

Studies on Collagen Post-Translational Modifications

Gareth William Langley

Kellogg College

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

Trinity Term 2017

Abstract

The collagen proteins are the principal component of the extracellular matrix and are the most abundant protein family in humans. Therefore, collagens form the major protein component of bone, tendon and cartilage. During collagen biosynthesis, proline and lysine residues are enzymatically converted to (*2S,4R*)-hydroxyproline, (*2S,3S*)-hydroxyproline and (*2S,5R*)-hydroxylysine by the collagen prolyl-4-hydroxylases (C-P4Hs), prolyl-3-hydroxylases (P3Hs) and procollagen-lysine, 2-oxoglutarate 5-dioxygenases (PLODs) respectively. The C-P4Hs, P3Hs and PLODs are members of the Fe^{II} and 2-oxoglutarate dependent oxygenase (2OG oxygenase) superfamily. Despite being amongst the first 2OG oxygenases to be identified, the collagen hydroxylases are poorly understood from structural and biochemical perspectives, a consequence of inefficient recombinant production methods.

Initial efforts focussed on developing a robust expression system for recombinant C-P4H. An *E. coli* expression system was pursued to allow for relatively large quantities of protein to be produced, as required for structural studies. A number of novel expression constructs were generated and used in small-scale expression trials. However, attempts to isolate purified C-P4H tetramer have ultimately proved unsuccessful.

A prolyl hydroxylase from *Paramecium bursaria* Chlorella Virus I (vCPH), was pursued as a model collagen prolyl hydroxylase; production of vCPH in *E. coli* affords large quantities of high purity recombinant protein for subsequent analyses. vCPH catalyses the formation of (*2S,4R*)-hydroxyproline at proline 3 and proline 5 of the short 8-mer peptide PAPKPAPK and kinetic analyses have revealed that the ascorbate dependence of vCPH-catalysis at these two positions differs significantly, indicating that ascorbate dependence of the collagen prolyl hydroxylases varies dependent upon the residues adjacent to the target residue.

A vCPH inhibition assay was developed and a library of potential inhibitors was screened. This screen included compounds that inhibit a different subfamily of 2OG dependent prolyl-4-hydroxylases, the hypoxia inducible factor prolyl hydroxylases (HIF PHDs), which also catalyse the formation of (*2S,4R*)-hydroxyproline. The studies revealed that inhibitors of the HIF PHDs that are currently in clinical trials also inhibit vCPH. However, analysis of novel HIF PHD inhibitors and inhibitors of the prolyl-3-hydroxylase OGFOD1 indicate that achieving selectivity between these prolyl hydroxylases is possible.

The inhibition studies have been complemented with amino acid analyses of collagen samples. No evidence for the existence of hydroxylated residues other than (*2S,4R*)-hydroxyproline, (*2S,3S*)-hydroxyproline and (*2S,5R*)-hydroxylysine was acquired. These amino acid analysis approaches may be taken further to analyse collagen samples derived from tissue to assess hydroxylation levels in diseased tendon.

Additionally, metathesis-derived lysine analogues have been synthesised and used to analyse 2OG-dependent histone demethylase (KDM) catalysis. These analogues have revealed new modes of activity with KDM3A and provided insight into the charge and conformational requirements of 2OG-dependent KDM catalysis.

Overall, the work presented in this thesis has provided insight into the biochemistry of a collagen prolyl hydroxylase and demonstrated the potential for the development of selective inhibitors of prolyl hydroxylases. Additionally, novel reactions catalysed by histone demethylases have been observed.

Contents

Abstract i

Contents iii

Acknowledgements vii

Abbreviations ix

Chapter 1 Introduction 1

- The Extracellular Matrix 3
- Collagen 5
- Collagen Post-Translational Modifications 11
- The 2OG Oxygenases 15
- Prolyl and Lysyl Hydroxylation is Not Limited to Collagen 26
- 2OG Oxygenases as Therapeutic Targets 30
- Thesis Objectives 34

Chapter 2 Human Collagen Prolyl-4-Hydroxylase 35

- Recombinant C-P4H Production 41
- Previous Biochemical Studies with C-P4H 42
- Aims 43
- C-P4H Production with pETDuet-1_PDI_HisCPH1 43
- Constructs Designed by OPPF-UK 49
- Arabinose Inducible Vectors 54
- Dual-Tag Approach 55
- Conclusions 56
- Summary 57

Chapter 3 Characterisation of a Prolyl Hydroxylase from *Paramecium bursaria* Chlorella Virus I 59

- vCPH Structure 62
- Expression and Purification of vCPH 62
- Enzyme Activity Assays 65
- 2OG Turnover 74
- Cofactor and Co-substrate Binding to vCPH 76
- Substrate Selectivity of vCPH 78
- Conclusions 86
- Summary 89

Chapter 4 vCPH Inhibition and Structural Studies 91

- Inhibition Assay Development 96
- Compound Library 97
- Differential Scanning Fluorometry 101
- Initial Screening 102
- IC_{50} Determination 104
- Mechanism of vCPH Inhibition 107
- Counter-screening Novel PHD2 Inhibitors 112
- TCA Cycle Intermediates 116

Structural Studies 119
Conclusions 122
Summary 126
Chapter 5 Amino Acid Analysis of Collagen 129
Tendon Structure 131
The Rotator Cuff 132
Analysing Hydroxylation in Collagen 134
Proyl- and Lysyl- Hydroxylation is Not Necessarily Limited to Proline C4, Proline C3 and Lysine C5 136
Aims 137
Amino Acid Analysis of Commercially Sourced Collagen 137
Results of Amino Acid Analysis with Rat Tail Tendon and Bovine Achilles Tendon 139
Processing of Tissue Samples 145
Conclusions 146
Summary 148
Chapter 6 Lysine Analogues as Probes for 2OG Oxygenases 149
Lysine is a Common Residue for Post-Translational Modifications 153
Lysine Methylation and Demethylation 155
Manipulation of the Lysine Side Chain and Replacement of the N^{ϵ}-Group as a Strategy to Probe Catalysis 156
Aims 157
Synthetic Strategy and Implementation 158
Peptide Library Expansion 160
The Role of Backbone Alignment in KDM4 and KDM7B Catalysis 161
Analysis of Charge Requirements for KDM4 Catalysis 163
Investigation of KDM3A Catalysis with Lysine Analogues 164
Lysyl Hydroxylases 166
Conclusions 170
Summary 172
Chapter 7 Summary and Conclusions 175
Chapter 8 Materials and Methods 181
General Methods 183
Growth Media 186
Electrophoresis and Immunoblotting 188
Plasmid Preparation 191
In-Fusion™ Cloning 196
Peptide Synthesis 198
Recombinant Protein Production and Protein Purification 199
In Vitro Biochemical Assays 205
Mass Spectrometry 209
Nuclear Magnetic Resonance Spectroscopy 214
Differential Scanning Fluorimetry 217
Crystallography 218
Isothermal Titration Calorimetry 219

Organic Synthesis 220
Infra-Red Spectroscopy 221
Melting Point Determination 221
Optical Rotation 221
Chapter 9 References 223
Appendix 1 Human 2OG Oxygenase Names and Reactions 243
Appendix 2 Synthetic Details for Lysine Analogues 251
Appendix 3 Crystallisation Screen Compositions 269
PACT Premier D11 Optimisation Screen 271
Appendix 4 Peptide, Protein and DNA Sequences 273
Peptide Sequences 275
Protein Sequences 276
DNA Sequences 280

Acknowledgements

I would like to thank Prof Chris Schofield for allowing me to work in his lab for the course of the project and for all the help and advice throughout. Additionally, I would like to thank Prof Andrew Carr for agreeing to co-supervise me as well as Dr Tim Johnson, Dr Christophe Genicot and Dr Luis Castro at UCB for their input over the course of the project. I would also like to thank UCB and the DTC/EPSRC for their financial support over the last four years.

I would like to thank all members of the CJS group past, present and future for making the lab an enjoyable place to work and share ideas. A special thank you goes to the people who have helped me on a day to day basis and I would like to pay particular thanks to Dr Richard Hopkinson, Dr Joan Gannon and Dr Adam Hardy for their constant support over the past 4 years.

I would also like to say thanks to Dr Cyrille Thinnes for all of his awesome advice as well as Dr Rebecca Hancock, Dr Michael McDonough, Dr Suzana Markolovic, Dr Sarah Wilkins, Dr Louise Walport, Dr Martine Abboud, Kate Dunne and Dr Christopher Lohans for their friendship, help and support.

In addition, I would like to acknowledge the work of Dr Rachel Shamo Schiller, Anne Brinkø, Dr Tony Tumber and Jimmy Holt-Martyn who have contributed to this work by completing assays and synthesising compounds.

Thanks are also due to Dr Stephanie Dakin who helped me during my time working at the Botnar and who has continued to be a helpful source of information and advice.

I would also like to acknowledge all of the hard work of the support staff in the Department of Chemistry, in particular the Stores, MS and NMR services, who have allowed for my research to be conducted with minimal delay. I would also like to say a special thanks to Dr Adam Hardy, Wendy Sobey, Jenny Houlsby, Val de Newton and Zahn Gower for the roles they have all played in keeping the CJS lab running on a daily basis.

A huge thank you should go to my friends and family who have had to put up with me as I've gone on my crazy Oxford adventure over the past 8 years. To Jimmy Beestone, Jess Parkinson, Stefan Turner, Alice Cantell-Hynes, Alice Jones, Emma Knowles, Hannah Hobson, Corran Pickering, Alison Wilson, Matthew Key, David Mould, Grant Hanson-Vaux, Jenny Saxton, Poppy Roworth, Martin Strohmeier, Anna Niedbala, George Taylor (and many others) – you have all made my time at Oxford special and I hope I can spend a little more time in your company moving forward.

Thank you to all my colleagues at Oxford Cricket Club for welcoming me into the club and making Saturday a day to look forward to as well as to the Oxfordshire Referees Society for being so accommodating and providing me with the opportunity to develop. I'd also like to thank all of the officials and volunteers at Pontypool RFC for keeping our great club alive and I look forward to having more opportunities to support the team in person.

Finally, I would like to pay particular thanks to my parents who have always been there for me and supported me, to my sister Emily, because I suppose I should really say something nice about her and last, but certainly not least, to Alison, who has provided me with her unwavering support and companionship throughout.

Diolch yn fawr pawb!

Abbreviations

Standard one and three letter amino acid codes are used throughout the text. Single letter codes for nucleobases are utilised.

2,4-PDCA	2,4-Pyridine dicarboxylic acid
2HG	2-Hydroxyglutarate
2OG	2-Oxoglutarate
2TY	2 x Tryptone/yeast
AI	Autoinduction
AQC	6-Aminoquinolyl-*N*-hydroxysuccinimidyl carbamate
BBOX	γ-Butyrobetaine hydroxylase
Boc	*tert*-Butyloxycarbonyl
CID	Collision induced dissociation
CODD	*C*-terminal oxygen degradation domain
C-P4H	Collagen prolyl-4-hydroxylase
CPMG	Carr-Purcell-Meiboom-Gill
CRTAP	Cartilage associated protein
Da	Daltons (1 Da = 1 atomic mass unit)
ddH_2O	Doubly distilled H_2O
DIC	*N,N'*-Diisopropylcarbodiimide
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
DSBH	Double-stranded β-helix
DTT	Dithiothreitol
E. coli	*Escherichia coli*
ECM	Extracellular matrix
eRF1	Eukaryotic release factor 1
FACIT	Fibril-associated collagens with interrupted triple helicies
FDH	Fomaldehyde dehydrogenase
FIH	Factor inhibiting hypoxia-inducible factor
Fmoc	Fluorenylmethyloxycarbonyl
Gly-Xaa-Yaa	Collagen repeat of glycine followed by two amino acids
H1	Histone 1
H2A	Histone 2A
H2B	Histone 2B
H3	Histone 3
H3K9	Histone 3, Lysine 9
H4	Histone 4
HEPES	4-(2-Hydroxyethyl)-1-piperazine ethanesulfonic acid
HIF	Hypoxia-inducible factor
HIF-1α	Hypoxia-inducible factor 1α
HIF-2α	Hypoxia-inducible factor 2α
His_6	Hexahistidine
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
HSPBAP1	27 kDa heat shock protein-associated protein 1
Hyp	(*2S,4R*)-Hydroxyproline

IC_{50}	Half maximal inhibitory concentration
IDH	Isocitrate dehydrogenase
IPF	Idiopathic pulmonary fibrosis
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Infra-red
ITC	Isothermal titration calorimetry
JmjC	Jumonji *C*
JMJD4	JmjC domain-containing protein 4
JMJD6	JmjC domain-containing protein 6
JMJD7	JmjC domain-containing protein 7
k_{cat}	Catalytic rate constant
KDM	Lysine demethylase
KDM1	Lysine-specific histone demethylase 1
KDM3A	Lysine-specific histone demethylase 3A
KDM4A	Lysine-specific histone demethylase 4A
KDM4B	Lysine-specific histone demethylase 4B
KDM4C	Lysine-specific histone demethylase 4C
KDM4D	Lysine-specific histone demethylase 4D
KDM4E	Lysine-specific histone demethylase 4E
KDM7B	Lysine-specific histone demethylase 7B
K_i	Inhibition constant
K_M	Michaelis constant
K_M^{app}	Apparent Michaelis constant
LC-MS	Liquid chromatography-mass spectrometry
LH	Lysyl hydroxylase
LIC	Ligation independent cloning
LOX	Lysyl oxidase
m/z	Mass to charge ratio
MACIT	Membrane-associated collagens with interrupted triple helices
MALDI	Matrix-assisted laser desorption/ionisation
MBP	Maltose binding protein
MCS	Multiple Cloning Site
MS	Mass spectrometry
MS-MS	Mass spectrometry-mass spectrometry
Multiplexin	Multiple triple-helix domains and interruptions
NAD^+	Oxidised nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide
NMR	Nuclear magnetic resonance
NODD	*N*-terminal oxygen degradation domain
NOG	*N*-oxalyl glycine
OD_{600}	Optical density at λ = 600 nm
Ofd1	Prolyl 3,4-dihydroxylase Ofd1
OGFOD1	2-Oxoglutarate and iron-dependent oxygenase domain containing 1
OPPF-UK	Oxford Protein Production Facility
P3H	Collagen prolyl-3-hydroxylases
P4H-TM	Transmembrane prolyl-4-hydroxylase
PAGE	Polyacrylamide gel electrophoresis
Pbf	2,3,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl
PCR	Polymerase chain reaction
PDB	Protein data bank
PDI	Protein disulfide isomerase

PHD	Plant homeodomain
PHD1	Hypoxia-inducible factor prolyl hydroxylase 1
PHD2	Hypoxia-inducible factor prolyl hydroxylase 2
PHD3	Hypoxia-inducible factor prolyl hydroxylase 3
PLOD	Procollagen-lysine, 2-Oxoglutarate 5-Dioxygenases
PMSF	Phenylmethylsulfonyl fluoride
PTM	Post-translational modification
RNA	Ribonucleic acid
rpm	Revolutions per minute
S. griseus	*Streptomyces griseus*
SAXS	Small angle x-ray scattering
SDS	Sodium dodecylsulfate
SDS-PAGE	Sodium dodecylsulfate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TB	Terrific broth
TCA	Tricarboxylic acid
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
TLC	Thin layer chromatography
T_m	Melting temperature
TMLH	ε-Trimethyllysine hydroxylase
TOF	Time-of-flight
TPR	Tetratricopeptide repeat
Tris	*Tris*(hydroxymethyl)aminomethane
U2AF65	U2 Small Nuclear RNA Auxiliary Factor 2
UKUFF	United Kingdom rotator cuff trial
UPLC	Ultra performance liquid chromatography
UV	Ultra-violet
vCPH	Proyl-4-hydroxylase from *Paramecium bursaria* chlorella virus 1
VHL	von Hippel–Lindau tumour suppressor protein
Xaa	Undefined amino acid at the second position of the collagen repeat
Yaa	Undefined amino acid at the third position of the collagen repeat

Chapter 1
Introduction

The Extracellular Matrix

Extracellular matrix (ECM) is a broad term that is used to describe the biochemical environment that is present between cells in multi-cellular organisms[1,2]. The ECM is key in determining the structural and biochemical function upon a range of tissues and plays a significant role mediating inter-cell signalling pathways[1]. The ECM is primarily made from protein and carbohydrate components that are arranged to form an ordered environment between cells[1,2]. The principal components of the ECM can be classified into three broad classes:

- Fibrous Proteins (*e.g.* collagens and elastins),
- Proteoglycans, and
- Glycoproteins.

This classification is somewhat prescriptive and the boundaries between classes are blurred; for example, some collagen classes that are part of fibrils can also be classed as proteoglycans.

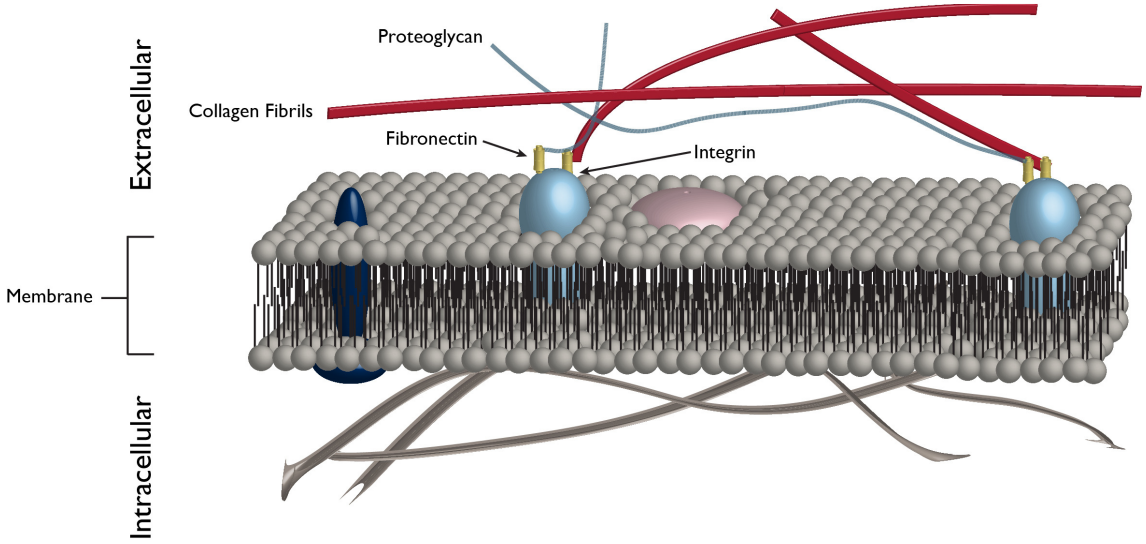


Figure 1 - Cartoon schematic showing a variety of extracellular matrix components and their relation to a cell membrane. Collagen fibrils and proteoglycan chains are highly organised in the ECM.

The ECM provides the scaffold upon which cellular components can exist and therefore ECM composition varies significantly between different tissues and organs. For example, musculoskeletal structures such as bone and cartilage are predominantly formed of fibrous ECM[3,4], the fibrous proteins are arranged to form structures that confer mechanical strength. In

contrast to the fibrous nature of bone, cartilage and connective tissues, the basement membrane is a sheet-like ECM structure that anchor epithelial cells and separates epithelial layers from stroma[5,6].

Tissues which constitute primarily of ECM, such as bone, tendon and cartilage, have cells that are sparsely embedded in an extensive ECM structure[3,7]. These cells are responsible for the production of new matrix proteins and are grouped together under the umbrella term 'fibroblasts'. Fibroblasts are tissue specific and are responsible for the synthesis of the ECM required for the tissue in which they are embedded. For example, tendonocytes and osteoblasts are responsible for the production of tendon and bone ECM respectively, and hence produce primarily fibrous ECM proteins[2,8,9].

ECM production is tightly regulated; ECM building blocks are secreted from nearby cells and then arranged, often through a 'self-assembly' network of receptor interactions and extra-cellular post-translational modification, to form a highly ordered structure that is observed in mature ECM[8,10]. In most tissues the ECM is continuously modelled and remodelled[11] and therefore changes in ECM may develop over time[12-14].

Whilst abhorrent production of ECM may be largely benign, as in the case of scarring, incorrect ECM formation plays a role in a variety of different diseases; examples of fibrotic disease include liver cirrhosis, cystic fibrosis and idiopathic pulmonary fibrosis (IPF)[11]. Fibrotic disease is often characterised by the deposition of excess ECM[15,16] and in a number of cases, such as IPF, can cause irreversible histological changes to the organ[15]. Chronic inflammation is often associated with subsequent fibrosis, as seen in intestinal fibrosis that results from inflammatory bowel disease[17,18].

The interplay between tumor cells and the ECM is important in cancer progression; tumor growth is associated with the abhorrent deposition of ECM, whilst during metastasis (which

vastly reduces the chance of patient survival), tumor cells detach from the ECM at their initial locus before reattaching to the ECM and proliferating at an alternative location[11,19,20].

Collagen

The collagens are the most abundant family of proteins in the human body[9,21]. To date, 28 members of the collagen family have been identified (Table 1), with collagens I, II and III being of the most abundant forms, making up approximately 80% to 90% of collagen proteins in the body[22]. Although 28 collagens are defined, there are other proteins that contain collagen-like domains, such as adiponectin[23].

Collagen proteins are the principal protein component of the ECM and are produced in the endoplasmic reticulum prior to secretion to the ECM[24,25]. The wide range of networked and fibrillar structures that the collagens can form are reflected in the tissues in which collagen is the principal component, this includes structurally significant tissues such as bone, tendon and cartilage[22].

Collagen Structure

The primary structure of collagens is characterised by a triplet repeating unit of glycine followed by two other amino acids (Gly-Xaa-Yaa). The Yaa position is usually a proline residue and the most commonly observed triplet is GPP[26]. The Yaa prolyl residue in this triplet is hydroxylated in the endoplasmic reticulum to (*2S,4R*)-hydroxyproline (commonly referred to as *trans*-4-hydroxyproline) approximately 40% of the time during collagen biosynthesis[26,27]. The presence of (*2S,4R*)-hydroxyproline has long been known to greatly stabilise collagen[28] through the stabilisation of the collagen triple helix[29] (the molecular mechanism of collagen stabilisation by (*2S,4R*)-hydroxyproline is discussed in more detail later in Chapter 1).

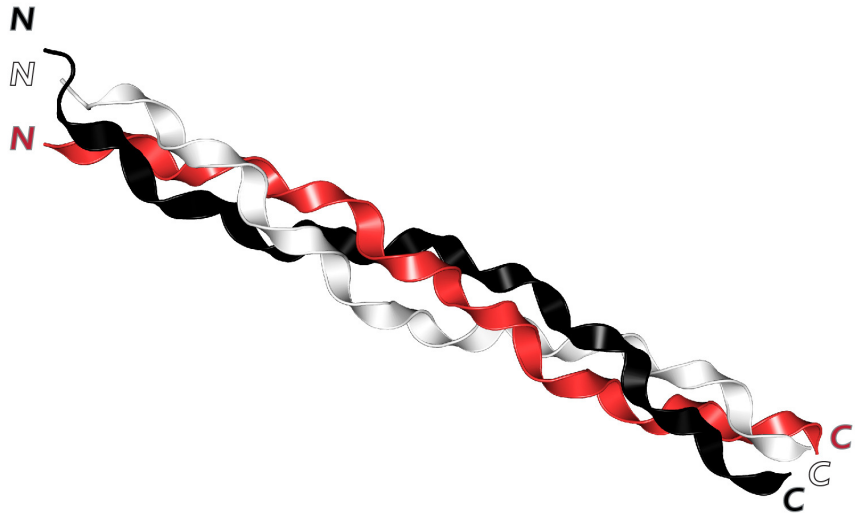


Figure 2 - A view from a crystal structure of an idealised, synthetic collagen (PDBID: 1CAG[30]). Three strands of (Glycine-Proline-(*2S,4R*)-hydroxyproline)$_n$ form individual helicies which arrange themselves in a right-handed superhelix, making up the distinctive collagen triple helix.

The collagen triple helix, a structure first proposed in 1955[31-33], is the characteristic structural feature of the collagen superfamily. Nearly 40 years later, high resolution X-ray diffraction data of a synthetic triple helix region revealed the triple helix in more detail[30]. The triple helix consists of three helical polypeptide chains arranged in a right handed superhelix (Figure 2)[30]. All collagen proteins feature at least one triple helical region, also referred to as the COL domain, a rigid structure that is an ideal building block for connective tissues.

Collagen proteins can either be homo- or hetero-trimeric, the precise nature of which will depend on the collagen variant in question; the trimeric make up of the collagens is detailed in Table 1. *In vivo*, the symmetry of the triple helix deviates from the idealised triple helix and has been found to be dependent on primary structure[34]. In addition, some collagen types feature interruptions to the triple helix, which allow for flexibility that often permits association with other fibrils or ECM components[22].

Collagen Sub-Classes and Variants

In order to form the fibrillar and networked structures that in turn make up the structural component of ECM, collagen proteins are linked together and organised in the extracellular

matrix, eventually resulting in a collagen scaffold that is appropriate for the tissue in question[1,2]. Collagens can be broadly classified into sub-classes that reflect their domain architecture, which helps to dictate the extracellular organisation and structure (some examples of supramolecular collagen organisation are given in Figure 3). Collagens VII and XXVIII have not been fully characterised within these collagen sub-classes.

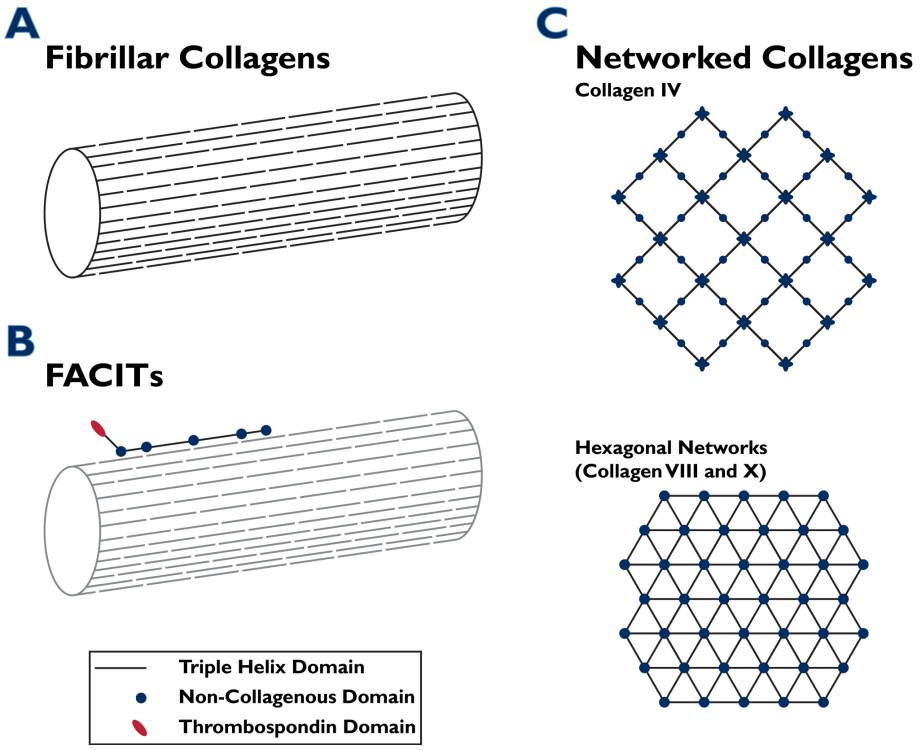


Figure 3 - Schematic showing examples of different supramolecular assemblies that can be formed by members of the collagen family[23,24].

Fibrillar Collagens - Types I, II, III, V, XI, XXIV and XXVII

Fibrillar collagens contain only triple helix domains, which allows for the formation of a rigid fibril (Figure 3A). This allows the fibrillar collagens to be key structural proteins and form the major component of fibrils. Therefore, the fibrillar collagens, particularly types I, II and III, are amongst the most abundant collagen forms and are the major component of bone, tendon and cartilage ECM[4,35]. The fibrillar collagens are highly conserved across *animalia* and stereotypically consist of a triple helix domain flanked by globular *N*- and *C*- terminal propeptides. During the biosynthesis of the fibrillar collagens, both the *N*- and *C*- terminal propeptides are cleaved prior to the formation of larger fibrillar structures[24,35], which occurs following transport to the ECM.

Networked Collagens - Types IV, VI, VIII and X

Similar to the fibrillar collagens, the networked or network-forming collagens are primarily composed of triple helix domains. In the ECM, rather than a fibril, the proteins are organised into a network, often mediated by globular non-collagenous domains at the *N*- or *C*- terminus[36]. Collagens VIII and X form a hexagonal network[23] whilst the more abundant collagen IV forms an expansive network that is the principal component of basal membranes[23,37] (Figure 3C).

Meanwhile, collagen VI forms a network structure that associates with collagen fibrils, helping to bind the fibrils together to form larger fibrils[22]. The collagen VI structure is described as 'beaded filaments'[24]; triple-helical regions of approximately 60 nm in length are separated by globular domains about 40 nm long, giving the impression of beads on a string[22].

Fibril-Associated Collagens with Interrupted Triple Helicies (FACIT) - Types IX, XII, XIV, XVI, XIX, XX, XXI, XXII and XXVI

The fibril-associated collagens with interrupted triple helicies (FACIT) contain multiple, interrupted triple helix domains. FACITs are heavily glycosylated and are often considered to be proteoglycans[22]. The FACITs do not form fibrils themselves and associate themselves with the surface of collagen fibrils[23]. One or more of the non-collagenous domains are flexible and allow for a kink that allows a globular domain to protrude away from the fibril (Figure 3B). The globular domain is usually highly glycosylated and is often involved in the recruitment and binding to other ECM components[22].

Membrane-Associated Collagens with Interrupted Triple Helices (MACIT) - Types XIII, XVII, XXIII and XXV

The membrane-associated with interrupted triple helices (MACIT) or 'membrane collagens' are homotrimeric collagens[38-40] and feature three extracellular triple helix domains broken by non-collagenous domains. MACITs contain a short *N*-terminal intracellular domain which precedes a single pass transmembrane domain, which anchors the collagen to cells[41,42]. The physiological function of the MACITs have not been well resolved. However, collagen XXV has been linked to neuron development[43].

Type	Composition(s)	Structure/Class	Example Tissues/Notes
I	$[\alpha1(I)]_2[\alpha2(I)]$	Fibrillar	Tendon, bone, skin. The most abundant collagen form.
II	$[\alpha1(II)]_3$	Fibrillar	Cartilage, vitreous humor.
III	$[\alpha1(III)]_3$	Fibrillar	Skin, muscle, blood vessels. Often upregulated in damaged tissues.
IV	$[\alpha1(IV)]_2[\alpha2(IV)]$ $[\alpha3(IV)][\alpha4(IV)][\alpha5(IV)]$ $[\alpha6(IV)]_2[\alpha5(IV)]$	Networked	Basal lamina. 6 different chains form at least 3 different collagen IV forms.
V	$[\alpha1(V)]_3$	Fibrillar	Similar to type I. Often found in foetal tissues. Forms hybrids with collagen XI. Globular *N*-terminal domain.
VI	$[\alpha1(VI)][\alpha2(VI)][\alpha3(VI)]$ $[\alpha1(VI)][\alpha2(VI)][\alpha4(VI)]$	Networked	Cartilage, bone, cornea, basal lamina.
VII	$[\alpha1(VII)]_2[\alpha2(VII)]$	Anchored to fibrillar collagen	Dermis, bladder.
VIII	$[\alpha1(VIII)]_3$ $[\alpha2(VIII)]_3$ $[\alpha1(VIII)]_2[\alpha2(VIII)]$	Networked	Dermis, brain, heart, kidney.
IX	$[\alpha1(IX)][\alpha2(IX)][\alpha3(IX)]$	FACIT	Cartilage, cornea, vitreous humor.
X	$[\alpha1(X)]_3$	Networked	Cartilage.
XI	$[\alpha1(XI)][\alpha2(XI)][\alpha3(XI)]$	Fibrillar	Cartilage, intervertebral disc.
XII	$[\alpha1(XII)]_3$	FACIT	Dermis, tendon.
XIII	$[\alpha1(XIII)]_3$	MACIT	Endothelium, dermis, eye, heart. α-chains fold from *N* to *C*, classical fibrillar collagens fold from *C* to *N*.
XIV	$[\alpha1(XIV)]_3$	FACIT	Widespread: bone, dermis, cartilage.
XV	$[\alpha1(XV)]_3$	Multiplexin	Found in a wide range of tissues. Down-regulated in some cancers.
XVI	$[\alpha1(XVI)]_3$	FACIT	Dermis, kidney. Highly stabilised by disulfide bonding motifs.
XVII	$[\alpha1(XVII)]_3$	MACIT	Hemidesmosomes in epithelia.
XVIII	$[\alpha1(XVIII)]_3$	Multiplexin	Widespread in basement membranes. Similar to collagen XV.
XIX	$[\alpha1(XIX)]_3$	FACIT	Basement membranes.
XX		FACIT	Cornea.
XXI		FACIT	Stomach, kidney.
XXII		FACIT	Tissue junctions.
XXIII	$[\alpha1(XXIII)]_3$	MACIT	Heart, retina. Binds $\alpha_2\beta_1$ integrins.
XXIV		Fibrillar	Bone, cornea.
XXV		MACIT	Brain, heart, testis.
XXVI		FACIT	Testis, ovary.
XXVII		Fibrillar	Cartilage. Forms thin fibres. High lysine content.
XXVIII	$[\alpha1(XXVIII)]_3$	None Assigned	Dermis, sciatic nerve. Localised to non-myelinated regions in the sciatic nerve.

Table 1 - Reported collagen types. Data collated from[21,22,24,27,38-40,44-50]. FACIT - Fibril-associated collagens with interrupted triple helicies. MACIT - Membrane-associated collagens with interrupted triple helices. Multiplexin - Multiple triple-helix domains and interruptions.

Multiple Triple-Helix Domains and Interruptions (Multiplexin) Collagens - Types XV and XVIII

The multiplexin collagens, similar to the FACITs, contain a large number of triple helix domains that are disrupted by non-collagenous domains. The two Multiplexin collagens, types XV and XVIII, are highly similar in seven of their collagen domains and have a highly conserved *C*-terminal non-collagenous domain, which distinguishes the Multiplexins from the other collagen classes[8]. The Multiplexin collagens are associated with the basement membranes, with tissue-dependent differences in the abundance of the two Multiplexin collagen forms[8,24].

Collagen Biosynthesis

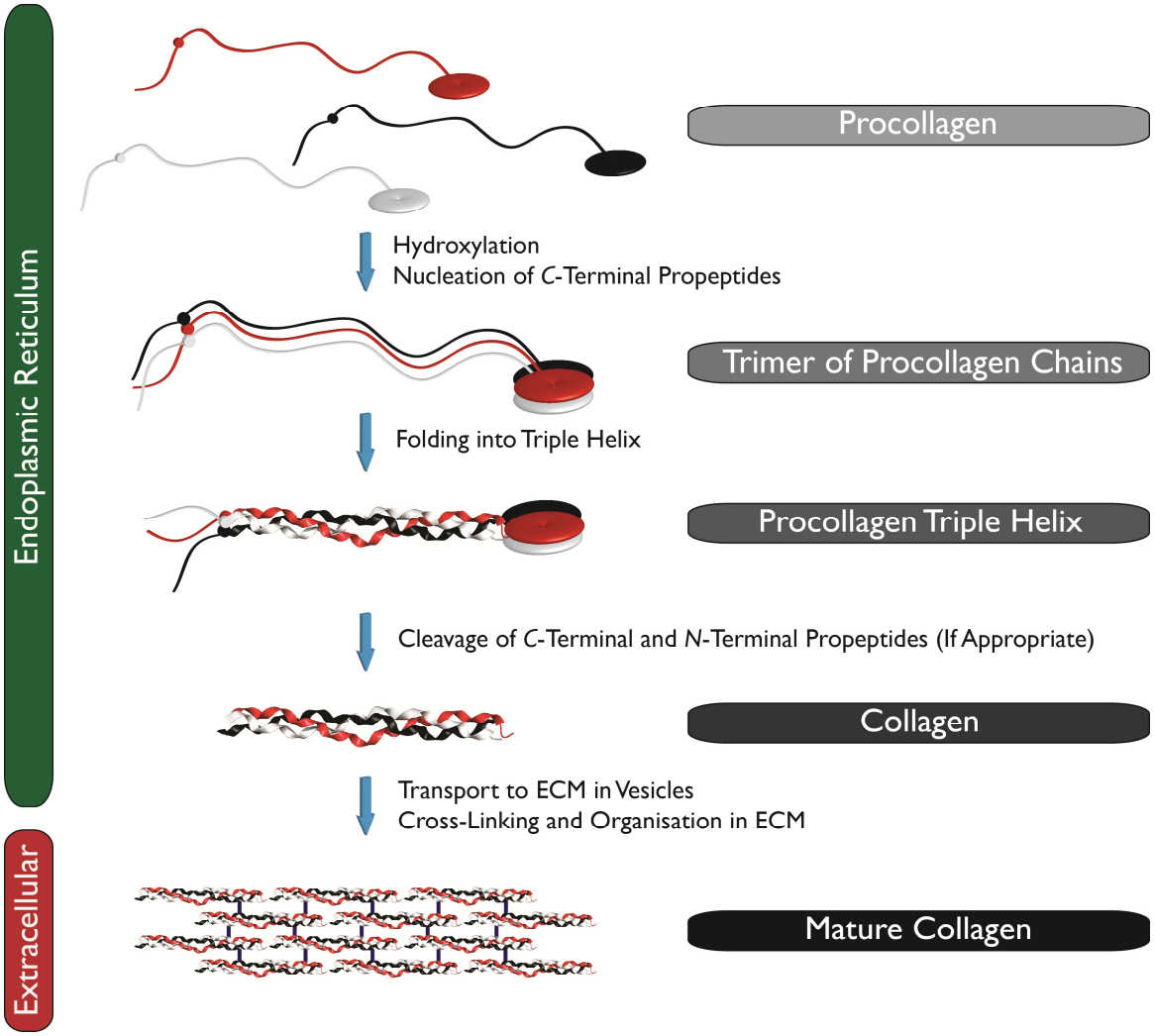


Figure 4 - Overview of collagen biosynthesis. The example shows the production of a collagen fibril. *N*-terminal and *C*-terminal propeptides are cleaved during the formation of fibrillar collagens, but can play an important role in the formation of networked collagens.

Procollagen, which at first is the simple amino acid chain product of protein expression, is formed in the endoplasmic reticulum during translation. Procollagen is post-translationally modified in the endoplasmic reticulum, primarily through the formation of hydroxylated proline and lysine residues. Procollagen chains trimerise, which allows for formation of the collagen triple helix[22]. Correctly folded collagen proteins are secreted into the ECM (*N*-terminal and *C*-terminal propeptides may be cleaved depending on collagen type).

Once secreted, collagen proteins are subsequently arranged and organised into biological scaffolds. This process includes further post-translational modification, such as the formation of allysine (*i.e.* conversion of the lysyl N^{ε} to an aldehyde[22,51]). Collagen proteins are often cross-linked to other collagen units, particularly during fibril formation, and organised through interactions with receptor proteins[10,27]. This extracellular network of post-translational modification, receptor interactions and crosslinking allows for the formation of mature ECM through "self-assembly"[5,10,22] (Figure 4).

Collagen Post-Translational Modifications

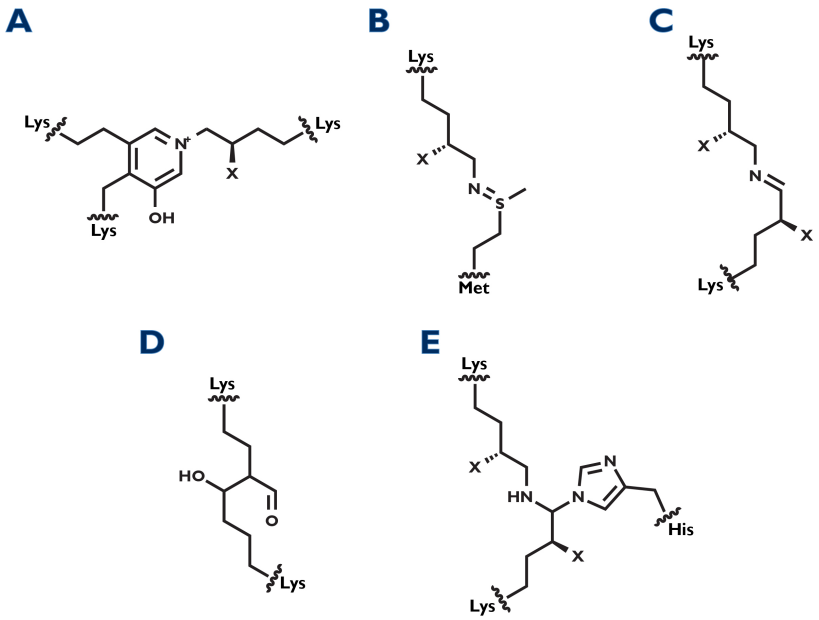


Figure 5 - Examples of cross-links found in mature collagen. Procollagen residues are indicated and are often post-translationally modified prior to crosslink formation. X = H or OH throughout. A - A pyridinoline cross link[52]. B - A sulfilimine cross link, found in collagen IV[36]. C - Lysine adduct formed with allysine[53]. D - Aldol product formed by two allysine side chains[53]. E - Result of histidine addition to the cross link presented in C[23,53].

During collagen biosynthesis, a range of post-translational modifications (PTMs) are made to procollagen, which are key in allowing collagen to form and function correctly. Collagen PTMs occur both intra- and extra-cellularly[22]. Hydroxylation is the most common collagen PTM and occurs in the endoplasmic reticulum; prolyl and lysyl residues in procollagen are enzymatically converted to (*2S,4R*)-hydroxyproline, (*2S,3S*)-hydroxyproline and (*2S,5R*)-hydroxylysine residues[24].

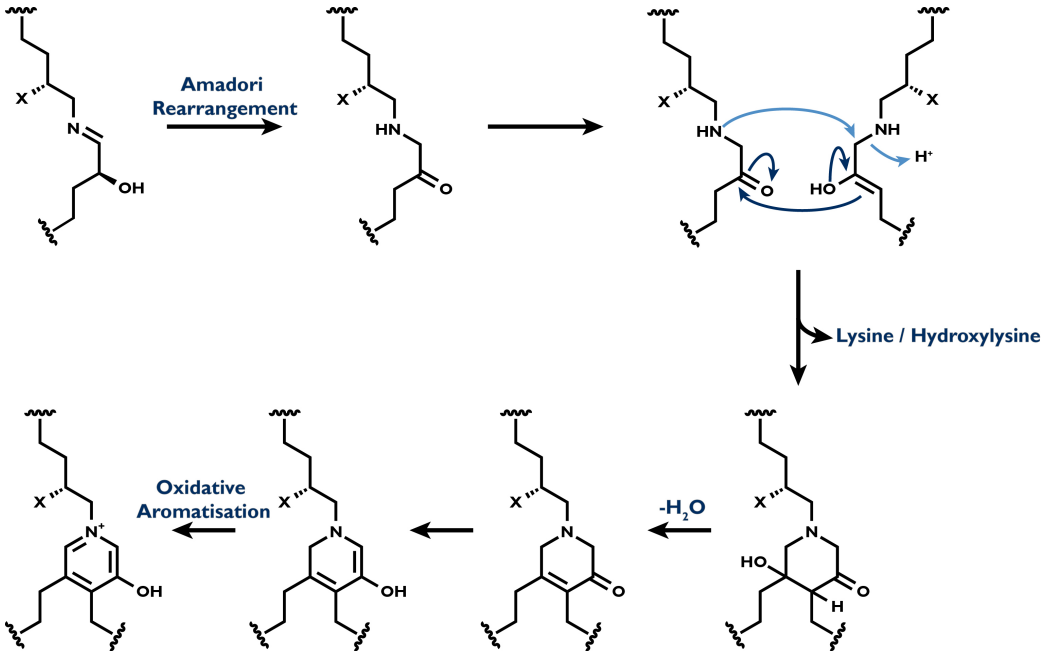


Figure 6 - Formation of pyridinoline cross links from lysine and hydroxylysine residues[52]. X = H or OH. Adducts formed between lysyl residues and allysine[53] rearrange to the corresponding ketone. Aldol condensation (dark blue arrows) and nucleophilic attack (light blue arrows) lead to the formation of a six-membered ring that undergoes subsequent dehydration, enolisation and oxidation to result in the aromatic pyrodinoline cross link.

Extra-cellular processing of procollagen is often centred around the formation of cross-links between proteins. One such example are the family of lysyl oxidase (LOX) enzymes, which convert the lysine N^{ϵ} to an aldehyde (*i.e.* conversion of lysyl residues to allysine)[51]. This in turn facilitates the formation of a variety of cross links[53] (examples of some collagen cross links are presented in Figure 5). The identity and abundance of different cross links varies according to tissue, not collagen type, indicative that extracellular processing is key in collagen cross-linking[23,52,53]. In a number of instances, (*2S,5R*)-hydroxylysine is required for cross link

formation; one such example is the formation of pyridinoline cross links (Figure 6), which are highly abundant in bone[52] and cartilage[23].

(2*S*,4*R*)-Hydroxyproline and Triple Helix Stability

The formation of (*2S,4R*)-hydroxyproline is one of the most significant post-translational modifications in terms of abundance; approximately 40% of residues at the Yaa position of collagen Gly-Xaa-Yaa repeats are (*2S,4R*)-hydroxyproline[26], making (*2S,4R*)-hydroxyproline the most common PTM in collagen and in turn, the most abundant PTM in animals. Absence of (*2S,4R*)-hydroxyproline formation is embryonically lethal in mice[54], whilst reduced levels of (*2S,4R*)-hydroxyproline formation in humans is implicated in scurvy[55,56].

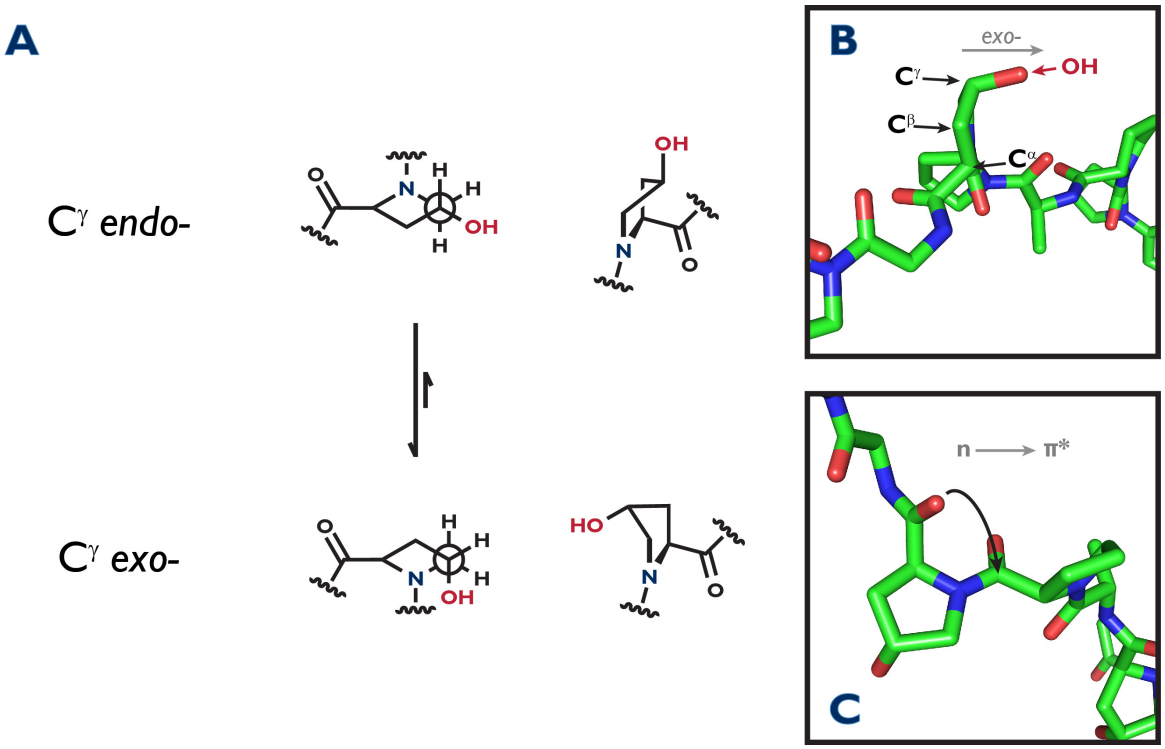


Figure 7 - Following hydoxylation of proline to form (*2S,4R*)-hydroxyproline, the *C$^{\gamma}$ exo-* ring pucker, where the nitrogen (blue) and hydroxyl group (red) are gauche to one another, is favoured over the *endo-* pucker (which is the favoured arrangement in proline). This allows for an *n* to π^* back donation that stabilises the formation of a *trans-* amide bond required for triple helix formation. A - Fisher projections and structures of the *endo-* and *exo-* pucker. B - A view from a crystal structure of a synthetic collagen (PDBID: 1CAG[30]), showing the *C$^{\gamma}$ exo-* ring pucker. C - A view from a crystal structure of a synthetic collagen (PDBID: 1CAG[30]), the alignment of the adjacent carbonyls allows for an *n* to π^* bonding interaction.

The presence of (*2S,4R*)-hydroxyproline at the Yaa position of the Gly-Xaa-Yaa collagen repeat has long been known to stabilise the collagen triple helix compared to a hypothetical

non-hydroxylated equivalent (protocollagen)[29]. This stabilisation was first proposed to arise *via* the formation of a hydrogen-bonding network[57], mediated by the addition of the hydrogen bond donor/acceptor -OH group. However, more recent studies have shown that the presence of (*2S,4R*)-hydroxyproline confers triple helix stability *via* a stereoelectronic effect rather than by increased hydrogen bonding (Figure 7)[58,59].

Conversion of proline to (*2S,4R*)-hydroxyproline induces a change in the ring pucker of proline. Prior to hydroxylation, the C^{γ} *endo*- form is favoured over the C^{γ} *exo*- form; following hydroxylation the C^{γ} *exo*- ring pucker is favoured because of a gauche effect (Figure 7A, the ring pucker can be observed in Figure 7B). This alignment mediates an $n \rightarrow \pi^*$ donation (Figure 7C) between adjacent backbone carbonyl groups. The C^{γ} *exo*- ring pucked more readily allows for the *trans*- amide bond conformation that is required for triple helix formation[58,59].

This effect was confirmed by studies with *4R*-methyl[60], *4R*-fluoro[58] and *4R*-chloro[61] substituted prolines. Due to the electron withdrawing effect of fluorine and chlorine, *4R*-fluoro and *4R*-chloro substituted prolines exhibit the same *ex*o- ring pucker as (*2S,4R*)-hydroxyproline but are not hydrogen bond donors/acceptors. Following incorporation of these hydroxyproline analogues into model peptides (*i.e*. Gly-Xaa-Yaa repeats with proline at the Xaa position and the analogue in the Yaa position), near identical stabilisation was observed when compared to analogous (*2S,4R*)-hydroxyproline scaffolds. Replacement of (*2S,4R*)-hydroxyproline with (*2S,4R*)-fluoroproline stabilises the triple helix to a greater extent than the hydroxylated form since (*2S,4R*)-fluoroproline favours the *endo*- pucker to a greater extent than (*2S,4R*)-hydroxyproline, this is due to the greater electron withdrawing effect (and hence greater preference for the gauche alignment) of F compared to OH[27,58].

(*2S,3S*)-Hydroxyproline

Despite being discovered in 1962[62], the precise role of (*2S,3S*)-hydroxyproline in collagen is poorly understood, particularly from a molecular perspective and when compared to the role

that (*2S,4R*)-hydroxyproline plays in collagen[63]. This is in part due to the lower abundance of (*2S,3S*)-hydroxyproline in collagen compared to (*2S,4R*)-hydroxyproline.

(*2S,3S*)-Hydroxyproline is a crucial amino acid in mature collagen, particularly in type IV collagen[64]. (*2S,3S*)-Hydroxyproline is most frequently observed at the Xaa position of the Gly-Xaa-Yaa repeats and has been linked to the folding of the collagen triple helix[65]. The complete absence of enzymes that are responsible for catalysing the formation of (*2S,3S*)-hydroxyproline, (similar to (*2S,4R*)-hydroxyproline) are embryonically lethal in mice[64]. Reduction in the abundance of (*2S,3S*)-hydroxyproline is linked to a range of musculoskeletal disorders such as *osteogenesis imperfecta*[66] (which is more commonly referred to as brittle bone disease), particularly *osteogenesis imperfecta* types VII, VIII and IX[67].

(2*S*,5*R*)-Hydroxylysine

In addition to prolyl hydroxylation, lysine side chains in procollagen are enzymatically hydroxylated in the endoplasmic reticulum to (*2S,5R*)-hydroxylysine[68]. This hydroxylation event can allow for subsequent glycosylation and can also play a direct role in crosslinking; exemplified by the formation of sulfilimines in basal membranes[69] and pyridinoline cross links in bone and cartilage[52] (Figure 6).

A reduction in levels of (*2S,5R*)-hydroxylysine has been linked to a number of disease states, such as Ehlers-Danlos syndrome type VI[70-72] and Bruck syndrome[73], whilst genetic knockout of (*2S,5R*)-hydroxylysine formation in collagen IV has been found to be embryonically lethal in mice[24].

The 2OG Oxygenases

Hydroxylation of procollagen proline and lysine residues are catalysed by the collagen prolyl-4-hydroxylases (C-P4Hs), collagen prolyl-3-hydroxylases (P3Hs)[74] and lysyl hydroxylases (Procollagen-Lysine, 2-Oxoglutarate 5-Dioxygenases, PLODs)[75]. These three enzyme families are all members of the 2-oxoglutarate dependent oxygenases (2OG oxygenases), a superfamily of

non-haem iron enzymes. The 2OG oxygenases are not confined to *metazoa* and are able to catalyse a range of reactions including desaturation[76], ring closure[77] and halogenations[78]. Most commonly, the 2OG oxygenases catalyse the hydroxylation of an unactivated C-H bond.

Structural and Mechanistic Features of the 2OG Oxygenases

All characterised 2OG oxygenases utilise Fe^{II} as a cofactor, alongside molecular O_2 and 2-oxoglutarate (2OG, **1**), an intermediate in the tricarboxylic acid (TCA) cycle, as co-substrates. The cofactor and co-substrate requirements for the 2OG oxygenases were first established for the C-P4H enzymes[79] and are conserved across the enzyme family. Structural studies have shown that the 2OG oxygenases have a conserved active site fold, the double stranded β-helix (DSBH) or jelly-roll fold[80], first observed in the related isopenicillin *N*-synthase[77] (Figure 8). A variant of the DSBH, the Jumanji C (JmjC) domain is found in a subset of the 2OG oxygenases[80].

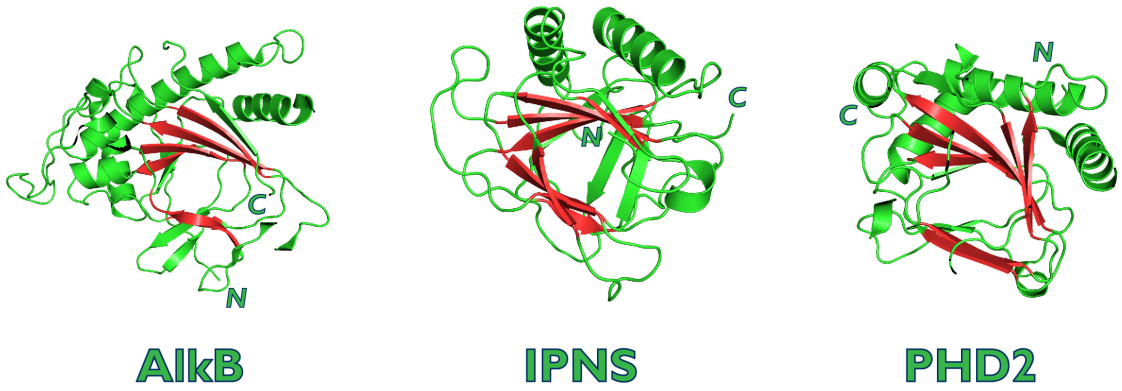


Figure 8 - The 2OG oxygenases have a double stranded beta helix. Views from crystal structures of AlkB (PDBID: 3I3Q[81]), IPNS (PDBID: 2VAU[77]) and PHD2 (PDBID: 3HQR[82]) with the DSBH highlighted in red.

For nearly all 2OG oxygenases, Fe^{II} is co-ordinated by a facial triad of two histinyl residues and a further residue, which is most commonly either an aspartic acid or glutamic acid residue[80], within the active site. These residues are arranged in a HXD/E…H motif (Figure 9). The other three co-ordination sites are available for co-ordination by 2OG and O_2. In the absence of co-substrates, the co-ordination sites are occupied by H_2O from the bulk[80]. The distal carboxylate of 2OG interacts with a positively charged region that forms a 2OG binding pocket; this is often

via an electrostatic interaction with either a lysyl or argininyl residue, which has been observed crystallographically (an example is shown in Figure 9)[80,83].

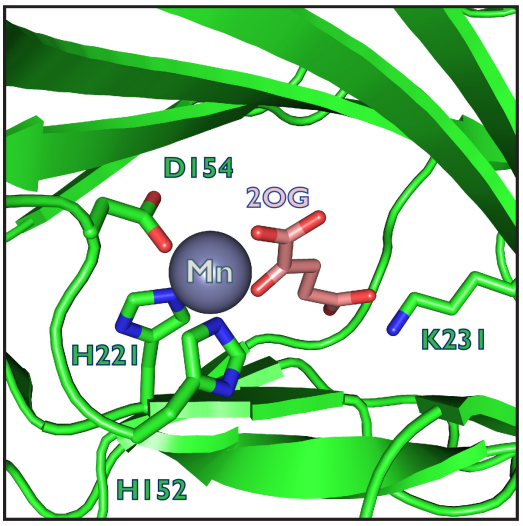
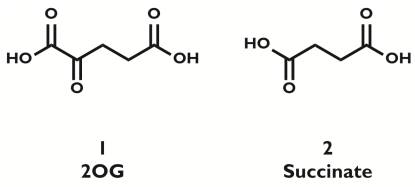


Figure 9 - A view from a crystal scructure of a 2OG oxygenase from *Paramecium bursaria* Chlorella Virus I in complex with Mn and 2OG (PDBID: 5C5T[84]). The active site is shown. The metal binding triad of H152, D154 and H221 co-ordinates the metal (for crystallography, Fe was substituted with Mn). The distal carboxylate of 2OG forms an electrostatic interaction with K231 in this example (which interacts with 2OG).

During catalysis, sequential binding of 2OG (**1**), substrate and O_2 precedes a reductive decarboxylation that oxidises Fe^{II} to an Fe^{IV}-oxo intermediate (Figure 11). This process leads to production of CO_2 and succinate (**2**). The Fe^{IV}-oxo species then abstracts an appropriately aligned hydrogen atom, resulting in a radical centred on the substrate. In the case of hydroxylation reactions, the hydroxylated product is then formed in a radical-rebound type process[78,85].



1
2OG

2
Succinate

Figure 10 - Structures of 2OG (**1**) and succinate (**2**). Both are shown as the acid form.

Turnover of 2OG (**1**) to succinate (**2**) may also occur in an 'uncoupled' fashion, *i.e.* oxidative decarboxylation of 2OG occurs without product formation. This process is referred to as 'uncoupled turnover', and is a process that also leads to the production of reactive oxygen species such as peroxide[78]. Monitoring of 2OG (**1**) and succinate (**2**) resonances by nuclear

magnetic resonance (NMR) spectroscopy has shown that uncoupled turnover can increase in the presence of substrate[86].

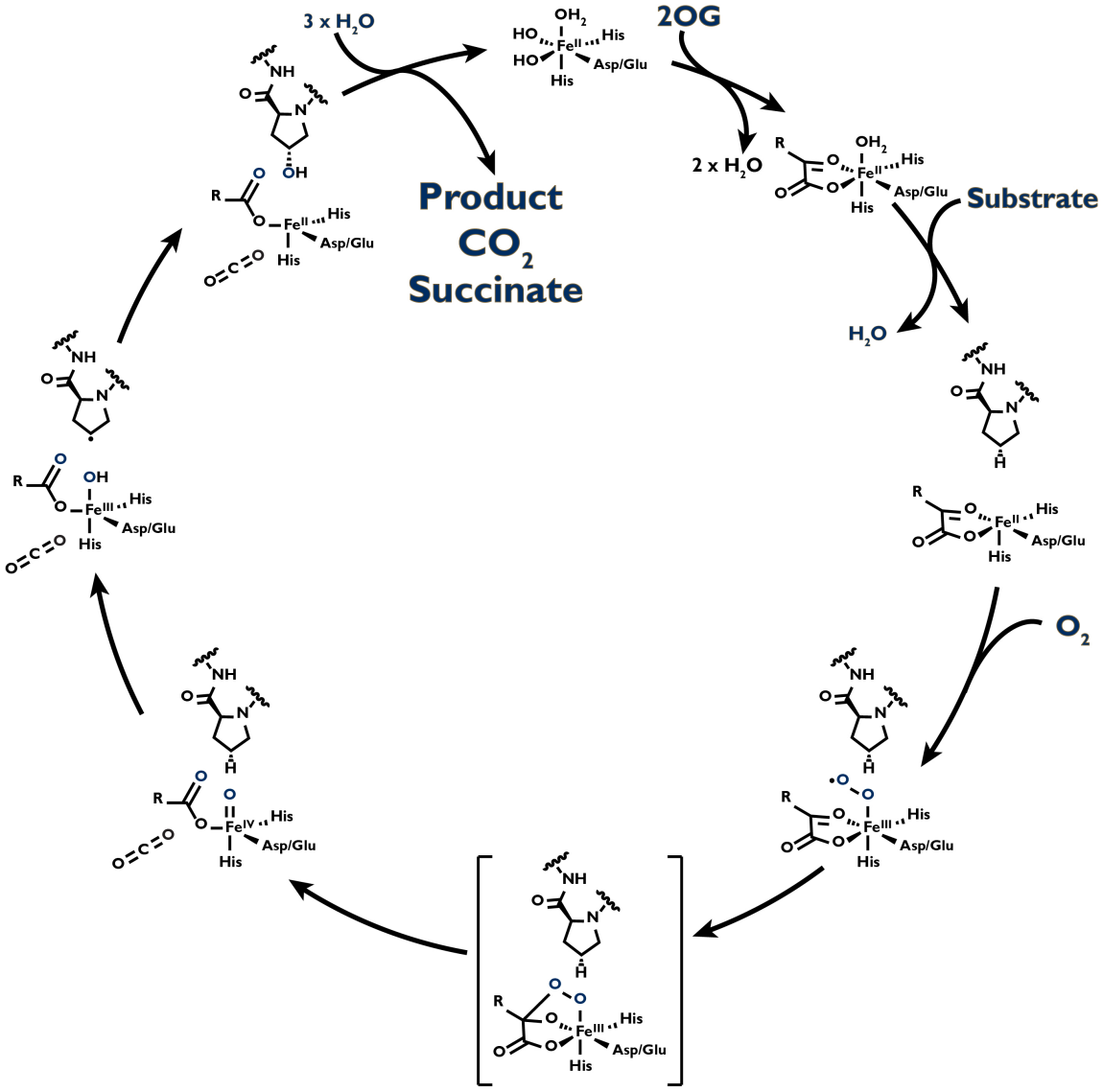
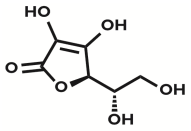


Figure 11 - 2OG oxygenase catalytic cycle. The example given shows the conversion of a proline residue to a (*2S,4R*)-hydroxyproline residue.

The release of CO_2 during the catalytic cycle has been used to assay the activity of 2OG oxygenases[87]. Through the use of radiolabelled 2OG, the liberation of radiolabelled CO_2 is monitored and related to 2OG oxygenase activity. This indirect approach to monitoring 2OG oxygenase activity does not always accurately take uncoupled turnover into account, and an increase in uncoupled turnover is impossible to deconvolute from coupled turnover in radiolabelled 2OG assays.

The 2OG Oxygenases and Ascorbate

The role of ascorbate (Vitamin C, **3**, shown as the acid form in Figure 12) on the catalytic cycle of the 2OG oxygenases is currently unresolved. Ascorbate has long been linked with the development of scurvy[56]. A reduction of C-P4H and PLOD activity in the absence of ascorbate leads to a reduction in the abundance of (*2S,4R*)-hydroxyproline and (*2S,5R*)-hydroxylysine in collagen; a reduction in the abundance of hydroxylated residues decreases collagen stability and brings about the onset of disease[88].



3
L-Ascorbic Acid

Figure 12 - Structure of L-ascorbic acid, **3**.

Ascorbate is not required for catalysis to take place; however, rates of 2OG turnover, coupled and uncoupled (*i.e.* turnover of 2OG leading product formation and oxidative decarboxylation of 2OG without product formation), are increased by most 2OG oxygenases in the presence of ascorbate. The precise role of ascorbate during catalysis has not been conclusively determined, it appears that part of the role that ascorbate plays is as a reducing agent, maintaining the Fe^{II} oxidation state (particularly after uncoupled 2OG turnover). However, studies with PLOD1 (and other 2OG oxygenases) using alternative reducing agents shows that other reducing agents increase the rate of 2OG turnover but not to the same extent as ascorbate[89], indicating a more complex role for ascorbate.

The Collagen Hydroxylases

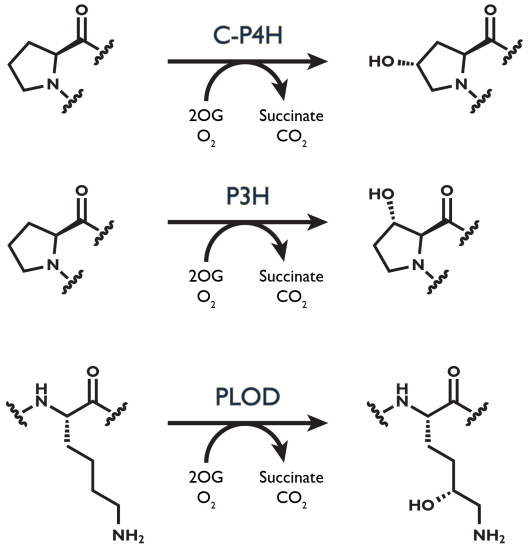


Figure 13 - Post-translational hydroxylation of procollagen is catalysed by three groups of 2OG oxygenases. 2OG (**1**) and O_2 are used as co-substrates during catalysis to give succinate (**2**) and CO_2.

C-P4Hs

C-P4H was the first 2OG-dependent oxygenase to be identified[79,90,91]. The C-P4Hs, in animals, are 240 kDa $\alpha_2\beta_2$ tetramers which catalyse the hydroxylation of prolyl residues to (*2S,4R*)-hydroxyproline at the Yaa position of the collagen Gly-Xaa-Yaa triplet. The α-subunit contains the catalytic Fe^{II}/2OG oxygenase domain, whilst the β-subunit is identical to a protein disulfide isomerase (PDI)[92].

In humans, and most other animals, there are three α-subunit isoforms, of which the α(I) form is the most abundant at the protein level. There is no evidence for the formation of 'mixed' tetramers[93] and thus there are only three tetramer forms, and as a consequence, the $\alpha(I)_2\beta_2$ tetramer is the most abundant C-P4H form[94]. The relative abundance of C-P4H tetramer forms varies between tissues in the body; for example, the $\alpha(II)_2\beta_2$ tetramer predominates in chondrocytes and capillary endothelial cells[94]. The localisation of the $\alpha(III)_2\beta_2$ has not been fully established. In addition, the binding properties of the three C-P4H tetramers differ; the $\alpha(I)_2\beta_2$ form binds to poly-L-proline whereas the $\alpha(II)_2\beta_2$ and $\alpha(III)_2\beta_2$ forms do not[95-97].

Due to the important role that the presence of (*2S,4R*)-hydroxyproline plays in conferring stability on collagen, it is unsurprising that any abnormalities in the activity of the C-P4Hs has a

major physiological effect[98]. Complete knockout of the α(I) subunit is embryonically lethal in mice[54], whilst mutations in both the α and β-subunits have been implicated in a wide range of connective tissue disorders[98] such as Cole-Carpenter syndrome, which arises due to mutations in the CPHB gene[99].

Despite being of significant physiological importance, the C-P4Hs have not been biochemically characterised in detail and no high-resolution structure of the complete tetramer is available. Even though smaller sections of the complex have been structurally characterised, the *N*-terminal region of the α(I) subunit has been shown to form a homodimer[100] (discussed in more detail in Chapter 2), there is no structural insight into the enzyme active site. Biochemical studies have analysed the turnover of radiolabelled 2OG[101,102], such analyses have shown that the C-P4H enzymes have a high requirement for ascorbate (**3**)[87,103] and that the C-P4H enzymes catalyse uncoupled turnover to a much greater extent than for many other 2OG oxygenases[103]. Due to the high levels of uncoupled 2OG turnover catalysed by the C-P4Hs, analysis of 2OG turnover cannot afford reliable quantitative kinetic analyses.

P3Hs

Despite the discovery that collagen proline residues are post-translationally modified to give (*2S,3S*)-hydroxyproline in 1962[62], the P3H enzymes (as distinct entitites) were not identified until 1979[74] and the genes encoding for the P3Hs were not identified until 2004[104]. Similar to the C-P4Hs, there are also three forms of the collagen P3H (P3H1, P3H2 and P3H3). The expression profiles of P3H1 and P3H2 differ at the RNA level; P3H1 is most commonly transcribed in tissues with high proportions of fibrillar collagens, P3H2 transcription is more prevalent in tissues that are rich in collagen IV[64,103].

P3H1 forms a heterotrimeric complex with cartilage associated protein (CRTAP) and peptidyl-prolyl *cis-trans* isomerases (CypB)[66,67,105-107]; knockout of CRTAP reduces levels of

(*2S,3S*)-hydroxyproline in collagen and mice that do not express CRTAP exhibit symptoms associated with a lack of (*2S,3S*)-hydroxyproline, such as *osteogenesis imperfecta*[66].

Biochemical characterisation of the P3Hs is, like the C-P4Hs, limited in scope. No high-resolution structural information is available about any of the P3Hs, whilst studies with P3H2 have shown that presence of CRTAP does not affect the 2OG turnover of P3H2[108]. P3H3 is even more poorly understood than P3H1 and P3H2.

P3H2 has been linked with prolyl-3-hydroxylation in collagen IV and P3H2 knockdown has been shown to be embryonically lethal in mice[64]. Since lethality was not observed in CRTAP knockout studies and that CRTAP does not increase P3H2 activity *in vitro*[108], this suggests that the P3H isoforms may exist in different supramolecular states.

PLODs

Formation of (*2S,5R*)-hydroxylysine in procollagen is catalysed by the PLODs, which are also referred to as the lysyl hydroxylases (LHs). In animals, three PLODs have been identified[103], which exist as homodimers[109] of around 160 kDa. Due to the key role played by (*2S,5R*)-hydroxylysine in cross-link formation[110] and in allowing for glycosylation[103], the PLODs have been implicated in a number of disease states, including Ehlers-Danlos syndrome type VIA (which is associated with mutations in the PLOD1 gene)[70-72] and Bruck syndrome[73].

As with the other collagen hydroxylases, there is no high-resolution structural data available for the PLODs. Limited *in vitro* studies have been conducted and hydroxylase activity for all three PLOD enzymes. *In vitro* glycosyltransferase activity has also been observed for PLOD3, but not for PLOD1 or PLOD2[109], this is despite all three PLODs containing a glycosyltransferase domain. Similar to the C-P4Hs, the PLODs have shown a strong requirement for ascorbate (**3**)[89] .

Human 2OG Oxygenases

In humans, the 2OG oxygenase superfamily consists of just over 60 enzymes (Figure 14) that catalyse the oxidation of an unactivated C-H bond in a variety of contexts (further details for the

2OG oxygenases such as alternative names and their reactivity are detailed in Appendix 1). The enzymes often feature multiple domains (Figure 15) alongside the conserved DSBH fold.

In a number of cases, such as the C-P4Hs, P3Hs and PLODs, 2OG oxygenase catalysis results in the formation of the hydroxylated (often an alcohol) product; however, should this product be unstable, the hydroxylated product may undergo a further rearrangement or fragmentation. This is observed in the case of the histone lysine demethylases where the oxidation of a methyl C-H results in the formation of an unstable hemiaminal intermediate which subsequently fragments to the demethylated lysine residue and a molecule of formaldehyde[111].

The human 2OG oxygenases catalyse hydroxylation in a wide range of contexts that span across a number of key biological processes including hypoxic sensing[112], ribosome function[113,114] and epigenetic gene regulation[115]. The precise role of a number of the human 2OG oxygenases is yet to be fully resolved. In a few cases, the functional assignments have proven to be controversial, such as in the case of Jumonji domain-containing protein 6 (JMJD6). JMJD6 has been assigned as both a lysyl hydroxylase and as an arginine demethylase[116-118], with the assigned arginine demethylase activity of proving to be controversial[116].

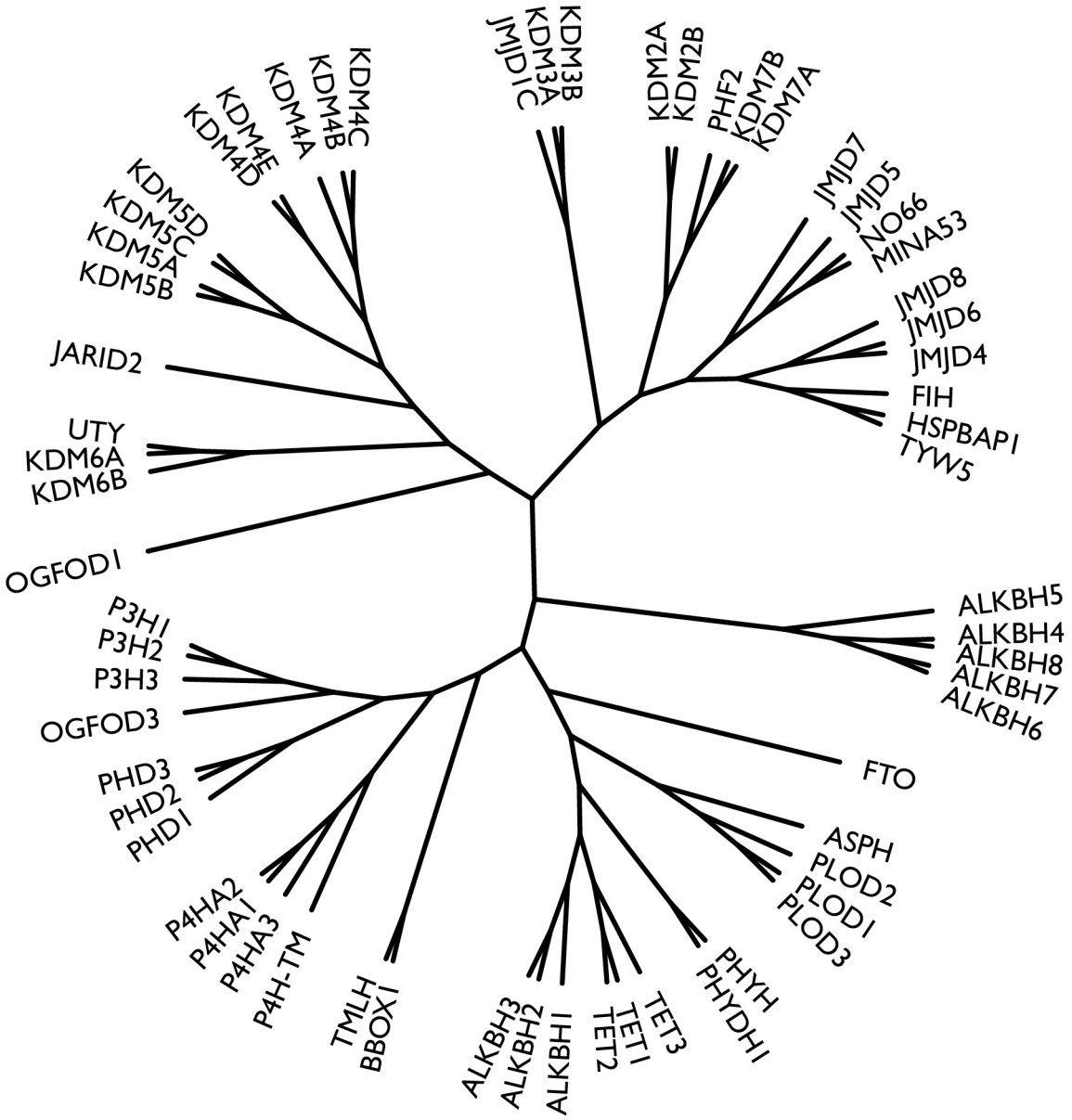


Figure 14 - The human 2OG oxygenases. Sequences of the catalytic domain were aligned using Clustal Omega[119] before the tree was generated using FigTree v1.4.3[120].

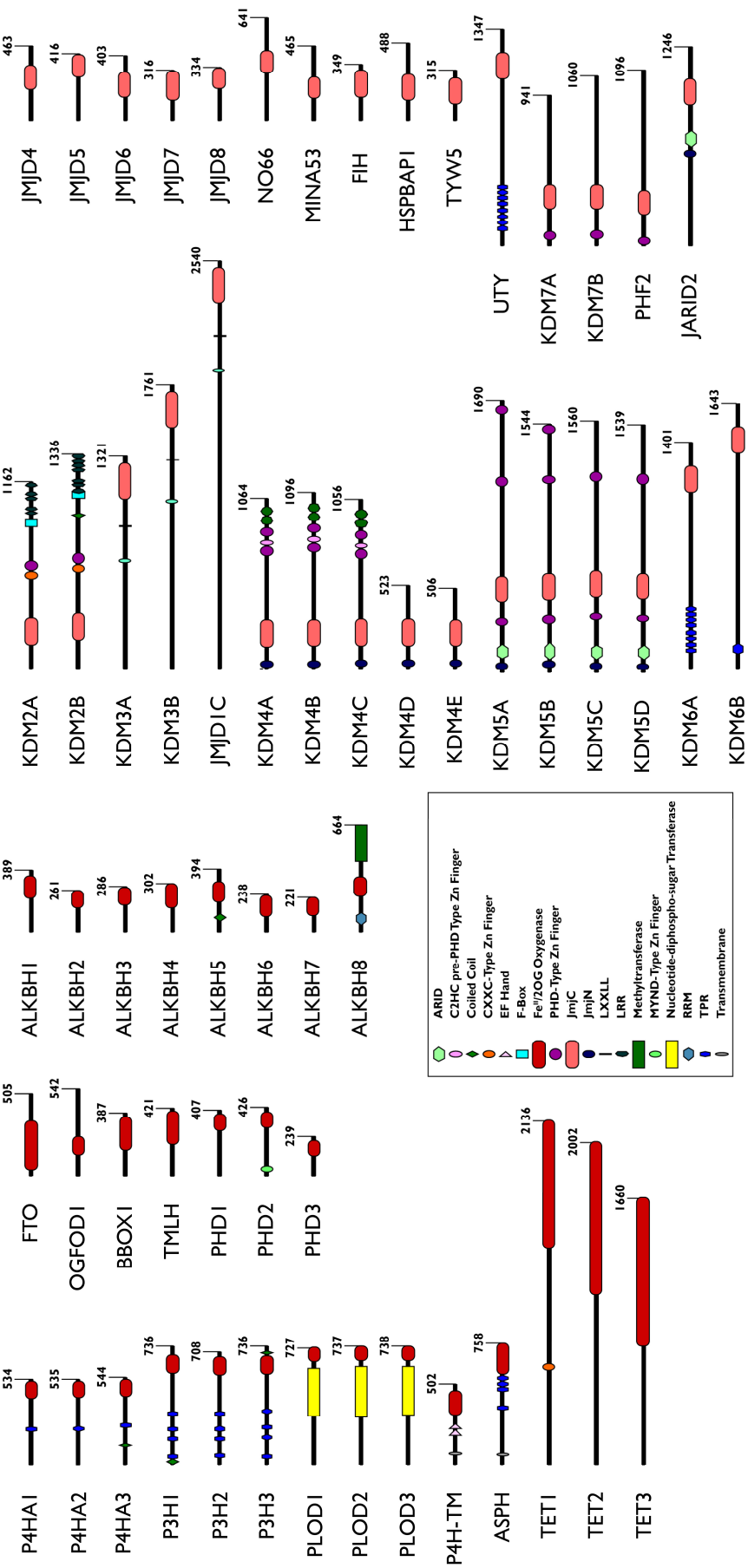


Figure 15 - Domain architecture of the human 2OG oxygenases.

Prolyl and Lysyl Hydroxylation is Not Limited to Collagen

Alongside the hydroxylation of prolyl- and lysyl- residues in procollagen, 2OG-dependent oxygenase catalysed prolyl- and lysyl- hydroxylation is observed in different contexts. For example, prolyl-hydroxylation is implicated in the hypoxic response[112] and is a PTM that is found in the ribosome[121].

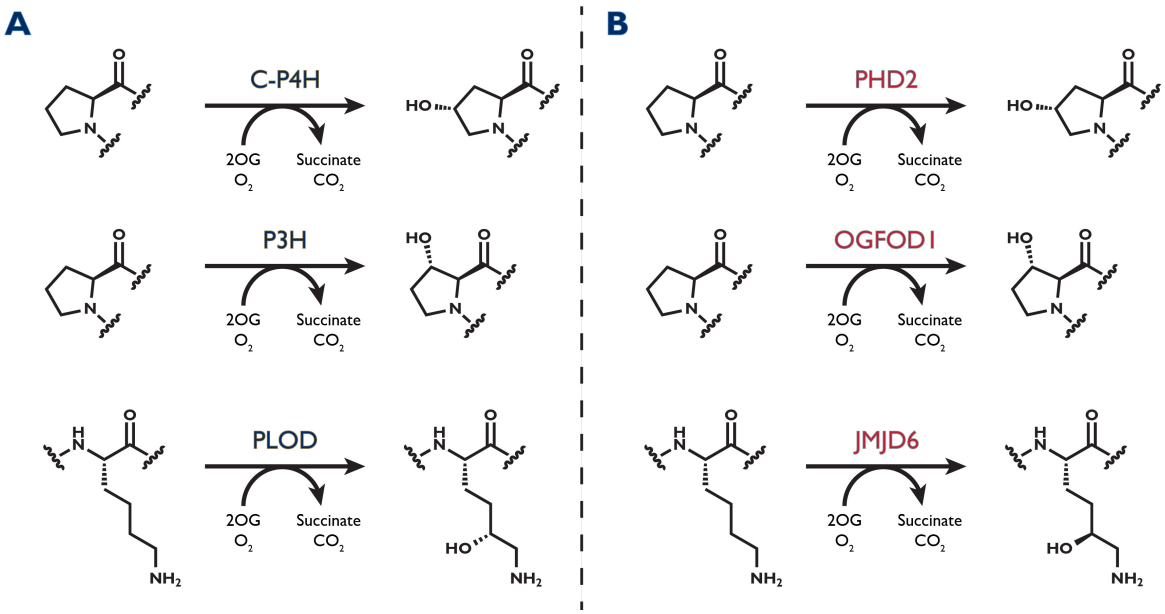


Figure 16 - Prolyl- and lysyl- hydroxylation is not limited to collagen. Comparison of reactions of the collagen 2OG oxygenases (A) and PHD2, OGFOD1 and JMJD6 (B). Fe^{II} is a cofactor for all reactions and is not indicated.

HIF Prolyl Hydroxylases and the Hypoxic Response

The formation of (*2S,4R*)-hydroxyproline in the *N*-terminal and *C*-terminal oxygen degradation domains (NODD and CODD) of the α-subunit of the hypoxia inducible factor 1α and 2α (HIF-1α and HIF-2α). HIF-1α and HIF-2α are transcription factors that play an important role in the hypoxic response[122].

Normoxia

In the presence of oxygen, the hydroxylation of HIF-1α prolyl residues at Pro402 and Pro564 is mediated by the HIF prolyl hydroxylases (PHD1, PHD2 and PHD3)[112,123], of which PHD2 is the most significant[124]. Following PHD-catalysed formation of (*2S,4R*)-hydroxyproline at Pro402, in the *N*-terminal oxygen-sensing degradation domain (NODD), or Pro564, in the *C*-terminal

oxygen-sensing degradation domain (CODD)[123], HIF-1α binds to the von Hippel–Lindau tumour suppressor protein (VHL). VHL forms part of a ubiquitination complex which catalyses the poly-ubiquitination of HIF-α[112]. Following poly-ubiquitination, HIF-α is proteasomally degraded, preventing HIF-α translocation to the nucleus, thereby preventing the transcription of HIF target genes[123,124].

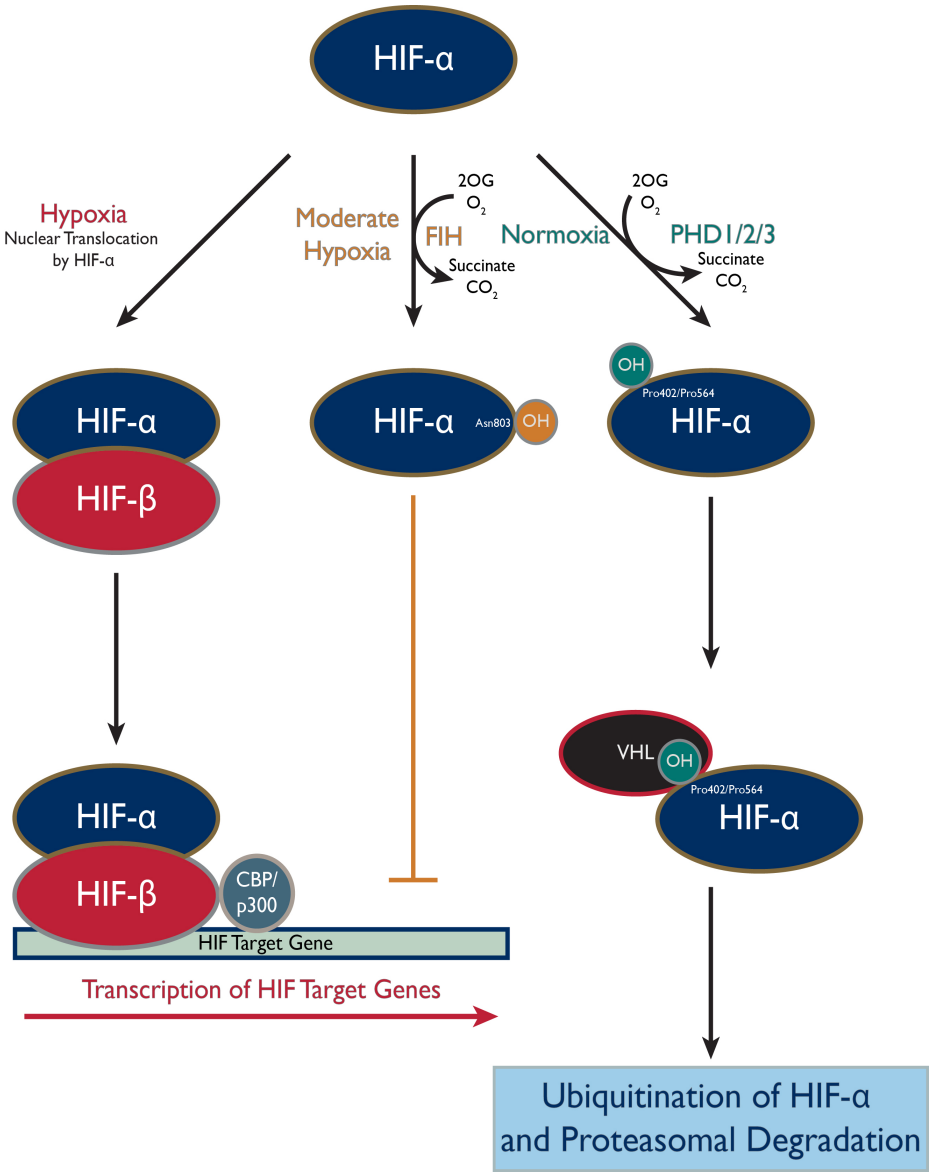


Figure 17 - Schematic representation of the hypoxic response. Residue numbers are given for HIF-1α. In the presence of oxygen, PHD-catalysed hydroxylation and subsequent binding by VHL leads to poly-ubiquitination of HIF-α-subunits and subsequent proteasomal degradation. In moderate hypoxia, PHD-catalysed hydroxylation of HIF-α is reduced whilst FIH-catalysed hydroxylation of asparagine (N803 in HIF-1α) still occurs, preventing transcription as N803-hydroxylated HIF-α is not able to bind to the p300/CBP co-activators. In hypoxia, the PHD enzymes and FIH are unable to catalyse hydroxylation of HIF-α; HIF-α translocates to the nucleus and dimerises with HIF-β, initiating transcription of HIF target genes with p300/CBP co-activators.

As with prolyl-hydroxylation in collagen, the formation of (*2S,4R*)-hydroxyproline induces a shift in the equilibrium from the C^{γ} *endo*- ring pucker towards the C^{γ} *exo*- form[125]. The C^{γ} *exo*-pucker, along with a hydrogen bonding motif enabled by the presence of the newly installed hydroxyl group, allows HIF-α to bind to VHL (Figure 18)[82,126].

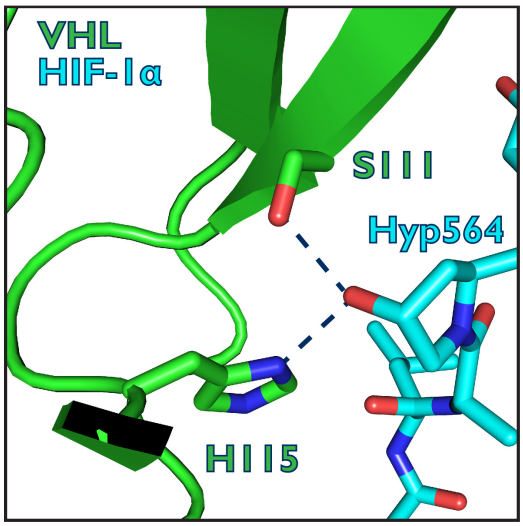


Figure 18 - A view from a crystal structure of VHL bound to a HIF-1α peptide (PDBID: 1LH8[126]). The (*2S,4R*)-hydroxyproline residue, which under physiological conditions and normoxia is installed by the HIF prolyl hydroxylases, at position 564 of HIF-1α forms hydrogen bonds with S111 and H115 of VHL[126].

In addition to the HIF/PHD2 mechanism, HIF-1α may also be hydroxylated at Asn803, catalysed by the asparagnyl hydroxylase factor inhibiting HIF (FIH). Hydroxylation of Asn 803 blocks binding of the p300/CBP co-activators and inhibits transcription of the multiple HIF target genes[123].

Hypoxia

PHD2 has a high level of sensitivity to variations in oxygen availability[127]. The level PHD2-mediated hydroxylation of HIF-1α is thus reflective of cellular O_2 availability. Therefore, under hypoxic conditions, HIF-1α is not degraded by the proteasome and is able to translocate to the nucleus. Once in the nucleus, HIF-1α complexes with HIF-β and, in combination with the p300/CBP co-activators, initiates transcription of HIF target genes[123].

In mild hypoxia, PHD-catalysed prolyl hydroxylation is reduced more rapidly than FIH activity, allowing for a ranged response to decreases in O_2 availability[128].

There are a wide range of HIF target genes implicated in a number of biological processes, such as angiogenesis, matrix deposition and changes in cell metabolism[124]. The most famous HIF target gene is EPO[129], which encodes erythropotein, a protein that initiates the production of red blood cells. Remodelling of the ECM and collagen deposition have been associated with hypoxia[130], and genes encoding for two of the C-P4H α-subunits (P4HA1 and P4HA2), as well as the PLOD2 gene have been identified as HIF target genes[131].

Prolyl-3-Hydroxylation and OGFOD1

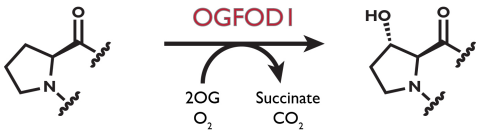


Figure 19 - OGFOD1 catalyses the formation of (*2S,3S*)-hydroxyproline. OGFOD1 requires Fe^{II} as a cofactor for activity.

2-oxoglutarate and iron-dependent oxygenase domain-containing protein 1 (OGFOD1) is a protein which catalyses the formation of (*2S,3S*)-hydroxyproline at Pro64 of ribosomal protein Rps23, which is found in the 40S ribosomal subunit[114,121]. OGFOD1-mediated formation of (*2S,3S*)-hydroxyproline enhances translational efficiency; the modification is conserved across eukaryotes[114]. Knock down of OGFOD1 leads to translational arrest and stress granule formation[121], indicative that the formation of (*2S,3S*)-hydroxyproline in Rps23 is an important process in healthy cells. The yeast homologue of OGFOD1, Ofd1, has been found to catalyse a second hydroxylation and form 3,4-dihydroxyproline (this activity is not observed in OGFOD1)[114].

Lysyl Hydroxylases

Alongside from the PLOD-catalysed formation of (*2S,5R*)-hydroxylysine[68], 2OG-oxygenases have been found to catalyse the hydroxylation of lysine to form (*2S,5S*)-hydroxylysine[116] and 4-hydroxylysine (absolute stereochemistry has not been assigned to date)[113].

JMJD6 catalyses the formation of (*2S,5S*)-hydroxylysine[116] in the U2 small nuclear ribonucleoprotein auxiliary factor 65-kilodalton subunit (U2AF65)[117]. JMJD6 has also, somewhat controversially, been assigned as an arginine demethylase[118].

2OG oxygenase catalysis has been found to catalyse the hydroxylation of the lysine side chain at other positions. ε-Trimethyllysine hydroxylase (TMLH), a 2OG oxygenase involved in the carnitine biosynthesis pathway[132,133], which catalyses the conversion of N^{ϵ}-trimethyllysine to (*2S,3S*)-3-hydroxy-N^{ϵ}-trimethyllysine[134].

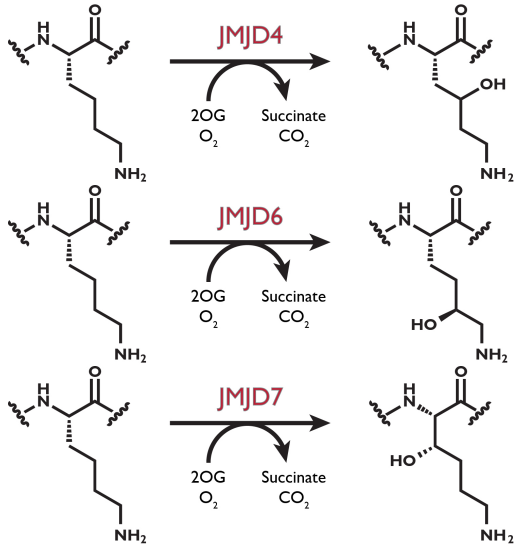
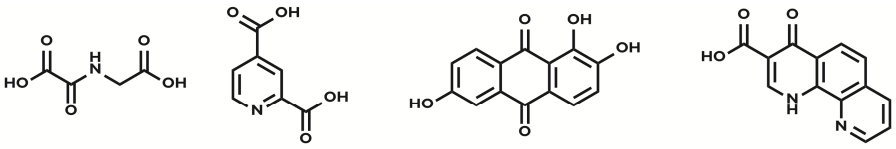


Figure 20 - Reactions catalysed by the 2OG dependent lysyl hydroxylases JMJD4[113], JMJD6[116] and JMJD7[135]. The stereochemistry of JMJD4-catalysed hydroxylation has not been resolved.

The formation of alternative hydroxylysine regioisomers has also been observed in proteins. Jumonji domain-containing protein 7 (JMJD7) has recently been found to catalyse the post-translational transformation of lysine to (*2S,3S*)-hydroxylysine[135] and Jumonji domain-containing protein 4 (JMJD4) catalyses the formation of (*2S*)-4-hydroxylysine in eukaryotic release factor 1 (eRF1). Lysyl-4-hydroxylation of eRF1 enables more efficient stop codon recognition (the stereochemistry has not been resolved)[113,135].

2OG Oxygenases as Therapeutic Targets

The 2OG oxygenases have attracted interest as potential therapeutic targets[136]. For example, the collagen processing C-P4H, P3H and PLOD families of enzymes appear to offer interesting potential for the treatment of fibrotic disease. Initially, this interest was centred around the C-P4Hs[16,102,137,138]. Early inhibitors developed for the C-P4Hs were discontinued, most probably due to off-target inhibition on, then unknown, other 2OG oxygenases (most likely PHD2)[55,136].



4 NOG | 5 2,4-PDCA | 6 2,7,8-Trihydroxyanthraquinone | 7 FG-0041

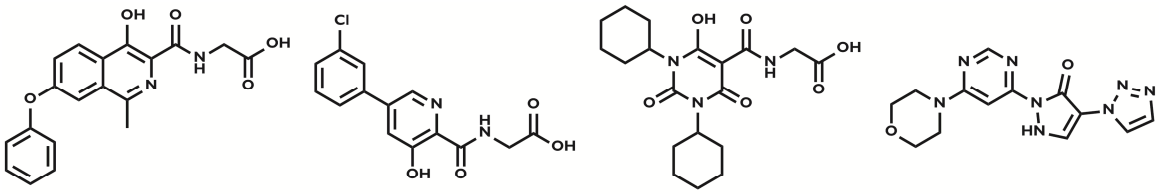
Figure 21 - Structures of broad-spectrum inhibitors of 2OG oxygenases, NOG (**4**) and 2,4-PDCA (**5**), alongside compounds developed as inhibitors of C-P4H, 2,7,8-trihydroxyanthraquinone (**6**)[138] and FG-0041 (**7**)[137].

These attempts at C-P4H inhibition have resulted in the identification of broad 2OG oxygenase inhibitors, such as *N*-oxalyl glycine (NOG, **4**) and 2,4-pydridine dicarboxylic acid (2,4-PDCA, **5**)[136], alongside a series of anthraquinones[102], such as 2,7,8-trihydroxyanthraquinone[138] (**6**), and compounds such as FG-0041 (**7**)[137]. To date, no compounds have successfully passed through clinical trials and attention from industry has turned away from the C-P4Hs as a clinical target.

Despite the interest in C-P4Hs, the P3Hs and PLODs have not been investigated as potential therapeutic targets to date. In part, this is due to the lack of recombinant expression systems and high-throughput assays.

HIF Prolyl Hydroxylases

Manipulation of the hypoxic response, through the upregulation or downregulation of the HIF transcription factors, provides interesting therapeutic potential; the HIF prolyl hydroxylases, in particular PHD2, have been the subject of studies by the pharmaceutical industry[136,139-142]. Inhibition of PHD2 allows for its substrates, HIF-1α and HIF-2α, to translocate to the nucleus and activate the HIF target genes (*i.e.* inducing a 'false' hypoxia). EPO is one such target gene; hypoxic upregulation of EPO transcription and subsequent production of erythropoietin promotes erythrogenesis and therefore PHD2 inhibition has been identified as a potential therapeutic target for the treatment of anaemia[55].



8
FG-4592

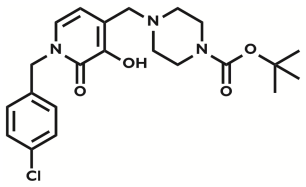
9
Vadadustat

10
GSK1278863

11
Molidustat

Figure 22 - Structures of inhibitors of the HIF prolyl hydroxylase PHD2. These compounds are currently in clinical trials for the treatment of anaemia.

There are currently clinical candidates in phase III trials from FibroGen (FG-4592, **8**)[143], Akebia (Vadadustat, **9**)[144,145]and GSK (GSK1278863, **10**)[146] whilst a candidate from Bayer (Molidstat, **11**)[147] is currently in phase II (the structures of these compounds are detailed in Figure 22). The effect of these compounds, however, on extracellular matrix production, in particular their effect on the C-P4Hs are currently unknown.



12
AKB-4924

Figure 23 - Structure of AKB-4924 (**12**). AKB-4924 is a PHD2 inhibitor currently in clinical trials for the treatment of inflammatory bowel disease.

PHD2 inhibition has also been explored as a potential therapeutic target in inflammatory disease[148]; a clinical candidate for the treatment inflammatory bowel disease, AKB-4924 (**12**, Aerpio Pharmaceuticals)[111], is currently in phase IIA clinical trials.

Other 2OG Oxygenases

Therapeutic interest is not limited to the human 2OG-dependent prolyl hydroxylases. For example, inhibiting carnitine biosynthesis through inhibition of γ-butyrobetaine hydroxylase (BBOX) with mildronate (**13**) has been used as a cardioprotective treatment[149] and for the treatment of seizures and alcohol intoxication[150]; however, the precise mechanism of action is unclear[136].

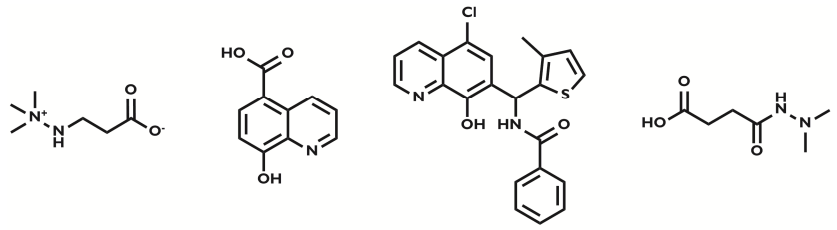


Figure 24 - Structures of Mildronate (**13**), IOX1 (**14**), CCT2-66 (**15**) and Daminozide (**16**).

The JmjC subset of the 2OG oxygenases have been implicated in a range of disease states[151]. In particular, the JmjC histone lysine demethylases (KDMs) have been implicated in a range of cancers[152,153] and cardiovascular disease[154]. The KDMs have therefore attracted significant interest from industry and academia[155,156] and has led to the development of probe compounds such as IOX1 (**14**)[157] alongside more selective compounds such as CCT2-66 (**15**)[158].

Plant 2OG oxygenases, particularly those implicated in the biosynthesis of gibberellins, have attracted interest as inhibitors of plant growth[136,159]. Daminozide (**16**), a plant growth regulator, has been found to inhibit the JmjC KDMs[160] in addition to inhibiting the giberellin biosynthesis enzymes.

Effect of 2OG Oxygenase Inhibitors on the Collagen Hydroxylases

Although clinical candidates are cross-screened against selected 2OG oxygenases, they are not routinely screened against the collagen hydroxylases due to the lack of a reliable screening platform. The effect of 2OG oxygenase inhibitors, outside of compounds that were developed as C-P4H inhibitors such as 2,4-PDCA (**5**), on the collagen hydroxylases, and hence collagen biosynthesis, is unknown. In addition, the mechanism of any anti-fibrotic activity exhibited by 2OG oxygenase inhibitors cannot be fully elucidated without an assessment on their impact on the collagen hydroxylases.

Thesis Objectives

The principal aim of the work described in this thesis was to contribute to biochemical characterisation the collagen prolyl hydroxylases and their inhibition. Specific objectives were:

- Development of a robust recombinant expression system for the human C-P4H enzymes.
- Biochemically characterise collagen prolyl hydroxylases and develop collagen prolyl hydroxylase inhibition assays.
- Assess clinically relevant HIF prolyl hydroxylase inhibitors for inihibition of collagen prolyl hydroxylases.

In addition to these studies, studies were initiated in order to study the hydroxylated residues in collagen and also to probe 2OG oxygenase catalysed post-translational modifications of lysine.

Chapter 2
Human Collagen Prolyl-4-Hydroxylase

The post-translational (*4R*)-hydroxylation of collagen proline residues is an enzyme-mediated process that was first identified in 1965[90] and the presence of (*2S,4R*)-hydroxyproline in collagen has been shown to be key in conferring stability upon the collagen triple helix[29]. Subsequent analysis of the cofactor and co-substrate requirements for prolyl hydroxylation led to the identification of the first collagen prolyl-4-hydroxylase (C-P4H)[91], which was the first member of the 2OG oxygenase superfamily to be defined. Subsequently, two further human C-P4Hs have been discovered[93,95,96].

The three human collagen prolyl hydroxylases are $\alpha_2\beta_2$ tetramers of around 240 kDa[103] and are localised to the endoplasmic reticulum[25]. The α-subunit contains the catalytic domain whilst the β-subunit is identical to the protein disulfide isomerase (PDI)[92]. There are three different forms of the α-subunit (α(I), α(II) and α(III)), encoded by P4HA1, P4HA2 and P4HA3 genes located on chromosomes 10, 5 and 11 respectively. Meanwhile, the β-subunit is encoded for by the P4HB gene that is located on chromosome 17. The C-P4H tetramer consists of two copies of the same form of α-subunit (*i.e.* mixed tetramers are not formed)[93] and therefore there are only three C-P4H isoforms.

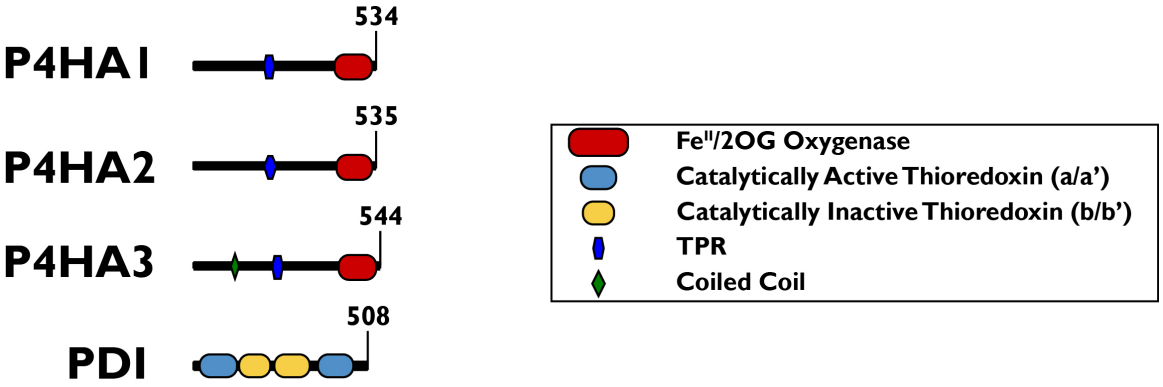


Figure 25 - Domain architecture of the three C-P4H α-subunits (proteins resulting from expression of P4HA1, P4HA2 and P4HA3) and the β-subunit (PDI, which results from expression of P4HB).

In addition to the catalytic Fe^{II}/2OG oxygenase domain (residues 245 to 517 in the α(I) isoform), a tetratricopeptide repeat (TPR) domain and, in the case of the α(III) isoform, a coiled coil domain (Figure 25), the α-subunit consists of an *N*-terminal dimerisation domain (residues 1 to

144 in the α(I) isoform) and peptide substrate binding domain (residues 145 to 244 in the α(I) isoform)[161,162]. The binding properties of the peptide substrate binding domains differ between the three α-subunit isoforms, since the α(I) form is known to bind poly-L-proline whereas the α(II) and α(III) forms do not[95-97].

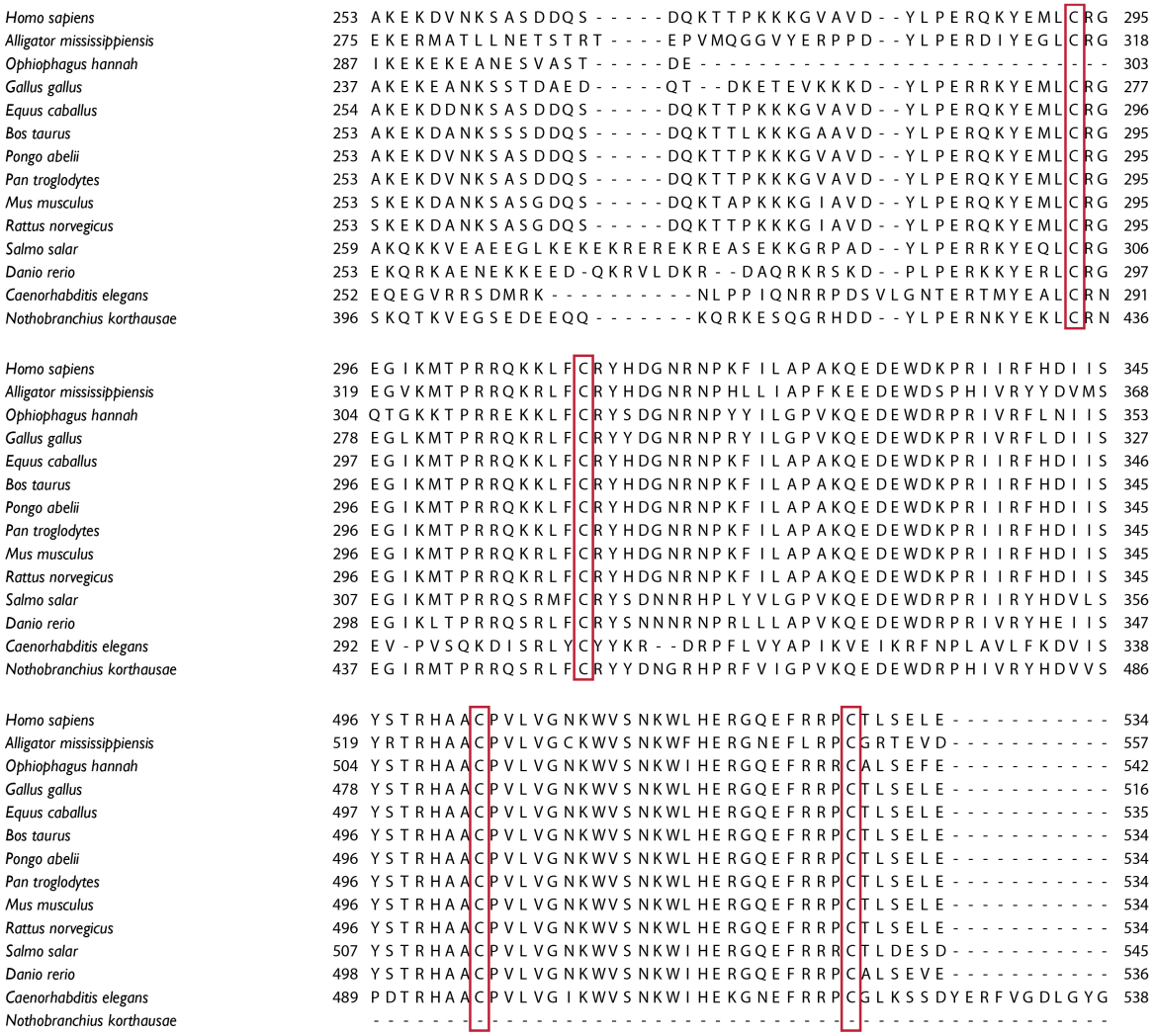


Figure 26 - Alignment of primary structures of C-P4H α(I) subunits from a selection of animals. The highly conserved cysteine residues believed to be involved in disulfide bonding are highlighted. Sequences resulting from the expression of P4HA1 from *Nothobranchius korthausae* (Uniprot ID:A0A1A8F9H4), *Caenorhabditis elegans* (Uniprot ID:Q10576), *Danio rerio* (Uniprot ID:Q7ZVE9), *Salmo salar* (Uniprot ID:C0H9T9), *Rattus norvegicus* (Uniprot ID:P54001), *Mus musculus* (Uniprot ID:Q60715), *Pan troglodytes* (Uniprot ID:H2Q231), *Pongo abelii* (Uniprot ID:Q5RAG8), *Bos taurus* (Uniprot ID:Q1RMU3), *Equus caballus* (Uniprot ID:F6PUV4), *Gallus gallus* (Uniprot ID:P16924), *Ophiophagus hannah* (Uniprot ID:V8NKS7), *Alligator mississippiensis* (Uniprot ID:A0A151M6V0) and *Homo sapiens* (Uniprot ID:P13674) were aligned using Clustal Omega[119].

PDI, meanwhile, contains four thioredoxin domains (a, b, b' and a'), which feature a C-X-X-C motif in the catalytic a and a' domains that flank the catalytically inactive b and b' domains

(Figure 25). The precise role of the PDI in the C-P4H tetramer is yet to be fully elucidated, but it is important for endoplasmic reticulum localisation and for ensuring that the otherwise insoluble α-subunit remains in solution and does not aggregate[163].

Tetramer formation is reported to be dependent upon the formation of at least one, and most likely two, intra-strand disulfide bridges that flank the catalytic domain in the α-subunit[164]; these disulfide bonds are between C293/C310 and C503/C527 in the human α(I) subunit. The cysteine residues that form these disulfide bonds are highly conserved across eukaryotes (Figure 26), indicating that their presence is important in allowing for C-P4H tetramer assembly.

Previous Structural Studies with Collagen Prolyl Hydroxylases

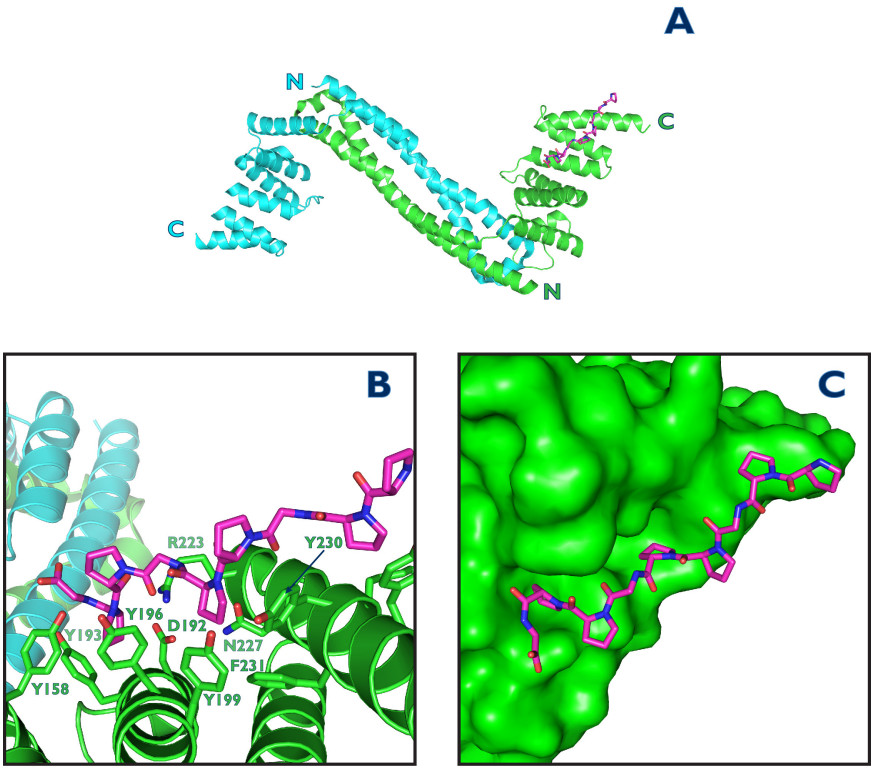


Figure 27 - Views from a crystal structure of the *N*-terminal dimerisation domain and peptide substrate binding domain of human C-P4H bound to a short peptide $(GPP)_3$ (PDBID: 4BTA[162]). A - Cartoon representation showing the dimerisation of two C-P4H α-subunit fragments (residues 1-238), the two subunits (blue and green respectively) dimerise about a pair of α helices. B - Close up of the binding of $(GPP)_3$ to the peptide substrate binding domain; side chains of residue within 5 Å of the peptide are shown and labelled. C - Binding mode of $(GPP)_3$, the protein is shown as a surface and a clear binding groove is observed.

To date, a structure of the complete human C-P4H tetramer has not been successfully solved. In addition, no structure of the catalytic domain of C-P4H has been determined. There are,

however, solved structures of the *N*-terminal dimerisation domain and peptide binding domains of the α(I) subunit (residues 1-238, Figure 27), along with a structure of PDI, the β-subunit, in isolation (Figure 28). PDI is a reasonably flexible structure; this flexibility, in particular the loop between b' and a', has been characterised in greater detail by NMR spectroscopy[165].

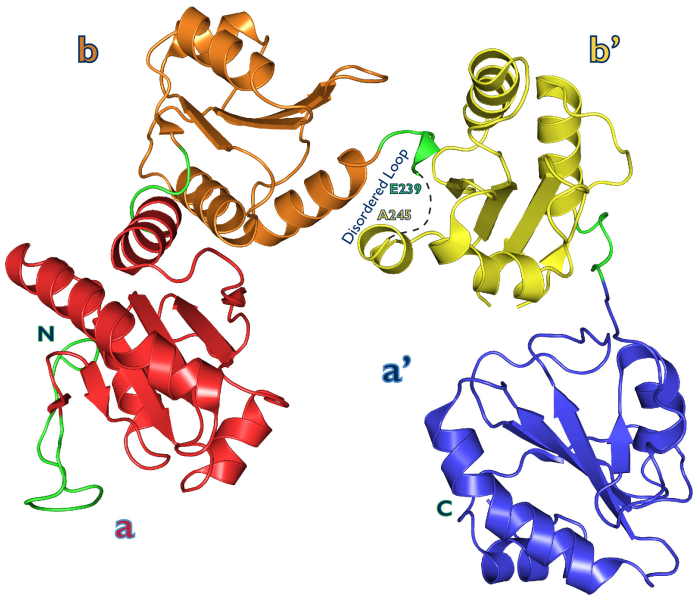


Figure 28 - A view from a crystal structure of reduced PDI (PDBID: 4EKZ[166]). The two catalytically active thioredoxin domains a and a' are indicated in red and blue respectively, while the catalytically inactive b and b' domains are indicated in orange and yellow. PDI has an abb'a' arrangement of domains, with a linker between b' and a', referred to as the x domain.

Small angle X-ray scattering (SAXS) has been used to analyse the C-P4H tetramer in solution; these studies[100] have, consistent with the crystallographic studies[162] (Figure 27), shown that C-P4H tetramer is a dimer of αβ dimers about the *N*-terminal dimerisation domain.

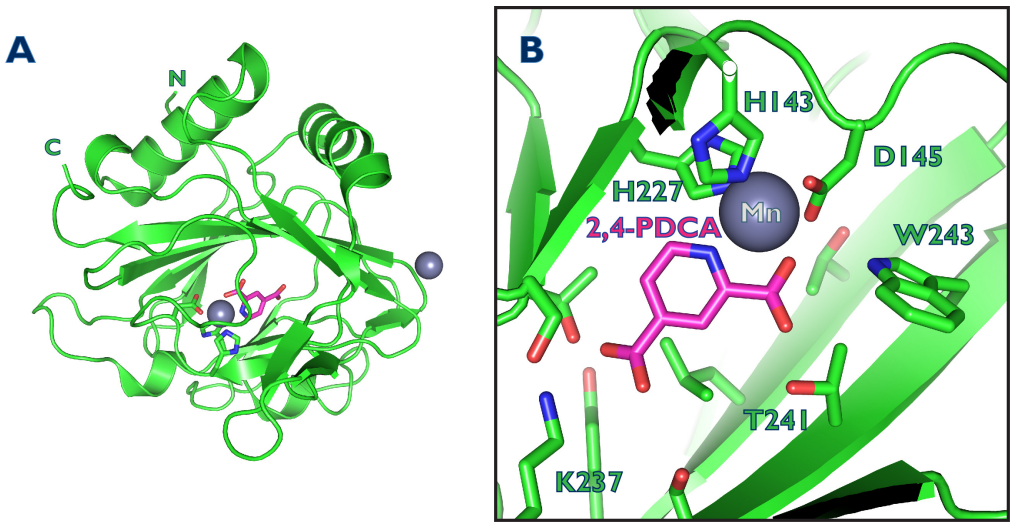


Figure 29 - Views from a crystal structure of a prolyl-4-hydroxylase from *Chlamydomonas reinhardtii* (residues 36-251, PDBID: 2JIG[167]). A - Cartoon representation of the whole protein, showing the DSBH fold. B - Close up view of the enzyme active site, side chains of residues within 5 Å of the 2OG mimic and bound inhibitor, 2,4-PDCA (**5**, indicated in pink) and Mn (grey). The HXD…H metal binding triad is labelled. K237 forms a dipole-dipole interaction with the carboxylate at the 4-position of 2,4-PDCA (**5**).

Collagen prolyl hydroxylases from other organisms have also been studied; structures of C-P4H-like proteins from *Chlamydomonas reinhardtii* (Figure 29)[167] and *Paramecium bursaria* Chlorella Virus I[84] have been solved following crystallographic studies. Both structures feature a DSBH and HXD…H metal co-ordination triad; with a lysine residue co-ordinating the distal carboxylate of 2OG (2OG analogues were used in structural studies). The structure of the prolyl-4-hydroxylase from *Paramecium bursaria* Chlorella Virus I[84] is discussed in greater detail in Chapter 3 and Chapter 4.

Recombinant C-P4H Production

Large scale recombinant production of the C-P4H tetramer is challenging; the α-subunit is insoluble in isolation, therefore necessitating co-expression. Additionally, the requirement for intrastrand disulfide bonds in the α-subunit to allow for tetramer assembly requires a non-reducing environment. C-P4H tetramer was purified from chick embryos[168] prior to the development of recombinant expression systems. Recombinant C-P4H tetramer expression has been achieved in insect cells[163,169], *in vitro* mammalian cultures[96] and more recently, expression systems in *Escherichia coli* (*E. coli*) have been reported in strains with mutated thioredoxin and

glutathione reductase genes (Origami strains), which creates a cytosolic environment more amenable for disulfide formation[170,171].

Two approaches have been used when designing and generating expression constructs; either two separate plasmids (encoding separately for the α and β-subunits) are co-transformed, or a single construct encoding for both subunits is generated. Both approaches have been reported in *E. coli* expression systems[170,171]. The use of different promoters in the dual-plasmid approach allows for a delay in the expression of the α-subunit into a 'pool' of β-subunit[171]. Subsequent purification of recombinantly produced enzyme has previously been achieved using poly-L-proline affinity chromatography, affording purification of $\alpha(I)_2\beta_2$ tetramer but not the $\alpha(II)_2\beta_2$ and $\alpha(III)_2\beta_2$ forms, due to the differing binding properties of the α(II) and α(III) subunits[93,96].

Previous studies in the Schofield group have seen the generation of a single expression vector, based on pETDuet-1, which encodes for both subunits of C-P4H[172], with the β-subunit in the first multiple cloning site (MCS) and an *N*-terminally hexahistidine (His_6) tagged α(I) subunit in the second MCS. Due to the intrinsic *N*-terminal His_6 tag in pETDuet-1, both the α(I) and β-subunits have *N*-terminal His_6 tags when expressed. Using this vector, small quantities of C-P4H were isolated (approx 2 mg from 12 L of culture) following a four-step purification process[173,174].

Previous Biochemical Studies with C-P4H

Biochemical studies have been conducted with the $\alpha(I)_2\beta_2$ and $\alpha(II)_2\beta_2$ tetramers that were produced in baculovirus[93]. 2OG turnover analyses have been conducted and suggest that the C-P4H enzymes have a high K_M^{app} for ascorbate[93,96,103]. 2OG turnover analyses have also indicated that the $\alpha(I)_2\beta_2$ and $\alpha(II)_2\beta_2$ and $\alpha(III)_2\beta_2$ tetramers have differing substrate affinities[87], which is consistent with the different binding peptide properties of the three isoforms[95-97].

Previous analyses of C-P4H activity have used the release of $^{14}CO_2$ from radio-labelled 2OG to analyse enzymatic activity[75,91,93,96,103,169,175]; such assays do not account for uncoupled turnover of

2OG and thus the development of a more reliable assay is desirable, not only to improve accuracy but also should allow for higher throughput and require a smaller quantity of enzyme than radio-labelled 2OG turnover experiments.

Aims

The aim of the work described in this chapter was to work towards the reliable and robust production of recombinant, active C-P4H tetramer that could subsequently be used for biochemical and structural studies. Development of an *E. coli* expression system was desirable since relatively large quantities of protein are required for structural studies.

Additionally, development of recombinant expression systems for the production of all three C-P4H isoforms was sought since recombinant C-P4H production in *E. coli* has only been reported for the $\alpha(I)_2\beta_2$ form[170,171].

C-P4H Production with pETDuet-1_PDI_HisCPH1

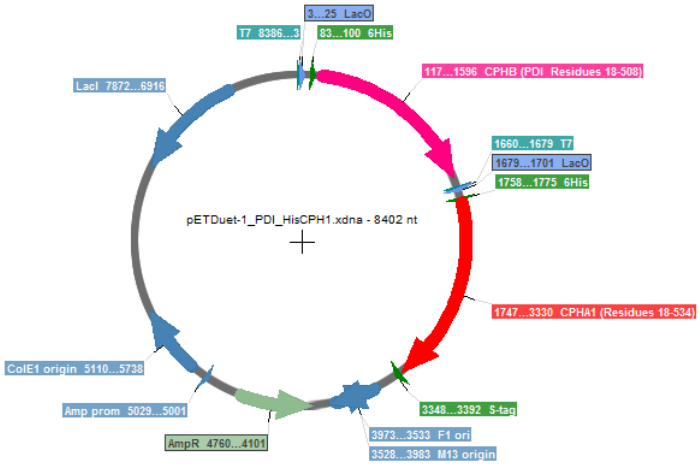


Figure 30 - Vector map for pETDuet-1_PDI_HisCPH1[173,174].

The procedures established in the Schofield group[173,174] were repeated; these conditions consist of expression of the pETDuet-1_PDI_HisCPH1 construct in *E. coli* Origami 2 (DE3) followed by purification consisting of Ni^{II}-affinity chromatography and two ion exchange steps (using Heparin and MonoQ columns), prior to size exclusion chromatography. Before each ion exchange step, the protein was buffer exchanged into NaCl-free buffer.

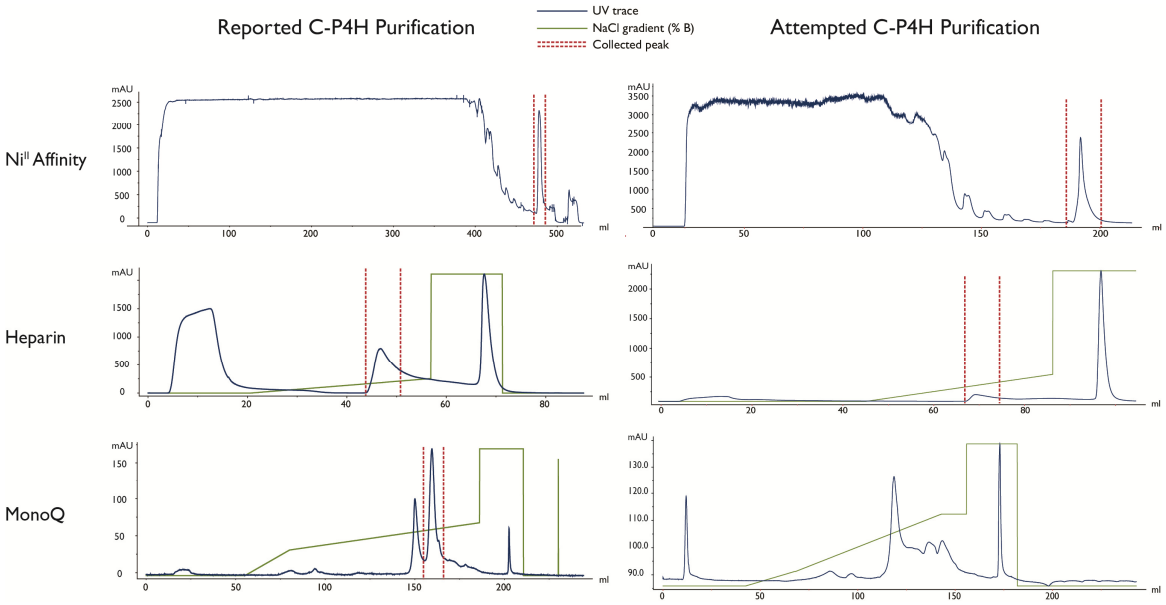


Figure 31 - UV traces (detection at λ = 280 nm) of a C-P4H purification reported previously[173,174] compared to representative traces of repeats of the reported procedure using the same expression construct and *E. coli* strain[173,174].

No C-P4H tetramer was isolated using the reported method (Figure 31)[173,174]. Although some protein was eluted during the Ni^{II} affinity step, it appears that, given the absence of peaks in the following ion exchange steps, no tetramer was formed. Analysis by gel electrophoresis revealed no presence of C-P4H tetramer in the collected Heparin or MonoQ peaks (Figure 32). Yields were not improved by repeating the method and changing buffer NaCl concentrations or using different buffer salts.

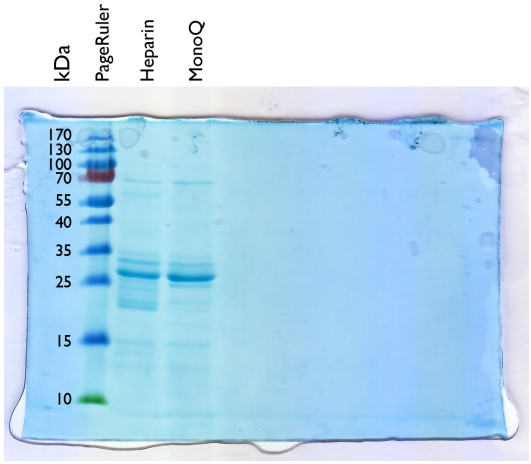


Figure 32 - SDS-PAGE analysis of the peaks eluted from ion-exchange steps during attempted purification of C-P4H. No clear bands corresponding to the C-P4H subunits were observed.

Expression Trials with *E. coli* Origami 2 (DE3)

Following unsuccessful attempts to produce C-P4H using the previous method[173,174], small scale (20 mL culture) expression trials with the pETDuet-1_CPH_HisPDI vector were conducted. SDS-PAGE analyses of crude lysates did not show the presence of any bands in the region of interest, suggesting either no expression of C-P4H, or that expression was too low to be identified.

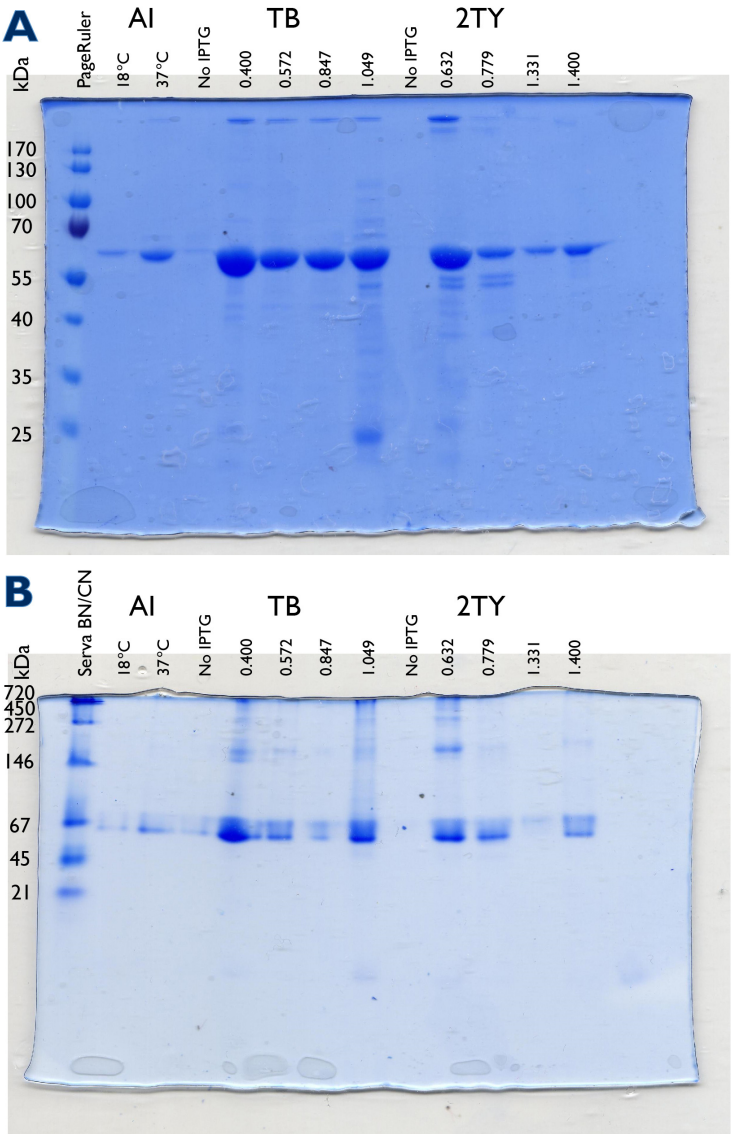


Figure 33 - SDS-PAGE (A) and non-denaturing PAGE (B) analysis of fractions eluted following small scale expression trials and Ni^{II} affinity purification in auto induction (AI), terrific broth (TB) and 2 x tryptone/yeast (2TY) media. The OD_{600} at which C-P4H expression was induced is indicated in lanes corresponding to expression trials TB and 2TY.

Therefore, to improve detection, expression trials were conducted on a larger scale (100 mL cultures) and a small scale Ni^{II} affinity purification (using a 1 mL HisTrap column) was conducted on the resulting lysate. The optical density at λ = 600 nm (OD_{600}) at which the cells were induced *via* addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG, **17**) was varied (0.4-1.4) in cultures grown in terrific broth (TB) and 2 x tryptone/yeast (2TY) media. Alongside trials in 2TY and TB, expression trials in autoinduction media (AI) were conducted and cells were incubated at both 37°C and 18°C post-induction.

SDS-PAGE and non-denaturing PAGE analyses of the purified cell lysates are presented in Figure 33. A band was observed corresponding to protein(s) of around 60 kDa; when expressed by pETDuet-1_PDI_HisCPH1, the α(I)and β-subunits have masses of 60295 Da and 56767 Da respectively. MS-MS analysis of excised bands indicated that this band was the soluble β-subunit. Non-denaturing PAGE showed that the eluted protein was predominantly monomeric, although some higher molecular weight bands were observed when expression was induced earlier in the log phase (Figure 33B).

In addition to PAGE analyses, protein samples isolated following Ni^{II} affinity chromatography were concentrated and then incubated with $(NH_4)_2Fe(SO_4)_2$ (50 μM), 2OG (disodium salt, 5 mM), sodium-L-ascorbate (10 mM) and Peptide 1 ($(GPP)_{10}$, 100 μM), a collagen triplet repeat peptide and known C-P4H substrate[25,103]. The peptide was subsequently analysed by mass spectrometry (MS), using a matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) MS instrument; MALDI predominantly produces singly charged ions (*e.g.* $[M+H]^+$), simplifying analysis. These assays showed that the sample of protein isolate from the cells grown in TB medium and induced at OD_{600} 0.400 may have very limited prolyl hydroxylase activity; a slight peak with a Δ*m/z* of +16 was present in the MS spectra after a 16 hour incubation at 25°C (Figure 34). No peak with Δ*m/z* of +16 was observed when an incubation was conducted with a

protein sample from cells that were not induced (*i.e.* no IPTG, **17**, was added); no change was observed when peptide was incubated with only cofactors/co-substrate.

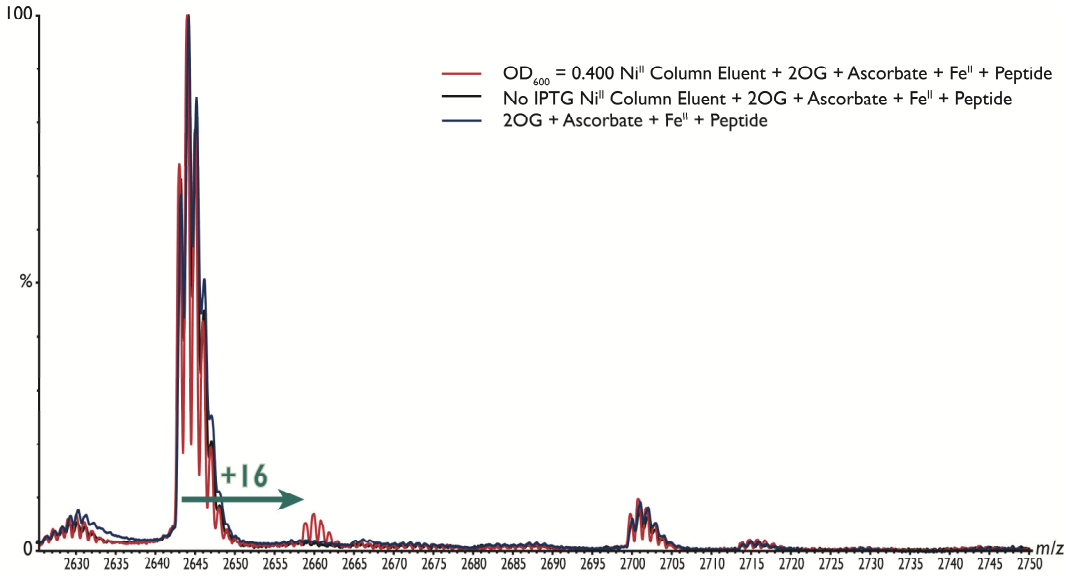


Figure 34 - Overlaid mass spectra showing the emergence of a peak with Δ*m/z* of +16 when Peptide 1 was incubated with cofactors and protein eluted following Ni^{II} affinity chromatography (red). No product peak was observed with a sample produced from cells that were not treated with IPTG (**17**, blue) or with a peptide sample treated with cofactors/co-substrates but not treated with enzyme (black).

Scale up of the cell growth in TB medium was then conducted using the conditions informed by expression trials (induction of expression with 500 µM IPTG (**17**) at an OD_{600} of 0.400 and overnight incubation at 18°C). However, subsequent attempted protein purification by Ni^{II} affinity chromatography did not yield any active C-P4H tetramer. Purification was attempted using the previously reported four-step purification protocol[173,174] as well as an approach employing Ni^{II} affinity chromatography followed by size exclusion chromatography. SDS-PAGE and non-denaturing PAGE (Figure 35) did not indicate successful C-P4H production and purification; the resultant protein was not active under the same conditions applied previously (Figure 34).

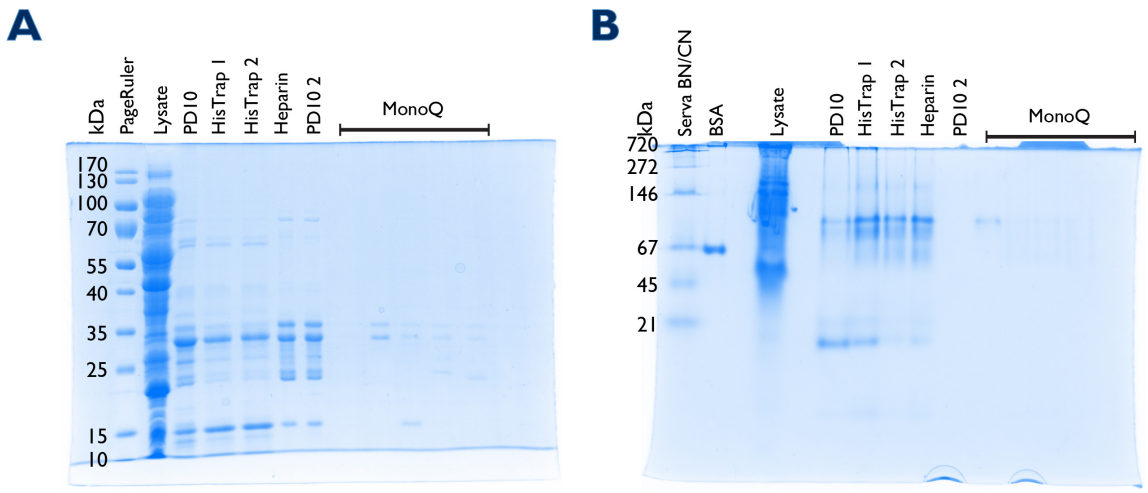


Figure 35 - SDS-PAGE (A) and non-denaturing PAGE (B) analysis of fractions during an attempted four step purification of C-P4H from *E. coli* Origami 2 (DE3). No bands corresponding to the subunits were observed by SDS-PAGE and the tetramer was not observed by non-denaturing PAGE.

Expression Trials with *E. coli* Origami B (DE3) pLysS

Following the lack of success with the *E.* coli Origami 2 (DE3) expression strain, expression trials were conducted in *E. coli* Origami B (DE3) pLysS with the pETDuet-1_CPH_HisPDI vector. *E. coli* Origami B cells are derived from a *lacZY* mutant of BL21 making it possible to control expression by varying IPTG (**17**) concentration upon induction (*i.e.* expression is slower at lower concentrations of IPTG (**17**)). The concentration of IPTG (**17**) and the OD_{600} of induction in cultures grown in either 2TY or TB media were varied, and cells were incubated at 18°C post-induction. As with the expression trials with *E. coli* Origami 2 (DE3), samples were purified by Ni^{II} affinity chromatography prior to analysis of eluted protein by SDS-PAGE and non-denaturing PAGE (Figure 36). While the appearance of a double band by SDS-PAGE was promising, no bands corresponding to the tetrameric form were observed by non-denaturing PAGE.

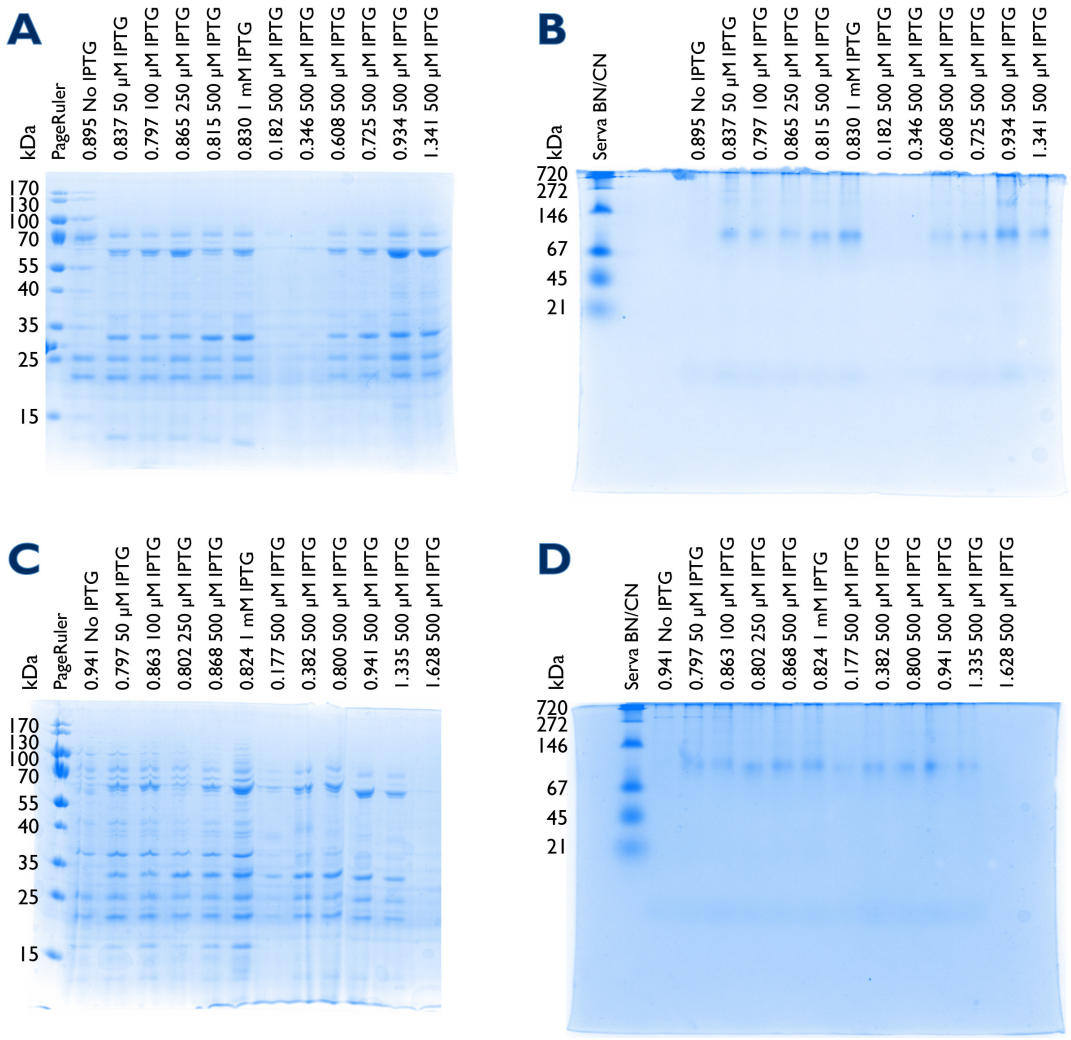


Figure 36 - SDS-PAGE and non-denaturing PAGE analysis of expression trials with *E. coli* Origami B (DE3) pLysS and pETDuet-1_CPH_HisPDI, samples were subjected to a small scale Ni^{II} affinity purification prior to analysis. The OD_{600} at induction and final concentration of IPTG (**17**) are indicated above each lane. A - SDS-PAGE of samples cultured in TB. B - Non-denaturing PAGE of samples cultured in TB. C - SDS-PAGE of samples cultured in 2TY. D - Non-denaturing PAGE of samples cultured in 2TY.

Following the inability to successfully express and purify sufficient quantities of C-P4H tetramer using the pETDuet-1_PDI_HisCPH1 vector, it was decided to generate a range of new expression vectors.

Constructs Designed by OPPF-UK

In collaboration with the Oxford Protein Production Facility (OPPF-UK), a range of expression vectors for protein production in *E. coli* were produced *via* In-Fusion™ cloning, making use of the popIN vector suite and methods developed by OPPF-UK[176]. An expression system that utilises two expression vectors was designed (Table 2) and separate constructs for the α-subunit and

β-subunit were produced, which would then be co-transformed into an appropriate *E. coli* strain for protein production. Both plasmids contain the *lacI* operon and therefore production of both proteins would be simultaneously induced by addition of IPTG (**17**).

Name	Backbone	Protein Encoded (Residues)	Tag	Antibiotic Resistance
pOPINRSE_α(I)_Stop	pOPINRSE	CPHα(I) (18-534)	No Tag	Kanamycin
pOPINRSE_α(I)	pOPINRSE	CPHα(I) (18-534)	*C*-terminal His_6	Kanamycin
pOPINRSF_α(I)	pOPINRSF	CPHα(I) (18-534)	*N*-terminal His_6	Kanamycin
pOPINRSE_α(I)C167S_Stop	pOPINRSE	CPHα(I) (18-534) C167S variant	No Tag	Kanamycin
pOPINRSE_α(I)C167S	pOPINRSE	CPHα(I) (18-534) C167S variant	*C*-terminal His_6	Kanamycin
pOPINRSF_α(I)C167S	pOPINRSF	CPHα(I) (18-534) C167S variant	*N*-terminal His_6	Kanamycin
pOPINRSE_α(II)_Stop	pOPINRSE	CPHα(II) (22-535)	No Tag	Kanamycin
pOPINRSE_α(II)	pOPINRSE	CPHα(II) (22-535)	*C*-terminal His_6	Kanamycin
pOPINRSF_α(II)	pOPINRSF	CPHα(II) (22-535)	*N*-terminal His_6	Kanamycin
pOPINRSE_α(III)_Stop	pOPINRSE	CPHα(III) (19-544)	No Tag	Kanamycin
pOPINRSE_α(III)	pOPINRSE	CPHα(III) (19-544)	*C*-terminal His_6	Kanamycin
pOPINRSF_α(III)	pOPINRSF	CPHα(III) (19-544)	*N*-terminal His_6	Kanamycin
pOPINCDE_β_Stop	pOPINCDE	CPHβ (18-508)	No Tag	Spectinomycin
pOPINCDE_β	pOPINCDE	CPHβ (18-508)	*C*-terminal His_6	Spectinomycin
pOPINCDF_β	pOPINCDF	CPHβ (18-508)	*N*-terminal His_6	Spectinomycin

Table 2 - List of expression constructs that were designed in collaboration with OPPF-UK. Production of constructs was not successful for red entries.

α-Subunit

Genes were incorporated into pOPINRSE both with and without an in-frame stop codon (to produce untagged protein and *C*-terminally His_6 tagged protein respectively) alongside pOPINRSF (encoding for an *N*-terminally His_6 tagged protein). Vectors were designed for all the three α-subunit isoforms with the endoplasmic reticulum localisation domain removed. In addition, a mutated α(I) form that omits the non-disulfide-forming cysteine (C167 was mutated to encode for serine) was produced reduce formation of an incorrect disulfide bonding pattern.

Cloning was successful for all α(I) and α(I) C167S vectors; site directed mutagenesis of pOPINRSF_α(I) was required to install a stop codon in frame with the gene. However, production of pOPINRSE_α(II) and all α(III) constructs were not successful as gene amplification *via* the polymerase chain reaction (PCR) failed.

β-Subunit

The popINCDE and popINCDF vectors were combined with genes encoding the β-subunit, with the localisation domain removed, to produce constructs that encoded for *N*- and *C*- terminally His_6 tagged proteins; a popINCDE construct with a stop codon following the CPHB gene (which encodes for the β-subunit) was also produced, encoding for untagged protein. All three constructs were successfully produced.

Expression Trials at OPPF-UK

α-subunit Construct	β-subunit Construct	Resulting Tetramer Tag
pOPINRSE_α(I)	pOPINCDE_β_Stop	α-subunit *C*-terminal His_6
pOPINRSF_α(I)	pOPINCDE_β_Stop	α-subunit *N*-terminal His_6
pOPINRSE_α(I)_Stop	pOPINCDE_β	β-subunit *C*-terminal His_6
pOPINRSE_α(I)_Stop	pOPINCDF_β	β-subunit *N*-terminal His_6
pOPINRSE_α(I) C167S	pOPINCDE_β_Stop	α-subunit *C*-terminal His_6
pOPINRSF_α(I) C167S	pOPINCDE_β_Stop	α-subunit *N*-terminal His_6
pOPINRSE_α(I) C167S _Stop	pOPINCDE_β	β-subunit *C*-terminal His_6
pOPINRSE_α(I) C167S _Stop	pOPINCDF_β	β-subunit *N*-terminal His_6
pOPINRSF_α(II)	pOPINCDE_β_Stop	α-subunit *N*-terminal His_6
pOPINRSE_α(II)_Stop	pOPINCDE_β	β-subunit *C*-terminal His_6
pOPINRSE_α(II)_Stop	pOPINCDF_β	β-subunit *N*-terminal His_6

Table 3 - Combinations of expression vectors used in expression trials.

Expression trials were conducted following OPPF-UK protocols[176]; following co-transformation of the generated pOPIN vectors into either *E. coli* BL21 (DE3) or *E. coli* Lemo21 (DE3) in the combinations detailed in Table 3. Expression trials were then conducted with both cell types in AI and Power Broth media, cells grown in Power Broth were induced with IPTG (**17**) at a final concentration of 1 mM when OD_{600} reached 0.6 and were then incubated at 20°C overnight and AI media was incubated at 20°C overnight after reaching an OD_{600} of 0.6. Following a small scale Ni^{II} affinity purification, SDS-PAGE analysis did not show any bands consistent with expression of

C-P4H tetramer; a green fluorescent protein control however did show good expression. This result was unsurprising since both the *E. coli* BL21 (DE3) and *E. coli* Lemo21 (DE3) expression strains do not have a cytosolic environment amenable to the formation of disulfide bonds.

Expression Trials in *E. coli* Origami 2 (DE3) pLysS

Expression trials were then attempted in *E. coli* Origami 2 (DE3) pLysS cells, which, unlike *E. coli* BL21 (DE3) and *E. coli* Lemo21 (DE3), possess a more oxidising environment that is amenable to disulfide formation. The vector combinations in Table 3 were co-transformed into *E. coli* Origami 2 (DE3) pLysS and small scale expression trials were conducted. Cultures (1 mL) were pelleted and then lysed before the soluble fractions were analysed by dot blot (Figure 37), using an anti-His_6 antibody. In many cases, evidence for production was His_6 tagged protein was obtained, as chemiluminescent spots were observed during analysis following treatment with substrate for the horseradish peroxidise-conjugated secondary antibody (Figure 37A).

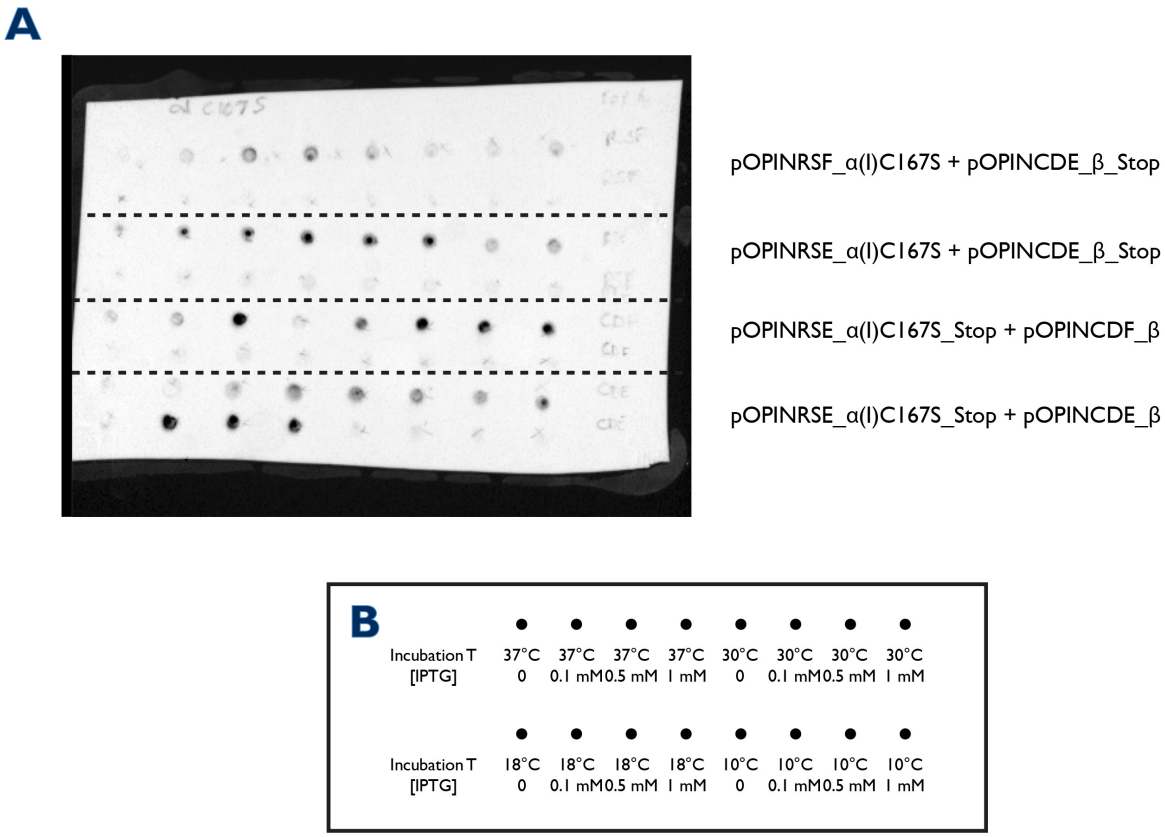


Figure 37 - Dot blot analysis of expression trials with pOPIN α(I)C167S expression vectors in combination with pOPIN β vectors. A - Merge of chemiluminescene and colorimetric channels. B - Arrangement of conditions for each expression trial in A.

SDS-PAGE analyses were then initiated with samples from the pOPINRSF_α(I)C167S/pOPINCDE_β_Stop, pOPINRSE_α(I)C167S/pOPINCDE_β_Stop and pOPINRSE_α(I)C167S _Stop/pOPINCDF_β combinations (Figure 38); however, there were no clear bands corresponding to the molecular weights of the two C-P4H subunits.

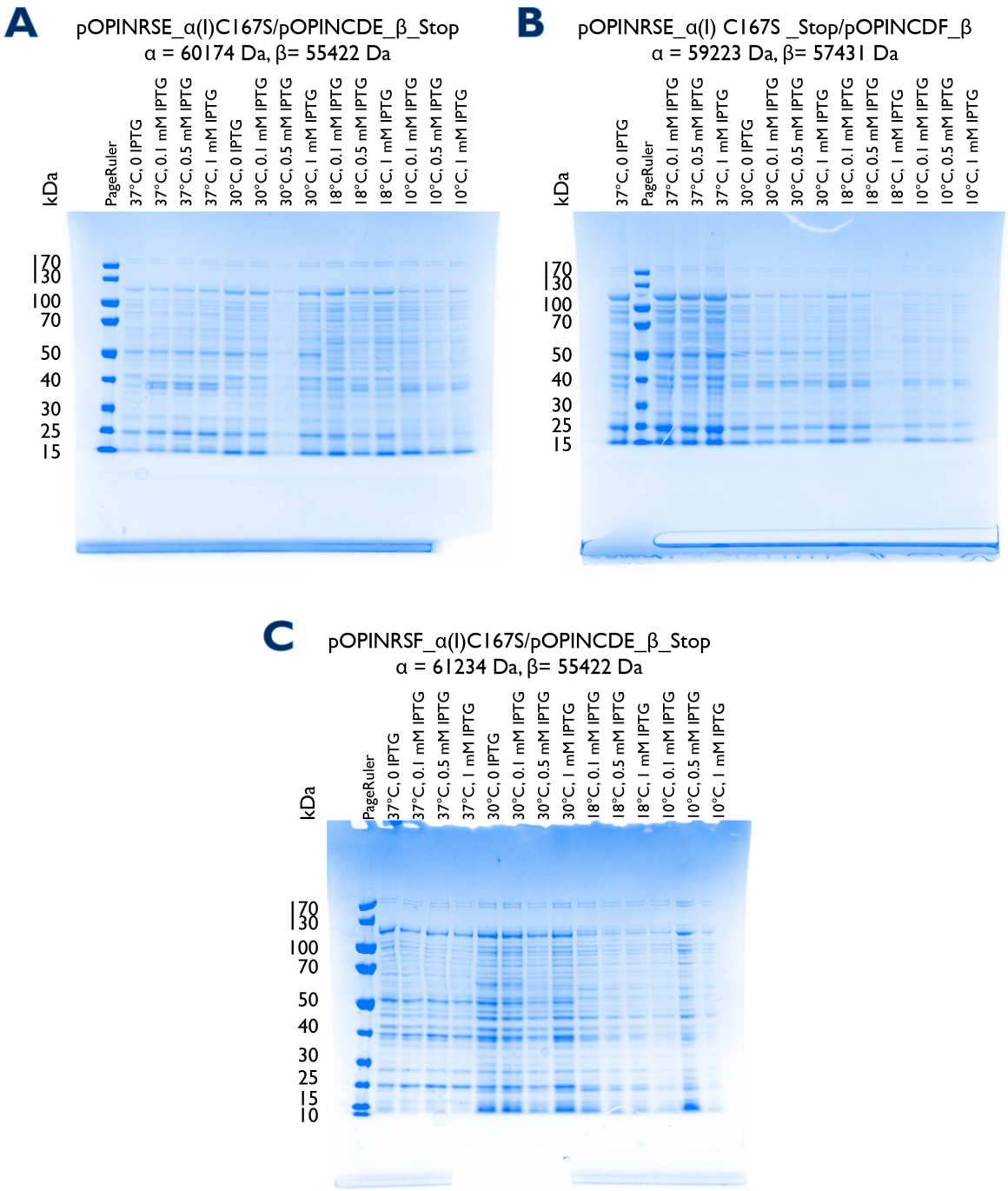


Figure 38 - SDS-PAGE analysis of expression trials conducted with pOPIN vectors encoding for the C-P4H α(I)C167S variant. No clear bands for the α or β-subunits were observed.

The observation of His_6 tagged protein in the pOPINRSE_α(I)C167S/pOPINCDE_β_Stop dot blot (Figure 37) indicated the successful expression of the α(I)C167S subunit. Therefore, small scale (100 mL) cultures were initiated and small scale Ni^{II} affinity chromatography purifications were attempted. However, no protein was isolated following Ni^{II} affinity chromatography.

It was hypothesised that the lysis methodology, or potentially the salt content of the lysis buffer, may be causing the protein to precipitate. Therefore, alternative lysis techniques (homogenisation and mechanical lysis) were investigated. In all cases, His_6 tagged protein (*i.e.* the α(I)C167S subunit) was observed in the insoluble fraction by western blot; no His_6 tagged protein was observed in the soluble fraction (Figure 39).

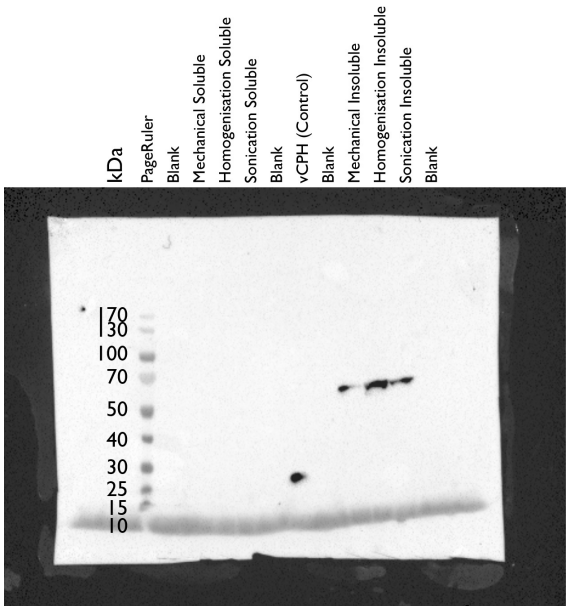


Figure 39 - Merged colorimetric and chemiluminescence channels of a western blot. The blot was probed for the presence of His_6 tagged protein (and by implication, the α(I)C167S subunit) using an anti-His_6 antibody. His_6 tagged vCPH was run as a standard. The α(I)C167S subunit was only observed in the insoluble fraction.

Expression trials in *E. coli* Origami B (DE3) pLysS were not possible due to the intrinsic kanamycin resistance of *E. coli* Origami B (DE3) pLysS.

Arabinose Inducible Vectors

Arabinose inducible vectors, using the pBAD promoter, were then produced using ligation independent cloning (LIC) methodologies[177]. This approach was pursued in order to allow for

differential induction timepoints of the α- and β-subunits, to mimic an approach that has previously been reported as a successful route to C-P4H tetramer[171].

The α(I)C167S, α(II) and α(III) subunits were separately introduced into the pBADHis vector and these vectors were separately co-transformed with pOPINCDE_β_stop into *E. coli* Origami 2 (DE3) and used in expression trials. Expression of the β-subunit was initiated with 500 µM IPTG (**17**) prior to the induction of the α-subunit with arabinose. Dot blot analysis did not show the presence of any His_6 tagged protein in the lysate; therefore, no α-subunit and no C-P4H tetramer was found in the soluble fraction.

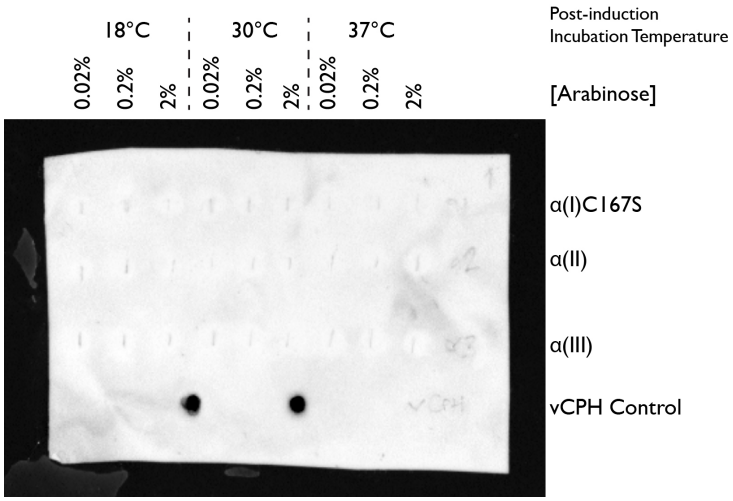


Figure 40 - Dot blot analysis of expression trials with pBADHis α-subunit vectors in combination with pOPINCDE_β_stop in *E. coli* Origami 2 (DE3). Colorimetric and chemiluminesence channels are merged to produce the final image. No His_6 tagged protein (*i.e.* α-subunit) was observed, indicating that no C-P4H tetramer is present in the soluble fraction.

Dual-Tag Approach

A dual-tagging system (*i.e.* different recombinant tags on the α and β-subunits)[178] was considered as a route to efficiently purify the C-P4H complex (provided that soluble protein can be obtained). The use of alternative recombinant tags (*i.e.* tags other than His_6) can also improve the solubility of the subunits. Therefore, expression vectors which encoded for an *N*-terminal maltose binding protein (MBP) fusion tag were produced using LIC to generate separate plasmids encoding for MBP tagged α(I), α(I)C167S, α(II) and α(III) subunits. These vectors are

awaiting sequence verification prior to expression trials in combination with the pOPIN β-subunit constructs.

Conclusions

The C-P4H enzymes play a key role in collagen biosynthesis[25,103] and their activities have a clear link to medicine and physiology. Although some basic characterisation of the $\alpha(I)_2\beta_2$ tetramer has been conducted *via* the monitoring of the turnover of radioactive 2OG[89], detailed biochemical studies into the differences between the three isoforms of the α-subunit have not been conducted, despite studies showing that the isoforms have differing tissue localisation[94] and binding properties[95-97].

Such studies require a reliable and reproducible method for the recombinant production of C-P4H tetramer. Mammalian[96] and viral[97] expression systems have previously been used to express the C-P4H tetramer, but do not yield sufficient quantities of pure protein for large scale biochemical and structural studies. Therefore, recombinant C-P4H production has been pursued using an *E. coli* expression system, in order to develop a system that affords large quantities of enzyme that in turn would allow for in-depth biochemical and structural studies.

Due to the requirement for a disulfide bonding pattern in the α-subunit[164], combined with the necessity for a co-expression system (due to insolubility of the α-subunit in isolation), the recombinant expression of C-P4H is a major challenge. Recombinant production of C-P4H(I) in *E. coli* has been reported[170,171]; however, recombinant production of the $\alpha(II)_2\beta_2$ and $\alpha(III)_2\beta_2$ C-P4H tetramers in *E. coli* has not been achieved to date.

Despite some reported success with the pETDuet-1_PDI_HisCPH1 construct in *E. coli* Origami 2 (DE3)[173,174], these protocols did not yield any C-P4H tetramer when replicated. Expression trials using the pETDuet-1_PDI_HisCPH1 plasmid in different *E. coli* strains were conducted and ultimately proved to be unsuccessful.

These expression trials were followed by the generation of a range of new plasmid expression constructs in collaboration with OPPF-UK, which were used in expression trials. Expression of the α-subunit was achieved, but the protein was insoluble. Altering salt concentration, buffer and lysis methods did not result in the isolation of soluble C-P4H tetramer. In addition, a dual plasmid system utilising different promoters on separate plasmids for the α- and β-subunits was implemented in a series of expression trials; again, these trials yielded no soluble C-P4H tetramer.

The combined failure of all three approaches (single plasmid, two plasmids and two plasmids with the differing induction conditions) suggests that the poor solubility of the α-subunit is the greatest challenge to overcome in the production of recombinant C-P4H; the lack of robust recombinant production methods is most likely the primary reason that the enzyme remains poorly understood on a biochemical level. Given the lack of success with *E. coli* in these studies, a more complex expression system appears to be required for large scale C-P4H production. The requirement for disulfides limits the range of *E. coli* strains that can be used. Therefore, pursuit of an insect cell expression system may provide a more tractable approach.

Once a reliable expression system for active C-P4H is established, it is hoped that a high-throughput screening platform, most likely *via* a MS-based assay, can be implemented. Additionally, structural studies will be pursued. In the absence of a robust C-P4H production protocol, a model system may be sought to study the action of the C-P4Hs.

Summary

The recombinant production of human C-P4H tetramer in *E. coli* is highly challenging, with the requirement for a disulfide bonding pattern in the α-subunit, coupled with the insolubility of the α-subunit playing a key role in this challenge. Establishment of a reliable system for the recombinant production of C-P4H tetramer is required before the development of high-throughput inhibition assays and structural studies can commence.

Attempts to produce sufficient quantities of recombinant C-P4H in *E. coli* have not been successful to date. Attempts to reproduce previously reported C-P4H production method[173,174] were fruitless and despite some positive indications resulting from small scale expression trials, no tetramer was isolated. A range of new expression constructs have been produced and used in small scale expression trails; despite evidence for the expression of the α-subunit, no tetramer was isolated following scale-up. Expression systems beyond *E. coli*, such as insect cells, may be a more tractable route towards the recombinant production of C-P4H tetramer.

Chapter 3 Characterisation of a Prolyl Hydroxylase from *Paramecium bursaria* Chlorella Virus 1

In light of the experimental challenge presented of working with the human C-P4H enzymes (Chapter 2), a model system was sought to better understand the biochemistry of collagen hydroxylases. A 27 kDa prolyl hydroxylase from *Paramecium Bursaria* Chlorella Virus I (vCPH) has previously been identified as a prolyl-4-hydroxylase[179]. Studies have reported that vCPH catalyses the production of 4-hydroxyproline in $(GPP)_n$ repeat peptides and additionally in $(PAPK)_n$ peptides[179]. However, the stereochemistry of this enzymatically-catalysed hydroxylation has not been definitively determined. vCPH has previously been used as a representative 2OG oxygenase, and has been used in stopped-flow experiments to allow for time resolved analysis of the metal-ligand charge transfer complex that forms during the catalytic cycle[84].

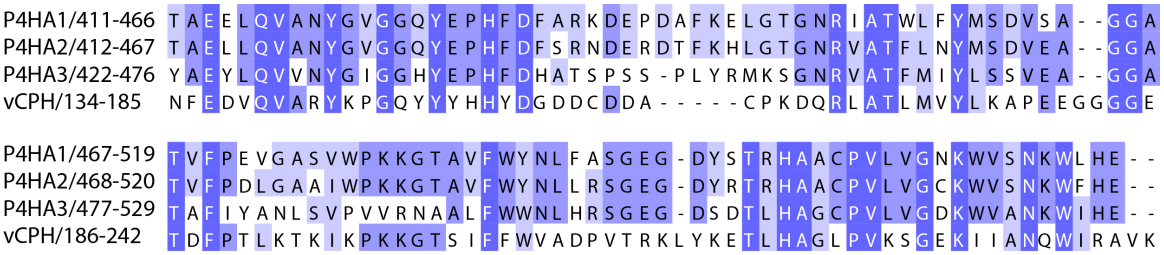


Figure 41 - Sequence alignment of vCPH with the core Fe^{II}/2OG oxygenase region of the human collagen prolyl-4-hydroxylase α-subunits. Protein sequences of P4HA1 (C-P4H α(I) subunit, Uniprot ID: P13674), P4HA2 (C-P4H α(II) subunit, Uniprot ID: O15460), P4HA3 (C-P4H α(III) subunit, Uniprot ID: Q7Z4N8) and vCPH (Uniprot ID: Q84406) were aligned using ClustalOmega[119].

Analysis of the primary sequences of vCPH and the human prolyl-4-hydroxylases indicates that vCPH is more similar to the C-P4Hs than the HIF prolyl-4-hydroxylases. Alignment of the core region of vCPH with those of the human 2OG oxygenases indicates that vCPH most closely aligns to the human C-P4Hs and P4H-TM. vCPH shares 32% identity with the core oxygenase regions of P4HA1 and P4HA2, and 35% identity with that of P4HA3; vCPH also shares 35% identity with the core regions of P4H-TM. vCPH aligns less closely with the HIF prolyl hydroxylases (PHD1, PHD2 and PHD3), with the core regions sharing around 20% identity; the core region of P4HA1, meanwhile, shares around 17% identity with that of PHD2.

vCPH has been identified as a 2OG oxygenase[84,179] and the rate of 2OG turnover catalysed by vCPH is increased in the presence of collagen-like repeat peptides. Therefore, due to its greater

similarity with the collagen prolyl hydroxylases versus the other human prolyl hydroxylases, such as the HIF prolyl hydroxylases, alongside its reported activity, [84,179] mean that vCPH has potential as a model for the human C-P4H.

vCPH Structure

The structure of vCPH has previously been solved in complex with Mn and, through soaking crystals, 2OG (**1**) by X-ray crystal diffraction[84]. No structure in complex with a vCPH substrate or inhibitor has been successfully solved and reported. Like other 2OG oxygenases, vCPH has a core DSBH fold (Figure 42A). The metal is co-ordinated by two histidinyl residues and an aspartic acid residue in a HXD…H motif (H152, D154 and H221, Figure 42B), consistent with most, but not all other 2OG oxygenases[80]. 2OG co-ordinates the metal centre in a bidentate fashion and is also positioned to form an electrostatic interaction with N^{ϵ} of K231 (Figure 42B).

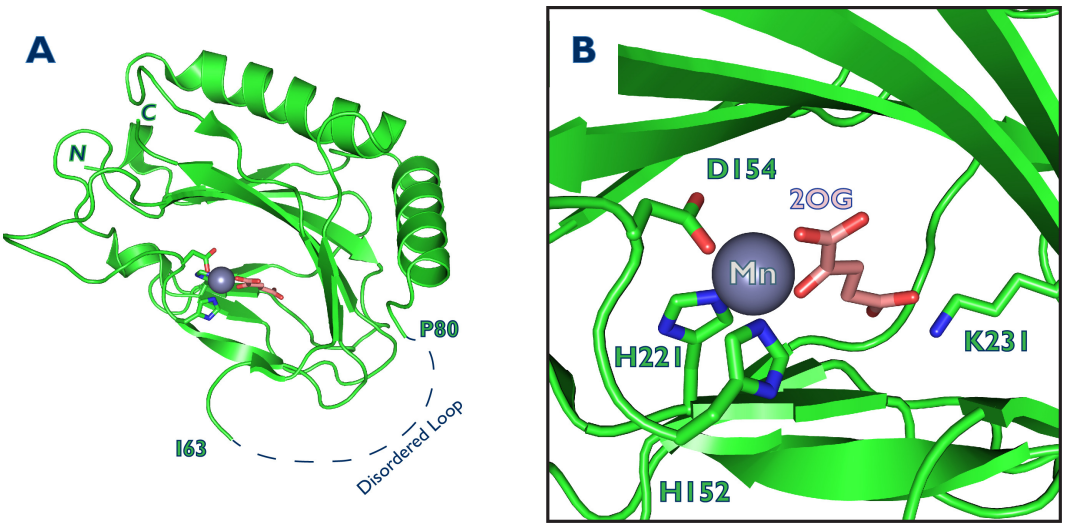


Figure 42 - Views from a crystal structure of vCPH in complex with Mn and 2OG (PDBID: 5C5T[84]). A - Cartoon representation of the structure of vCPH. B - The active site of vCPH. The metal binding triad, Mn, 2OG and K231 (which interacts with 2OG) are indicated.

Expression and Purification of vCPH

Expression trials using a previously reported vCPH construct (residues 36-242 in pET28a)[84] that encodes for *N*-terminally His_6 tagged vCPH were conducted in *E. coli* Rosetta(DE3) pLysS cells. A clear band corresponding to a protein of approximately 25 kDa (the predicted mass of the expressed protein is 25822 Da) was observed in induced samples but not in lanes corresponding

to non-induced cells (*i.e.* cells that were not treated with IPTG, **17**, Figure 43). Conditions for large scale vCPH production were selected based upon the conditions which gave the most prominent SDS-PAGE bands at approximately 25 kDa (Figure 43, induction at OD_{600} of approximately 1.000 with 0.5 mM IPTG, **17**).

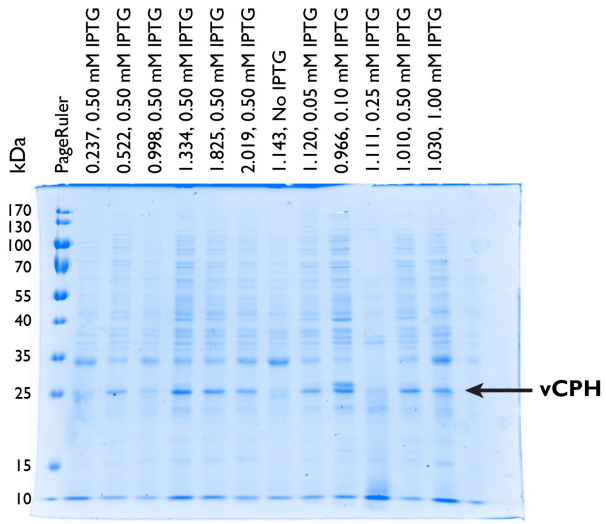


Figure 43 - SDS-PAGE analysis of small scale expression trials for the pET28a_vCPH36-242 construct in *E. coli* Rosetta(DE3) pLysS cells. Expression trials were conducted in 2TY medium; the OD_{600} values at induction and final IPTG (**17**) concentrations are indicated in lane labels. For the lanes shown, cells were incubated at 18°C for 16 hours following induction with IPTG (**17**).

Subsequent large scale purification was carried out based on a previously reported approach[84], adapting the buffer conditions to optimise the purification process (by removing the need for buffer exchange steps). Purification of vCPH from cell lysate consisted of three chromatographic steps: Ni^{II} affinity chromatography, followed by size exclusion chromatography (SEC, Superdex 75 resin) and anion exchange chromatography (Q Sepherose resin).

Two peaks were observed during SEC and despite the presence of vCPH in both peaks, impurities (possibly arising due to aggregation or degradation) at 15 kDa and 70 kDa were observed by SDS-PAGE (Figure 44). The second SEC peak (peak B), which did not contain any of the 70 kDa impurity was pooled and taken forward for further purification *via* anion exchange chromatography, resulting in high purity protein (greater than 95% as assessed by SDS-PAGE). For protocols that required a specific buffer for downstream analyses, the protein was then buffer exchanged following the final anion exchange step by dialysis.

For the production of *apo*-vCPH (*i.e.* without any co-purified metal ions present), the enzyme was treated with EDTA (50 mM, 2 hours) between the SEC and anion exchange steps to remove any co-purifying metal ions. The molecular weight of purified, *N*-terminally His_6-tagged protein was subsequently confirmed by whole-protein MS.

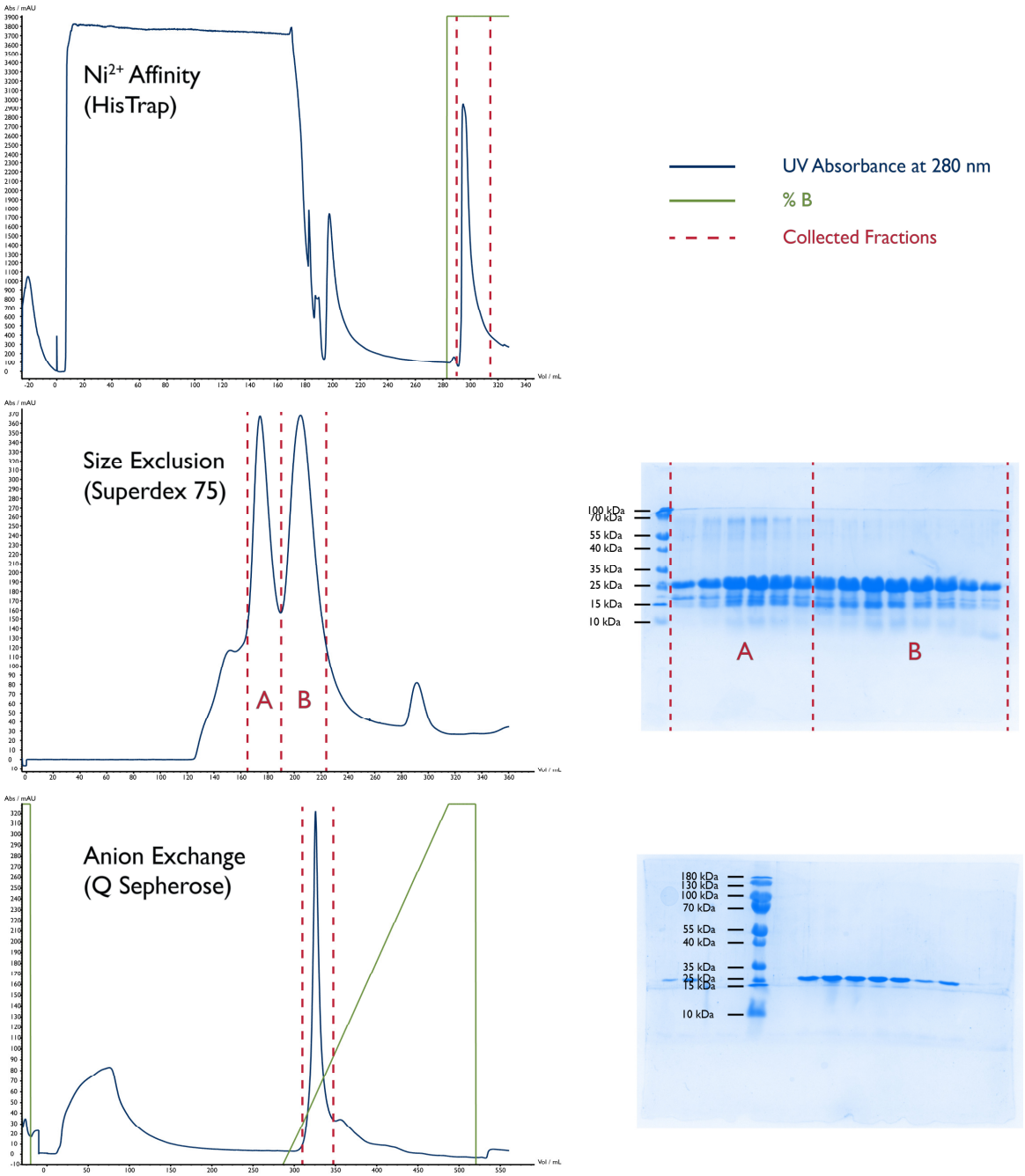


Figure 44 - Purification of vCPH. Following Ni^{II} affinity purification, the collected fractions were purified by SEC before Peak B was selected and further purified by anion exchange chromatography. SDS-PAGE gels are shown following SEC and anion exchange steps.

Enzyme Activity Assays

vCPH activity was assayed by monitoring the enzyme-dependent hydroxylation of peptide substrates. This was routinely achieved by analysing vCPH-treated peptide samples *via* MALDI-TOF MS. This approach allows for direct monitoring of product formation and has relatively low peptide and substrate requirements. This is in contrast to radioactive 2OG turnover assays, which were first used to characterise vCPH[179]. In addition to having higher peptide and substrate requirements, radioactive 2OG turnover assays are also unable to account accurately for uncoupled 2OG turnover (*i.e.* enzymatically-catalysed oxidative decarboxylation of 2OG which does not result in product formation), a process which is often increased in the presence of substrate[86].

vCPH Catalyses the Incorporation of Oxygen Atoms Into a Substrate Peptide

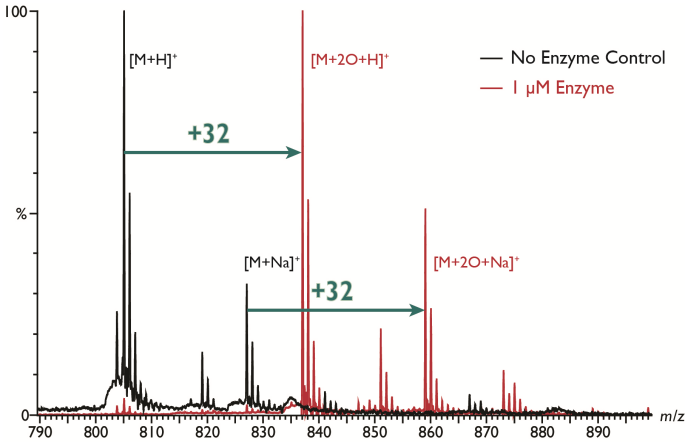


Figure 45 - Overlaid mass spectra showing emergence of peaks corresponding to enzymatically dependent incorporation of two oxygen atoms into Peptide 2. This is consistent with two vCPH-catalysed hydroxylation events.

Enzymatic activity was confirmed *via* the incubation of purified vCPH (1 µM) with 2OG (**1**, disodium salt, 5 mM), Fe^{II} (as $(NH_4)_2Fe(SO_4)_2$, 50 µM), sodium-L-ascorbate (**3**, 1 mM) and a previously reported peptide substrate PAPKPAPK (Peptide 2, 100 µM), on a 20 µL scale. Following analysis of the peptide by MALDI-TOF MS, peaks with Δ*m/z* of +32 compared to both the $[M+H]^+$ and $[M+Na]^+$ species versus a no enzyme control were observed, whilst starting material peaks were no longer present (Figure 45). This observation is consistent with

vCPH-catalysed incorporation of two oxygen atoms into Peptide 2, and by implication, double hydroxylation of Peptide 2.

vCPH Catalyses the Formation of (*2S,4R*)-Hydroxyproline

Previous studies have identified that hydroxyproline is produced during vCPH catalysis[179], but have not identified the stereochemistry of the resultant hydroxyproline. Since prolyl hydroxylases that catalyse the formation of 4-[90,112] and 3-hydroxyproline[74,121] are known, studies to define the stereochemistry of hydroxylation, along with identification of the residues in Peptide 2 that are modified, were undertaken. The target residues in Peptide 2 were investigated using MS-MS (Figure 46) and the stereochemistry of vCPH-catalysed hydroxylation was investigated using amino acid analysis (Figure 47).

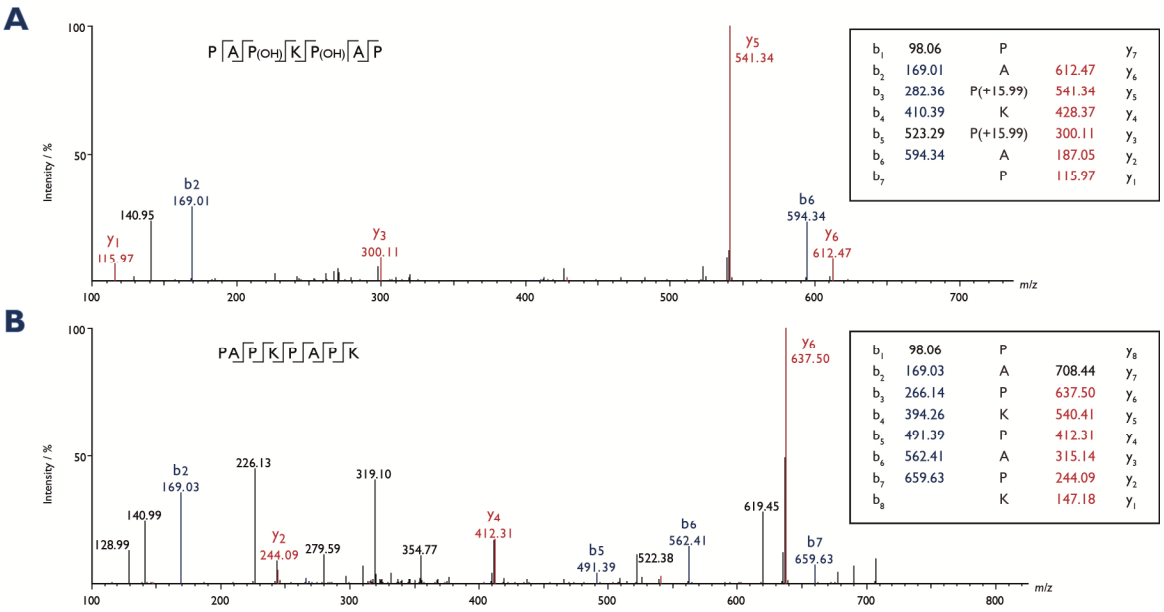


Figure 46 - MS-MS analysis of Peptide 2. Product ion spectra from ions corresponding to peptides of interest are shown. Red indicates the presence of a y-ion and blue the presence of a b-ion. Tables of predicted ions, coloured if they were observed, are inset. A - Product ion spectrum from a precursor ion with *m/z* = 355.1977, corresponding to $[M+2H]^{2+}$ of di-hydroxylated Peptide 2, following loss of the *C*-terminal lysine residue in the first stage of analysis. B - Product ion spectrum from a precursor ion with *m/z* = 403.2505, $[M+2H]^{2+}$ of unreacted Peptide 2.

The *in vitro* reaction was conducted on a 100 µL scale; following the completion of the reaction, as assayed by MALDI-TOF MS, the apparently di-hydroxylated product peptide was purified from enzyme and cofactors/co-substrates by SEC using a Superdex Peptide resin. MS-MS of the hydroxylated peptide confirmed that vCPH-catalysed hydroxylation of Peptide 2 occurs at both

proline 3 and proline 5 (Figure 46). There was no evidence for the hydroxylation of any other residue.

A portion of the purified product peptide was acid-hydrolysed and amino acid analysis of hydrolysed peptide was then used to investigate the regio- and stereo-chemistries of vCPH-catalysed prolyl hydroxylation (Figure 47). When compared to amino acid standards, samples of vCPH-treated Peptide 2 sample show a peak corresponding to the mass of derivatised hydroxyproline at a retention time of 3 minutes in a 10 minute amino acid analysis method. Comparison to amino acid standards show this peak corresponds to (*2S,4R*)-hydroxyproline (Figure 47A). This peak was not observed in an identically treated no enzyme control sample (Figure 47B). This study confirmed the proposal that vCPH catalyses the formation of (*2S,4R*)-hydroxyproline (henceforth referred to as Hyp in sequences), previous studies had not definitively shown that was the case[84,179].

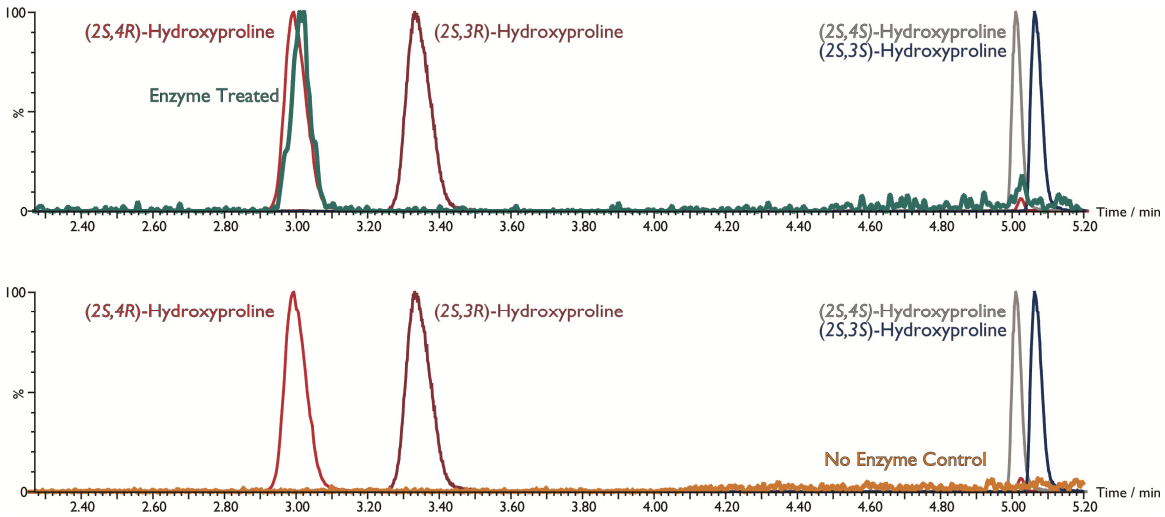


Figure 47 - Overlaid extracted ion chromatograms for *m/z* of 302, corresponding to the $[M+H]^+$ ion of derivatised hydroxyproline. Samples were analysed using the 10 minute amino acid analysis method.

Site Preference of Peptide 2 Hydroxylation

	Sequence
Peptide 2 (Parent)	PAPKPAPK
Peptide 3	Hyp-APKPAPK
Peptide 4	PA-Hyp-KPAPK
Peptide 5	PAPK-Hyp-APK
Peptide 6	PAPKPA-Hyp-K
Peptide 7	PA-Hyp-K-Hyp-APK

Table 4 - Peptides synthesised with sequential replacement of proline with (*2S,4R*)-hydroxyproline.

Previously, it was found that vCPH-mediated hydroxylation is blocked when a (*2S,4R*)-hydroxyproline residue is introduced at a target residue[84]. Therefore a series of peptides were synthesised, where each proline residue of Peptide 2 was sequentially replaced by (*2S,4R*)-hydroxyproline (Table 4). Peptides were subsequently incubated with vCPH and cofactors/co-substrates prior to MS analysis (Figure 48).

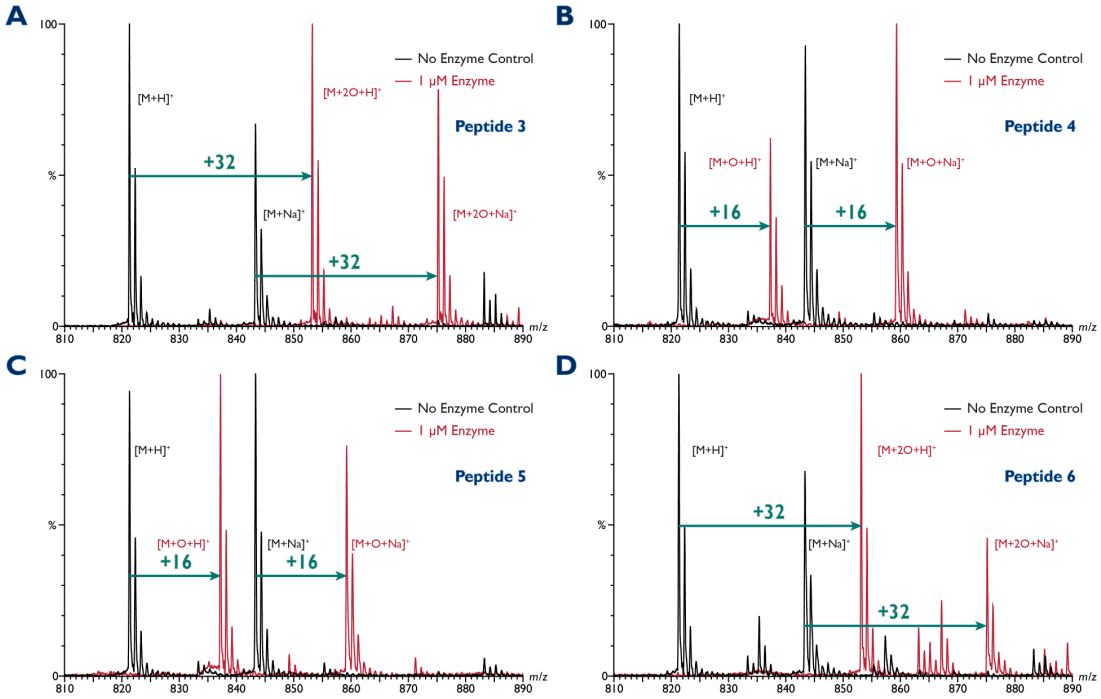


Figure 48 - Mass spectra of enzyme incubations with Peptide 3, Peptide 4, Peptide 5 and Peptide 6; no enzyme controls are overlaid. Double hydroxylation is still observed with Peptide 3 and Peptide 6, whereas replacement of the target proline residues of Peptide 2 with (*2S,4R*)-hydroxyproline (Peptide 4 and Peptide 5) results in substrates that are mono-hydroxylated.

The replacement of prolines at position 1 (Peptide 3) and 8 (Peptide 6) of the peptide did not prevent vCPH-catalysed double hydroxylation. Replacement of the substrate proline residues at positions 3 (Peptide 4) and 5 (Peptide 5) of Peptide 2 with (*2S,4R*)-hydroxyproline resulted in a

singly hydroxylated product when incubated with vCPH (Figure 48). In addition, no activity was observed between vCPH and a peptide where both proline residues at positions 3 and 5 were replaced by hydroxyproline (PA-Hyp-K-Hyp-APK, Peptide 7). These studies therefore suggest that, consistent with MS-MS analysis (Figure 46), vCPH-catalysed hydroxylation occurs selectively at proline 3 and proline 5 of Peptide 2.

Since Peptide 4 and Peptide 5 only underwent a single vCPH-catalysed hydroxylation, these two peptides potentially allowed for a more robust quantitative analysis of vCPH activity. By analysing product formation by MALDI-TOF MS, the linear range of hydroxylation was determined to allow for the determination of K_M and k_{cat} values for vCPH-catalysed hydroxylation using Peptide 4 and Peptide 5.

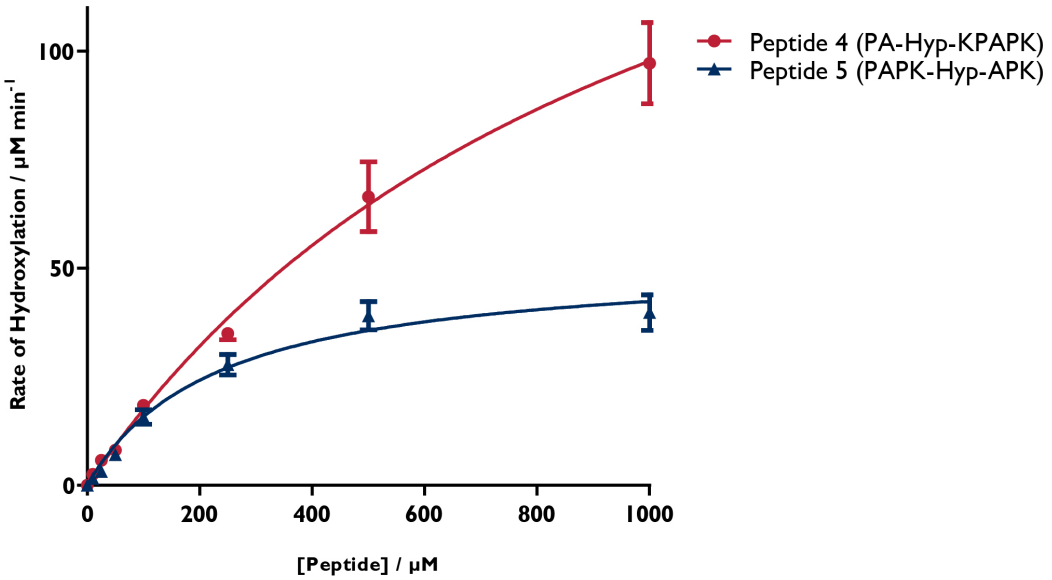


Figure 49 - Initial rates of vCPH-catalysed hydroxylation of Peptide 4 or Peptide 5 over a range of peptide concentrations. Peptides were incubated with vCPH (1 µM), 2OG (**1**, as the disodium salt, 5 mM), $(NH_4)_2Fe(SO_4)_2$ (50 µM) and sodium-L-ascorbate (**3**, 1 mM) and analysed using MALDI-TOF MS.

Peptide 4 was found to be more efficiently hydroxylated than Peptide 5 (Figure 49). However, a maximal rate of hydroxylation was not observed for Peptide 4, hence a fully quantitative analysis is not possible in this instance; for Peptide 5 K_M = 230.4 µM ± 45.1 µM and k_{cat} = 52.1 s^{-1} ± 3.7 s^{-1}.

No maximal rate was observed for vCPH-mediated hydroxylation of Peptide 4. This may, in part, be due to efficient binding of substrate and dissociation of the product, and may reflect the fast

kinetics observed in early experiments with the C-P4Hs (*via* monitoring of radioactive 2OG turnover)[89]. The observation that Peptide 4 is more efficiently hydroxylated than Peptide 5 is consistent with previous studies which have suggested that vCPH preferentially hydroxylates the proline at position 5[179] of Peptide 2 (PAPKPAPK).

Fe^{II}, 2OG and Ascorbate Dependence of vCPH-Catalysed Hydroxylation

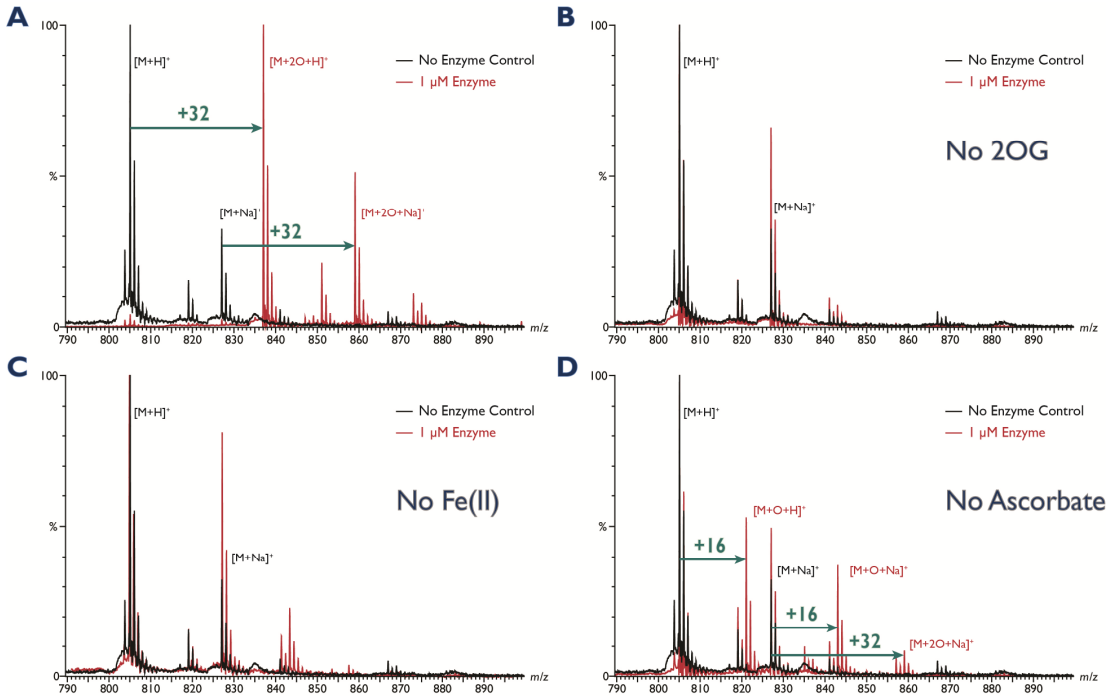


Figure 50 - Mass Spectra showing the effect of removal of cofactors/co-substrates. Enzyme-treated reactions (red) are overlaid with no enzyme controls (black). A concurrently run positive control is indicated in (A), emergence of peaks with Δ*m/z* of +32 is indicate of complete double hydroxylation of Peptide 2. Upon removal of 2OG (**1**) and Fe^{II}, enzymatic activity is lost. Removal of ascorbate does not prevent vCPH mediated hydroxylation. However, catalytic efficiency is reduced in the absence of ascorbate (D); ions with Δ*m/z* of +16 were observed, which correspond to a single hydroxylation.

In order to investigate the cofactor and co-substrate dependence of vCPH catalysis, an *in vitro* hydroxylation assay was then conducted using Peptide 2 with the systematic removal of Fe^{II}, 2OG (**1**) and ascorbate (**3**, Figure 50). All reactions were conducted for 14 hours at 25°C, prior to analysis of the resulting reaction mixture by MALDI-TOF MS. The study revealed that 2OG (**1**) and Fe^{II} are required for catalysis (Figure 50B and C). Hydroxylation of Peptide 2 was still observed in the absence of ascorbate (**3**); however, the extent of hydroxylation was substantially

reduced, with a peak with Δ*m/z* of +16, corresponding to the singly hydroxylated product, being observed after 16 hours incubation (Figure 50D).

A time-course study was then conducted in order to further probe the role of ascorbate in promoting vCPH catalysis (Figure 51). When ascorbate is present in the reaction mixture, following an initial build-up of mono-hydroxylated product over the first few minutes, catalysis proceeds efficiently to complete hydroxylation to the di-hydroxylated product (Figure 51A). In the absence of ascorbate, catalysis proceeds to the mono-hydroxylated product, which accumulates to greater levels than is observed when ascorbate is present (where the mono-hydroxylated product is efficiently turned over). These results imply that the ascorbate dependence of vCPH catalysis is substrate dependent.

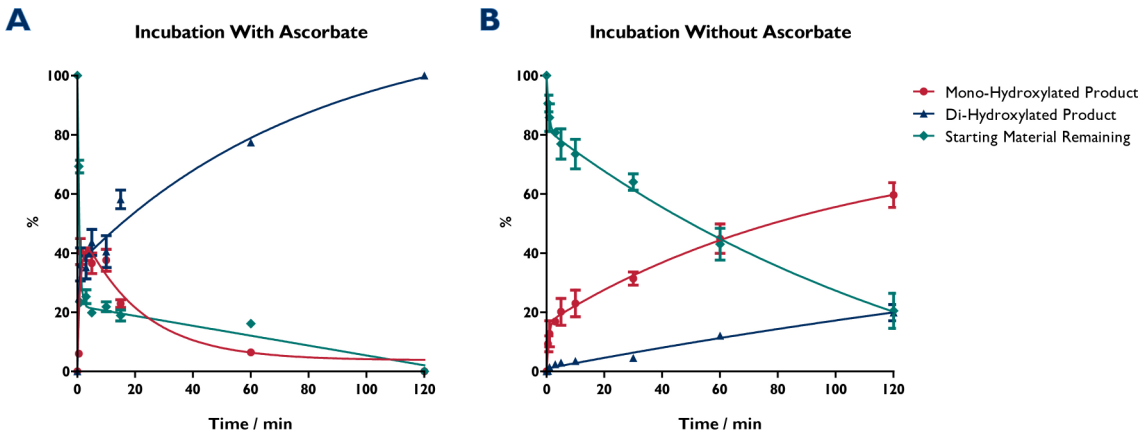


Figure 51 - Time course analysis of vCPH-catalysed hydroxylation of Peptide 2 in the presence and abscence of ascorbate. Relative abundances of starting material, mono-hydroxylated product and di-hydroxylated product are plotted as a function of time.

In order to elucidate the order of the two hydroxylations, an assay was conducted in the absence of ascorbate and the resultant peptide was separated from enzyme and cofactors/co-substrate by SEC. MS-MS analysis of the mono-hydroxylated product revealed that the first vCPH-catalysed hydroxylation of Peptide 2 takes place at P5 (Figure 52). This is consistent with the kinetics of Peptide 4 hydroxylation when compared to Peptide 5 (Figure 49).

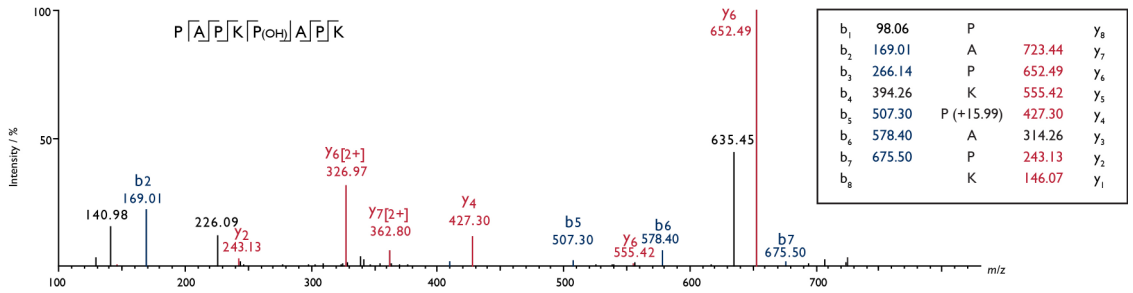


Figure 52 - Product ion spectrum from a precursor ion with *m/z* = 410.7562, corresponding to $[M+2H]^{2+}$ of mono-hydroxylated Peptide 2. Red indicates the presence of a y-ion and blue the presence of a b-ion. A table of predicted ions, coloured if they were observed, are inset.

Peptides that undergo only a single hydroxylation, Peptide 4 and Peptide 5, were then used for the quantitative determination of the substrate cofactor and co-substrate dependence of vCPH-catalysed hydroxylation (Figure 53). The values for K_M^{app} for Fe^{II} and K_M for 2OG (**1**) were within error of one another; the K_M^{app} for Fe^{II} was 0.9 µM ± 0.1 µM for Peptide 4 and 1.1 µM ± 0.1 µM for Peptide 5. The K_M values for 2OG (**1**) were 29.5 µM ± 8.7 µM and 27.4 µM ± 6.1 µM for Peptide 4 and Peptide 5 respectively.

However, the K_M^{app} for ascorbate (**3**) differs significantly between the two peptides (K_M^{app} for ascorbate (**3**) was 1.7 µM ± 0.3 µM for Peptide 4 and 13.8 µM ± 4.4 µM for Peptide 5), and by implication, the two different hydroxylation sites of Peptide 2. The differing ascorbate values are consistent with the build-up of mono-hydroxylated Peptide 2 product that was observed in time course (Figure 51) and MS-MS analysis of this mono-hydroxylated product (Figure 52).

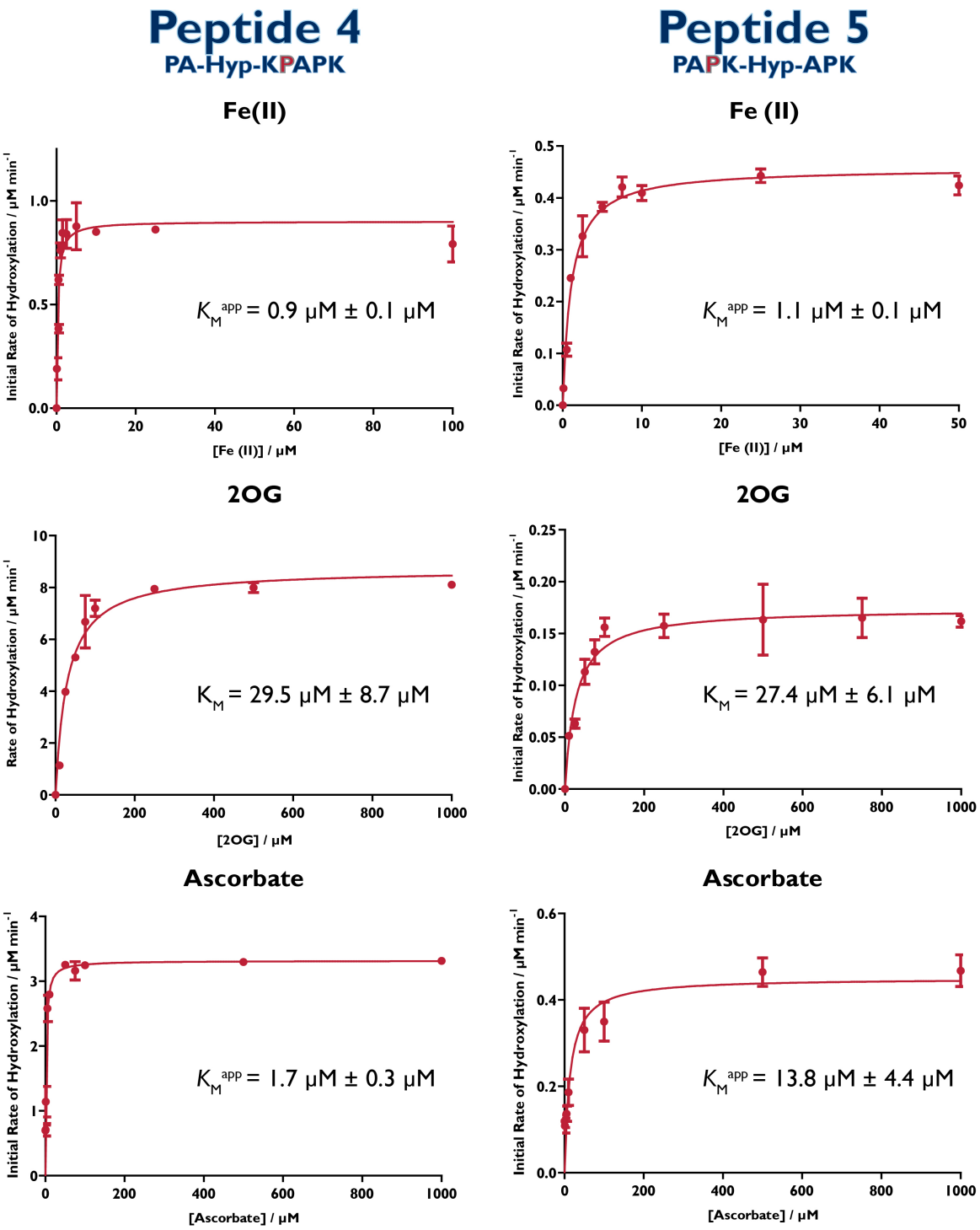


Figure 53 - Cofactor and co-substrate dependence of vCPH-catalysed hydroxylation of Peptide 4 (Left) and Peptide 5 (Right), sequences are indicated, with substrate prolyl residues highlighted in red. Similar kinetic parameters for Fe^{II} and 2OG (**1**) were observed, where as the K_M^{app} for ascorbate (**3**) differed tenfold.

2OG Turnover

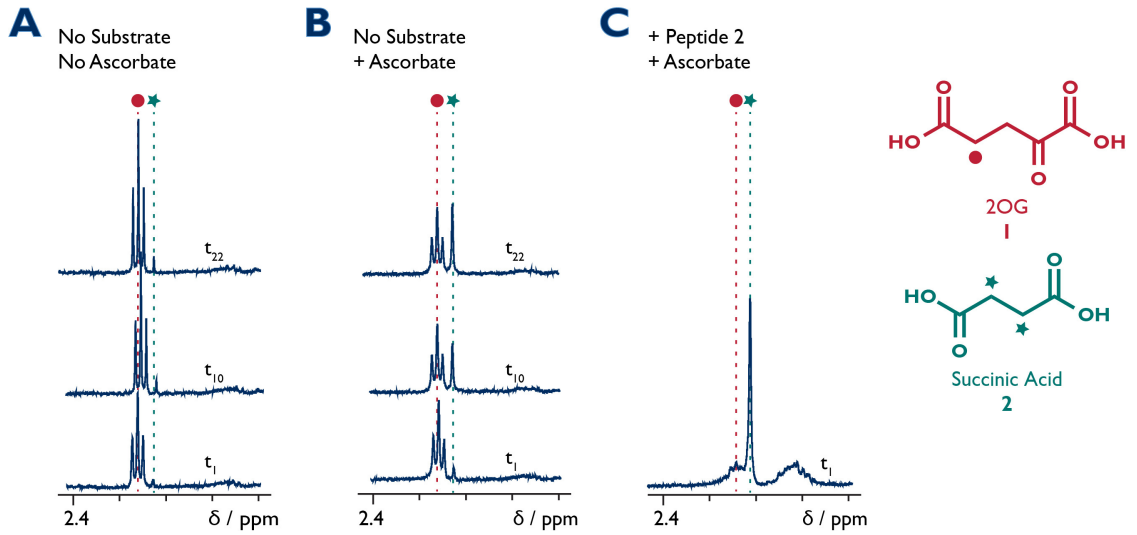


Figure 54 - ^{1}H NMR spectra showing turnover of 2OG (**1**, 600 µM) to succinate in the presence of vCPH (20 µM) and $(NH_4)_2Fe(SO_4)_2$ (50 µM) in 50 mM *tris*(hydroxymethyl)aminomethane (Tris) -D_{11}, pH 7.5, in 90% H_2O and 10% D_2O. ^{1}H resonances corresponding to 2OG (triplet peak indicated with a red circle) and succinate (singlet peak indicated with a green star); the position of the appropriate hydrogens are indicated on the structures of 2OG and succinate (the acid forms are depicted). A - Monitoring of uncoupled 2OG turnover in the presence of sodium-L-ascorbate (1 mM). B - Monitoring of uncoupled 2OG turnover in the absence of sodium-L-ascorbate. C - Monitoring of uncoupled 2OG turnover in the presence of sodium-L-ascorbate (1 mM) and Peptide 2 (500 µM), nearly all 2OG has been turnover at the first acquisition timepoint (t = 7.2 minutes).

^{1}H NMR spectroscopy was used to monitor the conversion of 2OG (**1**) to succinate (**2**) during vCPH catalysis. When vCPH was incubated with Fe^{II} and 2OG (**1**) but without substrate, 2OG (**1**) was turned over to succinate (**2**) at a relatively slow rate (*i.e.* uncoupled 2OG turnover was monitored). The rate of uncoupled 2OG turnover was enhanced in the presence of ascorbate (**3**, Figure 54A/B and Figure 55). In the presence of a vCPH substrate peptide, Peptide 2, the rate of 2OG turnover is significantly increased and the reaction rapidly proceeds to completion (Figure 54C).

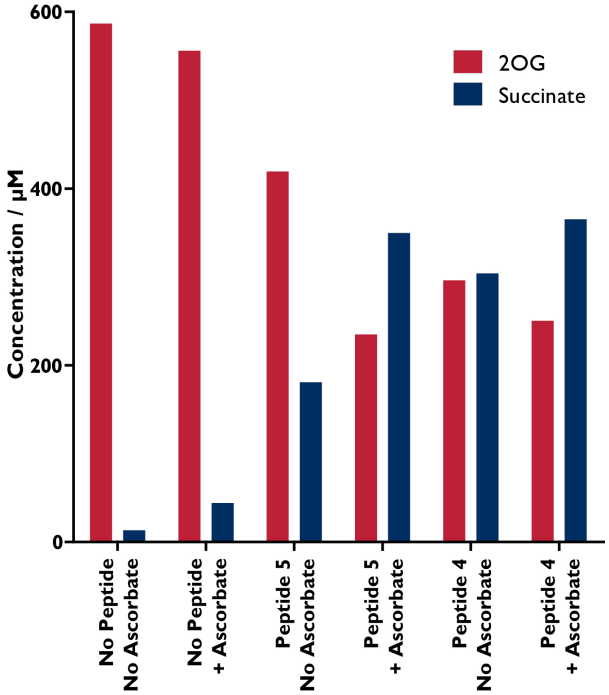


Figure 55 - Concentrations of 2OG (**1**, red) and succinate (**2**, blue) at timepoint 4 (t ≈ 13 minutes) of a vCPH (20 µM) incubation with peptide (500 µM), $(NH_4)_2Fe(SO_4)_2$ (50 µM), 2OG (disodium salt, Initial [2OG] = 600 µM) and, if appropriate, sodium-L-ascorbate (1 mM). ^{1}H NMR spectroscopy was used to monitor vCPH-catalysed 2OG turnover.

Analysis of 2OG turnover after 13 minutes, in the presence of Peptide 4 or Peptide 5 (Figure 55), shows that in the presence of ascorbate (**3**), 2OG turnover is broadly equal. In the absence of ascorbate (**3**), but in the presence Peptide 4, 2OG is turned over to a similar degree as when ascorbate is present, albeit slightly less efficiently than when ascorbate is present. However, in the presence of Peptide 5 and in absence of ascorbate (**3**), 2OG turnover is lower (but well above levels of uncoupled turnover) than the equivalent turnover in the presence of ascorbate. Consistent with MS-based kinetic studies (Figure 53), these data indicate that the ascorbate requirement for vCPH-catalysed hydroxylation of Peptide 5 is greater than Peptide 4 and therefore that the ascorbate requirement for vCPH-mediated hydroxylation is sequence dependent.

Cofactor and Co-substrate Binding to vCPH

Differential Scanning Fluorometry

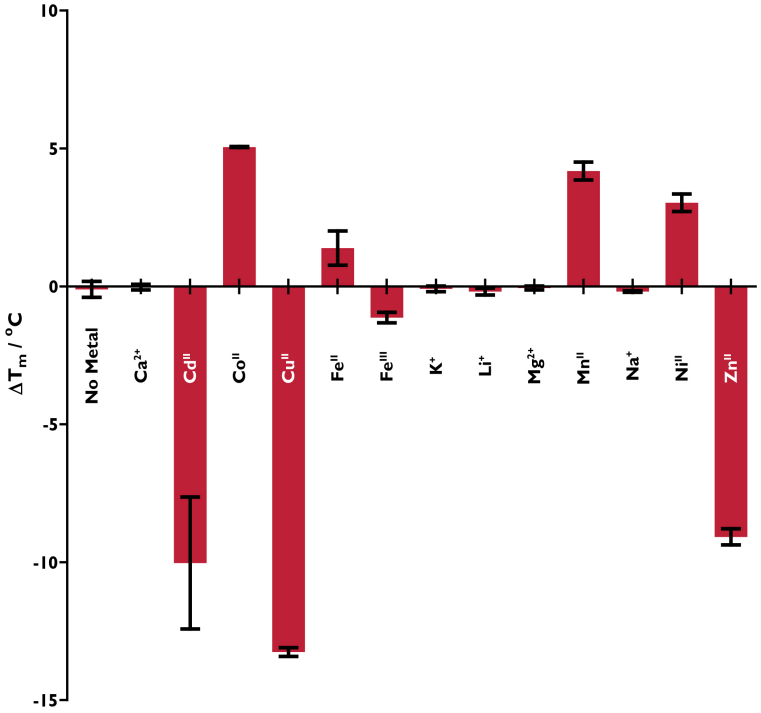


Figure 56 - Change in T_m for *apo*-vCPH supplemented with a range of metal ions. Positive values indicate an increase in stability of the enzyme and imply binding of the metal ion.

In order to investigate cofactor binding and to inform future structural studies, the binding of different metal ions to *apo*-vCPH was assessed by differential scanning fluorometry, *via* analysis of changes in the melting temperature (T_m) of vCPH in the presence of different metal ions (Figure 56). The protein was stabilised (*i.e.* a positive ΔT_m) in the presence of Mn^{II} (+4.18°C ± 0.32°C), Ni^{II} (+3.03°C ± 0.32°C), Fe^{II} (+1.39°C ± 0.62°C) and Co^{II} (+5.05°C ± 0.02°C), consistent with observations with other 2OG oxygenases[180] and indicative of metal binding to the protein, presumably through co-ordination to the metal binding histidine and aspartic acid residues (H152, D154 and H221) of the facial triad (binding to Mn is shown in Figure 42B).

Isothermal Titration Calorimetry

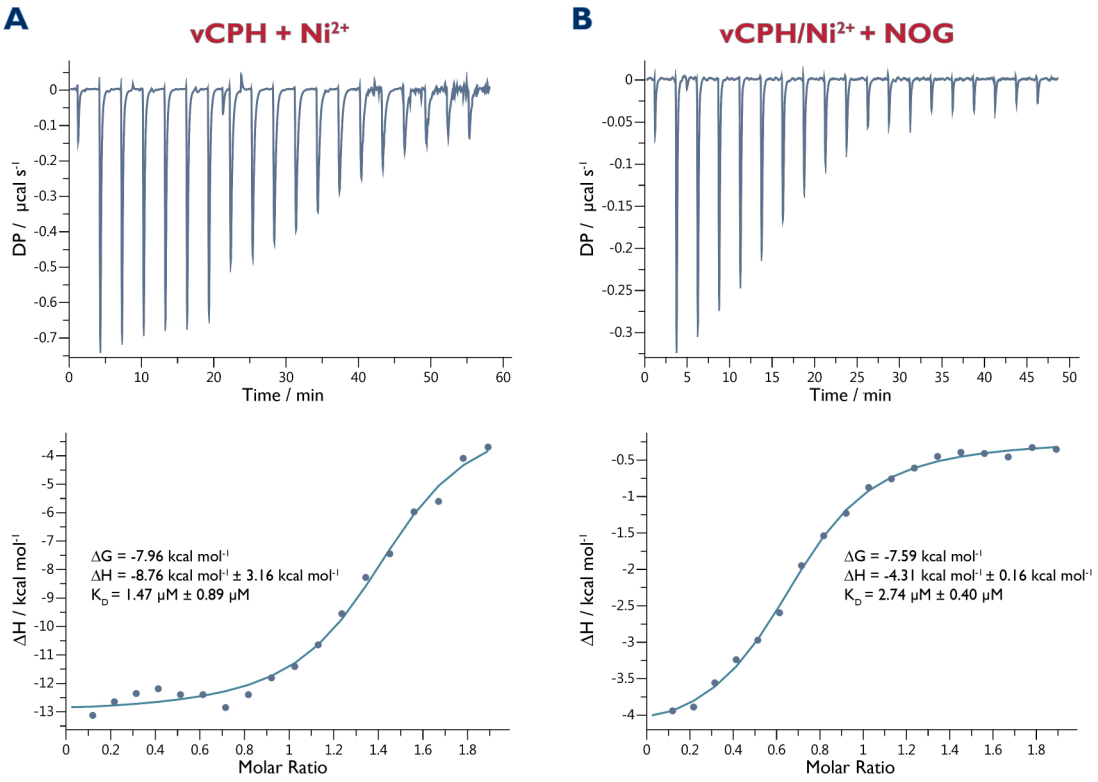


Figure 57 - Results of isothermal calorimetry experiments to assess the binding of vCPH with Ni^{II} (A) and NOG to the vCPH/Ni^{II} complex (B). Calculated binding and thermodynamic properties are indicated on the graph.

The binding of *apo*-vCPH to Mn^{II}, Ni^{II}, Fe^{II} and Co^{II} was then investigated by isothermal titration calorimetry (ITC). Evidence for binding to *apo*-vCPH was observed with Ni^{II} (Figure 57A), Ni^{II} was found to bind with high affinity, K_D = 1.47 µM ± 0.89 µM. ITC experiments that investigated the binding of Mn^{II} to *apo*-vCPH yielded poor quality data and no thermodynamic data could be extracted from these experiments (Figure 58A), this is similar experiments with other 2OG oxygenases which have also proved to be technically challenging. No evidence for binding between *apo*-vCPH and Co^{II} was observed by ITC (Figure 58B).

Binding of NOG (**4**), an inactive 2OG analogue[136], to the vCPH/Ni^{II} complex was also analysed by ITC (Figure 57B), with a K_D of 2.74 µM ± 0.40 µM. This observation is consistent with strong binding of 2OG (**1**), which is suggested by a ΔT_m of +9.97°C ± 0.27°C when 2OG binds to the vCPH/Mn^{II} complex. High quality ITC data was not obtained for 2OG binding to the vCPH/Ni^{II} complex.

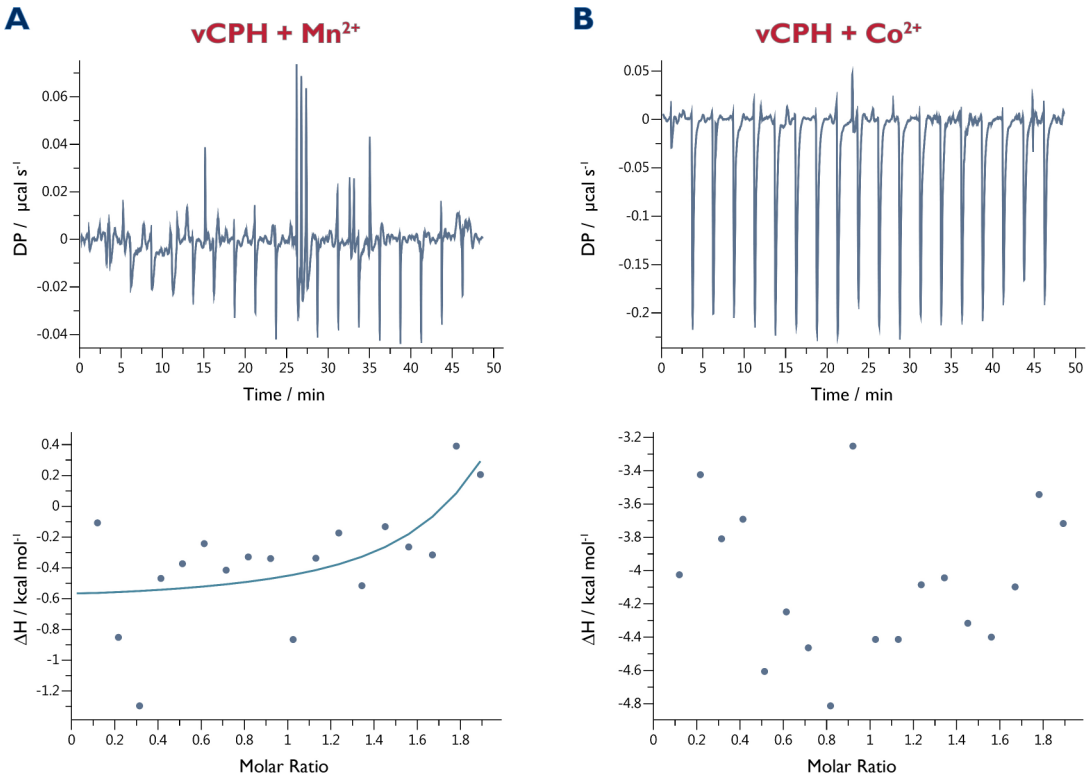


Figure 58 - Representative ITC data for the binding of *apo*-vCPH to Mn^{II} (A) and Co^{II} (B). No binding to Co^{II} was observed

Substrate Selectivity of vCPH

To investigate the sequence selectivity of vCPH, and in part to assess the potential of vCPH as a model for C-P4H, a range of other possible peptide substrates were screened for activity with vCPH *in vitro*.

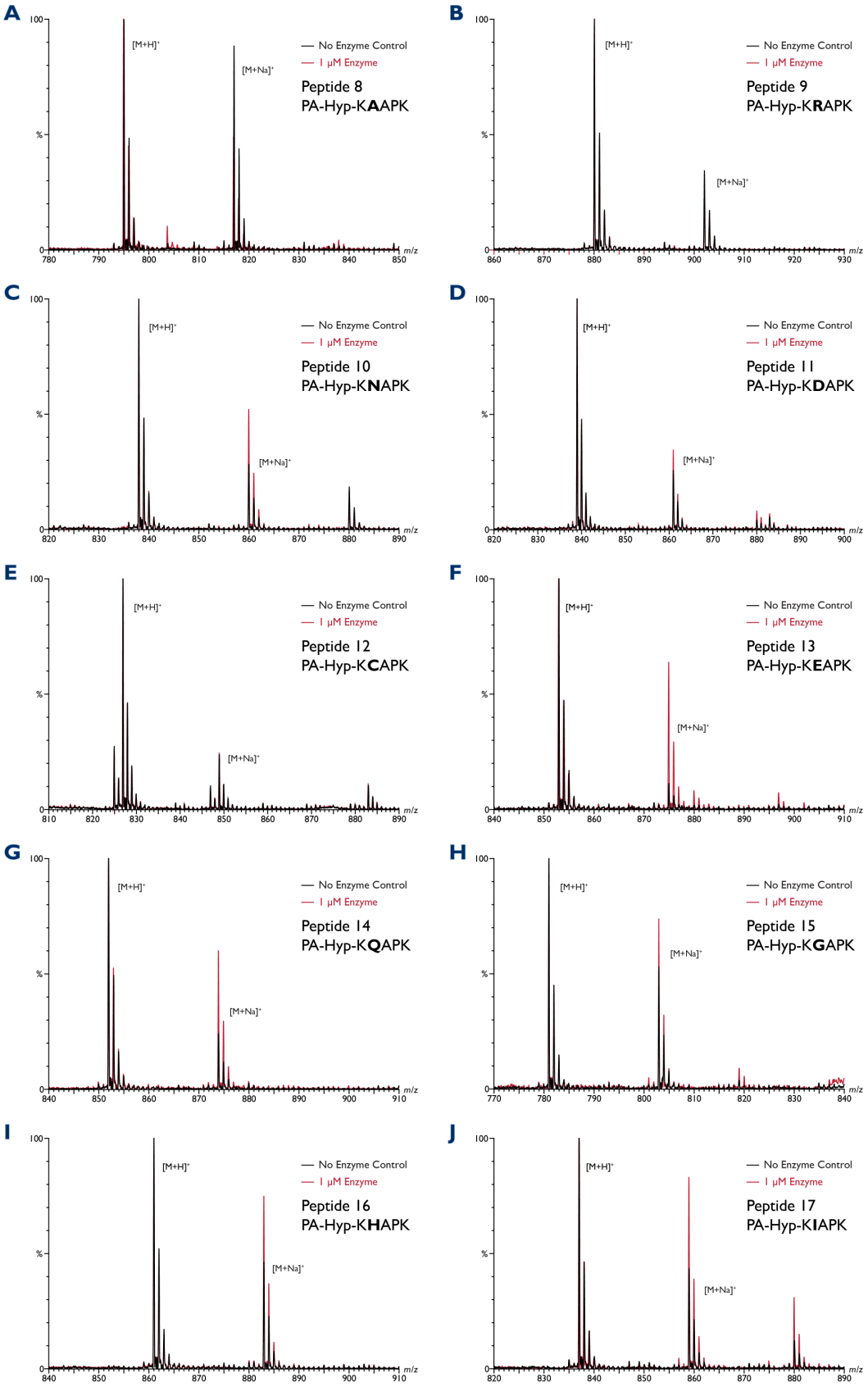
vCPH Specifically Hydroxylates Prolyl Residues

Using the mono-hydroxylated peptide substrate Peptide 4 as a template, a library of peptides with the general sequence PA-Hyp-KXAPK were synthesised on a small scale. Each peptide featured a different proteinogenic amino acid at the X position (Table 5) and the peptides were then tested as vCPH substrates in order to test if vCPH accepts alternative residues as a substrate (this type of substrate promiscuity is exhibited by FIH[123]). Peptide 4 was selected as a template for the peptide library since the peptide only undergoes a single vCPH-catalysed modification and target residue is more efficiently hydroxylated by vCPH than Peptide 5 (Figure 49).

Following successful solid-phase peptide synthesis, the peptides were incubated with vCPH and saturating cofactors/co-substrates and reactions were analysed by MS (Figure 59). A reaction was only observed when X = proline (Figure 59T). vCPH therefore appears to be highly selective for proline residues amongst proteinogenic amino acids, supportive of the assignment of vCPH as a prolyl hydroxylase.

	Sequence
Peptide 8	PA-Hyp-K**A**APK
Peptide 9	PA-Hyp-K**R**APK
Peptide 10	PA-Hyp-K**N**APK
Peptide 11	PA-Hyp-K**D**APK
Peptide 12	PA-Hyp-K**C**APK
Peptide 13	PA-Hyp-K**E**APK
Peptide 14	PA-Hyp-K**Q**APK
Peptide 15	PA-Hyp-K**G**APK
Peptide 16	PA-Hyp-K**H**APK
Peptide 17	PA-Hyp-K**I**APK
Peptide 18	PA-Hyp-K**L**APK
Peptide 19	PA-Hyp-K**K**APK
Peptide 20	PA-Hyp-K**M**APK
Peptide 21	PA-Hyp-K**F**APK
Peptide 22	PA-Hyp-K**S**APK
Peptide 23	PA-Hyp-K**T**APK
Peptide 24	PA-Hyp-K**W**APK
Peptide 25	PA-Hyp-K**Y**APK
Peptide 26	PA-Hyp-K**V**APK
Peptide 4	PA-Hyp-K**P**APK

Table 5 - Peptides synthesised to investigate the selectivity of vCPH catalysed reactions.


A
[M+H]+
[M+Na]+
No Enzyme Control
1 µM Enzyme
Peptide 8
PA-Hyp-KAAPK
B
[M+H]+
[M+Na]+
No Enzyme Control
1 µM Enzyme
Peptide 9
PA-Hyp-KRAPK
C
[M+H]+
[M+Na]+
No Enzyme Control
1 µM Enzyme
Peptide 10
PA-Hyp-KNAPK
D
[M+H]+
[M+Na]+
No Enzyme Control
1 µM Enzyme
Peptide 11
PA-Hyp-KDAPK
E
[M+H]+
[M+Na]+
No Enzyme Control
1 µM Enzyme
Peptide 12
PA-Hyp-KCAPK
F
[M+H]+
[M+Na]+
No Enzyme Control
1 µM Enzyme
Peptide 13
PA-Hyp-KEAPK
G
[M+H]+
[M+Na]+
No Enzyme Control
1 µM Enzyme
Peptide 14
PA-Hyp-KQAPK
H
[M+H]+
[M+Na]+
No Enzyme Control
1 µM Enzyme
Peptide 15
PA-Hyp-KGAPK
I
[M+H]+
[M+Na]+
No Enzyme Control
1 µM Enzyme
Peptide 16
PA-Hyp-KHAPK
J
[M+H]+
[M+Na]+
No Enzyme Control
1 µM Enzyme
Peptide 17
PA-Hyp-KIAPK
m/z

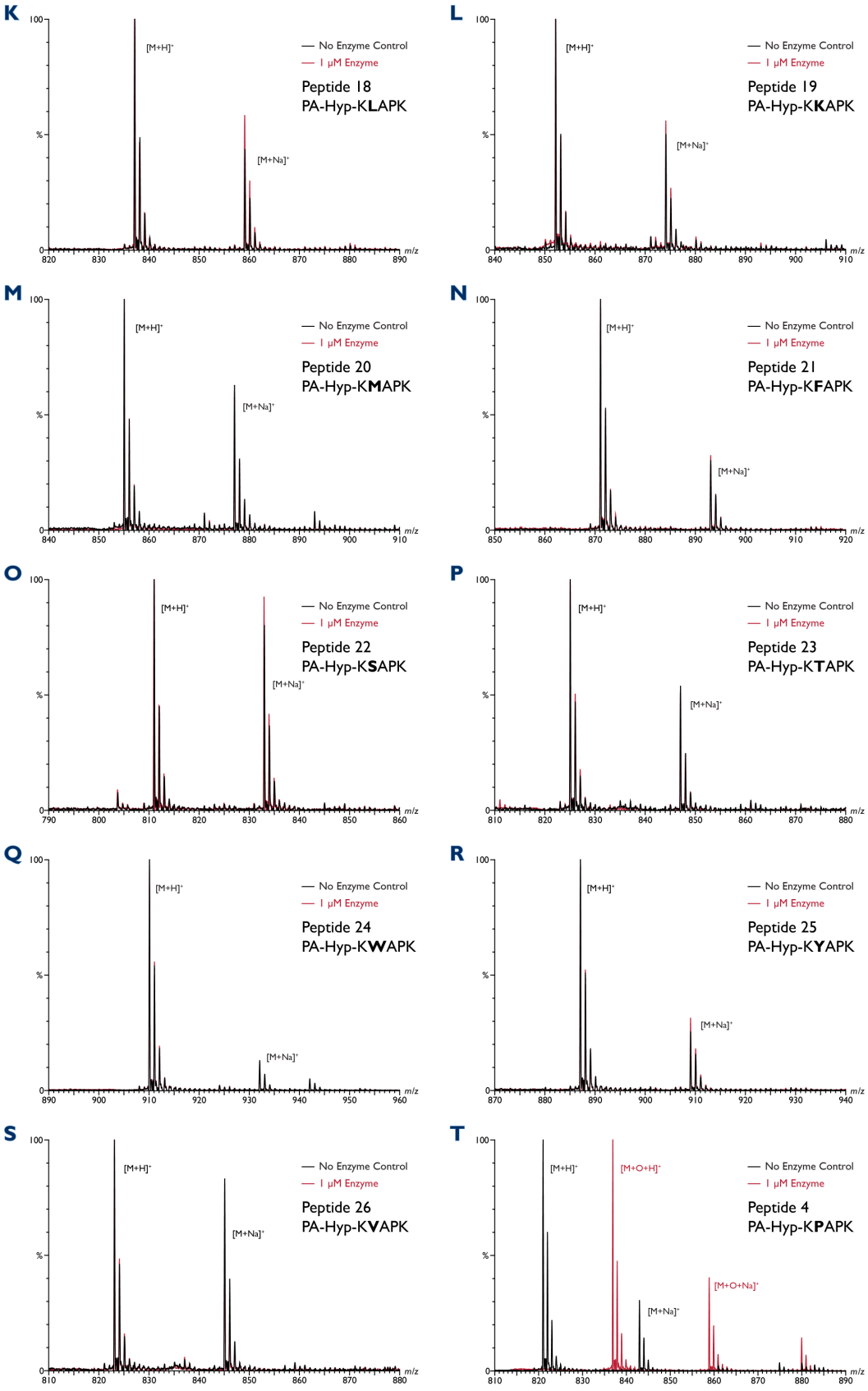


Figure 59 - Overlaid mass spectra of peptides (100 µM) incubated with vCPH (1 µM), 2OG (disodium salt, 500 µM), $(NH_4)_2Fe(SO_4)_2$ (50 µM) and sodium-L-ascorbate (1 mM). The spectra corresponding to enzyme incubations are indicated in red and a no enzyme control in black. vCPH-catalysed reaction was only observed with a proline residue at the substrate position 5 (T).

vCPH Catalyses Hydroxylation of (*2S,4S*)-Fluoroproline, Leading to F^- release

Previous studies with another 2OG oxygenase, BBOX, have utilised a fluorinated substrate to monitor hydroxylation[181]; hydroxylation of a vicinal CH bond leads to subsequent release of F^-. It was hypothesised that such a route would be applicable to monitor vCPH catalysis; following hydroxylation of an appropriately fluorinated substrate analogue, ketone formation (Figure 60A) should lead to loss of F^-[181]. If successful, this approach could then be used to develop an inhibition assay that monitors F^- release as a measure of product formation or to use ^{19}F NMR spectroscopy to analyse peptide binding.

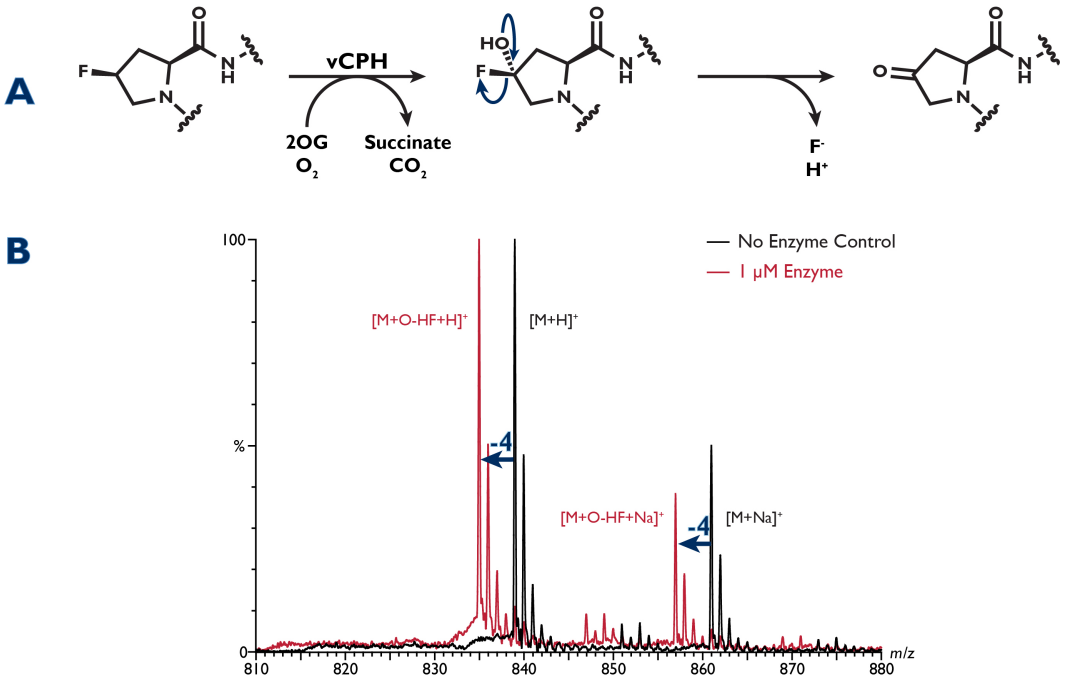


Figure 60 - Following vCPH-catalysed hydroxylation of (*2S,4S*)-fluoroproline containing Peptide 27, release of F^- is observed. A - Mechanism of F^- release following vCPH-mediated hydroxylation. B - Mass spectrum of Peptide 27 incubation with vCPH, overlaid with a no enzyme control; a Δ*m/z* of -4 was observed, consistent with F^- loss follwowing hydroxylation.

A peptide incorporating (*2S,4S*)-fluoroproline at position 5 of Peptide 4 was synthesised (Peptide 27, PA-Hyp-K-(*2S,4S*)-fluoroproline-APK) and tested as a vCPH substrate. Following vCPH-mediated hydroxylation, a Δ*m/z* of -4 between the observed enzyme-treated and no enzyme $[M+X]^+$ ions was observed by mass spectrometry (Figure 60, where X is H, Na, K *etc.*), consistent with hydroxylation followed by of release of F^-. F^- release was also observed by ^{19}F

NMR spectroscopy (Figure 61). The successful hydroxylation and F^- loss from Peptide 27 is further evidence that vCPH-catalyses the formation of (*2S,4R*)-hydroxyproline; this process is not possible with other fluoro-L-proline regio- or stereo-isomers.

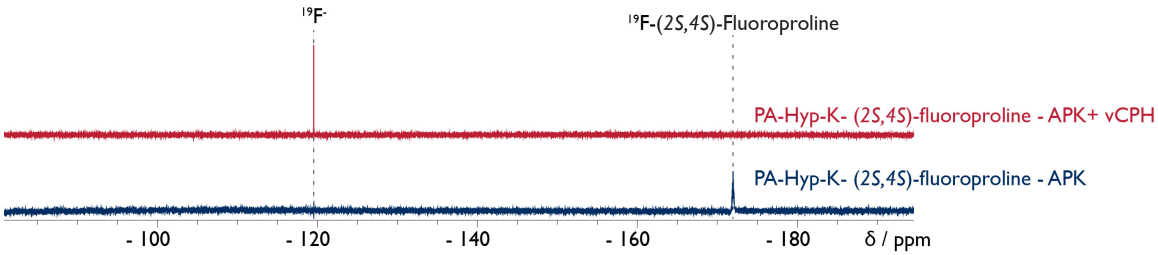


Figure 61 - ^{19}F NMR spectra showing release of F^- when Peptide 28 (500 µM) is incubated with vCPH (20 µM), $(NH_4)_2Fe(SO_4)_2$ (50 µM), 2OG (disodium salt, 600 µM) and sodium-L-ascorbate (1 mM) in 50 mM Tris-D_{11}, pH 7.5, in 90% H_2O and 10% D_2O. After 1 hour (red), a peak corresponding to F^- at -120 ppm was observed and the depletion of the peak corresponding to (*2S,4S*)-fluoroproline at -172 ppm. F^- release was confirmed by spiking the sample with NaF. Spectra were referenced to the internal standard trifluoroacetic acid (TFA) at -75.2 ppm.

Kinetic experiments revealed that, similar to Peptide 4, Peptide 27 was also efficiently hydroxylated by vCPH and the initial rate did not plateau when the peptide concentration was 1 mM (a 1000:1 substrate to enzyme ratio). The peptide was, however, not hydroxylated as efficiently as Peptide 4 (Figure 62).

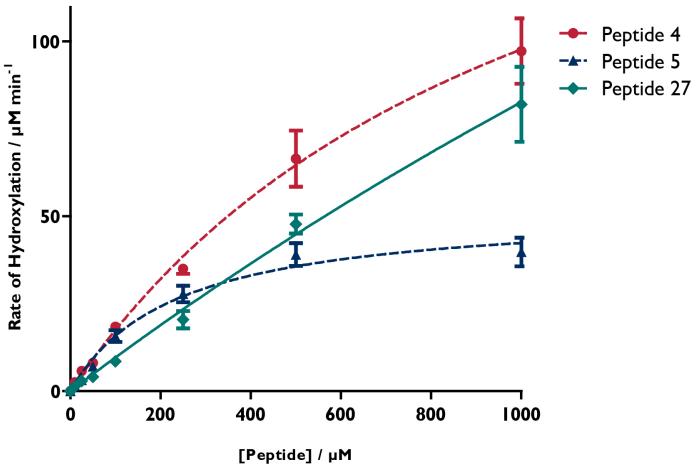


Figure 62 - Initial rate of vCPH-catalysed hydroxylation for Peptide 27 at a range of peptide concentrations. A maximal rate was not achieved and therefore K_M and k_{cat} were not determined. Curves for Peptide 4 and Peptide 5, which were analysed concurrently, are also indicated. The maximal rate of vCPH-catalysed hydroxylation of Peptide 27 is greater than Peptide 5 but less than Peptide 4.

2OG turnover studies with the Peptide 27 showed that, similar to Peptide 4 and Peptide 5, the rate of 2OG turnover is enhanced in the presence of ascorbate (Figure 63). Turnover at 13

minutes was comparable to Peptide 5, a poorer vCPH substrate than compared to Peptide 2 and Peptide 4; in the presence of Peptide 27 and after 13 minutes, vCPH turned over 315 µmoles of 2OG in the presence of ascorbate and 168 µmol in the absence of ascorbate. For Peptide 5 these values were 305 µmol and 181 µmol respectively (Figure 55). This, in combination with the observations with Peptide 5, suggests that ascorbate dependence of vCPH catalysis is greater for weaker substrates.

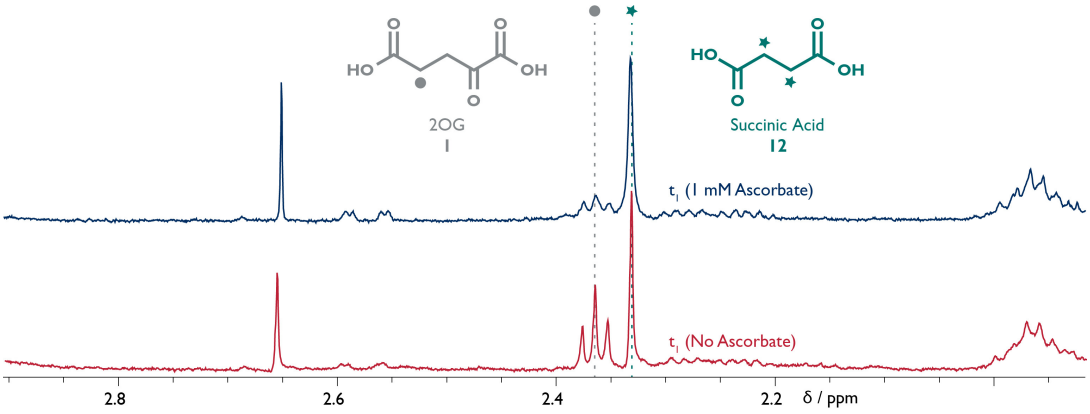


Figure 63 - ^{1}H NMR spectra of incubations of Peptide 27 (400 µM) with vCPH (20 µM), $(NH_4)_2Fe(SO_4)_2$ (50 µM) and 2OG (disodium salt, 600 µM) in 50 mM Tris-D_{11} pH 7.5, 10% D_2O at t_1 = 7 minutes. The triplet peak at (corresponding to 2OG) is decreased and singlet (corresponding to succinate) is correspondingly increased when the incubation was conducted in the presence of 1 mM sodium-L-ascorbate (blue spectrum).

Collagen Peptides

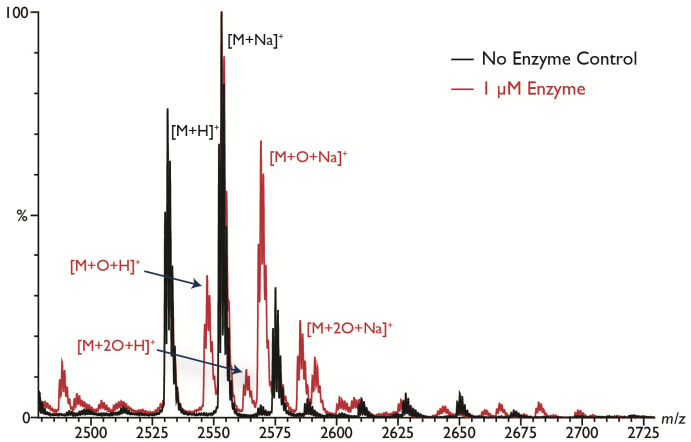


Figure 64 - Overlaid mass spectra showing the appearance of peaks with Δ*m/z* of +16 and +32 when Peptide 1 ($(GPP)_{10}$, 100 µM) was incubated with vCPH (1 µM), 2OG (disodium salt, 5 mM), $(NH_4)_2Fe(SO_4)_2$ (50 µM) and sodium-L-ascorbate (1 mM) for 4 hours. This indicates that vCPH accepts $(GPP)_{10}$ as a substrate.

In order to investigate whether vCPH is a *bone fide* collagen hydroxylase, a synthetic collagen repeat peptide, $(GPP)_{10}$ (Peptide 1) was tested as a potential vCPH substrate. The appearance of

peaks at Δ*m/z* of +16 and +32 versus a no enzyme control were consistent with hydroxylation reactions catalysed by vCPH. Peptide 1 was not hydroxylated as efficiently as Peptide 2 (or its derivatives). Since multiple hydroxylation products are formed (with multiple reactants), quantitative kinetic analyses of the hydroxylation of Peptide 1 are non-trivial.

HIF CODD/NODD Peptides

To test whether vCPH is able to behave like the HIF prolyl hydroxylases and hydroxylate proline residues in the oxygen-sensing regions of HIF-1, vCPH was incubated with cofactors, co-substrates and 19-mer peptides that correspond to the *N*-terminal (NODD, Peptide 29) and *C*-terminal (CODD, Peptide 30) oxygen degradation domains of HIF-1α. No reaction with either peptide was observed, even when the peptide:enzyme ratio was increased to 1:1. Additionally, Peptide 29 and Peptide 30 were tested for inhibition of vCPH in Chapter 4; no inhibition was observed, indicative of no binding of Peptide 29 or Peptide 30 to the vCPH active site.

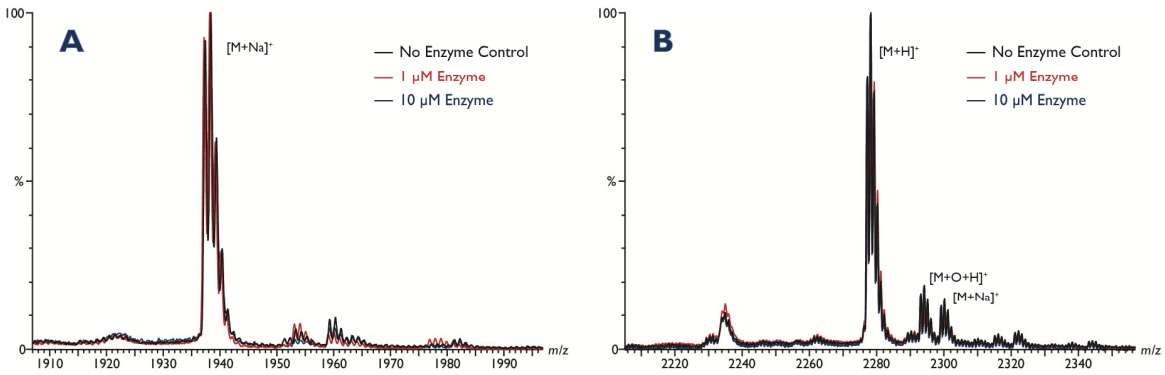


Figure 65 - Overlaid mass spectra showing no reaction between Peptide 29 and Peptide 30 (substrates for PHD1, PHD2 and PHD3) and vCPH. Peptide (100 μM) was incubated vCPH (1 μM or 10 μM), 2OG (disodium salt, 5 mM), $(NH_4)_2Fe(SO_4)_2$ (50 μM) and sodium-L-ascorbate (1 mM) A - Overlaid mass spectra for enzyme and non-enzyme treated NODD peptide (Peptide 29). B - Overlaid mass spectra for enzyme and non-enzyme treated CODD peptide (Peptide 30).

^{1}H NMR spectroscopy was used to monitor vCPH catalysed 2OG turnover in the presence of the CODD (Peptide 30) and NODD (Peptide 29) peptides. This showed that turnover in the presence of Peptide 29 and Peptide 30 was comparable to uncoupled turnover (Figure 66, in the case of CODD peptide the turnover was lower than uncoupled turnover). In comparison, peptide

substrates $(GPP)_{10}$ and PA-Hyp-KPAPK (Peptide 1 and Peptide 4) of vCPH enhanced the rate of 2OG turnover.

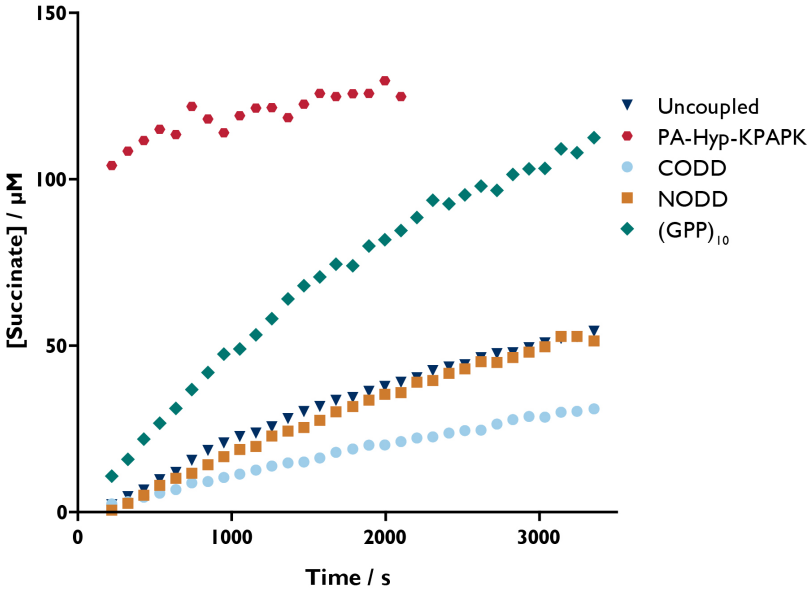


Figure 66 - vCPH-catalysed monitoring of 2OG turnover by ^{1}H NMR spectroscopy with peptides tested as alternative vCPH substrates. Succinate is produced at a greater rate in the presence of the substrate peptides $(GPP)_{10}$ and PA-Hyp-KPAPK (Peptide 1 and Peptide 4); CODD (Peptide 30) and NODD (Peptide 29) peptides, which are not substrates of vCPH, induce no enhancement of 2OG turnover compared to uncoupled turnover.

Conclusions

vCPH has been shown to catalyse the formation of (*2S,4R*)-hydroxyproline (Figure 47), and therefore catalyses the same transformation as the C-P4H enzymes[90] and the HIF prolyl hydroxylases[112]; this finding is consistent with previous assumptions as to the activity of vCPH[84,179]. The high similarity of the protein to the active site of the CPH enzymes, along with the substrate selectivity of the enzyme, means that vCPH can be a useful model hydroxylase, especially given the ease of recombinant production. Additionally, the small size of vCPH (vCPH is a 27 kDa monomer) potentially allows for studies of the enzyme, such as analysis of the protein by NMR spectroscopy, that would not be possible with a 240 kDa tetramer.

vCPH selectively catalyses the hydroxylation of prolyl residues and catalyses the hydroxylation of synthetic collagen repeat peptides (Figure 64) but not the substrates of the HIF prolyl hydroxylases (Figure 65). This activity is consistent with vCPH being a *bone fide* collagen prolyl

hydroxylase. The apparent high-selectivity of vCPH catalysis (Figure 59) contrasts with some other 2OG dependent hydroxylases, such as FIH, which accepts multiple residues as substrates[123].

vCPH is a highly active enzyme relative to some other 2OG dependent prolyl hydroxylases (such as the HIF prolyl hydroxylases). This is particularly apparent with the $(PAPK)_n$-type peptide substrates. The hydroxylation of Peptide 2 is rapidly catalysed by vCPH at two positions (prolines 3 and 5). Introduction of a (*2S,4S*)-fluoroproline residue at position 5 of Peptide 4 resulted in a peptide (Peptide 27) that was accepted by vCPH as a substrate, and releases F^- following vCPH-catalysed hydroxylation, as confirmed by NMR spectroscopy (Figure 61) and MS analysis of the peptide (Figure 60).

The (*2S,4S*)-fluoroproline containing peptide (Peptide 27) is a weaker substrate for vCPH than Peptide 4 (Figure 62), indicating that the proline ring conformation may be important in substrate binding and release. (*2S,4S*)-Fluoroproline has a strong preference for the *endo*-conformation due to the strong electron-withdrawing effect of F and upon hydroxylation, the stronger electron withdrawing effect of F (Figure 67B). Depending on the conformation of the substrate proline residue during catalysis pucker, this may affect catalytic efficiency. The resulting, short-lived, (*4S*)-fluoro, (*4R*)-hydroxyl product does not readily adopt the *exo*-ring pucker (Figure 67C) that is readily adopted by (*2S,4R*)-hydroxyproline (Figure 67A) due to the stronger electron withdrawing effect of F compared to O. This effect therefore may prevent efficient product release and reducing catalytic efficiency (Figure 62). However, further structural studies are required to fully elucidate this effect.

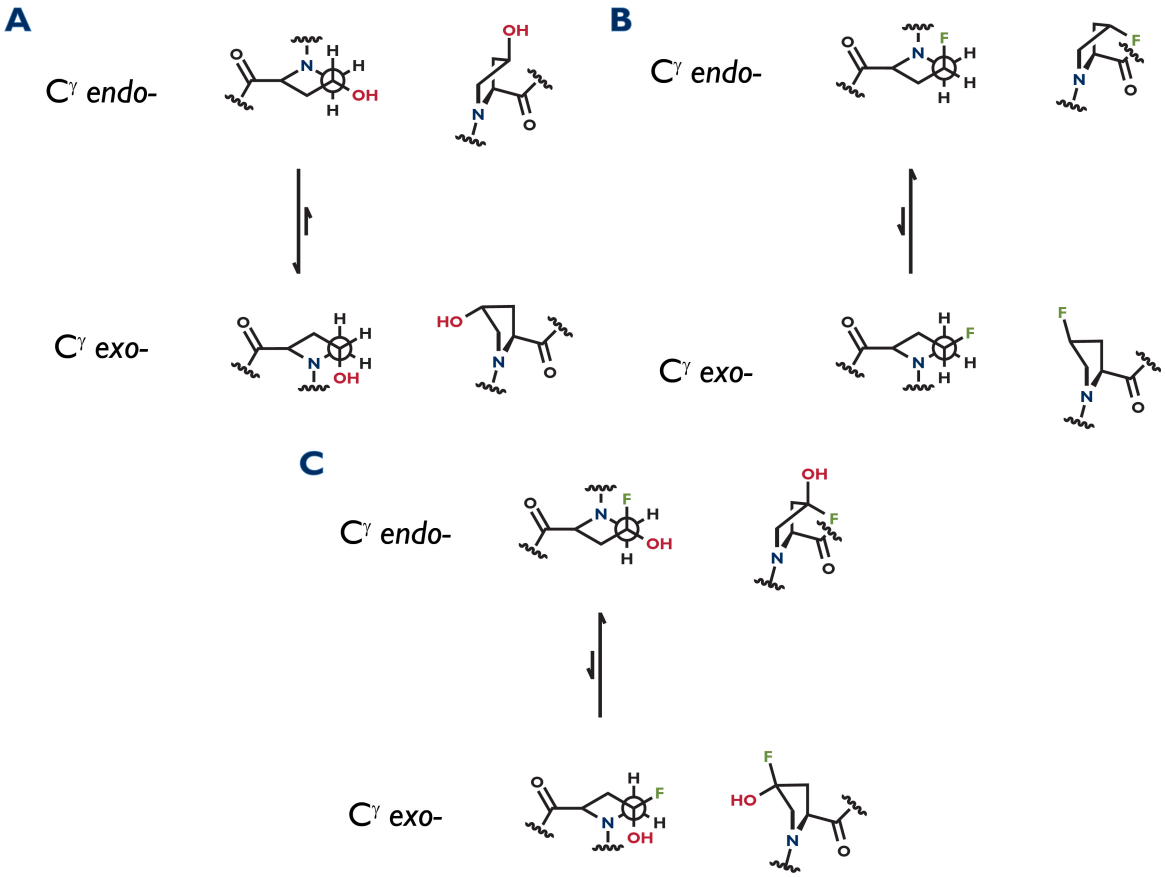


Figure 67 - Ring conformations adopted by (*2S,4R*)-hydroxyproline (A), (*2S,4S*)-fluoroproline (B) and (*4R*)-hydroxy-(*2S,4S*)-fluoroproline (C) residues. (*2S,4R*)-Hydroxyproline residues favour the *exo*- ring pucker and (*2S,4S*)-fluoroproline residues favour the *endo*- ring pucker, this is due to a gauche effect. Upon hydroxylation of (*2S,4S*)-fluoroproline, the resulting, albeit short-lived, (*4R*)-hydroxy-(*2S,4S*)-fluoroproline residue still favours the *endo*- conformation thanks to the greater electron withdrawing effect of fluorine when compared to oxygen.

Ascorbate (**3**) is not required for the vCPH catalytic cycle to progress. However, in the absence of ascorbate there is a build-up of mono-hydroxylated product is observed when Peptide 2 is incubated with vCPH. Kinetic analyses of vCPH catalysed hydroxylation of the two mono-hydroxylated peptides (PA-Hyp-KPAPK, Peptide 4, and PAPK-Hyp-APK, Peptide 5) reveal that the two hydroxylation events have near identical requirements for 2OG (**1**) and Fe^{II}, but have differing requirements for ascorbate (**3**); vCPH mediated hydroxylation at position 3 of Peptide 2 is less efficient and the K_M^{app} for ascorbate (**3**) is ten times greater than the ascorbate K_M^{app} for hydroxylation at position 5 of Peptide 2. vCPH 2OG turnover in the presence of Peptide 27 shows a similar ascorbate dependence to the weaker substrate Peptide 5. These observations have raised the distinct possibility that the requirement for ascorbate is sequence dependent

and is related to the affinity of the substrate. Given the requirement for ascorbate (**3**) of the human C-P4H enzymes and their historical links to scurvy[55,56,88], vCPH provides exciting potential as a model system from which to further elucidate the ascorbate (**3**) dependence of collagen prolyl hydroxylases, and the more general role of ascorbate (**3**) in 2OG oxygenase catalysis.

vCPH has been characterised as a collagen prolyl hydroxylase and the kinetic parameters determined now allow for the development of an MS-based inhibiton assay to study the effects of clinically relevant inhibitors on a collagen prolyl hydroxylase. Alternatively, monitoring of vCPH catalysis may also be achieved by assaying F^- release from Peptide 27, an approach that has previously been successfully applied to monitor the activity of a 2OG oxygenase involved in the carnitine biosynthesis pathway[181].

Summary

A hydroxylase from *Paramecium Bursaria* Chlorella Virus I (vCPH) has been shown to catalyse the formation of (*2S,4R*)-hydroxyproline in collagen substrate peptides but not in the substrates for HIF-1α. vCPH also efficiently hydroxylates $(PAPK)_n$ repeat peptides[179], which have been used in this study to biochemically characterise the enzyme. The requirement for ascorbate of vCPH-catalysed hydroxylation has been shown to differ depending on the position of the target proline within the PAPKPAPK peptide (Peptide 2), which is hydroxylated at prolines 3 and 5).vCPH is therefore an interesting model for understanding the role of ascorbate in 2OG oxygenase catalysis and also as a model collagen hydroxylase.

Chapter 4
vCPH Inhibition and Structural Studies

Work in this chapter has been conducted in collaboration with James Holt-Martyn (synthesis of compounds **34**-**43**) and Dr Anthony Tumber (LC-MS PHD2 inhibition assays). Studies with compounds **34**-**43** are currently being prepared for publication in James Holt-Martyn's DPhil thesis.

Although interest in the development of collagen prolyl hydroxylase inhibitors has subsided following unsuccessful attempts in the 1980s and early 1990s[16,102,137,138], the HIF prolyl hydroxylases, most notably PHD2, have attracted significant interest from a therapeutic perspective[55,139,141,143-146]. Currently, there are a number of clinical candidates that target PHD2, most often for the treatment of anaemia[143,144], whilst the hypoxic response has also been suggested as a potential target for anti-inflammatory agents[142,148].

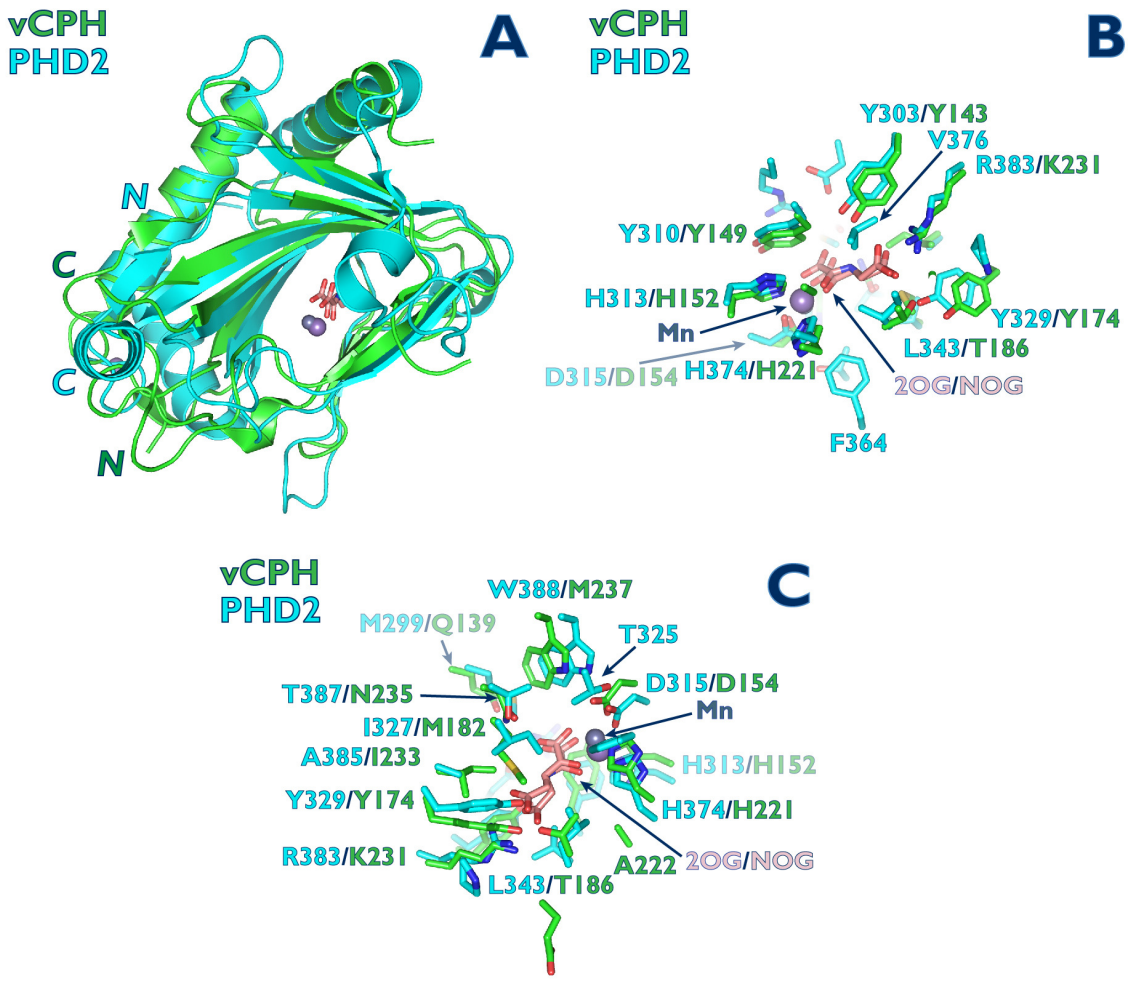


Figure 68 - Overlaid views from structures of vCPH (green, PDBID: 5C5T[84]) and PHD2 (blue, PDBID: 5L9R[182]). A - Overlaid structures showing general similarity between the two enzymes, particularly the DSBH. B and C – Different views of the 2OG binding residues within 5 Å of the 2OG (or an inactive 2OG mimic, NOG (**4**), in the PHD2 structure).

PHD2 and vCPH share approximately 20% sequence identity within their core Fe^{II}/2OG oxygenase regions. Analysis of the structures of both vCPH and PHD2 shows strong similarity between the 2OG binding sites of the two enzymes (Figure 68). There are, however, some

differences in the 2OG binding site, with the most apparent being the positively charged residue that binds the distal carboxylate of 2OG. vCPH, the C-P4Hs and P4H-TM feature a lysine residue at this position, whereas the HIF prolyl hydroxylases, P3Hs and OGFOD1 bind the distal carboxylate of 2OG with an arginine residue (Figure 69).



Figure 69 - Alignment of human prolyl hydroxylases with vCPH. The C-P4H enzymes, vCPH and P4H-TM bind to the distal carboxylate of 2OG with a lysine residue, whereas the HIF prolyl hydroxylases, OGFOD1 and P3H enzymes employ an arginine at this position (highlighted in red).

In spite of these differences, compounds that inhibit PHD2 may also inhibit vCPH and the effect of PHD2 inhibitors on collagen biosynthesis is currently unknown. In the absence of a robust and reliable assay for the human C-P4H enzymes, vCPH was chosen as a suitable model enzyme to investigate the potential impact of these inhibitors on the collagen prolyl hydroxylases.

Inhibition Assay Development

Studies described in Chapter 3 have provided the basis for inhibition assay development, utilising either a MALDI-TOF MS assay, directly monitoring of peptide hydroxylation, or by monitoring F^- release following vCPH-catalysed hydroxylation of Peptide 27 applying methodologies that were used for studies with BBOX[181]. Both approaches require that reactions are quenched prior to analysis (*i.e.* continuous monitoring is not possible with either assay platform). The MALDI-TOF MS based approach was chosen as it allows for direct monitoring of product formation.

Using the K_M value for 2OG, alongside the K_M^{app} values for ascorbate and Fe^{II} (determined in Chapter 3), a peptide hydroxylation assay was developed, directly monitoring the hydroxylation of Peptide 4 (PA-Hyp-KPAPK). Directly monitoring product formation allows for a more accurate

measure of enzymatic activity than the monitoring of 2OG turnover, which was previously used to monitor vCPH activity[179].

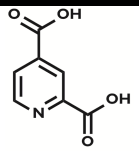
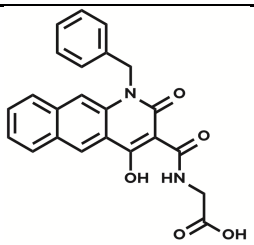
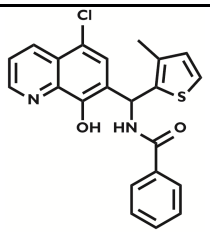
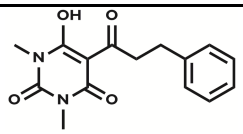
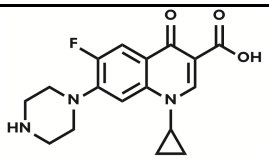
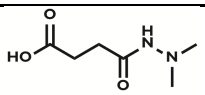
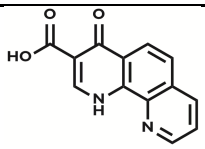
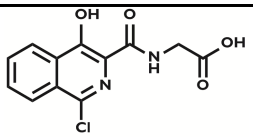
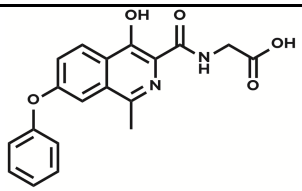
The assay was conducted with an excess of Fe^{II} and ascorbate (**3**), whilst the 2OG concentration was kept at the K_M to allow for inhibitor binding, as most inhibitors were expected to compete with 2OG. Time courses with different peptide concentrations were conducted to ensure the reaction was optimised for maximal hydroxylation within the linear range. The optimised assay conditions use a final vCPH concentration of 1 μM, therefore the detection limit (for assessing IC_{50} values) of the assay is 0.5 μM.

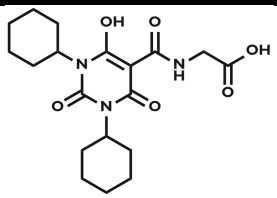
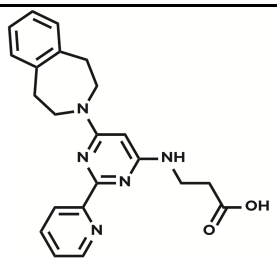
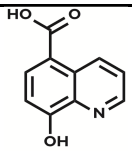
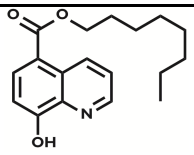
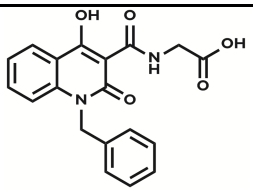
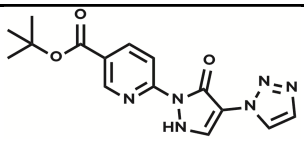
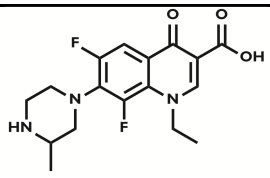
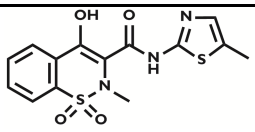
Component	Final Concentration
vCPH	1 μM
$(NH_4)_2Fe(SO_4)_2$	50 μM
Sodium-L-Ascorbate	100 μM
2OG (Disodium Salt)	30 μM
Peptide 4	10 μM
HEPES pH 7.5	50 mM
Inhibitor	Varied

Table 6 - Final concentrations of components in vCPH inhibition assay.

Compound Library

A number of compounds known to inhibit 2OG oxygenases were assembled into a library for testing against vCPH (Table 7). Since the active sites of PHD2 and VCPH align closely (20% identity, Figure 68), a number of known PHD2 inhibitors, including compounds currently in clinical trials (*e.g.* FG-4592, **8**), were included in this library. In addition to the small molecules detailed in Table 7, NODD and CODD peptides (Peptide 29 and Peptide 30), substrates for PHD2, were also tested as potential vCPH inhibitors.

Name	Compound Number	Structure	
2,4-PDCA	**5**		Broad spectrum inhibitor of the 2OG oxygenases, originally from early C-P4H studies[183].
BIOX2	**18**		Fluorescent PHD2 inhibitor[184].
CCT 2-66	**15**		Histone demethylase inhibitor with selectivity over the HIF prolyl hydroxylases[185].
CCT 6-140	**19**		Inhibitor of the 2OG-dependent prolyl-3-hydroxylase OGFOD1[158].
Ciprofloxacin	**20**		Antibiotic used for treatment of urinary tract infections. Linked to increase in likelihood of tendonitis and tendon rupture[186].
Daminozide	**16**		Plant growth inhibitor[159] that has also been identified as a KDM inhibitor[160].
FG-0041	**7**		Developed as an inhibitor of C-P4H[137] by FibroGen.
FG-2216	**21**		First generation FibroGen PHD2 inhibitor.
FG-4592 Roxadustat	**8**		FibroGen PHD2 inhibitor, currently in Phase III clinical trials for the treatment of renal anaemia[143].

Name	Compound Number	Structure	
GSK1278863 Daprodustat	**10**		GSK PHD2 inhibitor in Phase III clinical trials for the treatment of renal anaemia[146].
GSKJ1	**22**		Inhibitor of KDM6A and KDM6B developed by GSK[187]. GSKJ4 (**33**), a prodrug version (ethyl ester) has an effect on collagen deposition by chondrocytes[188].
IOX1	**14**		Broad spectrum inhibitor of 2OG oxygenases, in particular the histone demethylases[157]. Developed by the Schofield group in collaboration with the Structural Genomics Consortium.
IOX1 Prodrug	**23**		IOX1 prodrug (*n*-octyl ester) [189].
IOX2	**24**		Probe for the HIF prolyl hydroxylases (PHD1, PHD2 and PHD3)[190]. Developed by the Schofield group in collaboration with the Structural Genomics Consortium.
IOX4	**25**		Probe for the HIF prolyl hydroxylases (PHD1, PHD2 and PHD3)[191], structurally related to Molidustat (**11**). Developed by the Schofield group in collaboration with the Structural Genomics Consortium.
Lomefloxacin	**26**		Fluoroquinolone antibiotic, linked to increase in likelihood of tendonitis and tendon rupture[186].
Meloxicam	**27**		Non-steroidal anti-inflammatory drug. Cyclooxygenase (COX) inhibitor[192].

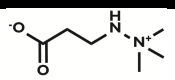
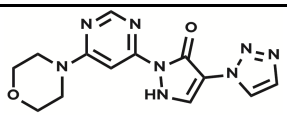
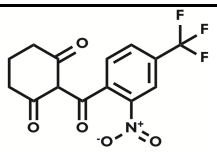
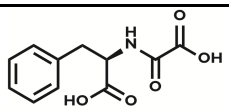
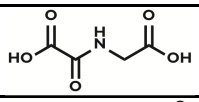
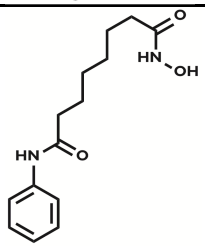
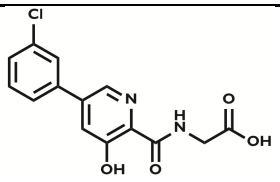
Name	Compound Number	Structure	
Mildronate	**13**		Inhibitor of BBOX (a 2OG oxygenase that is part of the carnitine biosynthesis pathway)[193]. Mildronate is a substrate mimic.
Molidustat	**11**		PHD2 inhibitor developed by Bayer[147] and in phase II clinical trials[194].
Nitisinone	**28**		Inhibitor of 4-hydroxyphenyl pyruvate dioxygenase (HPPD).
NOFD	**29**		Inhibitor of FIH with selectivity against the HIF prolyl hydroxylases[195].
NOG	**4**		Catalytically inactive 2OG mimic.
SAHA Vorinostat	**30**		A histone deacetylase inhibitor approved by the FDA for the treatment of T-cell lymphoma[196].
Vadadustat	**9**		PHD2 inhibitor developed by Akebia therapeutics (AKB-6548). Currently in Phase III clinical trials for the treatment of renal anaemia[144,145].

Table 7 - Compound library tested against vCPH for inhibition activity.

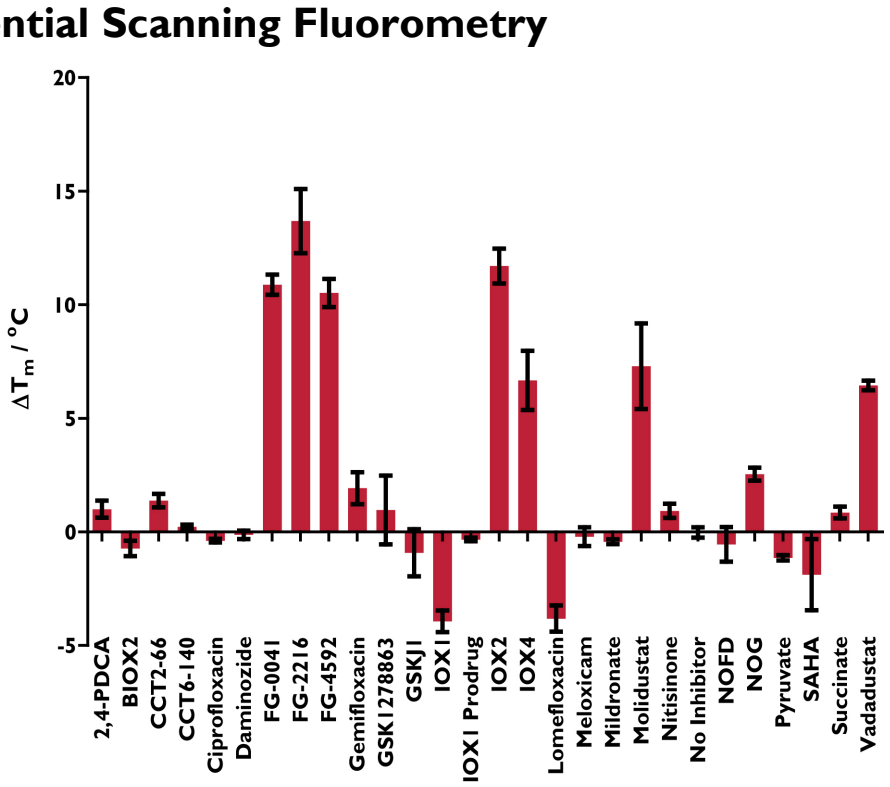


Figure 70 - Changes in melting temperature of the vCPH/Mn^{II} complex in the presence of a range of compounds. ΔT_m is given with respect to a no inhibitor control. Bars are plotted as the mean average of three measurements with the standard error indicated. A positive ΔT_m is indicative of protein stabilisation.

Prior to testing of compounds in the peptide hydroxylation assay, differential scanning fluorometry was used to assess any changes in the thermal stability of vCPH in the presence of the compounds from Table 7, to test for potential inhibitors binding. Mn^{II} was added to *apo*-vCPH and the resulting complex was treated with compound. The T_m of the mixture was assessed and compared to that of the untreated vCPH/Mn^{II} complex; changes in T_m are shown in Figure 70, where a positive ΔT_m is indicative of binding.

These data show that the vCPH/Mn^{II} complex is stabilised by known prolyl hydroxylase inhibitors. FG-0041 (**7**), which was developed as a C-P4H inhibitor[137] was found to stabilise the complex strongly, as did the compounds that target PHD2 (FG-2216 (**21**), FG-4592 (**8**), IOX2 (**24**), IOX4 (**25**), Molidustat (**11**) and Vadadustat (**9**)). However, little or no change in ΔT_m does not rule out binding[197] and therefore all compounds were screened as potential inhibitors of vCPH.

Initial Screening

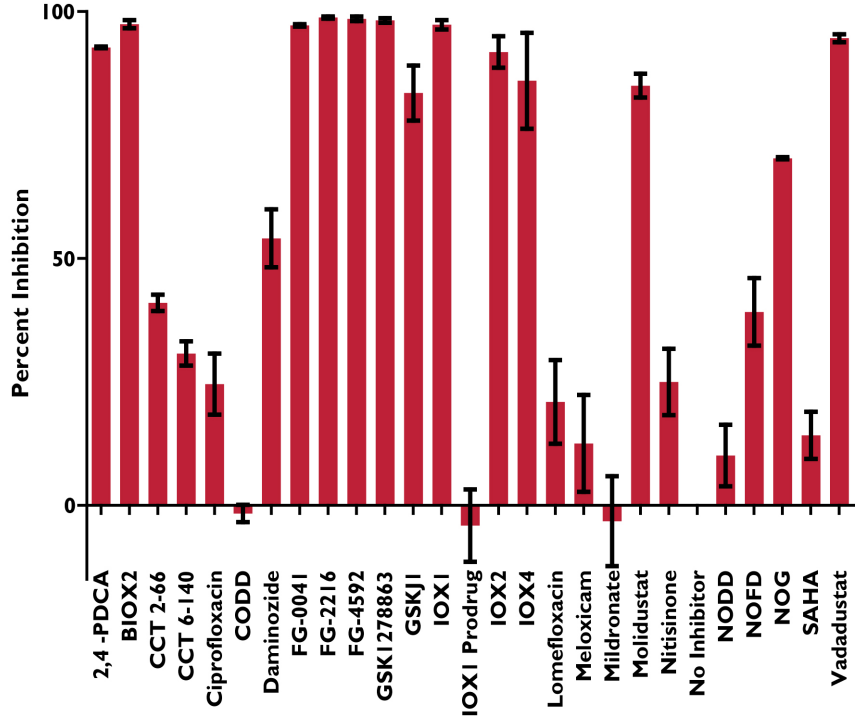


Figure 71 - Results of single shot screen at an inhibitor concentration of 100 µM. Bars are plotted as the mean average of three measurements with the standard error indicated. Percent inhibition is indicated, with 0% corresponding to vCPH activity in a no inhibitor control and 100% corresponding to complete inhibition of vCPH activity. IC_{50} values were subsequently calculated for any compounds exhibiting greater than 50% inhibition at 100 µM.

Potential inhibitors were screened against vCPH at a final inhibitor concentration of 100 µM. The data for the compounds listed in Table 7, along with the CODD and NODD peptides (Peptide 30 and Peptide 29 respectively), are shown in Figure 71. All compounds that exhibited greater than 50% inhibition at 100 µM were then taken forward for IC_{50} determination.

The PHD2 inhibitors which conferred stabilised the vCPH/Mn^{II} complex *via* differential scanning fluorimetry were also found to be inhibitors of vCPH at a concentration of 100 µM, with near 100% inhibition exhibited by FG-2216 (**21**), FG-4592 (**8**), IOX2 (**24**), IOX4 (**25**), Molidustat (**11**) and Vadadustat (**9**). 2,4-PDCA (**5**), BIOX2 (**18**) and IOX1 (**14**), which are known to inhibit other 2OG oxygenases but were not found to stabilise the vCPH/Mn^{II} complex, also showed significant vCPH inhibition.

No inhibition of vCPH was exhibited by either the CODD or NODD peptides (Peptide 30 and Peptide 29), which are substrates for HIF prolyl hydroxylases. Together, the absence of inhibition activity and the absence of any reaction when incubated with vCPH (Figure 65) indicates that there is no interaction between the HIF prolyl hydroxylase substrates and vCPH, further indication that vCPH is a collagen prolyl hydroxylase.

Counter-screening of compounds which target alternative enzyme classes (*i.e.* compounds that have targets that are not 2OG oxygenases) such as SAHA (**30**) and NOFD (**29**) did not exhibit any significant vCPH inhibition.

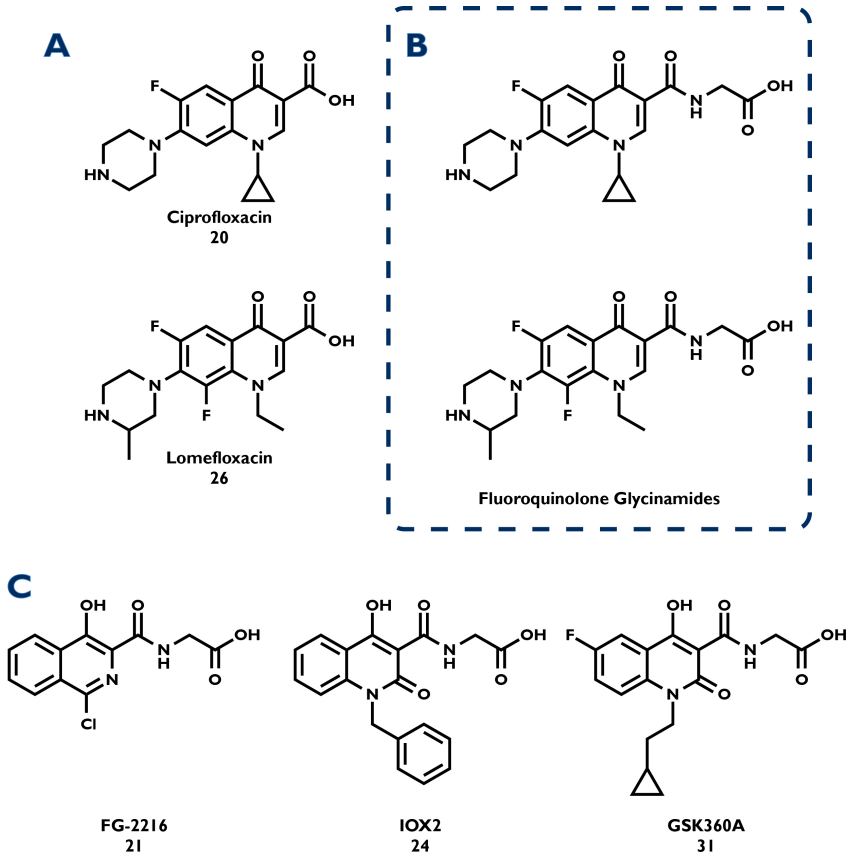


Figure 72 - Structural comparison of fluoroquinolone antibiotics and known PHD2 inhibitiors.

Low level inhibition by the two tested fluoroquinolone antibiotics (ciprofloxacin (**20**) and lomefloxacin (**26**), Figure 72A) was observed. Given the link between the fluoroquinolone antibiotics and tendon rupture[186,198], this result does not suggest direct collagen prolyl-4-hydroxylase inhibition by these antibiotics is the direct cause. Although the glycinamide

products (Figure 72B) have not been observed during metabolite analysis[199], these proposed structures strongly resemble another PHD2 inhibitor GSK360A (**31**), which contains the flouroquinolone core and has been observed to reduce fibrosis in diseased shoulder tendons[200]. GSK360A (**32**), as well as FG-2216 (**21**) and IOX2 (**24**) have a clear structural similarity to the proposed fluoroquinolone inhibitors (Figure 72B/C). The fluoroquinolone glycinamides may be metabolites and therefore further study, commencing with the synthesis of the structures presented in Figure 72B for testing as potential inhibitors of 2OG-dependent prolyl hydroxylases, are warranted.

IC_{50} Determination

The vCPH inhibition assay was subsequently conducted across a range of inhibitor concentrations for all inhibitors that showed suitable inhibitory activity in the 100 μM screen. 100% response was defined by a no-inhibitor control that was run concurrent to the inhibition assays. Dose-response curves were then drawn and IC_{50} values subsequently derived (Figure 73).

The strongest inhibitors of vCPH, with single-digit micromolar IC_{50} values, were Vadadustat (**9**, IC_{50} < 1 μM), IOX4 (**25**, IC_{50} = 2.5 μM ± 0.2 μM), 2,4-PDCA (**5**, IC_{50} = 5.3 μM ± 0.3 μM), FG-4592 (**8**, IC_{50} = 5.0 μM ± 0.1 μM) and FG-2216 (**21**, IC_{50} = 4.5 μM ± 0.1 μM). These compounds, with the exception of 2,4-PDCA (a broad-spectrum inhibitor of the 2OG oxygenases), were originally designed as inhibitors for PHD2, and were more potent inhibitors of vCPH than FG-0041 (**7**, IC_{50} = 15.7 μM ± 1.4 μM), which was designed as a C-P4H inhibitor[137]. Two probe compounds for PHD2, IOX2 (**24**, IC_{50} = 8.5 μM ± 0.2 μM) and BIOX2 (**18**, IC_{50} = 12.1 μM ± 0.5 μM), were also found to be potent vCPH inhibitors.

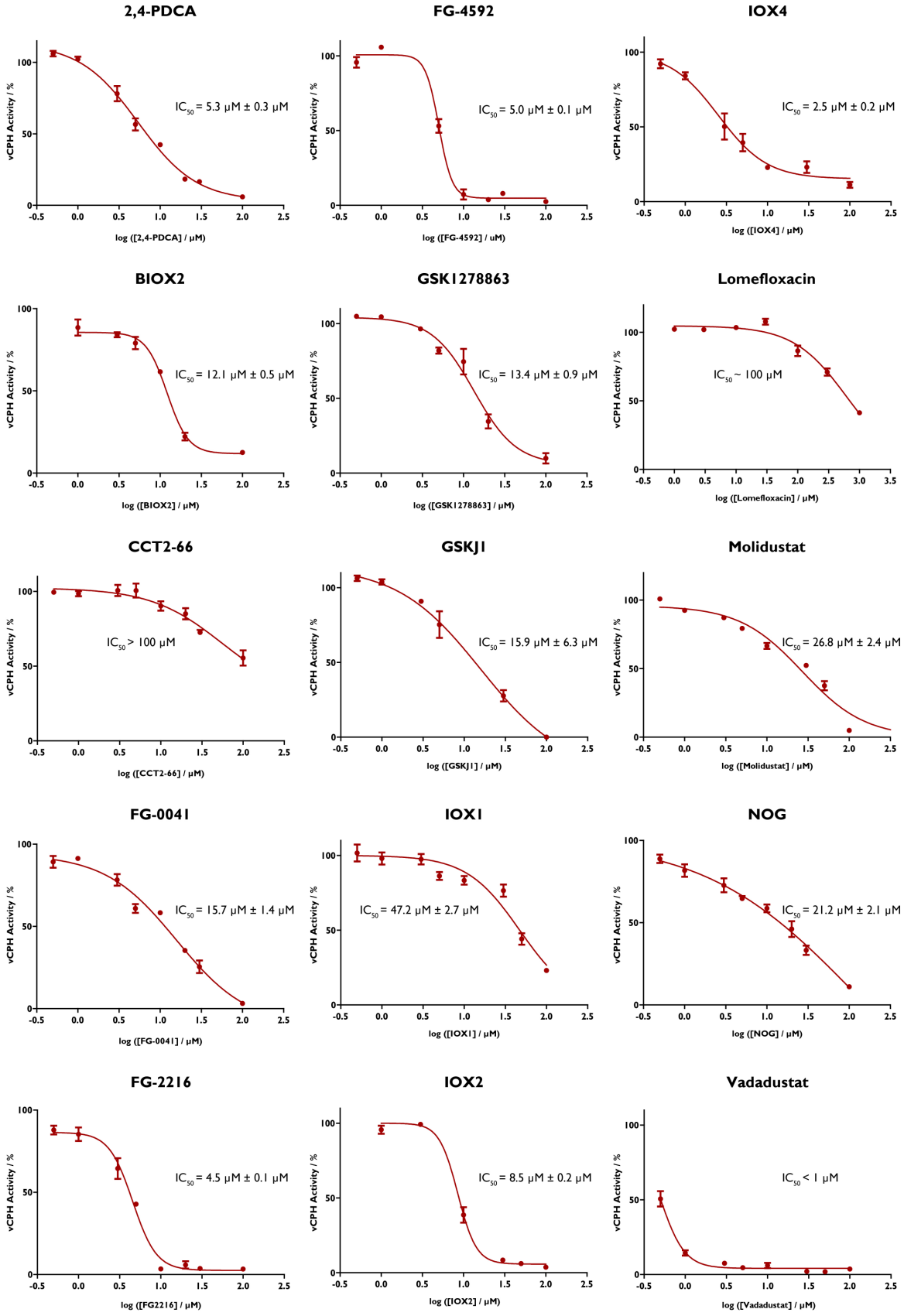


Figure 73 – Dose-response curves for compounds that inhibited vCPH. Points are plotted as the mean average of three repeats with the standard error indicated. Non-linear regression was conducted with GraphPad Prism. IC_{50} values are indicated on the curve with the standard error also indicated.

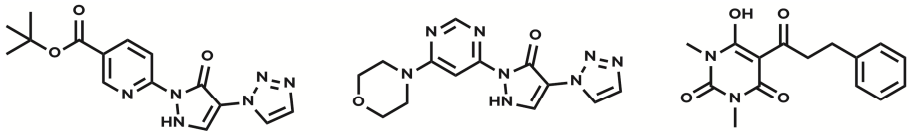


Figure 74 - Structures of IOX4 (**25**), Molidustat (**11**) and CCT6-140 (**19**).

Two other PHD2 inhibitors that are currently in clinical trials, GSK1278863 (**10**, IC_{50} = 13.4 μM ± 0.9 μM) and Molidustat (**11**), were less potent inhibitors of vCPH. In the case of Molidustat (**11**, IC_{50} = 26.8 μM ± 2.4 μM) this is somewhat in contrast to the greater potency exhibited by the structurally similar IOX4 (**25**, IC_{50} = 2.5 μM ± 0.2 μM, structures are indicated in Figure 74). IOX4 (**25**), GSK1278863 (**10**) and Molidustat (**11**) are more potent inhibitors of PHD2 than FG-4592 (**8**) and Vadadustat (**9**), which are the most potent vCPH inhibitors. This indicates that although the clinical PHD2 inhibitors are vCPH inhibitors, there is not a direct correlation between PHD2 and vCPH potency.

The broad spectrum inhibitors of 2OG oxygenases were also found to act as vCPH inhibitors, but to a lesser extent than the more potent PHD2 inhibitors. NOG (**4**, IC_{50} = 21.2 μM ± 2.1 μM), an inactive 2OG mimic had 50% inhibition when the concentration of NOG and 2OG were roughly equal, whilst IOX1 (**14**), which is a broad spectrum KDM inhibitor[157], was a poor inhibitor of vCPH (IC_{50} = 47.2 μM ± 2.7 μM).

CCT6-140 (**19**), an inhibitor that is selective for OGFOD1 versus PHD2[158,201] has an IC_{50} of greater than 100 μM with vCPH. This further supports the selectivity profile of CCT6-140 (**19**) and indicates that OGFOD1 inhibition can be achieved without inhibition of the collagen prolyl-4-hydroxylases.

The inhibition of vCPH by GSKJ1 (**22**) is also notable (IC_{50} = 15.9 μM ± 6.3 μM) and further illustrates that GSKJ1 (**22**) is less selective than initially claimed[187]. Cell-based experiments with GSKJ4 (**33**), an ethyl ester prodrug form of GSKJ1 (**22**), has shown a significant reduction in

collagen production by cells cultured *in vitro* when treated with GSKJ4[188], consistent with the inhibition of a collagen prolyl hydroxylase.

Mechanism of vCPH Inhibition

The mechanism of inhibition for a subset of vCPH inhibitors was then investigated. This subset of compounds consisted of the PHD2 inhibitors that have entered clinical trials (GSK1278863 (**10**), FG-4592 (**8**), FG-2216 (**21**), Molidustat (**11**) and Vadadustat (**9**)) alongside the broad spectrum inhibitors NOG (**4**) and 2,4-PDCA (**5**).

2OG Binding and Displacement Studies

The majority of the inhibitors tested are expected to compete with 2OG (**1**) in the active site of vCPH. To test for 2OG displacement by inhibitors, ^{1}H NMR spectroscopy, using the Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence, was used to monitor binding of 2OG to vCPH. 2OG resonances are observed when the molecule is in solution but depleted when bound to protein.

2OG (**1**), does not bind to *apo*-vCPH in the absence of metal, as evidenced by the continued observation of a sharp triplet peak when *apo*-vCPH is added to 2OG (Figure 75A). This signal is depleted upon addition of Ni^{II}, Zn^{II} and Mn^{II}. The signal is completely ablated upon addition of Mn^{II} but not Ni^{II} and Zn^{II}, indicating that either Mn^{II} binds with stronger affinity to *apo*-vCPH than Ni^{II} and Zn^{II} (consistent with differential scanning fluorometry analyses, Figure 56 [Chapter 3]) or that 2OG is unable to bind to the vCPH/Ni^{II} or vCPH/Zn^{II} complexes.

Since Mn^{II} is paramagnetic, the 2OG resonances are broadened upon addition of Mn^{II} (Figure 75B) and are not observed upon addition of vCPH. Addition of NOG (**4**) displaces 2OG and restores the 2OG resonance (Figure 75B), at higher metal concentrations, addition of NOG does not restore the resonance due to chelation of the free Mn^{II} by 2OG and/or NOG (Figure 75B).

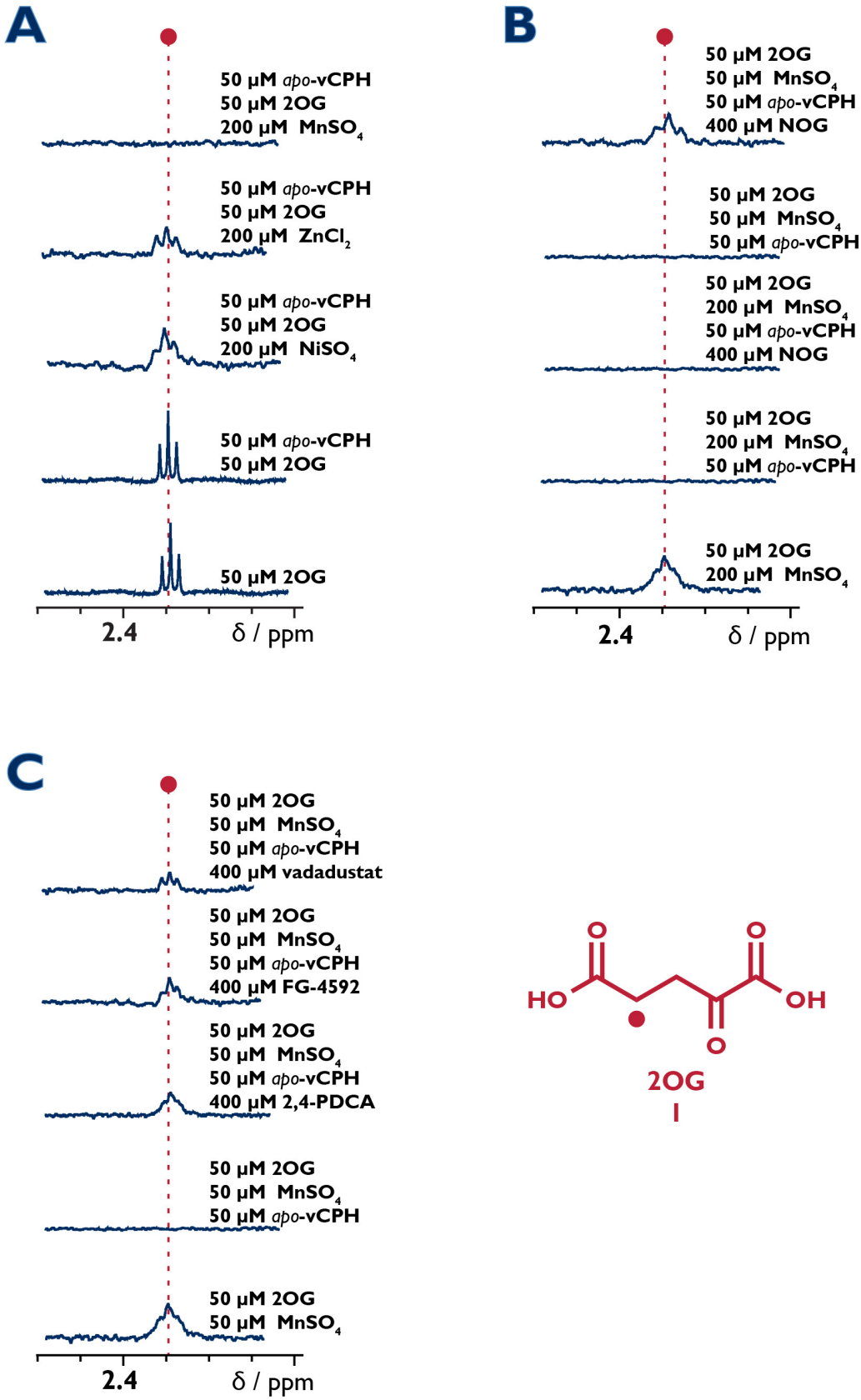


Figure 75 - Results of ^{1}H NMR spectroscopy analysing the binding of 2OG (**1**) to *apo*-vCPH. A - The 2OG resonance is unchanged upon addition of *apo*-vCPH but is not observed on addition of $MnSO_4$, indicative of binding to vCPH. B - 2OG (**1**) is displaced from the vCPH/Mn^{II} complex by NOG (**4**) as the resonance is restored upon addition of NOG. C - 2OG (**1**) is displaced from the vCPH/Mn^{II} complex by 2,4-PDCA (**5**), FG-4592 (**8**) and Vadadustat (**9**).

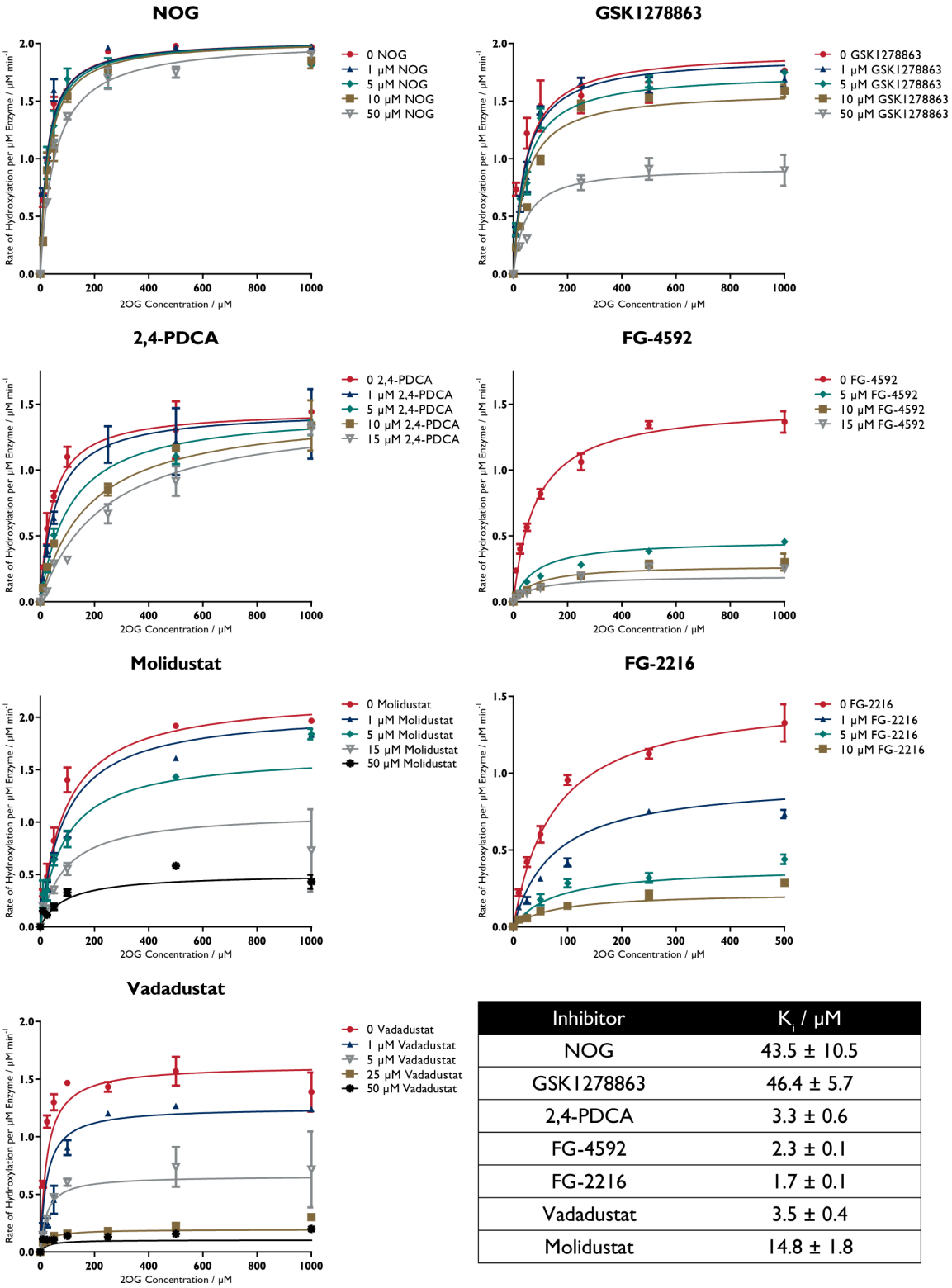
Upon addition of 2,4-PDCA (**5**), FG-4592 (**8**) and Vadadustat (**9**), resonances corresponding to 2OG were observed, indicating that 2OG (**1**) had been displaced from vCPH upon addition of these inhibitors. These data indicate that, consistent with the expected binding modes of the compounds, NOG (**4**), 2,4-PDCA (**5**), FG-4592 (**8**) and Vadadustat (**9**) all compete with 2OG and by implication, occupy the 2OG binding site of vCPH. However, due to the broadening induced by the paramagnetism of Mn^{II}, a quantitative assessment of 2OG displacement was not possible.

K_i Determination for Clinically Relevant PHD2 Inhibitors

To better understand the competition between the inhibitors and 2OG, the inhibition constant K_i was calculated by assessing changes in the K_M for 2OG at a range of inhibitor concentrations. In all cases, the K_M for 2OG was increased in the presence of inhibitor.

In the case of NOG, a competitive inhibition model was fitted, whilst for the other compounds, an uncompetitive inhibition model was applied since for compounds other than NOG, the V_{max} was not achieved in the presence of inhibitor[202]. Broadly, the rank order of the inhibitors was reflective of the IC_{50} values that were obtained; NOG (**4**, K_i = 43.5 μM ± 10.5 μM) and GSK1278863 (**10**, K_i = 46.4 μM ± 5.7 μM) were found to be the weakest inhibitors of vCPH in this subset, while the most potent were FG-2216 (**21**, K_i = 1.7 μM ± 0.1 μM) and FG-4592 (**8**, K_i = 2.3 μM ± 0.1 μM). Despite despite being the most potent vCPH inhibitor, Vadadustat (**9**, K_i = 3.5 μM ± 0.4 μM) was found to bind less tightly than FG-2216 (**21**) and FG-4592 (**8**).

Molidustat (**11**, K_i = 14.8 μM ± 1.8 μM), similar to the order observed when ranking IC_{50} values, has weaker binding that FG-2216 (**21**), FG-4592 (**8**) and 2,4-PDCA (**5**) but exhibits stronger vCPH binding than NOG (**4**) and GSK1278863 (**10**).



Inhibitor	K_i / μM
NOG	43.5 ± 10.5
GSK1278863	46.4 ± 5.7
2,4-PDCA	3.3 ± 0.6
FG-4592	2.3 ± 0.1
FG-2216	1.7 ± 0.1
Vadadustat	3.5 ± 0.4
Molidustat	14.8 ± 1.8

Figure 76 - K_i determination for clinical PHD2 inhibitors. Values for 2,4-PDCA (**5**) and NOG (**4**) was also calculated. Points are plotted as the mean of three repeats with the standard error measurement indicated. Curves were fitted using either an uncompetitive inhibition model or, in the case of NOG (**4**), a competitive inhibition model in GraphPad Prism. The derived K_i values and standard error are indicated in the table.

Inhibitor Competition with Peptide

In order to test for inhibitor competition with the peptide substrate, IC_{50} values were calculated in the presence of different peptide concentrations. At higher concentrations of peptide substrate, sensititity of the MALDI-TOF MS assay decreases rapidly as the relative abundance of hydroxylated product is much smaller at higher peptide concentrations. Therefore, the data obtained have greater susceptibility to uncertainty due to variation in data acquisition.

	IC_{50}, 10 µM Peptide	IC_{50}, 25 µM Peptide	IC_{50}, 100 µM Peptide
GSK1278863 (10)	5.0 µM ± 0.6 µM	5.6 µM ± 0.4 µM	8.6 µM ± 3.2 µM
FG-2216 (21)	4.7 µM ± 0.1 µM	4.6 µM ± 1.0 µM	Not Determined
NOG (4)	9.0 µM ± 1.2 µM	8.8 µM ± 1.7 µM	8.1 µM ± 4.1 µM

Table 8 - IC_{50} values obtained for NOG (**4**), FG-2216 (**21**) and GSK1278863 (**10**) at different peptide substrate concentrations. The dose-response curves are shown in Figure 77. Values are reported along with standard error measurement.

Initial studies with NOG (**4**), FG-2216 (**21**) and GSK1278863 (**10**) exhibited no change in IC_{50} when peptide concentration was increased from 10 µM to 25 µM (Table 8 and Figure 77), the values obtained for GSK1278863 (**10**) are within error. The absolute vCPH IC_{50} values for GSK1278863 (**10**) and NOG (**4**) vary from those obtained under analogous conditions (Figure 73); however, the quality of the data obtained in this instance is relatively poor (as indicated by the large error bars).

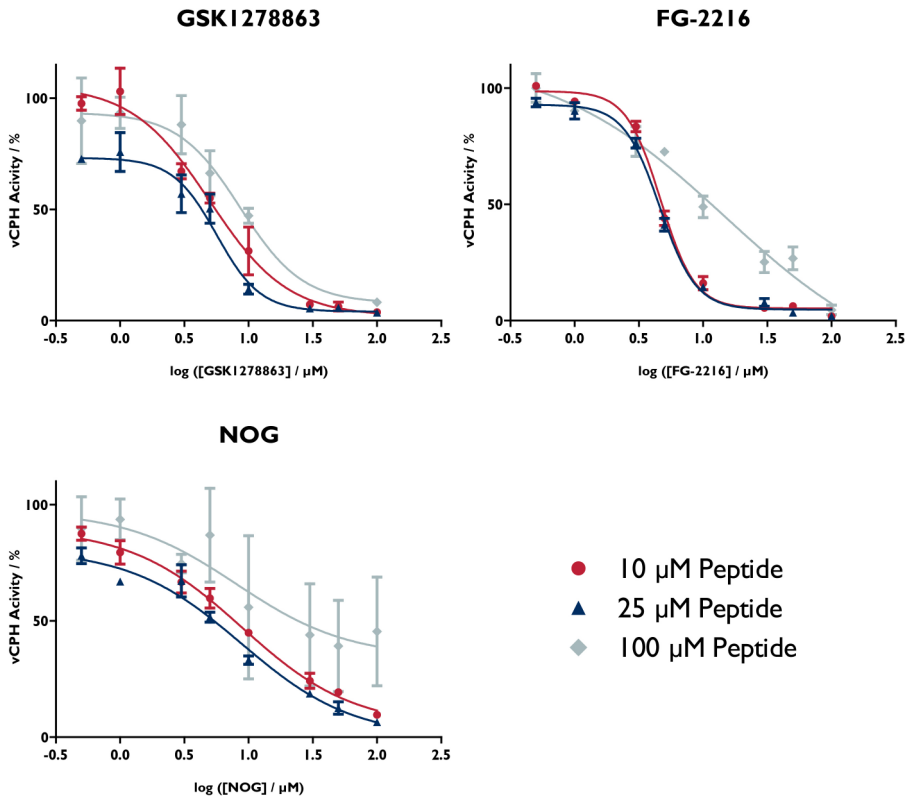


Figure 77 - Dose-response curves for vCPH treatment with GSK1278863 (**10**), FG-2216 (**21**) and NOG (**4**) in the presence of different peptide substrate concentrations.

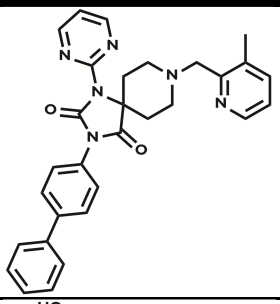
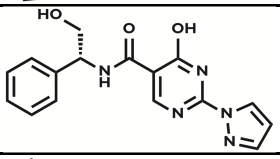
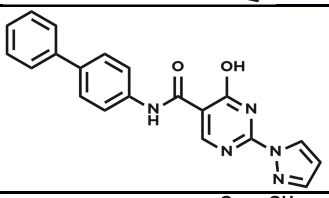
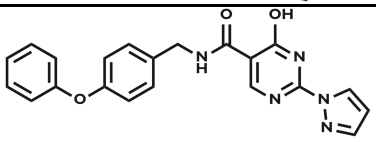
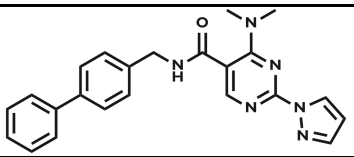
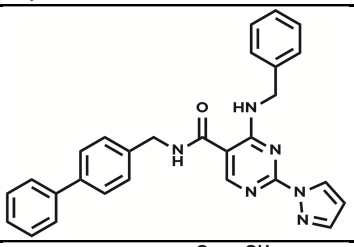
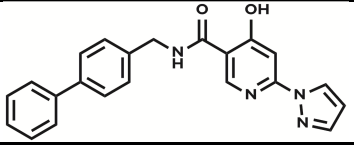
Assays conducted in the presence of 100 µM peptide did not return reliable data, an alternative assay platform is required to fully assess the impact of increasing peptide concentration on inhibition.

Counter-screening Novel PHD2 Inhibitors

The compounds used in this study were synthesised by Mr James Holt-Martyn. PHD2 assays were conducted by Dr Antony Tumber and OGFOD1 assays were conducted by Dr Christopher Lohans.

Research conducted in the Schofield group has sought to develop new potent and selective inhibitors of the HIF prolyl hydroxylases, specifically PHD2[203]. Therefore, to assess the selectivity of the compounds versus a collagen prolyl hydroxylase, the vCPH inhibition assay was used to counter-screen compound **34** and the series of compounds **35** to **43**. The structures of the compounds are shown in Table 9.

As with the initial compound library (Table 7), compounds were first analysed *via* differential scanning fluorometry and also screened for vCPH inhibition at a final concentration of 100 µM (Figure 78A and B). These studies revealed that JHM2-174 (**35**) and JHM388 (**39**) stabilised the vCPH-Mn^{II} complex, with a positive ΔT_m in the range of 3°C to 5°C. However, this is not to the same extent as the some of the compounds tested in Figure 70, where a ΔT_m in excess of +10°C was observed for the FibroGen compounds (**7**, **21** and **8**) and IOX2 (**24**).

Name	Compound Number	PHD2 IC_{50} / nM[203]	Structure
JHM2-167	**34**	500	
JHM2-174	**35**	- 92% inhibition at 10 µM by MALDI	
JHM3-201	**36**	258	
JHM3-202	**37**	- 100% inhibition at 10 µM by MALDI	
JHM377	**38**	- PHD2 Activation (61%)	
JHM388	**39**	- PHD2 Activation (50%)	
JHM532	**40**	145	

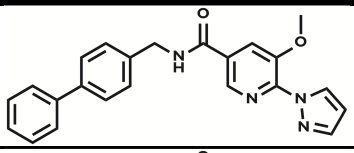
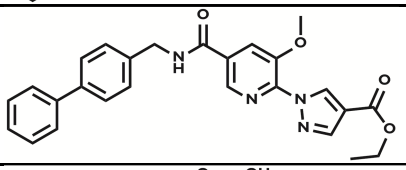
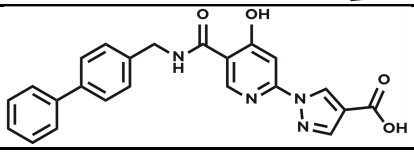
Name	Compound Number	PHD2 IC_{50} / nM[203]	Structure
JHM631	**41**	640	
JHM634	**42**	25	
JHM642	**43**	87	

Table 9 - Structures of novel PHD2 inhibitors that were assessed for inhibition against vCPH. IC_{50} values for PHD2 were assessed using a LC-MS peptide hydroxylation assay, with a final PHD2 concentration of 150 nM[203].

A single-shot screen of the compounds at a final concentration of 100 µM (Figure 78B) revealed no inhibition from JMJ2-167 (**34**) or the PHD2 activator compound JHM377 (**38**). The other PHD2 activator, JHM388 (**39**), exhibited less than 50% inhibition of vCPH at a concentration 100 µM despite differential scanning fluorometry indicating binding of JHM388 (**39**) to vCPH.

JHM2-174 (**35**, IC_{50} = 1.1 µM ± 0.1 µM) and JHM3-202 (**37**, IC_{50} = 2.0 µM ± 0.1 µM) exhibited single digit micromolar IC_{50} values, making them amongst the most potent inhibitors of vCPH that have been tested as part of this study; only Vadadustat (**9**, IC_{50} < 1 µM) was found to be more potent.

JHM634 (**42**) and JHM532 (**40**) have vCPH IC_{50} values of approximately 100 µM despite being potent inhibitors of PHD2 (particularly JHM634). These data therefore suggest that achieving selectivity between vCPH and PHD2 is feasible.

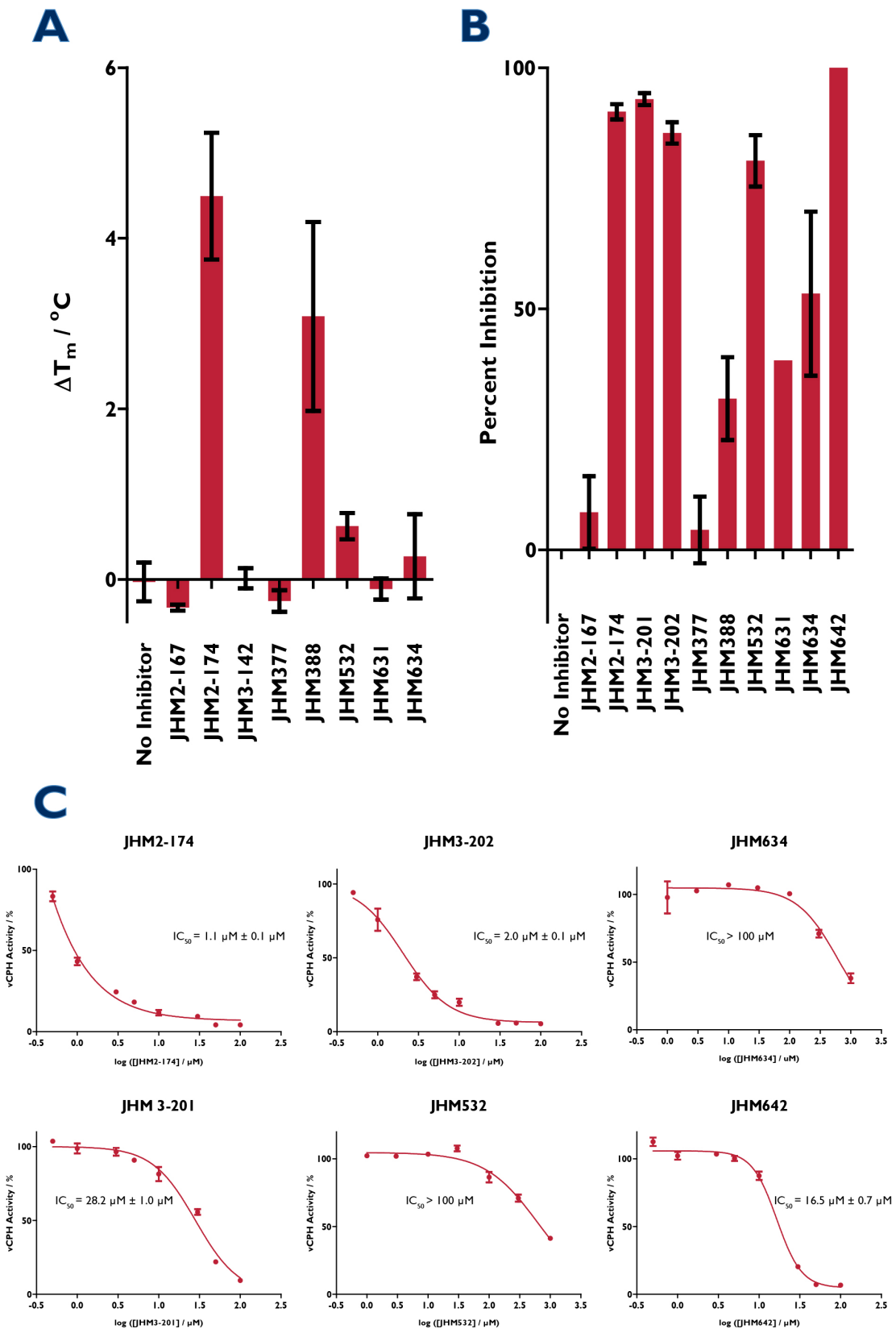


Figure 78 - Screening of the inhibitors detailed in Table 9 against vCPH. All points are shown as the mean of three readings with the standard error measurement indicated. A - Results of T_m analyses conducted *via* differential scanning fluorometry. B - Results of inhibition screen with vCPH at an inhibitor concentration of 100 μM. C - Dose-response curves for vCPH-catalysed hydroxylation of Peptide 4 in the presence of inhibitor. Non-linear regression was conducted in GraphPad Prism and IC_{50} values are indicated with standard error measurement.

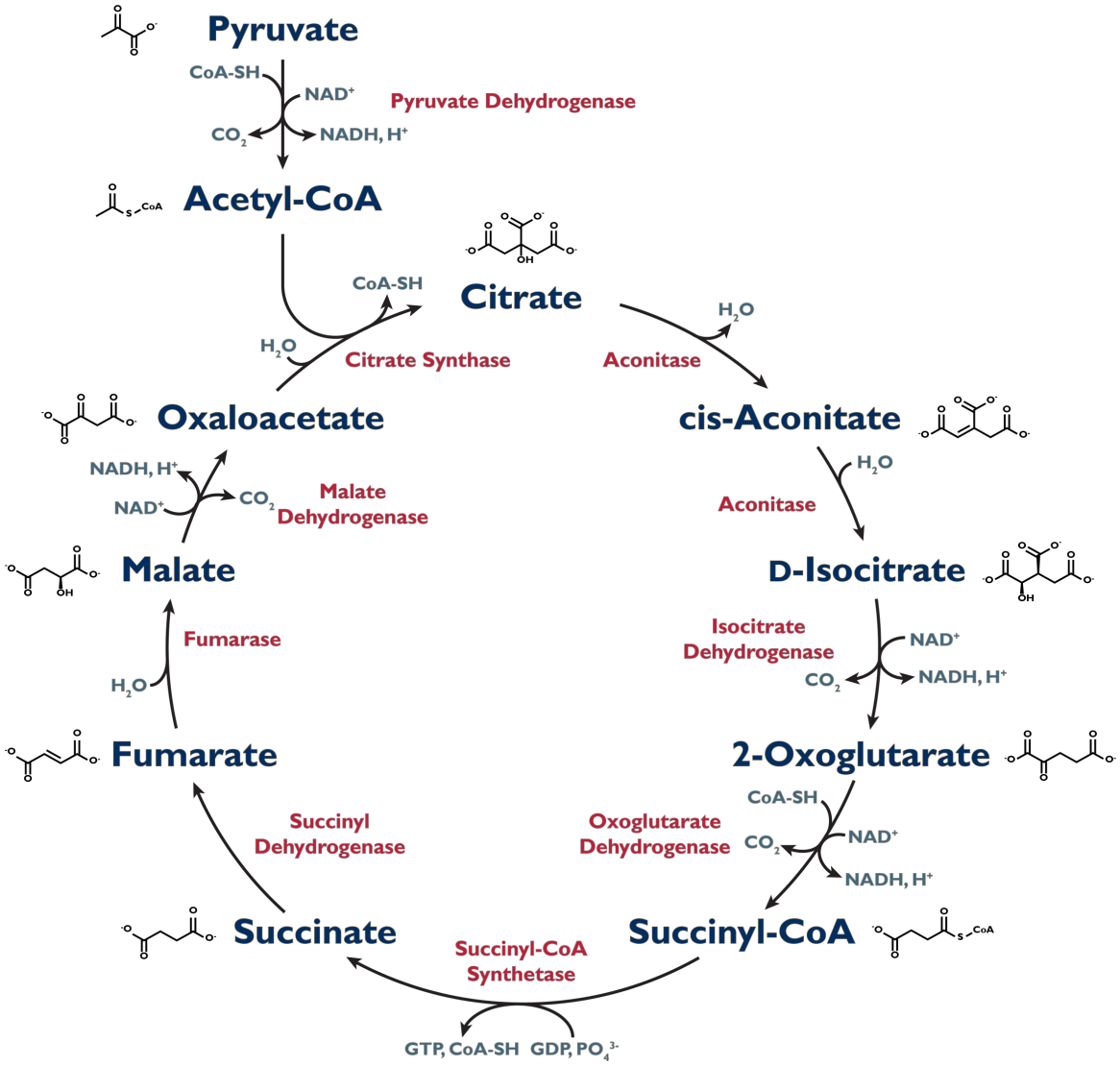


Figure 79 - The TCA Cycle. Enzymes/enzyme complexes are indicated in red. Cofactors and products/reactants are indicated in light blue. Pyruvate and acetyl-CoA enter the TCA cycle following glycolysis.

Intermediates in the TCA cycle (Figure 79) have been found to bind to and inhibit other members of the 2OG oxygenase superfamily[204]. Pyruvate (**44**), succinate (**2**) and glutamate (**45**) were tested as inhibitors of the collagen hydroxylases (the $\alpha(I)_2\beta_2$ C-P4H tetramer, P3H1 and PLOD1) using a radioactive 2OG turnover assay, and did not exhibit any significant inhibition of the collagen hydroxylases[205]. The TCA cycle intermediates can accumulate in disease[206], whilst mutations in genes encoding for isocitrate dehydrogenase (IDH) lead to a change of function variant, resulting in a build up of the two enantiomers of 2-hydroxyglutarate (2HG, (*R*)-2HG, **46**,

and (*S*)-2HG, **47**); these IDH variants and 2HG build up are associated with glioma and acute myeloid leukemia[207].

Pyruvate (**44**), alongside the small molecule intermediates in the TCA cycle (Citrate (**48**), *cis*-aconitate (**49**), D-isocitrate (**50**), succinate (**2**), fumarate (**51**), malate (**52**) and oxaloacetate (**53**)), with the exception of 2OG (**1**), were all assessed for inhibition using the MALDI-TOF MS assay at a final concentration of 100 µM and for stabilisation of the vCPH/Mn^{II} complex by differential scanning fluorometry (Figure 80A/B). Additionally, the two enantiomers of 2HG ((*R*)-2HG (**46**) and (*S*)-2HG (**47**)) and glutamate (**45**) were tested using the MALDI-TOF MS inhibition assay. 2,4-PDCA (**5**) was included in the 100 µM screen as a positive control.

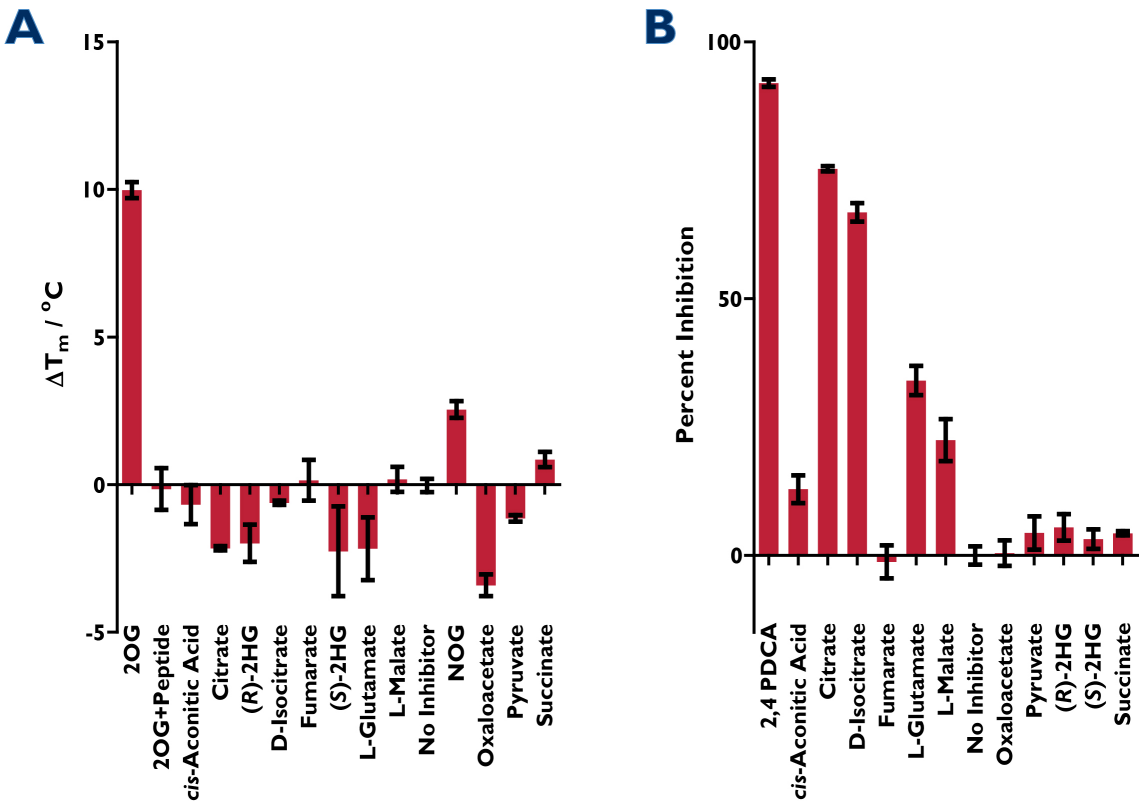


Figure 80 - Screening of TCA cycle intermediates as vCPH inhibitors. Compounds were tested as sodium salts with the exception of *cis*-aconitic acid and both 2HG isomers, which were tested as the acid form. A - Results of T_m analyses conducted *via* differential scanning fluorometry with vCPH and Mn^{II}. B - Results of inhibition screen with vCPH at an inhibitor concentration of 100 µM.

No stabilisation of the vCPH/Mn^{II} complex was observed by diferential scanning fluorometry (Figure 80A) by any of any of the TCA cycle intermediates with the exception of 2OG (**1**, ΔT_m = 9.97°C ± 0.27°C), the natural vCPH co-substrate. vCPH stabilisation conferred by the

presence of 2OG was lost upon the addition of Peptide 4, a vCPH substrate. Destabilisation of the vCPH/Mn^{II} complex was observed in the presence of citrate (**48**, ΔT_m = -2.16°C ± 0.07°C) and glutamate (**45**, ΔT_m = -2.17°C ± 1.06°C), which is also indicative of binding to vCPH.

Compounds which showed less than 50% inhibition at 100 µM were also tested at a final concentration of 1 mM. Both (*R*)-2HG (**46**) and (*S*)-2HG (**47**) showed greater than 50% inhibition at 1 mM and IC_{50} determination was conducted for these compounds. The remaining TCA cycle intermediates exhibited less than 50% inhibition at a final concentration of 1 mM and therefore have vCPH IC_{50} values greater than 1 mM (Figure 81).

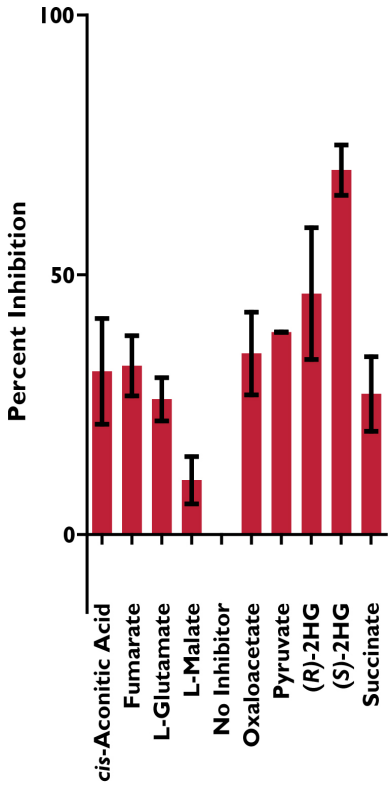


Figure 81 - Percent inhibition of vCPH observed at an inhibitor concentration of 1 mM. (*R*)-2HG (**46**) and (*S*)-2HG (**47**) exhibited greater than 50% inhibition at 1 mM.

Both citrate (**48**) and D-isocitrate (**50**) were found to have greater than 50% inhibition of vCPH at 100 µM (Figure 80B) and therefore IC_{50} values were subsequently calculated. Neither citrate (**48**, IC_{50} = 212.6 µM ± 7.1 µM) nor D-isocitrate (**50**, IC_{50} = 284.0 µM ± 26.8 µM) were potent vCPH inhibitors. Dose-response analysis showed that the IC_{50} values for both (*R*)-2HG (**46**) and (*S*)-2HG (**47**) were greater than 500 µM (Figure 82).

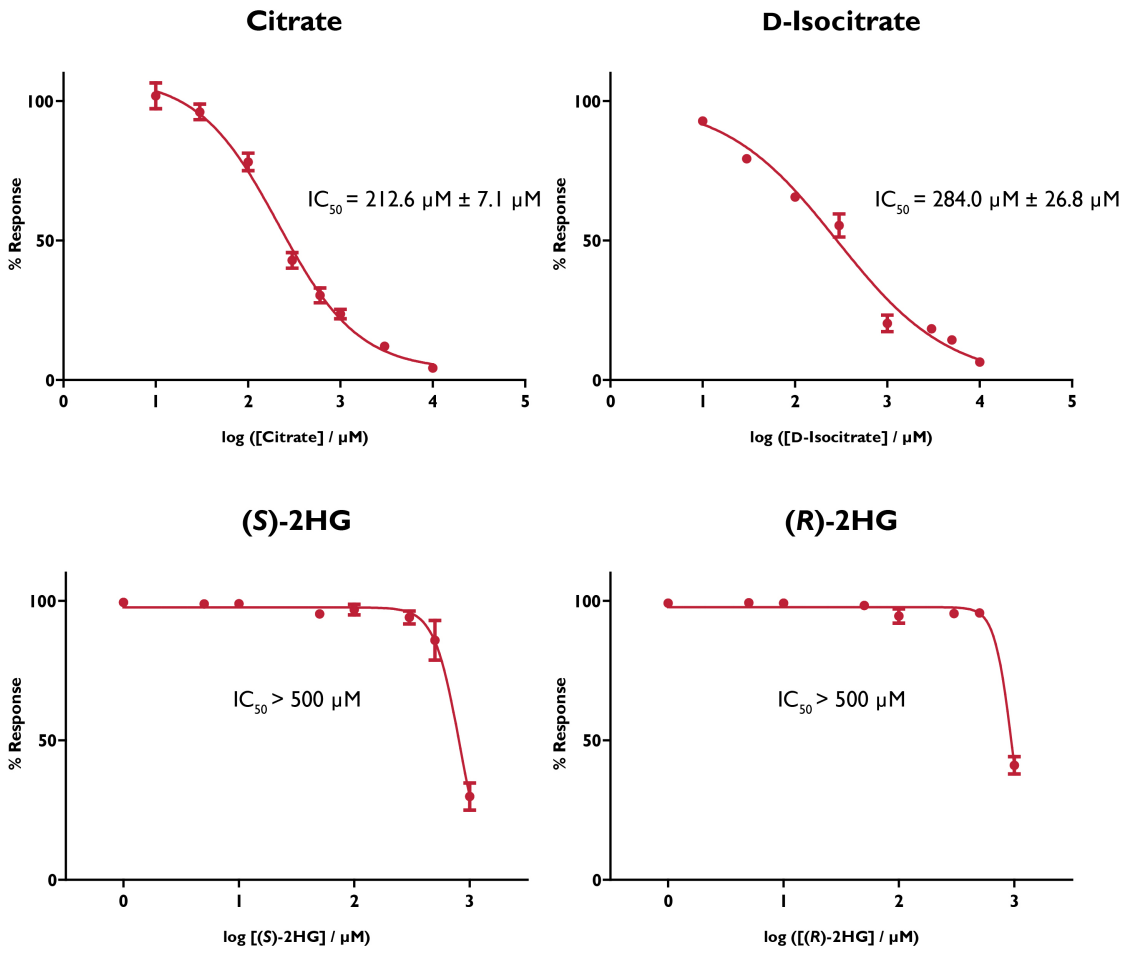


Figure 82- Dose response curves for vCPH-catalysed hydroxylation of Peptide 4 in the presence of citrate (**48**), D-isocitrate (**50**), (*R*)-2HG (**46**) and (*S*)-2HG (**47**).

High IC_{50} values for vCPH inhibition by (*R*)-2HG (**46**) and (*S*)-2HG (**47**) implies that collagen prolyl-4-hydroxylase activity is not affected by accumulation of 2HG, which builds up in some cancers due to change-of-function IDH variants[207]. This is contrast to observations with histone demethylases[204,208], but consistent with observations with the HIF prolyl hydroxylases[208].

Structural Studies

Although a structure of vCPH has been successfully solved *via* X-ray crystallography and deposited to the protein data bank[84], there are no available structures of vCPH in complex with inhibitors or with substrate. Since application of the vCPH purification protocol established in Chapter 3 afforded high purity protein that could be used in crystallisation trials, structural studies with vCPH were initiated and co-crystallisation of vCPH with peptide substrate and/or inhibitors was attempted. Since the reported crystallisation conditions resulted in twinned

crystals[84], novel crystal forms of vCPH were pursued and therefore crystallisation trials were conducted with a range of commercially available broad screens.

vCPH Final Concentration	Metal	Compound/Peptide Added	Screens Used (Temperatures)
22 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	NOG (**4**, 10 mM), Peptide 4 (10 mM)	PEGIon (25°C and 4°C) JCSG+ (25°C and 4°C) Salt Rx (25°C and 4°C) Xtal Screen (25°C and 4°C) PACT Premier (25°C and 4°C)
20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	FG-4592 (**8**, 10 mM)	JCSG+ (25°C and 4°C) PACT Premier (25°C and 4°C)
20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	2,4-PDCA (**5**, 10 mM)	JCSG+ (25°C and 4°C) PACT Premier (25°C and 4°C) PEGIon (4°C)
20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	NOG (**4**, 10 mM), Peptide 4 (10 mM)	PACT Premier (25°C and 4°C)
20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	2,4-PDCA (5, 10 mM)	PACT Premier D11 Optimisation Screen (25°C and 4°C)
20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	NOG (**4**, 10 mM)	PACT Premier D11 Optimisation Screen (25°C and 4°C)
20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	FG-4592 (**8**, 10 mM)	PACT Premier D11 Optimisation Screen (25°C and 4°C)
20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	NOG (**4**, 10 mM), Peptide 4 (10 mM)	PACT Premier D11 Optimisation Screen (25°C and 4°C)
20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	BIOX2 (**18**, 1mM)	JCSG+ (25°C and 4°C)
20 mg mL^{-1}	Ni^{II} (as $NiCl_2$)	-	-
20 mg mL^{-1}	Co^{II} (as $CoCl_2$)	-	-

Table 10 - Details of crystallisation trials conducted with vCPH. Treatment of vCPH with Ni^{II} and Co^{II} led to precipitation of protein and therefore crystallisation trials with these metals were not completed.

Crystallisation trials were conducted with compounds that were found to inhibit vCPH and were available in sufficient quantities for testing with multiple 96-well crystallisation screens. NOG (**4**), FG-4592 (**8**), 2,4-PDCA (**5**) and BIOX2 (**18**) were all tested and NOG (**4**) was also tested in combination with Peptide 4. vCPH crystallisation trials were conducted in the presence of Mn^{II} since compound binding was indicated by differential scanning fluorometry (Figure 70) and implied by NMR spectroscopy (Figure 75). Attempts to conduct crystallisation trials in the presence of Ni^{II} and Co^{II} (Ni^{II} and Co^{II} were found to stabilise vCPH by differential scanning fluorometry (Chapter 3, Figure 56)) resulted in the precipitation of the protein upon metal

addition to concentrated *apo*-vCPH (greater than 20 mg mL^{-1}) and therefore crystallisation trials were not possible.

All crystallisation trials were conducted at 4°C and 25°C. Any crystals that were obtained were subsequently harvested and screened; any protein crystals were then analysed using a synchrotron X-ray source. The conditions under which protein crystals were obtained are detailed in Table 11 and images of exemplar crystals are shown in Figure 83.

As crystals obtained from the PACT Premier D11 condition were observed after one week, optimisation of the D11 condition was conducted; the composition of this screen is detailed in Appendix 3.

Screen, Well & Temperature	Compound & Metal	Composition of Well Solution
PACT Premier C8 25°C	2,4-PDCA (5, 10 mM) Mn^{II} (as $MnCl_2$, 4 mM)	0.2 M Ammonium chloride 0.1 M HEPES pH 7.0 20% (w/v) PEG 6000
PACT Premier D11 25°C	2,4-PDCA (5, 10 mM) Mn^{II} (as $MnCl_2$, 4 mM)	0.2 M Calcium chloride dihydrate 0.1 M Tris 8.0 20% (w/v) PEG 6000
PACT Premier E9 25°C	2,4-PDCA (5, 10 mM) Mn^{II} (as $MnCl_2$, 4 mM)	0.2 M Potassium sodium tartrate tetrahydrate 20% (w/v) PEG 3350
PEGION D1 25°C	2,4-PDCA (5, 10 mM) Mn^{II} (as $MnCl_2$, 4 mM)	0.2 M Potassium sodium tartrate tetrahydrate 20% (w/v) PEG 3350, pH 7.4
PEGION C4 25°C	2,4-PDCA (5, 10 mM) Mn^{II} (as $MnCl_2$, 4 mM)	0.2 M Calcium acetate hydrate 20% (w/v) PEG 3350, pH 7.5

Table 11 - Conditions under which protein crystals were obtained.

Data sets were obtained with 3.51 Å and 3.97 Å resolution with crystals obtained from the PEGION C4 condition. The crystals were of the P62 2 2 space group and the unit cell dimensions are detailed in Table 12. Crystals originating from the PEGION D1 condition yielded data sets of 4.5 Å and 3.84 Å in the P63 2 2 and P61 2 2 space groups, respectively. Both of these crystal forms are different from the previously reported crystal form which was in the P1 2 1 1 space group and yielded data sets with a resolution of 1.6 Å[84]. Additionally, initial analyses have indicated that the crystals are not twinned, unlike the published structure[84].

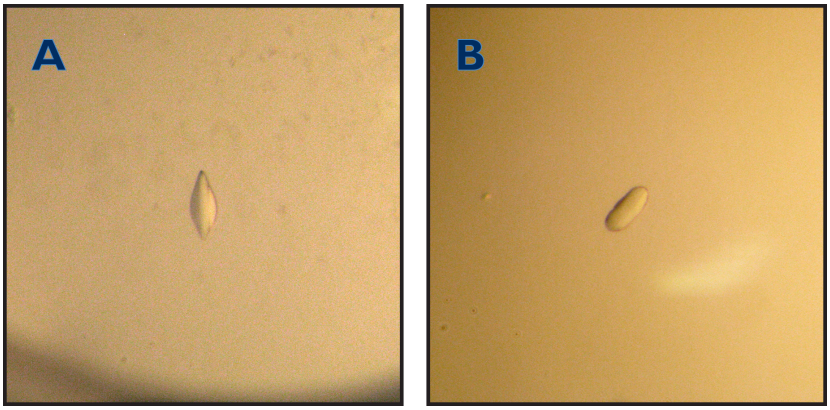


Figure 83 - Images of crystals observed in PEGION D1 (A) and PEGION C4 (B). Crystals grown under these conditions were harvested and afforded the data sets detailed in Table 12.

Ideally, data sets with a resolution of 2.5 Å and lower are required for analysis of inhibitor binding modes. Therefore, further optimisation of the crystallisation conditions is required to allow for analysis of inhibitor binding to vCPH.

Condition	Resolution	Space Group	Unit Cell
PEGION C4	3.97 Å	P62 2 2	A = 70.20 Å B = 70.20 Å C =201.04 Å α = 90° β = 90° ɣ = 120°
PEGION C4	3.51 Å	P62 2 2	A = 70.16 Å B = 70.16 Å C =200.59 Å α = 90° β = 90° ɣ = 120°
PEGION D1	4.50 Å	P63 2 2	A = 70.70 Å B = 70.70 Å C =201.04 Å α = 90° β = 90° ɣ = 120°
PEGION D1	3.84 Å	P61 2 2	A = 141.24 Å B = 141.24 Å C =204.35 Å α = 90° β = 90° ɣ = 120°

Table 12 - Resolution, space groups and unit cell details for protein crystals that successfully diffracted X-rays from a synchrotron source.

Conclusions

Following the development of a MS-based peptide hydroxylation assay for vCPH, three compound libraries were screened against vCPH for inhibition. These three libraries were a selection of commercially available inhibitors, many of which were known inhibitors of other 2OG oxygenases (Table 7), a selection of novel inhibitors of the HIF prolyl-4-hydroxylase PHD2 (Table 9) and the intermediates in the TCA cycle (Figure 79). All inhibitors were screened at a final concentration 100 µM prior to IC_{50} determination for any compounds that exhibited greater than 50% inhibition. Compounds were also assessed for vCPH binding by differential scanning fluorometry. 2,4-PDCA (**5**) and FG-0041 (**7**), which were developed as inhibitors for the

C-P4Hs[137,183] are potent inhibitors of vCPH, which supports the use of vCPH as a model for C-P4H inhibition.

Name		PHD2 IC_{50} / µM	vCPH IC_{50} / µM	vCPH K_i / µM
2,4-PDCA	**5**	1.25[203]	5.3 ± 0.3	3.3 ± 0.6
BIOX2	**18**	-	12.1 ± 0.5	-
CCT 2-66	**15**	*96*[158,185]	> 100*	-
CCT 6-140	**19**	*> 100*[158]	> 100	-
Ciprofloxacin	**20**	-	> 100	-
***cis*-Aconitinic Acid**	**49**	-	> 1000*	-
Citrate	**48**	-	212.6 ± 7.1	-
Daminozide	**16**	-	> 100*	-
D-Isocitrate	**50**	-	284.0 ± 26.8	-
FG-0041	**7**	-	15.7 ± 1.4	-
FG-2216	**21**	-	4.5 ± 0.1	1.7 ± 0.1
FG-4592	**8**	2.76[203]	5.0 ± 0.1	2.3 ± 0.1
Fumarate	**51**	-	> 1000*	-
Glutamate	**45**	-	> 1000*	-
GSK1278863	**10**	0.093[203]	13.4 ± 0.9	46.4 ± 5.7
GSKJ1	**22**	-	15.9 ± 6.3	-
IOX1	**14**	*14.3*[157]	47.2 ± 2.7	-
IOX1 Prodrug	**23**	-	> 100	-
IOX2	**24**	0.09[203]	8.5 ± 0.2	-
IOX4	**25**	0.54[203]	2.5 ± 0.2	-
JHM2-167	**34**	0.540[203]	> 100	-
JHM2-174	**35**	*92% inhibition at 10 µM*[203]	1.1 ± 0.1	-
JHM3-201	**36**	0.258[203]	28.2 ± 1.0	-
JHM3-202	**37**	*100% inhibition at 10 µM*[203]	2.0 ± 0.1	-
JHM377	**38**	PHD2 Activation (61%)[203]	> 100	-
JHM388	**39**	PHD2 Activation (50%)[203]	> 100*	-
JHM532	**40**	0.145[203]	> 100*	-
JHM631	**41**	0.640[203]	> 100	-
JHM634	**42**	0.025[203]	> 100*	-
JHM642	**43**	0.087[203]	16.5 ± 0.7	-
Lomefloxacin	**26**	-	> 100*	-
Malate	**52**	-	> 1000*	-
Meloxicam	**27**	-	> 100	-
Mildronate	**13**	-	> 100	-
Molidustat	**11**	1.32[203]	26.8 ± 2.4	14.8 ± 1.8
Nitisinone	**28**	-	> 100	-
NOFD	**29**	-	> 100*	-
NOG	**4**	2.38[203]	21.2 ± 2.1	43.5 ± 10.5
Oxaloacetate	**53**	-	> 1000*	-

Name		PHD2 IC_{50} / μM	vCPH IC_{50} / μM	vCPH K_i / μM
Pyruvate	**44**	-	> 1000*	-
(*R*)-2HG	**46**	*7300*[208]	>500*	-
(*S*)-2HG	**47**	*419*[208]	>500*	-
SAHA	**30**	-	> 100	-
Succinate	**2**	-	> 1000*	-
Vadadustat	**9**	1.33[203]	< 1	3.5 ± 0.4

Table 13 - Collated vCPH inhibition data. vCPH IC_{50} values and K_i values are indicated with the standard error measurement. PHD2 IC_{50} values were either obtained using a LC-MS assay (PHD2 (150 nM), 2OG (10 μM), Fe (10 μM), CODD (5 μM))[203] or a MALDI-TOF MS peptide hydroxylation assay[158]. Values obtained using a MALDI-TOF assay are italicised and are not directly comparable to values obtained using an LC-MS peptide hydroxylation assay. If some vCPH inhibition was observed at an inhibitor concentration of 100 μM (or 1 mM for TCA cycle intermediates), values are indicated with an asterisk. Dashes indicate where values have not been determined.

The PHD2 inhibitors that are currently in clinical trials, FG-4592 (**8**), GSK1278863 (**10**), Molidustat (**25**) and Vadadustat (**11**), along with IOX4 (**9**), have been shown to inhibit other 2OG dependent prolyl-hydroxylases[194] and, with the exception of FG-4592, also inhibit the aspartyl hydroxylase FIH[194]. Given the lack of selectivity exhibited by these compounds, it is unsurprising that they are also strong inhibitors of vCPH (Table 14). Full comparison of the vCPH inhibition data to that for the HIF prolyl hydroxylase PHD2 is hampered by the lack of a complete data set. Furthermore, PHD2 inhibition assays have been conducted using different assay platforms that do not allow for cross-comparison (mainly since they were conducted with differing enzyme concentrations). Values listed in Table 13 were obtained using MALDI-TOF MS[158] and LC-MS[194,203] peptide hydroxylation assays whilst studies monitoring 2OG turnover[87] and using AlphaScreen[190] have also been reported.

		IC_{50} / μM			
Inhibitor		**PHD2**[203]	**FIH**[194]	**OGFOD1**[194]	**vCPH**
FG-4592	**8**	2.760	> 100	< 1	5.0 ± 0.1
GSK1278863	**10**	0.093	21	2.1	13.4 ± 0.9
IOX4	**25**	0.540	66	5.1	2.5 ± 0.2
Molidustat	**11**	1.320	31	< 1	26.8 ± 2.4
Vadadustat	**9**	1.330	29	1.4	< 1

Table 14 - Selectivity data for clinical PHD2 inhibition candidates[194,203] collated with vCPH inhibition values. Assay platforms differ and therefore quantitative cross-comparisons cannot be made.

Comparison of the potency of GSK1278863 (**10**), FG-4592 (**8**), IOX4 (**25**), Molidustat (**11**) and Vadadustat (**9**) (Table 14) indicates that potency is not the same across the tested prolyl

hydroxylases PHD2, FIH and vCPH. For example, whereas Molidustat (**11**) and Vadadustat (**9**) have similar inhibition of PHD2, FIH and OGFOD1, Vadadustat (**9**) is a highly potent inhibitor of vCPH whilst Molidustat (**11**) is the least potent of the PHD2 inhbitors that are in clinical trials (Table 14). It should, however be noted that the structurally similar IOX4 (**25**) is a potent inhibitor of vCPH, PHD2 and OGFOD1, illustrating that small modifications could be made to the pharmacophore to tune-in desired selectivity profiles.

NMR studies have indicated that NOG (**4**), 2,4-PDCA (**5**), FG-4592 (**8**) and Vadadustat (**9**) compete with 2OG since they were found to displace 2OG from the vCPH/Mn^{II} complex. K_i determination for GSK1278863 (**10**), FG-4592 (**8**), FG-2216 (**21**), Molidustat (**11**), Vadadustat (**9**), NOG (**4**) and 2,4-PDCA (**5**) revealed that the relative affinities of the compounds were broadly in line with the vCPH IC_{50} values. Molidustat (**11**, K_i = 14.83 μM ± 1.81 μM, IC_{50} = 26.8 μM ± 2.4 μM) was found to bind more strongly than the more potent GSK1278863 (**10**, K_i = 46.36 μM ± 5.69 μM, IC_{50} = 13.4 μM ± 0.9 μM) and Vadadustat (**9**, K_i = 3.45 μM ± 0.42 μM, IC_{50} < 1 μM). To test for competition between inhibitor and substrate an alternative assay platform is required, since the MALDI-TOF MS assay is not sensitive at high peptide concentrations.

The inhibition of vCPH by GSKJ1 (**22**) suggests that the in-cell antifibrotic activity of GSKJ4[209] (**33**), an ethyl ester prodrug form of GSKJ1 (**22**), may be due to C-P4H inhibition. Inhibition of the HIF-prolyl hydroxylases has also been reported to lead to an anti-fibrotic response[200]. Understanding the impact of these inhibitors on collagen biosynthesis is required before a full understanding of the underlying mechanism of action can be achieved. Furthermore, analysis of the structures of the fluoroquinolone antibiotics, which have been implicated in tendon rupture[186], and comparison to known inhibitors of PHD2 and vCPH (Figure 72) reveals that the glycinamide form of the fluoroquinolones may be inhibitors of 2OG-dependent prolyl hydroxylases. These proposed inhibitors (Figure 72B) pay a striking resemblance to GSK360A

(**31**), which has been shown to reduce collagen deposition and fibrosis following tendon rupture[200].

Analysis of the novel PHD2 inhibitors[203] and the OGFOD1 inhibitor CCT6-140 (**19**)[158] suggests that the selectivity profile of the clinical candidates can be improved. CCT6-140 (**19**) is an *in vitro* sub-micromolar inhibitor of OGFOD1 but does not inhibit PHD2 or vCPH[158] whilst JHM634 (**42**) and JHM642 (**43**) are potent inhibitors of PHD2, with IC_{50} values similar to clinical candidates such as FG-4592 (**8**) and GSK1278863 (**10**) but do not inhibit vCPH or OGFOD1[203]. These studies show that selectivity between PHD2, OGFOD1 and collagen prolyl-4-hydroxylases can be achieved. Attempts to develop selective compounds for the human C-P4Hs first require the development of reliable enzyme production and assay systems for C-P4H. The effect of the inhibitors on the collagen P3Hs is currently unknown as, similar to the C-P4Hs, reliable recombinant protein production protocols and assays have not been developed. Understanding the effect of prolyl hydroxylase inhibitors on the collagen P3Hs is required before they can be described as truly selective.

Structural studies with vCPH have been initiated as no structure of vCPH in complex with a substrate or an inhibitor has been obtained. Co-crystallisation trials have been conducted, using the results of differential scanning fluorometry analyses to inform the choice of compounds and metals to test. Novel crystal forms have been obtained and X-ray diffraction data have been collected for these crystals. However, crystallisation conditions require optimisation since the best quality data sets had a resolution of 3.51 Å. These initial studies give an ideal basis from which to optimise crystallisation conditions.

Summary

A MALDI-TOF MS based inhibition assay has been developed for a viral model collagen prolyl-4-hydroxylase, and was used alongside differential scanning fluorometry to analyse a range of potential vCPH inhibitors. These analyses revealed that a number of compounds that

are known to inhibit other 2OG oxygenases also inhibit vCPH. In particular, inhibitors of the HIF prolyl-4-hydroxylase PHD2 in clinical trials, such as GSK1278863 (**10**), FG-4592 (**8**), Molidustat (**11**) and Vadadustat (**9**) have all been shown to inhibit vCPH. In addition to IC_{50} determination, studies have been conducted to calculate K_i values for these clinical candidates, the clinical PHD2 inhibitors bind to vCPH with high affinity, particularly in the cases of FG-4592 (**8**) and Vadadustat (**9**).

Assessment of novel PHD2 inhibitors[203] and CCT6-140, an OGFOD1 inhibitor[158], has also been conducted with vCPH. This work has revealed that compounds that inhibit PHD2[203] and OGFOD1[158] have been developed that do not inhibit vCPH (or the other prolyl hydroxylase). Therefore, achieving selectivity between PHD2, OGFOD1 and the collagen prolyl-4-hydroxylases may be possible. However, to develop a selective inhibitor for human C-P4H, the development of a reliable human C-P4H assay is first required. Additionally, an inhibition assay for the collagen P3Hs would further aid the development of selective inhibitors of 2OG-dependent prolyl hydroxylases.

Structural studies with vCPH have also been initiated with a view to obtain co-crystal structures with inhibitors and substrate. Novel crystal forms have been identified but high-resolution data sets have not yet been obtained. Therefore, further optimisation of crystallisation conditions are required before the binding modes of vCPH inhibitors can be analysed in greater detail.

Chapter 5
Amino Acid Analysis of Collagen

Tendon Structure

Tendon is a connective tissue which connects muscle to bone and is predominantly formed from ECM. Approximately 86% of the dry mass of tendon is collagen, with type I collagen making up in excess of 95% of the collagen component[3,7]. In tendon, collagen fibrils are often orientated in parallel in order to efficiently transmit the forces generated by muscular contractions. The fibrils are organised into fibres that are arranged around embedded cells (tenocytes), which in turn make up the fascicle, and the final tendon structures (Figure 84).

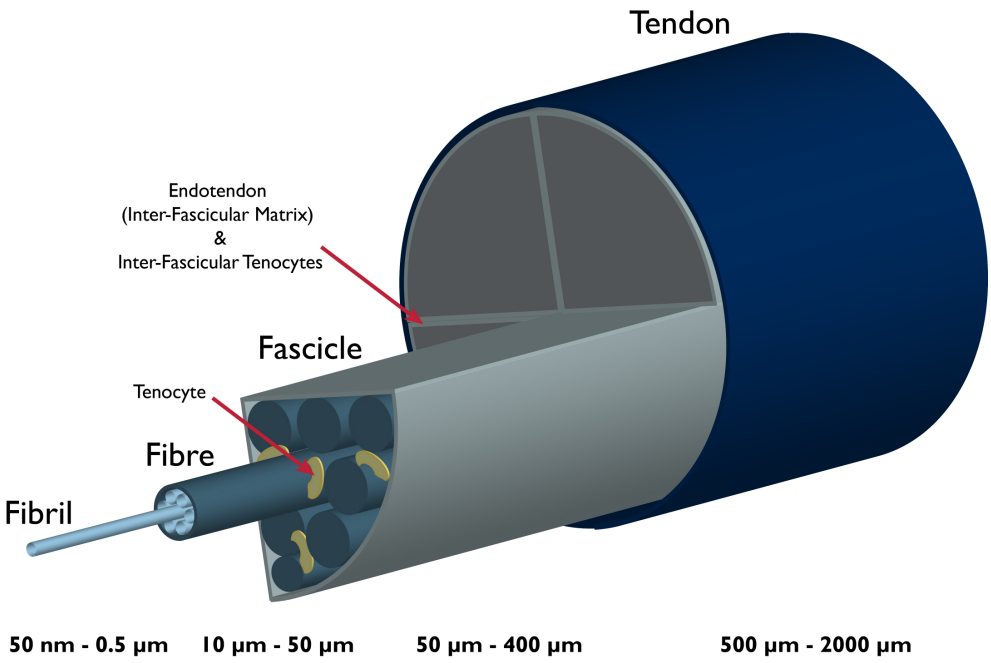


Figure 84 - Schematic Representation of Tendon Structure. Fibrils consist primarily of collagen type I and and are aligned to construct fibres that are in turn arranged in one of a number of possible manners (*e.g.* parallel/crossed *etc.*) to create fascicles that are the final building block of tendon. Tenocytes are embedded in the fascicles and inter-fascicular matrix, and are responsible for the production of new ECM.

Mature tenocytes or maturing tenoblast cells[3] account for more than 85% of the cells which are embedded in the tendon ECM, and are responsible for the production of fresh ECM components to replace and regenerate the tissue as it ages. This process does not replace ECM in a like-for-like manner; for example, the collagen content in tendon reduces in animals with aging[210]. This phenomenon is not limited to tendon; incidence of disease-related changes in collagen-rich structures tends to increase over time, as exemplified by increases in the incidence of osteoporosis with age[211-213].

The Rotator Cuff

The rotator cuff consists of the four muscles (the *infraspinatus*, *subscapularis*, *supraspinatus* and *teres minor*) and associated tendons that connect to the *scapula* (the shoulder blade); these four muscles are also attached to the *humerus*. There are three other scapulohumeral muscles that do not form part of the rotator cuff: the deltoid muscle, *coracobrachialis* muscle and *teres major* muscle[214].

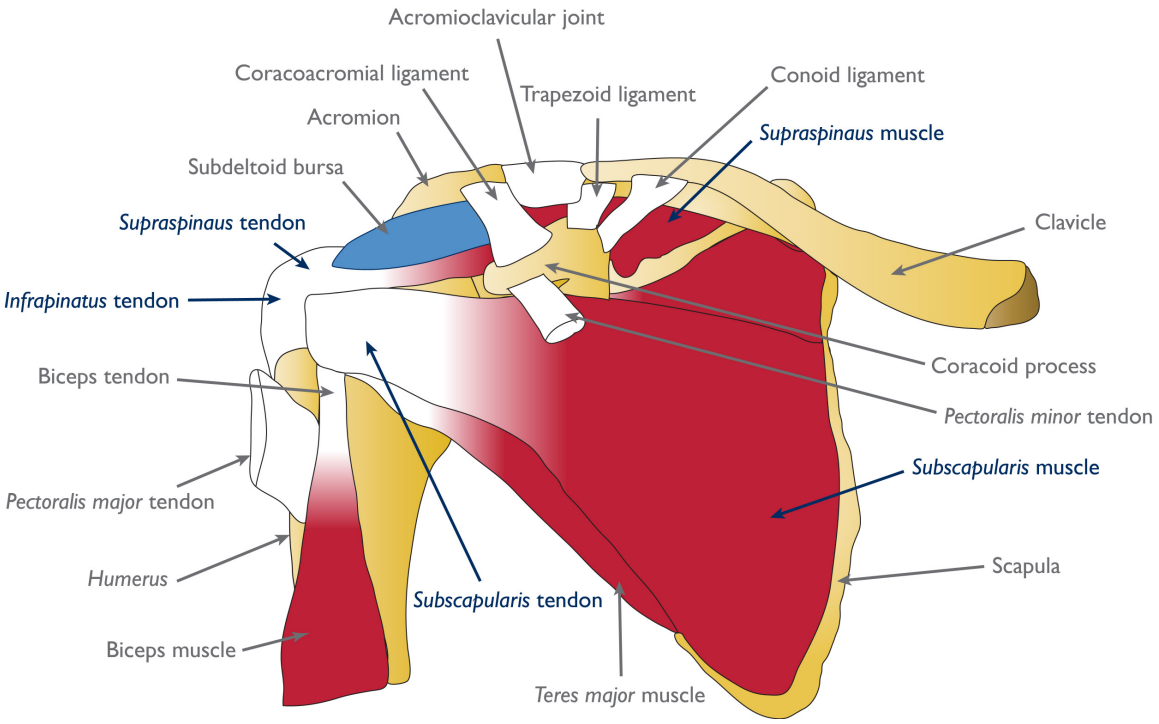


Figure 85 - Musculoskeletal view of the anterior of the right shoulder. Muscles and tendons that form the rotator cuff are indicated in blue. The *infapinatus* muscle, *teres* minor muscle and *teres* minor tendon are obscured and are behind the joint.

The rotator cuff is, as the name suggests, responsible for rotational motion about the shoulder joint. Additionally, the rotator cuff plays a key role in maintaining stability of the joint during movements mediated by the non-rotator cuff scapulohumeral muscles. Therefore, rotator cuff tendons are important in maintaining the integrity of the shoulder joint and for ensuring that the forces exerted by the scapulohumeral muscles result in motion about the shoulder.

Rotator Cuff Disease

Rotator cuff disease ranges from mild inflammation through to complete tearing of the tendon. Changes in the ECM composition and the makeup of the collagen component appear to have an effect on the likelihood of tendon injury[13,215]. For example, analysis of collagen content in injured rotator cuff has suggested that there is an increase in collagens type II and III, despite an overall reduction in collagen content injured tendon[216].

The collagen proteins themselves also differ between diseased and healthy tendon, with the thermal stability decreasing in tendons with full or partial rupture[217], indicative of underlying weakness in the collagen. This may arise due to alterations in the cross linking of the collagen structures, or is reflective of a reduction in the levels of hydroxylated proline residues in the collagen of diseased tendons, since the presence of (*2S,4R*)-hydroxyproline has long been known to increase the thermal stability of collagen[29].

Incidence of rotator cuff disease, in particular tearing or rupture of the tenson, increases with age[218,219]; this is indicative that age-related changes in tendon ECM play a role in the weakening of the rotator cuff and subsequent presentation of adverse symptoms. Non-invasive treatment of rotator cuff disease may be successful[220], but in many cases surgical intervention is required. A range of different surgical approaches, such as patches and supports, have been used to aid repair and healing[221]. Regularly, patients present with severe tears and/or shoulder fibrosis, potentially due to late referral and diagnosis[222].

Due to the wide range of motions within the shoulder that the rotator cuff tendons and muscles transmit, repairing damaged rotator cuff, in particular torn rotator cuff, is non-trivial; scar tissue is not as strong as healthy tissue and failure rates of repaired rotator cuff tendon are high[223,224]. In performance athletes, such as pitchers in professional baseball, rotator cuff injury is one of the most feared ailments[225]. Recently, a study was conducted to assess the effectiveness of surgical intervention[224,226], making a range of tissue samplespotentially available for analysis.

Analysing Hydroxylation in Collagen

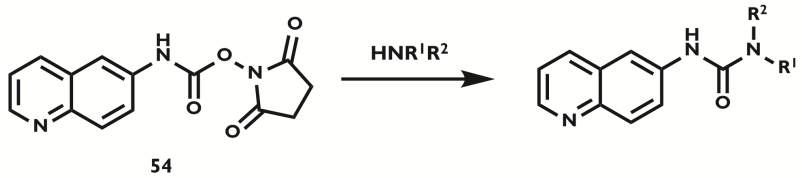
Proteomic analyses, particularly MS-MS based approaches, have been used to analyse the ECM[14] and identify a range of post-translational modifications and alterations to the environment. Such approaches have been applied to analyse the collagen component of ECM and identify hydroxylation sites[14,227,228].

MS-MS analyses, despite allowing for localisation of PTMs (subject to adequate ion coverage in product ion spectra), do not allow for accurate determination of the stereo- and regio-chemistry of the modification. This can be achieved *via* amino acid analysis, but can only inform on the global levels of the amino acids in question since the sample is first hydrolysed completely before chromatographic analysis.

This is particularly relevant for the analysis of collagen since MS-MS analyses are unable to distinguish between the different isomers of hydroxyproline and hydroxylysine. Within the Gly-Xaa-Yaa repeat of collagen, it is possible to infer that any hydroxyproline residues at the Yaa position are (*2S,4R*)-hydroxyproline. However, the potential presence of both 3- and 4-hydroxyproline at the Xaa position of the Gly-Xaa-Yaa repeat[27] makes it impossible to draw fully quantitative conclusions about the levels of the different hydroxylation products from MS-MS experiments. Therefore, appropriate chromatographic analyses of a fully hydrolysed protein sample is required in order to separate all of the post-translationally modified amino acids and quantify the global abundances of each modification in a sample.

Such chromatographic amino acid analyses regularly consist of complete hydrolysis of the protein prior to derivitisation, where a functional group such as the fluorenylmethyoxycarbonyl (Fmoc) group is added to a nucleophilic site on the amino acid (usually the free amine) to improve chromatographic separation and increase retention time. Derivatisation-free methods have also been developed, but often provide poorer separation; such derivatisation-free methods have applied for δ-^{13}C analyses of collagen samples[229].

Amino acid analyses have historically been conducted using high performance liquid chromatography (HPLC) methodologies. For analyses of collagen samples, the concentration of hydroxyproline has been used to calculate collagen content of samples[230,231], therein assuming the level hydroxylation of procollagen proline residues is constant across tissues, and removing the ability to assess hydroxylation levels in a quantitative manner.



54

Figure 86 - Derivatisation of a primary or secondary amine by 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate (AQC, **54**).

Improvements to HPLC components have given rise to ultra-high performance liquid chromatography (UPLC), which utilises higher pressures and improved columns to afford greater resolution than more traditional HPLC methods[232]. Through the use of the appropriate column conditions and utilising derivatisation reagents such as 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate (AQC, **54**), which reacts with all primary and secondary amines (Figure 86); UPLC-based analyses can be used to separate the different regio- and stereo- isomers of hydroxyproline and hydroxylysine[113,116]. This makes it possible to use UPLC amino acid analysis (an exemplar liquid chromatogram is shown in Figure 87) to investigate post-translational modifications in greater detail than more traditional HPLC/Fmoc based analyses.

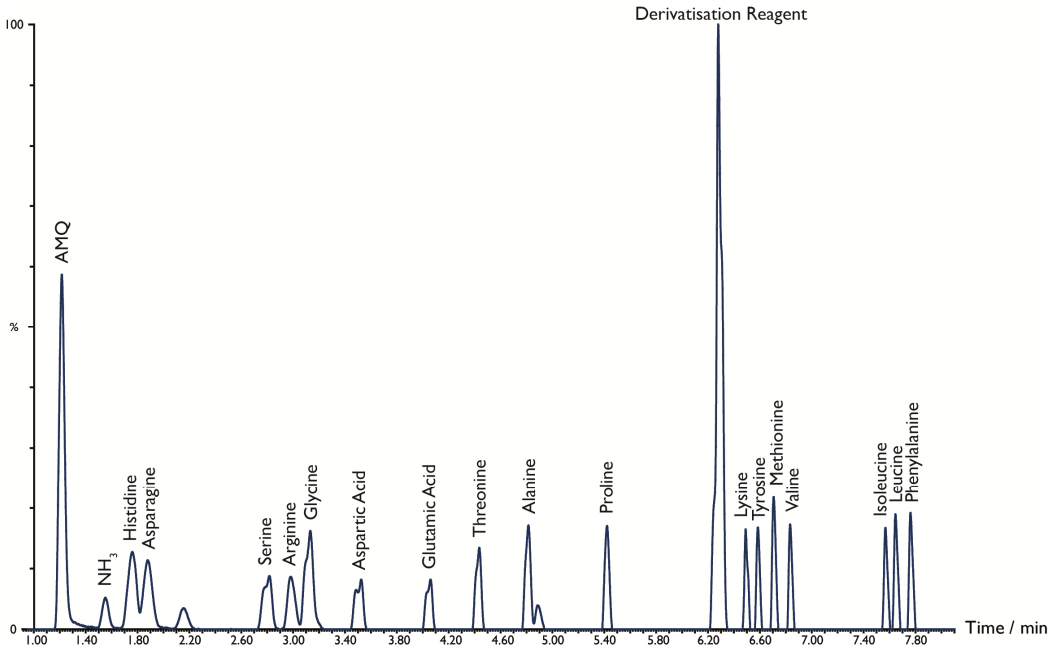


Figure 87 - Liquid chromatogram (absorbance at λ = 260 nm) following UPLC-MS AAA of amino acid standards. A 10 minute amino acid protocol was used (Chapter 8). Amino acid peaks are labelled and were identified by analysis of the corresponding mass spectra.

Proyl- and Lysyl- Hydroxylation is Not Necessarily Limited to Proline C4, Proline C3 and Lysine C5

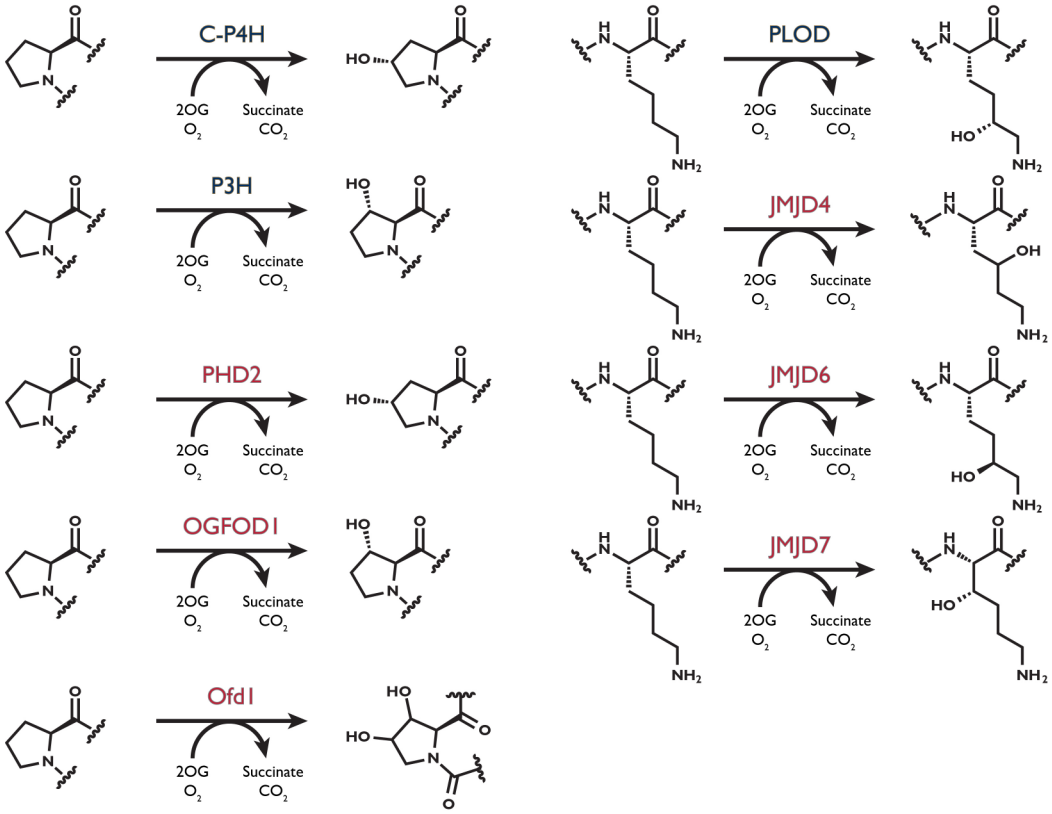


Figure 88 - Post-translational modifactions of lysyl- and prolyl- residues. Enzymes that catalyse the formation of the PTM are indicated. Stereochemistry is indicated where it has been successfully resolved.

Prolyl- and lysyl- hydroxylation is not limited to collagen (Chapter 1) and a range of different regio- and stereo- isomers of both hydroxyproline and hydroxylysine are formed post-translationally (Figure 88). Although the presence of (*2S,4R*)-hydroxyproline[28], (*2S,3S*)-hydroxyproline[62] and (*2S,5R*)-hydroxylysine[68] in collagen have been well established, the presence of alternative regio- and stereo- isomers has not been investigated.

In addition to the singly hydroxylated residues, prolyl residues can be 3,4-dihydroxylated (Figure 88), formation of *trans*-3,4-dihydroxy-L-proline is catalysed by the yeast OGFOD1 homologue Ofd1[114] and *trans*-3,4-dihydroxy-L-proline residues have also been identified in fungi[233,234]. 3,4-dihydroxyproline has not been identified in animals to date.

Aims

The goals of this study were to investigate the feasibility of conducting amino acid analysis on collagen isolated from tissue samples originating from the United Kingdom Rotator Cuff Trial (UKUFF) trial[224]. The study was therefore split into two branches:

1. Optimise hydrolysis of collagen samples and derivatisation conditions to allow for the reliable collection of good quality spectra.
2. Establish a protocol for the removal of non-collagen components from tendon samples.

Amino Acid Analysis of Commercially Sourced Collagen

Commercially-sourced collagen samples from bovine achilles tendon and rat tail tendon were used for initial studies to optimise hydrolysis and sample preparation. Two approaches were used to hydrolyse the samples: acid hydrolysis and enzymatic hydrolysis. Enzymatic hydrolysis was initially conducted with a commercially-sourced protease from *Streptomyces griseus*[235], which can be used for complete degradation of proteins/peptides to their constituent amino acids (and has been used for successful hydrolysis of samples that contain modified lysine residues[113,135]). Acid hydrolysis was conducted with a solid sample in 6M HCl (aq) vapour at 110°C. Following hydrolysis, the samples were treated according to the AccQ-Tag™ protocol and

then analysed using one of two UPLC-MS amino acid analysis methods (either a 10 or 30 minute method, detailed in Chapter 8). Amino acids were not observed by UPLC-MS for any of the samples due to incomplete hydrolysis and therefore hydrolysis protocols required optimisation.

Optimisation of Hydrolysis Conditions

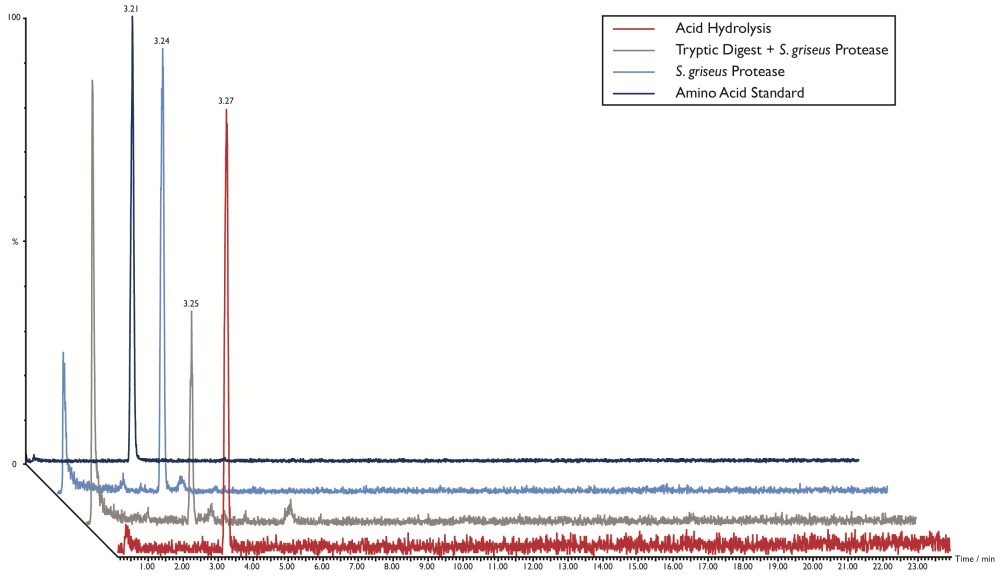


Figure 89 - Extracted ion chromatograms for *m/z* = 246, corresponding to derivatised glycine, from 30 minute amino acid analysis of hydrolysed bovine achilles tendon samples. Peaks corresponding to the presence of derivatised glycine are observed at approximately 3.27 minutes in each sample.

Hydrolysis of collagen was attempted using three approaches, either enzymatic digest using a protease from *S. griseus*, acid hydrolysis or a two-enzyme approach using trypsin followed by the *S. griseus* protease. Analysis of samples processed by all three hydrolysis methods resulted in the appearance of a peak corresponding to derivatised glycine, which was observed in the extracted ion chromatograms (Figure 89) for all three hydrolysis approaches. A peak corresponding to derivatised glycine was observed in the liquid chromatogram from the acid hydrolysed rat tail tendon sample (Figure 90).

The enzymatic hydrolysis was not optimised since hydrolysis with *S. griseus* protease of short, hydroxyproline-containing peptides, conducted as part of studies with prolyl-4-hydroxylases from *Dictyostelium discoideum* failed to fully hydrolyse the sample, implying that the *S. griseus* protease is unable to hydrolyse amide bonds adjacent to hydroxyproline.

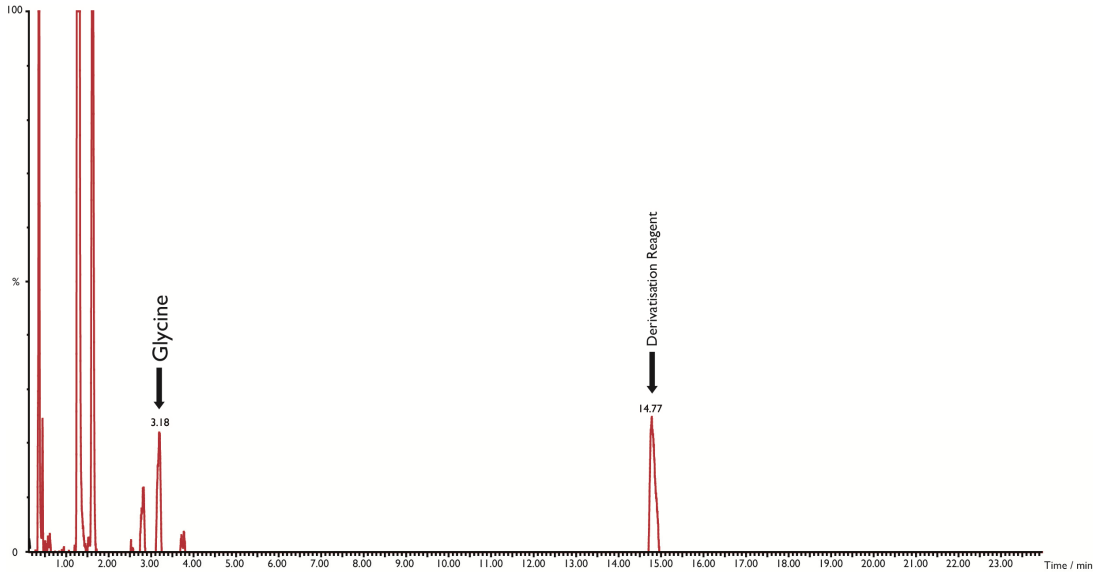


Figure 90 - Liquid chromatogram (λ = 260 nm) resulting from a 30 minute amino acid analysis method run with an acid hydrolysed sample of rat tail tendon.

To allow the derivatisation reaction (Figure 86) to successfully proceed, the pH of the reaction solution needs to be greater than 8. However, samples that were acid hydrolysed were still strongly acidic when the derivatisation was initiated, leading to a poor progression of the reaction. Acid hydrolysis protocols were therefore modified to ensure that samples were pH adjusted (if required) prior to resuspension and derivatisation. Following this, the derivatisation reaction was successful and high quality data (peaks corresponding to a range of amino acids were observed) was achieved using both the 10 and 30 minute amino acid analysis methods (column conditions and gradients detailed in Chapter 8).

Results of Amino Acid Analysis with Rat Tail Tendon and Bovine Achilles Tendon

Using the optimised hydrolysis approach, UPLC-MS amino acid analysis of both commercial collagen samples (collagen from bovine Achilles tendon and Rat tail tendon) was achieved. Peaks corresponding to AQC-derivatised amino acids were observed by UV absorbance (Figure 91) and were assigned by comparison of retention times to an amino acid standard and by MS analyses of the liquid chromatogram. These data were obtained by hydrolysing 125 µg of collagen which was then resuspended in 50 µL of resuspension buffer/derivatisation reagent mix (Chapter 8). 1 µL of the resulting sample was subsequently injected for UPLC-MS amino acid analysis.

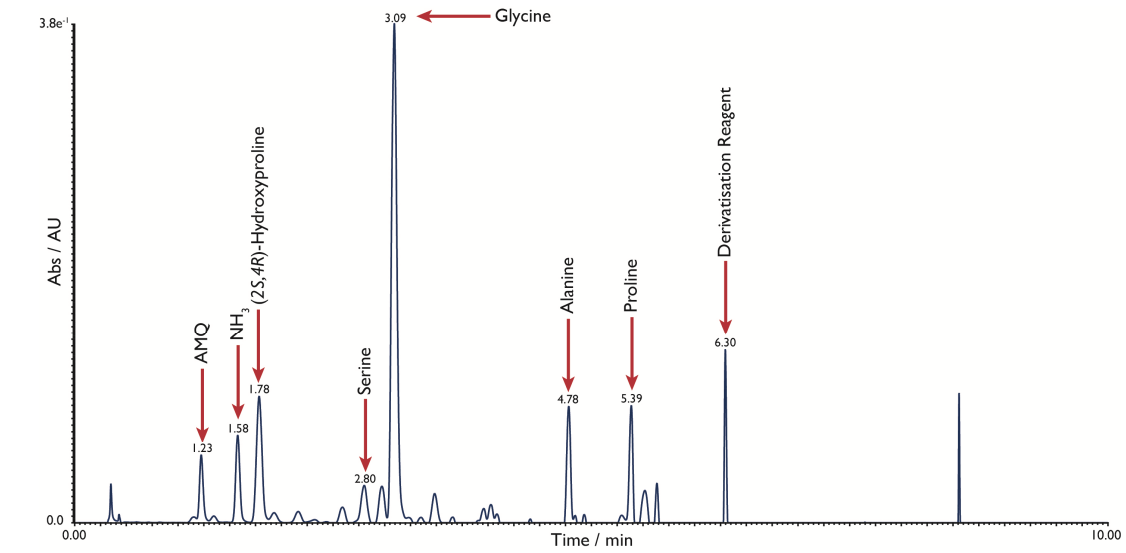


Figure 91 - Liquid chromatogram of a 10 minute UPLC-MS AAA method used to analyse acid hydrolysed and AQC (**54**) derivatised bovine Achilles tendon. Monitoring was conducted at λ = 260 nm. Peaks were assigned following comparison to amino acid standards and analysis of corresponding mass spectra.

The amino acid composition of the commercially-sourced bovine Achilles and rat tail tendon samples was broadly similar (Figure 92), which was unsurprising due the high similarity of collagen across species (The peptide sequences resulting from expression of COL1A1 in *Rattus norvegicus* and *Bos taurus* share 92% identity). Integration of liquid chromatogram peaks allows for calculation of ratios of the various amino acids to one another, while an internal standard is required for fully quantitative analysis.

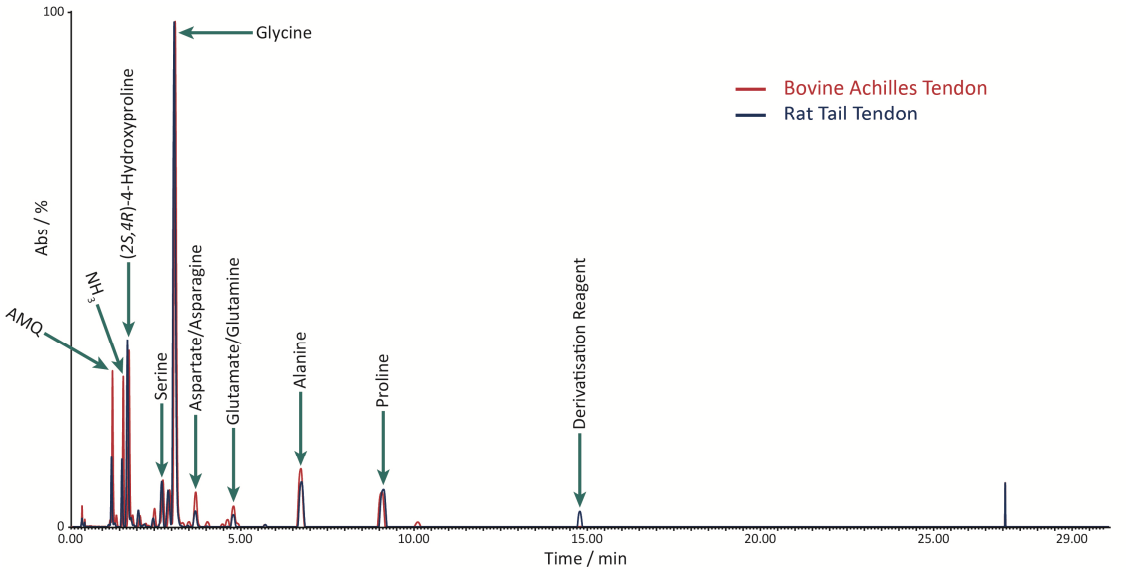


Figure 92 - Overlaid liquid chromatograms (monitoring absorbance at λ = 260 nm) of 30 minute UPLC-MS amino acid analysis run with hydrolysed samples of bovine achilles tendon (green) and rat tail tendon (blue). Samples were acid hydrolysed and derivatised using AQC (**54**). Peaks were assigned following comparison to amino acid standards and analysis of corresponding mass spectra.

Hydroxyproline Sterochemistry

The results from 10 minute UPLC-MS AAA of the bovine Achilles tendon and rat tail tendon samples were compared to standards for all possible 3- and 4- hydroxy-L-proline isomers (*i.e.* (*2S,3R*)-, (*2S,3S*)-, (*2S,4R*)- and (*2S,4S*)-hydroxyproline) in order to determine the identity of the hydroxyproline regio- and stereo- isomers were present.

Analysis of time-dependent ion abundance of $[M+H]^+$ for *m/z* = 302 (corresponding to derivatised hydroxyproline, Figure 93), shows peaks at 1.73 minutes in the two samples and standard (slight drift was observed for rat tail tendon), which corresponds to (*2S,4R*)-hydroxyproline. Small peaks at a retention time of around 4.60 minutes align closely with the standard for (*2S,3S*)-hydroxyproline.

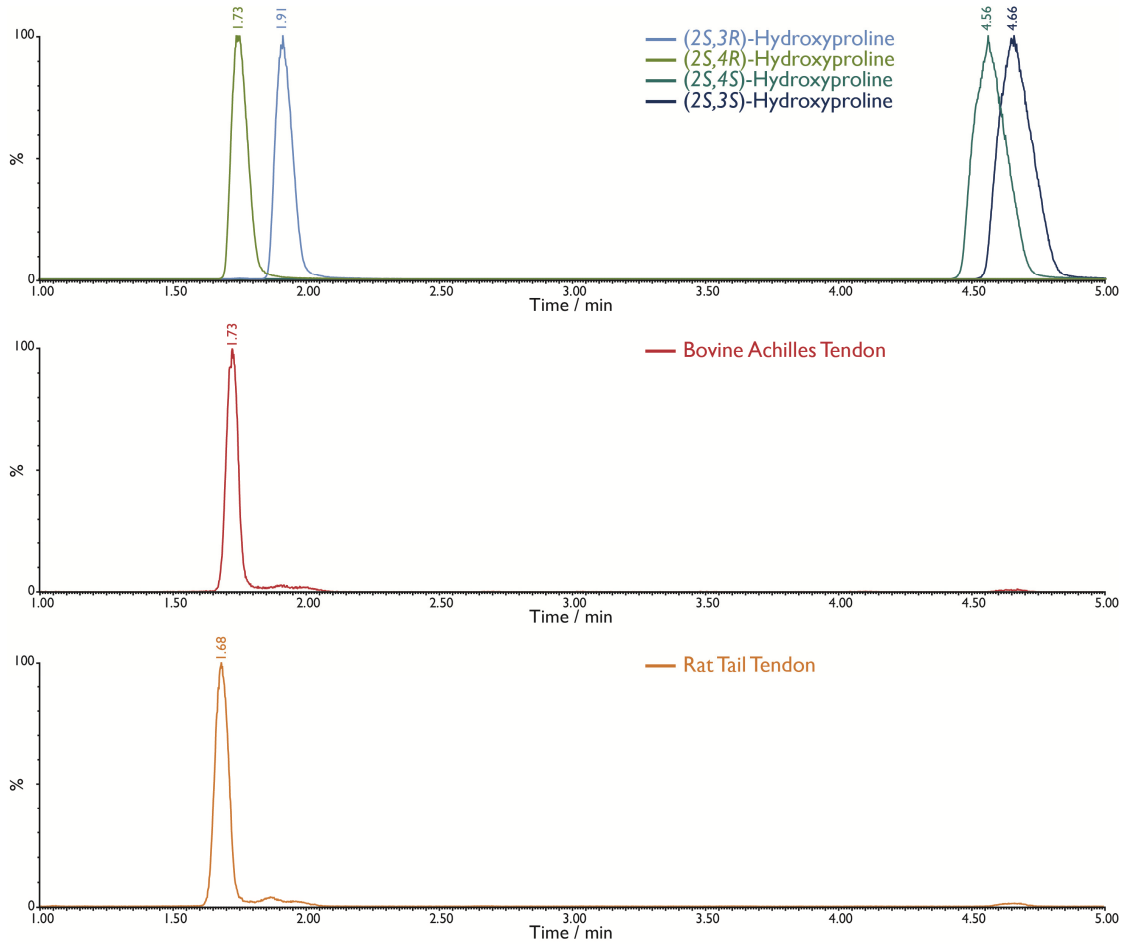


Figure 93 - Extracted ion chromatograms for *m/z* = 302, which corresponds to the $[M+H]^+$ ion of AQC-derivatised hydroxyproline. Comparison to hydroxyproline amino acid standards indicated the presence of both (*2S,4R*)-hydroxyproline and (*2S,3S*)-hydroxyproline, which is consistent with the reported activities of C-P4H and P3H[103].

The presence of (*2S,4R*)-hydroxyproline and (*2S,3S*)-hydroxyproline in collagen samples is consistent with the activity of the C-P4Hs and P3Hs[28,62,175,236]. Additionally, the greater abundance of (*2S,4R*)-hydroxyproline when compared to (*2S,3S*)-hydroxyproline (evidenced by stronger signals both by UV (λ = 260 nm) and ion abundance is consistent with the known collagen hydroxylation patterns[26].

Hydroxylysine Stereochemistry

The 30 minute UPLC-MS amino acid analysis method has previously been used to resolve the regio-isomers of hydroxylysine[113]. Samples were compared to 3-, 4- and 5- hydroxylysine amino acid standards (3- and 4-[113] hydroxylysine standards were mixtures of the *R*- and *S*- isomers, the stereochemistry of which has yet to be resolved). Analysis of the extracted ion chromatograms for the mass of doubly derivatised hydroxylysine indicates the presence of 5-hydroxylysine. The extracted ion chromatogram (Figure 94) shows no peaks that align with the standards for 3- and 4- hydroxylysine, which therefore indicates that the only hydroxylysine isomer that is present in collagen is 5-hydroxylysine.

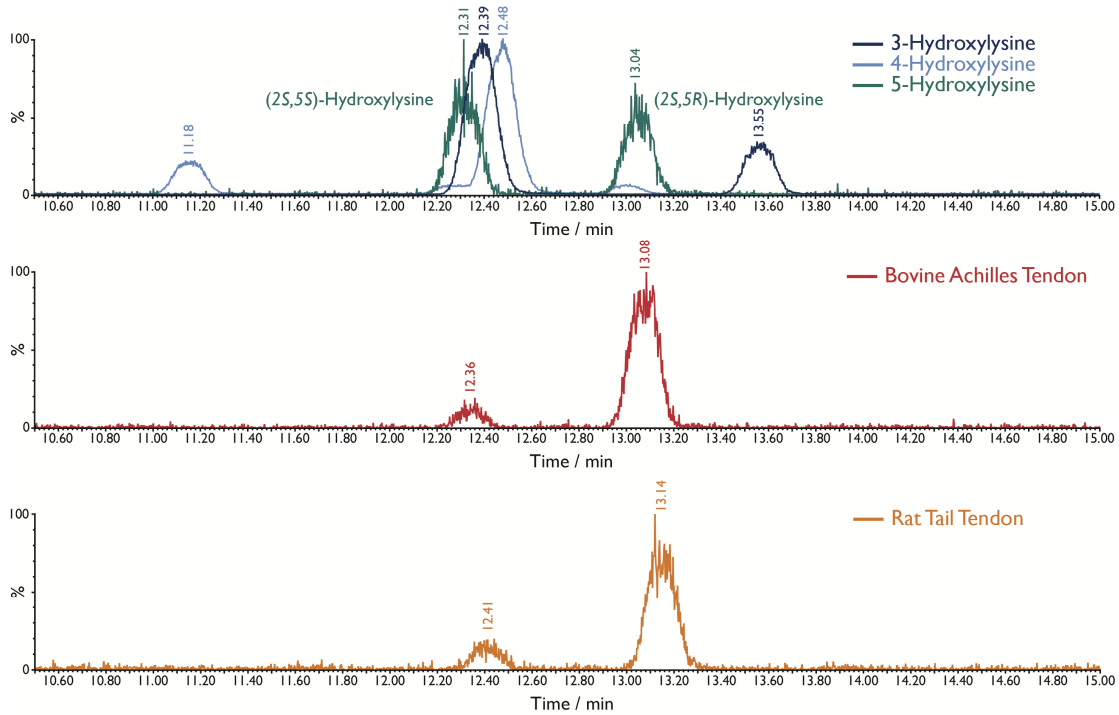


Figure 94 - Extracted ion chromatograms for *m/z* of 503, corresponding to doubly derivatised hydroxylysine (lysine residues are derivatised at both N^{α} and N^{ε}).

Comparison with previous amino acid analyses of 5-hydroxylysine[116] suggest that the predominant form (retention time of 13.10 minutes) is most likely to be the (*2S,5R*)-hydroxylysine form. Acid-catalysed racemisation to (*2R,5R*)-hydroxylysine, which proceeds *via* the mechanism in Figure 95, is proposed to lead to the peak at 12.40 minutes in the rat tail tendon and bovine Achilles tendon samples. Combined with poor abundance of hydroxylysine, acid-catalysed racemisation makes quantification of lysyl hydroxylation more challenging than quantification of prolyl hydroxylation in the absence of suitable enzymatic hydrolysis methodologies.

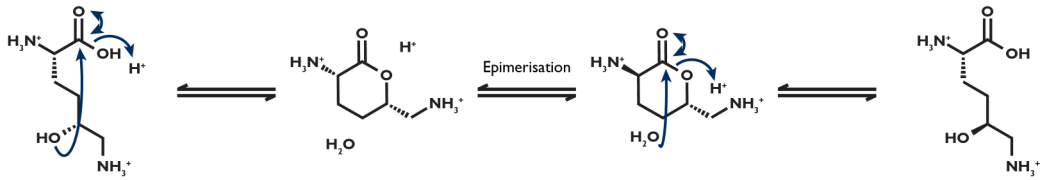


Figure 95 - (*2S,5R*)-hydroxylysine exists in equilibrium with (*2R,5R*)-hydroxylysine during acid hydrolysis. The epimerisation at C^{α} follows the lactone formation.

Comparison of bovine Achilles and rat tail tendon samples to amino acid standards does not show the presence of 3- or 4- hydroxylysine in collagen on a large scale and hence suggests that the JmjC hydroxylases (JMJD4[113], JMJD6[116] and JMJD7[135]) do not act on collagen. This does not, however, preclude the presence of these modifications at a single site, since amino acid analysis provides insight into bulk properties of the sample that is being analysed.

Comparison to Dihydroxyproline Standard

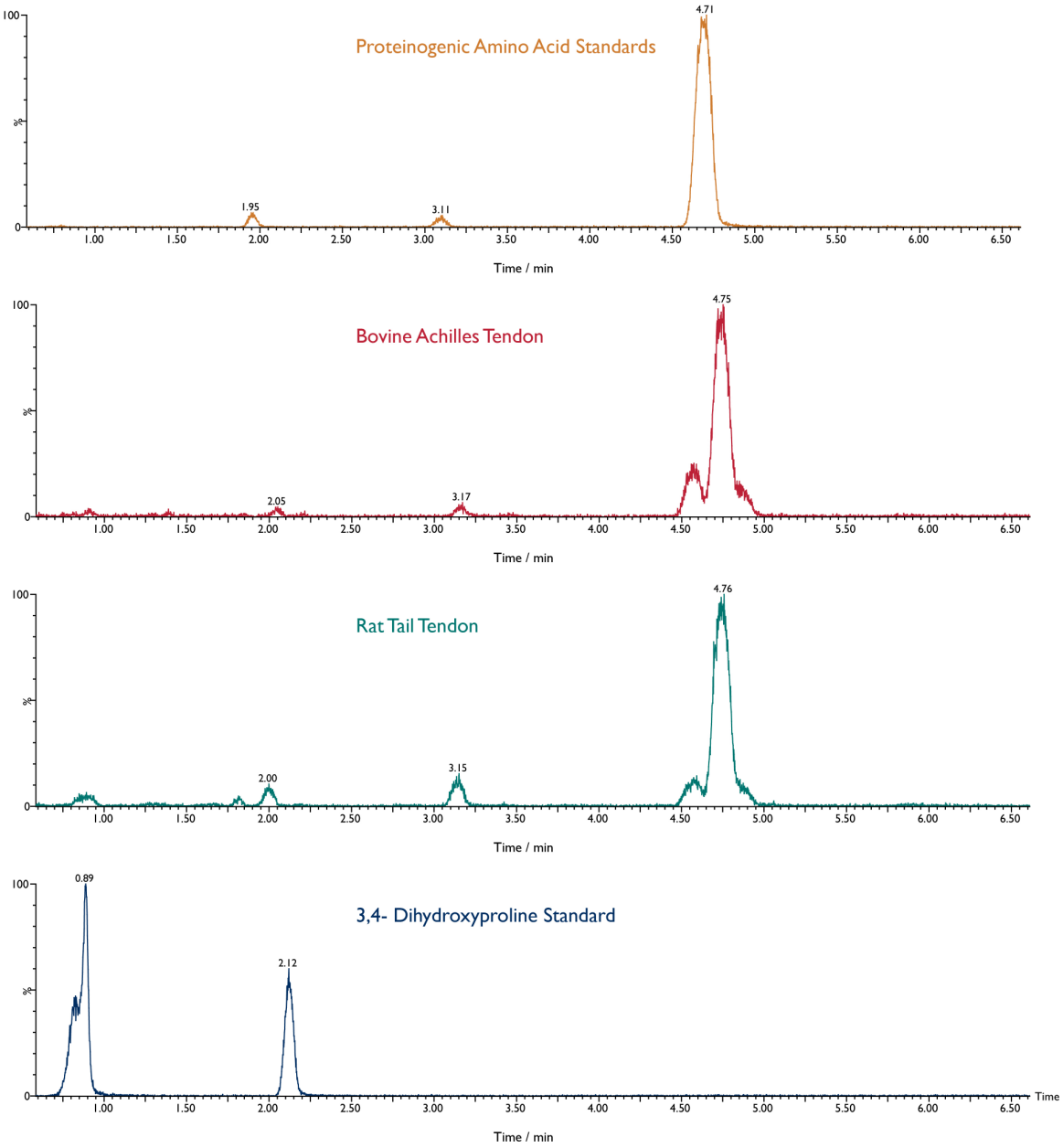


Figure 96 - Ion extracted chromatograms for *m/z* = 318, corresponding to AQC-derivatised dihydroxyproline and AQC-derivatised glutamate. An AQC-derivatised 3,4-dihydroxyproline standard had peaks at 0.89 minutes and 2.12 minutes, which were not observed in the AQC-derivatised amino acid standards, hydrolysed bovine Achilles tendon or rat tail tendon samples. The peaks at 4.75 minutes are consistent with the presence of derivatised glutamate.

The peaks that appeared from analysis of ion extracted chromatograms for the mass of AQC-derivatised 3,4-dihydroxyproline standard[237] (Figure 96) did not align with peaks in the AQC-derivatised and acid hydrolysed bovine Achilles and rat tail tendon samples and therefore provide no evidence for the presence of 3,4-dihydroxyproline in collagen.

Processing of Tissue Samples

Hamstring tendon from *Ovis aries* was used for initial attempts to purify the collagen component from tendon. Successful development of a reliable tissue processing technique would then allow for the testing of a range of samples. Amino acid analysis of the collagen samples from rat tail tendon and bovine Achilles tendon was successful following analysis of around 2.5 µg of hydrolysed and derivatised sample (1 µL injection of a 125 µg sample resuspended in 50 µL). Therefore, a suitable purification method would afford in excess of 50 µg of purified material to allow for sample handling and multiple injections if required.

Removing non-collagenous components from various tissues is a non-trivial process and a wide range of approaches have been reported in the literature[14,228,238-259], whilst the patent literature is constantly being updated with collagen extraction methodologies for the cosmetic industry. All of these approaches begin with large quantities of material (*i.e.* more than 1 g of dry mass).

Uniprot ID	Protein Name
P00760	Beta-trypsin
P02465	Collagen alpha-2(I) chain
P85154	Collagen alpha-2(I) chain
P02453	Collagen alpha-1(I) chain
Q9XSJ7	Collagen alpha-1(I) chain
P14639	Serum albumin
Q3ZC07	Actin, alpha cardiac muscle 1
Q9BE40	Myosin-1
P49824	Myosin-7
Q5XQN5	Keratin-5
O62702	Keratocan
Q3SWX2	Acyl-CoA thioesterase 9
P11615	Alpha-1B adrenergic receptor
Q9TTE2	Bone proteoglycan II
Q4W1W1	3-hydroxyacyl-CoA dehydratase 1
Q862B1	Calcitonin receptor-stimulating peptide 1
D2HZB0	Calcium-binding atopy-related autoantigen 1 homolog

Table 15 - List of proteins identified by MS-MS analysis of a tendon sample following a purifcation protocol[248]. Entries are listed in order of the number of peptides that were identified by searching in MASCOT.

An adapted procedure used for the extraction of collagen from rat tail tendon was trialled[248] and purity was assessed by MS-MS. Additionally, a more traditional method was attempted where collagen was precipitated from a suspension by increasing the concentration of NaCl[258] prior to MS-MS analysis of the constituent proteins. In both instances, a 2 mg biopsy (wet weight) was used as a starting point. MS-MS analysis showed that the samples were predominantly collagen proteins but were not pure (Table 15). Further development of a purification method is therefore needed in order to progress onto analysis of human rotator cuff samples.

Conclusions

Acid hydrolysis has proven to be the most efficient method for hydrolysis of collagen samples for amino acid analysis. For successful derivatisation, it is imperative that any residual acid is removed and, if required, the pH of the resuspended sample is adjusted to pH 8.5 before treatment of amino acid samples with the AQC derivatisation reagent. For an enzymatic hydrolysis approach to be successful, alternative enzymes need to be investigated since the *S. griseus* protease used in previous studies appears not to be suitable for samples that contain hydroxylated proline residues[235].

Reversed phase UPLC-MS amino acid analysis of commercially sourced collagen samples (from bovine achilles and rat tail tendon respectively) was successful and yielded data with good signal to noise and clearly resolved peaks for derivatised amino acids. All expected hydroxylated residues ((*2S,4R*)-hydroxyproline, (*2S,3S*)-hydroxyproline and (*2S,5R*)-hydroxylysine) were observed; these observations are consistent with the known activities of the C-P4H, PLOD and P3H enzymes.

Extracted ion chromatograms for hydroxylysine (Figure 94) were not smooth, indicative of either poor ionisation efficiency to the $[M+H]^+$ ion, or a low abundance of hydroxylysine in the hydrolysed amino acid sample. A low abundance of hydroxylysine may also reflect that lysine residues are further modified and may not be hydrolysed back to lysine/hydroxylysine[52,110].

The finding that the 3- or 4- hydroxylysine are likely not present in collagen suggests that only PLOD-catalysed formation of (*2S,5R*)-hydroxylysine occurs in collagen and that selective inhibition of lysyl-4-hydroxylases such as JMJD4[113] and lysyl-3-hydroxylases such as JMJD7[135] should not affect collagen biosynthesis.

By analysing samples both *via* UV monitoring at 260 nm and by MS, peaks can be readily identified and quantified. Previous analyses have used hydroxyproline as a measure of collagen content[231], ignoring variation in collagen hydroxylation across different tissue types and disease states. An initial approach, which affords relative quantification of amino acids, would be to calculate simple ratios of amino acids in a sample *via* analysis of the liquid chromatogram. For example, the presence of clear, well resolved peaks for (*2S,4R*)-hydroxyproline allows for calculation of the ratios of proline to (*2S,4R*)-hydroxyproline. However, absolute quantification of samples cannot be achieved in the absence of an external standard.

Analysis of tissue-derived samples requires a reliable approach to obtaining purified collagen samples and thus requires a robust purification protocol that reliably affords high purity collagen samples that can then be hydrolysed and analysed. Initial, unsuccessful attempts to purify collagen from a hamstring tendon from *Ovis aries* were limited in scope and therefore there is potential to further this study. Once a reliable protocol is in place, samples that are in storage from the UKUFF trial[224,226] could be processed in parallel and analysed. In addition, the development of fast approaches to assess the purity of collagen samples would aid the development of sample preparation protocols.

Amino acid analysis of tendon samples would ideally complement other studies of diseases tendon, such as proteomic analyses[14] and *in vitro* cell culture studies[260], in order to build up a better understanding of rotator cuff disease and by implication, tendon disease in general.

Summary

Amino acid analysis provides interesting potential as a complement to proteomic analyses and may provide interesting insight into bulk properties of protein samples. Commercially-sourced collagen samples have been successfully hydrolysed following development of a robust acid hydrolysis protocol. Subsequent derivatisation and UPLC-MS amino acid analysis has been successful and hydroxylated residues, in particular the presence of (*2S,4R*)-hydroxyproline, can clearly be identified using both UV absorbance and by MS. (*2S,5R*)-hydroxylysine was observed but was only visible by MS whilst there is no evidence from these studies for the presence of 3-hydroxylysine, 4-hydroxylysine or 3,4-dihydroproline in collagen.

However, the UPLC-MS amino acid analysis protocol for the analysis of tendon samples requires the development of a robust sample preparation method. Further development of an appropriate sample preparation method is required before analysis of a large range of collagen samples is initiated.

Chapter 6
Lysine Analogues as Probes for 2OG Oxygenases

The work described in this chapter was conducted in collaboration with Anne Brinkø (synthesis of compounds **59**, **61**, **63**, **62**, **74** and **73**), Dr Rachel Shamo Schiller (purification of KDM3A and KDM3A assays), Dr Joan Gannon (purification of HSPBAP1 and HSPBAP1 assays) and Dr Richard Hopkinson (NMR studies).

KDM4A, KDM4B, KDM4C and KDM4E were prepared by Dr Louise Walport. JMJD4 was prepared by Dr Sarah Wilkins and JMJD6 was prepared by Dr Monica Mantri.

Sections of this work (studies with KDM7B and the KDM4 subfamily) have been reported previously[261-263]. Sections have been reprinted and adapted with permission from Langley, G. W. et al. Analysis of JmjC Demethylase-Catalyzed Demethylation Using Geometrically-Constrained Lysine Analogues. *ACS Chem. Biol.* **11**, 755-762[261]. Copyright (2016) American Chemical Society.

Results from this chapter have been submitted as part of the following theses:

- Anne Brinkø PhD Thesis (Aarhus University)
  - Synthesis of compounds **59**, **61**, **63**, **62**, **74** and **73**.
- Rachel Shamo Schiller DPhil Thesis (University of Oxford, 2016)
  - KDM3A assays.
- Gareth William Langley Part II Thesis (University of Oxford, Honour School of Chemistry, 2013)
  - Development of synthetic methods and synthesis of all compounds detailed in Figure 102 with the exception of **59**, **61**, **63**, **62**, **74** and **73**.
  - Synthesis of Peptide 38, Peptide 32, Peptide 31, Peptide 33, Peptide 41, Peptide 39 and Peptide 40.
  - MALDI-TOF MS and NMR spectroscopy studies with Peptide 31 and KDM4E.
  - Studies with JMJD4 and JMJD6, initial assay and time course with HSPBAP1 (Figure 109).

Lysine is a Common Residue for Post-Translational Modifications

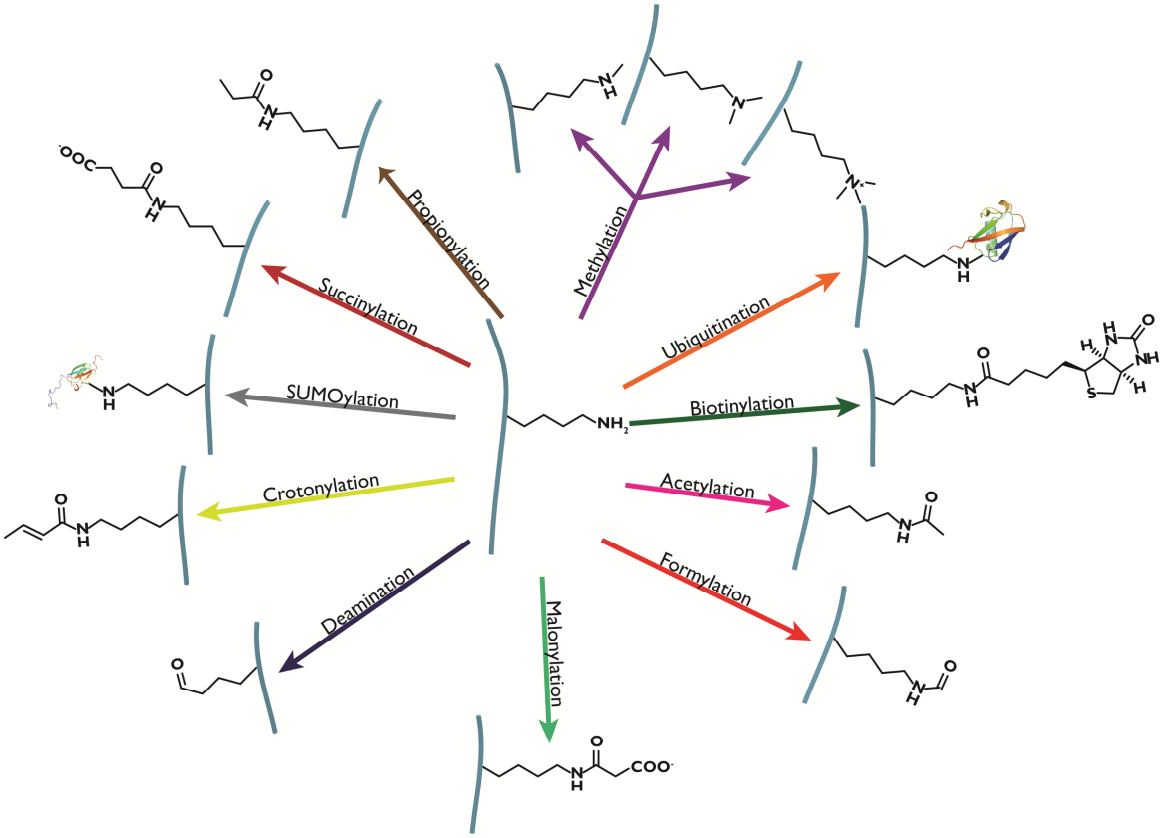


Figure 97 - Examples of post-translational modifications found on the ε-nitrogen of a lysine residue[115,264-271].

Lysine residues are subjected to a wide range of post-translational modifications[272]; many of these modifications are centred on the ε-nitrogen[115,264-271], taking advantage of the nucleophilicity of the nitrogen lone pair and, in many cases, the stability of a resultant amide bond in order to form stable linkages. Some of these modifications are detailed in Figure 97.

Epigenetic Regulation and Lysine PTMs

Epigenetic regulation of genes, *i.e.* the regulation of genes without changes in the underlying DNA sequence[273], is reliant on the organisation and packaging of DNA. The nucleosome (Figure 50) is the base unit of DNA packaging within the nucleus; each nucleosome consists of approximately 150 base pairs of DNA wrapped around an octamer of histone proteins[274]. Two copies of four histone proteins, histone 2A (H2A), histone 2B (H2B), histone 3 (H3) and histone 4 (H4), make up this octameric core. Nucleosomes are further organised and, with histone 1 (H1),

packed together to form chromatin; the nature of this packaging influences gene expression[267]. Loosely packaged sections (euchromatin) allow for the transcriptional machinery of the cell to access the DNA in this region, leading to increased expression of any genes in this region. Meanwhile, genes in more tightly packed heterochromatin regions are more likely to be silenced[269,271]. The arrangement of nucleosomes, and the formation of heterochromatin and euchromatin can be altered by the presence, or absence of PTMs to either the DNA itself, or on the histone core of the nucleosome.

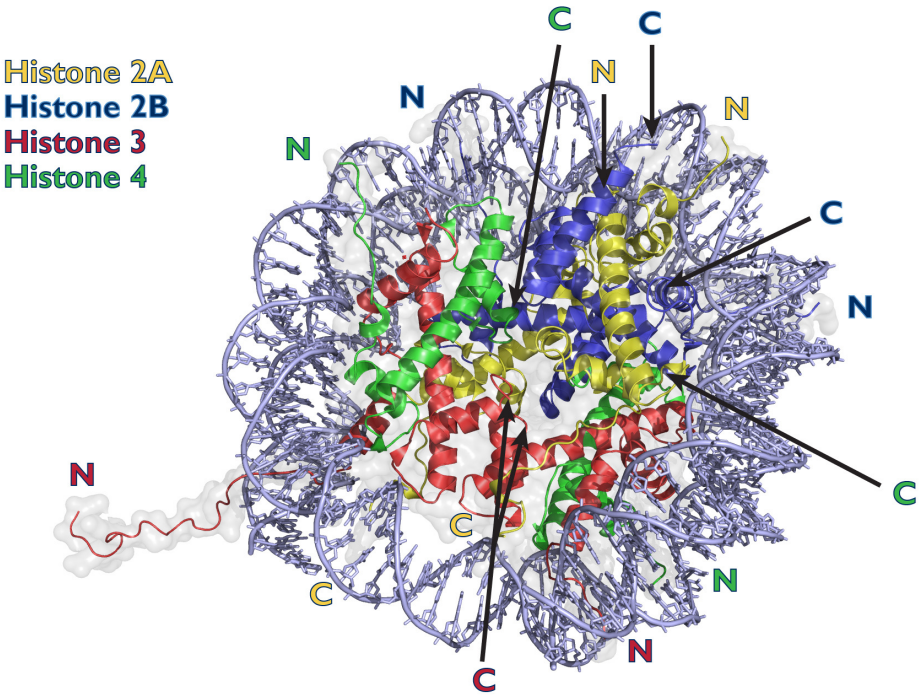


Figure 98 - A view from a crystal structure of the nucleosome, PDBID: 1AOI[274]. The structure is rendered as a cartoon, H2A is indicated in yellow, H2B in blue, H3 in red and H4 in green. The *N*-terminal tail of H3 can be seen extending from the core.

The *N*-terminal tail of histone H3 is subjected to a wide range of post-translational modifications that affect the packing of histones and consequently the expression of any gene encoded in this region[115]. Modification of histone tails, in particular the alteration of lysine residues to some of the modified forms detailed in Figure 97, alongside the modification of the DNA bases themselves are amongst the key mechanisms behind epigenetic gene regulation. Changes to the levels (*i.e.* removal or addition) of epigenetic PTMs, is therefore important in dynamic gene regulation.

The residues on which the post-translational modifications reside interact with whole proteins or protein domains that recognise the modification (reader domains), mediate the addition of the PTM (writer domains), or remove the PTM (eraser domains). Epigenetic readers, writers and erasers play a key role in the dynamic regulation of the PTMs, and by implication, gene regulation. In the case of lysine N^{ε}-acetylation, the presence of the PTM is mediated by histone acetyltransferases and deacetylases and recognised by readers such as bromodomains[115,267].

Lysine Methylation and Demethylation

The dynamic methylation and demethylation of histone lysine residues is implicated in the epigenetic regulation of gene expression[115,267]. Methylation is mediated by *S*-adenosyl methionine dependent methyltransferases, whilst demethylation is catalysed by the histone demethylases (KDMs). Methylated lysine residues in histones may be recognised by a range of domains including plant homeodomain (PHD) fingers, chromodomains and tudor domains.

There are two classes of histone demethylases, the flavin-dependent LSD1 and LSD2 and the larger family of 2OG and Fe^{II} dependent histone demethylases (JmjC KDMs)[111]. The JmjC KDMs have a high degree of structural similarity and all contain a JmjC fold, a form of the double stranded β-helix found in approximately half of the Fe^{II}/2OG dependent oxygenases. Not all JmjC-containing enzymes are demethylases; for example, the oxidation of side-chain C-H by 2OG oxygenases to form 4-[113] and 5-[116] hydroxylysine are catalysed by JMJD4 and JMJD6 respectively.

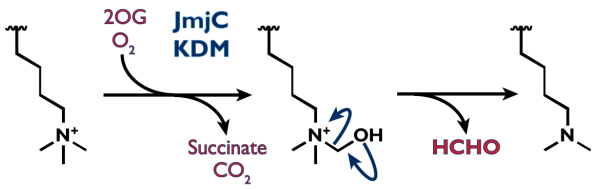


Figure 99 - Demethylation by JmjC KDMs proceeds *via* oxidation of a N^{ε} methyl C-H. This resultant hemiaminal collapses to give the demthylated product and formaldehyde[85].

When a N^{ε}-methyl C-H is oxidised to C-OH, the resultant hemiaminal product is unstable and collapses to form the demethylated product and formaldehyde (Figure 99)[85,111]. Previously it has been established, through the use of lysine analogues, that the reaction scope of the JmjC KDMs

is not limited to lysine demethylation[275] (*e.g.* hydroxylation of alkyl chains without subsequent dealkylation) and some of the JmjC KDMs have also exhibited arginine demethylase activity *in vitro*[276].

Manipulation of the Lysine Side Chain and Replacement of the N^{ε}-Group as a Strategy to Probe Catalysis

Backbone Alignment in Catalysis

In order for stereo- and regio-specific JmjC KDM catalysis to occur, the C-H bond that is oxidised needs to be presented towards the Fe^{II} centre in an orientation amenable to catalysis; therefore, for side chains such as lysine, which have the ability to exist in a wide range of conformations, the backbone alignment of the residue in the active site is important for enabling catalysis.

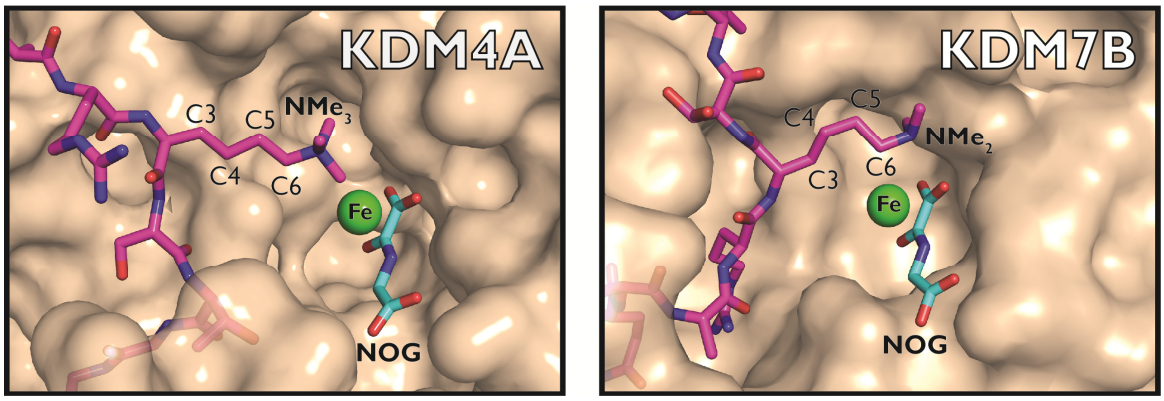


Figure 100 - Views from crystal structures of KDM4A (PDBID: 2OQ6)[277] and KDM7B (PDBID: 3KV4)[83]. Backbone alignments of H3K9 differ between the two demethylases. Backbone carbons are numbered, whilst the metal centre and NOG, a 2OG analogue, are indicated.

Closer analysis of the substrate structures of KDM4A[277] and KDM7B[83] indicate that the lysine backbone alignments differ during catalysis (Figure 100). The C3-C4-C5-C6 dihedral angle of the H3K9 residue in the KDM4A structure is -62°, effectively an *anti*- alignment, whilst a *syn*- type alignment in KDM7B is observed, with a C3-C4-C5-C6 dihedral angle of -174°. From this crystallographic data, therefore, it appears that the introduction of geometric constraints in the backbone of the target lysine may differentially affect demethylation efficiency of KDM7B and the KDM4 subfamily.

By introducing a double bond into the lysine backbone, rigidity can be achieved without the addition of the steric bulk that a ring system (*e.g.* a cyclopropane) introduces. This approach has previously been explored in order to compare the rates of trypsin-mediated peptide bond cleavage of substrates containing *E*- and *Z*- 4,5-dehydrolysine[278]; however, the synthetic routes employed did not allow for use of solid phase peptide synthesis methodologies[278]. Therefore, a more robust synthetic approach is required to introduce geometric constraint and the correct protective groups to allow for subsequent incorporation into peptide substrates *via* automated solid phase peptide synthesis, and in turn subsequent biochemical analyses.

The Charge State of During JmjC KDM Catalysis

N^{ϵ}-tri-methylated lysine residues have an inherent positive charge; however, the N^{ϵ} protonation state (and hence N^{ϵ} charge) of di-methylated and mono-methylated lysine residues during JmjC catalysis is unknown. In contrast, demethylation by the KDM1 family of demethylases mechanistically requires involvement N^{ϵ} lone pair and as a consequence, ε-nitrogen must be neutral. This mechanism means that the KDM1 enzymes cannot demethylate N^{ϵ}-trimethyl lysine residues[85].

Therefore, testing of substrates with alternative charge at N^{ϵ} (*i.e.* replacing the N^{ϵ} nitrogen with another atom such as carbon or oxygen) may provide an insight into charge state selectivity during JmjC KDM catalysis; by extension, such an approach could also be readily extended to other enzymes that act on a lysine residue. Similar to the introduction of backbone constraint, these studies require a robust synthetic method to reach the appropriately protected amino acids for incorporation into peptide substrates.

Aims

The principal aim of the project was to analyse the effect that backbone rigidity has on JmjC KDM catalysis. Additionally, the ability to access analogues with alternative atoms was desired in order to analyse the effect of charge at the ε position on JmjC KDM catalysis.

Therefore, the development of robust synthetic methods to produce appropriately protected amino acids was required before the synthesis of peptides for testing as substrates. Accessing novel probes, as well as improving upon previous syntheses[278,279] would allow for synthesis of a range of peptide probes for the JmjC KDMs.

These probes could then be readily utilised in *in vitro* biochemical studies. It was hoped that subsequent biochemical analyses would provide insight into the role of the lysine backbone and charge requirement of JmjC KDM catalysis.

Synthetic Strategy and Implementation

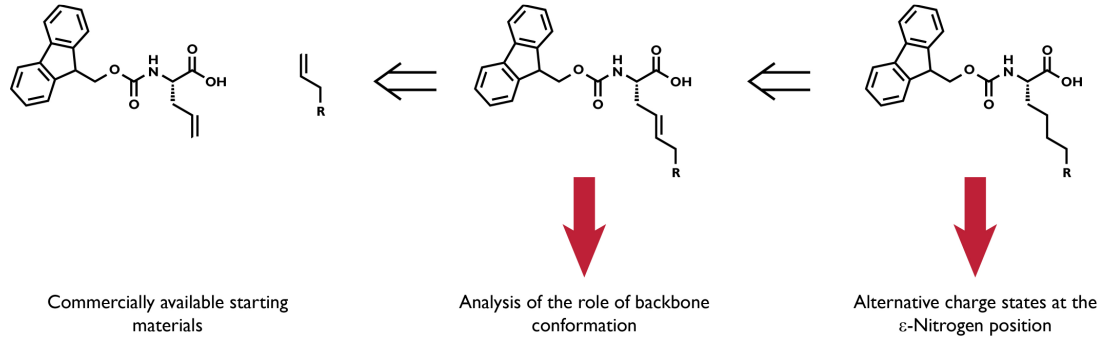


Figure 101 - Retrosynthetic analysis of the two classes of probe that could be accessed by olefin cross metathesis.

Manipulating the lysine side chain and producing lysine analogues with altered backbones or alternative atoms at the ε-nitrogen are hindered by inefficient synthetic methods. It was envisaged that olefin cross-metathesis would afford a route to a number of novel probes, as well as providing the potential to improve upon existing routes to lysine analogues (Figure 101). This approach is advantageous as it allows for one-step access from readily accessible or commercially available starting materials such as *N*-Fmoc Allyl-L-Glycine (**55**) and *N*-Boc Allylamine (**56**) to lysine analogues that are appropriately protected for direct incorporation into solid phase peptide synthesis[262].

Synthesis of Fmoc-protected Lysine Analogues

Using the Hoveyda-Grubbs second generation catalyst (**57**)[280], N^{α}-Fmoc-4,5-dehydro-L-lysine, **58**, was successfully synthesised from **55** and **56**; this approach was then successfully implemented

to produce a library of lysine analogues (Figure 102)[262] that could be directly incorporated into peptides.

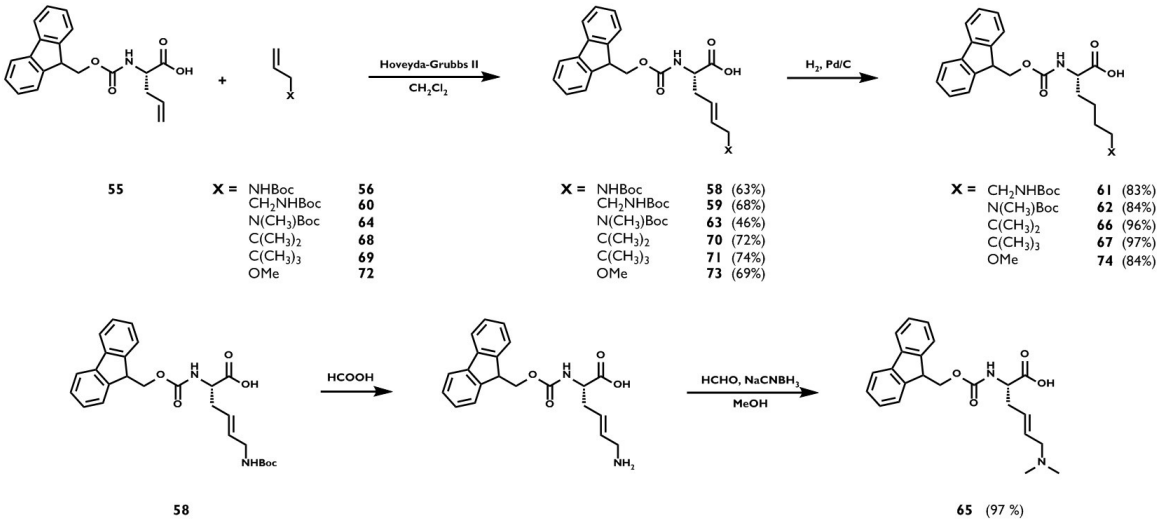


Figure 102 - Compounds synthesised using the metathesis-hydrogenation approach. Compound **58** was then used to produce a dimethylated lysine analogue, **65**, which could then be used to probe JmjC KDM catalysis. Isolated yields are indicated in parentheses.

When unprotected amines were used as metathesis partners, the reaction did not progress and the reaction mixture lost its green colour, suggestive of catalyst degradation. Therefore 4,5-dehydrohomolysine, **59**, was accessed from **60** before being hydrogenated to N^{α}-Fmoc Homolysine, **61**. Similarly, protected monomethyl lysine, **62**, was synthesised using the metathesis-hydrogenation methodology *via* **63** using **64** as the cross-metathesis partner.

In order to obtain a dimethylated lysine analogue, **65**, allowing for analysis of demethylase activity, a simple one pot synthesis was employed.

Replacement of N^{ε}- with carbon to give rise to dimethyl- (**66**) and trimethyl- (**67**) lysine analogues was achieved by using commercially available starting materials **68** and **69** (*via* the (*E*)-4,5-dehydro forms **70** and **71**). This approach vastly improved upon a previously reported synthesis[279] of the trimethyl-L-lysine analogue where the ε-nitrogen was replaced with a carbon, **67**. Furthermore, production of analogues where N^{ε}- is replaced by oxygen was achieved by coupling **55** and **72** to form analogue **73**, which was subsequently hydrogenated to **74**.

Peptide Synthesis and Initial Biochemical Studies

Lysine analogues **65** and **67** were incorporated into the 9 position of a 15mer peptide encoding residues 1-15 of histone H3 (Peptide 31 and Peptide 32). This was achieved using standard Fmoc-based solid phase peptide synthesis methodologies.

Peptide 31 and Peptide 32 were then tested as substrates of KDM4E. Demethylation of Peptide 31 was observed by MS (Figure 104A) and formaldehyde release was monitored by NMR spectroscopy[261,262]. Meanwhile, no reaction was observed with an analogous peptide which incorporated analogue **67** at the K9 position, Peptide 32. These experiments were run concurrently with a natural substrate mimic, Peptide 33, which featured N^{ε}-dimethyl-L-lysine, **75**, at the K9 position.

Peptide Library Expansion

Other lysine analogues were incorporated into histone peptides to create a library of peptide probes (Table 16) that could then be tested with JmjC KDMs. In addition, a 15mer substrate peptide for KDM7B, Peptide 35, which features N^{ε}-trimethyl-L-lysine, **76**, at the K4 position was synthesised, incorporating compound **65** at the K9 position.

	Residue at K9	Sequence
Peptide 31	**65**	ARTKQTAR-X-STGGKA
Peptide 32	**67**	ARTKQTAR-X-STGGKA
Peptide 33	N^{ε}-Dimethyl-L-lysine, **75**	ARTKQTAR-K(Me_2)-STGGKA
Peptide 34	N^{ε}-Dimethyl-L-lysine, **75**	ART-K(Me_3)-QTAR-K(Me_2)-STGGKA
Peptide 35	**65**	ART-K(Me_3)-QTAR-X-STGGKA
Peptide 36	Homolysine, **61**	ARTKQTAR-X-STGGKA
Peptide 37	**74**	ARTKQTAR-X-STGGKA
Peptide 38	**67**	KQLATKAAR-X-SAPAT

Table 16 - Library of histone 3 (1-15) peptides synthesised as probes for JmjC KDM catalysis.

Successful incorporation of the lysine analogues into the peptide sequence and double bond geometry (where appropriate) were confirmed by MS and NMR spectroscopy.

The Role of Backbone Alignment in KDM4 and KDM7B Catalysis

Peptide 31 was tested for activity against other members of the KDM4 subfamily and was, as predicted following structural analyses (Figure 100), accepted as a substrate by KDM4A, KDM4B and KDM4C. Reactions were monitored through MS analysis of peptides, NMR spectroscopy or through monitoring the reaction using a formaldehyde dehydrogenase (FDH) coupled demethylation assay[281]. The FDH assay monitors the liberation of formaldehyde during JmjC KDM-catalysed demethylation. FDH and the oxidised form of the cofactor nicotinamide adenine dinucleotide (NAD^+, **77**) are added to the reaction mixture and as formaldehyde is produced, FDH converts formaldehyde to formate, reducing NAD^+ to the reduced form, NADH (**78**) in the process. NADH (**78**) concentrations in the reaction mixture are therefore correlated to production of formaldehyde (and by implication, demethylation). Therefore, monitoring NADH (**78**) production *via* fluorescence (excitation at λ = 355 nm and emission at λ = 460 nm) can be used to calculate a rate of formaldehyde production and thus demethylation activity[281].

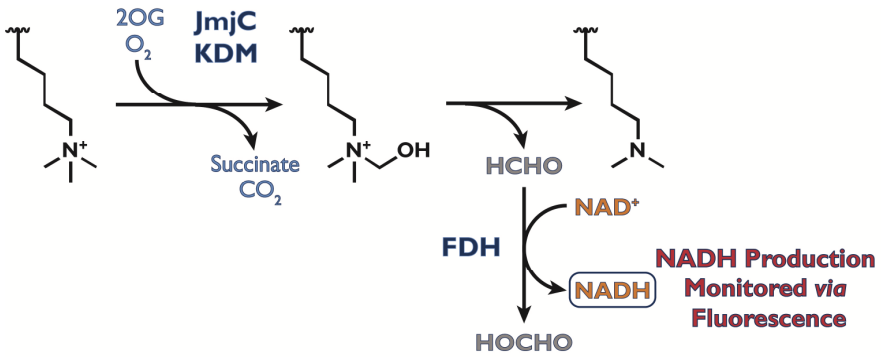


Figure 103 - Schematic representation of FDH-coupled assay. FDH converts the formaldehyde released following demethylation into formate, this process is coupled to the reduction of NAD^+ (**77**) to NADH (**78**). NADH levels are used to assay formaldehyde liberation by monitoring fluorescence, through excitation at 355 nm and emission at 460 nm[281].

KDM7B was prepared following previously reported procedures[282]. For kinetic analyses, the *N*-terminal thioredoxin tag was cleaved with a protease from tobacco etch virus (TEV protease). Removal of the thioredoxin tag is required to as the thioredoxin fusion protein, if present, will oxidise NADH (**78**) back to NAD^+ (**77**), thereby interfering with the assay platform.

Equivalent analyses with KDM7B and Peptide 35 showed that Peptide 35 was, counter to the predictions made following structural analyses in Figure 100, also a substrate for KDM7B (Figure 104). However, KDM7B-catalysed demethylation of Peptide 35 appeared to not be as efficient as KDM4-catalysed demethylation of Peptide 31; higher concentrations of KDM7B were required for incomplete conversion to the monomethyl product, whereas KDM4 catalysed complete conversion of Peptide 31 to the demethylated products.

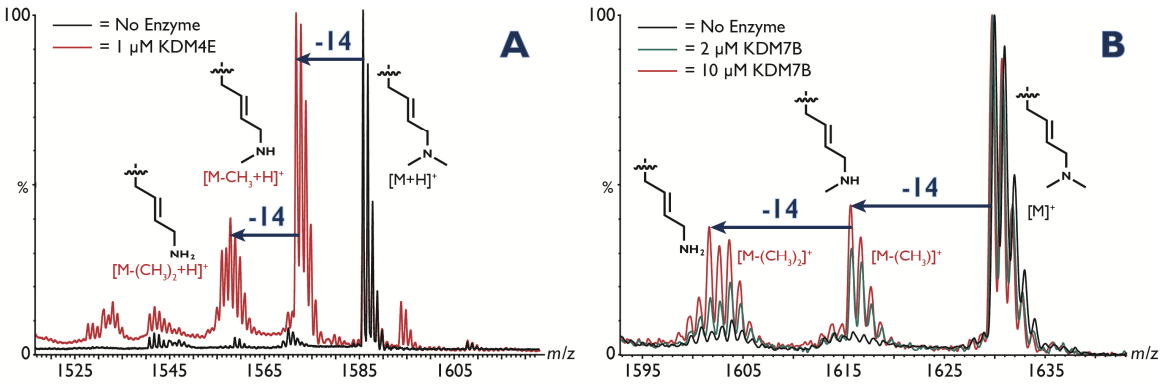


Figure 104 - Mass spectra showing demethylation of **65** at the K9 position of Peptide 31 and Peptide 35 by KMD4E and KDM7B respectively. All ions observed correspond to the $[M+H]^+$ species for Peptide 31 and $[M]^+$ for Peptide 35; product ions are labelled as losing methyl groups for clarity. Δ*m/z* of -14 is consistent with demethylation.

In order to better assess the relative efficiency of demethylation, an FDH-coupled demethylation assay[281] was employed to calculate kinetic parameters. For both KDM4E and KDM7B, the peptides that contained N^{ϵ}-dimethyl-4,5-dehydro-L-lysine, **65**, at the K9 position were worse substrates than their N^{ϵ}-dimethyl lysine, **75**, equivalents (Table 17). This indicates that the KDM7B active site is more flexible than first envisaged, and that a degree of lysine-backbone flexibility is required for optimal binding of substrate for KDM4E and KDM7B.

	Peptides containing N^{ϵ}-dimethyl lysine, 75 k_{cat}/K_M / $\mu M^{-1} s^{-1}$	**Peptides Containing N^{ϵ}-dimethyl-4,5-dehydro-L-lysine, 65 k_{cat}/K_M / $\mu M^{-1} s^{-1}$**
KDM4E	0.00215	0.00177
KDM7B	0.00059	0.00020

Table 17 - k_{cat}/K_M values for the demethylation of peptides containing analogues **75** (Peptide 33 and Peptide 34) and **65** (Peptide 31 and Peptide 35).

Analysis of Charge Requirements for KDM4 Catalysis

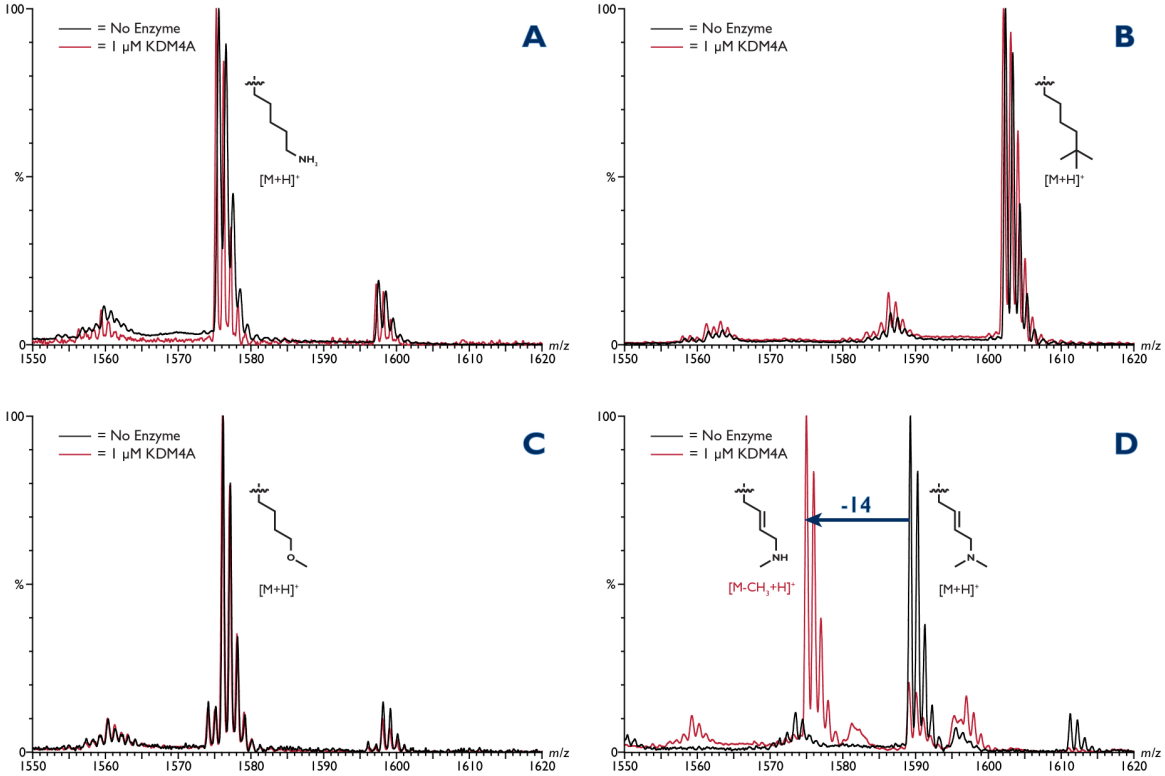


Figure 105 - Mass spectra of peptides incorporating alternative atoms at the N^{ε}- position following incubation with KDM4A (1 μM) and cofactors/co-substrates. No enzyme controls are overlaid in black. No reaction was observed with Peptide 32 (A), Peptide 36 (B) and Peptide 37 (C). Peptide 33, a positive control, was run concurrently and is shown for comparison (D).

In order to probe the charge requirements of KDM4 catalysis, Peptide 32, where N^{ε}- at the K9 position was replaced by carbon, was incubated with cofactors/co-substrates and members of the KDM4 family (KDM4A, KDM4B, KDM4C and KDM4E). Peptide 36 and Peptide 37 were also tested as substrates with the KDM4 activity to investigate the effect of lengthening the lysyl backbone and tolerance to alternative heteroatoms at the ε-position respectively. In each case, no reaction was observed; mass spectra for reactions with KDM4A are shown in Figure 105.

The peptides were then tested for inhibitory activity against KDM4E using the FDH assay (Peptide 33). Although at high concentrations some inhibition was observed, the peptides were not potent enough to be considered as inhibitors of KDM4E. Only Peptide 32 exhibited more than 50% inhibition at a concentration of 1mM (IC_{50} = 203.8 μM ± 1.2 μM); for comparison, NOG (**4**) has an IC_{50} of 24 μM[283].

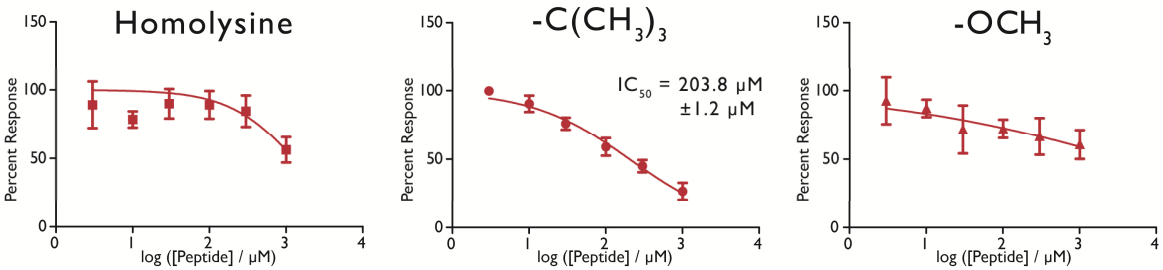


Figure 106 - KDM4E dose-response for Peptide 36, Peptide 32 and Peptide 37. A FDH-coupled assay was used to monitor the demethylation of Peptide 33 by KDM4E.

In combination, the weak inhibition of H3K9 demethylation by KDM4E by Peptide 32, Peptide 36 or Peptide 37, and the absence of reaction between the peptides and any of the KDM4 enzymes show that the presence of N^{ε} is required for substrate binding. Furthermore, these observations strongly indicate that a positively charged ε-nitrogen is important in substrate binding to KDM4 and subsequent KDM activity.

Investigation of KDM3A Catalysis with Lysine Analogues

These studies were conducted in collaboration with Dr Rachel Shamo Schiller, who also presented these findings in her DPhil Thesis.

In addition to the KDM4 subfamily, the KDM3 family of KDMs also catalyse the demethylation of methylated H3K9 residues[85]. Therefore, to ascertain if there are any inter-subfamily differences between the JmjC KDMs, Peptide 31, Peptide 32, Peptide 36 and Peptide 37 were tested as substrates of KDM3A (Figure 107). Similar to KDM4E, Peptide 31, which features N^{ε}-dimethyl-4,5-dehydro-L-lysine, **65**, at the K9 position, was demethylated when incubated with KDM3A (Figure 107A, Δ*m/z* of -28 corresponds to the removal of two methyl groups). In addition to this demethylation event, KDM3A-mediated reactions were also observed with Peptide 32, Peptide 36 and Peptide 37, in stark contrast to the studies with the KDM4 subfamily (Figure 106).

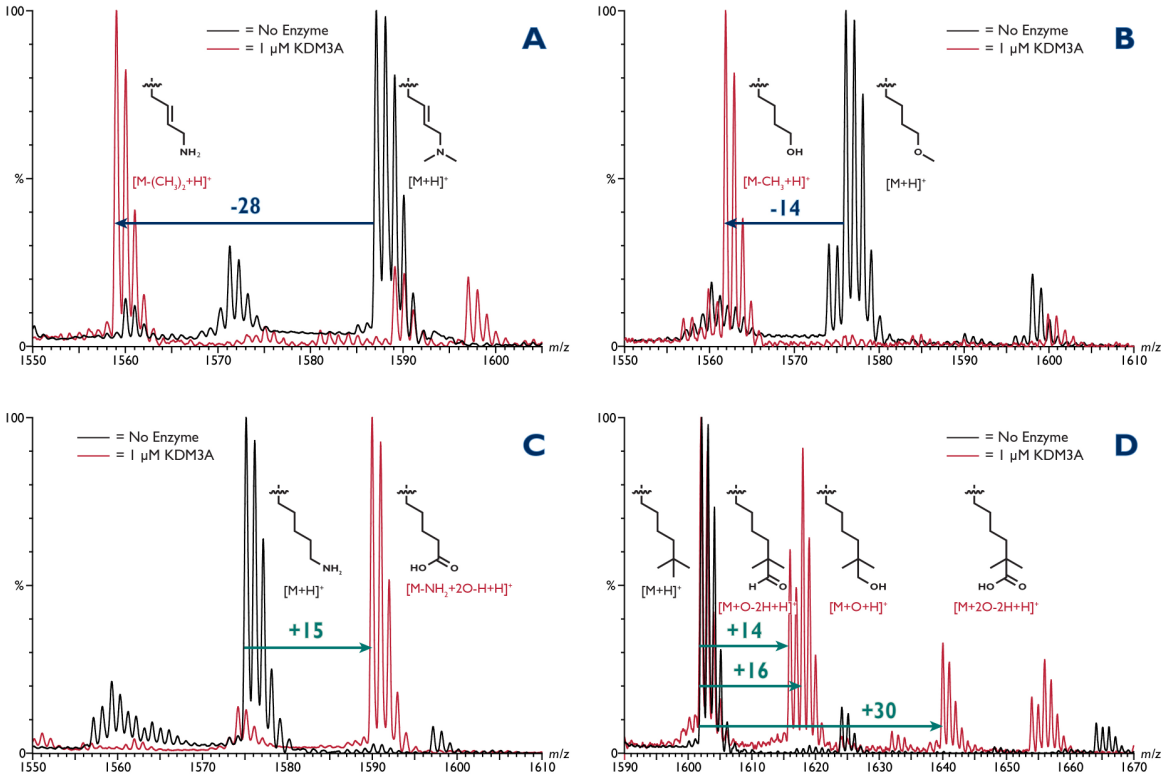


Figure 107 - Mass spectra showing KDM3A-catalysed modification of Histone 3 (1-15) peptides containing analogues 61,65, 67 and 74 at the K9 position. A - Demethylation of Peptide 31 B - Demethylation of Peptide 37 C - Sequential oxidation of Peptide 36 to a carboxylic acid D - Oxidation of Peptide 32 to the carboxylic acid *via* hydroxylated and ketone species. Further oxidation from this product is observed (peaks at *m/z* of 1654 and 1656 in the red spectrum).

Demethylation of the -OMe species in Peptide 37 to the alcohol was observed (Figure 107B); this suggests that a positive charge at the substrate ε-position is not required for KDM3A catalysis to proceed. This is further supported by the reaction observed between KDM3A and Peptide 32.

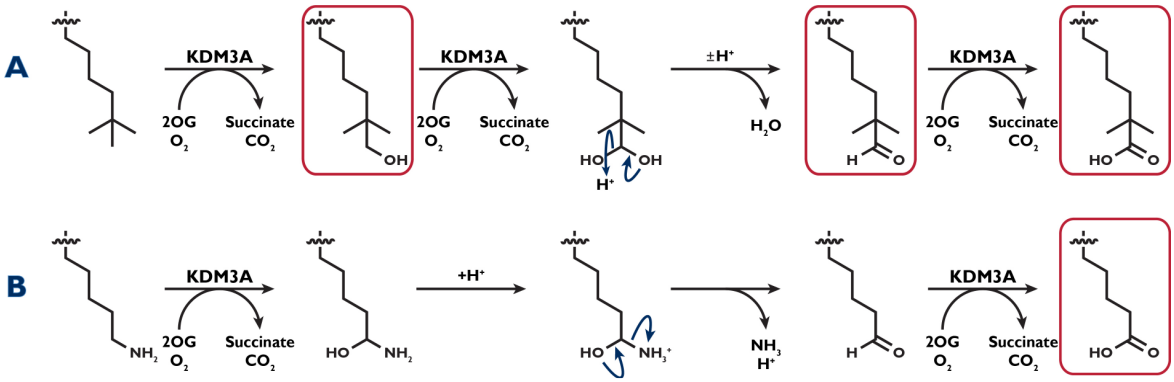


Figure 108 - Proposed reaction scheme for sequential hydroxylation of lysine analogue **67** at the K9 position of Peptide 32 (A) and analogue **61** at the K9 position of Peptide 36 (B) by KDM3A. Species observed in Figure 107 are indicated.

Peptide 36 and Peptide 32 undergo sequential KDM3A-catalysed oxidations to give the corresponding carboxylic acids (Figure 107C and Figure 107D). In the case of Peptide 36, it is

proposed that the amino group is lost to give an aldehyde which is then oxidised to the carboxylic acid (Figure 108B). Following the reaction of Peptide 32 with KDM3A, ions consistent with the hydroxylated, aldehyde (formed following a double hydroxylation and dehydration, Figure 108A) and carboxylate species are all observed. Additionally, ions consistent with hydroxylation of the other methyl groups are observed; the peaks at *m/z* of 1654 and 1656 are consistent with the hydroxylation and aldehyde formation when the final product of Figure 108A is accepted as a substrate.

KDM3A has exhibited the ability to act upon a wide range of potential substrates and these unprecedented KDM3A reaction modes reveal that KDM3A may have greater reaction scope than previously envisaged. The ability of KDM3A to catalyse sequential oxidations is, however, not unique amongst the 2OG oxygenases. The TET family of enzymes are known to catalyse the formation of 5-formylcytosine and 5-carboxylcytosine from 5-methylcytosine[284]. Additionally, these findings imply that the lysine residue is not positively charged during KDM3A catalysis, unlike the KDM4 subfamily.

Lysyl Hydroxylases

These studies were conducted in collaboration with Dr Sarah Wilkins (JMJD4), Dr Monica Mantri (JMJD6) and Dr Joan Gannon (HSPBAP1).

	Sequence	Enzyme
Peptide 39	NA-(L-4,5-dehydrolysine)-AMSARAEAIKALVKPKE	JMJD4
Peptide 40	NP-(L-4,5-dehydrolysine)-RSRSREHRR	JMJD6
Peptide 41	(L-4,5-dehydrolysine)-YGIICHDLIHEIYTVGKR	HSPBAP1

Table 18 - Peptides synthesised for testing with JMJD4, JMJD6 and HSPBAP1.

In addition to the studies with demethylases, substrate peptides were synthesised for three lysyl hydroxylases, JMJD4, JMJD6 and HSPBAP1 (which is proposed to be a lysyl-3-hydroxylase[285]), incorporating (*E*)-4,5-dehydro-L-Lysine in place of the substrate lysine (Table 18). Previous work had indicated that there was no reaction between Peptide 39 and JMJD4, or between Peptide 40 and JMJD6[262]. However, it was found that HSPBAP1 reacts with Peptide 41 and catalyses the

formation of a range of products (Figure 109A). Time course studies showed these products appear simultaneously (Figure 109B), suggesting that they are formed in parallel, potentially from a single intermediate[262].

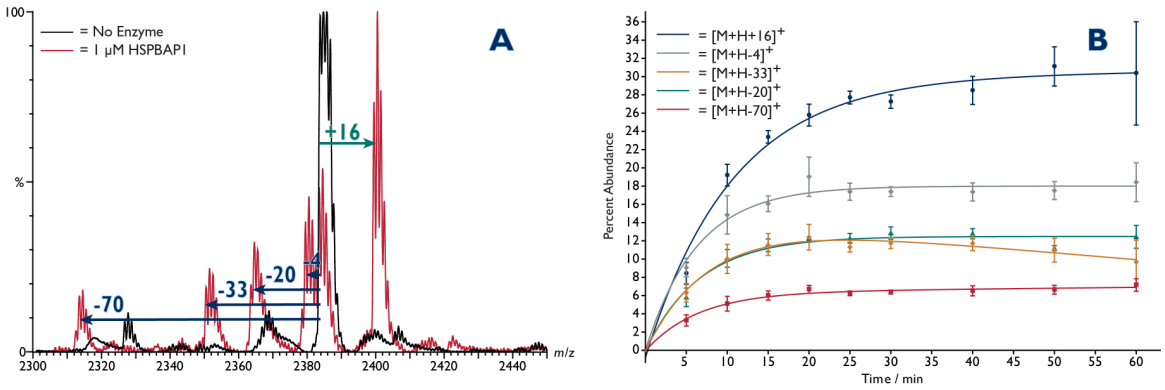


Figure 109 - Incubation of Peptide 41 with HSPBAP1 (and cofactors/co-substrates) leads to the formation of a range of products. A - Overlaid mass spectra B - Time course analysis showing that the different products form concurrently.

The resulting products and a no enzyme control were then analysed using MS-MS (MALDI-TOF-TOF). This study localised the modifications to the *N*-terminal (*E*)-4,5-dehydro-L-lysine residue of Peptide 41 (Table 19 and Table 20). Whilst more detailed characterisation, possibly by NMR spectroscopy or LC-MS, is required to fully identify the products, the [M+16] and [M-70] species are consistent with hydroxylation and side chain loss respectively.

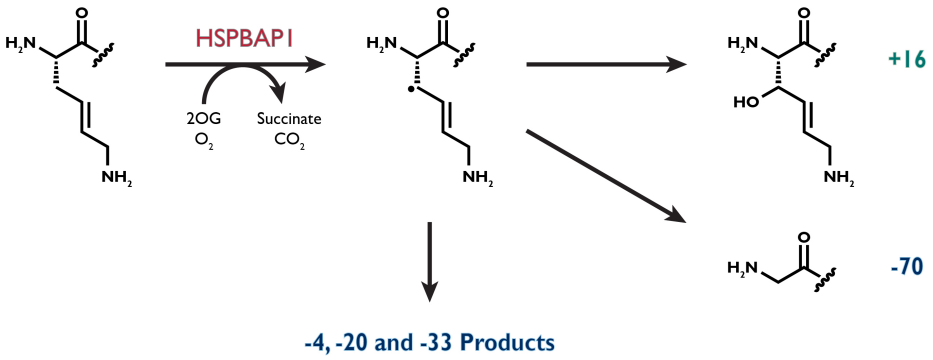


Figure 110 - Proposed products with observed masses of M+16 and M-70 formed when Peptide 41 is incubated with HSPBAP1. The identity of the products with masses of M-4, M-20 and M-33 has not been resolved and further study is required to identify the structure of these products.

	Observed *m/z*	Theoretical *m/z*		Theoretical *m/z*	Observed *m/z*	
b_1			K*			y_{20}
b_2	307.00	307.15	Y	2258.18		y_{19}
b_3		364.17	G	2095.11		y_{18}
b_4	477.45	477.26	I	2038.09	2037.92	y_{17}
b_5	590.15	590.34	I	1925.01	1925.27	y_{16}
b_6	693.35	693.35	C	1811.92		y_{15}
b_7	830.12	830.41	H	1708.91	1708.64	y_{14}
b_8	959.30	959.45	E	1571.85		y_{13}
b_9	1074.25	1074.48	D	1442.81		y_{12}
b_{10}		1187.56	L	1327.78	1327.57	y_{11}
b_{11}		1300.65	I	1214.70	1214.47	y_{10}
b_{12}	1437.57	1437.71	H	1101.62	1101.35	y_9
b_{13}		1566.75	E	964.56	964.38	y_8
b_{14}		1679.83	I	835.51	835.34	y_7
b_{15}	1842.85	1842.90	Y	722.43	722.24	y_6
b_{16}	1944.10	1943.94	T	559.37	559.24	y_5
b_{17}		2043.01	V	458.32	458.21	y_4
b_{18}	2099.74	2100.03	G	359.25	359.15	y_3
b_{19}		2228.13	K	302.23	302.17	y_2
b_{20}			R	174.13	174.12	y_1

Table 19 - MS-MS analysis of product ions formed following collision-induced dissociation of a peptide with *m/z* of 2400, corresponding to $[M+H]^+$ ion of the M+16 product. Theoretical *m/z* values were calculated assuming that an amino acid with a mass corresponding to the 3-hydroxy-(4,5)-dehydrolysine (K*), which was though to be the *N*-terminal residue. Good b and y ion coverage is consistent with this modification.

	Observed *m/z*	Theoretical *m/z*		Theoretical *m/z*	Observed *m/z*	
b_1			G			y_{20}
b_2	221.07	221.09	Y	2258.18		y_{19}
b_3		278.11	G	2095.11		y_{18}
b_4	391.27	391.20	I	2038.09	2038.41	y_{17}
b_5	504.11	504.28	I	1925.01		y_{16}
b_6	607.00	607.29	C	1811.92	1811.66	y_{15}
b_7	744.19	744.35	H	1708.91		y_{14}
b_8		873.39	E	1571.85	1571.58	y_{13}
b_9		988.42	D	1442.81		y_{12}
b_{10}		1101.50	L	1327.78	1327.46	y_{11}
b_{11}		1214.59	I	1214.70	1214.35	y_{10}
b_{12}		1351.65	H	1101.62	1101.34	y_9
b_{13}		1480.69	E	964.56	964.33	y_8
b_{14}		1593.77	I	835.51	835.27	y_7
b_{15}		1756.84	Y	722.43	722.21	y_6
b_{16}		1857.88	T	559.37	559.22	y_5
b_{17}		1956.95	V	458.32	458.18	y_4
b_{18}		2013.97	G	359.25	359.15	y_3
b_{19}	2142.18	2142.07	K	302.23	302.15	y_2
b_{20}			R	174.13	174.10	y_1

Table 20 - MS-MS analysis of product ions formed following collision-induced dissociation of a peptide with *m/z* of 2314, corresponding to $[M+H]^+$ ion of the M-70 product. Theoretical *m/z* values were calculated assuming the M-70 product was formed by loss of the lysyl side chain (resulting in an *N*-terminal glycine residue), which was though to be the *N*-terminal residue. Good y ion coverage and observation of ions with *m/z* matching those of the b_2, b_4, b_5, b_6 and b_7 ions are consistent with this modification.

Conclusions

Olefin cross metathesis has proved to be an efficient route to a range of lysine analogues that were successfully used to probe enzymatic catalysis centred on a lysine residue. The studies presented here have focussed on the reactivity of the 2OG oxygenases, but the approach could readily be applied to other enzyme classes and probe a wide range of lysine-centred PTMs. With this in mind, however, similar studies using analogues with alternative atoms (O/C) at the N^{ε}-position may not be feasible in a number of cases due to the decreased nucleophilicity at the ε-position that results if the nitrogen atom is exchanged.

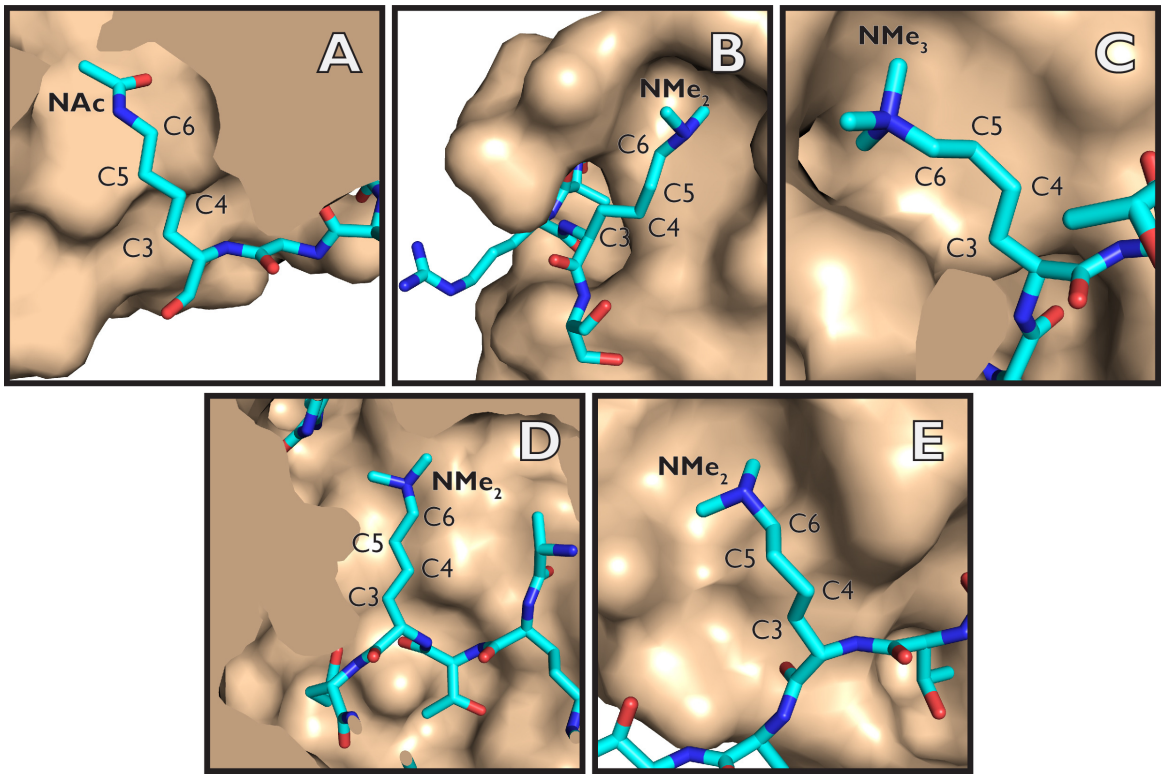


Figure 111 - Views from crystal structures of a range of lysyl residues in active sites and reader domains. A - Acetyl lysine in CREBBP (PDBID: 4N4F[286]). B - *Drosophila melanogaster* chromodomain recognising a methylated lysine residue (PDBID: 1KNA[287]) C - Binding of methylated H3K9 to human MPP8 chromodomain (PDBID: 3R93[288]) D - KDM1A active site (PDBID: 2UXN[289]). E - Set3 PHD finger in complex with dimethylated H3K4 (PDBID: 5TDR[290]).

Analyses of substrate structures for a selection of reader domains and that of KDM1A shows that lysine backbones adopt a range of alignments (Figure 111 and Table 21). The analogues could also be used to analyse the binding modes of reader domains in greater detail; for example,

some computational and structural analyses with peptides that incorporate analogue **67** have already been reported[279,288].

	PDBID	Protein/Domain (Species)	C3-C4-C5-C6 Dihedral
A	4N4F[286]	CBP Bromodomain (*Homo sapiens*)	-172.381
B	1KNA[287]	Chromodomain from HP1 (*Drosophila melanogaster*)	-89.494
C	3R93[288]	MPP8 chromodomain (*Homo sapiens*)	84.488
D	2UXN[289]	KDM1A (*Homo sapiens*)	166.759
E	5TDR[290]	Set3 PHD Domain (*Homo sapiens*)	-64.150

Table 21 - Dihedral angles of lysine residues highlighted in Figure 111.

As exemplified by the case of KDM4/KDM7B, the alignment of the backbone in a crystallographic setting may not entirely reflect the behaviour in solution. Therefore, the metathesis derived analogues could readily be used to characterise the binding of lysine to a range of proteins, including KDM1A, *in vitro*. For studies with reader domains, this could be achieved using techniques such as isothermal titration calorimetry.

Results with the KDM4 family suggest that a positively charged side chain is important for substrate binding and subsequent catalysis. A study with KDM6B and Peptide 38[262] also saw no reaction when N^{ε}- was replaced by carbon and suggests that the importance of N^{ε}- and, by implication, the positive charge, is not unique to the KDM4 subfamily of JmjC KDMs. In addition, analysing the role of lysyl backbone conformation in KDM4E and KDM7B catalysis suggest that there is significantly more flexibility in the lysine binding sites of both enzymes than was envisaged during analysis of substrate structures (Figure 100).

The reactivity exhibited by KDM3A with a number of the analogues is indicative of reactive modalities far beyond those that are currently understood. Similar to ɣ-butyrobetaine hydroxylase (BBOX)[291], KDM3A has been found to accept a nitrogen to carbon analogue (Peptide 32) as a substrate (Figure 107D). KDM3A was also found to catalyse the demethylation of a methyl ether (Peptide 37, Figure 107B) and oxidation of the ε-carbon of homolysine (Peptide 36, Figure 107C). Further studies investigating the reaction scope of KDM3A have originated from

these findings; leading to the discovery of novel reactivity that may not have been observed using other biochemical approaches[189]. Furthermore, these studies suggest that the lysyl side chain is, or at least can be, neutral during KDM3A catalysis, in contrast to the KDM4 subfamily. This observation is of significant help with respect to development of selective inhibitors.

In addition, reaction of Peptide 41 with HSPBAP1 led to the formation of a range of interesting products (Figure 109); these modifications have been localised to the 4,5-dehydrolysine residue at the *N*-terminus of the peptide although further studies are required to fully elucidate the products that are formed. The absence of reactivity with JMJD4 and JMJD6 is not necessarily due to backbone alignment, but may be due to incorrect projection of the C-H bond from the sp^2 C4 and C5 centres towards the reactive core of the enzyme.

The synthetic routes used were principally designed with the synthesis of peptide substrates in mind. However, slight modification of the synthetic methodology (*e.g.* altering the protecting group strategy) would readily allow for the production of substrate analogues for testing with enzymes that act upon lysine or methylated lysine. One such enzyme is TMLH, a 2OG oxygenase in the carnitine biosynthesis pathway that catalyses the formation of (*2S,3S*)-hydroxy-N^{ϵ}-trimethyllysine from N^{ϵ}-trimethyllysine[134].

Summary

Following the development of a novel, robust route to peptide probes *via* olefin cross-metathesis, a range of novel and unexpected reactivities have been observed, particularly with KDM3A and with HSPBAP1, that would not have been accessible using a top-down approach. Studies with the geometrically constrained lysine analogues, **65** and **58** have revealed that the KDM7B active site has a degree of flexibility and additionally that some flexibility in the K9 backbone is required for optimal binding to KDM4 subfamily. The observation that the KDM4 subfamily requires a positively charged side chain, whereas this is not required for KDM3A, is also of interest and will aid the development of selective inhibitors. The work presented here

has helped to inform the development of further studies with KDM3A and can be readily adapted and applied to a range of other biologically significant targets.

Chapter 7
Summary and Conclusions

The 28 collagen proteins are the most abundant in the human body and are the principal protein component of the extracellular matrix (ECM)[2,23]. Collagen is subjected to a range of post-translational modifications during collagen biosynthesis[24]. The formation of (*2S,4R*)-hydroxyproline, (*2S,3S*)-hydroxyproline and (*2S,5R*)-hydroxylysine are key in allowing for the formation of stable, mature collagen proteins and is mediated by the collagen prolyl-4-hydroxylases (C-P4Hs), prolyl-3-hydroxylases (P3Hs) and procollagen-lysine, 2-oxoglutarate 5-dioxygenases (PLODs), which are all members of the Fe^{II}/2-oxoglutarate dependent oxygenase superfamily[103]. The C-P4Hs, P3Hs and PLODs are all poorly characterised, with limited biochemical and structural information available. For example, there are no available structures of the active site of any of the human collagen hydroxylases[24,103].

The C-P4Hs are $\alpha_2\beta_2$ tetramers with a molecular weight of around 240 kDa. To allow for tetramer formation, two intrastrand disulfide bonds are required in the α-subunit[164], which is insoluble in isolation (thereby necessitating co-expression of the α- and β-subunits). These two factors make the recombinant expression of C-P4H tetramer a significant technical challenge. The work in Chapter 2 presents efforts towards the development of a reliable methodology for the recombinant production of C-P4H tetramer in *E. coli.* An *E. coli* system was chosen to afford quantities of protein suitable for subsequent biochemical and structural studies.

Three approaches were trialled using either a single expression vector encoding both subunits (building upon previously reported approaches[173,174]), two separate vectors under the control of the same promoter or two separate vectors under the control of different promoters. Each of these methodologies was used in small scale expression trials using *E. coli* Origami expression strains, which provide a cytosolic environment that allows for disulfide formation. A range of new C-P4H expression vectors have been produced and despite evidence for the expression of the α-subunit in some instances, large quantities of C-P4H tetramer have not been isolated. The

use of alternative expression systems, such as insect cell expression, may therefore be required for establishment of an appropriate method for recombinant C-P4H production.

In the absence of an expression system for the recombinant C-P4H enzymes, a prolyl hydroxylase from *Paramecium bursaria* chlorella virus I (vCPH), a 25 kDa monomeric protein, was investigated as a model collagen prolyl hydroxylase. Despite being reported as a collagen prolyl hydroxylase[84,179], the stereochemistry and regiochemistry of vCPH catalysis was previously unknown. The work in Chapter 3 has characterised vCPH as a prolyl-4-hydroxylase that catalyses the formation of (*2S,4R*)-hydroxyproline. vCPH catalyses the formation of (*2S,4R*)-hydroxyproline in collagen repeat peptides but not substrates of the hypoxia inducible factor (HIF) prolyl hydroxylases. This activity, along with sequence analysis, indicates that vCPH is a *bone fide* collagen prolyl-4-hydroxylase.

Similar to the human C-P4Hs, vCPH does not require ascorbate (**3**) for catalysis. However, vCPH (and the C-P4Hs[103]) exhibits a strong ascorbate requirement to reach maximum reaction velocity. The ascorbate requirement of vCPH catalysis has been shown to vary with sequence location of the target proline and the ability of vCPH to catalyse the formation of (*2S,4R*)-hydroxyproline in weaker substrates is reduced in the absence of ascorbate (**3**). The high ascorbate dependence of C-P4H has been implicated in the development of scurvy[56] and the precise role that ascorbate (**3**) plays in 2OG oxygenase catalysis is still not understood on a molecular level. vCPH therefore may provide a model system for the understanding of collagen prolyl-4-hydroxylase activity and also for the role that ascorbate plays in catalysis.

Using the characterisation of vCPH conducted in Chapter 3 as a basis, a vCPH inhibition assay was developed in Chapter 4 and was used to assess the effect of a library of inhibitors on vCPH. This inhibitor library featured a number of compounds that have been developed as inhibitors of the HIF prolyl hydroxylase PHD2, including compounds that are currently in phase II (Molidustat (**11**)[147]) and phase III (FG-4592 (**8**)[143], Vadadustat (**9**)[144,145]and GSK1278863 (**10**)[146]) clinical trials.

Inhibitors of PHD2, including the clinical candidates, were found to inhibit vCPH, which is consistent with the limited selectivity versus other 2OG-dependant prolyl hydroxylases that is observed with these compounds[194]. Analysis of novel PHD2 inhibitors (notably JHM634 (**42**) and JHM532 (**40**)) and compounds developed for the inhibition of the prolyl-3-hydroxylase OGFOD1 (CCT6-140 (**19**)[158,201]) have shown that it is possible to achieve selectivity between PHD2, OGFOD1 and vCPH. Therefore, the development of a selective C-P4H inhibitor is a realistic possibility, and will only be aided by the successful development of C-P4H and P3H assays.

Structural studies have also been initiated with vCPH as no structures of vCPH in complex with an inhibitor or substrate are available. These studies have yielded novel crystal forms and data sets have been collected to a resolution of 3.51 Å. Further optimisation of crystallisation conditions are required for structural analysis of inhibitor binding and comparison to the binding modes observed in PHD2[194].

In addition to these studies, the presence of (*2S,4R*)-hydroxyproline, (*2S,3S*)-hydroxyproline and (*2S,5R*)-hydroxylysine in collagen samples has been observed by UPLC-MS amino acid analysis in Chapter 5. These studies have successfully analysed acid-hydrolysed collagen samples from rat tail tendon and bovine Achilles tendon. These analyses provide no evidence for the presence of 3- or 4- hydroxylysine (which are formed by JMJD7[135] and JMJD4[113,135] in different contexts) in collagen and do not provide any evidence for the presence of 3,4-dihydroxyproline in collagen. Attempts to purify collagen from tissue samples and apply the amino acid analysis methods to clinically-derived samples have been hampered by the lack of an efficient purification procedure.

Finally, in Chapter 6, a range of cross-metathesis derived lysine analogues have been synthesised, incorporated into peptides and used to analyse 2OG-dependent histone demethylase catalysis. These approaches have been used to analyse the charge and conformational requirements of KDM4 and KDM7B catalysis and have revealed that both enzyme active sites, particularly the KDM7B active site, have a degree of flexibility that allows

for substrate binding. Further studies with the KDM4 subfamily have revealed that the KDM4 enzymes require a positive charge at N^{ε} since neutral substrate analogues are not accepted. This is contrasted by KDM3A, which accepts neutral substrates; the lysine analogues synthesised as part of this study have revealed a number of new modes of reactivity for KDM3A, far beyond what was previously perceived, and have initiated a new line of study into KDM3A catalysis. In addition to the novel reactivity modes of KDM3A that have been observed, a range of interesting products were observed when HSPBAP1 was reacted with a peptide containing 4,5-dehydrolysine. These two examples have shown that the use of substrate analogues can allow for study of underlying chemical mechanisms and to discover new modes of reactivity.

Overall, the studies presented in this thesis have provided biochemical insight into the action of collagen prolyl-4-hydroxylases and investigated the effect of a range of inhibitors on a model collagen prolyl-4-hydroxylase. Additionally, UPLC-MS analysis of collagen samples and studies utilising cross metathesis-derived lysine analogues to probe histone demethylase and lysyl hydroxylase catalysis. The key findings are:

- Large-scale recombinant production of mammalian C-P4Hs remains a significant technical challenge that hampers further biological study.
- A viral collagen prolyl hydroxylase can be used as a model C-P4H. Inhibition studies have revealed that clinical candidates that inhibit the HIF PHDs also inhibit collagen prolyl hydroxylases, and that development of selective C-P4H, PHD2 and OGFOD1 inhibitors is possible.
- There is no evidence for the presence of 3-hydroxylysine, 4-hydroxylysine or 3,4-dihydroxyproline (PTMs that are catalysed by other 2OG oxygenases) in collagen.
- Studies with metathesis-derived lysine analogues have revealed that the reaction scope of the histone demethylase KDM3A is far wider than previously envisaged.

Chapter 8
Materials and Methods

General Methods

All reagents and solvents were purchased from Sigma-Aldrich unless otherwise stated. Yeast extract, tryptone and agar were obtained from Oxoid or ThermoFisher Scientific. Single use 0.5 mL, 1.5 mL, 15 mL and 50 mL centrifuge tubes were supplied by Sarstedt.

Water was purified using a Millipore MilliQ system. Sterilisation of non-sterile media and equipment was conducted *via* autoclaving in at 121°C for 20 minutes using either a LTE TouchClave® II benchtop autoclave (LTE Scientific) or a LTE Falcon benchtop autoclave (LTE Scientific). Other solutions and buffers were sterilised by passing through a 0.22 µm filter (Merck).

All solid reagents were weighed using an appropriate scientific balance.

Volumes of aqueous solution were measured using appropriate pipettes (Gilson) for volumes less than 1 mL, graduated pipettes for volumes between 1 mL and 25 mL and measuring cylinders for volumes greater than 25 mL.

Sequence Alignments

Sequence alignments were conducted using the Clustal Omega server[119] and visualised using JalView.

pH Measurements

pH measurements were conducted using a Hanna Instruments HI 9321 pH meter and a 5 mm diameter electrode. Prior to each use, the pH meter was calibrated between pH 7 and pH 4 or between pH 7 and pH 10 using commercially sourced phthalate (pH 4), phosphate (pH 7) and borate (pH 10) buffer solutions (ThermoFisher Scientific). The electrode was stored in saturated KCl solution between uses.

Determination of Protein Concentration

Protein concentration was determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc.) by monitoring the absorbance at 280 nm, using the 'Protein A280' module of the controlling software. The absorbance was normalised to that of sample buffer to zero and the absorption coefficient, ε, of the protein of interest was calculated using the ProtParam tool on the ExPASy server[292].

Determination of DNA Concentration

DNA concentration was determined by monitoring absorbance at λ = 260 nm using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc.); DNA concentration was determined using the 'Nucleic Acid' software module, with the absorbance of sterile ddH_2O being set to 0.

Culture Density Measurements

Cell density was measured by assessing the optical density of culture medium at λ = 600 nm (OD_{600}). This was carried out by measuring the absorbance of a 1 mL sample in a cuvette (Fisher Scientific) using an Amersham Biosciences Novaspec III visible spectrophotometer; with the absorbance of the uninoculated growth medium being set to 0.

Lyophilisation

Aqueous samples were lyophilised by snap freezing in liquid N_2 and placing the sample in a vessel attached to a Scania coolsafe chiller unit; a vacuum was achieved and maintained using a vacuubrand RC6 vacuum pump.

Concentration of Proteins

Proteins were concentrated using an Amicon Ultra-15 Centrifugal Filter Units (Merck) of appropriate molecular weight cut off and volume, according to manufacturer's instructions. Centrifuges and samples were kept at 4°C during protein concentration.

Buffer Exchange

Samples were buffer exchanged using either a PD10 column (GE healthcare), Micro BioSpin column (BioRad) or Slide-A-Lyzer dialysis cassette (ThermoFisher Scientific), following the manufacturer's instructions.

Antibiotic Selection

Antibiotic selection was achieved by treating growth media with 1000 x stocks of antibiotic, resulting in the final concentrations detailed in Table 22. All antibiotics were obtained from Apollo Scientific. 1000 x stocks were prepared in ddH_2O (unless otherwise stated) and sterile filtered through a 0.45 µm filter (Grenier) prior to use.

Antibiotic	Final Concentration / µg mL^{-1}
Ampicillin	100
Kanamycin	30
Tetracycline	15 (in 50% EtOH)
Spectinomycin	50
Chloramphenicol	35 (in EtOH)

Table 22 - Final concentrations used for antibiotic selection.

Transformations

Plasmid DNA was incorporated into *E. coli* by incubating plasmid DNA with 40 µL of competent cells on ice for 15 minutes prior to a 30 second heat shock at 42°C in a Grant W14 Waterbath. The cells were then incubated on ice for 1 minute prior to the addition of 300 µL 2TY medium. The cells were incubated for one hour at 37°C with 200 rpm shaking in a New Brunswick Scientific G25 environmental shaker. Colonies were then selected on LB agar plates that had been treated with the appropriate antibiotic.

Starter Cultures

Starter cultures were initiated in antibiotic treated 2TY (100 mL medium in a 500 mL conical flask) *via* inoculation with a single colony from an agar plate, or from a single scraping from a glycerol stock. The culture was then incubated overnight at 37°C with 180 rpm shaking in a New Brunswick Scientific G25 environmental shaker.

Production of Chemically Competent Cells

Commercially sourced cells were streaked onto an antibiotic free LB-agar plate. A single colony was used to initiate a 100 mL starter culture in antibiotic free 2TY. 200 µL of starter culture was used to inoculate 100 mL antibiotic free 2TY in a 500 mL flask. This culture was grown to an OD_{600} of 0.6-0.8 at 37°C with 200 rpm shaking in a New Brunswick Scientific G25 environmental shaker. Cells were then harvested *via* centrifugation (750 x *g* for 5 minutes) in pre-chilled centrifuge tubes. The pellets were resuspended in 25 mL sterile 100 mM $CaCl_2$, 20% (v/v) glycerol and kept on ice for 20 minutes. Cells were then repelleted *via* centrifugation at 750 x *g* for 5 minutes, before being resuspended in 1 mL sterile 100 mM $CaCl_2$, 20% (v/v) glycerol. The resuspended cells were then divided into 40 µL aliquots, snap frozen in liquid nitrogen and stored at -80°C. For strains with intrinsic antibiotic resistance, the LB agar and subsequent growth media were treated with the appropriate antibiotic.

Growth Media

Unless otherwise stated, all media components were sterilised or autoclaved prior to use.

2 x Tryptone / Yeast Extract (2TY) Medium

Bacto Tryptone	16 g L^{-1}
Yeast Extract	10 g L^{-1}
NaCl	5 g L^{-1}

Table 23 - Composition of 2TY medium.

The medium was prepared in accordance with Table 23 and the pH adjusted to 7.0 with 5 M NaOH.

Luria Broth (LB) Medium

Bacto Tryptone	10 g L^{-1}
Yeast Extract	5 g L^{-1}
NaCl	10 g L^{-1}

Table 24 - Composition of LB medium.

The medium was prepared according to Table 24 and the pH adjusted to 7.0 with 5 M NaOH.

Terrific Broth (TB) Medium

Bacto Tryptone	12 g L^{-1}
Yeast Extract	24 g L^{-1}
Glycerol	4 mL L^{-1}

Table 25 - Composition of TB medium. Final concentrations are indicated. These components were first prepared in 900 mL volume before addition of 100 mL phosphate buffer.

Bacto tryptone, yeast extract and glycerol (Table 25) were dissolved in 900 mL H_2O and autoclaved. To this, 100 mL of sterile filtered 0.17 M KH_2PO_4, 0.72 M K_2HPO_4 solution was added.

Autoinduction Medium

1 M $MgSO_4$, 50 x 5052 and 50 x M buffers were prepared and sterile filtered. For 1 L of autoinduction medium, 20 mL 50 x 5052, 20 mL 50 x M and 2 mL 1 M $MgSO_4$ were added to 958 mL of autoclaved ZY mix.

ZY Mix

Bacto Tryptone	10 g L^{-1}
Yeast Extract	5 g L^{-1}

Table 26 - Composition of ZY mix.

50 x 5052 Buffer

Glycerol	25% (v/v)
D-Glucose	2.5% (w/v)
α-Lactose	10% (w/v)

Table 27 - Composition of 50 x 5052 buffer.

50 x M Buffer

$Na_2HPO_4.7H_2O$	1.25 M (335 g L^{-1})
KH_2PO_4	1.25 M (170 g L^{-1})
NH_4Cl	2.50 M (134 g L^{-1})
Na_2SO_4	0.25 M (35.5 g L^{-1})

Table 28 - Composition of 50 x M buffer.

PowerBroth™

PowerBroth™ was prepared from a commercial stock (Molecular Dimensions) according to manufacturer's instructions and autoclaved by staff at OPPF-UK.

Electrophoresis and Immunoblotting

SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gels were produced according to the recipes in Table 29 and Table 30.

Samples were prepared by mixing 4 parts sample with 1 part 5 x loading buffer (250 mM Tris-HCl pH 6.8, 30% (v/v) Glycerol, 5% (w/v) SDS, 5% (v/v) β-mercaptoethanol, 0.02% (w/v) Bromophenol Blue) and heated to 100°C for 10 minutes before being loaded alongside the PageRuler molecular weight marker (ThermoFisher Scientific). The gel was then run at a potential difference of 190 V in 25 mM Tris, 192 mM Glycine, 0.1% (w/w) SDS, pH 8.8 until the dye front reached the bottom of the gel.

Separating Gel

ddH_2O	0.79 mL
1.5 M Tris-HCl pH 8.8	6.00 mL
30% (w/v) Acrylamide	5.00 mL
10% (w/v) SDS	0.12 mL
10% (w/v) Ammonium Persulfate	0.10 mL
Tetraethylenediamine	0.01 mL

Table 29 - Composition of separating gel for SDS-PAGE.

Stacking Gel

ddH_2O	2.30 mL
0.5 M Tris-HCl pH 6.8	0.40 mL
30% (w/v) Acrylamide	0.32 mL
10% (w/v) SDS	0.032 mL
10% (w/v) Ammonium Persulfate	0.06 mL
Tetraethylenediamine	0.01 mL

Table 30 - Composition of stacking gel for SDS-PAGE.

NuPAGE™ 4-12% (w/v) Bis-Tris gradient gels (Invitrogen) were also used. After peparing a 2 x loading buffer from commercial 4 x stock (Invitrogen) through the addition of water and BME, 2 x loading dye was added to samples prior to boiling for 15 minutes. Samples were then loaded and the gel was run in 1 x MOPS-SDS buffer, which was freshly prepared from a commercial 20 x stock (Invitrogen). Gels were run at a potential difference of 190 V for 35 minutes.

Once the run was complete, the gel was washed with water and was stained by shaking in stain buffer (0.25% (w/v) Coomassie Brilliant Blue, 10% (v/v) AcOH, 50% (v/v) MeOH in ddH_2O) for 20 minutes. Destaining of gels was acheived by shaking in destain buffer (10% (v/v) AcOH, 10% (v/v) MeOH in H_2O) until bands were visible. Alternatively, gels were visualised by shaking in InstantBlue (Expedeon), following manufacturers' instructions.

Native PAGE

Non-denaturing polyacrylamide gel electrophoresis (Native PAGE) was conducting using gels prepared using the recipes detailed in Table 31 and Table 32. 8 µL of sample was mixed with 2 µL of 5 x loading buffer (250 mM Tris-HCl pH 6.8, 30% (v/v) Glycerol, 0.02% (w/v) Bromophenol Blue) before being loaded onto the gel. Samples were run against the SERVA BN/CN native marker (SERVA Electrophoresis GmbH). The gel was run at a potential difference of 90 V in 25 mM Tris, 192 mM Glycine, pH 8.8 at 4˚C until the dye front reached the bottom of the gel. Gels were stained and destained using the same procedure used for SDS-PAGE gels.

Separating Gel

ddH_2O	0.91 mL
1.5 M Tris-HCl pH 8.8	6.00 mL
30% (w/v) Acrylamide	5.00 mL
10% (w/v) Ammonium Persulfate	0.10 mL
Tetraethylenediamine	0.01 mL

Table 31 - Composition of separating gel for native PAGE.

Stacking Gel

ddH_2O	2.33 mL
0.5 M Tris-HCl pH 6.8	0.40 mL
30% (w/v) Acrylamide	0.32 mL
10% (w/v) Ammonium Persulfate	0.06 mL
Tetraethylenediamine	0.01 mL

Table 32 - Composition of stacking gel for native PAGE.

Agarose Gel Electrophoresis

Tris Base	242 g L^{-1}
Glacial Acetic Acid	57.1 mL L^{-1}
0.5 M EDTA	100 mL L^{-1}

Table 33 - Composition of 50 x TAE buffer.

1% (w/v) agarose gels were prepared using 0.5 g agarose in 50 mL of 1 x TAE buffer, which was dissolved by heating the mixture using a SuperWave 750 microwave (Tecnolec). 5 µL of SYBR Safe™ DNA gel stain (Invitrogen) was added to the dissolved gel before pouring; gels were cast in an H2-SET gel tray (Anachem), with a comb inserted at one end of the gel to allow for sample loading. The gels were allowed to cool at room temperature until the gel had set. Samples were mixed with 6 x Loading Dye (New England Biolabs) and run against a 1 kb DNA ladder (New England Biolabs) or 100 bp DNA ladder (New England Biolabs). Gels were run in 1 x TAE buffer (the composition of 50 x TAE is detailed in Table 33) on a Bio-Rad system at a constant potential difference of 100 V for approximately 40 minutes.

Western Blotting

Samples were separated by SDS-PAGE; the resulting unstained gel was blotted onto Hybond ECL membrane (0.2 µm pore size, GE Helthcare) in fresh transfer buffer (25 mM Tris, 192 mM glycine, 20% MeOH) for 16 hours at a constant current of 15 mA at 4°C using a Mini Trans-Blot Electrophore (Bio-Rad). Ponceau S solution was used to visualise successful protein transfer before the membrane was destained by shaking in ddH_2O.

Tris Base	3.152 g L^{-1}
NaCl	35.064 g L^{-1}
Tween-20 (Qiagen)	1 mL L^{-1}
HCl	To pH 8.5

Table 34 - Composition of 2 x TBS-T buffer.

The membrane was washed, with shaking, in 1 x TBS-T buffer (10 mM Tris-HCl pH 8.5, 300 mM NaCl, 0.05% (v/v) Tween-20; prepared from a 2 x stock, Table 34) at room temperature. The blot was subsequently blocked by washing in 5% skimmed milk powder (w/v, Tesco Stores UK) in 1 x TBS-T for 2 - 3 hours at room temperature.

Blots were then incubated in primary antibody at room temperature for 2 hours prior to washing three times in 0.3% (w/v) skimmed milk powder in 1 x TBS-T for ten minutes at room temperature with shaking.

Target	Primary Antibody	Secondary Antibody	Detection
Hexahistidine Tag	Monoclonal mouse anti-his (abcam, ab18184)	Goat anti-mouse horseradish peroxidase conjugate (Invitrogen)	Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare)

Table 35 - Antibodies and detection reagents used for probing dot blots and western blots.

The blot was then incubated in secondary antibody for 45 minutes at room temperature prior to further washing; three times in 0.3% (w/v) skimmed milk powder in 1 x TBS-T for ten minutes at room temperature with shaking. Where horseradish peroxidase-conjugated secondary antibodies were used, detection was achieved using Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare), following manufacturer's instructions, prior to imaging with a BioRad ChemiDock MP gel imager.

Dot Blotting

Sample (1 µL, most commonly lysate from expression trials) was spotted directly onto nitrocellulose blotting paper (GE Healthcare). The blot was then washed in TBS-T prior to processing following the same procedure as western blots, starting from the blocking step.

Plasmid Preparation

E. coli XL10 Gold (prepared from a commercial stock, Agilent) were used for all transformation steps involved in the preparation of plasmids unless otherwise indicated. Primers were purchased from Sigma-Aldrich.

DNA AmplificationddH_2O	11.50 µL
Q5 Reaction Buffer (5 x stock, New England Biolabs)	5.00 µL
Q5 High GC Enhancer (5 x stock, New England Biolabs)	5.00 µL
Forwards Primer, 10 µM	1.25 µL
Reverse Primer, 10 µM	1.25 µL
Template DNA	0.50 µL
dNTP mix (New England Biolabs)	1.00 µL
Q5 DNA Polymerase (New England Biolabs)	1.00 µL

Table 36 - Reaction mixture for a typical PCR.

DNA was amplified using the polymerase chain reaction (PCR). A reaction mixture of 25 µL was prepared according to Table 36. The PCR was run in a BioRad MJ Mini thermal cycler.

Samples were maintained at 95°C for 2 minutes to denature the template DNA. 40 cycles then followed:

i. Denaturing at 95°C for 30 s
ii. Annealing at 5°C below the lowest primer T_m for 30 s
iii. Extension at 68°C for 1 min per kb of plasmid

Following these cycles, reactions were held at 68°C for a 10 min final extension prior to holding at 4°C.

Site Directed Mutagenesis

Primers containing the desired mutation were designed following the protocols outlined in the QuikChangeTM site-directed mutagenesis kit (Agilent), and used to amplify the full vector and insert sequence. A reaction mixture was set up as outlined in Table 37 and the mutagenesis reaction was run in a BioRad MJ Mini thermal cycler.

ddH_2O	8.50 µL
Q5 Reaction Buffer (5 x stock, New England Biolabs)	5.00 µL
Q5 High GC Enhancer (5 x stock, New England Biolabs)	5.00 µL
Forwards Primer, 10 µM	1.25 µL
Reverse Primer, 10 µM	1.25 µL
Template DNA	1.00 µL (approx. 50 ng)
dNTP mix (New England Biolabs)	2.50 µL
Q5 DNA Polymerase (New England Biolabs)	0.50 µL

Table 37 - Reaction mixture for typical mutagenesis PCR.

Samples were heated to 94°C for 120 s to denature the template DNA. 18 cycles then followed:

i. Denaturing at 94°C for 45 s
ii. Annealing at 5°C below the lowest primer T_m for 60 s
iii. Extension at 68°C for 1 min per kb of plasmid

Following these cycles, reactions were held at 68°C for a 20 min final extension prior to holding at 4°C.

1 µL of DpnI (New England Biolabs) was added to the mixture to digest any supercoiled DNA; the digest was conducted for 3 hours at 37°C. 5 µL of the resultant mixture was used in a transformation and colonies selected on LB agar plates that were treated with the appropriate anitbiotic.

DNA Digests

DNA was digested using restriction enzymes purchased from New England Biolabs. When possible, a double digest (*i.e.* simultaneous digestion with two different restriction enzymes) was conducted using the conditions recommended by New England Biolabs. The digest was set up and conducted at 37°C for 3 hours; standard double digest conditions are detailed in Table 38. Products were purified using a 1% agarose gel in 1 x TAE buffer. Bands were excised and DNA extracted using a gel extraction kit (ThermoFisher Scientific) according to manufacturers' instructions and stored at -20°C.

DNA for Digestion	1.00 µg
Restriction Enzyme 1	1.00 µL
Restriction Enzyme 2	1.00 µL
10 x Buffer (New England Biolabs)	2.50 µL
Sterile H_2O	To a total volume of 25 µL

Table 38 - Typical conditions for an enzymatic double digest.

Ligations

Ligations were carried out using 1:1, 3:1 and 6:1 molar ratios of insert to cut vector. In a total reaction volume of 10 µL, 7 µL of vector/insert mix was added to 1 µL T4 DNA Ligase (New England Biolabs) and 1 µL 10 x T4 Ligase buffer (New England Biolabs). Reactions were incubated at 16°C overnight in a BioRad MJ Mini thermal cycler.

5 µL of the resultant ligation reaction mixture was used in a transformation with 40 µL of competent *E. coli* XL10 Gold. Colonies were selected on the appropriate antibiotic; If no colonies were observed then the transformation was repeated with commercially sourced NEB 5-α

Competent *E. coli*. Resulting colonies were used for initiation of starter cultures, plasmid purification and sequencing prior to downstream use.

Colony PCR

A colony was picked from a LB agar plate using a sterile 200 µL pipette tip. The tip was then scraped against the bottom of a PCR tube to transfer cells. The tip was subsequently used to initiate a 10 mL 2TY overnight culture in a 50 mL centrifuge tube.

BioMix Red (Bioline)	12.50 µL
Forwards Primer, 10 µM	1.25 µL
Reverse Primer, 10 µM	1.25 µL
Sterile H_2O	10.00 µL

Table 39 - Reaction mixture for colony PCR.

25 µL of PCR mix (Table 39) was then added to the PCR tube and was subjected to the following programme in a BioRad MJ Mini thermal cycler. Following initial denaturation of template DNA at 94°C for 120 s, 18 cycles then followed:

i. Denaturing at 95°C for 45 s
ii. Annealing at 5°C below the lowest primer T_m for 60 s
iii. Extension at 72°C for 1 min per kb of plasmid

Following these cycles, reactions were held at 68°C for a 20 minute final extension prior to holding at 4°C.

The resulting mixture was loaded directly onto an agarose gel and analysed by electrophoresis. Plasmid was then purified (by miniprep) from the overnight cultures that were initiated from colonies that corresponded to a band of the correct size on the resulting agarose gel. Purified plasmid was subsequently sequenced by Sanger sequencing.

Ligation Independent Cloning

Plasmid	Forward Primer Overhang	Reverse Primer Overhang
pNIC28-BSA4	TACTTCCAATCC*ATG*	TATCCACCTTTACTG*CTA*
pBAD LIC	TTTAAGAAGGAGATATAGATC*ATG*	TTATGGAGTTGGGATCTTA*TTA*
pBAD HIS$_6$ LIC	TACTTCCAATCCAATGCA	TTATCCACTTCCAATGTTA*TTA*
pET His$_6$ MBP TEV LIC	TACTTCCAATCCAATGCA	TTATCCACTTCCAATGTTATTA
pET His$_6$ MBP N$_{10}$ TEV LIC	TACTTCCAATCCAATGCA	TTATCCACTTCCAATGTTATTA

Table 40 - LIC overhangs added to the 5' end of complimentary forwards and reverse sequence primers. Sequences corresponding to the initiator methionine and stop codon are indicated in italics.

Ligation independent cloning was carried out following the methods described methods[177]. The gene of interest was amplified *via* PCR, the overhangs detailed in Table 40 were added to the 5' end of primer sequences (start and stop codons included and indicated in italics).

Plasmid	Restriction Enzyme	Insert: Nucleotide for LIC treatment	Vector: Nucleotide for LIC treatment
pNIC28-BSA4	BsaI	dCTP	dGTP
pBAD LIC	EcoRV	dGTP	dCTP
pBAD HIS6 LIC	SspI	dCTP	dGTP
pET His$_6$ MBP TEV LIC	SspI	dCTP	dGTP
pET His$_6$ MBP N$_{10}$ TEV LIC	SspI	dCTP	dGTP

Table 41 – Restriction enzyme and nucleotide requirements for LIC treatment of insert and vector.

The LIC plasmid was digested with the enzyme detailed in Table 41 (restriction enzymes sourced from New England Biolabs), following the manufacturer's recommended conditions. Cut vector was purified using a PCR cleanup kit (ThermoFisher Scientific) or by running on a 1% agarose gel followed by band excision and use of a Gel Extraction Kit (ThermoFisher Scientific); kits were used following manufacturer's instructions, substituting elution buffer for autoclaved ddH_2O.

PCR Product	5.00 µL
10 x T4 Polymerase Buffer (Invitrogen)	1.00 µL
10 x BSA (New England Biolabs)	1.00 µL
DTT (100 µM)	0.50 µL
Nucleotide (25 mM, New England Biolabs)	1.00 µL
T4 DNA Polymerase (Invitrogen)	0.25 µL
Sterile H_2O	1.25 µL

Table 42 - Reaction mixture for LIC treatment of PCR product.

BsaI/BsaRI Digested Vector	50.00 µL
10 x T4 Polymerase Buffer (Invitrogen)	10.00 µL
10 x BSA (New England Biolabs)	10.00 µL
DTT (100 µM)	5.00 µL
Nucleotide (25 mM, New England Biolabs)	10.00 µL
T4 DNA Polymerase (Invitrogen)	2.50 µL
Sterile H_2O	12.50 µL

Table 43 - Reaction mixture for LIC treatment of digested vector.

The purified PCR product and digested vector were used to set up further reactions ("LIC treatment") using the conditions detailed in Table 42 and Table 43 respectively. Reactions were incubated at 22°C for 30 minutes, then at 75°C for 20 minutes (in order to heat inactivate the enzyme), prior to cooling and storage at 4°C.

1 µL of treated vector was added to 2 µL treated insert and incubated at room temperature for at least ten minutes. The resulting mixture was then transformed in XL10 following standard procedures. Colonies were subsequently selected on LB-agar treated with the appropriate antibiotic; for plasmids that allowed for selection (*e.g.* pNIC28-BSA4) on sucrose, LB-agar was supplemented with 5% sucrose.

In-Fusion™ Cloning

Cloning into pOPINRSF, pOPINRSE, pOPINCDF and pOPINCDE vectors was achieved with the assistance of Dr Heather Rata of OPPF-UK, following OPPF-UK standard operating procedures[176].

PCR was conducted using 25 µL Phusion Flash 2 x Master Mix (ThermoFisher Scientific), 3 µL of forwards primer, 3 µL of reverse primer, 2 µL template and 17 µL ddH_2O. The first step was to hold at 98°C for 10 s to denature the template DNA. 30 cycles then followed:

i. Denaturing at 98°C for 1 s
ii. Annealing at 60°C for 5 s
iii. Extension at 72°C for 15 s per kb of plasmid

Following these cycles, reactions were maintained at 72°C for a 2 min final extension period prior to holding at 4°C. 5 µL of product was used to analyse the success of the PCR by agarose gel electrophoresis, using a 0.8% (w/v) agarose gel.

Template DNA was then was then digested *via* addition of 5 µL of DpnI solution (New England Biolabs, 5 units in 1 x CutSmart buffer) to 45 µL of PCR product before incubating at 37°C for 60 minutes. The PCR product was purified using an Agencourt AMPure XP magnetic bead purification kit (Beckman Coulter), following manufacturer's instructions.

1 µL of linearised pOPIN vector was added to PCR product (10 - 250 ng) and ddH_2O (to a final volume of 10 µL) which was then mixed with lyophilised InFusion™ reagent and incubated at 42°C for 30 minutes. The reaction mixture was then put on ice and diluted *via* the addition of 40 µL 10 mM Tris-HCL pH 8.0, 1 mM EDTA.

3 µL of the diluted reaction was then used in a transformation and colonies were selected on LB agar treated with antibiotic, X-Gal and IPTG (**17**). Plates were incubated at 37°C overnight and white colonies were picked, overnight cultures initiated and plasmid purified before verification by PCR and Sanger DNA sequencing.

Plasmid Purification

From an overnight 100 mL culture, plasmid DNA was extracted by pelleting 5 mL of culture followed by purification using a Miniprep kit (ThermoFisher Scientific), following the manufacturer's instructions. When larger quantities of DNA were required for downstream uses, 50 mL of culture was pelleted followed by the use of a Midiprep kit (ThermoFisher Scientific). Autoclaved ddH_2O was used as a direct substitute for elution buffer in all instances.

DNA Sequencing

Sanger DNA Sequencing was conducted by Source BioScience. Sequencing primers are detailed in Table 44.

Primer	Sequence
T7 Forwards	*TAATACGACTCACTATAGGG*
T7 Reverse	*GCTAGTTATTGCTCAGCGG*
DuetDOWN1	*GATTATGCGGCCGTGTACAA*
DuetUP2	TTGTACACGGCCGCATAATC
pETUpstream	ATGCGTCCGGCGTAGA
ARA F	CATTCTGTAACAAAGCGGGA
ARA R	CTGTTTTATCAGACCGCTTC
MBP F	GGTCGTCAGACTGTCGATGAAGCC

Table 44 - Primers used for Sanger DNA sequencing. Primers in italics were supplied by Source Bioscience, all other primers were supplied by Sigma Aldrich.

Peptide Synthesis

Solid Phase Peptide Synthesis

Peptides were synthesised using a LibertyBlue microwave peptide synthesiser (CEM). Synthesis proceeded from *C* to *N* terminus using *N*-terminal Fmoc protection and, where appropriate, side chain protection was achieved with acid-labile protecting groups (*e.g. tert*-butyloxycarbonyl (Boc) protection of lysine N^{ε}- or 2,3,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl (Pbf) protection of the arginine guanidino group).

Fmoc-protected proteinogenic amino acids were purchased from CS Bio, Novabiochem, Sigma-Aldrich, TCI, Alfa Aesar, Merck and AGTC Bioproducts and used to produce 0.2 M amino acid stock solutions in DMF (peptide synthesis grade, AGTC Bioproducts). *N,N'*-diisopropylcarbodiimide (TCI Europe) was used as coupling reagent in combination with either Hydroxybenzotriazole (AGTC Bioproducts) or Oxyma Pure (Merck). Fmoc deprotection was achieved using 20% piperidine in DMF (AGTC Bioproducts). Coupling and deprotection steps were microwave assisted and used the instrument's default methods. Peptides were synthesised on a rink amide-MBHA resin (100-200 mesh, 0.6-0.8 mmol g^{-1} loading, AGTC Bioproducts) on either a 100 µM or 50 µM scale.

The resin was washed with 25 mL dichloromethane and dried. Side chain deprotection and cleavage from the resin was conducted *via* treatment with 4 mL of 92.5% trifluoroacetic acid, 2.5% triisopropylsilane, 2.5% 1,3-dimethoxybenzene, 2.5% H_2O for 3 hours at 25°C. The resulting

mixture was filtered and the peptide was precipitated from the filtrate with cold Et_2O (40 mL). The peptides were pelleted *via* centrifugation and the supernatant was discarded. Peptides were subsequently resuspended in H_2O (small amounts of MeCN were added used to aid solubility in some cases) and lyophilised prior to purification.

Reversed Phase High Performance Liquid Chromatography

Peptide was dissolved in 2 mL H_2O, with MeCN added to aid solubility if required, before HPLC purification. A Grace Vydac 218TP 250 mm x 22 mm C18 column was fitted to either a Dionex 3000 Ultimate HPLC system or to a Jasco HPLC system (PU-2086 Plus Pumps, MX-2080-32 Dynamic Mixer and DG-2080-54 Degasser). Systems were controlled with Chromeleon (Dionex) or Jasco ChromNAC Control Center software.

The mobile phase was a gradient of MeCN + 0.1% TFA in H_2O + 0.1% TFA. Injection and fraction collection was conducted manually when the Dionex system was used; a Jasco AS-2055 Plus autosampler and CHF1225C fraction collector were used with the Jasco system. Following purification, fractions were analysed by MALDI MS or LC-MS; product containing fractions were pooled according to purity and lyophilised.

Recombinant Protein Production and Protein Purification

All chromatographic steps were conducted using one of the following FPLC systems:

- A GE Healthcare Äkta FPLC with a GE Healthcare Frac-920 fraction collector, controlled by Unicorn 5.1.
- A GE Healthcare Äkta Purifier with a GE Healthcare Frac-940 fraction collector, controlled by Unicorn 5.1.
- A GE Healthcare Äkta Pure with a GE Healthcare F9-C fraction collector controlled by Unicorn 6.4.1.

All columns, loops and column media were sourced from GE healthcare. UV monitoring was conducted at λ = 280 nm for protein purification and at λ = 220 nm for peptide purification.

Expression Trials

Plasmids were transformed into appropriate *E. coli* strains and an overnight culture was initiated following standard procedures. 1 mL of starter culture was then used to inoculate a 100 mL flask of growth medium. Cells were grown at 37°C with shaking in either an I26 or Innova 44 incubator shaker (New Brunswick Scientific) until the OD_{600} was between 0.6 and 0.8. At this point, 10 mL cells were added to a 50 mL Falcon tube, treated with inducing agent and then left shaking at the appropriate temperature for 16 hours.

Cells (1 mL) were then pelleted *via* centrifugation (20 minutes, 4000 x *g*) and were subsequently resuspended in BugBuster™ (Merck). The cells were lysed and insoluble fraction pelleted (10 minutes, 16000 x *g*) before being analysed by SDS-PAGE, dot blot or western blot.

For small scale Ni^{II} affinity purification, a 100 mL culture of cells was initiated from an overnight starter culture and, upon reaching the appropriate OD_{600}, induced. Cells were then left shaking at the appropriate temperature before cells were pelleted. Cells were resuspended in 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM imidazole (for C-P4H expression trials with pETDuet_PDI_HisCPH) or 50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, prior to sonication and purification by Ni^{II} affinity chromatography.

Large Scale Cell Growth and Protein Expression

The expression vector was transformed into an appropriate *E. coli* strain and a starter culture was initiated using standard procedures. 1% (v/v) of starter culture was used to inoculate a large growth in antibiotic-treated growth medium.

Cells were then cultured at 37°C with 180 rpm shaking in either an I26 or Innova 44 incubator shaker (New Brunswick Scientific) until the appropriate OD_{600} was achieved (Table 45). At this point the temperature was reduced and recombinant expression induced through the addition of freshly prepared, sterile filtered 1M IPTG (**17**) to reach the desired final concentration (Table

45). Cells were then cultured overnight at the post-induction temperature with 180 rpm shaking prior to harvesting *via* centrifugation (cycles of 11325 x *g*, 8 minutes) and storage at -80°C.

Protein	Growth Medium	Cell Type	Inducing Agent	[Inducing Agent] / mM	OD_{600} at Induction	Post-Induction Temp /°C
TEV Protease	2TY	*E. coli* BL21 (DE3)	IPTG (**17**)	0.5	0.6-0.8	18
KDM7B	TB	*E. coli* BL21 (DE3)	IPTG (**17**)	0.5	0.6-0.8	18
vCPH	2TY	*E. coli* Rosetta 2 (DE3) pLysS	IPTG (**17**)	0.5	0.8-1.1	18
C-P4H(I) C167S	2TY	*E. coli* Origami 2 (DE3)	IPTG (**17**)	0.5	0.6-0.8	30
C-P4H	TB	*E. coli* Origami 2 (DE3)	IPTG (**17**)	0.5	0.4	18

Table 45 - Induction conditions for large scale cell growth. Growth conditions for procedures that were not successful are indicated in red.Cell Lysis and Ni^{II} Affinity Chromatography

Cell pellets were resuspended in the appropriate binding buffer (Table 46). Phenylmethylsulfonyl fluoride (PMSF, 100 mM in EtOH, 1 mL per 100 mL binding buffer) and DNAseI (1 mg, Roche) were added prior to cell lysis. Lysis conditions for large scale preparations are outlined in Table 47.

Lysis by sonication was achieved using a Sonics VibraCell CV33 sonicator (Sonics & Materials Inc.) at 60% amplitude, smaller samples (*i.e.* a sample volume of less than 50 mL) were were lysed using a Sonics VibraCell CV334 sonicator (Sonics & Materials Inc.) fitted with a microtip and at an amplitude of 20%. Samples were kept on ice during sonication.

Lysis *via* liquid homogenisation was achieved using a pre-chilled Potter-Elvehjem homogeniser (glass mortar with a PTFE pestle, 20 strokes). Mechanical lysis was conducted by using a BeadBeater (BioSpec Products), with six 10 second cycles using 0.1 mm diameter glass beads; the entire apparatus was kept at 4°C during cell lysis and the apparatus was left for 30 seconds between cycles to cool.

Following cell lysis, the insoluble fraction was pelleted *via* centrifugation (58545 x *g*, 25 minutes); lysate was then filtered through a 0.45 µm syringe filter and loaded onto a pre-equilibrated HisTrap HP column; a 1 mL column was used for small scale preparations (less than 5 g cell pellet) and a 5 mL column was used for large scale purifications.

Protein	Buffer Type	Buffer Composition
TEV Protease	Binding	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM Imidazole
	Wash	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 40 mM Imidazole
	Elution	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 250 mM Imidazole
vCPH	Binding	50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM Imidazole
	Wash	50 mM HEPES pH 7.5, 500 mM NaCl, 40 mM Imidazole
	Elution	50 mM HEPES pH 7.5, 500 mM NaCl, 250 mM Imidazole
vCPH (*apo*)	Binding	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM Imidazole
	Wash	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 40 mM Imidazole
	Elution	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 250 mM Imidazole
KDM7B	Binding	50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM TCEP, 5% Glycerol
	Wash	50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM TCEP, 5% Glycerol, 20 mM Imidazole
	Elution	50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM TCEP, 5% Glycerol, 250 mM Imidazole
C-P4H (pETDuet_CPH_HisPDI)	Binding	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM Imidazole
	Wash	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 40 mM Imidazole
	Elution	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 250 mM Imidazole
C-P4H (pOPIN Vectors)	Binding	50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM Imidazole
	Wash	50 mM HEPES pH 7.5, 500 mM NaCl, 40 mM Imidazole
	Elution	50 mM HEPES pH 7.5, 500 mM NaCl, 250 mM Imidazole

Table 46 - Buffers used for Ni^{II} affinity chromatography. Methods that formed part of unsuccessful protocols are indicated in red. For C-P4H pOPIN vectors, salt concentrations were varied in one set of experiments between 0 and 1 M, this did not affect protein solutbility and no C-P4H tetramer was isolated.

Following sample application, the column was washed with 25 CV of wash buffer prior to elution with elution buffer. Fractions that contained eluted protein, as assessed by SDS-PAGE, were pooled and purified further using additional chromatographic steps.

Protein	Lysis Conditions
vCPH	Sonication, Cycles of 2 s on 3 s off
TEV Protease	Sonication, Cycles of 2 s on 3 s off
KDM7B	Sonication, Cycles of 2 s on 3 s off
C-P4H (pETDuet_CPH_HisPDI)	Sonication, Cycles of 2 s on 3 s off
C-P4H (pOPIN Vectors)	Sonication, Cycles of 2 s on 3 s off Homogenisation, 20 Strokes Mechanical lysis, 6 x 10 seconds

Table 47 - Lysis conditions used. Procedures that did not result in production of soluble protein are indicated in red.

Ion Exchange Chromatography

The sample was first buffer exchanged into a salt-free buffer (Table 48, Buffer A), most commonly using a PD10 column (GE Healthcare). The buffer exchanged sample was loaded onto a clean, pre-equilibrated ion exchange column using a 10 mL superloop (GE Healthcare). Following sample application, the column was washed with a minimum of 5 column volumes of buffer A, before the protein was eluted in a linear gradient of Buffer B (Table 48), to a maximum NaCl concentration of 1 M.

Protein	Ion Exchange Column	Buffer A (Loading Buffer)	Buffer B (Elution Buffer)
vCPH	Q Sepherose FF 25 mL	50 mM HEPES pH 7.5	50 mM HEPES pH 7.5, 1M NaCl
vCPH (*apo*)	Q Sepherose FF 25 mL	50 mM Tris pH 7.5	50 mM Tris pH 7.5, 1M NaCl
C-P4H (pETDuet_CPH_HisPDI)	MonoQ 50 mL	25 mM Tris-HCl pH 7.5	25 mM Tris-HCl pH 7.5, 1M NaCl
C-P4H (pETDuet_CPH_HisPDI)	HiTrap Heparin (5 mL x 2 in series)	25 mM Tris-HCl pH 7.5	25 mM Tris-HCl pH 7.5, 1M NaCl

Table 48 - Column media and buffers used for ion exchange chromatography. Procedures that did not result in production of soluble protein are indicated in red.

Product fractions, assessed by SDS-PAGE, were then pooled and carried forward to further purification steps or used for biochemical assays; vCPH was buffer exchanged into loading buffer *via* dialysis. For the production of *apo*-vCPH, the enzyme was treated with 200 mM EDTA

following ion exchange chromatography, prior to dialysis. Following buffer exchange, protein was concentrated and used directly in crystallographic studies, or was aliquoted and snap frozen in liquid N_2.

Size Exclusion Chromatography

Protein	Resin Used	Size Exclusion Buffer Composition
vCPH	Superdex 75	50 mM HEPES pH 7.5
vCPH (*apo*)	Superdex 75	50 mM Tris-HCl pH 7.5
TEV Protease	Superdex 75	25 mM Tris-HCl pH 7.5
KDM7B	Superdex 200	50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM TCEP, 5% Glycerol
C-P4H	Superdex 200	25 mM Tris-HCl pH 7.5
Peptide Samples	Superdex Peptide	25 mM NH_4HCO_3 pH 7.5

Table 49 - Buffer and resins used for size exclusion chromatography. Procedures that were part of unsuccessful protocols are indicated in red.

A Superdex 75 or Superdex 200 size exclusion column was first equilibrated with 1 column volume of size exclusion buffer. Depending on the scale of the protein purification, either a XK16/60 (Column volume = 120 mL) or a XK26/60 (column volume = 320 mL) column was used.

Following the previous chromatographic step, the sample was concentrated until the sample volume was less than 1% of the column volume. The sample was then injected into a clean 10 mL superloop (GE Healthcare) and applied to the column. Protein was eluted with 1.25 column volumes of size exclusion buffer. For proteins that required no further purification, fractions were collected and analysed by SDS-PAGE; product fractions were combined according to purity prior to being concentrated, aliquoted, snap frozen in liquid N_2 and stored at -80°C.

In the case of TEV protease, protein was concentrated to 8 mg mL^{-1} prior to dilution to 5 mg mL^{-1} in sterile 50% (v/v) glycerol.

Tag Cleavage with TEV Protease

Tagged protein that contained a TEV cleavage site was incubated overnight with hexahistidine-tagged TEV protease at 4°C (10:1 protein:TEV Protease) immediately following the first affinity chromatography step. Cleavage efficiency was assessed by whole protein mass

spectrometry or SDS-PAGE; following complete cleavage, the protein was buffer exchanged into Ni^{II} affinity binding buffer. The protein was then purified *via* Ni^{II} affinity chromatography, with the flow through and wash fractions being collected and retained. These fractions assed by SDS-PAGE, product fractions were pooled and then subjected to further chromatographic steps.

In Vitro Biochemical Assays

Peptide Hydroxylation Assays

Viral Collagen Prolyl Hydroxylase

Stocks of 2OG (disodium salt, 500 mM) and sodium-L-ascorbate (500 mM) were prepared in 50 mM HEPES pH 7.5. 100 mM Fe^{II} (as $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$) was prepared in 20 mM HCl. 10 x stocks were then produced by diluting the 2OG and ascorbate stocks in 50 mM HEPES pH 7.5, and the 10 x Fe^{II} stock was prepared by diluting the 100 mM solution in ddH_2O. In addition, a 10 x enzyme stock in 50 mM HEPES pH 7.5 and 10 x peptide substrate in ddH_2O solution was prepared.

Activity Assays

100 µM of peptide was incubated with 500 µM 2OG (disodium salt), 1 mM sodium-L-ascorbate, 50 µM $(NH_4)_2Fe(SO_4)_2$ and 1 µM vCPH. The reaction buffer was 50 mM HEPES pH 7.5. Reaction mixtures were prepared by combining the 2OG and enzyme, before adding a mixture of peptide, sodium-L-ascorbate and $(NH_4)_2Fe(SO_4)_2$. Reactions were conducted in triplicate alongside a no-enzyme control and incubated at 37°C for 16 hours. Reactions were quenched by adding an equal volume of MeOH and samples were analysed by MALDI-TOF MS.

K_M Determination

The concentration of only one component was varied whilst the other concentrations were maintained at those detailed in Table 50. Reactions were conducted at 25°C on a 20 µL scale in a 96 well plate (Greiner) in triplicate and against a no enzyme control. Reactions were initiated by combining a mix of 2OG (disodium salt), sodium-L-ascorbate and peptide with enzyme and $(NH_4)_2Fe(SO_4)_2$. Reactions were halted after 3 minutes *via* the addition of 20 µL MeOH prior to

analysis by MALDI-TOF MS. Percent hydroxylation was calculated by comparing the ion intensities of peaks corresponding to $[M+H]^+$ and $[M+Na]^+$ to $[M+O+H]^+$ and $[M+O+Na]^+$. Initial rates of product formation were calculated and plotted using GraphPad Prism, indicating the mean and standard error, before fitting a Michaelis-Menten curve.

Component	Final Concentration / μM
$(NH_4)_2Fe(SO_4)_2$	50
vCPH	1
Peptide	25
Sodium-L-Ascorbate	1000
2OG (Disodium Salt)	500

Table 50 - Final concentration of reaction components for vCPH K_M determination.

Inhibition Assays

Compounds **10**, **15** and **19** were synthesised by Dr Cyrille Thinnes[158,194], compounds **34**-**43** were synthesised by Mr James Holt-Martyn, compound **11** was synthesised by Dr Onur Atasoylu[194] and compound **9** was synthesised by Mr Dong Zhang[194]. Compound **18** was provided by Dr Nicolai Burzlaff[184]. Compound **8** was purchased from Selleck Chemicals.

Component	Final Concentration / μM
$(NH_4)_2Fe(SO_4)_2$	50
vCPH	1
Peptide	10
Sodium-L-Ascorbate	100
2OG (Disodium Salt)	50
Inhibitor	100 for broad screen, varied for IC_{50} determination

Table 51 - Reaction composition of vCPH inhibition assay. Reactions were conducted in 50 mM HEPES pH 7.5.

Final reaction concentrations are indicated in Table 51. Inhibitor and enzyme were incubated for 5 minutes prior to the addition of a solution of 2OG, Ascorbate, Fe^{II} and peptide. Reactions were conducted in triplicate at room temperature and quenched after 5 minutes *via* the addition of 20 μL MeOH. Reactions were analysed by MALDI-TOF MS; percent hydroxylation was calculated by comparing the ion intensities of peaks corresponding to $[M+H]^+$ and $[M+Na]^+$ to $[M+O+H]^+$ and $[M+O+Na]^+$. Percentage response was then calculated by comparing the reaction mixture to a no inhibitor control. Data was fitted using GraphPad Prism.

Preparation of Peptides Samples for MS-MS or Amino Acid Analysis

Component	Final Concentration / µM
$(NH_4)_2Fe(SO_4)_2$	50
vCPH	1
Peptide	100
Sodium-L-Ascorbate	1000*
2OG(Disodium Salt)	2000
HEPES pH 7.5	50000

Table 52 - Assay conditions for large scale enzyme activity assays. Peptide was subsequently purified and analysed by MS-MS or hydrolysed prior to amino acid analysis. *Ascorbate was excluded where appropriate.

Enzymatic activity assays were set up in a 1.5 mL microcentrifuge tube on a 100 µL scale, according to the composition detailed in Table 52. The reaction was conducted at 25°C for 16 hours and then analysed by MALDI-TOF MS to ensure the reaction had gone to completion. The reaction was not quenched and the peptide was then purified from enzyme and reaction components using size exclusion chromatography. The resultant purified peptide was lyophilised prior to resuspension in ddH_2O (50 µL). The peptide sample was then analysed by MS-MS or amino acid analysis.

Histone Demethylase Assays

KDM4 and KDM7B Demethylation Assays

Component	Final Concentration / µM
$(NH_4)_2Fe(SO_4)_2$	10
Enzyme	2
Peptide	10
Sodium-L-Ascorbate	100
2OG (Disodium Salt)	50

Table 53 – Final concentrations of reaction components for KDM4/KDM7B assays.

Reaction mixtures (final volume 20 µL) were prepared according to Table 53 in either 50 mM HEPES pH 7.5 (for samples with KDM4A, KDM4B, KDM4C, KDM4D and KDM4E) or 50 mM HEPES 500 mM NaCl pH 7.5 (for samples with KDM7B). Reactions were quenched by addition of 20 µL MeOH prior to analysis *via* MALDI-TOF MS.

KDM3A Demethylation Assays

Component	Final Concentration / µM
$(NH_4)_2Fe(SO_4)_2$	50
KDM3A	2
Peptide	10
Sodium-L-Ascorbate	100
2OG (Disodium Salt)	100

Table 54 – Final concentrations of reaction components for KDM3A assays.

Assays were conducted by Dr Rachel Shamo Schiller. Reaction mixtures were prepared according to Table 54 in 50 mM HEPES pH 7.5. Reactions were quenched after 1 hour through addition of 1% formic acid prior to analysis *via* MALDI-TOF MS.

KDM4 Kinetic Analyses

Kinetic parameters K_M and k_{cat} were obtained *via* monitoring of formaldehyde release, using a FDH/NAD coupled assay[281]. Peptide K_M and k_{cat} were determined by conducting the assay (with saturating cofactors and co-substrates) with a range of peptide concentrations, along with the cofactors and co-substrates detailed in Table 55, in 50 mM HEPES pH 7.5 with 0.01% Tween 20. Reactions were conducted on a 20 µL scale in separate wells of a clear bottomed 384 well plate; fluorescence was monitored during the course of the reaction using a PHERAstar FS (BMG Labtech) plate reader (355 nm excitation and 460 nm emission).

Component	Final Concentration
$(NH_4)_2Fe(SO_4)_2$	10 µM
Enzyme	2 µM
Sodium-L-Ascorbate	100 µM
2OG (Disodium Salt)	50 µM
FDH	0.025 U
NAD	250 mM

Table 55 – Final concentrations of reaction components for KDM4/KDM7B FDH-coupled assays.

Initial rates of formaldehyde production were calculated by comparing to a standard curve; values for K_M and k_{cat} were calculated by fitting to these initial rates to the Michaelis-Menten model, using the GraphPad Prism software suite.

Mass Spectrometry

Amino Acid Analysis

Acid Hydrolysis of Samples

Collagen from rat tail tendon was supplied as a serum in 20 mM AcOH and was freeze dried prior to preparation for amino acid analysis. Hydroxylysine and dihydroxyproline amino acid standards were obtained in previous studies[113,116,134,237].

Acid hydrolysis was conducted by placing a small solid sample in a pre-weighed sample tube. The sample tube was then placed in a reaction vial and a few drops of 6M HCl added to the vial. The vessel was heated to 110°C and held at 110°C for 16 hours under a N_2 atmosphere. The acid was removed and the vessel was heated to 110°C under a constant stream of N_2 for 12 hours in order to drive off any residual HCl. The samples were then dried further in a dessicator for a minimum of 24 hours.

UPLC-MS Analysis of Hydrolysed Samples

Time / min	A / %	B / %
0.00	99.90	0.10
0.54	99.90	0.10
5.74	90.90	9.10
7.74	78.80	21.20
8.04	40.40	59.60
8.05	10.00	90.00
8.64	10.00	90.00
8.73	99.90	0.10
9.50	99.90	0.10

Table 56 - UPLC gradient used in 10 minute AAA.

Hydrolysed samples were dissolved in 40 µL of pH 9 borate buffer (Waters). The pH of the resuspended samples was checked and adjusted with NaOH until the pH was greater than 8.5. 10 µL of 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate (**54**, Waters) in MeCN was added to the resuspended sample. Analysis was carried out using a Waters Acquity UPLC system coupled to a Waters Xevo G2-S quadruple-time-of-flight mass spectrometer, using an electrospray

ionisation (ESI) source. Instrument control and data processing was conducted using MassLynx 4.5 (Waters).

A Waters AccQ-Tag Ultra column (1.7 µm, 100 x 2.1 mm) was kept at 55°C; the solvent flow rate was increased in increments of 0.1 mL min^{-1} to the final flow rate of 0.7 mL min^{-1}. UV absorbance was monitored at a wavelength of 260 nm. Phase A was produced by diluting 5% of AccQ-Tag Ultra Eluent A concentrate (Waters) in ddH_2O, phase B consisted of 1.3% formic acid in acetonitrile.

Time / min	A / %	B / %
0.00	99.90	0.10
25.00	90.00	10.00
25.50	40.00	60.00
27.50	40.00	60.00
28.00	100.00	0.00
30.00	100.00	0.00

Table 57 - UPLC gradient used in 30 minute AAA.

1 µL of sample was injected into a 2 µL injection loop; solvent gradients used were the same as those used in previous analyses[113,116] and are detailed in Table 56 and Table 57.

MS analysis was carried out using positive ESI with a source temperature of 100°C and a desolvation temperature of 250°C. The capillary voltage was held at a potential difference of 3000 V and the sample cone was held at 100 V. Gas flows were set to 3 L $hour^{-1}$ for the cone and 1000 L $hour^{-1}$ for desolvation. The mass analyser was used in resolution mode.

Liquid Chromatography-Mass Spectrometry (LC-MS)

Time / min	A / %	B / %
0.00	98.00	2.00
0.50	98.00	2.00
7.50	25.00	75.00
8.50	0.00	100.00
9.00	0.00	100.00
9.25	98.00	2.00
10.00	98.00	2.00

Table 58 – Mobile phase conditions for standard small molecule LC-MS analysis. Phase A was ddH_2O with 0.01% TFA, Phase B was MeCN with 0.01% TFA.

Routine analysis of peptides and small molecules was achieved using an Agilent 1200 Series HPLC system coupled to an Agilent 6120 quadrupole mass spectrometer. Small molecule analyses were conducted using a Merck Chromolith Performance RP-18e 100 mm x 4.6 mm C18 column; mobile phase passed through the column at a flow rate of 1.50 mL min^{-1} (mobile phase composition is detailed in Table 58).

Time / min	A / %	B / %
0.00	98.00	2.00
0.50	98.00	2.00
5.00	50.00	50.00
7.50	15.00	85.00
8.50	0.00	100.00
9.00	0.00	100.00
9.25	98.00	2.00
10.00	98.00	2.00

Table 59 - Mobile phase for standard peptide LC-MS analysis. Phase A was ddH2O with 0.01% TFA, Phase B was MeCN with 0.01% TFA.

Peptides were analysed using a Grace Vydac 218TP 50 mm x 4.6 mm C18 column and mobile phase was run at a flow rate of 1.50 mL min^{-1}; the mobile phase conditions are detailed in Table 59. System control and data analysis was conducted using Agilent ChemStation.

Matrix Assisted Laser-Desorption/Ionisation (MALDI) Mass Spectrometry

MALDI matrix was prepared by dissolving α-cyano-4-hydroxycinnamic acid (CHCA) in 50% H_2O, 50% MeCN with 0.1% TFA at a concentration of 5 mg mL^{-1}. 1 µL of sample was spotted with 1 µL matrix on a MALDI target plate (Waters). Spots were left to dry before being analysed using a Micromass MALDI MX Micro (Waters) MALDI-TOF mass spectrometer, using the positive ion ionisation mode and the reflectron analysis mode. Instrument control and data analysis was conducted using Masslynx 4.0 (Waters).

Whole Protein Mass Spectrometry

Purified protein (0.01 mg mL^{-1} in ddH_2O) was analysed using a LCT Premier mass spectrometer (Waters) coupled to a NanoAcquity-UPLC system (Waters), equipped with a ProSwift™ RP-4H

1 mm x 50 mm column (ThermoFisher Scientific). Samples were loaded onto the column in 5% 5 mM ammonium formate pH 7.0, 5% MeCN and eluted in a gradient to 95% MeCN, 5% 5 mM ammonium formate pH 7.0 and was introduced directly into the electrospray ion source. Data was analysed in Masslynx 4.0 (Waters), with deconvolution using the MaxEnt1 algorithm.

Mass Spectrometry-Mass Spectrometry

Excision and Tryptic Digest of SDS-PAGE Bands

Bands of interest on a SDS-PAGE gel were exised and stored in a 1.5 mL microcentrifuge tube. The band was destained by shaking for 16 hours in 200 µL of wash solution (50% (v/v) MeOH, 45% (v/v) H_2O, 5% (v/v) AcOH) before being washed for a further 2 hours in fresh wash solution. The solution was discarded and the gel pieces dehydrated by treating with 200 µL MeCN for 5 minutes.

The MeCN was removed and 30 µL 10 mM DTT was added to the gel piece. The solution was removed after 30 minutes and 30 µL of 50 mM iodoacetamide was added to the gel pieces. After 30 minutes the gel pieces were dehydrated by treating with 200 µL MeCN for 5 minutes. Gel pieces were rehydrated for 10 minutes in 200 µL 100 mM NH_4HCO_3 prior to dehydration (200 µL MeCN for 5 minutes). Gel pieces were rehydrated for exactly 10 minutes in 30 µL of 20 ng μL^{-1} Trypsin (Proteomics grade, Promega) in 50 mM NH_4HCO_3. Excess trypsin solution was removed and 5 µL of 50 mM NH_4HCO_3 was added prioir to digestion overnight at 37°C overnight.

Following digestion, peptides were extracted by sequential 10 minute incubations in 50 µL of 50 mM NH_4HCO_3; 50% (v/v) MeCN, 45% (v/v) H_2O, 5% (v/v) formic acid and 85% (v/v) MeCN, 10% (v/v) H_2O, 5% formic acid. Extraction buffers were combined and concentrated to dryness using an Eppendorf Concentrator plus vacuum concentrator. The resultant pellet was resuspended in 20 µL of 98% (v/v) H_2O, 2% (v/v) MeCN, 0.1% (v/v) formic acid. Digested peptides were taken forward for analysis.

UPLC-MS-MS Analysis

UPLC-MS-MS analysis was conducted by Miss Elisabete Pires.

Tryptically digested peptides were analysed using an NanoAcquity-UPLC system (Waters) connected to an Orbitrap Elite mass spectrometer (ThermoFischer Scientific) equipped with an EASY-Spray nano-electrospray ion source (ThermoFischer Scientific). Separation was achieved using an EASY-spray Acclaim PepMap™ analytical column (75 µm internal diameter × 15 mm, RSLC C18, 3 µm, 100 Å) with a linear gradient (3% to 60% in 48 minues) of solvent B (0.1% formic acid in acetonitrile) in solvent A (0.1% formic acid in ddH_2O) at a flow on 300 nL min^{-1}. Peptides were directly introduced into the mass spectrometer *via* electrospray. The mass spectrometer was operated in a data-dependent mode using a collision induced dissociation (CID)-based method. Full scan MS spectra (scan range 350-1500 *m/z*, resolution 30000, AGC target 10^6, maximum injection time 250 ms) and subsequent CID MS-MS spectra (resolution 15000, AGC target 3×10^4, maximum injection time 500 ms) of 5 most intense peaks were acquired in the Orbitrap. CID fragmentation was performed at 35% of normalized collision energy and the signal intensity threshold was kept at 500 counts. The raw MS file was searched against the Uniprot database.

For digested samples, the appropriate protease was entered during database searching; for peptide samples 'NO ENZYME' was selected as the protease and hydroxylation was set as variable modification. Precursor mass tolerance was set as 15 ppm. Fragment mass tolerances for CID were set to 0.8 Da respectively. All spectra were manually validated.

MALDI-TOF-TOF MS-MS

MALDI-TOF-TOF MS-MS analysis was conducted by Miss Elisabete Pires. Data processing was conducted in collaboration with Dr Joan Gannon.

MALDI matrix was prepared by dissolving α-Cyano-4-hydroxycinnamic acid (CHCA) in 50% H_2O, 50% MeCN with 0.1% TFA at a concentration of 5 mg mL^{-1}. 1 µL of sample was spotted with 1 µL

matrix onto a 384-well target plate (Applied Biosystems) and analysed using a Sciex 4800 MALDI-TOF-TOF mass spectrometer. Fragmentation of peptides was achieved *via* CID and the *m/*z values of resulting ions reported. Data were analysed by manual comparison with theoretical *m/z* values which were calculated using the MS-Product module of ProteinProspector v 5.20.0[293].

Nuclear Magnetic Resonance Spectroscopy

Routine Nuclear Magnetic Resonance (NMR) spectra were obtained using either a Bruker AVIII400 HD spectrometer or a Bruker AVANCE 400 spectrometer. Variable temperature NMR experiments were conducted on a Bruker DRX500 spectrometer.

NMR spectroscopy of enzyme samples was conducted with the help of Dr Richard Hopkinson, Dr Martine Abboud and Dr Christopher Lohans. NMR spectra were obtained using either:

- A Bruker AVIII 600 MHz NMR spectrometer equipped with a broad band BB-^{19}F/^{1}H Prodigy N_2 cryoprobe controlled by Topspin 3.2 software (Bruker).

  Or

- A Bruker AVIII 700 MHz NMR Spectrometer equipped with an inverse TCI cryoprobe optimised for ^{1}H observation, controlled by Topspin 3 software (Bruker).

Data were analysed using Topspin 3.1 software (Bruker).

vCPH

2OG Turnover Monitoring

Component	Final Concentration / µM
$(NH_4)_2Fe(SO_4)_2$	50
vCPH	10
Peptide	400
Sodium-L-Ascorbate	1000
2OG (Disodium Salt)	600

Table 60 - Reaction mixtures used for time course monitoring of 2OG turnover using a 700 MHz NMR spectrometer.

vCPH was buffer exchanged into 50 mM Tris-d_{11} pH 7.5, 10% D_2O, 0.02% NaN_3 and reaction mixtures (160 µL) were prepared in a 1.5 mL microcentrifuge tube according to the compositions detailed in Table 60 and Table 61. Reactions were initiated *via* the addition of 2OG (t = 0). After 60 s, the reaction mixture was transferred into 5 mm diameter NMR tubes (Norell) for analysis with the 600 MHz NMR spectrometer or 2 mm MATCH NMR tubes (Hilgenberg) for analysis using the 700 MHz NMR spectrometer.

Component	Final Concentration / µM
$(NH_4)_2Fe(SO_4)_2$	50
vCPH	20
Peptide	400
Sodium-L-Ascorbate	1000*
2OG (Disodium Salt)	600ᵗ

Table 61 - Reaction mixtures used for time course monitoring of 2OG turnover using a 600 MHz NMR spectrometer. *Ascorbate was ommitted from some samples to allow for analysis of the effect of ascorbate on reaction progression. ᵗFor Peptide 2, 1800 µM 2OG was used to take into account that two vCPH-catalysed hydroxylation reactions take place with this substrate.

^1H excitation scuplting suppression with perfect echo NMR spectra were typically obtained using 32 scans, a relaxation delay of 1 s, and a pre-scan delay of 10 µs. A 2 ms sinc pulse was used for water suppression. Data were processed with a line broadening of 0.3 Hz using TopSpin 3.2 software (Bruker). Spectra were obtained at 135 s intervals and integrated using absolute intensity scaling.

^{19}F NMR

^{19}F NMR experiments were typically obtained using 2000 scans and a recovery delay of 2 s. Data were processed with 1 Hz Lorentzian line broadening using TopSpin 3.2 software (Bruker). Assay mixtures contained 20 µM vCPH, 50 µM Fe^{II}, 600 µM 2OG, 1 mM ascorbate and 500 µM Peptide 27 and were buffered in 50 mM Tris-D11, pH 7.5, in 90% H_2O and 10% D_2O.

2OG Displacement

Nuclear Magnetic Resonance (NMR) spectra were recorded using a Bruker AVIII 600 MHz NMR spectrometer equipped with a 5-mm inverse cryoprobe using 3 mm MATCH NMR tubes (Cortectnet). 2OG displacement was monitored by Carr-Purcell-Meiboom-Gill (CPMG) NMR.

Typical experimental parameters for CPMG NMR spectroscopy were as follows: total echo time, 40 ms; relaxation delay, 2 s; and number of transients, 264. The PROJECT-CPMG sequence as described by Augilar *et al.* was applied (90°x–[τ–180°y–τ–90°y–τ–180°y–τ]*n*–acq)[294]. Water suppression was achieved by pre-saturation. Assay mixtures were buffered in 50 mM Tris-D_{11} (pH 7.5) and 0.02% NaN_3 in 90% H_2O and 10% D_2O.

KDM4E

Component	Final Concentration / µM
$(NH_4)_2Fe(SO_4)_2$	100
KDM4E	15.5
Peptide	1000
Sodium-L-Ascorbate	1000
2OG (Disodium Salt)	500

Table 62 - Assay conditions for NMR samples with KDM4E. Peptide was substituted for ddH_2O in control samples.

For the NMR experiments with KDM4E, samples (75 µL final volume) were prepared according to Table 62 in deuterated ammonium formate buffer[86], peptide was substituted with H_2O in control samples. Samples were prepared in 1.5 mL microcentrifuge tubes and transferred to 2 mm MATCH NMR tubes (Hilgenberg) and then analysed using ^1H NMR on a Bruker AVIII 700 MHz equipped with an inverse TCI cryoprobe optimised for ^1H observation and equipped with Topspin 3 software.

Data were acquired using automated routines; 20 experiments were collected consisting of 16 transients, corresponding to 89 seconds acquisition time per experiment. The solvent deuterium signal was used as an internal lock and the solvent water resonance was suppressed by presaturation. ^1H chemical shifts are reported relative to the solvent resonance (δ_H 4.7 ppm).

In the sample with peptide, partial overlap of the succinate and glutamine β-proton resonances of the peptide was observed; the succinate concentration was calculated by determining the concentration of the overlapped peak (as if entirely corresponding to succinate), and then

subtracting the apparent succinate concentration of the same peak in a control sample without enzyme (thus containing peptide but no succinate).

Differential Scanning Fluorimetry

Differential scanning fluorimetry was conducted by adapting previously reported methods[295]. T_m values for proteins were assessed using a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad). Reactions were conducted on a 50 µL scale in a 96 well white PCR plate (StarLabs).

A 2 x stock of SyproOrange was produced *via* the addition of ddH_2O. This was added to a 2 x mixture of *apo* enzyme and any relevant additive (*e.g.* metal ions or inhibitor) in 50 mM HEPES pH 7.5. Temperature was increased in steps of 0.2°C from 20°C to 95°C; a reading was taken at each temperature. The system was monitored using an excitation at λ = 492 nm and emission at λ = 610 nm.

Component	Final concentration
***Apo*-vCPH**	5 µM
SyproOrange	1 in 2500 from commercial stock
HEPES pH 7.5	25 mM

Table 63 - Final concentration of reaction components.

Concentrations of *apo*-protein and SyproOrange were first varied in order to optimise the obtained signal. These data were then used to inform final protein and dye concentrations, which are detailed in Table 63.

To assess the effect of inhibitor binding on vCPH T_m, 50 µM Mn^{II} and 100 µM inhibitor were added to the reaction mixture. For metal binding experiments, 100 µM of the appropriate metal was added to the mixture.

Crystallography

Broad screens were commercially sourced from Molecular Dimensions, Hampton Research and Rigaku. Optimisation screen compositions are detailed in Appendix 3. Broad screens and optimisation screens were prepared and stored in sealed 96 well blocks at 4°C prior to use.

Crystallisation Trials

Protein and Final Concentration	Metal	Compound/Peptide Added	Screens Used (Temperatures)
vCPH, 22 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	NOG (**4**, 10 mM), Peptide 4 (10 mM)	PEGIon (25°C and 4°C) JCSG+ (25°C and 4°C) Salt Rx (25°C and 4°C) Xtal Screen (25°C and 4°C) PACT Premier (25°C and 4°C)
vCPH, 20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	FG-4592 (**8**, 10 mM)	JCSG+ (25°C and 4°C) PACT Premier (25°C and 4°C)
vCPH, 20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	2,4-PDCA (**5**, 10 mM)	JCSG+ (25°C and 4°C) PACT Premier (25°C and 4°C) PEGIon (4°C)
vCPH, 20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	NOG (**4**, 10 mM), Peptide 4 (10 mM)	PACT Premier (25°C and 4°C)
vCPH, 20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	2,4-PDCA (5, 10 mM)	PACT Premier D11 Optimisation Screen (25°C and 4°C)
vCPH, 20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	NOG (**4**, 10 mM)	PACT Premier D11 Optimisation Screen (25°C and 4°C)
vCPH, 20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	FG-4592 (**8**, 10 mM)	PACT Premier D11 Optimisation Screen (25°C and 4°C)
vCPH, 20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	NOG (**4**, 10 mM), Peptide 4 (10 mM)	PACT Premier D11 Optimisation Screen (25°C and 4°C)
vCPH, 20 mg mL^{-1}	Mn^{II} (as $MnCl_2$, 4 mM)	BIOX2 (**18**, 1mM)	JCSG+ (25°C and 4°C)

Table 64 - Details of crystallisation trials.

PEGIon, Salt Rx and Xtal Screen solutions were purchased from Hampton Research. JCSG+ and PACT Premier solutions were purchased from Molecular Dimensions. Screens were transferred to 96 well polycarbonate blocks and stored at 4°C.

Purified protein (purity greater that 95% as assessed by SDS-PAGE) was concentrated and combined with metal, compounds and peptides (as appropriate) in a total volume of 60 µL and incubated for a minimum of 15 minutes; samples were kept on ice throughout. Samples were

then centrifuged for 2 minutes at 16000 x *g* to pellet any precipitate. The sample was then transferred to a fresh tube and used to set up crystallisation trials.

Crystallisation trials were set up in low profile 96 well Intelli-Plates (Art Robbins Instruments) using a Phoenix/RE instrument (Art Robbins Instruments) equipped with a 96 syringe head and a single nano dispenser. Plate reservoirs were filled with 80 µL of crystallisation condition solution. Three sitting drop experiments per condition were conducted with well solution:protein solution ratios of 0.1 µL:0.2 µL, 0.1 µL:0.1 µL, and 0.2 µL:0.1 µL. Plates were sealed with Starseal advanced polyolefin film (Starlab). Plates were subsequently stored at either 25°C or 4°C and monitored at regular intervals.

Crystal Harvesting

Crystals were soaked in a cryoprotection solution (20% glycerol in well solution) and fished using pin-mounted nylon loops (Hampton Research) before being immediately cryo-cooled in liquid N_2. Cryo-cooled crystals were stored in cryovials (Hampton Research) and kept in liquid N_2 until required.

X-ray Diffraction

Crystals were first screened by analysing diffraction in-house. Crystals were analysed at 100K using a Rigaku FR/E+ Superbright copper rotating anode diffractometer equipped with Osmic HF optics and a Saturn 944+ charge-coupled device detector. Diffracting protein crystals were saved and stored in liquid N_2. Protein crystals, along with any other crystals from the same well were then analysed using synchrotron radiation. Single crystals were analysed at 100K at Diamond Light Source beamline I02, I04 or I04-1; all beamlines were equipped with a Pilatus 6M-F detector.

Isothermal Titration Calorimetry

Isothermal titration calorimetry was conducted using a Malvern MicroCal PEAQ-ITC instrument with automated liquid handling. The cell was kept at 25°C and contained 40 µM enzyme; 19

injections of 400 μM binding partner (*e.g.* metal ion, inhibitor compound, peptide) were added following standard protocols. All components were dissolved in 50 mM HEPES pH 7.5. For compound and peptide binding, the buffer in both cell and syringe was supplemented with 200 μM Ni^{II}. The system was controlled using Malvern MicroCal PEAQ-ITC automation software and data analysis and baseline correction was conducted using the Malvern MicroCal PEAQ-ITC analysis suite.

Organic Synthesis

Generalities

All reactions were carried out in round bottomed flasks and were stirred magnetic stirrers. Thin layer chromatography (TLC) was conducted using aluminium plates pre-coated with Kieselgel 60 F254 (Merck). Samples were visualised by UV fluorescence at 254 nm or by staining with $KMnO_4$ stain. Column chromatography was conducted using a Biotage SP4 automated flash chromatography system and appropriate Biotage SNAP silica cartridges.

Detailed procedures and compound characterisation of the compounds synthesised as part of the study in Chapter 6 are presented in Appendix 2.

Cross Metathesis – General Procedure

CH_2Cl_2 (50 mL) was degassed *via* three freeze-thaw cycles and used to dissolve *N*-Fmoc allyl glycine, **55**, (1 equiv., Alfa Aesar) and the appropriate cross metathesis partner (1.5 equiv.). Hoveyda-Grubbs second generation catalyst (15 mol %) was added to the mixture before the reaction was brought to reflux, under a N_2 atmosphere, for 6 hours. The reaction was monitored by TLC (2:1 EtOAc:cyclohexane with 0.2% AcOH) and by LC-MS.

Olefin Hydrogenation – General Procedure

The metathesis product was dissolved in EtOH. Palladium on carbon (10% by mass) was separately suspended in EtOH before being added to the reaction mixture. The reaction vessel was flushed with N_2 before being flushed with H_2. The reaction mixture was then left stirring

under a H_2 atmosphere at room temperature for 24 hours. Upon completion (assessed by LC-MS), the reaction mixture was filtered through a Celite™ plug under a stream of N_2. The filtrate was concentrated to dryness *in vacuo*.

Infra-Red Spectroscopy

Infra-red spectra were recorded using a Bruker Tensor 27 instrument from the solid state.

Melting Point Determination

Melting points were obtained using a Stuart SMP40 automated melting point apparatus.

Optical Rotation

$[\alpha]_D$ values were calculated from optical rotation measurements using a Perkin Elmer 431 series polarimeter.

Chapter 9
References

1 Engel, J. & Chiquet, M. in *The Extracellular Matrix: an Overview* (ed Mecham, R. P.) 1-39 (Springer Berlin Heidelberg, 2011).
2 Frantz, C., Stewart, K. M. & Weaver, V. M. The Extracellular Matrix at a Glance. *Journal of Cell Science* **123**, 4195-4200, (2010).
3 Kannus, P. Structure of the Tendon Connective Tissue. *Scandinavian Journal of Medicine & Science in Sports* **10**, 312-320, (2000).
4 Terajima, M., Perdivara, I., Sricholpech, M., Deguchi, Y., Pleshko, N. *et al.* Glycosylation and Cross-linking in Bone Type I Collagen. *Journal of Biological Chemistry* **289**, 22636-22647, (2014).
5 M, M. P. Basement Membrane Proteins: Structure, Assembly, and Cellular Interactions. *Critical Reviews in Biochemistry and Molecular Biology* **27**, 93-127, (1992).
6 Kalluri, R. Basement Membranes: Structure, Assembly and Role in Tumour Angiogenesis. *Nature Reviews Cancer* **3**, 422-433, (2003).
7 Józsa, L. G. & Kannus, P. *Human Tendons : Anatomy, Physiology, and Pathology*. (Human Kinetics, 1997).
8 Kielty, C. M. & Grant, M. E. in *Connective Tissue and Its Heritable Disorders* 159-221 (John Wiley & Sons, Inc., 2003).
9 Canty, E. G. & Kadler, K. E. Procollagen Trafficking, Processing and Fibrillogenesis. *Journal of Cell Science* **118**, 1341-1353, (2005).
10 Yurchenco, P. D., Smirnov, S. & Mathus, T. in *Methods in Cell Biology* Vol. 69 111-144 (Academic Press, 2002).
11 Cox, T. R. & Erler, J. T. Remodeling and Homeostasis of the Extracellular Matrix: Implications for Fibrotic Diseases and Cancer. *Disease Models & Mechanisms* **4**, 165-178, (2011).
12 Verzijl, N., DeGroot, J., Oldehinkel, E., Bank, R. A., Thorpe, S. R. *et al.* Age-Related Accumulation of Maillard Reaction Products in Human Articular Cartilage Collagen. *Biochemical Journal* **350**, 381-387, (2000).
13 Riley, G. P., Harrall, R. L., Constant, C. R., Chard, M. D., Cawston, T. E. *et al.* Tendon Degeneration and Chronic Shoulder Pain: Changes in the Collagen Composition of the Human Rotator Cuff Tendons in Rotator Cuff Tendinitis. *Annals of the Rheumatic Diseases* **53**, 359-366, (1994).
14 Dakin, S. G., Smith, R. K. W., Heinegård, D., Önnerfjord, P., Khabut, A. *et al.* Proteomic Analysis of Tendon Extracellular Matrix Reveals Disease Stage-Specific Fragmentation and Differential Cleavage of COMP (Cartilage Oligomeric Matrix Protein). *Journal of Biological Chemistry* **289**, 4919-4927, (2014).
15 Meltzer, E. B. & Noble, P. W. Idiopathic Pulmonary Fibrosis. *Orphanet Journal of Rare Diseases* **3**, 8, (2008).
16 Franklin, T. J. Therapeutic Approaches to Organ Fibrosis. *The International Journal of Biochemistry & Cell Biology* **29**, 79-89, (1997).
17 Rieder, F. & Fiocchi, C. Intestinal Fibrosis in IBD - A Dynamic, Multifactorial Process. *Nature Reviews Gastroenterology and Hepatology* **6**, 228-235, (2009).
18 Gordon, I. O., Agrawal, N., Goldblum, J. R., Fiocchi, C. & Rieder, F. Fibrosis in Ulcerative Colitis: Mechanisms, Features, and Consequences of a Neglected Problem. *Inflammatory Bowel Diseases* **20**, 2198-2206, (2014).
19 Mehlen, P. & Puisieux, A. Metastasis: A Question of Life or Death. *Nature Reviews Cancer* **6**, 449-458, (2006).
20 Ponta, H., Sherman, L. & Herrlich, P. A. CD44: From Adhesion Molecules to Signalling Regulators. *Nature Reviews Molecular Cell Biology* **4**, 33-45, (2003).
21 Franchi, M., Trirè, A., Quaranta, M., Orsini, E. & Ottani, V. Collagen Structure of Tendon Relates to Function. *The Scientific World Journal* **7**, 404-420, (2007).

22 Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D. *et al.* in *Molecular Cell Biology* Ch. 22, (W. H. Freeman, 2000).
23 Ricard-Blum, S. The Collagen Family. *Cold Spring Harbor Perspectives in Biology* **3**, (2011).
24 Myllyharju, J. & Kivirikko, K. I. Collagens, Modifying Enzymes and their Mutations in Humans, Flies and Worms. *Trends in Genetics* **20**, 33-43, (2004).
25 Myllyharju, J. Prolyl 4-Hydroxylases, the Key Enzymes of Collagen Biosynthesis. *Matrix Biology* **22**, 15-24, (2003).
26 Ramshaw, J. A. M., Shah, N. K. & Brodsky, B. Gly-X-Y Tripeptide Frequencies in Collagen: A Context for Host–Guest Triple-Helical Peptides. *Journal of Structural Biology* **122**, 86-91, (1998).
27 Shoulders, M. D. & Raines, R. T. Collagen Structure and Stability. *Annual Review of Biochemistry* **78**, 929-958, (2009).
28 Gustavson, K. H. The Function of Hydroxyproline in Collagens. *Nature* **175**, 70-74, (1955).
29 Berg, R. A. & Prockop, D. J. The Thermal Transition of a Non-Hydroxylated Form of Collagen. Evidence for a Role for Hydroxyproline in Stabilizing the Triple-Helix of Collagen. *Biochemical and Biophysical Research Communications* **52**, 115-120, (1973).
30 Bella, J., Eaton, M., Brodsky, B. & Berman, H. M. Crystal and Molecular Structure of a Collagen-Like Peptide at 1.9 Å Resolution. *Science* **266**, 75-81, (1994).
31 Cowan, P. M., McGavin, S. & North, A. C. T. The Polypeptide Chain Configuration of Collagen. *Nature* **176**, 1062-1064, (1955).
32 Rich, A. & Crick, F. H. C. The Structure of Collagen. *Nature* **176**, 915-916, (1955).
33 Ramachandran, G. N. & Kartha, G. Structure of Collagen. *Nature* **176**, 593-595, (1955).
34 Kramer, R. Z., Bella, J., Mayville, P., Brodsky, B. & Berman, H. M. Sequence Dependent Conformational Variations of Collagen Triple-Helical Structure. *Nature Structural & Molecular Biology* **6**, 454-457, (1999).
35 Exposito, J.-Y., Valcourt, U., Cluzel, C. & Lethias, C. The Fibrillar Collagen Family. *International Journal of Molecular Sciences* **11**, 407-426, (2010).
36 Fidler, A. L., Vanacore, R. M., Chetyrkin, S. V., Pedchenko, V. K., Bhave, G. *et al.* A Unique Covalent Bond in Basement Membrane is a Primordial Innovation for Tissue Evolution. *Proceedings of the National Academy of Sciences* **111**, 331-336, (2014).
37 Robertson, W. E., Rose, K. L., Hudson, B. G. & Vanacore, R. M. Supramolecular Organization of the α121-α565 Collagen IV Network. *Journal of Biological Chemistry* **289**, 25601-25610, (2014).
38 Heikkinen, A., Tu, H. & Pihlajaniemi, T. Collagen XIII: A Type II Transmembrane Protein with Relevance to Musculoskeletal Tissues, Microvessels and Inflammation. *The International Journal of Biochemistry & Cell Biology* **44**, 714-717, (2012).
39 Veit, G., Zwolanek, D., Eckes, B., Niland, S., Käpylä, J. *et al.* Collagen XXIII, Novel Ligand for Integrin $\alpha_2\beta_1$ in the Epidermis. *Journal of Biological Chemistry* **286**, 27804-27813, (2011).
40 Snellman, A., Keränen, M.-R., Hägg, P. O., Lamberg, A., Hiltunen, J. K. *et al.* Type XIII Collagen Forms Homotrimers with Three Triple Helical Collagenous Domains and Its Association into Disulfide-bonded Trimers Is Enhanced by Prolyl 4-Hydroxylase. *Journal of Biological Chemistry* **275**, 8936-8944, (2000).
41 Tu, H., Huhtala, P., Lee, H.-M., Adams, J. C. & Pihlajaniemi, T. Membrane-Associated Collagens With Interrupted Triple-Helices (MACITs): Evolution from a Bilaterian Common Ancestor and Functional Conservation in *C. elegans*. *BMC Evolutionary Biology* **15**, 281, (2015).
42 Hägg, P., Rehn, M., Huhtala, P., Väisänen, T., Tamminen, M. *et al.* Type XIII Collagen Is Identified as a Plasma Membrane Protein. *Journal of Biological Chemistry* **273**, 15590-15597, (1998).

43 Tanaka, T., Wakabayashi, T., Oizumi, H., Nishio, S., Sato, T. *et al.* CLAC-P/Collagen Type XXV is Required for the Intramuscular Innervation of Motoneurons During Neuromuscular Development. *The Journal of Neuroscience* **34**, 1370-1379, (2014).
44 Kassner, A., Tiedemann, K., Notbohm, H., Ludwig, T., Mörgelin, M. *et al.* Molecular Structure and Interaction of Recombinant Human Type XVI Collagen. *Journal of Molecular Biology* **339**, 835-853, (2004).
45 Clementz, A. G. & Harris, A. Collagen XV: Exploring Its Structure and Role within the Tumor Microenvironment. *Molecular Cancer Research* **11**, 1481-1486, (2013).
46 Ackley, B. D., Crew, J. R., Elamaa, H., Pihlajaniemi, T., Kuo, C. J. *et al.* The NC1/Endostatin Domain of *Caenorhabditis elegans* Type XVIII Collagen Affects Cell Migration and Axon Guidance. *The Journal of Cell Biology* **152**, 1219-1232, (2001).
47 Rehn, M. & Pihlajaniemi, T. Type XV and XVIII Collagens, A New Subgroup Within the Family of Collagens. *Seminars in Cell & Developmental Biology* **7**, 673-679, (1996).
48 Seppinen, L. & Pihlajaniemi, T. The Multiple Functions of Collagen XVIII in Development and Disease. *Matrix Biology* **30**, 83-92, (2011).
49 Grimal, S., Puech, S., Wagener, R., Ventéo, S., Carroll, P. *et al.* Collagen XXVIII is a Distinctive Component of the Peripheral Nervous System Nodes of Ranvier and Surrounds Nonmyelinating Glial Cells. *Glia* **58**, 1977-1987, (2010).
50 Veit, G., Kobbe, B., Keene, D. R., Paulsson, M., Koch, M. *et al.* Collagen XXVIII, a Novel von Willebrand Factor A Domain-containing Protein with Many Imperfections in the Collagenous Domain. *Journal of Biological Chemistry* **281**, 3494-3504, (2006).
51 Smith-Mungo, L. I. & Kagan, H. M. Lysyl Oxidase: Properties, Regulation and Multiple Functions in Biology. *Matrix Biology* **16**, 387-398, (1998).
52 Anastasia, L., Rota, P., Anastasia, M. & Allevi, P. Chemical Structure, Biosynthesis and Synthesis of Free and Glycosylated Pyridinolines Formed by Cross-Link of Bone and Synovium Collagen. *Organic & Biomolecular Chemistry* **11**, 5747-5771, (2013).
53 Siegel, R. C. Collagen Cross-Linking. Synthesis of Collagen Cross-Links *in vitro* With Highly Purified Lysyl Oxidase. *Journal of Biological Chemistry* **251**, 5786-5792, (1976).
54 Aro, E., Salo, A. M., Khatri, R., Finnilä, M., Miinalainen, I. *et al.* Severe Extracellular Matrix Abnormalities and Chondrodysplasia in Mice Lacking Collagen Prolyl 4-Hydroxylase Isoenzyme II in Combination with a Reduced Amount of Isoenzyme I. *Journal of Biological Chemistry*, (2015).
55 Smith, T. G. & Talbot, N. P. Prolyl Hydroxylases and Therapeutics. *Antioxididants & Redox Signaling* **12**, 431-433, (2010).
56 Bartholomew, M. James Lind's *Treatise of the Scurvy* (1753). *Postgraduate Medical Journal* **78**, 695-696, (2002).
57 Suzuki, E., Fraser, R. D. B. & MacRae, T. P. Role of Hydroxyproline in the Stabilization of the Collagen Molecule *via* Water Molecules. *International Journal of Biological Macromolecules* **2**, 54-56, (1980).
58 Bretscher, L. E., Jenkins, C. L., Taylor, K. M., DeRider, M. L. & Raines, R. T. Conformational Stability of Collagen Relies on a Stereoelectronic Effect. *Journal of the American Chemical Society* **123**, 777-778, (2001).
59 Vitagliano, L., Berisio, R., Mazzarella, L. & Zagari, A. Structural Bases of Collagen Stabilization Induced by Proline Hydroxylation. *Biopolymers* **58**, 459-464, (2001).
60 Shoulders, M. D., Hodges, J. A. & Raines, R. T. Reciprocity of Steric and Stereoelectronic Effects in the Collagen Triple Helix. *Journal of the American Chemical Society* **128**, 8112-8113, (2006).
61 Shoulders, M. D., Guzei, I. A. & Raines, R. T. 4-Chloroprolines: Synthesis, Conformational Analysis, and Effect on the Collagen Triple Helix. *Biopolymers* **89**, 443-454, (2008).
62 Ogle, J. D., Arlinghaus, R. B. & Logan, M. A. 3-Hydroxyproline, a New Amino Acid of Collagen. *Journal of Biological Chemistry* **237**, 3667-3673, (1962).

63 Hudson, D. M., Werther, R., Weis, M., Wu, J.-J. & Eyre, D. R. Evolutionary Origins of C-Terminal $(GPP)_n$ 3-Hydroxyproline Formation in Vertebrate Tendon Collagen. *PloS One* **9**, e93467, (2014).

64 Pokidysheva, E., Boudko, S., Vranka, J., Zientek, K., Maddox, K. *et al.* Biological Role of Prolyl 3-Hydroxylation in Type IV Collagen. *Proceedings of the National Academy of Sciences* **111**, 161-166, (2014).

65 Hudson, D. M. & Eyre, D. R. Collagen Prolyl 3-Hydroxylation: A Major Role for a Minor Post-Translational Modification? *Connective Tissue Research* **54**, 245-251, (2013).

66 Morello, R., Bertin, T. K., Chen, Y., Hicks, J., Tonachini, L. *et al.* CRTAP Is Required for Prolyl 3-Hydroxylation and Mutations Cause Recessive Osteogenesis Imperfecta. *Cell* **127**, 291-304, (2006).

67 Marini, J. C., Reich, A. & Smith, S. M. Osteogenesis Imperfecta Due to Mutations in Non-Collagenous Genes: Lessons in the Biology of Bone Formation. *Current Opinion in Pediatrics* **26**, 500-507, (2014).

68 Witkop, B. 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* **12**, 372-374, (1956).

69 Vanacore, R., Ham, A.-J. L., Voehler, M., Sanders, C. R., Conrads, T. P. *et al.* A Sulfilimine Bond Identified in Collagen IV. *Science* **325**, 1230-1234, (2009).

70 Byers, P. H. & Murray, M. L. Ehlers–Danlos Syndrome: A Showcase of Conditions That Lead to Understanding Matrix Biology. *Matrix Biology* **33**, 10-15, (2014).

71 Abdalla, E., Rohrbach, M., Bürer, C., Kraenzlin, M., El-Tayeby, H. *et al.* Kyphoscoliotic Type of Ehlers-Danlos Syndrome (EDS VIA) in Six Egyptian Patients Presenting with a Homogeneous Clinical Phenotype. *European Journal of Pediatrics*, 1-8, (2014).

72 Steinmann, B., Royce, P. M. & Superti-Furga, A. in *Connective Tissue and Its Heritable Disorders* 431-523 (John Wiley & Sons, Inc., 2003).

73 van der Slot, A. J., Zuurmond, A.-M., Bardoel, A. F. J., Wijmenga, C., Pruijs, H. E. H. *et al.* Identification of PLOD2 as Telopeptide Lysyl Hydroxylase, an Important Enzyme in Fibrosis. *Journal of Biological Chemistry* **278**, 40967-40972, (2003).

74 Tryggvason, K., Risteli, J. & Kivirikko, K. I. Separation of Prolyl 3-Hydroxylase and 4-Hydroxylase Activities and the 4-Hydroxyproline Requirement for Synthesis of 3-Hydroxyproline. *Biochemical and Biophysical Research Communications* **76**, 275-281, (1977).

75 Kivirikko, K. I. & Myllylä, R. in *Methods in Enzymology* Vol. Volume 82 (ed Leon W. Cunningham, D. W. F.) 245-304 (Academic Press, 1982).

76 Zhang, Z., Ren, J.-S., Clifton, I. J. & Schofield, C. J. Crystal Structure and Mechanistic Implications of 1-Aminocyclopropane-1-Carboxylic Acid Oxidase - The Ethylene-Forming Enzyme. *Chemistry & Biology* **11**, 1383-1394, (2004).

77 Roach, P. L., Clifton, I. J., Hensgens, C. M. H., Shibata, N., Schofield, C. J. *et al.* Structure of Isopenicillin-*N*-Synthase Complexed With Substrate and the Mechanism of Penicillin Formation. *Nature* **387**, 827-830, (1997).

78 Bollinger Jr, J. M., Chang, W.-c., Matthews, M. L., Martinie, R. J., Boal, A. K. *et al.* in *2-Oxoglutarate-Dependent Oxygenases* (eds Schofield, C. J. & Hausinger, R. P.) 95-122 (The Royal Society of Chemistry, 2015).

79 Hutton Jr, J. J., Tappel, A. L. & Udenfriend, S. Requirements for α-Ketoglutarate, Ferrous Ion and Ascorbate by Collagen Proline Hydroxylase. *Biochemical and Biophysical Research Communications* **24**, 179-184, (1966).

80 Aik, W. S., Chowdhury, R., Clifton, I. J., Hopkinson, R. J., Leissing, T. *et al.* in *2-Oxoglutarate-Dependent Oxygenases* (eds Schofield, C. J. & Hausinger, R. P.) 59-94 (The Royal Society of Chemistry, 2015).

81 Yu, B. & Hunt, J. F. Enzymological and Structural Studies of the Mechanism of Promiscuous Substrate Recognition by the Oxidative DNA Repair Enzyme AlkB. *Proceedings of the National Academy of Sciences* **106**, 14315-14320, (2009).
82 Chowdhury, R., McDonough, M. A., Mecinović, J., Loenarz, C., Flashman, E. *et al.* Structural Basis for Binding of Hypoxia-Inducible Factor to the Oxygen-Sensing Prolyl Hydroxylases. *Structure* **17**, 981-989, (2009).
83 Horton, J. R., Upadhyay, A. K., Qi, H. H., Zhang, X., Shi, Y. *et al.* Enzymatic and Structural Insights for Substrate Specificity of a Family of Jumonji Histone Lysine Demethylases. *Nature Structural & Molecular Biology* **17**, 38-43, (2010).
84 Longbotham, J. E., Levy, C., Johannissen, L. O., Tarhonskaya, H., Jiang, S. *et al.* Structure and Mechanism of a Viral Collagen Prolyl Hydroxylase. *Biochemistry* **54**, 6093-6105, (2015).
85 Walport, L. J., Hopkinson, R. J. & Schofield, C. J. Mechanisms of Human Histone and Nucleic Acid Demethylases. *Current Opinion in Chemical Biology* **16**, 525-534, (2012).
86 Hopkinson, R. J., Hamed, R. B., Rose, N. R., Claridge, T. D. W. & Schofield, C. J. Monitoring the Activity of 2-Oxoglutarate Dependent Histone Demethylases by NMR Spectroscopy: Direct Observation of Formaldehyde. *ChemBioChem* **11**, 506-510, (2010).
87 Hirsilä, M., Koivunen, P., Günzler, V., Kivirikko, K. I. & Myllyharju, J. Characterization of the Human Prolyl 4-Hydroxylases That Modify the Hypoxia-Inducible Factor. *Journal of Biological Chemistry* **278**, 30772-30780, (2003).
88 Szarka, A. & Lőrincz, T. The Role of Ascorbate in Protein Folding. *Protoplasma* **251**, 489-497, (2014).
89 Puistola, U., Turpeenniemi-Hujanen, T. M., Myllylä, R. & Kivirikko, K. I. Studies on the Lysyl Hydroxylase Reaction. I. Initial Velocity Kinetics and Related Aspects. *Biochimica et Biophysica Acta (BBA) - Enzymology* **611**, 40-50, (1980).
90 Peterkofsky, B. & Udenfriend, S. Enzymatic Hydroxylation of Proline in Microsomal Polypeptide Leading to Formation of Collagen. *Proceedings of the National Academy of Sciences of the United States of America* **53**, 335-342, (1965).
91 Hutton Jr, J. J., Tappel, A. L. & Udenfriend, S. Cofactor and Substrate Requirements of Collagen Proline Hydroxylase. *Archives of Biochemistry and Biophysics* **118**, 231-240, (1967).
92 Koivu, J., Myllylä, R., Helaakoski, T., Pihlajaniemi, T., Tasanen, K. *et al.* A Single Polypeptide Acts Both as the Beta Subunit of prolyl 4-Hydroxylase and as a Protein Disulfide-Isomerase. *Journal of Biological Chemistry* **262**, 6447-6449, (1987).
93 Annunen, P., Helaakoski, T., Myllyharju, J., Veijola, J., Pihlajaniemi, T. *et al.* Cloning of the Human Prolyl 4-Hydroxylase α Subunit Isoform α(II) and Characterization of the Type II Enzyme Tetramer: The α(I) And α(II) Subunits do not Form a Mixed α(I)α(II)β_2 Tetramer. *Journal of Biological Chemistry* **272**, 17342-17348, (1997).
94 Annunen, P., Autio-Harmainen, H. & Kivirikko, K. I. The Novel Type II Prolyl 4-Hydroxylase Is the Main Enzyme Form in Chondrocytes and Capillary Endothelial Cells, whereas the Type I Enzyme Predominates in Most Cells. *Journal of Biological Chemistry* **273**, 5989-5992, (1998).
95 Van Den Diepstraten, C., Papay, K., Bolender, Z., Brown, A. & Pickering, J. G. Cloning of a Novel Prolyl 4-Hydroxylase Subunit Expressed in the Fibrous Cap of Human Atherosclerotic Plaque. *Circulation* **108**, 508-511, (2003).
96 Kukkola, L., Hieta, R., Kivirikko, K. I. & Myllyharju, J. Identification and Characterization of a Third Human, Rat, and Mouse Collagen Prolyl 4-Hydroxylase Isoenzyme. *Journal of Biological Chemistry* **278**, 47685-47693, (2003).
97 Helaakoski, T., Annunen, P., Vuori, K., MacNeil, I. A., Pihlajaniemi, T. *et al.* Cloning, Baculovirus Expression, and Characterization of a Second Mouse Prolyl 4-Hydroxylase Alpha-Subunit Isoform: Formation of an Alpha 2 Beta 2 Tetramer with the Protein

Disulfide-Isomerase/Beta Subunit. *Proceedings of the National Academy of Sciences* **92**, 4427-4431, (1995).

98 Zou, Y., Donkervoort, S., Salo, A. M., Foley, A. R., Barnes, A. M. *et al.* P4HA1 Mutations Cause a Unique Congenital Disorder of Connective Tissue Involving Tendon, Bone, Muscle and the Eye. *Human Molecular Genetics* **26**, 2207-2217, (2017).

99 Rauch, F., Fahiminiya, S., Majewski, J., Carrot-Zhang, J., Boudko, S. *et al.* Cole-Carpenter Syndrome is Caused by a Heterozygous Missense Mutation in P4HB. *The American Journal of Human Genetics* **96**, 425-431, (2015).

100 Koski, M. K., Anantharajan, J., Kursula, P., Dhavala, P., Murthy, A. V. *et al.* Assembly of the Elongated Collagen Prolyl 4-Hydroxylase $\alpha_2\beta_2$ Heterotetramer Around a Central α_2 Dimer. *Biochemical Journal* **474**, 751-769, (2017).

101 Vuori, K., Pihlajaniemi, T., Marttila, M. & Kivirikko, K. I. Characterization of the Human Prolyl 4-Hydroxylase Tetramer and its Multifunctional Protein Disulfide-Isomerase Subunit Synthesized in a Baculovirus Expression System. *Proceedings of the National Academy of Sciences* **89**, 7467-7470, (1992).

102 Cunliffe, C. J. & Franklin, T. J. Inhibition of Prolyl 4-Hydroxylase by Hydroxyanthraquinones. *Biochemical Journal* **239**, 311-315, (1986).

103 Myllyharju, J. in *2-Oxoglutarate-Dependent Oxygenases* (eds Schofield, C. J. & Hausinger, R. P.) 149-168 (The Royal Society of Chemistry, 2015).

104 Vranka, J. A., Sakai, L. Y. & Bächinger, H. P. Prolyl 3-Hydroxylase 1, Enzyme Characterization and Identification of a Novel Family of Enzymes. *Journal of Biological Chemistry* **279**, 23615-23621, (2004).

105 Ishikawa, Y., Wirz, J., Vranka, J. A., Nagata, K. & Bächinger, H. P. Biochemical Characterization of the Prolyl 3-Hydroxylase 1·Cartilage-associated Protein·Cyclophilin B Complex. *Journal of Biological Chemistry* **284**, 17641-17647, (2009).

106 Cabral, W. A., Perdivara, I., Weis, M., Terajima, M., Blissett, A. R. *et al.* Abnormal Type I Collagen Post-translational Modification and Crosslinking in a Cyclophilin B KO Mouse Model of Recessive Osteogenesis Imperfecta. *PLoS Genetics* **10**, e1004465, (2014).

107 Ishikawa, Y. & Bächinger, H. P. An Additional Function of the Rough Endoplasmic Reticulum Protein Complex Prolyl 3-Hydroxylase 1·Cartilage-associated Protein Cyclophilin B: The CXXXC Motif Reveals Disulfide Isomerase Activity *In Vitro*. *Journal of Biological Chemistry* **288**, 31437-31446, (2013).

108 Tiainen, P., Pasanen, A., Sormunen, R. & Myllyharju, J. Characterization of Recombinant Human Prolyl 3-Hydroxylase Isoenzyme 2, an Enzyme Modifying the Basement Membrane Collagen IV. *Journal of Biological Chemistry* **283**, 19432-19439, (2008).

109 Rautavuoma, K., Takaluoma, K., Passoja, K., Pirskanen, A., Kvist, A.-P. *et al.* Characterization of Three Fragments That Constitute the Monomers of the Human Lysyl Hydroxylase Isoenzymes 1–3: The 30-kDa *N*-Terminal Fragment Is Not Required For Lysyl Hydroxylase Activity. *Journal of Biological Chemistry* **277**, 23084-23091, (2002).

110 Gjaltema, R. A. F., van der Stoel, M. M., Boersema, M. & Bank, R. A. Disentangling Mechanisms Involved in Collagen Pyridinoline Cross-Linking: The Immunophilin FKBP65 is Critical for Dimerization of lysyl hydroxylase 2. *Proceedings of the National Academy of Sciences* **113**, 7142-7147, (2016).

111 Okumura, C. Y. M., Hollands, A., Tran, D. N., Olson, J., Dahesh, S. *et al.* A New Pharmacological Agent (AKB-4924) Stabilizes Hypoxia Inducible Factor-1 (HIF-1) and Increases Skin Innate Defenses Against Bacterial Infection. *Journal of Molecular Medicine* **90**, 1079-1089, (2012).

112 Jaakkola, P., Mole, D. R., Tian, Y.-M., Wilson, M. I., Gielbert, J. *et al.* Targeting of HIF-α to the von Hippel-Lindau Ubiquitylation Complex by O_2-Regulated Prolyl Hydroxylation. *Science* **292**, 468-472, (2001).

113 Feng, T., Yamamoto, A., Wilkins, S. E., Sokolova, E., Yates, L. A. *et al.* Optimal Translational Termination Requires C4 Lysyl Hydroxylation of eRF1. *Molecular Cell* **53**, 645-654, (2014).
114 Loenarz, C., Sekirnik, R., Thalhammer, A., Ge, W., Spivakovsky, E. *et al.* Hydroxylation of the Eukaryotic Ribosomal Decoding Center Affects Translational Accuracy. *Proceedings of the National Academy of Sciences* **111**, 4019-4024, (2014).
115 Rose, N. R., Schofield, C. J. & Heightman, T. D. in *New Frontiers in Chemical Biology: Enabling Drug Discovery* Vol. 5 151-203 (Royal Society of Chemistry, 2011).
116 Mantri, M., Loik, N. D., Hamed, R. B., Claridge, T. D. W., McCullagh, J. S. O. *et al.* The 2-Oxoglutarate-Dependent Oxygenase JMJD6 Catalyses Oxidation of Lysine Residues to Give *5S*-Hydroxylysine Residues. *ChemBioChem* **12**, 531-534, (2011).
117 Webby, C. J., Wolf, A., Gromak, N., Dreger, M., Kramer, H. *et al.* JMJD6 Catalyses Lysyl-Hydroxylation of U2AF65, a Protein Associated with RNA Splicing. *Science* **325**, 90-93, (2009).
118 Chang, B., Chen, Y., Zhao, Y. & Bruick, R. K. JMJD6 is a Histone Arginine Demethylase. *Science* **318**, 444-447, (2007).
119 Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K. *et al.* Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments using Clustal Omega. *Molecular Systems Biology* **7**, (2011).
120 FigTree. Tree Figure Drawing Tool v. 1.4.3. (Institute of Evolutionary Biology, University of Edinburgh, 2016).
121 Singleton, R. S., Liu-Yi, P., Formenti, F., Ge, W., Sekirnik, R. *et al.* OGFOD1 Catalyzes Prolyl Hydroxylation of RPS23 and is Involved in Translation Control and Stress Granule Formation. *Proceedings of the National Academy of Sciences* **111**, 4031-4036, (2014).
122 Wang, G. L., Jiang, B. H., Rue, E. A. & Semenza, G. L. Hypoxia-Inducible Factor 1 is a Basic-Helix-Loop-Helix-PAS Heterodimer Regulated by Cellular O_2 Tension. *Proceedings of the National Academy of Sciences* **92**, 5510-5514, (1995).
123 Wilkins, S. E., Flashman, E., Scotti, J. S., Hopkinson, R. J., Chowdhury, R. *et al.* in *2-Oxoglutarate-Dependent Oxygenases* (eds Schofield, C. J. & Hausinger, R. P.) 169-209 (The Royal Society of Chemistry, 2015).
124 Schofield, C. J. & Ratcliffe, P. J. Oxygen Sensing by HIF Hydroxylases. *Nature Reviews Molecular Cell Biology* **5**, 343-354, (2004).
125 Hon, W.-C., Wilson, M. I., Harlos, K., Claridge, T. D. W., Schofield, C. J. *et al.* Structural Basis for the Recognition of Hydroxyproline in HIF-1α by pVHL. *Nature* **417**, 975-978, (2002).
126 Min, J.-H., Yang, H., Ivan, M., Gertler, F., Kaelin, W. G. *et al.* Structure of an HIF-1α-pVHL Complex: Hydroxyproline Recognition in Signaling. *Science* **296**, 1886-1889, (2002).
127 Flashman, E., Hoffart, L. M., Hamed, R. B., Bollinger Jr, J. M., Krebs, C. *et al.* Evidence for the Slow Reaction of Hypoxia-Inducible Factor Prolyl Hydroxylase 2 with Oxygen. *FEBS Journal* **277**, 4089-4099, (2010).
128 Tarhonskaya, H., Hardy, A. P., Howe, E. A., Loik, N. D., Kramer, H. B. *et al.* Kinetic Investigations of the Role of Factor Inhibiting Hypoxia-Inducible Factor (FIH) as an Oxygen Sensor. *Journal of Biological Chemistry*, (2015).
129 Semenza, G. L. & Wang, G. L. A Nuclear Factor Induced by Hypoxia *via De Novo* Protein Synthesis Binds to the Human Erythropoietin Gene Enhancer at a Site Required for Transcriptional Activation. *Molecular and Cellular Biology* **12**, 5447-5454, (1992).
130 Halberg, N., Khan, T., Trujillo, M. E., Wernstedt-Asterholm, I., Attie, A. D. *et al.* Hypoxia-Inducible Factor 1α Induces Fibrosis and Insulin Resistance in White Adipose Tissue. *Molecular and Cellular Biology* **29**, 4467-4483, (2009).
131 Gilkes, D. M., Bajpai, S., Chaturvedi, P., Wirtz, D. & Semenza, G. L. Hypoxia-Inducible Factor 1 (HIF-1) Promotes Extracellular Matrix Remodeling under Hypoxic Conditions by

Inducing P4HA1, P4HA2, and PLOD2 Expression in Fibroblasts. *Journal of Biological Chemistry* **288**, 10819-10829, (2013).

132 Vaz, F. M. & van Vlies, N. in *2-Oxoglutarate-Dependent Oxygenases* (eds Schofield, C. J. & Hausinger, R. P.) 324-337 (The Royal Society of Chemistry, 2015).

133 Vaz, F. M., Ofman, R., Westinga, K., Back, J. W. & Wanders, R. J. A. Molecular and Biochemical Characterization of Rat ε-*N*-Trimethyllysine Hydroxylase, the First Enzyme of Carnitine Biosynthesis. *Journal of Biological Chemistry* **276**, 33512-33517, (2001).

134 Lesniak, R. K., Markolovic, S., Tars, K. & Schofield, C. J. Human Carnitine Biosynthesis Proceeds *via* (*2S,3S*)-3-hydroxy-N^{ϵ}-trimethyllysine. *Chemical Communications* **53**, 440-442, (2017).

135 Markolovic, S. *Biochemical and Biophysical Studies on the Jumonji-C Hydroxylases JMJD7 and JMJD4* DPhil thesis, University of Oxford, (2017).

136 Rose, N. R., McDonough, M. A., King, O. N. F., Kawamura, A. & Schofield, C. J. Inhibition of 2-Oxoglutarate Dependent Oxygenases. *Chemical Society Reviews* **40**, 4364-4397, (2011).

137 Franklin, T. J., Morris, W. P., Edwards, P. N., Large, M. S. & Stephenson, R. Inhibition of Prolyl 4-Hydroxylase *in vitro* and *in vivo* by Members of a Novel Series of Phenanthrolinones. *Biochemical Journal* **353**, 333-338, (2001).

138 Franklin, T. J. & Hitchen, M. Inhibition of Collagen Hydroxylation by 2,7,8-Trihydroxyanthraquinone in Embryonic-Chick Tendon Cells. *Biochemical Journal* **261**, 127-130, (1989).

139 Rahtu-Korpela, L., Määttä, J., Dimova, E. Y., Hörkkö, S., Gylling, H. *et al.* Hypoxia-Inducible Factor-Prolyl 4-Hydroxylase-2 Inhibition Protects Against Development of Atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology*, (2016).

140 Maxwell, P. H. & Eckardt, K.-U. HIF Prolyl Hydroxylase Inhibitors for the Treatment of Renal Anaemia and Beyond. *Nature Reviews Nephrology* **12**, 157-168, (2016).

141 Shalwitz, R., Hartman, C., Buch, A., Shalwitz, I., Janusz, J. *et al.* Compositions and Methods for Treating Anemia. WO2014200773A2 (2014).

142 Fraisl, P., Aragones, J. & Carmeliet, P. Inhibition of Oxygen Sensors as a Therapeutic Strategy for Ischaemic and Inflammatory Disease. *Nature Reviews: Drug Discovery* **8**, 139-152, (2009).

143 Besarab, A., Chernyavskaya, E., Motylev, I., Shutov, E., Kumbar, L. M. *et al.* Roxadustat (FG-4592): Correction of Anemia in Incident Dialysis Patients. *Journal of the American Society of Nephrology* **27**, 1225-1233, (2016).

144 Martin, E. R., Smith, M. T., Maroni, B. J., Zuraw, Q. C. & deGoma, E. M. Clinical Trial of Vadadustat in Patients with Anemia Secondary to Stage 3 or 4 Chronic Kidney Disease. *American Journal of Nephrology* **45**, 380-388, (2017).

145 Pergola, P. E., Spinowitz, B. S., Hartman, C. S., Maroni, B. J. & Haase, V. H. Vadadustat, a Novel Oral HIF Stabilizer, Provides Effective Anemia Treatment in Nondialysis-Dependent Chronic Kidney Disease. *Kidney International* **90**, 1115-1122, (2016).

146 Holdstock, L., Meadowcroft, A. M., Maier, R., Johnson, B. M., Jones, D. *et al.* Four-Week Studies of Oral Hypoxia-Inducible Factor–Prolyl Hydroxylase Inhibitor GSK1278863 for Treatment of Anemia. *Journal of the American Society of Nephrology* **27**, 1234-1244, (2016).

147 Flamme, I., Oehme, F., Ellinghaus, P., Jeske, M., Keldenich, J. *et al.* Mimicking Hypoxia to Treat Anemia: HIF-Stabilizer BAY 85-3934 (Molidustat) Stimulates Erythropoietin Production without Hypertensive Effects. *PloS One* **9**, e111838, (2014).

148 Eltzschig, H. K., Bratton, D. L. & Colgan, S. P. Targeting Hypoxia Signalling for the Treatment of Ischaemic and Inflammatory Diseases. *Nature Reviews Drug Discovery* **13**, 852-869, (2014).

149 Dambrova, M., Liepinsh, E. & Kalvinsh, I. Mildronate: Cardioprotective Action through Carnitine-Lowering Effect. *Trends in Cardiovascular Medicine* **12**, 275-279, (2002).
150 Zvejniece, L., Svalbe, B., Makrecka, M., Liepinsh, E., Kalvinsh, I. *et al.* Mildronate Exerts Acute Anticonvulsant and Antihypnotic Effects. *Behavioural Pharmacology* **21**, 548-555, (2010).
151 Johansson, C., Tumber, A., Che, K. H., Cain, P., Nowak, R. *et al.* The Roles of Jumonji-Type Oxygenases in Human Disease. *Epigenomics* **6**, 89-120, (2014).
152 Dawson, Mark A. & Kouzarides, T. Cancer Epigenetics: From Mechanism to Therapy. *Cell* **150**, 12-27, (2012).
153 Højfeldt, J. W., Agger, K. & Helin, K. Histone Lysine Demethylases as Targets for Anticancer Therapy. *Nature Reviews Drug Discovery* **12**, 917-930, (2013).
154 Hancock, R. L., Dunne, K., Walport, L. J., Flashman, E. & Kawamura, A. Epigenetic Regulation by Histone Demethylases in Hypoxia. *Epigenomics* **7**, 791-811, (2015).
155 McAllister, T. E., England, K. S., Hopkinson, R. J., Brennan, P. E., Kawamura, A. *et al.* Recent Progress in Histone Demethylase Inhibitors. *Journal of Medicinal Chemistry* **59**, 1308-1329, (2016).
156 Thinnes, C. C., England, K. S., Kawamura, A., Chowdhury, R., Schofield, C. J. *et al.* Targeting Histone Lysine Demethylases - Progress, Challenges, and the Future. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1839**, 1416-1432, (2014).
157 Hopkinson, R. J., Tumber, A., Yapp, C., Chowdhury, R., Aik, W. *et al.* 5-Carboxy-8-Hydroxyquinoline is a Broad Spectrum 2-Oxoglutarate Oxygenase Inhibitor Which Causes Iron Translocation. *Chemical Science* **4**, 3110-3117, (2013).
158 Thinnes, C. C. *Chemical and Biological Studies on Human Oxygenases* DPhil thesis, University of Oxford, (2014).
159 Rademacher, W. Growth Retardants: Effects on Gibberellin Biosynthesis and Other Metabolic Pathways. *Annual Review of Plant Physiology and Plant Molecular Biology* **51**, 501-531, (2000).
160 Rose, N. R., Woon, E. C. Y., Tumber, A., Walport, L. J., Chowdhury, R. *et al.* Plant Growth Regulator Daminozide Is a Selective Inhibitor of Human KDM2/7 Histone Demethylases. *Journal of Medicinal Chemistry* **55**, 6639-6643, (2012).
161 Gorres, K. L. & Raines, R. T. Prolyl 4-Hydroxylase. *Critical Reviews in Biochemistry and Molecular Biology* **45**, 106-124, (2010).
162 Anantharajan, J., Koski, M. K., Kursula, P., Hieta, R., Bergmann, U. *et al.* The Structural Motifs for Substrate Binding and Dimerization of the α Subunit of Collagen Prolyl 4-Hydroxylase. *Structure* **21**, 2107-2118, (2013).
163 Vuori, K., Pihlajaniemi, T., Myllyla, R. & Kivirikko, K. I. Site-directed Mutagenesis of Human Protein Disulphide Isomerase: Effect on the Assembly, Activity and Endoplasmic Reticulum Retention of Human Prolyl 4-Hydroxylase in *Spodoptera frugiperda* Insect Cells. *EMBO Journal* **11**, 4213-4217, (1992).
164 John, D. C. A. & Bulleid, N. J. Prolyl 4-Hydroxylase: Defective Assembly of α-Subunit Mutants Indicates That Assembled α-Subunits Are Intramolecularly Disulfide Bonded. *Biochemistry* **33**, 14018-14025, (1994).
165 Curtis-Marof, R., Doko, D., Rowe, M. L., Richards, K. L., Williamson, R. A. *et al.* ^{19}F NMR Spectroscopy Monitors Ligand Binding to Recombinantly Fluorine-Labelled b'x From Human Protein Disulphide Isomerase (hPDI). *Organic & Biomolecular Chemistry* **12**, 3808-3812, (2014).
166 Wang, C., Li, W., Ren, J., Fang, J., Ke, H. *et al.* Structural Insights into the Redox-Regulated Dynamic Conformations of Human Protein Disulfide Isomerase. *Antioxidants & Redox Signaling* **19**, 36-45, (2012).

167 Koski, M. K., Hieta, R., Böllner, C., Kivirikko, K. I., Myllyharju, J. *et al.* The Active Site of an Algal Prolyl 4-Hydroxylase Has a Large Structural Plasticity. *Journal of Biological Chemistry* **282**, 37112-37123, (2007).
168 Berg, R. A. & Prockop, D. J. Affinity Column Purification of Protocollagen Proline Hydroxylase from Chick Embryos and Further Characterization of the Enzyme. *Journal of Biological Chemistry* **248**, 1175-1182, (1973).
169 Lamberg, A., Pihlajaniemi, T. & Kivirikko, K. I. Site-Directed Mutagenesis of the α Subunit of Human Prolyl 4-Hydroxylase: Identification of Three Histidine Residues Critical for Catalytic Activity. *Journal of Biological Chemistry* **270**, 9926-9931, (1995).
170 Kersteen, E. A., Higgin, J. J. & Raines, R. T. Production of Human Prolyl 4-Hydroxylase in *Escherichia coli*. *Protein Expression and Purification* **38**, 279-291, (2004).
171 Neubauer, A., Neubauer, P. & Myllyharju, J. High-Level Production of Human Collagen Prolyl 4-Hydroxylase in *Escherichia coli*. *Matrix Biology* **24**, 59-68, (2005).
172 Scotti, J. S. & Aik, W. S. *Unpublished Results*.
173 Aik, W. S. *Crystallographic Studies on 2-Oxoglutarate Dependent Oxygenases* DPhil thesis, University of Oxford, (2014).
174 Scotti, J. S. *Structural, Functional and Evolutionary Studies on Prolyl-Hydroxylases* DPhil thesis, University of Oxford, (2014).
175 Gorres, K. L., Edupuganti, R., Krow, G. R. & Raines, R. T. Conformational Preferences of Substrates for Human Prolyl 4-Hydroxylase†. *Biochemistry* **47**, 9447-9455, (2008).
176 Bird, L. *Oxford Protein Production Facility - Standard Protocols: Cloning and Expression Screening*, www.oppf.rc-harwell.ac.uk/OPPF/protocols/sop/OPPF-UK%20SOP%20Cloning%20and%20Small%20Scale%20Expression%2020161005.pdf (2016).
177 Gileadi, O., Burgess-Brown, N. A., Colebrook, S. M., Berridge, G., Savitsky, P. *et al.* in *Structural Proteomics: High-Throughput Methods* (eds Kobe, B., Guss, M., & Huber, T.) 221-246 (Humana Press, 2008).
178 Giannone, R. J., McDonald, W. H., Hurst, G. B., Huang, Y., Wu, J. *et al.* Dual-Tagging System for the Affinity Purification of Mammalian Protein Complexes. *BioTechniques* **43**, 296,298,300,302, (2007).
179 Eriksson, M., Myllyharju, J., Tu, H., Hellman, M. & Kivirikko, K. I. Evidence for 4-Hydroxyproline in Viral Proteins: Characterization of a Viral Prolyl 4-Hydroxylase and Its Peptide Substrates. *Journal of Biological Chemistry* **274**, 22131-22134, (1999).
180 Aik, W., Demetriades, M., Hamdan, M. K. K., Bagg, E. A. L., Yeoh, K. K. *et al.* Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO). *Journal of Medicinal Chemistry* **56**, 3680-3688, (2013).
181 Rydzik, A. M., Chowdhury, R., Kochan, G. T., Williams, S. T., McDonough, M. A. *et al.* Modulating Carnitine Levels by Targeting its Biosynthesis - Selective Inhibition of γ-Butyrobetaine Hydroxylase. *Chemical Science* **5**, 1765-1771, (2014).
182 Chowdhury, R., Leung, I. K. H., Tian, Y.-M., Abboud, M. I., Ge, W. *et al.* Structural Basis for Oxygen Degradation Domain Selectivity of the HIF Prolyl Hydroxylases. *Nature Communications* **7**, 12673, (2016).
183 Tschank, G., Raghunath, M., Günzler, V. & Hanauske-Abel, H. M. Pyridinedicarboxylates, the First Mechanism-Derived Inhibitors for Prolyl 4-Hydroxylase, Selectively Suppress Cellular Hydroxyprolyl Biosynthesis. Decrease in Interstitial Collagen and Clq Secretion in Cell Culture. *Biochemical Journal* **248**, 625-633, (1987).
184 Burzlaff, N. *Unpublished Results* (Friedrich Alexander University Erlangen-Nürnberg, 2016).
185 Thinnes, C. C., Tumber, A., Yapp, C., Scozzafava, G., Yeh, T. *et al.* Betti Reaction Enables Efficient Synthesis of 8-Hydroxyquinoline Inhibitors of 2-Oxoglutarate Oxygenases. *Chemical Communications* **51**, 15458-15461, (2015).

186 Childs, S. J. Safety of the Fluoroquinolone Antibiotics: Focus on Molecular Structure *Infections in Urology* **13**, 3-10, (2000).
187 Kruidenier, L., Chung, C.-w., Cheng, Z., Liddle, J., Che, K. *et al.* A Selective Jumonji H3K27 Demethylase Inhibitor Modulates the Proinflammatory Macrophage Response. *Nature* **488**, 404-408, (2012).
188 Yapp, C., Carr, A. J., Price, A., Oppermann, U. & Snelling, S. J. B. H3K27me3 Demethylases Regulate *in vitro* Chondrogenesis and Chondrocyte Activity in Osteoarthritis. *Arthritis Research & Therapy* **18**, 158, (2016).
189 Shamo Schiller, R. *Investigating the Inhibitor and Substrate Diversity of the JmjC Histone Demethylases* DPhil thesis, University of Oxford, (2016).
190 Chowdhury, R., Candela-Lena, J. I., Chan, M. C., Greenald, D. J., Yeoh, K. K. *et al.* Selective Small Molecule Probes for the Hypoxia Inducible Factor (HIF) Prolyl Hydroxylases. *ACS Chemical Biology* **8**, 1488-1496, (2013).
191 Chan, M. C., Atasoylu, O., Hodson, E., Tumber, A., Leung, I. K. H. *et al.* Potent and Selective Triazole-Based Inhibitors of the Hypoxia-Inducible Factor Prolyl-Hydroxylases with Activity in the Murine Brain. *PloS One* **10**, e0132004, (2015).
192 Auvinet, B., Ziller, R., Appelboom, T. & Vélicitat, P. Comparison of the Onset and Intensity of Action of Intramuscular Meloxicam and Oral Meloxicam in Patients in Acute Sciatica. *Clinical Therapeutics* **17**, 1078-1090, (1995).
193 Leung, I. K. H., Krojer, T. J., Kochan, G. T., Henry, L., von Delft, F. *et al.* Structural and Mechanistic Studies on γ-Butyrobetaine Hydroxylase. *Chemistry & Biology* **17**, 1316-1324, (2010).
194 Yeh, T.-L., Leissing, T. M., Abboud, M. I., Thinnes, C. C., Atasoylu, O. *et al.* Molecular and Cellular Mechanisms of HIF Prolyl Hydroxylase inhibitors in Clinical Trials. *Chemical Science*, (2017).
195 McDonough, M. A., McNeill, L. A., Tilliet, M., Papamicaël, C. A., Chen, Q.-Y. *et al.* Selective Inhibition of Factor Inhibiting Hypoxia-Inducible Factor. *Journal of the American Chemical Society* **127**, 7680-7681, (2005).
196 Grant, S., Easley, C. & Kirkpatrick, P. Vorinostat. *Nature Reviews Drug Discovery* **6**, 21-22, (2007).
197 Simeonov, A. Recent Developments in the Use of Differential Scanning Fluorometry in Protein and Small Molecule Discovery and Characterization. *Expert Opinion on Drug Discovery* **8**, 1071-1082, (2013).
198 Badal, S., Her, Y. F. & Maher, L. J. Nonantibiotic Effects of Fluoroquinolones in Mammalian Cells. *Journal of Biological Chemistry* **290**, 22287-22297, (2015).
199 Maia, A. S., Ribeiro, A. R., Amorim, C. L., Barreiro, J. C., Cass, Q. B. *et al.* Degradation of Fluoroquinolone Antibiotics and Identification of Metabolites/Transformation Products by Liquid Chromatography–Tandem Mass Spectrometry. *Journal of Chromatography A* **1333**, 87-98, (2014).
200 Gumucio, J. P., Flood, M. D., Bedi, A., Kramer, H. F., Russell, A. J. *et al.* Inhibition of Prolyl 4-Hydroxylase Decreases Muscle Fibrosis Following Chronic Rotator Cuff Tear. *Bone and Joint Research* **6**, 57-65, (2017).
201 Horita, S., Scotti, John S., Thinnes, C., Mottaghi-Taromsari, Yousef S., Thalhammer, A. *et al.* Structure of the Ribosomal Oxygenase OGFOD1 Provides Insights into the Regio- and Stereoselectivity of Prolyl Hydroxylases. *Structure* **23**, 639-652, (2015).
202 Copeland, R. A. in *Evaluation of Enzyme Inhibitors in Drug Discovery* 57-121 (John Wiley & Sons, Inc., 2013).
203 Holt-Martyn, J., Tumber, A. & Schofield, C. J. *Unpublished Results* (2017).
204 Tarhonskaya, H., Nowak, R. P., Johansson, C., Szykowska, A., Tumber, A. *et al.* Studies on the Interaction of the Histone Demethylase KDM5B with Tricarboxylic Acid Cycle Intermediates. *Journal of Molecular Biology*, (2017).

205 Majamaa, K., Turpeenniemi-Hujanen, T. M., Latipää, P., Günzler, V., Hanauske-Abel, H. M. *et al.* Differences Between Collagen Hydroxylases and 2-Oxoglutarate Dehydrogenase in their Inhibition by Structural Analogues of 2-Oxoglutarate. *Biochemical Journal* **229**, 127, (1985).

206 Baysal, B. E., Ferrell, R. E., Willett-Brozick, J. E., Lawrence, E. C., Myssiorek, D. *et al.* Mutations in *SDHD*, a Mitochondrial Complex II Gene, in Hereditary Paraganglioma. *Science* **287**, 848-851, (2000).

207 Dang, L., Jin, S. & Su, S. M. IDH Mutations in Glioma and Acute Myeloid Leukemia. *Trends in Molecular Medicine* **16**, 387-397, (2010).

208 Chowdhury, R., Yeoh, K. K., Tian, Y. M., Hillringhaus, L., Bagg, E. A. *et al.* The Oncometabolite 2-Hydroxyglutarate Inhibits Histone Lysine Demethylases. *EMBO Reports* **12**, 463-469, (2011).

209 Snelling, S., Rout, R., Davidson, R., Clark, I., Carr, A. J. *et al.* A Gene Expression Study of Normal and Damaged Cartilage in Anteromedial Gonarthrosis, a Phenotype of Osteoarthritis. *Osteoarthritis and Cartilage* **22**, 334-343, (2014).

210 Haut, R. C., Lancaster, R. L. & DeCamp, C. E. Mechanical Properties of the Canine Patellar Tendon: Some Correlations with Age and the Content of Collagen. *Journal of Biomechanics* **25**, 163-173, (1992).

211 Kraenzlin, M. E., Kraenzlin, C. A., Meier, C., Giunta, C. & Steinmann, B. Automated HPLC Assay for Urinary Collagen Cross-links: Effect of Age, Menopause, and Metabolic Bone Diseases. *Clinical Chemistry* **54**, 1546-1553, (2008).

212 Parkkari, J., Kannus, P., Niemi, S., Pasanen, M., JÄrvinen, M. *et al.* Secular Trends in Osteoporotic Pelvic Fractures in Finland: Number and Incidence of Fractures in 1970–1991 and Prediction for the Future. *Calcified Tissue International* **59**, 79-83, (1996).

213 Brewer, L., Williams, D. & Moore, A. Current and Future Treatment Options in Osteoporosis. *European Journal of Clinical Pharmacology* **67**, 321-331, (2011).

214 Clark, J. M. & Harryman, D. T. Tendons, Ligaments, and Capsule of the Rotator Cuff. Gross and Microscopic Anatomy. *The Journal of Bone & Joint Surgery* **74**, 713-725, (1992).

215 Chaudhury, S., Dicko, C., Burgess, M., Vollrath, F. & Carr, A. J. Fourier Transform Infrared Spectroscopic Analysis of Normal and Torn Rotator-Cuff Tendons. *The Journal of Bone and Joint Surgery, British Volume* **93**, 370-377, (2011).

216 Dean, B. J. F., Franklin, S. L. & Carr, A. J. A Systematic Review of the Histological and Molecular Changes in Rotator Cuff Disease. *Bone and Joint Research* **1**, 158-166, (2012).

217 Chaudhury, S., Holland, C., Porter, D., Tirlapur, U. K., Vollrath, F. *et al.* Torn Human Rotator Cuff Tendons Have Reduced Collagen Thermal Properties on Differential Scanning Calorimetry. *Journal of Orthopaedic Research* **29**, 1938-1943, (2011).

218 Sher, J. S., Uribe, J. W., Posada, A., Murphy, B. J. & Zlatkin, M. B. Abnormal Findings on Magnetic Resonance Images of Asymptomatic Shoulders. *Journal of Bone & Joint Surgery - American Volume* **77**, 10-15, (1995).

219 Linsell, L., Dawson, J., Zondervan, K., Rose, P., Randall, T. *et al.* Prevalence and Incidence of Adults Consulting for Shoulder Conditions in UK Primary Care; Patterns of Diagnosis and Referral. *Rheumatology* **45**, 215-221, (2006).

220 Berry, H., Fernandes, L., Bloom, B., Clark, R. J. & Hamilton, E. B. D. Clinical Study Comparing Acupuncture, Physiotherapy, Injection and Oral Anti-Inflammatory Therapy in Shoulder-Cuff Lesions. *Current Medical Research and Opinion* **7**, 121-126, (1980).

221 Chaudhury, S., Holland, C., Thompson, M. S., Vollrath, F. & Carr, A. J. Tensile and Shear Mechanical Properties of Rotator Cuff Repair Patches. *Journal of Shoulder and Elbow Surgery* **21**, 1168-1176, (2012).

222 Mitchell, C., Adebajo, A., Hay, E. & Carr, A. Shoulder Pain: Diagnosis and Management in Primary Care. *British Medical Journal* **331**, 1124-1128, (2005).

223 Harryman, D. T. n., Mack, L. A., Wang, K. Y., Jackins, S. E., Richardson, M. L. *et al.* Repairs of the Rotator Cuff. Correlation of Functional Results With Integrity of the Cuff. *Journal of Bone and Joint Surgery (American Volume)* **73**, 982-989, (1991).
224 Carr, A. J., Rees, J. L., Ramsay, C. R., Fitzpatrick, R., Gray, A. *et al.* Protocol for the United Kingdom Rotator Cuff Study (UKUFF) - A Randomised Controlled Trial of Open and Arthroscopic Rotator Cuff Repair. *Bone and Joint Research* **3**, 155-160, (2014).
225 Molony, J. *A Torn Rotator Cuff: You Don't Want This*, http://m.mlb.com/news/article/2032977// (2007).
226 Carr, A., Cooper, C., Campbell, M. K., Rees, J., Moser, J. *et al.* Effectiveness of Open and Arthroscopic Rotator Cuff Repair (UKUFF). *The Bone & Joint Journal* **99-B**, 107-115, (2017).
227 Arsenault, P. R., Heaton-Johnson, K. J., Li, L.-s., Song, D., Ferreira, V. S. *et al.* Identification of Prolyl Hydroxylation Modifications in Mammalian Cell Proteins. *Proteomics* **15**, 1259-1267, (2015).
228 Montgomery, H., Rustogi, N., Hadjisavvas, A., Tanaka, K., Kyriacou, K. *et al.* Proteomic Profiling of Breast Tissue Collagens and Site-specific Characterization of Hydroxyproline Residues of Collagen Alpha-1-(I). *Journal of Proteome Research* **11**, 5890-5902, (2012).
229 Raghavan, M., McCullagh, J. S. O., Lynnerup, N. & Hedges, R. E. M. Amino Acid $\delta^{13}C$ Analysis of Hair Proteins and Bone Collagen Using Liquid Chromatography/Isotope Ratio Mass Spectrometry: Paleodietary Implications from Intra-Individual Comparisons. *Rapid Communications in Mass Spectrometry* **24**, 541-548, (2010).
230 Bank, R. A., TeKoppele, J. M., Oostingh, G., Hazleman, B. L. & Riley, G. P. Lysylhydroxylation and Non-Reducible Crosslinking of Human Supraspinatus Tendon Collagen: Changes with Age and in Chronic Rotator Cuff Tendinitis. *Annals of the Rheumatic Diseases* **58**, 35-41, (1999).
231 Bank, R. A., Beekman, B., Tenni, R. & TeKoppele, J. M. Pre-Column Derivatisation Method for the Measurement of Glycosylated Hydroxylysines of Collagenous Proteins. *Journal of Chromatography B: Biomedical Sciences and Applications* **703**, 267-272, (1997).
232 Gumustas, M., Kurbanoglu, S., Uslu, B. & Ozkan, S. A. UPLC versus HPLC on Drug Analysis: Advantageous, Applications and Their Validation Parameters. *Chromatographia* **76**, 1365-1427, (2013).
233 Nakajima, T. & Volcani, B. E. 3,4-Dihydroxyproline: A New Amino Acid in Diatom Cell Walls. *Science* **164**, 1400-1401, (1969).
234 Buku, A., Faulstich, H., Wieland, T. & Dabrowski, J. 2,3-Trans-3,4-Trans-3,4-Dihydroxy-L-proline: An Amino Acid in Toxic Peptides of *Amanita virosa* Mushrooms. *Proceedings of the National Academy of Sciences* **77**, 2370-2371, (1980).
235 Edrissi, B., Taghizadeh, K. & Dedon, P. C. Quantitative Analysis of Histone Modifications: Formaldehyde Is a Source of Pathological N^6-Formyllysine That Is Refractory to Histone Deacetylases. *PLoS Genetics* **9**, e1003328, (2013).
236 Risteli, J., Tryggvason, K. & Kivirikko, K. I. A Rapid Assay for Prolyl 3-Hydroxylase Activity. *Analytical Biochemistry* **84**, 423-431, (1978).
237 Cummings, T. *Chemical Biology Studies of an Unusual Ribosomal Protein Dihydroxylation* MChem thesis, University of Oxford, (2015).
238 Jabaiah, A., Wang, X., Raman, S. K., Ragan, R., Da Silva, N. A. *et al.* Nanoscale Architecture and Cellular Adhesion of Biomimetic Collagen Substrates. *Journal of Biomaterials Applications* **28**, 1354-1365, (2014).
239 Pokidysheva, E., Zientek, K. D., Ishikawa, Y., Mizuno, K., Vranka, J. A. *et al.* Posttranslational Modifications in Type I Collagen from Different Tissues Extracted from Wild Type and Prolyl 3-Hydroxylase 1 Null Mice. *Journal of Biological Chemistry* **288**, 24742-24752, (2013).

240 Walton, R. S. & Czernuszka, J. T. Recent Developments in Collagen Patents. *Recent Patents on Regenerative Medicine* **2**, 1-10, (2012).
241 Tripp, J. A. & McCullagh, J. S. O. in *Amino Acid Analysis* Vol. 828 *Methods in Molecular Biology* (eds Alterman, M. A. & Hunziker, P.) Ch. 27, 339-350 (Humana Press, 2012).
242 Schwarz, S., Koerber, L., Elsaesser, A. F., Goldberg-Bockhorn, E., Seitz, A. M. *et al.* Decellularized Cartilage Matrix as a Novel Biomatrix for Cartilage Tissue-Engineering Applications. *Tissue Engineering, Part A* **18**, 2195-2209, (2012).
243 Gómez, K. K., Del Prado, M. L., Piña, M. C. & García de León, M. C. Extraction and Characterization of Collagen from Different Biological Tissues. *AIP Conference Proceedings* **1494**, 149-151, (2012).
244 Pacak, C. A., Powers, J. M. & Cowan, D. B. Ultrarapid Purification of Collagen Type I for Tissue Engineering Applications. *Tissue Engineering, Part C* **17**, 879-885, (2011).
245 Eyre, D. R., Weis, M., Hudson, D. M., Wu, J.-J. & Kim, L. A Novel 3-Hydroxyproline (3Hyp)-rich Motif Marks the Triple-helical *C* Terminus of Tendon Type I Collagen. *Journal of Biological Chemistry* **286**, 7732-7736, (2011).
246 Breiter, R., Koerber, L. & Schwarz, S. Method for Preparing a Highly Pure Collagen and its Use as an Implant. EP2236544A1 (2010).
247 Xiong, X., Ghosh, R., Hiller, E., Drepper, F., Knapp, B. *et al.* A New Procedure for Rapid, High Yield Purification of Type I Collagen for Tissue Engineering. *Process Biochemistry* **44**, 1200-1212, (2009).
248 Rajan, N., Habermehl, J., Cote, M.-F., Doillon, C. J. & Mantovani, D. Preparation of Ready-to-Use, Storable and Reconstituted Type I Collagen from Rat Tail Tendon for Tissue Engineering Applications. *Nature Protocols* **1**, 2753-2758, (2007).
249 Yu, J.-C., Jang, J.-D., Chang, C.-H., Lee, S.-B., Yeo, S.-G. *et al.* Separation and Purification of Collagen from Animal Tissues. WO2006096027A1 (2006).
250 Mayne, J. & Robinson, J. J. Comparative Analysis of the Structure and Thermal Stability of Sea Urchin Peristome and Rat Tail Tendon Collagen. *Journal of Cellular Biochemistry* **84**, 567-574, (2002).
251 Abraham, G. A., Carr, R. M., Jr., Kemp, P. D., Mercer, R. & Baker, L. Chemical Cleaning of Biological Material. WO9849969A1 (1998).
252 Nimni, M. E. & Cheung, D. T. Process for Purifying Collagen and Generating Bioprosthesis. US5374539A (1994).
253 Monera, O. D., Kay, C. M. & Hodges, R. S. Protein Denaturation with Guanidine Hydrochloride or Urea Provides a Different Estimate of Stability Depending on the Contributions of Electrostatic Interactions. *Protein Science* **3**, 1984-1991, (1994).
254 Burke, R. D., Bouland, C. & Sanderson, A. I. Collagen Diversity in the Sea Urchin, *Strongylocentrotus purpuratus*. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **94**, 41-44, (1989).
255 Ramshaw, J. A. M., Bateman, J. F. & Cole, W. G. Precipitation of Collagens by Polyethylene Glycols. *Analytical Biochemistry* **141**, 361-365, (1984).
256 Murray, L. W., Herbert Waite, J., Tanzer, M. L. & Hauschka, P. V. in *Methods in Enzymology* Vol. Volume 82 (ed Leon W. Cunningham, D. W. F.) 65-96 (Academic Press, 1982).
257 Miller, E. J. & Kent Rhodes, R. in *Methods in Enzymology* Vol. Volume 82 (ed Leon W. Cunningham, D. W. F.) 33-64 (Academic Press, 1982).
258 Trelstad, R. L., Catanese, V. M. & Rubin, D. F. Collagen Fractionation: Separation of Native Types I, II and III by Differential Precipitation. *Analytical Biochemistry* **71**, 114-118, (1976).
259 Longin, R. New Method of Collagen Extraction for Radiocarbon Dating. *Nature* **230**, 241-242, (1971).

260 Dakin, S. G., Martinez, F. O., Yapp, C., Wells, G., Oppermann, U. *et al.* Inflammation Activation and Resolution in Human Tendon Disease. *Science Translational Medicine* **7**, 311ra173-311ra173, (2015).
261 Langley, G. W., Brinkø, A., Munzel, M., Walport, L. J., Schofield, C. J. *et al.* Analysis of JmjC Demethylase-Catalyzed Demethylation Using Geometrically-Constrained Lysine Analogues. *ACS Chemical Biology* **11**, 755-762, (2016).
262 Langley, G. W. *Cyclic Peptide Probes for Histone Demethylases* MChem thesis, University of Oxford, (2013).
263 Langley, G. W., Brinkø, A., Münzel, M., Walport, L. J., Schofield, C. J. *et al.* Probing the Charge and Conformational Requirements of JmjC Demethylases. *The FASEB Journal* **31**, 767.761, (2017).
264 Xie, Z., Dai, J., Dai, L., Tan, M., Cheng, Z. *et al.* Lysine Succinylation and Lysine Malonylation in Histones. *Molecular & Cellular Proteomics* **11**, 100-107, (2012).
265 Herranz, N., Dave, N., Millanes-Romero, A., Morey, L., Díaz, Víctor M. *et al.* Lysyl Oxidase-like 2 Deaminates Lysine 4 in Histone H3. *Molecular Cell* **46**, 369-376, (2012).
266 Tan, M., Luo, H., Lee, S., Jin, F., Yang, Jeong S. *et al.* Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* **146**, 1016-1028, (2011).
267 Craig, J. W. N. C. *Epigenetics : A Reference Manual*. (Caister Academic Press, 2011).
268 Wiśniewski, J. R., Zougman, A. & Mann, M. N^{ϵ} -Formylation of Lysine is a Widespread Post-Translational Modification of Nuclear Proteins Occurring at Residues Involved in Regulation of Chromatin Function. *Nucleic Acids Research* **36**, 570-577, (2008).
269 Tost, J. *Epigenetics*. (Caister Academic Press, 2008).
270 Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B. *et al.* Lysine Propionylation and Butyrylation are Novel Post-translational Modifications in Histones. *Molecular & Cellular Proteomics* **6**, 812-819, (2007).
271 Allis, C. D., Jenuwein, T., Reinberg, D., Caparros, M.-L. & Editors. *Epigenetics*. (Cold Spring Harbor Laboratory Press, 2007).
272 Khoury, G. A., Baliban, R. C. & Floudas, C. A. Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database. **1**, 90, (2011).
273 Dupont, C., Armant, D. R. & Brenner, C. A. Epigenetics: Definition, Mechanisms and Clinical Perspective. *Seminars in Reproductive Medicine* **27**, 351-357, (2009).
274 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal Structure of the Nucleosome Core Particle at 2.8Å Resolution. *Nature* **389**, 251-260, (1997).
275 Hopkinson, R. J., Walport, L. J., Münzel, M., Rose, N. R., Smart, T. J. *et al.* Is JmjC Oxygenase Catalysis Limited to Demethylation? *Angewandte Chemie International Edition* **52**, 7709-7713, (2013).
276 Walport, L. J., Hopkinson, R. J., Chowdhury, R., Schiller, R., Ge, W. *et al.* Arginine Demethylation is Catalysed by a Subset of JmjC Histone Lysine Demethylases. *Nature Communications* **7**, 11974, (2016).
277 Ng, S. S., Kavanagh, K. L., McDonough, M. A., Butler, D., Pilka, E. S. *et al.* Crystal Structures of Histone Demethylase JMJD2A Reveal Basis for Substrate Specificity. *Nature* **448**, 87-91, (2007).
278 Mizusaki, K., Sugahara, Y., Tsunematsu, H. & Makisumi, S. The Conformation of the Lysyl Side Chain of Substrates at the Active Center of Trypsin. *Journal of Biochemistry* **100**, 21-25, (1986).
279 Hughes, R. M., Wiggins, K. R., Khorasanizadeh, S. & Waters, M. L. Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect. *Proceedings of the National Academy of Sciences* **104**, 11184-11188, (2007).

280 Garber, S. B., Kingsbury, J. S., Gray, B. L. & Hoveyda, A. H. Efficient and Recyclable Monomeric and Dendritic Ru-Based Metathesis Catalysts. *Journal of the American Chemical Society* **122**, 8168-8179, (2000).
281 Lizcano, J. M., Unzeta, M. & Tipton, K. F. A Spectrophotometric Method for Determining the Oxidative Deamination of Methylamine by the Amine Oxidases. *Analytical Biochemistry* **286**, 75-79, (2000).
282 Loenarz, C., Ge, W., Coleman, M. L., Rose, N. R., Cooper, C. D. O. *et al.* PHF8, a Gene Associated with Cleft Lip/Palate and Mental Retardation, Encodes for an N^{ϵ}-Dimethyl Lysine Demethylase. *Human Molecular Genetics* **19**, 217-222, (2010).
283 Rose, N. R., Woon, E. C. Y., Kingham, G. L., King, O. N. F., Mecinović, J. *et al.* Selective Inhibitors of the JMJD2 Histone Demethylases: Combined Nondenaturing Mass Spectrometric Screening and Crystallographic Approaches. *Journal of Medicinal Chemistry* **53**, 1810-1818, (2010).
284 Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B. *et al.* Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **333**, 1300-1303, (2011).
285 Gannon, J. & Schofield, C. J. *Unpublished Results* (University of Oxford, 2017).
286 Plotnikov, Alexander N., Yang, S., Zhou, Thomas J., Rusinova, E., Frasca, A. *et al.* Structural Insights into Acetylated-Histone H4 Recognition by the Bromodomain-PHD Finger Module of Human Transcriptional Coactivator CBP. *Structure* **22**, 353-360, (2014).
287 Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 Chromodomain Bound to a Lysine 9-Methylated Histone H3 Tail. *Science* **295**, 2080-2083, (2002).
288 Li, J., Li, Z., Ruan, J., Xu, C., Tong, Y. *et al.* Structural Basis for Specific Binding of Human MPP8 Chromodomain to Histone H3 Methylated at Lysine 9. *PloS One* **6**, e25104, (2011).
289 Yang, M., Culhane, J. C., Szewczuk, L. M., Gocke, C. B., Brautigam, C. A. *et al.* Structural Basis of Histone Demethylation by LSD1 Revealed by Suicide Inactivation. *Nature Structural & Molecular Biology* **14**, 535-539, (2007).
290 Gatchalian, J., Ali, M., Andrews, F. H., Zhang, Y., Barrett, A. S. *et al.* Structural Insight into Recognition of Methylated Histone H3K4 by Set3. *Journal of Molecular Biology* **429**, 2066-2074, (2017).
291 Ziering, D. L. & Pascal, R. A. Mechanism-Based Inhibition of Bacterial γ-Butyrobetaine Hydroxylase. *Journal of the American Chemical Society* **112**, 834-838, (1990).
292 Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R. *et al.* in *The Proteomics Protocols Handbook* (ed Walker, J. M.) 571-607 (Humana Press, 2005).
293 ProteinProspector v. 5.20.0 (2017).
294 Aguilar, J. A., Nilsson, M., Bodenhausen, G. & Morris, G. A. Spin Echo NMR Spectra Without J Modulation. *Chemical Communications* **48**, 811-813, (2012).
295 Niesen, F. H., Berglund, H. & Vedadi, M. The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions that Promote Protein Stability. *Nature Protocols* **2**, 2212-2221, (2007).
296 Muller, T. A. & Hausinger, R. P. in *2-Oxoglutarate-Dependent Oxygenases* (eds Schofield, C. J. & Hausinger, R. P.) 246-262 (The Royal Society of Chemistry, 2015).
297 Zheng, G. & He, C. in *2-Oxoglutarate-Dependent Oxygenases* 263-274 (The Royal Society of Chemistry, 2015).
298 Falnes, P. O. & Yen Ho, A. Y. in *2-Oxoglutarate-Dependent Oxygenases* (eds Schofield, C. J. & Hausinger, R. P.) 275-288 (The Royal Society of Chemistry, 2015).
299 Cheng, X. & Trievel, R. C. in *2-Oxoglutarate-Dependent Oxygenases* (eds Schofield, C. J. & Hausinger, R. P.) 210-245 (The Royal Society of Chemistry, 2015).
300 Chowdhury, R., Sekirnik, R., Brissett, N. C., Krojer, T., Ho, C.-h. *et al.* Ribosomal Oxygenases are Structurally Conserved from Prokaryotes to Humans. *Nature* **510**, 422-426, (2014).

301 Koivunen, P., Tiainen, P., Hyvärinen, J., Williams, K. E., Sormunen, R. *et al.* An Endoplasmic Reticulum Transmembrane Prolyl 4-Hydroxylase Is Induced by Hypoxia and Acts on Hypoxia-Inducible Factor α. *Journal of Biological Chemistry* **282**, 30544-30552, (2007).

302 Watt, S. A., Dayal, J. H. S., Wright, S., Riddle, M., Pourreyron, C. *et al.* Lysyl Hydroxylase 3 Localizes to Epidermal Basement Membrane and Is Reduced in Patients with Recessive Dystrophic Epidermolysis Bullosa. *PloS One* **10**, e0137639, (2015).

303 Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: Biological Roles of TET Proteins in DNA Demethylation and Transcription. *Nature Reviews Molecular Cell Biology* **14**, 341-356, (2013).

304 Pavlyuk, O., Teller, H. & McMills, M. C. An efficient synthesis of nitrogen-containing heterocycles via a tandem carbenoid N–H insertion/ring-closing metathesis sequence. *Tetrahedron Letters* **50**, 2716-2718, (2009).

305 Smith, V., Hongo, J.-A. S., Stinson, S. & Hazen, M. Neuronal leucine-rich repeat-1 antagonists for cancer therapy. WO2009046123A2 (2009).

306 Kapust, R. B., Tözsér, J., Fox, J. D., Anderson, D. E., Cherry, S. *et al.* Tobacco Etch Virus Protease: Mechanism of Autolysis and Rational Design of Stable Mutants with Wild-Type Catalytic Proficiency. *Protein Engineering, Design and Selection* **14**, 993-1000, (2001).

Appendix 1
Human 2OG Oxygenase Names and Reactions

Name Synonyms	Full Name(s)	Reaction Catalysed
ALKBH1 ABH ABH1 ALKBH	Nucleic acid dioxygenase ALKBH1 Alkylated DNA repair protein alkB homologue 1 α-ketoglutarate-dependent dioxygenase ABH1 DNA 6mA demethylase DNA N^6-methyl adenine demethylase DNA lyase ABH1 DNA oxidative demethylase ALKBH1 tRNA N^1-methyl adenine demethylase	Demethylation of 3-methyl cytosine[296].
ALKBH2 ABH2	DNA oxidative demethylase ALKBH2 Alkylated DNA repair protein alkB homologue 2 α-ketoglutarate-dependent dioxygenase alkB homologue 2 Oxy DC1	Demethylation of 1-methyl adenine and 3-methyl cytosine[296].
ALKBH3 ABH3 DEPC1 hABH3	α-ketoglutarate-dependent dioxygenase alkB homologue 3 Alkylated DNA repair protein alkB homologue 3 DEPC-1 Prostate cancer antigen 1	Demethylation of 1-methyl adenine and 3-methyl cytosine[296].
ALKBH4 ABH4	α-ketoglutarate-dependent dioxygenase alkB homologue 4 Alkylated DNA repair protein alkB homologue 4	
ALKBH5 ABH5 OFOXD1	RNA demethylase ALKBH5 Alkylated DNA repair protein alkB homologue 5 α-ketoglutarate-dependent dioxygenase alkB homologue 5	Demethylation of N^6-methyl adenine[296,297].
ALKBH6 ABH6	α-ketoglutarate-dependent dioxygenase alkB homologue 6 Alkylated DNA repair protein alkB homologue 6	
ALKBH7 ABH7 SPATA11	α-ketoglutarate-dependent dioxygenase alkB homologue 7 Alkylated DNA repair protein alkB homologue 7 Spermatogenesis cell proliferation-related protein Spermatogenesis-associated protein 11	
ALKBH8 ABH8	Alkylated DNA repair protein alkB homologue 8 Probable α-ketoglutarate-dependent dioxygenase ABH8 *S*-adenosyl-L-methionine-dependent tRNA methyltransferase ABH8 tRNA (carboxymethyluridine(3,4)-5-*O*)-methyltransferase ABH8	5-methoxycarbonyl methyluridine-(*S*)-hydroxylation[298].
ASPH BAH	Aspartyl/asparaginyl β-hydroxylase Aspartate β-hydroxylase Peptide-aspartate β-dioxygenase	Formation of (*2S,3R*)-hydroxyaspartate and (*2S,3R*)-hydroxyasparagine.
BBOX1 BBH BBOX	ɣ-butyrobetaine dioxygenase ɣ-butyrobetaine hydroxylase ɣ-butyrobetaine,2-oxoglutarate dioxygenase	Formation of 3-hydroxy 4-trimethylammonio butanoate[193].
FIH HIF1AN FIH-1	Hypoxia-inducible factor 1-α inhibitor Factor inhibiting HIF-1 Hypoxia-inducible factor asparagines hydroxylase	Formation of (*2S,3S*)-hydroxyasparagine, (*2S,3S*)-hydroxyaspartate and (*2S,3S*)-hydroxyhistidine in HIF subunits and ankyrin repeats[123].

Name Synonyms	Full Name(s)	Reaction Catalysed
FTO KIAA1752	α-ketoglutarate-dependent dioxygenase FTO Fat mass and obesity-associated protein	Demethylation of N^6-methyl adenosine[180].
HSPBAP1 HSPBAP1	HSPB1-associated protein 1 27 kDa heat shock protein-associated protein 1 Protein associated with small stress protein 1	3-hydroxylysine formation.
JARID2 JMJ	Protein Jumonji Jumonji/ARID domain-containing protein 2	
JMJD1C JHDM2C KIAA1380 TRIP8	Probable JmjC domain-containing histone demethylation protein 2C Jumonji domain-containing protein 1C Thyroid receptor-interacting protein 8	
JMJD4	JmjC domain-containing protein 4 Jumonji domain-containing protein 4	Formation of 4-hydroxylysine[113].
JMJD5 KDM8	Lysine-specific demethylase 8 JmjC domain-containing protein 5 Jumonji domain-containing protein 5	Formation of (*2S,3R*)-hydroxyarginine.
JMJD6 KIAA0585 PTDSR	Bifunctional arginine demethylase and lysyl-hydroxylase JMJD6 Histone arginine demethylase JMJD6 JmjC domain-containing protein 6 Jumonji domain-containing protein 6 Lysyl-hydroxylase JMJD6 Peptide-lysine 5-dioxygenase JMJD6 Phosphatidylserine receptor	Formation of (2S,5S)-hydroxylysine[116]. JMJD6 has also been assigned, albeit controversially, as an arginine demethylase[118].
JMJD7	JmjC domain-containing protein 7	Formation of (2S,3S)-hydroxylysine[135].
JMJD8 C16orf20	JmjC domain-containing protein 8 Jumonji domain-containing protein 8	
KDM2A CXXC8 FBL7 FBXL11 JHDM1A KIAA1004	Lysine-specific demethylase 2A CXXC-type zinc finger protein 8 F-box and leucine-rich repeat protein 11 F-box protein FBL7 F-box protein Lilina F-box/LRR-repeat protein 11 JmjC domain-containing histone demethylation protein 1A [Histone-H3]-lysine-36 demethylase 1A	Demethylation of N^{ϵ}-dimethylated and monomethylated H3K36 residues[299].
KDM2B CXXC2 FBL10 FBXL10 JHDM1B PCCX2 JEMMA	Lysine-specific demethylase 2B CXXC-type zinc finger protein 2 F-box and leucine-rich repeat protein 10 F-box protein FBL10 F-box/LRR-repeat protein 10 JmjC domain-containing histone demethylation protein 1B Jumonji domain-containing EMSY-interactor methyltransferase motif protein Protein-containing CXXC domain 2 [Histone-H3]-lysine-36 demethylase 1B	Demethylation of N^{ϵ}-dimethylated and monomethylated H3K36 residues[299].

Name Synonyms	Full Name(s)	Reaction Catalysed
KDM3A JHDM2A JMJD1 JMJD1A KIAA0742 TSGA	Lysine-specific demethylase 3A JmjC domain-containing histone demethylation protein 2A Jumonji domain-containing protein 1A	Demethylation of N^{ϵ}-dimethylated and monomethylated H3K9 residues[299].
KDM3B C5orf7 JHDM2B JMJD1B KIAA1082	Lysine-specific demethylase 3B JmjC domain-containing histone demethylation protein 2B Jumonji domain-containing protein 1B Nuclear protein 5qNCA	Demethylation of N^{ϵ}-dimethylated and monomethylated H3K9 residues[299].
KDM4A JHDM3A JMJD2 JMJD2A KIAA0677	Lysine-specific demethylase 4A JmjC domain-containing histone demethylation protein 3A Jumonji domain-containing protein 2A	Demethylation of N^{ϵ}- tri- and di- methylated H3K9 and H3K36 residues[299].
KDM4B JHDM3B JMJD2B KIAA0876	Lysine-specific demethylase 4B JmjC domain-containing histone demethylation protein 3B Jumonji domain-containing protein 2B	Demethylation of N^{ϵ}- tri- and di- methylated H3K9 and H3K36 residues[299].
KDM4C GASC1 JHDM3C JMJD2C KIAA0780	Lysine-specific demethylase 4C Gene amplified in squamous cell carcinoma 1 protein JmjC domain-containing histone demethylation protein 3C Jumonji domain-containing protein 2C	Demethylation of N^{ϵ}- tri- and di- methylated H3K9 and H3K36 residues[299].
KDM4D JHDM3D JMJD2D	Lysine-specific demethylase 4D JmjC domain-containing histone demethylation protein 3D Jumonji domain-containing protein 2D	Demethylation of N^{ϵ}- tri- and di- methylated H3K9 residues[299].
KDM4E KDM4DL	Lysine-specific demethylase 4E KDM4D-like protein Lysine-specific demethylase 4D-like	Demethylation of N^{ϵ}- tri- and di- methylated H3K9 residues[299] (psuedogene).
KDM5A JARID1A RBBP2 RBP2	Lysine-specific demethylase 5A Histone demethylase JARID1A Jumonji/ARID domain-containing protein 1A Retinoblastoma-binding protein 2	Demethylation of N^{ϵ}- tri- and di- methylated H3K4 residues[299].
KDM5B JARID1B PLU1 CT31 RBBP2H1	Lysine-specific demethylase 5B Cancer/testis antigen 31 Histone demethylase JARID1B Jumonji/ARID domain-containing protein 1B PLU-1 Retinoblastoma-binding protein 2 homologue 1	Demethylation of N^{ϵ}- tri- and di- methylated H3K4 residues[299].
KDM5C DXS1272E JARID1C SMCX XE169	Lysine-specific demethylase 5C Histone demethylase JARID1C Jumonji/ARID domain-containing protein 1C Protein SmcX Protein Xe169	Demethylation of N^{ϵ}- tri- and di- methylated H3K4 residues[299].

Name Synonyms	Full Name(s)	Reaction Catalysed
KDM5D HY HYA JARID1D KIAA0234 SMCY	Lysine-specific demethylase 5D Histocompatibility Y antigen Histone demethylase JARID1D Jumonji/ARID domain-containing protein 1D Protein SmcY	Demethylation of N^{ϵ}- tri- and di- methylated H3K4 residues[299].
KDM6A UTX	Lysine-specific demethylase 6A Histone demethylase UTX Ubiquitously-transcribed TPR protein on the X chromosome Ubiquitously-transcribed X chromosome tetratricopeptide repeat protein	Demethylation of N^{ϵ}- tri- and di- methylated H3K27 residues[299].
KDM6B JMJD3 KIAA0346	Lysine-specific demethylase 6B JmjC domain-containing protein 3 Jumonji domain-containing protein 3 Lysine demethylase 6B	Demethylation of N^{ϵ}- tri- and di- methylated H3K27 residues[299].
KDM7A JHDM1D KDM7 KIAA1718	Lysine-specific demethylase 7A JmjC domain-containing histone demethylation protein 1D Lysine-specific demethylase 7	Demethylation of N^{ϵ}- di- and mono- methylated H3K27 residues[299].
KDM7B PHF8 KIAA1111 ZNF422	Histone lysine demethylase PHF8 PHD finger protein 8	Demethylation of N^{ϵ}- di- and mono- methylated H3K9 residues[299].
MINA53 MINA MDIG MINA53 NO52 ROX	Bifunctional lysine-specific demethylase and histidyl-hydroxylase MINA 60S ribosomal protein L27a histidine hydroxylase Histone lysine demethylase MINA MYC-induced nuclear antigen Mineral dust-induced gene protein Nucleolar protein 52 Ribosomal oxygenase MINA	Formation of (*2S,3S*)-hydroxyhistidine in RPL27A[300].
NO66 C14orf169 MAPJD hsNO66 ROX	Bifunctional lysine-specific demethylase and histidyl-hydroxylase NO66 60S ribosomal protein L8 histidine hydroxylase Histone lysine demethylase NO66 Myc-associated protein with JmjC domain Nucleolar protein 66 Ribosomal oxygenase NO66	Formation of (*2S,3S*)-hydroxyhistidine in RPL8[300].
OGFOD1 KIAA1612 TPA1	Prolyl 3-hydroxylase OGFOD1 2-oxoglutarate and iron-dependent oxygenase domain-containing protein 1 Termination and polyadenylation 1 homologue	Formation of (*2S,3R*)-hydroxyproline in RPS23[121].
OGFOD3 C17orf101	2-oxoglutarate and iron-dependent oxygenase domain-containing protein 3	
P3H1 GROS1 LEPRE1 PSEC0109	Prolyl 3-hydroxylase 1 Growth suppressor 1 Leucine- and proline-enriched proteoglycan 1 Leprecan-1	Formation of (*2S,3S*)-hydroxyproline in procollagen[104].
P3H2 LEPREL1 MLAT4	Prolyl 3-hydroxylase 2 Leprecan-like protein 1 Myxoid liposarcoma-associated protein 4	Formation of (*2S,3S*)-hydroxyproline in procollagen[108].

Name Synonyms	Full Name(s)	Reaction Catalysed
P3H3 LEPREL2	Prolyl 3-hydroxylase 3 Leprecan-like protein 2	Formation of (*2S,3S*)-hydroxyproline in procollagen[64].
P4HA1 P4HA	Prolyl 4-hydroxylase subunit α-1 Procollagen-proline, 2-oxoglutarate-4-dioxygenase subunit α-1	Formation of (*2S,4R*)-hydroxyproline in procollagen[24].
P4HA2 UNQ290/PRO 330	Prolyl 4-hydroxylase subunit α-2 Procollagen-proline,2-oxoglutarate-4-dioxygenase subunit α-2	Formation of (*2S,4R*)-hydroxyproline in procollagen[24].
P4HA3 UNQ711/PRO 1374	Prolyl 4-hydroxylase subunit α-3 Procollagen-proline,2-oxoglutarate-4-dioxygenase subunit α-3	Formation of (*2S,4R*)-hydroxyproline in procollagen[24].
P4H-TM PH4	Transmembrane prolyl 4-hydroxylase Hypoxia-inducible factor prolyl hydroxylase 4	Formation of (*2S,4R*)-hydroxyproline in HIF α subunits[301]. This assignment is controversial.
PHD1 HPH-1 EGLN2 EIT6	Egl nine homologue 2 Estrogen-induced tag 6 HPH-3 Hypoxia-inducible factor prolyl hydroxylase 1 Prolyl hydroxylase domain-containing protein 1	Formation of (*2S,4R*)-hydroxyproline in HIF α subunits[123].
PHD2 HPH-2 EGLN1 C1orf12 PNAS-118 PNAS-137	Egl nine homologue 1 Hypoxia-inducible factor prolyl hydroxylase 2 Prolyl hydroxylase domain-containing protein 2 SM-20	Formation of (*2S,4R*)-hydroxyproline in HIF α subunits[112].
PHD3 HPH-3 EGLN3	Egl nine homologue 3 HPH-1 Hypoxia-inducible factor prolyl hydroxylase 3 Prolyl hydroxylase domain-containing protein 3	Formation of (*2S,4R*)-hydroxyproline in HIF α subunits[123].
PHF2 CENP-351 KIAA0662	Lysine-specific demethylase PHF2 GRC5 PHD finger protein 2	Demethylation of N^{ε}-methylated lysine residues[299].
PHYDH1	Phytanoyl-CoA dioxygenase domain-containing protein 1	
PHYH PAHX	Phytanoyl-CoA dioxygenase Phytanic acid oxidase Phytanoyl-CoA α-hydroxylase	Phytnoyl CoA hydroxylation.
PLOD1 LLH PLOD LH1	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1 Lysyl hydroxylase 1	Formation of (*2S,5R*)-hydroxylysine in procollagen[103].
PLOD2 LH2	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 Lysyl hydroxylase 2	Formation of (*2S,5R*)-hydroxylysine in procollagen[73].
PLOD3 LH3	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3 Lysyl hydroxylase 3	Formation of (*2S,5R*)-hydroxylysine in procollagen[302].

Name Synonyms	Full Name(s)	Reaction Catalysed
TET1 CXXC6 KIAA1676 LCX	Methylcytosine dioxygenase TET1 CXXC-type zinc finger protein 6 Leukemia-associated protein with a CXXC domain Ten-eleven translocation 1 gene protein	Demethylation of 5-methyl cytosine. Formation of 5-formylcytosine and 5-carboxylcytosine[284,303].
TET2 KIAA1546 Nbla00191	Methylcytosine dioxygenase TET2	Demethylation of 5-methyl cytosine. Formation of 5-formylcytosine and 5-carboxylcytosine[284,303].
TET3 KIAA0401	Methylcytosine dioxygenase TET3	Demethylation of 5-methyl cytosine. Formation of 5-formylcytosine and 5-carboxylcytosine[284,303].
TMLH TMLHE TMLD	Trimethyllysine dioxygenase, mitochondrial ε-trimethyllysine 2-oxoglutarate dioxygenase ε-trimethyllysine hydroxylase TML hydroxylase TML-α-ketoglutarate dioxygenase	Formation of (*2S,3S*)-hydroxy N^{ϵ}-trimethyllysine[134].
TYW5 C2orf60	tRNA wybutosine-synthesizing protein 5	Hydroxylation of yW-72, an intermediate in wybutosine biosynthesis.
UTY KDM6C	Histone demethylase UTY Ubiquitously-transcribed TPR protein on the Y chromosome Ubiquitously-transcribed Y chromosome tetratricopeptide repeat protein	

Table A 1 - Names and reactions of the human 2OG oxygenases.

Appendix 2
Synthetic Details for Lysine Analogues

Reprinted (adapted) with permission from Langley, G. W. *et al.* Analysis of JmjC Demethylase-Catalyzed Demethylation Using Geometrically-Constrained Lysine Analogues. *ACS Chem. Biol.* **11**, 755-762[261]. Copyright (2016) American Chemical Society.

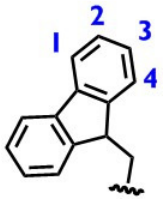
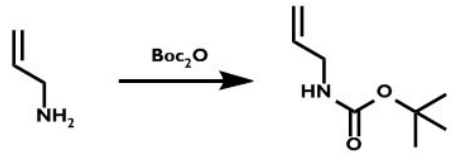


Figure A 1 - Numbering convention for aromatic Fmoc carbon atoms in NMR assignments.

Couplings are reported to the nearest 0.5 Hz and the following abbreviations were used to describe multiplicities; s = singlet, d = doublet, t = triplet, m = multiplet, dd = double doublet, td = triple doublet, tq = triple quartet.

General procedures for olefin cross-metathesis and olefin hydrogenation are given in Chapter 8.

tert-Butyl *N*-allylcarbamate, 56[304]

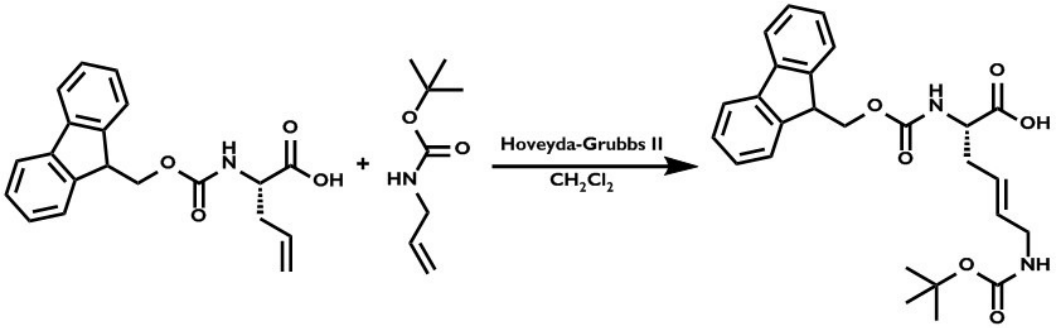


56

Triethylamine (2.40 mL, 17.2 mmol, 3.2 equiv.) was added to a solution of allyl amine hydrochloride (500 mg, 5.35 mmol, 1 equiv.) dissolved in dry CH_2Cl_2 (10 mL) under a N_2 atmosphere. The mixture was cooled to 0°C before Boc_2O (1.290 g, 5.89 mmol, 1.1 equiv.) was added in two portions. The reaction mixture was allowed to warm to 25°C and left to stir for 16 hours. When the reaction had run to completion, CH_2Cl_2 was evaporated *in vacuo* and the residue was purified by column chromatography (gradient of 0 to 20% ethyl acetate in cyclohexane). *tert*-Butyl *N*-allylcarbamate (**56**) was obtained as colourless oil (725 mg, 4.24 mmol, 79%); NMR and MS values were consistent with literature values[304].

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.81 (m, 1H, C*H*=CH_2), 5.11 (m, 2H, CH=C*H*$_2$), 4.67 (br s, 1H, N*H*), 3.72 (s, 2H, NHC*H*$_2$), 1.42 (s, 9H, 3 *x* Boc C*H*$_3$).
^{13}C NMR (101 MHz, $CDCl_3$) δ 155.9, 135.0, 115.8, 79.4, 43.2, 28.5
LRMS (Positive Ion ESI) $[M + Na]^+$ Theoretical *m/z* 180.1, Observed *m/z* 180.1

(2*S*,4*E*)-2-(((9*H*-fluoren-9-ylmethoxy)carbonyl)amino)-6-((((2-methyl-2-propanyl)oxy)carbonyl)amino)-4-hexenoic acid, 58

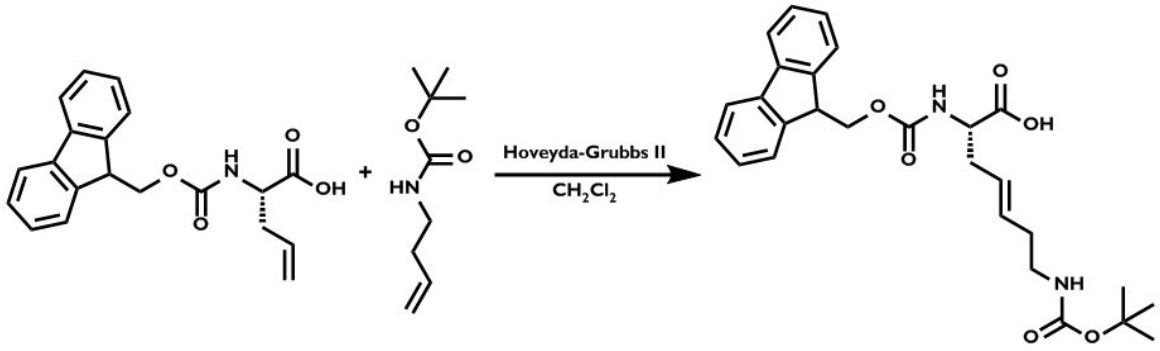


55 56 58

The standard metathesis procedure (Chapter 8) was applied, using **55** (1.000 g, 2.81 mmol, 1 equiv.) and *tert*-butyl *N*-allylcarbamate (**56**, 680 mg, 4.22 mmol, 1.5 equiv.) as the cross metathesis partner. Upon reaction completion (adjudged by TLC and LC-MS analyses), the reaction mixture was then concentrated *in vacuo*. The product was purified *via* column chromatography (33% to 50% EtOAc gradient in cyclohexane, 0.2% AcOH). Product fractions were combined and the solvent removed *in vacuo* to yield a brown oil. This oil was re-dissolved in $CHCl_3$ (10 mL) before solvent removal *in vacuo* to yield **58** as a brown solid (824 mg, 63%).

Melting Point 48°C – 50°C
$[\alpha]_D^{25}$ +24 (c = 0.01 g mL^{-1}, MeOH)
IR ν_{max}/cm^{-1} 1713, 2929, 2977
^{1}H NMR (500 MHz, *DMSO*-d_6) δ 8.22 (s, 1H, COO*H*), 7.87 (d, *J* = 7.5 Hz, 2H, Fmoc C_1*H*), 7.70 (d, *J* = 7.0 Hz, 2H, Fmoc C_4*H*), 7.42 (app t, *J* = 7.5 Hz, 2H, Fmoc C_2*H*), 7.34 (app t, *J* = 7.5 Hz, 2H, Fmoc C_3*H*), 6.42 (br. s., 1H, BocN*H*), 5.39 - 5.63 (m, 2H, Lys γ-C*H*, Lys δ-C*H*), 4.32 (d, *J* = 7.5 Hz, 2H, Fmoc-CH_2), 4.24 (t, *J* = 7.0 Hz, 1H, Fmoc*H*), 4.01 - 4.10 (m, 1H, Lys α-C*H*), 2.46 - 2.66 (m, 2H), 2.24 - 2.48 (m, 2H), 1.40 (s, 9H, 3 *x* Boc CH_3)
^{13}C NMR (101 MHz, *DMSO*-d_6) δ/ppm 173.3, 156.1, 155.5, 143.8, 140.7, 130.4, 127.7, 127.1, 126.3, 125.4, 120.2, 77.6, 65.7, 57.6, 54.0, 46.7, 41.6, 37.5, 33.6, 28.3
HRMS (Positive ion ESI) $[M + Na]^+$ Theoretical *m/z* 489.1996, Observed *m/z* 489.1997

(2*S*,4*E*)-2-(((9H-fluoren-9-ylmethoxy)carbonyl)amino)-7-((((2-methyl-2-propanyl)oxy)carbonyl)amino)-4-heptenoic acid, 59

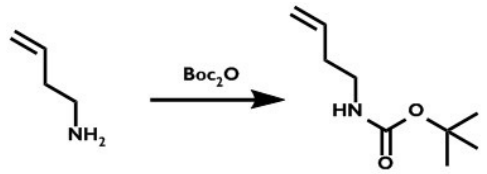


55 60 59

The standard metathesis procedure (Chapter 8) was applied using Hoveyda-Grubbs second generation catalyst (**57**, 30 mg, 0.0517 mmol, 0.15 equiv.), **55**, (116 mg, 0.345 mmol, 1 equiv.) and *tert*-butyl but-3-en-1-ylcarbamate, (**60**, 89 mg, 0.517 mmol, 1.5 equiv.). The mixture was purified by column chromatography (gradient of 5% to 40% ethyl acetate in cyclohexane + 1% formic acid), to give **59** as a brown oil (112 mg; 0.233 mmol, 68%). The product was then used directly for synthesis of **61** without further purification.

$[\alpha]_D^{25}$ = +58 (c = 0.50 g mL^{-1}, $CHCl_3$)
IR ν_{max}/cm^{-1} 1699, 2978
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.73 (br s, 2H, Fmoc $C_1\underline{H}$), 7.66-7.49 (m, 2H, Fmoc $C_4\underline{H}$), 7.51 – 7.16 (m, 4H, Fmoc $C_2\underline{H}$ and Fmoc $C_3\underline{H}$), 5.46 (br s, 2H, Lys γ-C*H*, Lys δ-C*H*), 4.91 – 4.06 (m, 4H, Fmoc*H*, Fmoc-C$\underline{H_2}$, Lys α-C*H*), 3.49 – 2.93 (m, 2H), 2.81 – 2.06 (m, 4H), 1.47 (m, 9H, 3 *x* Boc C$\underline{H_3}$).
^{13}C NMR (101 MHz, $CDCl_3$) δ 156.0, 143.8, 141.4, 132.2, 127.8, 127.1, 125.2, 120.0, 81.4, 66.9, 53,1, 47.2, 41.2, 36.2, 32.6, 28.5
HRMS (Positive Ion ESI) $[M+H]^+$ Theoretical *m/z* 481.23331, Observed *m/z* 481.23211

2-methyl-2-propanyl 3-buten-1-ylcarbamate, 60

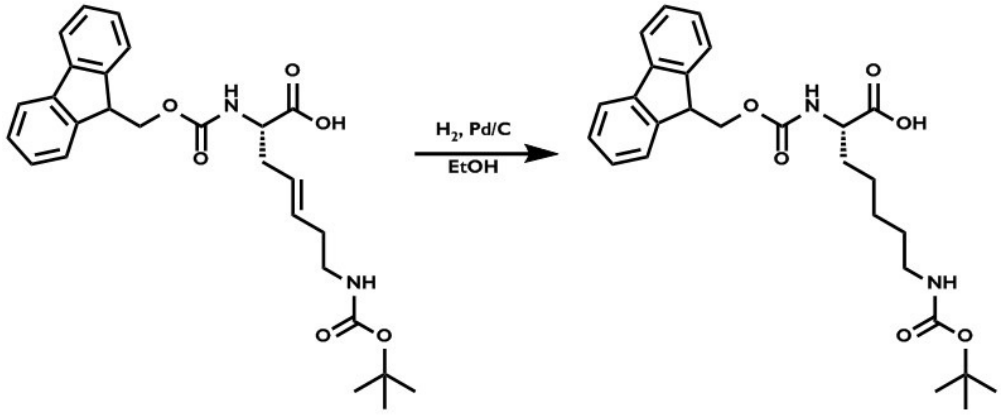


60

3-Butenylamine (0.39 mL, 4.22 mmol, 1 equiv.) and triethylamine (1.30 mL, 9.28 mmol, 2.2 equiv.) were dissolved in dry CH_2Cl_2 (10 mL) and cooled to 0°C under a N_2 atmosphere. Di-*t*-butyl dicarbonate (1.010 g, 4.64 mmol, 1.1 equiv.) was added in two portions and the reaction mixture was allowed to warm to 25°C and stirred for 16 hours. When the reaction had run to completion CH_2Cl_2 was evaporated *in vacuo* and the residue was purified by column chromatography (10% to 20% ethyl acetate in cyclohexane gradient). The product, **60**, was isolated as clear oil (713 mg, 4.16 mmol, 98%). NMR and MS values were consistent with literature[304].

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.74 (ddt, *J* = 17.0, 10.0, 7.0 Hz, 1H, C*H*=CH_2), 5.17 – 4.97 (m, 2H, CH=C*H*$_2$), 4.56 (br s, 1H, N*H*), 3.15 (dd, *J* = 12.5, 6.0 Hz, 2H, NHC*H*$_2$).2.23 (q, *J* = 7.0 Hz, 2H, $NHCH_2CH_2$), 1.43 (s, 9H, 3 *x* Boc C*H*$_3$)
^{13}C NMR (101 MHz, $CDCl_3$) δ 156.0, 135.4, 117.1, 79.1, 39.7, 34.3, 28.5
LRMS (Positive Ion ESI) $[M+H]^+$ Theoretical *m/z* 194.1, Observed *m/z* 194.1

(2*S*)-2-(((9*H*-fluoren-9-ylmethoxy)carbonyl)amino)-7-((((2-methyl-2-propanyl)oxy)carbonyl)amino)heptanoic acid, 61

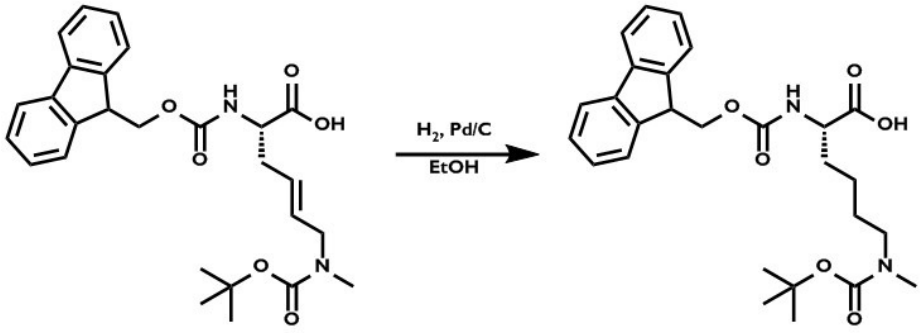


59 61

The general hydrogenation procedure (Chapter 8) was conducted on 59 (203 mg, 0.423 mmol) and **61** was obtained as light brown oil (189 mg, 0.392 mmol, 93%).

$[\alpha]_D^{25}$ = +18 (c = 0.50 g mL^{-1}, $CHCl_3$)
IR ν_{max}/cm^{-1} 1697, 2932, 3334
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.75 (d, *J* = 7.5 Hz, 2H, Fmoc $C_1\underline{H}$), 7.64 – 7.48 (m, 2H, Fmoc $C_4\underline{H}$), 7.38 (app t, *J* = 7.5 Hz, 2H, Fmoc $C_2\underline{H}$), 7.34 – 7.23 (m, 2H, Fmoc $C_3\underline{H}$), 4.53 – 4.21 (m, 4H, Fmoc$\underline{H}$, Fmoc-C$\underline{H_2}$ and Lys α-C$\underline{H}$), 3.09 (s, 2H), 1.88 – 1.55 (m, 2H), 1.56 – 1.13 (m, 15H)
^{13}C NMR (101 MHz, $CDCl_3$) δ 177.4, 156.3, 148.8, 144.0, 143.9, 141.4, 127.8, 127.2, 125.3, 120.1, 77.4, 67,1, 52.7, 47.3, 40.4, 32.3, 28.5, 26.3, 24.7
HRMS (Positive ion ESI) $[M + H]^+$ Theoretical *m/z* 483.24896, Observed *m/z* 483.24785

(2*S*)-2-(((9H-fluoren-9-ylmethoxy)carbonyl)amino)-6-(methyl(((2-methyl-2-propanyl)oxy)carbonyl)amino)hexanoic acid, 62

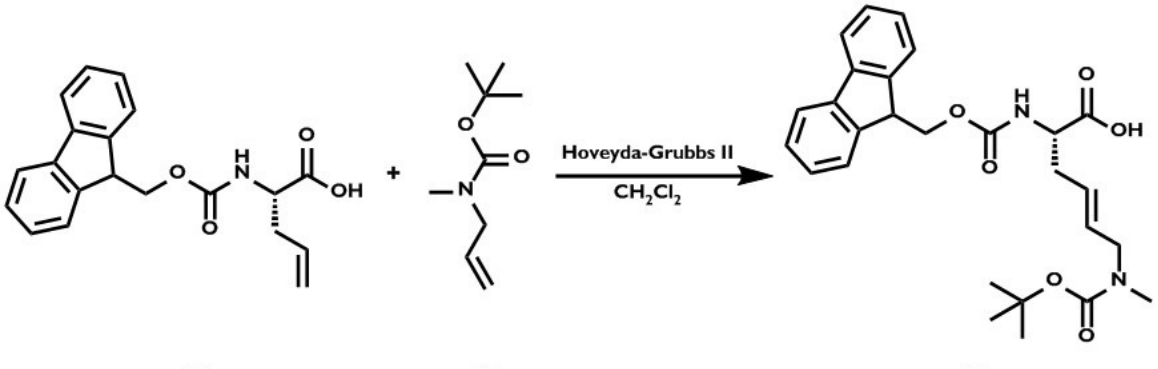


63 62

63 (56 mg, 0.210 mmol, 1 equiv.) was hydrogenated with H_2 (g) and Pd/C (10 mol%) according to the general hydrogenation procedure (Chapter 8). **62** was obtained as a light brown oil (53 mg, 0.109 mmol, 84%).

$[\alpha]_D^{25}$ = +4 (c = 0.50 g mL^{-1}, $CHCl_3$)
IR ν_{max}/cm^{-1} 1718, 2931
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.75 (d, *J* = 7.5 Hz, 2H, Fmoc $C_1\underline{H}$), 7.64 – 7.51 (m, 2H, Fmoc $C_4\underline{H}$), 7.38 – 7.13 (m, 4H, Fmoc $C_2\underline{H}$ and Fmoc $C_3\underline{H}$), 4.43 – 4.02 (m, 4H, Fmoc$\underline{H}$, Fmoc-C$\underline{H_2}$ and, Lys α-C$\underline{H}$), 3.21 – 3.02 (m, 2H), 2.78 (s, 3H, NC$\underline{H_3}$),1.44 (s, 15H).
^{13}C NMR (101 MHz, $CDCl_3$) δ 175.8, 156.3, 144.0, 143.9, 141.4, 127.8, 127.2, 125.3, 120.1, 79.9, 67.2, 54.0, 47.2, 34.3, 28.6
HRMS (Positive ion ESI) $[M + H]^+$ Theoretical *m/z* 483.24896, Observed *m/z* 483.24804

(2*S*,4*E*)-2-(((9H-fluoren-9-ylmethoxy)carbonyl)amino)-6-(methyl(((2-methyl-2-propanyl)oxy)carbonyl)amino)-4-hexenoic acid, 63

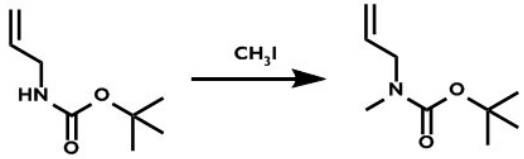


55 **64** **63**

According to the general metathesis procedure (Chapter 8), Hoveyda-Grubbs second generation catalyst (**57**, 20 mg, 0.0321 mmol, 0.15 equiv.) was reacted with **55** (72 mg, 0.214 mmol, 1 equiv.) and *tert*-butyl allyl(methyl)carbamate (**64**, 55 mg, 0.321 mmol, 1.5 equiv.). The product was purified by column chromatography (5% to 30% ethyl acetate in cyclohexane gradient with 1% formic acid) and isolated as greenish oil, **63** (47 mg, 0.0976 mmol, 46%).

$[\alpha]_D^{25}$= +22 (c = 0.50 g mL^{-1}, $CHCl_3$)
IR ν_{max}/cm^{-1} 1695, 2107, 2928
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.75 (d, *J* = 7.5 Hz, 2H, Fmoc $C_1\underline{H}$), 7.63 – 7.54 (m, 2H, Fmoc $C_4\underline{H}$), 7.39 (t, *J* = 7.5 Hz, 2H, Fmoc $C_2\underline{H}$), 7.30 (t, *J* = 7.5 Hz, 2H, Fmoc $C_3\underline{H}$), 5.68-38 (m, 2H, Overlapping Lys γ-C*H* and Lys δ-C*H*), 4.55 - 4.31 (m, 3H, Fmoc*H* and Fmoc-C$\underline{H_2}$), 4.29 – 4.12 (m, 1H, Lys α-C*H*), 3.89 – 3.60 (m, 2H), 2.79 (s, 3H, NC$\underline{H_3}$), 2.68 – 2.53 (m, 2H) 1.46 (s, 9H, 3 *x* Boc C$\underline{H_3}$)
13C NMR (101 MHz, CDCl3) δ 168.8, 156.2, 147.8, 143.8, 141.5, 127.8, 127.2, 125.2, 120.1, 80.2, 67.3, 53.6, 47.2, 35.1, 28.5
HRMS (Positive Ion ESI) [M+H]+ Theoretical *m/z* 481.23331, Observed *m/z* 481.23211

2-methyl-2-propanyl methyl(2-propen-1-yl)carbamate, 64

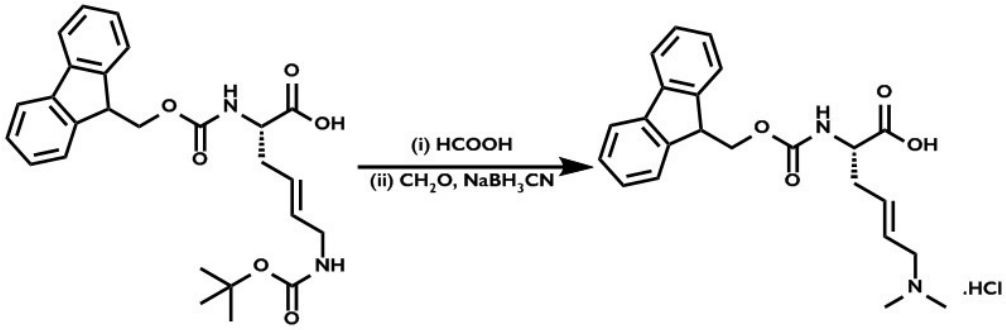


64

Powdered potassium hydroxide pellets (1.470 g, 26.6 mmol; 1.3 equiv.) was added to a solution of allylamine carbamate (3.290 g, 20.5 mmol, 1 equiv.) dissolved in dry DMSO (25 mL), under a N_2 atmosphere. The mixture was stirred for 10 minutes before methyl iodide (1.66 mL, 26.6 mmol, 1.3 equiv.) was added dropwise. The mixture was stirred for 16 hours, after which additional potassium hydroxide (517 mg) was added, and the solution was heated to 40°C. When the reaction had run to completion it was quenched by addition of saturated NH_4Cl (aq). The aqueous phase was extracted with ethyl acetate, and the combined organic phases were washed with brine, dried over magnesium sulfate, filtered and evaporated. The crude product, **64**, was isolated as yellow oil (2.650 g, 15.5 mmol, 74%) and used in the proceeding reaction without further purification. NMR and MS values were consistent with the literature values[305].

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.74 (ddd, *J* = 22.0, 11.0, 5.5 Hz, 1H, C*H*=CH_2), 5.22 – 5.00 (m, 2H, CH=C*H*$_2$), 3.79 (br s, 2H, NHC*H*$_2$), 2.80 (s, 3H, NHC*H*$_3$), 1.44 (s, 9H, 3 *x* Boc C*H*$_3$)
^{13}C NMR (101 MHz, $CDCl_3$) δ 155.8, 133.8, 79.5, 33.8, 28.5
LRMS (Positive Ion ESI) $[M + Na]^+$ Theoretical *m/z* 194.1, Observed *m/z* 194.1

(2*S*,4*E*)-6-(dimethylamino)-2-(((9*H*-fluoren-9-lmethoxy)carbonyl)amino)-4-hexenoic acid, 65

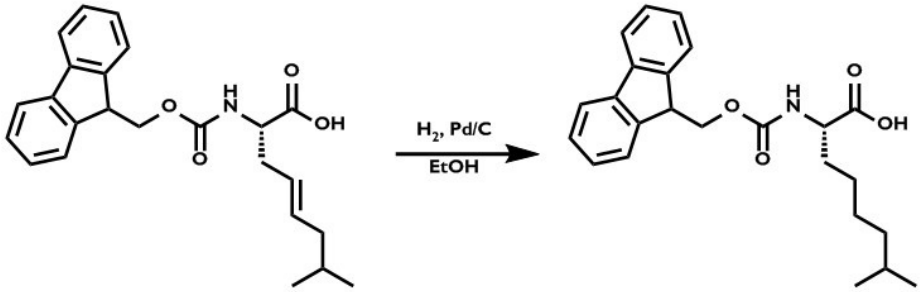


58 65

58 (600 mg, 1.29 mmol, 1 equiv.) was dissolved in formic acid (5 mL) and stirred at 25°C for 2 hours. Upon completion of this step (assessed *via* LC-MS), the formic acid was removed *in vacuo*. The residue was then dissolved in MeOH (15 mL) and formalin (37% formaldehyde in water solution, 0.27 mL, 5 equiv.) was added. $NaBH_3CN$ (8.3 mg, 2.71 mmol, 2.1 equiv.) was pre-dissolved in MeOH and was then added to the reaction mixture. The mixture was left to stir at 25°C for 4 hours. Concentrated HCl (1 mL) was added and the solvent was evaporated *in vacuo* and the product was re-dissolved in acetone (10 mL). After filtering, the solvent was removed *in vacuo* to yield **65** as a brown solid (495 mg, 89%), which was used directly in solid phase peptide synthesis.

Melting Point 97 °C – 103 °C
$[\alpha]_D^{25}$ +21 (c = 0.01 g mL^{-1}, MeOH)
IR ν_{max}/cm^{-1} 1697, 2953
^{1}H NMR (400 MHz, *CDCl*$_3$) δ/ppm 10.70 - 11.25 (br. s., 1 H, COO*H*), 7.68 - 7.76 (m, 2 H, Fmoc C_1*H*), 7.43 - 7.65 (m, 2 H, Fmoc C_4*H*), 7.32 - 7.41 (m, 2 H, Fmoc C_2*H*), 7.22 - 7.32 (m, 2 H, Fmoc C_2*H*), 6.40 (d, *J*=6.0 Hz, 1 H, α-C*H*), 6.06 - 5.63 (m, 2 H, overlapping δ-C*H* and γ-C*H*), 4.40 - 4.55 (m, 1 H, ε-CH_2'), 4.37 (d, *J*=10.0 Hz, 1 H, ε-CH_2''), 4.20 - 4.32 (m, 1 H, β-CH_2'), 4.06 - 4.20 (m, 1 H, β-CH_2''), 3.69 - 3.83 (m, 1 H, Fmoc*H*), 3.42 - 3.68 (m, 2 H, Fmoc-CH_2), 2.59 - 2.87 (m, 6 H, 2 *x* N-CH_3)
^{13}C NMR (101 MHz, *CDCl*$_3$) δ/ppm 173.0, 156.0, 143.5, 141.1, 137.3, 127.7, 127.0, 125.1, 122.0, 119.9, 119.9, 67.0, 59.1, 53.2, 52.7, 46.9, 42.3, 41.9, 35.3
HRMS (Positive ion ESI) $[M + Na]^+$ Theoretical *m/z* 395.1965, Observed *m/z* 395.1948

(2*S*)-2-(((9H-fluoren-9-ylmethoxy)carbonyl)amino)-7-methyloctanoic acid, 66

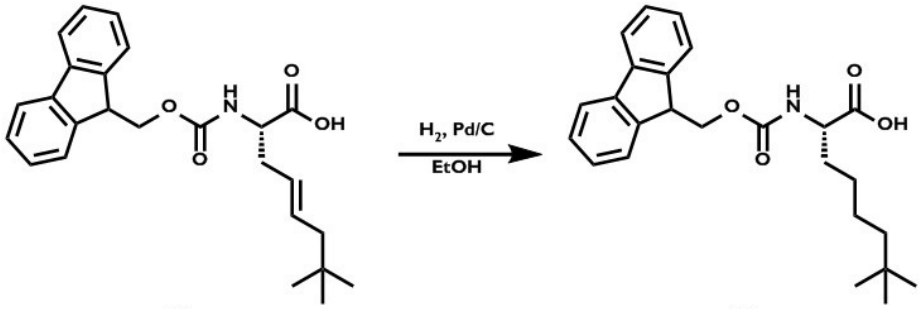


70 **66**

The general olefin hydrogenation procedure (Chapter 8) was conducted on **70** (300 mg, 0.760 mmol) to yield **66** as a white solid (290 mg, 96%).

Melting Point 43°C – 47°C
$[\alpha]_D^{25}$ +93 (c = 0.01 g mL^{-1}, MeOH)
IR ν_{max}/cm^{-1} 1714, 2955
^{1}H NMR (400 MHz, *DMSO-d*$_6$) δ/ppm 11.46 (br. s., 1 H), 7.90 (d, *J*=7.5 Hz, 1 H, Fmoc C$_1$*H*), 7.74 (d, *J*=7.0 Hz, 2 H, Fmoc C$_4$*H*), 7.39 - 7.46 (m, 2 H Fmoc C$_2$*H*), 7.29 - 7.37 (m, 2 H, Fmoc C$_3$*H*), 5.54 (br. s., 1 H, N*H*), 4.40 - 4.60 (m, 1 H, Fmoc*H*), 4.19 - 4.35 (m, 2 H, FmocC*H*$_2$), 3.90 - 4.00 (m, 1 H, α-C*H*), 1.57 - 1.74 (m, 2 H, β-C*H*$_2$), 1.44 - 1.55 (m, 1 H, ζ-C*H*), 1.20 - 1.39 (m, 4 H, Overlapping γ-C*H*$_2$, δ-C*H*$_2$), 1.09 - 1.19 (m, 2 H, ε-C*H*$_2$), 0.85 (d, *J*=6.5 Hz, 6 H, 2 *x* η-C*H*$_3$)
^{13}C NMR (101 MHz, *DMSO-d*$_6$) δ/ppm 174.5, 141.2, 128.7, 128.1, 127.5, 125.8, 123.8, 120.6, 66.1, 54.3, 47.1, 38.7, 31.3, 27.8, 26.7, 26.2, 23.0
HRMS (Positive ion ESI) [M + Na]$^+$ Theoretical *m/z* 418.1989, Observed *m/z* 418.1975

(2*S*)-2-(((9H-fluoren-9-ylmethoxy)carbonyl)amino)-7,7-dimethyloctanoic acid, 67

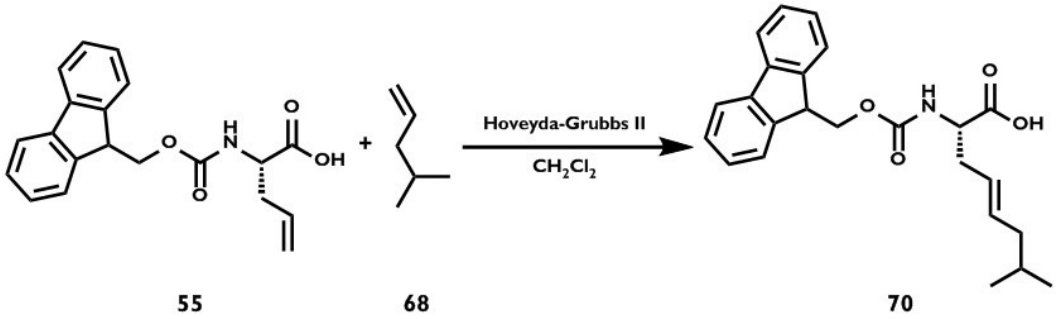


70 66

The general hydrogenation procedure (Chapter 8) was conducted using **71** (300 mg, 0.740 mmol) as the staring material to yield **67** as an off-white solid (294 mg, 97%). NMR and MS values were consistent with literature[279].

Melting Point 45°C – 48°C
$[\alpha]_D^{25}$ +73 (c = 0.01 g mL^{-1}, MeOH)
IR ν_{max}/cm^{-1} 1714, 2955
^{1}H NMR (400 MHz, *DMSO-d*$_6$) δ/ppm 7.89 (d, *J*=7.5 Hz, 2 H, Fmoc $C_1\underline{H}$), 7.74 (dd, *J*=7.5, 2.7 Hz, 2 H, Fmoc $C_4\underline{H}$), 7.64 (d, *J*=8.0 Hz, 1 H, N$\underline{H}$), 7.42 (t, *J*=7.0 Hz, 2 H, Fmoc $C_2\underline{H}$), 7.33 (t, *J*=7.5 Hz, 2 H, Fmoc $C_3\underline{H}$), 4.19 - 4.35 (m, 3 H, Overlapping Fmoc$\underline{H}$ and Fmoc-C$\underline{H_2}$), 3.95 (td, *J*=9.0, 4.5 Hz, 1 H, α-C$\underline{H}$), 1.57 - 1.79 (m, 2 H, β-C$\underline{H_2}$), 1.10 - 1.38 (m, 6 H, Overlapping γ-C$\underline{H_2}$, δ-C$\underline{H_2}$ and ε-C$\underline{H_2}$), 0.84 (s, 9 H, 3 *x* η-C$\underline{H_3}$)
^{13}C NMR (101 MHz, *DMSO-d*$_6$) δ/ppm 174.1, 156.2, 143.9, 143.8, 140.7, 127.6, 127.1, 125.3, 120.1, 65.6, 53.8, 46.7, 43.5, 30.9, 30.0, 29.2, 26.5, 23.6
HRMS (Positive ion ESI) $[M + Na]^+$ Theoretical *m/z* 432.2145, Observed *m/z* 432.2128

(2*S*,4*E*)-2-(((9H-fluoren-9-ylmethoxy)carbonyl)amino)-7-methyl-4-octenoic acid, 70

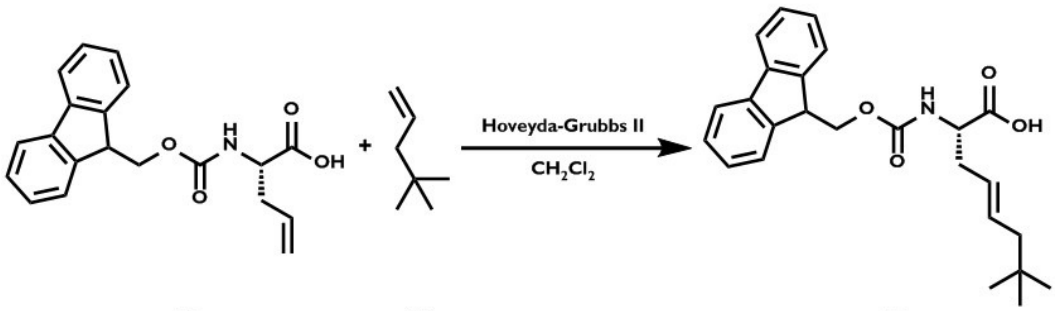


55 68 70

The standard metathesis procedure (Chapter 8) was applied using **55** (1.000 g) and 4-methyl-1-pentene, (**68**, 620 µL, 423 mg, 2.11 mmol, 1.5 equiv.) as the cross metathesis partner. Upon completion, the reaction mixture was then concentrated *in vacuo* before the product was separated from the crude mixture by column chromatography (0 to 33% EtOAc in Cyclohexane gradient, 0.2% AcOH). Catalyst (**57**, 75 mg) was recovered and the product, **70**, was isolated as an off-white solid (798 mg, 72%).

Melting Point 113 °C – 115 °C
$[\alpha]_D^{25}$ +96 (c = 0.01 g mL^{-1}, MeOH)
IR ν_{max}/cm^{-1} 1714, 2954 ^{1}H NMR (400 MHz, *DMSO*-d_6) δ/ppm 12.65 (br. s., 1 H, COO*H*), 7.90 (d, *J*=7.5 Hz, 2 H, Fmoc C_1*H*), 7.74 (d, *J*=7.5 Hz, 2 H, Fmoc C_4*H*), 7.43 (t, *J*=7.5 Hz, 2 H, Fmoc C_2*H*), 7.33 (t, *J*=7.5 Hz, 2 H, Fmoc C_3*H*), 5.44 - 5.58 (m, 1 H, δ-C*H*), 5.29 - 5.44 (m, 1 H, γ-C*H*), 4.17 - 4.34 (m, 3 H, Overlapping Fmoc*H* and Fmoc-CH_2), 3.95 - 4.04 (m, 1 H, α-C*H*), 3.38 (br. s., 1 H, N*H*), 2.26 - 2.49 (m, 2 H, β-CH_2), 1.76 - 1.88 (m, 2 H, ε-CH_2), 1.55 (tq, *J*=13.5, 7.0 Hz, 1 H, ζ-C*H*), 0.77 - 0.94 (m, 6 H, 2 *x* η-CH_3)
^{13}C NMR (101 MHz, *DMSO*-d_6) δ/ppm 173.4, 156.0, 143.8, 143.8, 140.7, 131.8, 130.8, 127.6, 127.1, 126.6, 125.7, 125.3, 120.1, 65.7, 54.1, 46.7, 41.4, 35.8, 34.1, 28.0, 27.8, 22.2, 22.1, 22.0
HRMS (Positive ion ESI) $[M + Na]^+$ Theoretical *m/z* 416.1832, Observed *m/z* 416.1819

(2*S*,4*E*)-2-(((9H-fluoren-9-ylmethoxy)carbonyl)amino)-7,7-dimethyl-4-octenoic acid, 71

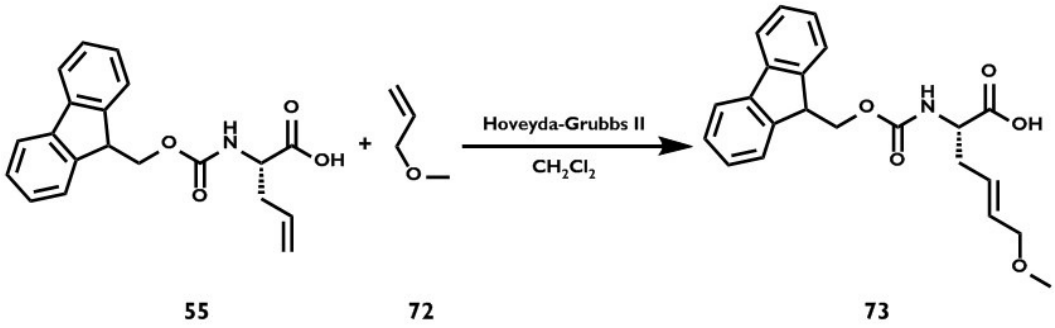


55 69 71

The standard metathesis procedure (Chapter 8) was applied, using **55** (1.000 g) and 4,4-Dimethyl-1-Pentene, (**69**, 0.620 mL, 423 mg, 2.11 mmol, 1.5 equiv.) as the cross metathesis partner. Upon completion, the reaction mixture was then concentrated *in vacuo* before the product was separated from the crude mixture by column chromatography (0 to 33% EtOAc gradient in cyclohexane, 0.2% AcOH). Appropriate fractions were combined and concentrated to dryness *in* vacuo. Catalyst, (**57**, 75 mg) was collected alongside the product, **71** (844 mg, 74%), which was obtained as an off-white solid.

Melting Point 43°C – 45°C
$[\alpha]_D^{25}$ +79 (c = 0.01 g mL^{-1}, MeOH)
IR ν_{max}/cm^{-1} 1714, 2955
^{1}H NMR (400 MHz, *DMSO*-d_6) δ/ppm 12.62 (br. s., 1 H, COO*H*), 7.90 (d, *J*=7.5 Hz, 2 H, Fmoc C_1*H*), 7.73 (d, *J*=7.0 Hz, 2 H, Fmoc C_4*H*), 7.63 (d, *J*=8.0 Hz, 1 H, N*H*), 7.43 (t, *J*=7.5 Hz, 2 H, Fmoc C_2*H*), 7.33 (t, *J*=7.0 Hz, 2 H, Fmoc C_3*H*), 5.47 - 5.63 (m, 1 H, δ-C*H*), 5.31 - 5.46 (m, 1 H, γ-C*H*), 4.17 - 4.32 (m, 3 H, Overlapping Fmoc*H* and Fmoc-CH_2), 4.00 (br. s., 1 H, α-C*H*), 2.41 - 2.54 (m, 1 H, β-CH_2'), 2.29 - 2.41 (m, 1 H, β-CH_2''), 1.83 (d, *J*=7.0 Hz, 2 H, ε-CH_2), 0.83 (s, 9 H, 3 *x* η-CH_3)
^{13}C NMR (101 MHz, *DMSO*-d_6) δ/ppm 173.4, 156.0, 143.8, 143.8, 140.7, 130.3, 127.6, 127.1, 125.3, 120.1, 65.7, 54.1, 46.7, 46.5, 40.6, 34.2, 30.6, 29.1
HRMS (Positive ion ESI) $[M + Na]^+$ Theoretical *m/z* 430.1993, Observed *m/z* 430.1989

(2*S*,4*E*)-2-(((9H-fluoren-9-ylmethoxy)carbonyl)amino)-6-methoxy-4-hexenoic acid, 73

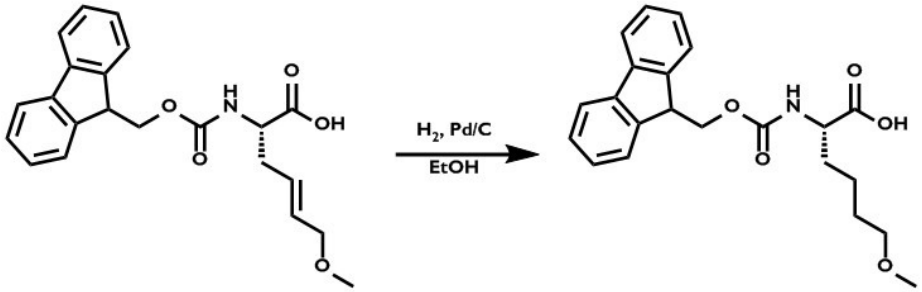


55 72 73

The general metathesis procedure (Chapter 8) was followed, using Hoveyda-Grubbs second generation catalyst (**57**, 28 mg, 0.0446 mmol, 0.15 equiv.), **55,** (102 mg, 0.297 mmol, 1 equiv.) and allyl methylether, (**72**, 0.04 mL, 0.446 mmol, 1.5 equiv.). The product was purified by column chromatography (20% to 60% ethyl acetate in cyclohexane gradient with 1% formic acid). **73** was isolated as an oil (80 mg, 0.210 mmol, 69%).

$[\alpha]_D^{25}$ = +19 (c = 0.50 g mL^{-1}, $CHCl_3$)
IR ν_{max}/cm^{-1} 1715, 2361, 2927
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.07 (br s, 1H, COO*H*), 7.66 (d, *J* = 7.5 Hz, 2H, Fmoc C_1*H*), 7.54 – 7.41 (m, 2H, Fmoc C_4*H*), 7.30 (t, *J* = 7.5 Hz, 2H, Fmoc C_2*H*), 7.21 (t, *J* = 7.5 Hz, Fmoc C_3*H*), 5.55 (dd, *J* = 19.5, 6.5 Hz, 2H, Overlapping Lys γ-C*H* and Lys δ-C*H*), 4.51 – 4.20 (m, 3H, Fmoc*H* and Fmoc-CH_2), 4.20 – 4.06 (m, 1H, Lys α-C*H*), 3.89 – 3.71 (m, 2H, Lys ε-C*H*), 3.21 (s, 3H, OCH_3), 2.50 (ddd, *J* = 19.5, 13.5, 7.0 Hz, 2H, Lys β-C*H*).
^{13}C NMR (101 MHz, $CDCl_3$) δ 175.0, 156.1, 143.9, 143.8, 141.4, 130.9, 127.8, 127.6, 127.2, 125.2, 120.1, 72.6, 67.2, 57.8, 53.3, 47.2, 35.0
HRMS (Positive ion ESI) $[M + H]^+$ Theoretical *m/z* 382.16490, Observed m/z 382.16415

(2*S*)-2-(((9H-fluoren-9-ylmethoxy)carbonyl)amino)-6-methoxyhexanoic acid, 74



73 **74**

The general hydrogenation procedure (Chapter 8) was conducted on **73** (80 mg, 0.210 mmol, 1 equiv.). The crude product, **74**, was obtained as light brown oil (68 mg, 0.176 mmol, 84% yield).

$[\alpha]_D^{25}$ = +11 (c = 0.50 g mL^{-1}, $CHCl_3$)
IR ν_{max}/cm^{-1} 1713, 2935
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.75 (d, *J* = 7.5 Hz, 2H, Fmoc $C_1\underline{H}$), 7.67-7.47 (m, 2H, Fmoc $C_4\underline{H}$), 7.39 (t, *J* = 7.0 Hz, 2H, Fmoc $C_2\underline{H}$), 7.30 (t, *J* = 7.0 Hz, 2H, Fmoc $C_3\underline{H}$), 4.39 (d, *J* = 6.5 Hz, 2H, Fmoc-$C\underline{H_2}$), 4.21 (t, *J* = 6.5 Hz, 1H, Fmoc$\underline{H}$), 3.48 – 3.22 (m, 4H, Lys α-$C\underline{H}$ and $OC\underline{H_3}$), 2.27 – 1.12 (m, 8H).
^{13}C NMR (101 MHz, $CDCl_3$) δ 176.0, 156.3, 143.9, 143.7, 141.3, 127.7, 127.1, 125.1, 112.0, 72.4, 67.1, 58.4, 47.1, 32.0, 28.9, 22.0
HRMS (Positive ion ESI) $[M + H]^+$ Theoretical *m/z* 384.18055, Observed *m/z* 384.17960

Appendix 3
Crystallisation Screen Compositions

PACT Premier D11 Optimisation Screen

	1	2	3	4	5	6	7	8	9	10	11	12
A	% PEG 3350	15	15	15	15	15	15	15	15	15	15	15
	[CaCl2] / mM	0	50	100	150	175	200	225	250	275	300	350	400
	Buffer	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0
B	% PEG 3350	17.5	17.5	17.5	17.5	17.5	17.5	17.5	17.5	17.5	17.5	17.5	17.5
	[CaCl2] / mM	0	50	100	150	175	200	225	250	275	300	350	400
	Buffer	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0
C	% PEG 3350	18.75	18.75	18.75	18.75	18.75	18.75	18.75	18.75	18.75	18.75	18.75	18.75
	[CaCl2] / mM	0	50	100	150	175	200	225	250	275	300	350	400
	Buffer	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0
D	% PEG 3350	20	20	20	20	20	20	20	20	20	20	20	20
	[CaCl2] / mM	0	50	100	150	175	200	225	250	275	300	350	400
	Buffer	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0
E	% PEG 3350	21.25	21.25	21.25	21.25	21.25	21.25	21.25	21.25	21.25	21.25	21.25	21.25
	[CaCl2] / mM	0	50	100	150	175	200	225	250	275	300	350	400
	Buffer	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0
F	% PEG 3350	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5
	[CaCl2] / mM	0	50	100	150	175	200	225	250	275	300	350	400
	Buffer	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0
G	% PEG 3350	23.75	23.75	23.75	23.75	23.75	23.75	23.75	23.75	23.75	23.75	23.75	23.75
	[CaCl2] / mM	0	50	100	150	175	200	225	250	275	300	350	400
	Buffer	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0
H	% PEG 3350	25	25	25	25	25	25	25	25	25	25	25	25
	[CaCl2] / mM	0	50	100	150	175	200	225	250	275	300	350	400
	Buffer	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0	0.1M Tris HCl pH 8.0

Table A 2 - Composition of PACT premier D11 optimisation screen.

Appendix 4
Peptide, Protein and DNA Sequences

Peptide Sequences

Peptide	Sequence	
Peptide 1	GPPGPPGPPGPPGPPGPPGPPGPPGPPGPP	
Peptide 2	PAPKPAPK	
Peptide 3	Hyp-APKPAPK	
Peptide 4	PA-Hyp-KPAPK	
Peptide 5	PAPK-Hyp-APK	
Peptide 6	PAPKPA-Hyp-K	
Peptide 7	PA-Hyp-K-Hyp-APK	
Peptide 8	PA-Hyp-KAAPK	
Peptide 9	PA-Hyp-KRAPK	
Peptide 10	PA-Hyp-KNAPK	
Peptide 11	PA-Hyp-KDAPK	
Peptide 12	PA-Hyp-KCAPK	
Peptide 13	PA-Hyp-KEAPK	
Peptide 14	PA-Hyp-KQAPK	
Peptide 15	PA-Hyp-KGAPK	
Peptide 16	PA-Hyp-KHAPK	
Peptide 17	PA-Hyp-KIAPK	
Peptide 18	PA-Hyp-KLAPK	
Peptide 19	PA-Hyp-KKAPK	
Peptide 20	PA-Hyp-KMAPK	
Peptide 21	PA-Hyp-KFAPK	
Peptide 22	PA-Hyp-KSAPK	
Peptide 23	PA-Hyp-KTAPK	
Peptide 24	PA-Hyp-KWAPK	
Peptide 25	PA-Hyp-KYAPK	
Peptide 26	PA-Hyp-KVAPK	
Peptide 27	PA-Hyp-K- (*2S,4S*)-fluoroproline -APK	
Peptide 29	AALTLLAPAAGDTIISLDF	
Peptide 30	DLDLEMLAPYIPMDDDFQL	
Peptide 31	ARTKQTAR-X-STGGKA	X = **65**
Peptide 32	ARTKQTAR-X-STGGKA	X = **67**
Peptide 33	ARTKQTAR-K(Me_2)-STGGKA	X = N^{ε}-Dimethyl-L-lysine, **75**
Peptide 34	ART-K(Me_3)-QTAR-K(Me_2)-STGGKA	X = N^{ε}-Dimethyl-L-lysine, **75**
Peptide 35	ART-K(Me_3)-QTAR-X-STGGKA	X = **65**
Peptide 36	ARTKQTAR-X-STGGKA	X = Homolysine, **61**
Peptide 37	ARTKQTAR-X-STGGKA	X = **74**
Peptide 38	KQLATKAAR-X-SAPAT	X = **67**
Peptide 39	NA-X-AMSARAEAIKALVKPKE	X = **65**
Peptide 40	NP-X-RSRSREHRR	X = **65**
Peptide 41	X-YGIICHDLIHEIYTVGKR	X = **65**

Table A 3 - Peptides described in this thesis. Hyp = (*2S,4R*)-hydroxyproline.

Protein Sequences

Collagen Prolyl-4-Hydroxylase (*Homo sapiens*)

α(I)-Subunit

Canonical Sequence (Uniprot ID: P13674)

MIWYILIIGILLPQSLAHPGFFTSIGQMTDLIHTEKDLVTSLKDYIKAEEDKLEQIKKWAE
KLDRLTSTATKDPEGFVGHPVNAFKLMKRLNTEWSELENLVLKDMSDGFISNLTIQRQYFP
NDEDQVGAAKALLRLQDTYNLDTDTISKGNLPGVKHKSFLTAEDCFELGKVAYTEADYYHT
ELWMEQALRQLDEGEISTIDKVSVLDYLSYAVYQQGDLDKALLLTKKLLELDPEHQRANGN
LKYFEYIMAKEKDVNKSASDDQSDQKTTPKKKGVAVDYLPERQKYEMLCRGEGIKMTPRRQ
KKLFCRYHDGNRNPKFILAPAKQEDEWDKPRIIRFHDIISDAEIEIVKDLAKPRLRRATIS
NPITGDLETVHYRISKSAWLSGYENPVVSRINMRIQDLTGLDVSTAEELQVANYGVGGQYE
PHFDFARKDEPDAFKELGTGNRIATWLFYMSDVSAGGATVFPEVGASVWPKKGTAVFWYNL
FASGEGDYSTRHAACPVLVGNKWVSNKWLHERGQEFRRPCTLSELE

Residues 18-534 (Minus Endoplasmic Reticulum Localisation Domain)

HPGFFTSIGQMTDLIHTEKDLVTSLKDYIKAEEDKLEQIKKWAEKLDRLTSTATKDPEGFV
GHPVNAFKLMKRLNTEWSELENLVLKDMSDGFISNLTIQRQYFPNDEDQVGAAKALLRLQD
TYNLDTDTISKGNLPGVKHKSFLTAEDCFELGKVAYTEADYYHTELWMEQALRQLDEGEIS
TIDKVSVLDYLSYAVYQQGDLDKALLLTKKLLELDPEHQRANGNLKYFEYIMAKEKDVNKS
ASDDQSDQKTTPKKKGVAVDYLPERQKYEMLCRGEGIKMTPRRQKKLFCRYHDGNRNPKFI
LAPAKQEDEWDKPRIIRFHDIISDAEIEIVKDLAKPRLRRATISNPITGDLETVHYRISKS
AWLSGYENPVVSRINMRIQDLTGLDVSTAEELQVANYGVGGQYEPHFDFARKDEPDAFKEL
GTGNRIATWLFYMSDVSAGGATVFPEVGASVWPKKGTAVFWYNLFASGEGDYSTRHAACPV
LVGNKWVSNKWLHERGQEFRRPCTLSELE

Residues 18-534 C167S Variant

HPGFFTSIGQMTDLIHTEKDLVTSLKDYIKAEEDKLEQIKKWAEKLDRLTSTATKDPEGFV
GHPVNAFKLMKRLNTEWSELENLVLKDMSDGFISNLTIQRQYFPNDEDQVGAAKALLRLQD
TYNLDTDTISKGNLPGVKHKSFLTAEDSFELGKVAYTEADYYHTELWMEQALRQLDEGEIS
TIDKVSVLDYLSYAVYQQGDLDKALLLTKKLLELDPEHQRANGNLKYFEYIMAKEKDVNKS
ASDDQSDQKTTPKKKGVAVDYLPERQKYEMLCRGEGIKMTPRRQKKLFCRYHDGNRNPKFI
LAPAKQEDEWDKPRIIRFHDIISDAEIEIVKDLAKPRLRRATISNPITGDLETVHYRISKS
AWLSGYENPVVSRINMRIQDLTGLDVSTAEELQVANYGVGGQYEPHFDFARKDEPDAFKEL
GTGNRIATWLFYMSDVSAGGATVFPEVGASVWPKKGTAVFWYNLFASGEGDYSTRHAACPV
LVGNKWVSNKWLHERGQEFRRPCTLSELE

α(II)-Subunit

Canonical Sequence (Uniprot ID: O15460)

MKLWVSALLMAWFGVLSCVQAEFFTSIGHMTDLIYAEKELVQSLKEYILVEEAKLSKIKSW
ANKMEALTSKSAADAEGYLAHPVNAYKLVKRLNTDWPALEDLVLQDSAAGFIANLSVQRQF
FPTDEDEIGAAKALMRLQDTYRLDPGTISRGELPGTKYQAMLSVDDCFGMGRSAYNEGDYY
HTVLWMEQVLKQLDAGEEATTTKSQVLDYLSYAVFQLGDLHRALELTRRLLSLDPSHERAG
GNLRYFEQLLEEEREKTLTNQTEAELATPEGIYERPVDYLPERDVYESLCRGEGVKLTPRR
QKRLFCRYHHGNRAPQLLIAPFKEEDEWDSPHIVRYYDVMSDEEIERIKEIAKPKLARATV
RDPKTGVLTVASYRVSKSSWLEEDDDPVVARVNRRMQHITGLTVKTAELLQVANYGVGGQY
EPHFDFSRNDERDTFKHLGTGNRVATFLNYMSDVEAGGATVFPDLGAAIWPKKGTAVFWYN
LLRSGEGDYRTRHAACPVLVGCKWVSNKWFHERGQEFLRPCGSTEVD

Residues 22-535 (Minus Endoplasmic Reticulum Localisation Domain)

SEFFTSIGHMTDLIYAEKELVQSLKEYILVEEAKLSKIKSWANKMEALTSKSAADAEGYLA
HPVNAYKLVKRLNTDWPALEDLVLQDSAAGFIANLSVQRQFFPTDEDEIGAAKALMRLQDT
YRLDPGTISRGELPGTKYQAMLSVDDCFGMGRSAYNEGDYYHTVLWMEQVLKQLDAGEEAT
TTKSQVLDYLSYAVFQLGDLHRALELTRRLLSLDPSHERAGGNLRYFEQLLEEEREKTLTN
QTEAELATPEGIYERPVDYLPERDVYESLCRGEGVKLTPRRQKRLFCRYHHGNRAPQLLIA

PFKEEDEWDSPHIVRYYDVMSDEEIERIKEIAKPKLARATVRDPKTGVLTVASYRVSKSSW
LEEDDDPVVARVNRRMQHITGLTVKTAELLQVANYGVGGQYEPHFDFSRNDERDTFKHLGT
GNRVATFLNYMSDVEAGGATVFPDLGAAIWPKKGTAVFWYNLLRSGEGDYRTRHAACPVLV
GCKWVSNKWFHERGQEFLRPCGSTEVD

α(III)-Subunit

Canonical Sequence (Uniprot ID: Q7Z4N8)

MGPGARLAALLAVLALGTGDPERAAARGDTFSALTSVARALAPERRLLGLLRRYLRGEEAR
LRDLTRFYDKVLSLHEDSTTPVANPLLAFTLIKRLQSDWRNVVHSLEASENIRALKDGYEK
VEQDLPAFEDLEGAARALMRLQDVYMLNVKGLARGVFQRVTGSAITDLYSPKRLFSLTGDD
CFQVGKVAYDMGDYYHAIPWLEEAVSLFRGSYGEWKTEDEASLEDALDHLAFAYFRAGNVS
CALSLSREFLLYSPDNKRMARNVLKYERLLAESPNHVVAEAVIQRPNIPHLQTRDTYEGLC
QTLGSQPTLYQIPSLYCSYETNSNAYLLLQPIRKEVIHLEPYIALYHDFVSDSEAQKIREL
AEPWLQRSVVASGEKQLQVEYRISKSAWLKDTVDPKLVTLNHRIAALTGLDVRPPYAEYLQ
VVNYGIGGHYEPHFDHATSPSSPLYRMKSGNRVATFMIYLSSVEAGGATAFIYANLSVPVV
RNAALFWWNLHRSGEGDSDTLHAGCPVLVGDKWVANKWIHEYGQEFRRPCSSSPED

Residues 19-544 (Minus Endoplasmic Reticulum Localisation Domain)

SGDPERAAARGDTFSALTSVARALAPERRLLGLLRRYLRGEEARLRDLTRFYDKVLSLHED
STTPVANPLLAFTLIKRLQSDWRNVVHSLEASENIRALKDGYEKVEQDLPAFEDLEGAARA
LMRLQDVYMLNVKGLARGVFQRVTGSAITDLYSPKRLFSLTGDDCFQVGKVAYDMGDYYHA
IPWLEEAVSLFRGSYGEWKTEDEASLEDALDHLAFAYFRAGNVSCALSLSREFLLYSPDNK
RMARNVLKYERLLAESPNHVVAEAVIQRPNIPHLQTRDTYEGLCQTLGSQPTLYQIPSLYC
SYETNSNAYLLLQPIRKEVIHLEPYIALYHDFVSDSEAQKIRELAEPWLQRSVVASGEKQL
QVEYRISKSAWLKDTVDPKLVTLNHRIAALTGLDVRPPYAEYLQVVNYGIGGHYEPHFDHA
TSPSSPLYRMKSGNRVATFMIYLSSVEAGGATAFIYANLSVPVVRNAALFWWNLHRSGEGD
SDTLHAGCPVLVGDKWVANKWIHEYGQEFRRPCSSSPED

β-subunit

Canonical Sequence (Uniprot ID: P07237)

MLRRALLCLAVAALVRADAPEEEDHVLVLRKSNFAEALAAHKYLLVEFYAPWCGHCKALAP
EYAKAAGKLKAEGSEIRLAKVDATEESDLAQQYGVRGYPTIKFFRNGDTASPKEYTAGREA
DDIVNWLKKRTGPAATTLPDGAAAESLVESSEVAVIGFFKDVESDSAKQFLQAAEAIDDIP
FGITSNSDVFSKYQLDKDGVVLFKKFDEGRNNFEGEVTKENLLDFIKHNQLPLVIEFTEQT
APKIFGGEIKTHILLFLPKSVSDYDGKLSNFKTAAESFKGKILFIFIDSDHTDNQRILEFF
GLKKEECPAVRLITLEEEMTKYKPESEELTAERITEFCHRFLEGKIKPHLMSQELPEDWDK
QPVKVLVGKNFEDVAFDEKKNVFVEFYAPWCGHCKQLAPIWDKLGETYKDHENIVIAKMDS
TANEVEAVKVHSFPTLKFFPASADRTVIDYNGERTLDGFKKFLESGGQDGAGDDDDLEDLE
EAEEPDMEEDDDQKAVKDEL

Residues 18-508 (Minus Endoplasmic Reticulum Localisation Domain)

DAPEEEDHVLVLRKSNFAEALAAHKYLLVEFYAPWCGHCKALAPEYAKAAGKLKAEGSEIR
LAKVDATEESDLAQQYGVRGYPTIKFFRNGDTASPKEYTAGREADDIVNWLKKRTGPAATT
LPDGAAAESLVESSEVAVIGFFKDVESDSAKQFLQAAEAIDDIPFGITSNSDVFSKYQLDK
DGVVLFKKFDEGRNNFEGEVTKENLLDFIKHNQLPLVIEFTEQTAPKIFGGEIKTHILLFL
PKSVSDYDGKLSNFKTAAESFKGKILFIFIDSDHTDNQRILEFFGLKKEECPAVRLITLEE
EMTKYKPESEELTAERITEFCHRFLEGKIKPHLMSQELPEDWDKQPVKVLVGKNFEDVAFD
EKKNVFVEFYAPWCGHCKQLAPIWDKLGETYKDHENIVIAKMDSTANEVEAVKVHSFPTLK
FFPASADRTVIDYNGERTLDGFKKFLESGGQDGAGDDDDLEDLEEAEEPDMEEDDDQKAVK
DEL

Residues 18-508 C53S C56S C312S C343S C397S C400S Variant

DAPEEEDHVLVLRKSNFAEALAAHKYLLVEFYAPWSGHSKALAPEYAKAAGKLKAEGSEIR
LAKVDATEESDLAQQYGVRGYPTIKFFRNGDTASPKEYTAGREADDIVNWLKKRTGPAATT
LPDGAAAESLVESSEVAVIGFFKDVESDSAKQFLQAAEAIDDIPFGITSNSDVFSKYQLDK
DGVVLFKKFDEGRNNFEGEVTKENLLDFIKHNQLPLVIEFTEQTAPKIFGGEIKTHILLFL

PKSVSDYDGKLSNFKTAAESFKGKILFIFIDSDHTDNQRILEFFGLKKEESPAVRLITLEE
EMTKYKPESEELTAERITEFSHRFLEGKIKPHLMSQELPEDWDKQPVKVLVGKNFEDVAFD
EKKNVFVEFYAPWSGHSKQLAPIWDKLGETYKDHENIVIAKMDSTANEVEAVKVHSFPTLK
FFPASADRTVIDYNGERTLDGFKKFLESGGQDGAGDDDDLEDLEEAEEPDMEEDDDQKAVK
DEL

Prolyl Hydroxylase from *Paramecium Bursaria* Chlorella Virus I

Sequence (Uniprot ID: Q84406)

MTNKFISYNKMETREYLLTILFVIACFMVLNLERREGFETSDRPGVCDGKYYEKIDGFLS
DIECDVLINAAIKKGLIKSEVGGATENDPIKLDPKSRNSEQTWFMPGEHEVIDKIQKKTR
EFLNSKKHCIDKYNFEDVQVARYKPGQYYYHHYDGDDCDDACPKDQRLATLMVYLKAPEE
GGGGETDFPTLKTKIKPKKGTSIFFWVADPVTRKLYKETLHAGLPVKSGEKIIANQWIRA
VK

Residues 36-242

GFETSDRPGVCDGKYYEKIDGFLSDIECDVLINAAIKKGLIKSEVGGATENDPIKLDPKS
RNSEQTWFMPGEHEVIDKIQKKTREFLNSKKHCIDKYNFEDVQVARYKPGQYYYHHYDGD
DCDDACPKDQRLATLMVYLKAPEEGGGGETDFPTLKTKIKPKKGTSIFFWVADPVTRKLY
KETLHAGLPVKSGEKIIANQWIRAVK

Protein Encoded by pET28a_vCPH_32-242

MGSSHHHHHHSSGLVPRGSHMEGFETSDRPGVCDGKYYEKIDGFLSDIECDVLINAAIKK
GLIKSEVGGATENDPIKLDPKSRNSEQTWFMPGEHEVIDKIQKKTREFLNSKKHCIDKYN
FEDVQVARYKPGQYYYHHYDGDDCDDACPKDQRLATLMVYLKAPEEGGGGETDFPTLKTK
IKPKKGTSIFFWVADPVTRKLYKETLHAGLPVKSGEKIIANQWIRAVK

KDM7B

Protein Encoded by pNH-TrxT_PHF8

MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQ
GKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGTE
NLYFQSMPVKTGSPTFVRELRSRTFDSSDEVILKPTGNQLTVEFLEENSFSVPILVLKKDG
LGMTLPSPSFTVRDVEHYVGSDKEIDVIDVTRQADCKMKLGDFVKYYYSGKREKVLNVISL
EFSDTRLSNLVETPKIVRKLSWVENLWPEECVFERPNVQKYCLMSVRDSYTDFHIDFGGTS
VWYHVLKGEKIFYLIRPTNANLTLFECWSSSSNQNEMFFGDQVDKCYKCSVKQGQTLFIPT
GWIHAVLTPVDCLAFGGNFLHSLNIEMQLKAYEIEKRLSTADLFRFPNFETICWYVGKHIL
DIFRGLRENRRHPASYLVHGGKALNLAFRAWTRKEALPDHEDEIPETVRTVQLIKDLAREI
RLVEDIFQQN

Protein after Cleavage with TEV Protease

SMPVKTGSPTFVRELRSRTFDSSDEVILKPTGNQLTVEFLEENSFSVPILVLKKDGLGMTL
PSPSFTVRDVEHYVGSDKEIDVIDVTRQADCKMKLGDFVKYYYSGKREKVLNVISLEFSDT
RLSNLVETPKIVRKLSWVENLWPEECVFERPNVQKYCLMSVRDSYTDFHIDFGGTSVWYHV
LKGEKIFYLIRPTNANLTLFECWSSSSNQNEMFFGDQVDKCYKCSVKQGQTLFIPTGWIHA
VLTPVDCLAFGGNFLHSLNIEMQLKAYEIEKRLSTADLFRFPNFETICWYVGKHILDIFRG
LRENRRHPASYLVHGGKALNLAFRAWTRKEALPDHEDEIPETVRTVQLIKDLAREIRLVED
IFQQN

TEV Protease

Protein Encoded by TEV-DM-Prk793 Plasmid [306]

MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDII
FWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDL
LPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVG
VDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYG
VTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVAL
KSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQ

TNSSSNNNNNNNNNNLGIEGRGENLYFQGHHHHHHHGESLFKGPRDYNPISSTICHLTNES
DGHTTSLYGIGFGPFIITNKHLFRRNNGTLVVQSLHGVFKVKNTTTLQQHLIDGRDMIIIR
MPKDFPPFPQKLKFREPQREERICLVTTNFQTKSMSSMVSDTSCTFPSGDGIFWKHWIQTK
DGQCGSPLVSTRDGFIVGIHSASNFTNTNNYFTSVPKNFMELLTNQEAQQWVSGWRLNADS
VLWGGHKVFMVKPEEPFQPVKEATQLMNRRRRR

Self-Cleavage Product[306]

GHHHHHHHGESLFKGPRDYNPISSTICHLTNESDGHTTSLYGIGFGPFIITNKHLFRRNNG
TLVVQSLHGVFKVKNTTTLQQHLIDGRDMIIIRMPKDFPPFPQKLKFREPQREERICLVTT
NFQTKSMSSMVSDTSCTFPSGDGIFWKHWIQTKDGQCGSPLVSTRDGFIVGIHSASNFTNT
NNYFTSVPKNFMELLTNQEAQQWVSGWRLNADSVLWGGHKVFMVKPEEPFQPVKEATQLMN
RRRRR

DNA Sequences

Sequences of expressed regions are indicated below.

pETDuet-1_PDI_HisCPH1

MCS1 (PDI)

CATGGGCAGCAGCCATCACCATCATCACCACAGCCAGGATCCGAATTCGGATGCACCGGAA
GAAGAAGATCACGTTCTGGTTCTGCGTAAAAGCAATTTTGCAGAAGCACTGGCAGCACATA
AATATCTGCTGGTTGAGTTTTATGCACCGTGGTGTGGTCATTGTAAAGCCCTGGCACCGGA
ATATGCAAAAGCAGCAGGTAAACTGAAAGCAGAAGGTAGCGAAATTCGTCTGGCAAAAGTT
GATGCAACCGAAGAAAGCGATCTGGCACAGCAGTATGGTGTTCGTGGTTATCCGACCATTA
AATTCTTTCGTAATGGTGATACCGCAAGCCCGAAAGAATATACCGCAGGTCGTGAAGCAGA
TGATATTGTTAATTGGCTGAAAAAACGTACCGGTCCGGCAGCAACCACCCTGCCGGATGGT
GCAGCAGCAGAAAGCCTGGTTGAAAGCAGCGAAGTTGCAGTTATTGGCTTTTTCAAAGATG
TGGAAAGCGATAGCGCAAAACAGTTTCTGCAGGCAGCCGAAGCAATTGATGATATTCCGTT
TGGTATTACCAGCAACAGTGATGTGTTTAGCAAATACCAGCTGGATAAAGATGGTGTGGTG
CTGTTCAAAAAATTCGATGAAGGTCGCAATAACTTTGAAGGCGAAGTGACCAAAGAAAACC
TGCTGGATTTTATCAAACACAATCAGCTGCCGCTGGTGATTGAATTTACCGAACAGACCGC
ACCGAAAATCTTTGGTGGTGAAATCAAAACCCATATCCTGCTGTTTCTGCCGAAAAGCGTT
AGCGATTATGATGGTAAACTGTCCAACTTTAAAACCGCAGCCGAAAGCTTTAAAGGCAAAA
TCCTGTTTATCTTCATCGATAGCGATCACACCGATAATCAGCGTATCCTGGAATTTTTCGG
TCTGAAAAAAGAAGAATGTCCGGCAGTTCGTCTGATTACCCTGGAAGAAGAGATGACCAAA
TACAAACCGGAAAGCGAAGAACTGACCGCAGAACGTATTACCGAATTTTGTCATCGTTTTC
TGGAAGGTAAAATCAAACCGCATCTGATGAGCCAAGAACTGCCTGAAGATTGGGATAAACA
GCCGGTTAAAGTTCTGGTGGGTAAAAACTTTGAGGATGTGGCCTTTGACGAGAAAAAAAAC
GTGTTCGTCGAGTTCTATGCCCCTTGGTGCGGACATTGCAAACAGCTGGCACCGATTTGGG
ACAAACTGGGTGAAACCTATAAAGATCATGAGAACATTGTGATCGCCAAAATGGATAGCAC
CGCAAATGAAGTTGAAGCCGTTAAAGTGCATAGCTTTCCGACACTGAAATTTTTCCCTGCA
AGCGCAGATCGTACCGTGATTGATTATAATGGTGAACGTACCCTGGATGGCTTCAAAAAAT
TCCTGGAAAGCGGTGGTCAGGATGGTGCCGGTGATGATGATGACCTGGAAGATCTGGAAGA
GGCCGAAGAACCGGATATGGAAGAGGATGATGATCAGAAAGCAGTGAAAGATGAACTGTAA

MCS2 (CPHA1)

ATGACCGATCTGATTCATACCGAAAAAGATCTGGTTACCAGCCTGAAAGATTATATCAAAG
CCGAAGAAGATAAACTGGAACAAATCAAAAAATGGGCAGAGAAACTGGATCGTCTGACCAG
CACCGCAACCAAAGATCCGGAAGGTTTTGTTGGTCATCCGGTTAATGCATTTAAACTGATG
AAACGTCTGAACACCGAATGGTCAGAACTGGAAAATCTGGTTCTGAAAGATATGAGCGACG
GCTTTATTAGCAATCTGACAATTCAGCGTCAGTATTTTCCGAACGATGAAGATCAGGTTGG
TGCAGCAAAAGCACTGCTGCGTCTGCAGGATACCTATAATCTGGATACCGATACCATTAGC
AAAGGTAATCTGCCTGGTGTGAAACATAAAAGTTTTCTGACCGCAGAAGATTGCTTTGAAC
TGGGTAAAGTTGCATATACCGAGGCCGATTATTATCATACAGAACTGTGGATGGAACAGGC
ACTGCGTCAGCTGGATGAAGGTGAAATCAGCACCATTGATAAAGTTAGCGTTCTGGATTAT
CTGAGCTATGCAGTTTATCAGCAGGGTGATCTGGATAAAGCCCTGCTGCTGACCAAAAAAC
TGCTGGAACTGGATCCGGAACATCAGCGTGCCAATGGTAATCTGAAATACTTCGAATACAT
CATGGCCAAAGAAAAAGATGTGAACAAAAGCGCAAGTGATGATCAGAGCGATCAGAAAACC
ACCCCGAAAAAAAAAAGGTGTTGCAGTTGATTATCTGCCGGAACGTCAGAAATATGAAATGC
TGTGTCGTGGTGAAGGCATTAAAATGACACCGCGTCGTCAGAAAAAACTGTTTTGTCGTTA
TCATGATGGCAACCGCAATCCGAAATTTATCCTGGCACCGGCAAAACAAGAGGATGAATGG
GATAAACCGCGTATTATTCGCTTTCACGATATTATCAGTGATGCCGAAATCGAAATTGTTA
AAGACCTGGCAAAACCGCGTCTGCGTCGTGCAACCATTAGTAATCCGATTACAGGCGATCT
GGAAACCGTTCATTATCGTATTAGCAAAAGCGCCTGGCTGAGCGGTTATGAAAATCCGGTT
GTTAGCCGCATTAATATGCGTATTCAGGATCTGACCGGTCTGGATGTTAGCACAGCAGAAG
AACTGCAGGTTGCAAATTATGGTGTTGGTGGTCAGTATGAACCGCATTTTGATTTTGCACG
TAAAGATGAACCGGATGCCTTTAAAGAACTGGGCACCGGTAATCGTATTGCCACCTGGCTG
TTTTATATGAGTGATGTTAGTGCCGGTGGTGCAACCGTTTTTCCGGAAGTTGGTGCCAGCG
TTTGGCCGAAAAAAGGCACCGCAGTGTTTTGGTATAACCTGTTTGCAAGCGGTGAAGGTGA

TTATAGCACCCGTCATGCAGCATGTCCGGTTCTGGTTGGTAATAAATGGGTTAGCAATAAA
TGGCTGCATGAACGCGGTCAAGAATTTCGTCGTCCGTGTACCCTGAGCGAACTGGAATAA

pOPINRSE_α(I)_Stop

ATGCATCCGGGTTTTTTTACCAGCATTGGTCAGATGACCGATCTGATTCATACCGAAAAAG
ATCTGGTTACCAGCCTGAAAGATTATATCAAAGCCGAAGAAGATAAACTGGAACAAATCAA
AAAATGGGCAGAGAAACTGGATCGTCTGACCAGCACCGCAACCAAAGATCCGGAAGGTTTT
GTTGGTCATCCGGTTAATGCATTTAAACTGATGAAACGTCTGAACACCGAATGGTCAGAAC
TGGAAAATCTGGTTCTGAAAGATATGAGCGACGGCTTTATTAGCAATCTGACAATTCAGCG
TCAGTATTTTCCGAACGATGAAGATCAGGTTGGTGCAGCAAAAGCACTGCTGCGTCTGCAG
GATACCTATAATCTGGATACCGATACCATTAGCAAAGGTAATCTGCCTGGTGTGAAACACA
AAAGCTTTCTGACCGCAGAAGATTGTTTTGAACTGGGTAAAGTTGCATATACCGAGGCCGA
TTATTATCATACAGAACTGTGGATGGAACAGGCACTGCGTCAGCTGGATGAAGGTGAAATC
AGCACCATTGATAAAGTTAGCGTTCTGGATTATCTGAGCTATGCAGTTTATCAGCAGGGTG
ATCTGGATAAAGCCCTGCTGCTGACCAAAAAACTGCTGGAACTGGATCCGGAACATCAGCG
TGCCAATGGTAATCTGAAATACTTCGAATACATCATGGCCAAAGAAAAAGATGTGAACAAA
AGCGCAAGTGATGATCAGAGCGATCAGAAAACCACCCCGAAAAAAAAAGGTGTTGCAGTTG
ATTATCTGCCGGAACGTCAGAAATATGAAATGCTGTGTCGTGGTGAAGGCATTAAAATGAC
ACCGCGTCGTCAGAAAAAACTGTTTTGTCGTTATCATGATGGCAACCGCAATCCGAAATTT
ATCCTGGCACCGGCAAAACAAGAGGATGAATGGGATAAACCGCGTATTATTCGCTTTCACG
ATATTATCAGTGATGCCGAAATCGAAATTGTTAAAGACCTGGCAAAACCGCGTCTGCGTCG
TGCAACCATTAGTAATCCGATTACAGGCGATCTGGAAACCGTTCATTATCGTATTAGCAAA
AGCGCCTGGCTGAGCGGTTATGAAAATCCGGTTGTTAGCCGCATTAATATGCGTATTCAGG
ATCTGACCGGTCTGGATGTTAGCACAGCAGAAGAACTGCAGGTTGCAAATTATGGTGTTGG
TGGTCAGTATGAACCGCATTTTGATTTTGCACGTAAAGATGAACCGGATGCCTTTAAAGAA
CTGGGCACAGGTAATCGTATTGCCACCTGGCTGTTTTATATGAGTGATGTTAGTGCCGGTG
GTGCAACCGTTTTTCCGGAAGTTGGTGCCAGCGTTTGGCCGAAAAAAGGCACCGCAGTGTT
TTGGTATAACCTGTTTGCAAGCGGTGAAGGTGATTATAGCACCCGTCATGCAGCATGTCCG
GTTCTGGTTGGTAATAAATGGGTTAGCAATAAATGGCTGCATGAACGCGGTCAAGAATTTC
GTCGTCCGTGTACCCTGAGCGAACTGGAATAA

pOPINRSE_α(I)

ATGCATCCGGGTTTTTTTACCAGCATTGGTCAGATGACCGATCTGATTCATACCGAAAAAG
ATCTGGTTACCAGCCTGAAAGATTATATCAAAGCCGAAGAAGATAAACTGGAACAAATCAA
AAAATGGGCAGAGAAACTGGATCGTCTGACCAGCACCGCAACCAAAGATCCGGAAGGTTTT
GTTGGTCATCCGGTTAATGCATTTAAACTGATGAAACGTCTGAACACCGAATGGTCAGAAC
TGGAAAATCTGGTTCTGAAAGATATGAGCGACGGCTTTATTAGCAATCTGACAATTCAGCG
TCAGTATTTTCCGAACGATGAAGATCAGGTTGGTGCAGCAAAAGCACTGCTGCGTCTGCAG
GATACCTATAATCTGGATACCGATACCATTAGCAAAGGTAATCTGCCTGGTGTGAAACACA
AAAGCTTTCTGACCGCAGAAGATTGTTTTGAACTGGGTAAAGTTGCATATACCGAGGCCGA
TTATTATCATACAGAACTGTGGATGGAACAGGCACTGCGTCAGCTGGATGAAGGTGAAATC
AGCACCATTGATAAAGTTAGCGTTCTGGATTATCTGAGCTATGCAGTTTATCAGCAGGGTG
ATCTGGATAAAGCCCTGCTGCTGACCAAAAAACTGCTGGAACTGGATCCGGAACATCAGCG
TGCCAATGGTAATCTGAAATACTTCGAATACATCATGGCCAAAGAAAAAGATGTGAACAAA
AGCGCAAGTGATGATCAGAGCGATCAGAAAACCACCCCGAAAAAAAAAGGTGTTGCAGTTG
ATTATCTGCCGGAACGTCAGAAATATGAAATGCTGTGTCGTGGTGAAGGCATTAAAATGAC
ACCGCGTCGTCAGAAAAAACTGTTTTGTCGTTATCATGATGGCAACCGCAATCCGAAATTT
ATCCTGGCACCGGCAAAACAAGAGGATGAATGGGATAAACCGCGTATTATTCGCTTTCACG
ATATTATCAGTGATGCCGAAATCGAAATTGTTAAAGACCTGGCAAAACCGCGTCTGCGTCG
TGCAACCATTAGTAATCCGATTACAGGCGATCTGGAAACCGTTCATTATCGTATTAGCAAA
AGCGCCTGGCTGAGCGGTTATGAAAATCCGGTTGTTAGCCGCATTAATATGCGTATTCAGG
ATCTGACCGGTCTGGATGTTAGCACAGCAGAAGAACTGCAGGTTGCAAATTATGGTGTTGG
TGGTCAGTATGAACCGCATTTTGATTTTGCACGTAAAGATGAACCGGATGCCTTTAAAGAA
CTGGGCACAGGTAATCGTATTGCCACCTGGCTGTTTTATATGAGTGATGTTAGTGCCGGTG
GTGCAACCGTTTTTCCGGAAGTTGGTGCCAGCGTTTGGCCGAAAAAAGGCACCGCAGTGTT
TTGGTATAACCTGTTTGCAAGCGGTGAAGGTGATTATAGCACCCGTCATGCAGCATGTCCG
GTTCTGGTTGGTAATAAATGGGTTAGCAATAAATGGCTGCATGAACGCGGTCAAGAATTTC
GTCGTCCGTGTACCCTGAGCGAACTGGAAAAACATCACCATCACCATCACTAA

pOPINRSF_α(I)

ATGGCACACCATCACCACCATCACAGCAGCGGTCTGGAAGTTCTGTTTCAGGGCCCGCATC
CGGGTTTTTTTACCAGCATTGGTCAGATGACCGATCTGATTCATACCGAAAAAGATCTGGT
TACCAGCCTGAAAGATTATATCAAAGCCGAAGAAGATAAACTGGAACAAATCAAAAAATGG
GCAGAGAAACTGGATCGTCTGACCAGCACCGCAACCAAAGATCCGGAAGGTTTTGTTGGTC
ATCCGGTTAATGCATTTAAACTGATGAAACGTCTGAACACCGAATGGTCAGAACTGGAAAA
TCTGGTTCTGAAAGATATGAGCGACGGCTTTATTAGCAATCTGACAATTCAGCGTCAGTAT
TTTCCGAACGATGAAGATCAGGTTGGTGCAGCAAAAGCACTGCTGCGTCTGCAGGATACCT
ATAATCTGGATACCGATACCATTAGCAAAGGTAATCTGCCTGGTGTGAAACACAAAAGCTT
TCTGACCGCAGAAGATTGTTTTGAACTGGGTAAAGTTGCATATACCGAGGCCGATTATTAT
CATACAGAACTGTGGATGGAACAGGCACTGCGTCAGCTGGATGAAGGTGAAATCAGCACCA
TTGATAAAGTTAGCGTTCTGGATTATCTGAGCTATGCAGTTTATCAGCAGGGTGATCTGGA
TAAAGCCCTGCTGCTGACCAAAAAACTGCTGGAACTGGATCCGGAACATCAGCGTGCCAAT
GGTAATCTGAAATACTTCGAATACATCATGGCCAAAGAAAAAGATGTGAACAAAAGCGCAA
GTGATGATCAGAGCGATCAGAAAACCACCCCGAAAAAAAAAGGTGTTGCAGTTGATTATCT
GCCGGAACGTCAGAAATATGAAATGCTGTGTCGTGGTGAAGGCATTAAAATGACACCGCGT
CGTCAGAAAAAACTGTTTTGTCGTTATCATGATGGCAACCGCAATCCGAAATTTATCCTGG
CACCGGCAAAACAAGAGGATGAATGGGATAAACCGCGTATTATTCGCTTTCACGATATTAT
CAGTGATGCCGAAATCGAAATTGTTAAAGACCTGGCAAAACCGCGTCTGCGTCGTGCAACC
ATTAGTAATCCGATTACAGGCGATCTGGAAACCGTTCATTATCGTATTAGCAAAAGCGCCT
GGCTGAGCGGTTATGAAAATCCGGTTGTTAGCCGCATTAATATGCGTATTCAGGATCTGAC
CGGTCTGGATGTTAGCACAGCAGAAGAACTGCAGGTTGCAAATTATGGTGTTGGTGGTCAG
TATGAACCGCATTTTGATTTTGCACGTAAAGATGAACCGGATGCCTTTAAAGAACTGGGCA
CAGGTAATCGTATTGCCACCTGGCTGTTTTATATGAGTGATGTTAGTGCCGGTGGTGCAAC
CGTTTTTCCGGAAGTTGGTGCCAGCGTTTGGCCGAAAAAAGGCACCGCAGTGTTTTGGTAT
AACCTGTTTGCAAGCGGTGAAGGTGATTATAGCACCCGTCATGCAGCATGTCCGGTTCTGG
TTGGTAATAAATGGGTTAGCAATAAATGGCTGCATGAACGCGGTCAAGAATTTCGTCGTCC
GTGTACCCTGAGCGAACTGGAATAA

pOPINRSE_α(I)C167S_Stop

ATGCATCCGGGTTTTTTTACCAGCATTGGTCAGATGACCGATCTGATTCATACCGAAAAAG
ATCTGGTTACCAGCCTGAAAGATTATATCAAAGCCGAAGAAGATAAACTGGAACAAATCAA
AAAATGGGCAGAGAAACTGGATCGTCTGACCAGCACCGCAACCAAAGATCCGGAAGGTTTT
GTTGGTCATCCGGTTAATGCATTTAAACTGATGAAACGTCTGAACACCGAATGGTCAGAAC
TGGAAAATCTGGTTCTGAAAGATATGAGCGACGGCTTTATTAGCAATCTGACAATTCAGCG
TCAGTATTTTCCGAACGATGAAGATCAGGTTGGTGCAGCAAAAGCACTGCTGCGTCTGCAG
GATACCTATAATCTGGATACCGATACCATTAGCAAAGGTAATCTGCCTGGTGTGAAACACA
AAAGCTTTCTGACCGCAGAAGATTCATTTGAACTGGGTAAAGTTGCATATACCGAGGCCGA
TTATTATCATACAGAACTGTGGATGGAACAGGCACTGCGTCAGCTGGATGAAGGTGAAATC
AGCACCATTGATAAAGTTAGCGTTCTGGATTATCTGAGCTATGCAGTTTATCAGCAGGGTG
ATCTGGATAAAGCCCTGCTGCTGACCAAAAAACTGCTGGAACTGGATCCGGAACATCAGCG
TGCCAATGGTAATCTGAAATACTTCGAATACATCATGGCCAAAGAAAAAGATGTGAACAAA
AGCGCAAGTGATGATCAGAGCGATCAGAAAACCACCCCGAAAAAAAAAGGTGTTGCAGTTG
ATTATCTGCCGGAACGTCAGAAATATGAAATGCTGTGTCGTGGTGAAGGCATTAAAATGAC
ACCGCGTCGTCAGAAAAAACTGTTTTGTCGTTATCATGATGGCAACCGCAATCCGAAATTT
ATCCTGGCACCGGCAAAACAAGAGGATGAATGGGATAAACCGCGTATTATTCGCTTTCACG
ATATTATCAGTGATGCCGAAATCGAAATTGTTAAAGACCTGGCAAAACCGCGTCTGCGTCG
TGCAACCATTAGTAATCCGATTACAGGCGATCTGGAAACCGTTCATTATCGTATTAGCAAA
AGCGCCTGGCTGAGCGGTTATGAAAATCCGGTTGTTAGCCGCATTAATATGCGTATTCAGG
ATCTGACCGGTCTGGATGTTAGCACAGCAGAAGAACTGCAGGTTGCAAATTATGGTGTTGG
TGGTCAGTATGAACCGCATTTTGATTTTGCACGTAAAGATGAACCGGATGCCTTTAAAGAA
CTGGGCACAGGTAATCGTATTGCCACCTGGCTGTTTTATATGAGTGATGTTAGTGCCGGTG
GTGCAACCGTTTTTCCGGAAGTTGGTGCCAGCGTTTGGCCGAAAAAAGGCACCGCAGTGTT
TTGGTATAACCTGTTTGCAAGCGGTGAAGGTGATTATAGCACCCGTCATGCAGCATGTCCG
GTTCTGGTTGGTAATAAATGGGTTAGCAATAAATGGCTGCATGAACGCGGTCAAGAATTTC
GTCGTCCGTGTACCCTGAGCGAACTGGAATAA

pOPINRSE_α(I)C167S

ATGCATCCGGGTTTTTTTACCAGCATTGGTCAGATGACCGATCTGATTCATACCGAAAAAG
ATCTGGTTACCAGCCTGAAAGATTATATCAAAGCCGAAGAAGATAAACTGGAACAAATCAA
AAAATGGGCAGAGAAACTGGATCGTCTGACCAGCACCGCAACCAAAGATCCGGAAGGTTTT
GTTGGTCATCCGGTTAATGCATTTAAACTGATGAAACGTCTGAACACCGAATGGTCAGAAC
TGGAAAATCTGGTTCTGAAAGATATGAGCGACGGCTTTATTAGCAATCTGACAATTCAGCG
TCAGTATTTTCCGAACGATGAAGATCAGGTTGGTGCAGCAAAAGCACTGCTGCGTCTGCAG
GATACCTATAATCTGGATACCGATACCATTAGCAAAGGTAATCTGCCTGGTGTGAAACACA
AAAGCTTTCTGACCGCAGAAGATTCATTTGAACTGGGTAAAGTTGCATATACCGAGGCCGA
TTATTATCATACAGAACTGTGGATGGAACAGGCACTGCGTCAGCTGGATGAAGGTGAAATC
AGCACCATTGATAAAGTTAGCGTTCTGGATTATCTGAGCTATGCAGTTTATCAGCAGGGTG
ATCTGGATAAAGCCCTGCTGCTGACCAAAAAACTGCTGGAACTGGATCCGGAACATCAGCG
TGCCAATGGTAATCTGAAATACTTCGAATACATCATGGCCAAAGAAAAAGATGTGAACAAA
AGCGCAAGTGATGATCAGAGCGATCAGAAAACCACCCCGAAAAAAAAAAGGTGTTGCAGTTG
ATTATCTGCCGGAACGTCAGAAATATGAAATGCTGTGTCGTGGTGAAGGCATTAAAATGAC
ACCGCGTCGTCAGAAAAAACTGTTTTGTCGTTATCATGATGGCAACCGCAATCCGAAATTT
ATCCTGGCACCGGCAAAACAAGAGGATGAATGGGATAAACCGCGTATTATTCGCTTTCACG
ATATTATCAGTGATGCCGAAATCGAAATTGTTAAAGACCTGGCAAAACCGCGTCTGCGTCG
TGCAACCATTAGTAATCCGATTACAGGCGATCTGGAAACCGTTCATTATCGTATTAGCAAA
AGCGCCTGGCTGAGCGGTTATGAAAATCCGGTTGTTAGCCGCATTAATATGCGTATTCAGG
ATCTGACCGGTCTGGATGTTAGCACAGCAGAAGAACTGCAGGTTGCAAATTATGGTGTTGG
TGGTCAGTATGAACCGCATTTTGATTTTGCACGTAAAGATGAACCGGATGCCTTTAAAGAA
CTGGGCACAGGTAATCGTATTGCCACCTGGCTGTTTTATATGAGTGATGTTAGTGCCGGTG
GTGCAACCGTTTTTCCGGAAGTTGGTGCCAGCGTTTGGCCGAAAAAAGGCACCGCAGTGTT
TTGGTATAACCTGTTTGCAAGCGGTGAAGGTGATTATAGCACCCGTCATGCAGCATGTCCG
GTTCTGGTTGGTAATAAATGGGTTAGCAATAAATGGCTGCATGAACGCGGTCAAGAATTTC
GTCGTCCGTGTACCCTGAGCGAACTGGAAAAACATCACCATCACCATCACTAA

pOPINRSF_α(I)C167S

ATGGCACACCATCACCACCATCACAGCAGCGGTCTGGAAGTTCTGTTTCAGGGCCCGCATC
CGGGTTTTTTTACCAGCATTGGTCAGATGACCGATCTGATTCATACCGAAAAAGATCTGGT
TACCAGCCTGAAAGATTATATCAAAGCCGAAGAAGATAAACTGGAACAAATCAAAAAATGG
GCAGAGAAACTGGATCGTCTGACCAGCACCGCAACCAAAGATCCGGAAGGTTTTGTTGGTC
ATCCGGTTAATGCATTTAAACTGATGAAACGTCTGAACACCGAATGGTCAGAACTGGAAAA
TCTGGTTCTGAAAGATATGAGCGACGGCTTTATTAGCAATCTGACAATTCAGCGTCAGTAT
TTTCCGAACGATGAAGATCAGGTTGGTGCAGCAAAAGCACTGCTGCGTCTGCAGGATACCT
ATAATCTGGATACCGATACCATTAGCAAAGGTAATCTGCCTGGTGTGAAACACAAAAGCTT
TCTGACCGCAGAAGATTCATTTGAACTGGGTAAAGTTGCATATACCGAGGCCGATTATTAT
CATACAGAACTGTGGATGGAACAGGCACTGCGTCAGCTGGATGAAGGTGAAATCAGCACCA
TTGATAAAGTTAGCGTTCTGGATTATCTGAGCTATGCAGTTTATCAGCAGGGTGATCTGGA
TAAAGCCCTGCTGCTGACCAAAAAACTGCTGGAACTGGATCCGGAACATCAGCGTGCCAAT
GGTAATCTGAAATACTTCGAATACATCATGGCCAAAGAAAAAGATGTGAACAAAAGCGCAA
GTGATGATCAGAGCGATCAGAAAACCACCCCGAAAAAAAAAGGTGTTGCAGTTGATTATCT
GCCGGAACGTCAGAAATATGAAATGCTGTGTCGTGGTGAAGGCATTAAAATGACACCGCGT
CGTCAGAAAAAACTGTTTTGTCGTTATCATGATGGCAACCGCAATCCGAAATTTATCCTGG
CACCGGCAAAACAAGAGGATGAATGGGATAAACCGCGTATTATTCGCTTTCACGATATTAT
CAGTGATGCCGAAATCGAAATTGTTAAAGACCTGGCAAAACCGCGTCTGCGTCGTGCAACC
ATTAGTAATCCGATTACAGGCGATCTGGAAACCGTTCATTATCGTATTAGCAAAAGCGCCT
GGCTGAGCGGTTATGAAAATCCGGTTGTTAGCCGCATTAATATGCGTATTCAGGATCTGAC
CGGTCTGGATGTTAGCACAGCAGAAGAACTGCAGGTTGCAAATTATGGTGTTGGTGGTCAG
TATGAACCGCATTTTGATTTTGCACGTAAAGATGAACCGGATGCCTTTAAAGAACTGGGCA
CAGGTAATCGTATTGCCACCTGGCTGTTTTATATGAGTGATGTTAGTGCCGGTGGTGCAAC
CGTTTTTCCGGAAGTTGGTGCCAGCGTTTGGCCGAAAAAAGGCACCGCAGTGTTTTGGTAT
AACCTGTTTGCAAGCGGTGAAGGTGATTATAGCACCCGTCATGCAGCATGTCCGGTTCTGG
TTGGTAATAAATGGGTTAGCAATAAATGGCTGCATGAACGCGGTCAAGAATTTCGTCGTCC
GTGTACCCTGAGCGAACTGGAATAA

pOPINRSE_α(II)_Stop

ATGGAATTTTTCACCAGCATTGGTCACATGACCGATCTGATTTATGCAGAAAAAGAACTGG
TTCAGAGCCTGAAAGAATATATCCTGGTTGAAGAAGCCAAACTGAGCAAAATCAAAAGCTG
GGCCAATAAAATGGAAGCACTGACCAGCAAAAGCGCAGCAGATGCAGAAGGTTATCTGGCA
CATCCGGTTAATGCATATAAACTGGTTAAACGCCTGAATACCGATTGGCCTGCACTGGAAG
ATCTGGTTCTGCAGGATAGCGCAGCCGGTTTTATTGCAAATCTGAGCGTTCAGCGTCAGTT
TTTTCCGACCGATGAAGATGAAATTGGTGCAGCCAAAGCACTGATGCGTCTGCAGGACACC
TATCGTCTGGATCCGGGTACAATTAGCCGTGGTGAACTGCCTGGCACCAAATATCAGGCAA
TGCTGAGCGTTGATGATTGTTTTGGTATGGGTCGTAGCGCATATAATGAGGGCGATTATTA
TCATACCGTTCTGTGGATGGAACAGGTTCTGAAACAGCTGGATGCCGGTGAAGAAGCAACC
ACCACCAAAAGCCAGGTTCTGGATTATCTGAGCTATGCAGTTTTTCAGCTGGGTGATCTGC
ATCGTGCCCTGGAACTGACCCGTCGTCTGCTGAGCCTGGATCCGAGCCATGAACGTGCCGG
TGGTAATCTGCGTTATTTTGAACAGCTGCTGGAAGAAGAACGCGAAAAAACCCTGACCAAT
CAGACCGAAGCAGAACTGGCAACACCGGAAGGTATTTATGAACGTCCTGTTGATTATCTGC
CGGAACGTGATGTTTATGAAAGCCTGTGTCGTGGTGAAGGTGTTAAACTGACACCGCGTCG
TCAGAAACGTCTGTTTTGTCGTTATCATCATGGTAATCGTGCACCGCAGCTGCTGATTGCA
CCGTTTAAAGAAGAGGATGAATGGGATAGTCCGCATATTGTGCGTTATTATGATGTTATGA
GCGACGAAGAAATCGAGCGCATTAAAGAAATTGCAAAACCGAAACTGGCACGTGCAACCGT
TCGTGATCCGAAAACCGGTGTTCTGACCGTTGCAAGCTATCGTGTTAGCAAAAGCAGTTGG
CTGGAAGAGGACGATGATCCGGTTGTTGCCCGTGTTAATCGTCGTATGCAGCATATTACCG
GTCTGACCGTGAAAACCGCAGAACTGCTGCAGGTTGCAAATTATGGTGTTGGTGGTCAGTA
TGAACCGCACTTTGATTTTTCACGTAATGATGAACGCGACACCTTTAAACATCTGGGTACA
GGTAATCGCGTTGCAACCTTTCTGAATTATATGAGTGATGTTGAAGCCGGTGGTGCCACCG
TTTTTCCGGATCTGGGTGCAGCAATTTGGCCGAAAAAAGGCACCGCAGTTTTTTGGTATAA
CCTGCTGCGTAGCGGTGAAGGCGATTATCGTACCCGTCATGCAGCATGTCCGGTTCTGGTT
GGTTGTAAATGGGTTAGCAATAAATGGTTTCATGAACGCGGTCAAGAATTTCTGCGTCCGT
GTGGTAGCACCGAAGTTGATTAA

pOPINRSF_α(II)

ATGGCACACCATCACCACCATCACAGCAGCGGTCTGGAAGTTCTGTTTCAGGGCCCGGAAT
TTTTCACCAGCATTGGTCACATGACCGATCTGATTTATGCAGAAAAAGAACTGGTTCAGAG
CCTGAAAGAATATATCCTGGTTGAAGAAGCCAAACTGAGCAAAATCAAAAGCTGGGCCAAT
AAAATGGAAGCACTGACCAGCAAAAGCGCAGCAGATGCAGAAGGTTATCTGGCACATCCGG
TTAATGCATATAAACTGGTTAAACGCCTGAATACCGATTGGCCTGCACTGGAAGATCTGGT
TCTGCAGGATAGCGCAGCCGGTTTTATTGCAAATCTGAGCGTTCAGCGTCAGTTTTTTCCG
ACCGATGAAGATGAAATTGGTGCAGCCAAAGCACTGATGCGTCTGCAGGACACCTATCGTC
TGGATCCGGGTACAATTAGCCGTGGTGAACTGCCTGGCACCAAATATCAGGCAATGCTGAG
CGTTGATGATTGTTTTGGTATGGGTCGTAGCGCATATAATGAGGGCGATTATTATCATACC
GTTCTGTGGATGGAACAGGTTCTGAAACAGCTGGATGCCGGTGAAGAAGCAACCACCACCA
AAAGCCAGGTTCTGGATTATCTGAGCTATGCAGTTTTTCAGCTGGGTGATCTGCATCGTGC
CCTGGAACTGACCCGTCGTCTGCTGAGCCTGGATCCGAGCCATGAACGTGCCGGTGGTAAT
CTGCGTTATTTTGAACAGCTGCTGGAAGAAGAACGCGAAAAAACCCTGACCAATCAGACCG
AAGCAGAACTGGCAACACCGGAAGGTATTTATGAACGTCCTGTTGATTATCTGCCGGAACG
TGATGTTTATGAAAGCCTGTGTCGTGGTGAAGGTGTTAAACTGACACCGCGTCGTCAGAAA
CGTCTGTTTTGTCGTTATCATCATGGTAATCGTGCACCGCAGCTGCTGATTGCACCGTTTA
AAGAAGAGGATGAATGGGATAGTCCGCATATTGTGCGTTATTATGATGTTATGAGCGACGA
AGAAATCGAGCGCATTAAAGAAATTGCAAAACCGAAACTGGCACGTGCAACCGTTCGTGAT
CCGAAAACCGGTGTTCTGACCGTTGCAAGCTATCGTGTTAGCAAAAGCAGTTGGCTGGAAG
AGGACGATGATCCGGTTGTTGCCCGTGTTAATCGTCGTATGCAGCATATTACCGGTCTGAC
CGTGAAAACCGCAGAACTGCTGCAGGTTGCAAATTATGGTGTTGGTGGTCAGTATGAACCG
CACTTTGATTTTTCACGTAATGATGAACGCGACACCTTTAAACATCTGGGTACAGGTAATC
GCGTTGCAACCTTTCTGAATTATATGAGTGATGTTGAAGCCGGTGGTGCCACCGTTTTTCC
GGATCTGGGTGCAGCAATTTGGCCGAAAAAAGGCACCGCAGTTTTTTGGTATAACCTGCTG
CGTAGCGGTGAAGGCGATTATCGTACCCGTCATGCAGCATGTCCGGTTCTGGTTGGTTGTA
AATGGGTTAGCAATAAATGGTTTCATGAACGCGGTCAAGAATTTCTGCGTCCGTGTGGTAG
CACCGAAGTTGATTAA

pOPINCDE_β_Stop

ATGGATGCACCGGAAGAAGAAGATCACGTTCTGGTTCTGCGTAAAAGCAATTTTGCAGAAG
CACTGGCAGCACATAAATATCTGCTGGTTGAGTTTTATGCACCGTGGTGTGGTCATTGTAA
AGCCCTGGCACCGGAATATGCAAAAGCAGCAGGTAAACTGAAAGCAGAAGGTAGCGAAATT
CGTCTGGCAAAAGTTGATGCAACCGAAGAAAGCGATCTGGCACAGCAGTATGGTGTTCGTG
GTTATCCGACCATTAAATTCTTTCGTAATGGTGATACCGCAAGCCCGAAAGAATATACCGC
AGGTCGTGAAGCAGATGATATTGTTAATTGGCTGAAAAAACGTACCGGTCCGGCAGCAACC
ACCCTGCCGGATGGTGCAGCAGCAGAAAGCCTGGTTGAAAGCAGCGAAGTTGCAGTTATTG
GCTTTTTCAAAGATGTGGAAAGCGATAGCGCAAAACAGTTTCTGCAGGCAGCCGAAGCAAT
TGATGATATTCCGTTTGGTATTACCAGCAACAGTGATGTGTTTAGCAAATACCAGCTGGAT
AAAGATGGTGTGGTGCTGTTCAAAAAATTCGATGAAGGTCGCAATAACTTTGAAGGCGAAG
TGACCAAAGAAAACCTGCTGGATTTTATCAAACACAATCAGCTGCCGCTGGTGATTGAATT
TACCGAACAGACCGCACCAAAAATCTTTGGCGGTGAAATCAAAACCCATATCCTGCTGTTT
CTGCCGAAAAGCGTTAGCGATTATGATGGTAAACTGTCCAACTTTAAAACCGCAGCCGAAA
GCTTTAAAGGCAAAATCCTGTTTATCTTCATCGATAGCGATCACACCGATAATCAGCGTAT
CCTGGAATTTTTCGGTCTGAAAAAAGAAGAATGTCCGGCAGTTCGTCTGATTACCCTGGAA
GAAGAGATGACCAAATACAAACCGGAAAGCGAAGAACTGACCGCAGAACGTATTACCGAAT
TTTGTCATCGTTTTCTGGAAGGTAAAATCAAACCGCATCTGATGAGCCAAGAACTGCCTGA
AGATTGGGATAAACAGCCGGTTAAAGTTCTGGTGGGTAAAAACTTTGAGGATGTGGCCTTT
GACGAGAAAAAAAACGTGTTCGTCGAGTTCTATGCCCCTTGGTGCGGACATTGCAAACAGC
TGGCACCGATTTGGGACAAACTGGGTGAAACCTATAAAGATCATGAGAACATTGTGATCGC
CAAAATGGATAGCACCGCAAATGAAGTTGAAGCCGTTAAAGTGCATAGCTTTCCGACACTG
AAATTTTTCCCTGCAAGCGCAGATCGTACCGTGATTGATTATAATGGTGAACGTACCCTGG
ATGGCTTCAAAAAATTCCTGGAAAGCGGTGGTCAGGATGGTGCCGGTGATGATGATGACCT
GGAAGATCTGGAAGAGGCCGAAGAACCGGATATGGAAGAGGATGATGATCAGAAAGCAGTG
AAAGATGAACTGTAA

pOPINCDE_β

ATGGATGCACCGGAAGAAGAAGATCACGTTCTGGTTCTGCGTAAAAGCAATTTTGCAGAAG
CACTGGCAGCACATAAATATCTGCTGGTTGAGTTTTATGCACCGTGGTGTGGTCATTGTAA
AGCCCTGGCACCGGAATATGCAAAAGCAGCAGGTAAACTGAAAGCAGAAGGTAGCGAAATT
CGTCTGGCAAAAGTTGATGCAACCGAAGAAAGCGATCTGGCACAGCAGTATGGTGTTCGTG
GTTATCCGACCATTAAATTCTTTCGTAATGGTGATACCGCAAGCCCGAAAGAATATACCGC
AGGTCGTGAAGCAGATGATATTGTTAATTGGCTGAAAAAACGTACCGGTCCGGCAGCAACC
ACCCTGCCGGATGGTGCAGCAGCAGAAAGCCTGGTTGAAAGCAGCGAAGTTGCAGTTATTG
GCTTTTTCAAAGATGTGGAAAGCGATAGCGCAAAACAGTTTCTGCAGGCAGCCGAAGCAAT
TGATGATATTCCGTTTGGTATTACCAGCAACAGTGATGTGTTTAGCAAATACCAGCTGGAT
AAAGATGGTGTGGTGCTGTTCAAAAAATTCGATGAAGGTCGCAATAACTTTGAAGGCGAAG
TGACCAAAGAAAACCTGCTGGATTTTATCAAACACAATCAGCTGCCGCTGGTGATTGAATT
TACCGAACAGACCGCACCAAAAATCTTTGGCGGTGAAATCAAAACCCATATCCTGCTGTTT
CTGCCGAAAAGCGTTAGCGATTATGATGGTAAACTGTCCAACTTTAAAACCGCAGCCGAAA
GCTTTAAAGGCAAAATCCTGTTTATCTTCATCGATAGCGATCACACCGATAATCAGCGTAT
CCTGGAATTTTTCGGTCTGAAAAAAGAAGAATGTCCGGCAGTTCGTCTGATTACCCTGGAA
GAAGAGATGACCAAATACAAACCGGAAAGCGAAGAACTGACCGCAGAACGTATTACCGAAT
TTTGTCATCGTTTTCTGGAAGGTAAAATCAAACCGCATCTGATGAGCCAAGAACTGCCTGA
AGATTGGGATAAACAGCCGGTTAAAGTTCTGGTGGGTAAAAACTTTGAGGATGTGGCCTTT
GACGAGAAAAAAAACGTGTTCGTCGAGTTCTATGCCCCTTGGTGCGGACATTGCAAACAGC
TGGCACCGATTTGGGACAAACTGGGTGAAACCTATAAAGATCATGAGAACATTGTGATCGC
CAAAATGGATAGCACCGCAAATGAAGTTGAAGCCGTTAAAGTGCATAGCTTTCCGACACTG
AAATTTTTCCCTGCAAGCGCAGATCGTACCGTGATTGATTATAATGGTGAACGTACCCTGG
ATGGCTTCAAAAAATTCCTGGAAAGCGGTGGTCAGGATGGTGCCGGTGATGATGATGACCT
GGAAGATCTGGAAGAGGCCGAAGAACCGGATATGGAAGAGGATGATGATCAGAAAGCAGTG
AAAGATGAACTGAAACATCACCATCACCATCACTAA

pOPINCDF_β

ATGGCACACCATCACCACCATCACAGCAGCGGTCTGGAAGTTCTGTTTCAGGGCCCGGGAT
GCACCGGAAGAAGAAGATCACGTTCTGGTTCTGCGTAAAAGCAATTTTGCAGAAGCACTGG
CAGCACATAAATATCTGCTGGTTGAGTTTTATGCACCGTGGTCAGGTCATAGCAAAGCCCT

GGCACCGGAATATGCAAAAGCAGCAGGTAAACTGAAAGCAGAAGGTAGCGAAATTCGTCTG
GCAAAAGTTGATGCAACCGAAGAAAGCGATCTGGCACAGCAGTATGGTGTTCGTGGTTATC
CGACCATTAAATTCTTTCGTAATGGTGATACCGCAAGCCCGAAAGAATATACCGCAGGTCG
TGAAGCAGATGATATTGTTAATTGGCTGAAAAAACGTACCGGTCCGGCAGCAACCACCCTG
CCGGATGGTGCAGCAGCAGAAAGCCTGGTTGAAAGCAGCGAAGTTGCAGTTATTGGCTTTT
TCAAAGATGTGGAAAGCGATAGCGCAAAACAGTTTCTGCAGGCAGCCGAAGCAATTGATGA
TATTCCGTTTGGTATTACCAGCAACAGTGATGTGTTTAGCAAATACCAGCTGGATAAAGAT
GGTGTGGTGCTGTTCAAAAAATTCGATGAAGGTCGCAATAACTTTGAAGGCGAAGTGACCA
AAGAAAACCTGCTGGATTTTATCAAACACAATCAGCTGCCGCTGGTGATTGAATTTACCGA
ACAGACCGCACCAAAAATCTTTGGCGGTGAAATCAAAACCCATATCCTGCTGTTTCTGCCG
AAAAGCGTTAGCGATTATGATGGTAAACTGTCCAACTTTAAAACCGCAGCCGAAAGCTTTA
AAGGCAAAATCCTGTTTATCTTCATCGATAGCGATCACACCGATAATCAGCGTATCCTGGA
ATTTTTCGGTCTGAAAAAAGAAGAAAGTCCGGCAGTTCGTCTGATTACCCTGGAAGAAGAG
ATGACCAAATACAAACCGGAAAGCGAAGAACTGACCGCAGAACGTATTACCGAATTTTCTC
ATCGTTTTCTGGAAGGTAAAATCAAACCGCATCTGATGAGCCAAGAACTGCCTGAAGATTG
GGATAAACAGCCGGTTAAAGTTCTGGTGGGTAAAAACTTTGAGGATGTGGCCTTCGACGAG
AAAAAAAACGTGTTTGTGGAGTTCTATGCCCCTTGGAGCGGTCATAGTAAACAGCTGGCAC
CGATTTGGGACAAACTGGGTGAAACCTATAAAGATCATGAGAACATTGTGATCGCCAAAAT
GGATAGCACCGCAAATGAAGTTGAAGCCGTTAAAGTGCATAGCTTTCCGACACTGAAATTT
TTCCCTGCAAGCGCAGATCGTACCGTGATTGATTATAATGGTGAACGTACCCTGGATGGCT
TCAAAAAATTCCTGGAAAGCGGTGGTCAGGATGGTGCCGGTGATGATGATGACCTGGAAGA
TCTGGAAGAGGCCGAAGAACCGGATATGGAAGAGGATGATGATCAGAAAGCAGTGAAAGAT
GAACTGTAA

pBADLIC_α(I)_C167S

ATGAAATCTTCTCACCATCACCATCACCATGAAAACCTGTACTTCCAATCCAATGCACATC
CGGGTTTTTTTACCAGCATTGGTCAGATGACCGATCTGATTCATACCGAAAAAGATCTGGT
TACCAGCCTGAAAGATTATATCAAAGCCGAAGAAGATAAACTGGAACAAATCAAAAAATGG
GCAGAGAAACTGGATCGTCTGACCAGCACCGCAACCAAAGATCCGGAAGGTTTTGTTGGTC
ATCCGGTTAATGCATTTAAACTGATGAAACGTCTGAACACCGAATGGTCAGAACTGGAAAA
TCTGGTTCTGAAAGATATGAGCGACGGCTTTATTAGCAATCTGACAATTCAGCGTCAGTAT
TTTCCGAACGATGAAGATCAGGTTGGTGCAGCAAAAGCACTGCTGCGTCTGCAGGATACCT
ATAATCTGGATACCGATACCATTAGCAAAGGTAATCTGCCTGGTGTGAAACACAAAAGCTT
TCTGACCGCAGAAGATTCATTTGAACTGGGTAAAGTTGCATATACCGAGGCCGATTATTAT
CATACAGAACTGTGGATGGAACAGGCACTGCGTCAGCTGGATGAAGGTGAAATCAGCACCA
TTGATAAAGTTAGCGTTCTGGATTATCTGAGCTATGCAGTTTATCAGCAGGGTGATCTGGA
TAAAGCCCTGCTGCTGACCAAAAAACTGCTGGAACTGGATCCGGAACATCAGCGTGCCAAT
GGTAATCTGAAATACTTCGAATACATCATGGCCAAAGAAAAAGATGTGAACAAAAGCGCAA
GTGATGATCAGAGCGATCAGAAAACCACCCCGAAAAAAAAAGGTGTTGCAGTTGATTATCT
GCCGGAACGTCAGAAATATGAAATGCTGTGTCGTGGTGAAGGCATTAAAATGACACCGCGT
CGTCAGAAAAAACTGTTTTGTCGTTATCATGATGGCAACCGCAATCCGAAATTTATCCTGG
CACCGGCAAAACAAGAGGATGAATGGGATAAACCGCGTATTATTCGCTTTCACGATATTAT
CAGTGATGCCGAAATCGAAATTGTTAAAGACCTGGCAAAACCGCGTCTGCGTCGTGCAACC
ATTAGTAATCCGATTACAGGCGATCTGGAAACCGTTCATTATCGTATTAGCAAAAGCGCCT
GGCTGAGCGGTTATGAAAATCCGGTTGTTAGCCGCATTAATATGCGTATTCAGGATCTGAC
CGGTCTGGATGTTAGCACAGCAGAAGAACTGCAGGTTGCAAATTATGGTGTTGGTGGTCAG
TATGAACCGCATTTTGATTTTGCACGTAAAGATGAACCGGATGCCTTTAAAGAACTGGGCA
CAGGTAATCGTATTGCCACCTGGCTGTTTTATATGAGTGATGTTAGTGCCGGTGGTGCAAC
CGTTTTTCCGGAAGTTGGTGCCAGCGTTTGGCCGAAAAAAGGCACCGCAGTGTTTTGGTAT
AACCTGTTTGCAAGCGGTGAAGGTGATTATAGCACCCGTCATGCAGCATGTCCGGTTCTGG
TTGGTAATAAATGGGTTAGCAATAAATGGCTGCATGAACGCGGTCAAGAATTTCGTCGTCC
GTGTACCCTGAGCGAACTGGAATAA

pBADLIC_α(II)

ATGAAATCTTCTCACCATCACCATCACCATGAAAACCTGTACTTCCAATCCAATGCAGAAT
TTTTCACCAGCATTGGTCACATGACCGATCTGATTTATGCAGAAAAAGAACTGGTTCAGAG
CCTGAAAGAATATATCCTGGTTGAAGAAGCCAAACTGAGCAAAATCAAAAGCTGGGCCAAT
AAAATGGAAGCACTGACCAGCAAAAGCGCAGCAGATGCAGAAGGTTATCTGGCACATCCGG
TTAATGCATATAAACTGGTTAAACGCCTGAATACCGATTGGCCTGCACTGGAAGATCTGGT

TCTGCAGGATAGCGCAGCCGGTTTTATTGCAAATCTGAGCGTTCAGCGTCAGTTTTTTCCG
ACCGATGAAGATGAAATTGGTGCAGCCAAAGCACTGATGCGTCTGCAGGACACCTATCGTC
TGGATCCGGGTACAATTAGCCGTGGTGAACTGCCTGGCACCAAATATCAGGCAATGCTGAG
CGTTGATGATTGTTTTGGTATGGGTCGTAGCGCATATAATGAGGGCGATTATTATCATACC
GTTCTGTGGATGGAACAGGTTCTGAAACAGCTGGATGCCGGTGAAGAAGCAACCACCACCA
AAAGCCAGGTTCTGGATTATCTGAGCTATGCAGTTTTTCAGCTGGGTGATCTGCATCGTGC
CCTGGAACTGACCCGTCGTCTGCTGAGCCTGGATCCGAGCCATGAACGTGCCGGTGGTAAT
CTGCGTTATTTTGAACAGCTGCTGGAAGAAGAACGCGAAAAAACCCTGACCAATCAGACCG
AAGCAGAACTGGCAACACCGGAAGGTATTTATGAACGTCCTGTTGATTATCTGCCGGAACG
TGATGTTTATGAAAGCCTGTGTCGTGGTGAAGGTGTTAAACTGACACCGCGTCGTCAGAAA
CGTCTGTTTTGTCGTTATCATCATGGTAATCGTGCACCGCAGCTGCTGATTGCACCGTTTA
AAGAAGAGGATGAATGGGATAGTCCGCATATTGTGCGTTATTATGATGTTATGAGCGACGA
AGAAATCGAGCGCATTAAAGAAATTGCAAAACCGAAACTGGCACGTGCAACCGTTCGTGAT
CCGAAAACCGGTGTTCTGACCGTTGCAAGCTATCGTGTTAGCAAAAGCAGTTGGCTGGAAG
AGGACGATGATCCGGTTGTTGCCCGTGTTAATCGTCGTATGCAGCATATTACCGGTCTGAC
CGTGAAAACCGCAGAACTGCTGCAGGTTGCAAATTATGGTGTTGGTGGTCAGTATGAACCG
CACTTTGATTTTTCACGTAATGATGAACGCGACACCTTTAAACATCTGGGTACAGGTAATC
GCGTTGCAACCTTTCTGAATTATATGAGTGATGTTGAAGCCGGTGGTGCCACCGTTTTTCC
GGATCTGGGTGCAGCAATTTGGCCGAAAAAAGGCACCGCAGTTTTTTGGTATAACCTGCTG
CGTAGCGGTGAAGGCGATTATCGTACCCGTCATGCAGCATGTCCGGTTCTGGTTGGTTGTA
AATGGGTTAGCAATAAATGGTTTCATGAACGCGGTCAAGAATTTCTGCGTCCGTGTGGTAG
CACCGAAGTTGATTAA

pBADLIC_α(III)

ATGAAATCTTCTCACCATCACCATCACCATGAAAACCTGTACTTCCAATCCAATGCAGGTG
ATCCGGAACGTGCAGCAGCACGTGGTGATACCTTTAGCGCACTGACCAGCGTTGCACGTGC
CCTGGCACCGGAACGTCGTCTGCTGGGTCTGCTGCGTCGTTATCTGCGTGGTGAAGAAGCA
CGTCTGCGCGATCTGACCCGTTTTTATGATAAAGTTCTGAGCCTGCATGAAGATAGCACCA
CACCGGTTGCAAATCCGCTGCTGGCATTTACCCTGATTAAACGTCTGCAGAGCGATTGGCG
TAATGTTGTTCATAGCCTGGAAGCAAGCGAAAACATTCGTGCACTGAAAGATGGTTATGAG
AAAGTTGAACAGGATCTGCCTGCATTTGAAGATCTGGAAGGTGCAGCCCGTGCCCTGATGC
GTCTGCAGGATGTTTATATGCTGAATGTTAAAGGTCTGGCACGCGGTGTTTTTCAGCGTGT
TACCGGTAGCGCAATTACCGATCTGTATAGCCCGAAACGTCTGTTTAGCCTGACCGGTGAT
GATTGTTTTCAGGTTGGTAAAGTTGCCTATGATATGGGCGATTATTATCATGCAATTCCGT
GGCTGGAAGAAGCAGTTAGCCTGTTTCGTGGTAGTTATGGTGAATGGAAAACCGAAGATGA
AGCAAGTCTGGAAGATGCACTGGATCATCTGGCATTTGCATATTTTCGTGCAGGTAATGTT
AGCTGTGCACTGAGCCTGAGCCGTGAATTTCTGCTGTATAGTCCGGATAATAAACGTATGG
CACGTAACGTGCTGAAATATGAACGCCTGCTGGCAGAAAGCCCGAATCATGTTGTTGCAGA
AGCCGTTATTCAGCGTCCGAATATTCCGCATCTGCAGACCCGTGATACCTATGAAGGTCTG
TGTCAGACCCTGGGTAGCCAGCCGACCCTGTATCAGATTCCGAGCCTGTATTGTAGCTATG
AAACCAATAGCAATGCATATCTGCTGCTGCAGCCGATTCGTAAAGAAGTTATTCATCTGGA
ACCGTATATCGCGCTGTATCATGATTTTGTTAGCGATAGCGAAGCCCAGAAAATTCGTGAA
CTGGCAGAACCGTGGCTGCAGCGTAGCGTTGTTGCAAGCGGTGAAAAACAGCTGCAGGTTG
AATATCGTATTAGCAAAAGCGCATGGCTGAAAGATACCGTTGATCCGAAACTGGTTACCCT
GAACCATCGTATTGCAGCACTGACCGGTCTGGATGTTCGTCCGCCTTATGCAGAATATCTG
CAGGTAGTTAATTATGGCATTGGTGGTCATTATGAACCGCATTTTGATCATGCAACCAGCC
CGAGCAGTCCGCTGTATCGTATGAAAAGCGGTAATCGTGTTGCCACCTTTATGATTTATCT
GAGCAGCGTTGAAGCCGGTGGTGCAACCGCCTTTATCTATGCAAATCTGAGCGTTCCGGTT
GTTCGTAATGCAGCCCTGTTTTGGTGGAATCTGCATCGTAGCGGTGAAGGTGATAGCGATA
CCCTGCATGCCGGTTGTCCGGTTCTGGTTGGTGATAAATGGGTTGCCAATAAATGGATTCA
CGAATATGGTCAAGAATTTCGTCGTCCGTGTAGCAGCAGTCCGGAAGATTAA

pET28a_vCPH_36-242

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATA
TGGAGGGGTTTGAAACCAGCGATCGTCCGGGCGTCTGCGATGGCAAATATTATGAAAAAAT
TGACGGTTTCCTGTCCGATATTGAATGTGACGTGCTGATTAACGCGGCAATTAAAAAGGGC
CTGATCAAATCTGAAGTCGGGGGTGCGACAGAAAATGACCCGATTAAACTGGACCCGAAAA
GCCGTAATTCGGAACAGACCTGGTTTATGCCGGGTGAGCATGAAGTTATCGACAAAATCCA
GAAAAAGACACGTGAATTTCTGAACTCCAAAAAACATTGTATTGATAAATATAATTTCGAG

GACGTGCAGGTGGCACGCTACAAGCCGGGTCAGTATTATTATCACCACTATGACGGAGATG
ATTGTGATGATGCGTGCCCGAAAGACCAGCGCCTGGCGACGCTGATGGTGTATCTGAAAGC
GCCGGAGGAGGGTGGAGGCGGGGAGACTGATTTTCCAACTCTGAAAACCAAAATCAAGCCG
AAAAAGGGTACTTCTATCTTCTTTTGGGTTGCCGATCCGGTGACCCGCAAACTGTACAAAG
AAACGCTGCACGCCGGTCTGCCGGTAAAATCGGGAGAGAAAATCATCGCAAACCAATGGAT
CCGTGCTGTTAAATGA

pNX-TrxT_PHF8

ATGCACCATCATCATCATCATTCTTCTGGTATGAGCGATAAAATTATTCACCTGACTGACG
ACAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATCCTCGTCGATTTCTGGGCAGA
GTGGTGCGGTCCGTGCAAAATGATCGCCCCGATTCTGGATGAAATCGCTGACGAATATCAG
GGCAAACTGACCGTTGCAAAACTGAACATCGATCAAAACCCTGGCACTGCGCCGAAATATG
GCATCCGTGGTATCCCGACTCTGCTGCTGTTCAAAAACGGTGAAGTGGCGGCAACCAAAGT
GGGCGCACTGTCTAAAGGTCAGTTGAAAGAGTTCCTCGACGCTAACCTGGCCGGTACCGAG
AACTTGTACTTCCAATCCATGCCAGTGAAGACCGGGAGCCCTACGTTCGTCAGAGAGCTCC
GGAGTAGGACTTTTGACAGCTCAGATGAAGTGATTCTGAAGCCCACTGGAAATCAACTGAC
CGTGGAATTCCTGGAAGAAAATAGCTTCAGTGTGCCCATCCTGGTCCTGAAGAAGGATGGG
TTGGGCATGACGCTGCCCTCGCCATCATTCACTGTGAGGGATGTTGAACACTATGTTGGTT
CTGACAAAGAGATTGATGTGATTGATGTGACCCGCCAGGCTGACTGCAAGATGAAGCTTGG
TGATTTTGTGAAATACTATTACAGCGGGAAGAGGGAGAAAGTCCTCAATGTCATTAGTTTG
GAATTCTCTGATACCAGACTTTCTAACCTTGTGGAGACACCGAAGATTGTTCGAAAGCTGT
CATGGGTCGAAAACTTGTGGCCAGAGGAATGTGTCTTTGAGAGACCCAATGTACAGAAGTA
CTGCCTCATGAGTGTGCGAGATAGCTATACAGACTTTCACATTGACTTTGGTGGCACCTCT
GTCTGGTACCATGTACTCAAGGGTGAAAAGATCTTCTACCTGATCCGCCCAACAAATGCCA
ATCTGACTCTCTTTGAGTGCTGGAGCAGTTCCTCTAATCAGAATGAGATGTTCTTTGGGGA
CCAGGTGGACAAGTGCTACAAGTGTTCCGTGAAGCAAGGACAGACACTTTTCATTCCCACA
GGGTGGATCCATGCTGTGCTGACGCCTGTGGACTGCCTTGCCTTTGGAGGGAACTTCTTAC
ACAGCCTTAACATCGAGATGCAGCTCAAAGCCTATGAGATTGAGAAGCGGCTGAGCACAGC
AGACCTCTTCAGATTCCCCAACTTTGAGACCATCTGTTGGTATGTGGGAAAGCACATCCTG
GACATCTTTCGCGGTTTGCGAGAGAACAGGAGACACCCTGCCTCCTACCTGGTCCATGGTG
GCAAAGCCTTGAACTTGGCCTTTAGAGCCTGGACAAGGAAAGAAGCTCTGCCAGACCATGA
GGATGAGATCCCGGAGACAGTGCGAACCGTACAGCTCATTAAAGATCTGGCCAGGGAGATC
CGCCTGGTGGAAGACATCTTCCAACAGAACTGACAGTAAAGGTGGATACGGATCCGAATTC
GAGCTCCGTCGACAAGCTT

TEV-DM-Prk793 Plasmid[306]

ATGAAAACTGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTC
TCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCC
GGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATC
TTCTGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCC
CGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGG
CAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTG
CTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGA
AAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGC
TGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGC
GTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAAC
ACATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGC
GATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGT
GTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCG
CAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCT
GCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTG
AAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCC
AGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTAC
TGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAG
ACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGGATCGAGGGAA
GGGGAGAAAATCTTTATTTTCAAGGTCATCATCATCATCATCATCATGGAGAAAGCTTGTT
TAAGGGGCCGCGTGATTACAACCCGATATCGAGCACCATTTGTCATTTGACGAATGAATCT
GATGGGCACACAACATCGTTGTATGGTATTGGATTTGGTCCCTTCATCATTACAAACAAGC
ACTTGTTTAGAAGAAATAATGGAACACTGTTGGTCCAATCACTACATGGTGTATTCAAGGT

CAAGAACACCACGACTTTGCAACAACACCTCATTGATGGGAGGGACATGATAATTATTCGC
ATGCCTAAGGATTTCCCACCATTTCCTCAAAAGCTGAAATTTAGAGAGCCACAAAGGGAAG
AGCGCATATGTCTTGTGACAACCAACTTCCAAACTAAGAGCATGTCTAGCATGGTGTCAGA
CACTAGTTGCACATTCCCTTCATCTGATGGCATATTCTGGAAGCATTGGATTCAAACCAAG
GATGGGCAGTGTGGCAGTCCATTAGTATCAACTAGAGATGGGTTCATTGTTGGTATACACT
CAGCATCGAATTTCACCAACACAAACAATTATTTCACAAGCGTGCCGAAAAACTTCATGGA
ATTGTTGACAAATCAGGAGGCGCAGCAGTGGGTTAGTGGTTGGCGATTAAATGCTGACTCA
GTATTGTGGGGGGGCCATAAAGTTTTCATGGTGAAACCTGAAGAGCCTTTTCAGCCAGTTA
AGGAAGCGACTCAACTCATGAATCGTCGTCGCCGTCGCTAA