kato, M., Nagao, A., Nureki, O., and Suzuki, T. (2010) Expanding role of the Jumonji C domain as an RNA hydroxylase. *J Biol Chem 285*, 34503–34507.
(117) Waas, W. F., Druzina, Z., Hanan, M., and Schimmel, P. (2007) Role of a tRNA base modification and its precursors in frameshifting in eukaryotes. *J Biol Chem 282*, 26026–26034.
(118) Songe-Møller, L., van den Born, E., Leihne, V., Vågbø, C. B., Kristoffersen, T., Krokan, H. E., Kirpekar, F., Falnes, P. Ø., and Klungland, A. (2010) Mammalian ALKBH8 Possesses tRNA methyltransferase activity required for the biogenesis of multiple wobble uridine modifications implicated in translational decoding. *Mol Cell Biol 30*, 1814–1827.
(119) Fu, Y., Dai, Q., Zhang, W., Ren, J., Pan, T., and He, C. (2010) The AlkB domain of mammalian ABH8 catalyzes hydroxylation of 5-methoxycarbonylmethyluridine at the wobble position of tRNA. *Angew Chemie Int Ed 49*, 8885–8888.
(120) van den Born, E., Vågbø, C. B., Songe-Møller, L., Leihne, V., Lien, G. F., Leszczynska, G., Malkiewicz, A., Krokan, H. E., Kirpekar, F., Klungland, A., and Falnes, P. Ø. (2011) ALKBH8-mediated formation of a novel diastereomeric pair of wobble nucleosides in mammalian tRNA. *Nat Commun 2*, 172.
(121) Lu, Y., Chang, Q., Zhang, Y., Beezhold, K., Rojanasakul, Y., Zhao, H., Castranova, V., Shi, X., and Chen, F. (2009) Lung cancer-associated JmjC domain protein mdig suppresses formation of tri-methyl lysine 9 of histone H3. *Cell Cycle 8*, 2101–2109.
(122) Ge, W., Wolf, A., Feng, T., Ho, C., Sekirnik, R., Zayer, A., Granatino, N., Cockman, M. E., Loenarz, C., Loik, N. D., Hardy, A. P., Claridge, T. D. W., Hamed, R. B., Chowdhury, R., Gong, L., Robinson, C. V, Trudgian, D. C., Jiang, M., Mackeen, M. M., Mccullagh, J. S., Gordiyenko, Y., Thalhammer, A., Yamamoto, A., Yang, M., Liu-Yi, P., Zhang, Z., Schmidt-Zachmann, M., Kessler, B. M., Ratcliffe, P. J., Preston, G. M., Coleman, M. L., and Schofield, C. J. (2012) Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat Chem Biol 8*, 960–962.
(123) Teichmann, M., Dumay-Odelot, H., and Fribourg, S. (2012) Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes. *Transcription 3*, 2–7.
(124) Polacek, N., and Mankin, A. S. (2005) The ribosomal peptidyl transferase center: structure, function, evolution, inhibition. *Crit Rev Biochem Mol Biol 40*, 285–311.
(125) Thakur, C., Lu, Y., Sun, J., Yu, M., Chen, B., and Chen, F. (2014) Increased expression of mdig predicts poorer survival of the breast cancer patients. *Gene 535*, 218–224.
(126) Tsuneoka, M., Koda, Y., Soejima, M., Teye, K., and Kimura, H. (2002) A novel Myc target gene, mina53, that is involved in cell proliferation. *J Biol Chem 277*, 35450–35459.
(127) Suzuki, C., Takahashi, K., Hayama, S., Ishikawa, N., Kato, T., Ito, T., Tsuchiya, E., Nakamura, Y., and Daigo, Y. (2007) Identification of Myc-associated protein with JmjC domain as a novel therapeutic target oncogene for lung cancer. *Mol Cancer Ther 6*, 542–551.
(128) Xing, J., Wang, K., Liu, P.-W., Miao, Q., and Chen, X.-Y. (2014) Mina53, a novel molecular marker for the diagnosis and prognosis of gastric adenocarcinoma. *Oncol Rep 31*, 634–640.
(129) Singleton, R. S., Liu-Yi, P., Formenti, F., Ge, W., Sekirnik, R., Fischer, R., Adam, J., Pollard, P. J., Wolf, A., Thalhammer, A., Loenarz, C., Flashman, E., Yamamoto, A., Coleman, M. L., Kessler, B. M., Wappner, P., Schofield, C. J., Ratcliffe, P. J., and Cockman, M. E.

(2014) OGFOD1 catalyzes prolyl hydroxylation of RPS23 and is involved in translation control and stress granule formation. *Proc Natl Acad Sci 111*, 4031–4036.
(130) Loenarz, C., Sekirnik, R., Thalhammer, A., Ge, W., Spivakovsky, E., Mackeen, M. M., McDonough, M. A., Cockman, M. E., Kessler, B. M., Ratcliffe, P. J., Wolf, A., and Schofield, C. J. (2014) Hydroxylation of the eukaryotic ribosomal decoding center affects translational accuracy. *Proc Natl Acad Sci 111*, 4019–4024.
(131) Katz, M. J., Acevedo, J. M., Loenarz, C., Galagovsky, D., Liu-Yi, P., Perez-Pepe, M., Thalhammer, A., Sekirnik, R., Ge, W., Melani, M., Thomas, M. G., Simonetta, S., Boccaccio, G. L., Schofield, C. J., Cockman, M. E., Ratcliffe, P. J., and Wappner, P. (2014) Sudestada1, a Drosophila ribosomal prolyl-hydroxylase required for mRNA translation, cell homeostasis, and organ growth. *Proc Natl Acad Sci 111*, 4025–4030.
(132) Wilson, D. N. (2014) Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat Rev Microbiol 12*, 35–48.
(133) Feng, T., Yamamoto, A., Wilkins, S. E., Sokolova, E., Yates, L. A., Münzel, M., Singh, P., Hopkinson, R. J., Fischer, R., Cockman, M. E., Shelley, J., Trudgian, D. C., Schödel, J., Mccullagh, J. S. O., Ge, W., Kessler, B. M., Gilbert, R. J., Frolova, L. Y., Alkalaeva, E., Ratcliffe, P. J., Schofield, C. J., and Coleman, M. L. (2014) Optimal translational termination requires C4 lysyl hydroxylation of eRF1. *Mol Cell 53*, 645–654.
(134) Chavatte, L., Seit-Nebi, A., Dubovaya, V., and Favre, A. (2002) The invariant uridine of stop codons contacts the conserved NIKSR loop of human eRF1 in the ribosome. *EMBO J 21*, 5302–5311.
(135) Frolova, L., Seit-Nebi, A., and Kisselev, L. E. V. (2002) Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. *RNA 8*, 129–136.
(136) Scotti, J. S., Leung, I. K. H., Ge, W., Bentley, M. A., Paps, J., Kramer, H. B., Lee, J., Aik, W., Choi, H., Paulsen, S. M., Bowman, L. A. H., Loik, N. D., Horita, S., Ho, C., Kershaw, N. J., Tang, C. M., Claridge, T. D. W., Preston, G. M., McDonough, M. A., and Schofield, C. J. (2014) Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylation. *Proc Natl Acad Sci 111*, 13331–13336.
(137) Le Quesne, J. P. C., Spriggs, K. A., Bushell, M., and Willis, A. E. (2010) Dysregulation of protein synthesis and disease. *J Pathol 220*, 140–151.
(138) Horton, J. R., Upadhyay, A. K., Hashimoto, H., Zhang, X., and Cheng, X. (2011) Structural basis for human PHF2 Jumonji domain interaction with metal ions. *J Mol Biol 406*, 1–8.
(139) de Ligt, J., Willemsen, M. H., van Bon, B. W. M., Kleefstra, T., Yntema, H. G., Kroes, T., Vulto-van Silfhout, A. T., Koolen, D. A., de Vries, P., Gilissen, C., del Rosario, M., Hoischen, A., Scheffer, H., de Vries, B. B. A., Brunner, H. G., Veltman, J. A., and Vissers, L. E. L. M. (2012) Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med 367*, 1921–1929.
(140) Matsunami, N., Hensel, C. H., Baird, L., Stevens, J., Otterud, B., Leppert, T., Varvil, T., Hadley, D., Glessner, J. T., Pellegrino, R., Kim, C., Thomas, K., Wang, F., Otieno, F. G., Ho, K., Christensen, G. B., Li, D., Prekeris, R., Lambert, C. G., Hakonarson, H., and Leppert, M. F. (2014) Identification of rare DNA sequence variants in high-risk autism families and their prevalence in a large case/control population. *Mol Autism 5*, 5.
(141) Cheng, Y., Wang, Y., Li, J., Chang, I., and Wang, C.-Y. (2017) A novel read-through transcript JMJD7-PLA2G4B regulates head and neck squamous cell carcinoma cell proliferation and survival. *Oncotarget 8*, 1972–1982.
(142) International Human Genome Sequencing Consortium. (2004) Finishing the euchromatic sequence of the human genome. *Nature 431*, 931–945.
(143) Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., Bourexis, D., Brister, J. R., and Bryant, S. H. (2016) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res 44*, D7–D19.
(144) Youn, M. Y., Yokoyama, A., Fujiyama-Nakamura, S., Ohtake, F., Minehata, K., Yasuda, H., Suzuki, T., Kato, S., and Imai, Y. (2012) JMJD5, a Jumonji C (JmjC) domain-containing protein, negatively regulates osteoclastogenesis by facilitating NFATc1 protein degradation. *J*

Biol Chem 287, 12994–13004.
(145) Wu, C., Jin, X., Tsueng, G., Afrasiabi, C., and Su, A. I. (2016) BioGPS: building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Res 44*, D313–D316.
(146) Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P.-H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., and Pontén, F. (2015) Tissue-based map of the human proteome. *Science 347*, 1260419.
(147) Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., and Schultz, N. (2012) The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov 2*, 401–404.
(148) Gao, J., Aksoy, B., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., and Schultz, N. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science 6*, 1–20.
(149) Szász, A. M., Lánczky, A., Nagy, Á., Förster, S., Hark, K., Green, J. E., Boussioutas, A., Busuttil, R., Szabó, A., and Győrffy, B. (2016) Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. *Oncotarget 7*, 49322–49333.
(150) The Cancer Genome Atlas Research Network. (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature 511*, 543–550.
(151) Lewis, J. K., Wei, J., and Siuzdak, G. (2000) Matrix-assisted laser desorption/ionization mass spectrometry in peptide and protein analysis, in *Encyclopedia of Analytical Chemistry* (Meyers, R. A., Ed.), pp 1–14. John Wiley & Sons Ltd, Chichester.
(152) Marvin, L. F., Roberts, M. A., and Fay, L. B. (2003) Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry. *Clin Chim Acta 337*, 11–21.
(153) Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., and Schwede, T. (2008) Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc 4*, 1–13.
(154) Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., and Schwede, T. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res 42*, W252–W258.
(155) Lee, C., Kim, S. J., Jeong, D. G., Lee, S. M., and Ryu, S. E. (2003) Structure of human FIH-1 reveals a unique active site pocket and interaction sites for HIF-1 and von Hippel-Lindau. *J Biol Chem 278*, 7558–7563.
(156) Del Rizzo, P. A., Krishnan, S., and Trievel, R. C. (2012) Crystal structure and functional analysis of JMJD5 indicate an alternate specificity and function. *Mol Cell Biol 32*, 4044–4052.
(157) Wang, H., Zhou, X., Wu, M., Wang, C., Zhang, X., Tao, Y., Chen, N., and Zang, J. (2013) Structure of the JmjC-domain-containing protein JMJD5. *Acta Crystallogr D Biol Crystallogr 69*, 1911–1920.
(158) Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol 7*, 539.
(159) Rose, N. R., McDonough, M. A., King, O. N. F., Kawamura, A., and Schofield, C. J. (2011) Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev 40*, 4364–4397.
(160) Cockman, M. E., Webb, J. D., Kramer, H. B., Kessler, B. M., and Ratcliffe, P. J. (2009) Proteomics-based identification of novel factor inhibiting hypoxia-inducible factor (FIH) substrates indicates widespread asparaginyl hydroxylation of ankyrin repeat domain-containing proteins. *Mol Cell Proteomics 8*, 535–546.
(161) Bornhorst, J. A., and Falke, J. J. (2000) Purification of proteins using polyhistidine

affinity tags. *Methods Enzymol 326*, 245–254.
(162) Gileadi, O., Burgess-Brown, N. A., Colebrook, S. M., Berridge, G., Savitsky, P., Smee, C. E. A., Loppnau, P., Johansson, C., Salah, E., and Pantic, N. H. (2008) High throughput production of recombinant human proteins for crystallography, in *Methods in Molecular Biology* (Kobe, B., Guss, M., and Huber, T., Eds.), pp 221–246. Humana Press, Totowa, NJ.
(163) Hausinger, R. P. (2004) Fe(II)/α-ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol Biol 39*, 21–68.
(164) Cunliffe, C. J., Franklin, T. J., Hales, N. J., and Hill, G. B. (1992) Novel inhibitors of prolyl 4-hydroxylase. 3. Inhibition by the substrate analogue N-oxaloglycine and its derivatives. *J Med Chem 35*, 2652–2658.
(165) McNeill, L. A., Flashman, E., Buck, M. R. G., Hewitson, K. S., Clifton, I. J., Jeschke, G., Claridge, T. D. W., Ehrismann, D., Oldham, N. J., and Schofield, C. J. (2005) Hypoxia-inducible factor prolyl hydroxylase 2 has a high affinity for ferrous iron and 2-oxoglutarate. *Mol Biosyst 1*, 321–324.
(166) Francis, S. M., Gas, M.-E., Daugeron, M.-C., Bravo, J., and Seraphin, B. (2012) Rbg1-Tma46 dimer structure reveals new functional domains and their role in polysome recruitment. *Nucleic Acids Res 40*, 11100–11114.
(167) Wilkins, S. E., Karttunen, S., Hampton-Smith, R. J., Murchland, I., Chapman-Smith, A., and Peet, D. J. (2012) Factor inhibiting HIF (FIH) recognizes distinct molecular features within hypoxia-inducible factor-α (HIF-α) versus ankyrin repeat substrates. *J Biol Chem 287*, 8769–8781.
(168) Song, H., Mugnier, P., Das, A. K., Webb, H. M., Evans, D. R., Tuite, M. F., Hemmings, B. A., and Barford, D. (2000) The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell 100*, 311–321.
(169) Hewitson, K. S., Holmes, S. L., Ehrismann, D., Hardy, A. P., Chowdhury, R., Schofield, C. J., and McDonough, M. A. (2008) Evidence that two enzyme-derived histidine ligands are sufficient for iron binding and catalysis by factor inhibiting HIF (FIH). *J Biol Chem 283*, 25971–25978.
(170) Rutherfurd, S. M., and Gilani, G. S. (2009) Amino acid analysis. *Curr Protoc Protein Sci 11.9*, 1–37.
(171) Salazar, C., Armenta, J. M., and Shulaev, V. (2012) An UPLC-ESI-MS/MS assay using 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate derivatization for targeted amino acid analysis. *Metabolites 2*, 398–428.
(172) Leśniak, R. K., Markolovic, S., Tars, K., and Schofield, C. J. (2017) Human carnitine biosynthesis proceeds via (2S,3S)-3-hydroxy-N ε -trimethyllysine. *Chem. Commun. 53*, 440–442.
(173) Li, B., and Trueb, B. (2000) DRG represents a family of two closely related GTP-binding proteins. *Biochim Biophys Acta 1491*, 196–204.
(174) Leipe, D. D., Wolf, Y. I., Koonin, E. V., and Aravind, L. (2002) Classification and evolution of P-loop GTPases and related ATPases. *J Mol Biol 317*, 41–72.
(175) Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002) The Michaelis-Menten model accounts for the kinetic properties of many enzymes, in *Biochemistry* 5th ed. W H Freeman, New York.
(176) McNeill, L. A., Bethge, L., Hewitson, K. S., and Schofield, C. J. (2005) A fluorescence-based assay for 2-oxoglutarate-dependent oxygenases. *Anal Biochem 336*, 125–131.
(177) Williams, S. T., Walport, L. J., Hopkinson, R. J., Madden, S. K., Chowdhury, R., Schofield, C. J., and Kawamura, A. (2014) Studies on the catalytic domains of multiple JmjC oxygenases using peptide substrates. *Epigenetics 9*, 1596–1603.
(178) Horita, S., Scotti, J. S., Thinnes, C., Mottaghi-Taromsari, Y. S., Thalhammer, A., Ge, W., Aik, W., Loenarz, C., Schofield, C. J., and McDonough, M. A. (2015) Structure of the ribosomal oxygenase OGFOD1 provides insights into the regio- and stereoselectivity of prolyl hydroxylases. *Structure 23*, 639–652.
(179) Reid, D. G., MacLachlan, L. K., Edwards, A. J., Hubbard, J. A., and Sweeney, P. J. (1997) Introduction to the NMR of proteins. *Methods Mol Biol 60*, 1–28.
(180) Sinha, K. M., Yasuda, H., Coombes, M. M., Dent, S. Y. R., and de Crombrugghe, B.

(2010) Regulation of the osteoblast-specific transcription factor Osterix by NO66, a Jumonji family histone demethylase. *EMBO J 29*, 68–79.
(181) Hsia, D. A., Tepper, C. G., Pochampalli, M. R., Hsia, E. Y. C., Izumiya, C., Huerta, S. B., Wright, M. E., Chen, H.-W., Kung, H.-J., and Izumiya, Y. (2010) KDM8, a H3K36me2 histone demethylase that acts in the cyclin A1 coding region to regulate cancer cell proliferation. *Proc Natl Acad Sci U S A 107*, 9671–9676.
(182) Zentner, G. E., and Henikoff, S. (2013) Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol 20*, 259–266.
(183) Ishikawa, K., Azuma, S., Ikawa, S., Semba, K., and Inoue, J. I. (2005) Identification of DRG family regulatory proteins (DFRPs): specific regulation of DRG1 and DRG2. *Genes to Cells 10*, 139–150.
(184) Witkop, B. (1956) The application of Hudson's lactone rule to γ- and δ-hydroxyamino acids and the question of the configuration of δ-hydroxy-L-lysine from collagen. *Experientia XII*, 372–374.
(185) Kapuscinski, J. (1995) DAPI: a DNA-specific fluorescent probe. *Biotech Histochem 70*, 220–233.
(186) Bourne, H. R., Sanders, D. A., and McCormick, F. (1991) The GTPase superfamily: conserved structure and molecular mechanism. *Nature 349*, 117–127.
(187) Sazuka, T., Tomooka, Y., Ikawa, Y., Noda, M., and Kumar, S. (1992) DRG: a novel developmentally regulated GTP-binding protein. *Biochem Biophys Res Commun 189*, 363–370.
(188) Kumar, S., Iwao, M., Yamagishi, T., Noda, M., and Asashima, M. (1993) Expression of GTP-binding protein gene drg during Xenopus laevis development. *Int J Dev Biol 37*, 539–546.
(189) Etheridge, N., Trusov, Y., Verbelen, J. P., and Botella, J. R. (1999) Characterization of ATDRG1, a member of a new class of GTP-binding proteins in plants. *Plant Mol Biol 39*, 1113–1126.
(190) Sazuka, T., Kinoshita, M., Tomooka, Y., Ikawa, Y., Noda, M., and Kumar, S. (1992) Expression of DRG during murine embryonic development. *Biochem Biophys Res Commun 189*, 371–377.
(191) Ishikawa, K., Azuma, S., Ikawa, S., Morishita, Y., Gohda, J., Akiyama, T., Semba, K., and Inoue, J. (2003) Cloning and characterization of Xenopus laevis drg2, a member of the developmentally regulated GTP-binding protein subfamily. *Gene 322*, 105–112.
(192) Wei, D., Yao, J., Yang, X., Cheng, L., Lu, D., and Xue, J. (2004) Molecular cloning and expression of two closely related GTP-binding proteins from zebrafish. *DNA Seq 15*, 246–250.
(193) Devitt, M. L., Maas, K. J., and Stafstrom, J. P. (1999) Characterization of DRGs, developmentally regulated GTP-binding proteins, from pea and Arabidopsis. *Plant Mol Biol 39*, 75–82.
(194) Lu, L., Lv, Y., Dong, J., Hu, S., and Peng, R. (2016) DRG1 is a potential oncogene in lung adenocarcinoma and promotes tumor progression via spindle checkpoint signaling regulation. *Oncotarget 7*, 72795–72806.
(195) Jang, S. H., Kim, A.-R., Park, N.-H., Park, J. W., and Han, I.-S. (2016) DRG2 regulates G2/M progression via the cyclin B1-Cdk1 complex. *Mol Cells 39*, 699–704.
(196) Daugeron, M.-C., Prouteau, M., Lacroute, F., and Séraphin, B. (2011) The highly conserved eukaryotic DRG factors are required for efficient translation in a manner redundant with the putative RNA helicase Slh1. *Nucleic Acids Res 39*, 2221–2233.
(197) Nelson, B. J., Maas, K. J., Dekeyser, J. L., and Stafstrom, J. P. (2009) Association of DRG1 and DRG2 with ribosomes from pea, arabidopsis, and yeast. *Int J Plant Sci 170*, 834–844.
(198) Ishikawa, K., Akiyama, T., Ito, K., Semba, K., and Inoue, J. (2009) Independent stabilizations of polysomal Drg1/Dfrp1 complex and non-polysomal Drg2/Dfrp2 complex in mammalian cells. *Biochem Biophys Res Commun 390*, 552–556.
(199) Kiniwa, Y., Li, J., Wang, M., Sun, C., Lee, J. E., Wang, R.-F., and Wang, H. Y. (2015) Identification of DRG-1 as a melanoma-associated antigen recognized by CD4+ Th1 cells. *PLoS One 10*, e0124094.
(200) Delassus, G. S., Cho, H., and Eliceiri, G. L. (2011) New signaling pathways from cancer progression modulators to mRNA expression of matrix metalloproteinases in breast cancer

cells. *J Cell Physiol 226*, 3378–3384.
(201) Ko, M. S., Lee, U. H., Kim, S. I., Kim, H. J., Park, J. J., Cha, S. J., Kim, S. B., Song, H., Chung, D. K., Han, I. S., Kwack, K., and Park, J. W. (2004) Overexpression of DRG2 suppresses the growth of Jurkat T cells but does not induce apoptosis. *Arch Biochem Biophys 422*, 137–144.
(202) Vlangos, C. N., Das, P., Patel, P. I., and Elsea, S. H. (2000) Assignment of developmentally regulated GTP-binding protein (DRG2) to human chromosome band 17p11.2 with somatic cell hybrids and localization to the Smith-Magenis syndrome critical interval. *Cytogenet Cell Genet 88*, 283–285.
(203) Kavaliauskas, D., Nissen, P., and Knudsen, C. R. (2012) The busiest of all ribosomal assistants: elongation factor Tu. *Biochemistry 51*, 2642–2651.
(204) Myllyharju, J., and Kivirikko, K. I. (2001) Collagens and collagen-related diseases. *Ann Med 33*, 7–21.
(205) Azevedo, C., and Saiardi, A. (2016) Why always lysine? The ongoing tale of one of the most modified amino acids. *Adv Biol Regul 60*, 144–150.
(206) Bannister, A. J., and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Res 21*, 381–395.
(207) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol 276*, 307–326.
(208) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr 40*, 658–674.
(209) Mantri, M., Krojer, T., Bagg, E. A., Webby, C. J., Butler, D. S., Kochan, G., Kavanagh, K. L., Oppermann, U., McDonough, M. A., and Schofield, C. J. (2010) Crystal structure of the 2-oxoglutarate- and Fe(II)-dependent lysyl hydroxylase JMJD6. *J Mol Biol 401*, 211–222.
(210) Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr 54*, 905–921.
(211) Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr 66*, 213–221.
(212) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr 60*, 2126–2132.
(213) Lancaster, D. E., McNeill, L. A., McDonough, M. A., Aplin, R. T., Hewitson, K. S., Pugh, C. W., Ratcliffe, P. J., and Schofield, C. J. (2004) Disruption of dimerization and substrate phosphorylation inhibit factor inhibiting hypoxia-inducible factor (FIH) activity. *Biochem J 383*, 429–437.
(214) van Staalduinen, L. M., Novakowski, S. K., and Jia, Z. (2014) Structure and functional analysis of YcfD, a novel 2-oxoglutarate/Fe^{2+}-dependent oxygenase involved in translational regulation in Escherichia coli. *J Mol Biol 426*, 1898–1910.
(215) Wang, C., Zhang, Q., Hang, T., Tao, Y., Ma, X., Wu, M., Zhang, X., and Zang, J. (2015) Structure of the JmjC domain-containing protein NO66 complexed with ribosomal protein Rpl8. *Acta Crystallogr D Biol Crystallogr 71*, 1955–1964.
(216) Geoghegan, K. F., Dixon, H. B. F., Rosner, P. J., Hoth, L. R., Lanzetti, A. J., Borzilleri, K. A., Marr, E. S., Pezzullo, L. H., Martin, L. B., LeMotte, P. K., McColl, A. S., Kamath, A. V, and Stroh, J. G. (1999) Spontaneous α-N-6-phosphogluconoylation of a "His tag" in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins. *Anal Biochem 267*, 169–184.
(217) Kelly, S. M., Jess, T. J., and Price, N. C. (2005) How to study proteins by circular dichroism. *Biochim Biophys Acta 1751*, 119–139.
(218) Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc 1*, 2876–2890.
(219) Greenfield, N. J. (2009) Using circular dichroism collected as a funcion of temperature to

determine the thermodynamics of protein unfolding and binding interactions. *Nat Protoc 1*, 2527–2535.
(220) Lorenzen, K., and van Duijn, E. (2010) Native mass spectrometry as a tool in structural biology. *Curr Protoc Protein Sci 17.12*, 1–17.
(221) Sahin, E., and Roberts, C. J. (2012) Size-exclusion chromatography with multi-angle light scattering for elucidating protein aggregation mechanism. *Methods Mol Biol 899*, 403-423.
(222) Huang, Y., Fang, J., Bedford, M. T., Zhang, Y., and Xu, R. (2006) Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science 312*, 748–751.
(223) Lee, J., Thompson, J. R., Botuyan, M. V., and Mer, G. (2008) Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nat Struct Mol Biol 15*, 109–111.
(224) Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., and Bedford, M. T. (2006) Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep 7*, 397–403.
(225) Horton, J. R., Upadhyay, A. K., Qi, H. H., Zhang, X., Shi, Y., and Cheng, X. (2010) Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat Struct Mol Biol 17*, 38–43.
(226) Goda, S., Isagawa, T., Chikaoka, Y., Kawamura, T., and Aburatani, H. (2013) Control of histone H3 lysine 9 (H3K9) methylation state via cooperative two-step demethylation by jumonji domain containing 1A (JMJD1A) homodimer. *J Biol Chem 288*, 36948–36956.
(227) Shin, S., and Janknecht, R. (2007) Diversity within the JMJD2 histone demethylase family. *Biochem Biophys Res Commun 353*, 973–977.
(228) Hong, X., Zang, J., White, J., Wang, C., Pan, C.-H., Zhao, R., Murphy, R. C., Dai, S., Henson, P., Kappler, J. W., Hagman, J., and Zhang, G. (2010) Interaction of JMJD6 with single-stranded RNA. *Proc Natl Acad Sci U S A 107*, 14568–14572.
(229) Han, G., Li, J., Wang, Y., Li, X., Mao, H., Liu, Y., and Chen, C. D. (2012) The hydroxylation activity of Jmjd6 is required for its homo-oligomerization. *J Cell Biochem 113*, 1663–1670.
(230) Wolf, A., Mantri, M., Heim, A., Müller, U., Fichter, E., Mackeen, M. M., Schermelleh, L., Dadie, G., Leonhardt, H., Vénien-Bryan, C., Kessler, B. M., Schofield, C. J., and Böttger, A. (2013) The polyserine domain of the lysyl-5 hydroxylase Jmjd6 mediates subnuclear localization. *Biochem J 453*, 357–370.
(231) Hahn, P., Wegener, I., Burrells, A., Böse, J., Wolf, A., Erck, C., Butler, D., Schofield, C. J., Böttger, A., and Lengeling, A. (2010) Analysis of Jmjd6 cellular localization and testing for its involvement in histone demethylation. *PLoS One 5*, e13769.
(232) Ng, S. S., Kavanagh, K. L., McDonough, M. A., Butler, D., Pilka, E. S., Lienard, B. M. R., Bray, J. E., Savitsky, P., Gileadi, O., von Delft, F., Rose, N. R., Offer, J., Scheinost, J. C., Borowski, T., Sundstrom, M., Schofield, C. J., and Oppermann, U. (2007) Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature 448*, 87–91.
(233) Cheng, Z., Cheung, P., Kuo, A. J., Yukl, E. T., Wilmot, C. M., Gozani, O., and Patel, D. J. (2014) A molecular threading mechanism underlies Jumonji lysine demethylase KDM2A regulation of methylated H3K36. *Genes Dev 28*, 1758–1771.
(234) Horton, J. R., Engstrom, A., Zoeller, E. L., Liu, X., Shanks, J. R., Zhang, X., Johns, M. A., Vertino, P. M., Fu, H., and Cheng, X. (2016) Characterization of a linked Jumonji domain of the KDM5/JARID1 family of histone H3 lysine 4 demethylases. *J Biol Chem 291*, 2631–2646.
(235) Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., Kok, C. Y., Jia, M., De, T., Teague, J. W., Stratton, M. R., McDermott, U., and Campbell, P. J. (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res 43*, D805–D811.
(236) Laumonnier, F., Holbert, S., Ronce, N., Faravelli, F., Lenzner, S., Schwartz, C. E., Lespinasse, J., Van Esch, H., Lacombe, D., Goizet, C., Phan-Dinh Tuy, F., van Bokhoven, H., Fryns, J.-P., Chelly, J., Ropers, H.-H., Moraine, C., Hamel, B. C. J., and Briault, S. (2005) Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. *J*

Med Genet 42, 780–786.
(237) Koivisto, A. M., Ala-Mello, S., Lemmelä, S., Komu, H. A., Rautio, J., and Järvelä, I. (2007) Screening of mutations in the PHF8 gene and identification of a novel mutation in a Finnish family with XLMR and cleft lip/cleft palate. *Clin Genet 72*, 145–149.
(238) Qiu, J., Shi, G., Jia, Y., Li, J., Wu, M., Li, J., Dong, S., and Wong, J. (2010) The X-linked mental retardation gene PHF8 is a histone demethylase involved in neuronal differentiation. *Cell Res 20*, 908–918.
(239) Santos, C., Rodriguez-Revenga, L., Madrigal, I., Badenas, C., Pineda, M., and Milà, M. (2006) A novel mutation in JARID1C gene associated with mental retardation. *Eur J Hum Genet 14*, 583–586.
(240) Santos-Rebouças, C. B., Fintelman-Rodrigues, N., Jensen, L. R., Kuss, A. W., Ribeiro, M. G., Campos Jr., M., Santos, J. M., and Pimentel, M. M. G. (2011) A novel nonsense mutation in KDM5C/JARID1C gene causing intellectual disability, short stature and speech delay. *Neurosci Lett 498*, 67–71.
(241) Singh, V., Ram, M., Kumar, R., Prasad, R., Roy, B. K., and Singh, K. K. (2017) Phosphorylation: implications in cancer. *Protein J 36*, 1–6.
(242) Giglione, C., Boularot, A., and Meinnel, T. (2004) Protein N-terminal methionine excision. *Cell Mol Life Sci 61*, 1455–1474.
(243) Frottin, F., Martinez, A., Peynot, P., Mitra, S., Holz, R. C., Giglione, C., and Meinnel, T. (2006) The proteomics of N-terminal methionine cleavage. *Mol Cell Proteomics 5*, 2336–2349.
(244) Niesen, F. H., Berglund, H., and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc 2*, 2212–2221.
(245) Ali, M. H., and Imperiali, B. (2005) Protein oligomerization: how and why. *Bioorganic Med Chem 13*, 5013–5020.
(246) Morera, L., Lübbert, M., and Jung, M. (2016) Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. *Clin Epigenetics 8*, 57.
(247) Dever, T. E., and Green, R. (2012) The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb Perspect Biol 4*, a013706.
(248) Frolova, L., Le Goff, X., Rasmussen, H. H., Cheperegin, S., Drugeon, G., Kress, M., Arman, I., Haenni, A.-L., Celis, J. E., Phllippe, M., Justesen, J., and Kisselev, L. (1994) A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature 372*, 701–703.
(249) Drugeon, G., Jean-Jean, O., Frolova, L., Le Goff, X., Philippe, M., Kisselev, L., and Haenni, A.-L. (1997) Eukaryotic release factor 1 (eRF1) abolishes readthrough and competes with suppressor tRNAs at all three termination codons in messenger RNA. *Nucleic Acids Res 25*, 2254–2258.
(250) Scolnick, E., Tompkins, R., Caskey, T., and Nirenberg, M. (1968) Release factors differing in specificity for terminator codons. *Proc Natl Acad Sci U S A 61*, 768–774.
(251) Zhouravleva, G., Frolova, L., Le Goff, X., Le Guellec, R., Inge-Vechtomov, S., Kisselev, L., and Philippe, M. (1995) Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO J 14*, 4065–4072.
(252) Cheng, Z., Saito, K., Pisarev, A. V., Wada, M., Pisareva, V. P., Pestova, T. V., Gajda, M., Round, A., Kong, C., Lim, M., Nakamura, Y., Svergun, D. I., Ito, K., and Song, H. (2009) Structural insights into eRF3 and stop codon recognition by eRF1. *Genes Dev 23*, 1106–1118.
(253) Alkalaeva, E. Z., Pisarev, A. V., Frolova, L. Y., Kisselev, L. L., and Pestova, T. V. (2006) In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. *Cell 125*, 1125–1136.
(254) Taylor, D., Unbehaun, A., Li, W., Das, S., Lei, J., Liao, H. Y., Grassucci, R. A., Pestova, T. V, and Frank, J. (2012) Cryo-EM structure of the mammalian eukaryotic release factor eRF1-eRF3-associated termination complex. *Proc Natl Acad Sci 109*, 18413–18418.
(255) des Georges, A., Hashem, Y., Unbehaun, A., Grassucci, R. A., Taylor, D., Hellen, C. U. T., Pestova, T. V, and Frank, J. (2014) Structure of the mammalian ribosomal pre-termination complex associated with eRF1*eRF3*GDPNP. *Nucleic Acids Res 42*, 3409–3418.
(256) Shoemaker, C. J., and Green, R. (2011) Kinetic analysis reveals the ordered coupling of

translation termination and ribosome recycling in yeast. *Proc Natl Acad Sci U S A 108*, E1392–E1398.
(257) Frolova, L., Le Goff, X., Zhouravleva, G., Davydova, E., Philippe, M., and Kisselev, L. (1996) Eukaryotic polypeptide chain release factor eRF3 is an eRF1- and ribosome-dependent guanosine triphosphatase. *RNA 2*, 334–341.
(258) Seit-Nebi, A., Frolova, L., and Kisselev, L. (2002) Conversion of omnipotent translation termination factor eRF1 into ciliate-like UGA-only unipotent eRF1. *EMBO Rep 3*, 881–886.
(259) Frolova, L. Y., Tsivkovskii, R. Y., Sivolobova, G. F., Oparina, N. Y., Serpinsky, O. I., Blinov, V. M., Tatkov, S. I., and Kisselev, L. L. (1999) Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA 5*, 1014–1020.
(260) Seit-Nebi, A., Frolova, L., Justesen, J., and Kisselev, L. (2001) Class-1 translation termination factors: invariant GGQ minidomain is essential for release activity and ribosome binding but not for stop codon recognition. *Nucleic Acids Res 29*, 3982–3987.
(261) Merkulova, T. I., Frolova, L. Y., Lazar, M., Camonis, J., and Kisselev, L. L. (1999) C-terminal domains of human translation termination factors eRF1 and eRF3 mediate their in vivo interaction. *FEBS Lett 443*, 41–47.
(262) Kolosov, P., Frolova, L., Seit-Nebi, A., Dubovaya, V., Kononenko, A., Oparina, N., Justesen, J., Efimov, A., and Kisselev, L. (2005) Invariant amino acids essential for decoding function of polypeptide release factor eRF1. *Nucleic Acids Res 33*, 6418–6425.
(263) Wong, L. E., Li, Y., Pillay, S., Frolova, L., and Pervushin, K. (2012) Selectivity of stop codon recognition in translation termination is modulated by multiple conformations of GTS loop in eRF1. *Nucleic Acids Res 40*, 5751–5765.
(264) Bulygin, K. N., Khairulina, Y. S., Kolosov, P. M., Ven'yaminova, A. G., Graifer, D. M., Vorobjev, Y. N., Frolova, L. Y., Kisselev, L. L., and Karpova, G. G. (2010) Three distinct peptides from the N domain of translation termination factor eRF1 surround stop codon in the ribosome. *RNA 16*, 1902–1914.
(265) Pillay, S., Li, Y., Wong, L. E., and Pervushin, K. (2016) Structural characterization of eRF1 mutants indicate a complex mechanism of stop codon recognition. *Sci Rep 6*, 18644.
(266) Blanchet, S., Rowe, M., Von der Haar, T., Fabret, C., Demais, S., Howard, M. J., and Namy, O. (2015) New insights into stop codon recognition by eRF1. *Nucleic Acids Res 43*, 3298–3308.
(267) Preis, A., Heuer, A., Barrio-Garcia, C., Hauser, A., Eyler, D. E., Berninghausen, O., Green, R., Becker, T., and Beckmann, R. (2014) Cryoelectron microscopic structures of eukaryotic translation termination complexes containing eRF1-eRF3 or eRF1-ABCE1. *Cell Rep 8*, 59–65.
(268) Keeling, K. M., Xue, X., Gunn, G., and Bedwell, D. M. (2014) Therapeutics based on stop codon readthrough. *Annu Rev Genomics Hum Genet 15*, 371–394.
(269) Mort, M., Ivanov, D., Cooper, D. N., and Chuzhanova, N. A. (2008) A meta-analysis of nonsense mutations causing human genetic disease. *Hum Mutat 29*, 1037–1047.
(270) Polshakov, V. I., Eliseev, B. D., Birdsall, B., and Frolova, L. Y. (2012) Structure and dynamics in solution of the stop codon decoding N-terminal domain of the human polypeptide chain release factor eRF1. *Protein Sci 21*, 896–903.
(271) Polshakov, V. I., Eliseev, B. D., Frolova, L. Y., Chang, C.-F., and Huang, T.-H. (2015) Backbone (1)H, (13)C and (15)N resonance assignments of the human eukaryotic release factor eRF1. *Biomol NMR Assign 9*, 37–42.
(272) Higman, V. A. (2016) Protein NMR: a practical guide. (http://www.protein-nmr.org.uk/).
(273) Ohki, S., and Kainosho, M. (2008) Stable isotope labeling methods for protein NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc 53*, 208–226.
(274) Tong, H., Bell, D., Tabei, K., and Siegel, M. M. (1999) Automated data massaging, interpretation, and e-mailing modules for high throughput open access mass spectrometry. *J Am Soc Mass Spectrom 10*, 1174–1187.
(275) Strupat, K., Scheibner, O., and Bromirski, M. (2013) High-resolution, accurate-mass Orbitrap mass spectrometry – definitions, opportunities , and advantages. *Thermo Fischer*.
(276) Steen, H., and Mann, M. (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat Rev*

Mol Cell Biol 5, 699–711.
(277) Edrissi, B., Taghizadeh, K., and Dedon, P. C. (2013) Quantitative analysis of histone modifications: formaldehyde is a source of pathological N6-formyllysine that is refractory to histone deacetylases. *PLoS Genet 9*, e1003328.
(278) Jones, O. (2016) Part II thesis: Evidence for a novel oxygenase-catalysed post-translational modification. University of Oxford, FHS Molecular and Cellular Biochemistry.
(279) Huynh, K., and Partch, C. L. (2015) Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protoc Protein Sci 79*, 28.9.1-28.9.14.
(280) Whitehead, B., Craven, C. J., and Waltho, J. P. (1997) Double and triple resonance NMR methods for protein assignment. *Methods Mol Biol 60*, 29–52.
(281) Yao, J., Dyson, H. J., and Wright, P. E. (1997) Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Lett 419*, 285–289.
(282) Wishart, D. S., and Sykes, B. D. (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR 4*, 171–180.
(283) Ishima, R., and Torchia, D. A. (2000) Protein dynamics from NMR. *Nat Struct Biol 7*, 740–743.
(284) Columbia University Department of Chemistry NMR Facility. (2017) HMBC. (http://www.columbia.edu/cu/chemistry/groups/nmr/HMBC.html).
(285) Matheisl, S., Berninghausen, O., Becker, T., and Beckmann, R. (2015) Structure of a human translation termination complex. *Nucleic Acids Res 43*, 8615–8626.
(286) Brown, A., Shao, S., Murray, J., Hegde, R. S., and Ramakrishnan, V. (2015) Structural basis for stop codon recognition in eukaryotes. *Nature 524*, 493–496.
(287) Arnstein, H. R., and Clubb, M. E. (1958) The biosynthesis of penicillin. 8. Investigation of cyclic cysteinylvaline peptides as precursors. *Biochem J 68*, 528–535.
(288) Rutledge, P. J. (2015) Isopenicillin N Synthase, in *2-Oxoglutarate-dependent oxygenases* (Hausinger, R. P., and Schofield, C. J., Eds.) RSC Metall., pp 414–424. Royal Society of Chemistry, Cambridge.
(289) Bush, K., and Bradford, P. A. (2016) β-Lactams and β-lactamase inhibitors: an overview. *Cold Spring Harb Perspect Med 6*, a025247.
(290) Janc, J. W., Egan, L. A., and Townsend, C. A. (1995) Purification and characterization of clavaminate synthase from Streptomyces antibioticus. A multifunctional enzyme of clavam biosynthesis. *J Biol Chem 270*, 5399–5404.
(291) Yang, M., Hardy, A. P., Chowdhury, R., Loik, N. D., Scotti, J. S., McCullagh, J. S. O., Claridge, T. D. W., McDonough, M. A., Ge, W., and Schofield, C. J. (2013) Substrate selectivity analyses of factor inhibiting hypoxia-inducible factor. *Angew Chemie Int Ed 52*, 1700–1704.
(292) World Health Organisation (2016) Antibiotic Resistance. (http://www.who.int/mediacentre/factsheets/antibiotic-resistance/en/)
(293) Baldwin, J. E., Lowe, C., and Schofield, C. J. (1986) A γ-Lactam Analogue of Penems Possessing Antibacterial Activity. *Tetrahedron Lett 27*, 3461–3464.
(294) Indelicato, J. M., and Pasini, C. E. (1988) The acylating potential of γ-lactam antibacterials: base hydrolysis of bicyclic pyrazolidinones. *J Med Chem 31*, 1227–1230.
(295) Baldwin, J. E., Lynch, G. P., and Schofield, C. J. (1992) Isopenicillin N synthase: a new mode of reactivity. *Tetrahedron 48*, 9085–9100.
(296) Baldwin, J. E., Adlington, R. M., Bryans, J. S., Lloyd, M. D., and Sewell, T. J. (1997) Chemo-enzymatic synthesis of bicyclic γ-lactams using clavaminic acid synthase. *Tetrahedron 53*, 7011–7020.
(297) Pundir, S., Magrane, M., Martin, M. J., O'Donovan, C., and The UniProt Consortium (2015) Searching and navigating the UniProt databases. *Curr Protoc Bioinformatics 50*, 1.27.1-1.27.10.
(298) Gault, J., Donlan, J. A. C., Liko, I., Hopper, J. T. S., Gupta, K., Housden, N. G., Struwe, W. B., Marty, M. T., Mize, T., Bechara, C., Zhu, Y., Wu, B., Kleanthous, C., Belov, M., Damoc, E., Makarov, A., and Robinson, C. V. (2016) High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nat Methods 13*, 333–336.

(299) Parsons, T. B., Struwe, W. B., Gault, J., Yamamoto, K., Taylor, T. A., Raj, R., Wals, K., Mohammed, S., Robinson, C. V., Benesch, J. L. P., and Davis, B. G. (2016) Optimal synthetic glycosylation of a therapeutic antibody. *Angew Chemie Int Ed 55*, 2361–2367.
(300) Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A 103*, 8060–8065.
(301) Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., and Bricogne, G. (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr D Biol Crystallogr 68*, 368–380.
(302) Chen, V. B., Arendall 3rd, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr 66*, 12–21.
(303) Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins Struct Funct Bioinforma 59*, 687–696.
(304) Manta, B., and Gladyshev, V. N. (2017) Regulated methionine oxidation by monooxygenases. *Free Radic Biol Med*, http://dx.doi.org/10.1016/j.freeradbiomed.2017.02.010
(305) Kaya, A., Lee, B. C., and Gladyshev, V. N. (2015) Regulation of protein function by reversible methionine oxidation and the role of selenoprotein MsrB1. *Antioxid Redox Signal 23*, 814–822.
(306) Lee, B. C., Péterfi, Z., Hoffmann, F. W., Moore, R. E., Kaya, A., Avanesov, A., Tarrago, L., Zhou, Y., Weerapana, E., Fomenko, D. E., Hoffmann, P. R., and Gladyshev, V. N. (2013) MsrB1 and MICALs regulate actin assembly and macrophage function via reversible stereoselective methionine oxidation. *Mol Cell 51*, 397–404.
(307) Achilli, C., Ciana, A., and Minetti, G. (2015) The discovery of methionine sulfoxide reductase enzymes: a historical account and future perspectives. *BioFactors 41*, 135–152.
(308) Jagannadham, M. V. (2009) Identifying the sequence and distinguishing the oxidized — Methionine from phenylalanine peptides by MALDI TOF/TOF mass spectrometry in an antarctic Bacterium Pseudomonas syringae. *Proteomics Insights 2*, 27–31.
(309) Erlandsson Harris, H., and Andersson, U. (2004) The nuclear protein HMGB1 as a proinflammatory mediator. *Eur J Immunol 34*, 1503–1512.
(310) Gu, S. X., Stevens, J. W., and Lentz, S. R. (2015) Regulation of thrombosis and vascular function by protein methionine oxidation. *Blood 125*, 3851–3859.
(311) Tossounian, M.-A., Pedre, B., Wahni, K., Erdogan, H., Vertommen, D., Van Molle, I., and Messens, J. (2015) Corynebacterium diphtheriae Methionine Sulfoxide Reductase A exploits a unique mycothiol redox relay mechanism. *J Biol Chem 290*, 11365–11375.

A Appendix: Supplementary Information

A1. JMJD7 co-immunoprecipitation results

*Table A.1. JMJD7-interacting proteins enriched in the DMOG-treated versus non-DMOG-treated HEK293T cells.**

Protein	Gene Symbol	SwissProt ID	Length (AA)	% Coverage	Ratio: JMJD7+DMOG/JMJD7
IPI00031836	DRG1 Developmentally-regulated GTP-binding protein 1	Q9Y295	367	44.7	JMJD7+DMOG Only
IPI00219097	HMGB2 High mobility group protein B2	P26583	209	20.9	JMJD7+DMOG Only
IPI00009865	KRT10 Keratin, type I cytoskeletal 10	P13645	584	9.6	JMJD7+DMOG Only
IPI00106491	MRTO4 mRNA turnover protein 4 homolog	Q9UKD2	239	9.2	JMJD7+DMOG Only
IPI00219018	GAPDH Glyceraldehyde-3-phosphate dehydrogenase	P04406	335	10	JMJD7+DMOG Only
IPI00013917	RPS12 40S ribosomal protein S12	P25398	132	25	JMJD7 Only
IPI00004839	CRKL Crk-like protein	P46109	303	11.2	JMJD7 Only
IPI00220327	KRT1 Keratin, type II cytoskeletal 1	P04264	644	11.5	39.34
IPI00019359	KRT9 Keratin, type I cytoskeletal 9	P35527	623	2.7	14.49
IPI00215637	DDX3X ATP-dependent RNA helicase DDX3X	O00571	662	11	8.64
IPI00465044	RCC2 Protein RCC2	Q9P258	522	18.6	6.98
IPI00006203	ZBED1 Zinc finger BED domain-containing protein 1	O96006	694	4	6.79
IPI00054042	GTF2I Isoform 1 of General transcription factor II-I	P78347-1	998	5.9	3.93
IPI00024320	RBM3 Putative RNA-binding protein 3	P98179	157	21	3.16
IPI00022334	OAT Ornithine aminotransferase, mitochondrial	P04181	439	10.9	2.97
IPI00646304	PPIB Peptidyl-prolyl cis-trans isomerase B	P23284	216	46.3	2.9
IPI00329791	DDX46 Probable ATP-dependent RNA helicase DDX46	Q7L014	1031	7.5	2.27
IPI00099810	ZCCHC17 Isoform 1 of Nucleolar protein of 40 kDa	Q9NP64-1	241	12.9	2.22
IPI00554737	PPP2R1A Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	P30153	589	5.8	2.14
IPI00339381	HLTF Isoform 1 of Helicase-like transcription factor	Q14527-1	1009	13.8	1.94
IPI00071318	LUC7L Isoform 1 of Putative RNA-binding protein Luc7-like 1	Q9NQ29-1	371	24.3	1.85
IPI00443657	CIRBP Uncharacterized protein	N/A	135	9.6	1.82
IPI00291467	SLC25A6 ADP/ATP translocase 3	P12236	298	23.4	1.73
IPI00180730	Similar to Elongation factor 1-alpha 1	N/A	462	20.2	1.7
IPI00006932	LUC7L2 cDNA FLJ55988, highly similar to RNA-binding protein Luc7-like 2	N/A	458	28.3	1.63
IPI00004237	PHKA2 Phosphorylase b kinase regulatory subunit alpha, liver isoform	P46019	1235	5.1	1.53
IPI00007188	SLC25A5 ADP/ATP translocase 2	P05141	298	17.8	1.52
IPI00163505	RBM39 Isoform 1 of RNA-binding protein 39	Q14498-1	530	12.4	1.51
IPI00025277	PDCD6 Programmed cell death protein 6	O75340	191	12.7	1.49
IPI00221354	FUS Isoform Short of RNA-binding	P35637-2	525	21.9	1.48

Protein	Gene Symbol	SwissProt ID	Length (AA)	% Coverage	Ratio: JMJD7+DMOG/JMJD7
	protein FUS				
IPI00218829	GSPT1 Eukaryotic peptide chain release factor GTP-binding subunit ERF3A	P15170	499	9.2	1.34
IPI00000874	PRDX1 Peroxiredoxin-1	Q06830	199	59.3	1.34
IPI00293746	C1orf35 Isoform 1 of Multiple myeloma tumor-associated protein 2	Q9BU76-1	263	12.7	1.29
IPI00033561	RNPS1 Isoform 1 of RNA-binding protein with serine-rich domain 1	Q15287-1	305	16	1.26
IPI00419258	HMGB1 High mobility group protein B1	P09429	215	27.2	1.26
IPI00017289	SNRNP27 U4/U6.U5 small nuclear ribonucleoprotein 27 kDa protein	Q8WVK2	155	36.8	1.18
IPI00221222	SUB1 Activated RNA polymerase II transcriptional coactivator p15	P53999	127	48.8	1.16
IPI00018971	TRIM21 Isoform 1 of E3 ubiquitin-protein ligase TRIM21	P19474-1	475	30.7	1.15
IPI00414676	HSP90AB1 Heat shock protein HSP 90-beta	P08238	724	4.4	1.13
IPI00007752	TUBB2C Tubulin beta-2C chain	P68371	445	32.1	1.1
IPI00003865	HSPA8 Isoform 1 of Heat shock cognate 71 kDa protein	P11142-1	646	17.9	1.09
IPI00004247	PHKA1 Isoform 1 of Phosphorylase b kinase regulatory subunit alpha, skeletal muscle isoform	P46020-1	1223	6.8	1.07
IPI00218343	TUBA1C Tubulin alpha-1C chain	Q9BQE3	449	41.9	1.04
IPI00302850	SNRPD1 Small nuclear ribonucleoprotein Sm D1	P62314	119	26.9	1.03

*Experiment was performed by Dr Ming Yang (University of Oxford).

A2. ^{13}Cα chemical shift index (CSI)

The ^{13}Cα CSI is an indirect, non-experimental technique that predicts protein secondary structure based on ^{13}C chemical shifts, given that α-carbons tend to experience a downfield chemical shift in an α-helix and an upfield chemical shift in a β-strand[282]. The CSI approach uses derived 'random coil' reference values and thresholds specific to the ^{13}Cα nuclei of each of the 20 natural amino acids to predict protein secondary structure by applying a two-stage method: (1) Assignment of a CSI [**-1** (Cα < ref), **0** (Cα ~ ref), or **1** (Cα > ref), where the difference must be > 0.7 ppm] on the basis of observed chemical shifts, and (2) secondary structure assignment on the basis of the local densities [e.g. nonzero CSI for over 4-5 residues must be > 70% for defining α-helices and β-strands, and the first appearance of a CSI opposite in magnitude (or two consecutive zero values) marks a termination point][282]. **Table A.2** and **Table A.3** outline the ^{13}Cα CSI values assigned to each amino acid residue and the respective predicted secondary structure.

Table A.2. $^{13}C\alpha$ chemical shift index of wildtype $eRF1_{ND}$ and $eRF1_{ND}$-30.

Residue		$^{13}C\alpha$ Reference (ppm)[282]	$^{13}C\alpha$ $eRF1_{ND}$ (ppm)	$\Delta^{13}C\alpha$ ($eRF1_{ND}$ - reference)	$^{13}C\alpha$ $eRF1_{ND}$-30 (ppm)	$\Delta C\alpha$ ($eRF1_{ND}$-30 - reference)	$eRF1_{ND}$ CSI	$eRF1_{ND}$-30 CSI
1	MET	56.6	-	-	-	-	-	-
2	ALA	52.5	52.88	0.38	52.64	0.14	0	0
3	ASP	51.4	54.34	2.94	54.13	2.73	1	1
4	ASP	51.4	52.22	0.82	51.95	0.55	1	0
5	PRO	62.9	64.12	1.22	63.97	1.07	0	0
6	SER	58.3	59.17	0.87	59.06	0.76	1	1
7	ALA	52.5	53.62	1.12	-	-	1	-
8	ALA	52.5	53.83	1.33	53.48	0.98	1	1
9	ASP	51.4	55.33	3.93	55.17	3.77	1	1
10	ARG	56.3	58.02	1.72	57.75	1.45	1	1
11	ASN	53.6	55.52	1.92	55.27	1.67	1	1
12	VAL	63.0	65.86	2.86	65.58	2.58	1	1
13	GLU	56.7	58.84	2.14	58.60	1.90	1	1
14	ILE	62.6	65.41	2.81	-	-	1	-
15	TRP	57.8	60.84	3.04	60.67	2.87	1	1
16	LYS	56.7	60.29	3.59	60.21	3.51	1	1
17	ILE	62.6	63.77	1.17	63.65	1.05	1	1
18	LYS	56.7	61.21	4.51	60.99	4.29	1	1
19	LYS	56.7	59.42	2.72	-	-	1	-
20	LEU	55.7	58.16	2.46	57.83	2.13	1	1
21	ILE	62.6	66.68	4.08	66.55	3.95	1	1
22	LYS	56.7	59.16	2.46	59.00	2.30	1	1
23	SER	58.3	61.54	3.24	61.27	2.97	1	1
24	LEU	55.7	57.27	1.57	57.09	1.39	1	1
25	GLU	56.7	59.08	2.38	58.84	2.14	1	1
26	ALA	52.5	52.45	-0.05	52.26	-0.24	0	0
27	ALA	52.5	52.90	0.40	52.65	0.15	1	0
28	ARG	56.3	54.44	-1.86	54.28	-2.02	-1	-1
29	GLY	45.0	44.90	-0.10	44.65	-0.35	0	0
30	ASN	53.6	52.22	-1.38	51.88	-1.72	-1	-1
31	GLY	45.0	45.55	0.55	45.35	0.35	0	0
32	THR	63.1	59.94	-3.16	60.43	-2.67	-1	-1
33	SER	58.3	58.01	-0.29	57.80	-0.50	0	0
34	MET	56.6	55.20	-1.40	55.14	-1.46	-1	-1
35	ILE	62.6	60.90	-1.70	60.71	-1.89	-1	-1
36	SER	58.3	57.50	-0.80	57.46	-0.84	-1	-1
37	LEU	55.7	53.66	-2.04	53.31	-2.39	-1	-1
38	ILE	62.6	60.96	-1.64	60.80	-1.80	-1	-1
39	ILE	62.6	57.78	-4.82	57.63	-4.97	-1	-1
40	PRO	62.9	61.99	-0.91	-	-	0	-
41	PRO	62.9	64.43	1.53	64.28	1.38	1	0
42	LYS	56.7	57.23	0.53	57.10	0.40	0	0
43	ASP	51.4	53.12	1.72	52.84	1.44	1	1
44	GLN	56.2	54.32	-1.88	54.25	-1.95	-1	-1
45	ILE	62.6	63.64	1.04	63.60	1.00	1	1
46	SER	58.3	61.13	2.83	60.64	2.34	1	1
47	ARG	56.3	59.45	3.15	59.15	2.85	1	1
48	VAL	63.0	66.41	3.41	66.26	3.26	1	1
49	ALA	52.5	55.67	3.17	55.52	3.02	1	1
50	LYS	56.7	59.29	2.59	-	-	1	-
51	MET	56.6	59.35	2.75	59.12	2.52	1	1
52	LEU	55.7	57.73	2.03	57.52	1.82	1	1
53	ALA	52.5	55.22	2.72	55.01	2.51	1	1
54	ASP	51.4	57.62	6.22	57.30	5.90	1	1
55	GLU	56.7	59.01	2.31	58.86	2.16	1	1
56	PHE	57.9	62.00	4.10	61.63	3.73	1	1
57	GLY	45.0	47.45	2.45	47.06	2.06	1	1
58	THR	63.1	65.83	2.73	65.37	2.27	1	1
59	ALA	52.5	54.30	1.80	54.18	1.68	1	1
60	SER	58.3	61.67	3.37	60.83	2.53	1	1
61	ASN	53.6	53.42	-0.18	53.25	-0.35	0	0
62	ILE	62.6	63.68	1.08	62.22	-0.38	1	0
63	LYS	56.7	59.07	2.37	-	-	1	-
64	SER	58.3	57.13	-1.17	-	-	-1	-
65	ARG	56.3	60.16	3.86	56.62	0.32	1	0
66	VAL	63.0	65.76	2.76	63.08	0.08	1	0
67	ASN	53.6	55.77	2.17	52.52	-1.08	1	-1
68	ARG	56.3	60.19	3.89	58.36	2.06	1	1
69	LEU	55.7	58.34	2.64	58.01	2.31	1	1
70	SER	58.3	62.03	3.73	61.09	2.79	1	1
71	VAL	63.0	67.16	4.16	65.86	2.86	1	1

Residue		^{13}Cα Reference (ppm)[282]	^{13}Cα eRF1$_{ND}$ (ppm)	Δ^{13}Cα (eRF1$_{ND}$ - reference)	^{13}Cα eRF1$_{ND}$-30 (ppm)	ΔCα (eRF1$_{ND}$-30 - reference)	eRF1$_{ND}$ CSI	eRF1$_{ND}$-30 CSI
72	LEU	55.7	58.65	2.95	58.28	2.58	1	1
73	GLY	45.0	47.34	2.34	47.09	2.09	1	1
74	ALA	52.5	55.87	3.37	55.60	3.10	1	1
75	ILE	62.6	66.87	4.27	-	-	1	-
76	THR	63.1	67.03	3.93	-	-	1	-
77	SER	58.3	62.92	4.62	62.78	4.48	1	1
78	VAL	63.0	67.15	4.15	67.05	4.05	1	1
79	GLN	56.2	59.92	3.72	59.65	3.45	1	1
80	GLN	56.2	58.90	2.70	58.74	2.54	1	1
81	ARG	56.3	57.03	0.73	56.81	0.51	1	0
82	LEU	55.7	57.77	2.07	57.41	1.71	1	1
83	LYS	56.7	58.12	1.42	57.79	1.09	1	1
84	LEU	55.7	55.72	0.02	55.58	-0.12	0	-1
85	TYR	58.6	58.09	-0.51	57.90	-0.70	0	0
86	ASN	53.6	54.12	0.52	54.00	0.40	0	0
87	LYS	56.7	54.59	-2.11	54.36	-2.34	-1	-1
88	VAL	63.0	62.39	-0.61	62.20	-0.80	0	-1
89	PRO	62.9	62.66	-0.24	-	-	0	-
90	PRO	62.9	66.38	3.48	66.11	3.21	0	0
91	ASN	53.6	52.35	-1.25	52.20	-1.40	-1	-1
92	GLY	45.0	43.40	-1.61	43.15	-1.85	-1	-1
93	LEU	55.7	54.17	-1.53	54.07	-1.63	-1	-1
94	VAL	63.0	60.87	-2.13	60.53	-2.47	-1	-1
95	VAL	63.0	60.69	-2.31	-	-	-1	-
96	TYR	58.6	56.67	-1.93	56.37	-2.23	-1	-1
97	CYS	58.3	56.80	-1.50	56.46	-1.84	-1	-1
98	GLY	45.0	46.37	1.37	46.17	1.17	1	1
99	THR	63.1	62.51	-0.59	62.44	-0.66	0	0
100	ILE	62.6	59.25	-3.35	59.09	-3.51	-1	-1
101	VAL	63.0	60.99	-2.01	60.76	-2.24	-1	-1
102	THR	63.1	61.02	-2.08	60.71	-2.39	-1	-1
103	GLU	56.7	59.34	2.64	59.12	2.42	1	1
104	GLU	56.7	56.64	-0.06	56.38	-0.32	0	0
105	GLY	45.0	45.52	0.52	45.36	0.36	0	0
106	LYS	56.7	54.93	-1.77	54.63	-2.07	-1	-1
107	GLU	56.7	55.91	-0.79	55.71	-0.99	-1	-1
108	LYS	56.7	55.42	-1.28	55.34	-1.36	-1	-1
109	LYS	56.7	56.24	-0.46	56.13	-0.57	0	0
110	VAL	63.0	61.76	-1.24	61.53	-1.47	-1	-1
111	ASN	53.6	53.62	0.02	53.42	-0.18	0	0
112	ILE	62.6	60.52	-2.08	60.33	-2.27	-1	-1
113	ASP	51.4	52.60	1.20	52.41	1.01	1	1
114	PHE	57.9	56.95	-0.95	56.78	-1.12	-1	-1
115	GLU	56.7	52.00	-4.70	51.86	-4.84	-1	-1
116	PRO	62.9	62.86	-0.04	62.56	-0.34	0	0
117	PHE	57.9	58.93	1.03	58.96	1.06	1	1
118	LYS	56.7	53.01	-3.69	52.81	-3.89	-1	-1
119	PRO	62.9	63.27	0.37	62.92	0.02	0	0
120	ILE	62.6	60.20	-2.41	60.11	-2.49	-1	-1
121	ASN	53.6	53.35	-0.25	53.13	-0.47	0	0
122	THR	63.1	61.13	-1.97	60.89	-2.21	-1	-1
123	SER	58.3	56.72	-1.58	56.59	-1.71	-1	-1
124	LEU	55.7	55.34	-0.36	55.06	-0.64	0	0
125	TYR	58.6	58.68	0.08	58.20	-0.40	0	0
126	LEU	55.7	54.26	-1.44	54.09	-1.61	-1	-1
127	CYS	58.3	57.45	-0.85	57.08	-1.22	-1	-1
128	ASP	51.4	52.92	1.52	-	-	1	-
129	ASN	53.6	53.17	-0.43	52.83	-0.77	0	-1
130	LYS	56.7	54.53	-2.17	54.32	-2.38	-1	-1
131	PHE	57.9	57.12	-0.78	56.86	-1.04	-1	-1
132	HIS	55.8	57.00	1.20	-	-	1	-
133	THR	63.1	62.06	-1.05	-	-	-1	-
134	GLU	56.7	61.08	4.38	60.80	4.10	1	1
135	ALA	52.5	54.81	2.31	54.51	2.01	1	1
136	LEU	55.7	56.76	1.06	56.47	0.77	1	-1
137	THR	63.1	67.02	3.92	66.88	3.78	1	1
138	ALA	52.5	55.20	2.70	54.95	2.45	1	1
139	LEU	55.7	57.65	1.95	57.37	1.67	1	1
140	LEU	55.7	56.83	1.13	56.77	1.07	1	1
141	SER	58.3	60.75	2.45	60.52	2.22	1	1
142	ASP	51.4	56.03	4.63	55.84	4.44	1	1

* **1** = ΔCα > 0.7; **-1** = ΔCα < -0.7; **0** = ΔCα = -0.7 – 0.7

Table A.3. $eRF1_{ND}$-30 secondary structure prediction based on the $^{13}C\alpha$ chemical shift index, Table A.2 (Note: A = α-helix, B = β-sheet, C = random coil).

Residue		Secondary Structure Prediction		
		$eRF1_{ND}$	$eRF1_{ND}$-30	$eRF1_{ND}$ (PDB ID: 2LLX[270])
1	MET	-	-	C
2	ALA	C	C	C
3	ASP	A	C	C
4	ASP	A	C	C
5	PRO	A	C	C
6	SER	A	A	C
7	ALA	A	A	C
8	ALA	A	A	A
9	ASP	A	A	A
10	ARG	A	A	A
11	ASN	A	A	A
12	VAL	A	A	A
13	GLU	A	A	A
14	ILE	A	A	A
15	TRP	A	A	A
16	LYS	A	A	A
17	ILE	A	A	A
18	LYS	A	A	A
19	LYS	A	A	A
20	LEU	A	A	A
21	ILE	A	A	A
22	LYS	A	A	A
23	SER	A	A	A
24	LEU	A	A	A
25	GLU	A	A	A
26	ALA	A	C	A
27	ALA	A	C	C
28	ARG	A	C	C
29	GLY	C	C	C
30	ASN	C	C	C
31	GLY	C	C	C
32	THR	C	C	C
33	SER	C	C	C
34	MET	B	B	B
35	ILE	B	B	B
36	SER	B	B	B
37	LEU	B	B	B
38	ILE	B	B	B
39	ILE	B	B	C
40	PRO	C	-	C
41	PRO	C	C	C
42	LYS	C	C	C
43	ASP	C	C	C
44	GLN	C	A	A
45	ILE	A	A	A
46	SER	A	A	A
47	ARG	A	A	A
48	VAL	A	A	A
49	ALA	A	A	A
50	LYS	A	A	A
51	MET	A	A	A
52	LEU	A	A	A
53	ALA	A	A	A
54	ASP	A	A	A
55	GLU	A	A	A
56	PHE	A	A	A
57	GLY	A	A	A
58	THR	A	A	A
59	ALA	A	A	A
60	SER	A	A	C
61	ASN	A	C	C
62	ILE	A	C	C
63	LYS	A	-	C
64	SER	A	-	A
65	ARG	A	C	A
66	VAL	A	C	A
67	ASN	A	A	A
68	ARG	A	A	A
69	LEU	A	A	A
70	SER	A	A	A

Residue		Secondary Structure Prediction		
		$eRF1_{ND}$	$eRF1_{ND}$-30	$eRF1_{ND}$ (PDB ID: $2LLX^{270}$)
71	VAL	A	A	A
72	LEU	A	A	A
73	GLY	A	A	A
74	ALA	A	A	A
75	ILE	A	A	A
76	THR	A	A	A
77	SER	A	A	A
78	VAL	A	A	A
79	GLN	A	A	A
80	GLN	A	A	A
81	ARG	A	A	A
82	LEU	A	A	A
83	LYS	A	A	A
84	LEU	C	C	C
85	TYR	C	C	C
86	ASN	C	C	C
87	LYS	C	C	C
88	VAL	C	C	C
89	PRO	C	-	C
90	PRO	C	C	C
91	ASN	B	B	C
92	GLY	B	B	C
93	LEU	B	B	B
94	VAL	B	B	B
95	VAL	B	-	B
96	TYR	B	B	B
97	CYS	B	B	B
98	GLY	C	C	B
99	THR	C	C	B
100	ILE	B	B	B
101	VAL	B	B	C
102	THR	B	B	C
103	GLU	C	C	C
104	GLU	C	C	C
105	GLY	C	C	C
106	LYS	B	B	B
107	GLU	B	B	B
108	LYS	B	B	B
109	LYS	C	B	B
110	VAL	C	B	B
111	ASN	C	B	B
112	ILE	C	B	B
113	ASP	C	C	B
114	PHE	C	C	C
115	GLU	C	C	C
116	PRO	C	C	C
117	PHE	C	C	C
118	LYS	C	C	C
119	PRO	C	C	C
120	ILE	C	C	C
121	ASN	C	C	C
122	THR	C	C	C
123	SER	C	C	B
124	LEU	C	C	B
125	TYR	C	C	B
126	LEU	C	B	B
127	CYS	C	B	C
128	ASP	C	B	C
129	ASN	C	B	C
130	LYS	C	B	C
131	PHE	C	B	C
132	HIS	C	-	C
133	THR	C	-	A
134	GLU	A	A	A
135	ALA	A	A	A
136	LEU	A	A	A
137	THR	A	A	A
138	ALA	A	A	A
139	LEU	A	A	A
140	LEU	A	A	A
141	SER	A	A	A
142	ASP	A	A	A

B APPENDIX: INVESTIGATION OF JMJD7 AS A METHIONINE OXIDASE

B.1. Introduction: Physiological relevance of methionine oxidation

** The work described in **Appendix B** formed part of a successful application made for an EMBO Short-Term Fellowship, which funded a two week visit to the laboratory of Professor Joris Messens (Vrije Universiteit Brussel (VIB), Belgium). Where specified, experiments were done in collaboration with Maria Tossounian (VIB). All MsrA/B enzymes used in this chapter were kindly prepared and provided by Maria.*

The reversible, post-translational oxidation of methionine to methionine sulfoxide has long been considered as a regulatory mechanism in cellular redox control and oxidative stress signalling[304]. In cells, sulphur-containing amino acids such as methionine are susceptible to oxidation by small molecules like hydrogen peroxide, peroxynitrite, and hypochlorous acid, often leading to a loss of biological activity[305]. In addition to these small molecules, there are a group of flavin-dependent monooxygenases, the Mical proteins, that enzymatically oxidise methionine to methionine sulfoxide[306]. These enzymes are distinguished by their ability to stereospecifically oxidise methionine, as distinct from the non-enzymatic oxidising agents that yield a mixture of methionine sulfoxide diastereomers. Methionine oxidation can be reversed by the methionine sulfoxide reductases, MsrA and MsrB, which stereospecifically reduce L-methionine-*S*- and *R*-sulfoxide (Met- *S*-/*R*-SO) to methionine, respectively[307] (**Figure B.1**).

While methionine oxidation has been thought to be a consequence of aerobic life and a signal to activate oxidative stress responses, recent studies suggest that this reversible modification may also play a regulatory role in cells. This proposal is supported by work with the Mical monooxygenases, which catalyse the formation of methionine-*R*-sulfoxide in actin, thus

regulating actin assembly and macrophage function[306]. While enzymatic methionine oxidation is not well understood, the discovery of the Mical proteins raises the possibility that reversible, enzyme-catalysed methionine oxidation may be a common mechanism of protein regulation and may further play a larger biological role than once envisaged.

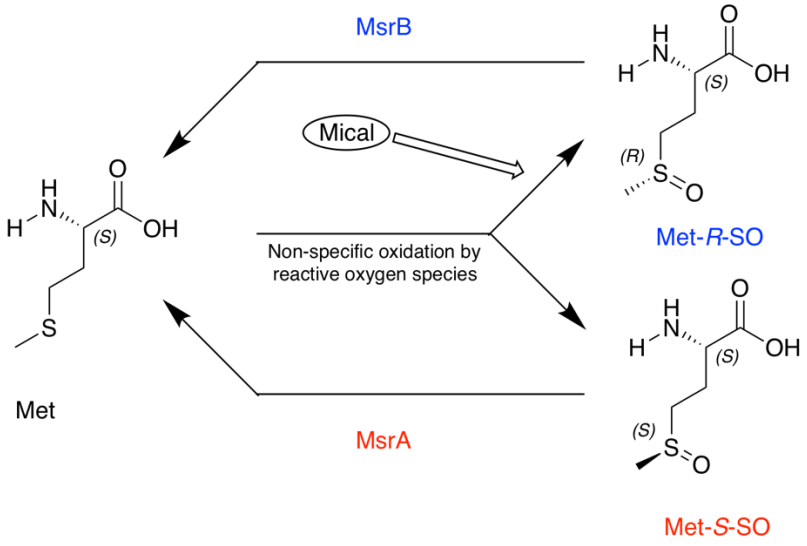


Figure B.1. Methionine oxidation and reduction. Reactive oxygen species cause the non-specific oxidation of Met to methionine-*R*/*S*-sulfoxide (Met-*R*/*S*-SO). The methionine sulfoxide reductases, MsrA and MsrB, catalyse the reduction of Met-*S*-SO and Met-*R*-SO, respectively. The Mical oxygenases have been reported to catalyse stereoselective formation of Met-*R*-SO.

As described in **Chapter 2**, JMJD7 substrate hunting identified two 20-mer peptides from two related proteins, High mobility group box proteins 1 and 2 (HMGB1-2), as JMJD7 substrates. These peptides ($HMGB1_{1\text{-}20}$, MGKGDPKKPRGKMSSYAFFV-NH_2; and $HMGB2_{1\text{-}20}$, MAKGDPKKPKGKMSAYAFFV-NH_2) were identified to undergo +16 and +32 Da modifications, consistent with two oxidations (e.g. hydroxylation and/or methionine oxidation; **Figure B.1**). A series of *C*-terminal truncations were made to narrow down the modified residue (data not shown), and a purified 16-mer peptide ($HMGB1_{1\text{-}16}$, MGKGDPKKPRGKMSSY-NH_2; GL Biochem, China) was used for subsequent experiments, which aimed to identify the target residue and characterise JMJD7 catalysis.

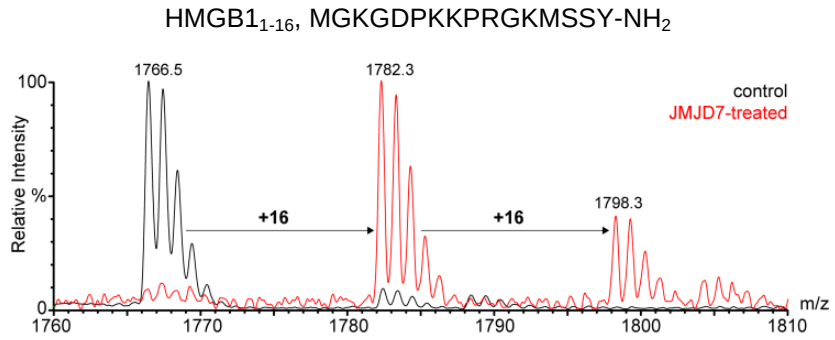


Figure B.2. MALDI-TOF mass spectra of an HMGB1$_{1-16}$ peptide incubated in the presence (red) and absence (black) of JMJD7. JMJD7 catalyses oxidation of a peptide fragment derived from HMGB1 (HMGB1$_{1-16}$, MGKGDPKKPRGKMSSY-NH$_2$). Note that there are two methionine residues in this peptide sequence. Assay conditions: 100 µM Fe(II), 200 µM 2OG, 500 µM ascorbate, 50 µM peptide, ± 10 µM His$_6$-JMJD7 in 50 mM HEPES pH 7.5, 37°C, 1 h.

B.1.1. Aims

Given the biochemical assignment of JMJD7 as a (3*S*)-lysyl hydroxylase described in **Chapter 2**, it was of interest to determine whether HMGB1-2, like DRG1-2, were also genuine JMJD7 substrates. The work described in this appendix chapter thus aims to (1) identify the HMGB1-2 target residues of JMJD7 catalysis; (2) assign JMJD7 catalytic activity; and (3) determine whether HMGB1-2 are genuine cellular JMJD7 substrates.

B.2. JMJD7 catalyses methionine oxidation on HMGB1 is an enzyme-dependent manner

B.2.1. Alanine scanning mutagenesis

Alanine scanning (described in **Chapter 2.3.1**) was employed on the HMGB1$_{1-16}$ peptide to identify the target residue(s). 16-mer peptides with an alanine substitution at each position of the HMGB1$_{1-16}$ peptide were synthesised by SPPS and screened for JMJD7-catalysed hydroxylation using MALDI-TOF-MS. Alanine scanning revealed that the +16 Da species was significantly reduced in the M1A variant (24%) and double M1A, M13A variant (2%), as compared to the wildtype HMGB1 peptide (control, 81%), thus suggesting that M1 is the primary target residue (**Figure B.3**). The results further indicated that the +32 Da peak likely

corresponds to non-specific oxidation of M13, given that a small amount of an oxidised species (+16 Da) is observed in the no-enzyme control [Fe(II), 2OG, and ascorbate, only].

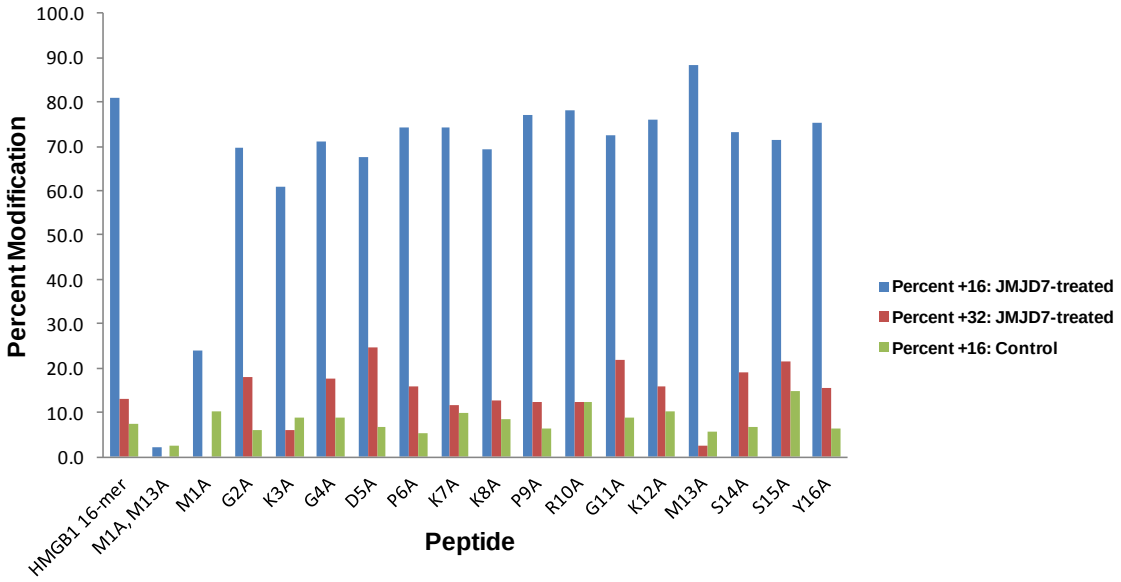


Figure B.3. Alanine scanning suggests that M1 is the target residue. Mutation of M1 (M1A) results in near ablation of the oxidised product. Mutation of M1 and M13 (M1A, M13A) results in complete ablation of oxidised products, suggesting that M1 is the target residue and M13 undergoes background, non-specific oxidation under the assay conditions described in **Figure B.2**. Blue bars represent the percentage of oxidised product in the JMJD7-treated sample (+16); red bars represent doubly oxidised product in the JMJD7-treated sample (+32); green bars represent percentage of oxidised product in the no-enzyme control (background oxidation; +16).

The identification of M1 as the target residue was surprising given the assignment of JMJD7 as a (3*S*)-lysyl hydroxylase (**Chapter 2**). Furthermore, methionine oxidation is a common artefact of hydroxylation assays, conditions of which promote oxidation. While it may well be that methionine oxidation in the presence of JMJD7 is an artefact, especially considering that the initiator methionine of HMGB1 is likely cleaved in cells (according to an initiation methionine cleavage prediction model; http://www.isv.cnrs-gif.fr/terminator2/index.html[243]), the extent of oxidation observed in the JMJD7-treated sample versus the no-enzyme control was striking. Thus, MALDI-TOF-MS/MS analyses were performed to investigate M1 as the target residue.

B.2.2. MALDI-TOF-MS/MS analyses

Analysis of the JMJD7-modified $HMGB1_{1-16}$ peptide by MALDI-TOF-MS/MS (performed with assistance from Dr Holger Kramer, University of Oxford; currently at Imperial College London) revealed a dominant 1718.1 Da peak in the JMJD7-treated sample that was not present in the no-enzyme control (**Figure B.4**).

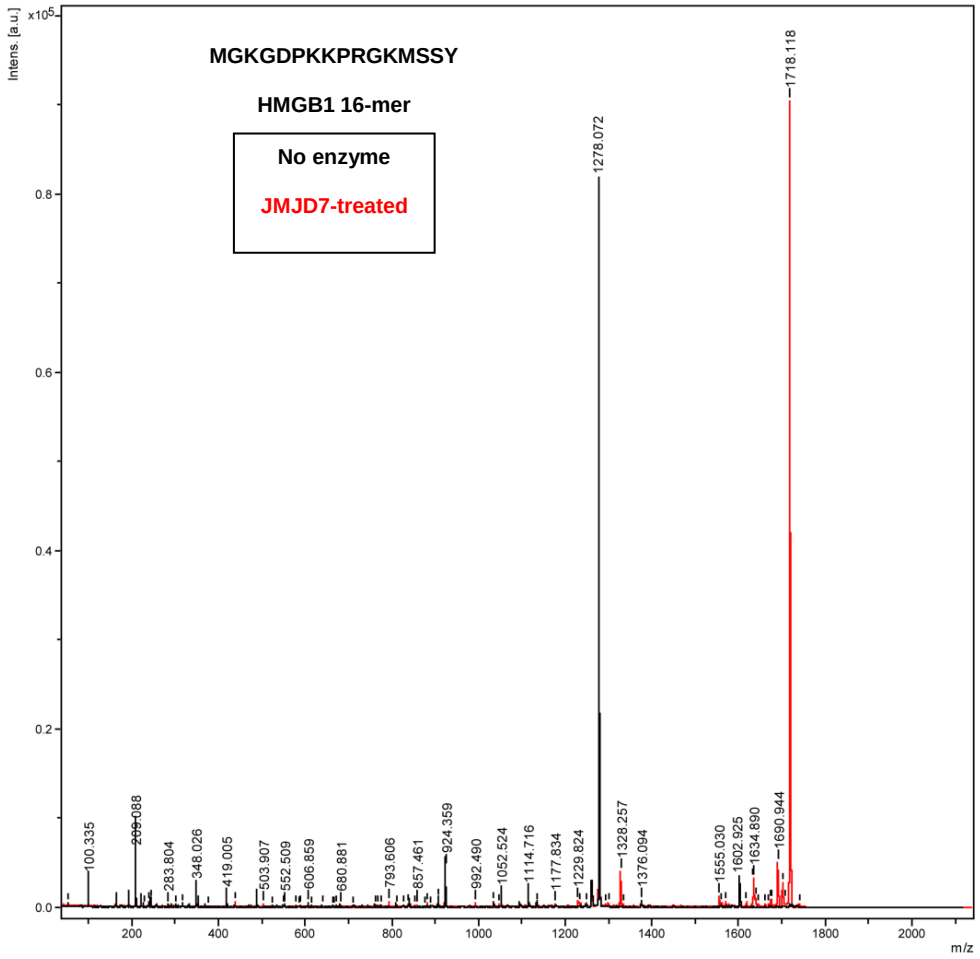


Figure B.4. MALDI-TOF-MS/MS spectra of $HMGB1_{1-16}$ following incubation with (red) and without (black) JMJD7. The spectra suggest that the +16 Da modification corresponds to methionine oxidation, given the presence of a dominant peak, which is -64 Da from the mass of the intact, oxidised peptide ($[M+H]^+$ = 1781.9). A -64 peak corresponds to commonly observed loss of methane sulphenate (CH_3SOH) from oxidised methionine at high laser energies[308].

Notably, the peak at 1718.1 Da, which is -64 Da from the mass of the intact, oxidised peptide ($[M+H]^+$ = 1781.9), is indicative of the elimination of methane sulphenate (CH_3SOH) from oxidised methionine at high laser energies[308]. The appearance of a peak (typically the most abundant ion in the spectrum) corresponding to $[M+H]^+$-64 is a signature mark of methionine

oxidation[308]. Given the poor intensity of the other peaks, assignment of b and y ions could not be reliably performed. However, the combined alanine scanning and MS/MS data indicated that methionine oxidation of residue M1 was significantly increased in the presence of JMJD7.

B.2.3. Co-factor dependence assays

Given the large of extent of methionine oxidation observed in the presence of JMJD7, co-substrate/factor dependence assays were performed to investigate whether the modification was in fact enzyme-dependent. Consistent with 2OG oxygenase catalysis, methionine oxidation was apparently dependent on 2OG and Fe(II), as evidenced by the reduction in methionine oxidation observed in their absence. Furthermore, the presence of NOG, a broad spectrum 2OG oxygenase inhibitor, and substitution of JMJD7 with the Fe(II)-binding H178A variant, similarly reduced the extent of methionine oxidation. As observed for DRG1 co-factor/substrate dependence assays, ascorbate was not required, though enhanced catalytic activity. Overall, the preliminary results from the co-factor/substrate dependence assays should be taken with caution, given that Fe(II) and ascorbate can perform Fenton chemistry, and in the presence of oxygen form reactive oxygen species, including hydrogen peroxide(H_2O_2). It is likely that the production of H_2O_2 (either JMJD7-mediated or not) is causing the background methionine oxidation observed in the no enzyme control.

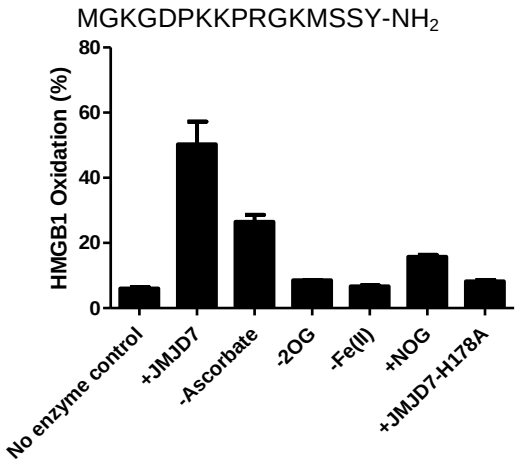


Figure B.5. HMGB1$_{1-16}$ methionine oxidation is JMJD7-dependent. Consistent with 2OG oxygenase catalysis, HMGB1 oxidation relies on 2OG and Fe(II) and is inhibited by NOG (pan-2OG oxygenase inhibitor). A catalytically less active variant (JMJD7-H178A) does not catalyse HMGB1 oxidation. Note the background oxidation present in the no-enzyme control. Assay conditions were as described in **Figure B.2**. Data represent mean ± s.d. (n = 3).

B.3. Investigation of the stereospecificity of JMJD7-catalysed methionine oxidation

**Assays described in this section were performed in collaboration with Maria Tossounian (VIB), who kindly produced and provided all MsrA/B enzymes used in this study.*

It was hypothesised that JMJD7 may be a genuine methionine oxidase should its methionine oxidase catalysis be stereoselective (i.e. the observed methionine oxidation is not non-specific). The stereoselective methionine sulfoxide reductases, MsrA and MsrB, were thus exploited to investigate JMJD7 stereospecificity.

B.3.1. MALDI-TOF-MS assays

An $HMGB1_{1-12}$ peptide (MGKGDPKKPRGK-NH_2) containing only one methionine residue (M1) to simplify analysis, was synthesised by SPPS and purified by HPLC. The peptide was oxidised by JMJD7 in phosphate buffered saline (PBS) buffer (as described in **Figure B.2**) and quenched with EDTA, an Fe(II) chelator (final concentration of 400 μM). The JMJD7-oxidised peptides were then incubated with MsrA and MsrB at 37°C in the presence of 330 μM DTT to investigate whether the stereoselective enzymes could reduce the oxidised peptide. As shown in **Figure B.6**, MsrA was unable to reduce the peptide; however, MsrB reduced ~40% of oxidised peptide, indicating that JMJD7 *may* catalyse formation of Met-*R*-SO.

It is important to note that these preliminary results should be interpreted with caution given that MsrB does not completely reduce the peptide (as would be expected if JMJD7 catalyses only formation of Met-*R*-SO). While it is possible that MsrB does not completely reduce the peptide because a higher concentration of MsrB is required, a positive control (e.g. a peptide oxidised by H_2O_2 to yield a mixture of Met-*R/S*-SO) is also required to prove that MsrA is in fact active. This would to rule out the possibility that the reason MsrB does not completely reduce the peptide is because the JMJD7-catalysed product may be a mixture of Met-*R/S*-SO.

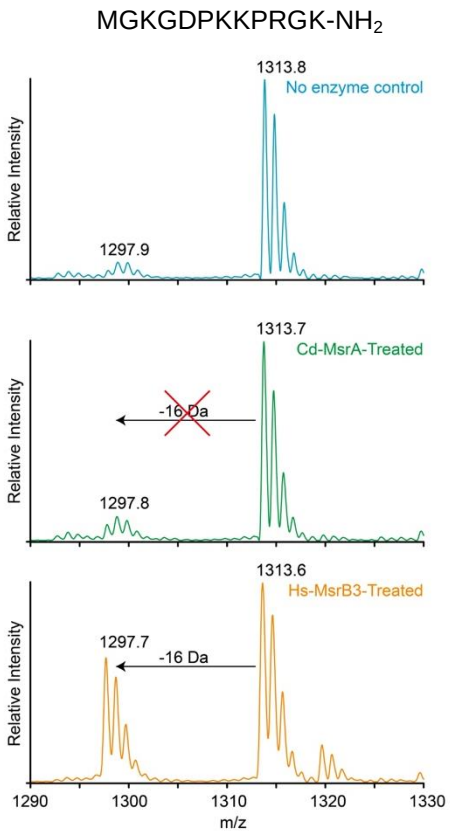


Figure B.6. MALDI-TOF mass spectra corresponding to HMGB1$_{1-12}$ JMJD7-oxidised peptide (blue) in the absence of a reductase, (green) in the presence of MsrA, and (yellow) in the presence of MsrB. Note that MsrB reduces the JMJD7-oxidised peptide (though not completely). Performed in collaboration with Maria Tossounian (VIB).

B.3.2. Amino acid analysis

Amino acid analyses using a non-specific protease from *S. griseus* and UPLC-MS (described in **Section 2.3.4.1**) were performed on an HMGB1$_{1-16}$-M13A (MGKGDPKKPRGK**A**SSY-NH$_2$) variant peptide to investigate whether JMJD7 catalyses stereoselective methionine oxidation. As shown in the extracted ion chromatograms corresponding to the mass of AQC-derivatised Met-SO (m/z = 336.1; **Figure B.7A**), the two peaks arising from the diastereomeric, commercially available L-Met-*R*/*S*-SO standard (Sigma-Aldrich) were resolved following incubation of the standard with MsrA. Given that MsrA stereoselectively catalyses reduction of L-Met-*S*-SO (**Figure B.1**), it could be concluded that the first peak (earlier retention time) corresponds to L-Met-*R*-SO, given that the second peak is no longer present following incubation with MsrA.

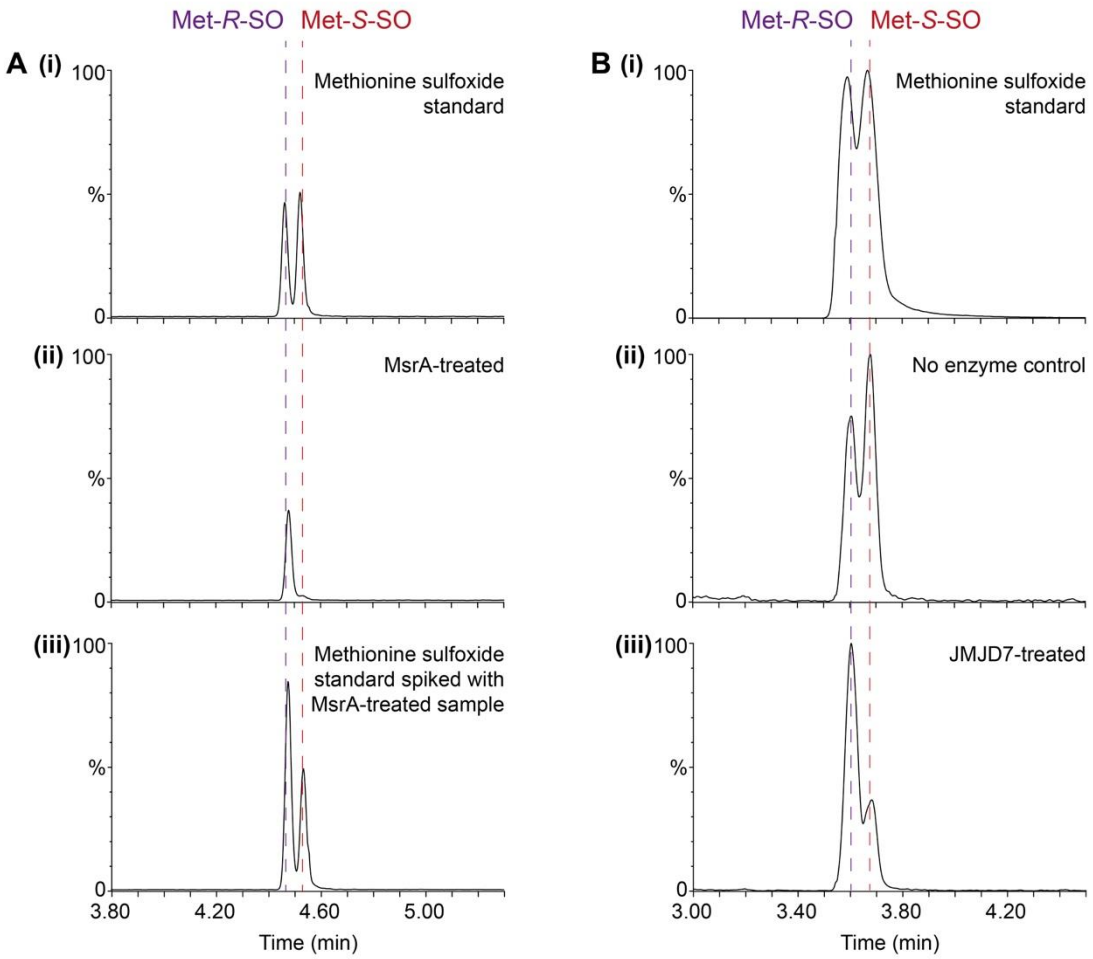


Figure B.7. Extracted ion chromatograms (XID) corresponding to the mass of derivatised Met-SO (m/z = 336) for the (A) resolution of Met-R/S-SO diastereomers and (B) determination of the stereochemistry of the JMJD7-catalysed product. (A) XID for the ***(i)*** commercially available L-methionine sulfoxide standard (a mixture of diastereomers); ***(ii)*** Met-SO standard following incubation with MsrA; and ***(iii)*** Met-SO standard spiked with the MsrA-treated sample. Because MsrA reduces Met-*S*-SO to Met, the data suggest that the second peak corresponds to Met-*S*-SO, while the first peak corresponds to Met-*R*-SO. ***(B)*** XID for the ***(i)*** the Met-SO standard; ***(ii)*** $HMGB1_{1-16}$-M13A no-enzyme control; and ***(iii)*** $HMGB1_{1-16}$-M13A JMJD7-treated sample, revealing that Met-*R*-SO is present in a ~3:1 excess over the *S*-stereoisomer. Note the differences in absolute retention times arise from having performed (A) and (B) on different days. MsrA and MsrB were kindly prepared and provided Maria Tossounian (VIB).

Analysis of unmodified $HMGB1_{1-16}$-M13A post-hydrolysis indicated that methionine sulfoxide was indeed present in the no-enzyme control (**Figure B.7B**), indicating that non-specific methionine oxidation (near 1:1 ratio of *R*:*S*) had occurred, likely due to the oxidative assay conditions and/or hydrolysis conditions (overnight incubation with a non-specific protease at 37°C). Notably, however, analysis of the JMJD7-treated peptide indicated a large excess of *R*-stereoisomer (3:1 ratio of *R*:*S*), suggesting that JMJD7 catalyses selective formation of Met-*R*-SO. While the amino acid analyses corroborate the MALDI-TOF-MS assays described in

Section B.3.1, these results can only be regarded as preliminary because background methionine oxidation must be fully eliminated to conclusively define JMJD7 catalysis as stereoselective. Thus, the experiment warrants further optimisation, which might include protease or acid hydrolysis under inert gas (e.g. nitrogen) to prevent non-specific methionine oxidation. Additionally, quantitative amino acid analysis may provide a better alternative method.

B.3.3. Reversed-phase chromatography analyses

Analytical reversed-phase chromatography (RPC) analyses were also performed [in collaboration with Maria Tossounian (VIB) at the University of Gent] to monitor the oxidation/reduction of the $HMGB1_{1-12}$ peptide using the retention times of the oxidised/reduced peptides. **Figure B.8** shows representative chromatograms for the $HMGB1_{1-12}$ peptide under the following conditions: (A) reduced/untreated; (B) JMJD7-treated; (C) H_2O_2-treated; and (D) JMJD7-treated following incubation with DTT. These chromatograms not only indicate the retention times of the reduced and oxidised peptides, but also show that the presence of DTT (required to maintain the reductase activity of the Msr enzymes) does not reduce the oxidised the peptide. These chromatograms further indicate that the use of a C18 ZipTip column to purify the peptide from H_2O_2 and the enzymes results in large peptide loss, likely due to the hydrophilic nature of the peptide.

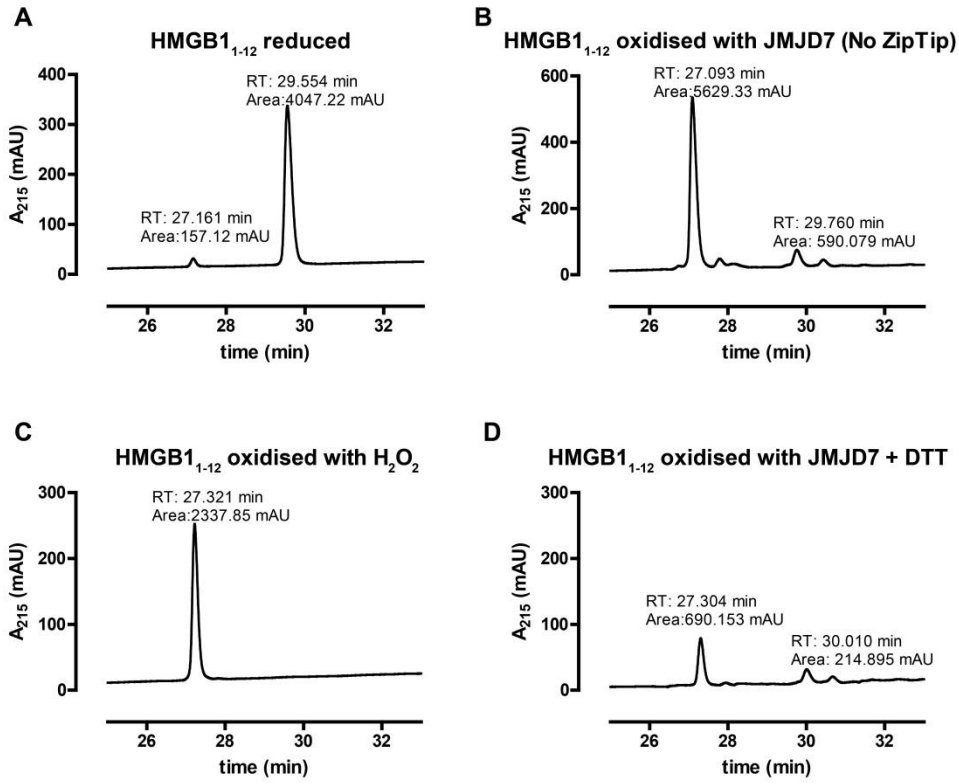


Figure B.8. RPC assay chromatograms corresponding to (A) $HMGB1_{1-12}$ peptide; (B) $HMGB1_{1-12}$ peptide treated with JMJD7 (no ZipTip purification); (C) $HMGB1_{1-12}$ peptide treated with H_2O_2 (with ZipTip purification); (D) JMJD7-treated $HMGB1_{1-12}$ peptide incubated with DTT (with ZipTip purification). The oxidised peptide elutes at ~27 minutes, whereas the reduced peptide elutes at ~30 minutes. Assays were performed in collaboration with Maria Tossounian (VIB) at the University of Gent.

Chromatograms corresponding to the JMJD7-oxidised peptide post-incubation with MsrA and MsrB revealed that while the oxidised peptide underwent minimal (if any) reduction in the presence of MsrA, it underwent nearly 50% reduction in the presence of MsrB (**Figure B.9**). Positive controls were also performed by incubating MsrA and MsrB with H_2O_2-oxidised peptides (which would yield a mixture of L-Met-*R*/*S*-SO). While the signal-to-noise is the chromatograms is poor for these peptide samples due to loss of peptide following C18 ZipTip purification, they do show that ~50% reduction of the H_2O_2-oxidised peptide took place in both samples, consistent with non-specific oxidation yielding a 1:1 mixture of L-Met-*R*/*S*-SO. Together with the MALDI-TOF-MS assays (**Section B.3.1**) and amino acid analyses (**Section B.3.2**), these preliminary experiments all suggest that JMJD7 catalyses formation of L-Met-*R*-SO. However, as for the MALDI-TOF-MS assays and amino acid analyses, conclusive evidence for stereoselective JMJD7-catalysed Met-*R*-SO formation by RPC would require near

complete reduction (>80%) of the JMJD7-oxidised peptide by MsrB and higher quality positive/negative controls (i.e. optimising ZipTip purification).

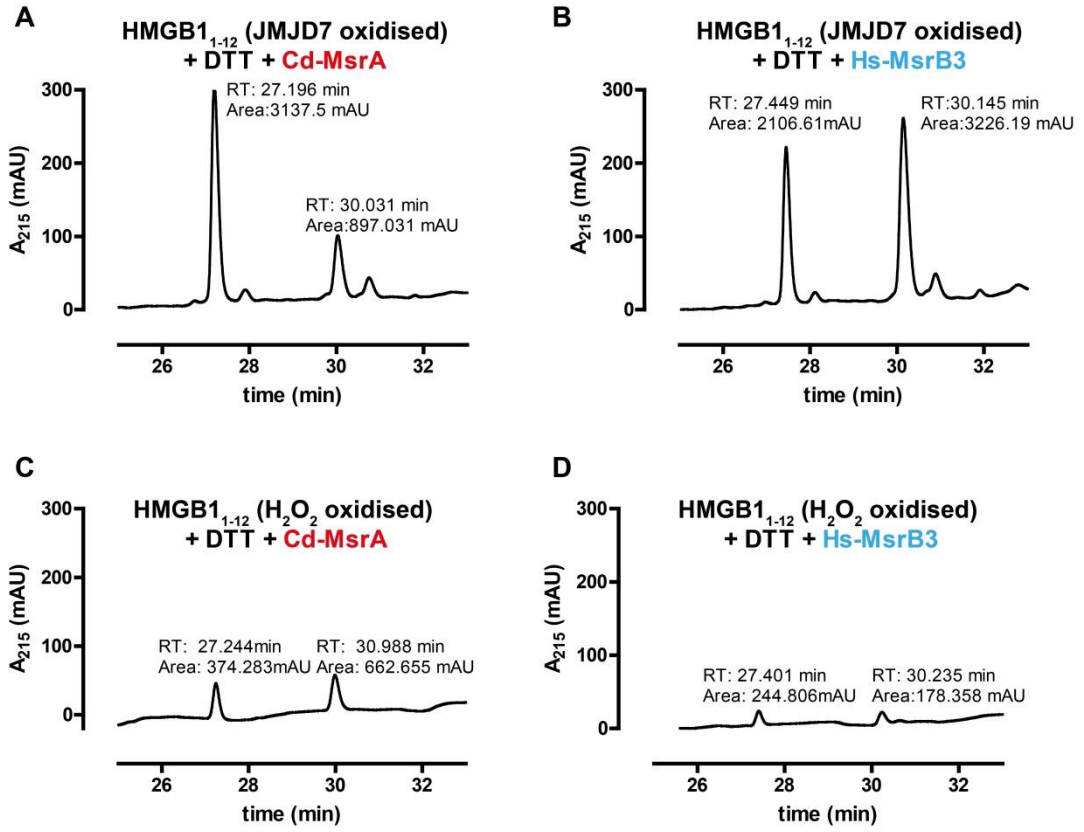


Figure B.9. RPC assays performed with methionine sulfoxide reductases. (A) JMJD7-oxidised $HMGB1_{1-12}$ peptide (~27 min) incubated with MsrA reveals small reduction. ***(B)*** JMJD7-oxidised $HMGB1_{1-12}$ peptide incubated with MsrB indicates that ~50% of the oxidised peptide (~27 min) is reduced (~30 min), suggesting that JMJD7 oxidises Met-R-SO. ***(C)*** H_2O_2-oxidised $HMGB1_{1-12}$ peptide treated with MsrA (following ZipTip purification to remove H_2O_2). ***(D)*** H_2O_2-oxidised $HMGB1_{1-12}$ peptide treated with MsrB (following ZipTip purification to remove H_2O_2). Assays were performed in collaboration with Maria Tossounian (VIB) at the University of Gent.

B.4. Summary and future work

The work described in this appendix chapter shows that JMJD7, a previously uncharacterised human enzyme, is a 2OG oxygenase capable of catalysing methionine sulfoxidation. This is of interest to the 2OG oxygenase field, given that no 2OG oxygenase has been reported catalyse methionine oxidation to date. While enzyme-catalysed methionine sulfoxidation is currently less well-understood than non-enzymatic methionine oxidation, a growing body of evidence

suggests that this modification has an important regulatory role in cells[304]. This is exemplified by the recently discovered role of the Mical proteins, which catalyse formation of L-methionine-*R*-sulfoxide[306]. Interestingly, the preliminary evidence presented in this chapter suggests that JMJD7 also catalyses formation of L-methionine-*R*-sulfoxide, though as noted, these preliminary data require optimisation and validation. It would, in a way, be surprising if JMJD7 was found to catalyse methionine-*R*-sulfoxidation, given that stereochemical outcome of JMJD7-catalysed (3*S*)-lysyl hydroxylation is the opposite relative to the amide backbone. The future investigation of JMJD7-catalysed methionine oxidation will require optimisation of the MS-based experiments described, which will be challenging given that methionine oxidation is a common artefact of hydroxylation/oxidation assays. However, the preliminary results in this chapter offer a foundation for these future studies.

The identification of HMGB1 as a physiologically-relevant JMJD7 substrate may provide clues as to other biological pathways in which JMJD7 is involved. Thus, future studies might also investigate the potential for JMJD7-catalysed methionine oxidation in cells, though consider that this may be difficult given that the target initiator methionine of HMGB1 is predicted to be cleaved in cells following protein synthesis[243]. Though the mechanisms underlying initiator methionine cleavage are not fully understood, it is possible that JMJD7-catalysed methionine oxidation is a signal for such cleavage. Alternatively, as HMGB1 is involved in cytokine release[309], it is possible that JMJD7-catalysed methionine oxidation plays a regulatory role in the immune and/or inflammatory response. Interestingly, there is some evidence that methionine oxidation regulates signalling pathways involved in inflammation[310]. Given that oxygen is an absolute requirement for oxygenase catalysis, JMJD7 may provide a new link between cellular redox control and oxygen availability.

B.5. Materials and methods

B.5.1. JMJD7 oxidation assays

JMJD7 oxidation assays were performed as described in **Section 7.7.1.1** and analysed by MALDI-TOF MS. For assays undergoing subsequent reduction by MsrA/B, the reactions were quenched with 400 μM EDTA prior to the addition of MsrA/B as described in **Section B.5.2**.

B.5.2. Reductase assays with MsrA and MsrB

B.5.2.1. MALDI-TOF-MS assays

Reductase assays were performed using 33.3 μM oxidised peptide, 16.7 μM MsrA/B, and 330 μM DTT in argon-flushed PBS buffer, pH 7.4. DTT is required for optimal reductase activity, and all buffer and water solutions were argon (or nitrogen)- flushed for 10 minutes prior to use. Note that these conditions require optimisation.

B.5.2.2. Amino acid analysis

To resolve L-methionine sulfoxide stereoisomers by amino acid analysis, a 300 μL reaction containing 50 μM of L-methionine sulfoxide, 1.25 μM *Cd*-MsrA, and 500 μM DTT was incubated for 30 minutes at 37°C in nitrogen-flushed 50 mM HEPES, pH 7.5 (incubation conditions provided by Maria Tossounian, VIB[311]). The reaction was quenched and the enzyme was precipitated following the addition of 300 μL acetonitrile and formic acid (1% final concentration). The solution was incubated on ice for 30 minutes prior to pelleting the precipitate by centrifugation. The sample was dried in a speed vacuum and stored at -20°C prior to analysis.

B.5.2.3. Reversed-phase chromatography assays

Positive controls (H_2O_2-oxidised) were prepared by incubating ~30 nmol of $HMGB1_{1-12}$ peptide with 0.6% (160 mM) H_2O_2 for 30 min at 30°C. H_2O_2 was removed from the solution using C18 ZipTip (Pierce™ C18 Tips, 100 μL bed, Thermo Scientific) pipette tips according to the

manufacturer's protocol. Note that a 0.1% TFA solution (0% acetonitrile) was used as the rinse/wash solution because the peptide was extremely hydrophilic (consequently resulting in large peptide loss). JMJD7-treated peptides (~3 nmol) were incubated with 1 nmol of MsrA or MsrB and 200 μM DTT for 30 min at 37°C prior to ZipTip purification and analysis by RPC at the University of Gent. Notably, these reported conditions, particularly the reductase conditions and ZipTip purification, require further optimisation.

B.5.3. Solid phase peptide synthesis

All peptides were synthesised with *C*-terminal amides as described in **Sections 7.5.1** or **7.5.2**.

B.5.4. MALDI-TOF-MS/MS

MALDI-TOF-MS/MS was done in collaboration with Dr Holger Kramer (Centre for Gene Function, *OX*ION, University of Oxford) according to the following protocol courtesy of Dr Kramer. Reactions for MALDI-TOF-MS/MS analysis were prepared as described in **Section B.5.1**. MALDI-TOF-MS was carried out on a Bruker Ultraflex TOF/TOF mass spectrometer. The instrument was calibrated in MS mode directly before data acquisition with Peptide Calibration Standard II (Bruker Daltonics, Coventry, UK) using monoisotopic peptide masses. Samples were prepared by mixing CHCA matrix solution with sample in a ratio of 1:1 and spotting directly onto a Bruker Daltonics 384-spot MALDI plate. Sample ionisation was achieved with a nitrogen laser (337 nm) at 45-55% laser energy, and MS/MS spectra were acquired by Laser-induced fragmentation (LIFT) method. Data was analysed using Flex Analysis (Bruker Daltonics).

B.5.5. Amino acid analysis

Sample preparation, protease hydrolysis, and UPLC-MS detection of standards and HMGB1 samples were performed as described for the JMJD7 DRG substrate peptides in **Section 7.8.5**.

C APPENDIX: COPYRIGHT PERMISSIONS

11200 Rockville Pike
Suite 302
Rockville, Maryland 20852

August 19, 2011

American Society for Biochemistry and Molecular Biology

To whom it may concern,

It is the policy of the American Society for Biochemistry and Molecular Biology to allow reuse of any material published in its journals (the Journal of Biological Chemistry, Molecular & Cellular Proteomics and the Journal of Lipid Research) in a thesis or dissertation at no cost and with no explicit permission needed. Please see our copyright permissions page on the journal site for more information.

Best wishes,

Sarah Crespi

American Society for Biochemistry and Molecular Biology
11200 Rockville Pike, Rockville, MD
Suite 302
240-283-6616
JBC | MCP | JLR

Tel: 240-283-6600 • Fax: 240-881-2080 • E-mail: asbmb@asbmb.org

NATURE PUBLISHING GROUP LICENSE
TERMS AND CONDITIONS

Nov 23, 2016

This Agreement between Suzana Markolovic ("You") and Nature Publishing Group ("Nature Publishing Group") consists of your license details and the terms and conditions provided by Nature Publishing Group and Copyright Clearance Center.

License Number	3994890365353
License date	Nov 23, 2016
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Nature
Licensed Content Title	Ribosomal oxygenases are structurally conserved from prokaryotes to humans
Licensed Content Author	Rasheduzzaman Chowdhury, Rok Sekirnik, Nigel C. Brissett, Tobias Krojer, Chia-hua Ho, Stanley S. Ng
Licensed Content Date	May 11, 2014
Licensed Content Volume Number	510
Licensed Content Issue Number	7505
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
High-res required	no
Figures	Extended Data Figure 10: Phylogenetic relationships of human JmjC 2OG-dependent oxygenases. (doi:10.1038/nature13263)
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Biochemical and structural characterisation of Jumonji-C hydroxylases
Expected completion date	Apr 2017
Estimated size (number of pages)	300
Requestor Location	Suzana Markolovic Chemistry Research Laboratory 12 Mansfield Road

	Oxford, OX1 3TA United Kingdom Attn: Suzana Markolovic
Billing Type	Invoice
Billing Address	Suzana Markolovic Chemistry Research Laboratory 12 Mansfield Road Oxford, United Kingdom OX1 3TA Attn: Suzana Markolovic
Total	0.00 USD

Terms and Conditions

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use,subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run).NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
   For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

   Note: For republication from the *British Journal of Cancer*, the following credit lines apply.
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)For AOP papers, the

credit line should read:
Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   Note: For adaptation from the *British Journal of Cancer,* the following credit line applies.
   Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information.Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

   Note: For translation from the *British Journal of Cancer,* the following credit line applies.
   Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.
Special Terms:
v1.1

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.