Orthogonal proteomics approaches to discover novel non-haem iron- and 2-oxoglutarate-dependent dioxygenase-catalysed prolyl hydroxylation in the mammalian proteome

Hamish Pegg

University College London
and
The Francis Crick Institute
PhD Supervisor: Professor Sir Peter Ratcliffe

A thesis submitted for the degree of
Doctor of Philosophy
University College London
May 2022

Declaration

I, Hamish Pegg, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Oxygen is used in a variety of biochemical processes, including regulation of the transcription factor hypoxia inducible factor (HIFα). Oxygen-sensitive prolyl hydroxylation of HIFα, which is catalysed by the HIFα prolyl 4-hydroxylases (PHDs), enables von Hippel Lindau protein (pVHL) to bind and ubiquitylate HIFα leading to its degradation by the proteasome. The PHDs belong to a large family of non-haem iron- and 2-oxoglutarate-dependent dioxygenases (2OGDs), which includes several other prolyl hydroxylases. Thus, the integration of oxygen into proteins, particularly on proline residues, is a fundamental post-translational modification (PTM). Despite this, the number of known prolyl hydroxylase substrates is relatively small, so orthogonal proteomics approaches using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) were performed to survey the proteome for prolyl hydroxylation in a variety of cellular contexts, including hypoxia, PHD- and pVHL-defective cells, pan-2OGD inhibition, and cells treated with a proteasome inhibitor. Qualitative analysis of hydroxyproline assignments highlighted the confounding role of oxidation occurring on residues in proximity to the reported site of hydroxyproline. With this in mind, the sensitivity of hydroxyproline detection in LC-MS/MS experiments was improved by extensive high pH off-line fractionation and the development of a hydroxyproline affinity purification reagent that enabled endogenous hydroxyproline-containing HIFα peptides to be identified. In total, 49 novel hydroxyproline sites were identified on 34 proteins, including prolyl hydroxylation of 8 protein disulphide isomerases. Further research is required to generate hydroxylase assignments that might provide the framework for functional studies. In addition to identification of novel hydroxyproline sites, oxygen decreased the abundance of 14 proteins and further research is required to determine the mechanism underpinning this apparent decay. Overall, the application of orthogonal LC-MS/MS approaches enabled the identification of many novel hydroxyproline sites and proteins that might be involved in dynamic oxygen-sensing pathways.

Impact Statement

Prolyl hydroxylation is a post-translational modification (PTM) that integrates oxygen into multiple biochemical processes. Despite the identification of several prolyl hydroxylases, the total number of substrates for these enzymes is relatively small, especially when compared other modifying enzymes (e.g., kinases and ubiquitin E3 ligases). With this in mind, the work performed in this thesis aimed to identify novel prolyl hydroxylase substrates and other proteins that might be involved in dynamic oxygen signalling pathways. Research methods were developed to achieve this aim and the identification of 49 secure hydroxyproline sites is likely to benefit future scholarship. The initial results of this thesis indicated that secure hydroxyproline assignments in mass spectrometry data are rare. Analysis of publicly deposited "benchmark" proteomics data emphasised that the interrogation of MS1 and MS2 spectra is important to corroborate the assignments made by the software, which can generate incorrect assignments. Therefore, methodological approaches to discover hydroxyproline were emphasised in this work.

Experiments using precursor ion quantification demonstrated altered prolyl hydroxylation in responses to hypoxia and inhibitor treatment. Importantly, these experiments were adapted to capture both short and long half-life proteins. Since prolyl hydroxylation is a rare PTM, I sought an enrichment method that would enable the identification of low abundance hydroxyproline sites in the proteome. The development of recombinant von Hippel Lindau protein (pVHL) as an affinity purification reagent to enrich hydroxyproline-containing peptides was successful, as demonstrated by the identification of hydroxyproline-containing HIFα peptides. This particular result represents a clear methodological advancement in the field and enabled other hydroxyproline-containing peptides to be identified that were otherwise beyond the limits of detection.

VHL disease affects 1 in 36,000 people globally and is characterised by mutations affecting the hydroxyproline binding pocket of pVHL. These mutations manifest as multiple tumour types and it is therefore of significant interest to understand the mechanistic basis of disease progression to aid therapeutic development. Importantly, there could be an oncogenic function of novel pVHL interactors in the context of inactivating *VHL* mutations. The novel hydroxyproline-containing interactors of pVHL

observed in this thesis could therefore be involved in the mechanistic pathway in which failure of pVHL to recognize prolyl hydroxylation contributes to disease progression in VHL disease.

At the commencement of my PhD project, inhibitors of the HIFα prolyl hydroxylases (PHDs) were in development for clinical use. Additionally, other prolyl hydroxylases may be targeted for therapeutic benefit. Therefore, the identification of additional substrates to these enzymes, which was the focus of this thesis, might lead to greater understanding of the effects of pharmacological inhibitors targeting the prolyl hydroxylases.

Overall, the biological discoveries and methodological advances made in this body of work are likely to have an immediate benefit to the academic community. Follow-up studies using this body of work as the basis for future research could advance the understanding of therapeutic opportunities and tractability of prolyl hydroxylase-related pathways.

Acknowledgements

I would like to offer my deepest gratitude to Professor Sir Peter Ratcliffe, who has supervised me through my PhD project. The scientific education and pastoral support I have received has been exemplary. I would also like to express my sincere thanks to Dr Matthew Cockman, who has overseen my technical training.

To the Hypoxia Biology Laboratory, I am grateful for your companionship and the professional standards you set as an example to learn from. I would like to thank Dr Yoichiro Sugimoto for his perceptive discussions and contributions to data analysis.

I am grateful to the Proteomics, Structural biology, and Peptide chemistry science technology platforms (STPs) for their assistance throughout my research projects. In particular, I would like to extend my gratitude to Dr Joanna Kirkpatrick, who provided training and expert advice in mass spectrometry related techniques that were critical to the work I performed in this thesis.

For the valuable contributions to the research projects performed in this thesis, I extend my sincere appreciation to Professor Johanna Myllyharju, Dr Antti Salo, and Dr Ya-Min Tian.

My sincere gratitude is offered to Dr Mia Monnius, Dr Marcel Tuppi, and Dr Graeme Hewitt for their insightful comments and suggestions.

I would like to extend my thanks to Professor Katrin Rittinger, Dr Bram Snijders, and the late Dr John Timms, for their contributions as my PhD student committee.

For the opportunity they have afforded me by funding this research, I am grateful to the Medical Research Council, Cancer research UK, and The Wellcome Trust.

I am deeply grateful to the friends, family, and the Academic Training team who have supported me throughout the PhD programme. Particular thanks go to Mia Monnius, who is my rock.

Table of Contents

Table of figures

Figures 1, 3, 7, 22, 23, 48, and 72 were created with Biorender.com

Figure 13 was created with WebLogo (Crooks et al., 2004).

List of tables

Abbreviations

Genes and gene products

Human

Table 1 | Human gene and gene products described in this thesis

Gene ID	Uniprot Accession	Protein name(s) used in this thesis
ACACB	O00763	**ACACB**, Acetyl-CoA carboxylase 2
ACSL4	O60488	**ACSL4**, Long-chain-fatty-acid-CoA ligase 4
ACTB	P60709	**ACTB**, Beta-actin
ADO	Q96SZ5	**ADO**, 2-aminoethanethiol dioxygenase
ADRB2	P07550	**ADRB2**, Beta-2 adrenergic receptor
ADSL	P30566	**ADSL**, Adenylosuccinate lyase
AGR2	O95994	**AGR2**, Anterior gradient protein 2
AGR3	Q8TD06	**AGR3**, Anterior gradient protein 3
AKT1	P31749	**AKT1**, RAC-alpha serine/threonine-protein kinase
ARG2	P78540	**ARG2**, Arginase 2
ARHGAP35	Q9NRY4	**ARHGAP35**, Rho GTPase-activating protein 35
ARNT	P27540	**HIF1β**, Hypoxia inducible factor 1 beta
ARRB2	P32121	**ARRB2**, Beta-arrestin-2
ATE1	O95260	**ATE1**, Arginyl-tRNA-protein transferase 1
ATF4	P18848	**ATF4**, Activating transcription factor 4
B4GALT1	P15291	**B4GALT1**, Beta-1,4-galactosyltransferase 1
BCKDK	O14874	**BCKDK**, Branched-chain alpha-ketoacid dehydrogenase kinase
BCL2L11	O43521	**BCL2L11**, Bcl-2-like protein 11
BRD4	O60885	**BRD4**, Bromodomain-containing protein 4
BTN3A1	O00481	**BTN3A1**, Butyrophilin subfamily 3 member A1
CAPN2	P17655	**CAPN2**, Calpain-2 catalytic subunit
CASC4	Q6P4E1	**CASC4**, Cancer susceptibility candidate gene 4 protein
CBX4	O00257	**CBX4**, Chromobox protein homolog 4
CDKN1C	P49918	**CDKN1C**, Cyclin-dependent kinase inhibitor 1C
CENPN	Q96H22	**CENPN**, Centromere protein N
CEP192	Q8TEP8	**CEP192**, Centrosomal protein of 192 kDa
CERKL	Q49MI3	**CERKL**, Ceramide kinase-like protein

CLTC	Q00610	**CLTC**, Clathrin heavy chain 1
COL18A1	P39060	**COL18A1**, Collagen alpha-1 (XVIII) chain
COL19A1	Q14993	**COL19A1**, Collagen alpha-1 (XIX) chain
COL1A1	P02452	**COL1A1**, Collagen alpha-1 (I) chain
COL22A1	Q8NFW1	**COL22A1**, Collagen alpha-1 (XXII) chain
COL2A1	P02458	**COL2A1**, Collagen alpha-1 (II) chain
COL3A1	P02461	**COL3A1**, Collagen alpha-1 (III) chain
COL4A1	P02462	**COL4A1**, Collagen alpha-1 (IV) chain
COL4A2	P08572	**COL4A2**, Collagen alpha-2 (IV) chain
COL4A4	P53420	**COL4A4**, Collagen alpha-4 (IV) chain
COL4A6	Q14031	**COL4A6**, Collagen alpha-6 (IV) chain
COL5A3	P25940	**COL5A3**, Collagen alpha-3 (V) chain
COL6A1	P12109	**COL6A1**, Collagen alpha-1 (VI) chain
COL6A2	P12110	**COL6A2**, Collagen alpha-2 (VI) chain
COL7A1	Q02388	**COL7A1**, Collagen alpha-1 (VII) chain
COL8A1	P27658	**COL8A1**, Collagen alpha-1 (VIII) chain
CPA4	Q9UI42	**CPA4**, Carboxypeptidase 4
CPT1B	Q92523	**CPT1B**, Carnitine O-palmitoyltransferase 1B
CTHRC1	Q96CG8	**CTHRC1**, Collagen triple helix repeat-containing protein 1
DGKI	O75912	**DGKI**, Diacyl glycerol kinase iota
DNAJC10	Q8IXB1	**DNAJC10**, DnaJ homolog subfamily C member 10
DOCK1	Q14185	**DOCK1**, Dedicator of cytokinesis protein 1
DYNC1LI1	Q9Y6G9	**DYNC1LI1**, Cytoplasmic dynein 1 light intermediate chain 1
DYRK1A	Q13627	**DYRK1A**, Dual specificity tyrosine-phosphorylation-regulated kinase 1A
DYRK1B	Q9Y463	**DYRK1B**, Dual specificity tyrosine-phosphorylation-regulated kinase 1B
EEF2	P13639	**EEF2**, Eukaryotic elongation factor 2
EEF2K	O00418	**EEF2K**, Eukaryotic elongation factor 2 kinase
EGLN1	Q9GZT9	**PHD2**, Prolyl hydroxylase domain containing protein 2
EGLN2	Q96KS0	**PHD1**, Prolyl hydroxylase domain containing protein 1
EGLN3	Q9H6Z9	**PHD3**, Prolyl hydroxylase domain containing protein 3
EHMT2	Q96KQ7	**EHMT2**, Euchromatic histone-lysine N-methyltransferase 2
ELOB	Q15370	**Elongin B**
ELOC	Q15369	**Elongin C**
EPAS1	Q99814	**HIF2α**, EPAS1, Hypoxia inducible factor 2 alpha

EPO	P01588	**EPO**, Erythropoietin
EPOR	P19235	**EPOR**, Erythropoietin receptor
ERP44	Q9BS26	**ERP44**, Endoplasmic reticulum resident protein 44
F3	P13726	**F3**, Tissue factor
FKBP10	Q96AY3	**FKBP10**, Peptidyl-prolyl cis-trans isomerase FKBP10
FLNA	P21333	**FLNA**, Filamin A
FLNB	O75369	**FLNB**, Filamin B
FLOT2	Q14254	**FLOT2**, Flotilin 2
FN1	P02751	**FN1**, Fibronectin
FOXO3	O43524	**FOXO3**, Forkhead box protein O3
GOLM1	Q8NBJ4	**GOLM1**, Golgi membrane protein 1
HIF1A	Q16665	**HIF1α**, Hypoxia inducible factor 1 alpha
HIF1AN	Q9NWT6	**FIH**, Factor inhibiting hypoxia inducible factor
HIF3A	Q9Y2N7	**HIF3α**, Hypoxia inducible factor 3 alpha
HSPH1	Q92598	**HSPH1**, Heat shock protein 105 kDa
HYOU1	Q9Y4L1	**HYOU1**, Hypoxia up-regulated protein 1
IGFBP3	P17936	**IGFBP3**, Insulin-like growth factor-binding protein 3
IKBKB	O14920	**IKBKB**, Inhibitor of nuclear factor kappa-B kinase subunit beta
IL32	P24001	**IL-32**, Interleukin 32
JMJD6	Q6NYC1	**JMJD6**, Jumonji domain-containing protein 6
JMJD8	Q96S16	**JMJD8**, Jumonji domain-containing protein 8
KDM2A	Q9Y2K7	**KDM2A**, Lysine-specific demethylase 2A
KDM5A	P29375	**KDM5A**, Lysine-specific demethylase 5A
KDM5B	Q9UGL1	**KDM5B**, Lysine-specific demethylase 5B
KHDRBS1	Q07666	**KHDRBS1**, KH domain-containing, RNA-binding, signal transduction-associated protein 1
LMAN1	P49257	**LMAN1**, Protein ERGIC-53
MAGT1	Q9H0U3	**MAGT1**, Magnesium transporter protein 1
MAPK6	Q16659	**MAPK6**, Mitogen-activated protein kinase 6
MAPK7	Q13164	**MAPK7**, Mitogen-activated protein kinase 7
MDC1	Q14676	**MDC1**, Mediator of DNA damage checkpoint protein 1
MESD	Q14696	**MESD**, Mesoderm development protein
MLXIPL	Q9NP71	**MLXIPL**, Carbohydrate-responsive element-binding protein
MMP2	P08253	**MMP2**, 72 kDa type IV collagenase
MRC2	Q9UBG0	**MRC2**, C-type mannose receptor 2
MUC5B	Q9HC84	**MUC5B**, Mucin 5B

MYB	P10242	**MYB**, Transcriptional activator Myb
MYO10	Q9HD67	**MYO10**, Unconventional myosin X
NDRG3	Q9UGV2	**NDRG3**, N-myc downstream-regulated gene 3 protein
NDUFA4L2	Q9NRX3	**NDUFA4L2**, NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 4-like 2
NUDT1	P36639	**NUDT1**, Oxidized purine nucleoside triphosphate hydrolase
NUP214	P35658	**NUP214**, 214 kDa nucleoporin
OGFOD1	Q8N543	**OGFOD1**, 2-oxoglutarate- and iron-dependent oxygenase domain-containing protein 1
OS9	Q13438	**OS9**, Amplified in osteosarcoma 9
OTUB1	Q96FW1	**OTUB1**, OUT domain-containing ubiquitin aldehyde-binding protein 1
P3H1	Q32P28	**Prolyl 3-hydroxylase 1**
P3H2	Q8IVL5	**Prolyl 3-hydroxylase 2**
P3H3	Q8IVL6	**Prolyl 3-hydroxylase 3**
P3H4	Q92791	**Prolyl 3-hydroxylase 4**
P4HA1	P13674	**C-P4Hα1**, Collagen prolyl 4-hydroxylase subunit alpha 1
P4HA2	O15460	**C-P4Hα2**, Collagen prolyl 4-hydroxylase subunit alpha 2
P4HA3	Q7Z4N8	**C-P4Hα3**, Collagen prolyl 4-hydroxylase subunit alpha 3
P4HB	P07237	**P4HB**, prolyl 4-hydroxylase subunit beta
P4HTM	Q9NXG6	**P4H-TM**, Transmembrane prolyl 4-hydroxylase
PAM	P19021	**PAM**, Peptidyl-glycine alpha-amidating monooxygenase
PARP2	Q9UGN5	**PARP2**, Poly [ADP-ribose] polymerase 2
PAX2	Q02962	**PAX2**, Paired box protein Pax-2
PCOLCE	Q15113	**PCOLCE**, Procollagen C-endopeptidase enhancer 1
PDE4D	Q08499	**PDE4D**, cAMP-specific 3',5'-cyclic phosphodiesterase 4D
PDIA2	Q13087	**PDIA2**, Protein disulphide isomerase A2
PDIA3	P30101	**PDIA3**, Protein disulphide isomerase A3
PDIA4	P13667	**PDIA4**, Protein disulphide isomerase A4
PDIA5	Q14554	**PDIA5**, Protein disulphide isomerase A5
PDIA6	Q15084	**PDIA6**, Protein disulphide isomerase A6
PDILT	Q8N807	**PDILT**, Protein disulphide-isomerase-like protein of the testis
PFKFB2	O60825	**PFKFB2**, 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 2
PFKFB4	Q16877	**PFKFB4**, 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 4
PFKP	Q01813	**PFKP**, ATP-dependent 6-phosphofructokinase, platelet type
PIAS4	Q8N2W9	**PIAS4**, Protein inhibitor of activated STAT protein 4

PKM	P14618	**PKM**, Pyruvate kinase muscle isozyme
PLOD1	Q02809	**PLOD1**, Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1
PLOD2	O00469	**PLOD2**, Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2
PLOD3	O60568	**PLOD3**, Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3
PNRC1	Q12796	**PNRC1**, Proline-rich nuclear receptor coactivator 1
POLR2A	P24928	**POLR2A**, DNA-directed RNA polymerase II subunit RPB1
PPP2R2A	P63151	**PPP2R2A**, Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform
PRKCSH	P14314	**PRKCSH**, Glucosidase 2 subunit beta
RAB20	Q9NX57	**RAB20**, Ras-related protein Rab-20
RGS4	P49798	**RGS4**, regulator of G-signalling protein 4
RGS5	O15539	**RGS5**, regulator of G-signalling protein 5
RNF4	P78317	**RNF4**, RING finger protein 4
RPS23	P62266	**RPS23**, 40S ribosomal protein S23
RYDEN	Q9NUL5	**RYDEN**, Repressor of yield of DENV protein
SCYL2	Q6P3W7	**SCYL2**, SCY1-like protein 2
SDF2L1	Q9HCN8	**SDF2L1**, Stromal cell-derived factor 2-like protein 1
SERPINH1	P50454	**SERPINH1**, 47 kDa heat shock protein
SFMBT1	Q9UHJ3	**SFMBT1**, Scm-like with four MBT domains protein 1
SIPA1L2	Q9P2F8	**SIPA1L2**, Signal-induced proliferation-associated 1-like protein 2
SKP1	P63208	**SKP1**, S-phase kinase-associated protein 1
SKP2	Q13309	**SKP2**, S-phase kinase-associated protein 2
SLC38A10	Q9HBR0	**SLC38A10**, Putative sodium-coupled neutral amino acid transporter 10
SPRY2	O43597	**SPRY2**, Protein sprouty homolog 2
SREBP	P36956	**SREBP**, Sterol regulatory element-binding protein 1
SUN2	Q9UH99	**SUN2**, SUN domain-containing protein 2
TBK1	Q9UHD2	**TBK1**, TANK-binding kinase 1
TELO2	Q9Y4R8	**TELO2**, Telomere length regulation protein TEL2 homolog
TET2	Q6ND21	**TET2**, Methylcytosine dioxygenase TET2
TET3	O43151	**TET3**, Methylcytosine dioxygenase TET3
THRA	P10827	**THRA**, Thyroid hormone receptor alpha
TMEM109	Q9BVC6	**TMEM109**, Transmembrane protein 109
TMX1	Q9H3N1	**TMX1**, Thioredoxin-related transmembrane protein 1
TMX2	Q9Y320	**TMX2**, Thioredoxin-related transmembrane protein 2
TMX3	Q96JJ7	**TMX3**, Thioredoxin-related transmembrane protein 3

TMX4	Q9H1E5	**TMX4**, Thioredoxin-related transmembrane protein 4
TNS3	Q68CZ2	**TNS3**, Tensin 3
TP53	P04637	**TP53**, Cellular tumor antigen p53
TRPA1	O75762	**TRPA1**, Transient receptor potential cation channel subfamily A member 1
TUBB3	Q13509	**TUBB3**, Tubulin beta-3 chain
TUSC3	Q13454	**TUSC3**, Tumour suppressor candidate 3
TXNDC12	O95881	**TXNDC12**, Thioredoxin domain-containing protein 12
TXNDC15	Q96J42	**TXNDC15**, Thioredoxin domain-containing protein 15
TXNDC5	Q8NBS9	**TXNDC5**, Thioredoxin domain-containing protein 5
UBC	P0CG48	**Ubiquitin**, Polyubiquitin C
UBE2N	P61088	**UBE2N**, Ubiquitin-conjugating enzyme E2 N
UBE2V1	Q13404	**UBE2V1**, Ubiquitin-conjugating enzyme E2 variant 1
URB1	O60287	**URB1**, Nucleolar pre-ribosomal associated protein 1
USP38	Q8NB14	**USP38**, Ubiquitin carboxyl-terminal hydrolase 38
VCAN	P13611	**VCAN**, Versican core protein
VHL	P40337	**pVHL**, von Hippel Lindau protein
WIPF2	Q8TF74	**WIPF2**, WAS/WASL-interacting protein family member 2
ZHX2	Q9Y6X8	**ZHX2**, Zinc fingers and homeoboxes protein 2
ZNF395	Q9H8N7	**ZNF395**, Zinc finger protein 395

Mouse

Table 2 | Mouse genes and gene products described in this thesis

Gene ID	Uniprot Accession	Protein name(s) used in this thesis
Egln1	Q91YE3	**Phd2**, Prolyl hydroxylase domain-containing protein 2
Egln2	Q91YE2	**Phd1**, Prolyl hydroxylase domain-containing protein 1
Egln3	Q91UZ4	**Phd3**, Prolyl hydroxylase domain-containing protein 3
Fkbp10	Q61756	**Fkbp10**, FK506-binding protein 10
Hif1a	Q61221	**Hif1α**, Hypoxia inducible factor 1 alpha
Os9	Q8K2C7	**Os9**, Amplified in osteosarcoma 9
P4hb	P09103	**P4hb**, Prolyl 4-hydroxylase subunit beta
Pkm	P52480	**Pkm**, Pyruvate kinase muscle isozyme
Rps23	P62267	**Rps23**, 40S ribosomal protein S23
Serpinh1	P19324	**Serpinh1**, 47 kDa heat shock protein

Amino acids

Table 3 | Amino acids

Single letter code	Three letter code	Definition
A	Ala	Alanine
C	Cys	Cysteine
C’	-	Carbamidomethylated cysteine
D	Asp	Aspartate
E	Glu	Glutamate
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
HyK	-	Hydroxylysine
HyP	-	Hydroxyproline
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
X	Xaa	Any amino acid in the X position
Y	Yaa	Any amino acid in the Y position
Y	Tyr	Tyrosine

Other abbreviations

Table 4 | Abbreviations

Abbreviation	Definition

-10lgP	-10 * $\log_{10}$(p-value), where the p-value represents the likelihood the PSM is assigned by chance. Note, the term "-10lgP" is vendor-specific terminology from PEAKS®.
2,2-DIP	2,2'-dipyridine
2OG	2-oxoglutarate
2OGD	Non-haem iron- and 2-oxoglutarate-dependent dioxygenase
ACN	Acetonitrile
AGC	Automatic gain control
AHA	Azidohomoalanine
AMP	Adenosine monophosphate
AP-MS	Affinity purification-mass spectrometry
AScore	Ambiguity score; -10 * $\log_{10}$(p-value), where the p-value represents the likelihood the PTM is assigned by chance. Note, the term "AScore" is vendor-specific terminology from PEAKS®.
BONCAT	Biorthogonal noncanonical amino acid tagging
BSA	Bovine serum albumin
BV	Bed volume
C-P3H	Collagen prolyl 3-hydroxylase Note, this refers to all four of the collagen prolyl 3-hyroxylases.
C-P4H	Collagen prolyl 4-hydroxylase Note, this refers to the active heterotetramer composed of two alpha subunits (e.g., C-P4Hα1) and two beta subunits (e.g., P4HB).
ccRCC	Clear cell renal cell carcinoma
CHAPS	3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate
CHX	Cycloheximide
CID	Collision induced dissociation
CRISPR	Clustered regularly interspaced short palindromic repeats
DDA	Data-dependent acquisition
DdGNT1	*Dictyostelium discoideum* [Skp1-protein]-hydroxyproline N-acetylglucosaminyltransferase (Uniprot accession Q8T1C6)
DdSKP1	*Dictyostelium discoideum* SCF ubiquitin ligase complex protein SKP1a (Uniprot accession P52285)
DFO	Deferoxamine
DIA	Data-independent acquisition
DMEM	Dulbecco's minimal essential media
DMEM:F12	Dulbecco's modified essential media/Nutrient mixture F-12

DMOG	Dimethyloxalylglycine
ΔRT	Change in retention time
DTT	Dithiothreitol
ECM	Extracellular matrix
eHAP	Fully-haploid engineered HAP1 cells
ER	Endoplasmic reticulum
ERT	Estimated retention time
ETD	Electron transfer dissociation
FA	Formic acid
FBS	Foetal bovine serum
FDR	False discovery rate
FRET	Förster resonance energy transfer
GASP	Gel aided sample preparation method
GlcNAc	N-acetylglucosamine
GLM	Generalised linear model
GST-TS-VBC	Trimeric complex comprising GST-TS-VHL, Elongin B, and Elongin C
GST-TS-VHL	Glutathione S-transferase-Twin-Strep-tag®-von Hippel Lindau protein
GST-VBC	Trimeric complex comprising GST-VHL, Elongin B, and Elongin C
GST-VHL	Glutathione S-transferase-von Hippel Lindau protein
HAB	Haemangioblastoma
HCD	Higher energy C-trap dissociation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIF	Hypoxia inducible factor Note, this refers to the active heterodimer composed of an alpha subunit (e.g., HIF1α) and a beta subunit (e.g., HIF1β).
HIF1α MMAA	Synthetic tryptic HIF1α HyP564 peptide with Methionine to alanine mutations
HIFα	Hypoxia inducible factor alpha Note, this refers to HIF1α, HIF2α, and HIF3α.
HILIC	Hydrophilic interaction chromatography
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
IAA	Iodoacetamide
IEF	Isoelectric focusing
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-d1-thiogalactopyranoside

IQR	Interquartile range
iRT	Indexed retention time
IVT	*In vitro* transcription
IVTT	*In vitro* transcription coupled to translation
LB media	Luria-Bertani media
LC	Liquid chromatography
LC-MS/MS	Liquid chromatography coupled to tandem mass spectrometry
LDF	Linear discriminative function Note, the term "Linear discriminative function" is vendor-specific terminology from PEAKS®.
LRT	Likelihood ratio test
LysC	Endopeptidase LysC Note, this enzyme cleaves peptide bonds at the carboxyl side of lysine residues.
m/z	Mass to charge ratio
MALDI-TOF	Matrix assisted laser desorption ionisation-time of flight
MEF	Mouse embryonic fibroblast
MEM	Minimum essential media
MG132	Carbobenzoxy-l-leucyl-l-leucyl-l-leucinal
MIPS	Monoisotopic precursor selection
MoMo software	Modification motifs software
MS	Mass spectrometry
MS1	Mass spectrum of the precursor ion
MS2	Mass spectrum of fragment ions derived from the precursor ion
MS3	Mass spectrum of fragment ions derived from fragment ions in the MS2 spectrum
MSigDB	Molecular signature database
O	Oxygen Note, the term "O" refers to a single oxygen atom, whereas "O_2" refers to atmospheric oxygen.
Ofd1	*Saccharomyces cerevisiae* Prolyl 3,4-dihydroxylase ofd1 (Uniprot accession Q11120)
$^{\bullet}OH$	Hydroxyla radical
ORF	Open reading frame
p-adj	Adjusted p-value
PAL	Peptidyl-α-hydroxylglycine α-amidating lyase
PDI	Protein disulphide isomerase Note, this refers to all protein disulphide isomerase genes.

PE	Proline effect
PF	Peptide Features Note, the term “Peptide Features” is vendor-specific terminology from PEAKS®.
PHD	HIFα prolyl 4-hydroxylase Note, this refers to PHD1, PHD2, and PHD3.
PHM	Peptidylglycine α-hydroxylating monooxygenase
PRIDE database	Proteomics Identification database
PSM	Peptide spectrum match
PTM	Post translational modification
PVDF	Polyvinylidene fluoride
pVHL19	pVHL (aa54-213)
pVHL30	pVHL (aa1-213)
PX	ProteomeXchange
R0/K0	L-Arginine-HCl / L-Lysine-2HCl
R10/K8	L-Arginine-HCl, 13C6, 15N4 / L-Lysine-2HCl, 13C6, 15N2
R6/K4	L-Arginine-HCl, 13C6 / L-Lysine 2-HCl, 4,4,5,5-D4
RF lens	Radiofrequency lens
RNAi	Ribonucleic acid interference
ROS	Reactive oxygen species
rpm	Rotations per minute
RT	Retention time
RT-qPCR	Reverse transcriptase quantitative polymerase chain reaction
SCX	Strong cation exchange
SDS-PAGE	Sodium dodecylsulphate-polyacrylamide gel electrophoresis
SILAC	Stable isotope labelling of amino acids in cell culture
siRNA	Small interfering ribonucleic acid
SP4	Solvent precipitation SP3
SPR	Surface plasmon resonance
STP	Science technology platform
SUMO	Small ubiquitin-like modifier
TDA	Targeted data acquisition
TIC	Total ion chromatogram
timsTOF	Trapped ion mobility separation-time of flight
TKO	Triple knock-out Note, this refers to PHD1-3 TKO cells.
TMT	Tandem mass tag

UBD	Ubiquitin binding domain
UHPLC	Ultra-high performance liquid chromatography
VEP	VHL elution peptide
VHLdb	pVHL interactions database
WT	Wild-type
XIC	Extracted ion chromatogram
Y2H	Yeast two-hybrid

Chapter 1. Introduction

1.1 Protein hydroxylation is central to oxygen-sensing mechanisms

Animals utilise oxygen in every cell for oxidative phosphorylation and other biochemical reactions (Semenza, 1999). In mammals, oxygen is transported from the lungs to respiring tissues in erythrocytes. The proliferation and differentiation of erythroid precursors is regulated by the hormone erythropoietin (EPO) and this, in turn, is under regulation by the transcription factor hypoxia-inducible factor (HIF) (Wenger and Kurtz, 2011). HIF is a heterodimer comprised of a constitutively active HIF1β protein, and one of the three oxygen-labile HIFα proteins (HIF1α-3α). As indicated by the name, HIF transactivation is sensitive to oxygen tension. This is because oxygen is the rate-limiting factor for the catalytic activity of the HIF prolyl hydroxylases (PHDs). In normoxic conditions, PHDs catalyse prolyl hydroxylation of HIFα. This post-translational modification (PTM) (i.e., hydroxyproline) enables the ubiquitin E3 ligase von Hippel Lindau protein (pVHL) to interact with and ubiquitylate HIFα, leading to its proteasomal degradation. In hypoxia, the PHDs are inhibited, leading to HIFα stabilisation and dimerization with HIFβ, which activates a transcriptional cascade (Schofield and Ratcliffe, 2004). EPO and many other genes are upregulated by HIF, and this exerts control over a range of cellular and physiological processes (Bishop and Ratcliffe, 2015). Therefore, the role of prolyl hydroxylation of HIFα in co-ordinating systems-wide responses to oxygen is fundamental (Ratcliffe, 2013).

1.1.1 Transcriptional programming of the cell is co-ordinated by the PHD-HIFα-pVHL axis

A key discovery that triggered the identification of the PHD-HIFα-pVHL axis was the finding that EPO was produced in hepatoma cell lines, HepG2 and HeP3B, in response to hypoxia (Goldberg et al., 1987). This led to the discovery of the transcription factor that bound to the 3' enhancer of the *EPO* gene to induce expression of EPO in hypoxia (i.e., HIF) (Semenza and Wang, 1992; Wang et al., 1995). HIF-dependent gene expression in hypoxia was observed in virtually all mammalian cell lines and was found to occur irrespectively of EPO production in the cell lines tested (Maxwell et al., 1993).

Subsequent studies have identified over 500 HIF target genes in a variety of cell types that are upregulated in hypoxia (Mole et al., 2009; Schödel et al., 2013; Schödel et al., 2011).

Oxygen-dependent HIFα turnover was found to be regulated by pVHL (Maxwell et al., 1999). This led to the discovery that pVHL acts as the ubiquitin E3 ligase to prime HIFα for proteasomal degradation, and that tumour-associated mutations of *VHL* abrogated this function (Cockman et al., 2000). Several key reports subsequently demonstrated specific binding of HIFα to pVHL in a prolyl hydroxylation dependent manner (Ivan et al., 2001; Jaakkola et al., 2001; Yu et al., 2001). This interaction was found to occur at HIF1α P564, which was within a minimal pVHL-binding sequence in the oxygen-dependent degradation domain (ODD) (Cockman *et al.*, 2000; Tanimoto et al., 2000). Additionally, another hydroxyproline-dependent interaction between HIF1α and pVHL was observed at another site, P402, which shared a common motif with P564, 'LXXLAP' (Masson et al., 2001). The domains in which these hydroxylation sites are located are therefore referred to as the N-terminus and C-terminus ODD (NODD and CODD, respectively).

Three closely related non-haem iron- and 2-oxoglutarate-dependent dioxygenases (2OGDs), known as prolyl-hydroxylase domain-containing protein 1-3 (PHD1-3), were identified as the enzymes that hydroxylate HIF1α at P402 and P564 (Bruick and McKnight, 2001; Epstein et al., 2001). Biochemical characterization of these enzymes indicated that the K_m values were in the region of the concentration of atmospheric oxygen, meaning that even slight decreases in oxygen tension would inhibit their activity (Hirsilä et al., 2003). These kinetic parameters aligned with cellular data in which HIFα stability was tracked over a range of oxygen tensions, whereby graded hypoxic stabilisation of HIFα was observed (Jiang et al., 1996). This property indicates that the PHDs are effective oxygen sensors in cells.

Together, these discoveries provided the mechanism through which cells sense oxygen and achieve adaptive responses to hypoxia through transcriptional reprogramming of the cell. In recognition of the impact and legacy of these discoveries, Dr Gregg Semenza, Dr William Kaelin Jr., and Dr Peter Ratcliffe received the Nobel prize for Physiology or Medicine in 2019 (Kupferschmidt, 2019). It is therefore important to stress that the discovery of the PHDs led to the hypothesis that there might be other, non-HIFα, PHD

substrates with analogous functions in oxygen-sensing. This is a topic of contention that has been addressed throughout this thesis.

1.1.2 Other oxygen-sensing mechanisms

Protein hydroxylation coupled to ubiquitin-mediated proteasomal degradation is an evolutionary conserved phenomenon for oxygen sensing *via* HIFα (Loenarz et al., 2011; Schofield and Ratcliffe, 2004). This mode of oxygen sensing has also been reported to be conserved in plants but does not involve HIFα; oxygen instead regulates the stability of ethylene-response transcription factors through enzymatic cysteine oxidation coupled to the N-degron pathway (Gibbs et al., 2011; Licausi et al., 2011; Weits et al., 2014; White et al., 2017). In humans, a similar mechanism regulates the stability of RGS4, RGS5, and IL-32 (Lee et al., 2005; Masson et al., 2019). It is therefore possible that other dioxygenases, or enzymatic hydroxylation of non-prolyl residues, are involved in oxygen sensing and may achieve this function through proteolytic regulation of their substrates. This is considered to some extent in this thesis.

1.2 Protein hydroxylases of the 2OGD superfamily

Over 60 members of the 2OGD superfamily have been described, with many reported to catalyse protein hydroxylation on a range of residues (DHKNP) (Ploumakis and Coleman, 2015). Hydroxylation of these residues is believed to be irreversible (Singleton et al., 2011), so dynamic responses to hydroxylation are achieved through protein turnover and re-synthesis. However, only the hydroxylation of HIFα has so far proven to destabilise 2OGD substrates (Markolovic et al., 2015), which means these proteins exist in cells in their hydroxylated proteoforms and are unlikely to be involved in highly dynamic oxygen signalling pathways.

The 2OGD protein hydroxylases share a common catalytic fold and mechanism for hydroxylating their substrates (Markolovic *et al.*, 2015). These enzymes contain a catalytic triad ('His-Xaa-Asp/Glu…His') that binds Fe^{2+} in an octahedral manner. The other sites are co-ordinated by 2-oxoglutarate (2OG) and a water molecule that is displaced by oxygen when the protein substrate binds to the active site. In the absence of 2OG and the substrate, water molecules are co-ordinated to the Fe^{2+} ion (Markolovic *et al.*, 2015). When oxygen is introduced to the active site, oxidative decarboxylation

occurs; 2OG is converted to succinate, releasing one molecule of CO_2, and the substrate is hydroxylated (Markolovic *et al.*, 2015). Therefore, the 2OGDs can be inhibited by a range of interventions, which includes iron chelators (e.g., deferoxamine, DFO), iron displacement (e.g., $CoCl_2$), hypoxia, and small-molecule analogues of 2OG (e.g., dimethyloxalylglycine, DMOG) (Tian et al., 2011). Additionally, these enzymes can perform uncoupled decarboxylation, in which 2OG is reduced to succinate and CO_2 is produced, but the protein substrate is not hydroxylated (Counts et al., 1978; Li et al., 2004). Within the 2OGD superfamily there are 5 known types of prolyl hydroxylase, that catalyse the irreversible hydroxylation of proline residues and utilise oxygen, iron, and 2-oxoglutarate as cofactors (Figure 1.A). These enzymes are briefly discussed below.

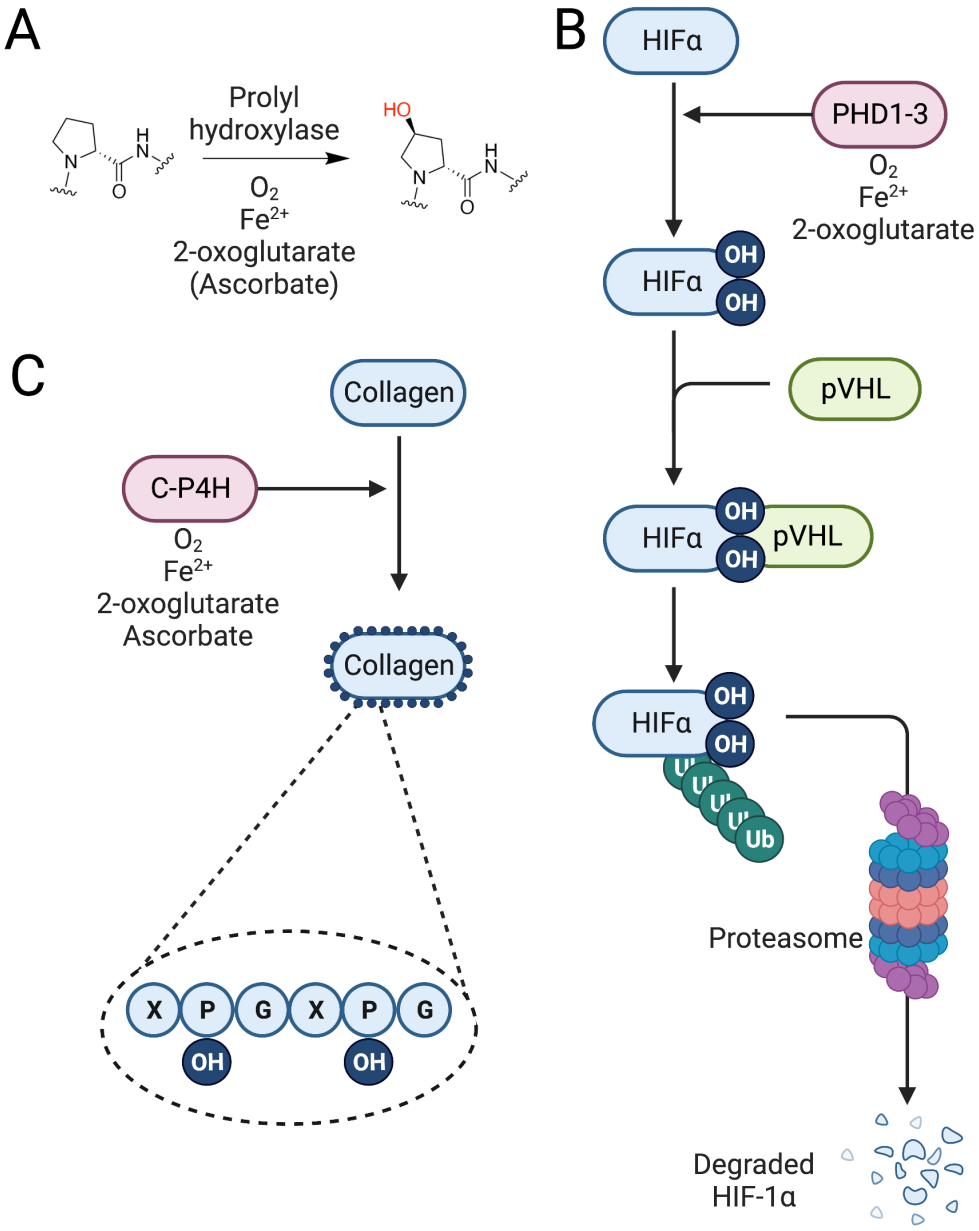


Figure 1 | Overview of prolyl hydroxylation.

A) Prolyl hydroxylation is catalysed by a subset of the non-haem iron- and 2-oxoglutarate-dependent dioxygenases (2OGDs) that utilise oxygen, iron, and 2-oxoglutarate as cofactors. For some enzymes, ascorbate is also required as a cofactor. Note, the schematic depicts the formation of trans-4-hydroxyproline; some prolyl hydroxylases generate other regioisomers, such as cis-3-hydroxyproline. B) Schematic depiction of HIF-1a prolyl hydroxylation at two distinct sites by PHD1-3. pVHL can bind to both hydroxyproline sites and ubiquitylate HIF-1a, which leads to its degradation by the proteasome. C) Schematic depiction of C-P4H catalysed prolyl hydroxylation of collagen. C-P4H uses ascorbate as a cofactor and targets proline residues in the 'Yaa' position of 'Xaa-Yaa-Gly' repeating triplets.

1.2.1 PHD1-3

As described in Section 1.1.1, the three PHD isoenzymes catalyse prolyl hydroxylation of HIFα in an oxygen-sensitive manner (Figure 1.B). These proteins are located in the nucleus and cytoplasm (Depping et al., 2015; Metzen et al., 2003; Steinhoff et al., 2009; Yasumoto et al., 2009), which facilitates regulation of HIFα in these compartments of the cell. Many non-HIFα substrates have been reported for the PHDs (Zurlo et al., 2016) but work at that time in the Ratcliffe laboratory had not confirmed activity of the PHD enzymes on many of these proteins in *in vitro* hydroxylation assays using recombinant enzymes and substrates. At the commencement of my thesis project, PHD inhibitors were in clinical trials to develop therapeutic agents for the treatment of anaemia and other hypoxia-related diseases (Yeh et al., 2017). Non-HIFα PHD substrates would be expected to contribute to off-target (i.e., HIF-independent) effects, which could be deleterious in the clinical setting. Therefore, the existence or otherwise of non-HIFα PHD substrates is a specific focus of research in this project.

1.2.2 C-P4H

Collagen prolyl 4-hydroxylase (C-P4H) is an endoplasmic reticulum (ER)-resident enzyme complex that contains two alpha and two beta subunits (Myllyharju and Kivirikko, 2001). The alpha subunits are encoded for by three genes, *P4HA1-3*, and the beta subunit is encoded for by one gene, *P4HB*. The respective gene products, prolyl 4-hydroxylase subunit alpha 1-3 (P4Hα1-3) and prolyl 4-hydroxylase subunit beta (P4HB), are assembled into a heterotetramer that is composed of two alpha subunits and two beta subunits (Gorres and Raines, 2010).

C-P4H catalyses prolyl 4-hydroxylation of proline residues occupying the ‘Yaa’ position of repeating ‘Xaa-Yaa-Gly’ triplets in collagen proteins (Figure 1.C) (Gorres and Raines, 2010). The terminology indicates the repeating triplet does not always contain a proline residue in the ‘Yaa’ position. However, when proline is present in this position it is normally hydroxylated by C-P4H. A number of non-collagen C-P4H substrates have been reported that contain collagen-like repeating ‘Xaa-Yaa-Gly’ sequences (Myllyharju and Kivirikko, 2001).

There are at least 28 collagen types that share a common structural motif, the triple helix, which is characterised by repeating 'Xaa-Yaa-Gly' triplets (Ricard-Blum, 2011). The triple helix is comprised of three collagen proteins, with both homotrimers and heterotrimers known to exist. It is common for fibrillar collagens to exhibit continuous 'Xaa-Yaa-Gly' repeat sequences but other collagen types (e.g., network-forming collagens) exhibit imperfections and interruptions to the repeating triplet. These interruptions are associated with increased flexibility (Ricard-Blum, 2011). Importantly, the persistence of prolyl hydroxylation in the interrupted triple-helical domains indicates that the repetition of 'Xaa-Yaa-Gly' sequences is not a strict requirement within C-P4H substrates. C-P4H-catalysed prolyl hydroxylation of elastin has also been reported, whereby hydroxylation occurs in 'Val-Pro-Gly-Val-Gly' sequences (Bhatnagar et al., 1978; Schmelzer et al., 2016), which is consistent with the interrupted 'Xaa-Yaa-Gly' sequences of collagens. Furthermore, the minimum substrate required for C-P4H-catalysed hydroxylation has been reported to be an 'Xaa-Pro-Gly' peptide, with 'Pro-Pro-Gly' being the preferred substrate (Gorres and Raines, 2010). C-P4H-catalysed prolyl hydroxylation has been reported for a number of substrates in which hydroxylation occurs at a single 'Pro-Gly' site in the protein sequence (Jiang et al., 2018; Ono et al., 2009; Qi et al., 2008; Rhoads and Udenfriend, 1969). Given the prevalence of 'Pro-Gly' sequences in the proteome, it is highly likely C-P4H has more substrates than have currently been reported.

1.2.3 C-P3H

Collagen prolyl 3-hydroxylase (C-P3H) is an ER resident enzyme that catalyses prolyl hydroxylation in the 'Xaa' position of 'Xaa-Yaa-Gly' sequences, where 'Yaa' is 4-hydroxyproline (Gorres and Raines, 2010). Despite the relatively simple substrate requirements, 3-hydroxyproline is a rare PTM (Hudson and Eyre, 2013) and, to the best of my knowledge, no non-collagen substrates have been reported.

1.2.4 P4H-TM

Transmembrane prolyl 4-hydroxylase (P4H-TM) is a prolyl 4-hydroxylase with a transmembrane domain that spans the ER membrane. *In vitro* hydroxylation assays have suggested that P4H-TM might be a fourth prolyl hydroxylase targeting HIFα (Koivunen

et al., 2007), but this has not yet been verified in cells. The catalytic domain is inside the ER lumen and would therefore be expected to catalyse hydroxylation of substrates within the ER (Koivunen *et al.*, 2007), which suggests HIFα is unlikely to be a cellular substrate of P4H-TM. It has been reported that *P4H-TM*-null mice secrete more EPO than their wild-type counterparts (Laitala et al., 2012), and P4H-TM deficient zebrafish displayed higher mRNA levels of *HIF1A* and some HIF target genes, including *EPO* (Hyvärinen et al., 2010). These results suggest that P4H-TM might catalyse prolyl hydroxylation of a non-HIFα protein to regulate gene expression in a HIF-independent manner, but that some crosstalk between these pathways exists.

1.2.5 OGFOD1

OGFOD1 is a nuclear 2OGD that catalyses prolyl hydroxylation of RPS23 P62 (Singleton et al., 2014). This reaction is conserved in yeast (Loenarz et al., 2014) and fruit flies (Katz et al., 2014). To the best of my knowledge, OGFOD1 is not known to catalyse prolyl hydroxylation on any other proteins.

For each prolyl hydroxylase, there is a possibility that other substrates exist, and this is considered throughout this thesis. Additionally, there may be other prolyl hydroxylases in the proteome that have not yet been characterised. Prolyl hydroxylation, like hydroxylation on other residues catalysed by the different 2OGDs, is understood to be an irreversible post-translational modification. Hydroxylation can directly lead to an altered interactome (e.g., HIF1α) (Jaakkola *et al.*, 2001), increased stability (e.g., collagen) (Shoulders and Raines, 2009), or fine-tune the catalytic activity of the substrate (e.g., RPS23 and the coding accuracy within the ribosome) (Loenarz *et al.*, 2014). This means that few novel prolyl hydroxylase substrates, should they exist, would be expected to be involved in dynamic oxygen-sensing pathways like HIFα. Instead, there may be alternate functional consequences of the hydroxyproline proteoforms that have not been observed before.

1.3 VHL disease

pVHL recognises the prolyl hydroxylated HIFα proteoforms and primes them for destruction by the proteasome (Jaakkola *et al.*, 2001). pVHL has been described as a tumour suppressor, and autosomal dominant inheritance of *VHL* mutations affects approximately 1 in 36,000 humans. Mutations of the *VHL* gene predispose affected individuals to a range of tumours, including clear-cell renal-cell carcinoma (ccRCC), hemangioblastoma (HAB), and pheochromocytoma (Kaelin, 2002). Different patterns of disease were recognised and subsequently associated with different types of mutations. Type 1 mutations result in loss of pVHL expression or affect protein folding. This results in HIFα stabilisation and typically manifests as hemangioblastoma or ccRCC. Type 2A, 2B, and 2C mutations are missense mutations, each with different clinical manifestations. Type 2A and 2B mutations stabilise HIFα and lead to upregulation of HIF target genes. They both lead to hemangioblastoma and pheochromocytoma. Type 2A mutations are linked to a low risk of ccRCC, while the risk for individuals with type 2B mutations is high. On the other hand, type 2C mutations do not perturb the ability of pVHL to target HIFα for proteolytic destruction in an oxygen-sensitive manner and these mutations appear to only manifest as pheochromocytomas (Kaelin, 2002).

HIFα has been described as the main oncogenic driver of VHL disease, and more recently it has been established that HIF2α is pro-oncogenic, whilst HIF1α might have tumour suppressor functions in VHL disease (Schödel et al., 2016). However, since type 2C mutations do not impact oxygen-sensitive HIFα proteolysis, the possibility of non-HIFα pVHL substrates contributing to tumour development in VHL disease has been explored (Ratcliffe, 2003). Although several interactions have been described for pVHL that are perturbed by mutations associated with VHL disease (Grosfeld et al., 2007; Hergovich et al., 2003; Hoffman et al., 2001), the precise mechanism of how these interactions contribute to the clinical manifestations in VHL disease has not been elucidated. Therefore, there remains a possibility that an as-yet undiscovered prolyl hydroxylase substrate is a pVHL target and contributes to the clinical manifestations of VHL disease patients with type 2C mutations. This hypothesis is explored in this thesis (Chapter 5).

1.4 The rationale and significance of this thesis

Oxygen sensing, both at the cellular level and with co-ordination of physiological responses to oxygen, occurs *via* PHD-catalysed prolyl hydroxylation (section 1.1). Other, non-2OGD, dioxygenases catalyse protein oxidation leading to proteolysis, which is a conserved signalling mechanism in plants and animals (Section 1.1.2). Additionally, many prolyl hydroxylases have been identified within the 2OGD superfamily, but the number of substrates is relatively few (Section 1.2). This contrasts with enzymes catalysing other PTMs involved in signalling processes, such as kinases and ubiquitin E3 ligases. For example, more than 38,000 phosphopeptides (Erickson et al., 2015) and over 23,000 ubiquitylation sites (van der Wal et al., 2018) have been reported, suggesting that the number of substrates for each enzyme is much higher than that of the prolyl hydroxylases.

Therefore, I hypothesised that there are more prolyl hydroxylase substrates that remain to be discovered. The functional consequences of prolyl hydroxylation are likely to be broad, in accordance with the diversity of functions of known prolyl hydroxylase substrates (section 1.2) and could be involved in oxygen signalling pathways.

A particular focus is applied to PHD1-3 because i) these dioxygenases display the greatest sensitivity to oxygen in the 2OGD superfamily, which could implicate PHD substrates in rapid oxygen-sensing processes, ii) dysregulation of HIFα appears to be oncogenic and this may be true of other PHD substrates (Section 1.3), iii) the existence of additional PHD substrates may regulate HIFα stability by competing for the active site of PHD1-3, and iv) the clinical inhibitors of PHD may have off-target (i.e., HIFα–independent) effects that ought to be understood.

With the above in mind, the discovery of novel prolyl hydroxylation sites in the proteome could lead to a greater understanding of dynamic signalling responses to hypoxia, as well as other biochemical processes in the cell that incorporate oxygen.

Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) has been applied to study a large variety of PTMs in the proteome and is the most suitable method to discover novel prolyl hydroxylation sites. Since this technology is used

comprehensively throughout this thesis, a brief introduction is provided below, with more detailed introductions to specific methodology provided in the relevant results chapters.

1.5 Mass spectrometry enables direct identification of site-specific protein hydroxylation in high-throughput workflows

As described in the previous section, the aim of this thesis is to discover novel sites of prolyl hydroxylation in the proteome that are catalysed by 2OGDs. LC-MS/MS provides direct interrogation of PTMs *in situ* from complex protein samples and is high-throughput (Aebersold and Mann, 2003). Therefore, this technology provides the most pragmatic detection method of as-yet undiscovered hydroxyproline sites in the proteome. The following section provides a brief overview of the workflows relevant to the work performed in this thesis.

1.5.1 LC-MS/MS overview

1.5.1.1 Sample preparation

Proteins are extracted from cell lysates and desalted to remove ions and detergents that ionise well in the mass spectrometer and would therefore cause relative suppression of the analyte. Following this, the proteins are digested into peptides of approximately 5-30 residues, which are the most suitable length for peptide sequencing (Steen and Mann, 2004). Trypsin is the most commonly used endopeptidase because it is cheap, robust, reproducible, and relatively stable. Trypsin specifically cleaves at the C-terminus of arginine and lysine residues, which generates peptides that protonate easily under acidic conditions (i.e., increasing the ionisation efficiency of these peptides) (Gillet et al., 2016). LysC, which is stable in conditions that denature trypsin (e.g., 7 M urea), cleaves specifically at the C-terminus of lysine residues and is sometimes used prior to trypsin to improve trypsin digestion efficiency (Steen and Mann, 2004). Trypsin digest, or the combination of LysC pre-treatment followed by trypsin digest, has been used throughout this thesis.

1.5.1.2 Peptide separation

Digested lysates are highly complex samples and there is a large range in the relative abundance of each peptide (i.e., dynamic range). Many peptides exhibit identical mass-to-charge ratio (*m/z*) and using MS alone would not allow detection of low-abundance species (Zhang et al., 2014). Therefore, peptides are separated prior to mass spectrometry. Liquid chromatography (LC) enables continuous separation of peptides, which can be performed on-line (i.e., peptides elute directly into the ionisation source of the mass spectrometer) or off-line (i.e., peptide fractions are collected in vials for further applications). Orthogonal fractionation methods can be combined to maximise peptide separation (Yates et al., 2009). In this thesis, high pH fractionation was performed off-line and acetonitrile gradients at low pH are coupled to the mass spectrometer. The functionality of this approach has been discussed elsewhere (Bekker-Jensen et al., 2017; Gilar et al., 2005).

1.5.1.3 Precursor ion selection

Partially separated ionised peptides are analysed in the mass spectrometer to determine the *m/z* of the intact precursor (MS1) and fragmented product (MS2) ions. Fragment ions can themselves be isolated and further fragmented (MS3) to provide additional information that is not available in the MS2 spectrum (Steen and Mann, 2004). Several different precursor ion selection methods are available, including data-dependent acquisition (DDA), data-independent acquisition (DIA), and targeted precursor ion selection. These have been reviewed elsewhere (Gillet *et al.*, 2016). The experiments performed in this thesis implement precursor ion selection in DDA mode. The acquisition of precursor ions and fragment ions occurs in iterative cycles whereby the instrument software makes real-time decisions about fragmentation of the precursor ion according to pre-defined criteria (Gillet *et al.*, 2016).

1.5.1.4 Tandem mass spectra

The series of fragment ions relating to a single precursor ion is known as the MS2 spectrum. In this thesis, collision-induced dissociation (CID) and higher energy C-trap dissociation (HCD) were used to break the precursor ion into fragment ions. The

fragmentation of precursor ions is typically achieved by collisions with an inert gas, such as nitrogen or argon, to break the amide bonds (Steen and Mann, 2004). In this mode of fragmentation, the product ions are typically b- and y-ions, in which the charge is retained at the N- and C-terminus, respectively. Additional ions might be generated that contain neutral losses. For example, the loss of ammonia or water is common during fragmentation. The MS2 spectra can be compared to a reference proteome containing the amino acid sequences of all proteins in that database, which enables peptide sequencing to be performed. (Aebersold and Mann, 2003).

1.5.1.5 Proteomics software

The raw data from the mass spectrometer can be analysed by various proteomic software programmes. These programmes compare the MS2 spectra with a reference proteome to determine the amino acid sequence of the precursor ion (Steen and Mann, 2004). PTMs can be included in the search algorithm, which increases the complexity of the proteome search but can lead to more successful peptide spectrum matches (PSMs) (Cox and Mann, 2011). Furthermore, the software includes algorithms to generate confidence scores (Eng et al., 2011), which can later be used to filter the data for peptides that have been confidently assigned by the software (Aebersold and Mann, 2003). PEAKS® software was used throughout this thesis to interpret mass spectrometry data. PEAKS® software performs *de novo* peptide sequencing and compares this sequence to the reference database, which improves the accuracy of PSMs (Tran et al., 2017). Various versions of this software have been used throughout this thesis as the software is updated.

1.5.1.6 Data analysis

As previously mentioned, proteomics software utilises probability-based scoring of PSMs, which provides a degree of confidence the peptide assignment is correct. PEAKS® includes a confidence score for PTMs (Beausoleil et al., 2006), as well as providing a filter for the relative ion intensity of the diagnostic ions. The diagnostic ions are fragment ions that i) contain, and ii) immediately precede the modified residue. This means the exact site at which a PTM is incorporated into the peptide sequence can be identified. If these ions are low abundance, or even non-existent, the confidence of the assignment is reduced. Since the majority of studies in this thesis are focused on identifying

hydroxyproline sites, manual inspection of MS2 spectra is performed throughout. These analyses consider the quality of the MS2 spectrum for peptide sequencing, with a particular focus on the diagnostic fragment ions, and are vital to identifying secure prolyl hydroxylation assignments. Therefore, a brief overview of the factors influencing peptide fragmentation is provided in the section below.

1.5.1.7 Peptide fragmentation during MS/MS and the relative proton mobility model

This section will briefly cover the factors influencing peptide fragmentation in MS/MS, which ultimately determines the quality of the MS2 spectrum for peptide sequencing and PTM localisation.

The proton mobility model was first introduced to account for proton migration in a peptide and the subsequent cleavage of peptide bonds (Wysocki et al., 2000). The mobile proton theory suggests cleavage of peptide bonds is initiated by the migration of a proton from the initial site of protonation to the peptide backbone. Basic residues (His, Lys, and Arg) sequester the proton so more energy is required to transfer the proton from the basic side chain to the peptide backbone to induce fragmentation (Barton and Whittaker, 2009). A relative proton mobility model was developed in which the proton mobility of any given peptide could be classified as “non-mobile”, “partially mobile”, or “mobile” (Kapp et al., 2003). Peptides containing more arginine residues than protons are defined as non-mobile. A partially mobile peptide contains fewer arginine residues than the number of protons, but the total number of basic residues in the peptide is greater than or equal to the number of protons. A mobile peptide contains fewer arginine residues (and total basic residues) than the number of protons in the peptide. The authors detail how both N- and C-terminal cleavage of any amino acid pair could be accurately predicted according to their model of relative proton mobility, but some inaccuracies remained. The relative proton mobility model was therefore integrated into a linear model that accounts for relative proton mobility in combination with all other factors affecting peptide fragmentation (Barton and Whittaker, 2009) to improve scoring of peptide sequence assignments in database searches (Barton et al., 2007). An understanding of these factors aids PTM localisation determination during manual MS2 inspection and is discussed further in Chapter 3.

1.5.2 Applications of LC-MS/MS to study protein hydroxylation: a historical perspective

The following section provides a brief overview of the application of LC-MS/MS to identifying enzymatic protein hydroxylation. The development of new technologies, reagents, and orthogonal methodologies have enabled several key discoveries of 2OGD-catalysed protein hydroxylation.

The development of electrospray ionisation and its application to the analysis of biomolecules by mass spectrometry (Fenn et al., 1989) soon led to the discovery of methionine oxidation in peptides and the neutral loss ion (-64 Da) that is characteristic of this modification (Lagerwerf et al., 1996). Oxidations were also reported to occur on other residues in peptides, including cysteine, tryptophan, tyrosine, and histidine (Kotiaho et al., 2000; Schöneich, 2000; Steen and Mann, 2001). For the purposes of this thesis, oxidation that is not catalysed by 2OGDs is referred to as "artefactual oxidation", and the exact function of non-2OGD-catalysed protein oxidation is not always clear. This type of oxidation is isobaric to 2OGD-catalysed protein hydroxylation, which can increase the difficulty of analysis for peptides containing residues prone to artefactual oxidation. This is discussed in more detail in Chapter 3.

Identification of hydroxyproline on HIFα peptides was achieved by comparing the chromatographic retention of synthetic analogues (Epstein et al., 2001) and by analysing the MS1 spectrum of HIFα-derived peptides using MALDI-TOF mass spectrometry (Ivan et al., 2001). These studies were closely followed by the identification of factor inhibiting HIF (FIH)-catalysed asparaginyl hydroxylation of HIFα by LC-MS/MS (Lando et al., 2002). The use of $^{18}O_2$ demonstrated that hydroxylation involves incorporation of an O atom from atmospheric oxygen, not water (Hewitson et al., 2002). Thus, mass spectrometry has been used to identify and discover enzymatic protein hydroxylation. This set the precedent for further, more complex studies of hydroxylase-substrate assignments.

Proteomic analyses of cells treated with siRNA targeting *HIF1AN*, the gene coding for FIH, led to the discoveries of IκBα and notch as novel substrates (Cockman et al., 2006; Coleman et al., 2007). Importantly, RNAi experiments targeting *HIF1AN* expression enabled quantitative analysis of site-specific hydroxylation in the novel substrates.

A novel interactomics method was developed to identify putative substrates of FIH (Cockman et al., 2009). This method used the pan-2OGD inhibitor, DMOG, to "trap" substrates as catalytic intermediates. This stabilises the interaction between the substrate and the catalytic domain of FIH throughout the affinity purification pipeline. To distinguish between DMOG-treated and untreated populations, stable isotope labelling of amino acids in cell culture (SILAC) was used (Cockman *et al.*, 2009; Steen and Mann, 2004). This method incorporates heavy lysine and arginine isotopes into the protein sequence so that all tryptic peptides from the samples being compared can be distinguished and enables relative quantitation of the precursor ions to be performed. After identifying candidates through the "substrate trapping" experiment, the hydroxylase assignments were generated by suppressing *HIF1AN* expression with siRNA (Cockman *et al.*, 2009). This exemplified the application of orthogonal proteomic approaches to identify novel 2OGD substrates.

In an alternative application of SILAC, a pulse-chase experiment revealed FIH-catalysed asparaginyl hydroxylation is not reversible (Singleton *et al.*, 2011). Although this study did not lead to the identification of novel substrates, it provided an important framework in which protein hydroxylation could be tracked over time.

Identification of other 2OGD substrates has been achieved using the methods described above and conventional interaction assays such as immunoprecipitation of endogenous proteins (e.g., identification of OGFOD1-catalysed hydroxylation of RPS23 P62) (Singleton *et al.*, 2014). Therefore, these methods, or similarly orthogonal proteomic assays, could lead to the discovery of novel prolyl hydroxylase substrates.

1.6 Aims and scope of this thesis

The aim of this body of work was to discover novel sites of prolyl hydroxylation that are catalysed by 2OGDs. Both the possibility that the known prolyl hydroxylases have as-yet undiscovered substrates and that additional 2OGD enzymes display prolyl hydroxylase activity were considered. LC-MS/MS-based proteomics approaches were employed to analyse prolyl hydroxylation throughout the proteome, and orthogonal approaches were performed in order to identify hydroxyproline sites using complementary assays. In particular, the sensitivity of hydroxylation sites to hypoxia and the pan-2OGD inhibitor, DMOG, was investigated. As described above, potential oncogenic pathways involving

prolyl hydroxylation might be suppressed by pVHL and escape this regulation in VHL disease, as suggested by type 2C mutations. Therefore, prolyl hydroxylation of pVHL-interacting proteins and peptides was investigated. Additionally, the potential for proteins to act as oxygen sensors through oxygen-dependent proteolytic regulation was studied. This body of work achieved the aims of this thesis and laid the foundation for further research into prolyl hydroxylation and oxygen-sensing mechanisms.

Chapter 2. Methods and materials

There are many bespoke LC-MS/MS experiments performed in this thesis. These are described in detail in the relevant results chapters. With this in mind, this chapter contains a description of a western blot protocol that has been used throughout this thesis and lists the reagents used in this body of work.

2.1 Methods

2.1.1 Western blot

For the western blot assays performed in this thesis, 20 µg total protein lysate and 5 µL PageRuler™ Plus Prestained protein ladder (10 to 250 kDa) was resolved by SDS-PAGE at 140 V for 45 minutes using a Tris-glycine-SDS buffer system. Immobilon®-P PVDF membrane (0.45 µm) was activated in methanol and proteins were transferred onto the membrane in a Tris-glycine buffer system at 100 V for 1 hour. Membranes were incubated in blocking buffer (PBS, 0.1 % (v/v) Tween®20, 1 % (w/v) BSA) for 1 hour at room temperature with gentle agitation (15 rpm). The blocking buffer was replenished, and primary antibody was incubated at 1:1,000 dilution for 18 hours at 4 °C with gentle agitation. The membrane was washed three times in wash buffer (PBS, 0.1 % (v/v) Tween®20) and incubated in blocking buffer containing secondary antibody at a dilution of 1:5,000 for 1 hour at room temperature with gentle agitation. The membrane was washed three times in wash buffer and incubated with the chemiluminescent substrate for 5 minutes. Western blots were visualised using the ChemiDoc™ MP Imaging System.

2.1.2 LC-MS/MS

The LC-MS/MS-based proteomics experiments were performed at the Francis Crick Institute in collaboration with the Proteomics science technology platform (STP).
Throughout this thesis, PEAKS® software is used to process mass spectrometry data. Some of the experiments performed in this body of work utilise precursor ion or reporter ion quantification. According to the PEAKS® software design, supporting peptides for each protein that are processed in the quantification (Q) search tool are ranked according to fold-change between conditions and the median value is used as the protein ratio.

Missing values are generated when a peptide is not detected in one condition. When a missing value is identified PEAKS® Q imputes the fold-change as 256 (i.e., $\log_2(256) = 8$). If the median is taken from two supporting peptides, one of which has a zero value, the fold-change can be calculated at approximately 128 (i.e., $\log_2(130) = 7$). Therefore, any fold-change value >6 indicates 'infinite increase', even though the biological regulation is likely to be many orders of magnitude lower. This explains the apparent cut-offs in the data. These 'infinite' values can also be generated when comparing the fold-change of peptide abundance.

2.2 Materials

2.2.1 Cell lines

Table 5 | Cell lines

Name	Tissue of origin
786-O	Renal cell carcinoma
A549	Lung adenocarcinoma
eHAP	Engineered from the near haploid KBM-7 cell line to be fully haploid
GAMG	Glioblastoma
HCT116	Colon carcinoma
HEK293T	Human embryonic kidney
HeLa	Human papillomavirus-related endocervical adenocarcinoma
Hep-G2	Hepatoblastoma
HeP3B	Paediatric hepatocellular carcinoma
Jurkat	Precursor T-cell acute lymphoblastic leukaemia
K562	Chronic myeloid leukaemia
KBM-7	Chronic myeloid leukaemia
LNCaP	Prostate carcinoma
MCF7	Invasive breast carcinoma
MEF	Mouse embryonic fibroblast (breed C57BL/6)
RCC4	Clear cell renal cell carcinoma
RKO	Colon carcinoma
SH-SY5Y	Neuroblastoma
U2OS	Osteosarcoma

U-87 MG	Glioblastoma

2.2.2 Antibodies

Table 6 | Primary antibodies

Target protein	Vendor	Catalogue number
CDKN1C	Cell signalling technology®	2557
HIF1α	BD Biosciences	610959
HIF1α HyP564	Cell signalling technology®	3434
KDM5B	Abcam	ab181089
HIF2α	Cell signalling technology®	7096
RGS4	Cell signalling technology®	15129
RGS5	Santa Cruz Biotechnology	sc-514184
TUBB	Proteintech®	10094-1-AP

Table 7 | Secondary antibodies

Target protein	Vendor	Catalogue number
Anti-mouse-HRP	Agilent Dako	P044701-2
Anti-rabbit-HRP	Agilent Dako	P039901-2

2.2.3 Peptides

The peptides used in this thesis were synthesised at the Francis Crick Institute by the Peptide chemistry STP.

Table 8 | Synthetic peptides

Name	Sequence
AKT1 HyP313	TFCGT(HyP)EYLAPEVLEDNDYGR
AKT1 HyP313 HyP318	TFCGT(HyP)EYLA(HyP)EVLEDNDYGR
Biotinylated HIF1α peptide	Biotin-eahx-NPFSTQDTDLDLEMLA(HyP)YIPMDDDFQLR
HIF1α MMAA HyP564	NPFSTQDTDLDLEALA(HyP)YIPADDDFQLR
HIF1α MMAA P564	NPFSTQDTDLDLEALAPYIPADDDFQLR
HIF1α HyP564	NPFSTQDTDLDLEMLA(HyP)YIPMDDDFQLR

Oligopeptides	An equimolar mixture of *trans*-4-hydroxy-L-proline, (HyP)Y, A(HyP)Y, and (HyP)YI
P4HB HyP395	NVFVEFYA(HyP)W(C')GH(C')K
P4HB P395	NVFVEFYAPW(C')GH(C')K
PDIA3 HyP404	DVLIEFYA(HyP)W(C')GH(C')K
PDIA3 P404	DVLIEFYAPW(C')GH(C')K
PDIA4 HyP89	DTVLLEFYA(HyP)W(C')GH('C)K
PDIA4 P89	DTVLLEFYAPW(C')GH(C')K
VEP	DEALA(HyP)YIPD

2.2.4 Plasmids

Table 9 | Plasmids

Plasmid	**Recombinant protein**
pGEX-4T-3 *VHL* & *Elongin B*	GST-VHL (aa54-213) & Elongin B (aa1-118)
pBB75 *Elongin C*	Elongin C (aa17-112)
pGEX-4T-3 *Twin-Strep-tag® VHL* & *Elongin B*	GST-TS-VHL (aa54-213) & Elongin B (aa1-118)

The plasmids used for recombinant expression of GST-VBC, pGEX-4T-3 *VHL* and *Elongin B* and pBB75 *Elongin C*, were gifts from Dr Ya-Min Tian. The construction of these plasmids has been previously reported (Stebbins et al., 1999).

The pGEX-4T-3 *VHL* and *Elongin B* plasmid was edited to express recombinant pVHL with a Twin-Strep-tag® (TS) inserted between the GST tag and pVHL (GST-TS-VHL). This was procured from Integrated DNA Technologies™ using their custom gene service. The following nucleotide sequence was inserted between the thrombin cleavage site and the VHL ORF (Figure 2):

5'

TGGAGTCATCCTCAATTCGAGAAAGGTGGAGGTTCTGGCGGTGGATCGGGA

GGTTCAGCGTGGAGCCACCCGCAGTTCGAAAAAACCGGTACCGCTAGC 3'

This corresponds to the Twin-Strep-tag® and a short linker sequence that contains 5 non-redundant restriction endonuclease cleavage sites.

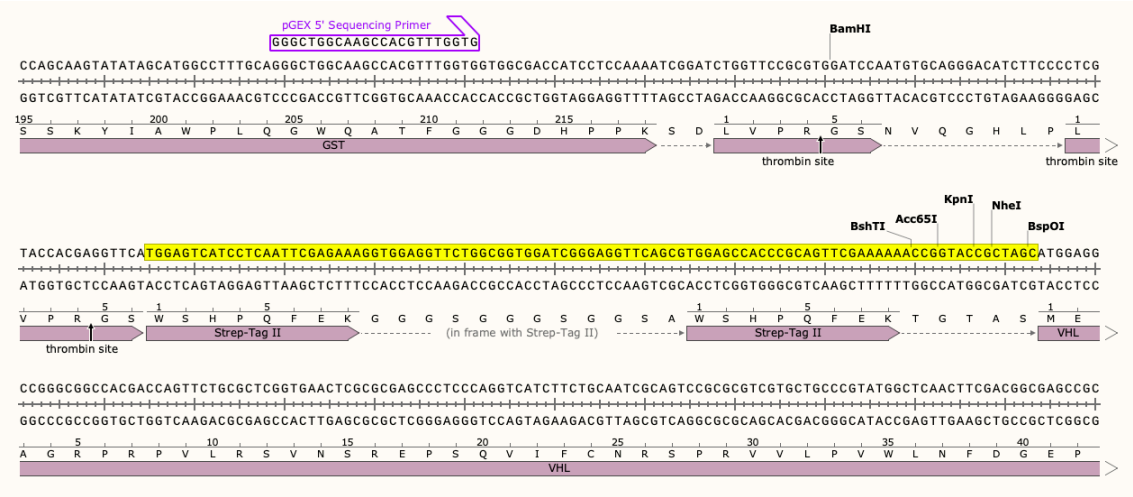


Figure 2 | DNA sequence and translated coding sequences of the GST-TS-VHL ORF.

When compared to the GST-VBC plasmid, the 'inserted sequence' is highlighted in the yellow box.

The expression and purification of recombinant GST-VBC and GST-TS-VBC complexes has been described in Chapter 5.

2.2.5 Primers

Sequencing primers were used to validate the plasmids used for recombinant GST-VBC and GST-TS-VBC expression.

Table 10 | Primers

Primer	Sequence
pGEX 5'	5' GGGCTGGCAAGCCACGTTTGGTG 3'
pGEX 3'	5' CCGGGAGCTGCATGTGTCAGAGG 3'
M13 forward	5' CCCAGTCACGACGTTGTAAAACG 3'
M13 reverse	5' CAGGAAACAGCTATGAC 3'

2.2.6 Reagents

Table 11 | Reagents

Reagent	Vendor	Catalogue number
4-20% Mini-PROTEAN® TGX™ Precast Protein Gels	Bio-Rad™	4561096

9-13 µm glass spheres	Merck™	440345
Acclaim™ Pepmap™ 100 C18 HPLC column (3 µm x 0.075 mm x 500 mm)	Thermo Scientific™	164570
Acclaim™ Pepmap™ 100 C18 HPLC column (5 µm x 0.1 mm x 20 mm)	Thermo Scientific™	164199
Amine coupling kit	Cytiva™	BR100050
Ampicillin	Merck™	171254
Benzonase®	Merck™	E1014
Biacore™ CM5 sensor chip	Cytiva™	BR100399
BSA	Merck™	A2153
CHAPS Detergent	Thermo Scientific™	28300
DMOG	Biotechne® TOCRIS	4408
DTT	Merck™	111474
EASY-Spray™ HPLC Column (2 µm x 75 µm x 500 mm)	Thermo Scientific™	ES903
E. coli BL21 (DE3) GOLD	Agilent	230132
Evotip C18	Evosep™	EV2001
Gibco™ DMEM	Thermo Scientific™	11965092
Gibco™ DMEM for SILAC	Thermo Scientific™	A33822
Gibco™ DMEM:F12	Thermo Scientific™	11320033
Gibco™ FBS	Thermo Scientific™	26140079
Gibco™ Foetal Bovine Serum, dialyzed	Thermo Scientific™	26400044
Gibco™ IMDM	Thermo Scientific™	12440053
Gibco™ MEM	Thermo Scientific™	11095080
Glutathione sepharose 4B beads	GE healthcare™	17-0756-01
HALT™ protease inhibitor cocktail	Thermo Scientific™	78429
IAA	Merck™	804744
Immobilon®-P PVDF membrane	Merck™	IPVH00010
IPTG	Merck™	16758
iRT peptide solution	Biognosys™	Ki-3002-1
Kanamycin	Merck™	K1876
L-Arginine-HCl, 13C6 for SILAC	Thermo Scientific™	88210
L-Arginine-HCl, 13C6, 15N4 for SILAC	Thermo Scientific™	89990
L-Arginine-HCl, for SILAC	Thermo Scientific™	89989
L-Lysine-2HCl, 13C6, 15N2 for SILAC	Thermo Scientific™	88209
L-Lysine-2HCl, 4,4,5,5-D4 for SILAC	Thermo Scientific™	88437

L-Lysine-2HCl, for SILAC	Thermo Scientific™	89987
L-proline for SILAC	Thermo Scientific™	88430
Laemmli buffer	Bio-Rad™	1610747
Lysyl Endopeptidase®, Mass spectrometry grade (LysC)	FUJIFILM Wako Pure Chemical Corporation	4548995075888
MG132	Merck™	474790
NP-40 Surfact-Amps™ Detergent Solution	Thermo Scientific™	85124
Oasis HLB 96-well plate, 30 mg sorbent per well	Waters™	WAT058951
Oasis HLB µelution plate, 2 mg sorbent per well	Waters™	186001828BA
PageRuler™ Plus Prestained protein ladder, 10 to 250 kDa	Thermo Scientific™	26619
Pierce™ 660 nm protein assay kit	Thermo Scientific™	22662
Pierce™ glutathione DynaBeads	Thermo Scientific™	88821
Pierce™ HeLa Protein Digest Standard	Thermo Scientific™	88328
Pierce™ high pH reversed-phase peptide fractionation kit	Thermo Scientific™	84868
Pierce™ Trypsin Protease, MS grade	Thermo Scientific™	90058
Quick Coomassie stain	Generon™	NB-45-00078-1L
*Rapi*Gest™ SF Surfactant	Waters™	186001861
Reduced glutathione	Merck™	G4251
Sep-Pak C18 1 cc Vac Cartridge, 50 mg Sorbent per Cartridge, 55-105 µm	Waters™	WAT054955
SuperSignal™ West Dura Extended Duration Substrate	Thermo Scientific™	34075
Tetracycline-free FBS	PAN™-Biotech	P30-3602
TMT10plex™ isobaric label reagent set	Thermo Scientific™	90110
Trans-4-hydroxy-L-proline	Merck™	H54409
Tween®20	Merck™	P1379
Twin-Strep-tag® Capture Kit	IBA Lifesciences GmbH™	2-4370-000
VH298	Merck™	SML1896
XBridge BEH C18 Column, 130Å, 2.5 µm, 3 mm X 150 mm	Waters™	186006710
XBridge BEH C18 Column, 130Å, 3.5 µm, 1 mm X 150 mm	Waters™	186003128

2.2.7 Equipment

Table 12 | Equipment

Instrument	Vendor
Biacore™ S200	Cytiva™
ChemiDoc™ MP Imaging System	Bio-Rad™
Evosep™ One	Evosep™
InvivO$_2$	Baker Ruskin™
nanoElute®	Bruker™
Orbitrap Fusion™ Lumos™ Tribrid™	Thermo Scientific™
Q Exactive™ Hybrid Quadrupole-Orbitrap™	Thermo Scientific™
timsTOF	Bruker™
UltiMate™ 3000	Thermo Scientific™

Chapter 3. *In silico* analysis of hydroxyproline in publicly deposited deep proteome datasets

3.1 Introduction

The overall aim of the work described in this thesis was to determine the existence or otherwise of sites of 2OGD-catalysed hydroxyproline in the proteome, particularly those that might show functional similarities to hydroxyproline sites defined in the transcription factor HIFα. At the time of commencing the work a large number of such sites had been reported in the literature, but not corroborated in the laboratory by reacting recombinant HIFα prolyl hydroxylases (PHD 1, 2 and 3) with peptides or full-length polypeptides representing the reported prolyl hydroxylation substrates.

To investigate this unexpected result, I first sought to interrogate large proteomic datasets from publicly available data repositories for the existence of prolyl hydroxylation at the reported sites in endogenous proteins in cells where conditions for prolyl hydroxylation might have been present, but not met in the *in vitro* assays that had been performed using recombinant proteins.

In this chapter, I will first describe these analyses together with surveys of the proteome for oxidations on different residues that might be confused with enzyme-catalysed prolyl hydroxylation. I then proceed to survey the proteome in general for sites of prolyl hydroxylation that can be defined at high confidence levels and outline some of the difficulties in doing this.

3.1.1 Human proteome datasets retrieved from the PRIDE database

Public proteomic data repositories have been developed to facilitate collaboration and communication of proteomics research through assessment and re-use of published LC-MS/MS data. The ProteomeXchange (PX) consortium aims to provide a common framework and infrastructure for dissemination of all MS-based proteomics data (Vizcaíno et al., 2014). The proteomics identification (PRIDE) database is one of the repositories employed by PX that enables storage and exchange of raw LC-MS/MS data files (Hermjakob and Apweiler, 2006). Peer review journals increasingly require submission of proteomics data to public repositories and the number of datasets submitted

to PX has shown sustained growth (Deutsch et al., 2019; Deutsch et al., 2016). The availability and variety of data on PX provides a valuable resource for discovery of novel PTM sites that were not considered in the original investigation.

Benchmark proteomics datasets were retrieved from PRIDE (Bekker-Jensen *et al.*, 2017; Davis et al., 2017; Geiger et al., 2012) and processed in PEAKS® software to study hydroxyproline in the steady state proteome. The datasets were selected for their extensive coverage of the "deep proteome" across 13 human cell lines: A549, GAMG1, HCT116, HEK293, HeLa, HepG2, K562, Jurkat, LNCaP, MCF7, RKO, SH-SY5Y, and U2OS.

3.1.2 Factors influencing hydroxyproline assignments in LC-MS/MS data

PTM assignment by LC-MS/MS is direct and semi-quantitative. MS1 data, which includes *m/z*, retention time (RT) and peak intensities, provides important data pertaining to the parent ion acquisition but does not, in and of itself, enable PTM assignment. MS2 data provides the intensity and *m/z* of fragment ions derived from the parent ion, which facilitates peptide sequencing and localisation of PTMs.

Throughout this thesis I have used PEAKS® software to process raw LC-MS/MS data files. In PEAKS® software, *de novo* sequencing is performed on the raw MS2 spectrum with the application of "local confidence score" (LCS), which indicates the likelihood of a residue being correctly assigned in the peptide sequence. The fragment ions of the MS2 spectrum are also aligned against a reference proteome to generate a peptide spectrum match (PSM), taking into account user-defined modifications (Tran et al., 2016). A "linear discriminative function" (LDF) score is applied to measure the quality of the PSM. Many factors are considered in this calculation, such as the matching of fragment ions and peaks in the spectrum, and the similarity of the *de novo* peptide and the database sequence. The LDF score is converted to a p-value and estimates the probability that the PSM is assigned by chance. Finally, the p-value is converted to a -10lgP score ($-10*\log_{10}(\text{p-value})$) to make it more "human-friendly". In this format, -10lgP = 20 indicates a 1 % probability that the PSM has been assigned by chance.

The presence of diagnostic ions in the MS2 spectrum enables the exact site of modification to be determined. PEAKS® software calculates ion intensity of the so-called "position-determining ions" (i.e., diagnostic ions) relative to the most intense fragment

ion in the MS2 spectrum (Beausoleil *et al.*, 2006). PEAKS® also generates a -10lgP score for PTM localisation, which is termed the ambiguity score (AScore). For the AScore calculation, the p-value indicates the probability that the PTM localisation has been assigned by chance (Beausoleil *et al.*, 2006). Both AScore and ion intensity can be filtered to identify PTM sites that meet user-defined confidence thresholds. Whilst these filters increase the likelihood of a correct PTM assignment, manual inspection of the MS2 spectrum is necessary to provide a final verdict on the assignment.

3.1.2.1 Factors impacting peptide fragmentation during MS/MS

Several factors influence peptide fragmentation (Section 1.5.1.7), which impacts the security of PSMs and PTM localisation scores, along with corroboration of PTM assignments by manual inspection of the MS2 spectra. As described by Simpson and colleagues, the presence of basic amino acids in the parent ion can impact peptide sequencing and PTM localisation (Kapp *et al.*, 2003). Generally, y-ions are more abundant than b-ions in tryptic fragments, and the ratio of b- to y-ions is likely to increase if there is a basic residue at the N-terminus or internally (i.e., within the peptide), such as might occur after a trypsin missed cleavage event (Elias et al., 2004; Khatun et al., 2007; Tabb et al., 2003). The changes to relative intensity of the fragment ions are predicted by the relative proton mobility model: the intensity of b-ions increases if the peptide has a non-mobile or partially mobile proton, and decreases if the peptide has a mobile proton (Barton *et al.*, 2007). The impact of increased b-ion intensity on PTM localisation is context dependent; on the one hand it may generate diagnostic ions that are otherwise unavailable through the y-ion series but, on the other hand, could also suppress fragmentation efficiency of the diagnostic fragment ions.

The proline effect (PE) is an established phenomenon in which selective cleavage at the N-terminus of proline residues generates a highly abundant y-ion (Tabb *et al.*, 2003). The PE is particularly strong in the MS2 spectrum of peptides with mobile or partially mobile protons as opposed to peptides with non-mobile protons (Huang et al., 2005). The PE is reported to be predictable; ‘Asp’, ‘His’, ‘Ile’, ‘Leu’, or ‘Val’ occupancy of the ‘Xaa’ position of ‘Xaa-Pro’ amino acid pairs favours N-terminal cleavage at ‘Pro’, whereas ‘Gly’ or ‘Pro’ occupancy of the ‘Xaa’ position suppresses fragmentation (Breci et al., 2003). The PE also includes a negative impact on the relative intensity of fragment ions

containing non-prolyl residues, which is exacerbated for poly-prolyl peptides. The prolyl residue nearest the N-terminus exhibits the highest probability of selective cleavage in poly-proline peptides. If prolyl residues are dispersed along the peptide sequence (e.g., the repeating 'Xaa-Pro-Gly' triplet of collagen), then the PE applies to each proline. For peptides containing consecutive prolyl residues, fragmentation efficiency is reduced throughout this sequence, and the proline nearest the N-terminus is most likely to fragment with the highest efficiency (Dong et al., 2011). The PE may negatively impact the identification of the adjacent N-terminal residue, which could be detrimental for probability-based PTM localisation statistics. Whilst the PE can be detrimental to peptide sequencing and statistics-based scoring of PTM localisation, the relatively high intensity fragment ion aids PTM localisation of prolyl hydroxylation by manual inspection so long as the adjacent N-terminal fragment ion can be detected.

3.1.2.2 Prolyl hydroxylation is isobaric to oxidation at non-prolyl residues

Hydroxylation of proline residues (+15.9949 Da) is isobaric to both enzymatic hydroxylation and artefactual oxidation of other residues. Protein oxidation can occur through a variety of enzymatic and non-enzymatic mechanisms (Markolovic *et al.*, 2015; Silva et al., 2013; Stadtman and Levine, 2003). For the purposes of this thesis, "artefactual" oxidation refers to the oxidation of amino acids independently of the 2OGDs. Unimod (https://www.unimod.org/; last accessed 01/04/22) has defined 11 residues (CDFHKMNPRWY) that are susceptible to oxidation, which could generate incorrect hydroxyproline assignments through misassigned oxidation of these residues. It is therefore important to consider oxidation of all 11 residues indicated by Unimod in order to limit the generation of false positive hydroxylation assignments. Some proteins may be susceptible to oxidation on multiple residues in the same tryptic fragment and this could be confused with dioxidation, which is an artefactual PTM that can occur on F, M, W, and Y. Dioxidation of these residues should also be considered during data processing in proteomic software to limit false assignments of hydroxylation on another residue such as proline. This has been exemplified previously, whereby methionine dioxidation was misassigned by the proteomics software as two separate hydroxylation events on the same peptide; hydroxyproline and methionine oxidation (Verrastro et al., 2015). Oxidation of certain residues is associated with neutral loss ions, but these are not assigned by PEAKS®

software (Table 13). Knowledge of these ions may aid identification of the precise site of oxidation during manual MS2 inspection.

Table 13 | Monoisotopic mass of neutral loss ions in the MS2 spectra of oxidised residues. C(cam), carbamidomethylated cysteine.

Oxidised residue	Neutral loss ion (Da)	Reference
C(cam)	-107.0283	(Na et al., 2012; Steen and Mann, 2001)
M	-63.9983	(Kotiaho *et al.*, 2000)
W	-18.0106	(Lioe et al., 2004)

3.1.3 Reported PHD-catalysed prolyl hydroxylation of 45 non-HIFα protein substrates

The three alpha isoforms of hypoxia inducible factor (HIF1α-3α) are substrates of the prolyl hydroxylase domain containing proteins (PHD1-3) (Epstein *et al.*, 2001). Prolyl hydroxylation of HIFα facilitates von Hippel Lindau protein (pVHL) binding and proteasomal degradation of HIFα (Jaakkola *et al.*, 2001). Together, these interactions perform a signalling function that upregulates the expression of HIF target genes in response to hypoxia (Metzen and Ratcliffe, 2004). Significant research efforts have been directed towards identifying HIF-independent oxygen-sensing mechanisms that involve the PHDs.

To date, 43 non-HIFα PHD substrates have been reported, while EEF2K and ZHX2 have been implicated as PHD substrates by intervention studies but the PHD isoenzyme that catalyses the putative hydroxylation was not reported (Table 14). The site of prolyl hydroxylation was not reported for CERKL, EEF2, PAX2, and MAPK7. Whilst these assignments are mainly published by peer reviewed journals, reports of PHD-catalysed prolyl hydroxylation of ACTB P70 (Zi et al., 2021) and CENPN P311 (Moser et al., 2015) are available in the pre-publication repository, bioRxiv.

Different methods have been used to identify the 45 reported substrates and typically use either immunoblotting with pan-hydroxyproline antibody or LC-MS/MS, as evidence for hydroxyproline. Some uncertainties exist with respect to hydroxyproline detection. For instance, the specificity of pan-hydroxyproline antibodies has not been studied in depth and may generate false positive results by reacting with epitopes that do not contain

hydroxyproline. LC-MS/MS can generate false positives by incorrectly assigning a peptide sequence to an MS2 spectrum, or by misassigning an isobaric mutation (e.g., A > S, F > Y, P > I/L, V > D), or misassigning oxidation of a nearby residue as hydroxyproline, as outlined above. The latter is expected to be particularly prone to generating incorrect hydroxyproline assignments.

The methods indirectly used to support hydroxyproline detection are typically cellular interaction studies with PHD or pVHL, which utilise ectopic expression of epitope-tagged constructs (Di Conza et al., 2017) or immunoprecipitation of endogenous proteins (Kuznetsova et al., 2003). Overexpression of proteins increases the likelihood of detecting non-specific interactions in such studies, which may, for example, be artefacts of non-physiological concentrations of the protein or a consequence of mis-folding at higher expression levels. Transient overexpression studies exacerbate these potential inaccuracies because of greater variability between samples when compared to experiments targeting endogenous proteins or utilising cell lines with stable ectopic expression of the protein.

Novel methods have been developed and refined to study interactions between substrates of PHD and/or pVHL in order to reduce non-specific interactions. Factor inhibiting hypoxia inducible factor (FIH) is a 2OGD that catalyses asparaginyl hydroxylation of HIFα and other ankyrin-repeat domain proteins (Cockman *et al.*, 2006; Lando *et al.*, 2002). "Substrate trapping" – the application of dimethyloxalylglycine (DMOG) to stabilise interactions between a 2OGD and its substrates – was first established for FIH (Cockman *et al.*, 2009). This has subsequently become an increasingly popular method to study alternative substrates of the PHDs (Rodriguez et al., 2018; Rodriguez et al., 2016; Zurlo et al., 2019). The application of DMOG stabilises HIFα, which manifests as HIF transactivation. Therefore, spurious results could be generated by non-specific interactions between the bait enzyme and proteins that are upregulated after DMOG treatment.

pVHL capture assays were established to infer the presence of enzymatic prolyl hydroxylation in RCC4 and reticulocyte lysates (Jaakkola *et al.*, 2001). Recently a novel genome-wide *in vitro* expression system using reticulocyte lysates, which has been used to screen cDNA libraries of the human genome, was coupled to a pVHL capture assay with competitive HIFα peptide elution and has generated novel putative PHD substrate

assignments (Liu et al., 2020a; Zhang et al., 2018). These assays do not provide evidence of direct enrichment of prolyl hydroxylated proteins and proteins may interact with pVHL independently of hydroxyproline (Minervini et al., 2015).

Other approaches include yeast two-hybrid (Y2H) assays (Köditz et al., 2007). It has been reported that the yeast strains used in traditional Y2H systems lack expression of PHD orthologues (Alcaide-German et al., 2008; Bex et al., 2007). Therefore, ectopic expression of these isoenzymes could lead to an interaction between PHDs and target proteins that is not physiologically relevant, which includes the putative prolyl hydroxylation reaction.

Candidate PHD substrates have also been identified by mining publicly deposited proteomic data (Scholz et al., 2013) and screening for the 'LXXLAP' motif, which is conserved in all human HIFα proteins (Cummins et al., 2006), but this motif may not represent the recognition site of all PHD substrates (Loenarz *et al.*, 2011).

Prolyl hydroxylation has sometimes been inferred by 2-oxoglutarate (2OG) turnover assays, in which [^{14}C]-CO_2 is measured as a proxy for PHD catalytic activity (German et al., 2016; Zheng et al., 2014). Since 2OG conversion to succinate and CO_2 precedes substrate hydroxylation, uncoupled decarboxylation can occur and therefore does not necessarily represent substrate hydroxylation.

Another method of assay for hydroxyproline measures the conversion of [^{3}H]-P to [^{3}H]-HyP in *in vitro* assays using recombinant hydroxylases (Ullah et al., 2017). Briefly, this radio-assay involves *in vitro* expression of the target protein and hydrolysis of the hydroxylation assay product. Proline and hydroxyproline are extracted and resolved by liquid chromatography coupled to scintillation counting. Whilst this method directly measures hydroxyproline accrual, the sensitivity is limited by high background [^{3}H]-P signal and low [^{3}H]-HyP signal intensity in proteins for which i) the hydroxylation reaction is inefficient and sub-stoichiometric, and ii) hydroxylation occurs on a low proportion of proline residues in the protein sequence. Additionally, this method lacks the accuracy of mass spectrometry that would enable site-specific prolyl hydroxylation to be identified.

Given the potential inaccuracies in the methods used to detect and correctly assign PHD-catalysed prolyl hydroxylation, further research is required to validate these reports. Prior to the experiments described in this thesis, experiments investigating PHD-catalysed

hydroxylation of 23 reported non-HIFα PHD substrates had been performed on peptides and full-length protein substrates using several different *in vitro* hydroxylation assays, whereby LC-MS/MS and [^{3}H]-P to [^{3}H]-HyP conversion assays were used to detect hydroxyproline (Cockman et al., 2019). The LC-MS/MS experiments demonstrated that under conditions where PHD-catalysed prolyl hydroxylation of HIFα is driven to extremely high stoichiometry (i.e., > 90 %), PHD-dependent prolyl hydroxylation is not detected on any non-HIFα substrate. An interesting observation was that artefactual oxidation was common in these assays, which suggests that some of the initial assignments by LC-MS/MS might have misassigned these isobaric modifications (Cockman *et al.*, 2019). Similarly, hydroxyproline was readily detected in the [^{3}H]-P to [^{3}H]-HyP conversion assay with HIFα as the substrate but not for any of the non-HIFα candidates. These results suggested the initial reports may have misassigned these isobaric PTMs as hydroxyproline or that additional factors are required for hydroxylation that are present in cells.

Table 14 | Reported non-HIFα substrates of PHD1-3.

45 non-HIFα PHD1-3 substrates (81 target sites) have been reported and are detailed here. The protein name used by the authors was searched on Uniprot to retrieve the Gene ID and canonical Uniprot accession of the human isoform. The target sites and PHD isoenzyme specified by the authors are listed. In two instances, the PHD isoenzyme is not defined (ND) but regulation by PHD1-3 is inferred by intervention studies on these enzymes.

Protein name	Gene ID	Uniprot Accession	Target site(s)	PHD Isoenzyme	Reference
ACC2	*ACACB*	O00763-1	P343*; P450*; P2131*	PHD3	German *et al.* (2016)
			P450*	PHD3	Yoon et al. (2020)
β-actin	*ACTB*	P60709-1	P70	PHD2, 3	(Zi *et al.*, 2021)
			P307*; P322*	PHD3	Luo et al. (2014)
β(2)AR	*ADRB2*	P07550-1	P382*; P395*	PHD3	Xie et al. (2009)
ADSL	*ADSL*	P30566-1	P24	PHD1	Zurlo *et al.* (2019)
Akt	*AKT1*	P31749-1	P125*; P313*; P318*; P423*	PHD2	Guo et al. (2016)
β-arrestin 2	*ARRB2*	P32121-1	P176; P179; P181	PHD2	Yan et al. (2011a)
ATF-4	*ATF4*	P18848-1	P156*; P162*; P164*; P167*; P174*	PHD3	Köditz *et al.* (2007)

BIM-EL	*BCL2L11*	O43521-1	P67; P70	PHD3	(Li et al., 2019)
BRD4	*BRD4*	O60885-1	P536	PHD2	Erber et al. (2019)
CENP-N	*CENPN*	Q96H22-1	P311*	PHD2	Moser *et al.* (2015)
Cep192	*CEP192*	Q8TEP8-3	P2313*	PHD1	Moser et al. (2013)
CERKL	*CERKL*	Q49MI3-1	ND	PHD1, 3	Chen et al. (2015)
CPT1B	*CPT1B*	Q92523-1	P295	PHD2, 3	Angelini et al. (2021)
DGKι	*DGKI*	O75912-1	P903	PHD3	(Bex *et al.*, 2007)
DYRK1	*DYRK1A*	Q13627-1	ND P380	PHD1	Lee et al. (2016) Lee et al. (2020)
	DYRK1B	Q9Y463-1	ND P332	PHD1	Lee *et al.* (2016) Lee *et al.* (2020)
eEF2	*EEF2*	P13639-1	ND	PHD2	Romero-Ruiz et al. (2012)
eEF2K	*EEF2K*	O00418-1	P98*	ND	Moore et al. (2015)
G9a	*EHMT2*	Q96KQ7	P676; P1207	PHD1	Casciello et al. (2017)
EPOR	*EPOR*	P19235-1	P443*; P450*	PHD3	Heir et al. (2016)
FLNA	*FLNA*	P21333-1	P2317*; P2324*	PHD2	Segura et al. (2016)
FOXO3a	*FOXO3*	O43524-1	P426*; P437*	PHD1	Zheng *et al.* (2014)
IKKβ	*IKBKB*	O14920-1	P191*	PHD1	Cummins *et al.* (2006)
MAPK6	*MAPK6*	Q16659-1	P25*	PHD3	Rodriguez *et al.* (2016)
ERK5	*MAPK7*	Q13164-1	ND	PHD3	Arias-González et al. (2013)
ChREBP	*MLXIPL*	Q9NP71-1	P15; P141; P526**	PHD3	Heidenreich et al. (2020)
NDRG3	*NDRG3*	Q9UGV2-1	P294*	PHD2	Lee et al. (2015)
OTUB1	*OTUB1*	Q96FW1-1	P210; P263	PHD1	Scholz *et al.* (2013)
Pax2	*PAX2*	Q02962-1	ND	PHD3	Yan et al. (2011b)
PDE4D	*PDE4D*	Q08499-1	P29*; P382*; P419*	PHD2	Huo et al. (2012)
PKM2	*PKM*	P14618-1	P403*; P408*	PHD3	Luo et al. (2011)
Rpb1	*POLR2A*	P24928-1	P1465*	PHD1	Mikhaylova et al. (2008)
B55α	*PPP2R2A*	P63151-1	P319*	PHD2	Di Conza *et al.* (2017)
SFMBT1	*SFMBT1*	Q9UHJ3-1	P106; P651	PHD2	Liu *et al.* (2020a)
Spry2	*SPRY2*	O43597-1	P18*; P144*; P160*	PHD1, 2, 3	Anderson et al. (2011)

TBK1	*TBK1*	Q9UHD2-1	P48	PHD2	Hu et al. (2020)
HCLK2	*TELO2*	Q9Y4R8-1	P374*; P419*; P422*	PHD3	Xie et al. (2012)
TET2	*TET2*	Q6ND21-1	P1335**; P1342**	PHD2, 3	Fan et al. (2020)
TET3	*TET3*	O43151-1	P1030**; P1037**	PHD2, 3	Fan *et al.* (2020)
TR-α	*THRA*	P10827-1	P160*; P162*	PHD2, 3	Xie et al. (2015)
p53	*TP53*	P04637-1	P142*	PHD1	Ullah *et al.* (2017)
			P359*	PHD3	Rodriguez *et al.* (2018)
TRPA1	*TRPA1*	O75762-1	P394*	PHD2	Takahashi et al. (2011)
UBC13	*UBE2N*	P61088-1	P19; P21; P59	PHD1	Scholz *et al.* (2013)
UEV1A	*UBE2V1*	Q13404-4	P79***; P80***	PHD1	Scholz *et al.* (2013)
ZHX2	*ZHX2*	Q9Y6X8-1	P427; P440; P464	ND	Zhang *et al.* (2018)

*These target sites were tested in assays of PHD-catalysed hydroxylation by Cockman *et al.* (2019). **The target sites listed here differ from those in the original publication because mouse and/or zebrafish analogues were tested by the authors. ***The Target sites listed here differ from the original publication because the authors reported prolyl hydroxylation of the non-canonical Uniprot accession, Q13404-1.

3.1.4 *In silico* analysis of hydroxyproline in publicly deposited deep proteome data

With the above in mind, an *in silico* analysis of hydroxyproline was performed on publicly available benchmark deep proteome datasets (Bekker-Jensen *et al.*, 2017; Davis *et al.*, 2017; Geiger *et al.*, 2012). The prevalence and security of hydroxyproline assignments was first assessed in the context of oxidation of all residues that are susceptible to this modification, either through enzymatic or non-enzymatic mechanisms. A targeted survey of oxidation on peptides containing reported sites of PHD-catalysed prolyl hydroxylation was performed to evaluate the likelihood of these proteins being genuine substrates of PHD1-3. Finally, an unbiased discovery pipeline was performed to identify novel hydroxyproline sites in the proteome for which the hydroxyproline sites are highly secure and are consistent with enzymatic hydroxylation. These approaches were subsequently evaluated to consider the difficulties concerning hydroxyproline assignments in LC-MS/MS data.

3.2 Methods

3.2.1 LC-MS/MS data retrieval and processing in PEAKS® X

Benchmark 'deep proteome' datasets (PXD002395, PXD003977, PXD004452) were downloaded from the PRIDE repository and processed in PEAKS®. Where possible, raw files were grouped according to publication, cell type, fragmentation mode and replicate status. Table 15 outlines the parameters used when processing the raw data.

Table 15 | Processing parameters for publicly deposited deep proteome datasets.

Parameter	Setting
PEAKS® version	X
PEAKS® search tool	DB
Reference proteome	Human, canonical (Uniprot id: UP000005640)
Precursor ion error tolerance	±10 ppm
Fragment ion error tolerance	±0.5 Da (Ion trap) ±0.02 Da (Orbitrap)
Enzyme	Trypsin
Enzyme specificity	Specific
Maximum number of miscleavages per peptide	2
Fixed modification	Carbamidomethylation (+57.0214 Da; C)
Variable modifications	Oxidation (+15.9949 Da; CDFHKMNPRWY) Dioxidation (+31.9899 Da; FMWY) Deamidation (+0.9840 Da; NQ) Acetylation (+42.0367 Da; protein N-terminus)
Maximum number of PTMs per peptide	3

3.2.2 Security of PSM and PTM assignments

A filter was applied to select for confidently assigned peptide spectrum matches (PSMs). This filter, -10lgP ≥ 20, indicates that there is a ≤ 1 % probability the PSM was assigned by chance.

Because of the possibility of misassignment amongst different residues with the potential for oxidation, different filters were applied to censor assignments on the basis of computationally determined confidence scores. Ion intensity ≥ 5 % requires the

diagnostic fragment ions to be detected at 5 % relative intensity or higher in the MS2 spectrum. AScore is a -10lgP value which indicates that the probability the PTM assignment is generated by chance. An AScore value of 20 is equivalent to a p-value of 0.01. PTM localisation filters (ion intensity ≥ 5 % and AScore ≥ 20) were applied to increase the likelihood of correct PTM assignments.

3.2.3 Assessing the abundance and security of global protein oxidation assignments

To provide a framework for considering enzymatic and non-enzymatic oxidation assignments, analyses were conducted to define the prevalence of oxidations across the proteome. For these analyses the abundance of nominally assigned oxidised peptides was deduced by spectral counting (i.e., counting the number of oxidised Peptide Features for each oxidised residue). "Peptide Features" is a vendor-specific term used by PEAKS® that describes the profile of any given precursor ion across a specific retention time range. As the precursor ion elutes from the liquid chromatography column, the peak intensities rise and fall to give a bell curve for each isotope of the charged peptide (i.e., the isotope cluster). A Peptide Feature therefore encompasses a series of *m/z* values within the isotope cluster, a retention time range for each isotope of the parent ion, and the peak intensities of each isotope. Different charge states, elution in multiple fractions, and distinct retention times in the same fraction can lead to multiple Peptide Features of a single PSM. The abundance of an individual PSM was inferred by summing the number of Peptide Features across all tryptic datasets.

To investigate the security of global protein oxidation, the PSM security (-10lgP ≥ 20), ion intensity (≥ 5 %) and AScore (≥ 20) filters were applied in combination and the Peptide Feature counts were repeated.

3.2.4 Security of oxidation assignments on reported non-HIFα PHD substrates

For proteins in which the target site of PHD catalysed prolyl hydroxylation was defined in the original publication, Peptide Features were grouped according to protein, target site, and the assigned oxidation site. The total number of Peptide Features was counted over a range of stringency filters. These filters were (1) -10lgP ≥ 20, (2a) -10lgP ≥ 20 & AScore ≥ 20, (2b) -10lgP ≥ 20 & ion intensity ≥ 5 %, and (3) -10lgP ≥ 20 & AScore ≥ 20

& ion intensity ≥ 5 %. All instances of target site oxidation are reported. All instances of oxidation occurring at a non-target site within the parent ion (i.e., "other" oxidation site) are grouped together and reported. The "Other" oxidation sites are specified when the modification is assigned at the highest stringency level (i.e., stringency filter 3).

3.2.5 Prospective survey for hydroxyproline in the tryptic proteome

To define sites of high-confidence prolyl hydroxylation across the proteome, peptides containing a hydroxyproline assignment were subject to a high stringency workflow to confirm the hydroxyproline assignment (Figure 3), including by manual inspection of the MS2 spectra.

Identical tryptic peptides assigned to more than one protein (i.e., non-unique) were filtered so only the first instance was retained. -10lgP ≥ 20, AScore ≥ 20, and ion intensity ≥ 5 % were employed to select for peptides with confident assignments from the software, as described in Section 3.2.3.

Prolyl hydroxylation is understood to be irreversible. So long as the rate of protein synthesis and hydroxylation exceeds protein turnover, a relatively high stoichiometry would be expected. The stoichiometry filter (≥ 5 %) selects for proteins with such characteristics. The gene expression profiles of the known prolyl hydroxylases exhibit low tissue specificity so hydroxylation would be expected in multiple datasets. The frequency filter (> 10 %) selects for hydroxyproline sites that occur in at least 4 of the 39 datasets.

Prolyl hydroxylation has been associated with increased polarity of the peptide (Cockman *et al.*, 2019; Kato et al., 1988; Zolg et al., 2018). A filter was therefore applied to identify peptides with increased polarity compared to the non-oxidised counterpart (i.e., -ΔRT). Peptides lacking a non-oxidised counterpart were accepted with this filter.

Prolyl hydroxylation should be an independent PTM. Therefore, peptides in which hydroxyproline colocalised with other PTMs (e.g., protein N-terminal acetylation, deamidation of NQ, and dioxidation of MWY) were not accepted unless those modifications were also observed on the non-hydroxyprolyl counterpart. Missed cleavages were similarly rejected unless the non-hydroxyprolyl counterpart exhibited the same peptide sequence. Peptides lacking a non-oxidised counterpart were accepted with these filters.

Manual inspection of the MS2 spectra was performed and compared with that of the non-oxidised peptide, if available. The MS2 spectrum was assessed for obvious misassignment of the peptide sequence (e.g., high background noise or highly dissimilar fragment ion assignments between hydroxyproline and non-oxidised counterparts). The diagnostic ions were investigated to confirm that their signal intensity was above the noise in the MS2 spectrum and the +15.99 Da mass shift was appropriately assigned to the indicated proline. Neutral loss ions corresponding to oxidation of carbamidomethylated cysteine, methionine, or tryptophan were considered to determine if the peptide contained a misassigned artefactual oxidation of an adjacent or proximal residue. Hydroxyproline assignments lacking a non-oxidised counterpart were required to exhibit exemplar MS2 spectra quality (i.e., near-perfect matching of a complete fragment ion series with few unmatched fragment ions in the spectrum) in order to be accepted because the security of assignment relies entirely on a single MS2 spectrum.

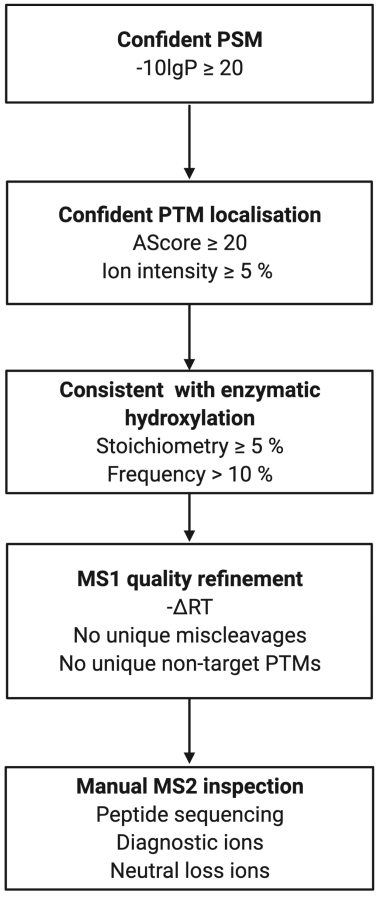


Figure 3 | High stringency workflow to identify hydroxyproline assignments that are consistent with enzymatic hydroxylation.

The first level of filtering identifies confidently assigned peptide spectrum matches (PSMs). The second level considers the confidence of correct PTM localisation. The third filtering level identifies hydroxyproline sites that occur at sufficiently high stoichiometry and occur in a sufficient number of datasets to be considered enzyme-catalysed. The fourth level considers properties of the precursor ion that must be consistent with prolyl hydroxylation. The fifth level of filtering identifies hydroxyproline-containing peptides for which the MS2 spectra give strong evidence in favour of the correct assignment.

3.2.6 Motif analysis

It is possible manual inspection of MS2 spectra, which is ultimately a subjective process, overlooks genuine hydroxyproline sites despite following objective guidelines. I therefore sought to identify possible hydroxylation consensus sequences to increase the confidence of hydroxyproline assignments that could not be corroborated by MS2 inspection. Since prolyl hydroxylation should occur *in situ*, flanking sequences surrounding the target site (i.e., not the tryptic peptide sequences) were used for the motif analysis.

Amino acid sequences containing 11 residues, wherein the reported hydroxyproline site is the central residue, were processed in MoMo software to identify plausible hydroxylation motifs (Cheng et al., 2019). Hydroxyproline sites assigned to collagen were removed because prolyl hydroxylation of repeating triplet sequences is a well-established motif (Gorres and Raines, 2010). Flanking sequences of at least 5 amino acids are required by the software to perform the motif analysis. For this reason, hydroxyproline sites that occur within the five residues at either the N- or C-terminus of a protein were removed from the analysis. The motif analysis was applied for each stringency level and the motifs with unadjusted p-value < 0.01 were accepted. This analysis was performed in collaboration with Dr Yoichiro Sugimoto.

3.2.7 Inspection of PDI proteins for hydroxylation assignments

As part of the analysis of novel prolyl hydroxylation sites across the proteome I also considered all proteins bearing a particular consensus within a functional thioredoxin-like domain that had been revealed to be hydroxylated on proline in one particular example. In this work peptides spanning thioredoxin-like domain active sites were extracted from the tryptic proteome and investigated for oxidation in a similar manner to that of the non-HIFα PHD substrates (Section 3.2.4). The exemplar hydroxyproline site occurred in the

-2 position of the thioredoxin-like domain active sites, so this position was assumed to be the ‘target site’ in all PDI proteins.

3.3 Results

The aim of this investigation was to identify and validate novel hydroxyproline sites in the proteome. Studies were performed to i) analyse global protein oxidation, ii) interrogate the steady state proteome for confidently assigned hydroxyproline in reported non-HIF PHD substrates, and iii) identify highly secure hydroxyproline sites from tryptic proteome datasets.

3.3.1 Analysis of protein oxidation in the human proteome

Direct detection of 2OGD-catalysed hydroxylation can be achieved through LC-MS/MS, but the software required to process the raw data can generate false positive assignments. One example of how incorrect assignments may arise is the misassignment of an isobaric modification on residues in proximity to the designated PTM site, either of a single oxidation or dioxidation as two separate oxidation events. Artefactual oxidation, including dioxidation, and non-2OGD-catalysed hydroxylation can arise from multiple pathways, so it is important to understand the prevalence of oxidation in the cellular proteome before performing studies to identify hydroxyproline sites consistent with 2OGD-catalysed prolyl hydroxylation.

In this analysis 12.2 million Peptide Features were nominally assigned in 39 tryptic datasets. Oxidation was assigned to 1.3 million Peptide Features, which represented 10.6 % of the total proteome. Methionine (7.3 % of the total proteome), tryptophan (1.3 %) and tyrosine (0.9 %) were the most abundant assigned sites of oxidation in the proteome (Figure 4). Oxidation of other residues was found to be relatively low.

There are a variety of reasons why oxidation was potentially underestimated at various residues in the proteome. For instance, proline oxidation (0.4 %) was likely to be underrepresented because collagen is secreted into the cell media, which was removed prior to sample collection. Reduction and alkylation of cysteine residues during sample preparation is a possible explanation for why cysteine oxidation (0.1 %) might have been underrepresented. Interestingly, arginine (0.1 %) and lysine (0.1 %) were the least prevalent sites of oxidation in the tryptic proteome. Trypsin cleaves peptide bonds at the

C-terminus of lysyl and arginyl residues, which means protein sequences rich in these residues are likely to be omitted from tryptic proteome datasets. This might explain the low number of oxidations that were observed at these residues.

Overall, the data shows that sites of postulated 2OGD-catalysed hydroxylation are vastly outnumbered by oxidation at sites associated with non-enzymatic oxidation. This is likely to impede correct assignment of 2OGD-catalysed hydroxylation, particularly if MWY are in proximity to the reported site of hydroxylation.

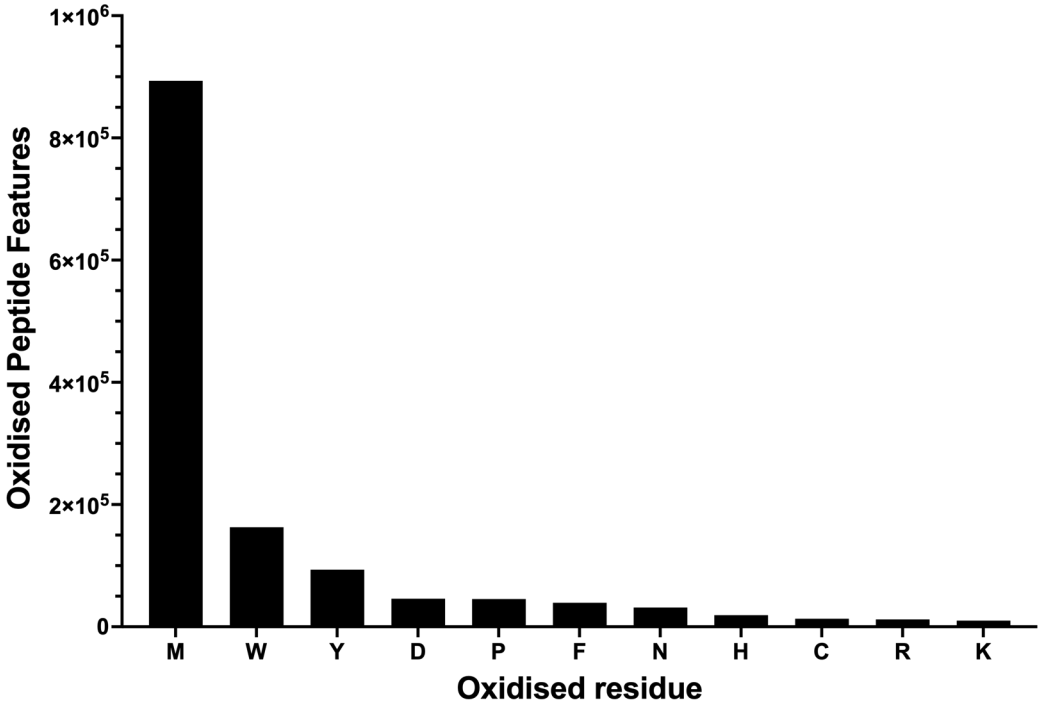


Figure 4 | Methionine, tryptophan and tyrosine oxidation are prevalent PTMs in the proteome.

Mass spectrometry data from 39 publicly deposited 'deep proteome' tryptic datasets were processed in PEAKS® X software. Spectral counting was performed for each oxidised residue in the proteome.

To further consider difficulties in correct assignment of 2OGD-catalysed prolyl hydroxylation, filters were applied to stratify oxidation sites by the level of PTM localisation confidence. Different proteomics software programmes generate different probability scores for PTM localisation, and the parameters used in PEAKS® software are detailed in Section 3.2.4. Filters for AScore (≥ 20) and ion intensity (≥ 5 %), were applied to identify Peptide Features for which the assigned oxidation could be confidently

localised. Oxidation of methionine (35 %), tyrosine (21 %), and tryptophan (18 %) was assigned with high confidence at a higher frequency than other residues (Figure 5). For instance, proline hydroxylation was assigned with high confidence by the software for 2.5 % of the nominal assignments, which was 13-fold lower than methionine oxidation and 7-fold lower than tryptophan or tyrosine oxidation.

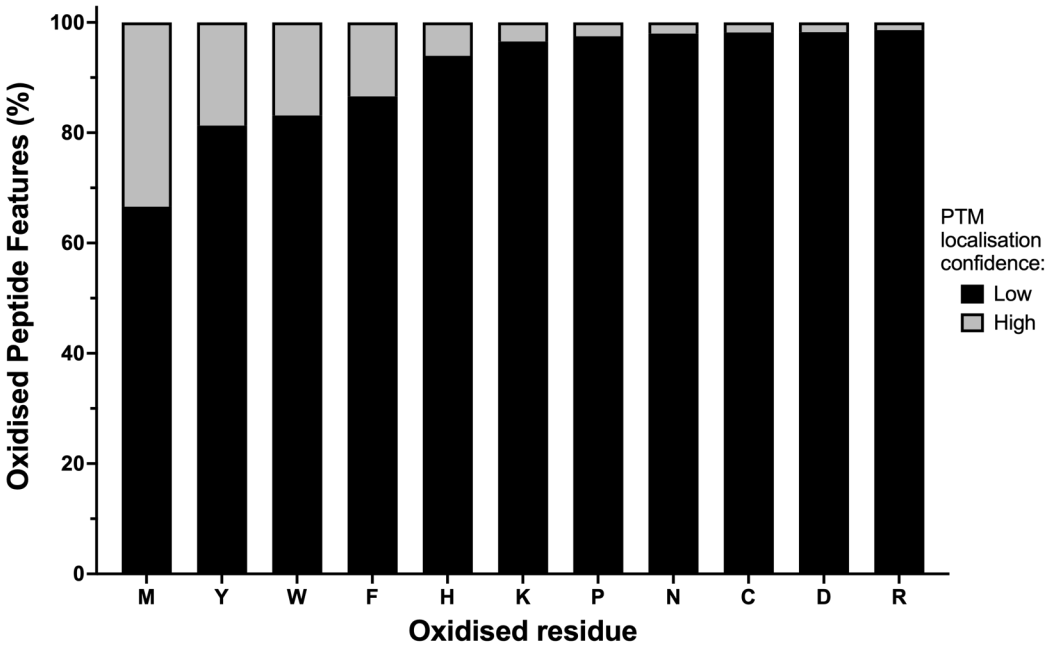


Figure 5 | PEAKS® X assigns oxidation of MYW residues with higher confidence than other residues in the proteome.

1.3 million oxidised Peptide Features were filtered for confident PTM localisation. In total, 0.3 million Peptide Features achieved high PTM localisation confidence, grey bars. 1.0 million oxidised Peptide Features did not meet the PTM localisation criteria (low confidence PTM localisation; black bars).

Given the greater prevalence of artefactual oxidation, these results mean that the risk of confounding oxidations is even greater despite applying confidence filters. For example, methionine oxidation is 18-times more prevalent than prolyl hydroxylation, and methionine oxidation satisfies the high PTM localisation confidence thresholds 13-times more frequently than hydroxyproline assignments. When combined, this indicates the number of methionine oxidations confidently localised by the software is 234-times higher than that of proline. The number of confidently localised oxidations on tryptophan (23-times) and tyrosine (15-times) are also much greater than those of proline. Therefore, a very high level of discriminatory power is required to assign proline hydroxylation

when an oxidised peptide also contains M, W, or Y. Peptides containing hydroxyproline assignments in proximity to residues prone to artefactual oxidation should be treated with extreme caution because oxidation is significantly more likely to occur artefactually than enzymatically.

3.3.2 Reported non-HIFα PHD1-3 substrates

In total, 45 non-HIFα substrates of PHD1-3 catalysed prolyl hydroxylation have been reported but ambiguity concerning the accuracy of the initial assignments remained after *in vitro* hydroxylation assays were unable to recapitulate any PHD-catalysed hydroxylation event on 23 of these substrates under conditions in which prolyl hydroxylation of HIFα was driven to near completion in the reaction (Section 3.1.3) (Cockman *et al.*, 2019). A possible explanation of this was that factors are required to promote hydroxylation in cells that were not present in the activity assays using recombinant enzymes. I therefore aimed to assess the existence of hydroxyproline at all of the reported target sites from the tryptic datasets retrieved from PX.

Proteins for which the target site was not defined (CERKL, PAX2, MAPK7, and EEF2) were not analysed. Oxidation of the target sites was considered before and after the same stringency filters from Section 3.3.1 were applied that would, if passed, increase the likelihood of correct PTM assignment. Non-oxidised peptides and those with oxidation assigned to other residues on the parent ion were also investigated for comparison. The PSM filter (-10lgP ≥ 20) (stringency filter 1) selects for confidently sequenced peptides. This was used in combination with the PTM localisation filters AScore ≥ 20 (stringency filter 2a) or ion intensity ≥ 5 % (stringency filter 2b), as described in Section 3.3.1. The final filter combines all three criteria (stringency filter 3) and peptides that meet this threshold are considered to be assigned with confidence by the software. The data presented here (Table 16) is an updated version of a dataset that was previously published (Cockman *et al.*, 2019).

Table 16 | Analysis of the cellular proteome for oxidised peptides containing reported sites of PHD1-3 catalysed prolyl hydroxylation.

The number of Peptide Features containing reported sites of PHD1-3 catalysed prolyl hydroxylation was counted using a range of stringency filters (1 – 3). Each reported site of prolyl hydroxylation was interrogated individually. The number of hydroxyprolines assigned at the

target site is indicated in red font. Non-oxidised Peptide Features and those containing oxidation at any other position within the parent ion (i.e., "Other") are also presented. The oxidation site is specified when the software generates a confident assignment (stringency filter 3). Stringency filters: -10lgP ≥ 20 (1); '1' and AScore ≥ 20 (2a); '1' and Ion Intensity ≥ 5 % (2b); '2a' and '2b' combined (3).

Protein	Uniprot Accession	Reported site	Number of Peptide Features				
			Assigned	1	2a	2b	3
ACACB	O00763-1	P343*	Non-oxidised	52	n/a	n/a	n/a
		P450*	Non-oxidised	3	n/a	n/a	n/a
			Other	1	1	0	0
		P2131*	Non-oxidised	3	n/a	n/a	n/a
ACTB	P60709-1	P70	Non-oxidised	866	n/a	n/a	n/a
			P70	6	0	0	0
			W79	76	40	58	36
			M82	216	94	185	76
			Other	72	1	8	0
		P307*	Non-oxidised	6806	n/a	n/a	n/a
			P307	3	0	1	0
			M305	3910	504	3789	482
			Y306	132	6	78	6
			Other	256	54	32	0
		P322*	Non-oxidised	1585	n/a	n/a	n/a
			P322	59	6	13	0
			M313	112	8	61	2
			K315	5	2	2	2
			M325	695	78	640	55
			Other	17	0	0	0
ADRB2	P07550-1	P382*; P395*	Non-oxidised	1	n/a	n/a	n/a
ADSL	P30566-1	P24	Non-oxidised	213	n/a	n/a	n/a
			P24	4	0	0	0
			M26	42	30	39	30
			Other	2	0	1	0
AKT1	P31749-1	P125*	Non-oxidised	18	n/a	n/a	n/a
		P313*; P318*	Non-oxidised	97	n/a	n/a	n/a
			Other	1	0	0	0
		P423*	Non-oxidised	43	n/a	n/a	n/a
ARRB2	P32121-1	P176; P179; P181	Non-oxidised	21	n/a	n/a	n/a

ATF4	P18848-1	P156*; P162*; P164*; P167*; P174*	Target peptides not detected			
BCL2L11	O43521-1	P67; P70	Target peptides not detected			
BRD4	O60885-1	P536	Non-oxidised	66	n/a	n/a	n/a
			P536	26	2	4	0
			K535	9	8	1	1
			K535 K537	20	19	13	13
			Other	45	11	19	0
CENPN	Q96H22-1	P311*	Non-oxidised	4	n/a	n/a	n/a
CEP192	Q8TEP8-3	P2313*	Target peptides not detected			
CPT1B	Q92523-1	P295	Target peptides not detected			
DGKI	O75912-1	P903	Non-oxidised	1	n/a	n/a	n/a
DYRK1A	Q13627-1	P380	Non-oxidised	11	n/a	n/a	n/a
DYRK1B	Q9Y463-1	P332	Target peptides not detected			
EEF2K	O00418-1	P98*	Non-oxidised	41	n/a	n/a	n/a
			Other	5	0	0	0
EHMT2	Q96KQ7-1	P676	Unmodified	9	n/a	n/a	n/a
		P1207	Unmodified	47	n/a	n/a	n/a
EPOR	P19235-1	P443*; P450*	Protein is not detected			
FLNA	P21333-1	P2317*; P2324*	Non-oxidised	563	n/a	n/a	n/a
			P2324	1	0	0	0
			Other	3	0	0	0
FOXO3	O43524-1	P426*; P427*	Non-oxidised	4	n/a	n/a	n/a
IKBKB	O14920-1	P191*	Target peptides not detected			
MAPK6	Q16659-1	P25*	Target peptides not detected			
MLXIPL	Q9NP71-1	P15	Target peptides not detected			
		P141	Non-oxidised	9	n/a	n/a	n/a
		P526	Target peptides not detected			
NDRG3	Q9UGV2-1	P294*	Non-oxidised	46	n/a	n/a	n/a
			Other	11	5	0	0
OTUB1	Q96FW1-1	P210	Non-oxidised	68	n/a	n/a	n/a
			M211	26	13	24	11
		P263	Non-oxidised	65	n/a	n/a	n/a
			Other	3	0	0	0
PDE4D	Q08499-1	P29*	Target peptides not detected			

		P382*	Non-oxidised	23	n/a	n/a	n/a
			M371	14	13	9	9
			Other	1	0	0	0
		P419*	Non-oxidised	7	n/a	n/a	n/a
PKM	P14618-1	P403*; P408*	Non-oxidised	3405	n/a	n/a	n/a
			P403	5	1	0	0
			P408	5	0	0	0
			Other	10	1	0	0
POLR2A	P24928-1	P1465*	Non-oxidised	21	n/a	n/a	n/a
			Other	3	0	2	0
PPP2R2A	P63151-1	P319*	Non-oxidised	118	n/a	n/a	n/a
			W311	24	19	10	10
			M315	65	24	36	21
			Other	21	6	0	0
SFMBT1	Q9UHJ3-1	P106	Non-oxidised	1	n/a	n/a	n/a
		P651	Non-oxidised	14	n/a	n/a	n/a
SPRY2	O43597-1	P18*	Non-oxidised	5	n/a	n/a	n/a
		P144*	Non-oxidised	1	n/a	n/a	n/a
		P160*	Non-oxidised	1	n/a	n/a	n/a
TBK1	Q9UHD2	P48	Non-oxidised	23	n/a	n/a	n/a
TELO2	Q9Y4R8-1	P374*	Non-oxidised	41	n/a	n/a	n/a
		P419*; P422*	Non-oxidised	17	n/a	n/a	n/a
TET2	Q6ND21-1	P1335	Non-oxidised	2	n/a	n/a	n/a
		P1342	Non-oxidised	7	n/a	n/a	n/a
TET3	O43151-1	P1030; P1037	Target peptides not detected				
THRA	P10827-1	P160*; P162*	Non-oxidised	7	n/a	n/a	n/a
TP53	P04637-1	P142*	Non-oxidised	19	n/a	n/a	n/a
		P359*	Non-oxidised	1	n/a	n/a	n/a
TRPA1	O75762-1	P394*	Target peptides not detected				
UBE2N	P61088-1	P19; P21	Non-oxidised	151	n/a	n/a	n/a
		P59	Non-oxidised	281	n/a	n/a	n/a
			P59	4	0	0	0
			Other	308	0	85	0
UBE2V1	Q13404-4	P79; P80	Non-oxidised	80	n/a	n/a	n/a
			Other	3	0	1	0

ZHX2	Q9Y6X8-1	P427	Non-oxidised	23	n/a	n/a	n/a
		P440	Non-oxidised	16	n/a	n/a	n/a
		P464	Non-oxidised	35	n/a	n/a	n/a

*These target sites were tested in assays of PHD-catalysed hydroxylation by Cockman *et al.* (2019).

The analysis of all public data sets revealed the following:

- Fewer than 0.1 % of the peptides assigned to the target sites were reported with hydroxyproline.
- None of the hydroxyproline assignments were generated with high confidence from the software.
- Other sites of oxidation were assigned to many of the peptides.
- Other sites of oxidation were generally assigned with more Peptide Features than the hydroxyproline peptides.
- Other sites of oxidation were generally assigned with greater PTM localisation confidence than the peptides containing hydroxyproline assignments.

One example of oxidation being more prevalent and assigned with greater confidence than the reported hydroxyproline site is ACTB P307; oxidation of M305 and Y306 are both more likely to be correct than the hydroxyproline assignment (Figure 6A). For all of the proteins that were examined in this analysis, oxidation was only confidently assigned to MWY, with one exception, BRD4. In contrast to HyP536, the assignments of BRD4 HyK535 and HyK537 were confidently localised by the software (Figure 6B). Together, the data generated from publicly deposited proteomic datasets did not provide support for the *bona fide* existence of prolyl hydroxylation at reported sites in proposed substrates of the HIFα prolyl hydroxylases in the cellular proteome.

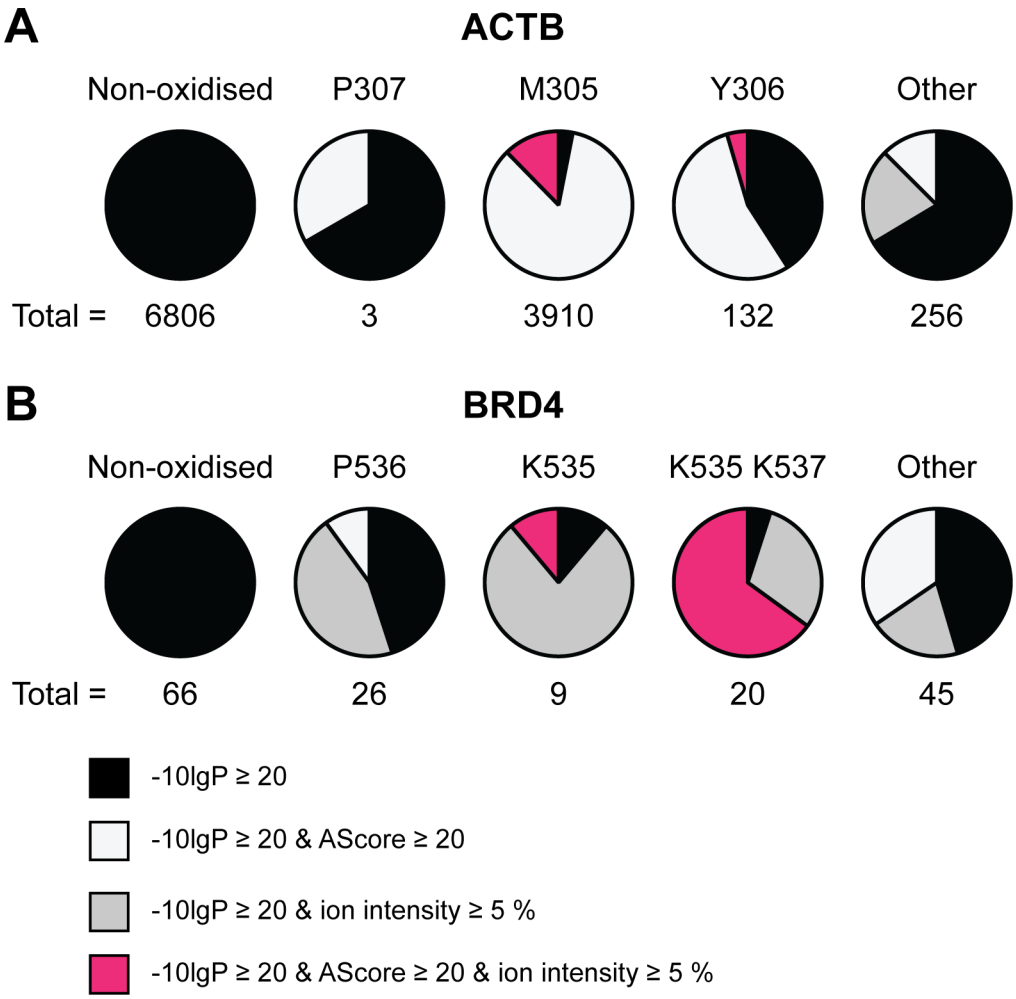


Figure 6 | Illustration of high confidence oxidation on residues close to the reported site of prolyl hydroxylation.

Oxidised and non-oxidised tryptic Peptide Features containing ACTB P307 (A) and BRD4 P536 (B) were counted over a range of PTM localisation stringency filters. The tryptic data from all of the public data sets was analysed. Peptide Features meeting the highest stringency criteria (-10lgP ≥ 20 & AScore ≥ 20 & ion intensity ≥ 5 %; pink segments) are accepted as nominally correct assignments and these oxidation sites are indicated. Peptide Features containing oxidation at any other position within the parent ion (i.e., "Other") are also presented. The total number of Peptide Features containing the indicated oxidations is reported below each pie chart.

The reported non-HIFα PHD substrates do not display confidently localised hydroxyproline but oxidation of other residues in the tryptic peptides are confidently assigned. This was predicted by the results of global protein oxidation (Section 3.3.1). Together with the work of Cockman *et al.* (2019), these results suggest some of the hydroxyproline assignments were probably incorrect in the original reports. This highlights the need for a cautious approach to the identification of novel hydroxyproline sites so as to limit incorrect assignments as far as possible.

3.3.3 Deep proteome analysis of proline oxidation consistent with 2OGD-catalysed hydroxylation

Despite the lack of support for reported non-HIFα PHD substrates it is possible other proteins are subject to prolyl hydroxylation by PHD or other prolyl hydroxylases. I therefore interrogated the same tryptic datasets for evidence of novel hydroxyproline assignments that might be catalysed by any prolyl hydroxylase.

The results of global protein oxidation surveys (Section 3.3.1) and interrogation of reported non-HIFα substrates (Section 3.3.2) suggest low confidence sites of proline hydroxylation are likely to be misassigned, particularly if residues susceptible to artefactual oxidation are present in the peptide. The application of PTM localisation filters might remove low confidence hydroxyproline assignments but manual inspection of MS2 spectra is ultimately required for accurate and specific PTM localisation by determination of the diagnostic ions. However, this process is time consuming. I therefore applied the same criteria as for the non-HIFα PHD substrates with additional filters to identify novel hydroxyproline sites. These filters were designed to objectively eliminate hydroxyproline-containing peptides that exhibit a high probability of false assignment. A single Peptide Feature that fulfilled all of the filtering criteria was considered to be a true positive site of prolyl hydroxylation.

In this analysis, 42,283 Peptide Features containing hydroxyproline passed the confident PSM criteria, which corresponded to 10,733 discrete hydroxyproline sites (Figure 7). The data was refined to identify confident sites of prolyl hydroxylation, at least according to PEAKS®. After applying the PTM localisation criteria used in Section 3.3.1 and 3.3.2 for confident PTM assignment, 1,350 hydroxyproline-containing Peptide Features (469 discrete HyP sites) remained. There were 585 Peptide Features (75 sites) that displayed characteristics consistent with irreversible 2OGD-catalysed hydroxylation. After refining hydroxyproline-containing peptides according to the quality of the MS1 spectra, 362 Peptide Features (52 sites) remained. Manual inspection of the MS2 spectra was performed on all 362 Peptide Features to identify candidates with strong supporting evidence for the hydroxyproline assignment. There were 214 Peptide Features (31 sites) that met these criteria.

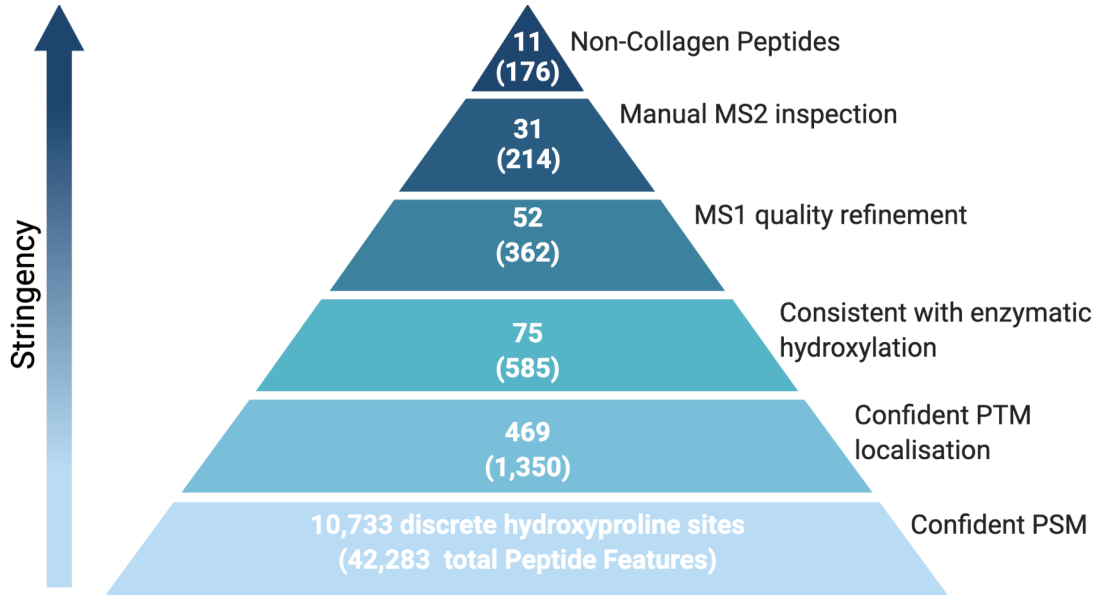


Figure 7 | Identification and validation of 11 unique hydroxyproline sites in the tryptic proteome.

A series of filters was applied to increase the confidence of correct hydroxyproline assignment in the tryptic proteome. The number of discrete hydroxyproline sites and the corresponding number of Peptide Features indicate the prevalence and abundance, respectively, of hydroxyproline that meet the criteria at each stringency level.

The confirmation of 20 collagen hydroxyproline sites by manual MS2 inspection provided proof of principle that this workflow can identify *bona fide* prolyl hydroxylase substrates. Finally, collagen peptides were removed to give 176 hydroxyproline-containing Peptide Features, corresponding to 11 discrete hydroxyproline sites (Table 17).

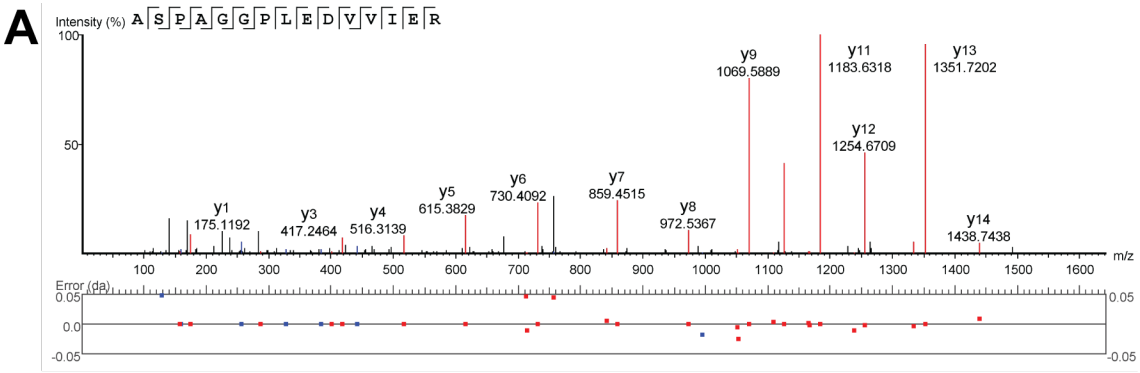
Table 17 | Candidates from deep proteome profiling.

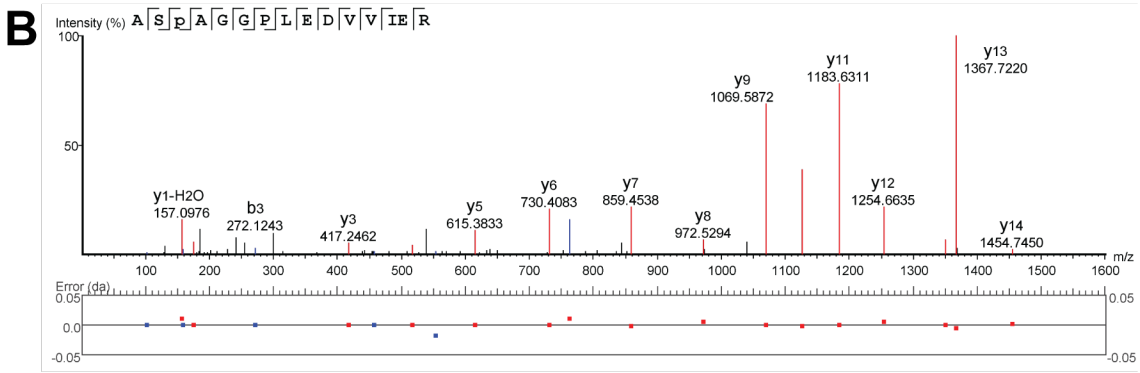
Summary table of representative MS2 spectra for each protein candidate and the corresponding hydroxyproline site (Site). The Confidence scores (-10lgP and AScore) indicate the software-designated confidence of the PSM and PTM localisation, respectively. Figures 14-20 are presented in the Appendix (Section 3.5)

Protein	Site	-10lgP	AScore	Source file	Scan	Fig.
CASC4	HyP252	38.45	26.02	20151022_QE5_UPLC10_DBJ_SA_HCT116_REP1_46frac_12.raw	7956	15
FKBP10	HyP36	36.72	59.3	20151022_QE5_UPLC10_DBJ_SA_293_REP2_46frac_19.raw	16696	8

HYOU1	HyP977	45.26	28.7	20151022_QE5_UPLC10_DBJ_SA_293_REP2_46frac_6.raw	13381	16
LMAN1	HyP378	60.96	26.02	20100611_Velos1_TaGe_SA_Hela_1.raw	13906	10
MUC5B	HyP2516	42.09	69.29	20151028_QE3_UPLC8_DBJ_SA_MCF7_Rep1_46frac_40.raw	16496	18
PDIA3	HyP404	48.27	23.1	20100723_Velos1_TaGe_SA_Hek293_03.raw	31398	11
PRKCSH	HyP290	44.98	39.76	20151028_QE5_UPLC10_DBJ_SA_MCF7_REP2_46frac_22.raw	18199	17
SERPINH1	HyP30	57.32	100.21	20101215_Velos1_TaGe_SA_A549_01.raw	1652	21
SLC38A10	HyP532	46.56	143.14	20151022_QE5_UPLC10_DBJ_SA_SH-SY5Y_REP2_46frac_20.raw	11319	20
SUN2	HyP315	45.9	55.39	20151023_QE3_UPLC8_DBJ_SA_SH-SY5Y_Rep1_46frac_43.raw	25586	19
TMEM109	HyP42	38.34	20	20101224_Velos1_TaGe_SA_HeLa_01.raw	2821	12

To exemplify the process of manual inspection, the MS2 spectra supporting FKBP10 P36 as a site of prolyl hydroxylation were illustrated in Figure 8. The MS2 spectra of the non-hydroxylated and hydroxyproline-containing peptides awee highly similar, suggesting the same peptide sequence had been fragmented in each mass spectrum. The fragment ions in the spectrum were matched throughout the y-ion series at high abundance relative to the noise of the MS2 spectrum, which provided security of correct peptide sequencing. The diagnostic ions (y12 and y13) are highly abundant in each MS2 spectrum. When comparing the two spectra, all y-ions up to and including the y12 ion were of equal mass, within the error tolerance limits of the analysis. The fragment ions demonstrated a mass shift of +15.99 Da on the y13 fragment ion of the hydroxylated peptide compared to that of the non-hydroxylated peptide, and this was also observed on the subsequent fragment ion. This peptide sequence did not contain any residues prone to artefactual oxidation that might have generated neutral loss ions, which simplified the analysis. Together, these factors provided strong evidence to accept the hydroxyproline assignment as correct. The same approach was taken for all MS2 spectra investigated throughout this thesis.





C

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	44.0499	72.0449	A				15
2	60.0449	159.0765	S	1438.7438	1420.7430	1421.7266	14
3	70.0656	256.1292	P	1351.7202	1333.7153	1334.6946	13
4	44.0499	327.1671	A	1254.6709	1236.6583	1237.6543	12
5	30.0343	384.1868	G	1183.6318	1165.6182	1166.6073	11
6	30.0343	441.2087	G	1126.6119	1108.5997	1109.5780	10
7	70.0656	538.2625	P	1069.5889	1051.5853	1052.5884	9
8	86.0969	651.3466	L	972.5367	954.5255	955.5090	8
9	102.0554	780.3892	E	859.4515	841.4346	842.4249	7
10	88.0398	895.4161	D	730.4092	712.3511	713.3936	6
11	72.0812	994.5040	V	615.3829	597.3718	598.3554	5
12	72.0812	1093.5529	V	516.3139	498.3034	499.2870	4
13	86.0969	1206.6370	I	417.2464	399.2350	400.2181	3
14	102.0554	1335.6796	E	304.1615	286.1509	287.1345	2
15	129.0652		R	175.1192	157.1097	158.0920	1

D

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	44.0499	72.0449	A				15
2	60.0449	159.0763	S	1454.7450	1436.7380	1437.7216	14
3	86.0605	272.1243	P(+15.99)	1367.7220	1349.7075	1350.6895	13
4	44.0499	343.1617	A	1254.6635	1236.6583	1237.6418	12
5	30.0343	400.1832	G	1183.6311	1165.6212	1166.6047	11
6	30.0343	457.2032	G	1126.6125	1108.5997	1109.5833	10
7	70.0656	554.2759	P	1069.5872	1051.5782	1052.5618	9
8	86.0969	667.3415	L	972.5294	954.5255	955.5090	8
9	102.0551	796.3840	E	859.4538	841.4413	842.4249	7
10	88.0398	911.4110	D	730.4083	712.3988	713.3824	6
11	72.0812	1010.4794	V	615.3833	597.3718	598.3554	5
12	72.0812	1109.5479	V	516.3141	498.3034	499.2870	4
13	86.0969	1222.6318	I	417.2462	399.2350	400.2186	3
14	102.0554	1351.6744	E	304.1615	286.1509	287.1345	2
15	129.1139		R	175.1190	157.0976	158.0919	1

Figure 8 | Representative MS2 spectra of FKBP10 HyP36.

Representative MS2 spectra, and corresponding fragment ion tables, of tryptic fragments containing the target site, FKBP10 P36, from the Bekker Jensen HEK293T Replicate 2 dataset. A) and C) MS2 spectrum of the Non-oxidised peptide with corresponding fragment ion table. B) and D) MS2 spectrum of the oxidised peptide with corresponding fragment ion table. Both MS2 spectra convey confident peptide sequencing through the y-ion series. Manual inspection of the diagnostic fragment ions (y12 and y13) enables confident PTM localisation of the +15.99 Da shift to P36.

3.3.4 Identification of false negative assignments

The majority of putative hydroxyproline sites were filtered out during the previous analysis to yield 31 high confidence sites of prolyl hydroxylation. The filters aimed to remove false positive hydroxyproline assignments as far as possible to increase the likelihood of identifying novel *bona fide* 2OGD substrates. The highly stringent parameters also eliminated true positives, as seen by the attrition of collagen; there were

453 discrete collagen hydroxyproline sites that met the confident PSM filter but only 20 of these sites were accepted after the high stringency workflow, including manual MS2 inspection. These results suggested that the high stringency workflow generated false negatives in addition to removing false positives. I therefore tested if this was true of the 11 novel high confidence hydroxyproline assignments. The number of Peptide Features assigned to these 11 hydroxyproline sites was diminished after application of the filters (Figure 9) but the attrition applied more strikingly to certain sites than others, as indicated below.

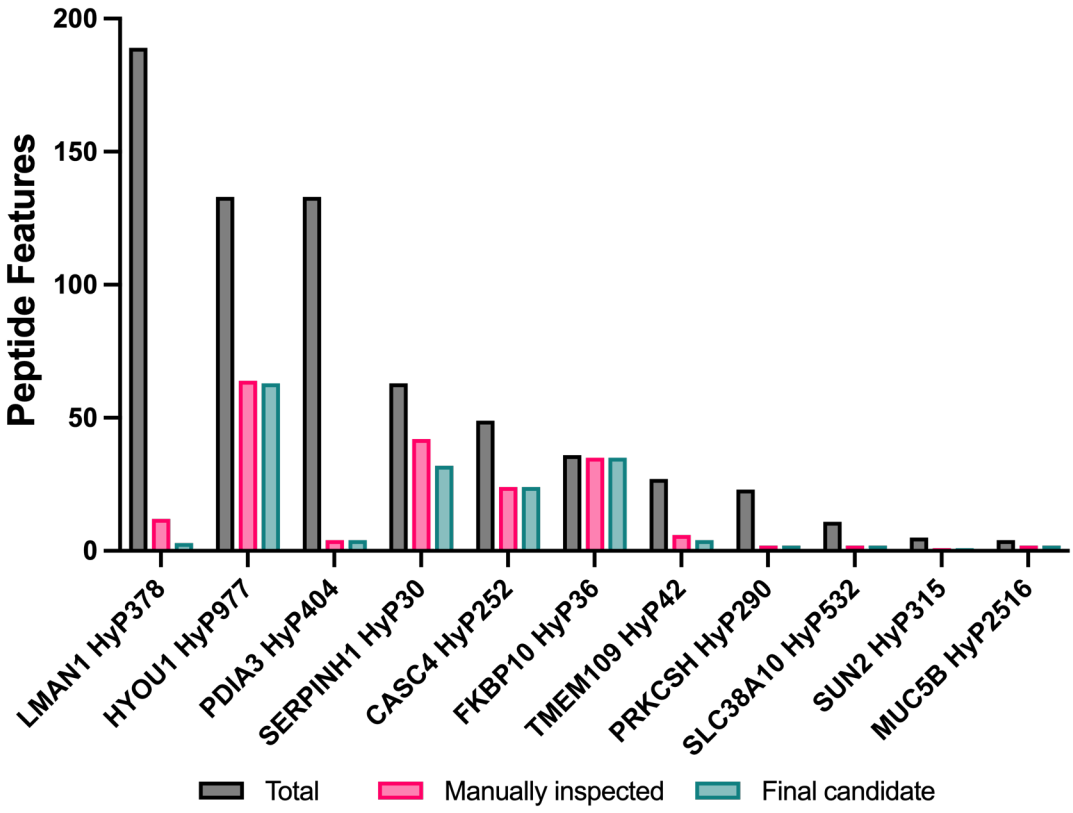


Figure 9 | Final candidates from high stringency hydroxyproline analysis.

11 non-collagen hydroxyproline sites were identified from the high stringency analysis. The bar chart presents the total number of Peptide Features (i.e., before filtering), those that were considered for manual MS2 inspection, and the final candidates that fulfilled all criteria.

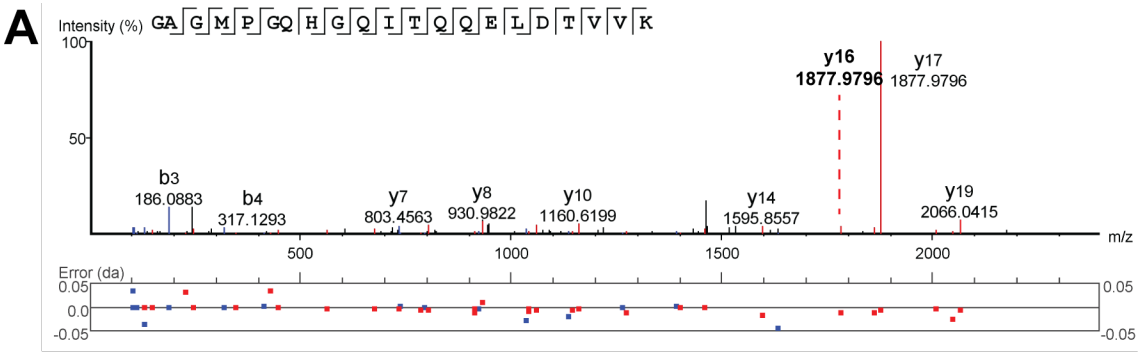
3.3.4.1 LMAN1 HyP378 exemplifies some complexities of hydroxyproline assignment towards the N-terminus of long peptides containing methionine

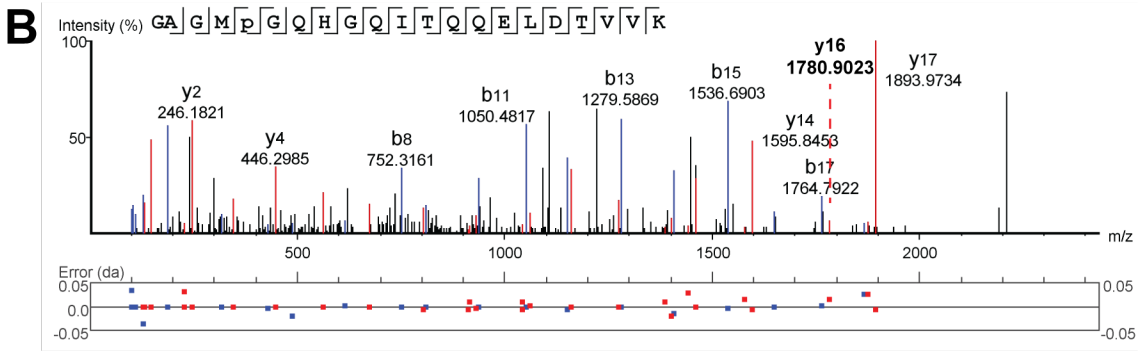
LMAN1 HyP378 was the most abundant of the final candidates before filtering; 189 nominally assigned Peptide Features were reported containing the hydroxyproline site. However, only 12 Peptide Features were considered for manual inspection and only 3 of these were accepted. Therefore, the features supporting hydroxylation or creating uncertainty over the assignment were re-examined in detail. A typical set of MS2 spectra corresponding to the hydroxylated and non-hydroxylated LMAN1 peptides are illustrated in Figure 10.

The diagnostic y17 fragment ion (i.e., the first y-ion containing P378) is readily identified in both MS2 spectra. The y16 ion of 12 Peptide Features met the 5 % ion intensity threshold but 9 MS2 spectra exhibited many unmatched fragment ions in this region of the MS2 spectra (i.e., the y16 ions exhibit a low signal to noise ratio), and this reduced the confidence of correct assignment for those MS2 spectra. This suggests the attrition of LMAN1 HyP378 peptides is most likely to be due to low relative ion intensity of the y16 ion, which is one of the diagnostic ions. Low relative intensity of the y16 fragment ion is most likely to be a consequence of both the proline effect and peptide length (Barton and Whittaker, 2009; Dong *et al.*, 2011), and can have a negative impact on AScore (Beausoleil *et al.*, 2006).

For the LMAN1 HyP378 MS2 spectra, -64 Da neutral loss ions were not identified, which would otherwise indicate artefactual oxidation of M377 that could have been misassigned by PEAKS®. LMAN1 was sometimes assigned with M377 oxidation instead of HyP378. The -64 Da neutral loss ions were present in the MS2 spectrum of oxidised M377 peptides but not HyP378 peptides, which was a pertinent discrepancy between the two assignments. Another important distinction between these isobaric peptides was the ΔRT; peptides assigned with M377 oxidation exhibited a greater -ΔRT relative to the non-oxidised parent ion than peptides assigned with HyP378, which was consistent with previous observations (Cockman *et al.*, 2019). This was one example of how methionine-oxidised peptides could be distinguished from those containing hydroxyproline by manually inspecting the MS1 and MS2 spectra.

Investigation of the filtering workflow indicated that LMAN1 HyP378 peptides were often disqualified on the basis of low ion intensity and low AScore (Figure 10E), which was partly due to the position of the hydroxyproline site towards the N-terminus of a long peptide. It is highly likely that other hydroxyproline-containing peptides with similar characteristics have been rejected on the basis of low signal to noise ratio of the diagnostic ions, which may present as low ion intensity or low AScore, or both, and increase the difficulty of manual MS2 inspection.





C

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	30.0343	58.0293	G				21
2	44.0499	129.1029	A	2137.0706	2119.0601	2120.0435	20
3	30.0343	186.0883	G	2066.0415	2048.0229	2049.0325	19
4	104.0532	317.1293	M	2009.0148	1991.0015	1991.9850	18
5	70.0656	414.1775	P	1877.9796	1859.9611	1860.9562	17
6	30.0343	471.2026	G	1780.9303	1762.9083	1763.8918	16
7	101.0712	599.2611	Q	1723.8973	1705.8868	1706.8704	15
8	110.0715	736.3162	H	1595.8557	1577.8282	1578.8118	14
9	30.0343	793.3398	G	1458.7819	1440.7693	1441.7528	13
10	101.0714	921.4053	Q	1401.7561	1383.7478	1384.7313	12
11	86.0969	1034.5125	I	1273.7113	1255.6893	1256.6729	11
12	74.0605	1135.5519	T	1160.6199	1142.6052	1143.5947	10
13	101.0714	1263.5925	Q	1059.5737	1041.5626	1042.5513	9
14	101.0714	1391.6444	Q	931.4983	913.5109	914.4858	8
15	102.0554	1520.6915	E	803.4563	785.4459	786.4239	7
16	86.0969	1633.8214	L	674.4125	656.3977	657.3813	6
17	88.0398	1748.8026	D	561.3287	543.3137	544.2972	5
18	74.0605	1849.8502	T	446.2992	428.2867	429.2334	4
19	72.0812	1948.9187	V	345.2513	327.2390	328.2226	3
20	72.0812	2047.9871	V	246.1823	228.1356	229.1542	2
21	101.0712		K	147.1136	129.1029	130.0865	1

D

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	30.0343	58.0293	G				21
2	44.0499	129.1026	A	2153.0654	2135.0549	2136.0383	20
3	30.0343	186.0879	G	2082.0283	2064.0178	2065.0012	19
4	104.0530	317.1281	M	2025.0070	2006.9965	2007.9800	18
5	86.0605	430.1799	P(+15.99)	1893.9734	1875.9268	1876.9395	17
6	30.0343	487.2175	G	1780.9023	1762.9083	1763.8918	16
7	101.0711	615.2523	Q	1723.8973	1705.8868	1706.8704	15
8	110.0713	752.3161	H	1595.8453	1577.8119	1578.8118	14
9	30.0343	809.3384	G	1458.7795	1440.7379	1441.7528	13
10	101.0714	937.3949	Q	1401.7784	1383.7478	1384.7190	12
11	86.0969	1050.4817	I	1273.6985	1255.6893	1256.6729	11
12	74.0605	1151.5338	T	1160.6156	1142.6052	1143.5887	10
13	101.0714	1279.5869	Q	1059.5638	1041.5446	1042.5476	9
14	101.0714	1407.6575	Q	931.5145	913.5068	914.4691	8
15	102.0554	1536.6903	E	803.4585	785.4403	786.4239	7
16	86.0969	1649.7732	L	674.4084	656.3977	657.3813	6
17	88.0398	1764.7922	D	561.3254	543.3137	544.2972	5
18	74.0605	1865.8177	T	446.2985	428.2867	429.2703	4
19	72.0812	1964.9136	V	345.2506	327.2390	328.2226	3
20	72.0812	2063.9819	V	246.1821	228.1715	229.1193	2
21	101.0711		K	147.1133	129.1026	130.0867	1

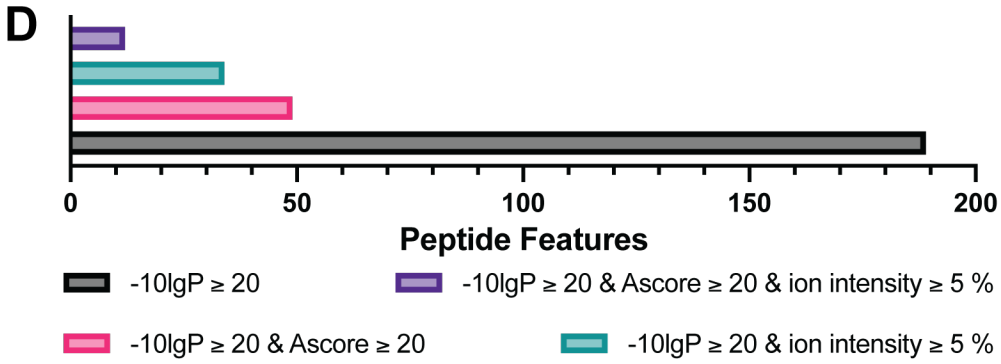


Figure 10 | Representative MS2 spectra of LMAN1 HyP387.

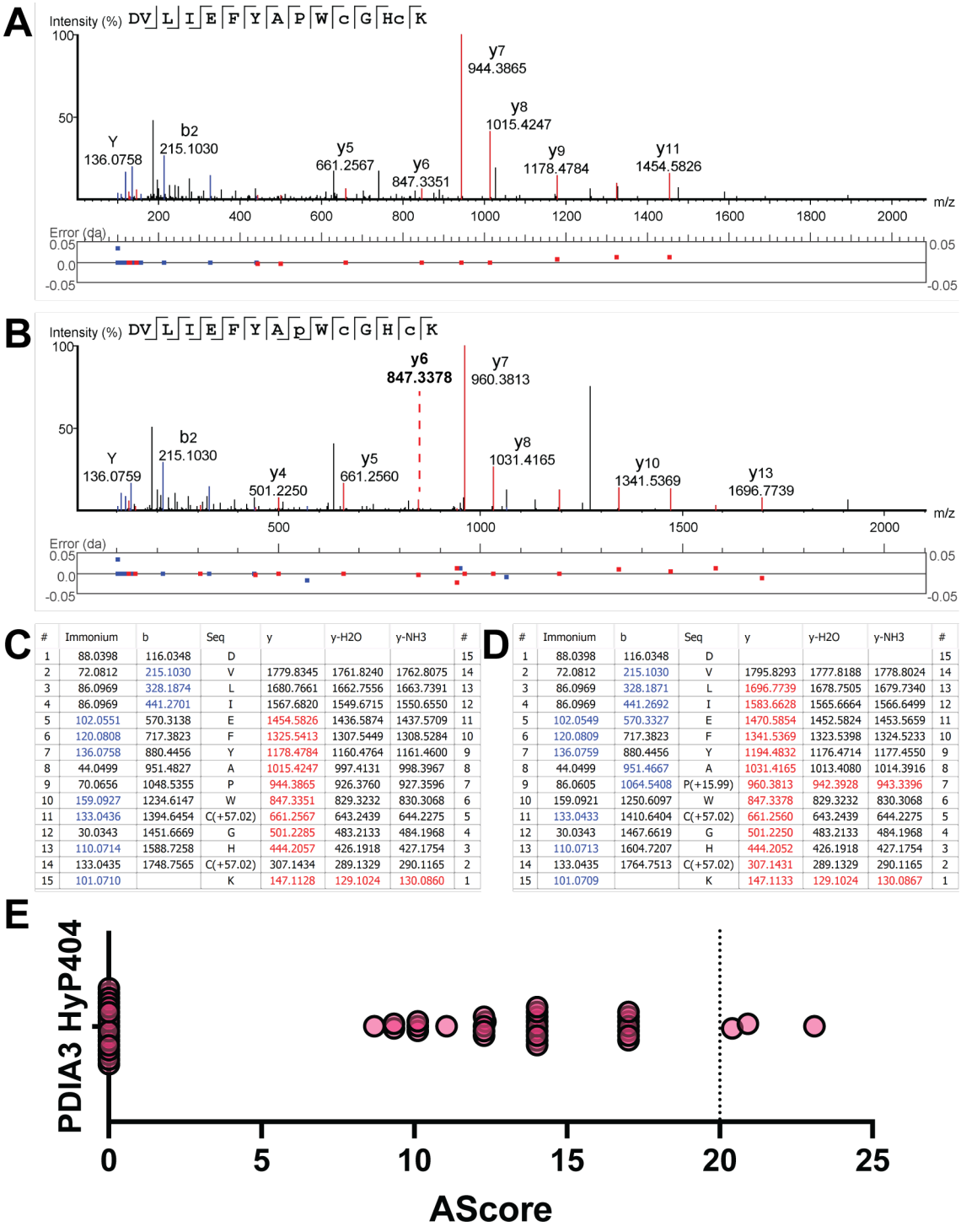
Representative MS2 spectra and corresponding ion tables of tryptic fragments containing the target site, P378, from the Geiger HeLa CID dataset. A) and C) MS2 spectrum of the Non-oxidised peptide with corresponding fragment ion table. B) and D) MS2 spectrum of the oxidised peptide with corresponding fragment ion table. The MS2 spectra have been annotated to indicate the y16 fragment ions (red dashed line, bold typeface). Both MS2 spectra convey confident

peptide sequencing through the y-ion series. Manual inspection of the diagnostic fragment ions (y16 and y17) enables confident PTM localisation of the +15.99 Da shift to P378. Note, the b4 and b5 fragment ions are matched in both MS2 spectra but the background noise in this region of the spectra means these ions cannot be used to add further confidence regarding the oxidation to P378. E) The number of Peptide Features containing LMAN1 HyP378 after filters were applied for different PTM localisation criteria.

3.3.4.2 PDIA3 HyP404 exemplifies computationally difficult PTM localisation in complex peptide sequences

For PDIA3 HyP404, all 4 of the Peptide Features that were manually inspected at the MS2 level were accepted. Manual inspection involved the identification of the y6 and y7 fragment ions (Figure 11), which corroborated the HyP404 assignment. Additional observations strengthened the security of this assignment. W405 was a possible site of artefactual oxidation but this was not observed either on the y6 fragment ion or from an increased abundance of -H_2O neutral loss ions after the y6 ion (Lioe *et al.*, 2004). There are two carbamidomethylated cysteine residues in proximity to the target site and misassigned oxidation of these residues was possibility (Steen and Mann, 2001). The neutral loss ions generated from oxidised carbamidomethylated cysteine (-107 Da) were not observed in the MS2 spectra assigned to PDIA3 HyP404. This was another example of manually inspecting MS2 spectra to distinguish between hydroxyproline assignments and isobaric artefactual oxidations.

Following this the attrition of hydroxyproline-bearing Peptide Features across the filtering steps was re-examined to consider why so many Peptide Features were considered to be insecure. This revealed that of the 48 PSMs (133 Peptide Features) reported containing PDIA3 HyP404, 45 PSMs (129 Peptide Features) were disqualified because of their low AScore (Figure 11E). Inspection of the peptide sequence revealed that 10 of the 15 residues in this peptide are sites to which PEAKS® had been instructed to consider oxidation as a variable modification. The complexity of PTM localisation of this peptide sequence had a negative impact on AScore, especially when putative oxidation sites were adjacent to the hydroxyproline assignment (Beausoleil *et al.*, 2006). These results suggest it was highly likely that other sites of prolyl hydroxylation were inappropriately disqualified on the basis of a low AScore value that is, in part, influenced by a peptide sequence containing many possible sites of oxidation.



#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	88.0398	116.0348	D				15
2	72.0812	215.1030	V	1779.8345	1761.8240	1762.8075	14
3	86.0969	328.1874	L	1680.7661	1662.7556	1663.7391	13
4	86.0969	441.2701	I	1567.6820	1549.6715	1550.6550	12
5	102.0551	570.3138	E	1454.5826	1436.5874	1437.5709	11
6	120.0808	717.3823	F	1325.5413	1307.5449	1308.5284	10
7	136.0758	880.4456	Y	1178.4784	1160.4764	1161.4600	9
8	44.0499	951.4827	A	1015.4247	997.4131	998.3967	8
9	70.0656	1048.5355	P	944.3865	926.3760	927.3596	7
10	159.0927	1234.6147	W	847.3351	829.3232	830.3068	6
11	133.0436	1394.6454	C(+57.02)	661.2567	643.2439	644.2275	5
12	30.0343	1451.6669	G	501.2285	483.2133	484.1968	4
13	110.0714	1588.7258	H	444.2057	426.1918	427.1754	3
14	133.0435	1748.7565	C(+57.02)	307.1434	289.1329	290.1165	2
15	101.0710		K	147.1128	129.1024	130.0860	1

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	88.0398	116.0348	D				15
2	72.0812	215.1030	V	1795.8293	1777.8188	1778.8024	14
3	86.0969	328.1871	L	1696.7739	1678.7505	1679.7340	13
4	86.0969	441.2692	I	1583.6628	1565.6664	1566.6499	12
5	102.0549	570.3327	E	1470.5854	1452.5824	1453.5659	11
6	120.0809	717.3823	F	1341.5369	1323.5398	1324.5233	10
7	136.0759	880.4456	Y	1194.4832	1176.4714	1177.4550	9
8	44.0499	951.4667	A	1031.4165	1013.4080	1014.3916	8
9	86.0605	1064.5408	P(+15.99)	960.3813	942.3928	943.3396	7
10	159.0921	1250.6097	W	847.3378	829.3232	830.3068	6
11	133.0433	1410.6404	C(+57.02)	661.2560	643.2439	644.2275	5
12	30.0343	1467.6619	G	501.2250	483.2133	484.1968	4
13	110.0713	1604.7207	H	444.2052	426.1918	427.1754	3
14	133.0435	1764.7513	C(+57.02)	307.1431	289.1329	290.1165	2
15	101.0709		K	147.1133	129.1024	130.0867	1

Figure 11 | Representative MS2 spectra of PDIA3 HyP404.

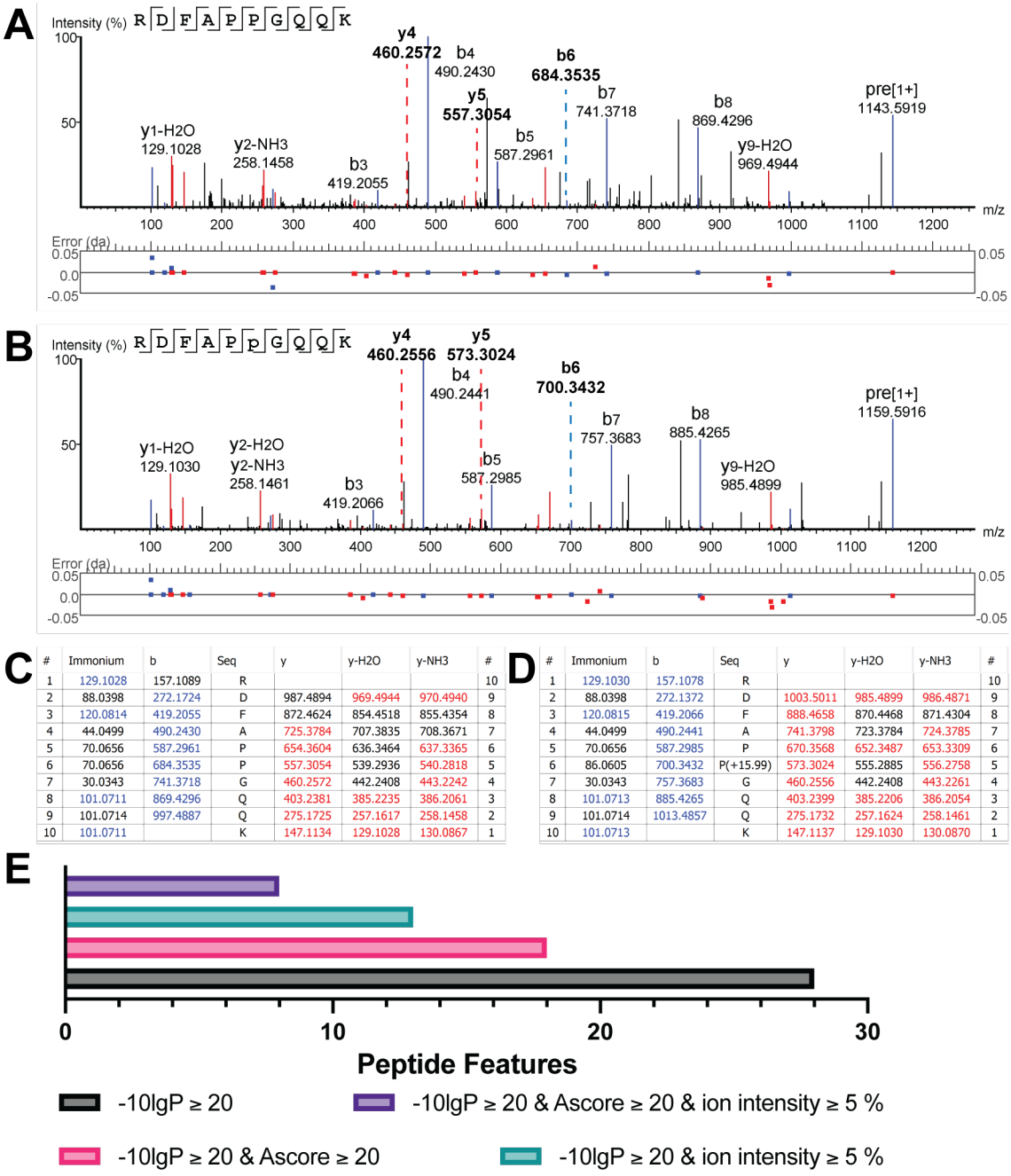
Representative MS2 spectra and corresponding ion tables of tryptic fragments containing the target site, P404, from the Geiger HEK293 HCD dataset. A) and C) MS2 spectrum of the Non-oxidised peptide with corresponding fragment ion table. B) and D) MS2 spectrum of the oxidised peptide with corresponding fragment ion table. Both MS2 spectra convey confident peptide sequencing through the y-ion series. The oxidised MS2 spectrum has been annotated to indicate the y6 fragment ion (red dashed line, bold typeface). Manual inspection of the diagnostic fragment

ions (y6 and y7) enables confident PTM localisation of the +15.99 Da shift to P404. E) AScore distribution for all PSMs containing PDIA3 HyP404. The dotted line indicates the threshold used in high stringency workflow.

3.3.4.3 TMEM109 HyP42 exemplifies complexities of N-terminal miscleavage and the proline effect in adjacent prolyl residues

For the TMEM109 HyP42 peptide, 27 Peptide Features were identified containing the reported site of hydroxylation, 8 of which were considered for manual inspection and only 4 were retained after manual inspection of the MS2 spectra. In TMEM109 the target tryptic peptide containing HyP42 contained arginine at the N-terminus, which explained the abundant b-ion series in the MS2 spectrum (Barton *et al.*, 2007). The target site was the second consecutive proline residue (^{40}APP^{42}) and cleavage of peptide bonds between Pro-Pro is inefficient compared to that of Ala-Pro (Breci *et al.*, 2003), which was observed in the MS2 spectra (Figure 12). Despite these limitations the diagnostic ions were present in both the b- and y-ion series, confident peptide sequencing was achieved in both fragment ion series, and the fragmentation pattern was highly similar between the non-hydroxylated and hydroxyproline peptides. The mass shift was observed on the y5 and b6 ions, which strengthened the security of the assignment.

The precursor ion was highly abundant, which suggested fragmentation of this peptide could be improved. Inefficient fragmentation, both for the peptide as a whole and either side of the target site, is likely to contribute to low ion intensity and AScore values. This was because the resulting fragment ions displayed low relative ion intensity. Consistent with this hypothesis, 10 Peptide Features were eliminated due to low AScore values and 9 Peptide Features were removed because the ion intensity filter criteria were not met (Figure 12E). Manual MS2 inspection enabled TMEM109 HyP42 to be accepted as a high confidence hydroxyproline assignment. This process also highlighted some of the difficulties concerning PTM localisation confidence in peptides that display N-terminal missed cleavage events and sequences containing consecutive prolyl residues.



#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	129.1028	157.1089	R				10
2	88.0398	272.1724	D	987.4894	969.4940	970.4940	9
3	120.0814	419.2055	F	872.4624	854.4518	855.4354	8
4	44.0499	490.2430	A	725.3784	707.3835	708.3671	7
5	70.0656	587.2961	P	654.3604	636.3464	637.3365	6
6	70.0656	684.3535	P	557.3054	539.2936	540.2818	5
7	30.0343	741.3718	G	460.2572	442.2408	443.2242	4
8	101.0711	869.4296	Q	403.2381	385.2235	386.2061	3
9	101.0714	997.4887	Q	275.1725	257.1617	258.1458	2
10	101.0711		K	147.1134	129.1028	130.0867	1

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	129.1030	157.1078	R				10
2	88.0398	272.1372	D	1003.5011	985.4899	986.4871	9
3	120.0815	419.2066	F	888.4658	870.4468	871.4304	8
4	44.0499	490.2441	A	741.3798	723.3784	724.3785	7
5	70.0656	587.2985	P	670.3568	652.3487	653.3309	6
6	86.0605	700.3432	P(+15.99)	573.3024	555.2885	556.2758	5
7	30.0343	757.3683	G	460.2556	442.2408	443.2261	4
8	101.0713	885.4265	Q	403.2399	385.2206	386.2054	3
9	101.0714	1013.4857	Q	275.1732	257.1624	258.1461	2
10	101.0713		K	147.1137	129.1030	130.0870	1

Figure 12 | Representative MS2 spectra of TMEM109 HyP42.

Representative MS2 spectra and corresponding ion tables of tryptic fragments containing the target site, P42, from the Geiger HeLa HCD dataset. A) and C) MS2 spectrum of the Non-oxidised peptide with corresponding fragment ion table. B) and D) MS2 spectrum of the oxidised peptide with corresponding fragment ion table. Both MS2 spectra convey confident peptide sequencing through the y- and b-ion series. The MS2 spectra have been annotated to indicate the y4 and y5 fragment ions (red dashed line, bold typeface), and the b6 fragment ion (blue dashed line, bold typeface). Manual inspection of the diagnostic fragment ions (y4 and y5; b5 and b6) enables confident PTM localisation of the +15.99 Da shift to P42. E) The number of Peptide Features containing TMEM109 HyP42 after filters were applied for different PTM localisation criteria.

The above examples, in combination with the attrition of collagen hydroxyproline sites demonstrated that false negatives were generated throughout the high stringency workflow. This highlights how the analysis of MS2 spectra alone has modest discriminatory power to distinguish prolyl hydroxylation from incorrect assignments. The application of orthogonal methods is therefore required to identify further hydroxyproline sites and to increase the confidence of enzymatic prolyl hydroxylation, and these are described in other chapters of this thesis. One orthogonal method applicable to the current dataset is to identify consensus sequences that might be genuine hydroxylation motifs or reveal a systematic effect on interpretation of the MS2 spectra.

3.3.5 Multiple sequence alignment analysis to identify possible hydroxylation consensus sequences

The aforementioned examples of manual MS2 spectra inspection support the hypothesis that there are certain physicochemical properties that result in inefficient fragmentation around the target site that can, in turn, negatively impact the security of PTM localisation. Additionally, peptide sequence composition can be detrimental to the AScore value on the basis of computational complexity for the proteomic software to process, which further reduces the confidence of correct hydroxyproline assignment. It is possible that hydroxyproline sites were overlooked on the basis of these factors, especially during manual inspection of MS2 spectra. The existence of a motif that was enriched in parallel with even a low confidence hydroxyproline site would raise the hypothesis that the motif might represent a consensus site of hydroxylation and hence increase confidence in hydroxylation occurring at those sites.

Therefore, multiple sequence alignment was performed at each stringency level to search for possible prolyl hydroxylation consensus motifs. Briefly, 11-mer non-collagen sequences centred on the reported hydroxyproline sites were aligned for each stringency level. Since hydroxyproline can occur at variable positions within tryptic peptides, the peptide sequences used in the motif analysis were derived from the protein sequence instead of tryptic peptides. The data presented here is not direct proof of novel prolyl hydroxylation consensus motifs but might aid discovery of hydroxyproline sites when the MS2 spectra are of sufficiently low quality to be accepted after manual inspection.

Peptides containing 'HyP-G', 'P-HyP' and 'K-X-X-X-HyP' 'sequences (where HyP indicates hydroxyproline and X represents any amino acid) were enriched in this analysis (Figure 13). 'HyP-G' sequences were enriched after manual MS2 inspection (Table 17), which suggests this is a *bona fide* hydroxylation consensus motif. One example of a 'P-HyP' site was also confirmed by manual inspection, which suggests this might be a novel prolyl hydroxylation consensus sequence.

No 'K-X-X-X-HyP' sequences were detected after the manual MS2 inspection. Assuming faithful trypsin cleavage of proteins with these peptide sequences, the diagnostic b-ions for the hydroxyproline site would be the b3 and b4 ions. It has been shown that b-ion intensity is generally suppressed from the b3 ion (Shao et al., 2014). It has also been highlighted for LMAN1 HyP378 that y-ion intensity at the N-terminus of tryptic peptides can be relatively low. Therefore, the physicochemical properties of 'K-X-X-X-HyP' sequences are likely to result in low fragmentation ion intensity of the diagnostic ions, which does not provide sufficient resolution to confidently localise the reported hydroxyproline site. In this way, 'K-X-X-X-HyP' sequences might be prone to incorrect assignment.

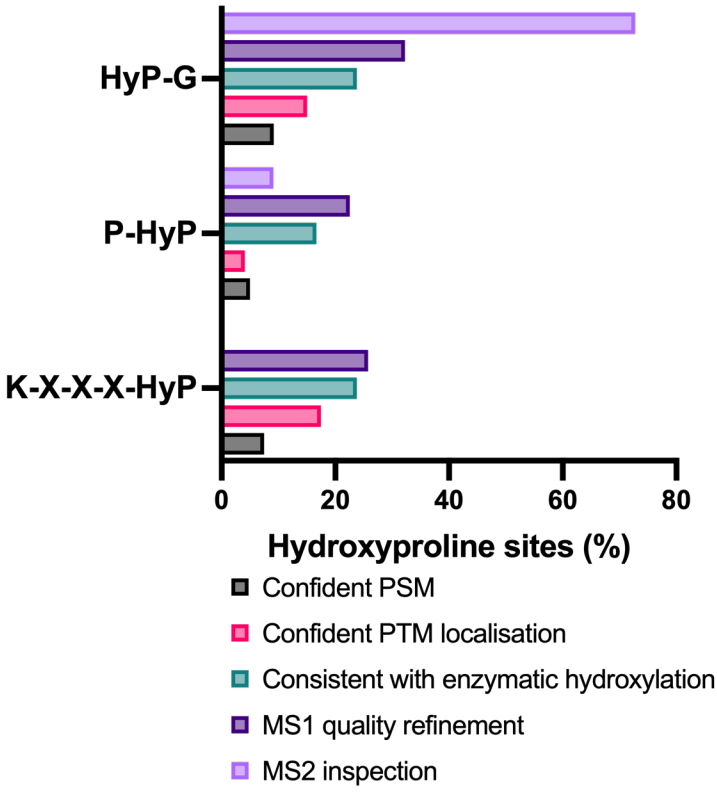


Figure 13 | Multiple sequence analysis identifies several motifs that are enriched with increased stringency.

11-mer sequences containing the reported hydroxyproline assignments as the central residue were analysed in MoMo software. Consensus sequences that are enriched as the stringency increases are presented. The proportion of total hydroxyproline assignments that were identified in the indicated sequence is displayed for each stringency level. HyP = hydroxyproline.

These results indicate that it is likely that other genuine sites of prolyl hydroxylation were disqualified and have not been identified in this analysis. As a reference, the 35 non-collagen hydroxyproline sites that were manually inspected are listed in Table 18. It is likely that hydroxyproline sites occurring in 'HyP-G' or 'P-HyP' sequences are correct, whilst those occurring in 'K-X-X-X-HyP' are not sites of enzymatic hydroxylation. Further analyses might provide more security that the reported hydroxylation site is a correct assignment by the proteomics software.

Table 18 | Motif annotation of the novel hydroxyproline sites considered for manual inspection.

35 non-collagen hydroxyproline sites were manually inspected at the MS2 level. The sites that were accepted are highlighted in red font. Each site is annotated with the relevant motif(s) identified using MoMo software.

Protein	Site	Motif(s)
B4GALT1	HyP159	-
BTN3A1	HyP447	-
CAPN2	HyP221	'P-HyP'
CAPN2	HyP222	'K-X-X-X-HyP'; 'P-HyP'
CASC4	HyP252	'HyP-G'
CLTC*	HyP1346	-
DOCK1	HyP1847	'K-X-X-X-HyP'; 'P-HyP'
DYNC1LI1	HyP180	'P-HyP'
FKBP10	HyP36	-
FLNB	HyP1429	'K-X-X-X-HyP'; 'HyP-G'
FLOT2	HyP374	'K-X-X-X-HyP'
HSPH1	HyP701	'K-X-X-X-HyP'
HYOU1	HyP977	'HyP-G'
KDM2A	HyP943	K-X-X-X-HyP'
LMAN1	HyP378	'HyP-G'
MDC1	HyP424	'K-X-X-X-HyP'

MUC5B	HyP2516	'HyP-G'
MYO10	HyP255	'HyP-G'
NUP214	HyP1120	-
OS9	HyP654	'HyP-G'
PARP2	HyP319	'P-HyP'
PDIA3**	HyP404	-
PRKCSH	HyP290	-
RYDEN	HyP235	-
SCYL2	HyP442	'P-HyP'
SDF2L1	HyP192	-
SERPINH1	HyP30	'HyP-G'
SLC38A10	HyP532	'HyP-G'
SUN2	HyP315	'HyP-G'
TMEM109	HyP42	'P-Hyp'; 'HyP-G'
TNS3	HyP565	'K-X-X-X-HyP'
TUBB3	HyP272	-

*The tryptic peptide containing CLTC HyP1346 is identical to that of CLTC1 HyP1346. **The tryptic peptide containing PDIA3 HyP404 is identical to that of PDIA4 HyP553.

3.3.6 Further analysis identifies hydroxyproline in PxCxxC sites across the proteome

The high stringency workflow identified PDIA3 HyP404 but a large number of Peptide Features for this site were disqualified. Subsequent inspection highlighted that the majority of peptides harbouring HyP404 exhibited low AScores that prevented more Peptide Features from being accepted through the high stringency workflow. The hydroxyproline site is located in the -2 position from the active site of the thioredoxin-like domain, which raised the question of whether this represents a conserved hydroxylation motif. Since many of the PDIA3 Hyp404 peptides were disqualified during AScore filtering, it was postulated that this might occur in the processing of peptides representing other thioredoxin-like domains so that they were unlikely to generate a motif that was enriched with stringency as in the previous multiple sequence alignment analysis, but nevertheless might represent genuine sites of prolyl hydroxylation.

To investigate this hypothesis a further search of the same tryptic data was performed to investigate hydroxylation of all 34 thioredoxin-like domain active sites on 23 reported

PDI proteins (Tannous et al., 2015). This analysis was similar to that performed on the reported non-HIFα PHD substrates (Section 3.3.2) and used the -2 position as the 'target site'.

The results are provided in Table 19 and show two characteristics. First, hydroxylation is only observed in the -2 position when a proline occupies this site. Second, hydroxylation only occurs in proximity to active sites in which the active site sequence is 'C-X-X-C'. For hydroxyproline assignments, the majority of peptides typically pass the 5 % ion intensity threshold, which contrasts with the assigned tryptophan oxidations that would otherwise be a cause for concern. None of the peptides passed the AScore threshold, except for PDIA3 HyP404 (4 Peptide Features) and TMX4 HyP62 (1 Peptide Feature). This supports the argument that the sequence complexity of the PDIA3 HyP404 peptide, and by extension all PDI hydroxyproline sites, negatively impacts its identification in the high stringency workflow, and that other hydroxyproline-containing peptides would be excluded for the same reasons.

Table 19 | Hydroxyproline is present at 16 unique sites on 8 PDI proteins in the tryptic proteome.

The tryptic datasets were interrogated for all peptides containing thioredoxin-like domain active sites. Thioredoxin-like domain active sites were identified for 20 PDI proteins. PSMs were filtered according to stringency of PTM localisation and the number of assigned Peptide Features (PFs) per stringency filter is reported. Where oxidation is assigned, the target site (-2 site) and the adjacent residue (typically tryptophan) are detailed. All other sites of oxidation are summarised as 'Other'. Stringency filters: -10lgP ≥ 20 (1); '1' and AScore ≥ 20 (2a); '1' and Ion Intensity ≥ 5 % (2b); '2a' and '2b' combined (3).

Protein	Domain	Active site	-2 site	Assigned	PFs per stringency filter			
					1	2a	2b	3
AGR2	a	CPHS	D79	Non-oxidised	64	n/a	n/a	n/a
				Other	42	3	24	0
AGR3	a	CQYS	E69	Non-oxidised	1	n/a	n/a	n/a
DNAJC10	a'	CSHC	P156	Target peptides not detected				
	a''	CPPC	P478	Target peptides not detected				
	a'''	CHPC	P586	Target peptides not detected				
	a''''	CGPC	P698	Target peptides not detected				
ERP44	a	CRFS	D56	Target peptides not detected				
MAGT1	a	CVVC	R85	Target peptides not detected				
P4HB	a	CGHC	P51	Non-oxidised	43	n/a	n/a	n/a
				P51	28	0	13	0
				W52	6	0	0	0

	a'	CGHC	P395	Non-oxidised	96	n/a	n/a	n/a
				P395	139	0	125	0
				W396	22	0	1	0
				Other	5	0	0	0
PDIA2	a	CGHC	P68	Target peptides not detected				
	a'	CTHC	P416	Target peptides not detected				
PDIA3	a	CGHC	P55	Non-oxidised	42	n/a	n/a	n/a
				P55	2	0	0	0
				W56	1	0	0	0
				Other	5	1	4	0
	a'	CGHC	P404*	Non-oxidised	156	n/a	n/a	n/a
				P404	133	4	90	4
				W405	36	6	0	0
				Other	7	0	0	0
PDIA4	a	CGHC	P89	Non-oxidised	22	n/a	n/a	n/a
				P89	84	0	75	0
				W90	7	0	0	0
				Other	1	0	0	0
	a'	CGHC	P204	Non-oxidised	4	n/a	n/a	n/a
				P204	3	0	1	0
	a''	CGHC	P553*	Non-oxidised	156	n/a	n/a	n/a
				P404	133	4	90	4
				W405	36	6	0	0
				Other	7	0	0	0
PDIA5	a	CSMC	P180	Non-oxidised	1	n/a	n/a	n/a
	a'	CGHC	P303	P303	1	0	0	0
	a''	CPHC	P424	P424	4	0	4	0
				W425	1	0	0	0
PDIA6	a	CGHC	P53	P53	3	0	0	0
				W54	1	0	0	0
	a'	CGHC	P188	Non-oxidised	42	n/a	n/a	n/a
				P188	13	0	9	0
				W189	13	0	0	0
				Other	1	0	0	0
PDILT	a	SKQS	P70	Target peptides not detected				
	a'	SKKC	P415	Target peptides not detected				
TMX1	a	CPAC	P54	Target peptides not detected				
TMX2	a	SNDC	N165	Target peptides not detected				
TMX3	a	CGHC	P51	Non-oxidised	1	n/a	n/a	n/a
				P51	10	0	10	0

				W52	1	0	0	0
				Other	1	0	0	0
TMX4	a	CPSC	P62	P62	2	1	0	0
				W63	3	0	0	0
TUSC3	a	CSVC	R97	Target peptides not detected				
TXNDC5	a	CGHC	P87	P87	1	0	0	0
	a'	CGHC	P215	Non-oxidised	44	n/a	n/a	n/a
				P215	8	0	6	0
	a''	CGHC	P348	Non-oxidised	52	n/a	n/a	n/a
				P348	13	0	13	0
				W349	3	2	2	2
TXNDC12	a	CGAC	S64	Non-oxidised	11	n/a	n/a	n/a
TXNDC15	a	CRFS	P218	Other	1	0	0	0

*The amino acid sequence of tryptic peptides containing PDIA3 P404 and PDIA4 P553 are identical.

There are many hydroxyproline sites assigned by PEAKS® in proximity to the active sites of thioredoxin-like domains. As predicted, very few of these sites passed the AScore filter, which most likely resulted from the abundance of residues prone to oxidation in the tryptic peptides. Despite this, many of the reported hydroxyproline sites pass the ion intensity threshold. This occurs much more frequently for hydroxyproline than oxidised tryptophan assignments, and for this reason the sites of PDI prolyl hydroxylation contrast with those of the reported non-HIFα substrates. The data strongly suggested that certain PDIs would harbour a consensus sequence for hydroxylation on proline.

To further investigate the primary sequence determinants for hydroxylation, multiple sequence alignment was performed on non-tryptic amino acid sequences that corresponded to the hydroxyproline-containing peptides. A 15-residue flanking sequence either side of the central hydroxyproline assignment was used. There is a common primary sequence shared between all hydroxylated peptides identified here (F-X-A-HyP-W-C-X-H-C-X-X-X-X-P) (Figure 14). This suggests there may be additional PDI hydroxylation sites in DNAJC10 and PDIA2 proteins, which share most of these features. Overall, this analysis identifies more sites of putative PDI prolyl hydroxylation than the high stringency workflow and suggests there may be additional sequences susceptible to hydroxylation that are currently beyond the limits of detection.

Figure 14 | Multiple sequence alignment identifies putative PDI hydroxylation consensus sequence.

Multiple sequence alignment was performed with 30-mer sequences of the PDI proteins identified in the tryptic datasets containing hydroxyproline in proximity to the thioredoxin-like domain active sites. Amino acids are coloured according to their chemical properties: Green (polar residues CGNQSTY), blue (basic residues HKR), red (acidic residues DE), and black (hydrophobic residues AFILMPVW).

3.4 Discussion

The work in this chapter aimed to determine the existence or otherwise of prolyl hydroxylation in publicly deposited deep proteome datasets. The analysis focused first on a series of reported non-HIFα substrates for one or more of the HIFα prolyl hydroxylases then secondly on proteome-wide assignment of hydroxyproline.

3.4.1 Protein oxidation is highly abundant but frequently assigned with low confidence throughout the human proteome

To guide the work a survey was first made of all sites of protein oxidation that have been defined – including on amino acid residues reported to support both enzyme-catalysed and non-enzymatic oxidation. This showed that oxidation was very prevalent but that the large majority was on residues susceptible to artefactual oxidations. The contrast was even greater when restricted to high confidence oxidation assignments. This indicates that for residues such as proline the confidence with which protein oxidation can be localized to that residue is of critical importance to corroborating the assignment.

3.4.2 Absence of *in cellulo* data supporting the assignment of reported non-HIFα PHD substrates

With that in mind PTM localisation filters were applied to examine the security of hydroxyproline assignments of 45 reported non-HIFα PHD1-3 substrates. No high confidence hydroxyproline sites were observed even though artefactual oxidation of residues in proximity to the target site passed the same filters. Given the complete absence of evidence of hydroxylase activity in assays with recombinant PHD (Cockman *et al.*, 2019), it seems likely that non-HIFα PHD substrates have been overreported. Hydroxyproline-containing peptides were not detected for 72 of the 81 target sites. These results were most likely due to the absence of prolyl hydroxylation of the protein or low protein abundance, which could be attributed to low expression levels or instability of the hydroxylated protein. The data suggests that non-HIFα PHD1-3 substrates are likely to be relatively uncommon and that there is a requirement for robust analysis in order to assign such enzymatic activity. Therefore, orthogonal data that demonstrates enzyme- and residue-dependence of the proposed oxidation would provide further confidence of the enzymatic hydroxylation assignment. Nevertheless, there might be other sites of PHD-catalysed prolyl hydroxylation, or prolyl hydroxylation catalysed by other enzymes, in the proteome.

3.4.3 Evaluation of the high stringency workflow employed to identify highly secure hydroxyproline sites

In light of the above, a high stringency approach was undertaken to identify novel sites of hydroxyproline in the same proteomic datasets. This strategy reduced the number of discrete hydroxyproline sites from 10,733 to 59, which were manually inspected. Of these, 30 hydroxyproline sites met the evaluation criteria and were therefore considered to exhibit robust evidence in support of the hydroxyproline assignment. Hydroxyproline was identified for 19 sites in collagen proteins, which confirms this workflow can identify *bona fide* prolyl hydroxylase substrates. Therefore, 11 novel hydroxyproline sites were identified in this analysis, 8 of which exhibit a common ‘HyP-G’ motif.

The reduction of 10,733 nominally assigned hydroxyproline sites to 11 suggests false positives are highly prevalent in proteomic datasets. The most likely explanation of these

results is misassigned isobaric oxidation. There may also be instances of misassigned PSMs because some peptides assigned with hydroxyproline exhibited a positive ΔRT, which is inconsistent with oxidation of any residue. Further analysis of the data also indicated false negatives. For example, collagen hydroxyproline sites are underrepresented at the highest confidence level and numerous Peptide Features containing the 11 novel hydroxyproline sites were also disqualified throughout the high stringency workflow.

The data was analysed to indicate if there may be particular peptide sequences for which the rate of corroborating hydroxyproline assignments by manual MS2 inspection is higher than others. Peptides containing 'HyP-G' were enriched throughout the high stringency workflow and comprised a high proportion of the hydroxyproline sites accepted after manual MS2 inspection. TMEM109 HyP42, an example of a 'P-HyP' sequence, was accepted after manual MS2 inspection, which confirmed this is a plausible hydroxylation consensus sequence. On the other hand, secure PTM localisation was not observed for any of the hydroxyproline assignments in 'K-X-X-X-HyP' sequences. Therefore, 'K-X-X-X-HyP' sequences were not pursued further in this thesis.

PDIA3 HyP404 was identified with high confidence and it was recognised that this site is in close proximity to the active site of a thioredoxin-like domain. Given other hydroxylation consensus sequences were identified and PDIA3 HyP404 is in proximity to the active site of a conserved domain in PDI proteins, it was hypothesised that other PDI proteins were subject to prolyl hydroxylation but were censored according to low AScore. Tryptic peptides containing PDIA3 HyP404 are identical to PDIA4 HyP553, and 14 additional hydroxyproline sites were identified on 8 total PDI proteins. Despite these peptides containing many residues prone to artefactual oxidation there was little evidence for this in the tryptic proteome. The hydroxyproline sites shared a common sequence ('F-X-A-HyP-W-C-X-H-C-X-X-X-X-P'), which further supports the hypothesis of enzymatic hydroxylation of the PDI proteins. Together the PDI hydroxyproline sites constitute 439 hydroxyproline-containing Peptide Features, only 4 of which surpassed all criteria in the high stringency workflow. Reducing the number of residues to which variable oxidation can be assigned is likely to increase the AScore values of the PDI hydroxyproline sites and generate more true positives, but this approach would also increase the number of false positives across the total proteome.

Not all *bona fide* hydroxyproline sites were identified in this analysis. COL14A1 HyP1643 (Brown et al., 1994) and FGA HyP565 (Ono *et al.*, 2009) displayed convincing evidence in support of the hydroxyproline assignments at the MS2 level (data not shown). However, they were disqualified in the high stringency workflow on the basis of low frequency (i.e., they were identified in ≤ 3 of the 39 datasets). This is most likely a consequence of secreted proteins displaying limited identification in the cellular proteome. Additionally, cell line-specific expression profiles can also limit the discovery of novel hydroxyproline sites. MUC5B HyP2516 was one of the 11 highly secure hydroxyproline sites but it was only identified in 4 datasets, which corresponded to A549 and MCF7 cells, and this is a consequence of a specific gene expression profile. It is probable other true positives have not been discovered since i) only 13 cell lines were analysed, and ii) identification of secreted proteins from these cells is likely to be difficult to reproduce without specific targeting or enrichment due to the method in which cells were harvested.

The distinction between false positives and genuine hydroxyproline assignments in LC-MS/MS data is difficult to determine and a trade-off between the two is expected. The approach taken here aimed to reduce the number of false positives as far as possible, which inadvertently increased the number of false negatives. Orthogonal methods that consider regulation of prolyl hydroxylation, or direct interaction with hydroxyproline, would be beneficial to consider additional criteria during analysis of LC-MS/MS data.

Oxidation of residues prone to artefactual oxidation (MWY) are identified with greater confidence than oxidation of other residues, including enzymatic prolyl hydroxylation. Consistently, the reporting of non-HIFα PHD substrates is most likely overrepresented because of misassigned artefactual oxidation of the target peptides.

This analysis identified 11 non-collagen hydroxyproline sites and these findings are largely consistent with previously published reports of hydroxyproline in the proteome (Table 62) (Arsenault et al., 2015; Henningsen et al., 2010; Stoehr et al., 2016; Zhou et al., 2016). During the course of this work a further report was published that also aimed to identify hydroxyproline in the steady state proteomes of publicly deposited data and there is significant overlap between their findings and those presented here (Onisko,

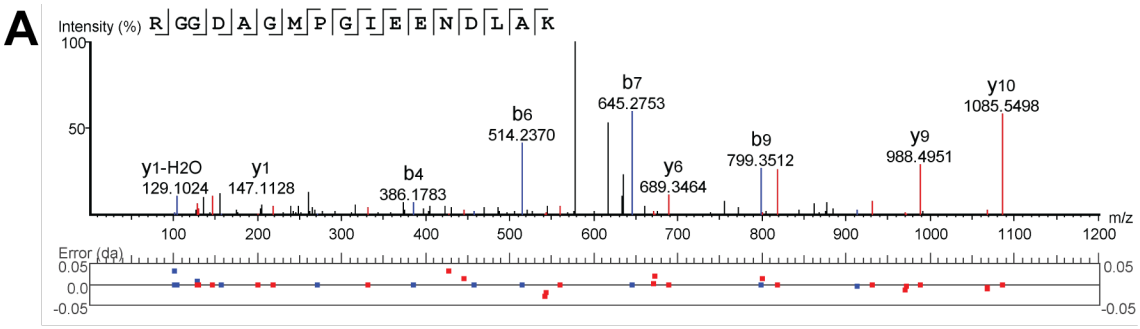
2020). There is also convincing data to suggest prolyl hydroxylation occurs in other consensus sequences, as previously described.

The majority of the hydroxyproline sites presented in this chapter occur on either ER resident proteins or proteins that are processed in the secretory pathway. It is therefore likely that ER resident prolyl hydroxylases catalyse these PTMs. The hydroxyproline sites of some of the candidates (e.g., SERPINH1, FKBP10 and TMEM109) occur at the N-terminus of the protein, in proximity to the signal peptide cleavage site. This raises the question of co-translational hydroxylation.

An important point to emphasise is that the analysis performed here relies on the quality of the MS2 spectrum. Peptide sequence appears to impart a great influence on the likelihood of artefactual oxidation and fragmentation of the peptide, both of which impact PTM localisation confidence scores and the ability to confidently localise the hydroxyproline assignment during manual MS2 inspection. Assessing the security of hydroxyproline assignments lacking a non-hydroxylated counterpart presents an even greater challenge when there is no comparator to evaluate the quality of the MS1 and MS2 spectra.

The factors outlined in this chapter highlight many difficulties associated with PTM localisation and demonstrate how certain criteria used to eliminate false positive hydroxyproline assignment will also generate false negatives. Overall, the high stringency of this approach to discover novel hydroxyproline sites was considered suitable because it generated assignments consistent with 2OGD-catalysed hydroxylation and limited the number of false positives. Additional security will likely be derived from further proteomic studies that consider orthogonal approaches to discover novel hydroxyproline sites, which have been reported in the subsequent chapters of this thesis.

3.5 Appendix



C

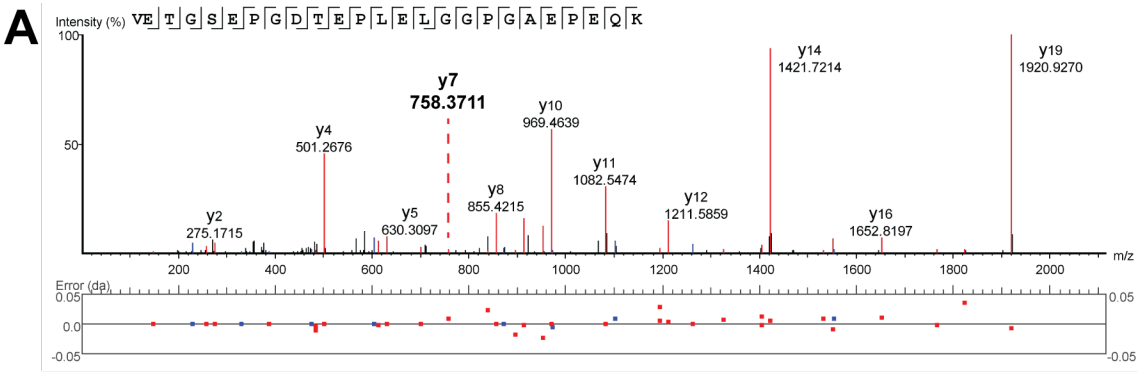
#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	129.1022	157.1336	R				17
2	30.0343	214.1304	G	1589.7111	1571.7006	1572.6841	16
3	30.0343	271.1519	G	1532.6897	1514.6792	1515.6627	15
4	88.0398	386.1766	D	1475.6682	1457.6577	1458.6412	14
5	44.0499	457.2164	A	1360.6412	1342.6307	1343.6143	13
6	30.0343	514.2371	G	1289.6041	1271.5936	1272.5771	12
7	104.0532	645.2754	M	1232.5826	1214.5721	1215.5557	11
8	86.0605	758.3255	P(+15.99)	1101.5415	1083.5312	1084.5273	10
9	30.0343	815.3461	G	988.4944	970.4783	971.4711	9
10	86.0969	928.4423	I	931.4720	913.4625	914.4460	8
11	102.0553	1057.4751	E	818.3868	800.3741	801.3620	7
12	102.0554	1186.5162	E	689.3433	671.3361	672.2861	6
13	87.0558	1300.5591	N	560.3020	542.2506	543.2769	5
14	88.0398	1415.5861	D	446.2589	428.2503	429.2339	4
15	86.0969	1528.6702	L	331.2335	313.2234	314.2070	3
16	44.0499	1599.7073	A	218.1503	200.1389	201.1244	2
17	101.0707		K	147.1126	129.1022	130.0862	1

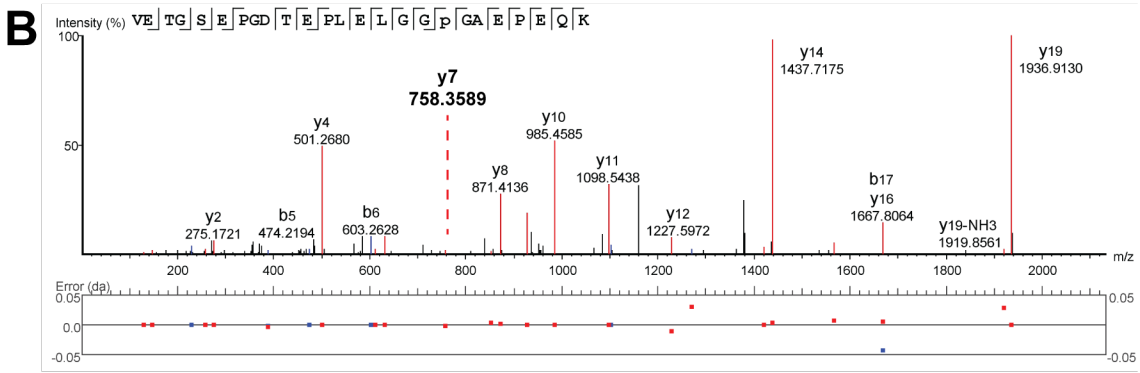
D

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	129.1024	157.1095	R				17
2	30.0343	214.1304	G	1573.7162	1555.7057	1556.6892	16
3	30.0343	271.1490	G	1516.6947	1498.6842	1499.6677	15
4	88.0398	386.1783	D	1459.6733	1441.6628	1442.6464	14
5	44.0499	457.2133	A	1344.6464	1326.6359	1327.6194	13
6	30.0343	514.2370	G	1273.6093	1255.5988	1256.5823	12
7	104.0532	645.2753	M	1216.5878	1198.5773	1199.5608	11
8	70.0656	742.3306	P	1085.5498	1067.5477	1068.5281	10
9	30.0343	799.3512	G	988.4951	970.4963	971.4719	9
10	86.0969	912.4399	I	931.4702	913.4625	914.4460	8
11	102.0552	1041.4788	E	818.3905	800.3627	801.3620	7
12	102.0554	1170.5212	E	689.3464	671.3312	672.2957	6
13	87.0558	1284.5642	N	560.3035	542.3221	543.2951	5
14	88.0398	1399.5912	D	446.2451	428.2160	429.2339	4
15	86.0969	1512.6753	L	331.2344	313.2234	314.2070	3
16	44.0499	1583.7124	A	218.1502	200.1399	201.1231	2
17	101.0721		K	147.1128	129.1024	130.0861	1

Figure 15 | Representative MS2 spectra of CASC4 HyP252.

Representative MS2 spectra and corresponding ion tables of tryptic fragments containing the target site, P252, from the Bekker Jensen HCT116 replicate 1 dataset. A) and C) MS2 spectrum of the Non-oxidised peptide with corresponding fragment ion table. B) and D) MS2 spectrum of the oxidised peptide with corresponding fragment ion table. Both MS2 spectra convey confident peptide sequencing through the y-ion series. Manual inspection of the diagnostic fragment ions (y9 and y10) enables confident PTM localisation of the +15.99 Da shift to P252.





C

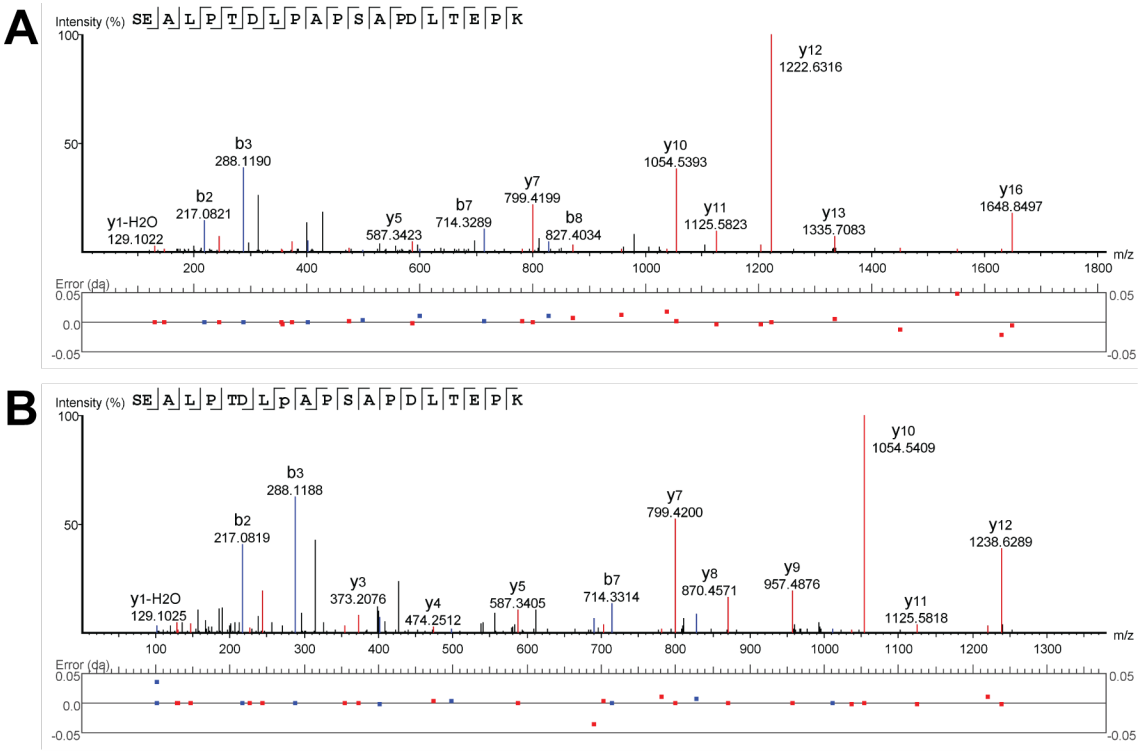
#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	72.0812	100.0762	V				25
2	102.0554	229.1192	E	2440.0996	2422.0891	2423.0725	24
3	74.0605	330.1665	T	2311.0571	2293.0466	2294.0300	23
4	30.0343	387.1911	G	2210.0095	2191.9990	2192.9824	22
5	60.0449	474.2194	S	2152.9880	2134.9775	2135.9609	21
6	102.0554	603.2628	E	2065.9561	2047.9453	2048.9290	20
7	70.0656	700.3153	P	1936.9130	1918.9028	1919.8561	19
8	30.0343	757.3368	G	1839.8606	1821.8501	1822.8336	18
9	88.0398	872.3637	D	1782.8391	1764.8286	1765.8121	17
10	74.0605	973.4114	T	1667.8064	1649.8016	1650.7852	16
11	102.0554	1102.4546	E	1566.7562	1548.7540	1549.7375	15
12	70.0656	1199.5067	P	1437.7175	1419.7125	1420.6949	14
13	86.0969	1312.5908	L	1340.6692	1322.6587	1323.6422	13
14	102.0554	1441.6334	E	1227.5972	1209.5746	1210.5581	12
15	86.0969	1554.7175	L	1098.5438	1080.5320	1081.5155	11
16	30.0343	1611.7389	G	985.4585	967.4479	968.4315	10
17	30.0343	1668.8049	G	928.4374	910.4265	911.4100	9
18	86.0605	1781.8081	P(+15.99)	871.4136	853.4005	854.3885	8
19	30.0343	1838.8295	G	758.3711	740.3573	741.3409	7
20	44.0499	1909.8666	A	701.3464	683.3358	684.3194	6
21	102.0554	2038.9092	E	630.3085	612.2988	613.2823	5
22	70.0656	2135.9619	P	501.2680	483.2561	484.2397	4
23	102.0554	2265.0046	E	404.2139	386.2034	387.1911	3
24	101.0714	2393.0632	Q	275.1721	257.1611	258.1458	2
25	101.1078		K	147.1122	129.1020	130.0858	1

D

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	72.0812	100.0762	V				25
2	102.0554	229.1190	E	2424.1047	2406.0942	2407.0776	24
3	74.0605	330.1669	T	2295.0623	2277.0518	2278.0352	23
4	30.0343	387.1881	G	2194.0144	2176.0039	2176.9873	22
5	60.0449	474.2185	S	2136.9932	2118.9827	2119.9661	21
6	102.0554	603.2633	E	2049.9612	2031.9506	2032.9341	20
7	70.0656	700.3153	P	1920.9270	1902.9080	1903.8915	19
8	30.0343	757.3368	G	1823.8298	1805.8552	1806.8387	18
9	88.0398	872.3633	D	1766.8467	1748.8337	1749.8173	17
10	74.0605	973.4181	T	1651.8058	1633.8068	1634.7903	16
11	102.0554	1102.4436	E	1550.7795	1532.7494	1533.7426	15
12	70.0656	1199.5067	P	1421.7214	1403.7186	1404.6863	14
13	86.0969	1312.5908	L	1324.6655	1306.6637	1307.6472	13
14	102.0554	1441.6334	E	1211.5859	1193.5496	1194.5562	12
15	86.0969	1554.7073	L	1082.5474	1064.5371	1065.5206	11
16	30.0343	1611.7389	G	969.4639	951.4530	952.4615	10
17	30.0343	1668.7604	G	912.4445	894.4315	895.4335	9
18	70.0656	1765.8131	P	855.4215	837.4100	838.3687	8
19	30.0343	1822.8346	G	758.3589	740.3573	741.3409	7
20	44.0499	1893.8717	A	701.3456	683.3358	684.3194	6
21	102.0554	2022.9143	E	630.3097	612.3007	613.2823	5
22	70.0656	2119.9670	P	501.2676	483.2604	484.2505	4
23	102.0554	2249.0098	E	404.2139	386.2034	387.1881	3
24	101.0714	2377.0681	Q	275.1715	257.1617	258.1450	2
25	101.1078		K	147.1129	129.1022	130.0858	1

Figure 16 | Representative MS2 spectra of HYOU1 HyP977.

Representative MS2 spectra and corresponding ion tables of tryptic fragments containing the target site, P977, from the Bekker Jensen HEK293 replicate 2 dataset. A) and C) MS2 spectrum of the Non-oxidised peptide with corresponding fragment ion table. B) and D) MS2 spectrum of the oxidised peptide with corresponding fragment ion table. The MS2 spectra have been annotated to indicate the y7 fragment ions (red dashed line, bold typeface). Both MS2 spectra convey confident peptide sequencing through the y-ion series. Manual inspection of the diagnostic fragment ions (y7 and y8) enables confident PTM localisation of the +15.99 Da shift to P977.



C

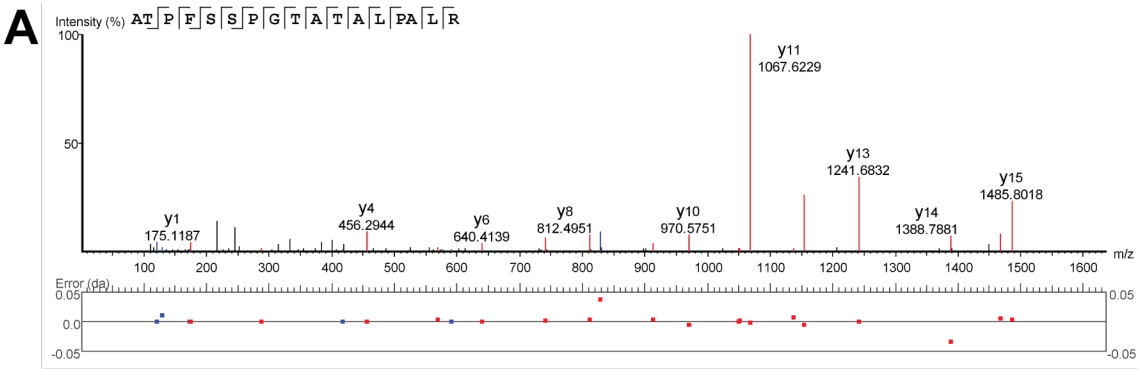
#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	60.0449	88.0399	S				20
2	102.0554	217.0821	E	1962.0065	1943.9960	1944.9795	19
3	44.0499	288.1190	A	1832.9640	1814.9535	1815.9370	18
4	86.0969	401.2024	L	1761.9269	1743.9164	1744.8999	17
5	70.0656	498.2527	P	1648.8497	1630.8544	1631.8158	16
6	74.0605	599.2924	T	1551.7417	1533.7795	1534.7631	15
7	88.0398	714.3289	D	1450.7550	1432.7318	1433.7153	14
8	86.0969	827.4034	L	1335.7083	1317.7048	1318.6884	13
9	70.0656	924.4678	P	1222.6316	1204.6257	1205.6044	12
10	44.0499	995.5049	A	1125.5823	1107.5681	1108.5516	11
11	70.0656	1092.5576	P	1054.5393	1036.5123	1037.5145	10
12	60.0449	1179.5897	S	957.4756	939.4781	940.4617	9
13	44.0499	1250.6268	A	870.4487	852.4461	853.4297	8
14	70.0656	1347.6796	P	799.4199	781.4055	782.3926	7
15	88.0398	1462.7065	D	702.3668	684.3563	685.3398	6
16	86.0969	1575.7905	L	587.3423	569.3293	570.3129	5
17	74.0605	1676.8383	T	474.2532	456.2452	457.2288	4
18	102.0554	1805.8809	E	373.2082	355.1975	356.1848	3
19	70.0656	1902.9336	P	244.1655	226.1550	227.1385	2
20	101.1078		K	147.1126	129.1022	130.0858	1

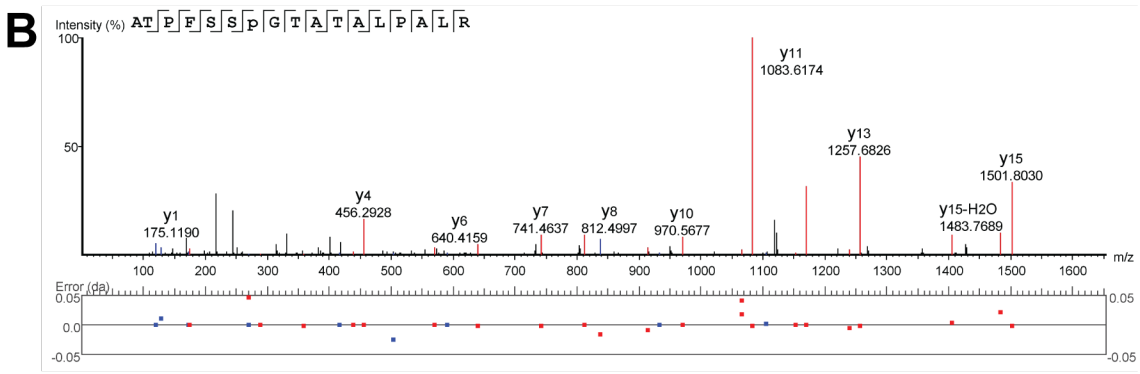
D

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	60.0449	88.0399	S				20
2	102.0554	217.0819	E	1978.0015	1959.9910	1960.9745	19
3	44.0499	288.1188	A	1848.9589	1830.9484	1831.9319	18
4	86.0969	401.2057	L	1777.9218	1759.9113	1760.8948	17
5	70.0656	498.2519	P	1664.8376	1646.8271	1647.8107	16
6	74.0605	599.3040	T	1567.7849	1549.7744	1550.7579	15
7	88.0398	714.3314	D	1466.7372	1448.7267	1449.7102	14
8	86.0969	827.4062	L	1351.7103	1333.6998	1334.6833	13
9	86.0605	940.4627	P(+15.99)	1238.6289	1220.6044	1221.5992	12
10	44.0499	1011.4985	A	1125.5818	1107.5681	1108.5516	11
11	70.0656	1108.5526	P	1054.5409	1036.5345	1037.5145	10
12	60.0449	1195.5846	S	957.4876	939.4781	940.4617	9
13	44.0499	1266.6217	A	870.4571	852.4461	853.4297	8
14	70.0656	1363.6744	P	799.4200	781.3978	782.3926	7
15	88.0398	1478.7014	D	702.3628	684.3563	685.3398	6
16	86.0969	1591.7855	L	587.3405	569.3293	570.3129	5
17	74.0605	1692.8331	T	474.2512	456.2452	457.2288	4
18	102.0554	1821.8757	E	373.2076	355.1967	356.1812	3
19	70.0656	1918.9285	P	244.1655	226.1549	227.1385	2
20	101.0706		K	147.1126	129.1025	130.0865	1

Figure 17 | Representative MS2 spectra of PRKCSH HyP290.

Representative MS2 spectra and corresponding ion tables of tryptic fragments containing the target site, P290, from the BJ MCF7 replicate 1 dataset. A) and C) MS2 spectrum of the Non-oxidised peptide with corresponding fragment ion table. B) and D) MS2 spectrum of the oxidised peptide with corresponding fragment ion table. Both MS2 spectra convey confident peptide sequencing through the y-ion series. Manual inspection of the diagnostic fragment ions (y11 and y12) enables confident PTM localisation of the +15.99 Da shift to P290.





C

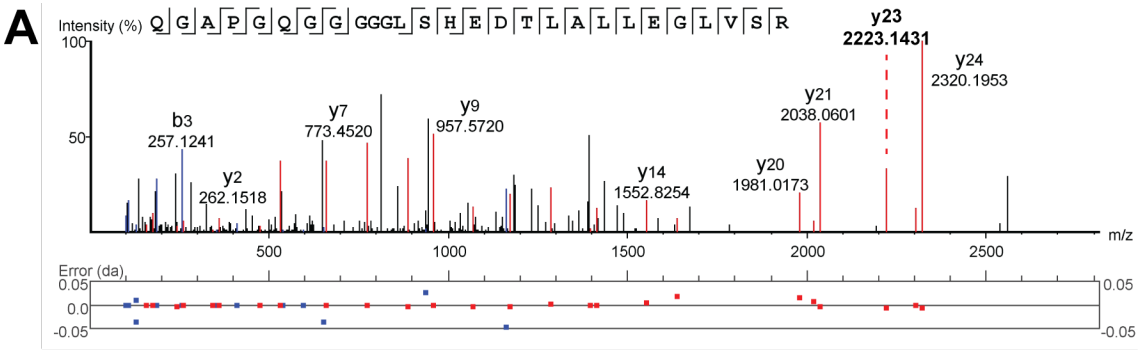
#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	44.0499	72.0449	A				17
2	74.0605	173.0927	T	1586.8536	1568.8431	1569.8267	16
3	70.0656	270.1454	P	1485.8018	1467.7883	1468.7789	15
4	120.0810	417.2146	F	1388.7881	1370.7427	1371.7262	14
5	60.0449	504.2458	S	1241.6832	1223.6743	1224.6578	13
6	60.0449	591.2769	S	1154.6592	1136.6349	1137.6257	12
7	70.0656	688.3306	P	1067.6229	1049.6096	1050.5903	11
8	30.0343	745.3521	G	970.5751	952.5574	953.5410	10
9	74.0605	846.3997	T	913.5429	895.5359	896.5195	9
10	44.0499	917.4368	A	812.4951	794.4883	795.4719	8
11	74.0605	1018.4845	T	741.4599	723.4512	724.4348	7
12	44.0499	1089.5216	A	640.4139	622.4034	623.3870	6
13	86.0969	1202.6057	L	569.3722	551.3663	552.3499	5
14	70.0656	1299.6584	P	456.2944	438.2823	439.2659	4
15	44.0499	1370.6956	A	359.2401	341.2296	342.2131	3
16	86.0969	1483.7797	L	288.2024	270.1924	271.1760	2
17	129.1024		R	175.1187	157.1084	158.0919	1

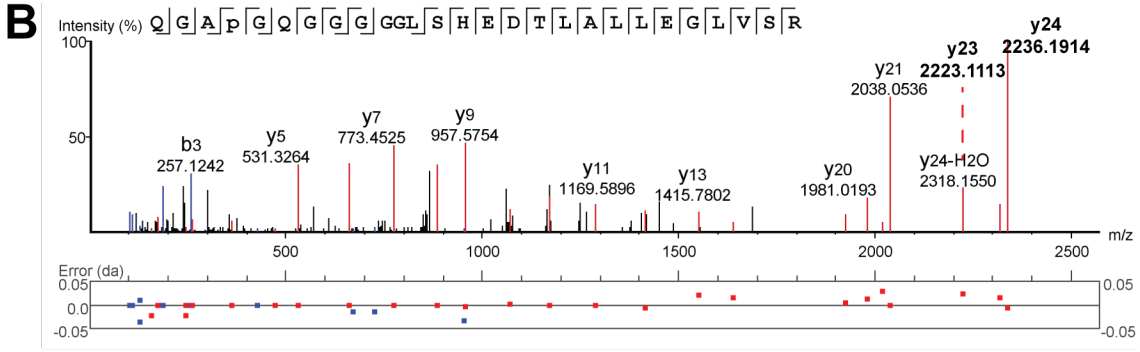
D

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	44.0499	72.0449	A				17
2	74.0605	173.0931	T	1602.8485	1584.8380	1585.8215	16
3	70.0656	270.1452	P	1501.8030	1483.7689	1484.7739	15
4	120.0810	417.2140	F	1404.7430	1386.7375	1387.7211	14
5	60.0449	504.2722	S	1257.6826	1239.6760	1240.6527	13
6	60.0449	591.2775	S	1170.6475	1152.6357	1153.6207	12
7	86.0605	704.3255	P(+15.99)	1083.6174	1065.5867	1066.5469	11
8	30.0343	761.3469	G	970.5677	952.5574	953.5410	10
9	74.0605	862.3947	T	913.5563	895.5359	896.5195	9
10	44.0499	933.4324	A	812.4997	794.4883	795.4719	8
11	74.0605	1034.4795	T	741.4637	723.4512	724.4348	7
12	44.0499	1105.5134	A	640.4159	622.4034	623.3870	6
13	86.0969	1218.6006	L	569.3760	551.3663	552.3499	5
14	70.0656	1315.6533	P	456.2928	438.2823	439.2671	4
15	44.0499	1386.6904	A	359.2427	341.2296	342.2131	3
16	86.0969	1499.7745	L	288.2028	270.1452	271.1760	2
17	129.1025		R	175.1190	157.1084	158.0919	1

Figure 18 | Representative MS2 spectra of MUC5B HyP2516.

Representative MS2 spectra and corresponding ion tables of tryptic fragments containing the target site, P2516, from the BJ MCF7 replicate 2 dataset. A) and C) MS2 spectrum of the Non-oxidised peptide with corresponding fragment ion table. B) and D) MS2 spectrum of the oxidised peptide with corresponding fragment ion table. Both MS2 spectra convey confident peptide sequencing through the y-ion series. Manual inspection of the diagnostic fragment ions (y11 and y12) enables confident PTM localisation of the +15.99 Da shift to P2516.





C

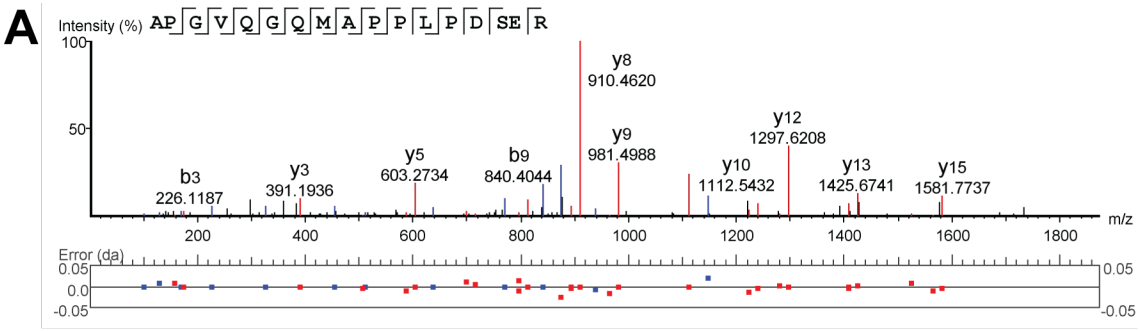
#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	101.0710	129.1027	Q				27
2	30.0343	186.0874	G	2448.2476	2430.2371	2431.2205	26
3	44.0499	257.1241	A	2391.2263	2373.2158	2374.1992	25
4	70.0656	354.1780	P	2320.1953	2302.1797	2303.1621	24
5	30.0343	411.1998	G	2223.1431	2205.1257	2206.1091	23
6	101.0714	539.2592	Q	2166.1150	2148.1045	2149.0879	22
7	30.0343	596.2810	G	2038.0601	2020.0374	2021.0293	21
8	30.0343	653.3367	G	1981.0173	1963.0243	1964.0078	20
9	30.0343	710.3221	G	1924.0134	1906.0029	1906.9865	19
10	30.0343	767.3436	G	1866.9919	1848.9814	1849.9650	18
11	30.0343	824.3651	G	1809.9705	1791.9600	1792.9435	17
12	86.0969	937.4219	L	1752.9490	1734.9385	1735.9220	16
13	60.0449	1024.4812	S	1639.8438	1621.8544	1622.8379	15
14	110.0716	1161.5881	H	1552.8254	1534.8224	1535.8059	14
15	102.0554	1290.5826	E	1415.7720	1397.7615	1398.7469	13
16	88.0398	1405.6096	D	1286.7283	1268.7209	1269.7045	12
17	74.0605	1506.6572	T	1171.7075	1153.6940	1154.6775	11
18	86.0969	1619.7413	L	1070.6598	1052.6462	1053.6298	10
19	44.0499	1690.7784	A	957.5720	939.5621	940.5457	9
20	86.0969	1803.8625	L	886.5394	868.5250	869.5086	8
21	86.0969	1916.9465	L	773.4520	755.4410	756.4246	7
22	102.0554	2045.9891	E	660.3698	642.3569	643.3405	6
23	30.0343	2103.0107	G	531.3253	513.3143	514.2979	5
24	86.0969	2216.0947	L	474.3049	456.2929	457.2765	4
25	72.0812	2315.1631	V	361.2195	343.2088	344.1950	3
26	60.0449	2402.1951	S	262.1518	244.1404	245.1291	2
27	129.1027		R	175.1192	157.1084	158.0924	1

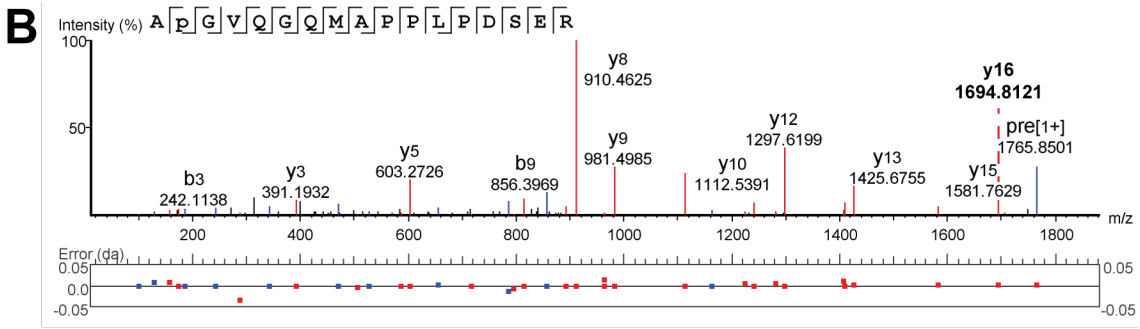
D

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	101.0710	129.1023	Q				27
2	30.0343	186.0879	G	2464.2427	2446.2322	2447.2156	26
3	44.0499	257.1242	A	2407.2212	2389.2107	2390.1941	25
4	86.0605	370.1726	P(+15.99)	2336.1914	2318.1550	2319.1570	24
5	30.0343	427.1938	G	2223.1113	2205.1257	2206.1091	23
6	101.0714	555.2527	Q	2166.1150	2148.1045	2149.0879	22
7	30.0343	612.2741	G	2038.0536	2020.0137	2021.0293	21
8	30.0343	669.3101	G	1981.0193	1963.0243	1964.0078	20
9	30.0343	726.3312	G	1924.0067	1906.0029	1906.9865	19
10	30.0343	783.3385	G	1866.9919	1848.9814	1849.9650	18
11	30.0343	840.3600	G	1809.9705	1791.9600	1792.9435	17
12	86.0969	953.4777	L	1752.9490	1734.9385	1735.9220	16
13	60.0449	1040.4761	S	1639.8479	1621.8544	1622.8379	15
14	110.0716	1177.5350	H	1552.8102	1534.8224	1535.8059	14
15	102.0554	1306.5775	E	1415.7802	1397.7634	1398.7469	13
16	88.0398	1421.6045	D	1286.7324	1268.7209	1269.7045	12
17	74.0605	1522.6522	T	1171.7046	1153.6940	1154.6775	11
18	86.0969	1635.7362	L	1070.6514	1052.6462	1053.6298	10
19	44.0499	1706.7733	A	957.5754	939.5621	940.5457	9
20	86.0969	1819.8574	L	886.5364	868.5250	869.5086	8
21	86.0969	1932.9415	L	773.4525	755.4410	756.4246	7
22	102.0554	2061.9841	E	660.3693	642.3569	643.3405	6
23	30.0343	2119.0056	G	531.3264	513.3143	514.2979	5
24	86.0969	2232.0896	L	474.3040	456.2929	457.2765	4
25	72.0812	2331.1580	V	361.2194	343.2088	344.1924	3
26	60.0449	2418.1899	S	262.1505	244.1646	245.1248	2
27	129.1023		R	175.1187	157.1325	158.0919	1

Figure 19 | Representative MS2 spectra of SUN2 HyP315.

Representative MS2 spectra and corresponding ion tables of tryptic fragments containing the target site, P315, from the BJ SH-SY5Y replicate 1 dataset. A) and C) MS2 spectrum of the Non-oxidised peptide with corresponding fragment ion table. B) and D) MS2 spectrum of the oxidised peptide with corresponding fragment ion table. Both MS2 spectra convey confident peptide sequencing through the y-ion series. The MS2 spectra have been annotated to indicate the y23 fragment ions (red dashed line, bold typeface). Manual inspection of the diagnostic fragment ions (y23 and y24) enables confident PTM localisation of the +15.99 Da shift to P315.





C

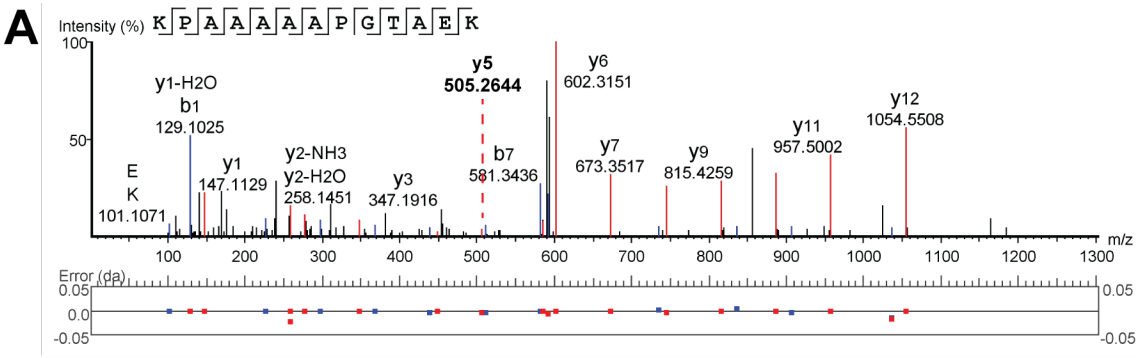
#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	44.0499	72.0449	A				17
2	70.0656	169.0978	P	1678.8217	1660.8112	1661.7947	16
3	30.0343	226.1187	G	1581.7737	1563.7584	1564.7513	15
4	72.0812	325.1863	V	1524.7363	1506.7369	1507.7205	14
5	101.0703	453.2463	Q	1425.6741	1407.6719	1408.6528	13
6	30.0343	510.2684	G	1297.6208	1279.6100	1280.5900	12
7	101.0714	638.3260	Q	1240.6033	1222.5885	1223.5850	11
8	104.0533	769.3656	M	1112.5432	1094.5299	1095.5134	10
9	44.0499	840.4044	A	981.4988	963.4894	964.4898	9
10	70.0656	937.4653	P	910.4620	892.4526	893.4418	8
11	70.0656	1034.5093	P	813.4102	795.3818	796.3927	7
12	86.0969	1147.5701	L	716.3508	698.3329	699.3303	6
13	70.0656	1244.6461	P	603.2734	585.2627	586.2571	5
14	88.0398	1359.6731	D	506.2245	488.2099	489.1935	4
15	60.0449	1446.7051	S	391.1936	373.1830	374.1666	3
16	102.0554	1575.7477	E	304.1615	286.1509	287.1345	2
17	129.1026		R	175.1190	157.0963	158.0919	1

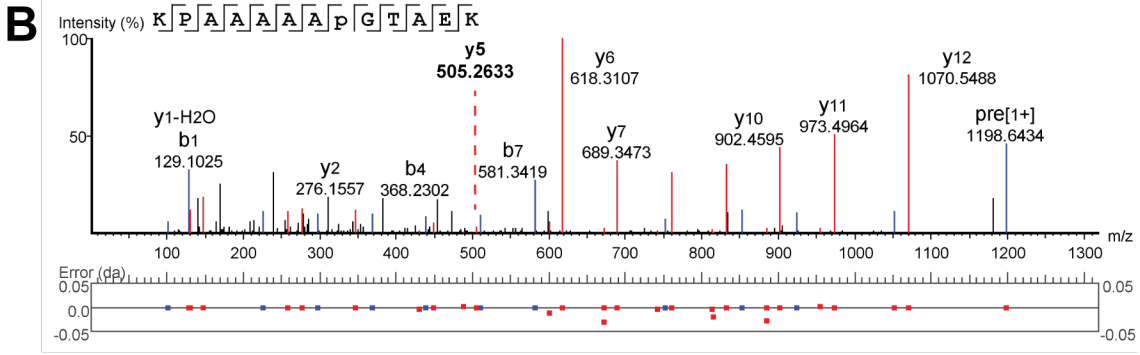
D

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	44.0499	72.0449	A				17
2	86.0605	185.0916	P(+15.99)	1694.8121	1676.8060	1677.7896	16
3	30.0343	242.1138	G	1581.7629	1563.7584	1564.7419	15
4	72.0812	341.1826	V	1524.7474	1506.7369	1507.7205	14
5	101.0713	469.2411	Q	1425.6755	1407.6555	1408.6523	13
6	30.0343	526.2613	G	1297.6199	1279.6100	1280.5872	12
7	101.0714	654.3164	Q	1240.5979	1222.5885	1223.5652	11
8	104.0533	785.3745	M	1112.5391	1094.5299	1095.5134	10
9	44.0499	856.3969	A	981.4985	963.4732	964.4742	9
10	70.0656	953.4514	P	910.4625	892.4540	893.4353	8
11	70.0656	1050.5042	P	813.4088	795.3995	796.3895	7
12	86.0969	1163.5911	L	716.3565	698.3467	699.3303	6
13	70.0656	1260.6410	P	603.2726	585.2627	586.2463	5
14	88.0398	1375.6680	D	506.2258	488.2099	489.1935	4
15	60.0449	1462.7000	S	391.1932	373.1830	374.1666	3
16	102.0554	1591.7426	E	304.1615	286.1509	287.1708	2
17	129.1021		R	175.1187	157.0970	158.0919	1

Figure 20 | Representative MS2 spectra of SLC38A10 HyP532.

Representative MS2 spectra and corresponding ion tables of tryptic fragments containing the target site, P532, from the BJ SH-SY5Y replicate 2 dataset. A) and C) MS2 spectrum of the Non-oxidised peptide with corresponding fragment ion table. B) and D) MS2 spectrum of the oxidised peptide with corresponding fragment ion table. Both MS2 spectra convey confident peptide sequencing through the y-ion series. The oxidised MS2 spectrum has been annotated to indicate the y16 fragment ions (red dashed line, bold typeface). Manual inspection of the diagnostic fragment ions (y15 and y16) enables confident PTM localisation of the +15.99 Da shift to P532.





C

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	101.1071	129.1025	K				13
2	70.0656	226.1560	P	1054.5508	1036.5585	1037.5258	12
3	44.0499	297.1904	A	957.5002	939.4894	940.4730	11
4	44.0499	368.2319	A	886.4609	868.4523	869.4359	10
5	44.0499	439.2707	A	815.4259	797.4152	798.3987	9
6	44.0499	510.3091	A	744.3924	726.3781	727.3616	8
7	44.0499	581.3436	A	673.3517	655.3409	656.3245	7
8	70.0656	678.3938	P	602.3151	584.3058	585.2874	6
9	30.0343	735.4125	G	505.2644	487.2511	488.2346	5
10	74.0605	836.4556	T	448.2380	430.2296	431.2132	4
11	44.0499	907.5032	A	347.1916	329.1819	330.1655	3
12	102.0548	1036.5585	E	276.1559	258.1451	259.1516	2
13	101.1078		K	147.1129	129.1025	130.0858	1

D

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	101.1073	129.1025	K				13
2	70.0656	226.1556	P	1070.5488	1052.5371	1053.5206	12
3	44.0499	297.1928	A	973.4964	955.4805	956.4678	11
4	44.0499	368.2302	A	902.4595	884.4476	885.4590	10
5	44.0499	439.2678	A	831.4227	813.4134	814.4148	9
6	44.0499	510.3045	A	760.3849	742.3771	743.3565	8
7	44.0499	581.3419	A	689.3473	671.3375	672.3516	7
8	86.0605	694.3887	P(+15.99)	618.3107	600.2987	601.2946	6
9	30.0343	751.4109	G	505.2633	487.2471	488.2346	5
10	74.0605	852.4584	T	448.2413	430.2334	431.2132	4
11	44.0499	923.4938	A	347.1932	329.1819	330.1655	3
12	102.0550	1052.5371	E	276.1557	258.1452	259.1284	2
13	101.1078		K	147.1131	129.1025	130.0865	1

Figure 21 | Representative MS2 spectra of SERPINH1 HyP30.

Representative MS2 spectra and corresponding ion tables of tryptic fragments containing the target site, P30, from the Geiger A549 HCD dataset. A) and C) MS2 spectrum of the Non-oxidised peptide with corresponding fragment ion table. B) and D) MS2 spectrum of the oxidised peptide with corresponding fragment ion table. Both MS2 spectra convey confident peptide sequencing through the y-ion series. The MS2 spectra have been annotated to indicate the y5 fragment ion (red dashed lines). Manual inspection of the diagnostic fragment ions (y5 and y6) enables confident PTM localisation of the +15.99 Da shift to P30.

Chapter 4. Orthogonal intervention strategies to identify 2OGD-catalysed prolyl hydroxylation in proteomic datasets

4.1 Introduction

The aim of the work in this thesis was to discover novel 2OGD-catalysed prolyl hydroxylation, with a particular focus towards the discovery of novel PHD1-3 substrates. *In silico* analysis of publicly deposited steady-state proteomic datasets provided some indication of novel 2OGD-catalysed hydroxyproline sites but an overreliance on the MS2 spectra in the absence of any biological intervention resulted in a limited capacity to identify 2OGD substrates. As an example of the difficulties of this 'MS2 alone' approach, some peptides do not fragment well due to intrinsic physicochemical properties (e.g., proline effect and non-mobile ions), which can negatively impact the PSM and PTM localisation confidence scores generated by the software. Isobaric oxidation of non-prolyl residues can be misassigned as hydroxyproline, and sequences containing many residues prone to non-enzymatic oxidation (e.g., tryptophan) have been shown to generate low PTM localisation confidence scores despite the MS2 spectra providing strong evidence in support of a correct hydroxyproline assignment (e.g., PDIA3 HyP404) (Figure 11). The use of high stringency PTM filters limits the number of incorrect hydroxyproline assignments but can also generate false negatives in which genuine sites of prolyl hydroxylation are inappropriately removed. In addition to these limitations, analysis of the steady-state proteome did not provide any factors to consider enzymatic hydroxylation beyond the degree of modification (i.e., stoichiometry) or the frequency of identity in multiple datasets. Neither of these descriptors provided direct evidence for 2OGD-catalysed prolyl hydroxylation. I therefore sought an orthogonal method in which prolyl hydroxylation was the direct subject of interrogation and for which the analysis could proceed independently of the quality of the MS2 spectra.

4.1.1 Applying interventions to study 2OGD-catalysed prolyl hydroxylation

The use of different interventions to compare hydroxylation in permissive or restrictive conditions might provide evidence for regulation of enzymatic hydroxylation that adds to the quality of evidence from the MS2 spectra, especially for peptides that could be prone

to poor fragmentation. Figure 22 outlines the premise of comparing samples from distinct biological conditions. Samples derived from different cell culture conditions are prepared separately, and data acquisition occurs independently for each sample. Comparison of precursor ion intensities enables regulation to be identified, with different methods available to determine the effect of the intervention. One option is to calculate the stoichiometry of hydroxylated peptide in each sample and compare these values to determine the effect of hydroxylase inhibition. Alternatively, peptide area can be normalised against the total ion chromatogram (TIC) so that relative peptide abundance changes are compared between samples.

Although both quantitation methods are consistent between each sample, the peptides have been prepared and analysed separately which can lead to inconsistent introduction of artefactual modifications in the instrument (e.g., oxidation and deamidation) and variable trypsinolysis. Such modifications might contribute to incorrect prolyl hydroxylation assignments and lead to incorrect conclusions regarding the experimental conditions and regulation of the hydroxyproline assignment.

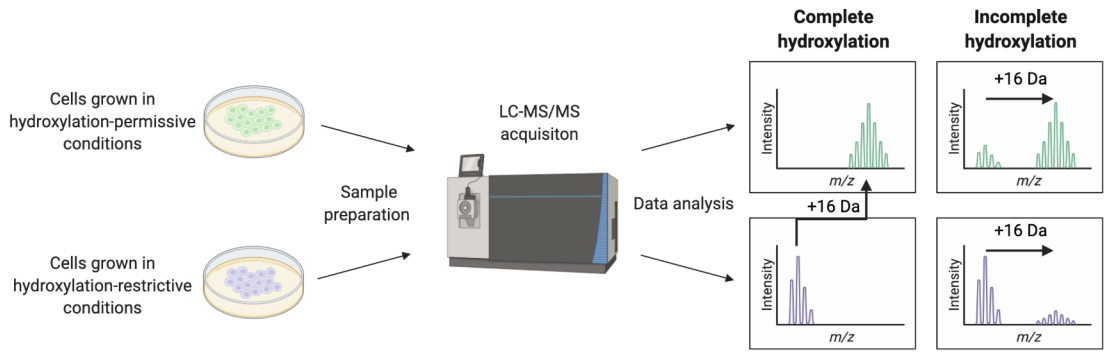


Figure 22 | Interventions to study prolyl hydroxylation.

Differential hydroxylation can be detected by LC-MS/MS after cells have been cultured in hydroxylation permissive or restrictive conditions. Hydroxylation may be complete or incomplete. Complete hydroxylation generates a single precursor ion and comparison of hydroxylated and non-hydroxylated peptides must be performed across conditions. Incomplete hydroxylation generates non-hydroxylated and hydroxylated ions that can be compared within a single sample.

4.1.2 Precursor ion quantitation with stable isotope labels

With the above in mind, a SILAC experiment was considered to improve the quantitation and sample handling (Figure 23). Stable isotope labelling of amino acids in cell culture

(SILAC) incorporates heavy isotopes into nascent proteins. Trypsin is the most commonly used protease in proteomics for its high fidelity to cleave the peptide bond at the C-terminus of arginine and lysine residues. For this reason, isotopically labelled arginine and lysine are used; R10, K8 are the most commonly used pair of isotopic labels in 2-plex experiments, whilst 3-plex experiments typically use R6, K4 for the third label. The nomenclature indicates the mass shift (Da) of the labelled amino acids.

Equal amounts of each protein lysate are combined after harvesting to control for downstream sample preparation steps. The increased complexity of SILAC lysates requires extensive off-line fractionation to improve proteome depth. Peptides are detected between samples according to a set of precursor ion peaks with similar retention times and the expected mass differences of the isotopically labelled residues. The intensities of these peaks are normalised against the TIC of each label, and the normalised peak areas are used to determine abundance changes between experimental conditions. This enables hydroxyproline regulation to be analysed as previously described.

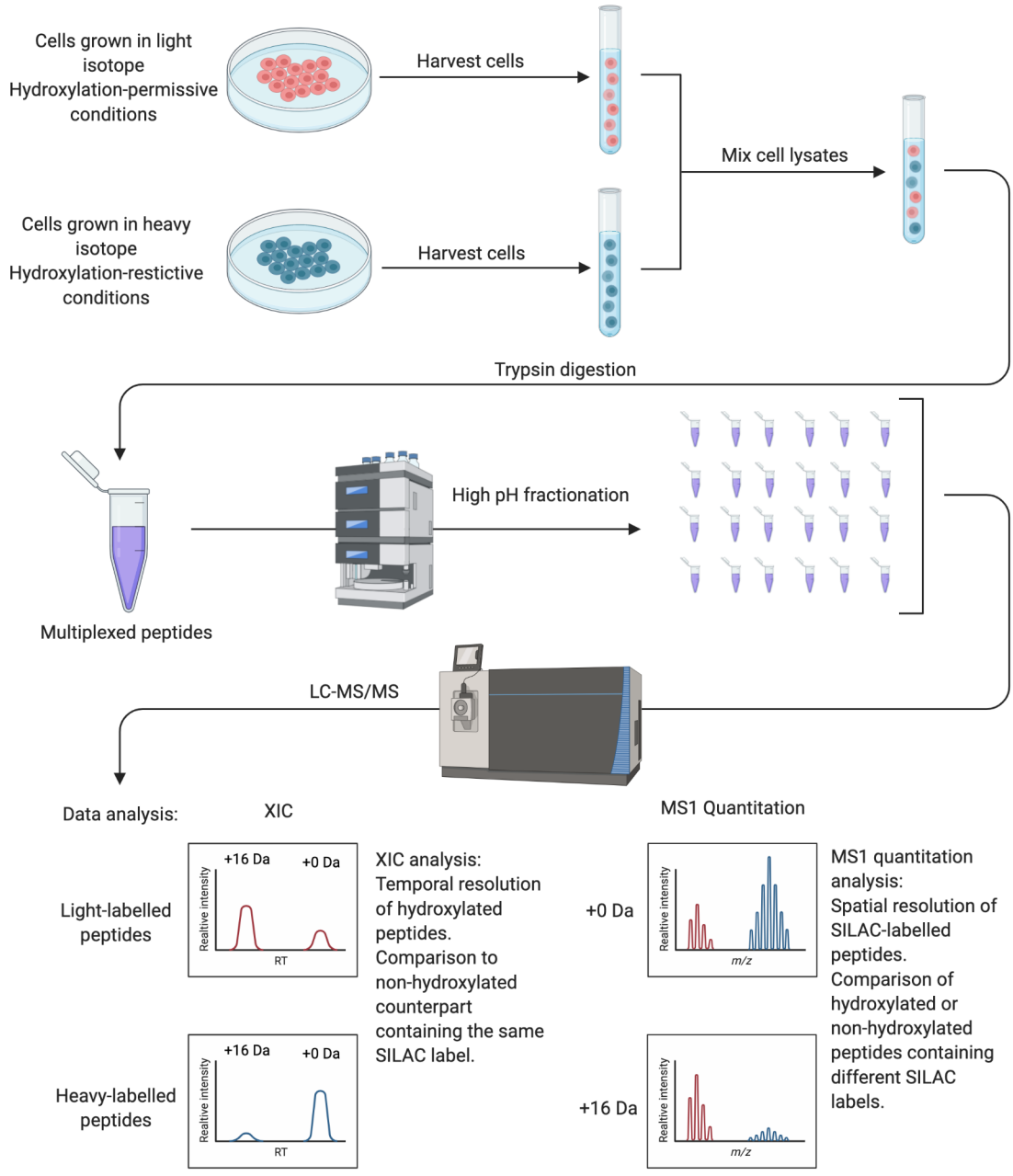


Figure 23 | SILAC overview for investigating prolyl hydroxylation.

Cells cultured in isotopically labelled media are exposed to hydroxylation-permissive or -restrictive conditions. Cell lysates are mixed so that proteins from each sample are exposed to identical conditions during sample handling. The multiplexed tryptic peptides are subject to high pH off-line fractionation to increase proteome depth and each fraction is analysed by LC-MS/MS. For data analysis, XICs and MS1 quantitation can be performed to determine regulation of prolyl hydroxylation.

4.1.3 Orthogonal interventions to study 2OGD-catalysed prolyl hydroxylation

The work in this chapter describes prolyl hydroxylation in response to hypoxia, pharmacologic inhibition, and genetic intervention using SILAC to label different experimental conditions.

Hypoxia is a physiologically relevant intervention that affects different hydroxylases according to their affinity for oxygen. Consequently, prolyl hydroxylation was assessed under modest (1 % O_2) and severe (0.1 % O_2) hypoxia relative to a steady-state normoxic (21 % O_2) control. The pan-2OGD inhibitor, DMOG, was employed in a second SILAC experiment to consider prolyl hydroxylation in the context of pharmacological inhibition of all 2OGDs.

Prolyl hydroxylation is an irreversible modification, which means that the changes in hydroxyproline are a function of nascent protein synthesis in the inhibitory condition coupled to degradation of the pre-existing hydroxylated population. For this reason, the interventions were performed over a sufficiently long period (24 hours) in order to detect changes in precursor ion intensity consistent with suppressed hydroxylation. A further experiment was performed with treatments beyond 24 hours to specifically examine prolyl hydroxylation dynamics of very long half-life proteins.

Hypoxia and DMOG are broad spectrum interventions with the potential to inhibit both prolyl and non-prolyl hydroxylase members of the 2OGD family. In this scenario, a non-prolyl hydroxylation site could be misassigned by the software as hydroxyproline. I therefore sought a genetic intervention to specifically target prolyl hydroxylation. Novel substrates of the PHDs are of significant interest, so I used a triple knock-out (TKO) cell line in which all three mammalian PHDs were genetically inactivated to specifically examine PHD-catalysed prolyl hydroxylation.

Hypoxia, DMOG treatment, and PHD ablation will inhibit HIFα prolyl hydroxylation but the hydroxylated HIFα proteoforms are rapidly degraded and are therefore unlikely to be detected in the control (i.e., oxygenated and non-inhibited) samples. It is possible that other hydroxylase substrates undergo similar functional dynamics, so a proteasome inhibitor was also added to facilitate detection of hydroxyproline-containing peptides of short half-life proteins.

With all of the above in mind, three experiments were performed to study prolyl hydroxylation in the mammalian proteome. A SILAC experiment, which was performed previously by Dr Matthew Cockman to compare protein abundance changes between hypoxia and normoxia, compared normoxia (21 % O_2) to 1 % and 0.1 % O_2. I re-analysed the data to consider prolyl hydroxylation across these conditions.

In a second SILAC experiment, MEF TKO cells were compared to their wild-type (WT) counterparts to assess the activity of PHD-catalysed prolyl hydroxylation. As a third condition, DMOG was applied to MEF WT cells to consider pan-2OGD-catalysed protein hydroxylation. A proteasome inhibitor, MG132, was applied in all experimental conditions to capture short half-life proteins (e.g., HIFα).

To study prolyl hydroxylation in long half-life proteins, rapidly proliferating cells were exposed to 1 % O_2 or DMOG for a minimum of 36 hours. A label-free quantification (LFQ) method was used to determine changes in hydroxylation between conditions. The conditions employed in these experiments are summarised in Table 20.

Table 20 | Interventions to study enzymatic prolyl hydroxylation.

Cell line	Conditions	Quantification	# Fractions	Data processing
U-87 MG	21 % O_2 (24 hours) 1 % O_2 (24 hours) 0.1 % O_2 (24 hours)	SILAC	20	PEAKS® 8.5
MEF	WT: 12.5 µM MG132 (4 hours) TKO: 12.5 µM MG132 (4 hours) WT: 1 mM DMOG (24 hours) and 12.5 µM MG132 (4 hours)	SILAC	30	PEAKS® X
eHAP	21 % O_2 (36 hours) 1 % O_2 (36 and 72 hours) 1 mM DMOG (36 hours)	LFQ	1 per condition	PEAKS® Xpro

4.2 Methods

4.2.1 Hypoxia SILAC screen in U-87 MG cells

The SILAC experiment comparing prolyl hydroxylation between hypoxia and normoxia in U-87 MG cells was performed by Dr Matthew Cockman. I performed the analysis of the processed data.

4.2.1.1 Cell culture conditions

U-87 MG cells were cultured in MEM supplemented with 10 % (v/v) dialysed FBS. SILAC-compatible amino acids were used to label each experimental condition:

- R0/K0 were used to label the light condition (21 % O_2).
- R6/K4 were used to label the medium condition (1 % O_2)
- R10/K8 were used to label the heavy condition (0.1 % O_2).

200 mg/mL L-proline was supplemented to minimise arginine to proline conversion. Cells were passaged for a minimum of 5 cell doublings and > 99% label incorporation was confirmed by LC-MS/MS of each sample. Labelled cells were seeded at 4 x10^6 cells on 150 mm^2 plates and exposed to the aforementioned oxygen concentrations for 24 hours.

4.2.1.2 Sample preparation

Samples were harvested and prepared for LC-MS/MS using an adapted protocol from a previously published report (Davis *et al.*, 2017). Briefly, cells were washed in ice-cold PBS and harvested in urea lysis buffer (7 M urea, 2 M thiourea, 40 mM Tris-HCl pH 8.0, 2 % (w/v) CHAPS, 250 U / mL Benzonase®, 1x HALT™ protease inhibitor cocktail, 20 mM DTT). Lysates were incubated for 1 hour with end-to-end rotation (15 rpm) at 4 °C. The lysates were clarified at 17,000 x *g* and protein quantitation of the soluble fraction was performed with a Pierce™ 660 nm protein assay kit.

Lysates were pooled 1:1:1 to yield 1 mg total protein. Proteins were digested with trypsin using the gel aided sample preparation (GASP) method (Fischer and Kessler, 2015). In this protocol, the reaction of reduced cysteine residues with monomeric acrylamide forms a propionamide adduct that replaces the alkylation steps of traditional sample preparation workflows (e.g., the use of iodoacetamide (IAA) to form carbamidomethylated

cysteines). Propionamide adducts can also form on primary amines, such as lysine residues and the N-termini of peptides, which was considered in the data processing. Peptides were extracted in acetonitrile (ACN) and lyophilised by vacuum centrifugation to remove the organic content.

4.2.1.3 High pH off-line fractionation

High pH off-line fractionation was performed with an XBridge BEH C18 XP column (2.5 µm x 3 mm × 150 mm), using the loading pump on a U3000 HPLC system with autosampler fraction collection. A 60-minute ACN gradient (1-35 % buffer B) was performed in NH_4OH, pH 10, at a flow rate of 200 µL/min, as described in Table 21. This resulted in 20 fractions that were lyophilised by vacuum centrifugation to remove the organic content.

Table 21 | Linear gradient for off-line high pH fractionation of SILAC-labelled peptides derived from U-87 MG extracts.

High pH fractionation was performed using buffer A (NH_4OH, pH 10) and buffer B (90 % ACN, NH_4OH, pH 10). A linear ACN gradient (1-35% buffer B) was performed over 60 minutes at a flow rate of 200 µL/min.

Time (minutes)	Buffer B (%)
0	1
12	1
72	35
80	95
90	95
90	1
100	1

4.2.1.4 LC-MS/MS data acquisition

The fractions were resuspended in 2 % ACN, 0.1 % FA and pooled pair-wise to generate 10 concatenated fractions (Table 22).

Table 22 | Fraction pooling scheme to generate 10 concatenated fractions.

Concatenated fraction #	Fraction A	Fraction B
F1	1	11
F2	2	12
F3	3	13
F4	4	14
F5	5	15
F6	6	16
F7	7	17
F8	8	18
F9	9	19
F10	10	20

Peptide fractions were analysed by LC−MS/MS in an Orbitrap Fusion™ Lumos™ Tribrid™ instrument. A 120-minute ACN gradient (2-35 % buffer B) in 0.1% FA, as described in Table 23, was performed with a U3000 UHPLC instrument with an EASY-Spray™ HPLC column (2 µm x 75 µm × 500 mm) at a flow rate of 0.250 µL/min.

Table 23 | Linear gradient for on-line low pH separation of SILAC-labelled peptides derived from U-87 MG extracts.

Peptide separation was performed on-line at low pH using buffer A (0.1 % FA) and buffer B (80 % ACN, 0.1 % FA). A linear ACN gradient (2-35 % buffer B) was performed over 120 minutes at a flow rate of 0.250 µL/min.

Time (minutes)	Buffer B (%)
0	2
3	2
6	5
123	35
130	99
135	99
138	2
145	99
148	99
149	2
175	2

The settings for precursor ion selection and fragmentation are defined in Table 24.

Table 24 | Settings for data-dependent acquisition SILAC-labelled peptides derived from U-87 MG extracts.

Entry	Parameter	Setting
MS1	Detector type	Orbitrap
	Resolution	120,000
	Mass range	Normal
	Use quadrupole isolation	True
	Scan range (*m/z*)	400-1,500
	RF lens (%)	30
	AGC target	4×10^5
	Maximum injection time (ms)	50
	Microscans	1
	Data type	Profile
	Polarity	Positive
MIPS	Monoisotopic peak determination	Peptide
Intensity	Filter type, intensity threshold	5×10^3
Charge state	Include charge state(s)	2-7
Dynamic exclusion	Exclude after *n* times	1
	Exclusion duration (s)	15
	Mass tolerance	ppm
	Low	5
	High	5
	Exclude isotopes	Yes
	If occurs within	30 s
	Exclusion duration	90 s
MS2	Isolation mode	Quadrupole
	Isolation window (*m/z*)	0.7
	Activation type	HCD
	Collision energy (%)	28
	Detector type	Linear ion trap
	Orbitrap resolution	30,000
	First mass (*m/z*)	110
	AGC target	4×10^3

	Maximum injection time (ms)	300
	Microscans	1
	Data type	Profile
Data dependent properties	Select: 'Cycle time'	3 s

4.2.1.5 Data processing

Raw data files were processed in PEAKS® using the parameters defined in Table 25.

Table 25 | Processing parameters for the hypoxia SILAC experiment performed with U-87 MG cells.

Parameter	Setting
PEAKS® version	8.5
PEAKS® search tool	DB
Reference proteome	Human, canonical (Uniprot id: UP000005640)
Precursor ion error tolerance	±10 ppm
Fragment ion error tolerance	±0.5 Da
Enzyme	Trypsin
Enzyme specificity	Specific
Maximum number of miscleavages per peptide	2
Fixed modification	Propionylation (+71.0371 Da; C)
Variable modifications	Oxidation (+15.9949 Da; CDFHKMNPRWY) Dioxidation (+31.9899 Da; FMWY) Deamidation (+0.9840 Da; NQ) Acetylation (+42.0367 Da; protein N-terminus) Propionylation (+71.0371 Da; peptide N-term, K) SILAC medium K (+ 4.0251 Da; K) SILAC heavy K (+ 8.0142 Da; K) SILAC medium R (+ 6.0201 Da; R) SILAC heavy R (+ 10.0083 Da; R) RtoP medium (+ 5.0168 Da; R) RtoP heavy (+ 6.0138 Da; K)
Maximum number of PTMs per peptide	3

Precursor ion quantification of SILAC-labelled peptides was performed using the Q search tool in PEAKS®, as described in Table 26.

Table 26 | Processing parameters for quantification of the hypoxia SILAC experiment performed with U-87 MG cells.

Quantitation was performed on the data generated using the DB search tool, as described in Table 25.

Parameter	Setting
PEAKS® version	8.5
PEAKS® search tool	Q
Quantification type	ICAT/SILAC
Quantification name	SILAC-3plex (R10,K8|R6,K4)
Quantification RT range	1 minute
FDR threshold	1 %
R to P conversion	True
Light (labelling efficiency)	R (100 %), K (100 %)
Medium (labelling efficiency)	R (+6.02) (100 %), K (+4.03) (100 %)
Heavy (labelling efficiency)	R (+10.01) (100 %), K (+8.01) (100 %)

4.2.1.6 Analysis

Filters were applied at the peptide level as previously described (Chapter 3). Briefly, confident PSMs were selected for (-10lgP ≥ 20). A second filter was applied for confident PTM assignments (AScore ≥ 20). The Q search tool does not enable a filter for ion intensity to be applied.

Two filters were applied at the protein level. A unique peptide is defined as a peptide with a -10lgP value above the peptide filtering threshold (-10lgP ≥ 20) that can be mapped to only one protein. Proteins were filtered for ≥ 3 unique peptide assignments. The protein -10lgP value is a weighted sum of the -10lgP scores of all peptides assigned to that protein. The supporting peptides are ranked in descending order according to their -10lgP value. The k-th ranked peptide contributes to the weighted sum with a weight of (-10lgP / k). Proteins were filtered to select for confident assignments (-10lgP ≥ 20).

To determine protein abundance in the PEAKS® Q search tool, the supporting peptides are ranked in descending order according to their abundance and the median value is used

as the protein abundance. This is repeated for each SILAC label. The protein abundance in each hypoxic condition (i.e., 1 % or 0.1 % O_2) was plotted relative to that of the normoxic condition on a log_2 scale to indicate the fold-change of protein abundance between the experimental groups.

HIF target genes were retrieved from the molecular signature database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb/; last accessed 25/03/2022). These were used to identify HIF target genes identified in U-87 MG cells in each condition of the SILAC experiment.

To determine hypoxic regulation of individual hydroxyproline sites, the abundance of peptides harbouring these PTMs must be compared. In SILAC experiments the abundance of a single peptide is determined in each experimental condition. This means a tryptic missed cleavage event, and other artefactual events (e.g., oxidation or deamidation) that generate discrete peptides can lead to missing values. Biological regulation could also result in completely diminished peptide abundance. Comparing ratios (e.g., abundance fold-change) between conditions can generate spurious results when one or more zero-values are present (i.e., infinitely increased or decreased peptide abundance from one condition to the next).

The stoichiometry for each hydroxyproline site is calculated independently of peptide sequence, which can consequently circumvent missing values by enabling discrete peptides to be compared. It also provides a relative value that can be used to compare the abundance of all hydroxyproline sites between conditions. I therefore plotted the delta stoichiometry, whereby the stoichiometry of hydroxyproline in the normoxic condition is subtracted from that of the hypoxic condition. Hydroxyproline sites were grouped according to the confidence of the peptide (PSM filter) and hydroxyproline (PTM filter) assignments.

For each oxidation site in the proteome, stoichiometry was calculated as:

$$Stoichiometry\ (\%) = 100\ \mathrm{X}\ \frac{\sum Area(peptides\ oxidised\ at\ the\ targetsite)}{\sum Area(peptides\ containing\ the\ target\ site)}$$

The data in Chapter 3 highlighted how stringency filters might generate false negatives. I therefore included peptides that did not meet the confidence filters (i.e., -10lgP < 20 and AScore < 20) because it is possible quantitation of the precursor ion might enable low

confidence hydroxyproline sites to be identified according to their regulation in the inhibitory conditions.

Candidates were identified on the basis of an absolute reduction in delta stoichiometry by ≥ 10 % in 0.1 % O_2. A simultaneous reduction at 1 % O_2 strengthened the candidacy but was not a requirement.

Once candidates were identified, the peptides were investigated for any inconsistencies with the peptide or hydroxyproline assignment, which have been described in the public data analysis (Chapter 3). Briefly, MS1 quality refinement required the hydroxyproline-containing peptide to i) exhibit a -ΔRT compared to the non-oxidised counterpart, ii) have no unique PTMs other than the hydroxyproline site(s), and iii) have no unique missed cleavage events. Manual inspection of the MS2 spectra was performed to consider correct peptide sequencing, identification of the diagnostic ions, and to ensure neutral loss ions consistent with oxidation of CMW (Table 13) were not present.

4.2.2 Genetic and pharmacologic intervention to MEF cells

4.2.2.1 Determination of optimal experimental conditions

A 'range-finding' experiment was performed to determine the optimal dose and duration of DMOG and MG132 treatments to capture prolyl hydroxylated HIF1α. Mouse embryonic fibroblasts (MEFs) were cultured in DMEM supplemented with 10 % (v/v) FBS. For the optimised experiment, cells were seeded at 2.5 x10^6 cells on 150 mm^2 plates. Wild-type (WT) and HIFα prolyl hydroxylase triple knock-out (TKO) MEF cells were treated with 12.5 μM MG132 (or the equivalent volume of vehicle, PBS) for 4 hours prior to harvest. Cells were treated with 1 mM DMOG (or the equivalent volume of vehicle, PBS) for 24 hours prior to harvest. The media was replenished 6 hours prior to harvest. The overall set of conditions tested are described in Table 27.

Table 27 | Experimental conditions for the optimised 'range-finding' experiment to confirm dose and duration of proteasome and pan-2OGD inhibitor treatments.

MEF = Mouse embryonic fibroblast; WT = wild-type; TKO = HIFα prolyl hydroxylase (PHD) triple knock-out

Cell line	Proteasome inhibition	pan-2OGD inhibition
WT MEF	Vehicle	Vehicle

WT MEF	12.5 μM MG132 (4 hours)	Vehicle
WT MEF	12.5 μM MG132 (4 hours)	1 mM DMOG (24 hours)
TKO MEF	Vehicle	Vehicle
TKO MEF	12.5 μM MG132 (4 hours)	Vehicle
TKO MEF	12.5 μM MG132 (4 hours)	1 mM DMOG (24 hours)

Cells were washed in ice-cold PBS and harvested in urea lysis buffer (7 M urea, 2 M thiourea, 40 mM Tris-HCl pH 8.0, 2 % (w/v) CHAPS, 250 U / mL Benzonase®, 1x HALT™ protease inhibitor cocktail, 20 mM DTT). Lysates were incubated for 1 hour with end-to-end rotation (15 rpm) at 4 °C. The lysates were clarified at 17,000 x *g* and protein quantitation of the soluble fraction was performed with Pierce™ 660 nm protein assay kit.

A western blot was performed to detect HIF1α and the prolyl hydroxylated CODD site, HyP564. For each sample, 20 μg total protein lysate was resolved by 7.5 % SDS-PAGE and the western blot protocol described in Section 2.1.1 was performed.

4.2.2.2 Cell culture conditions

WT and TKO MEF cells were cultured in DMEM supplemented with 10 % (v/v) dialysed FBS. SILAC compatible amino acids were used to label each experimental condition:

- R0/K0 were used to label the light condition (MEF WT cells with 12.5 μM MG132).
- R6/K4 were used to label the medium condition (MEF TKO cells with 12.5 μM MG132)
- R10/K8 were used to label the heavy condition (MEF WT cells with 1 mM DMOG and 12.5 μM MG132).

200 mg/mL L-proline was also supplemented to minimise arginine to proline conversion. Cells were passaged for a minimum of 8 cell doublings and > 99% label incorporation was confirmed by LC-MS/MS of each sample. Labelled cells were seeded at 2.5 $x10^6$ cells on 150 mm^2 plates. 1 mM DMOG treatment was performed for 24 hours prior to harvest. Cell culture media was replenished 6 hours prior to harvest for each experimental condition. 12.5 μM MG132 was added to each plate 4 hours prior to harvest.

4.2.2.3 Sample preparation

Cells were harvested using the same protocol described in Section 4.2.2.1. Lysates were pooled 1:1:1 to yield 1 mg total protein. Proteins were digested with trypsin using the GASP protocol, as described in Section 4.2.1.2.

4.2.2.4 High pH off-line fractionation

High pH off-line fractionation was performed on 800 µg of digested material with an XBridge BEH C18 XP column (2.5 µm x 3 mm x 150 mm), using the loading pump on a U3000 HPLC system with autosampler fraction collection. A 60-minute ACN gradient (1-35 % buffer B) was performed in NH_4OH, pH 10, at a flow rate of 200 µL/min, as described in Table 21. This resulted in 30 fractions that were lyophilised by vacuum centrifugation to remove the organic content.

4.2.2.5 LC-MS/MS data acquisition

The lyophilised fractions were resuspended in 2 % ACN, 0.1 % formic acid (FA). Peptide fractions were analysed by LC−MS/MS in an Orbitrap Fusion™ Lumos™ Tribrid™ instrument. A 28-minute gradient (2-40 % ACN buffer B) in 0.1 % FA, as described in Table 28, was performed with a U3000 HPLC instrument with an Acclaim™ Pepmap™ 100 C18 HPLC column (5 µm x 0.1 mm x 20 mm) at a flow rate of 0.275 µL/min.

Table 28 | Non-linear gradient for on-line low pH separation of SILAC-labelled peptides derived from the experiment comparing DMOG treatment and PHD inactivation in MEFs.

Peptide separation was performed on-line at low pH using buffer A (0.1 % FA) and buffer B (80 % ACN, 0.1 % FA). A non-linear ACN gradient (2-40 % buffer B) was performed over 28 minutes at a flow rate of 0.275 µL/min.

Time (minutes)	Buffer B (%)
0	2
5	2
5.5	8
33	40
34	95
44	95

45	2
60	2

The settings for precursor ion selection and fragmentation are defined in Table 29.

Table 29 | Settings for data-dependent acquisition in the SILAC experiment comparing DMOG treatment and PHD inactivation in MEFs.

Entry	Parameter	Setting
MS1	Detector type	Orbitrap
	Resolution	120,000
	Mass range	Normal
	Use quadrupole isolation	True
	Scan range (*m/z*)	350-1,500
	RF lens (%)	40
	AGC target	4.0×10^5
	Maximum injection time (ms)	50
	Microscans	1
	Data type	Profile
	Polarity	Positive
MIPS	Monoisotopic peak determination	Peptide
Intensity	Filter type, intensity threshold	1×10^4
Charge state	Include charge state(s)	2-8
Dynamic exclusion	Exclude after *n* times	1
	Exclusion duration (s)	30
	Mass tolerance	ppm
	Low	10
	High	10
	Exclude isotopes	True
MS2	Isolation mode	Quadrupole
	Isolation window (*m/z*)	1.2
	Activation type	CID
	Collision energy (%)	35
	Detector type	IonTrap
	First mass (*m/z*)	110

	AGC target	2 x10^3
	Maximum injection time (ms)	300
	Microscans	1
	Data type	Centroid
Data dependent properties	Select: 'Cycle time'	3 s

4.2.2.6 Data processing

Raw data files were processed in PEAKS® using the parameters defined in Table 30.

Table 30 | Processing parameters for the SILAC experiment comparing DMOG treatment and PHD inactivation in MEFs.

Parameter	Setting
PEAKS® version	X
PEAKS® search tool	DB
Reference proteome	Mouse, canonical (Uniprot id: UP000000589)
Precursor ion error tolerance	±8 ppm
Fragment ion error tolerance	±0.5 Da
Enzyme	Trypsin
Enzyme specificity	Specific
Maximum number of miscleavages per peptide	2
Fixed modification	Propionylation (+71.0371 Da; C)
Variable modifications	Oxidation (+15.9949 Da; CDFHKMNPRWY) Dioxidation (+31.9899 Da; FMWY) Deamidation (+0.9840 Da; NQ) Acetylation (+42.0367 Da; protein N-terminus) Propionylation (+71.0371 Da; peptide N-term, K) SILAC medium K (+ 4.0251 Da; K) SILAC heavy K (+ 8.0142 Da; K) SILAC medium R (+ 6.0201 Da; R) SILAC heavy R (+ 10.0083 Da; R) RtoP medium (+ 5.0168 Da; R) RtoP heavy (+ 6.0138 Da; K)
Maximum number of PTMs per peptide	3

Precursor ion quantification of SILAC-labelled peptides was performed using the Q search tool in PEAKS®, as described in Table 31.

Table 31 | Processing parameters for quantification of the hypoxia SILAC experiment comparing DMOG treatment and PHD inactivation in MEFs.

Parameter	Setting
PEAKS® version	X
PEAKS® search tool	Q
Quantification type	ICAT/SILAC
Quantification name	SILAC-3plex (R10,K8|R6,K4)
Quantification RT range	1 minute
FDR threshold	1 %
R to P conversion	True
Light (labelling efficiency)	R (100 %), K (100 %)
Medium (labelling efficiency)	R (+6.02) (100 %), K (+4.03) (100 %)
Heavy (labelling efficiency)	R (+10.01) (100 %), K (+8.01) (100 %)

4.2.2.7 Analysis

The data was analysed the same way as the U87MG hypoxia SILAC experiment (see Section 4.2.1.6) with the following exception: HIF target genes were not analysed in this assay. The hallmark HIF target genes retrieved from the molecular signature database (MSigDB) are specific to human cell lines and may not be regulated by HIF in mouse cells. Additionally, analysing protein abundance changes between conditions is complicated in the context of proteasome inhibition, especially given there was no control group lacking MG132. Since the aim of this assay was to prospect for novel sites of prolyl hydroxylation, I did not perform an analysis of HIF target genes.

4.2.3 Using synthetic peptides to estimate the chromatographic properties of PDI target peptides from cellular extracts

In the previous experiment the target sites of the protein disulphide isomerases (PDIs) were found to be largely unresponsive to DMOG treatment. Experiments were performed with synthetic peptides to determine the chromatographic properties of these peptides,

which would inform further assays studying prolyl hydroxylation of PDIs. Additionally, the MS2 spectra could be compared to the biological samples to provide additional security of the hydroxyproline assignments.

The following sections describe the methods used to perform this analysis. The settings for LC-MS/MS data acquisition (Table 33) and parameters of how the raw data was processed in PEAKS® (Table 34) was identical for the synthetic peptides and endogenous peptides derived from eHAP extracts.

4.2.3.1 Synthetic PDI target peptides

Peptide standards identical to the putative PDI hydroxylation sites resolved with trypsin were synthesised for use as retention time standards (Table 32). These peptides were synthesised with carbamidomethylated cysteines to prevent disulphide bridge formation and replicate the peptides derived from biological samples.

Table 32 | Peptide sequences and monoisotopic mass of the synthetic PDI peptides.
HyP = hydroxyproline. C' = carbamidomethylated cysteine.

Protein	Target site	Peptide sequence	$[M + H]^+$
PDIA3	P404	DVLIEFYAPW(C')GH(C')K	1894.8616
	HyP404	DVLIEFYA(HyP)W(C')GH(C')K	1910.8565
PDIA4	P89	DTVLLEFYAPW(C')GH(C')K	1995.9092
	HyP89	DTVLLEFYA(HyP)W(C')GH('C)K	2011.9041
P4HB	P395	NVFVEFYAPW(C')GH(C')K	1913.8462
	HyP395	NVFVEFYA(HyP)W(C')GH(C')K	1929.8411

4.2.3.2 LC-MS/MS acquisition

The lyophilised peptides were resuspended in 5 % ACN, 0.1 % FA and incubated in a sonicating water bath for 15 minutes to increase solubilisation. For each peptide, 100 fmol was loaded onto a pre-equilibrated Evosep™ tip with 1x iRT peptide solution and 500 ng tryptic Pierce™ HeLa Protein Digest Standard. Peptides were eluted using the proprietary Evosep™ 44-minute pre-formed gradient and the peptides were analysed by LC−MS/MS in an Orbitrap Fusion™ Lumos™ Tribrid™ instrument. The settings used for precursor ion selection and fragmentation are defined in Table 33.

Table 33 | Settings for label-free data-dependent acquisition of synthetic PDI peptides and endogenous peptides derived from eHAP extracts.

Entry	Parameter	Setting
MS1	Detector type	Orbitrap
	Resolution	60,000
	Mass range	Normal
	Use quadrupole isolation	True
	Scan range (*m/z*)	375-1,200
	RF lens (%)	30
	AGC target	1×10^6
	Maximum injection time (ms)	50
	Microscans	1
	Data type	Profile
	Polarity	Positive
MIPS	Monoisotopic peak determination	Peptide
Intensity	Filter type, intensity threshold	5×10^4
Charge state	Include charge state(s)	2-6
Dynamic exclusion	Exclude after *n* times	1
	Exclusion duration (s)	15
	Mass tolerance	ppm
	Low	10
	High	10
	Exclude isotopes	Yes
MS2	Isolation mode	Quadrupole
	Isolation window (*m/z*)	1.4
	Activation type	HCD
	HCD collision energy (%)	32
	Detector type	Orbitrap
	Resolution	15,000
	AGC target	1×10^6
	Maximum injection time (ms)	22
	Microscans	1
	Data type	Centroid

Data dependent properties	Select: 'Cycle time'	1 s

4.2.3.3 Data processing

Raw data was processed in PEAKS® using the parameters defined in Table 34.

Table 34 | Processing parameters for analysing the synthetic PDI peptides and endogenous peptides derived from eHAP extracts.

Parameter	Setting
PEAKS® version	Xpro
PEAKS® search tool	DB
Reference proteome	Human, canonical (Uniprot id: UP000005640)
Precursor ion error tolerance	±10 ppm
Fragment ion error tolerance	±0.02 Da
Enzyme	Trypsin
Enzyme specificity	Specific
Maximum number of miscleavages per peptide	2
Fixed modification	Carbamidomethylation (+57.0214 Da; C)
Variable modifications	Oxidation (+15.9949 Da; CDFHKMNPRWY) Dioxidation (+31.9899 Da; FMWY) Deamidation (+0.9840 Da; NQ) Acetylation (+42.0367 Da; protein N-terminus)
Maximum number of PTMs per peptide	3

4.2.3.4 Data analysis

The PEAKS® DB search identified the synthetic peptides. Peptide sequencing was confirmed by manual MS2 inspection, which included identification of the diagnostic ions for the hydroxyproline sites. The retention time of each peptide was normalised relative to the indexed retention time (iRT) peptides. This enabled the retention time (RT) window of each peptide to be determined in order to generate an estimated retention time (ERT) of these tryptic peptides for downstream experiments.

4.2.4 Analysis of PDI hydroxylation following hypoxia and DMOG treatment in eHAP cells

eHAP cells were selected to study the regulation of prolyl hydroxylation in long-half proteins, such as the PDIs. eHAP cells are derived from KBM-7 cells, which exhibit a higher proliferation rate than most adherent cell lines (Kotecki et al., 1999) and must therefore perform protein synthesis at a rate that facilitates cell growth and division. I therefore hypothesised changes of hydroxyproline stoichiometry would be greater in a rapidly dividing cell line compared to those used in the previous SILAC screens.

4.2.4.1 Cell culture

eHAP cells were seeded at 4 x10^6 cells on 150 mm^2 plates in IMDM media, 10 % tetracycline-free FBS. After 24 hours the media was replenished, and cells were exposed to the treatment conditions: 1 % O_2 or 1 mM DMOG. The media was replenished at 16 and 32 hours. For the 72-hour hypoxic treatment, 1 x10^6 cells were seeded on 150 mm^2 plates and the media was replenished at 16, 32, 48, and 64 hours.

4.2.4.2 Sample preparation

Cells were harvested in urea lysis buffer as previously described (Section 4.2.2.3). The lysates were clarified at 17,000 x *g* and the soluble fraction was incubated for 1 hour in the dark with 50 mM IAA at room temperature. Excess IAA was quenched by the addition of 20 mM DTT for 30 minutes with end-to-end rotation (15 rpm). Protein quantitation of the soluble fraction was achieved with Pierce™ 660 nm protein assay kit and normalised to 0.5 mg/mL.

To remove salts and detergents from the samples prior to LC-MS/MS acquisition 50 µg of total protein lysate was incubated for 18 hours in 800 µL ice-cold acetone. The samples were centrifuged at 17,000 x *g* for 20 minutes and the supernatant was discarded. The protein pellet was resuspended in 500 µL 80 % ice-cold acetone and the pellet was disturbed by vortex. The samples were centrifuged for 17,000 x *g* for 10 minutes and the supernatant was discarded. The protein pellet was resuspended in 400 µL ice-cold acetone and the wash step was repeated. The protein pellet was left to air dry for 5 minutes before resuspension in 100 µL 1 M GndHCl, 0.1 M Tris-HCl pH 8.0.

To improve protein solubilisation and digestion to tryptic peptides, 0.5 µg LysC was added to each sample and incubated at 37 °C for 3 hours with orbital rotation at 800 rpm. The samples were diluted in an equivalent volume of water and 0.5 µg trypsin was added to the reaction mixture. This was incubated for 18 hours at 37 °C with orbital rotation at 800 rpm.

The digestion reaction was stopped by the dropwise addition of 10 % FA to achieve pH 3.0. The samples were then loaded onto a pre-equilibrated Oasis HLB µElution plate (2 mg sorbent per well) and washed three times in 200 µL 0.1 % FA. Elution was achieved in a final volume of 100 µL of 80 % ACN, 0.1 % FA. The samples were lyophilised by vacuum centrifugation to remove the organic content.

4.2.4.3 LC-MS/MS acquisition and data processing

The lyophilised peptides were resuspended in 5 % ACN, 0.1 % FA. For each sample, 500 ng was loaded onto a pre-equilibrated Evosep™ tip with 1x iRT peptide solution. The peptides were eluted using the proprietary Evosep™ 44-minute pre-formed gradient and analysed by LC−MS/MS in an Orbitrap Fusion™ Lumos™ Tribrid™ instrument. The settings used for precursor ion selection and fragmentation are defined in Table 33. Raw data was processed in PEAKS® using the parameters defined in Table 34.

4.2.4.4 Analysis

Filters that selected for confident PSMs (-10lgP ≥ 20) and PTM assignments (ion intensity ≥ 1 %) were applied. The AScore filter was not applied because the analysis in Section 3.3.6 had demonstrated that the AScore values generated for the PDI target sites were often below the recommended threshold.

Peptides containing the target sites were not identified in all experimental conditions. These missing values were most likely due to a combination of analysing unfractionated lysates (i.e., extremely high complexity) and the stochastic nature of precursor ion selection in DDA mode, in which the most abundant precursor ions were selected for fragmentation in 1 second cycles. With this in mind, I measured the abundance of precursor ions (m/z = 3) corresponding to the unmodified and hydroxylated forms of PDIA3 and P4HB (Table 35). The specific site of hydroxylation could not be confirmed by investigating the XICs because this analysis only considers peptides at the MS1 level.

Therefore, the unmodified (+0.00 Da) and oxidised (+15.99 Da), instead of prolyl hydroxylated, peptides were considered.

Table 35 | Monoisotopic mass of precursor ions used to generate extracted ion chromatograms.

Protein	Target site	Δ mw (Da)	$[M + 3H]^{3+}$
PDIA3	P404	+0.00	632.2920
		+15.99	637.6237
	P89	+0.00	665.9746
		+15.99	671.3062
P4HB	P395	+0.00	638.6203
		+15.99	643.9519

4.3 Results

Analysis of steady state proteomic data retrieved from the PRIDE database (Chapter 3) demonstrated how false positives and false negatives are generated when prospecting for sites of highly secure prolyl hydroxylation. This means relying exclusively on the MS2 spectra for PTM analysis is a limited strategy to discover novel sites of enzymatic hydroxyproline in the proteome. Orthogonal approaches are required to provide alternative metrics of hydroxyproline formation, such as proteome-wide screens comparing hydroxylation-permissive and -restrictive conditions.

4.3.1 Investigation of prolyl hydroxylation at 1 % and 0.1 % O_2

Hypoxia is a physiological intervention that suppresses the catalytic activity of hydroxylases according to their affinity for oxygen. A SILAC experiment had previously been performed in the Ratcliffe group to ascertain protein abundance changes in hypoxia and normoxia. Exposure of cells to 21 %, 1 % and 0.1 % O_2 was performed for 24 hours to account for hydroxylase-specific differences in oxygen sensitivity. This experiment could inform on alternative strategies for hydroxylation assignment, such as comparing hydroxylase-restrictive and -permissive conditions. SILAC labelling provided a comparative approach to discriminate between artefactual (i.e., non-2OGD-catalysed)

oxidations that are abundant and present on peptides from all conditions (e.g., methionine oxidation) and 2OGD-catalysed protein hydroxylation that will be absent or suppressed in specific SILAC populations. Therefore, I re-analysed the data to prospect for prolyl hydroxylation sites that were suppressed by hypoxia.

4.3.1.1 Proof of principle

Prior to LC-MS/MS, the abundance of HIF1α and HIF2α was determined by western blot to validate the experimental conditions (Figure 24). The western blots show stabilisation of HIF1α and HIF2α after exposure to different hypoxic conditions for 24 hours.

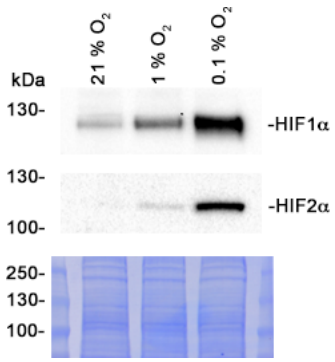


Figure 24 | HIF1α and HIF2α abundance at 21 %, 1 %, and 0.1 % O_2.

U-87 MG cells were exposed to 21 %, 1 % or 0.1 % O_2 for 24 hours prior to harvest. Western blots were performed with anti-HIF1α and anti-HIF2α antibodies, as indicated. Equivalent loading was confirmed by a Coomassie stain of the membrane after blotting.

The multiplexed protein lysates were digested with trypsin using the GASP method and high pH off-line fractionation was performed to increase total proteome depth. LC-MS/MS data was acquired for 10 concatenated fractions and processed in PEAKS® 8.5. The 'Q' search tool was used to interpret the SILAC labels and perform relative quantitation of peptides and proteins across the entire proteome.

Proteins were filtered for confident assignment (-10lgP ≥ 20) and ≥ 3 unique peptide assignments, which resulted in 4,075 confident unique protein identifications. A correlation between protein abundance changes for each of the hypoxic conditions relative to the normoxic control was noted (Figure 25). The majority of the proteome showed no change at 1 % O_2 but there are many proteins that decreased in abundance at 0.1 %.

Since hypoxia could lead to increased protein abundance in a HIF-dependent or -independent manner, the hallmark hypoxia gene set was retrieved from the molecular signatures database (MSigDB). In contrast to the rest of the proteome, HIF target genes displayed a trend for increased abundance in hypoxia, particularly at 0.1 % O_2. The abundance of IGFBP3 increased 5-fold at 0.1 % O_2, and F3 increased 3.5-fold. Some proteins displayed greater abundance increase in hypoxia than the target genes (e.g., PFKFB2, which displayed 7-fold increase in both hypoxic conditions). This might reflect cell-line specific HIF transactivation leading to increased protein synthesis in hypoxic cells, or it could represent hypoxic stabilisation.

The PHD-HIFα-pVHL pathway is an example of prolyl hydroxylation resulting in protein degradation by the ubiquitin proteasome system. It is therefore possible other proteins might be stabilised under conditions in which prolyl hydroxylase activity is perturbed. PFKFB2, URB1, NUDT1, ARG2 and CPA4 were the top 5 non-HIF target genes proteins that showed the greatest abundance increase at 1 % and 0.1 % O_2. Upon closer inspection these proteins were not found to be oxidised on any residue in an oxygen-sensitive manner. This reduces the probability these proteins undergo hydroxylation-mediated degradation, with the important caveat that proteasomal inhibitors were not applied in this screen.

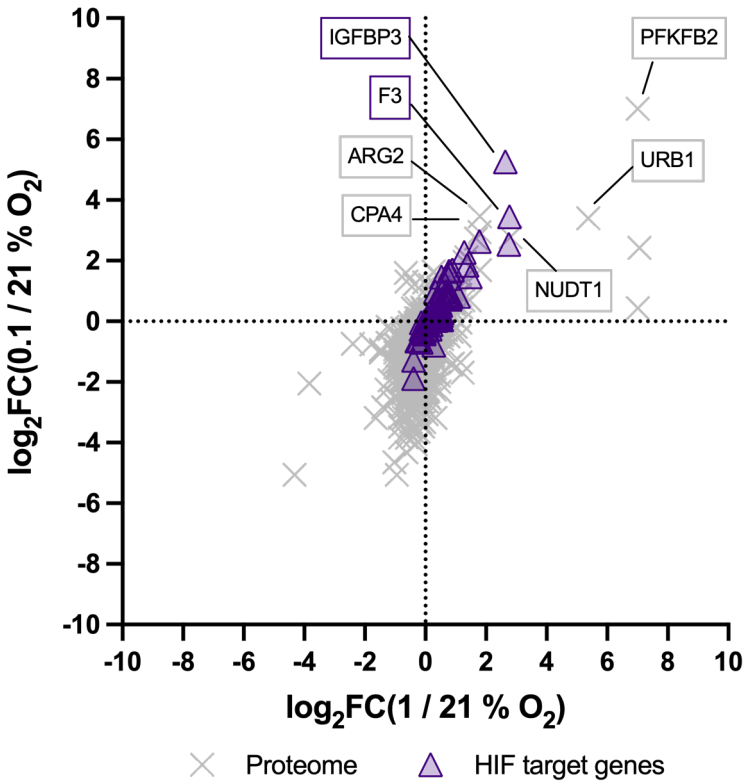


Figure 25 | Protein abundance changes in after exposure to 1 % or 0.1 % O_2 for 24 hours.

The abundance fold-change (FC) of 4,075 proteins in hypoxic conditions relative to normoxia was plotted on a log_2 scale. Proteins were grouped according to HIF target genes annotated in the MSigDB and non-target genes (i.e., the remaining proteome). The annotated proteins are discussed in the main text.

There have been 45 reports of non-HIFα PHD substrates (Table 14). As outlined in Chapter 3, previous studies within the laboratory found no evidence for PHD-catalysed prolyl hydroxylation towards 23 of these proteins by *in vitro* hydroxylase assays (Cockman *et al.*, 2019). Furthermore, work described in Chapter 3 of this thesis also found no evidence of confidently assigned prolyl hydroxylation on any of the 45 candidates in the steady state proteome. However, some of the reported non-HIFα PHD substrates have been suggested to undergo hydroxylation-dependent protein degradation, which would prevent identification of these sites in the steady state proteome. I therefore interrogated the data for these proteins to consider abundance changes in hypoxia, whereby increased abundance might represent hypoxic stabilisation.

Only 21 reported non-HIFα substrates were identified in this experiment (Table 36), of which MAPK7 showed the greatest induction in hypoxia (78 % increase at 1 % and 33 % increase at 0.1 % O_2). However, when compared to the induction of HIF1α and HIF2a in

the western blot above, the abundance change of MAPK7 is markedly less does not appear to be consistent with hypoxic stabilisation. Taken together, these results do not provide clear or consistent evidence of oxygen-dependent degradation of the reported non-HIFα PHD substrates.

Table 36 | Abundance changes of the reported non-HIFα PHD substrates in hypoxia relative to normoxia.

Some proteins did not meet the -10lgP (≥ 20) and unique peptides (≥ 3) filters and are therefore low confidence protein assignments. Coverage (%) gives an indication of how much of the protein sequence is identified in the multiplexed samples. The asterisk (*) indicates proteins that have been reported to undergo oxygen-sensitive proteolysis.

Protein	-10lgP	Intensity (21 % O_2)	Fold-change (% O_2)		Coverage (%)	#Peptides	#Unique
			1 / 21	0.1 / 21			
ACACB	64.39	2.57E+05	0.8	0.45	1	5	1
ACTB	246.69	3.58E+09	0.94	0.77	33	58	20
ADSL	135.83	9.47E+07	0.89	0.74	41	14	14
AKT1*	92.66	8.19E+06	1.07	0.65	25	11	9
ARRB2	54.33	2.13E+06	0.86	0.57	7	2	2
BRD4	79.53	3.16E+06	0.96	0.64	5	9	5
EEF2	249.31	6.88E+09	0.94	0.73	59	75	74
EEF2K	53.01	1.44E+06	1.4	0.89	8	4	4
FLNA*	316.49	4.91E+09	0.98	0.97	50	146	125
FOXO3*	54.01	4.97E+05	0.76	0.13	7	2	2
IKBKB	95.79	4.40E+06	0.83	0.7	10	5	5
MAPK7*	91.48	5.03E+05	1.78	1.33	7	3	3
NDRG3*	135.03	4.90E+07	0.86	0.68	29	9	9
OTUB1	117.24	1.11E+08	1	0.85	44	9	9
PKM	270.43	1.26E+09	1.08	0.99	9	68	8
POLR2A*	184.46	2.28E+07	0.98	0.63	25	30	30
PPP2R2A*	131.54	6.22E+07	1.08	0.83	37	14	10
SPRY2*	86.66	2.70E+06	0.56	0.17	15	3	3
TBK1	131.19	1.88E+07	0.98	0.79	25	14	14
TELO2	103.87	6.20E+06	0.91	0.62	12	7	7
UBE2N	86.39	5.41E+07	0.93	0.79	39	6	6

4.3.1.2 Global prolyl hydroxylation changes in hypoxia

The protein level changes manifest in the SILAC data were consistent with robust HIF pathway activation, and hence inhibition of HIFα prolyl hydroxylase activity in the hypoxic conditions that were assayed. I therefore analysed the data to prospect for novel sites of prolyl hydroxylation that were suppressed in hypoxia. To consider novel sites of prolyl hydroxylation, peptides were filtered for confident PSMs (-10lgP ≥ 20) and PTM localisation (AScore ≥ 20), as described in Chapter 3. There were 55,041 confidently assigned PSMs. Hydroxyproline was assigned to 115 peptides and 55 of these passed the minimum PTM localisation threshold (Figure 26).

The confidently assigned peptides that show sensitivity to hypoxia are mainly sensitive to 0.1 % O_2, but not 1 %. Interestingly, 35 of 88 collagen hydroxyproline sites show complete suppression at 0.1 % O2 but either no change, or an apparent increase in hydroxylation level at 1 %. This might reflect the oxygen affinity of the hydroxylases that catalyse these modifications. HIF-dependent upregulation of *P4HA1* and *P4HA2* most likely compensated to maintain efficient collagen hydroxylation at 1 % O_2. In agreement with this, the stoichiometry of some collagen hydroxyproline sites actually increased at 1 % O_2 (e.g., COL1A1 HyP178 and COL6A1 HyP484). However, at 0.1 % O_2 the activity of collagen prolyl 4-hydroxylase (C-P4H) appears to be inhibited. Overall, there were few hydroxyproline sites that exhibited large stoichiometry attrition at both 1 % and 0.1 % O_2. Several sites of oxygen-sensitive prolyl hydroxylation were observed on non-collagen proteins. These are discussed in more detail in Section 4.3.1.3.

The majority of hydroxyproline sites derived from confidently assigned peptides (PSM filter), including those with confident hydroxyproline assignments (PSM and PTM filters), cluster around Δ stoichiometry value of 0. The hydroxyproline sites in this region of the plot exhibited little regulation during the 24-hour hypoxic exposures and are therefore not oxygen sensitive sites of prolyl hydroxylation, at least in these experimental conditions.

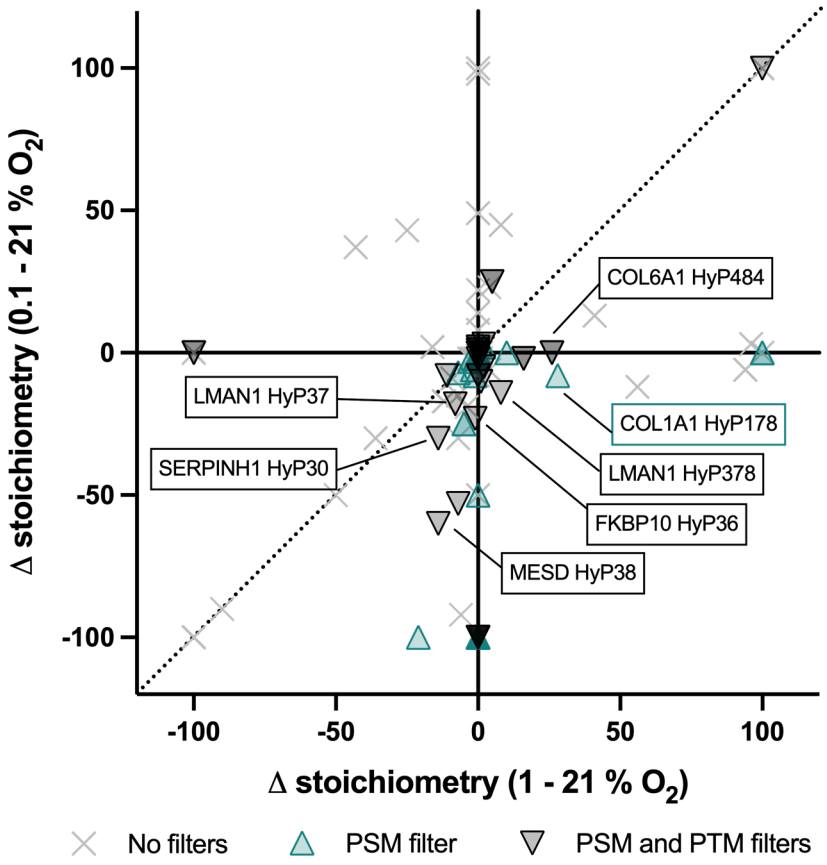


Figure 26 | Regulation of prolyl hydroxylation in hypoxia.

The change in hydroxyproline site stoichiometry was calculated for each condition relative to the normoxic control. The stoichiometry was calculated from peptides grouped according to the confidence of the PSM and PTM localisation. Lines at x = 0 and y = 0 indicate no change between the 1 % or 0.1 % O_2, respectively, and control. The dotted line at x = y indicates equal regulation in both hypoxic conditions relative to normoxia. The annotated hydroxyproline sites are discussed in the main text.

The stoichiometry of low confidence hydroxyproline sites were randomly distribute, consistent with many being mis-assigned. Some sites are equally suppressed in both hypoxic conditions, which includes a cluster of 9 hydroxyproline sites that show complete suppression in both hypoxic conditions with 100 % stoichiometry in the normoxic control. Manual inspection of these peptides revealed they were only present with additional PTMs, or missed cleavage events, or displayed +ΔRT when compared to the unmodified counterpart. These characteristics are consistent with incorrect peptide sequence determination or PTM assignments, or both. Peptides with these characteristics and displaying complete suppression in both hydroxylase-restrictive conditions are likely to have arisen from spurious assignments in the control group.

4.3.1.3 Identification of non-collagen hydroxyproline sites that display sensitivity to hypoxia

There were 9 non-collagen hydroxyproline sites that met the PSM confidence filter and displayed at least 10 % absolute stoichiometry suppression at 0.1 % O_2 (Figure 26). These were manually inspected to rule out hydroxyproline sites for which the tryptic peptides strongly suggested an incorrect assignment. The sites that passed the manual MS1 and MS2 spectra inspection are listed in Table 37. Of these FKBP10 HyP36, LMAN1 HyP378, and SERPINH1 HyP30 were identified in Chapter 3, which strengthens these assignments. In addition, the MS1 data showed that these peptides displayed -ΔRT compared to the unmodified counterparts. At the MS2 level the fragment ions were consistent with the PSM and the diagnostic ions were detected above the noise of the spectra (data not shown). LMAN1 HyP37 and MESD HyP38 also passed the MS1 and MS2 criteria. The assignment of MESD HyP38 was further strengthened by the hydroxylation being observed in a "HyP-G" motif, which was identified as an enriched prolyl hydroxylation consensus sequence of highly secure hydroxyproline sites in Section 3.3.5.

Table 37 | Hydroxyproline sites sensitive to 0.1 % O_2.

Protein	PTM Site	Stoichiometry			Δ stoichiometry	
		21% O_2	1% O_2	0.1 % O_2	1 - 21 % O_2	0.1 -21 % O_2
FKBP10	HyP36	93%	92%	70%	-1%	-23%
LMAN1	HyP378	65%	73%	50%	8%	-14%
LMAN1	HyP37	51%	43%	33%	-8%	-18%
MESD	HyP38	60%	46%	0%	-14%	-60%
SERPINH1	HyP30	68%	54%	38%	-14%	-30%

To corroborate the stoichiometric change in Figure 26, I examined the precursor ion abundance of the high confidence candidates with the hypothesis that a bona fide substrate would show simultaneous suppression of the hydroxylated peptide and increased abundance of the non-hydroxylated counterpart. FKBP10 HyP36 and SERPINH1 HyP30 exemplify the inverse regulation of hydroxyproline-containing and

non-hydroxylated tryptic peptides (Figure 27). As a protein is synthesised under inhibitory conditions, the abundance of non-hydroxylated peptides increases. The pre-existing population of prolyl hydroxylated protein is degraded, which results in reduced relative abundance of the hydroxyproline-containing peptides. When these factors are combined, the hydroxyproline stoichiometry decreases. LMAN1 HyP378 stoichiometry increases at 1 % O_2 and decreases at 0.1 %. This is consistent with observations at some collagen sites, as described above.

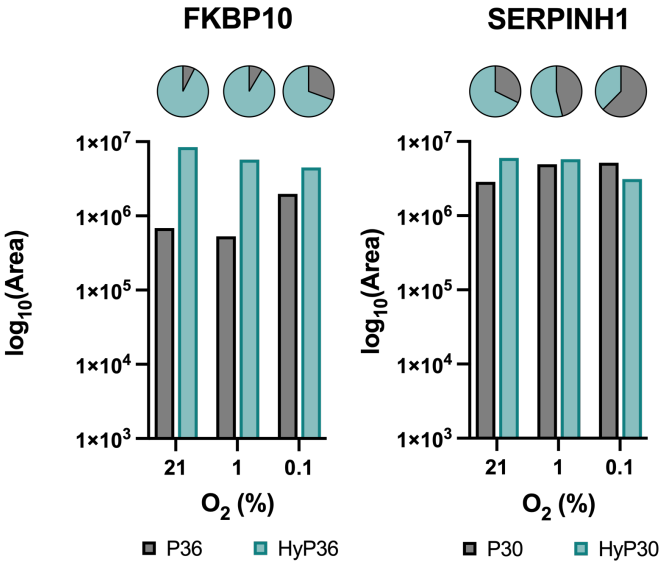


Figure 27 | Hydroxylation of FKBP10 P36 and SERPINH1 P30 is sensitive to hypoxia.

The summed area of proline- of hydroxyproline-containing peptides exhibit inverse regulation under hypoxia for FKBP10 P36 and SERPINH1 P30. Pie charts above each oxygen concentration indicate the stoichiometry of the hydroxylation sites in each condition.

The hydroxyproline sites identified in Chapter 3 were interrogated within this group because identification in multiple datasets could strengthen the security of these assignments. This resulted in the identification of TMEM109 HyP42 as a putative hypoxia-sensitive hydroxylation site (28 % stoichiometry at 21 % O_2 and 20 % stoichiometry at 0.1 % O_2; data not shown). The target site (^{41}PPG^{43}) is compatible as a C-P4H substrate, which would suggest that TMEM109 HyP42 should exhibit suppression in severe hypoxia. This was observed but the absolute suppression of HyP42 stoichiometry at 0.1 % O_2 did not meet the 10 % threshold. This could be due to long protein half-life, in which the pre-existing population of prolyl hydroxylated TMEM109

is not significantly diluted or turned over in 24 hours. Therefore, additional experiments would be required to demonstrate the oxygen-sensitivity of TMEM109 P42 hydroxylation.

Finally, the reported non-HIFα sites of PHD-catalysed prolyl hydroxylation were investigated. There was no evidence for prolyl hydroxylation at any of the target sites on peptides detected in this assay.

Overall, the analysis of U-87 MG cells exposed to increasingly stringent oxygen levels identified five novel sites of oxygen-sensitive prolyl hydroxylation. In addition, there were a number of sites of prolyl hydroxylation that appeared to be unresponsive to hypoxia treatment, at least in U-87 MG cells following incubation at 1 % or 0.1 % O_2 for 24 hours. Two technical aspects could be improved in further proteome-wide screens of prolyl hydroxylation: the identification of short half-life proteins and increased total proteome depth.

4.3.2 Investigation of prolyl hydroxylation following genetic and pharmacological inhibition

The primary purpose of the previous experiment was to prospect for protein abundance changes in hypoxia. Therefore, the experimental conditions were sub-optimal for comparing prolyl hydroxylation between hydroxylase permissive and restrictive conditions. Despite this, the analysis of U-87 MG cells provided the first demonstration of oxygen-sensitive prolyl hydroxylation for some hydroxyproline sites.
HIFα was not identified in the U-87 MG hypoxia SILAC screen, which suggested there was limited proteome depth. Even if proteome coverage had not been a limiting factor, proteasomal degradation of the prolyl hydroxylated proteoform would have prevented detection of the hydroxyproline-containing HIFα peptides. This experiment was therefore unlikely to identify other proteins that are subject to a similar regulatory process of hydroxylation-mediated proteasomal degradation. Therefore, a second SILAC experiment was designed with a different set of conditions in an attempt to identify novel substrates of 2OGDs, in particular, those of the PHDs.

To this end, I sought to compare mouse embryonic fibroblasts (MEFs) bearing defective copies of PHD1-3 genes, termed triple knock-out (TKO), with a wild-type (WT) counterpart line. As a complement to this approach, WT MEFs were treated with the pan-2OGD inhibitor dimethyloxalylglycine (DMOG) as a third condition. Importantly, all of the comparisons were made in cells treated with proteasomal inhibitor (MG132), thereby stabilising any hydroxylation events that would otherwise promote degradation (e.g., PHD-catalysed HIFα hydroxylation).

4.3.2.1 Experiment design

Prior to undertaking the SILAC screen I sought to define inhibitor conditions that would both inhibit hydroxylase activity (DMOG) and stabilise short half-life proteins (MG132). In particular, I wished to determine the maximal incubation time for MG132 and/or DMOG treatments in order to increase the dynamic range between hydroxylation-permissive and -restrictive conditions in the context of the SILAC labelling. The availability of antibodies specific to HIF1α and its hydroxylated form (Hyp564) allowed me to compare the expression and hydroxylation status at different doses (and duration) of inhibitor treatment.

Figure 28 shows the optimal conditions defined at 1 mM DMOG treatment for 24 hours and 12.5 μM MG132 treatment for 4 hours prior to cell harvest. Anti-HIF1α blots confirmed HIF1α stabilisation by proteasome blockade and PHD1-3 genetic inactivation. Inhibition of HIF1α prolyl hydroxylation by genetic inactivation of PHD1-3, and DMOG treatment was confirmed with an anti-HIF1α HyP564 antibody. This confirmed the conditions of the proteomic screen were suitable for interrogation of enzymatic prolyl hydroxylation across the proteome, including short half-life proteins.

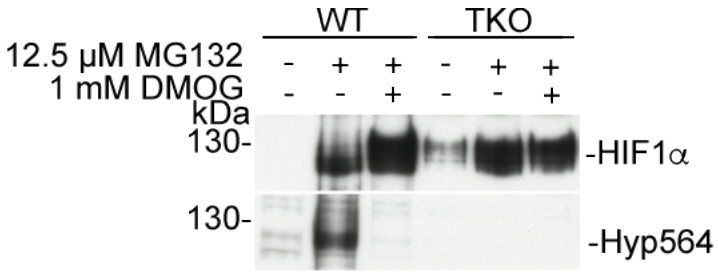


Figure 28 | HIF1α stabilisation and P564 hydroxylation dynamics in response to proteasome inhibition independently and in combination with pan-2OGD inhibition.

MEF WT and PHD1-3 triple knock-out (TKO) cells were exposed 12.5 μM MG132 for 4 hours prior to harvest. 1 mM DMOG treatment was performed for 24 hours prior to harvesting the cells. Western blotting with anti-HIF1α and anti-HIF1α HyP564 was performed to determine HIF1α stabilisation and prolyl hydroxylation, respectively, in response to these treatments.

4.3.2.2 LC-MS/MS method validation

For the proteomic screen, WT MEFs were labelled in the light media (R0/K0). TKO MEFs were cultured in the medium-labelled media (R6/K4), and DMOG-treated WT MEFs were labelled with the heavy isotopes (R10/K8). Heavy-labelled WT MEFs were treated with 1 mM DMOG for 24 hours prior to treatment. Additionally, each experimental group was treated with 12.5 μM MG132 for 4 hours prior to harvest.

SILAC samples exhibit a higher degree of complexity compared to non-multiplexed samples, which can hinder the number of unique peptide identifications and, ultimately, total proteome depth. In an attempt to circumvent this difficulty, high pH off-line fractionation was performed, which generated 30 fractions. These fractions were resolved on a 60-minute gradient with MS2 detection in CID mode in the linear ion trap to improve the sensitivity of the analysis.

Raw data were processed in PEAKS® X and quantitation was performed in the "Q" search tool. Proteins were filtered for confident assignment (-10lgP ≥ 20) and ≥ 3 unique peptides. This analysis identified 4,771 unique proteins that passed the minimum thresholds described above, which is an improvement on the U87MG hypoxia SILAC experiment by approximately 700 protein identifications. The increased coverage is most likely a consequence of greater fractionation. HIF1α was detected and did not show much regulation in any condition (Figure 29), which is consistent with proteasome inhibition sufficiently stabilising HIF1α (Figure 28).

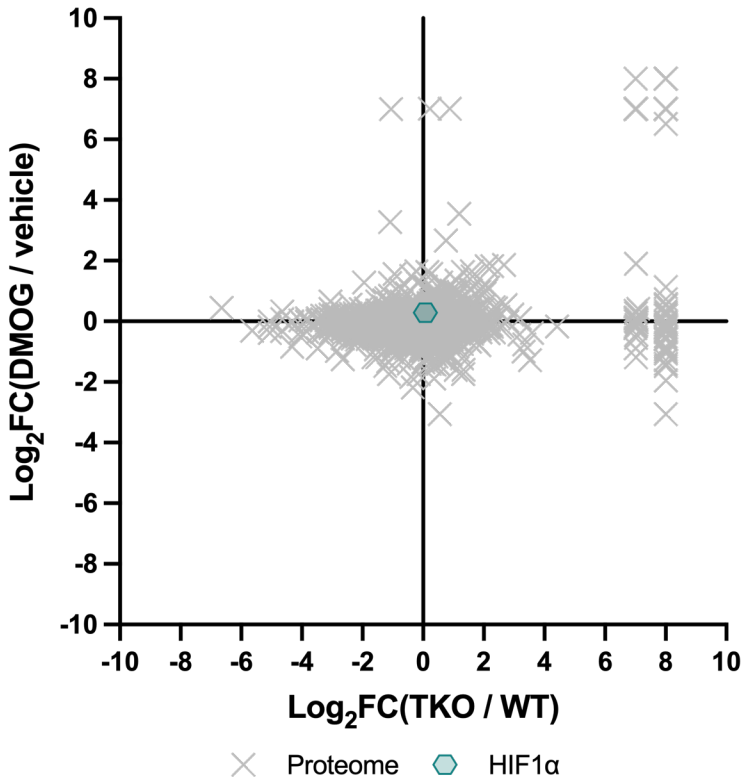


Figure 29 | Protein abundance changes in proteasome-inhibited cells after exposure to 1 mM DMOG for 24 hours or in TKO cells.

Abundance fold-change of 4,771 unique proteins in DMOG-treated wild-type (WT) cells or PHD1-3 triple knock-out (TKO) cells relative to vehicle treated wild type cells was plotted on a $\log_2$ scale.

4.3.2.3 Investigation of enzymatic prolyl hydroxylation

The principal aim of this experiment was to identify 2OGD-catalysed prolyl hydroxylation, particularly hydroxyproline sites generated by the catalytic activity of the PHDs. TKO cells were used to help define novel substrates of these isoenzymes and DMOG was applied to consider pan-2OGD-catalysed hydroxylation. Proteasome inhibition was applied in each condition to stabilise any proteins that undergo hydroxylation-mediated degradation.

Filters were applied to identify confident peptide spectrum matches (PSMs) (-10lgP ≥ 20) and hydroxyproline sites assigned with high PTM localisation confidence (AScore ≥ 20). In this analysis 66,332 confidently assigned unique peptides were identified. Hydroxyproline was assigned to 304 of these peptides and 93 were assigned with high confidence. Novel sites of prolyl hydroxylation were sought, therefore hydroxylated

peptides were grouped according to the assignment as a collagen or non-collagen hydroxyproline site.

Within each group of peptides, the stoichiometry was calculated for each hydroxyproline site in each condition. These sites were then grouped according to the confidence of the peptide with the highest -10lgP score as follows:

- Low confidence peptide assignments (-10lgP < 20).
- High confidence peptide assignments (-10lgP ≥ 20) with low confidence PTM assignment (AScore < 20).
- High confidence peptide assignment (-10lgP ≥ 20) with high confidence PTM localisation (AScore ≥ 20).

4.3.2.3.1 Novel PHD1-3 substrate candidates

The experimental conditions used for Figure 28 were repeated for the LC-MS/MS experiment. Therefore, the stoichiometry of the PHD substrates would be expected to exhibit complete suppression in TKO cells and complete, or nearly complete, suppression in DMOG-treated cells.

The stoichiometry of some hydroxyproline sites showed equivalent suppression by PHD1-3 gene inactivation and DMOG treatment (Figure 30), which would be consistent with PHD-catalysed prolyl hydroxylation. Inspection of the PHD substrate candidate peptides at the MS1 and MS2 level revealed these sites had weak evidence supporting both peptide and hydroxyproline assignments. For example, the MS2 spectra of the hydroxylated and non-hydroxylated peptides were not similar. Most of these peptides exhibited prolyl hydroxylation exclusively in combination with another PTM (e.g., deamidation, N-terminal propionylation, or oxidation at non-prolyl residues) that were not present on the unmodified counterparts. Trypsin missed cleavage events were also specific to some of the hydroxyproline-containing peptides. Some of these peptides exhibited a positive ΔRT when compared to the unmodified counterpart, which is inconsistent with oxidation of the peptide. Overall, there was no strong evidence a PHD substrate was identified in this analysis.

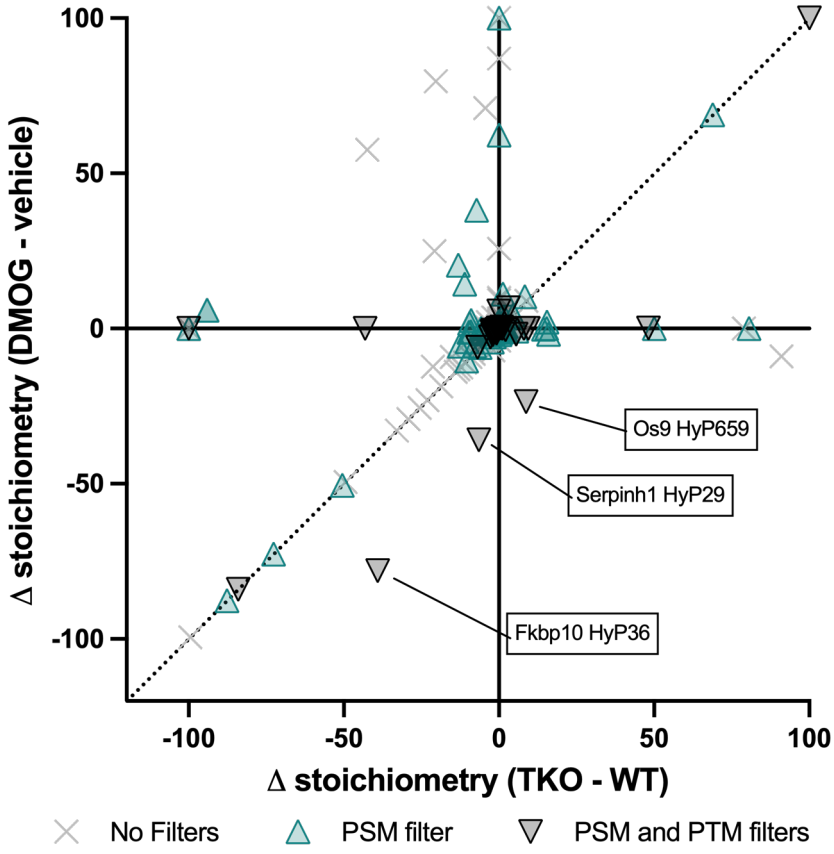


Figure 30 | Stoichiometry changes of non-collagen hydroxyproline sites.

The Δ stoichiometry of hydroxyproline sites from non-collagen proteins. Hydroxyproline sites were grouped according to the confidence of PSM (-10lgP ≥ 20) and PTM localisation (AScore ≥ 20). Solid lines at x = 0 and y = 0 indicate no change between the TKO cells or DMOG treated WT cells, respectively, and control. The dotted line at x = y indicates equal regulation in both hydroxylation-restrictive conditions relative to vehicle treated cells. The annotated hydroxyproline sites are discussed in the main text.

4.3.2.3.2 Novel non-PHD 2OGD substrate candidates

The stoichiometry of non-PHD 2OGD substrates would be expected to decrease following DMOG treatment and display minimal changes in TKO cells relative to the WT control. This was observed for most collagen hydroxyproline sites, which are discussed in Section 4.3.2.3.3. There were three hydroxyproline sites that showed greater suppression by DMOG treatment than by PHD1-3 inactivation (Figure 31). Manual inspection of these sites corroborated the peptide-level and hydroxyproline assignments. Importantly, these peptides displayed no additional modifications and the trypsin missed cleavage events observed for Os9 HyP659 and Serpinh1 HyP29 peptides were also

observed on the unmodified counterparts. Hydroxylated peptides were detected in either identical or adjacent fractions as unmodified counterparts with an anticipated -ΔRT, consistent with correct PTM assignment.

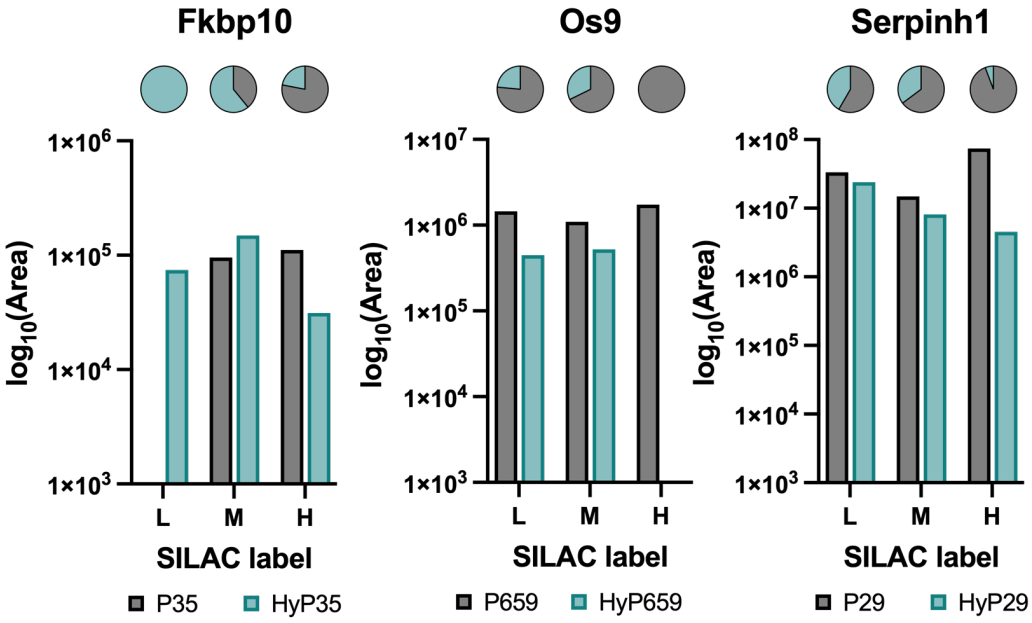
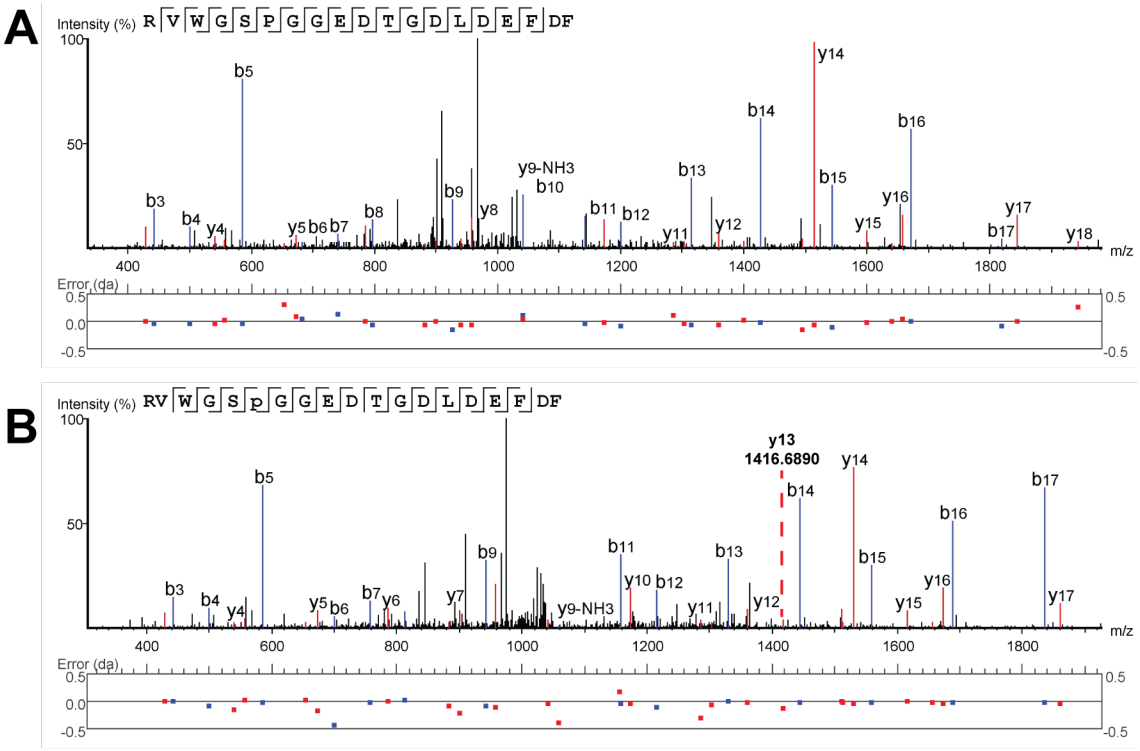


Figure 31 | SILAC label quantitation of peptides containing hydroxyproline sites sensitive to DMOG intervention that were confirmed by manual MS2 inspection.

Target site stoichiometry was plotted for each candidate in each experimental condition, which are indicated by the SILAC labels 'L', 'M', and 'H'. Pie charts above each SILAC label indicate the stoichiometry of the hydroxylation site. L = vehicle-treated WT. M = vehicle-treated TKO. H = DMOG-treated WT.

The human homologues of Fkbp10 HyP35 (FKBP10 HyP36) and Serpinh1 HyP29 (SERPINH1 HyP30) were identified in the public data analysis as sites of high confidence hydroxyproline assignment. The assignment of Os9 HyP659 (OS9 HyP654) was identified in the high stringency workflow but rejected on the basis of poor MS2 quality spectra that could not be used to confirm correct localisation of the hydroxylation. In this experiment, the MS2 spectra of Os9 HyP659 and P569 peptides were of sufficient quality to localise the hydroxylation to P659 according to the diagnostic b5 and b6 ions (Figure 32). The y13 and y14 ions also enable confident localisation of HyP659 but the y13 ion was not identified in the unmodified MS2 spectrum. The MS2 spectra of Fkbp10 HyP35 and Serpinh1 HyP29 also provided sufficient quality to corroborate the hydroxyproline

assignment generated by PEAKS® (data not shown). Together, this dataset provides evidence of DMOG-sensitive prolyl hydroxylation of Fkbp10 P35, Serpinh1 P29 and Os9 P659. Prolyl hydroxylation of these sites persisted in TKO cells, which suggests these sites are subject to non-PHD 2OGD-catalysed prolyl hydroxylation.



C

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	129.1139	157.1089	R				19
2	72.0812	256.1773	V	1942.5330	1924.7871	1925.7706	18
3	159.0921	442.3030	W	1843.7080	1825.7188	1826.7023	17
4	30.0343	499.3370	G	1657.5890	1639.6394	1640.6200	16
5	60.0449	586.3740	S	1600.6530	1582.6179	1583.6014	15
6	70.0656	683.3170	P	1513.6720	1495.7560	1496.5695	14
7	30.0343	740.2440	G	1416.5437	1398.5332	1399.4919	13
8	30.0343	797.4880	G	1359.5990	1341.5117	1342.4952	12
9	102.0554	926.6020	E	1302.5620	1284.4902	1285.3640	11
10	88.0398	1041.3580	D	1173.4830	1155.4476	1156.4312	10
11	74.0605	1142.5770	T	1058.4313	1040.4208	1041.3580	9
12	30.0343	1199.6340	G	957.4670	939.4500	940.3566	8
13	88.0398	1314.6490	D	900.3750	882.4290	883.3351	7
14	86.0969	1427.6970	L	785.3540	767.3246	768.3082	6
15	88.0398	1542.7970	D	672.1480	653.9340	655.2241	5
16	102.0554	1671.7430	E	557.1960	539.2136	540.2510	4
17	120.0812	1818.8890	F	428.1720	410.1710	411.1546	3
18	88.0398	1933.8203	D	281.1132	263.1026	264.0862	2
19	120.0812		F	166.0862	148.0757	149.0592	1

D

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	129.1139	157.1089	R				19
2	72.0812	256.1773	V	1958.7926	1940.7821	1941.7656	18
3	159.0921	442.2550	W	1859.7860	1841.7136	1842.6971	17
4	30.0343	499.3740	G	1673.6920	1655.6344	1656.6420	16
5	60.0449	586.3340	S	1616.6340	1598.6129	1599.5964	15
6	86.0605	699.7940	P(+15.99)	1529.6440	1511.5950	1512.5990	14
7	30.0343	756.4040	G	1416.6890	1398.5332	1399.5167	13
8	30.0343	813.3740	G	1359.5551	1341.5117	1342.4952	12
9	102.0554	942.5510	E	1302.5690	1284.4902	1285.7841	11
10	88.0398	1057.4702	D	1173.5110	1155.2710	1156.4312	10
11	74.0605	1158.5680	T	1058.4313	1040.4208	1041.4530	9
12	30.0343	1215.6620	G	957.5050	939.3730	940.3566	8
13	88.0398	1330.5840	D	900.5810	882.4570	883.3351	7
14	86.0969	1443.6760	L	785.3450	767.3246	768.3082	6
15	88.0398	1558.7150	D	672.4270	654.2180	655.2241	5
16	102.0554	1687.7560	E	557.1920	539.2136	540.3630	4
17	120.0812	1834.8260	F	428.1780	410.1710	411.1546	3
18	88.0398	1949.8153	D	281.1132	263.1026	264.0862	2
19	120.0812		F	166.0862	148.0757	149.0592	1

Figure 32 | Representative MS2 spectra of Os9 HyP659.

Representative MS2 spectra and corresponding ion tables of tryptic fragments containing the target site, P659, from light-labelled cells. A) and C) MS2 spectrum of the non-hydroxylated peptide with corresponding fragment ion table. B) and D) MS2 spectrum of the hydroxylated peptide with corresponding fragment ion table. Both MS2 spectra convey confident peptide sequencing through both fragment ion series. Manual inspection of the diagnostic fragment ions (b5 & b6, and y13 & y14) enables confident PTM localisation of the +15.99 Da shift to P659.

4.3.2.3.3 Collagen prolyl hydroxylation

Prolyl hydroxylation of collagen triple helical domains is catalysed by the related 2OGD, C-P4H (Myllyharju and Kivirikko, 2004). Therefore, peptides derived from these domains provide a useful reference with which to assess the potency of DMOG. The stoichiometry of the majority of collagen hydroxyproline sites decreased with DMOG treatment, as expected (Figure 33). Overall, collagen prolyl hydroxylation appears to faithfully represent the expected hydroxyproline dynamics in TKO cells and DMOG-treated WT cells, at least for the majority of collagen hydroxyproline sites.

Interestingly, I observed that some collagen hydroxyproline sites displayed dynamics that were inconsistent with C-P4H-catalysed prolyl hydroxylation. For example, an inverse relationship between increased stoichiometry in TKO cells and decreased stoichiometry after DMOG treatment. Closer inspection revealed several factors might contribute to these dynamics. The peptides containing these hydroxyproline sites often contained trypsin missed cleavage events, residues that are prone to artefactual oxidation, or the hydroxyproline was assigned in the 'Xaa' position of an 'Xaa-Yaa-Gly' sequence. Although C-P3H catalyses prolyl 3-hydroxylation in the 'Xaa' position of 'Xaa-Yaa-Gly' sequences, there is a pre-requisite for 4-hydroxyproline to occupy the 'Yaa' position (Gorres and Raines, 2010). This was not the case for some of the sites identified in the data, which are therefore likely to have been incorrectly assigned. In addition to random missed cleavage events, the stochastic nature of incorrectly localised artefactual oxidations and hydroxyproline assignments could also generate the spurious dynamics of prolyl hydroxylation observed in this data. This highlights some of the complexities of analysing hydroxyproline sites in proteomics data, even when the site is assigned to *bona fide* 2OGD substrates. This reiterates the requirement for careful qualitative analysis of hydroxyproline assignments in LC-MS/MS data.

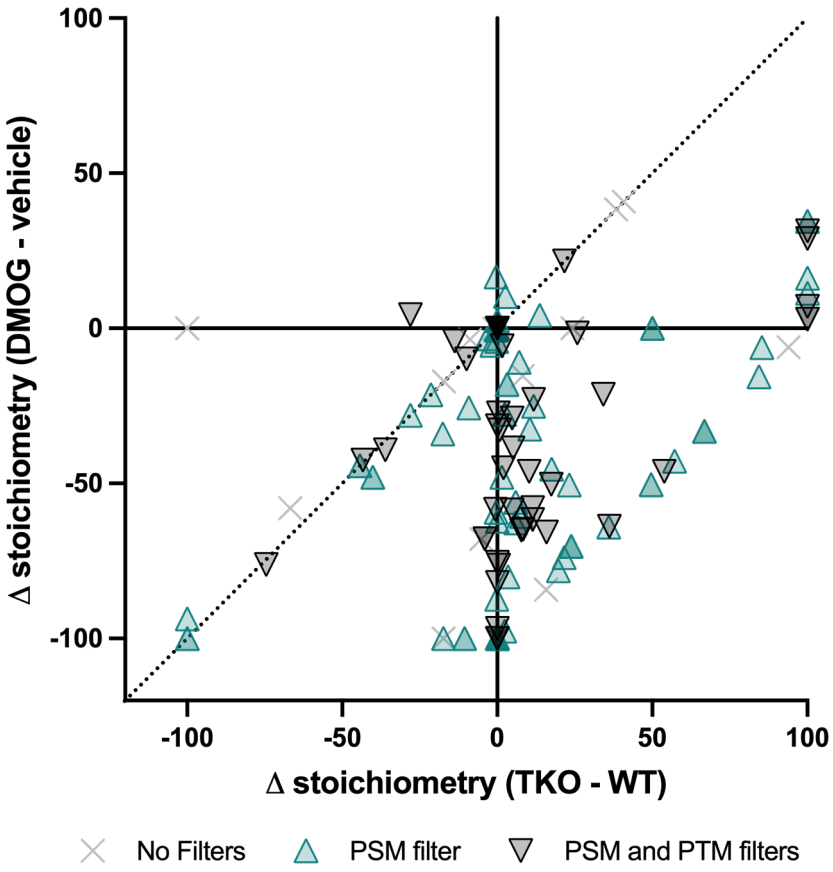


Figure 33 | Collagen hydroxyproline stoichiometry changes in TKO or DMOG-treated MEFs.

The Δ stoichiometry of hydroxyproline sites from the triple helical domains of collagen proteins. Hydroxyproline sites were grouped according to the confidence of PSM (-10lgP ≥ 20) and PTM localisation (AScore ≥ 20). Solid lines at x = 0 and y = 0 indicate no change between the TKO cells or DMOG-treated WT cells, respectively, and control. The dotted line at x = y indicates equal regulation in both hydroxylation-restrictive conditions relative to vehicle treated cells.

4.3.2.3.4 Sites of DMOG-insensitive prolyl hydroxylation

There are over 150 non-collagen hydroxyproline sites clustered around the origin that show minimal change in either condition (Figure 31). This includes RPS23 HyP62, which shows some suppression of the hydroxylated peptide after DMOG treatment (Figure 34), but the non-hydroxylated counterpart is not detected in any sample (i.e., stoichiometry remains at 100 % in each condition). A recent study of protein turnover in non-dividing primary cells predicts the half-life of RPS23 to be approximately 110 hours, or 4.5 days (Mathieson et al., 2018). Although dilution of pre-existing proteins must be accounted for in the proliferating MEF cells used in this investigation, it is still highly likely 24-hour

DMOG treatment is not sufficient to observe large suppression of prolyl hydroxylation. There is a possibility that the hydroxylation of other proteins with similarly long half-life will not be suppressed in the experimental conditions considered in this analysis.

This cluster of 150 hydroxyproline sites in MEFs was investigated for homologues of the candidates identified in the hypoxia SILAC experiment in human glioblastoma cells (Section 4.3.1.2) and the public data analysis (13 human cell lines) (Chapter 3). The hydroxyproline sites detected in this cluster were novel except for protein disulphide isomerases (PDIs). In this experiment, P4hb HyP51, P4hb HyP397, Pdia3 HyP55, Pdia3 HyP404, and Txndc5 HyP333 were identified. The half-life of these PDI proteins varies with cell type and is within the range of 3.5 to 25 days (Mathieson *et al.*, 2018). This analysis identified a reduction in P4hb HyP397 stoichiometry by 6 % after DMOG treatment. There was a slight increase in the abundance of peptides corresponding to P397 and a slight decrease in those containing HyP397 (Figure 34), which is consistent with DMOG-sensitive hydroxylation of a protein with a long half-life. However, the absolute change to HyP397 stoichiometry is small and, for the purposes of this analysis, cannot be distinguished from the stochastic variation expected between experimental groups. The results of P4hb HyP397 are representative of the other PDI hydroxyproline sites identified in this analysis in that changes were either very small or negligible. Further experiments were therefore designed to determine if P4hb HyP397 and the other PDI hydroxyproline sites were regulated by DMOG (Section 4.3.3).

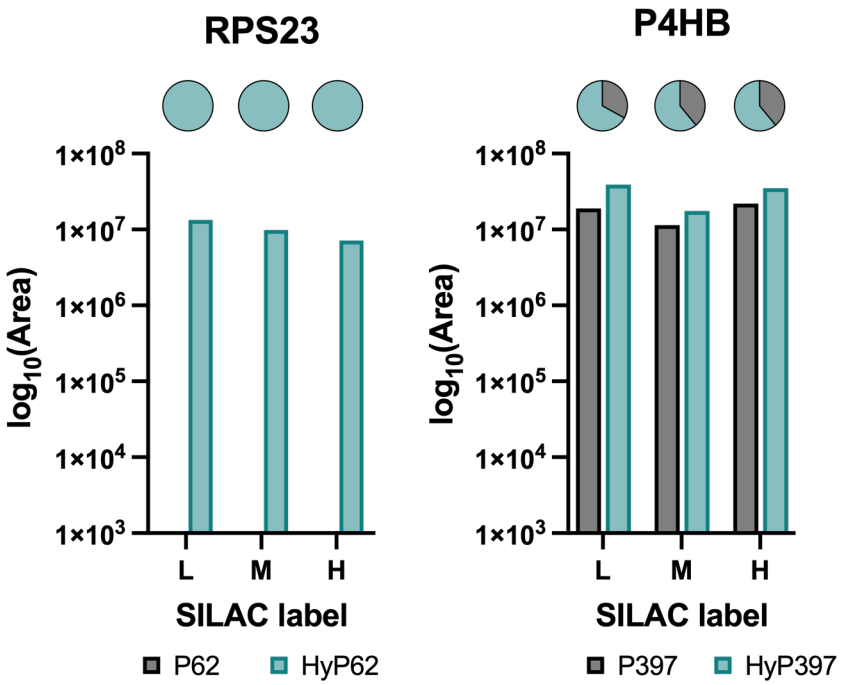


Figure 34 | SILAC quantitation of peptides containing hydroxyproline sites that were not sensitive to DMOG intervention.

Target site stoichiometry was plotted for each candidate in each experimental condition, which are indicated by the SILAC labels 'L', 'M', and 'H'. Pie charts above each SILAC label indicate the stoichiometry of the hydroxylation site. L = vehicle-treated WT MEFs. M = vehicle-treated TKO MEFs. H = DMOG-treated WT MEFs.

4.3.2.3.5 Reported non-HIFα PHD substrates

These SILAC data provided a useful resource to reference against the further interrogated for the reported non-HIFα PHD substrates. Pkm HyP403 was the only site identified from the 81 reported hydroxyproline sites. This site exhibited 0.2 % stoichiometry in hydroxylation permissive conditions and was fully suppressed in both TKO cells and after DMOG treatment (Figure 35), which would be compatible with PHD-catalysed hydroxylation. However, the diagnostic ions are not detected in the MS2 spectrum, which reduces the confidence of correct PTM assignment. Additionally, the low stoichiometry is unexpected for a putative site of enzymatic hydroxylation. The other reported hydroxyproline site of Pkm, HyP408, is not detected despite the residue existing on the same tryptic peptide. Hydroxyproline was not identified for any other reported site of PHD-catalysed prolyl hydroxylation.

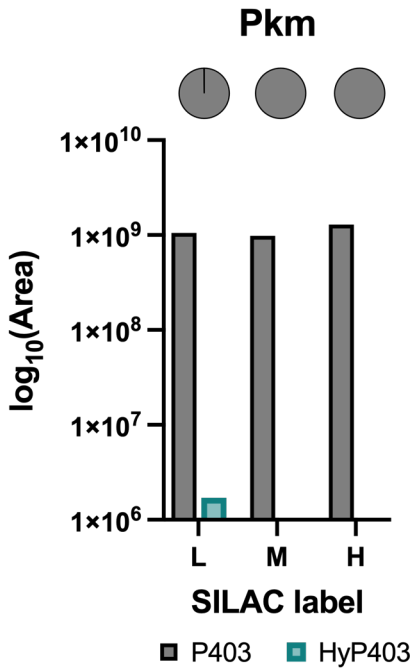


Figure 35 | SILAC data does not support prolyl hydroxylation of Pkm P403.

The area of all peptides containing Pkm P403 or HyP403 were summed and used to generate the stoichiometry of HyP403 in each experimental condition. Pie charts above each SILAC label indicate the stoichiometry of the hydroxylation site. L = vehicle-treated WT. M = vehicle-treated TKO. H = DMOG-treated WT.

Overall, genetic inhibition of the PHDs and pharmacological intervention targeting pan-2OGD hydroxylase activity revealed 3 novel sites of DMOG-sensitive prolyl hydroxylation on: Fkbp10 HyP35, Os9 HyP659, and Serpinh1 HyP29. The data do not exclude the possibility that a fourth class of substrate encompassing members of the PDI family of proteins are modified by an uncharacterised 2OGD. In particular, the DMOG intervention was insufficient to fully elucidate this effect on long half-life proteins, which was exemplified by RPS23 HyP62.

From a technical perspective, increasing the off-line fractionation resulted in improved proteome coverage compared to the U87MG hypoxia SILAC experiment and enabled detection of endogenous HIFα protein. However, it was still insufficient to detect the endogenous hydroxyproline-containing HIFα peptides.

4.3.3 Investigation of PDI hydroxylation in a rapidly dividing cell line

The results of the previous experiments indicate that the sensitivity of long half-life proteins to hypoxia and other interventions might be difficult to assess because the pre-existing population of hydroxylated proteoforms provide high background signal. To investigate this hypothesis, I devised a multifaceted strategy to determine factors that might regulate prolyl hydroxylation of long half-life proteins. PDIs were selected as the targets of the analysis because their high abundance makes it possible to detect these proteins in unfractionated samples. This reduced the time and cost of LC-MS/MS experiments.

The sequences flanking the hydroxyproline sites of PDIs could represent a novel hydroxylation motif that has not been assigned to substrates of any given enzyme to date. Moreover, the sensitivity of PDI hydroxyproline sites to any intervention, including hypoxia and DMOG, has not been described. I therefore considered experimental conditions that would determine if prolyl hydroxylation of PDIs is an oxygen-sensitive modification (incubation at 1 % O_2) and if the hydroxylase that catalyses the modification is a member of the 2OGD family (DMOG treatment).

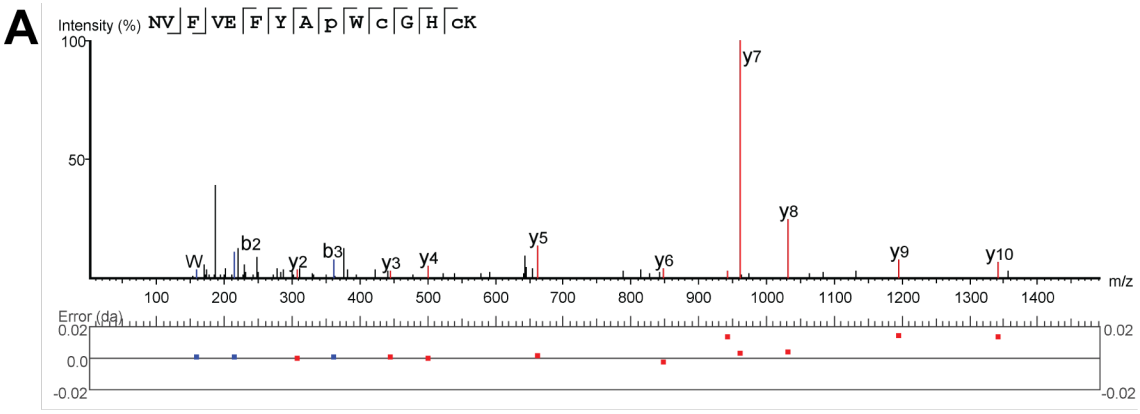
The PDIs are long half-life proteins so two components of the experimental setup were considered to facilitate the demonstration or otherwise of prolyl hydroxylation suppression in response to hypoxia or DMOG. Cells that exhibit a high proliferation rate must perform protein synthesis at a rate that facilitates cell growth and division. Under hydroxylase-inhibitory conditions, a high proliferation rate would be expected to dilute a pre-existing pool of hydroxylated proteins faster than the cells used in the previous experiments. For this reason, eHAP cells were selected, which were derived from KBM-7 cells that can maintain an exponential doubling time of as little as 12 hours depending on the cell culture conditions (Kotecki *et al.*, 1999). A long duration of the inhibitory conditions would also be hypothesised to increase the dynamic range of the experiment. Treatments were performed for 36 hours, instead of the 24 hours performed for the previous SILAC experiments. Additionally, cells were placed at 1 % O_2 for 72 hours to monitor PDI prolyl hydroxylation at a more prolonged period of hypoxia.

4.3.3.1 Synthetic peptides of PDI target sites

Peptide standards corresponding to the unmodified and hydroxyproline forms of P4HB P395, PDIA4 P89, and PDIA3 P404 were analysed to determine the *m/z* and retention time (RT) of these tryptic fragments (data not shown). These were used to predict the RT of the peptides derived from biological samples. The synthetic peptides were also used to validate the MS2 assignments the peptides derived from eHAP cells in Section 4.3.3.2.

4.3.3.2 Regulation of PDI hydroxylation by DMOG and hypoxia

The raw LC-MS/MS data of the biological samples was processed in PEAKS® Xpro. The target peptides were not assigned by PEAKS® in every condition, which generated missing values when calculating stoichiometry. However, when the peptides were identified in PEAKS® the MS2 assignments were consistent with those of the synthetic peptides, which was exemplified by the peptides pertaining to P4HB HyP395 (Figure 36). Importantly, the peptides are confidently sequenced in the y-ion series and the MS2 spectra of the endogenous peptide closely resembled that of the synthetic peptide. Additionally, the relatively high abundance of the diagnostic ions (y6 and y7) enabled confident PTM localisation. Finally, manual inspection for the relevant neutral loss ions did not provide evidence of misassigned oxidation at other residues, such as W396 or C397 (Lioe *et al.*, 2004; Steen and Mann, 2001). These factors increased the confidence of the assignments of the target peptides that were identified within the expected retention time (ERT) window (ERT ± 30 seconds) predicted by the synthetic peptides.



C

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	87.0558	115.0508	N				15
2	72.0812	214.1179	V	1815.7981	1797.7876	1798.7711	14
3	120.0812	361.1866	F	1716.7297	1698.7192	1699.7028	13
4	72.0812	460.2560	V	1569.6613	1551.6508	1552.6343	12
5	102.0554	589.2986	E	1470.5929	1452.5824	1453.5659	11
6	120.0812	736.3669	F	1341.5366	1323.5398	1324.5233	10
7	136.0761	899.4303	Y	1194.4668	1176.4714	1177.4550	9
8	44.0499	970.4674	A	1031.4138	1013.4080	1014.3916	8
9	86.0605	1083.5150	P(+15.99)	960.3776	942.3571	943.3545	7
10	159.0913	1269.5944	W	847.3367	829.3232	830.3068	6
11	133.0435	1429.6250	C(+57.02)	661.2525	643.2439	644.2275	5
12	30.0343	1486.6465	G	501.2242	483.2133	484.1968	4
13	110.0717	1623.7054	H	444.2011	426.1918	427.1754	3
14	133.0435	1783.7361	C(+57.02)	307.1432	289.1329	290.1165	2
15	101.1078		K	147.1128	129.1022	130.0858	1

D

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	87.0558	115.0508	N				15
2	72.0787	214.1183	V	1815.7981	1797.7876	1798.7711	14
3	120.0804	361.1868	F	1716.7242	1698.7192	1699.7028	13
4	72.0812	460.2552	V	1569.6484	1551.6508	1552.6343	12
5	102.0554	589.2975	E	1470.5811	1452.6002	1453.5854	11
6	120.0812	736.3567	F	1341.5474	1323.5398	1324.5233	10
7	136.0754	899.4214	Y	1194.4758	1176.4714	1177.4550	9
8	44.0499	970.4674	A	1031.4180	1013.4046	1014.4014	8
9	86.0597	1083.5150	P(+15.99)	960.3818	942.3718	943.3640	7
10	159.0911	1269.5944	W	847.3324	829.3232	830.3068	6
11	133.0424	1429.6250	C(+57.02)	661.2537	643.2452	644.2328	5
12	30.0343	1486.6465	G	501.2236	483.2142	484.1968	4
13	110.0707	1623.7054	H	444.2030	426.1871	427.1754	3
14	133.0435	1783.7361	C(+57.02)	307.1432	289.1328	290.1169	2
15	101.1078		K	147.1127	129.1019	130.0858	1

Figure 36 | Representative MS2 spectra of endogenous and synthetic tryptic peptides containing P4HB HyP395.

A and C) MS2 spectrum and ion table, respectively, of the endogenous tryptic P4HB HyP395-containing peptide. B and D) MS2 spectrum and ion table, respectively, of the synthetic tryptic P4HB HyP395-containing peptide. Confident peptide sequencing is achieved in the y-ion series. The diagnostic ions (y6 and y7) confirm the presence of hydroxyproline at the target site.

The missing values increased the difficulty of the analysis, so the RT and *m/z* of the precursor ions identified by PEAKS® were used to generate extracted ion chromatograms (XICs) of the target peptide, which is similar to previous analyses performed by this research group (Cockman *et al.*, 2019). The monoisotopic mass of the triply charged precursor ions were specified at ± 5.0 ppm error tolerance. Unmodified (+0.00 Da) and

oxidised (+15.99 Da) PDIA3 P404 peptides were identified in each sample within the ERT window (Figure 37).

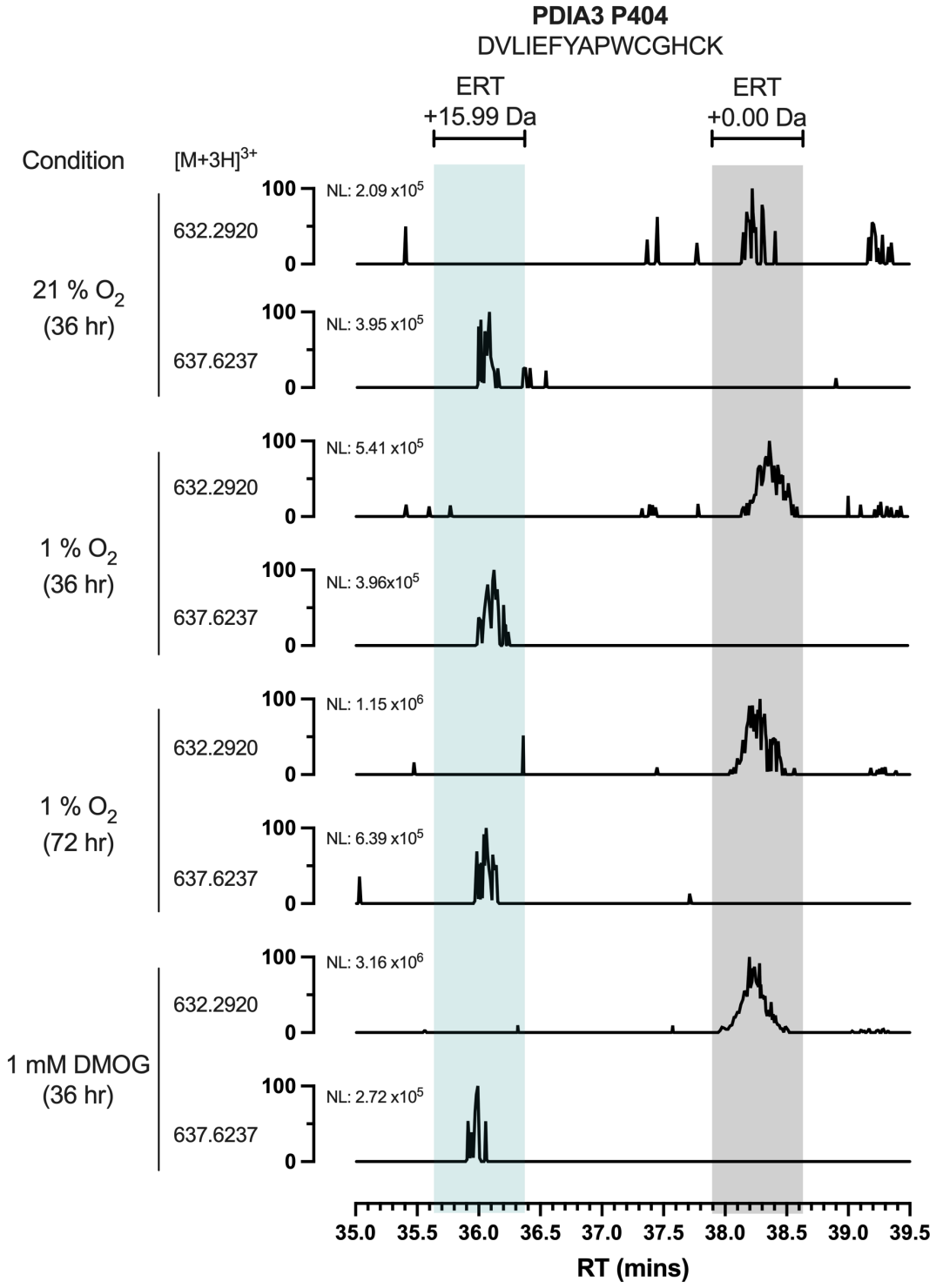


Figure 37 | Oxidation of PDIA3 aa396-410 in response to hypoxia or DMOG treatment.

Extracted ion chromatograms of the triply charged parent ion containing PDIA3 aa396-410. The non-oxidised (+0.00 Da; $[M + 3H]^{3+}$ = 632.2920) and oxidised (+15.99 Da; $[M + 3H]^{3+}$ = 637.6237) parent ions were identified at the estimated retention time (ERT) predicted by unmodified and hydroxyproline-containing synthetic peptide standards. The y-axis indicates intensity of precursor ions relative to the base peak, which is the most intense ion in the chromatogram window. The intensity of the base peak is indicated by the normalisation level (NL). Experimental conditions are indicated to the left of the chromatograms.

The area of each peak in the XICs was calculated and used to determine the abundance of the precursor ions. This subsequently enabled the stoichiometry of oxidised peptide to be calculated. This analysis was repeated for all PDI sites identified in this dataset (Figure 38). For each of the PDI sites, the abundance of the unmodified peptide increased with hydroxylation-restrictive conditions. The abundance of the oxidised peptide did not exhibit large changes in either hypoxic condition but decreased in DMOG-treated cells. These results demonstrate PDI prolyl hydroxylation is regulated by hypoxia or DMOG treatment, which suggests these proteins are substrates of a 2OGD enzyme.

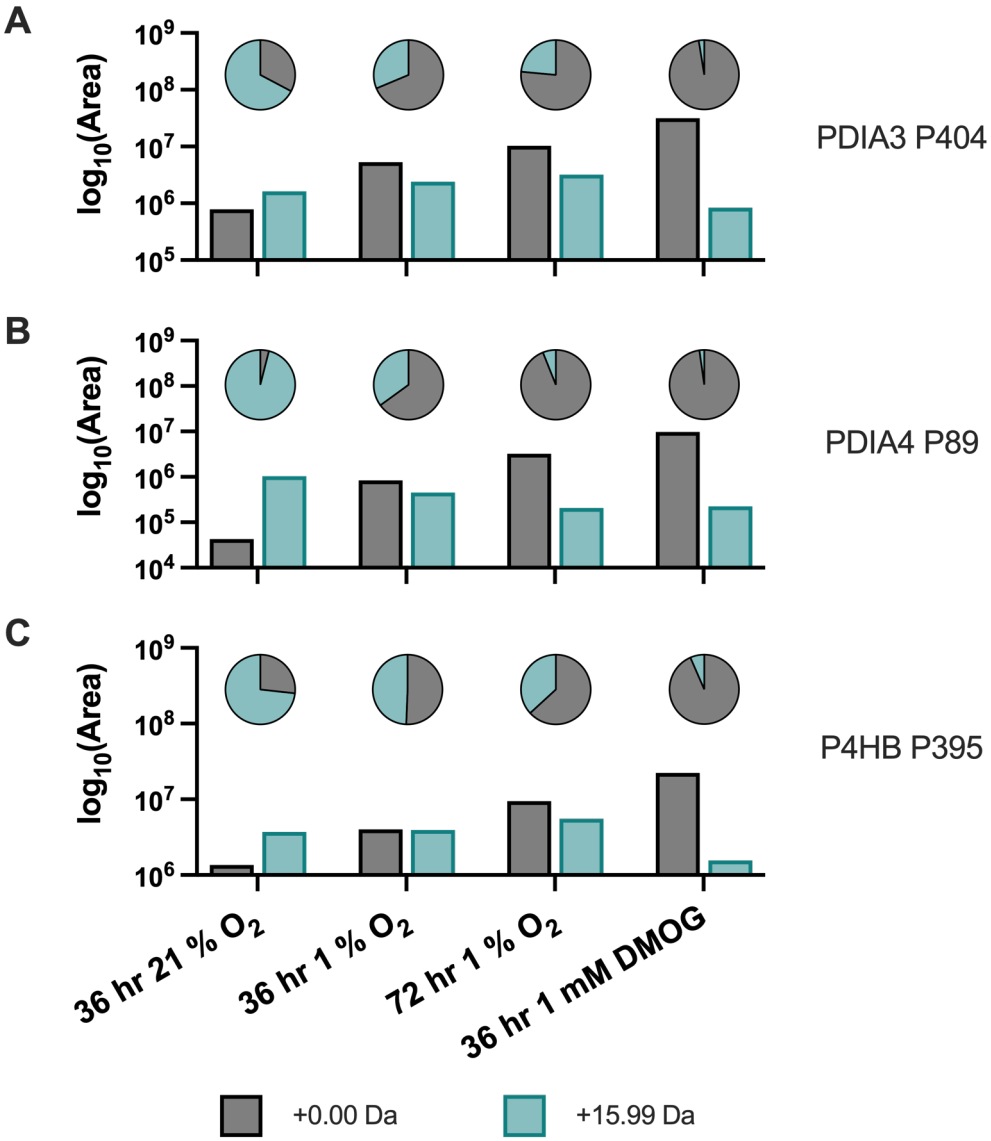


Figure 38 | Regulation of peptides containing putative prolyl hydroxylation sites of PDI proteins in response to hypoxia or DMOG treatment.

XICs were generated for each of the indicated peptides, as described in Figure 37. The $\log_{10}$(Area) of each peak in the XICs (bars) was used to calculate the stoichiometry (pie charts) of non-oxidised (+0.00 Da) and oxidised (+15.99 Da) peptides for each experimental condition.

4.4 Discussion

Three experiments were performed to identify sites of prolyl hydroxylation that are sensitive to hypoxia and DMOG, which would be consistent with substrates of 2OGDs. TKO MEFs were used to specifically consider prolyl hydroxylation catalysed by the HIFα prolyl hydroxylases. SILAC was used to label different experimental groups in the

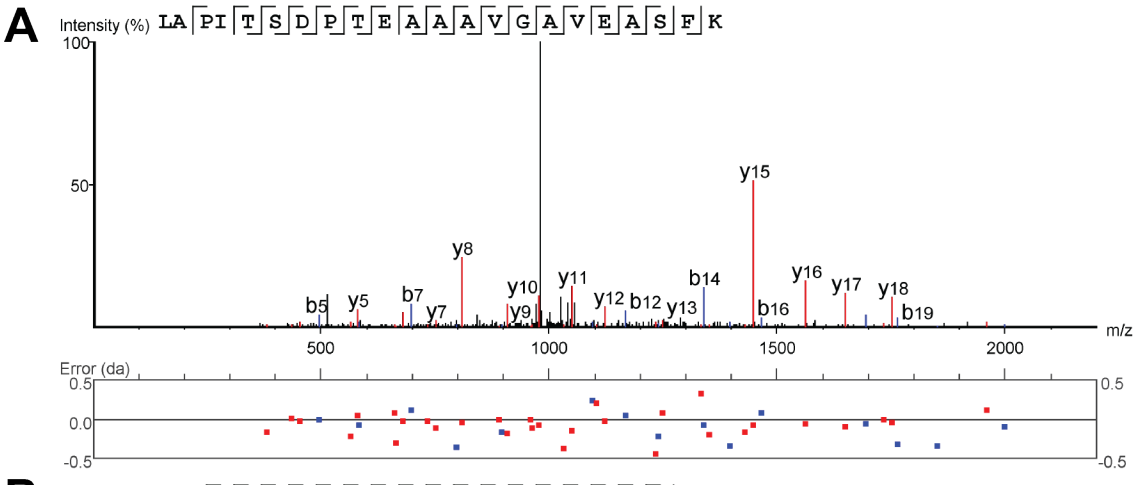
high-depth proteome-wide screens, and a label-free quantitation method was used to analyse PDI hydroxylation in a low-depth proteomic screen.

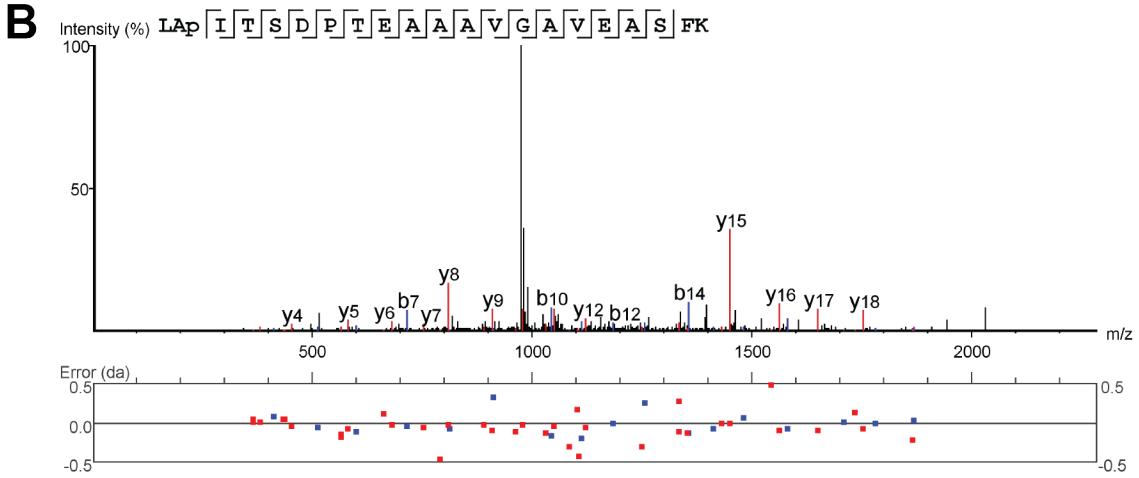
4.4.1 Weak evidence supporting PHD-catalysed prolyl hydroxylation

The hypoxia SILAC experiment aimed to identify PHD substrates through increased protein abundance at 1 % and 0.1 % O_2 compared to 21 %. HIFα was not identified so it is not clear whether relatively low abundance proteins manifesting similar regulation might have remained undetected in these experiments. Additionally, no sites of confidently localised prolyl hydroxylation were suppressed at 1 % O_2, which suggests that targets of PHD-catalysed hydroxylation are limited or of low abundance in the proteome.

The second SILAC experiment compared a PHD TKO cell line and a DMOG-treated WT cell line against vehicle-treated WT cells. Proteasome blockade was performed to stabilise proteins that might undergo hydroxylation-mediated proteolysis. No confidently localised sites of prolyl hydroxylation displayed complete suppression in the TKO cells, suggesting a lack of genuine PHD substrates. However, HIFα peptides representing sites of known PHD-catalysed hydroxylation were not identified in this analysis, allowing the possibility that such proteins might exist but were below thresholds for detection in these experiments.

Of the reported non-HIFα PHD substrates, only pyruvate kinase PKM (Pkm) displayed prolyl hydroxylation at a previously reported target site (HyP403) (Luo *et al.*, 2011). Additionally, the hydroxylated peptide was not detected in the TKO or DMOG-treated samples, which is consistent with PHD catalysed hydroxylation. However, it should be noted that the oxidation site was low stoichiometry (0.2 %) in the vehicle-treated WT sample and the hydroxylation could not be confidently localised by manual inspection of the MS2 spectra (Figure 39). The flanking sequence of P403 does not contain residues that are prone to oxidation, as was displayed in Figure 4, which most likely rules out misassigned oxidation of a nearby residue. Therefore, although there is ambiguity regarding the hydroxylation assignment, the data presented here does not provide an alternative explanation for the modified peptide.





C

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	86.0969	114.0919	L				22
2	44.0499	185.1290	A	2032.0232	2014.0127	2014.9962	21
3	70.0656	282.1817	P	1960.8621	1942.9756	1943.9591	20
4	86.0969	395.2658	I	1863.9333	1845.9229	1846.9064	19
5	74.0605	496.3070	T	1750.8920	1732.8459	1733.8224	18
6	60.0449	583.4310	S	1649.8900	1631.7911	1632.7747	17
7	88.0398	698.2350	D	1562.8361	1544.7592	1545.7427	16
8	70.0656	795.7920	P	1447.8140	1429.8940	1430.7157	15
9	74.0605	896.6450	T	1350.8870	1332.3400	1333.6630	14
10	102.0554	1025.5155	E	1249.5450	1232.0850	1232.6152	13
11	44.0499	1096.2900	A	1120.6260	1102.5891	1103.3459	12
12	44.0499	1167.5310	A	1049.7100	1031.5520	1032.9130	11
13	44.0499	1238.8540	A	978.6010	960.5250	961.6110	10
14	72.0812	1337.7810	V	907.6640	889.4777	890.4510	9
15	30.0343	1395.0649	G	808.4630	790.4094	791.3929	8
16	44.0499	1465.6570	A	751.5100	733.4180	734.3715	7
17	72.0812	1564.8223	V	680.3830	662.2500	663.6380	6
18	102.0554	1693.9250	E	581.2320	563.5050	564.2659	5
19	44.0499	1765.2200	A	452.2690	434.2060	435.2233	4
20	60.0449	1852.2700	S	381.3760	363.2026	364.1862	3
21	120.0812	1999.1050	F	294.1812	276.1706	277.1542	2
22	101.1078		K	147.1128	129.1022	130.0858	1

D

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	86.0969	114.0919	L				22
2	44.0499	185.1290	A	2048.0183	2030.0077	2030.9912	21
3	86.0605	298.1766	P(+15.99)	1976.9811	1958.9706	1959.9541	20
4	86.0969	411.1680	I	1864.1500	1845.9229	1846.9064	19
5	74.0605	512.3660	T	1750.9360	1732.6890	1733.8224	18
6	60.0449	599.4590	S	1649.8950	1631.7911	1632.7747	17
7	88.0398	714.4060	D	1562.8730	1544.2640	1545.7427	16
8	70.0656	811.4960	P	1447.7500	1429.7400	1430.7157	15
9	74.0605	912.1250	T	1350.8199	1332.3920	1333.7760	14
10	102.0554	1041.6790	E	1249.9470	1231.6317	1232.6152	13
11	44.0499	1112.7469	A	1120.6689	1102.3970	1103.9990	12
12	44.0499	1183.5811	A	1049.6140	1031.6790	1032.5355	11
13	44.0499	1254.3521	A	978.5530	960.5148	961.6080	10
14	72.0812	1353.8280	V	907.5840	889.5000	890.4613	9
15	30.0343	1410.7860	G	808.4470	790.8810	791.3929	8
16	44.0499	1481.6750	A	751.4610	733.3879	734.3715	7
17	72.0812	1580.9020	V	680.3860	662.2270	663.3344	6
18	102.0554	1709.8370	E	581.3780	563.4240	564.4430	5
19	44.0499	1780.8920	A	452.3000	434.1720	435.1540	4
20	60.0449	1867.8879	S	381.1870	363.1690	364.1220	3
21	120.0812	2014.9972	F	294.1812	276.1706	277.1542	2
22	101.1078		K	147.1128	129.1022	130.0858	1

Figure 39 | Representative MS2 spectra of Pkm HyP403.

Representative MS2 spectra and corresponding ion tables of tryptic fragments containing the target site, P403, from light-labelled cells. A) and C) MS2 spectrum of the non-hydroxylated peptide with corresponding fragment ion table. B) and D) MS2 spectrum of the hydroxylated peptide with corresponding fragment ion table. Both MS2 spectra convey confident peptide sequencing through both fragment ion series. The diagnostic fragment ions (b2 & b3, and y19 & y20) are not present, so the assignment of HyP403 cannot be corroborated.

Hydroxylation of Pkm P403 was observed as a low stoichiometry modification (0.2 %) that could not be localised in the MS2 spectrum. Additionally, the assignment did not meet the PTM filters (AScore ≥ 20 and ion intensity ≥ 5 %). This is considered to be weak evidence supporting the PTM assignment and non-enzymatic mechanisms might explain a relatively low abundance modification. Iron-catalysed protein oxidation can occur in cells (i.e., Fenton chemistry), whereby all amino acids are susceptible to hydroxyl radical (•OH) mediated oxidation (Xu and Chance, 2007). Additionally, •OH mediated oxidation can occur during electrospray ionisation (Boys et al., 2009; Morand et al., 1993). Given that Pkm is a highly abundant protein in the MEF proteome (Figure 40), it is highly likely that low levels of artefactual oxidation were detected for peptides on this protein.

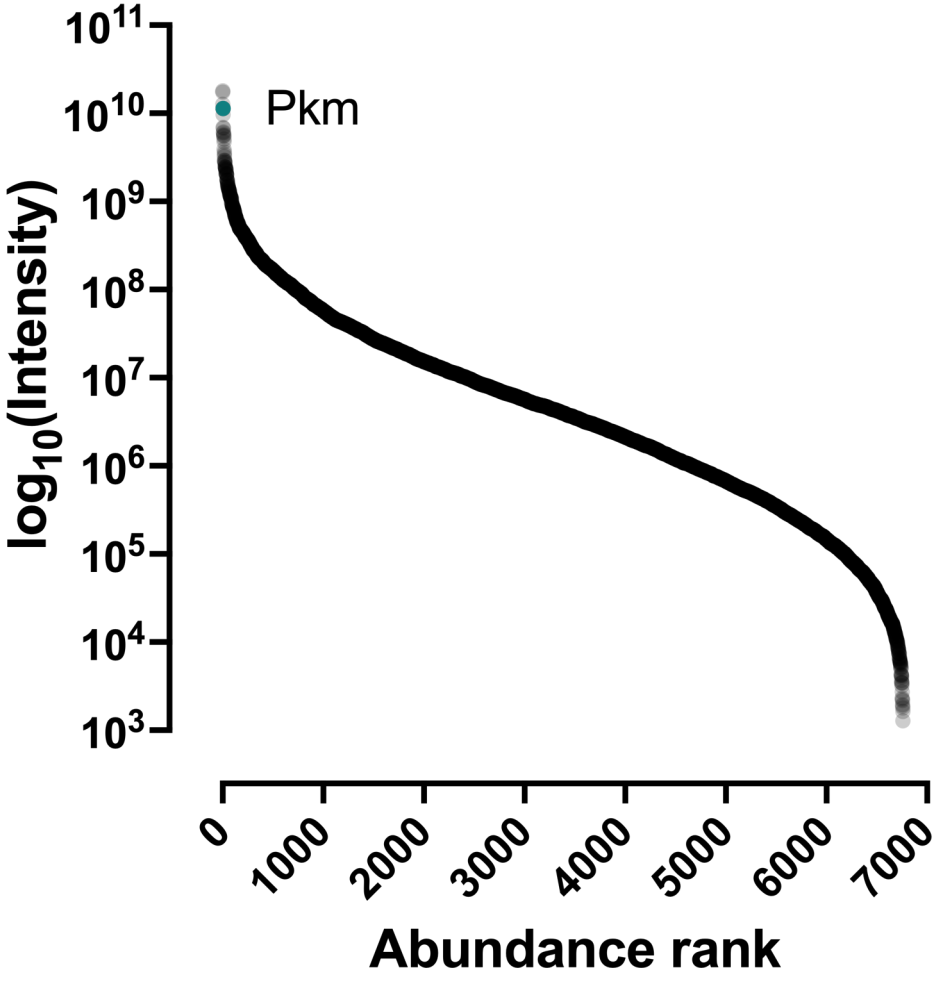


Figure 40 | Protein abundance.

Proteins were ranked according to their $\log_{10}$(Intensity) in vehicle treated MEFs (i.e., abundance rank). Pkm ranked as the 5th most abundant protein in the MEF proteome.

The current data was filtered to incorporate hydroxyproline sites with a minimum stoichiometry of 5 %. This filter was applied to remove hydroxyproline sites that cannot be distinguished from spurious oxidation based on relative intensity and thereby limit the number of false assignments by the software. The assignment of Pkm HyP403 is an example of a low stoichiometry, highly ambiguous hydroxyproline assignment. However, such sites on other proteins might be present as a consequence of incomplete degradation of the hydroxylated proteoform or represent another form of inefficient biological regulation. For this reason, a future analysis might consider analysing all hydroxyproline assignments, regardless of the stoichiometry.

4.4.2 Oxygen sensitivity of the HIFα and collagen prolyl hydroxylases.

When considering collagen, 35 of 88 hydroxyproline sites show complete suppression at 0.1 % O_2 but no change at 1 %. The K_m of C-P4Hα1 for oxygen is 40 µM (Hirsilä *et al.*, 2003). At 21 % O_2 (approximately 280 µM O_2) the hydroxylation reaction should proceed at maximum velocity. At 0.1 % O_2 (1.3 µM) the activity of collagen prolyl 4-hydroxylase (C-P4H) should be inhibited because the reaction only reaches 1 % of its maximum velocity (V_{max}). This was reflected in the data, whereby hydroxyproline stoichiometry of collagen target sites decreased at 0.1 % O_2. Despite the reduction, hydroxylation persisted for some collagen hydroxyproline sites. This is likely to be a consequence of the irreversible nature of prolyl hydroxylation and incomplete turnover of the pre-existing population of collagen proteins that had matured prior to hypoxic exposure.

Although at 1 % O_2 the oxygen concentration is below the K_m value for C-P4Hα1 (13 µM), the reaction can still proceed at reduced velocity (approximately 10 % of V_{max}). This would theoretically enable a substantial amount of hydroxylation to persist in the 24-hour exposure to low oxygen tension. Furthermore, HIF-dependent transcriptional upregulation of the *P4HA1* and *P4HA2* genes most likely compensate to maintain efficient collagen hydroxylation by increasing protein levels of C-P4H. In agreement with this, the stoichiometry of some collagen hydroxyproline sites actually increased at 1 % O_2 (e.g., COL1A1 HyP178 and COL6A1 HyP484). Together, it appears that collagen prolyl hydroxylation persists at 1 % O_2 and other substrates of C-P4H would be expected to display similar regulation.

It is difficult to relate K_m values, derived from *in vitro* experiments, to *in cellulo* reactions, whereby multiple factors influence hydroxylation. Some tissues exist at low oxygen concentrations and yet the proteolytic regulation of HIFα remains intact (e.g., bone marrow) (Pollard and Kranc, 2010). This physiological hypoxia suggests that there are other factors in the regulation of oxygen-sensitive hydroxylation in addition to the apparent affinity of the hydroxylases for oxygen. Despite this, the apparent affinity for oxygen and compensatory upregulation of *P4HA1* and *P4HA2* is consistent with persistence of collagen hydroxylation at most sites at 1 % O_2.

In contrast to C-P4H, the affinity of the HIFα hydroxylases for oxygen are approximately 230 μM (Hirsilä *et al.*, 2003), so PHD-catalysed prolyl hydroxylation should display similar suppression in both hypoxic conditions (i.e., < 0.5 % of V_{max} at 1 % O_2). This is reflected in the western blot, in which HIF1α and HIF2α were stabilised in both hypoxic conditions and strengthens the previous argument that PHD substrates should have been identified by significant regulation at 1 % O_2.

4.4.3 Strong evidence supporting the assignment of novel 2OGD substrates

Three non-collagen sites of prolyl hydroxylation (FKBP10 HyP36, LMAN1 HyP378, and SERPINH1 HyP30) displayed sensitivity to 0.1 % O_2 and DMOG, which is consistent with 2OGD-catalysed modification. These sites of prolyl hydroxylation had been observed previously (Chapter 3) and these assays provided the first direct evidence for involvement of a 2OGD enzyme.

OS9 HyP654 was also identified as a DMOG-sensitive site of prolyl hydroxylation but the peptide was not identified in the hypoxia SILAC screen. This observation was particularly interesting because it has been reported to interact with HIF1α and promote its degradation (Baek et al., 2005). Additionally, there has been a description of OS9 interacting with transiently expressed PHD3 in a DMOG-sensitive manner (Rodriguez *et al.*, 2016), which suggests PHD-catalysed prolyl hydroxylation of OS9. However, the interaction between OS9 and PHD has been contested (Brockmeier et al., 2011). Although transiently overexpressed OS9 and PHD2 interacted in immunoprecipitation experiments, there was no evidence for these proteins occurring in the same cellular compartments or interacting in a Förster resonance energy transfer (FRET)-based *in vivo* interaction assay (Brockmeier *et al.*, 2011). The data generated in these assays indicates

prolyl hydroxylation of OS9 occurs in a DMOG-sensitive manner independently of PHD activity, and hydroxylation occurs at a "P-G" site. This site was also identified in the analysis of publicly deposited proteomic data (Chapter 3), suggesting the modification is conserved, at least in mouse and human cells. Together with endoplasmic reticulum (ER) localisation of OS9, the data presented here suggests OS9 P654 is a substrate of an ER-resident (i.e., non-PHD) prolyl hydroxylase in the 2OGD family.

MESD HyP38 and LMAN1 HyP37 were also sensitive to 0.1 % O_2. Prolyl hydroxylation of LMAN1 HyP378, SERPINH1 HyP30, MESD HyP38 and OS9 P654 occurs at "P-G" sites. C-P4H is known to catalyse prolyl hydroxylation in the "Yaa" position of repeating "Xaa-Yaa-Gly" triplets (Gorres and Raines, 2010) and has been reported to catalyse hydroxylation of "P-G" sites in other proteins (Ono *et al.*, 2009; Rhoads and Udenfriend, 1969). In addition to the primary sequence, the low sensitivity to 1 % O_2 but high sensitivity to 0.1 %, and the subcellular localisation of these proteins (ER residents) suggests they may be substrates of C-P4H.

Prolyl hydroxylation of PDIs was described in the analysis of publicly deposited data (Chapter 3) and likely represents a novel hydroxylation consensus sequence. Hydroxyproline was identified at the target sites in the second SILAC screen and the modification persisted in TKO cells, suggesting PDIs are not PHD1-3 substrates. DMOG treatment generated an apparent decrease to hydroxyproline stoichiometry of all the PDI target sites detected in this analysis, but this was within the error tolerance limits (± 9.9 %) of the experiment.

Upon learning PDI proteins exhibit long half-life that might have precluded the short-term intervention with DMOG having a discernible effect on hydroxylation status, an experiment was performed to increase the duration of treatments and dilute the signal contribution of the pre-existing population of hydroxylated proteoforms. Together with the earlier experiments, these results demonstrated oxygen- and DMOG-sensitivity of several hydroxyproline sites in the proteome. This implicates these proteins as substrates of 2OGDs and could lead to further discoveries concerning the role of 2OGDs in adaptive responses to oxygen concentration. From a technical perspective, these results also illustrated how experimental conditions (including the choice of cell line and duration of inhibitor treatment) might be adjusted according to the half-life of the target proteins in order to reveal enzymatic hydroxylation.

The experiments here clearly identified several proteins as novel substrates of 2OGDs. Broad spectrum inhibition of prolyl hydroxylation (e.g., DMOG and hypoxia treatment) enabled these targets to be identified but did not provide direct evidence of the specific enzyme(s) responsible for catalysis. Genetic intervention would have enabled PHD substrates to be identified if such peptides were present within the limits of detection of the assay. To identify the enzyme(s) responsible for site-specific prolyl hydroxylation of the candidates identified in these analyses, these experiments could be repeated within a CRISPR library screen. Alternatively, enzyme-specific inhibitors (e.g., the HIF prolyl hydroxylase inhibitor Roxadustat), could be applied.

4.4.4 Interventions to study an irreversible PTM on long half-life proteins

The analysis of PDIs highlighted how protein half-life can impact the identification of 2OGD substrates following hydroxylase inhibition. Although eHAP cells did not display any visible signs of toxicity after 36 or 72 hours of treatment, it is possible that the cells were stressed and hydroxylation could have been affected non-specifically. Nevertheless, the reduction by both hypoxia and DMOG, and broad compatibility of the level of reduction of hydroxylation with knowledge of the half-life of the proteins and growth of cells strongly argue that PDIs are substrates of prolyl hydroxylases that belong to the 2OGD family.

In the experiments performed in this chapter, SILAC labels were used to compare hydroxylation-permissive and -restrictive conditions. An alternative strategy would be a pulse-chase experiment in which the SILAC label is altered following a change in cell culture conditions. In such an experiment, newly synthesised proteins would be distinguished from the pre-existing pool of hydroxylated proteins. This would enable prolyl hydroxylation of long half-life proteins to be determined without a requirement for prolonged treatments that could cause cell stress and toxicity.

A SILAC pulse-chase experiment has been performed to compare protein stability and gene expression changes following DMOG treatment (Stoehr *et al.*, 2016). Importantly, DMOG treatment was performed for 2, 6, and 18 hours, which minimised the potential toxicity of DMOG treatment. Although the authors report an inhibitory effect of DMOG on the totality of prolyl hydroxylation, they do not report site-specific regulation after

DMOG treatment. I therefore analysed the supplementary data attached to the publication and identified several hydroxyproline-containing peptides that are suppressed in the presence of DMOG and are relevant to this thesis (Table 38). The peptide abundance was inferred by the number of PSMs assigned to a single peptide (i.e., spectral counting). The MS1 and MS2 spectra for these peptides were not available so the analysis did not take into account these factors. Interestingly, DMOG treatment led to complete suppression of all PDI sites identified in their data. This is consistent with the data presented in this chapter and further strengthens the hypothesis that PDI proteins are 2OGD substrates.

An alternative method to study protein hydroxylation over time would be the application of biorthogonal noncanonical amino acid tagging (BONCAT) (Bagert et al., 2014; Dieterich et al., 2006). This method introduces azidohomoalanine (AHA) to newly synthesised proteins in place of the initiator methionine. This method therefore enables enrichment of nascent proteins by affinity purification targeting AHA, which would enable comparison of proteins synthesised in hydroxylase-permissive or restrictive conditions.

Table 38 | Hydroxyproline-containing peptides suppressed by DMOG.

Abundance changes of hydroxyproline-containing peptides between vehicle and DMOG treated cells were determined by spectral counting. The table presents a subset of peptides identified by Stoehr *et al.* (2016).

Protein	Site	Effect of DMOG
P4HB	P51	Complete suppression
P4HB	P395	Complete suppression
PDIA3	P55	Complete suppression
PDIA3	P404*	Complete suppression
PDIA4	P204	Complete suppression
SERPINH1	P30	Complete suppression

*The tryptic peptide sequence of PDIA3 P404 is identical to that of PDIA4 P553.

4.4.5 Improving total proteome depth

SILAC requires multiplexing and this generates 3 precursor ions per peptide ID, which dilutes the signal intensity of peptides derived from low abundance proteins and is likely to limit proteome depth.

The U-87 MG hypoxia SILAC screen identified a total of 55 confidently assigned hydroxyproline containing peptides. It was hypothesised that increased proteome depth would generate more assignments and increase the discovery potential of the experiment. The total number of confidently assigned hydroxyproline peptides in the second SILAC screen, 93, did not increase significantly despite requiring 20 hours more instrument time. This most likely reflects a generally low number of hydroxyproline sites in the proteome.

Given the improvement to proteome depth achieved by more extensive fractionation, I postulated a different strategy might lead to further discovery of prolyl hydroxylation. The samples used in the MEF SILAC experiment were also analysed using a slightly different strategy in an attempt to increase the total proteome depth. The 30 fractions generated by high pH off-line fraction were concatenated (e.g., F1 was combined with F16 (F1:F16), F2:F17, etc.). These fractions were then subjected to a 90-minute acetonitrile gradient (instead of 60 minutes) and precursor ion fragmentation was performed in higher energy C-trap dissociation (HCD) mode to increase the resolution of fragment ions. Despite these differences, the data was > 99 % identical to the assay described in this chapter and no additional high confidence hydroxyproline sites were identified. Therefore, it is not clear if a more extensive off-line fractionation protocol would generate many more confident assignments.

The endogenous hydroxylated HIFα peptides remained beyond the limits of detection of these experiments, so it is possible there is a population of low abundance hydroxyproline sites that remains to be revealed. During the course of this work the Proteomics science technology platform (STP) procured a Bruker™ trapped ion mobility separation time of flight (timsTOF) instrument, with ion mobility separation. It is possible the fourth dimension of peptide resolution increases the duty cycle and would result in more complete proteome coverage. Alternatively, affinity purification may be necessary to increase the signal to noise ratio of hydroxyproline-containing peptides.

Chapter 5. Hydroxyproline affinity purification

5.1 Introduction

Mass spectrometry is the only direct method to perform high-throughput identification of site-specific PTMs in complex samples. In previous analyses there were few hydroxyproline sites detected, even in highly fractionated proteomic datasets. This was particularly problematic in multiplexed SILAC experiments, in which combining multiple cell lysates dilutes low abundance signals. In addition to low signal, proteomic software often assigns hydroxyproline incorrectly, which further increases the difficulty of identifying novel hydroxyproline sites. It was therefore proposed that an enrichment strategy that specifically targets hydroxyproline would increase the efficiency of detection by mass spectrometry and increase the confidence of correct assignment by the software.

5.1.1 Enrichment methods to identify PTMs in proteomic screens

An important distinction between enrichment methods is whether proteins or proteolytic fragments are targeted. Enrichment at the protein level provides the greatest possibility of identifying a protein substrate of the modifying enzyme because many proteotypic peptides are generated during downstream proteolysis, increasing the likelihood of identification. However, protein enrichment might occur *via* indirect interactions, so that not all captured proteins are direct substrates of the modifying enzyme. Additionally, this approach does not guarantee the peptide(s) containing the modified residue(s) will be identified, which could lead to a false negative assignment based on the absence of modification in the peptides that are assigned. On the other hand, peptide affinity enrichment provides a means of specific enrichment of the modified peptide, but the discovery potential is limited by the sensitivity of the analysis because the identity of each protein can only be deduced from peptide(s) containing modified residue(s).

Several strategies have been applied to enrich proteins and peptides containing a variety of PTMs. Antibody-based enrichment strategies and affinity chromatography have been most commonly used, whilst chemical tagging methods are used less often due to inefficient labelling. These are briefly described in the following sections.

5.1.1.1 Antibody-based affinity purification

Antibodies exist for most commonly studied PTMs and can be applied at the protein or peptide level. Examples include acetylation, methylation, and phosphorylation (Černý et al., 2013). Ideally, such antibodies should recognize the specified PTM irrespective of its peptidic context. Several such pan-hydroxyproline antibodies have been developed and a proteomic screen has been performed following antibody-based enrichment of hydroxyproline-containing peptides (Zhou *et al.*, 2016). The authors did not identify endogenous HIFα peptides in their data, so they used enrichment of collagen peptides as a positive control to demonstrate that their methods did indeed enrich for hydroxyproline-containing peptides. Nevertheless, enrichment of HIFα peptides is the gold standard, since this could then be used to infer successful enrichment of other putative PHD substrates that might be of similarly low abundance.

It might be possible to use the commercially available pan-hydroxyproline antibodies to enrich for sites of prolyl hydroxylation in the proteome, but validation of their specificity must be performed prior to proteomic screens. I have attempted to demonstrate specific binding of these antibodies to hydroxylated HIFα peptides, which was unsuccessful despite the conditions enabling the anti-HIF1α HyP564 antibody bind to its target (data not shown). I therefore concluded that it was unlikely that these antibodies exhibit the necessary binding activity to be considered suitable reagents for affinity purification of hydroxyproline-containing peptides.

5.1.1.2 Non-antibody-based affinity purification

Non-antibody-based affinity enrichment typically utilise the recognition domains of proteins that specifically bind to modified proteins. This has been performed with ubiquitin binding domains (UBDs) to enrich ubiquitin (Scott et al., 2015) and lectins to enrich glycosylated proteins (Jung et al., 2009). These are typically used at the protein level, but it could be possible to take this approach to enrich peptides. Several proteins are known to bind to prolyl hydroxylated proteins, such as von Hippel Lindau protein (pVHL) binding to HIFα (Jaakkola *et al.*, 2001) or integrin binding to collagen (Zhang et al., 2003), so it might be possible to apply this strategy to the enrichment of hydroxyproline-containing peptides. The application of pVHL to enrich for

hydroxyproline-containing peptides has been reported (Arsenault *et al.*, 2015). However, the authors did not observe affinity enrichment of hydroxyproline-containing HIFα peptides. Therefore, this strategy remains to be validated against a *bone fide* pVHL substrate.

5.1.1.3 Affinity chromatography

Affinity chromatography has been used to enrich for peptides containing various PTMs. For instance, immobilised metal affinity chromatography (IMAC), strong cation exchange (SCX) and hydrophilic interaction chromatography (HILIC) have all been used to enrich phosphopeptides (Macek et al., 2009). SCX, HILIC and isoelectric focusing (IEF) have been used to enrich peptides containing methylated arginine (Uhlmann et al., 2012). Hydroxyproline has a weak effect on the physicochemical properties of a peptide; the hydroxylation slightly increases the polarity of a peptide. Therefore, I concluded that it was unlikely that an affinity chromatography method of this type would sufficiently enrich hydroxyproline-containing peptides.

5.1.1.4 Chemical labelling

Chemical labelling approaches have been used to derivatise PTMs so that they can be affinity purified according to the introduced functional group (Chuh and Pratt, 2015). In *Dictyostelium discoideum* (Dd), DdSKP1 HyP143 is modified by the ligation of a pentasaccharide chain (Teng-Umnuay et al., 1999). I postulated that DdGnt1-catalysed modification of hydroxyproline-containing peptides would facilitate affinity enrichment targeting the novel functional group. My preliminary data with a range of synthetic hydroxyproline-containing peptides suggested that the DdGNT1 (i.e., the enzyme that catalyses ligation of GlcNAc to DdSKP1 HyP143) is highly specific to its substrate (data not shown). Therefore, I concluded that this approach was unlikely to facilitate proteome-wide enrichment of hydroxyproline sites.

In addition to glycosylation, it has been reported that hydroxyproline can serve as a substrate of phosphorylation catalysed by cyclic-AMP-dependent kinase (Feramisco et al., 1979). It might therefore be possible to derivatise hydroxyproline and perform an affinity purification method using established phosphopeptide enrichment techniques.

Overall, the main difficulties with chemical labelling strategies are overcoming low labelling efficiency and limiting off-target derivatisation. Although chemical derivatisation of hydroxyproline could be successful, this strategy would have required substantial method development and appeared to have a low chance of success.

5.1.2 A pVHL-based enrichment method to identify novel hydroxyproline sites

Recombinant pVHL, in combination with the adapter proteins Elongin B and Elongin C (VBC) has been shown to bind synthetic hydroxyproline containing HIFα peptides in a variety of assays, including peptide enrichment (Hon et al., 2002; Jaakkola *et al.*, 2001). This reagent was used to detect HIFα and collagen in far-western analyses (Grosfeld *et al.*, 2007). As described above, this reagent has been applied to enrich hydroxyproline-containing peptides for LC-MS/MS analysis (Arsenault *et al.*, 2015). Hydroxylated HIFα peptides were not detected in the pVHL-bound fractions in this investigation, which suggested the method could be improved.

There have been reports of 16 non-HIFα PHD substrates that interact with pVHL (Table 39). Several other proteins have also been reported to interact with pVHL but there is no evidence to suggest these proteins are subject to PHD-catalysed prolyl hydroxylation. Key interactors discussed in this thesis are provided in Table 39. For other pVHL interactors the reader is directed to VHLdb (http://vhldb.bio.unipd.it/; last accessed 13/04/22), which a database of pVHL interacting proteins (Tabaro et al., 2016). It should be noted that although the majority of pVHL interactors described in the literature have been reported on VHLdb, this is not an exhaustive database of pVHL interactors. For example, an interaction between pVHL and MYB has been reported (Okumura et al., 2016) but is not present in VHLdb. Although it is not clear if protein interactions are direct or artefacts of ectopic expression, the peptide studies suggest recombinant pVHL can bind to multiple hydroxyproline-containing peptides, which increases the discovery potential of using this putative hydroxyproline affinity reagent in proteome-wide screens.

Table 39 | Non-HIFα protein and peptide interactors of pVHL.

Proteins marked with an asterisk (*) have not been reported to undergo PHD-catalysed prolyl hydroxylation.

Protein	Enrichment level	Binding assay	Reference
ACTB	Peptide	Dot blot peptide binding assay	Zi *et al.* (2021)

ADRB2	Protein	Co-IP of transiently expressed ADRB2 and pVHL	Xie *et al.* (2009)
AKT	Protein Peptide	Co-IP of transiently expressed AKT and pVHL On-bead peptide binding assay	Guo *et al.* (2016)
AR	Protein	Co-IP of transiently expressed AR and pVHL	Wang et al. (2014)
BCL2L11	Protein Peptide	Co-IP of transiently expressed BCL2L11 and endogenous pVHL On-bead peptide binding assay	Li *et al.* (2019)
CERKL	Protein	Co-IP of transiently expressed CERKL and pVHL	Chen *et al.* (2015)
COL4A2*	Protein	Co-IP of endogenous COL4A2 and stably expressed pVHL	Grosfeld *et al.* (2007)
EPOR	Protein Peptide	Co-IP of transiently expressed EPOR and pVHL On-bead peptide binding assay Peptide binding assay by SPR	Heir *et al.* (2016)
FN1*	Protein	Co-IP of endogenous FN1 and stably expressed pVHL	Iwai et al. (1999); Ohh et al. (1998)
FLNA	Protein	Co-IP of transiently expressed FLNA and pVHL	Segura *et al.* (2016)
MAPK7	Protein	Co-IP of transiently expressed MAPK7 and pVHL	Arias-González *et al.* (2013)
NDRG3	Protein	Co-IP of transiently expressed NDRG3 and pVHL	Lee *et al.* (2015)
POLR2A	Peptide	On-bead peptide binding assay	Kuznetsova *et al.* (2003)
SFMBT1	Protein	VHL pull-down of IVT-produced SFMBT1 Co-IP of endogenous SFMBT1 and pVHL	Liu *et al.* (2020a)
SPRY2	Protein	Co-IP of transiently expressed SPRY2 and pVHL	Anderson *et al.* (2011)
TBK1	Protein	Co-IP of transiently expressed TBK1 and pVHL	Hu *et al.* (2020)

	Peptide	On-bead peptide binding assay	
TET2/3	Protein	Co-IP of transiently expressed TET2/3 and pVHL	Fan *et al.* (2020)
TUBB1-8*	Protein	Co-IP of endogenous TUBB1-8 and pVHL	Hergovich *et al.* (2003)
ZHX2	Protein	VHL pull-down of IVT-produced ZHX2 Co-IP of endogenous ZHX2 and transiently expressed pVHL	Zhang *et al.* (2018)

Overall, the successful application of recombinant pVHL in different hydroxyproline peptide binding assays and affinity purification experiments reported by others suggested this could be a suitable reagent to enrich *bona fide* prolyl hydroxylase substrates in proteomic screens. Methods were developed to perform pVHL-based affinity enrichment at the protein and peptide level to identify novel hydroxyproline sites in the proteome.

5.2 Methods

5.2.1 Recombinant protein expression

Recombinant GST-VHL (aa54-213), Elongin B (aa1-118) and Elongin C (aa17-112) trimeric complex (GST-VBC) was expressed and purified in a similar manner to previously reports (Hon *et al.*, 2002; Stebbins *et al.*, 1999). Briefly, *E. coli* BL21 (DE3) GOLD cells were transformed with pGEX-4T-3 *VHL* & *Elongin B* and pBB75 *Elongin C* and incubated in LB media supplemented with 100 µg/mL ampicillin and 50 µg/mL kanamycin. When the cells reached an optical density at 600 nm (OD_{600}) of 0.6, Isopropyl β-d1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM and incubated for 18 hours at 18 °C. Cells were retrieved from the media by centrifugation at 4,000 x *g* for 30 minutes. Cells were lysed by sonication in a solution containing protease inhibitor cocktail, and the soluble fraction was incubated with glutathione sepharose 4B beads. The beads were loaded into a pre-equilibrated gravity column and washed six times, each with 5 bed volumes (BV) of wash buffer. GST-VBC was eluted in 1 BV of wash buffer supplemented with 10 mM reduced glutathione and this step was repeated with 100 mM reduced glutathione. A post-elution wash step was performed with 1 BV of wash buffer to improve the recovery of GST-VBC. The beads were prepared in 1x

Laemmli buffer and all of the samples were analysed by SDS-PAGE to confirm recovery (Figure 41). The 10 mM, 100 mM elution samples and the post-elution sample were pooled and dialysed to remove glutathione. In total, 15 mL 6.5 mg/mL GST-VBC was recovered from a 12 L batch expression.

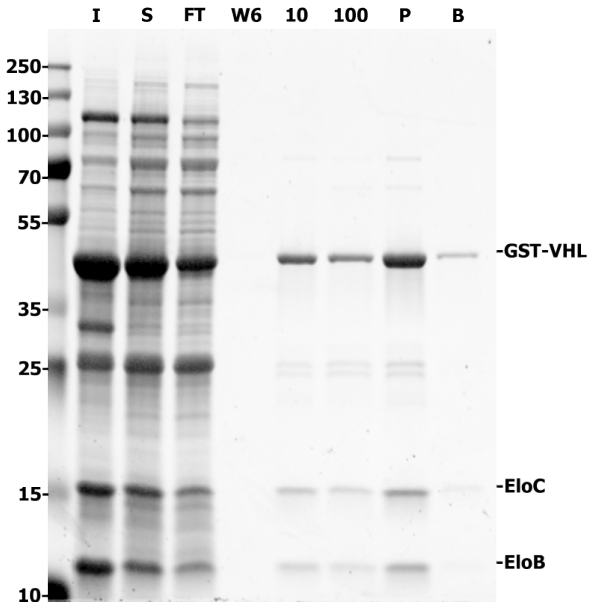


Figure 41 | Recombinant GST-VBC purification.

GST-VHL, Elongin B and Elongin C (GST-VBC) were co-expressed in *E. coli* by induction of the lac operon with 1 mM IPTG. Following cell harvest, the cells were lysed by sonication and the insoluble (I) fraction was pelleted by ultracentrifugation. The soluble (S) fraction was incubated with Glutathione sepharose 4B beads for 2 hours before being applied to a gravity flow column. The flow through (FT) was collected and beads were washed six times in 5 bed volumes (BVs) of Buffer A. The final wash (W6) fraction was collected and elution was performed in Buffer B + 10 mM reduced glutathione (10) followed by wash buffer + 100 mM reduced glutathione (100). The beads were washed in 1 BV of wash buffer to collect the post-elution (P) fraction. Finally, the beads (B) were prepared in Laemmli buffer to assess elution efficiency.

5.2.2 Protein affinity purification

To assess the fidelity of the expressed trimeric complex protein enrichment assays were performed. In the first instance, a western blot was performed to confirm HIF1α capture. In the second experiment a proteome-wide screen was performed to identify protein interactors whose binding was reduced after DMOG treatment. It was hypothesised these proteins would interact with pVHL directly *via* hydroxyproline residues.

5.2.2.1 Validation by western blot

RCC4 cells were seeded at 4 x10^6 in DMEM media supplemented with 10 % FBS on 150 mm^2 plates. The cells were incubated for 24 hours, at which point the media was replenished and the media was supplemented with vehicle (PBS) for 8 hours or 1 mM DMOG for 4 or 8 hours prior to harvest. Cells were washed in ice-cold PBS and lysed in lysis buffer (25 mM Tris HCl pH 7.4, 5 mM $MgCl_2$, 150 mM NaCl, 1 % (v/v) NP-40, 1x HALT™ protease inhibitor cocktail). Benzonase® was added to a final concentration of 250 U / mL and the samples were incubated at 4 °C for 1 hour with end-to-end rotation (15 rpm). The lysates were clarified by ultracentrifugation at 17,000 x *g* for 15 minutes and the soluble fraction was quantified using the Pierce™ 660 nm protein assay kit.
GST-VBC (150 µg) was immobilised on glutathione DynaBeads that had been pre-equilibrated in lysis buffer. The beads were split equally into three microcentrifuge tubes to perform affinity purification on 5 mg of each RCC4 lysate (i.e., 50 mg GST-VBC per 5 mg RCC4 lysate). The mixture was incubated for 18 hours at 4 °C with end-to-end rotation (30 rpm). The beads were washed for 1 minute in lysis buffer. This was repeated for a total of 4 washes.
The beads were resuspended in 100 µL 1x Laemmli buffer and proteins were resolved by 4 – 20 % SDS-PAGE. The western blot protocol described in 2.1.1 was used to visualise HIF1α protein enrichment.

5.2.2.2 Proteome-wide affinity pull down in response to DMOG treatment

To perform the proteome-wide protein enrichment assay the previous experiment was repeated with the following exceptions:

- RCC4 cells were treated with 1 mM DMOG for 16 hours prior to harvest.
- For the protein pull down, 100 µg GST-VBC was used with 5 mg of each RCC4 lysate.
- Quick Coomassie stain was applied immediately after the proteins were resolved by SDS-PAGE.

In total, 19 gel slices were collected for each lane (i.e., ± 1 mM DMOG). The gel slices were incubated in wash solution (50 % MeOH, 5 % CH_3COOH) for 18 hours at room temperature with gentle agitation. The was solution was replenished and incubated for 3

hours at room temperature. ACN was added for 5 minutes to dehydrate the gel pieces. This was repeated 3 times. Cysteines were reduced by the addition of 10 mM DTT for 30 minutes. Cysteines were then alkylated by incubating the gel pieces with 50 mM IAA for 30 minutes in the dark. ACN was added to dehydrate the gel pieces. Trypsin was dissolved in 50 mM NH_4CO_3 and added to each sample so that each gel slice was incubated with 2 µg trypsin for 18 hours at 37 °C with orbital rotation at 800 rpm. Peptides were extracted in 50 % ACN, 5 % CH_3COOH. A second extraction was performed in 85 % ACN, 5 % CH_3COOH.

This resulted in 19 fractions for each condition (± 1 mM DMOG) that were lyophilised by vacuum centrifugation to remove the organic content. The peptides were resuspended in 2 % ACN, 0.1 % FA and analysed by LC−MS/MS in an Orbitrap Fusion™ Lumos™ Tribrid™ instrument. A 28-minute ACN gradient (2-40 % buffer B) in 0.1% FA, as described in Table 40, was performed with a U3000 HPLC instrument with an Acclaim™ Pepmap™ 100 C18 HPLC column (5 µm x 0.1 mm x 20 mm) at a flow rate of 0.275 µL/min.

Table 40 | Linear gradient for on-line low pH separation of label-free peptides derived from the protein enrichment experiment performed on RCC4 extracts.

Peptide separation was performed on-line at low pH using buffer A (0.1 % FA) and buffer B (80 % ACN, 0.1 % FA). A linear ACN gradient (2-40% buffer B) was performed over 28 minutes at a flow rate of 0.275 µL/min.

Time (minutes)	Buffer B (%)
0	2
5	2
5.5	8
33	40
34	95
44	95
45	2
60	2

The settings for precursor ion selection and fragmentation are defined in Table 41.

Table 41 | Settings for label-free data-dependent acquisition in the protein enrichment experiment performed on RCC4 extracts.

Entry	Parameter	Setting
MS1	Detector type	Orbitrap
	Resolution	120,000
	Mass range	Normal
	Use quadrupole isolation	True
	Scan range (*m/z*)	350-1,500
	RF lens (%)	30
	AGC target	4×10^5
	Maximum injection time (ms)	50
	Microscans	1
	Data type	Profile
	Polarity	Positive
MIPS	Monoisotopic peak determination	Peptide
Intensity	Filter type, intensity threshold	9×10^3
Charge state	Include charge state(s)	2-6
Dynamic exclusion	Exclude after *n* times	1
	Exclusion duration (s)	15
	Mass tolerance	ppm
	Low	10
	High	10
	Exclude isotopes	Yes
MS2	Isolation mode	Quadrupole
	Isolation window (*m/z*)	1.2
	Activation type	CID
	collision energy (%)	35
	Detector type	IonTrap
	First mass (*m/z*)	110
	AGC target	1×10^4
	Maximum injection time (ms)	100
	Microscans	1
	Data type	Profile
Data dependent properties	Select: 'Cycle time'	3 s

The raw data was processed in PEAKS® using the parameters defined in Table 42.

Table 42 | Processing parameters of the protein affinity enrichment experiment performed on RCC4 extracts.

Parameter	Setting
PEAKS® version	X
PEAKS® search tool	DB
Reference proteome	Human, canonical (Uniprot id: UP000005640)
Precursor ion error tolerance	±10 ppm
Fragment ion error tolerance	±0.5 Da
Enzyme	Trypsin
Enzyme specificity	Specific
Maximum number of miscleavages per peptide	2
Fixed modification	Carbamidomethylation (+57.0214 Da; C)
Variable modifications	Oxidation (+15.9949 Da; CDFHKMNPRWY) Dioxidation (+31.9899 Da; FMWY) Deamidation (+0.9840 Da; NQ) Acetylation (+42.0367 Da; protein N-terminus)
Maximum number of PTMs per peptide	3

5.2.3 Surface plasmon resonance

Surface plasmon resonance (SPR) was performed to assess the binding kinetics of recombinant pVHL to synthetic tryptic peptides. The Twin-Strep-tag® is based on the streptavidin-biotin interaction and has a high affinity interaction (low pM range) with Srep-Tactin®XT resin (Schmidt et al., 2013). Importantly, binding is reversible, which facilitates its use in binding assays such as SPR whereby the sensor chip must be regenerated after each round of binding.

A modified pGEX-4T-3 *VHL* & *Elongin B* plasmid was procured from IDT-technologies™ that contained a Twin-Strep-tag® (TS) inserted between the *GST* and *VHL* ORFs (GST-TS-VHL). This enabled expression and purification of the trimeric complex (GST-TS-VBC) using the same protocols described in Section 5.2.1.

The Biacore™ CM5 sensor chip was coated with Strep-Tactin®XT using an amine coupling kit, following the manufacturer's instructions from the Twin-Strep-tag® capture kit.

The SPR experiments were performed on a Biacore™ S200 instrument using a running buffer comprised of 25 mM HEPES pH 7.5, 50 mM NaCl, 0.05 % (v/v) Surfactant P20. GST-TS-VBC capture was achieved with 1 μM of the recombinant trimer over 60 seconds with a flow rate of 30 μL/min, which achieved immobilisation levels of approximately 1,500 response units (RU). The efficacy of substrate binding was achieved by injection of 80 nM tryptic HIF1α HyP564 peptide for 60 seconds, which achieved RU_{max} = 25.3, which is within the 20-100 RU range recommended by the manufacturers. To regenerate the sensor chip between binding assays, three consecutive injections of 3 M GuHCl were each performed for 60 seconds.

Peptide binding assays were performed with 2-fold dilution series of tryptic HIF1α HyP564, AKT1 HyP313, and AKT1 HyP313 HyP318. The SPR experiments were performed in collaboration with Dr Simone Kunzelmann of the structural biology STP at the Francis Crick Institute.

5.2.4 Comparison of elution strategies

The purpose of comparing different elution strategies was to determine conditions that would liberate target peptides, minimise sample losses, and maintain the integrity of the recombinant GST-VBC complex on glutathione DynaBeads. For each experiment, 25 μg GST-VBC was immobilised on glutathione DynaBeads, unless otherwise stated. The solutions used in these experiments are indicated in the main text of the results section. Proteins were analysed by 4 – 20 % SDS-PAGE and visualised by Quick Coomassie stain.

5.2.5 Peptide affinity purification

RCC4 cells were washed in PBS and incubated with trypsin-versene until they detached from the flask. The cells were washed twice in PBS and pellets containing 1 – 5 x10^7 cells were stored in protein low-bind microcentrifuge tubes at -80 °C until required. The cell pellets were resuspended in 250 μL lysis buffer (6 M urea, 2 M thiourea, 40 mM Tris-HCl pH 8.0, 2 mM $MgCl_2$, 0.1 % (w/v) CHAPS, 20 mM DTT, 1x HALT™ protease

inhibitor cocktail, 250 U / mL Benzonase®) and incubated at 4 °C for 1 hour with end-to-end rotation. IAA was added to a final concentration of 50 mM and incubated in the dark for 30 minutes. The lysates were subsequently clarified by centrifugation at 17,000 x *g* for 30 minutes. The protein concentration of the soluble fraction was determined using the Pierce™ 660 nm protein assay kit.

In order to prepare tryptic peptides at scale for peptide enrichment assays, trypsin digestion was performed using the SP4 method (Johnston et al., 2021). Briefly, the protein lysates were incubated at a ratio of 10 µg 9-13 µm glass spheres per 1 µg lysate. ACN was added to a final concentration of 80 % and centrifuged at 16,000 x *g* for 5 minutes at 4 °C. The beads were washed 4 times, each for 30 seconds, in 80 % EtOH, with centrifugation at 16,000 x *g* for 2 minutes at 4 °C between each wash. The beads were resuspended in 100 mM NH_4CO_3, 50 mM HEPES pH 8.0. Trypsin was added at a ratio of 1 µg trypsin per 100 µg lysate and incubated for 18 hours at 37 °C with orbital rotation at 800 rpm. Peptides were retrieved by collecting the supernatant after centrifugation at 16,000 x *g* for 2 minutes. To increase the recovery of tryptic peptides, the beads were washed in 100 mM NH_4CO_3 and the centrifugation step was repeated. The samples were then desalted, using an Oasis HLB 96-well elution plate (30 mg sorbent per well) and the peptides were eluted in 80 % ACN, 0.1 % FA. The peptides were lyophilised by vacuum centrifugation to remove the organic content.

For the peptide affinity purification experiments, GST-VBC was incubated with the tryptic extract. Target peptides were liberated from the GST-VBC complex by performing on-bead trypsin digest. Further details specific to each of the three peptide enrichment experiments are provided below.

5.2.5.1 Experiment 1

The first experiment used 7 mg RCC4 lysate, digested with trypsin by performing the SP4 protocol. After digestion and desalting, peptides were resuspended in 25 mM HEPES pH 7.5, 50 mM NaCl at a final concentration of 7 mg/mL. The P564 and HyP564 versions of the HIF1α MMAA peptides were each spiked into the sample at a final concentration of 14 µM. GST-VBC (0.9 mg) was immobilised on pre-equilibrated glutathione DynaBeads and added to the mixture, which was incubated for 18 hours at 4 °C with end-

to-end rotation (15 rpm). The beads were washed 5 times, with each wash performed for 5 minutes.

To perform on-bead digestion, 10 μg trypsin was added to the mixture and incubated for 18 hours at 37 °C with orbital rotation at 1,000 rpm. The mixture was acidified by drop-wise addition of 10 % FA. Peptides were desalted on an Oasis HLB 96-well plate (30 mg sorbent per well) and eluted in 80 % ACN, 0.1 % FA. The samples were lyophilised by vacuum centrifugation to remove the organic content. The LC-MS/MS data acquisition and processing settings are described in Sections 5.2.5.4 and 5.2.5.5, respectively.

5.2.5.2 Experiment 2

The conditions for experiment 1 (Section 5.2.5.1) were repeated in experiment 2, with the following exceptions:

- 20 mg RCC4 lysate was digested, using the SP4 protocol.
- The peptides were resuspended in 25 mM HEPES pH 7.5, 50 mM NaCl at a concentration of 5 mg/mL.
- The peptide solution was split into two equal volumes (i.e., 2x 10 mg tryptic RCC4 extracts).
- The P564 and HyP564 versions of the HIF1α MMAA peptides were spiked into one of the samples, with each peptide at a final concentration of 5 nM (‘5 nM spike’). An equivalent volume of vehicle was added to the other sample (‘no spike’).
- 0.9 mg of glutathione DynaBead-coupled GST-VBC was added to each sample and incubated at room temperature for 6 hours with end-to-end rotation (30 rpm).
- Each sample was washed 6 times, with each wash performed for 10 minutes. These wash samples were desalted on the Oasis HLB 96-well plate (30 mg sorbent per well) using the same protocol as for the GST-VBC-bound fraction, and analysed by LC-MS/MS.

The LC-MS/MS data acquisition and processing settings are described in Sections 5.2.5.4 and 5.2.5.5, respectively.

5.2.5.3 Experiment 3

A greater degree of orthogonality was applied in experiment 3 to increase the quality of separation of target peptides from the tryptic GST-VBC mixture and non-specific peptide milieu.

The third peptide enrichment experiment used 30 mg RCC4 lysate, digested with trypsin by performing the SP4 protocol. After digestion and desalting, peptides were resuspended in 25 mM HEPES pH 7.5, 50 mM NaCl, 0.01 % (w/v) *Rapi*Gest™ SF Surfactant (*Rapi*Gest™) at a final concentration of 6 mg/mL.

To reduce non-specific interactions in the peptide enrichment assay, a pre-clearing step was performed in which the tryptic peptide mixture was incubated with 300 µL of pre-equilibrated glutathione DynaBeads for 2 hours. The unbound material was used as the 'pre-cleared input' for the peptide enrichment assay.

GST-VBC (3 mg) was immobilised on 300 µL of pre-equilibrated glutathione DynaBeads and added to the pre-cleared input, which was incubated for 18 hours at 4 °C with end-to-end rotation (15 rpm). The beads were split equally into three different protein low-bind microcentrifuge tubes, and a different washing protocol was performed on each sample.

- In the 'long wash' assay, the beads were washed 4 times, with each wash performed for 5 minutes in wash buffer (25 mM HEPES pH 7.5, 150 mM NaCl).
- In the 'short wash' assay, the beads were washed 4 times, with each wash performed for 30 seconds in wash buffer.
- For the short wash with *Rapi*Gest™, the beads were washed 4 times, with each wash performed for 30 seconds in wash buffer supplemented with 0.01 % (w/v) *Rapi*Gest™ SF Surfactant.

The down-stream processing steps (i.e., on-bead trypsin digest and desalting) were performed using the method outline in experiment 1 (Section 5.2.5.1). The LC-MS/MS data acquisition and processing settings are described in Sections 5.2.5.4 and 5.2.5.5, respectively.

5.2.5.4 LC-MS/MS data acquisition

The lyophilised peptides were resuspended in 300 µL 5 % ACN, 0.1 % FA and analysed in two steps. First, the un-fractionated GST-VBC-bound samples were analysed on an Orbitrap Fusion™ Lumos™ Tribrid™ instrument. Following this, the samples were subject to off-line high pH fractionation, either with a Pierce™ high pH reversed-phase peptide fractionation kit or using a U3000 HPLC.

For experiments 1 and 2, 20 µL of the un-fractionated samples were loaded onto pre-equilibrated Evosep™ tips with 1x iRT peptide solution. The peptides were eluted using the proprietary Evosep™ 44-minute pre-formed gradient and analysed by LC−MS/MS in an Orbitrap Fusion™ Lumos™ Tribrid™ instrument.

For experiment 3, 20 µL of the un-fractionated sample was analysed by LC−MS/MS in an Orbitrap Fusion™ Lumos™ Tribrid™ instrument. As described above (Section 5.2.5.3), greater peptide separation was desired for the samples in experiment 3. Therefore, a 140-minute ACN gradient (2-45 % buffer B) in 0.1 % FA, as described in Table 43, was performed with a U3000 HPLC instrument with an Acclaim™ Pepmap™ 100 C18 HPLC column (3 µm x 0.075 mm x 500 mm) at a flow rate of 0.200 µL/min.

Table 43 | Non-linear gradient for on-line low pH separation of label-free peptides derived from the un-fractionated peptide enrichment experiment 3.

Peptide separation was performed on-line at low pH using buffer A (0.1 % FA) and buffer B (80 % ACN, 0.1 % FA). A non-linear ACN (2-45% buffer B) gradient was performed over 140 minutes at a flow rate of 0.200 µL/min.

Time (minutes)	Buffer B (%)
0	2
5	2
5.5	8
12	35
145	45
150	95
160	95
161	2
180	2

The settings used for precursor ion selection and fragmentation in the peptide enrichment experiments 1-3 are defined in Table 44.

Table 44 | Settings for label-free data-dependent acquisition in the un-fractionated peptide enrichment samples derived from RCC4 extracts.

Entry	Parameter	Experiment 1	Experiment 2	Experiment 3
MS1	Detector type	Orbitrap	Orbitrap	Orbitrap
	Resolution	60,000	60,000	120,000
	Mass range	Normal	Normal	Normal
	Use quadrupole isolation	True	True	True
	Scan range (*m/z*)	400-1,600	400-1,600	350-1,800
	RF lens (%)	30	30	30
	AGC target	1×10^6	1×10^6	4×10^5
	Maximum injection time (ms)	50	50	50
	Microscans	1	1	1
	Data type	Profile	Profile	Profile
	Polarity	Positive	Positive	Positive
MIPS	Monoisotopic peak determination	Peptide	Peptide	Peptide
Intensity	Filter type, intensity threshold	5×10^4	5×10^4	5×10^4
Charge state	Include charge state(s)	2-6	2-6	2-7
Dynamic exclusion	Exclude after *n* times	1	1	1
	Exclusion duration (s)	15	15	15
	Mass tolerance	ppm	ppm	ppm
	Low	10	10	10
	High	10	10	10
	Exclude isotopes	True	True	True
MS2	Isolation mode	Quadrupole	Quadrupole	Quadrupole
	Isolation window (*m/z*)	1.4	1.4	1.6
	Activation type	HCD	HCD	HCD
	HCD collision energy (%)	32	32	30
	Detector type	Orbitrap	Orbitrap	Ion trap
	Resolution	15,000	15,000	15,000

	AGC target	1 x10^6	1 x10^6	1 x10^4
	Maximum injection time (ms)	22	22	35
	Microscans	1	1	1
	Data type	Centroid	Centroid	Centroid
Data dependent properties	Select: 'Cycle time' or 'Top *n* scans'	1 s cycle time	1 s cycle time	Top 20 scans

After the initial LC-MS/MS analysis of the un-fractionated samples from experiments 1 and 2, high pH fractionation was performed on 100 µg of the peptide mixtures using the Pierce™ high pH reverse-phase fractionation kit, following the manufacturer's protocol. The gradient is described in

Table 45.

Table 45 | Non-linear gradient for off-line high pH fractionation of label-free peptides derived from experiments 1 and 2 of the peptide enrichment assays.

High pH fractionation was performed using 0.1 % TEA, pH 10 and increasing concentrations of ACN (0-80 %).

Fraction number	**ACN (%)**
1	0
2	2
3	4
4	6
5	8
6	10
7	12
8	14
9	16
10	18
11	20
12	22
13	24

14	26
15	50
16	80

For the samples from experiments 1 and 2, the fractions generated from the Pierce™ high pH reverse-phase fractionation kit were combined to generate 8 concatenated fractions, as described in Table 46. Each fraction was lyophilised by vacuum centrifugation to remove the organic content.

Table 46 | Fraction pooling scheme to generate 8 concatenated fractions for samples derived from the GST-VBC-bound fractions of experiments 1 and 2.

Concatenated fraction #	Fraction A	Fraction B
F1	1	16
F2	2	15
F3	3	9
F4	4	10
F5	5	11
F6	6	12
F7	7	13
F8	8	14

For the concatenated fractions derived from experiments 1 and 2, each fraction was analysed by LC-MS/MS using a timsTOF instrument. A nanoElute® UPLC instrument was used to perform a 20-minute ACN gradient (5-40 %) in 0.1 % FA. Precursor ions were selected for fragmentation in DDA-PASEF (parallel accumulation-serial fragmentation) mode using a scan range of 350-2,200 (*m/z*). The collision energy was optimised for good quality MS2 spectra, which was determined by the relative abundance of the precursor ion in the MS2 spectra of target peptides.

For the un-fractionated sample derived from experiment 3, high pH off-line fractionation was performed on 100 µg of digested material with an XBridge BEH C18 XP column (3.5 µm x 1 mm x 150 mm), using the loading pump on a U3000 HPLC system with autosampler fraction collection. A 60-minute ACN gradient (1-35 % buffer B) was

performed in NH_4OH, pH 10, at a flow rate of 40 µL/min, as described in Table 47. This resulted in 96 fractions that were lyophilised by vacuum centrifugation to remove the organic content.

Table 47 | Linear gradient for off-line high pH fractionation of label-free peptides derived from experiment 3 of the peptide enrichment assays.

High pH fractionation was performed using buffer A (20 mM NH_4HCO_2, pH 10) and buffer B (90 % ACN, 20 mM NH_4HCO_2, pH 10). A linear ACN gradient (1-35 % buffer B) was performed over 60 minutes at a flow rate of 40 µL/min.

Time (minutes)	Buffer B (%)
0	1
12	1
72	35
80	95
90	95
90.1	1
100	1

Each fraction was resuspended in 2 % ACN, 0.1 % FA and combined to generate 24 concatenated fractions, as described in Table 48. The concatenated fractions were analysed by LC-MS/MS in an Orbitrap Fusion™ Lumos™ Tribrid™ instrument, using the same on-line ACN gradient (Table 43) and data acquisition settings (Table 44) that were used for the un-fractionated sample derived from experiment 3.

Table 48 | Fraction pooling scheme to generate 24 concatenated fractions.

Concatenated fraction #	Fraction A	Fraction B	Fraction C	Fraction D
F1	1	25	49	73
F2	2	26	50	74
F3	3	27	51	75
F4	4	28	52	76
F5	5	29	53	77
F6	6	30	54	78
F7	7	31	55	79
F8	8	32	56	80

F9	9	33	57	81
F10	10	34	58	82
F11	11	35	59	83
F12	12	36	60	84
F13	13	37	61	85
F14	14	38	62	86
F15	15	39	63	87
F16	16	40	64	88
F17	17	41	65	89
F18	18	42	66	90
F19	19	43	67	91
F20	20	44	68	92
F21	21	45	69	93
F22	22	46	70	94
F23	23	47	71	95
F24	24	48	72	96

5.2.5.5 Data processing

For data that was acquired on the Orbitrap Fusion™ Lumos™ Tribrid™ instrument, the data was processed in PEAKS® using the parameters defined in Table 49.

Table 49 | Processing parameters of data acquired on the Orbitrap Fusion™ Lumos™ Tribrid™ instrument from the peptide enrichment assays performed on RCC4 extracts.

Parameter	Setting
PEAKS® version	Xpro
PEAKS® search tool	DB
Reference proteome	Human, canonical (Uniprot id: UP000005640)
Precursor ion error tolerance	±10 ppm
Fragment ion error tolerance	±0.5 Da
Enzyme	Trypsin
Enzyme specificity	Specific
Maximum number of miscleavages per peptide	2
Fixed modification	Carbamidomethylation (+57.0214 Da; C)
Variable modifications	Oxidation (+15.9949 Da; CDFHKMNPRWY)

	Dioxidation (+31.9899 Da; FMWY) Deamidation (+0.9840 Da; NQ) Acetylation (+42.0367 Da; protein N-terminus)
Maximum number of PTMs per peptide	3

For data that was acquired on the timsTOF instrument, the data was processed in PEAKS® using the parameters defined in Table 50.

Table 50 | Processing parameters of data acquired on the timsTOF instrument from the peptide enrichment assays performed on RCC4 extracts.

Parameter	Setting
PEAKS® version	Xpro
PEAKS® search tool	DB
Reference proteome	Human, canonical (Uniprot id: UP000005640)
Precursor ion error tolerance	±15 ppm
Fragment ion error tolerance	±0.03 Da
Enzyme	Trypsin
Enzyme specificity	Specific
Maximum number of miscleavages per peptide	2
Fixed modification	Carbamidomethylation (+57.0214 Da; C)
Variable modifications	Oxidation (+15.9949 Da; CDFHKMNPRWY) Dioxidation (+31.9899 Da; FMWY) Deamidation (+0.9840 Da; NQ) Acetylation (+42.0367 Da; protein N-terminus)
Maximum number of PTMs per peptide	3

For experiments 1 and 2, which contained the HIF1α MMAA peptides, a search was performed with the PEAKS® SPIDER search tool to identify peptide mutations. The processing parameters are defined in Table 51.

Table 51 | Processing parameters to identify mutations on peptides derived from the peptide enrichment assays performed on RCC4 extracts

Parameter	Setting
PEAKS® version	Xpro
PEAKS® search tool	SPIDER

Query type	Homology match
Fragment ion tolerance	0.02 Da
L equals I	True
Q equals K	True
De novo score threshold	1 %
Peptide hit threshold	-10lgP = 30

5.2.5.6 Analysis

The aim of the peptide enrichment experiments was to identify hydroxyproline-containing peptides that were enriched in a GST-VBC-dependent manner. Therefore, a qualitative approach was taken to assess the hydroxyproline assignments generated in PEAKS®. Filters were applied at the peptide level to identify confidently matched PSMs (-10lgP ≥ 20) and hydroxyproline-containing peptides for which the diagnostic ions were detected in the MS2 spectrum (ion intensity ≥ 1 %). Manual inspection of the MS1 and MS2 spectra was performed using the criteria outlined in Section 3.2.5.

5.3 Results

To test the efficacy of recombinant GST-VBC for hydroxyproline binding, a protein pull down was performed to capture HIF1α. Following this, a proteome-wide pull-down was performed to identify proteins that interact with GST-VBC in a hydroxylation-dependent manner. The GST-VBC complex was then applied to peptide enrichment assays in an attempt to improve total enrichment of prolyl hydroxylase substrates.

5.3.1 Protein enrichment

Recombinant GST-VBC was used to enrich proteins to confirm hydroxyproline binding activity of the affinity reagent. In the first instance, enrichment of HIF1α confirmed hydroxyproline-binding activity of GST-VBC. This was succeeded by a proteome-wide screen of DMOG-treated pVHL-deficient cells to prospect for GST-VBC-binding proteins containing hydroxyproline sites that might be regulated by a member of the 2OGD family.

5.3.1.1 Proof of principle

The hydroxyproline-binding capacity of recombinant GST-VBC was confirmed by performing affinity purification at the protein level with lysates from the pVHL-defective cell line, RCC4. HIF1α was captured, albeit inefficiently, as indicated by the similar levels of HIF1α in the input and flow through (Figure 42). The pull-down experiment was also performed on RCC4 cells that had been treated with 1 mM DMOG for 4 or 8 hours. HIF1α enrichment after 4 hours of DMOG treatment was similar to vehicle-treated cells, which suggests there was a delayed onset of action of DMOG. HIF1α enrichment was significantly decreased at 8 hours, indicating prolyl hydroxylation is suppressed at this time point. This experiment provided proof of principle that the recombinant GST-VBC construct could be used to enrich hydroxyproline-containing proteins.

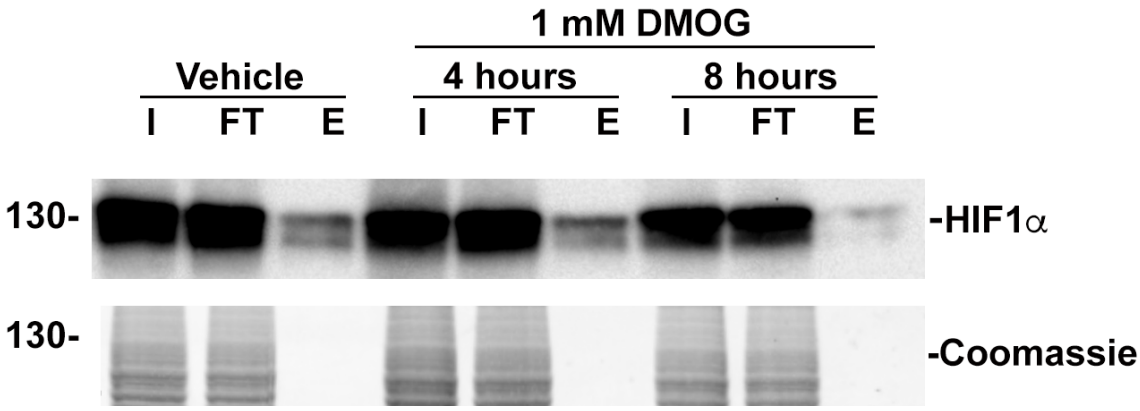


Figure 42 | Affinity purification of HIF1α by GST-VCB is perturbed by DMOG treatment.

RCC4 cells were treated with vehicle for 8 hours or 1 mM DMOG for 4, or 8 hours. A protein pull-down was performed with GST-VCB. HIF1α enrichment was visualised by immunoblot. Coomassie stain of the membrane is provided as a loading control. I = Input. FT = Flow through. E = Elution.

5.3.1.2 Proteome-wide protein enrichment

For the proteome-wide screen RCC4 cells were treated with DMOG or vehicle for 16 hours. A longer duration of DMOG treatment than the previous experiment was considered in order to be confident of prolyl hydroxylation suppression in the subsequent proteomic screen. Affinity purification was performed at the protein level and the samples were resolved by SDS-PAGE (Figure 43). Comparison of the input and flow through

shows little depletion between the cells treated with DMOG or vehicle. This indicates that prolyl hydroxylation of proteins that interact with GST-VBC is likely to be a low prevalence modification in the proteome.

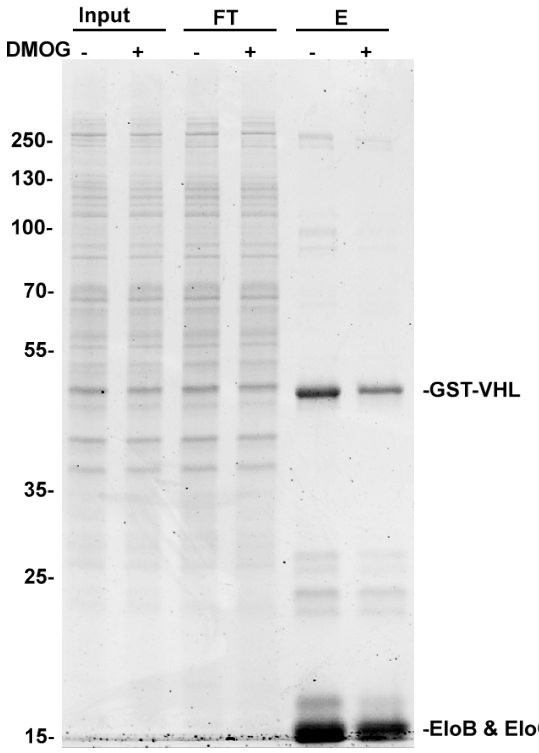


Figure 43 | Protein affinity purification with GST-VBC for proteome-wide investigation of DMOG sensitive prolyl hydroxylation.

RCC4 cells were treated with 1 mM DMOG or vehicle for 16 hours. Samples were resolved by SDS-PAGE and visualised by Quick Coomassie stain. The affinity purified samples were eluted (E) by boiling in Laemmli buffer. Gel slices covering the entirety of the elution lanes were used for down-stream LC-MS/MS analysis. The recombinant proteins of the GST-VBC complex are indicated.

To investigate DMOG-sensitive protein enrichment, LC-MS/MS analysis was performed on gel slices from the gel presented in Figure 43. The data was processed in PEAKS® X and a filter was applied for confident protein assignment (-10lgP ≥ 20). A filter for unique peptides (≥ 2) was applied to increase the security of assignments. The protein abundance fold-change after exposure to DMOG was plotted against the summed abundance from both conditions ('Intensity'). Proteins displaying fold-change ≥ 2 were considered to be significant. With this in mind, *bona fide* prolyl hydroxylase substrates should exhibit a 2-fold reduction in protein abundance following DMOG treatment.

The data indicates HIF1α and several collagen proteins are enriched in a DMOG-sensitive manner (Figure 44). Most proteins cluster around the centre and 2 groups of outliers are generated because of missing values (i.e., protein abundance = 0 in one condition). Although missing values can generate spurious enrichment statistics, some collagen proteins exhibit these dynamics, which suggests that proteins that were genuinely enriched by GST-VBC in a DMOG-sensitive manner might be identified in this group.

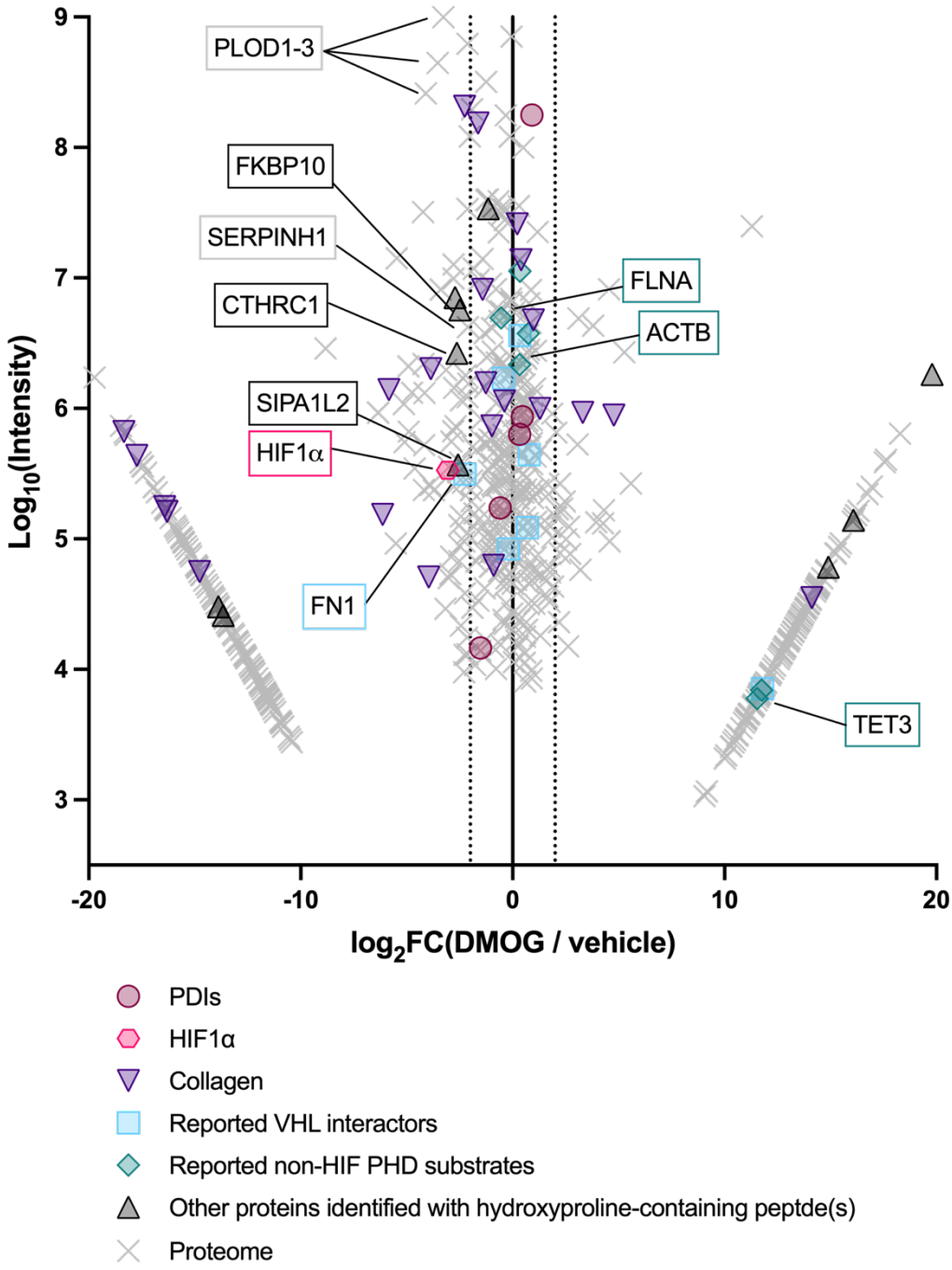


Figure 44 | DMOG sensitive protein enrichment by GST-VBC affinity purification.

Proteins were subject to GST-VBC affinity purification column, resolved by SDS-PAGE, and analysed by LC-MS/MS. The $\log_2$ fold change of protein abundance between DMOG and vehicle treated cells is plotted against the $\log_{10}$ summed protein abundance from both experimental groups (intensity). Dotted lines indicate 2-fold change in protein abundance. Annotated proteins are discussed in the main text and the box colour indicates the protein group.

Proteins containing confident sites of prolyl hydroxylation were surveyed in an attempt to identify novel 2OGD substrates. For this analysis, hydroxyproline-containing peptides

were filtered for confident PSM (-10lgP ≥ 20) and PTM localisation (AScore ≥ 20). Collagen peptides were excluded and MS1 quality refinement was subsequently performed, which resulted in 10 non-collagen peptides with confident hydroxyproline assignments. The proteins from which these peptides are derived were designated by grey triangles in Figure 44. CTHRC1 contains a collagen-like triple helical domain and the confidently localised hydroxylation site, HyP75, is within this domain. This raises the possibility CTHRC1 is a novel C-P4H substrate. FKBP10 HyP36 was identified in this dataset, which strengthens the assignment from the previous two chapters. SIPA1L2 was found to be hydroxylated at two residues, HyP1070 and HyP1072 and displayed a similar response to DMOG as HIF1α. These proteins most likely represent novel 2OGD substrates.

Some collagen-binding proteins also showed DMOG-sensitive enrichment (e.g., PLOD1-3, and SERPINH1). It is not possible to distinguish if the DMOG-sensitive enrichment is due to a direct interaction between GST-VBC and the enriched protein or if this effect occurs indirectly *via* interaction with collagen. SERPINH1, for example, is a known collagen chaperone and the work of this thesis has already established P30 as a site of 2OGD-catalysed prolyl hydroxylation. Tryptic peptides containing the target site were not identified in this analysis, so it is not known how hydroxylation was affected by DMOG treatment. The data from the previous SILAC screen suggest hydroxylation of SERPINH1 P30 would be suppressed following DMOG treatment (Chapter 4). It is therefore possible that this protein exhibits DMOG-dependent regulation as a function of direct GST-VBC interaction by HyP30 or indirect interaction with GST-VBC by collagen. FKBP10 is another collagen chaperone that was enriched in a DMOG-sensitive manner. Although DMOG-sensitive enrichment is consistent with the DMOG sensitivity of the hydroxyproline site identified in the MEF SILAC screen (Fkbp10 HypP35) (Chapter 4), there remains a possibility that the enrichment was due to an association with collagen.

Previous analyses have shown the PDI proteins to be subject to 2OGD-catalysed prolyl hydroxylation (Section 4.3.3). Some PDI proteins were identified after the protein enrichment, suggesting there was an interaction with the GST-VBC complex, but

DMOG-sensitive enrichment was not observed. This might have been due to the long half-life of PDI proteins relative to the duration of DMOG treatment. Thus, it is not known if the PDIs can bind directly to GST-VBC *via* the hydroxyproline sites. A negative control, such as GST alone or a GST-VBC mutant with a defective hydroxyproline-binding pocket could clarify this in future experiments.

Identification of DMOG-sensitive HIF1α enrichment confirms that this experiment is effective as an orthogonal method to identify proteins containing 2OGD-catalysed prolyl hydroxylation. With this in mind, I considered the possibility that some of the reported non-HIFα PHD substrates, which were beyond the limits of detection in previous experiments described in this thesis, might be sufficiently enriched by GST-VBC pull-down to permit identification. I therefore interrogated the data for all 45 reported non-HIFα PHD substrates.

This analysis identified 6 reported non-HIFα PHD substrates, all of which have been observed in highly fractionated publicly deposited proteomic datasets (Section 3.3.2). Although a hydroxyproline-dependent interaction with pVHL has been described for 3 of the 6 target proteins identified in this analysis (ACTB, FLNA, and TET3), the results of this analysis appear to contradict the previously reported findings. None of the reported non-HIFα PHD substrates displayed loss of enrichment in a DMOG-sensitive manner. Although it is possible these proteins are not dynamically enriched because of long half-life, proteomic data described in the previous chapters of this thesis suggest that these proteins are not prolyl hydroxylated.

The remaining pVHL interactors (i.e., those not previously reported as PHD substrates) were also surveyed in this analysis. Fibronectin (FN1) was found to be enriched with a similar response to DMOG as HIF1α and SIPA1L2. No hydroxyproline sites were identified for this protein. The tubulin proteins were identified in this analysis (blue squares) but did not show DMOG-sensitive enrichment.

Overall, this experiment has identified a number of novel candidates of DMOG-sensitive prolyl hydroxylation, which implicates them as candidate 2OGD substrates. Long half-life proteins, such as PDIs, may bind directly to GST-VBC *via* a hydroxyproline site but the duration of DMOG treatment is insufficient to determine if the enrichment is

dependent on prolyl hydroxylation. Although some of the reported non-HIFα pVHL substrates were identified in both GST-VBC-enriched fractions, only FN1 displayed loss of enrichment in a DMOG-sensitive manner. This suggests prolyl hydroxylation is not required for the interaction between pVHL and the reported non-HIFα pVHL substrates, which includes some of the reported non-HIFα PHD substrates. The possibility of indirect interaction with collagen, or other proteins that bind directly to pVHL, increases the difficulty of analysis. Furthermore, some hydroxyproline-containing proteins may not bind to pVHL due to structural constraints. For these reasons, a direct method of hydroxyproline-containing peptide enrichment was developed.

5.3.2 Peptide enrichment

The protein enrichment experiment identified pVHL binding to HIF1α, many collagen proteins, and some novel hydroxyproline-containing proteins in a DMOG-sensitive manner *ex vivo*. It was not possible to determine if these interactions occurred directly *via* the hydroxyproline-binding pocket of GST-VBC (e.g., FKBP10 and SERPINH1). Additionally, there was a possibility some hydroxyproline-containing proteins were not enriched by GST-VBC because of structural or topological restrictions. I therefore sought a method to directly enrich hydroxyproline sites irrespective of structure or *in cellulo* interactions with pVHL.

To directly enrich for sites of prolyl hydroxylation the GST-VBC affinity purification workflow was applied to tryptic peptides. This required method development to ensure that the samples were suitable for LC-MS/MS analyses (e.g., no detergents or contaminating proteins). In the first instance I demonstrated binding to tryptic peptides using an adapted construct, GST-TS-VBC, in surface plasmon resonance (SPR) assays. This work was followed by attempts to elute target peptides from the GST-VBC complex without forfeiting integrity of the affinity reagent, which would have otherwise contributed contaminating proteins to the LC-MS/MS samples. Finally, preparative purifications were performed to determine suitable conditions for hydroxyproline-containing peptide enrichment.

5.3.2.1 Recombinant GST-VBC binds synthetic peptides that contain hydroxyproline

There are reports of recombinant pVHL binding to synthetic hydroxyproline-containing peptides in on-bead peptide binding assays (Table 39), which provides some demonstration that pVHL can bind to a broad range of peptides when hydroxyproline is present. This might not necessarily be true of protein substrates, in which structural constraints may prevent pVHL binding. Although the previously reported peptide binding assays have used non-tryptic peptides, the binding of recombinant GST-VBC to hydroxyproline-containing tryptic peptides is expected to be possible for a number of substrates.

SPR was employed to study GST-TS-VBC binding to synthetic tryptic peptides (Figure 45). For this study, a pVHL construct containing a GST tag and a Twin-Strep-tag® (TS) was used and immobilised onto a CM5 sensor chip coated in Strep-Tactin®XT. Synthetic peptides corresponding to prolyl hydroxylated tryptic fragments were used to assess the potential of GST-TS-VBC to capture target peptides. The HIF1α HyP564 peptide was used as a positive control. In addition to HIF1α HyP564 peptide, two AKT1 peptides were also used. AKT1 is a reported non-HIFα PHD substrate that was proposed to be recognised by pVHL in its hydroxylated form (Guo *et al.*, 2016). The peptides used in this assay contained one (AKT1 HyP313) or two of the target hydroxyproline residues (AKT1 HyP313 HyP318). Although I have not found evidence for prolyl hydroxylation of AKT1 *in vivo*, these synthetic peptides have been shown to bind recombinant GST-VBC in fluorescence polarization assays by a collaborator in the Schofield laboratory (Lippl, 2018).

The kinetics of GST-TS-VBC binding are presented in Table 52. The results achieved with the HIF1α HyP564 peptide were highly similar to previously published data studying the interaction of recombinant GST-VBC and synthetic HIFα peptides (Hon *et al.*, 2002). The HIF1α HyP564 peptide displayed slow dissociation from GST-TS-VBC (Figure 45), and similar dynamics with the doubly hydroxylated AKT1 peptide were observed. In contrast to this, the singly hydroxylated AKT1 peptide displayed rapid association and dissociation from GST-TS-VBC. These results suggest GST-TS-VBC

might engage in transient interactions with some peptides (e.g., AKT1 HyP313) because of a high off-rate whilst others are more likely to be retained for longer (e.g., throughout an affinity purification experiment) on the basis of low off-rate (e.g., HIF1α HyP564). By introducing a second hydroxyproline (i.e., a second ligand), the doubly hydroxylated AKT1 peptide exhibited a much lower off-rate. This was probably due to increasing the local concentration of ligand, which increased the probability of re-binding the peptide substrate following its initial dissociation. In contrast, a non-hydroxylated HIF1α P564 control peptide did not demonstrate any binding. Together, these results demonstrated that some hydroxyproline-containing peptides (e.g., HIF1α HyP564 and AKT1 HyP313 HyP318) would be expected to be captured by GST-TS-VBC if they are present in the sample. However, their detection by LC-MS/MS would rely on the sensitivity of the assay.

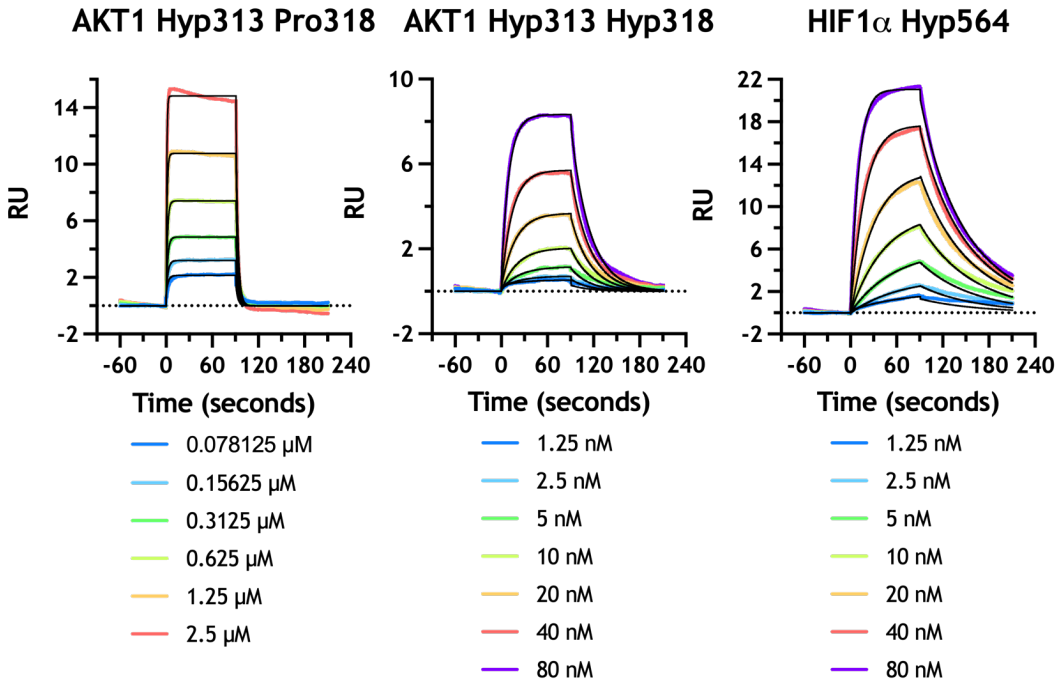


Figure 45 | GST-TS-VBC binding to hydroxylated peptides in SPR experiments.

Recombinant GST-TS-VBC was immobilised on a Strep-Tactin®XT chip and synthetic 'tryptic' peptides were introduced to measure the kinetics of GST-TS-VBC binding. The reported PHD substrates are indicated above each chart to indicate the hydroxylation status of the synthetic peptide.

These results show GST-TS-VBC can bind hydroxylated HIFα and non-HIFα peptides. The TS tag is unlikely to influence substrate binding. Therefore, the results of the SPR assays are applicable to the binding activity of GST-VBC, which is the affinity purification reagent used in the peptide enrichment assays performed in this thesis. Although GST-VBC may not bind all hydroxyproline-containing peptides in a proteome-wide affinity purification experiment, it has the potential to enrich 2OGD substrates if they are present in the analyte. The SPR data also generated a hypothesis that multiply hydroxylated peptides will be more efficiently enriched than singly hydroxylated peptides because of an increased local concentration of ligands. This would facilitate re-binding of the peptide and generate a longer complex half-life than that of the singly hydroxylated version of the peptide.

Table 52 | GST-TS-VBC kinetics from the SPR binding assays.

Peptide	K_D (M)	K_a ($M^{-1}s^{-1}$)	K_d (s^{-1})	$t_{1/2}$ (s)
HIF1α HyP564	2.57×10^{-8}	1.24×10^{6}	0.01947	35
AKT1 HyP313 Pro318	1.21×10^{-6}	3.83×10^{5}	0.36330	1
AKT1 HyP313 HyP318	7.99×10^{-8}	7.54×10^{5}	0.04051	17

5.3.2.2 Target peptide elution

The workflow that I have described above has established protein affinity enrichment by GST-VBC (Section 5.3.1). However, the preparative use of GST-VBC to enrich for hydroxyproline-containing peptides across the proteome required further method development. I demonstrated GST-VBC binding to hydroxyproline-containing tryptic peptides by SPR, but this configuration is different to that used for proteomic screens. For the proteomic screen GST-VBC would be immobilised onto magnetic beads in a microcentrifuge tube, instead of a sensor chip in a flow cell, and a complex mixture of tryptic peptides is added, rather than a pure synthetic peptide. Additionally, after target peptides dissociate from the complex, they are removed by buffer exchange rather than

by continuous flow. This facilitates re-binding of target peptides and could lead to lower resolution between high- and low-affinity interactors.

Ideally, target peptides are retained in the wash steps that remove non-specific interactions from GST-VBC and are subsequently eluted. This strategy would enrich hydroxyproline-containing peptides and avoid contamination from the recombinant GST-VBC, which will be in a large molar excess if it is introduced into the sample. In the previously described protein enrichment experiment, SDS-PAGE was used to resolve the GST-VBC proteins. Tryptic peptides may diffuse out of the gel because of their low molecular weight. In contrast to glycine-based SDS-PAGE, which effectively resolves proteins between 30 and 200 kDa, tricine-based SDS-PAGE allows resolution of proteins between 1 and 100 kDa (Schägger, 2006), which implies tryptic peptides could be retained in these gels.

I therefore considered tricine-SDS-PAGE as a method to separate GST-VBC proteins and target peptides. HIF1α MMAA peptides were spiked in to 50 µg tryptic HeLa extract at a final concentration of 1 µM with an equivalent concentration of GST-VBC. The separation of GST-VBC and target peptides was tested using a 15 % tricine gel and the peptide extraction method from an in-gel trypsin digest protocol was used. Although resolution by molecular weight was observed by Quick Coomassie stain, the overall yield of peptide extraction was low (< 1 %) and the spike-in HIF1α MMAA peptides were not detected (data not shown). It is possible this method could be optimised in the future, but I instead proceeded to consider elution strategies that would liberate target peptides from the bait and minimise sample losses.

In addition to GST-VBC contamination, detergents present a difficulty for any potential peptide enrichment method. Most affinity purification protocols use detergents in the sample buffers to reduce non-specific interactions and maintain solubility of proteins. Detergents can cause ion suppression in electrospray ionisation mass spectrometers, along with irreversible contamination to C18 columns, which are expensive to replace. Furthermore, detergents are concentrated in the peptide desalting methods that are typically used prior to LC-MS/MS analysis. The peptide enrichment method must therefore be optimised in solutions that lack detergent and the following experiments were performed in a detergent-free buffer (25 mM NaCl, 50 mM HEPES pH 7.5).

A simple elution strategy that retains all target peptides would have been to denature the trimeric complex so that target peptides are released from the GST-VBC complex. Peptides are subject to desalting on a C18 column prior to analysis by LC-MS/MS and it was hypothesised that these columns would retain the denatured GST-VBC at high acetonitrile concentration (e.g., ≥ 40 %) so that the target peptides could be eluted below this threshold. 500 ng GST-VBC was denatured by the stepwise addition of 10 % FA to pH 2. The sample was loaded onto a C18 column and solutions containing 0.1 % FA with increasing concentrations of ACN (5-60 %) were added to the column to elute the proteins. The three proteins of the GST-VBC complex eluted at 20 % ACN (data not shown), so I concluded that this approach would not successfully prevent GST-VBC contamination.

The previous experiments did not identify conditions in which GST-VBC could be resolved from target peptides for efficient detection of hydroxyproline-containing peptides by LC-MS/MS. I therefore considered chemical elution of the target peptides using a high salt concentration (5 M NaCl) and a variety of solutions covering a range of low and high pH conditions (pH 3, 10, and 11). GST-VBC was immobilised on glutathione DynaBeads and incubated in the different elution buffers for 10 minutes with gentle agitation. The beads were retained and boiled in Laemmli buffer to evaluate retention of the GST-VBC construct. GST-VBC remained intact in the sample buffer and the pH 10 solution but was liberated from the glutathione beads in the other buffers (data not shown). To further examine high pH elution a titration was performed from pH 9.0 – pH 11.0 (Figure 46). GST-VBC was liberated from the beads at pH 10.75. This suggests that high pH elution might be possible at pH ≤ 10.5.

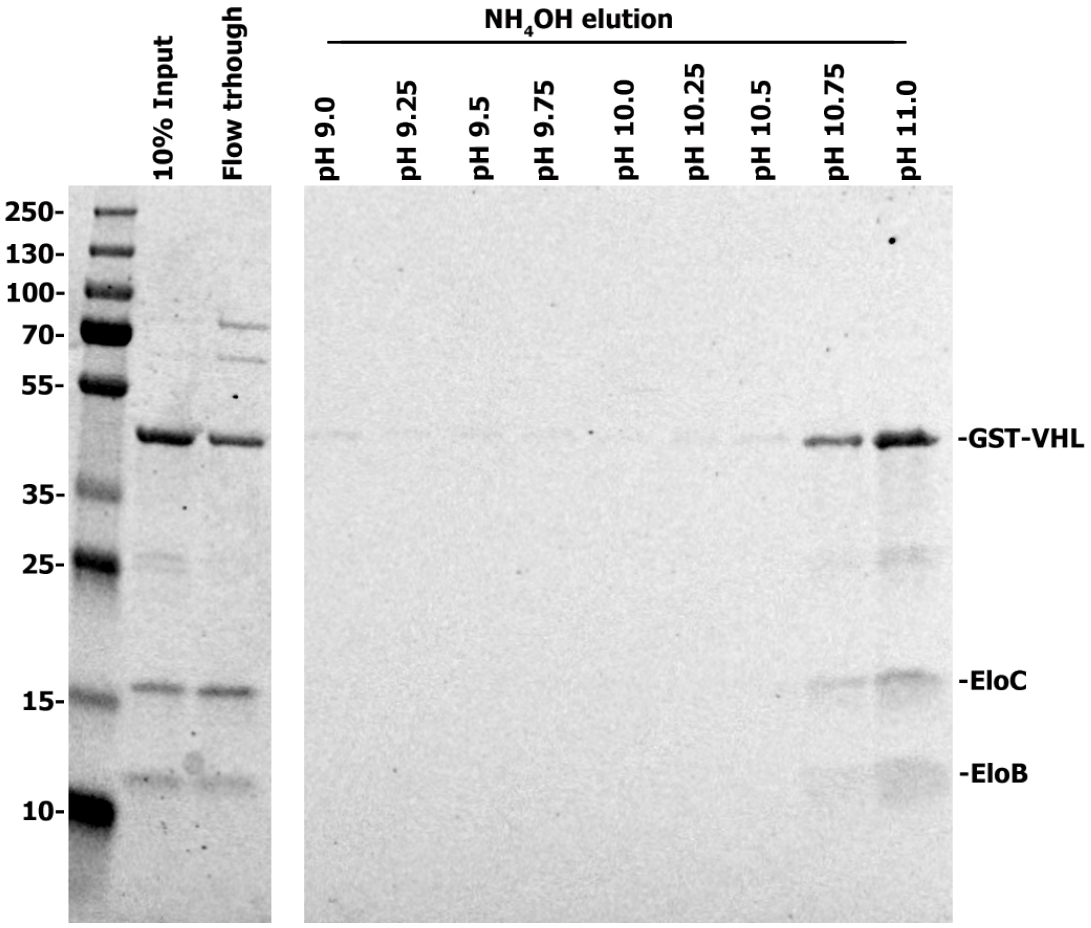


Figure 46 | Titration of high pH solutions to consider target peptide elution from GST-VBC in non-denaturing conditions.

In addition to non-specific elution strategies, it has been demonstrated that competitive elution would displace target peptides from GST-VBC (Zhang *et al.*, 2018). Competitive elution was therefore tested because it could elute target peptides from GST-VBC without liberating GST-VBC from the glutathione DynaBeads. This approach was taken by Lee and colleagues, whereby L-hydroxyproline, small hydroxyproline-containing oligopeptides, and gelatine were used in an attempt to competitively elute target peptides (Arsenault *et al.*, 2015). Additionally, a pVHL inhibitor, VH298, has been developed that would be expected to compete with target peptides to occupy the substrate-binding pocket of pVHL (Frost et al., 2016). Finally, a HIFα mimetic peptide, which had been designed to interact with pVHL (Lippl, 2018), termed pVHL elution peptide (VEP) was synthesised as an alternative ligand. These three ligands were each prepared at final concentration of 1 mM and compared to three high pH solutions (pH 9.5 – 10.5). The

experiment tested elution of GST-VBC bound to a biotinylated HIF1α HyP564 peptide that was immobilised on streptavidin DynaBeads. Each elution step was performed for 10 minutes with gentle agitation and repeated for a total of three incubations. None of the ligands liberated GST-VBC from the immobilised HIF1α peptide, whilst all three high pH solutions demonstrated inefficient elution (Figure 47). It might have been possible to engineer ligands that would be more effective, especially given the results of the SPR analysis, but this was not guaranteed to succeed and would have required a significant time investment. I therefore considered an alternative approach to liberate target peptides from GST-VBC whilst minimising sample losses.

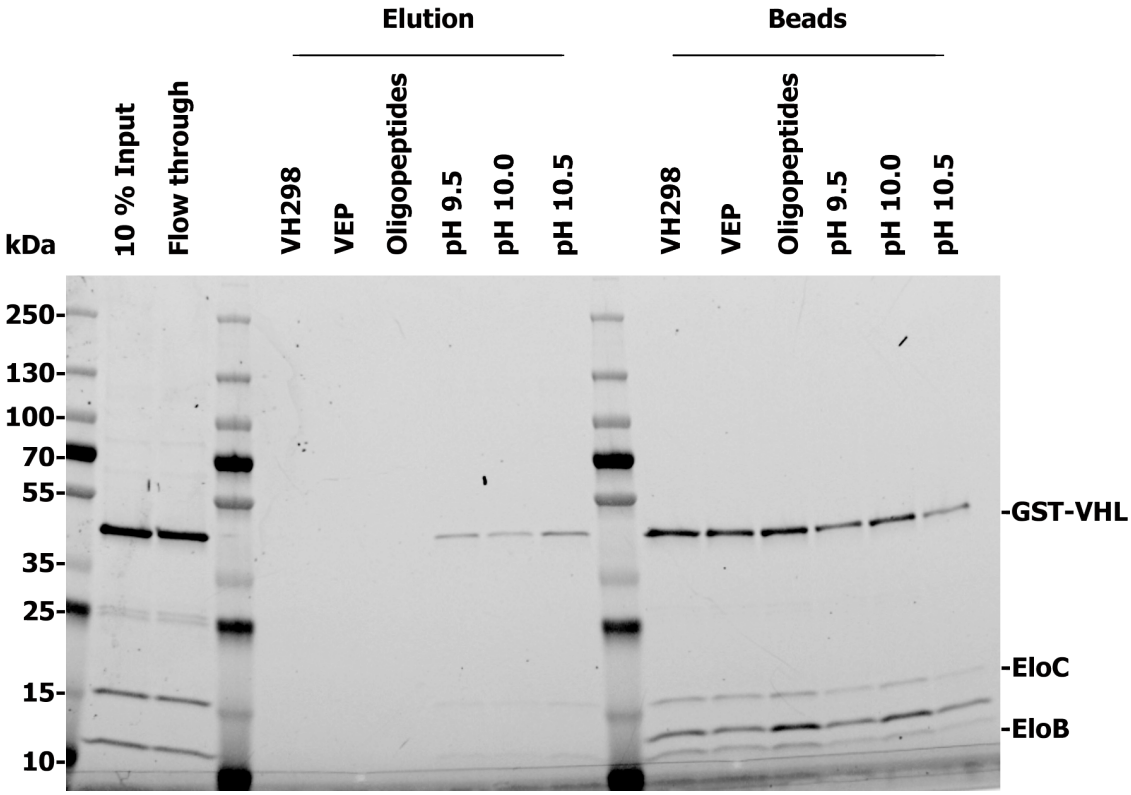


Figure 47 | Comparison of competitive ligand-based and high pH elution strategies to disrupt GST-VBC binding to hydroxylated HIF1α peptide.

Although there was a possibility the high pH elution strategy might be developed further to liberate GST-VBC from the biotinylated HIF1α peptide, the inefficiency of the elution meant that attempts to elute target peptides from GST-VBC were terminated. An alternative strategy was proposed whereby on-bead tryptic digest was performed to

liberate all target peptides from the bait. This approach should not result in losses of the target peptides, but high background (i.e., GST-VBC peptides) was expected in LC-MS/MS analyses, which could hinder the sensitivity of the assays for prolyl hydroxylated GST-VBC-binding peptides.

An initial experiment was performed with 1 µg GST-VBC immobilised on glutathione DynaBeads and incubated with 100 µg HeLa tryptic lysate. Following trypsin digest the peptides were desalted and analysed by LC-MS/MS. The results indicated peptides derived from GST-VBC were the most abundant in the experiment, but many other peptides were detected at a similar order of magnitude (data not shown). There were no apparent malfunctions to the instrumentation and peptides did not carry over into subsequent samples. This indicated that the on-bead digest method was suitable for liberating target peptides from the GST-VBC complex for LC-MS/MS applications and would theoretically minimise samples losses.

5.3.2.3 Hydroxyproline-containing peptide enrichment with recombinant pVHL

The previous experiments established that on-bead trypsinolysis was the most suitable method for liberating target peptides from GST-VBC, at least from the different elution strategies that were tested. I therefore sought to determine conditions that provided the greatest enrichment of hydroxyproline sites using this method. Three protocols were performed that considered i) GST-VBC binding to tryptic HIFα peptides, ii) dissociation of hydroxyproline-containing peptides from the GST-VBC complex, and iii) removal of nonspecific peptides in different wash strategies (Figure 48).

Collagen and elastin are components of the extracellular matrix (ECM) that contain many sites of prolyl hydroxylation (Gorres and Raines, 2010) that could interfere with the peptide enrichment assays. It is therefore important to note that a trypsin-versene solution was applied to RCC4 cells to aid their dissociated from the culture plates. The cells were then washed thoroughly in PBS to remove components of the ECM.

Peptide stringency filters were employed for each experiment. Filters for confident PSM (-10lgP $\geq$ 20) and PTM localisation (ion intensity $\geq$ 1 %) were employed. The PTM localisation filter was less stringent than in previous analyses because GST-VBC should exhibit specific capture of hydroxyproline-containing residues, which serves as an additional control in the assay design. Additionally, MS1 and MS2 spectra of

hydroxyproline-containing peptides were manually inspected to validate the assignments. The results of these experiments are summarised in Table 53.

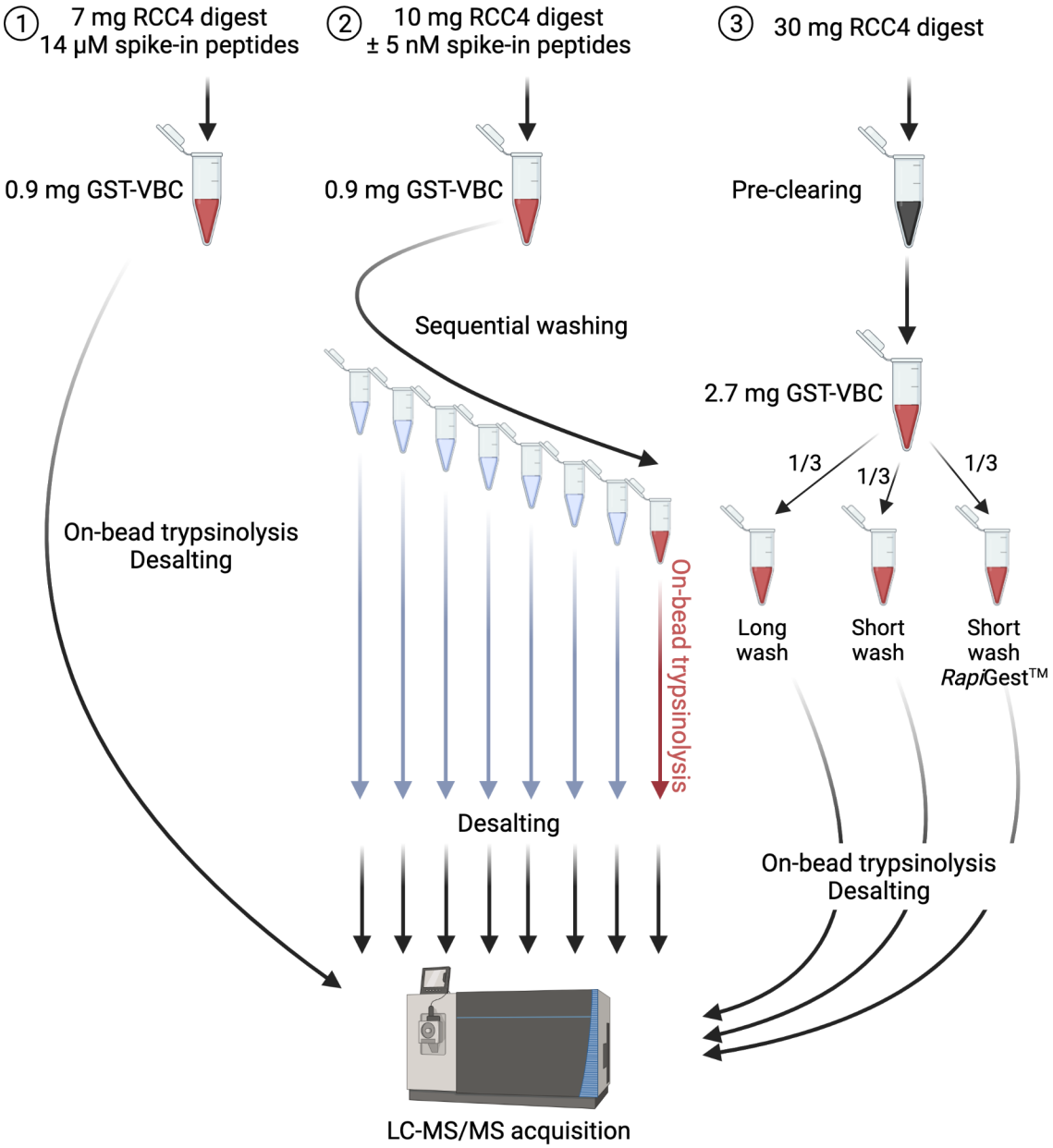


Figure 48 | Strategies for peptide enrichment with recombinant GST-VBC.

Peptide enrichment assays were performed on RCC4 lysates digested with trypsin. Where indicated, synthetic tryptic peptides containing HIF1α CODD target site (HyP564 and P564 versions) were spiked into the tryptic digests. These peptides contained mutation of both methionine residues to alanine (MMAA) to simplify the analysis. 1) Proof of principle experiment in which 7 mg RCC4 tryptic digest was spiked with 14 µM of HIF1α MMAA peptides. On-bead proteolysis was performed with trypsin to liberate target peptides. Desalting was performed on an Oasis HLB plate and LC-MS/MS acquisition was achieved using an Orbitrap Fusion™ Lumos™ Tribrid™ instrument. 2) Two GST-VBC peptide enrichment experiments were performed, each with 10 mg RCC4 tryptic digest. One experiment was spiked with 5 nM HIF1α

MMAA peptides. The beads were washed for six times, each for 10 minutes, and on-bead trypsinolysis was performed on the GST-VCB fraction. All fractions, including the washes, were desalted and analysed by LC-MS/MS as described in (1). 3) Peptide enrichment was performed with 30 mg RCC4 tryptic digest that had been pre-cleared by incubation with glutathione DynaBeads. The enriched material was separated into three equal fractions. Each fraction was subject to a different wash protocol before on-bead trypsinolysis, desalting and LC-MS/MS acquisition was performed as described in (1).

In the first experiment, HIF1α MMAA peptides were spiked into RCC4 tryptic digests to achieve an approximately equimolar ratio with GST-VBC. These peptides were designed to contain the exact same sequence as the tryptic HIF1α CODD peptide except both methionine residues were substituted for alanine, to simplify MS/MS analyses and enable distinction from endogenous peptides. Both hydroxyproline (HyP564) and non-hydroxylated (P564) peptides were used. Although some P564 peptide was detected in the VBC enriched fractions, the HIF1α MMAA HyP564 peptide was specifically enriched in this assay (Figure 49) and the identity was confirmed by manual inspection of the MS2 spectra (Figure 53).

In addition to the HIF1α MMAA HyP564 peptide, several COL4A2 hydroxyproline-containing peptides were identified in the GST-VBC-enriched fraction. These results served as proof of principle that the GST-VBC complex could be used to identify tryptic peptides containing hydroxyproline that were derived from HIFα and non-HIFα proteins. The samples were then subject to high pH fractionation using a Pierce™ high pH peptide fractionation kit. This increased the number of COL4A2 peptides identified and enabled detection of COL4A1 and non-collagen hydroxyproline-containing peptides. Despite this, the overall number of hydroxyproline-containing peptides identified was low. It was considered possible that the high concentration of HIF1α MMAA peptides competed with endogenous hydroxyproline-containing peptides and that this limited the number of novel hydroxyproline sites identified in this analysis.

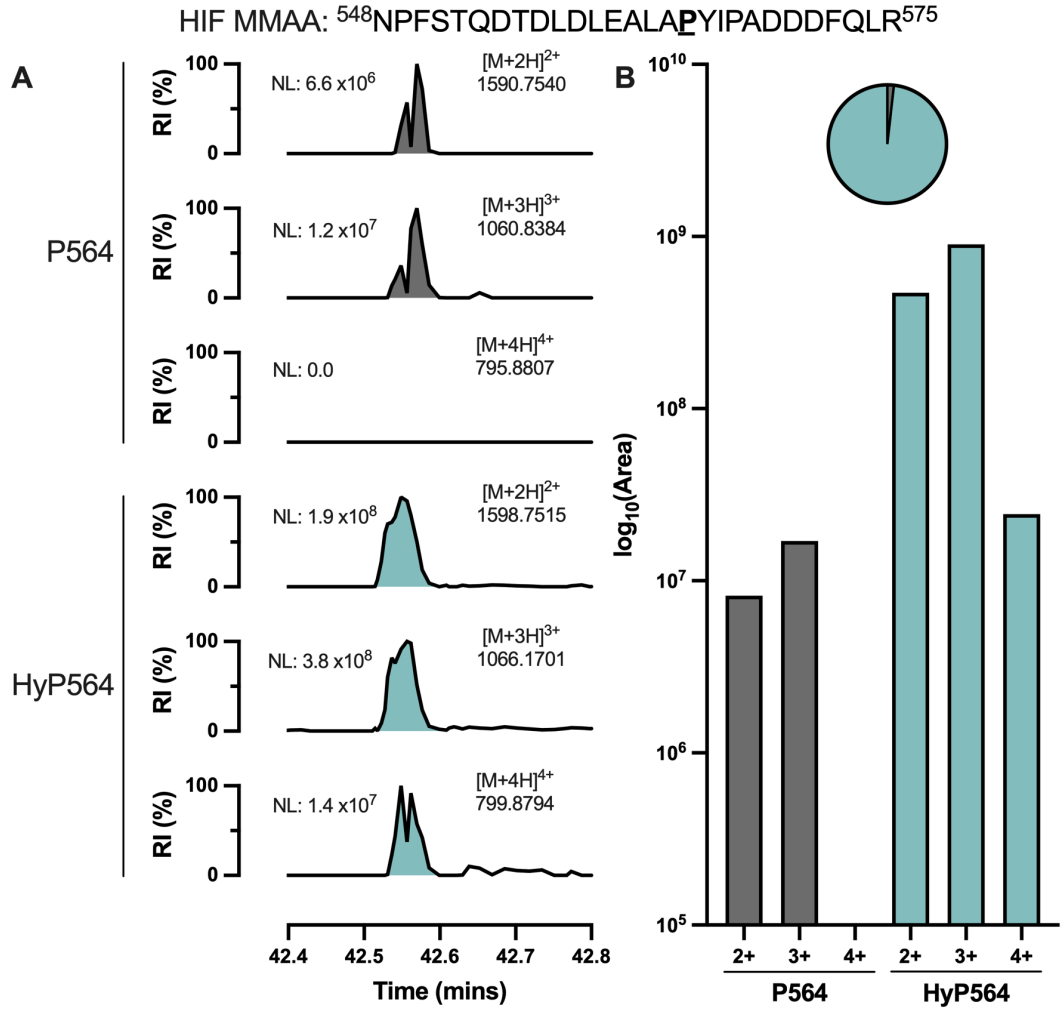


Figure 49 | XICs demonstrate specific enrichment of hydroxyproline-containing HIF1α MMAA peptide by GST-VBC affinity purification.

Hydroxylated (HyP564) and non-hydroxylated (P564) HIF1α MMAA peptides were spiked into RCC4 tryptic digests at equimolar concentrations prior to GST-VBC affinity purification. The GST-VBC-bound material was subject to a second round of trypsin digestion and analysed by LC-MS/MS. A) XICs of the target peptides in three charge states ($z = 2^+, 3^+, 4^+$). Peptide intensity relative to the base peak, the intensity of which is indicated by the normalisation level (NL), is plotted over time. B) The bar chart depicts the area under the curve that was calculated for the highlighted peaks from 'A' on a $\log_{10}$ axis. The pie chart indicates the relative abundance of the P564 and HyP564 peptides in the GST-VBC-enriched fraction on a log scale

For the second experiment a comparison of digests containing HIF1α MMAA peptides was performed. In an attempt to reduce competitive binding for the hydroxyproline-binding pocket of GST-VBC between the HIF1α MMAA peptides and endogenous peptides, HIF1α MMAA peptides were spiked into RCC4 digests at a molar ratio of approximately 1:3,000 to GST-VBC. An additional step was taken to consider

dissociation of hydroxyproline-containing peptides during washing; the results of the SPR analysis suggested that some peptides might dissociate from GST-VBC more quickly than others, so the wash fractions were retained for LC-MS/MS analysis. The number and duration of wash steps was increased from five 5-minute washes to six 10-minute washes.

The HIF1α MMAA peptides were not detected in any sample but there was evidence of increased retention of hydroxyproline-containing peptides throughout the washing protocol. For example, the peptide containing FKBP10 P36 was identified in Fraction 1 but the HyP36 peptide was identified in fractions 1-3 (Figure 50). The retention of FKBP10 HyP36 is probably due to specific binding with GST-VBC *via* the hydroxyproline binding pocket, whereas P36 peptides are likely to have associated in a non-specific manner.

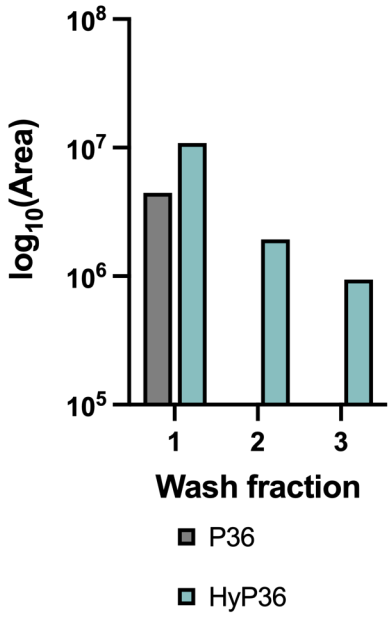


Figure 50 | Detection of tryptic peptides containing the target site of FKBP10 prolyl hydroxylation, P36, in hydroxylated and non-hydroxylated forms.

The dissociation of collagen hydroxyproline-containing peptides during the wash steps (data not shown) suggested peptide capture by GST-VBC is dependent on the affinity of the interaction. It was hypothesised that high affinity interactions would be identified in the GST-VBC-bound fraction. For instance, given the generally low number of collagen peptides identified in the wash steps, it was expected that many collagen peptides would

be detected in the GST-VBC-bound fraction. High pH fractionation was performed on the GST-VBC-bound material using the same kit as for experiment 1. Surprisingly, only one collagen peptide was identified in this sample, COL6A2 HyP357 HyP363, which was not identified in the input or any of the wash fractions. Although this was far fewer collagen peptides than had been expected, it indicated some enrichment of certain hydroxyproline-containing peptides. The most abundant hydroxyproline-containing peptides in the GST-VBC-bound fraction were SIPA1L2 HyP1070 HyP1072 and KHDRBS1 HyP297 HyP299 (Figure 51). Importantly, these peptides were not detected in the input or wash fractions, suggesting that their detection was entirely dependent on GST-VBC affinity purification. Additionally, these observations supported the hypothesis that peptides containing multiple hydroxyproline sites would exhibit greater enrichment by GST-VBC than singly hydroxylated peptides. Although identification of these peptides in the GST-VBC-bound fraction was an improvement on hydroxyproline-containing peptide detection from experiment 1, endogenous HIFα peptides were not identified. High background signal of GST-VBC and peptides that exhibit non-specific interactions with the DynaBeads most likely limited the sensitivity of this assay.

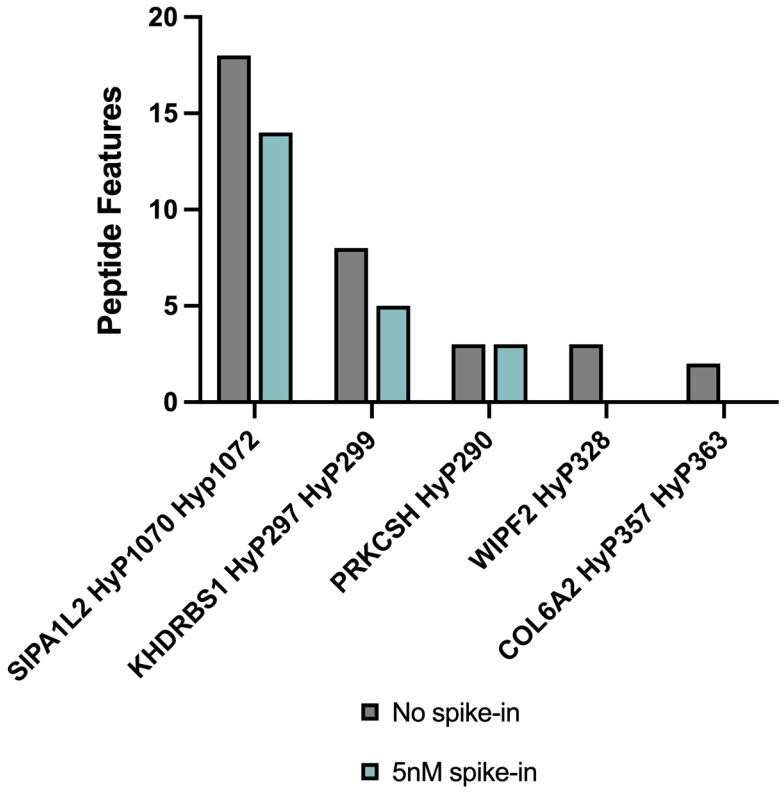


Figure 51 | Hydroxyproline peptides identified in the GST-VBC-bound fraction after peptide enrichment and sequential washing.

Hydroxyproline-containing peptides were manually inspected for correct PSM and PTM assignment for each peptide enrichment assay (± 5 nM HIF1α MMAA peptide). The abundance of each peptide is inferred by the number of Peptide Features assigned by PEAKS® Xpro.

For the third experiment, a single affinity purification was split equally, and several complementary adaptations to the protocol were performed to increase the overall sensitivity of the assay. A pre-clearing step was performed with the magnetic beads to reduce the number of non-specific interactions in the GST-VBC-bound fraction. Three different wash strategies (Section 5.2.5) were performed that might influence peptide dissociation from GST-VBC. Finally, these samples were subjected to a 180-minute ACN gradient to increase the dynamic range of the samples.

Analysis of the samples generated using the three different wash strategies revealed that the endogenous tryptic peptide containing HIF1α HyP564 (Figure 54) was detected in all three samples, and that of HIF2α HyP402 (Figure 57) was detected in the sample with the long washing protocol. This is the first time that the endogenous hydroxyproline-containing HIFα peptides have been identified by LC-MS/MS in unbiased affinity purification assays. Many collagen peptides were identified, which further validates the hydroxyproline-containing peptide enrichment method. CTHRC1 HyP75 HyP81 was identified in the long wash and short wash with *Rapi*Gest™ samples. SIPA1L2 HyP1070 HyP1072 was detected in the short wash with *Rapi*Gest™ sample. These results further strengthen the assignments that have appeared previously in the protein enrichment investigation.

The long wash sample from experiment 3 contained two endogenous hydroxyproline-containing HIFα peptides. This sample was therefore subjected to high pH off-line fractionation and analysis by LC-MS/MS to improve the total coverage of peptides in this sample and to prospect for peptides that might have been beyond the limits of detection without fractionation. HIFα peptides were not detected from the fractionated samples, which might have been due to sample losses during the high pH fractionation. These peptides are highly acidic and might have been too hydrophilic at pH 10 to bind to the C18 column. Despite this, collagen peptide enrichment was observed by an increase to

the total number of collagen proteins identified and the total number of Peptide Features (Figure 52). The number of discrete collagen hydroxyproline sites increased from 39 to 250 after high pH fractionation. Only 2 of the 39 sites (COL4A1 HyP1424 and HyP1425) were not identified after fractionation, which highlights the reproducibility of the approach.

High pH fractionation of the GST-VBC-bound material increased the number of Peptide Features containing CTHRC1 HyP75. This approach also facilitated identification of SIPA1L2 HyP1070 HyP1072, which had not been detected in the unfractionated sample.

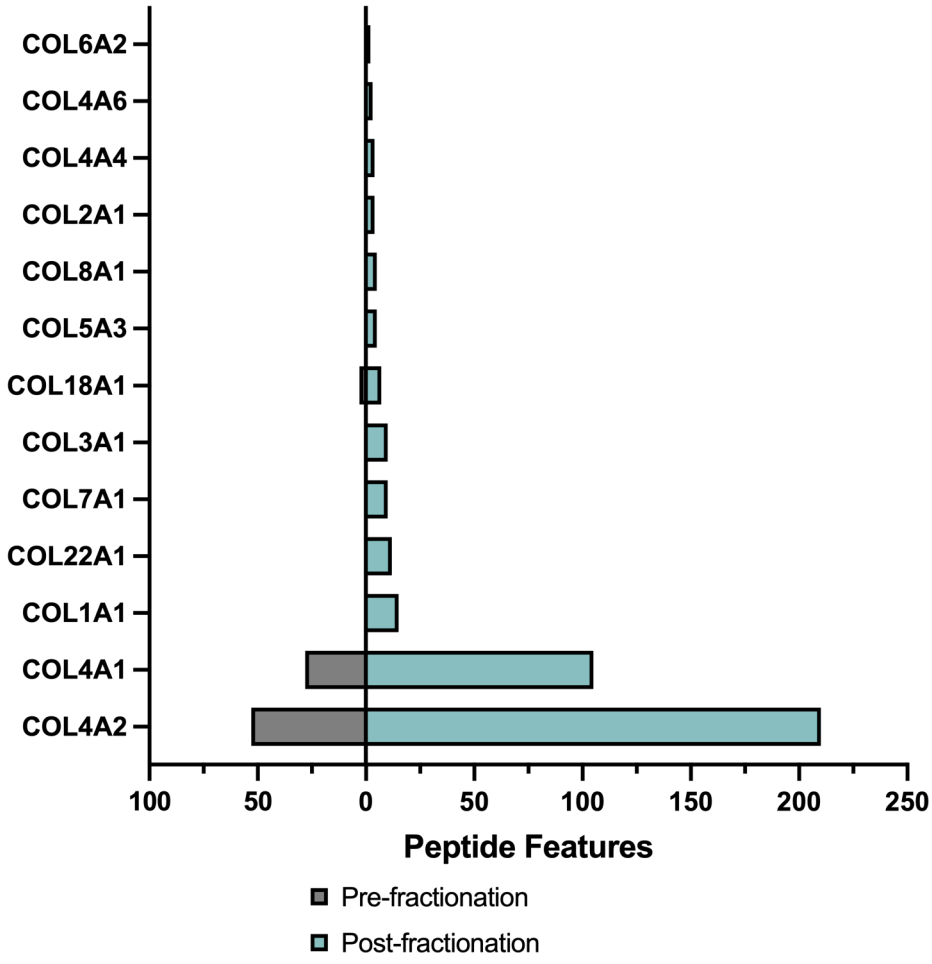


Figure 52 | Number of hydroxyproline-containing Peptide Features before and after high pH fractionation of the GST-VBC-bound material.

5.3.2.4 Combined peptide enrichment results to identify common hydroxyproline sites

The data from each peptide enrichment assay was assimilated to identify secure hydroxyproline-containing peptides that had been identified (Table 53). HIFα peptides (both endogenous and exogenous) were identified, which confirms the approach.

Although a large number of collagen peptides were enriched across all experiments, Table 53 only shows data relating to a triply hydroxylated COL4A2 peptide (aa109-129) (Figure 59) that was highly abundant in each assay. Enrichment of hydroxylated peptides derived from collagen was consistent between peptide enrichment assays and high pH fractionation identified prolyl hydroxylated peptides from less abundant collagen proteins (Figure 52). Additionally, only a small number of peptides that would be expected to derive from the large triple helical domains were enriched in these assays. These results suggest there is specific enrichment of certain collagen peptides. Importantly, these peptides are always derived from the triple helical domains, which are the sites of C-P4H-catalysed prolyl hydroxylation. This suggests collagen peptides derived from the triple helical domains could be used as a positive control in lieu of HIFα peptides.

After establishing that the assay does enrich for hydroxyproline-containing peptides, the novel sites of prolyl hydroxylation were analysed to discover possible 2OGD substrates. These are described in Table 53 and discussed in more detail in Section 5.4.1.

Table 53 | High confidence hydroxyproline-containing peptides detected in the peptide enrichment assays.

Peptides were detected in the GST-VBC-bound fraction (except for identification in the wash fraction of experiment 2). The abundance of each peptide is determined by the number of Peptide Features detected. Peptide Features = 0 indicates the peptide was detected but the abundance was below the minimum threshold (i.e., Area < 1 x10^3). '-' indicates the peptide was not detected in the sample. *PNRC1 HyP20 was only identified in combination with F23 oxidation.

Protein	HyP site(s)	Peptide Features detected						
		Exp. 1	Exp. 2		Exp. 3			
		14 µM spike	No spike	5 nM spike	Long wash	Short wash	Short wash Rapigest	Long wash with fractionation
HIF MMAA	HyP564	7	n/a	-	n/a	n/a	n/a	n/a
HIF1α	HyP564	-	-	-	3	3	3	-
HIF2α	HyP405	-	-	-	3	-	-	-

COL4A2	HyP114 HyP120 HyP123	5	6 (wash)	5 (wash)	6	6	6	13
FKBP10	HyP36		3 (wash)	4 (wash)	-	-	-	5
KHDRBS1	HyP297 HyP299	-	8	5	0	-	0	9
SIPA1L2	HyP1070 HyP1072	-	18	14	-	-	-	1
CTHRC1	HyP59 HyP63	-	-	-	-	-	1	8
	HyP75 HyP78 HyP81	-	-	-	3	2	3	8
PRKCSH	HyP290	1	3	3	-	-	-	-
WIPF2	HyP328	1	3	0	-	-	-	-
VCAN	HyP2675	-	-	-	2	2	2	-
ARHGAP35	HyP115	-	-	-	1	-	-	-
PNRC1	HyP20*	-	-	-	-	1	6	-

5.4 Discussion

The work comprising this chapter aimed to develop a method for directly enriching hydroxyproline-containing peptides. pVHL was used as the affinity reagent because it had been demonstrated to interact with a range of synthetic peptides containing hydroxyproline. Moreover, the use of pVHL as a tool to probe for hydroxyproline in the proteome might provide an insight into HIFα-independent signalling pathways regulated by pVHL that are perturbed in familial VHL disease, including clear cell renal cell carcinoma (ccRCC). The efficacy of recombinant pVHL was confirmed in peptide binding assays using SPR and protein affinity enrichment. In the protein enrichment experiments, DMOG was used to provide evidence for hydroxylation-dependent interactions. Both of these methods confirmed the recombinant trimeric GST-VBC complex was suitable for developing a peptide-based enrichment strategy.

5.4.1 Peptide enrichment by GST-VBC affinity purification

The results presented in this chapter indicate successful development of an affinity reagent to enrich hydroxyproline-containing peptides in an unbiased manner. The success

of this method was determined by identification of three positive controls. First, the affinity purification exhibited specific enrichment of the HIF1α MMAA HyP564 peptide that had been spiked into the tryptic RCC4 digest at an equimolar concentration to the non-hydroxylated, P564, counterpart. Secondly, many hydroxyproline-containing peptides derived from collagen were identified in all experiments. The identification of collagen peptides was important because it demonstrated enrichment of hydroxyproline sites that did not conform to the HIFα-like sequences that have previously been described as recognition sites for pVHL *in vivo* (Hon *et al.*, 2002; Jaakkola *et al.*, 2001). Finally, endogenous hydroxyproline-containing peptides derived from HIF1α and HIF2α were identified. These results were of particular significance because no other enrichment method has identified these peptides to date.

Despite successfully enriching endogenous hydroxyproline-containing HIFα peptides, they were not identified after high pH fractionation. These peptides contain methionine residues, which are prone to non-enzymatic oxidation. It is therefore possible that the additional sample handling increased the likelihood of artefactual oxidation occurring on these peptides and this would have effectively split the signal of the precursor ions such that they were not selected for fragmentation. Additionally, the tryptic peptides containing HIF1α HyP564 and HIF2α HyP402 contain many acidic residues, which become charged at high pH, increasing the hydrophilicity of the peptide. This could have contributed to reduced retention on the hydrophobic C18 columns used with off-line fractionation. Alternative off-line fractionation methods might have been better suited to resolving GST-VBC target peptides. For example, strong anion exchange might have been a more suitable method to retain the endogenous HIFα peptides since this has been shown to favour detection of acidic peptides (Ritorto et al., 2013). Despite these difficulties, the application of GST-VBC as a hydroxyproline affinity reagent facilitated identification of several novel hydroxyproline candidates, which are discussed below.

5.4.1.1 CTHRC1

CTHRC1 was identified in peptide enrichment experiment 3 with multiple hydroxyproline sites (HyP59, HyP63, HyP75, HyP78, HyP81), which were observed in a collagen-like repeating 'X-Y-G' triplet sequence. The proximity of multiple hydroxylation sites on two separate tryptic peptides (HyP59 and HyP63 (Figure 64), and

HyP75, HyP78, and HyP81 (Figure 65)) probably increases the apparent binding affinity to GST-VBC, which aids the enrichment of these peptides. The sequence similarity with collagens implicates C-P4H as the hydroxylase that catalyses prolyl hydroxylation of CTHRC1. One of the reported hydroxyproline sites, HyP59, bears similarity to FKBP10 HyP36 because the hydroxylation occurs in a 'PAG' sequence. Although a second prolyl hydroxylase might be involved in the modification of PAG sequences, the proximity of CTHRC1 HyP59 to the other hydroxyproline sites suggests C-P4H may also catalyse prolyl hydroxylation of PAG sequences.

In addition to peptide enrichment, CTHRC1 was enriched at the protein level with GST-VBC in a DMOG sensitive manner, which implies the interaction is dependent on prolyl hydroxylation.

The peptide assays indicate a direct interaction between CTHRC1 tryptic peptides and GST-VBC can occur *ex vivo*, but further research is required to determine if pVHL and CTHRC1 interact in cells. This might be relevant to VHL disease whereby pVHL-dependent regulatory processes are thought to be contribute to diseases arising from *VHL* mutations, such as ccRCC and phaeochromocytoma. With this in mind, the expression of CTHRC1 was examined in the human protein atlas (HPA) (https://www.proteinatlas.org/, accessed 31/03/22) (Uhlén et al., 2015). The expression of this gene was shown to be an unfavourable prognostic marker in renal cancer.

5.4.1.2 SIPA1L2

In the peptide enrichment assays, SIPA1L2 was identified with two hydroxyproline sites (HyP1070 and HyP1072) in experiments 2 and 3 (Figure 56). These hydroxyproline sites do not conform to HIFα-like ('LXXLAP') or collagen-like ('XPG') hydroxylation consensus motifs, so it is not apparent which prolyl hydroxylase would catalyse these modifications. There is some evidence for an interaction occurring between SIPA1L2 and transiently expressed PHD3 in a DMOG sensitive manner (Rodriguez *et al.*, 2016), which raises the possibility this could be a genuine non-HIFα PHD substrate. However, this protein did not reach the minimum threshold set by the authors to be considered a PHD substrate.

In addition to these results, SIPA1L2 was found to interact with GST-VBC at the protein level in a DMOG-sensitive manner, which also suggests the interaction at the protein

level is dependent on prolyl hydroxylation. In renal cancer, high SIPA1L2 has been shown to be an unfavourable prognostic marker, at least according to the HPA.

5.4.1.3 KHDRBS1

Two hydroxyproline sites were identified for KHDRBS1, HyP297 and HyP299, in the experiments 2 and 3 of the peptide enrichment assays (Figure 55). These hydroxyproline sites occur in a proline-rich domain and conform to the 'P-HyP' sequences identified as a putative hydroxylation consensus sequence in Chapter 3. KHDRBS1 has 4 proline-rich domains but hydroxylation was only observed in the region spanning aa295-301. The hydroxylation sites were observed on both a fully cleaved and mis-cleaved tryptic peptide. The identification of these hydroxylation sites on two independent peptide sequences increases the confidence of the assignments.

In the protein enrichment assay, KHDRBS1 was exclusively identified in the vehicle-treated sample, which suggests the interaction with GST-VBC is prolyl hydroxylation-dependent. However, KHDRBS1 was not initially recognised because there was only 1 peptide identified, which was below the threshold required for confident protein assignment (unique peptides ≥ 2). This peptide also contained the two hydroxyproline sites identified in the peptide enrichment assay, which further strengthens the security of the hydroxyproline assignments.

Interestingly, the hydroxylations are always reported with high confidence at P297 and P299 despite 6 prolyl residues being present in the proline-rich region (aa295-301). This suggests a high degree of specificity by the putative prolyl hydroxylase. According to the HPA, high expression of KHDRBS1 is an unfavourable prognostic marker of renal and liver cancer.

5.4.1.4 FKBP10

FKBP10 HyP36 was established as a 2OGD substrate (Chapter 4) and the enrichment of FKBP10 by GST-VBC decreased after DMOG treatment in the protein enrichment assay. FKBP10 has been reported to be a collagen chaperone (Ishikawa et al., 2017). It is possible, therefore, that the loss of FKBP10 binding following DMOG treatment was a consequence of binding to collagen, which is the DMOG-sensitive interactor of GST-VBC. The protein enrichment assay does not clarify this question in the analysis.

In the peptide enrichment assays, FKBP10 HyP36 peptides were identified in experiments 2 and 3. In experiment 2, FKBP10 HyP36 peptides dissociated from GST-VBC in later wash fractions than the P36 peptide. Although this does not prove it was captured by GST-VBC, these results suggest specific retention of the hydroxylated species. FKBP10 is a highly abundant protein, and several non-hydroxylated peptides were identified in the GST-VBC-bound fraction of experiment 3, including the P36 peptide. Despite this, the stoichiometry of HyP36 increased from 77 % (pre-cleared input) to 98 % (long wash with high pH fractionation), which suggests specific enrichment of the hydroxylated peptide. This contrasts with other peptides identified in the GST-VBC-bound fraction that did not appear to be specifically enriched because the stoichiometry of the hydroxyproline site decreased in the GST-VBC-bound fraction.

Consistent with the findings observed in this chapter, the FKBP10 HyP36 peptide was also identified in a previous report of GST-VBC-mediated hydroxyproline-containing peptide enrichment assay (Arsenault *et al.*, 2015). According to the HPA, high expression of FKBP10 is an unfavourable prognostic marker of renal, thyroid, cervical and urothelial cancers.

5.4.1.5 PRKCSH

PRKCSH HyP290 was previously identified by *in silico* analysis of 13 human cell lines (Chapter 3). The detection of this peptide in affinity purified material from experiments 2 and 3 (Figure 62) increases the security of the hydroxyproline assignment because it has now been identified in multiple experiments using orthogonal proteomic approaches. Although non-hydroxylated PRKCSH peptides are detected in all samples of experiment 3, including the GST-VBC-bound fraction, the HyP290 peptide was only detected in this fraction (i.e., it was not observed in the pre-cleared input). This suggests specific enrichment of a low stoichiometry modification in RCC4 cells. According to the HPA, high expression of PRKCSH is an unfavourable prognostic marker of renal cancer.

5.4.1.6 Other candidates

Several other hydroxyproline sites have been detected following GST-VBC affinity purification (Table 53). Although the hydroxyproline assignments are secure (Section 5.5), the identification of these hydroxyproline sites in a single experiment means these

are regarded as low confidence assignments. These proteins were not identified in the protein enrichment assay or in any previous hydroxyproline screen, so more data is required to increase the security of their candidacy as novel prolyl hydroxylase substrates. According to the HPA, VCAN has been reported to be an unfavourable prognostic marker of renal cancer.

PNRC1 HyP20 was identified in two of the affinity enriched samples of experiment 3 (Figure 58). Manual inspection of the MS2 spectra indicates this site of hydroxylation is confidently localised. However, the peptide was only detected in the presence of F23 oxidation. Throughout this thesis, the identification of prolyl hydroxylation in combination with another site of oxidation has been considered to reduce the confidence of the peptide assignment. However, the MS2 spectra suggest this is a confident assignment, and F23 oxidation is isobaric to a polymorphism that would generate a F23Y mutation. Furthermore, PNRC1 was enriched at the protein level and the same hydroxylated peptide was identified in that analysis. The protein interaction with GST-VBC was not significantly perturbed by DMOG treatment (log_2FC(DMOG/vehicle) = -1.2), which might be due to long half-life. Together, these results suggest a direct interaction between PNRC1 and GST-VBC might exist.

WIPF2 HyP328 (Figure 63) was detected in multiple experiments peptide enrichment assays but not at the protein level. VCAN HyP2675 (Figure 60) and ARHGAP35 HyP115 (Figure 61) were only detected in experiment 3 of the peptide enrichment assays. Although the MS1 and MS2 inspection provided confidence the assignments were correct, further data is required to increase the candidacy of these hydroxyproline sites. Despite a low overall number of non-HIFα and non-collagen hydroxyproline-containing peptides, several hydroxyproline sites identified by GST-VBC affinity purification were identified with high confidence of correct assignment.

Several sites of prolyl hydroxylation that have been established in previous chapters of this thesis were not enriched with GST-VBC. For example, peptides containing LMAN1 HyP378 and HYOU1 HyP977 were identified with secure hydroxyproline assignments in the input and flow through of experiment 3 but not in the GST-VBC-bound fraction.

This suggests these peptides are not recognised by GST-VBC. Similarly, several PDI hydroxyproline sites were identified in the input samples of experiments 2 and 3 but were not identified in the GST-VBC-bound fractions. However, it should be noted the PDIs are reduced and alkylated prior to trypsin digestion. Thus, a direct interaction between GST-VBC and PDI might occur but this interaction would be perturbed in *ex vivo* experiments by carbamidomethylated cysteines.

Together, these experiments indicate GST-VBC affinity purification can enrich HIFα and non-HIFα hydroxyproline-containing peptides. Interestingly, several of the candidate proteins were also enriched in a DMOG-sensitive manner, supporting the hypothesis that GST-VBC-binding is mediated by hydroxyproline. It also raises the possibility that these proteins might interact with pVHL *in cellulo*. HIFα ('LXXLAP') and collagen ('XPG') represent distinct prolyl hydroxylation consensus sequences. Although CTHRC1 is hydroxylated in a collagen-like domain, the other hydroxyproline sites enriched by GST-VBC did not conform to either hydroxylation consensus sequence. KHDRBS1 HyP297 and HyP299 are both hydroxyproline sites occurring within 'P-HyP' sequences, which were identified as a novel hydroxylation consensus sequence in Chapter 3. Together, this suggests GST-VBC is an effective affinity reagent to capture a broad range of hydroxyproline-containing peptides that might result from the activity of different prolyl hydroxylase enzymes.

5.4.1.7 Reported non-HIFα PHD substrates

The peptide enrichment assays demonstrated the application of GST-VBC to enrich for hydroxyproline-containing peptides. I therefore performed targeted analyses of the hydroxyproline sites in the reported non-HIF PHD substrates in these assays. Although some target peptides were identified in my screens, they were not specifically enriched when compared to the input, and manual inspection of the MS2 spectra indicated misassigned artefactual oxidation of methionine. Furthermore, none of the reported non-HIFα PHD substrates were found to be enriched in a DMOG-sensitive manner at the protein level. Taken together, the data presented in this chapter does not support any of the reported interaction between non-HIFα PHD substrates and pVHL. This could be due

to a necessity of additional factors that were not present in the RCC4 extracts used in this thesis, or an absence of hydroxyproline on the reported substrates.

5.4.1.8 Reported pVHL substrates

In addition to several reported non-HIFα PHD substrates there have been reports of non-HIFα pVHL substrates. These were analysed in all protein and peptide enrichment assays for evidence of direct association between the reported substrate and GST-VBC.

Data from the peptide enrichment assays indicates the interaction with collagen occurs directly between GST-VBC and the hydroxyproline residues of collagen. This supports the reported interaction between COL4A2 and pVHL (Grosfeld *et al.*, 2007; Kurban et al., 2008). It has been suggested that pVHL does not interact with fibrillar collagens when they have been incorporated into a triple helical structure (Grosfeld *et al.*, 2007). Consistent with this, the protein enrichment assay demonstrated the interaction of type IV, VI, VIII and IX collagens (i.e., non-fibrillar collagens) was reduced after treatment with DMOG, but type I and type II collagens (i.e., fibrillar collagens) (Ricard-Blum, 2011) did not display DMOG sensitive enrichment. Tryptic peptides are unlikely to exist in trimeric helixes, and hydroxyproline-containing peptides derived from fibril-forming collagens were enriched in the peptide enrichment assays. These results support the previous observation of an interaction between pVHL and denatured fibrillar collagens, and suggest pVHL can interact with collagens if the hydroxyprolyl residues are accessible.

To further explore the interactions between GST-VBC and collagen, the data from experiments 1-3 of the peptide enrichment assays was assimilated and peptides identified in multiple experiments were further analysed. Although more than 250 distinct hydroxyproline-containing peptides derived from collagen were identified in the totality of the data, only three peptides were identified in every experiment: these peptides corresponded to COL4A2 HyP114 HyP120 HyP123, COL4A2 HyP707 HyP716, and COL4A1 HyP399 HyP404. Although COL4A2 and COL4A1 are the most abundant collagens in RCC4 cells, it is surprising that so few peptides are identified in all experiments. The triple helical domains are expansive (COL4A1: 1,266aa; COL4A2: 1,299aa), so capture of only a relatively small number of hydroxyproline-containing

peptides indicates some specificity of GST-VBC binding. The interaction between pVHL and collagen is discussed in more detail in Chapter 7.

In contrast to collagen, there was no evidence of an interaction between GST-VBC and FN1 at the peptide level despite previous reports of a direct interaction (Hoffman *et al.*, 2001; Iwai *et al.*, 1999; Ohh *et al.*, 1998). However, FN1 was shown to be enriched at the protein level in a DMOG-sensitive manner. Mature ECM proteins, including FN1, were likely to be under-represented in the peptide enrichment assays given the method of cell harvest, which might explain its absence in the data presented here.

pVHL has also been reported to interact with tubulin in cells (Hergovich *et al.*, 2003). A hydroxyproline-containing peptide of TUBB3 has been identified in Chapter 3. However, intervention studies with hypoxia and DMOG did not indicate regulation of the hydroxylation site (Chapter 4). This peptide, corresponding to TUBB3 HyP272, was also observed in the peptide enrichment assays in this chapter. However, the peptide was more abundant in the input samples than the GST-VBC-bound fraction, and many other non-hydroxylated peptides derived from TUBB3 were also identified in this fraction. In the protein enrichment assay, the hydroxylated peptide was not detected and TUBB3 was not shown to decrease after treatment with DMOG, which suggests the interaction with GST-VBC is not dependent on prolyl hydroxylation. Whilst this protein is highly abundant and has been observed in the GST-VBC-bound fraction there was no evidence for specific enrichment of a hydroxyproline-containing peptide.

5.4.2 Evaluation of recombinant GST-VBC as an affinity reagent to enrich hydroxyproline-containing peptides

The application of recombinant GST-VBC to enrich hydroxyproline-containing peptides proved to be successful, as demonstrated by enrichment of endogenous HIFα peptides that contain target sites of PHD-catalysed prolyl hydroxylation. Several collagen peptides were enriched and could be considered as positive controls in future peptide enrichment assays.

GST-VBC has been used to enrich hydroxyproline-containing peptides in a previous report but HIFα and collagen peptides were not identified using this method (Arsenault *et al.*, 2015). Despite this, FKBP10 HyP36 was identified in their data, which is consistent with the data presented in this chapter. In another enrichment assay, Chen and colleagues

used an antibody to capture hydroxyproline-containing peptides (Zhou *et al.*, 2016). The authors reportedly observed 562 peptides containing hydroxyproline, which included collagen peptides and FKBP10 HyP36. In the data presented in this chapter, fewer non-collagen peptides were identified than in the data presented by Chen and colleagues. This might be cell-type specific or could represent greater substrate recognition specificity of GST-VBC compared to the antibody used in their assay.

Several hydroxyproline motifs ('HyP-G', 'P-HyP', 'HyP-X-P', 'P-X-HyP', and 'HyP-P-P') have been identified (Zhou *et al.*, 2016). Consistent with this, the results presented from the peptide enrichment assays in this chapter indicate novel candidates that are consistent with these motifs: CTHRC1 ('HyP-G'), KHDRBS1 ('P-HyP' and 'HyP-P-P') and SIPA1L2 ('P-X-HyP' and 'HyP-X-P'). I also identified 'HyP-G' and 'P-HyP' motifs in the *in silico* analysis of publicly deposited deep proteome data (Chapter 3). Together, the results of the peptide enrichment assays presented in this chapter show consistency with other reports of hydroxyproline-containing peptide affinity purification

5.4.3 Limitations of GST-VBC as an affinity reagent for peptide enrichment assays

5.4.3.1 Physiological relevance

The application of GST-VBC to enrich hydroxyproline-containing peptides has enabled identification of several novel prolyl hydroxylase substrate candidates. However, this does not equate to identification of novel cellular pVHL substrates. For example, COL4A2 was captured by GST-VBC in *ex vivo* protein enrichment assays but was not found to be subject to pVHL-catalysed ubiquitylation (Grosfeld *et al.*, 2007). Given its role as a tumour suppressor, it would be of interest to determine if any of the putative hydroxylase substrates are physiological targets for degradation by pVHL-catalysed ubiquitylation. This could be achieved directly by ubiquitylation assays or indirectly by assessing protein abundance in cells displaying different pVHL status.

5.4.3.2 Competitive binding to pVHL

The aim of these experiments was to apply GST-VBC to capture hydroxyproline-containing peptides. The hydroxyproline binding pocket of pVHL can interact with a single peptide at any time, so competition between target peptides exists. A large molar

excess of GST-VBC was used to limit this, but some evidence of competitive binding to GST-VBC remained.

In the first peptide enrichment experiment, HIF1α MMAA peptides (P564 and HyP564 forms) were spiked into the tryptic RCC4 digest at a roughly equimolar concentration to the GST-VBC complex. The abundance of the HyP564 peptide was enriched approximately 50-fold relative to the P564 peptide, which indicates specific enrichment *via* direct interaction with the hydroxyproline binding domain of GST-VBC. However, it is possible that the exogenous peptides competed with endogenous target peptides and consequently limited the number of targets that could be identified. In experiments 2 and 3 there was either no synthetic peptide, or the concentration was greatly reduced. In these experiments there was an increased number of hydroxyproline-containing peptides identified, suggesting the previous hypothesis was correct.

Trypsin was applied to aid RCC4 cell detachment during harvesting, which meant it was unlikely that the samples contained extracellular collagen. It was therefore surprising to observe highly abundant collagen proteins in these experiments, particularly COL4A2 and COL4A1, which suggests ccRCCs express high levels of type IV collagens and retain much of this protein intracellularly. It is possible the abundance of collagen peptides competed with target peptides. Although pVHL does not appear to bind fibrillar collagens in their native condition (Grosfeld *et al.*, 2007), the results in this chapter suggest trypsin digest facilitated GST-VBC binding. This further increased the number of collagen peptides occupying the hydroxyproline binding pocket of GST-VBC in the peptide enrichment assays. It is likely that the potential to discover novel hydroxyproline-containing peptides using GST-VBC affinity purification could be improved by performing the same experiments in different cell lines that exhibit lower total collagen expression.

5.4.3.3 Hydroxyproline-containing peptides captured by GST-VBC

The exact binding preferences of pVHL are not completely understood. pVHL is unlikely to bind all hydroxyproline sites, which could be due to steric hindrance or peptide sequences that are incompatible with the hydroxyproline-binding domain of pVHL. Several hydroxyproline-containing peptides previously assigned in this thesis were not observed in the peptide enrichment assays (e.g., PDIs and OS9). This suggests there is

some sequence specificity for GST-VBC binding to substrates. Lee and colleagues mutated residues outside the hydroxyproline binding pocket in an attempt to reduce the specificity of the substrate binding domain (Arsenault *et al.*, 2015). It is not clear if this approach was successful but a similarly mutated GST-VBC construct could be compared with the one used in these assays to determine if this would increase the number of hydroxyproline-containing peptides that are identified in the peptide enrichment assays.

5.4.3.4 pVHL-independent proteolysis

Some proteins might undergo prolyl hydroxylation-mediated proteolysis *via* an alternative ubiquitin E3 ligase and would not be stabilised in pVHL-defective cells. This has been reported for CEP192, in which SKP1 is the ubiquitin E3 ligase reported to initiate degradation of prolyl hydroxylated CEP192 (Moser *et al.*, 2013). Although it could be argued pVHL might not recognise these proteins, there remains a possibility of capturing the hydroxylated peptide *ex vivo*. Experiments using proteasome inhibition, such as the MEF SILAC samples from Chapter 4, would be required to identify such proteins.

5.4.3.5 Limited sensitivity

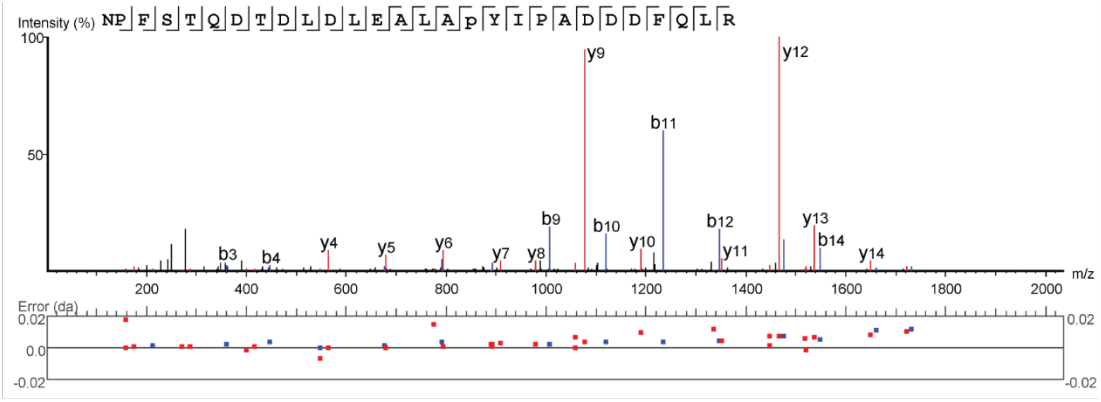
The application of GST-VBC to capture hydroxyproline-containing peptides was successful. However, on-bead tryptic digest causes high background in the analyte. In addition to the tryptic peptides derived from GST-VBC, many non-specific (i.e., non-hydroxylated) peptides are also present in the digested material. Although pre-clearing appeared to reduce the abundance of non-specific interactors, a competitive elution strategy could be optimised to reduce the non-hydroxylated background that is derived from the on-bead digest. This would result in a single peptide that would most likely be present at a high concentration in the LC-MS/MS analysis. However, this would still be preferable to the on-bead digest because the elution peptide would only occupy a single region of the chromatogram, thereby increasing the overall sensitivity of the assay.

Overall, the development of a hydroxyproline affinity reagent has been performed and was successfully applied to i) enrich endogenous HIFα peptides, ii) enrich previously established sites of prolyl hydroxylation (e.g., collagen and FKBP10 HyP36) and iii)

identify novel sites of prolyl hydroxylation. Several limitations have been discussed with suggestions for future method development. It should be possible to apply this strategy to continue to identify novel sites of hydroxyproline from a range of cell lines and tissues. Additionally, the candidates identified from this series of experiments might represent physiological substrates of pVHL that could be dysregulated in diseases containing *VHL* mutations.

5.5 Appendix

A



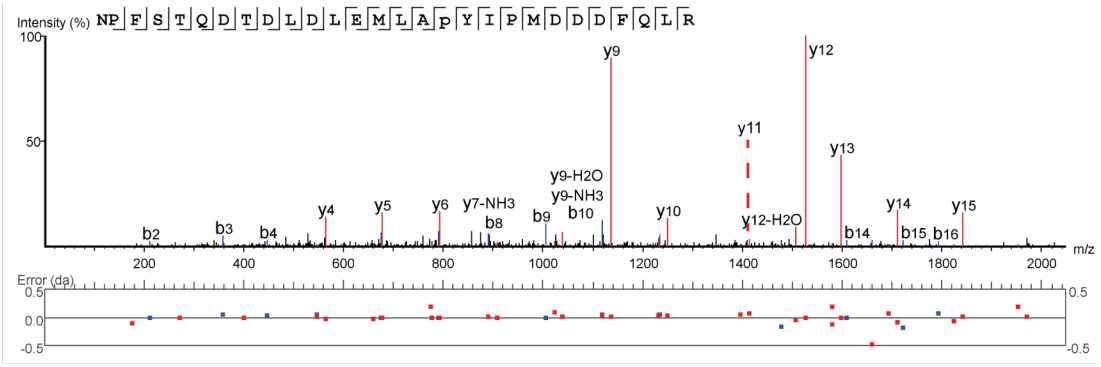
B

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	87.0558	115.0508	N				28
2	70.0656	212.1020	P	3082.4526	3064.4421	3065.4255	27
3	120.0812	359.1695	F	2985.3999	2967.3894	2968.3728	26
4	60.0449	446.1995	S	2838.3315	2820.3210	2821.3044	25
5	74.0605	547.2513	T	2751.2996	2733.2891	2734.2725	24
6	101.0714	675.3085	Q	2650.2517	2632.2412	2633.2246	23
7	88.0398	790.3328	D	2522.1931	2504.1826	2505.1660	22
8	74.0605	891.3826	T	2407.1663	2389.1558	2390.1392	21
9	88.0398	1006.4094	D	2306.1187	2288.1082	2289.0916	20
10	86.0969	1119.4918	L	2191.0916	2173.0811	2174.0645	19
11	88.0398	1234.5183	D	2078.0076	2059.9971	2060.9805	18
12	86.0969	1347.6018	L	1962.9806	1944.9701	1945.9536	17
13	102.0554	1476.6416	E	1849.8966	1831.8861	1832.8696	16
14	44.0499	1547.6813	A	1720.8431	1702.8435	1703.8270	15
15	86.0969	1660.7594	L	1649.8083	1631.8064	1632.7899	14
16	44.0499	1731.7954	A	1536.7258	1518.7158	1519.7080	13
17	86.0605	1844.8553	P(+15.99)	1465.6879	1447.6776	1448.6669	12
18	136.0761	2007.9187	Y	1352.6429	1334.6376	1335.6086	11
19	86.0969	2121.0027	I	1189.5751	1171.5742	1172.5577	10
20	70.0656	2218.0554	P	1076.4966	1058.4829	1059.4740	9
21	44.0499	2289.0925	A	979.4453	961.4373	962.4209	8
22	88.0398	2404.1196	D	908.4075	890.3976	891.3826	7
23	88.0398	2519.1465	D	793.3824	775.3580	776.3569	6
24	88.0398	2634.1733	D	678.3563	660.3464	661.3300	5
25	120.0812	2781.2419	F	563.3293	545.3194	546.3098	4
26	101.0714	2909.3003	Q	416.2607	398.2510	399.2367	3
27	86.0969	3022.3845	L	288.2021	270.1924	271.1752	2
28	129.1139		R	175.1180	157.0906	158.0913	1

Figure 53 | Representative MS2 spectrum of HIF1α MMAA HyP564 peptide.

A) MS2 spectrum and, B) ion table, of the HIF1α MMAA HyP564 peptide identified in the GST-VBC-bound fraction of the peptide enrichment assay (experiment 1). Confident peptide sequencing is achieved in the y-ion series. The diagnostic ions (y11 and y12) confirm the presence of hydroxyproline at the target site.

A



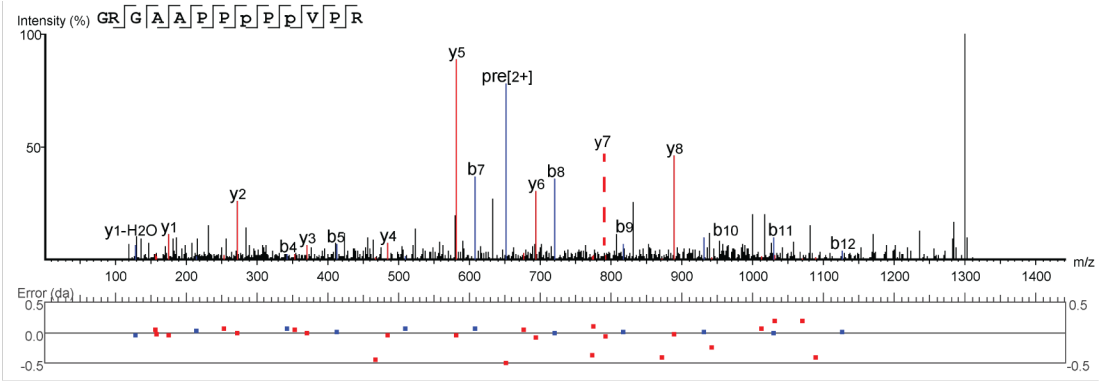
B

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	87.0558	115.0508	N				28
2	70.0656	212.1080	P	3202.4595	3184.4490	3185.4324	27
3	120.0812	359.1010	F	3105.4067	3087.3962	3088.3796	26
4	60.0449	446.1490	S	2958.3384	2940.3279	2941.3113	25
5	74.0605	547.1760	T	2871.3062	2853.2957	2854.2791	24
6	101.0714	675.3280	Q	2770.2585	2752.2480	2753.2314	23
7	88.0398	790.3270	D	2642.2000	2624.1895	2625.1729	22
8	74.0605	891.3540	T	2527.1731	2509.1626	2510.1460	21
9	88.0398	1006.3980	D	2426.1252	2408.1147	2409.0981	20
10	86.0969	1119.4360	L	2311.0984	2293.0879	2294.0713	19
11	88.0398	1234.4570	D	2198.0144	2180.0039	2180.9873	18
12	86.0969	1347.6068	L	2082.9873	2064.9768	2065.9602	17
13	102.0554	1476.8250	E	1969.8719	1951.6810	1952.8763	16
14	104.0533	1607.7010	M	1840.8361	1822.9150	1823.8337	15
15	86.0969	1720.9600	L	1709.9020	1691.8098	1692.7080	14
16	44.0499	1791.7130	A	1596.7520	1578.8480	1579.4940	13
17	86.0605	1904.8588	P(+15.99)	1525.6830	1507.7290	1508.6721	12
18	136.0761	2067.9221	Y	1412.5560	1394.6409	1395.5649	11
19	86.0969	2181.0061	I	1249.5360	1231.5360	1232.5612	10
20	70.0656	2278.0588	P	1136.4680	1118.4351	1119.4360	9
21	104.0533	2409.0994	M	1039.4210	1021.4407	1022.3190	8
22	88.0398	2524.1262	D	908.4050	890.4002	891.3540	7
23	88.0398	2639.1533	D	793.3730	775.1600	776.3720	6
24	88.0398	2754.1802	D	678.3440	660.3680	661.3300	5
25	120.0812	2901.2485	F	563.3550	545.3194	546.2820	4
26	101.0714	3029.3071	Q	416.2616	398.2510	399.2470	3
27	86.0969	3142.3914	L	288.2030	270.1924	271.1590	2
28	129.1139		R	175.2270	157.1084	158.0919	1

Figure 54 | Representative MS2 spectrum of HIF1α HyP564.

MS2 spectrum (A) and ion table (B) of the tryptic fragment containing HIF1α HyP564 from the GST-VBC-bound fraction of the peptide enrichment assay (experiment 3; long wash). Confident peptide sequencing is achieved in the y-ion series and the absence of -64 Da neutral loss ions confirms methionine oxidation is not present on this peptide. The diagnostic ions (y11 and y12) confirm the location of hydroxyproline at the target site. The MS2 spectrum is annotated to indicate the presence of the y11 ion (red dashed line).

A



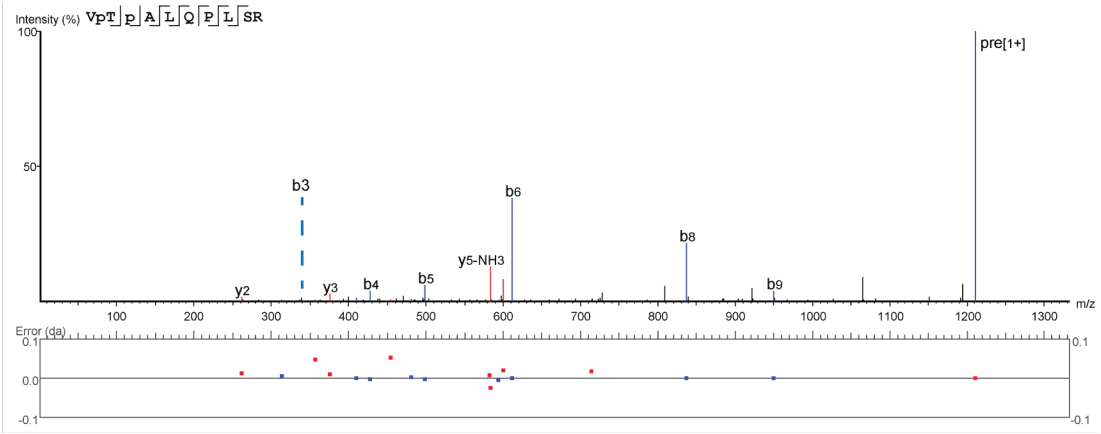
B

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	30.0343	58.0293	G				13
2	129.1610	214.0850	R	1243.6904	1225.6799	1226.6635	12
3	30.0343	271.1519	G	1088.0060	1069.5789	1070.3470	11
4	44.0499	342.1040	A	1030.3510	1012.4820	1013.5410	10
5	44.0499	413.1970	A	959.5308	941.7650	942.5038	9
6	70.0656	510.1960	P	888.5130	870.4832	871.8730	8
7	70.0656	607.2580	P	791.4970	773.7990	774.3030	7
8	86.0605	720.3920	P(+15.99)	694.4730	676.3110	677.3612	6
9	70.0656	817.4040	P	581.3770	563.3300	564.3135	5
10	86.0605	930.4540	P(+15.99)	484.3280	466.7260	467.2608	4
11	72.0812	1029.5620	V	371.2340	353.1720	354.1490	3
12	70.0656	1126.5820	P	272.1800	254.0730	255.1447	2
13	129.1139		R	175.1680	157.0500	158.1240	1

Figure 55 | Representative MS2 spectrum of KHDRBS1 HyP297 HyP299.

MS2 spectrum (A) and ion table (B) of the tryptic fragment containing KHDRBS1 HyP297 HyP299 from the GST-VBC-bound fraction of the peptide enrichment assay (experiment 2; no spike-in peptide). Confident peptide sequencing is achieved in the y-ion series. The diagnostic ions (HyP297: y3 and y4; HyP299; y5 and y6) confirm the location of hydroxyproline at the target site. The MS2 spectrum is annotated to indicate the presence of the y17 ion (red dashed line).

A



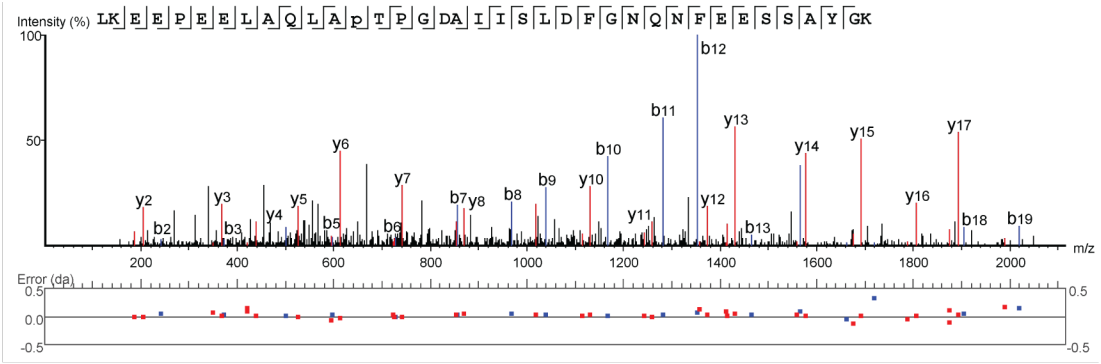
B

#	b	b-H2O	b-NH3	Seq	y	y-H2O	y-NH3	#
1	100.0762	82.0657	83.0492	V				11
2	213.1239	195.1134	196.0969	P(+15.99)	1111.6106	1093.6001	1094.5836	10
3	314.1650	296.1610	297.1446	T	998.5629	980.5523	981.5359	9
4	427.2220	409.2087	410.1900	P(+15.99)	897.5152	879.5046	880.4882	8
5	498.2600	480.2458	481.2250	A	784.4675	766.4570	767.4406	7
6	611.3420	593.3350	594.3134	L	713.4110	695.4199	696.4034	6
7	739.3990	721.3884	722.3720	Q	600.3260	582.3260	583.3460	5
8	836.4500	818.4412	819.4248	P	472.2878	454.2230	455.2608	4
9	949.5380	931.5253	932.5089	L	375.2240	357.1760	358.2080	3
10	1036.5679	1018.5573	1019.5409	S	262.1380	244.1404	245.1240	2
11				R	175.1189	157.1084	158.0919	1

Figure 56 | Representative MS2 spectrum of SIPA1L2 HyP1070 HyP1072.

MS2 spectrum (A) and ion table (B) of the tryptic fragment containing SIPA1L2 HyP1070 HyP1072 from the GST-VBC-bound fraction of the peptide enrichment assay (experiment 2; no spike-in peptide). Confident peptide sequencing is achieved in the b-ion series. The diagnostic ions (HyP1072; b3 and b4) confirm the location of hydroxyproline at the target site. The MS2 spectrum is annotated to indicate the presence of the b3 ion (blue dashed line).

A

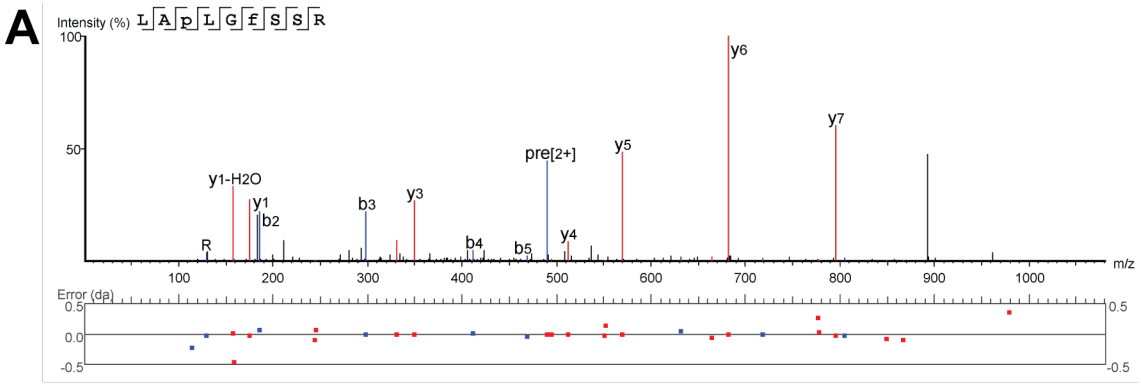


B

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	86.0969	114.0919	L				37
2	101.1078	242.1090	K	3910.8503	3892.8398	3893.8232	36
3	102.0554	371.1780	E	3782.7554	3764.7449	3765.7283	35
4	102.0554	500.2390	E	3653.7129	3635.7024	3636.6858	34
5	70.0656	597.2750	P	3524.6702	3506.6597	3507.6431	33
6	102.0554	726.3690	E	3427.6174	3409.6069	3410.5903	32
7	102.0554	855.3590	E	3298.5750	3280.5645	3281.5479	31
8	86.0969	968.4320	L	3169.5322	3151.5217	3152.5051	30
9	44.0499	1039.4740	A	3056.4482	3038.4377	3039.4211	29
10	101.0714	1167.5601	Q	2985.8291	2967.4006	2968.3840	28
11	86.0969	1280.6219	L	2857.3525	2839.3420	2840.3254	27
12	44.0499	1351.6310	A	2744.1509	2726.2581	2727.2415	26
13	86.0605	1464.7010	P(+15.99)	2673.2314	2655.2209	2656.2043	25
14	74.0605	1565.6960	T	2560.1838	2542.1733	2543.1567	24
15	70.0656	1662.9010	P	2459.0110	2440.6470	2442.1089	23
16	30.0343	1719.5389	G	2362.0833	2344.0728	2345.0562	22
17	88.0398	1834.9073	D	2305.0618	2287.0513	2288.0347	21
18	44.0499	1905.8700	A	2190.0349	2172.0244	2173.0078	20
19	86.0969	2018.8650	I	2118.9978	2100.9873	2101.9707	19
20	86.0969	2132.1125	I	2005.9137	1987.9032	1988.6970	18
21	60.0449	2219.1445	S	1892.7800	1874.9270	1875.6810	17
22	86.0969	2332.2288	L	1805.7610	1787.7871	1788.8170	16
23	88.0398	2447.2556	D	1692.6910	1674.8290	1675.6865	15
24	120.0812	2594.3240	F	1577.6520	1559.6320	1560.6597	14
25	30.0343	2651.3455	G	1430.5410	1412.4960	1413.5649	13
26	87.0558	2765.3884	N	1373.5450	1355.5863	1356.4290	12
27	101.0714	2893.4470	Q	1259.5520	1241.5433	1242.5010	11
28	87.0558	3007.4900	N	1131.4510	1113.4590	1114.4683	10
29	120.0812	3154.5583	F	1017.3940	999.4418	1000.4254	9
30	102.0554	3283.6008	E	870.3250	852.3330	853.3569	8
31	102.0554	3412.6436	E	741.3580	723.2730	724.3190	7
32	60.0449	3499.6755	S	612.3300	594.3660	595.2717	6
33	60.0449	3586.7075	S	525.2780	507.2561	508.2397	5
34	44.0499	3657.7446	A	438.2040	420.0670	421.1030	4
35	136.0761	3820.8081	Y	367.1590	349.0890	350.1706	3
36	30.0343	3877.8293	G	204.1250	186.1270	187.0940	2
37	101.1078		K	147.1128	129.1022	130.0858	1

Figure 57 | Representative MS2 of HIF2α HyP405.

MS2 spectrum (A) and ion table (B) of the tryptic fragment containing HIF2α HyP405 from the GST-VBC-bound fraction of the peptide enrichment assay (experiment 3; long wash). Confident peptide sequencing is achieved in the y-ion and b-ion series. The diagnostic ions (b12 and b13) confirm the location of hydroxyproline at the target site. There are no fragment ions corresponding to -64 Da neutral losses, which might have otherwise indicated the presence of methionine oxidation.

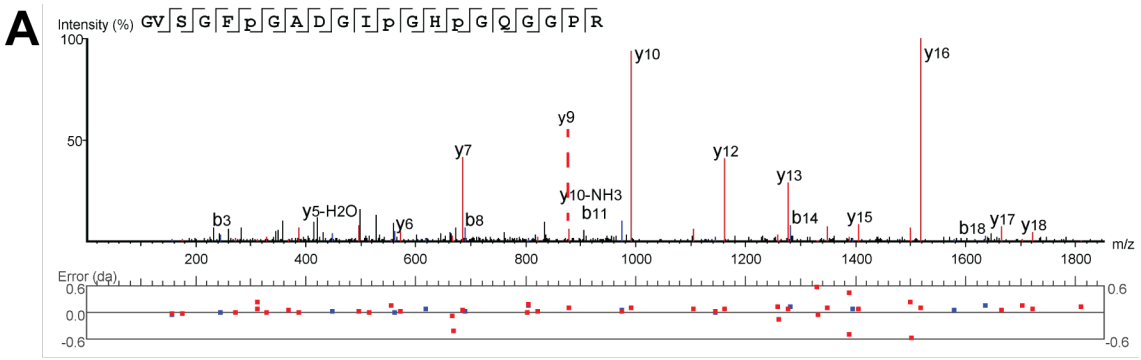


B

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	86.0969	114.3170	L				9
2	44.0499	185.0400	A	866.5400	848.4260	849.4880	8
3	86.0605	298.1650	P(+15.99)	795.4200	777.1020	778.3330	7
4	86.0969	411.2370	L	682.3420	664.4070	665.3248	6
5	30.0343	468.3240	G	569.2780	551.2900	552.0780	5
6	136.0761	631.2890	F(+15.99)	512.2340	494.2440	495.2193	4
7	60.0449	718.3660	S	349.1750	331.1850	332.1560	3
8	60.0449	805.4350	S	262.1510	244.2360	245.0410	2
9	129.1440		R	175.1530	157.0780	158.5520	1

Figure 58 | Representative MS2 spectrum of PNRC1 HyP20.

MS2 spectrum (A) and ion table (B) of the tryptic fragment containing PNRC1 HyP20 from the GST-VBC-bound fraction of the peptide enrichment assay (experiment 3; short wash with RapiGest). Confident peptide sequencing is achieved in the y-ion series. The diagnostic ions (y6 and y7) confirm the location of hydroxyproline at the target site.



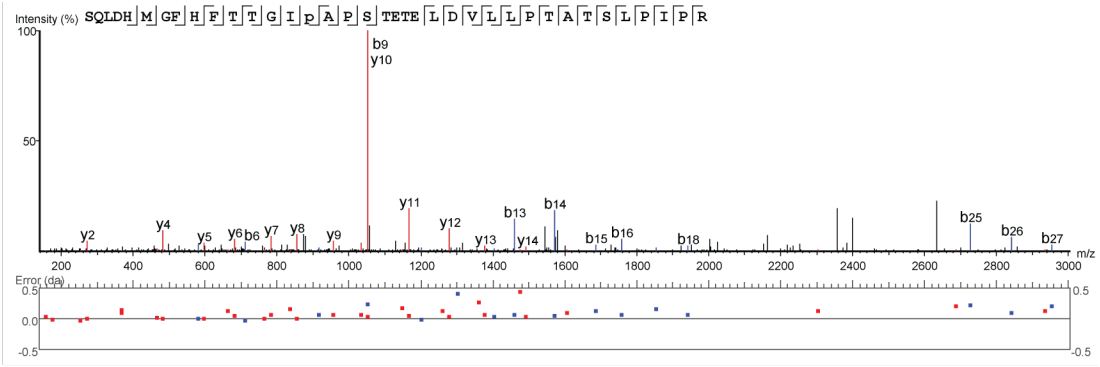
B

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	30.0343	58.0293	G				21
2	72.0812	157.1570	V	1907.8994	1889.8889	1890.8724	20
3	60.0449	244.1170	S	1808.6910	1790.8204	1791.8040	19
4	30.0343	301.1512	G	1721.7020	1703.6080	1704.7720	18
5	120.0812	448.1820	F	1664.7080	1646.7670	1647.7505	17
6	86.0605	561.2770	P(+15.99)	1517.5811	1499.4440	1501.2810	16
7	30.0343	618.1880	G	1404.5740	1387.1670	1387.1670	15
8	44.0499	689.2970	A	1347.5129	1329.0450	1330.6860	14
9	88.0398	804.1820	D	1276.5110	1258.4520	1259.7469	13
10	30.0343	861.3742	G	1161.4750	1143.5654	1144.5160	12
11	86.0969	974.3920	I	1104.4600	1086.5439	1087.5275	11
12	86.0605	1087.5060	P(+15.99)	991.3410	973.4598	974.3920	10
13	30.0343	1144.5160	G	878.3060	860.4121	861.3957	9
14	110.0717	1281.4530	H	821.3670	803.4000	804.1820	8
15	86.0605	1394.5360	P(+15.99)	684.2880	666.4240	667.7390	7
16	30.0343	1451.6555	G	571.2520	553.2841	554.0990	6
17	101.0714	1579.6610	Q	514.2480	496.2280	497.2462	5
18	30.0343	1636.5780	G	386.1900	368.2040	369.1210	4
19	30.0343	1693.7570	G	329.1680	311.1010	311.9210	3
20	70.0656	1790.8097	P	272.1660	254.1611	255.1447	2
21	129.1139		R	175.1690	157.1570	158.0919	1

Figure 59 | Representative MS2 spectrum of COL4A2 HyP114 HyP120 HyP123.

MS2 spectrum (A) and ion table (B) of the tryptic fragment containing COL4A2 HyP114 HyP120 HyP123 from the GST-VBC-bound fraction of the peptide enrichment assay (experiment 3; long wash). Confident peptide sequencing is achieved in the y-ion series. The diagnostic ions (HyP114: y156 and y16; HyP120: y9 and y10; HyP123; y6 and y7) confirm the location of hydroxyproline at the target site. The MS2 spectrum is annotated to indicate y6 ion (red dotted line).

A



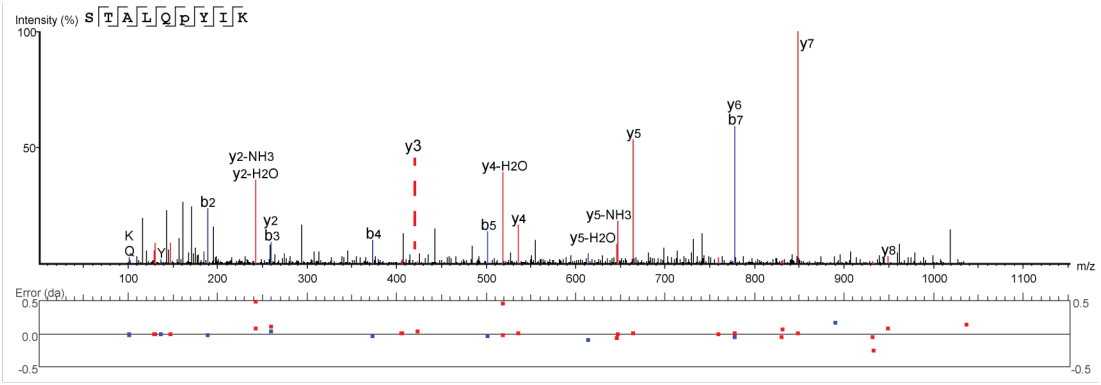
B

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	60.0449	88.0399	S				37
2	101.0714	216.0984	Q	3918.9946	3900.9841	3901.9675	36
3	86.0969	329.1825	L	3790.9360	3772.9255	3773.9089	35
4	88.0398	444.2094	D	3677.8518	3659.8413	3660.8247	34
5	110.0717	581.2560	H	3562.8250	3544.8145	3545.7979	33
6	104.0533	712.3430	M	3425.7661	3407.7556	3408.7390	32
7	30.0343	769.3303	G	3294.7256	3276.7151	3277.6985	31
8	120.0812	916.3270	F	3237.7041	3219.6936	3220.6770	30
9	110.0717	1053.2100	H	3090.6357	3072.6252	3073.6086	29
10	120.0812	1200.5430	F	2953.5767	2935.5662	2936.4170	28
11	74.0605	1301.1610	T	2806.5083	2788.4978	2789.4812	27
12	74.0605	1402.5811	T	2705.4607	2687.2400	2688.4336	26
13	30.0343	1459.5770	G	2604.4131	2586.4026	2587.3860	25
14	86.0969	1572.6750	I	2547.3916	2529.3811	2530.3645	24
15	86.0605	1685.6340	P(+15.99)	2434.3074	2416.2969	2417.2803	23
16	44.0499	1756.7340	A	2321.2598	2303.2493	2304.1040	22
17	70.0656	1853.6949	P	2250.2227	2232.2122	2233.1956	21
18	60.0449	1940.8280	S	2153.1699	2135.1594	2136.1428	20
19	74.0605	2041.9441	T	2066.1379	2048.1274	2049.1108	19
20	102.0554	2170.9868	E	1965.0902	1947.0797	1948.0632	18
21	74.0605	2272.0344	T	1836.0476	1818.0371	1819.0206	17
22	102.0554	2401.0769	E	1735.0000	1716.9895	1717.9730	16
23	86.0969	2514.1611	L	1605.8490	1587.9469	1588.9304	15
24	88.0398	2629.1880	D	1492.8270	1474.8628	1475.4070	14
25	72.0812	2728.0310	V	1377.7690	1359.8358	1360.5490	13
26	86.0969	2841.2390	L	1278.7321	1260.6281	1261.7510	12
27	86.0969	2954.2170	L	1165.6429	1147.5050	1148.6669	11
28	70.0656	3051.4773	P	1052.5690	1034.5310	1035.5829	10
29	74.0605	3152.5249	T	955.4840	937.5465	938.5301	9
30	44.0499	3223.5620	A	854.5140	836.4988	837.3250	8
31	74.0605	3324.6099	T	783.4070	765.4650	766.4453	7
32	60.0449	3411.6418	S	682.3700	664.2880	665.3976	6
33	86.0969	3524.7258	L	595.3930	577.3820	578.3656	5
34	70.0656	3621.7786	P	482.3070	464.2979	465.2650	4
35	86.0969	3734.8628	I	385.2557	367.0900	368.1280	3
36	70.0656	3831.9155	P	272.1610	254.1970	255.1447	2
37	129.1139		R	175.1410	157.0630	158.0919	1

Figure 60 | Representative MS2 spectrum of VCAN HyP2675.

MS2 spectrum (A) and ion table (B) of the tryptic fragment containing VCAN HyP2675 from the GST-VBC-bound fraction of the peptide enrichment assay (experiment 3; short wash with RapiGest). Confident peptide sequencing is achieved in the y-ion and b-ion series. The diagnostic ions (b14 and b15) confirm the location of hydroxyproline at the target site.

A



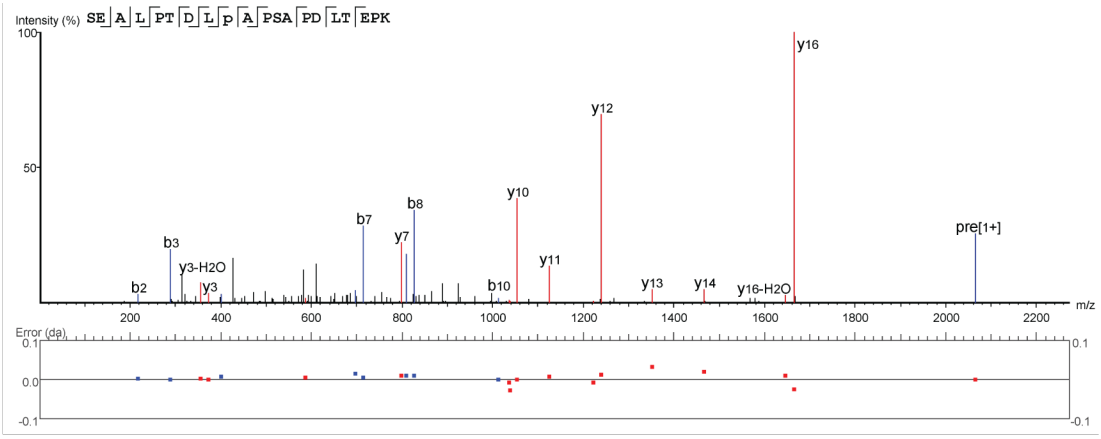
B

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	60.0449	88.0399	S				9
2	74.0605	189.1040	T	949.4380	931.5720	932.7600	8
3	44.0499	260.0700	A	848.4640	830.5240	831.3820	7
4	86.0969	373.2520	L	777.4310	759.4280	760.4235	6
5	101.0960	501.3020	Q	664.3430	646.4260	647.3470	5
6	86.0605	614.4100	P(+15.99)	536.2920	518.3230	519.2809	4
7	136.0740	777.4310	Y	423.2150	405.2230	406.2090	3
8	86.0969	890.2860	I	260.0700	242.0890	242.6720	2
9	101.0960		K	147.1210	129.1070	130.0780	1

Figure 61 | Representative MS2 spectrum of ARHGAP35 HyP113.

MS2 spectrum (A) and ion table (B) of the tryptic fragment containing ARHGAP35 HyP113 from the GST-VBC-bound fraction of the peptide enrichment assay (experiment 3; short wash with RapiGest). Confident peptide sequencing is achieved in the y-ion series. The diagnostic ions (y3 and y4) confirm the location of hydroxyproline at the target site. The MS2 spectrum is annotated to indicate the location of the y3 ion (red dashed line).

A



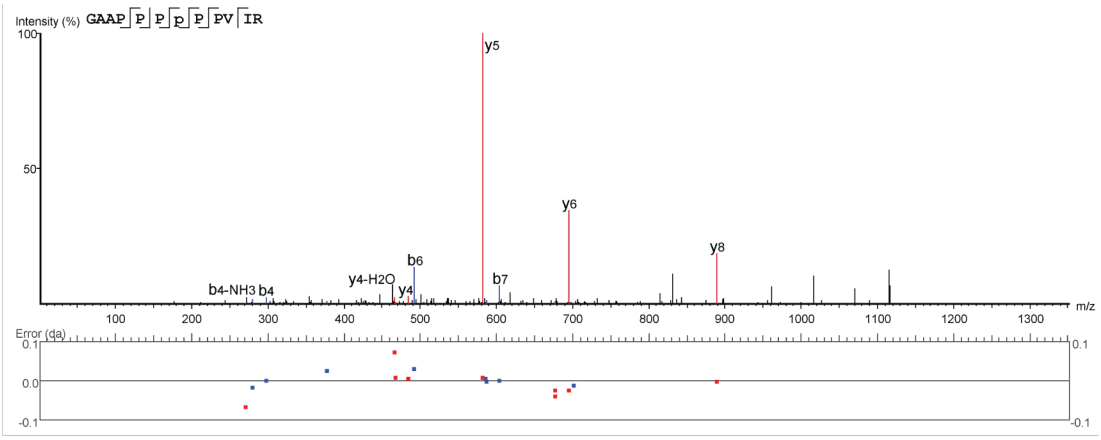
B

#	b	b-H2O	b-NH3	Seq	y	y-H2O	y-NH3	#
1	88.0399	70.0293	71.0129	S				20
2	217.0780	199.0719	200.0554	E	1978.0015	1959.9910	1960.9745	19
3	288.1180	270.1090	271.0926	A	1848.9589	1830.9484	1831.9319	18
4	401.1940	383.1931	384.1766	L	1777.9218	1759.9113	1760.8948	17
5	498.2563	480.2458	481.2294	P	1664.8650	1646.8149	1647.8107	16
6	599.3040	581.2935	582.2770	T	1567.7849	1549.7744	1550.7579	15
7	714.3260	696.3040	697.3040	D	1466.7150	1448.7267	1449.7102	14
8	827.4050	809.3930	810.3881	L	1351.6760	1333.6998	1334.6833	13
9	940.4627	922.4521	923.4357	P(+15.99)	1238.6121	1220.6157	1221.6080	12
10	1011.5000	993.4893	994.4728	A	1125.5699	1107.5681	1108.5516	11
11	1108.5526	1090.5421	1091.5256	P	1054.5430	1036.5400	1037.5430	10
12	1195.5846	1177.5741	1178.5576	S	957.4887	939.4781	940.4617	9
13	1266.6217	1248.6112	1249.5947	A	870.4567	852.4461	853.4297	8
14	1363.6744	1345.6639	1346.6475	P	799.4080	781.4090	782.3926	7
15	1478.7014	1460.6909	1461.6744	D	702.3668	684.3563	685.3398	6
16	1591.7855	1573.7750	1574.7585	L	587.3330	569.3293	570.3129	5
17	1692.8331	1674.8226	1675.8062	T	474.2558	456.2452	457.2288	4
18	1821.8757	1803.8652	1804.8488	E	373.2080	355.1930	356.1812	3
19	1918.9285	1900.9180	1901.9015	P	244.1655	226.1550	227.1385	2
20				K	147.1128	129.1022	130.0858	1

Figure 62 | Representative MS2 spectrum of PRKCSH HyP290.

MS2 spectrum (A) and ion table (B) of the tryptic fragment containing PRKCSH HyP290 from the GST-VBC-bound fraction of the peptide enrichment assay (experiment 2; no spike-in peptide). Confident peptide sequencing is achieved in the y-ion series. The diagnostic ions (y11 and y12) confirm the location of hydroxyproline at the target site.

A



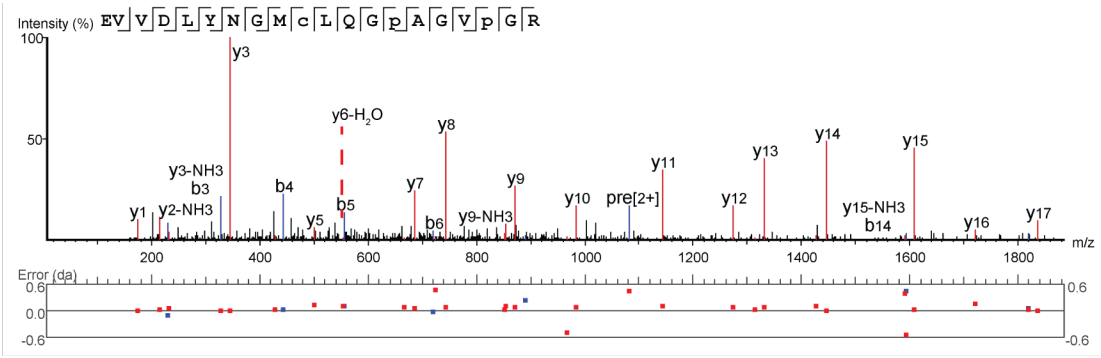
B

#	b	b-H2O	b-NH3	Seq	y	y-H2O	y-NH3	#
1	58.0293	40.0187	41.0023	G				12
2	129.0664	111.0558	112.0394	A	1127.6571	1109.6466	1110.6301	11
3	200.1035	182.0929	183.0765	A	1056.6200	1038.6095	1039.5930	10
4	297.1570	279.1457	280.1490	P	985.5829	967.5723	968.5559	9
5	394.2090	376.1985	377.1570	P	888.5350	870.5195	871.5031	8
6	491.2300	473.2512	474.2348	P	791.4774	773.4668	774.4504	7
7	604.3110	586.2920	587.2860	P(+15.99)	694.4510	676.4550	677.4240	6
8	701.3760	683.3516	684.3352	P	581.3680	563.3663	564.3499	5
9	798.4149	780.4044	781.3879	P	484.3170	466.2390	467.2890	4
10	897.4833	879.4728	880.4564	V	387.2714	369.2608	370.2444	3
11	1010.5674	992.5569	993.5405	I	288.2030	270.1924	271.2460	2
12				R	175.1189	157.1084	158.0919	1

Figure 63 | Representative MS2 spectrum of WIPF2 HyP328.

MS2 spectrum (A) and ion table (B) of the tryptic fragment containing WIPF2 HyP328 from the GST-VBC-bound fraction of the peptide enrichment assay (experiment 2; no spike-in peptide). Confident peptide sequencing is achieved in the y-ion series. The diagnostic ions (y5 and y6) confirm the location of hydroxyproline at the target site.

A



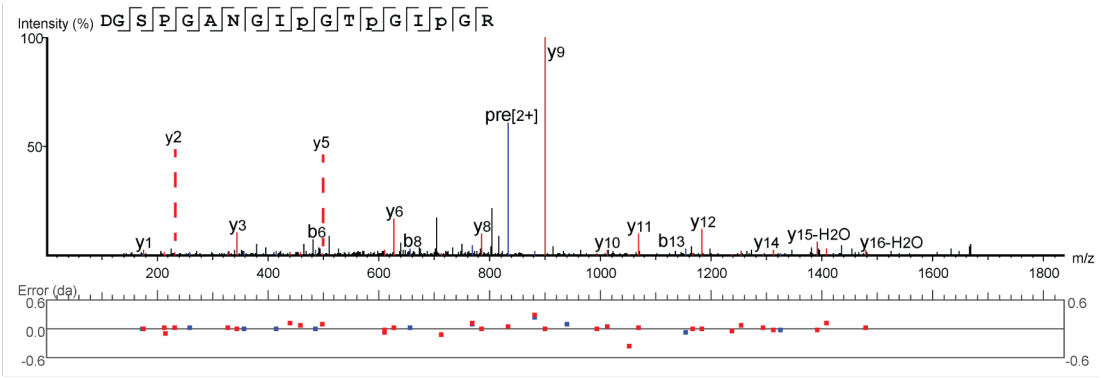
B

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	102.0554	130.0504	E				20
2	72.0812	229.2280	V	2034.9735	2016.9630	2017.9465	19
3	72.0812	328.1770	V	1935.9050	1917.8945	1918.8781	18
4	88.0398	443.1710	D	1836.8250	1818.8262	1819.7660	17
5	86.0969	556.1900	L	1721.6450	1703.7992	1704.7827	16
6	136.0761	719.3900	Y	1608.6740	1590.3080	1592.2480	15
7	87.0558	833.4045	N	1445.6450	1427.5229	1428.6354	14
8	30.0343	890.1740	G	1331.5190	1313.6089	1314.5660	13
9	104.0533	1021.4664	M	1274.5040	1256.5874	1257.5709	12
10	133.0435	1181.4971	C(+57.02)	1143.4480	1125.5470	1126.5305	11
11	86.0969	1294.5812	L	983.4460	965.5162	967.0020	10
12	101.0714	1422.6398	Q	870.3550	852.3900	853.3070	9
13	30.0343	1479.6611	G	742.3020	723.8900	725.3572	8
14	86.0605	1592.2480	P(+15.99)	685.3020	667.2600	668.3358	7
15	44.0499	1663.7460	A	572.3151	554.1910	555.2881	6
16	30.0343	1720.7675	G	501.1440	483.2674	484.2509	5
17	72.0812	1819.7660	V	444.2565	426.2459	427.1860	4
18	86.0605	1932.8835	P(+15.99)	345.1770	327.1775	328.1770	3
19	30.0343	1989.9050	G	232.0710	214.1298	215.0640	2
20	129.1139		R	175.1150	157.1084	158.0919	1

Figure 64 | Representative MS2 spectrum of CTHRC1 HyP59 HyP63.

MS2 spectrum (A) and ion table (B) of the tryptic fragment containing CTHRC1 HyP59 HyP63 from the GST-VBC-bound fraction of the peptide enrichment assay (experiment 3; long wash). Confident peptide sequencing is achieved in the y-ion series. The diagnostic ions (HyP59: y6-H_2O and y6) confirm the location of hydroxyproline at the target site. The MS2 spectrum is annotated to indicate the location of the y6-H_2O ion (red dashed line).

A



B

#	Immonium	b	Seq	y	y-H2O	y-NH3	#
1	88.0398	116.0348	D				18
2	30.0343	173.0540	G	1552.7714	1534.7609	1535.7444	17
3	60.0449	260.0640	S	1495.7499	1477.7130	1478.7229	16
4	70.0656	357.1480	P	1408.5940	1390.7531	1391.6909	15
5	30.0343	414.1750	G	1311.6899	1293.6300	1294.6381	14
6	44.0499	485.1900	A	1254.5620	1236.6332	1237.6680	13
7	87.0558	599.2425	N	1183.6169	1165.5961	1166.5920	12
8	30.0343	656.2390	G	1069.5190	1051.9170	1052.5366	11
9	86.0969	769.2470	I	1012.4810	994.5200	995.5151	10
10	86.0605	882.1410	P(+15.99)	899.4410	881.4475	882.1410	9
11	30.0343	939.3050	G	786.3950	768.3998	769.2470	8
12	74.0605	1040.4648	T	729.3889	711.3784	712.5050	7
13	86.0605	1153.5950	P(+15.99)	628.3000	610.4030	611.3390	6
14	30.0343	1210.5339	G	515.2936	497.2830	498.1680	5
15	86.0969	1323.6490	I	458.1970	440.2616	441.1150	4
16	86.0605	1436.6656	P(+15.99)	345.1810	327.1775	328.1260	3
17	30.0343	1493.6871	G	232.1040	214.0950	215.2120	2
18	129.1139		R	175.1230	157.1084	158.0919	1

Figure 65 | Representative MS2 spectrum of CTHRC1 HyP75 HyP78 HyP81.

MS2 spectrum (A) and ion table (B) of the tryptic fragment containing CTHRC1 HyP59 HyP63 from the GST-VBC-bound fraction of the peptide enrichment assay (experiment 3; long wash). Confident peptide sequencing is achieved in the y-ion series. The diagnostic ions (HyP59: y6-H_2O and y6) confirm the location of hydroxyproline at the target site. The MS2 spectrum is annotated to indicate the location of the y2 and y5 ions (red dashed lines).

Chapter 6. Functional screening for novel oxygen-labile proteins

6.1 Introduction

The aim of this work is to discover novel substrates of enzymatic hydroxylation that might contribute to cellular signalling pathways in response to hypoxia, with a primary focus on PHD-catalysed hydroxylation. The *bona fide* PHD substrates, HIF1α-3α, are subject to rapid proteolysis following their hydroxylation, which is mediated by pVHL. The experiments in Chapter 5.3 applied pVHL as an affinity reagent to enrich hydroxyproline-containing peptides in proteome-wide screens from pVHL-defective cells. Although this approach successfully identified novel prolyl hydroxylase substrate candidates, it is possible that a number of other 2OGD-catalysed hydroxyproline sites were not identified in these assays. pVHL is unlikely to bind to all hydroxyproline-containing peptides, and some oxygen-labile proteins might not be stabilised in pVHL-defective cells. Therefore, I proposed a functional proteome-wide screen in which the stability of proteins would be investigated at high and low oxygen tension.

6.1.1 Oxygen-dependent proteolysis regulates biological circuits

Bimodal signals in response to hypoxia are achieved by hydroxylation-mediated degradation of signalling proteins and re-synthesis of their non-hydroxylated proteoforms.

HIF transcriptional responses are regulated by the oxygen-sensitive prolyl hydroxylase activity of the PHDs directed towards HIFα, which results in pVHL-catalysed ubiquitylation and proteasomal degradation of the hypoxia inducible factors. It is possible that non-HIFα PHD substrates exist and are degraded by a different ubiquitin E3 ligase than pVHL. It would be unlikely for these proteins to be enriched in the GST-VHL, Elongin B and Elongin C (GST-VBC)-based affinity purification assays because the hydroxylation-dependent proteolysis would be expected to persist in the pVHL-deficient RCC4 cells, and such proteins would in any case not be predicted to bind to GST-VBC. I therefore sought an alternative strategy to identify proteolytic substrates of the PHDs

that considered changes to protein stability in conditions that restrict PHD-catalysed hydroxylation.

In addition to PHD-dependent pathways, other hydroxylation-mediated proteolytic mechanisms exist in cells. N-terminal cysteine dioxidation of regulator of G-protein signalling 4 (RGS4) and RGS5 is catalysed by 2-aminoethanethiol dioxygenase (ADO) (Masson *et al.*, 2019). Arginyl-tRNA-protein transferase 1 (ATE1) interacts with the oxidised cysteine and activates the N-end rule pathway, which leads to protein turnover mediated by the ubiquitin proteasome system (Lee *et al.*, 2005). The catalytic activity of ADO is sensitive to oxygen and has been suggested to perform physiological responses to hypoxia on a faster time scale than is observed for the HIF transcriptional response (Masson *et al.*, 2019).

The catalytic activity of peptidylglycine α-amidating monooxygenase (PAM) has been shown to be oxygen-sensitive (Simpson et al., 2015). The C-terminal '-RRG' sequence of secretory pathway proteins can be converted to an amidated product, '-RR-NH_2', by PAM, and glyoxylate is released in a two-step mechanism involving two domains of PAM. The copper-dependent peptidylglycine α-hydroxylating monooxygenase (PHM) domain initiates the conversion of substrates by hydroxylating the C-terminal glycine. This intermediate is subsequently converted to the amidated product by the zinc-dependent peptidyl-α-hydroxylglycine α-amidating lyase (PAL) domain, which expels glyoxylate as a by-product (Prigge et al., 2000). Importantly, this mechanism is irreversible and can lead to altered turnover of the amidated product. The role of PAM as an oxygen sensor has been proposed to function in a signalling mechanism that occurs by initiating α-amidation-dependent turnover of its substrates, with re-synthesis providing the counteraction (Simpson *et al.*, 2015).

In addition to the PHD-HIFα-pVHL axis, the recent discoveries of oxygen-sensitive pathways represent distinct mechanisms through which protein stability can be regulated in response to oxygen. First, there may be additional ADO and PAM substrates that could be discovered by studying protein stability in hypoxic conditions. Secondly, these pathways set a precedent for additional mechanisms of hydroxylation-mediated protein turnover that have not yet been discovered.

Therefore, it is possible that proteins other than substrates of the PHDs, ADO, and PAM are degraded in response to an oxygen-sensitive hydroxylation event. Such proteins could

have the capacity to function in hypoxia signalling pathways, which might occur independently of HIF. These proteins might be substrates of the aforementioned enzymes, or they might implicate other oxygen-sensitive enzymes. With this in mind, I sought a proteome-wide screen to identify proteins whose abundance decreased in the presence of oxygen.

6.1.2 Quantitative proteomics using reporter ion quantification

Quantitative proteomics was introduced in Chapter 4, in which K/R isotopes were used to label three samples in two different SILAC experiments. This method enabled samples to be multiplexed immediately after harvest so that the analytes were exposed to identical conditions during sample processing (e.g., trypsin digestion, high pH off-line fractionation and LC-MS/MS analysis). This controlled for spurious modifications that might occur between experimental groups during sample preparation and data acquisition (e.g., artefactual oxidation), which was an important factor to consider in the analysis of peptides containing hydroxyproline. However, the purpose of this analysis is to quantify protein abundance changes in many conditions. Therefore, controlling spurious oxidation at the peptide level is not such a priority. In addition, SILAC experiments generate greater complexity at the MS1 level and are typically limited to comparison of 2 or 3 conditions, which is likely to reduce the overall sensitivity of the assay. Since SILAC is limited to 3-plex experiments, an alternative method is required to generate the high number of comparisons desired.

Tandem mass tags (TMT) were developed as isobaric tags that yield reporter ions of distinct mass following fragmentation (Thompson et al., 2003). These tags consist of a reporter group, which provides the distinct mass after fragmentation, linked to a balance group to ensure the isobaric tags have the same mass. Finally, the peptide-reactive group enables these tags to specifically target the primary amine groups at the N-terminus of peptides and the ε-amine group on the side chains of lysine residues. This makes TMT a suitable reagent for labelling tryptic peptides. Labelled peptides coelute at the MS1 level and display identical charge states between conditions. These factors maintain sensitivity, which is not guaranteed in other labelling experiments (e.g., SILAC). The application of TMT to combine at least 10 samples was considered to be appropriate for the functional proteome-wide screen planned in this chapter (McAlister et al., 2012).

6.1.3 Reported non-HIFα PHD substrates

As outlined in Chapter 3, there have been reports of 45 non-HIFα PHD substrates to date (Table 14). Some of these have been reported to undergo hydroxylation-mediated proteolysis. Several studies have been performed to identify PHD-catalysed prolyl hydroxylation (Cockman *et al.*, 2019), including the data presented in chapters 3-5 of this thesis, but there has not been any strong evidence to support the reported PHD-catalysed hydroxylation events. It is possible that these proteins are subjected to oxygen-sensitive degradation, which may be a consequence of prolyl hydroxylation, or a different regulatory mechanism might be involved. Therefore, these proteins were considered in this chapter in an attempt to observe oxygen-dependent turnover.

6.1.4 Experiment design

In order to observe oxygen-sensitive proteolysis, proteins were first stabilised by the culture of cells under an atmosphere of 1 % O_2 for 24 hours. Target proteins were expected to undergo rapid degradation following re-oxygenation. A focusing experiment was performed to determine the period of re-oxygenation required for the proteomic screen. The samples were then prepared for a TMT-labelled 10-plex experiment, and this was used to identify candidates of oxygen-sensitive proteolysis in SH-SY5Y cells.

6.2 Methods

6.2.1 Cell culture and harvest

SH-SY5Y cells were selected because they exhibit expression of all proteins required for the oxygen sensitive degradation of HIFα and RGS4/5 (Masson *et al.*, 2019). For the validation western blot, 150 mm^2 culture dishes were each seeded with 4 x 10^6 SH-SY5Y cells and incubated in DMEM:F12 media supplemented with 10 % (v/v) FBS for 18 hours. At this point, the media was replenished, and the cells were incubated for 24 hours at 1 % O_2 in an Invivo$_2$ hypoxic chamber.

The samples were re-oxygenated by transferring the cells from the hypoxic chamber to an incubator set to ambient atmosphere (21 % O_2). The control samples were kept in the

hypoxic chamber. Cells were harvested at 0, 2, 4, 8, 16, 32, 48, and 64 minutes after the switch in oxygenation began.

Cells were washed in ice-cold PBS and lysed on the plate in urea lysis buffer (7 M urea, 2 M thiourea, 40 mM Tris-HCl pH 8.0, 2 % (w/v) CHAPS, 250 U / mL, 1x HALT protease inhibitor cocktail, 20 mM DTT). The lysates were incubated for 1 hour at 4 °C with end-to-end rotation (15 rpm). IAA was added to a final concentration of 50 mM and the debris was pelleted by ultracentrifugation at 17,000 x *g* for 30 minutes. The protein concentrations were calculated for the soluble fraction using the Pierce™ 660 nm protein assay kit.

6.2.2 Western blot

The protocol described in Chapter 2.1.1 was used to perform the western blots described in this chapter. I used 4-20 % SDS-PAGE to resolve CDKN1C, RGS4, RGS5, and TUBB. KDM5B, HIF1α and HIF2α were resolved by 7.5 % SDS-PAGE. After chemiluminescent detection, the total protein content of each lane was visualised by applying Coomassie stain to the membrane.

6.2.3 Sample preparation for LC-MS/MS

After analysing the protein abundance changes over time by western blot, the 0, 4, 8, 16 and 48-minute samples were selected for LC-MS/MS analysis. Fractions of these samples (200 µg) were subject to acetone precipitation. The protein precipitates were resuspended in 1 M GndHCl, 0.1 M HEPES, pH 8.0 and incubated in a sonicating water bath for 15 minutes to aid solubilisation. The sample was treated with 2 µg LysC for 3 hours at 37 °C with orbital rotation at 1,200 rpm. The samples were diluted in an equivalent volume of water and 2 µg Trypsin was added. Trypsin digest was performed for 18 hours at 37 °C with orbital rotation at 1,200 rpm.

TMT labelling and high pH off-line fractionation was performed in collaboration with Dr Joanna Kirkpatrick and Helen Flynn. Peptide labelling was performed on 100 µg of each sample with the TMT10plex kit (Lot number: UJ288603) according to the manufacturer's protocol. Label checks were performed by analysing 500 ng of each sample by LC-MS/MS using a Q Exactive™ Hybrid Quadrupole-Orbitrap™ mass spectrometer. After

correcting for protein N-terminal acetylation, ≥ 99 % of peptides were correctly labelled in each sample.

The reactions for each TMT channel were quenched with 5 % hydroxylamine and the samples were combined. The sample mixture was desalted with a Sep-Pak C18 1 cc Vac Cartridge (50 mg sorbent) and the organic content was evaporated by vacuum centrifugation. A fraction of the mixed sample (500 ng) was checked by LC-MS/MS using a Q Exactive™ Hybrid Quadrupole-Orbitrap™ mass spectrometer. There were similar ratios of each reporter ion, indicating equal amounts of each TMT channel had been combined.

6.2.4 High pH off-line fractionation

High pH off-line fractionation was performed on 100 µg of digested material with an XBridge BEH C18 Column (3.5 µm, 1 mm X 150 mm), using the loading pump on a U3000 HPLC system with autosampler fraction collection. A 60-minute ACN gradient (1-35 % buffer B) was performed in 20 mM NH_4HCO_2, pH 10, at a flow rate of 40 µL/min, as described in Table 54. This resulted in 96 fractions that were lyophilised by vacuum centrifugation to remove the organic content.

Table 54 | Linear gradient for off-line high pH fractionation of TMT-labelled peptides derived from SH-SY5Y extracts.

High pH fractionation was performed using buffer A (20 mM NH_4HCO_2, pH 10) and buffer B (90 % ACN, 20 mM NH_4HCO_2, pH 10). A linear ACN gradient (1-35 % buffer B) was performed over 60 minutes at a flow rate of 40 µL/min.

Time (minutes)	Buffer B (%)
0	1
12	1
72	35
80	95
90	95
90.1	1
100	1

6.2.5 LC-MS/MS

The lyophilised fractions were resuspended in 2 % ACN, 0.1 % FA and pooled to generate 32 concatenated fractions (Table 55).

Table 55 | Fraction pooling scheme to generate 32 concatenated fractions.

Concatenated fraction #	Fraction A	Fraction B	Fraction C
F1	1	33	65
F2	2	34	66
F3	3	35	67
F4	4	36	68
F5	5	37	69
F6	6	38	70
F7	7	39	71
F8	8	40	72
F9	9	41	73
F10	10	42	74
F11	11	43	75
F12	12	44	76
F13	13	45	77
F14	14	46	78
F15	15	47	79
F16	16	48	80
F17	17	49	81
F18	18	50	82
F19	19	51	83
F20	20	52	84
F21	21	53	85
F22	22	54	86
F23	23	55	87
F24	24	56	88
F25	25	57	89
F26	26	58	90
F27	27	59	91
F28	28	60	92

F29	29	61	93
F30	30	62	94
F31	31	63	95
F32	32	64	96

Peptide fractions were analysed by LC−MS/MS in an Orbitrap Fusion™ Lumos™ Tribrid™ instrument. A 120-minute ACN gradient (3-50 % buffer B) in 0.1 % FA, as described in Table 56, was performed with a U3000 HPLC instrument with an Acclaim™ Pepmap™ 100 C18 HPLC column (3 µm x 0.075 mm x 500 mm) at a flow rate of 0.275 µL/min.

Table 56 | Non-linear gradient for on-line peptide separation at low pH.

Low pH on-line peptide separation was performed using buffer A (0.1 % FA, pH 3) and buffer B (80 % ACN, 0.1 % FA, pH 3). A non-linear ACN gradient (3-50 % buffer B) was performed over 120 minutes at a flow rate of 0.275 µL/min.

Time (minutes)	Buffer B (%)
0	3
5	3
20	9
110	30
125	50
125.1	99
130	99
131	3
145	3

In order to gain maximum sensitivity in the samples, an adapted version of the multi-notch method was implemented (Table 57), which utilises synchronous precursor ion selection (SPS) from the MS2 spectrum (McAlister et al., 2014). Briefly, multiple fragment ions in the MS2 spectrum are co-isolated for further fragmentation to yield the reporter ions in the MS3 spectrum. Therefore, there is greater conversion of precursor ions to reporter ions in the SPS method than with traditional reporter ion quantitation methods. The SPS method has been demonstrated to improve the accuracy and sensitivity of reporter ion quantification (McAlister *et al.*, 2014). The Gygi group developed a method template to generate high quality SPS MS3 data (Paulo et al., 2016). This

template was optimised to capture the protein dynamics of HIF1α, RGS4, and RGS5 that had been observed in the validation western blot (Figure 66).

Table 57 | Settings for data-dependent acquisition of TMT-labelled peptides using an adapted version of the multi-notch method that utilises SPS.

Precursor (MS1) ion selection was consistent throughout the analysis. The parameters for the analysis of the fragment ions (MS2 and MS3) varied depending on the charge state of the precursor ion (e.g., MS2 scan event type 1-3).

Entry	Parameter	Setting
MS1	Detector type	Orbitrap
	Resolution	120,000
	Mass range	Normal
	Use quadrupole isolation	True
	Scan range (*m/z*)	375-1,500
	RF lens (%)	30
	AGC target	1×10^5
	Maximum injection time (ms)	50
	Microscans	1
	Data type	Profile
	Polarity	Positive
MIPS	Monoisotopic peak determination	Peptide
Intensity	Filter type, intensity threshold	5×10^3
Charge state	Include charge state(s)	2-6
Dynamic exclusion	Exclude after *n* times	1
	Exclusion duration (s)	60
	Mass tolerance	ppm
	Low	10
	High	10
	Exclude isotopes	True
MS2 scan event type 1	Include charge state(s)	2
	Include undetermined charge states	False
	Isolation mode	Quadrupole
	Isolation window (*m/z*)	1.2
	Activation type	CID

	Collision energy (%)	35
	Detector type	Linear ion trap
	First mass (*m/z*)	
	AGC target	1 x10^3
	Maximum injection time (ms)	50
	Microscans	1
	Data type	Centroid
MS3 settings for MS2 scan event type 1	Synchronous precursor selection	True
	Number of SPS precursors	5
	MS isolation window (*m/z*)	1.3
	MS2 isolation window (*m/z*)	2
	Activation type	HCD
	Collision energy (%)	65
	Detector type	Orbitrap
	Resolution	50,000
	Scan range (*m/z*)	100-500
	AGC target	1 x10^3
	Maximum injection time (ms)	105
	Microscans	1
	Data type	Profile
MS2 scan event type 2	Include charge state(s)	3
	Include undetermined charge states	False
	Isolation mode	Quadrupole
	Isolation window (*m/z*)	0.7
	Activation type	CID
	Collision energy (%)	35
	Detector type	Linear ion trap
	First mass (*m/z*)	
	AGC target	1 x10^3
	Maximum injection time (ms)	50
	Microscans	1
	Data type	Centroid
MS3 settings for MS2 scan event type 2	Synchronous precursor selection	True

	Number of SPS precursors	10
	MS isolation window (*m/z*)	0.7
	MS2 isolation window (*m/z*)	2
	Activation type	HCD
	Collision energy (%)	65
	Detector type	Orbitrap
	Resolution	50,000
	Scan range (*m/z*)	100-500
	AGC target	1 x10^3
	Maximum injection time (ms)	105
	Microscans	1
	Data type	Profile
MS2 scan event type 3	Include charge state(s)	4-6
	Include undetermined charge states	False
	Isolation mode	Quadrupole
	Isolation window (*m/z*)	0.5
	Activation type	CID
	Collision energy (%)	35
	Detector type	Linear ion trap
	First mass (*m/z*)	
	AGC target	1 x10^3
	Maximum injection time (ms)	50
	Microscans	1
	Data type	Centroid
MS3 settings for MS2 scan event type 3	Synchronous precursor selection	True
	Number of SPS precursors	10
	MS isolation window (*m/z*)	0.7
	MS2 isolation window (*m/z*)	2
	Activation type	HCD
	Collision energy (%)	65
	Detector type	Orbitrap
	Resolution	50,000
	Scan range (*m/z*)	100-500
	AGC target	1 x10^3

	Maximum injection time (ms)	105
	Microscans	1
	Data type	Profile
Data dependent properties (MS2)	Select: 'Cycle time'	3 s
Data dependent properties (MS3)	Select: 'Top *n* scans'	5 (MS2 scan event type 1) 10 (MS2 scan event type 2 and 3)

6.2.6 Data processing

Raw data files were processed in PEAKS® using the parameters defined in Table 58.

Table 58 | Processing parameters of the TMT-labelled peptides that were acquired using an adapted version of the multi-notch method.

Parameter	Setting
PEAKS® version	Xpro
Reference proteome	Human, canonical (Uniprot id: UP000005640)
Precursor ion error tolerance	±10 ppm
Fragment ion error tolerance	±0.5 Da
Enzyme	Trypsin
Enzyme specificity	Specific
Maximum number of miscleavages per peptide	2
Fixed modification	Carbamidomethylation (+57.0214 Da; C) TMT10plex (+229.1629 Da; K, N-term)
Variable modifications	Oxidation (+15.9949 Da; CDFHKMNPRWY) Acetylation (+42.0367 Da; protein N-terminus)
Maximum number of variable PTMs per peptide	3

Reporter ion quantification of TMT-labelled peptides was performed using the Q search tool in PEAKS®, as described in Table 59.

Table 59 | Processing parameters to perform quantification of the TMT-labelled peptides that were acquired using an adapted version of the multi-notch method.

Parameter	Setting

PEAKS® version	Xpro
PEAKS® search tool	Q
Quantification type	TMT-10plex (CID/HCD)
Quantification mass tolerance	0.02 Da
FDR threshold	1 %
Reporter ion type	MS3
Perform purity correction	True
TMT10-126	126.1277
TMT10-127N	127.1248
TMT10-127C	127.1311
TMT10-128N	128.1281
TMT10-128C	128.1344
TMT10-129N	129.1315
TMT10-129C	129.1378
TMT10-130N	130.1348
TMT10-130C	130.1411
TMT10-131	131.1382

6.2.7 Analysis

Filters were applied at the peptide level as previously described (Chapter 3). Briefly, confident PSMs were selected for (-10lgP ≥ 20). A second filter was applied for confident PTM assignments (AScore ≥ 20). The Q search tool does not enable a filter for ion intensity to be applied. An additional peptide filter was applied in which a missing value was permitted from a maximum of 1 TMT channel.

Two filters were applied at the protein level. A unique peptide is defined as a peptide with a -10lgP value above the peptide filtering threshold (-10lgP ≥ 20) that can be mapped to only one protein. Proteins were filtered for ≥ 2 unique peptide assignments.

6.2.8 Generalised linear model

The significance of abundance change following re-oxygenation was interrogated by a generalised linear model (GLM) using the *DESeq2* programme in R (Love et al., 2014). This model was compared against an intercept-only model to determine if the protein

abundance changes over time following re-oxygenation were significantly different from 0. This analysis was performed in collaboration with Dr Yoichiro Sugimoto.

For the GLM, two assumptions were made. First, protein abundance does not change significantly in hypoxia. This assumption was based on the contrast regarding the duration of the equilibration (24 hours) and the treatment (1 hour) of sustained hypoxia; the cells should have equilibrated to 1 % O_2 after 24 hours (including hypoxic induction of HIF target genes), so protein abundance changes after an additional hour were expected to be minimal. Therefore, each hypoxic time point was transformed to equal 0 minutes for the GLM. The second assumption is that protein abundance changed in re-oxygenation as a linear function of $\log_2$(time). Each time point was then $\log_2$(time + 1) transformed so that the 0-minute time points could be included in the analysis.

A line of best fit was applied to the re-oxygenated time series for each protein in the dataset. A likelihood ratio test (LRT) was performed to assess the goodness of fit for each protein, and this was converted to a p-value that described the probability that the abundance changes over time were not due to chance. This p-value was adjusted (p-adj) for multiple comparisons using the false discovery rate (FDR), which was set to 1 %.

To identify oxygen-labile candidates, two filters were applied to select for proteins with p-adj < 0.001. A second filter selected for proteins that exhibit oxygen-dependent decay:

1. Δ abundance change in re-oxygenation (48 – 0 minutes) < 0
   and
2. Δ abundance change in re-oxygenation (48 – 0 minutes) < Δ abundance change in sustained hypoxia (48 – 0 minutes).

6.2.9 One-phase decay model to analyse candidates

The GLM considered protein abundance changed as a linear function of $\log_2$(time) in re-oxygenation. However, this did not take into account the changes to the rate of decay as the substrate is degraded. Therefore, further analysis was performed on the proteins found to be significantly decreased following re-oxygenation by the GLM. This included the one-phase decay model, which was applied in Prism version 8.1.

6.2.10 Experimental design

Protein abundance was compared between hypoxia and re-oxygenation. A normoxic comparison was not included because protein abundance might increase as a consequence of the hypoxic stabilisation of HIFα (i.e., HIF transactivation). Long durations of re-oxygenation might therefore alter protein abundance as a function of transcriptional re-programming. A short period of re-oxygenation (i.e., up to an hour) was considered to study protein dynamics because HIFα and RGS4/5 proteins have been shown to undergo rapid proteolysis after the oxygen tension is increased (Masson *et al.*, 2019).
Although translation inhibitors enable protein half-life to be investigated, they were not used so that physiologically relevant protein dynamics were captured by the proteome-wide screen. Additionally, the study aimed to identify proteins whose half-life was short in relation to the time required to achieve effective protein synthesis inhibition.

6.3 Results

The aim of the work in this chapter to identify proteins whose stability is regulated by oxygen tension. Such proteins could function in cellular responses to hypoxia, which has been well established for HIFα. These proteins would most likely be regulated by a hydroxylation-dependent mechanism, as has been shown for substrates of PHD and ADO. In hypoxia, HIF transactivation leads to cellular re-programming, which could directly or indirectly alter protein abundance through gene expression changes. Therefore, cells were equilibrated in hypoxia and the protein decay following re-oxygenation was measured. A short time period (less than 1 hour) was tested to identify proteins that displayed rapid decay on a similar timescale to HIFα and RGS4/5.

6.3.1 Proof of principle

In the first instance I studied the dynamics of *bona fide* oxygen-labile proteins (HIF1α, HIF2α, RGS4, RGS5) to establish experimental conditions that would later be used for the proteomic screen.
Cells were equilibrated at 1 % O_2 for 24 hours. Rapid protein degradation of the positive controls was observed following exposure to 21 % O_2 at 8 time points from 0 – 64 minutes (Figure 66). HIF1α and HIF2α displayed complete degradation after 16 minutes. The

signal for RGS5 was completely diminished by 32 minutes. Faint signal remained for RGS4 between 32 and 64 minutes but the decrease in signal from 0 minutes of re-oxygenation was nearly completed in a similar time frame to RGS5. These results confirm oxygen-dependent proteolysis occurred in the samples prepared for mass spectrometry analysis.

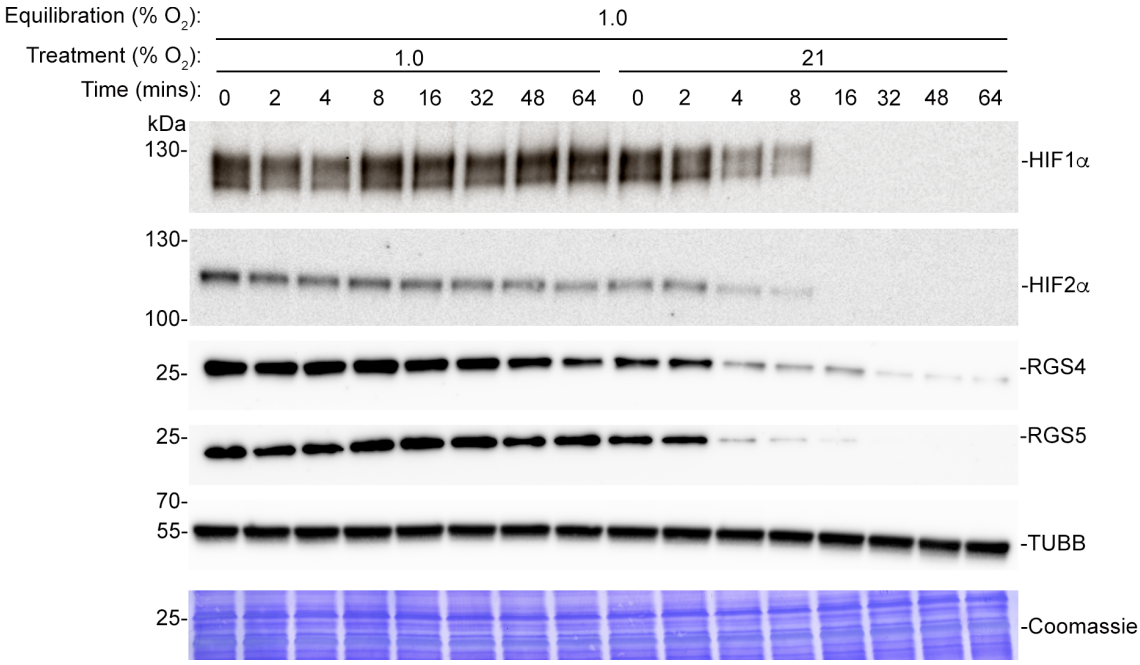


Figure 66 | Protein abundance changes in response to sustained hypoxia or re-oxygenation. SH-SY5Y cells were equilibrated at 1 % O_2 for 24 hours and treated for 0 – 64 minutes at 21 % O_2 or maintained at 1 % O_2.

A 10-plex TMT labelling kit was procured for the proteomic screen, which would facilitate 5 pair-wise comparisons across the time series. To determine the time points that would be used for the proteomic screen, I analysed the dynamics of the aforementioned positive controls. There was little difference observed after 2 minutes of re-oxygenation, so this time point was not included in the screen. Turnover of HIF1α, HIF2α, RGS4, and RGS5 appeared to be at its greatest between 4, 8 and 16 minutes, so these time points were included. Some proteins may display a slower degradation rate than the aforementioned positive controls, so a time point after 32 minutes was sought for the proteomic screen. At 64 minutes it is possible protein abundance decreases for non-proteolytic reasons. For example, the rate of synthesis for some proteins might not equal or exceed their synthesis, leading to dilution by cell division (Ross et al., 2020).

Therefore, I decided to use the following time points for the proteomic investigation: 0, 4, 8, 16, and 48 minutes.

6.3.2 Proteome-wide profiling of protein abundance changes following re-oxygenation

To investigate proteome-wide protein abundance changes over time following re-oxygenation, an LC-MS/MS-based experiment was performed using the TMT method that labels tryptic peptides with isobaric chemical tags to facilitate relative quantification between experimental conditions (McAlister *et al.*, 2012). Labelled peptides derived from different samples share an identical precursor mass, but each version of the tag provides a unique reporter ion, with distinct masses, following fragmentation. The relative abundance of these reporter ions enables quantification to be performed between experimental groups.

As previously stated, the lysates corresponding to sustained hypoxia and re-oxygenation at 0, 4, 8, 16, and 48 minutes were prepared for LC-MS/MS analysis. The samples were lysed, digested with trypsin, and labelled with the TMTpro 10-plex kit. High pH fractionation was performed, and 32 concatenated fractions were analysed by LC-MS/MS. The raw data was processed in PEAKS® Xpro and filters were applied to select for confidently assigned peptides. First, at the peptide level, the PSM must be confidently assigned (-10lgP ≥ 20). Second, a missing value (i.e., no peptide detection) was permitted for a maximum of one TMT channel. PTM assignments were also required to meet a confidence threshold (AScore ≥ 20). At the protein level, a confident assignment (-10lgP ≥ 20) was required. Additionally, a minimum number of unique peptides (unique peptides ≥ 2) was required for each protein assignment.

The overall proteome coverage was good (7,371 unique proteins) and the absolute change in relative abundance between 48 and 0 minutes (48 minutes – 0 minutes) was plotted for each condition. Three positive controls (HIF1α, RGS4 and RGS5) were identified and these proteins displayed abundance changes consistent with the validation blot (Figure 67). HIF1α and RGS5 showed the greatest abundance decrease following re-oxygenation. No other proteins displayed abundance changes similar to these proteins. The reduction of RGS4 was less marked than the other positive controls but still clearly demonstrated oxygen-sensitive attrition.

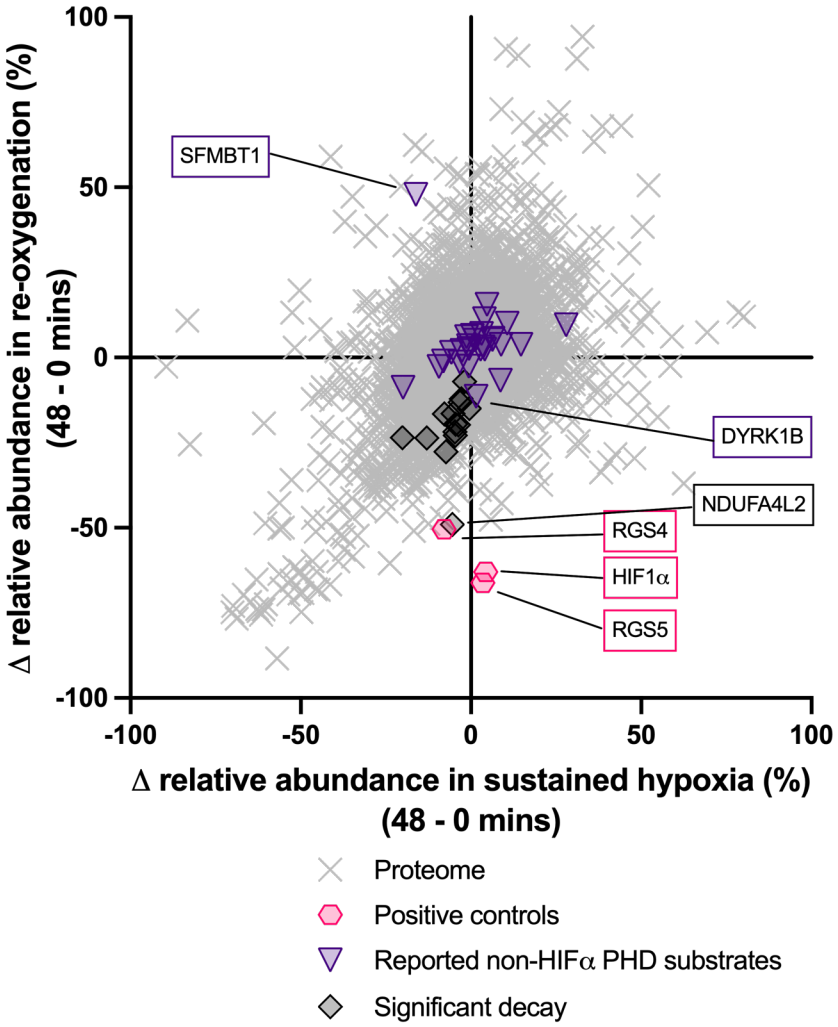


Figure 67 | Global protein abundance changes following re-oxygenation or sustained hypoxia.

The absolute change in relative abundance between 0 and 48 minutes was plotted for sustained hypoxia and re-oxygenation to visualise the abundance changes of 7,371 unique proteins in these conditions. Annotated proteins are discussed in the main text.

A generalised linear model (GLM) was used to determine proteins with significant decay. This model considered protein abundance at all time points in the re-oxygenation time series and predicted the probability that the observed protein abundance changes were due to chance. The p-value was then adjusted (p-adj) to account for the false discovery rate (1 %) and a threshold of p-adj < 0.001 was used to identify proteins that demonstrated significant decay in re-oxygenation (Table 60). Importantly, HIF1α, RGS5, and RGS4

were identified with a p-adj value exceeding the minimum threshold, which provides a proof of principle the GLM can identify genuine oxygen-labile proteins in the data.

Table 60 | The generalised linear model identified 17 proteins that exhibit significant decay in re-oxygenation.

The significance threshold of protein abundance changes in reoxygenation was set to p-adj < 0.001. Proteins were ranked in descending order of ΔΔ ratio, which is the absolute change in Δ relative abundance between re-oxygenated cells and those maintained in hypoxia.

Protein	p-adj	Δ relative abundance (48 – 0 minutes)		ΔΔ ratio
		Hypoxia	Reoxygenation	
HIF1A	6.12E-05	3%	-66%	-69%
RGS5	1.65E-107	4%	-63%	-67%
NDUFA4L2	8.98E-11	-6%	-49%	-43%
RGS4	4.65E-27	-8%	-50%	-42%
ZNF395	4.00E-06	-7%	-28%	-21%
KDM5B	5.12E-07	-5%	-23%	-18%
BCKDK	9.31E-04	-5%	-22%	-17%
RAB20	8.73E-04	-4%	-20%	-16%
MMP2	1.17E-06	-5%	-20%	-15%
GOLM1	6.89E-08	0%	-15%	-15%
MRC2	4.00E-06	-5%	-16%	-11%
CDKN1C	2.57E-07	-13%	-24%	-11%
PCOLCE	5.40E-06	-3%	-13%	-10%
COL6A1	1.11E-04	-8%	-17%	-9%
ACSL4	1.97E-04	-3%	-12%	-9%
PFKP	8.73E-04	-2%	-7%	-5%
PFKFB4	2.94E-06	-20%	-24%	-4%

NDUFA4L2 displayed a significant decrease following re-oxygenation and was the only significantly regulated protein to display a similar magnitude of protein attrition (ΔΔ ratio) to the positive controls. This suggests NDUFA4L2 may be regulated through a hydroxylation-mediated proteolytic pathway on a similar time scale to HIFα and RGS4/5. The other candidates are discussed in more detail in Section 6.3.3.

The data in Figure 67 showed a cluster of proteins that appeared to decrease in both hypoxia and following re-oxygenation. It is not clear why they decreased after 48 minutes

in both conditions. There is a possibility that the synthesis of these proteins was suppressed in hypoxia but did not recover on the timescale of the experiment. This would have facilitated protein abundance to decrease by two mechanisms that are independent of the oxygen tension. The first would be degradation and the second would be dilution with each cell division (Ross *et al.*, 2020). It would be expected that the protein abundance would recover in sustained re-oxygenation (e.g., > 24 hours). Whilst this might represent an interesting biological phenomenon in the context of adaptation to hypoxia or normoxia, the aim of this work was to identify novel substrates of protein hydroxylation, so further investigation of this cluster was not performed.

Some of the reported non-HIFα PHD substrates have been reported to be subject to oxygen-sensitive proteolysis. I therefore interrogated the data for all 45 reported substrates. None of the 26 reported non-HIFα PHD substrates that were identified in this analysis exhibited significant decay following re-oxygenation. DYRK1B displayed the greatest decrease in abundance (-11 %) in re-oxygenation but this was not a significant change. SFMBT1 is reported to undergo PHD2-catalysed prolyl hydroxylation leading to pVHL-catalysed ubiquitylation and degradation by the proteasome (Liu *et al.*, 2020a). In the analysis presented in this thesis, SFMBT1 protein abundance actually increased by 48 % following re-oxygenation, which is inconsistent with the reported protein dynamics in normoxia. Overall, there is no strong evidence to support hydroxylation-mediated degradation for any of the reported non-HIFα PHD substrates that were detected in this analysis.

6.3.3 Oxygen-labile proteins

The significantly regulated proteins were manually inspected for abundance change over the time series. A model of one-phase decay was fitted to the data. The time (t) was transformed (t + 1) and plotted against the fold-change of protein abundance relative to 0 minutes of re-oxygenation (Figure 68). Overall, a small number of proteins in SH-SY5Y cells that might have exhibited oxygen-dependent degradation. Furthermore, the rate of decay was likely to be slower than that of HIF1α, RGS4 and RGS5, with the exception of NDUFA4L2. The rate of attrition of ZNF395, KDM5B and RAB20 appeared to be increasing towards the end of the time series, so the reduction in protein abundance may

have been more substantial over a longer time period. This raised a hypothesis that some targets might display slower rates of oxygen-dependent degradation than HIF1α, RGS4 and RGS5. The other candidates appeared to have stabilised by the end of the 48-minute period of re-oxygenation, which suggests they might not be subject to oxygen-dependent degradation. Nevertheless, these candidates require validation before further studies are attempted to investigate the mechanism of oxygen-dependent degradation.

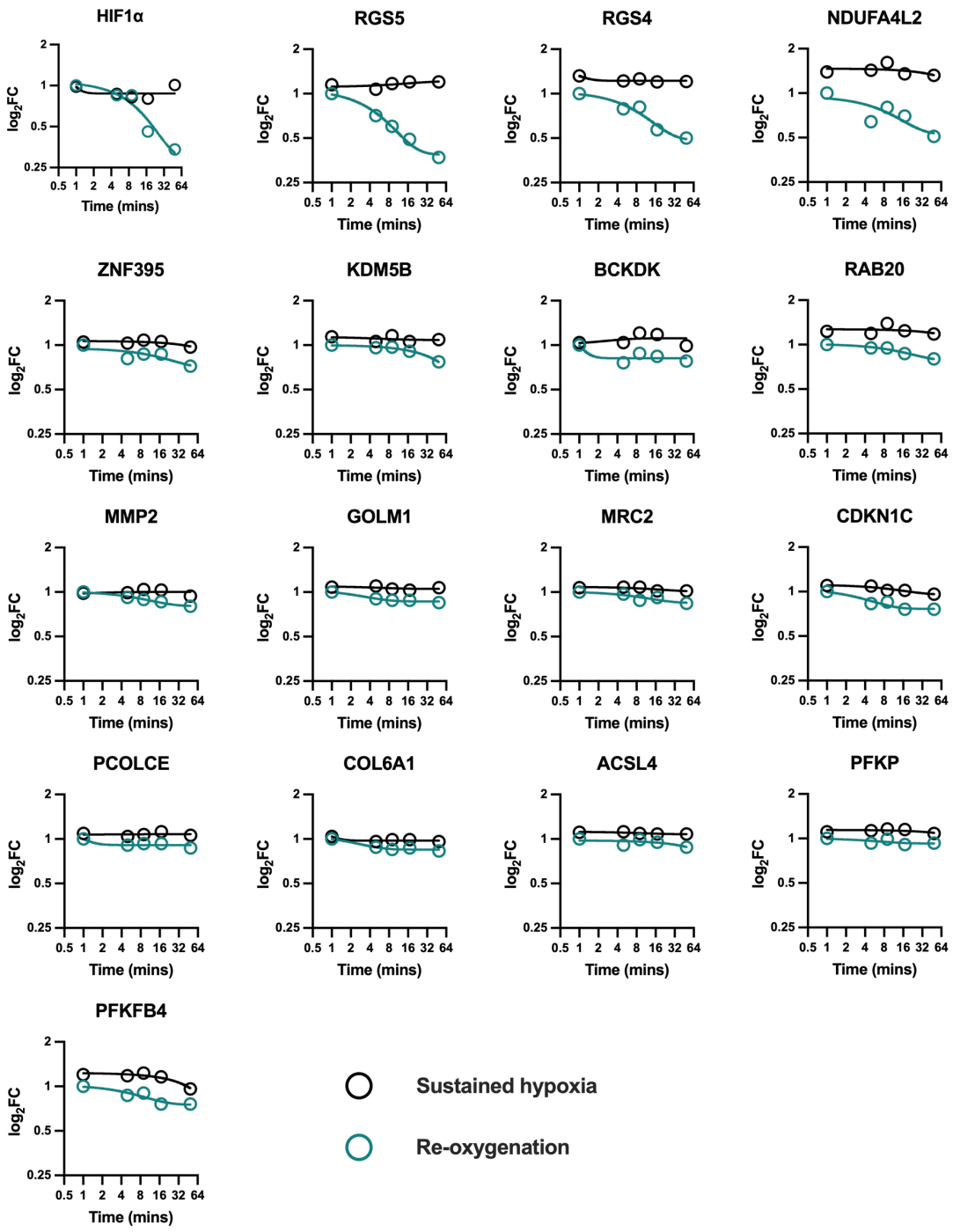


Figure 68 | Oxygen-labile candidates.

Log_2(time + 1) is plotted against log_2(Fold-change relative to 0 minutes re-oxygenation) for the positive controls and all of the significantly regulated candidates. A model of one-phase decay was fitted to the time-series of both experimental conditions.

6.3.4 KDM5B validation

The experiments so far have displayed rapid turnover of HIF1α, HIF2α, RGS4 and RGS5 in response to re-oxygenation. After statistical analysis, HIF1α, RGS4, RGS5, and 14 novel candidates were identified with significant attrition in re-oxygenation. Further analysis of these proteins indicated that only a few proteins were likely to have been subjected to oxygen-sensitive degradation. In order to rank the candidates for further inspection I cross-referenced the data from this analysis with the published literature and data in other chapters of this thesis.

None of the candidates identified in this proteome-wide screen displayed hydroxyproline, N-terminal cysteine, or other oxygen-sensitive PTMs (e.g., ubiquitylation) that might have increased the likelihood that they are oxygen-sensitive targets of degradation. Furthermore, they have not been identified in any of the previous chapters of this thesis with such modifications. The exception is COL6A1, which is subject to prolyl hydroxylation in the triple helical domain. This is unlikely to contribute to its degradation, at least based on the known role of prolyl hydroxylation on collagen proteins (Salo and Myllyharju, 2021). However, COL6A1 peptides were identified in the peptide enrichment assays, which suggests a direct interaction with pVHL, at least *ex vivo*. CDKN1C has also been reported to interact with pVHL (Minervini et al., 2017). In the MEF SILAC assay KDM5B was found to be upregulated at the protein level in MEFs defective for the PHDs, and ZNF395 was found to be upregulated in hypoxia in U-87 MG cells (Chapter 4). Neither protein exhibited hydroxylation or ubiquitylation in the hydroxylation-permissive condition, which suggests the these might be HIF target genes. Alternatively, increased protein coverage might be required to detect oxygen-dependent PTMs.

Together with reports of KDM5B and CDKN1C upregulation by HIF (Pollard et al., 2008; Wierenga et al., 2014; Xia et al., 2009) and the availability of a specific antibodies, these candidates were analysed further in an attempt to validate oxygen-dependent proteolysis. A repeat of the previous experiment was performed with the intention of visualising the protein dynamics by western blot. The repeat experiment was performed with a normoxic control to consider hypoxic induction of the proteins. KDM5B and CDKN1C abundance increased significantly after equilibration at 1 % O_2 for 24 hours (Figure 69), which is consistent with the reports suggesting these proteins are HIF target

genes. There was a decrease in KDM5B abundance after 48 minutes of re-oxygenation, but the level had not decreased to that of the normoxic cell. This is consistent with the hypothesis that KDM5B undergoes a slower rate of decay than HIF1α. A longer period of re-oxygenation is therefore required to study KDM5B dynamics. Unlike KDM5B, there was no evidence of oxygen-sensitive CDKN1C decay following re-oxygenation. Therefore, this candidate was not investigated further.

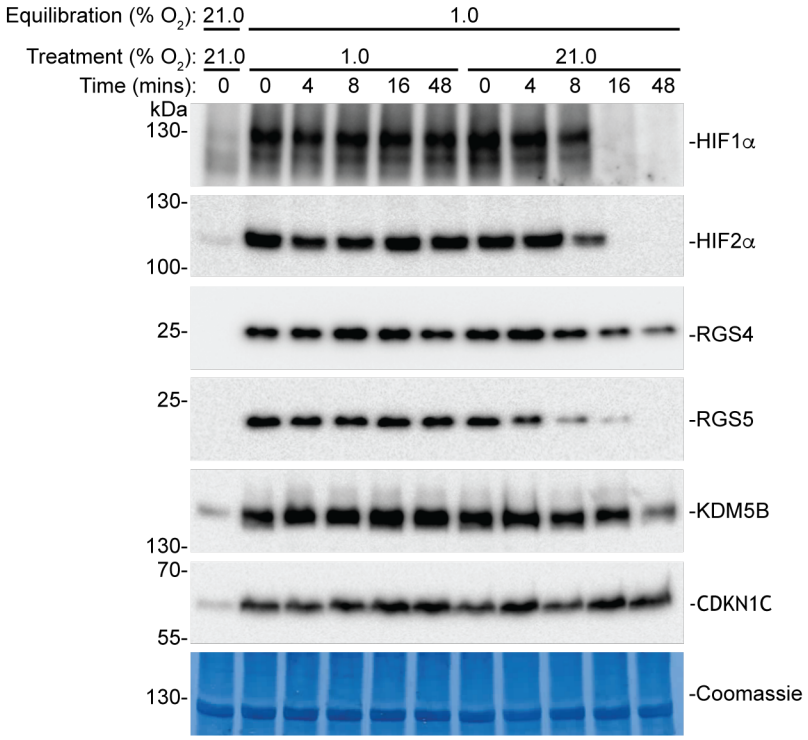


Figure 69 | Validation of KDM5B abundance change after 48 minutes of re-oxygenation.

A complicating factor to KDM5B regulation following re-oxygenation are the reports of KDM5B as a HIF target gene (Krieg et al., 2010; Pollard *et al.*, 2008; Xia *et al.*, 2009). Since HIFα is rapidly degraded upon re-exposure to 21 % O_2, it is possible that the abundance of short half-life HIF target genes decline upon re-oxygenation (but not in sustained hypoxia) as HIF transcriptional activity is lost. To address the possibility that decreased KDM5B abundance at 48 minutes of re-oxygenation is due to a loss of HIF transactivation, an experiment was performed with cycloheximide (CHX) to inhibit translation. HIF1α was used as the positive control for oxygen-dependent protein degradation.

Cells were equilibrated in normoxia for 24 hours before being treated for 3 hours in hypoxia or sustained normoxia with and without CHX. Both proteins show increased abundance after 3 hours of hypoxia (Figure 70). This could be due to increased protein synthesis or decreased degradation, or both, as is known for HIF1α. Inspection of the protein abundance levels after treatment with CHX revealed that decay was more rapid in normoxia than hypoxia. Therefore, at least part of the hypoxic induction appears to be due to stabilization.

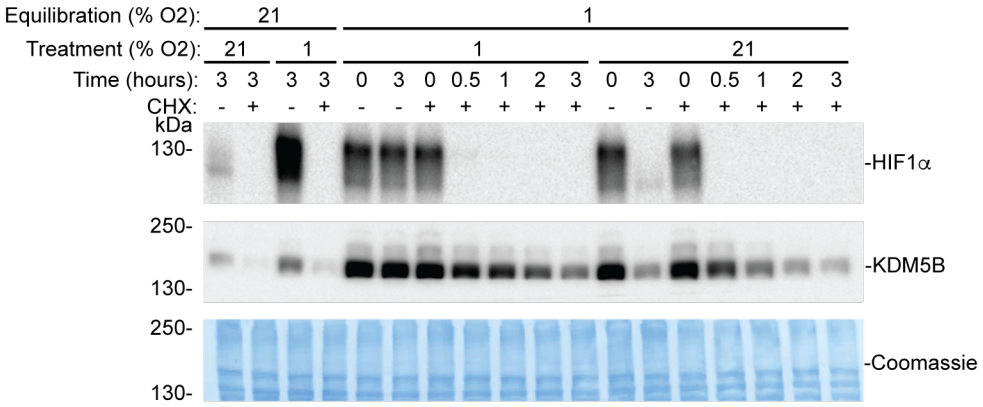


Figure 70 | KDM5B abundance change after reoxygenation in the presence of translation inhibition.

CHX = cycloheximide.

The cells were also equilibrated in hypoxia for 24 hours before re-oxygenation was performed for 3 hours with and without CHX. The cells were also maintained in hypoxia under these conditions to serve as a control. Without CHX, the abundance of both proteins remains high in hypoxia but decreases in normoxia. After CHX treatment both proteins decrease in both hypoxia and re-oxygenation. KDM5B abundance begins to decline in both conditions, but the effect is greater following re-oxygenation, as indicated by the densitometry plot (Figure 71). This suggests KDM5B turnover is reasonably fast in hypoxia ($t_{1/2}$ = 3 hours) when protein synthesis is inhibited in SH-SY5Y cells. However, KDM5B turnover appears to be even greater in CHX-treated SH-SY5Y cells when oxygen is abundant ($t_{1/2}$ = 1 hour). Since HIF transcriptional activity cannot account for these changes, it appears that KDM5B is a reasonably short half-life protein that exhibits greater turnover at high oxygen tension. Overall, these results have identified KDM5B as a target of oxygen-sensitive proteolysis.

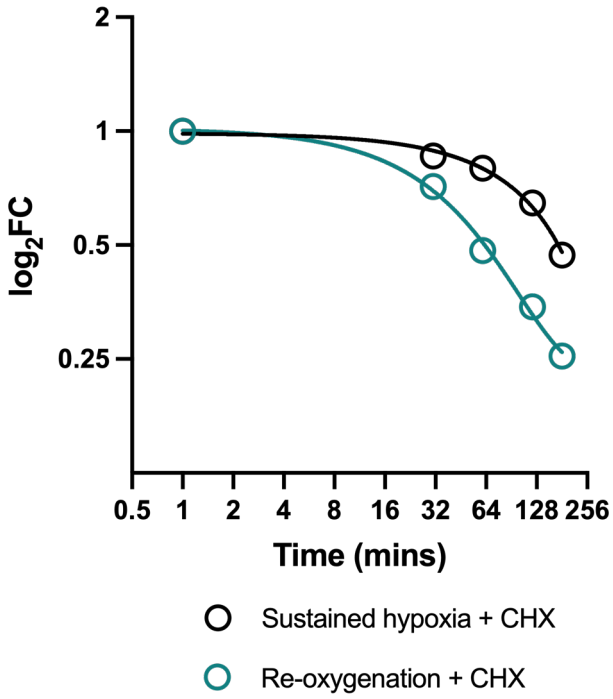


Figure 71 | KDM5B decay according to pixel densitometry plot from Figure 70.

Protein abundance was inferred from pixel density of the bands corresponding to KDM5B in Figure 70. Log_2(time + 1) is plotted against log_2(Fold-change relative to 0 minutes of re-oxygenation) for KDM5B in both conditions. A model of one-phase decay was fitted to demonstrate greater decay in re-oxygenated cells in the presence of cycloheximide (CHX).

6.4 Discussion

These experiments aimed to identify oxygen-labile proteins and used the regulation of HIF1α, HIF2α, RGS4, and RGS5 to inform the duration of re-oxygenation in which to assess protein turnover. The global profiling of protein dynamics in response to re-oxygenation identified 14 novel candidates that showed significant decay, and oxygen-sensitive protein turnover of KDM5B was validated. Although many proteins displayed large decay at 48 minutes of re-oxygenation, they were not considered to be significant by the GLM. Whilst it is possible a different statistical model could improve the identification of oxygen-labile proteins, the data in this analysis most likely reflects a genuine biological phenomenon in which relatively few proteins are degraded in an oxygen-sensitive manner, at least in SH-SY5Y cells. Several limitations of the proteomic study are also discussed (Section 6.4.3).

A literature review of the candidates was performed to determine if any of these proteins have previously been described to undergo rapid proteolysis. ZNF395 and KDM5B have been reported to be stabilised in the presence of MG132 (Chauvistré et al., 2020; Jordanovski et al., 2013; Xu et al., 2018). KDM5B, a lysine-specific demethylase that belongs to the 2OGD superfamily and specifically targets H3K4me3, has also been described as a HIF target gene and a study that was published in the course of this work found evidence for hypoxic stabilisation of KDM5B, independently of HIF transactivation (Zhou et al., 2021). An oxygen-sensitive function of KDM5A has recently been described in which H3K4me3 is rapidly demethylated following re-oxygenation (Batie et al., 2019). In this publication the contribution of KDM5B was dismissed because of its low abundance in normoxia. However, there is a possibility that the demethylase activity observed in this report could have been due to elevated KDM5B levels instead of an oxygen-sensitive catalytic property of KDM5A.

Given the number of studies reporting that KDM5B is stabilised by the proteasome inhibitor, MG132, and the link to hypoxic stabilisation of this protein, I decided to focus on validation of the protein dynamics observed for KDM5B in the proteome-wide functional screen. In conditions of translational arrest, KDM5B stability declined at a higher rate following re-oxygenation than in sustained hypoxia. This suggests KDM5B is subject to oxygen-dependent proteolysis, but the rate of attrition is slower than that of HIF1α. The stability of KDM5B, and factors that might regulate it, are discussed below (Section 6.4.1). Additionally, a strategy is proposed that would enable the other candidates identified in the proteomic screen to be studied independently of HIF transactivation (Section 6.4.2).

6.4.1 Factors that regulate KDM5B stability

The experiments outlined above identified KDM5B as a target of oxygen-sensitive proteolysis. However, the rate of degradation is slower than that of HIF1α. The stability of KDM5B has been reported to depend on both SUMOylation and ubiquitylation (Cuijpers et al., 2017), which is discussed in more detail below.

KDM5B is reported to be a substrate of a SUMO E3 ligase, CBX4 (Bueno and Richard, 2013). The authors reported K242 and K278 to be target sites of CBX4-catalysed SUMOylation, and that SUMOylated KDM5B is recognised by a ubiquitin E3 ligase, RNF4, and subsequently targeted for proteasomal degradation. A different study found that KDM5B protein abundance was increased when the deubiquitinase, USP38, was transiently overexpressed (Zhao et al., 2020), which is consistent with proteasomal turnover of KDM5B following its ubiquitylation.

The SUMOylation of KDM5B K242 and K278 have also been reported in additional proteomic screens (Hendriks et al., 2014; Hendriks et al., 2017; Xiao et al., 2015). However, in contrast to the results of Bueno and Richard (2013), SUMOylation was reported to protect KDM5B from RNF4-catalysed ubiquitination and subsequent proteasomal degradation. A separate proteome-wide screen investigated protein stability and found that KDM5B abundance increased following RNF4 knock-down (Kumar et al., 2017), which is in agreement with the previous reports of RNF4-catalysed ubiquitylation of KDM5B leading to its degradation. In addition, SUMOylation of KDM5B has been reported to decrease following DNA damage and is subsequently targeted for degradation by the proteasome, which was rescued by RNF4 knock-down (Hendriks et al., 2015). In contrast to these reports, KDM5B has been reported to undergo SKP2-catalysed ubiquitylation leading to proteasomal degradation (Liu and Zhou, 2022; Lu et al., 2015).

It has recently been found that hypoxia increases PIAS4-catalysed SUMOylation of KDM5B, which extends the half-life of KDM5B by preventing ubiquitylation and subsequent proteasomal degradation (Zhou *et al.*, 2021). The authors of this study compared the half-life of KDM5B in normoxia and hypoxia in the presence of CHX. In this report, there was a stabilising effect of hypoxia, which is consistent with the data provided in this thesis, but the authors did not consider re-oxygenation.

All of the above suggest KDM5B stability is governed by the ratio of SUMOylation (stabilising) and ubiquitylation (destabilising), and that SUMOylation is increased in hypoxia. Since the literature does not appear to have reached a consensus, more research would be required to clarify the ubiquitin and SUMO E3 ligases involved in the regulation of KDM5B. Together with the results from the analyses presented in this chapter, it is likely that the oxygen concentration influences the ratio of stabilising and destabilising

PTMs. However, it is unlikely that the catalytic activity of the KDM5B-specific ubiquitin and SUMO E3 ligases respond directly to oxygen. Therefore, a different, or perhaps additional, PTM (e.g., hydroxylation) is likely to be involved in the regulation of KDM5B stability.

Protein hydroxylation is the known mechanism by which protein stability is directly regulated by oxygen, but this modification has not been observed on KDM5B to date. Therefore, affinity purification of KDM5B coupled to LC-MS/MS should be performed to compare proteoforms from normoxia, hypoxia and re-oxygenation in the presence of proteasome inhibition. The use of orthogonal proteases would increase the coverage of the protein sequence (Giansanti et al., 2016) and a broad search for post translational modification would be considered. These experiments would aim to identify hydroxylation of KDM5B, and the regulation of other PTMs that might relay an oxygen-sensing signal (e.g., SUMOylation and ubiquitylation).

6.4.2 A strategy to study oxygen-labile candidates

The dynamics of KDM5B highlighted some complexities of the analysis. Firstly, the availability of specific and sensitive antibodies was fundamental to validation of endogenous KDM5B as an oxygen-sensitive protein. For other candidates another strategy might be desired in which epitope-tagged proteins are expressed to facilitate detection by western blot.

KDM5B is a HIF target gene with a seemingly short half-life, even in hypoxia (Figure 71). Such proteins are difficult to distinguish from targets of oxygen-sensitive proteolysis unless translational arrest is employed because the HIF transcriptional drive is removed shortly after re-oxygenation. Other strategies might circumvent HIF transcriptional activity, such as ectopic protein expression.

On the other hand, another difficulty is the prolonged half-life of KDM5B, at least when compared to the other positive controls. The duration of translation inhibitors must be extended to fully study the dynamics of long-lived proteins, but issues of cytotoxicity and cell death can be encountered if the treatments are performed for too long (Pajak et al., 2005).

With all of the above in mind a strategy was proposed to study the protein dynamics of the other candidates. In the first instance, gene expression in hypoxia would be tested by

RT-qPCR. This would indicate if it is possible that the loss of HIF transactivation contributes to protein abundance changes following re-oxygenation in SH-SY5Y cells. Proteins would then be analysed by stably expressing a fusion protein in which the target proteins contain a 3x-FLAG epitope tag. This would have the benefits of facile detection by western blot and removal from HIF transcription. The 3xFLAG tag also enables relatively easy immunoprecipitation, which would facilitate high coverage in LC-MS/MS experiments to prospect for oxygen-sensitive PTMs. Using these expressed fusion proteins would enable protein half-life to be determined in the presence of CHX. Additionally, a pulse-chase method could enable distinct protein populations to be analysed for PTMs by LC-MS/MS (Ross *et al.*, 2020). This would circumvent prolonged treatment of potentially toxic treatments to study protein half-life. A SILAC-based pulse-chase experiment would enable comparison of the proteins synthesised under hydroxylase restrictive conditions and the pre-existing population (Singleton *et al.*, 2011; Stoehr *et al.*, 2016).

Due to time constraints and interruptions to laboratory work from the Sars-CoV-2 pandemic, this work was not carried out to completion and is therefore not reported in this thesis.

6.4.3 Limitations to the proteomic screen

Several limitations of the proteomic screen and follow-up studies are discussed below.

Although extensive high pH off-line fractionation was performed to increase total proteome depth, HIF2α was not identified, which suggests there might be other oxygen-labile proteins beyond the limits of detection in this analysis.

It is also possible that the expression levels of oxygen-labile proteins might be higher in other cell lines, which would improve their detection. For example, IL-32, a substrate of ADO catalysed cysteine dioxidation is not expressed in SH-SY5Y cells (Masson *et al.*, 2019). The authors studied RKO cells to observe oxygen-sensitive proteolysis of IL-32, which indicates the choice of cell line is an important factor to consider in the identification of novel oxygen-labile proteins.

The data in this analysis was derived from a single experiment (n = 1). Many proteins had missing values (i.e., no peptides detected in one TMT channel). It is possible repeating this experiment would improve the accuracy of the data.

There were few sites of significant decay in re-oxygenation despite many proteins displaying -Δ relative abundance following re-oxygenation. This might indicate the statistical analysis was too stringent. The abundance of some proteins fluctuated across the time series, which results in poor goodness of fit despite an overall trend of attrition after re-oxygenation. This generates a low p-value, which prevents the protein from being considered to undergo significant abundance decrease following re-oxygenation. A second reason for a low p-value generated by the GLM is high abundance at the 0-minute time point. This can occur spuriously and means the data from 4, 8, and 16 minutes do not fit the line of best fit. Replicate data might improve the accuracy of protein abundance changes across the time series, which would increase the confidence of the analysis. However, repetition is expensive and time consuming. The candidates identified by these experiments require validation by orthogonal methods in any case.

Overall, the data presented here suggest KDM5B is an oxygen-labile protein, but the rate of decay is slower than HIF1α. The exact mechanism of degradation remains to be elucidated. Other candidates may yet be shown to be oxygen-labile proteins and further research is required to determine if these proteins are involved in physiological oxygen-sensing pathways.

Chapter 7. Discussion

The aim of this thesis was to discover novel sites of prolyl hydroxylation in the proteome. Several complementary proteomic approaches were performed to examine prolyl hydroxylation. Targets of the HIFα prolyl 4-hydroxylases (PHD1-3) were of particular focus because the role of these enzymes in cellular reprogramming *via* the HIF transcription factors demonstrates their function as critical oxygen sensors in cells (Pugh and Ratcliffe, 2017). Therefore, other substrates of these enzymes might also be expected to participate in oxygen-sensitive pathways that exert physiological control. The data presented in this thesis is derived from orthogonal proteomic methods to discover novel sites of prolyl hydroxylation or proteins that might be involved in oxygen-sensing pathways.

In total, 49 novel sites of hydroxyproline were discovered through this work (Table 62), and 9 of these were determined to be enzymatic prolyl hydroxylase substrates on the basis of suppressed hydroxylation when cells were treated with the pan-2OGD inhibitor, DMOG, or incubated in hypoxic conditions. Other candidates identified throughout this thesis might also display similar regulation in hydroxylase-restrictive conditions, but this was not observed because of the limited sensitivity of the assays. Although collagen prolyl hydroxylation is a widespread PTM, these results indicate prolyl hydroxylation of non-collagen targets is much more restricted in the proteome than might have been expected when compared to other PTMs, such as phosphorylation (Bekker-Jensen et al., 2020), ubiquitylation (Fulzele and Bennett, 2018), and SUMOylation (Hendriks *et al.*, 2017), for which thousands of substrates have been identified. This suggests a high degree of specificity of the prolyl hydroxylases for their substrates.

Some proteins are known to be processed in an oxygen-sensitive manner independently of the PHDs or pVHL (e.g., RGS4 and RGS5). Therefore, a functional screen was also included in this body of work that considered protein stability in response to fluctuating oxygen tension irrespective of protein modification. Several candidates were identified and further analysis of KDM5B indicated it is subject to oxygen-sensitive proteolysis. This result sets a precedent for more proteins to be discovered that are proteolytically processed in response to oxygen and may therefore participate in dynamic responses to oxygen tension fluctuation.

In this chapter I evaluate the orthogonal proteomic approaches performed in this thesis and propose further research that might immediately follow from this work. I also discuss the broader implications of this work, in particular the extent of prolyl hydroxylation in the proteome and the functional consequences of this PTM.

7.1 Orthogonal proteomics approaches to discover novel prolyl hydroxylase substrates

7.1.1 The application of LC-MS/MS to discover hydroxyproline

An *in silico* analysis of benchmark proteomic datasets deposited on PRIDE identified 11 sites of hydroxyproline for which there was strong evidence to support these assignments (Chapter 3). The analysis in Chapter 3 identified 14 further sites of prolyl hydroxylation that occurred in at least one of the 2 motifs (‘HyP-G’ and ‘P-HyP’) that were enriched with increasing stringency in the analysis. Throughout this thesis, several additional hydroxyproline sites have been identified that conformed to ‘HyP-G’ or ‘P-HyP’ sequences. Additionally, others have observed prolyl hydroxylation in ’HyP-G’ and ‘P-HyP’ motifs, along with similar motifs that can be supported by the data in this thesis (e.g., ‘HyP-X-P’, ‘P-X-HyP’, and ‘HyP-P-P’) (Zhou *et al.*, 2016). This suggests proline residues adjacent to glycine or other prolines can serve as minimal substrate sequences of prolyl hydroxylases. An additional putative hydroxylation consensus sequence was identified (‘F-X-A-P-W-C-X-H-C-X-X-X-P’) that corresponds to 16 distinct hydroxyproline sites on 8 PDIs (Chapter 3). Several of these sites were later shown to be suppressed in hypoxia or following DMOG treatment, or both, indicating likely involvement of 2OGDs in their formation (Chapter 4).

These results indicated the security of hydroxyproline assignment could be increased by i) high quality MS1 and MS2 spectra supporting the assignment, ii) identification of hydroxylation within a consensus sequences, and iii) suppression of hydroxylation in hydroxylase-restrictive conditions. However, these approaches also generated false negatives. For example, the high stringency workflow to identify hydroxyproline sites eliminated many collagen sites and the SILAC experiment comparing vehicle- and DMOG-treated MEFs did not display PDI hydroxyproline suppression because the duration of treatment was too short. Additionally, the sensitivity of pan-proteome

analyses was a limiting factor. This was exemplified by the absence of *bona fide* prolyl hydroxylase target sites (e.g., RPS23 HyP62 in Chapter 3 and HIF1α HyP564 in Chapter 4) despite performing the assays with experimental conditions that would be expected to detect these peptides. It was therefore necessary to apply alternative methods to identify more prolyl hydroxylase substrate candidates and proteins involved in oxygen-sensitive pathways.

Peptide enrichment has become an increasingly popular method to discover PTMs (Zhao and Jensen, 2009). I therefore sought to develop a peptide enrichment method that would increase the coverage of hydroxyproline-containing peptides in the proteome. pVHL was selected as the affinity reagent because of its known association with hydroxyprolyl HIFα proteoforms (Jaakkola *et al.*, 2001) and the recombinant trimeric complex, GST-VBC, has been demonstrated to bind a range of synthetic hydroxyproline-containing peptides (Guo *et al.*, 2016; Heir *et al.*, 2016; Kuznetsova *et al.*, 2003; Li *et al.*, 2019). Additionally, its role as a tumour suppressor has not been fully elucidated (Tarade and Ohh, 2018) and these assays were designed to identify proteins and peptides that bind directly to GST-VBC and could represent physiological substrates. Although some of these might interact with GST-VBC exclusively *ex vivo*, and therefore may not be physiological substrates of pVHL, there was a possibility that some *in cellulo* interactions might be identified that could lead to further understanding the physiological pVHL interactome, and hence provide new insights into its tumour suppressor function.

The application of GST-VBC as an affinity purification reagent identified over 200 distinct collagen hydroxyproline sites and 15 non-collagen sites (Chapter 5). The non-collagen hydroxyproline sites included endogenous HIF1α HyP564 and HIF2α HyP405, which provided the first demonstration of a pVHL-based reagent to affinity purify hydroxylated HIFα peptides (i.e., *bona fide* pVHL substrates). Some hydroxyproline sites that had been identified in previous chapters of the thesis were identified in the cellular extracts but not in the GST-VBC-bound fraction, which corroborates their assignment and indicates that binding of GST-VBC to hydroxylated peptides is likely to be sequence specific with binding restrictions extending beyond simply the presence of hydroxyproline.

Although the protein and peptide enrichment assays performed in Chapter 5 enabled novel hydroxyproline sites to be discovered, the sequence specificity of GST-VBC for

affinity enrichment was considered to be a potential limiting factor to the discovery of prolyl hydroxylase substrates. The data suggested that not all prolyl hydroxylase substrates would be detected by GST-VBC and these proteins might undergo proteolytic turnover as a consequence of prolyl hydroxylation (e.g., HIFα). A ubiquitin E3 ligase other than pVHL could be involved in regulating the oxygen-dependent degradation of such proteins in cells. Additionally, hydroxylation could occur on other residues (e.g., cysteine dioxidation of RGS4) that would not be recognised by GST-VBC. These proteins would be expected to be degraded in the pVHL-deficient cellular extracts used in Chapter 5. Therefore, a different method was required to identify such proteins.

A proteome-wide screen was performed to identify proteins that were destabilised in the presence of oxygen. Although 14 novel candidates were identified, data corroborating oxygen-sensitive stability was only generated for KDM5B, and the rate of degradation was much slower than that observed for HIF1α. The other candidates remain to be validated, so it is possible that some of the other candidates exhibit oxygen-sensitive protein turnover.

7.1.2 Prolyl hydroxylase substrates

A number of approaches were taken to identify novel prolyl hydroxylase substrates in the proteome (i.e., modifications that were catalysed by known or as-yet undiscovered 2OGD family members). One of the SILAC experiments performed in Chapter 4 considered oxygen-sensitive prolyl hydroxylation, which did not provide enzyme specificity *per se*. The other SILAC experiment considered prolyl hydroxylation in the context of PHD1-3 triple knock-out cells to identify PHD substrates or wild-type cells that had been treated with DMOG to indiscriminately search for 2OGD substrates. No substrates were identified for the PHDs, including the HIFα target sites, suggesting that such peptides were beyond the limits of detection in this assay. The prolyl hydroxylation of FKBP10, SERPINH1, LMAN1, and OS9 was suppressed by DMOG, suggesting these are substrates of a 2OGD. A further experiment in eHAP cells examined prolyl hydroxylation in hypoxia and after DMOG treatment, and suggested that PDIA3, PDIA4, and P4HB are also 2OGD substrates.

Given the involvement of 2OGD catalytic activity and the occurrence of prolyl hydroxylation in putative hydroxylation consensus sequences, it was possible to define

enzyme candidacy for several hydroxylation sites identified in this thesis. These are discussed in the sections below.

7.1.2.1 Putative C-P4H substrates

Collagen prolyl 4-hydroxylase (C-P4H) catalyses prolyl hydroxylation in the 'Yaa' positions of the repeating 'Xaa-Yaa-Gly' triplets of collagen and collagen-like proteins and in single 'P-G' sites (Gorres and Raines, 2010; Kiriakidis et al., 2017; McGee et al., 1971; Ono *et al.*, 2009; Yoneyama et al., 2013). Some of the candidates shown to undergo hydroxyproline suppression in the presence of DMOG display hydroxyproline in 'HyP-G' sequences (e.g., SERPINH1, LMAN1, and OS9), which suggest these candidates should be tested as C-P4H substrates. The prevalence of other candidates identified in this thesis in which hydroxyproline was identified in 'HyP-G' sites suggests these could also be C-P4H substrates (e.g., HYOU1 HyP977, CASC4 HyP252, and TMEM109 HyP42) even though DMOG sensitivity was not observed at these sites.

CTHRC1 was identified from the hydroxyproline enrichment experiments that use GST-VBC as the affinity reagent in Chapter 5. The interaction between CTHRC1 and GST-VBC was suppressed 2.5-fold in the presence of DMOG, which suggests that the interaction depends on hydroxylation of CTHRC1. It was therefore interesting to observe hydroxyproline-containing CTHRC1 peptides in the GST-VBC-bound fraction of the peptide enrichment assays. In both the protein and peptide enrichment assays CTHRC1 was found to be hydroxylated on multiple proline residues in the 'Yaa' position of repeating 'Xaa-Yaa-Gly' triplets, which is consistent with C-P4H-catalysed hydroxylation.

Prolyl hydroxylation of FKBP10 P36 was consistently identified in this thesis and has been documented elsewhere (Arsenault *et al.*, 2015; Onisko, 2020; Zhou *et al.*, 2016). The hydroxylation (^{36}HyPAG^{38}) does not appear to conform to a common hydroxylation consensus motif, but a similar hydroxylation site was observed in CTHRC1 (^{59}HyPAG^{61}). Given the likely relationship between CTHRC1 and C-P4H, it is possible that C-P4H catalyses hydroxylation of 'PAG' sequences in addition to 'PG' sites.

Collagen prolyl 3-hydroxylase (C-P3H) catalyses prolyl 3-hydroxylation in the 'Xaa' position of 'Xaa-Yaa-Gly' sequences when 'Yaa' is 4-HyP. It might therefore be possible for C-P3H to hydroxylate 'PAG' sequences. However, C-P4H has recently been

described to hydroxylate residues in the 'Xaa' position of 'Xaa-Yaa-Gly' triplets (van Huizen et al., 2019), whereby 3-Hyp and 4-HyP were distinguished by precursor ion fragmentation in ETD-HCD mode. In the method described in this report, peptides are first fragmented using ETD (to generate MS2 spectra) and then another round of fragmentation is performed in HCD mode (to generate the MS3 spectrum) (i.e., ETD-HCD fragmentation). Luider and colleagues report that distinct products form in the MS3 spectra, which enables 3-Hyp and 4-Hyp to be distinguished (van Huizen *et al.*, 2019). Interestingly, alanine was the residue that was observed to most frequently occupy the 'Yaa' position when 4-hydroxyproline was identified in the 'Xaa' position of 'Xaa-Yaa-Gly' sequences of collagen (van Huizen *et al.*, 2019). This supports the hypothesis that CTHRC1 P59 and FKBP10 P36 (i.e., 'PAG' sequences) are C-P4H substrates.

It therefore appears likely that C-P4H has a broad range of substrates beyond collagen. Although collagen prolyl hydroxylation is mainly considered to be a post translational modification (Salo and Myllyharju, 2021), there have been reports that it can occur in a co-translational manner (Ishikawa and Bächinger, 2013; Kirk et al., 1987). Nascent polypeptide chains do not have the same structural folds as matured proteins, which might enable C-P4H to catalyse prolyl hydroxylation on a broad range of substrates. To test this hypothesis, several *in vitro* hydroxylation assays were performed in collaboration with Dr Antti Salo and Dr Johanna Myllyharju (Oulu University) in which recombinant C-P4H was reacted with the products of *in vitro* transcription translation (IVTT) in rabbit reticulocyte lysates. The candidates tested included COL19A1, FKBP10, SERPINH1, LMAN1, OS9, and P4HB. The reactions were performed in duplicate, with one IVTT assay performed under normal conditions and the other incorporated [3H]-Pro into the polypeptide sequence of the substrates, which enabled hydroxyproline detection by LC-MS/MS or radiochemical assay, respectively. Although prolyl hydroxylation of COL19A1 was observed in a post translational reaction, none of the other candidates served as C-P4H substrates in these assays. However, addition of C-P4H during polypeptide synthesis (i.e., a co-translation hydroxylation assay) indicated that all of the candidates tested were substrates of C-P4H, except P4HB (data not shown). These results support the hypothesis C-P4H acts as a prolyl hydroxylase for a broad range of substrates by engaging with the nascent polypeptides.

In support of these findings, a recent report was published during the work described in this thesis, which indicated that C-P4H associates with polysomes and performs co-translation prolyl hydroxylation of flavivirus proteins (Aviner et al., 2021). Interestingly, several of the candidate substrates identified in this thesis were also identified in the polysome fractions (e.g., FKBP10 and SERPINH1) (Aviner *et al.*, 2021), which provides further evidence in support of co-translational hydroxylation performed by C-P4H.

Overall, it is likely that C-P4H represents the prolyl hydroxylase with the largest range of substrates identified in this thesis, but further research is required to validate these assignments in a cellular context. Whilst prolyl hydroxylation of triple helical proteins can occur in a post translational manner, it appears likely that C-P4H also engages in co-translational hydroxylation of nascent polypeptides, which exhibit reduced conformational constraints. The work in this thesis raises the possibility that this facilitates C-P4H to hydroxylate a broad range of substrates beyond the triplet repeat collagens and collagenous domain proteins. Further research is required to confirm this hypothesis.

7.1.2.2 Prolyl hydroxylation of PDIs

Prolyl hydroxylation of PDIs was identified in Chapter 3 and a putative hydroxylation consensus motif was determined. Hydroxylation of PDIs has also been observed elsewhere (Henningsen *et al.*, 2010; Onisko, 2020; Stoehr *et al.*, 2016) The assays in Chapter 4 demonstrated that prolyl hydroxylation of PDIs is suppressed by both hypoxia and DMOG treatment. The latter is particularly important because it suggests PDIs are substrates of a member of the 2OGD superfamily.

Prolyl hydroxylation of PDIs has also been reported previously, and a mechanism of C-P4H "auto-hydroxylation" has been proposed (Onisko, 2020). This hypothesis was most likely generated because P4HB, which is hydroxylated at P51 and P395, is the β subunit of C-P4H. However, *in vitro* assays of recombinant C-P4H formation suggest that the prolyl hydroxylase complex has highly specific requirements for P4HB to act as the β subunit (Koivunen et al., 1996). Therefore, auto-hydroxylation would be expected to only serve as a mechanism of P4HB hydroxylation. Furthermore, the peptide sequence surrounding the target sites of PDIs appears to be highly conserved and different from C-P4H target sequences. Finally, the *in vitro* hydroxylation assays described above did not

demonstrate C-P4H-catalysed prolyl hydroxylation of P4HB. This suggests an alternate prolyl hydroxylase is involved in the post translational modification of PDIs.

PDIs are largely restricted to the ER and secretory pathway. This would appear to exclude OGFOD1 and PHD1-3 catalysed prolyl hydroxylation of these proteins. Furthermore, PDI target sites were found to be hydroxylated in MEF TKO cells (Chapter 4), which provides additional evidence that PHD1-3 are not responsible for the catalysis of these sites. P4H-TM is an ER-resident prolyl 4-hydroxylase that has not been assigned any substrates *in cellulo* even though it can hydroxylate HIFα polypeptides produced in rabbit reticulocyte lysates (Koivunen *et al.*, 2007). Additionally, PDIA3 was identified as an interactor of JMJD8 by immunoprecipitation (Yeo et al., 2017). JMJD8 is an ER-resident member of the 2OGD superfamily that has not yet been fully characterised. Therefore, it is possible P4H-TM, or another 2OGD such as JMJD8, is the enzyme responsible for prolyl hydroxylation of PDI.

7.1.2.3 Methods to generate hydroxylase assignment

Hydroxylase assignments can be performed with *in vitro* hydroxylase assays using recombinant hydroxylases and synthetic peptide or IVTT-produced polypeptides as substrates. Whilst these assays provide direct evidence for enzyme-specific hydroxylation of the candidate (Cockman *et al.*, 2019), *in vitro* hydroxylation assays are not performed under physiologically relevant conditions. This can lead to hydroxylation of non-physiological substrates (i.e., false positive assignments), or additional factors may be required for hydroxylation that are not present in the *in vitro* reaction (i.e., false negative assignments) (Cockman *et al.*, 2019). Therefore, it is advisable to generate the assignment in cells using endogenous protein expression levels, whenever possible.

Hydroxylase assignments in physiologically relevant conditions can be achieved by observing suppression of hydroxylation in cells that are deficient for specific hydroxylase activity. This can be achieved with CRISPR-Cas9 technology to generate knock-out (KO) cells of the target hydroxylase (Cockman et al., 2022) or RNAi-based knock-down of specific transcripts (Singleton *et al.*, 2014). This approach would be suitable for screening substrate candidates of different hydroxylases. Some of the candidates are highly abundant and would be expected to be detected in unfractionated lysates by LC-MS/MS (e.g., PDIs in eHAP cells; Chapter 4). For less abundant proteins, off-line fractionation

of the extracts or affinity purification coupled to mass spectrometry (AP-MS) of the target proteins can increase the likelihood of detecting the target site.

Given the large number of hydroxylases that would need be interrogated, a multiplexed proteomic screen, for example TMTpro™ 18-plex labelling, could provide a time-efficient method that enables comparison of multiple prolyl hydroxylase-deficient cells (Li et al., 2021). Alternatively, targeted proteomics can facilitate identification of multiple target peptides in a single LC-MS/MS run (Doerr, 2012; Peterson et al., 2012).

AP-MS perhaps offers the most direct route to enzyme assignment of specific target sites. However, this would represent a time-consuming task to characterise all target sites from cells with perturbed candidate hydroxylase expression. The availability of specific antibodies can also be a limiting factor but this could be circumvented by applying CRISPR Cas9 technology to knock-in epitope tags to the endogenous proteins (Schmid-Burgk et al., 2016).

In the absence of securing an enzyme assignment, providing evidence of enzymatic hydroxylation can be a valuable result for directing further research. For example, treating cells with a generic iron chelator, 2,2-DIP, and 2OGD specific inhibitors, DFO and DMOG, informed researchers that RGS4 and RGS5 were subject to enzymatic hydroxylation but this was independent of 2OGD activity (Masson *et al.*, 2019). In a similar manner, experiments with DMOG in this thesis have identified PDIs, FKBP10, and several proteins containing ‘HyP-G’ sites as putative 2OGD substrates (Table 62).

7.1.3 Functional consequences of prolyl hydroxylation

Prolyl hydroxylation has been shown to be a key signalling mechanism in cells through the PHD-HIFα-pVHL axis (Ratcliffe et al., 2017). Since prolyl hydroxylation is believed to be irreversible, bimodal signals are expected to be achieved through protein degradation and re-synthesis. The interaction between HIFα and pVHL is specific to prolyl hydroxylated HIFα proteoforms (Jaakkola *et al.*, 2001). It is therefore likely that prolyl hydroxylation alters the interactome of the putative prolyl hydroxylase substrates, which could lead to proteasomal degradation by pVHL or another ubiquitin E3 ligase. Several proteins have been found to interact with collagen and the apparent affinity of the interaction increases when hydroxyproline is present on collagen (Rappu et al., 2019).

Therefore, prolyl hydroxylation could also alter the interactome of target proteins and lead to altered functions that are unrelated to protein turnover.

A second mode of achieving bimodal signals in response to oxygen tension has been proposed by Espenshade and colleagues (Clasen et al., 2017). The hydroxylase is likely to exhibit higher affinity for the non-hydroxylated substrate than the hydroxylated product. Oxygen-sensitive prolyl hydroxylation could therefore regulate protein complex formation, which achieves signalling through third party molecules. This has been demonstrated for the yeast analogues of OGFOD1 (Ofd1) and its substrate RPS23 (Rps23), which regulate the transcriptional activity of SREBP (Sre1) by sequestration of Ofd1 protein into a stable complex (Clasen *et al.*, 2017). Hydroxylation of Rps23 leads to dissociation of the complex, which enables Ofd1 and Sre1 to interact and initiate a transcriptional cascade (Clasen *et al.*, 2017; Hughes and Espenshade, 2008).

Prolyl hydroxylation may be involved in biochemical process that are not considered to be oxygen signalling pathways. It has been proposed that the initial phase of disulphide bond formation catalysed by PDIs is oxygen-independent, but a second phase of PDI-dependent disulphide bond formation and isomerisation is oxygen-dependent (Koritzinsky et al., 2013; Levitin et al., 2021). Prolyl hydroxylation occurs in proximity to the active site sequences of the thioredoxin-like domains of many PDIs (Chapter 3 and Chapter 4). It is therefore possible that oxygen-dependent prolyl hydroxylation is involved in regulating the catalytic activity of PDIs.

Finally, C-P4H catalysed prolyl hydroxylation contributes to thermal stability of collagen proteins (Salo and Myllyharju, 2021). With the exception of CTHRC1, which exhibits prolyl hydroxylation on 5 residues, the prolyl hydroxylase substrate candidates identified in this thesis contain a maximum of two hydroxyproline sites. It is unlikely that hydroxylation of one or two residues would significantly increase the thermal stability of proteins in a similar manner to collagen.

7.1.3.1 pVHL substrates

pVHL is a protein that exclusively binds to prolyl hydroxylated HIFα proteoforms. It is therefore integral to the PHD-HIFα-pVHL signalling axis that facilitates transcriptional responses to oxygen (Jaakkola *et al.*, 2001). As a ubiquitin E3 ligase, it targets prolyl hydroxylated HIFα for proteolytic processing in the proteasome and a number of

mutations associated with VHL disease perturb this function. In this context, HIFα is reported to be oncogenic (Schödel *et al.*, 2016). It is therefore possible that other prolyl hydroxylase substrates are modified by pVHL-dependent ubiquitylation and dysregulation of these proteins in VHL disease could be oncogenic.

Given its function as a tumour suppressor (Kaelin, 2017), a number of studies have attempted to identify novel prolyl hydroxylase substrates that are also pVHL targets (Zhang and Zhang, 2018; Zurlo *et al.*, 2016). In this thesis, recombinant GST-VBC was applied as an affinity reagent to directly enrich hydroxyproline-containing peptides with the aim of identifying novel prolyl hydroxylase substrates. It is possible that these interactions occur exclusively *ex vivo*. Therefore, an important addition to the discoveries in this thesis would be evidence of an interaction occurring *in cellulo*.

Lu and colleagues performed a functional assay of pVHL by comparing the ubiquitylation status of proteins from 786-O cells (i.e., pVHL-deficient) and cells expressing ectopic pVHL (Li et al., 2007; Wang et al., 2022). A proteasome inhibitor, MG132, was applied to both experimental groups and affinity purification of the ubiquitin remnant motif (Xu et al., 2010) was performed on tryptic peptides to identify ubiquitylated proteins. pVHL-dependent ubiquitylation was observed for HIF1α and HIF2α, which confirmed the approach taken could identify *bona fide* prolyl hydroxylated pVHL substrates. In addition to these proteins, 14 candidates identified in this thesis were also found to be ubiquitylated in a pVHL-dependent manner (Table 61), which is indicative of an interaction occurring *in cellulo*. COL6A2, COL18A1, and KHDRBS1 are direct targets of GST-VBC that were identified in the peptide enrichment assays (Chapter 5), and the data from Lu and colleagues strengthens the hypothesis that these proteins are direct targets of pVHL in cells.

Some caution should be applied when interpreting these results; although the proteins identified in this proteome-wide screen could be direct targets of pVHL-catalysed ubiquitylation, expression of pVHL in 786-O cells would be expected to cause re-programming of the cell because HIF transactivation is deactivated. Therefore, ubiquitylation could be an indirect consequence of pVHL re-expression in which other ubiquitin E3 ligases are upregulated in these cells.

Table 61 | Overlap between pVHL-dependent ubiquitylated proteins identified by Wang *et al.* (2022) and the targets of prolyl hydroxylation or oxygen-sensitive proteolysis identified in this thesis.

Protein	Discovery in this thesis	Reference
ACSL4	DMOG-sensitive interactor of GST-VBC; Significant protein decay following re-oxygenation in SH-SY5Y cells	Chapters 5 & 6
CAPN2	High confidence hydroxyproline sites (HyP221 HyP222) in 'p-HyP' sequences	Chapter 3
COL18A1	Protein interactor of GST-VBC; direct enrichment of prolyl hydroxylated peptides by GST-VBC	Chapter 5
COL6A2	DMOG-sensitive protein interactor of GST-VBC; direct enrichment of prolyl hydroxylated peptides by GST-VBC	Chapter 5
FLNB	Confident hydroxyproline site (HyP1429) but insufficient support from assignment from the MS2 spectrum	Chapter 3
HYOU1	High confidence hydroxyproline site (HyP977) in a 'HyP-G' sequence	Chapter 3
KHDRBS1	DMOG-sensitive interactor of GST-VBC; direct enrichment of prolyl hydroxylated peptide by GST-VBC	Chapter 5
MRC2	Significant protein decay following re-oxygenation in SH-SY5Y cells	Chapter 6
MYO10	Confident hydroxyproline site (HyP255) in a 'HyP-G' sequence but insufficient support from assignment from the MS2 spectrum	Chapter 3
LMAN1	2OGD substrate (HyP378)	Chapters 3 & 4
PDIA3	2OGD substrate (HyP404)	Chapters 3 & 4
PFKP	DMOG-sensitive interactor of GST-VBC; Significant protein decay following re-oxygenation in SH-SY5Y cells	Chapters 5 & 6
SERPINH1	2OGD substrate (HyP30)	Chapters 3 & 4
SUN2	High confidence hydroxyproline site (HyP315) in a 'HyP-G' sequence	Chapter 3

7.1.4 Other mass spectrometry methods that could be applied to study proteome-wide prolyl hydroxylation

The experiments in this thesis describe a range of orthogonal proteomics approaches to identify novel prolyl hydroxylase substrates. There is further discovery potential, and the

application of other complementary LC-MS/MS-based methods could be performed, which are briefly discussed in this section.

The application of GST-VBC as a reagent for hydroxyproline-containing peptide enrichment was successful (Chapter 5). Other hydroxyproline-binding proteins exist (Rappu *et al.*, 2019), which could be developed into hydroxyproline affinity reagents, or the development of hydroxyproline specific antibodies and nanobodies could lead to novel affinity purification reagents. Although direct enrichment of hydroxyproline-containing peptides can lead to the discovery of target sites in hydroxylase substrates, these assays do not provide the necessary level of information to determine the candidate enzyme. A method that provides proof of enzymatic hydroxylation is still required from additional assays.

An interactomics method for identifying putative hydroxylase substrates is "substrate trapping", which has been applied to a variety of 2OGDs (Cockman *et al.*, 2022; Cockman *et al.*, 2009; Rodriguez *et al.*, 2016). Candidates are identified according to increased abundance in the presence of DMOG, which suggests the interaction occurs between the substrate and active site of the hydroxylase. Methodological advances make it possible to generate epitope-tagged endogenous proteins (Schmid-Burgk *et al.*, 2016), which could increase the accuracy of interactomics assays. In DMOG-based "substrate trapping" experiments, hydroxyproline is unlikely to be observed because the hydroxylation reaction does not proceed to completion. Therefore, additional assays are required to demonstrate enzymatic hydroxylation.

Another interactomics method is proximity ligation, such as BioID or TurboID (May et al., 2020). These assays use promiscuous biotin ligase fusion proteins to tag neighbours in proximity to the target protein. This reaction is fast and can therefore capture transient interactions, which would enable prolyl hydroxylation to occur. Using endogenous bait protein has been shown to increase the sensitivity and accuracy of these assays by reducing the number of non-specific interactors (Santos-Barriopedro et al., 2021; Stockhammer et al., 2021). After biotinylation, affinity purification is performed at the protein level to enrich biotinylated proteins. Therefore, it might be possible to observe prolyl hydroxylation on the substrate, which increases the discovery potential of this method because putative substrates can be discriminated from other interacting partners

that are not involved in the hydroxylation reaction. Despite this, additional assays are required to demonstrate enzymatic hydroxylation.

Overall, the orthogonal proteomic approaches considered in this thesis could be expanded to discover more sites of prolyl hydroxylation and assign the relevant prolyl hydroxylases. Several 2OGDs have been shown to catalyse hydroxylation on a broad range of substrates. These include FIH (Cockman *et al.*, 2009) and C-P4H (Gorres and Raines, 2010). During the course of preparing this thesis, JMJD6 has been discovered to hydroxylate lysine residues in lysine-rich disordered regions on a large number of proteins (Cockman *et al.*, 2022). Therefore, there is a precedent for 2OGDs performing protein hydroxylation on a range of substrates.

7.2 Evaluating the substrate specificity of PHD1-3

At least 45 non-HIFα PHD substrates have been reported to date (Table 14). At the beginning of my PhD project, 23 of the reported non-HIFα PHD substrates were being tested by *in vitro* hydroxylation assays in which HIFα was faithfully hydroxylated by PHD. Despite this, there was no evidence for PHD-dependent prolyl hydroxylation of any of the 23 reported non-HIFα PHD substrates (Cockman *et al.*, 2019). Although this work provided no support for these reports of non-HIFα PHD substrates it remained possible that, unlike HIFα, additional factors are required for PHD-catalysed hydroxylation of non-HIFα substrates. Such factors would presumably be absent in the *in vitro* assays, so cellular data was required to resolve this question.

Throughout this thesis the reported non-HIFα PHD substrates have been analysed for evidence of prolyl hydroxylation or altered protein stability in hydroxylase-restrictive conditions. Table 63 (Chapter 8) provides a comprehensive summary of the data relating to the reported non-HIFα PHD substrates in each of the experiments performed throughout this thesis. From these results, there is sufficient cellular data to contest the assignments of the reported non-HIFα PHD substrates. Despite this, continued research is required to fully determine the specificity of the PHDs.

7.3 Implications for VHL disease

VHL disease is an autosomal dominant disorder that predisposes effected individuals to developing various cancers, including clear cell renal cell carcinoma (ccRCC), haemangioblastoma (HAB), and phaeochromocytoma (Kaelin, 2007; 2017). Multiple subtypes of *VHL* mutations (Type 1, 2A, 2B, and 2C) are present in VHL disease, and sporadic development of ccRCC or HAB have also been linked to *VHL* mutations. Type 2A and 2B mutations perturb the oxygen dependent degradation of HIFα. However, type 2C mutations permit functional oxygen sensitive HIFα proteolysis. This has, in part, led many to postulate non-HIFα proteins are also pVHL targets and contribute to tumorigenesis (Kaelin, 2002; Ratcliffe, 2003).

Although GST-VBC has been developed in this body of work as an affinity reagent for hydroxyproline-containing peptides, there is a possibility that candidates identified from these assays might be physiological pVHL substrates and could contribute to the VHL disease when the interaction with pVHL is perturbed. There was some evidence of non-HIFα proteins interacting with GST-VBC in Chapter 5, which included proteins that have been reported to interact with pVHL, such as type IV collagen (Grosfeld *et al.*, 2007; Kurban *et al.*, 2008) and FN1 (Clifford et al., 2001; Hoffman *et al.*, 2001; Ohh *et al.*, 1998). Several novel prolyl hydroxylase substrate candidates were also identified. Since DMOG suppressed these interactions, it would appear that GST-VBC binding is dependent on prolyl hydroxylation. Binding of GST-VBC to tryptic peptides provided evidence of a direct interaction between collagens and several novel putative prolyl hydroxylase substrates. However, since these interactions occurred *ex vivo* they did not provide proof of a direct interaction occurring *in cellulo*. It was interesting to note that the novel candidates have been reported to correlate with unfavourable prognosis in renal cancer, which further implicates these proteins as potential proto-oncogenes, at least in the context of VHL disease.

In order to test the *in cellulo* interactions of pVHL, a similar functional assay could be performed to the one Lu and colleagues presented (Wang *et al.*, 2022). pVHL-dependent ubiquitylation of the target proteins would increase the confidence of a cellular interaction occurring. Importantly, ubiquitylation in the context of type 2A, 2B or 2C *VHL* mutations

might provide insight into the involvement of these candidates in the tumorigenesis associated with VHL disease.

In addition to functional assays, the proximity ligation assays described in Section 7.1.4 would be applicable to defining pVHL substrates in cells. These studies would also enable the interactomes of different pVHL mutants to be studied. Cellular interactomics assays might also provide some information regarding the reported interactions of pVHL with other proteins (Tabaro *et al.*, 2016) that were not identified in the GST-VBC affinity purification experiments.

7.4 Outlook

7.4.1 New technologies that further improve data acquisition of hydroxyproline sites

There are three modes of precursor ion selection, which are outlined in Figure 72. Data dependent acquisition (DDA) has been used throughout this thesis. In this mode of precursor ion selection, a threshold such as 'TopN' or 'Cycle Time' is defined. In each MS cycle, precursor ions are stochastically selected for fragmentation until the threshold has been achieved. This mode favours more abundant precursor ions as they are more likely to be selected (Gillet *et al.*, 2016). The MS2 spectra generated in DDA mode are relatively easy to analyse because they are derived from a single precursor ion (except for chimaeric spectra). This makes DDA a highly suitable method for manual inspection of MS2 spectra to corroborate PTM assignments generated by search algorithms. However, DDA data is often inconsistent between samples and injections because of the stochastic nature of precursor ion selection (Liu et al., 2004).

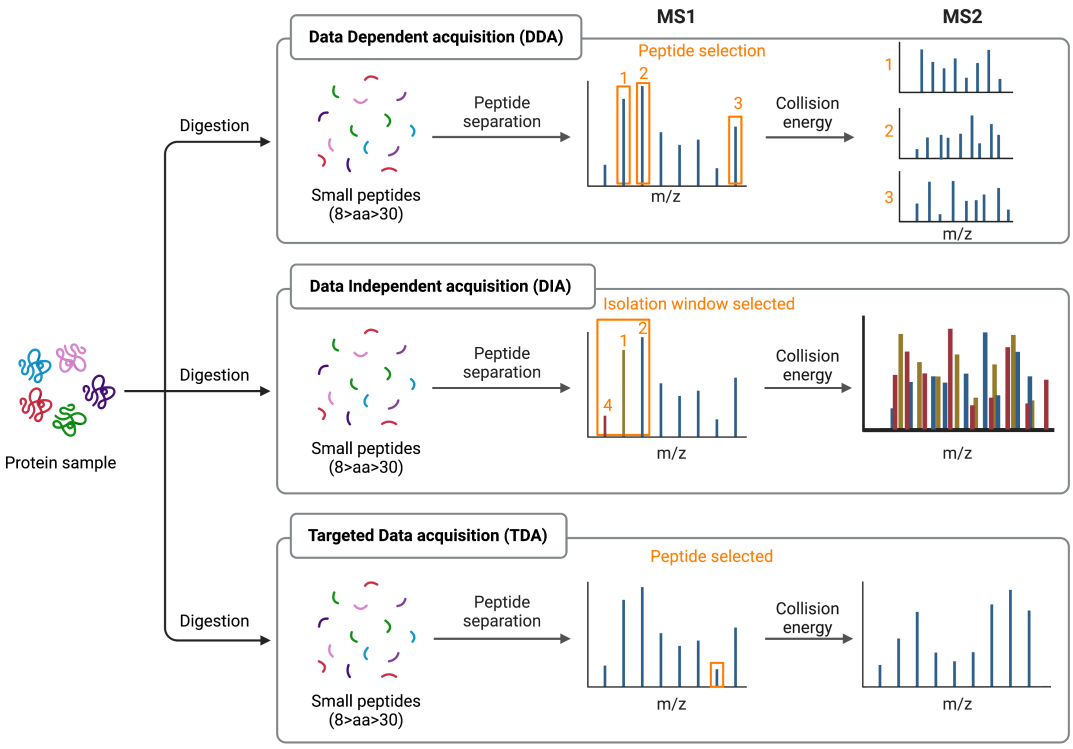


Figure 72 | Orthogonal methods for precursor ion selection.

Schematic outlining the 3 modes of precursor ion selection. Digested peptides are resolved by liquid chromatography and injected into the mass spectrometer. Three modes of precursor ion selection can be performed: data dependent acquisition (DDA), data independent acquisition (DIA) and targeted data acquisition (TDA). For DDA, precursor ions are selected stochastically and sequentially throughout the MS cycle. This generates a unique MS2 spectrum for each precursor ion. For DIA, all precursor ions in an isolation window are simultaneously selected for fragmentation. This generates a complex MS2 spectra containing the fragment ions of all precursor ions in the isolation window. For TDA, the instrument selects a user-defined precursor ion and generates a single MS2 spectrum.

Data independent acquisition (DIA) uses defined *m/z* isolation windows in each MS cycle and selects all the precursor ions in each window for fragmentation. This method ensures the generation of fragment ions for all precursor ions in a sample, which increases the reproducibility of peptide identification between samples. Moreover, the isolation windows are assigned such that DIA can be used as an alternative to DDA for proteome-wide shotgun analyses (Gillet et al., 2012). Despite these advantages, the MS2 spectra of DIA experiments are highly complicated owing to the simultaneous fragmentation of many precursor ions. Thus, deconvolution of the MS2 spectra to generate peptide and PTM assignments either require a spectral library, which is acquired in DDA mode, or else a sophisticated machine-learning programme (Demichev et al., 2020; Ludwig et al.,

2018). It is therefore difficult to manually inspect the MS2 spectra of DIA experiments to corroborate PTM assignments generated by the software.

Targeted data acquisition (TDA) is a mode of precursor ion selection that instructs the instrument to only select specific precursor ions for fragmentation. This is achieved by defining a so-called 'inclusion list', which contains the parameters required to identify the target precursor ions (e.g., *m/z*, *z*, and RT). TDA reproducibly generates MS/MS assignments for the target peptides and can often achieve more sensitive detection of peptides (Kiyonami et al., 2011). A major limitation of TDA mode is the number of measurable peptides, which is typically capped at several hundred to maintain sensitivity of the assay (Picotti and Aebersold, 2012). An additional difficulty is the variance of chromatographic behaviour between experiments, but this can be overcome by performing real time retention time calibration or internal standard-triggered acquisition of the endogenous peptides (van Bentum and Selbach, 2021).

The application of DDA to identify novel hydroxyproline sites has been demonstrated throughout this thesis. A key feature of this analytical method is the generation of MS2 spectra from a single precursor ion, which facilitates manual inspection and corroboration of software assigned PTMs. This is particularly important for studies of protein oxidation because the potential for an improperly localised PTM assignment is very high (Chapter 3.3.1). Although DIA and TDA have their advantages for increased consistency of peptide identification between samples, there is a reliance of prior knowledge of the MS2 spectra to generate the peptide and PTM assignment. Therefore, these methods are more suitable for validating the target sites and detecting them in functional studies, whereas DDA remains the preferable precursor ion selection mode during the initial discovery phase. The application of these orthogonal methods should facilitate discovery, validation, and quantification of hydroxyproline sites in future experiments.

7.4.2 Oxygen labile proteins

Several candidates of oxygen-dependent proteolysis include NDUFA4L2, ZNF395, and KDM5B. These proteins have also been described as HIF target genes (Jordanovski *et al.*, 2013; Tello et al., 2011; Xia *et al.*, 2009), which increases the difficulty of the validation experiments. This is because a short half-life HIF target would be expected to show similar dynamics to oxygen labile proteins in the conditions used in the proteome-

wide screen. I therefore performed a follow-up experiment to study the oxygen-dependent turnover of KDM5B in the presence of cycloheximide (CHX) to inhibit global translation, which includes translation of HIF target genes. This experiment demonstrated that KDM5B turnover is enhanced upon reoxygenation (Chapter 6.3.4). Further studies might consider the application of a proteasome inhibitor (e.g., MG132) to rescue protein stability. The application of CHX has an advantage of inhibiting all HIFα isoforms, whereas genetic intervention with RNAi or CRISPR-Cas9 strategies are unlikely to inhibit all HIF dependent responses to hypoxia.

Exogenous expression of the target genes is another strategy to discriminate between oxygen-dependent proteolysis and turnover of short half-life HIF target genes. The introduction of an epitope tag would facilitate detection by immunoblot and enrichment by immunoprecipitation. Analysis of the immunoprecipitated sample could lead to the identification of PTMs that are involved in the oxygen-dependent proteolytic pathway, such as hydroxylation and ubiquitination.

The purpose of the functional screens was to identify novel targets of oxygen dependent proteolysis that might be involved in acute cellular responses to hypoxia and reoxygenation. To this end, identification of an oxygen-labile protein must lead to further discoveries to elucidate any potential signalling axis. Knowledge of the PTM that primes the target for degradation would be a key finding, as would the protein that initiates degradation. By means of comparison, PHD catalysed prolyl hydroxylation of HIFα is recognised by the ubiquitin E3 ligase, pVHL, which initiates HIFα degradation by the ubiquitin proteasome system. It might therefore be of interest to test the stability of the candidates in cells deficient for known hydroxylases or pVHL. The experiments outlined above should enable protein half-life to be determined at different oxygen concentrations, along with the key PTM sites that regulate turnover of these proteins.

7.4.3 Low stoichiometry hydroxyproline assignments

Since protein oxidation is believed to be irreversible, it is logical that enzymatic hydroxylation would be a high stoichiometry modification. As a signalling event, prolyl hydroxylation of HIFα is near complete (i.e., >99 % stoichiometry) and leads to degradation of the hydroxylated proteoform. The data acquired for collagen in this thesis suggests the hydroxylation pattern is heterogenous and a broad range of stoichiometries

is observed (10 – 100 %) in hydroxylation permissive conditions (Figure 73). Despite this, the combined effect of many hydroxylation events is to increase the thermal stability of collagen. RPS23 HyP62, a substrate of the nuclear 2OGD, OGFOD1, was exclusively observed at 100 % stoichiometry in this thesis. Thus, bona fide prolyl hydroxylation occurs at high stoichiometry, regardless of its function.

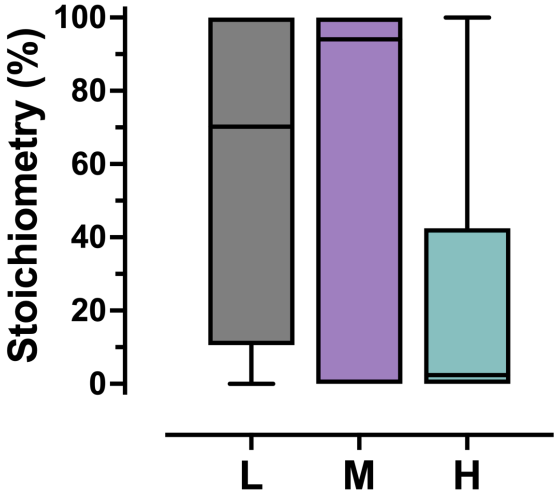


Figure 73 | Collagen hydroxyproline stoichiometry in MEF cells.

The distribution of stoichiometry of prolyl hydroxylation in the MEF SILAC experiment described in Chapter 4.3.2 is visualised by Tukey box plots. Boxes represent 1.5 x IQR, and whiskers represent values outside of this range. The median value is depicted by the horizontal bar. L = vehicle-treated WT. M = vehicle-treated TKO. H = DMOG-treated WT.

Low stoichiometry oxidation might reflect enzymatic hydroxylation for low efficiency substrates or substrate competition. Additionally, hydroxylated proteoforms might be present due to escape from hydroxylation-mediated proteolysis or improper release from a hydroxylation-restrictive complex. In the SILAC experiment comparing hydroxylase restrictive conditions in the context of proteasome inhibition (Chapter 4.3.2), one of the reported non-HIFα PHD substrates, Pkm P403, was identified as a low stoichiometry (0.2 %) site of prolyl hydroxylation. The regulation of this modification appears to be consistent with PHD catalysed hydroxylation and the ambiguity surrounding correct PTM localisation has already been discussed (Chapter 4.4.1). Supposing the hydroxyproline assignment were correct, the low stoichiometry modification presents an unusual conundrum with respect to function. Moreover, the high background of the non-hydroxylated proteoform means that the function of this modification would be difficult to study at the endogenous level. This is further complicated by distinguishing any

potential Pkm HyP403 function from HIF dependent effects, which will be induced in conditions that also inhibit Pkm P403 hydroxylation. *In vitro* assays are unlikely to overcome this difficulty because recombinant hydroxylation assays did not generate PHD-catalysed prolyl hydroxylation of human PKM (Cockman *et al.*, 2019). Thus, using Pkm HyP403 as an example, low stoichiometry hydroxyproline sites are i) difficult to analyse and validate at the endogenous level, ii) difficult to distinguish from the non-hydroxylated proteoform in functional studies, and iii) difficult to distinguish from other targets of the same 2OGDs in functional studies. Thus, it is intractable to study the function of low stoichiometry prolyl hydroxylation at the endogenous level in cells since neither the modification nor the function can be directly attributed to enzymatic hydroxylation at the reported target site.

Throughout this thesis, a filter (stoichiometry ≥ 5 %) has been applied to remove low stoichiometry modifications. Although this filter does not distinguish between enzymatic and non-enzymatic oxidation, non-enzymatic oxidation of proline is expected to be an inefficient and therefore low stoichiometry PTM (Liu et al., 2020b). Therefore, the rationale for applying this filter is to remove sites of hydroxylation that could be confused with artefactual oxidation derived from reactive oxygen species (ROS). Further experiments might remove this filter and implement an alternative strategy to reduce false positives in the dataset. One example would be to use $^{18}O_2$ and $H_2{}^{18}O_2$ treatments to distinguish between enzymatic and ROS-induced oxidation (Hawkins and Davies, 2019; Hewitson *et al.*, 2002), however such experiments are likely to be expensive.

7.5 Conclusions and future perspectives

To conclude, the data presented in this thesis provides unprecedented insight into the extent of prolyl hydroxylation in the proteome. Qualitative and quantitative analyses revealed that many hydroxyproline assignments generated by proteomic software were likely to be incorrect. This implied that the consideration of additional factors, including accurate localisation and demonstration of enzyme-, 2OG- or oxygen-dependence is required for assignment of 2OGD-catalysed prolyl hydroxylation. With this in mind, 49 novel hydroxyproline sites were assigned at high confidence using a range of orthogonal proteomics approaches. Significantly, these approaches did not provide any supporting evidence to corroborate the reported assignments of 45 non-HIFα PHD substrates despite

reliable detection of the unmodified target peptides across different datasets, which suggests PHD substrates have been overreported in the literature. An additional difficulty in hydroxyproline discovery is the sensitivity of mass spectrometry for interrogating low abundance proteins in a complex cell extract. Development of GST-VBC as a hydroxyproline-containing peptide enrichment method in this thesis should enable further hydroxyproline discoveries to be made in different cellular contexts. Functional proteomic studies, assaying protein abundance changes in response to hypoxia and reoxygenation, could lead to the discovery of hydroxylation-dependent pathways that respond to cellular oxygen levels. Together, the discoveries reported in this thesis should aid further research into the cellular biochemical process that impact on cellular oxygen homeostasis.

Chapter 8. Appendix

Table 62 | Summary of non-collagen candidates identified in this thesis.

Protein	HyP site(s)	Evidence	Reference
ACSL4	-	Significant protein decay following reoxygenation in SH-SY5Y cells	Chapter 6
ARHGAP35	P115	Peptide enrichment by GST-VBC	Chapter 5
BCKDK	-	Significant protein decay following reoxygenation in SH-SY5Y cells	Chapter 6
CAPN2	P221	Confidently assigned hydroxyproline in a 'P-HyP' sequence without MS2 validation.	Chapter 3
	P222	Confidently assigned hydroxyproline in a 'P-HyP' sequence without MS2 validation.	Chapter 3
CASC4	P252	Confidently assigned hydroxyproline in a 'HyP-G' sequence with MS2 validation. Published literature.	Chapter 3; Onisko (2020)
CDKN1C	-	Significant protein decay following reoxygenation in SH-SY5Y cells	Chapter 6
CTHRC1	P59 P63 P75 P78 P81	Protein enrichment is suppressed by DMOG Peptide enrichment by GST-VBC Prolyl hydroxylation occurs in the 'Yaa' position of collagen-like repeating 'Xaa-Yaa-Gly' sequences	Chapter 5
DOCK1	P1847	Confidently assigned hydroxyproline in a 'P-HyP' sequence without MS2 validation.	Chapter 3
DYNC1LI1	P180	Confidently assigned hydroxyproline in a 'P-HyP' sequence without MS2 validation.	Chapter 3
FKBP10	P36	Confidently assigned hydroxyproline with MS2 validation. Confidently assigned hydroxyproline in U-87 MG cells with suppression by 0.1 % O_2. Confidently assigned hydroxyproline in MEF cells with suppression by 1 mM DMOG. Protein enrichment by VBC is suppressed by DMOG. Peptide enrichment by VBC.	Chapters 3, 4 and 5; Arsenault *et al.* (2015); Onisko (2020); Zhou *et al.* (2016)

		Published literature.	
FLNB	P1429	Confidently assigned hydroxyproline in a 'HyP-G' sequence without MS2 validation.	Chapter 3
GOLM1	-	Significant protein decay following reoxygenation in SH-SY5Y cells	Chapter 6
HYOU1	P977	Confidently assigned hydroxyproline in a 'HyP-G' sequence with MS2 validation. Confidently assigned hydroxyproline in RCC4 cells but not enriched by GST-VBC. Published literature.	Chapter 3; Onisko (2020)
KDM5B	-	Significant protein decay following reoxygenation in SH-SY5Y cells Oxygen dependent proteolysis observed by western blot in cycloheximide experiments	Chapter 6
KHDRBS1	P297 P299	Peptide enrichment by GST-VBC Prolyl hydroxylation occurs in 'P-HyP' sequences	Chapter 5
LMAN1	P37	Confidently assigned hydroxyproline in U-87 MG cells with suppression by 0.1 % O_2.	Chapter 4
	HyP378	Confidently assigned hydroxyproline in a 'HyP-G' sequence with MS2 validation. Confidently assigned hydroxyproline in U-87 MG cells with suppression by 0.1 % O_2. Confidently assigned hydroxyproline in MEF cells with suppression by 1 mM DMOG. Published literature.	Chapter 3, 4 Onisko (2020)
MESD	P38	Confidently assigned hydroxyproline in U-87 MG cells with suppression by 0.1 % O_2.	Chapter 4
MMP2	-	Significant protein decay following reoxygenation in SH-SY5Y cells	Chapter 6
MRC2	-	Significant protein decay following reoxygenation in SH-SY5Y cells	Chapter 6
MUC5B	P2516	Confidently assigned hydroxyproline in a 'HyP-G' sequence with MS2 validation.	Chapter 3.
MYO10	P255	Confidently assigned hydroxyproline in a 'HyP-G' sequence without MS2 validation.	Chapter 3

NDUFA4L2	-	Significant protein decay following reoxygenation in SH-SY5Y cells	Chapter 6
OS9	P654	Confidently assigned hydroxyproline in a 'HyP-G' sequence without MS2 validation. Confidently assigned hydroxyproline in MEF cells with suppression by 1 mM DMOG.	Chapter 3, 4
P4HB	P51	Frequently assigned hydroxyproline in PDI consensus hydroxylation site. Confidently assigned hydroxyproline in MEFs but insensitive to DMOG Published literature	Chapter 3, 4 Onisko (2020); Stoehr *et al.* (2016)
	P395	Frequently assigned hydroxyproline in PDI consensus hydroxylation site. Confidently assigned hydroxyproline in MEFs but insensitive to DMOG. Prolyl hydroxylation is suppressed by hypoxia and DMOG in eHAP cells. Published literature	Chapter 3, 4 Henningsen *et al.* (2010); Onisko (2020); Stoehr *et al.* (2016)
PARP2	P319	Confidently assigned hydroxyproline in a 'P-HyP' sequence without MS2 validation.	Chapter 3
PCOLCE	-	Significant protein decay following reoxygenation in SH-SY5Y cells	Chapter 6
PDIA3	P55	Frequently assigned hydroxyproline in PDI consensus hydroxylation site Confidently assigned hydroxyproline in MEFs but insensitive to DMOG Published literature	Chapter 3, 4 Stoehr *et al.* (2016)
	P404*	Confidently assigned hydroxyproline with MS2 validation. Frequently assigned hydroxyproline in a PDI hydroxylation consensus sequence. Confidently assigned hydroxyproline in MEFs but insensitive to DMOG Prolyl hydroxylation is suppressed by hypoxia and DMOG in eHAP cells Published literature	Chapter 3, 4 Onisko (2020); Stoehr *et al.* (2016)

PDIA4	P89	Frequently assigned hydroxyproline in PDI consensus hydroxylation site Prolyl hydroxylation is suppressed by hypoxia and DMOG in eHAP cells Published literature	Chapter 3, 4 Onisko (2020)
	P204	Frequently assigned hydroxyproline in PDI consensus hydroxylation site Published literature	Chapter 3 Onisko (2020); Stoehr *et al.* (2016)
	P553*	Confidently assigned hydroxyproline with MS2 validation. Frequently assigned hydroxyproline in a PDI hydroxylation consensus sequence. Confidently assigned hydroxyproline in MEFs but insensitive to DMOG Prolyl hydroxylation is suppressed by hypoxia and DMOG in eHAP cells Published literature	Chapter 3, 4 Onisko (2020); Stoehr *et al.* (2016)
PDIA5	P303	Frequently assigned hydroxyproline in PDI consensus hydroxylation site	Chapter 3
	P424	Frequently assigned hydroxyproline in PDI consensus hydroxylation site	Chapter 3
PDIA6	P53	Frequently assigned hydroxyproline in PDI consensus hydroxylation site Published literature	Chapter 3 Onisko (2020)
	P188	Frequently assigned hydroxyproline in PDI consensus hydroxylation site	Chapter 3
PFKFB4	-	Significant protein decay following reoxygenation in SH-SY5Y cells	Chapter 6
PFKP	-	Significant protein decay following reoxygenation in SH-SY5Y cells	Chapter 6
PNRC1	P20	Peptide enrichment by GST-VBC	Chapter 5
PRKCSH	P290	Confidently assigned hydroxyproline with MS2 validation. Peptide enrichment by VBC.	Chapters 3, 5.
RAB20	-	Significant protein decay following reoxygenation in SH-SY5Y cells	Chapter 6

SCYL2	P442	Frequently assigned hydroxyproline in a 'P-HyP' sequence without MS2 validation.	Chapter 3
SERPINH1	P30	Confidently assigned hydroxyproline in a 'HyP-G' sequence with MS2 validation. Confidently assigned hydroxyproline in U-87 MG cells with suppression by 0.1 % O_2. Confidently assigned hydroxyproline in MEF cells with suppression by 1 mM DMOG. Published literature	Chapter 3, 4 Henningsen *et al.* (2010); Onisko (2020); Stoehr *et al.* (2016); Zhou *et al.* (2016)
SIPA1L2	P1070 P1072	Protein enrichment by GST-VBC is suppressed by DMOG Peptide enrichment by GST-VBC	Chapter 5
SLC38A10	P532	Confidently assigned hydroxyproline in a 'HyP-G' sequence with MS2 validation.	Chapter 3
SUN2	P315	Confidently assigned hydroxyproline in a 'HyP-G' sequence with MS2 validation.	Chapter 3
TMEM109	P42	Confidently assigned hydroxyproline in a 'Hyp-G' and 'P-HyP' sequence with MS2 validation.	Chapter 3
TMX3	P51	Frequently assigned hydroxyproline in PDI consensus hydroxylation site	Chapter 3
TMX4	P62	Frequently assigned hydroxyproline in PDI consensus hydroxylation site	Chapter 3
TXNDC5	P87	Frequently assigned hydroxyproline in PDI consensus hydroxylation site Confidently assigned hydroxyproline in MEFs but insensitive to DMOG	Chapter 3, 4
	P215	Frequently assigned hydroxyproline in PDI consensus hydroxylation site	Chapter 3
	P348	Frequently assigned hydroxyproline in PDI consensus hydroxylation site	Chapter 3
VCAN	P2675	Peptide enrichment by GST-VBC	Chapter 5
WIPF2	P328	Peptide enrichment by GST-VBC	Chapter 5
ZNF395	-	Significant protein decay following reoxygenation in SH-SY5Y cells	Chapter 6

Table 63 | Summary of the data relating to the reported non-HIFα PHD substrates generated in this thesis.

Protein	Reported site	Evidence
ACACB	P343 P450 P2131	Non-oxidised peptides for each target site were identified in publicly deposited datasets (chapter 3). ACACB protein abundance decreases 20 % and 55 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). ACACB is present in the VBC-bound fraction of the protein enrichment assay. This interaction increases 11-fold after DMOG treatment (chapter 5).
ACTB	P70 P307 P322	Oxidised peptides identified for each target site. Oxidation of non-prolyl residues accounts for >99% of oxidised PSMs for these peptides (chapter 3). ACTB protein abundance decreased 6 % and 23 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). ACTB protein abundance decreases 35 % and 20 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4). ACTB protein abundance increased 9 %, 102 % and 51 % at 1 % O_2 (36 hours), 1 % O_2 (72 hours), and with DMOG-treated cells, respectively, in the eHAP experiment (chapter 4). Target peptides were identified in non-oxidised forms and with oxidation at non-prolyl residues in the eHAP experiment (chapter 4). ACTB is present in the VBC-bound fraction of the protein enrichment assay and the abundance does not change after DMOG treatment (chapter 5). HyP322 observed in VBC enrichment. Peptides containing oxidation of non-prolyl residues are more abundant than the HyP322 peptide. Manual inspection of the MS2 spectra suggests methionine oxidation according to -64 Da neutral loss ions (chapter 5).
ADRB2	P382 P395	Non oxidised peptide detected in publicly deposited dataset (chapter 3). Protein is stable after 1 hour of reoxygenation (chapter 6).
ADSL	P24	Peptides were identified with confident localisation of oxidation to M26 but not P24 in publicly deposited datasets (chapter 3). ADSL protein abundance decreased 11 % and 26 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4).

		ADSL protein abundance decreased 38 % and 8 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4). ADSL protein abundance decreased 44 %, 5 %, and 46 % at 1 % O_2 (36 hours), 1 % O_2 (72 hours), and with DMOG-treated cells, respectively, in the eHAP experiment (chapter 4). The target peptide was identified in its non-oxidised form in the eHAP experiment (chapter 4). Protein is stable after 1 hour of reoxygenation, and peptides containing the target site were identified with oxidation at M26 (chapter 6).
AKT1	P125 P313 P318 P423	Non-oxidised peptides for each target site were identified in publicly deposited datasets (chapter 3). AKT1 protein abundance increased 7 % and decreased 35 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). AKT1 protein abundance decreases 15 % and 10% in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4). The peptide containing P125 was identified in its non-oxidised form in the U87MG hypoxia SILAC experiment (chapter 4). Protein is stable after 1 hour of reoxygenation (chapter 6).
ARRB2	P176 P179 P181	Non-oxidised peptides for each target site were identified in publicly deposited datasets (chapter 3). ARRB2 protein abundance decreased 14 % and 43 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). ARRB2 protein abundance decreased 50 % and 10 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4). Protein is stable after 1 hour of reoxygenation (chapter 6).
ATF4	P156 P162 P164 P167 P174	Peptides containing the target site were not detected in any of the experiments performed in this thesis. ATF4 protein abundance decreased 45 % and 22 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (Chapter 4).
BCL2L11	P67 P70	Peptides containing the target sites were not identified in the publicly deposited data (chapter 3).
BRD4	P536	In the publicly deposited data, oxidation at K535 and K537 was confidently localised, but oxidation at P536 was not (chapter 3).

		BRD4 protein abundance decreased 6 % and 36 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). The non-oxidised target peptide was present in the U87MG hypoxia SILAC experiment (chapter 4). BRD4 protein abundance did not change in the MEF SILAC experiment (chapter 4). BRD4 protein abundance increased 505 %, 112 %, and 161 % at at 1 % O_2 (36 hours), 1 % O_2 (72 hours), and with DMOG-treated cells, respectively, in the eHAP experiment (chapter 4). The target peptide was observed in its non-oxidised form in the eHAP experiment (chapter 4). BRD4 is stable after 1 hour of reoxygenation. The HyP536 peptide was detected but the discriminatory y2 fragment ion was not detected in the MS2 spectrum (chapter 6).
CENPN	P311	Non-oxidised peptides were identified in publicly deposited datasets (chapter 3). The CENPN protein was only identified in TKO cells of the MEF SILAC experiment (chapter 4). CENPN was stable after 1 hour of reoxygenation (chapter 6).
CEP192	P2313	CEP192 was stable after 1 hour of reoxygenation (chapter 6).
CERKL	ND	The target sites were not defined. Therefore, prolyl hydroxylation of CERKL was not analysed.
CPT1B	P295	Peptides containing the target site were not detected in any of the experiments performed in this thesis.
DGKI	P903	A non-oxidised peptide was identified in the publicly deposited datasets (chapter 3). DGKI was stable after 1 hour of reoxygenation (chapter 6).
DYRK1A	P380	Non-oxidised peptides were identified in publicly deposited datasets (chapter 3). DYRK1A protein abundance did not change in the MEF SILAC experiment (chapter 4). DYRK1A was stable after 1 hour of reoxygenation. A peptide containing the target site was identified in its non-oxidised form (chapter 6).
DYRK1B	P332	Peptides containing the target site were not identified in the publicly deposited data (chapter 3).

		DYRK1B abundance decreased by 10 % following 1 hour of reoxygenation. This was not determined to be significant. Peptides containing the target site were identified in the non-oxidised form (chapter 6).
EEF2	ND	The target sites were not defined. Therefore, prolyl hydroxylation of EEF2 was not analysed. EEF2 protein abundance decreased 6 % and 27 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). EEF2 protein abundance decreased by 27 % in both hydroxylase restrictive conditions of the MEF SILAC experiment (chapter 4). EEF2 protein abundance decreased 39 %, increased 18 %, and decreased 31 %% at 1 % O_2 (36 hours), 1 % O_2 (72 hours), and with DMOG-treated cells, respectively, in the eHAP experiment (chapter 4). EEF2 is present in the VBC-bound fraction of the protein enrichment assay and the abundance does not change after DMOG treatment (chapter 5). EEF2 was stable after 1 hour of reoxygenation (chapter 6).
EEF2K	P98	Peptides were detected with oxidation at non-prolyl residues in the publicly deposited data (chapter 3). EEF2K protein abundance increased 40 % and decreased 11 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). EEF2K protein abundance decreased by 47 % in TKO cells. The abundance did not change in DMOG-treated cells of the MEF SILAC experiment (chapter 4). EEF2K was stable after 1 hour of reoxygenation (chapter 6).
EHMT2	P676 P1207	Non-oxidised peptides were identified in publicly deposited datasets (chapter 3).
EPOR	P443 P450	Peptides containing the target site were not detected in any of the experiments performed in this thesis.
FLNA	P2317 P2324	Peptides containing oxidation at the target sites accounted for < 99.5 % of the target peptides detected in the publicly deposited data. The oxidation was not confidently localised (chapter 3). FLNA protein abundance did not change in the U87MG hypoxia SILAC experiment (chapter 4).

		Several non-oxidised peptides containing the target sites were identified in the U87MG hypoxia SILAC experiment (chapter 4). FLNA protein abundance decreased 10 % and 11 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4). Several non-oxidised peptides containing the target sites were identified in the MEF SILAC experiment (chapter 4). FLNA protein abundance increased 69 %, 209 %, and 95 %% at 1 % O_2 (36 hours), 1 % O_2 (72 hours), and with DMOG-treated cells, respectively, in the eHAP experiment (chapter 4). The target peptide was observed in its non-oxidised form in the eHAP experiment (chapter 4). FLNA is present in the VBC-bound fraction of the protein enrichment assay and the abundance does not change after DMOG treatment (chapter 5). The non-oxidised target peptide was identified in the VBC-bound fraction after peptide enrichment (Chapter 5). FLNA was stable after 1 hour of reoxygenation. A non-oxidised peptide containing the target sites was detected (chapter 6).
FOXO3	P426 P427	Non-oxidised peptides were identified in publicly deposited datasets (chapter 3). FOXO3 protein abundance decreased 24 % and 87 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). FOXO3 protein abundance increased 29 % and 35 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4). A non-oxidised peptide containing the target sites was detected in the MEF SILAC experiment (chapter 4).
IKBKB	P191	Peptides containing the target site were not detected in any of the experiments performed in this thesis. IKBKB protein abundance decreased 17 % and 30 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). IKBKB protein abundance did not change in the MEF SILAC experiment (chapter 4). IKBKB was stable after 1 hour of reoxygenation (chapter 6).
MAPK6	P25	Peptides containing the target site were not detected in any of the experiments performed in this thesis.

		MAPK6 protein abundance increased 38 % and decreased 28 % in TKo and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4).
MAPK7	ND	The target sites were not defined. Therefore, prolyl hydroxylation of MAPK7 was not analysed. MAPK7 protein abundance increased 78 % and 33 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). MAPK7 protein abundance decreased 68 % and 19 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4). MAPK7 was stable after 1 hour of reoxygenation (chapter 6).
MLXPL	P15 P141 P526	Non-oxidised peptides containing P141 were detected in the publicly deposited datasets (chapter 3). Peptides containing P15 and P526 were not detected in any of the experiments performed in this thesis.
NDRG3	P294	Target peptides containing oxidation at non-prolyl residues were identified in the publicly deposited datasets (chapter 3). NDRG3 protein abundance decreased 14 % and 32 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). NDRG3 protein abundance decreased 61 % and 7 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4). The NDRG3 protein was only detected in the normoxic sample of the eHAP experiment (chapter 4). NDRG3 was stable after 1 hour of reoxygenation. The target peptide was identified with methionine oxidation (chapter 6).
OTUB1	P210 P263	Peptides were detected with oxidation at non-prolyl residues in the publicly deposited data. Oxidation of M211 was assigned with high confidence (chapter 3). OTUB1 protein abundance was stable at 1 % O_2 and decreased 15 % at 0.1 % O_2 in the U87MG hypoxia SILAC experiment (chapter 4). A non-oxidised peptide containing P263 was identified in the U87MG hypoxia SILAC experiment (chapter 4). OTUB1 protein abundance increased 7 % and decreased 11 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4).

		Non-oxidised peptides corresponding to both target sites were identified in the MEF SILAC experiment (chapter 4). OTUB1 protein abundance decreased 18 %, increased 100 %, and increased 96 %% at 1 % O_2 (36 hours), 1 % O_2 (72 hours), and with DMOG-treated cells, respectively, in the eHAP experiment (chapter 4). The target peptide corresponding to P263 was identified in its non-oxidised form in the eHAP experiment (chapter 4). OTUB1 was stable after 1 hour of reoxygenation. The peptide containing P210 was identified with methionine oxidation. (chapter 6).
PAX2	ND	The target sites were not defined. Therefore, prolyl hydroxylation of PAX2 was not analysed.
PDE4D	P29 P382 P419	Target peptides containing oxidation at M371 were assigned with high confidence in the publicly deposited datasets. P29 target peptides were not detected and non-oxidised P419 peptides were detected (chapter 3). PDE4D abundance decreased by 10 % after 1 hour of reoxygenation. This was not significant (chapter 6).
PKM	P403 P408	Peptides containing oxidation at the target sites were identified in the publicly deposited datasets, but the assignments were low confidence (chapter 3). PKM protein abundance did not change in the U87MG hypoxia SILAC experiment (chapter 4). PKM protein abundance decreased 14 % and increased 22 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4). HyP403 was detected at 0.2 % stoichiometry in the MEF SILAC experiment. This was completely suppressed in the TKO cells and DMOG-treated cells (chapter 4). PKM protein abundance increased 121 %, 334 %, and 374 %% at 1 % O_2 (36 hours), 1 % O_2 (72 hours), and with DMOG-treated cells, respectively, in the eHAP experiment (chapter 4). The target peptide was identified in its non-oxidised form in the eHAP experiment (chapter 4). PKM is present in the VBC-bound fraction of the protein enrichment assay and the abundance does not change after DMOG treatment (chapter 5).

		The HyP403 peptide was observed in the VBC-bound fraction after peptide enrichment at a stoichiometry of 1.4 % (i.e., the non-oxidised peptide was present at significantly higher abundance) (chapter 5). PKM was stable after 1 hour of reoxygenation. The non-oxidised target peptide was observed (chapter 6).
POLR2A	P1465	Target peptides were identified with oxidation at non-prolyl residues (chapter 3). POLR2A protein abundance did not change at 1 % O_2 and decreased 37 % at 0.1 % O_2 in the U87MG hypoxia SILAC experiment (chapter 4). POLR2A was stable in TKO cells and decreased 8 % in DMOG-treated cells in the MEF SILAC experiment (chapter 4). POLR2A protein abundance increased 78 %, 1175 %, and 69 %% at 1 % O_2 (36 hours), 1 % O_2 (72 hours), and with DMOG-treated cells, respectively, in the eHAP experiment (chapter 4). POLR2A was stable after 1 hour of reoxygenation (chapter 6).
PPP2R2A	P319	Target peptides were identified with confidently localised oxidation at W311 and M315 in publicly deposited datasets (chapter 3). PPP2R2A protein abundance increased 8 % and decreased 17 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). PPP2R2A protein abundance decreased 17 % and 7 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4). PPP2R2A protein abundance decreased 68 %, 37 %, and 26 % at 1 % O_2 (36 hours), 1 % O_2 (72 hours), and with DMOG-treated cells, respectively, in the eHAP experiment (chapter 4). The target peptide was identified in its non-oxidised form in the MEF SILAC experiment (chapter 4). PPP2R2A was stable after 1 hour of reoxygenation (chapter 6).
SFMBT1	P106 P651	The target peptides were identified in the non-oxidised forms in the publicly deposited datasets (chapter 3). The abundance of SFMBT1 increases by 33 % after 1 hour of reoxygenation (chapter 6).
SPRY2	P18 P144 P160	The target peptides were identified in the non-oxidised forms in the publicly deposited datasets (chapter 3).

		SPRY2 protein abundance decreased 44 % and 83 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). SPRY2 was stable after 1 hour of reoxygenation. The target peptides corresponding to P18 and P144 were identified in their non-oxidised forms (chapter 6).
TBK1	P48	Non-oxidised peptides were identified in publicly deposited datasets (chapter 3). Protein abundance decreased 2 % and 21 % at 1 % and 0.1 % O_2, respectively, in the U87MG SILAC hypoxia experiment (chapter 4).
TELO2	P374 P419 P422	The target peptides were identified in the non-oxidised forms in the publicly deposited datasets (chapter 3). TELO2 protein abundance decreased 9 % and 21 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). TELO2 protein abundance decreases 25 % and 24 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4). The abundance of TELO2 increases by 10 % after 1 hour of reoxygenation. The target peptides were identified in their non-oxidised forms (chapter 6).
TET2	P1335 P1342	The target peptides were identified in the non-oxidised forms in the publicly deposited datasets (chapter 3). TET2 was stable after 1 hour of reoxygenation (chapter 6).
TET3	P1030 P1037	Peptides containing the target site were not detected in any of the experiments performed in this thesis. TET3 protein abundance increased 721 % and decreased 36 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4). ACTB is present in the VBC-bound fraction of the protein enrichment assay. The abundance increases 11.8-fold after DMOG treatment (chapter 5).
THRA	P160 P162	The target peptides were identified in the non-oxidised forms in the publicly deposited datasets (chapter 3).
TP53	P142 P359	The target peptides were identified in the non-oxidised forms in the publicly deposited datasets (chapter 3). TP53 protein abundance decreased 50 % and 15 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (chapter 4).

		The peptide corresponding to P142 was identified with cysteine oxidation in the MEF SILAC experiment (chapter 4). TP53 was stable after 1 hour of reoxygenation (chapter 6).
TRPA1	P394	Peptides containing the target site were not detected in any of the experiments performed in this thesis.
UBE2N	P19 P21 P59	The target peptides were identified in their non-oxidised forms in the publicly deposited datasets. Oxidation was assigned to P59 for < 1 % of the peptides containing this target site and did not pass the localisation filters. Oxidation at non-prolyl residues account for > 50 % of the target peptides (chapter 3). UBE2N protein abundance decreased 7 % and 21 % at 1% and 0.1 % O_2, respectively, in the U87MG hypoxia SILAC experiment (Chapter 4). UBE2N protein abundance did not change in the MEF SILAC experiment (Chapter 4). Target peptides were identified in their non-oxidised forms in the MEF SILAC chapter (chapter 4). UBE2N protein abundance decreased 18 %, increased 56 %, and increased 23 % at 1 % O_2 (36 hours), 1 % O_2 (72 hours), and with DMOG-treated cells, respectively, in the eHAP experiment (chapter 4). The target peptides were identified in their non-oxidised forms in the eHAP experiment (chapter 4). Target peptides corresponding to P59 and HyP59 were identified in the VBC bound fraction of the peptide enrichment assays. The stoichiometry of the HyP59 peptide was 8 % (chapter 5). UBE2N was stable after 1 hour of reoxygenation. The HyP59 peptide was identified at 1 % stoichiometry, which did not increase following reoxygenation. Methionine oxidation was also observed on the target peptide (chapter 6).
UBE2V1	P79 P80	Oxidation was observed at non-prolyl residues of the target peptide in the publicly deposited datasets (chapter 3). UBE2V1 protein abundance decreased 29 % and 8 % in TKO and DMOG-treated cells in the MEF SILAC experiment (chapter 4). UBE2V1 protein abundance decreased 1 %, increased 91 %, and increased 67 % at 1 % O_2 (36 hours), 1 % O_2 (72 hours), and with DMOG-treated cells, respectively, in the eHAP experiment (chapter 4)

	P427	The target peptides were identified in the non-oxidised forms in the publicly deposited datasets (chapter 3).
ZHX2	P440 P464	ZHX2 protein abundance increased 240 % and decreased 16 % in TKO and DMOG-treated cells, respectively, in the MEF SILAC experiment (Chapter 4).

Reference List

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature *422*, 198-207. 10.1038/nature01511.

Alcaide-German, M.L., Vara-Vega, A., Garcia-Fernandez, L.F., Landazuri, M.O., and del Peso, L. (2008). A yeast three-hybrid system that reconstitutes mammalian hypoxia inducible factor regulatory machinery. BMC Cell Biol *9*, 18. 10.1186/1471-2121-9-18.

Anderson, K., Nordquist, K.A., Gao, X., Hicks, K.C., Zhai, B., Gygi, S.P., and Patel, T.B. (2011). Regulation of Cellular Levels of Sprouty2 Protein by Prolyl Hydroxylase Domain and von Hippel-Lindau Proteins. Journal of Biological Chemistry *286*, 42027-42036. 10.1074/jbc.M111.303222.

Angelini, A., Saha, P.K., Jain, A., Jung, S.Y., Mynatt, R.L., Pi, X., and Xie, L. (2021). PHDs/CPT1B/VDAC1 axis regulates long-chain fatty acid oxidation in cardiomyocytes. Cell reports *37*, 109767. 10.1016/j.celrep.2021.109767.

Arias-González, L., Moreno-Gimeno, I., del Campo, A.R., Serrano-Oviedo, L., Valero, M.L., Esparís-Ogando, A., de la Cruz-Morcillo, M., Melgar-Rojas, P., García-Cano, J., Cimas, F.J., et al. (2013). ERK5/BMK1 is a novel target of the tumor suppressor VHL: implication in clear cell renal carcinoma. Neoplasia *15*, 649-659. 10.1593/neo.121896.

Arsenault, P.R., Heaton-Johnson, K.J., Li, L.-S., Song, D., Ferreira, V.S., Patel, N., Master, S.R., and Lee, F.S. (2015). Identification of prolyl hydroxylation modifications in mammalian cell proteins. Proteomics *15*, 1259-1267. 10.1002/pmic.201400398.

Aviner, R., Li, K.H., Frydman, J., and Andino, R. (2021). Cotranslational prolyl hydroxylation is essential for flavivirus biogenesis. Nature *596*, 558-564. 10.1038/s41586-021-03851-2.

Baek, J.H., Mahon, P.C., Oh, J., Kelly, B., Krishnamachary, B., Pearson, M., Chan, D.A., Giaccia, A.J., and Semenza, G.L. (2005). OS-9 interacts with hypoxia-inducible factor 1alpha and prolyl hydroxylases to promote oxygen-dependent degradation of HIF-1alpha. Mol Cell *17*, 503-512. 10.1016/j.molcel.2005.01.011.

Bagert, J.D., Xie, Y.J., Sweredoski, M.J., Qi, Y., Hess, S., Schuman, E.M., and Tirrell, D.A. (2014). Quantitative, time-resolved proteomic analysis by combining bioorthogonal noncanonical amino acid tagging and pulsed stable isotope labeling by amino acids in cell culture. Molecular & cellular proteomics : MCP *13*, 1352-1358. 10.1074/mcp.M113.031914.

Barton, S.J., Richardson, S., Perkins, D.N., Bellahn, I., Bryant, T.N., and Whittaker, J.C. (2007). Using statistical models to identify factors that have a role in defining the abundance of ions produced by tandem MS. Anal Chem *79*, 5601-5607. 10.1021/ac0700272.

Barton, S.J., and Whittaker, J.C. (2009). Review of factors that influence the abundance of ions produced in a tandem mass spectrometer and statistical methods for discovering these factors. Mass Spectrometry Reviews *28*, 177-187. https://doi.org/10.1002/mas.20188.

Batie, M., Frost, J., Frost, M., Wilson, J.W., Schofield, P., and Rocha, S. (2019). Hypoxia induces rapid changes to histone methylation and reprograms chromatin. Science *363*, 1222-1226. 10.1126/science.aau5870.

Beausoleil, S.A., Villén, J., Gerber, S.A., Rush, J., and Gygi, S.P. (2006). A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat. Biotechnol. *24*, 1285-1292. 10.1038/nbt1240.

Bekker-Jensen, D.B., Bernhardt, O.M., Hogrebe, A., Martinez-Val, A., Verbeke, L., Gandhi, T., Kelstrup, C.D., Reiter, L., and Olsen, J.V. (2020). Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. Nature Communications *11*, 787. 10.1038/s41467-020-14609-1.

Bekker-Jensen, D.B., Kelstrup, C.D., Batth, T.S., Larsen, S.C., Haldrup, C., Bramsen, J.B., Sorensen, K.D., Hoyer, S., Orntoft, T.F., Andersen, C.L., et al. (2017). An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. Cell Syst *4*, 587-599.e584. 10.1016/j.cels.2017.05.009.

Bex, C., Knauth, K., Dambacher, S., and Buchberger, A. (2007). A yeast two-hybrid system reconstituting substrate recognition of the von Hippel-Lindau tumor suppressor protein. Nucleic Acids Res *35*, e142. 10.1093/nar/gkm932.

Bhatnagar, R., Rapaka, R., and Urry, D. (1978). Interaction of polypeptide models of elastin with prolyl hydroxylase. FEBS letters *95*, 61-64.

Bishop, T., and Ratcliffe, P.J. (2015). HIF hydroxylase pathways in cardiovascular physiology and medicine. Circ Res *117*, 65-79. 10.1161/circresaha.117.305109.

Boys, B.L., Kuprowski, M.C., Noël, J.J., and Konermann, L. (2009). Protein oxidative modifications during electrospray ionization: solution phase electrochemistry or corona discharge-induced radical attack? Anal Chem *81*, 4027-4034. 10.1021/ac900243p.

Breci, L.A., Tabb, D.L., Yates, J.R., 3rd, and Wysocki, V.H. (2003). Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. Anal Chem *75*, 1963-1971. 10.1021/ac026359i.

Brockmeier, U., Platzek, C., Schneider, K., Patak, P., Bernardini, A., Fandrey, J., and Metzen, E. (2011). The function of hypoxia-inducible factor (HIF) is independent of the endoplasmic reticulum protein OS-9. PLoS One *6*, e19151. 10.1371/journal.pone.0019151.

Brown, J.C., Golbik, R., Mann, K., and Timpl, R. (1994). Structure and stability of the triple-helical domains of human collagen XIV. Matrix Biol *14*, 287-295. 10.1016/0945-053x(94)90194-5.

Bruick, R.K., and McKnight, S.L. (2001). A conserved family of prolyl-4-hydroxylases that modify HIF. Science *294*, 1337-1340. 10.1126/science.1066373.

Bueno, M.T., and Richard, S. (2013). SUMOylation negatively modulates target gene occupancy of the KDM5B, a histone lysine demethylase. Epigenetics *8*, 1162-1175. 10.4161/epi.26112.

Casciello, F., Al-Ejeh, F., Kelly, G., Brennan, D.J., Ngiow, S.F., Young, A., Stoll, T., Windloch, K., Hill, M.M., Smyth, M.J., et al. (2017). G9a drives hypoxia-mediated gene repression for breast cancer cell survival and tumorigenesis. Proceedings of the National Academy of Sciences *114*, 7077-7082. doi:10.1073/pnas.1618706114.

Černý, M., Skalák, J., Cerna, H., and Brzobohatý, B. (2013). Advances in purification and separation of posttranslationally modified proteins. J Proteomics *92*, 2-27. 10.1016/j.jprot.2013.05.040.

Chauvistré, H., Daignault, S.M., Shannan, B., Ju, R.J., Picard, D., Vogel, F.C.E., Egetemaier, S., Krepler, C., Rebecca, V.W., Sechi, A., et al. (2020). The Janus-faced role of KDM5B heterogeneity in melanoma: differentiation as a situational driver of both growth arrest and drug-resistance. bioRxiv, 2020.2004.2001.999847. 10.1101/2020.04.01.999847.

Chen, J., Liu, F., Li, H., Archacki, S., Gao, M., Liu, Y., Liao, S., Huang, M., Wang, J., Yu, S., et al. (2015). pVHL interacts with Ceramide kinase like (CERKL) protein and

ubiquitinates it for oxygen dependent proteasomal degradation. Cell Signal *27*, 2314-2323. 10.1016/j.cellsig.2015.08.011.
Cheng, A., Grant, C.E., Noble, W.S., and Bailey, T.L. (2019). MoMo: discovery of statistically significant post-translational modification motifs. Bioinformatics *35*, 2774-2782. 10.1093/bioinformatics/bty1058.
Chuh, K.N., and Pratt, M.R. (2015). Chemical methods for the proteome-wide identification of posttranslationally modified proteins. Curr Opin Chem Biol *24*, 27-37. 10.1016/j.cbpa.2014.10.020.
Clasen, S.J., Shao, W., Gu, H., and Espenshade, P.J. (2017). Prolyl dihydroxylation of unassembled uS12/Rps23 regulates fungal hypoxic adaptation. eLife *6*, e28563. 10.7554/eLife.28563.
Clifford, S.C., Cockman, M.E., Smallwood, A.C., Mole, D.R., Woodward, E.R., Maxwell, P.H., Ratcliffe, P.J., and Maher, E.R. (2001). Contrasting effects on HIF-1alpha regulation by disease-causing pVHL mutations correlate with patterns of tumourigenesis in von Hippel-Lindau disease. Hum Mol Genet *10*, 1029-1038. 10.1093/hmg/10.10.1029.
Cockman, M.E., Lancaster, D.E., Stolze, I.P., Hewitson, K.S., McDonough, M.A., Coleman, M.L., Coles, C.H., Yu, X., Hay, R.T., Ley, S.C., et al. (2006). Posttranslational hydroxylation of ankyrin repeats in IkappaB proteins by the hypoxia-inducible factor (HIF) asparaginyl hydroxylase, factor inhibiting HIF (FIH). Proc Natl Acad Sci U S A *103*, 14767-14772. 10.1073/pnas.0606877103.
Cockman, M.E., Lippl, K., Tian, Y.-M., Pegg, H.B., Figg, W.D.J., Abboud, M.I., Heilig, R., Fischer, R., Myllyharju, J., Schofield, C.J., and Ratcliffe, P.J. (2019). Lack of activity of recombinant HIF prolyl hydroxylases (PHDs) on reported non-HIF substrates. eLife *8*, e46490. 10.7554/eLife.46490.
Cockman, M.E., Masson, N., Mole, D.R., Jaakkola, P., Chang, G.W., Clifford, S.C., Maher, E.R., Pugh, C.W., Ratcliffe, P.J., and Maxwell, P.H. (2000). Hypoxia inducible factor-alpha binding and ubiquitylation by the von Hippel-Lindau tumor suppressor protein. The Journal of biological chemistry *275*, 25733-25741. 10.1074/jbc.M002740200.
Cockman, M.E., Sugimoto, Y., Pegg, H.B., Masson, N., Salah, E., Tumber, A., Flynn, H., Kirkpatrick, J.M., Schofield, C.J., and Ratcliffe, P. (2022). Widespread hydroxylation of unstructured lysine-rich protein domains by JMJD6. Proc Natl Acad Sci U S A *Manuscript in progress*.
Cockman, M.E., Webb, J.D., Kramer, H.B., Kessler, B.M., and Ratcliffe, P.J. (2009). Proteomics-based identification of novel factor inhibiting hypoxia-inducible factor (FIH) substrates indicates widespread asparaginyl hydroxylation of ankyrin repeat domain-containing proteins. Molecular & cellular proteomics : MCP *8*, 535-546. 10.1074/mcp.M800340-MCP200.
Coleman, M.L., McDonough, M.A., Hewitson, K.S., Coles, C., Mecinovic, J., Edelmann, M., Cook, K.M., Cockman, M.E., Lancaster, D.E., Kessler, B.M., et al. (2007). Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. The Journal of biological chemistry *282*, 24027-24038. 10.1074/jbc.M704102200.
Counts, D.F., Cardinale, G.J., and Udenfriend, S. (1978). Prolyl hydroxylase half reaction: peptidyl prolyl-independent decarboxylation of alpha-ketoglutarate. Proceedings of the National Academy of Sciences *75*, 2145-2149. doi:10.1073/pnas.75.5.2145.

Cox, J., and Mann, M. (2011). Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. Annual Review of Biochemistry *80*, 273-299. 10.1146/annurev-biochem-061308-093216.
Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190. 10.1101/gr.849004.
Cuijpers, S.A.G., Willemstein, E., and Vertegaal, A.C.O. (2017). Converging Small Ubiquitin-like Modifier (SUMO) and Ubiquitin Signaling: Improved Methodology Identifies Co-modified Target Proteins. Molecular & cellular proteomics : MCP *16*, 2281-2295. 10.1074/mcp.TIR117.000152.
Cummins, E.P., Berra, E., Comerford, K.M., Ginouves, A., Fitzgerald, K.T., Seeballuck, F., Godson, C., Nielsen, J.E., Moynagh, P., Pouyssegur, J., and Taylor, C.T. (2006). Prolyl hydroxylase-1 negatively regulates IkappaB kinase-beta, giving insight into hypoxia-induced NFkappaB activity. Proc Natl Acad Sci U S A *103*, 18154-18159. 10.1073/pnas.0602235103.
Davis, S., Charles, P.D., He, L., Mowlds, P., Kessler, B.M., and Fischer, R. (2017). Expanding Proteome Coverage with CHarge Ordered Parallel Ion aNalysis (CHOPIN) Combined with Broad Specificity Proteolysis. Journal of proteome research *16*, 1288-1299. 10.1021/acs.jproteome.6b00915.
Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods *17*, 41-44. 10.1038/s41592-019-0638-x.
Depping, R., Jelkmann, W., and Kosyna, F.K. (2015). Nuclear-cytoplasmatic shuttling of proteins in control of cellular oxygen sensing. J Mol Med (Berl) *93*, 599-608. 10.1007/s00109-015-1276-0.
Deutsch, E.W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J.J., Kundu, D.J., García-Seisdedos, D., Jarnuczak, A.F., Hewapathirana, S., Pullman, B.S., et al. (2019). The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. Nucleic Acids Research *48*, D1145-D1152. 10.1093/nar/gkz984.
Deutsch, E.W., Csordas, A., Sun, Z., Jarnuczak, A., Perez-Riverol, Y., Ternent, T., Campbell, D.S., Bernal-Llinares, M., Okuda, S., Kawano, S., et al. (2016). The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. Nucleic Acids Research *45*, D1100-D1106. 10.1093/nar/gkw936.
Di Conza, G., Trusso Cafarello, S., Zheng, X., Zhang, Q., and Mazzone, M. (2017). PHD2 Targeting Overcomes Breast Cancer Cell Death upon Glucose Starvation in a PP2A/B55α;-Mediated Manner. Cell reports *18*, 2836-2844. 10.1016/j.celrep.2017.02.081.
Dieterich, D.C., Link, A.J., Graumann, J., Tirrell, D.A., and Schuman, E.M. (2006). Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). Proc Natl Acad Sci U S A *103*, 9482-9487. 10.1073/pnas.0601637103.
Doerr, A. (2012). Targeting with PRM. Nat Methods *9*, 950. 10.1038/nmeth.2193.
Dong, N.-p., Zhang, L.-x., and Liang, Y.-z. (2011). A comprehensive investigation of proline fragmentation behavior in low-energy collision-induced dissociation peptide mass spectra. International Journal of Mass Spectrometry *308*, 89-97. https://doi.org/10.1016/j.ijms.2011.08.005.
Elias, J.E., Gibbons, F.D., King, O.D., Roth, F.P., and Gygi, S.P. (2004). Intensity-based protein identification by machine learning from a library of tandem mass spectra. Nat Biotechnol *22*, 214-219. 10.1038/nbt930.

Eng, J.K., Searle, B.C., Clauser, K.R., and Tabb, D.L. (2011). A face in the crowd: recognizing peptides through database search. Molecular & cellular proteomics : MCP *10*, R111.009522. 10.1074/mcp.R111.009522.
Epstein, A.C., Gleadle, J.M., McNeill, L.A., Hewitson, K.S., O'Rourke, J., Mole, D.R., Mukherji, M., Metzen, E., Wilson, M.I., Dhanda, A., et al. (2001). C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. Cell *107*, 43-54. 10.1016/s0092-8674(01)00507-4.
Erber, L.N., Luo, A., and Chen, Y. (2019). Targeted and Interactome Proteomics Revealed the Role of PHD2 in Regulating BRD4 Proline Hydroxylation. Molecular & Cellular Proteomics, mcp.RA119.001535. 10.1074/mcp.RA119.001535.
Erickson, B.K., Jedrychowski, M.P., McAlister, G.C., Everley, R.A., Kunz, R., and Gygi, S.P. (2015). Evaluating multiplexed quantitative phosphopeptide analysis on a hybrid quadrupole mass filter/linear ion trap/orbitrap mass spectrometer. Anal Chem *87*, 1241-1249. 10.1021/ac503934f.
Fan, S., Wang, J., Yu, G., Rong, F., Zhang, D., Xu, C., Du, J., Li, Z., Ouyang, G., and Xiao, W. (2020). TET is targeted for proteasomal degradation by the PHD-pVHL pathway to reduce DNA hydroxymethylation. The Journal of biological chemistry. 10.1074/jbc.RA120.014538.
Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science *246*, 64-71. 10.1126/science.2675315.
Feramisco, J.R., Kemp, B.E., and Krebs, E.G. (1979). Phosphorylation of hydroxyproline in a synthetic peptide catalyzed by cyclic AMP-dependent protein kinase. The Journal of biological chemistry *254*, 6987-6990.
Fischer, R., and Kessler, B.M. (2015). Gel-aided sample preparation (GASP)--a simplified method for gel-assisted proteomic sample generation from protein extracts and intact cells. Proteomics *15*, 1224-1229. 10.1002/pmic.201400436.
Frost, J., Galdeano, C., Soares, P., Gadd, M.S., Grzes, K.M., Ellis, L., Epemolu, O., Shimamura, S., Bantscheff, M., Grandi, P., et al. (2016). Potent and selective chemical probe of hypoxic signalling downstream of HIF-α hydroxylation via VHL inhibition. Nat Commun *7*, 13312. 10.1038/ncomms13312.
Fulzele, A., and Bennett, E.J. (2018). Ubiquitin diGLY Proteomics as an Approach to Identify and Quantify the Ubiquitin-Modified Proteome. Methods Mol Biol *1844*, 363-384. 10.1007/978-1-4939-8706-1_23.
Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012). Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Molecular & cellular proteomics : MCP *11*, M111.014050-M014111.014050. 10.1074/mcp.M111.014050.
German, N.J., Yoon, H., Yusuf, R.Z., Murphy, J.P., Finley, L.W., Laurent, G., Haas, W., Satterstrom, F.K., Guarnerio, J., Zaganjor, E., et al. (2016). PHD3 Loss in Cancer Enables Metabolic Reliance on Fatty Acid Oxidation via Deactivation of ACC2. Mol Cell *63*, 1006-1020. 10.1016/j.molcel.2016.08.014.
Giansanti, P., Tsiatsiani, L., Low, T.Y., and Heck, A.J.R. (2016). Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. Nature Protocols *11*, 993-1006. 10.1038/nprot.2016.057.
Gibbs, D.J., Lee, S.C., Isa, N.M., Gramuglia, S., Fukao, T., Bassel, G.W., Correia, C.S., Corbineau, F., Theodoulou, F.L., Bailey-Serres, J., and Holdsworth, M.J. (2011).

Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. Nature *479*, 415-418. 10.1038/nature10534.
Gilar, M., Olivova, P., Daly, A.E., and Gebler, J.C. (2005). Orthogonality of Separation in Two-Dimensional Liquid Chromatography. Analytical Chemistry *77*, 6426-6434. 10.1021/ac050923i.
Gillet, L.C., Leitner, A., and Aebersold, R. (2016). Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. Annual Review of Analytical Chemistry *9*, 449-472. 10.1146/annurev-anchem-071015-041535.
Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Molecular & cellular proteomics : MCP *11*, O111.016717. 10.1074/mcp.O111.016717.
Goldberg, M.A., Glass, G.A., Cunningham, J.M., and Bunn, H.F. (1987). The regulated expression of erythropoietin by two human hepatoma cell lines. Proc Natl Acad Sci U S A *84*, 7972-7976. 10.1073/pnas.84.22.7972.
Gorres, K.L., and Raines, R.T. (2010). Prolyl 4-hydroxylase. Crit Rev Biochem Mol Biol *45*, 106-124. 10.3109/10409231003627991.
Grosfeld, A., Stolze, I.P., Cockman, M.E., Pugh, C.W., Edelmann, M., Kessler, B., Bullock, A.N., Ratcliffe, P.J., and Masson, N. (2007). Interaction of hydroxylated collagen IV with the von hippel-lindau tumor suppressor. The Journal of biological chemistry *282*, 13264-13269. 10.1074/jbc.M611648200.
Guo, J., Chakraborty, A.A., Liu, P., Gan, W., Zheng, X., Inuzuka, H., Wang, B., Zhang, J., Zhang, L., Yuan, M., et al. (2016). pVHL suppresses kinase activity of Akt in a proline-hydroxylation-dependent manner. Science *353*, 929-932. 10.1126/science.aad5755.
Hawkins, C.L., and Davies, M.J. (2019). Detection, identification, and quantification of oxidative protein modifications. The Journal of biological chemistry *294*, 19683-19708. 10.1074/jbc.REV119.006217.
Heidenreich, S., Weber, P., Stephanowitz, H., Petricek, K.M., Schütte, T., Oster, M., Salo, A.M., Knauer, M., Goehring, I., Yang, N., et al. (2020). The glucose-sensing transcription factor ChREBP is targeted by proline hydroxylation. The Journal of biological chemistry. 10.1074/jbc.RA120.014402.
Heir, P., Srikumar, T., Bikopoulos, G., Bunda, S., Poon, B.P., Lee, J.E., Raught, B., and Ohh, M. (2016). Oxygen-dependent Regulation of Erythropoietin Receptor Turnover and Signaling. The Journal of biological chemistry *291*, 7357-7372. 10.1074/jbc.M115.694562.
Hendriks, I.A., D'Souza, R.C., Yang, B., Verlaan-de Vries, M., Mann, M., and Vertegaal, A.C. (2014). Uncovering global SUMOylation signaling networks in a site-specific manner. Nat Struct Mol Biol *21*, 927-936. 10.1038/nsmb.2890.
Hendriks, I.A., Lyon, D., Young, C., Jensen, L.J., Vertegaal, A.C., and Nielsen, M.L. (2017). Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. Nat Struct Mol Biol *24*, 325-336. 10.1038/nsmb.3366.
Hendriks, I.A., Treffers, L.W., Verlaan-de Vries, M., Olsen, J.V., and Vertegaal, A.C.O. (2015). SUMO-2 Orchestrates Chromatin Modifiers in Response to DNA Damage. Cell reports *10*, 1778-1791. 10.1016/j.celrep.2015.02.033.
Henningsen, J., Rigbolt, K.T., Blagoev, B., Pedersen, B.K., and Kratchmarova, I. (2010). Dynamics of the skeletal muscle secretome during myoblast differentiation. Molecular & cellular proteomics : MCP *9*, 2482-2496. 10.1074/mcp.M110.002113.

Hergovich, A., Lisztwan, J., Barry, R., Ballschmieter, P., and Krek, W. (2003). Regulation of microtubule stability by the von Hippel-Lindau tumour suppressor protein pVHL. Nat Cell Biol *5*, 64-70. 10.1038/ncb899.
Hermjakob, H., and Apweiler, R. (2006). The Proteomics Identifications Database (PRIDE) and the ProteomExchange Consortium: making proteomics data accessible. Expert review of proteomics *3*, 1-3. 10.1586/14789450.3.1.1.
Hewitson, K.S., McNeill, L.A., Riordan, M.V., Tian, Y.M., Bullock, A.N., Welford, R.W., Elkins, J.M., Oldham, N.J., Bhattacharya, S., Gleadle, J.M., et al. (2002). Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. The Journal of biological chemistry *277*, 26351-26355. 10.1074/jbc.C200273200.
Hirsilä, M., Koivunen, P., Günzler, V., Kivirikko, K.I., and Myllyharju, J. (2003). Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. The Journal of biological chemistry *278*, 30772-30780. 10.1074/jbc.M304982200.
Hoffman, M.A., Ohh, M., Yang, H., Klco, J.M., Ivan, M., and Kaelin, W.G., Jr. (2001). von Hippel-Lindau protein mutants linked to type 2C VHL disease preserve the ability to downregulate HIF. Hum Mol Genet *10*, 1019-1027. 10.1093/hmg/10.10.1019.
Hon, W.C., Wilson, M.I., Harlos, K., Claridge, T.D., Schofield, C.J., Pugh, C.W., Maxwell, P.H., Ratcliffe, P.J., Stuart, D.I., and Jones, E.Y. (2002). Structural basis for the recognition of hydroxyproline in HIF-1 alpha by pVHL. Nature *417*, 975-978. 10.1038/nature00767.
Hu, L., Xie, H., Liu, X., Potjewyd, F., James, L.I., Wilkerson, E.M., Herring, L.E., Xie, L., Chen, X., Cabrera, J.C., et al. (2020). TBK1 Is a Synthetic Lethal Target in Cancer with VHL Loss. Cancer Discov *10*, 460-475. 10.1158/2159-8290.Cd-19-0837.
Huang, Y., Triscari, J.M., Tseng, G.C., Pasa-Tolic, L., Lipton, M.S., Smith, R.D., and Wysocki, V.H. (2005). Statistical characterization of the charge state and residue dependence of low-energy CID peptide dissociation patterns. Anal Chem *77*, 5800-5813. 10.1021/ac0480949.
Hudson, D.M., and Eyre, D.R. (2013). Collagen prolyl 3-hydroxylation: a major role for a minor post-translational modification? Connect Tissue Res *54*, 245-251. 10.3109/03008207.2013.800867.
Hughes, B.T., and Espenshade, P.J. (2008). Oxygen-regulated degradation of fission yeast SREBP by Ofd1, a prolyl hydroxylase family member. Embo j *27*, 1491-1501. 10.1038/emboj.2008.83.
Huo, Z., Ye, J.-C., Chen, J., Lin, X., Zhou, Z.-N., Xu, X.-R., Li, C.-M., Qi, M., Liang, D., Liu, Y., and Li, J. (2012). Prolyl hydroxylase domain protein 2 regulates the intracellular cyclic AMP level in cardiomyocytes through its interaction with phosphodiesterase 4D. Biochemical and Biophysical Research Communications *427*, 73-79. https://doi.org/10.1016/j.bbrc.2012.09.005.
Hyvärinen, J., Hassinen, I.E., Sormunen, R., Mäki, J.M., Kivirikko, K.I., Koivunen, P., and Myllyharju, J. (2010). Hearts of hypoxia-inducible factor prolyl 4-hydroxylase-2 hypomorphic mice show protection against acute ischemia-reperfusion injury. The Journal of biological chemistry *285*, 13646-13657. 10.1074/jbc.M109.084855.
Ishikawa, Y., and Bächinger, H.P. (2013). A molecular ensemble in the rER for procollagen maturation. Biochim Biophys Acta *1833*, 2479-2491. 10.1016/j.bbamcr.2013.04.008.
Ishikawa, Y., Holden, P., and Bächinger, H.P. (2017). Heat shock protein 47 and 65-kDa FK506-binding protein weakly but synergistically interact during collagen folding in the

endoplasmic reticulum. Journal of Biological Chemistry *292*, 17216-17224. https://doi.org/10.1074/jbc.M117.802298.
Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J.M., Lane, W.S., and Kaelin, W.G., Jr. (2001). HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. Science *292*, 464-468. 10.1126/science.1059817.
Iwai, K., Yamanaka, K., Kamura, T., Minato, N., Conaway, R.C., Conaway, J.W., Klausner, R.D., and Pause, A. (1999). Identification of the von Hippel-lindau tumor-suppressor protein as part of an active E3 ubiquitin ligase complex. Proc Natl Acad Sci U S A *96*, 12436-12441. 10.1073/pnas.96.22.12436.
Jaakkola, P., Mole, D.R., Tian, Y.M., Wilson, M.I., Gielbert, J., Gaskell, S.J., von Kriegsheim, A., Hebestreit, H.F., Mukherji, M., Schofield, C.J., et al. (2001). Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. Science *292*, 468-472. 10.1126/science.1059796.
Jiang, B.H., Semenza, G.L., Bauer, C., and Marti, H.H. (1996). Hypoxia-inducible factor 1 levels vary exponentially over a physiologically relevant range of O2 tension. Am J Physiol *271*, C1172-1180. 10.1152/ajpcell.1996.271.4.C1172.
Jiang, W., Zhou, X., Li, Z., Liu, K., Wang, W., Tan, R., Cong, X., Shan, J., Zhan, Y., Cui, Z., et al. (2018). Prolyl 4-hydroxylase 2 promotes B-cell lymphoma progression via hydroxylation of Carabin. Blood *131*, 1325-1336. 10.1182/blood-2017-07-794875.
Johnston, H.E., Yadav, K., Kirkpatrick, J.M., Biggs, G.S., Oxley, D., Kramer, H.B., and Samant, R.S. (2021). Solvent Precipitation SP3 (SP4) enhances recovery for proteomics sample preparation without magnetic beads. bioRxiv, 2021.2009.2024.461247. 10.1101/2021.09.24.461247.
Jordanovski, D., Herwartz, C., Pawlowski, A., Taute, S., Frommolt, P., and Steger, G. (2013). The hypoxia-inducible transcription factor ZNF395 is controlled by IκB kinase-signaling and activates genes involved in the innate immune response and cancer. PLoS One *8*, e74911. 10.1371/journal.pone.0074911.
Jung, K., Cho, W., and Regnier, F.E. (2009). Glycoproteomics of plasma based on narrow selectivity lectin affinity chromatography. Journal of proteome research *8*, 643-650. 10.1021/pr8007495.
Kaelin, W.G. (2007). Von Hippel-Lindau disease. Annu Rev Pathol *2*, 145-173. 10.1146/annurev.pathol.2.010506.092049.
Kaelin, W.G., Jr. (2002). Molecular basis of the VHL hereditary cancer syndrome. Nat Rev Cancer *2*, 673-682. 10.1038/nrc885.
Kaelin, W.G., Jr. (2017). The VHL Tumor Suppressor Gene: Insights into Oxygen Sensing and Cancer. Trans Am Clin Climatol Assoc *128*, 298-307.
Kapp, E.A., Schütz, F., Reid, G.E., Eddes, J.S., Moritz, R.L., O'Hair, R.A., Speed, T.P., and Simpson, R.J. (2003). Mining a tandem mass spectrometry database to determine the trends and global factors influencing peptide fragmentation. Anal Chem *75*, 6251-6264. 10.1021/ac034616t.
Kato, H., Matsumura, Y., and Maeda, H. (1988). Isolation and identification of hydroxyproline analogues of bradykinin in human urine. FEBS Lett *232*, 252-254. 10.1016/0014-5793(88)80427-7.
Katz, M.J., Acevedo, J.M., Loenarz, C., Galagovsky, D., Liu-Yi, P., Pérez-Pepe, M., Thalhammer, A., Sekirnik, R., Ge, W., Melani, M., et al. (2014). Sudestada1, a Drosophila ribosomal prolyl-hydroxylase required for mRNA translation, cell

homeostasis, and organ growth. Proc Natl Acad Sci U S A *111*, 4025-4030. 10.1073/pnas.1314485111.
Khatun, J., Ramkissoon, K., and Giddings, M.C. (2007). Fragmentation characteristics of collision-induced dissociation in MALDI TOF/TOF mass spectrometry. Anal Chem *79*, 3032-3040. 10.1021/ac061455v.
Kiriakidis, S., Hoer, S.S., Burrows, N., Biddlecome, G., Khan, M.N., Thinnes, C.C., Schofield, C.J., Rogers, N., Botto, M., Paleolog, E., and Maxwell, P.H. (2017). Complement C1q is hydroxylated by collagen prolyl 4 hydroxylase and is sensitive to off-target inhibition by prolyl hydroxylase domain inhibitors that stabilize hypoxia-inducible factor. Kidney Int *92*, 900-908. 10.1016/j.kint.2017.03.008.
Kirk, T.Z., Evans, J.S., and Veis, A. (1987). Biosynthesis of type I procollagen. Characterization of the distribution of chain sizes and extent of hydroxylation of polysome-associated pro-alpha-chains. The Journal of biological chemistry *262*, 5540-5545.
Kiyonami, R., Schoen, A., Prakash, A., Peterman, S., Zabrouskov, V., Picotti, P., Aebersold, R., Huhmer, A., and Domon, B. (2011). Increased selectivity, analytical precision, and throughput in targeted proteomics. Molecular & cellular proteomics : MCP *10*, M110.002931. 10.1074/mcp.M110.002931.
Köditz, J., Nesper, J., Wottawa, M., Stiehl, D.P., Camenisch, G., Franke, C., Myllyharju, J., Wenger, R.H., and Katschinski, D.M. (2007). Oxygen-dependent ATF-4 stability is mediated by the PHD3 oxygen sensor. Blood *110*, 3610-3617. 10.1182/blood-2007-06-094441.
Koivunen, P., Helaakoski, T., Annunen, P., Veijola, J., Räisänen, S., Pihlajaniemi, T., and Kivirikko, K.I. (1996). ERp60 does not substitute for protein disulphide isomerase as the beta-subunit of prolyl 4-hydroxylase. Biochem J *316 (Pt 2)*, 599-605. 10.1042/bj3160599.
Koivunen, P., Tiainen, P., Hyvärinen, J., Williams, K.E., Sormunen, R., Klaus, S.J., Kivirikko, K.I., and Myllyharju, J. (2007). An endoplasmic reticulum transmembrane prolyl 4-hydroxylase is induced by hypoxia and acts on hypoxia-inducible factor alpha. The Journal of biological chemistry *282*, 30544-30552. 10.1074/jbc.M704988200.
Koritzinsky, M., Levitin, F., van den Beucken, T., Rumantir, R.A., Harding, N.J., Chu, K.C., Boutros, P.C., Braakman, I., and Wouters, B.G. (2013). Two phases of disulfide bond formation have differing requirements for oxygen. J Cell Biol *203*, 615-627. 10.1083/jcb.201307185.
Kotecki, M., Reddy, P.S., and Cochran, B.H. (1999). Isolation and characterization of a near-haploid human cell line. Exp Cell Res *252*, 273-280. 10.1006/excr.1999.4656.
Kotiaho, T., Eberlin, M.N., Vainiotalo, P., and Kostiainen, R. (2000). Electrospray mass and tandem mass spectrometry identification of ozone oxidation products of amino acids and small peptides. J. Am. Soc. Mass Spectrom. *11*, 526-535. https://doi.org/10.1016/S1044-0305(00)00116-1.
Krieg, A.J., Rankin, E.B., Chan, D., Razorenova, O., Fernandez, S., and Giaccia, A.J. (2010). Regulation of the histone demethylase JMJD1A by hypoxia-inducible factor 1 alpha enhances hypoxic gene expression and tumor growth. Mol Cell Biol *30*, 344-353. 10.1128/mcb.00444-09.
Kumar, R., González-Prieto, R., Xiao, Z., Verlaan-de Vries, M., and Vertegaal, A.C.O. (2017). The STUbL RNF4 regulates protein group SUMOylation by targeting the SUMO conjugation machinery. Nat Commun *8*, 1809. 10.1038/s41467-017-01900-x.

Kupferschmidt, K. (2019). Cellular oxygen sensor system earns Nobel for trio. Science *366*, 167. 10.1126/science.366.6462.167.
Kurban, G., Duplan, E., Ramlal, N., Hudon, V., Sado, Y., Ninomiya, Y., and Pause, A. (2008). Collagen matrix assembly is driven by the interaction of von Hippel-Lindau tumor suppressor protein with hydroxylated collagen IV alpha 2. Oncogene *27*, 1004-1012. 10.1038/sj.onc.1210709.
Kuznetsova, A.V., Meller, J., Schnell, P.O., Nash, J.A., Ignacak, M.L., Sanchez, Y., Conaway, J.W., Conaway, R.C., and Czyzyk-Krzeska, M.F. (2003). von Hippel-Lindau protein binds hyperphosphorylated large subunit of RNA polymerase II through a proline hydroxylation motif and targets it for ubiquitination. Proc Natl Acad Sci U S A *100*, 2706-2711. 10.1073/pnas.0436037100.
Lagerwerf, F.M., van de Weert, M., Heerma, W., and Haverkamp, J. (1996). Identification of oxidized methionine in peptides. Rapid Commun Mass Spectrom *10*, 1905-1910. 10.1002/(sici)1097-0231(199612)10:15<1905::Aid-rcm755>3.0.Co;2-9.
Laitala, A., Aro, E., Walkinshaw, G., Mäki, J.M., Rossi, M., Heikkilä, M., Savolainen, E.R., Arend, M., Kivirikko, K.I., Koivunen, P., and Myllyharju, J. (2012). Transmembrane prolyl 4-hydroxylase is a fourth prolyl 4-hydroxylase regulating EPO production and erythropoiesis. Blood *120*, 3336-3344. 10.1182/blood-2012-07-441824.
Lando, D., Peet, D.J., Whelan, D.A., Gorman, J.J., and Whitelaw, M.L. (2002). Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch. Science *295*, 858-861. 10.1126/science.1068592.
Lee, Dong C., Sohn, Hyun A., Park, Z.-Y., Oh, S., Kang, Yun K., Lee, K.-m., Kang, M., Jang, Ye J., Yang, S.-J., Hong, Young K., et al. (2015). A Lactate-Induced Response to Hypoxia. Cell *161*, 595-609. 10.1016/j.cell.2015.03.011.
Lee, M.J., Tasaki, T., Moroi, K., An, J.Y., Kimura, S., Davydov, I.V., and Kwon, Y.T. (2005). RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. Proc Natl Acad Sci U S A *102*, 15030-15035. 10.1073/pnas.0507533102.
Lee, S.B., Frattini, V., Bansal, M., Castano, A.M., Sherman, D., Hutchinson, K., Bruce, J.N., Califano, A., Liu, G., Cardozo, T., et al. (2016). An ID2-dependent mechanism for VHL inactivation in cancer. Nature *529*, 172-177. 10.1038/nature16475.
Lee, S.B., Ko, A., Oh, Y.T., Shi, P., D'Angelo, F., Frangaj, B., Koller, A., Chen, E.I., Cardozo, T., Iavarone, A., and Lasorella, A. (2020). Proline Hydroxylation Primes Protein Kinases for Autophosphorylation and Activation. Mol Cell *79*, 376-389.e378. 10.1016/j.molcel.2020.06.021.
Levitin, F., Lee, S.C.S., Hulme, S., Rumantir, R.A., Wong, A.S., Meester, M.R., and Koritzinsky, M. (2021). Oxygen-independent disulfide bond formation in VEGF-A and CA9. The Journal of biological chemistry *296*, 100505. 10.1016/j.jbc.2021.100505.
Li, D., Hirsilä, M., Koivunen, P., Brenner, M.C., Xu, L., Yang, C., Kivirikko, K.I., and Myllyharju, J. (2004). Many amino acid substitutions in a hypoxia-inducible transcription factor (HIF)-1alpha-like peptide cause only minor changes in its hydroxylation by the HIF prolyl 4-hydroxylases: substitution of 3,4-dehydroproline or azetidine-2-carboxylic acid for the proline leads to a high rate of uncoupled 2-oxoglutarate decarboxylation. The Journal of biological chemistry *279*, 55051-55059. 10.1074/jbc.M410287200.
Li, J., Cai, Z., Bomgarden, R.D., Pike, I., Kuhn, K., Rogers, J.C., Roberts, T.M., Gygi, S.P., and Paulo, J.A. (2021). TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for Sample Multiplexing. Journal of proteome research *20*, 2964-2972. 10.1021/acs.jproteome.1c00168.

Li, L., Zhang, L., Zhang, X., Yan, Q., Minamishima, Y.A., Olumi, A.F., Mao, M., Bartz, S., and Kaelin, W.G. (2007). Hypoxia-Inducible Factor Linked to Differential Kidney Cancer Risk Seen with Type 2A and Type 2B VHL Mutations. Molecular and Cellular Biology *27*, 5381-5392. doi:10.1128/MCB.00282-07.
Li, S., Rodriguez, J., Li, W., Bullova, P., Fell, S.M., Surova, O., Westerlund, I., Topcic, D., Bergsland, M., Stenman, A., et al. (2019). EglN3 hydroxylase stabilizes BIM-EL linking VHL type 2C mutations to pheochromocytoma pathogenesis and chemotherapy resistance. Proc Natl Acad Sci U S A *116*, 16997-17006. 10.1073/pnas.1900748116.
Licausi, F., Kosmacz, M., Weits, D.A., Giuntoli, B., Giorgi, F.M., Voesenek, L.A., Perata, P., and van Dongen, J.T. (2011). Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. Nature *479*, 419-422. 10.1038/nature10536.
Lioe, H., O'Hair, R.A., and Reid, G.E. (2004). A mass spectrometric and molecular orbital study of H2O loss from protonated tryptophan and oxidized tryptophan derivatives. Rapid Commun Mass Spectrom *18*, 978-988. 10.1002/rcm.1434.
Lippl, K. (2018). Structural, biochemical and inhibition studies on hypoxia-inducible factor (HIF) prolyl hydroxylases. (University of Oxford).
Liu, H., Sadygov, R.G., and Yates, J.R., 3rd (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem *76*, 4193-4201. 10.1021/ac0498563.
Liu, X., Simon, J.M., Xie, H., Hu, L., Wang, J., Zurlo, G., Fan, C., Ptacek, T.S., Herring, L., Tan, X., et al. (2020a). Genome-wide Screening Identifies SFMBT1 as an Oncogenic Driver in Cancer with VHL Loss. Mol Cell *77*, 1294-1306.e1295. 10.1016/j.molcel.2020.01.009.
Liu, X.R., Zhang, M.M., and Gross, M.L. (2020b). Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. Chem Rev *120*, 4355-4454. 10.1021/acs.chemrev.9b00815.
Liu, Y., and Zhou, Y. (2022). Circ_0087960 stabilizes KDM5B by reducing SKP2 mediated ubiquitination degradation and promotes osteogenic differentiation in periodontal ligament stem cells. Regen Ther *19*, 122-130. 10.1016/j.reth.2022.01.003.
Loenarz, C., Coleman, M.L., Boleininger, A., Schierwater, B., Holland, P.W.H., Ratcliffe, P.J., and Schofield, C.J. (2011). The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens. EMBO reports *12*, 63-70. https://doi.org/10.1038/embor.2010.170.
Loenarz, C., Sekirnik, R., Thalhammer, A., Ge, W., Spivakovsky, E., Mackeen, M.M., McDonough, M.A., Cockman, M.E., Kessler, B.M., Ratcliffe, P.J., et al. (2014). Hydroxylation of the eukaryotic ribosomal decoding center affects translational accuracy. Proc Natl Acad Sci U S A *111*, 4019-4024. 10.1073/pnas.1311750111.
Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. 10.1186/s13059-014-0550-8.
Lu, W., Liu, S., Li, B., Xie, Y., Adhiambo, C., Yang, Q., Ballard, B.R., Nakayama, K.I., Matusik, R.J., and Chen, Z. (2015). SKP2 inactivation suppresses prostate tumorigenesis by mediating JARID1B ubiquitination. Oncotarget *6*, 771-788. 10.18632/oncotarget.2718.
Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C., and Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol Syst Biol *14*, e8126. 10.15252/msb.20178126.

Luo, W., Hu, H., Chang, R., Zhong, J., Knabel, M., O'Meally, R., Cole, R.N., Pandey, A., and Semenza, G.L. (2011). Pyruvate kinase M2 is a PHD3-stimulated coactivator for hypoxia-inducible factor 1. Cell *145*, 732-744. 10.1016/j.cell.2011.03.054.
Luo, W., Lin, B., Wang, Y., Zhong, J., O'Meally, R., Cole, R.N., Pandey, A., Levchenko, A., and Semenza, G.L. (2014). PHD3-mediated prolyl hydroxylation of nonmuscle actin impairs polymerization and cell motility. Mol Biol Cell *25*, 2788-2796. 10.1091/mbc.E14-02-0775.
Macek, B., Mann, M., and Olsen, J.V. (2009). Global and site-specific quantitative phosphoproteomics: principles and applications. Annu Rev Pharmacol Toxicol *49*, 199-221. 10.1146/annurev.pharmtox.011008.145606.
Markolovic, S., Wilkins, S.E., and Schofield, C.J. (2015). Protein Hydroxylation Catalyzed by 2-Oxoglutarate-dependent Oxygenases. The Journal of biological chemistry *290*, 20712-20722. 10.1074/jbc.R115.662627.
Masson, N., Keeley, T.P., Giuntoli, B., White, M.D., Puerta, M.L., Perata, P., Hopkinson, R.J., Flashman, E., Licausi, F., and Ratcliffe, P.J. (2019). Conserved N-terminal cysteine dioxygenases transduce responses to hypoxia in animals and plants. Science (New York, N.Y.) *365*, 65-69. 10.1126/science.aaw0112.
Masson, N., Willam, C., Maxwell, P.H., Pugh, C.W., and Ratcliffe, P.J. (2001). Independent function of two destruction domains in hypoxia-inducible factor-alpha chains activated by prolyl hydroxylation. Embo j *20*, 5197-5206. 10.1093/emboj/20.18.5197.
Mathieson, T., Franken, H., Kosinski, J., Kurzawa, N., Zinn, N., Sweetman, G., Poeckel, D., Ratnu, V.S., Schramm, M., Becher, I., et al. (2018). Systematic analysis of protein turnover in primary cells. Nat Commun *9*, 689. 10.1038/s41467-018-03106-1.
Maxwell, P.H., Pugh, C.W., and Ratcliffe, P.J. (1993). Inducible operation of the erythropoietin 3' enhancer in multiple cell lines: evidence for a widespread oxygen-sensing mechanism. Proc Natl Acad Sci U S A *90*, 2423-2427. 10.1073/pnas.90.6.2423.
Maxwell, P.H., Wiesener, M.S., Chang, G.W., Clifford, S.C., Vaux, E.C., Cockman, M.E., Wykoff, C.C., Pugh, C.W., Maher, E.R., and Ratcliffe, P.J. (1999). The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. Nature *399*, 271-275. 10.1038/20459.
May, D.G., Scott, K.L., Campos, A.R., and Roux, K.J. (2020). Comparative Application of BioID and TurboID for Protein-Proximity Biotinylation. Cells *9*. 10.3390/cells9051070.
McAlister, G.C., Huttlin, E.L., Haas, W., Ting, L., Jedrychowski, M.P., Rogers, J.C., Kuhn, K., Pike, I., Grothe, R.A., Blethrow, J.D., and Gygi, S.P. (2012). Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. Anal Chem *84*, 7469-7478. 10.1021/ac301572t.
McAlister, G.C., Nusinow, D.P., Jedrychowski, M.P., Wühr, M., Huttlin, E.L., Erickson, B.K., Rad, R., Haas, W., and Gygi, S.P. (2014). MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. Anal Chem *86*, 7150-7158. 10.1021/ac502040v.
McGee, J.O., Rhoads, R.E., and Udenfriend, S. (1971). The substrate recognition site of collagen proline hydroxylase: the hydroxylation of -X-Pro-Gly- sequences in bradykinin analogs and other peptides. Arch Biochem Biophys *144*, 343-351. 10.1016/0003-9861(71)90487-5.
Metzen, E., Berchner-Pfannschmidt, U., Stengel, P., Marxsen, J.H., Stolze, I., Klinger, M., Huang, W.Q., Wotzlaw, C., Hellwig-Bürgel, T., Jelkmann, W., et al. (2003).

Intracellular localisation of human HIF-1 alpha hydroxylases: implications for oxygen sensing. J Cell Sci *116*, 1319-1326. 10.1242/jcs.00318.
Metzen, E., and Ratcliffe, P.J. (2004). HIF hydroxylation and cellular oxygen sensing. Biol Chem *385*, 223-230. 10.1515/bc.2004.016.
Mikhaylova, O., Ignacak Monika, L., Barankiewicz Teresa, J., Harbaugh Svetlana, V., Yi, Y., Maxwell Patrick, H., Schneider, M., Van Geyte, K., Carmeliet, P., Revelo Monica, P., et al. (2008). The von Hippel-Lindau Tumor Suppressor Protein and Egl-9-Type Proline Hydroxylases Regulate the Large Subunit of RNA Polymerase II in Response to Oxidative Stress. Molecular and Cellular Biology *28*, 2701-2717. 10.1128/MCB.01231-07.
Minervini, G., Lopreiato, R., Bortolotto, R., Falconieri, A., Sartori, G., and Tosatto, S.C.E. (2017). Novel interactions of the von Hippel-Lindau (pVHL) tumor suppressor with the CDKN1 family of cell cycle inhibitors. Sci Rep *7*, 46562. 10.1038/srep46562.
Minervini, G., Mazzotta, G.M., Masiero, A., Sartori, E., Corrà, S., Potenza, E., Costa, R., and Tosatto, S.C. (2015). Isoform-specific interactions of the von Hippel-Lindau tumor suppressor protein. Sci Rep *5*, 12605. 10.1038/srep12605.
Mole, D.R., Blancher, C., Copley, R.R., Pollard, P.J., Gleadle, J.M., Ragoussis, J., and Ratcliffe, P.J. (2009). Genome-wide association of hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha DNA binding with expression profiling of hypoxia-inducible transcripts. The Journal of biological chemistry *284*, 16767-16775. 10.1074/jbc.M901790200.
Moore, C.E., Mikolajek, H., Regufe da Mota, S., Wang, X., Kenney, J.W., Werner, J.M., and Proud, C.G. (2015). Elongation Factor 2 Kinase Is Regulated by Proline Hydroxylation and Protects Cells during Hypoxia. Mol Cell Biol *35*, 1788-1804. 10.1128/mcb.01457-14.
Morand, K., Talbo, G., and Mann, M. (1993). Oxidation of peptides during electrospray ionization. Rapid Commun Mass Spectrom *7*, 738-743. 10.1002/rcm.1290070811.
Moser, S.C., Bensaddek, D., Ortmann, B., Maure, J.F., Mudie, S., Blow, J.J., Lamond, A.I., Swedlow, J.R., and Rocha, S. (2013). PHD1 links cell-cycle progression to oxygen sensing through hydroxylation of the centrosomal protein Cep192. Dev Cell *26*, 381-392. 10.1016/j.devcel.2013.06.014.
Moser, S.C., Bensaddek, D., Ortmann, B., Rocha, S., Lamond, A.I., and Swedlow, J.R. (2015). The Prolylhydroxylase PHD2 Modulates Centromere Function through the Hydroxylation of CENP-N. bioRxiv, 030452. 10.1101/030452.
Myllyharju, J., and Kivirikko, K.I. (2001). Collagens and collagen-related diseases. Ann Med *33*, 7-21. 10.3109/07853890109002055.
Myllyharju, J., and Kivirikko, K.I. (2004). Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet *20*, 33-43. 10.1016/j.tig.2003.11.004.
Na, S., Bandeira, N., and Paek, E. (2012). Fast multi-blind modification search through tandem mass spectrometry. Molecular & cellular proteomics : MCP *11*, M111.010199. 10.1074/mcp.M111.010199.
Ohh, M., Yauch, R.L., Lonergan, K.M., Whaley, J.M., Stemmer-Rachamimov, A.O., Louis, D.N., Gavin, B.J., Kley, N., Kaelin, W.G., Jr., and Iliopoulos, O. (1998). The von Hippel-Lindau tumor suppressor protein is required for proper assembly of an extracellular fibronectin matrix. Mol Cell *1*, 959-968. 10.1016/s1097-2765(00)80096-9.
Okumura, F., Uematsu, K., Byrne, S.D., Hirano, M., Joo-Okumura, A., Nishikimi, A., Shuin, T., Fukui, Y., Nakatsukasa, K., and Kamura, T. (2016). Parallel Regulation of von

Hippel-Lindau Disease by pVHL-Mediated Degradation of B-Myb and Hypoxia-Inducible Factor α. Mol Cell Biol *36*, 1803-1817. 10.1128/mcb.00067-16.
Onisko, B.C. (2020). The Hydroxyproline Proteome of HeLa Cells with Emphasis on the Active Sites of Protein Disulfide Isomerases. Journal of proteome research *19*, 756-768. 10.1021/acs.jproteome.9b00625.
Ono, M., Matsubara, J., Honda, K., Sakuma, T., Hashiguchi, T., Nose, H., Nakamori, S., Okusaka, T., Kosuge, T., Sata, N., et al. (2009). Prolyl 4-hydroxylation of alpha-fibrinogen: a novel protein modification revealed by plasma proteomics. The Journal of biological chemistry *284*, 29041-29049. 10.1074/jbc.M109.041749.
Pajak, B., Gajkowska, B., and Orzechowski, A. (2005). Cycloheximide-mediated sensitization to TNF-alpha-induced apoptosis in human colorectal cancer cell line COLO 205; role of FLIP and metabolic inhibitors. J Physiol Pharmacol *56 Suppl 3*, 101-118.
Paulo, J.A., O'Connell, J.D., and Gygi, S.P. (2016). A Triple Knockout (TKO) Proteomics Standard for Diagnosing Ion Interference in Isobaric Labeling Experiments. J Am Soc Mass Spectrom *27*, 1620-1625. 10.1007/s13361-016-1434-9.
Peterson, A.C., Russell, J.D., Bailey, D.J., Westphall, M.S., and Coon, J.J. (2012). Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. Molecular & cellular proteomics : MCP *11*, 1475-1488. 10.1074/mcp.O112.020131.
Picotti, P., and Aebersold, R. (2012). Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods *9*, 555-566. 10.1038/nmeth.2015.
Ploumakis, A., and Coleman, M.L. (2015). OH, the Places You'll Go! Hydroxylation, Gene Expression, and Cancer. Mol Cell *58*, 729-741. 10.1016/j.molcel.2015.05.026.
Pollard, P.J., and Kranc, K.R. (2010). Hypoxia signaling in hematopoietic stem cells: a double-edged sword. Cell Stem Cell *7*, 276-278. 10.1016/j.stem.2010.08.006.
Pollard, P.J., Loenarz, C., Mole, D.R., McDonough, M.A., Gleadle, J.M., Schofield, C.J., and Ratcliffe, P.J. (2008). Regulation of Jumonji-domain-containing histone demethylases by hypoxia-inducible factor (HIF)-1alpha. Biochem J *416*, 387-394. 10.1042/bj20081238.
Prigge, S.T., Mains, R.E., Eipper, B.A., and Amzel, L.M. (2000). New insights into copper monooxygenases and peptide amidation: structure, mechanism and function. Cell Mol Life Sci *57*, 1236-1259. 10.1007/pl00000763.
Pugh, C.W., and Ratcliffe, P.J. (2017). New horizons in hypoxia signaling pathways. Exp Cell Res *356*, 116-121. 10.1016/j.yexcr.2017.03.008.
Qi, H.H., Ongusaha, P.P., Myllyharju, J., Cheng, D., Pakkanen, O., Shi, Y., Lee, S.W., Peng, J., and Shi, Y. (2008). Prolyl 4-hydroxylation regulates Argonaute 2 stability. Nature *455*, 421-424. 10.1038/nature07186.
Rappu, P., Salo, A.M., Myllyharju, J., and Heino, J. (2019). Role of prolyl hydroxylation in the molecular interactions of collagens. Essays Biochem *63*, 325-335. 10.1042/ebc20180053.
Ratcliffe, P., Koivunen, P., Myllyharju, J., Ragoussis, J., Bovée, J.V., Batinic-Haberle, I., Vinatier, C., Trichet, V., Robriquet, F., Oliver, L., and Gardie, B. (2017). Update on hypoxia-inducible factors and hydroxylases in oxygen regulatory pathways: from physiology to therapeutics. Hypoxia (Auckl) *5*, 11-20. 10.2147/hp.S127042.
Ratcliffe, P.J. (2003). New insights into an enigmatic tumour suppressor. Nat Cell Biol *5*, 7-8. 10.1038/ncb0103-7.

Ratcliffe, P.J. (2013). Oxygen sensing and hypoxia signalling pathways in animals: the implications of physiology for cancer. J Physiol *591*, 2027-2042. 10.1113/jphysiol.2013.251470.
Rhoads, R.E., and Udenfriend, S. (1969). Substrate specificity of collagen proline hydroxylase: hydroxylation of a specific proline residue in bradykinin. Arch Biochem Biophys *133*, 108-111. 10.1016/0003-9861(69)90493-7.
Ricard-Blum, S. (2011). The collagen family. Cold Spring Harb Perspect Biol *3*, a004978. 10.1101/cshperspect.a004978.
Ritorto, M.S., Cook, K., Tyagi, K., Pedrioli, P.G., and Trost, M. (2013). Hydrophilic strong anion exchange (hSAX) chromatography for highly orthogonal peptide separation of complex proteomes. Journal of proteome research *12*, 2449-2457. 10.1021/pr301011r.
Rodriguez, J., Herrero, A., Li, S., Rauch, N., Quintanilla, A., Wynne, K., Krstic, A., Acosta, J.C., Taylor, C., Schlisio, S., and von Kriegsheim, A. (2018). PHD3 Regulates p53 Protein Stability by Hydroxylating Proline 359. Cell reports *24*, 1316-1329. 10.1016/j.celrep.2018.06.108.
Rodriguez, J., Pilkington, R., Garcia Munoz, A., Nguyen, L.K., Rauch, N., Kennedy, S., Monsefi, N., Herrero, A., Taylor, C.T., and von Kriegsheim, A. (2016). Substrate-Trapped Interactors of PHD3 and FIH Cluster in Distinct Signaling Pathways. Cell reports *14*, 2745-2760. 10.1016/j.celrep.2016.02.043.
Romero-Ruiz, A., Bautista, L., Navarro, V., Heras-Garvín, A., March-Díaz, R., Castellano, A., Gómez-Díaz, R., Castro, M.J., Berra, E., López-Barneo, J., and Pascual, A. (2012). Prolyl hydroxylase-dependent modulation of eukaryotic elongation factor 2 activity and protein translation under acute hypoxia. The Journal of biological chemistry *287*, 9651-9658. 10.1074/jbc.M111.299180.
Ross, A.B., Langer, J.D., and Jovanovic, M. (2020). Proteome Turnover in the Spotlight: Approaches, Applications, and Perspectives. Molecular & cellular proteomics : MCP *20*, 100016. 10.1074/mcp.R120.002190.
Salo, A.M., and Myllyharju, J. (2021). Prolyl and lysyl hydroxylases in collagen synthesis. Exp Dermatol *30*, 38-49. 10.1111/exd.14197.
Santos-Barriopedro, I., van Mierlo, G., and Vermeulen, M. (2021). Off-the-shelf proximity biotinylation for interaction proteomics. Nat Commun *12*, 5015. 10.1038/s41467-021-25338-4.
Schägger, H. (2006). Tricine-SDS-PAGE. Nat Protoc *1*, 16-22. 10.1038/nprot.2006.4.
Schmelzer, C.E., Nagel, M.B., Dziomba, S., Merkher, Y., Sivan, S.S., and Heinz, A. (2016). Prolyl hydroxylation in elastin is not random. Biochim Biophys Acta *1860*, 2169-2177. 10.1016/j.bbagen.2016.05.013.
Schmid-Burgk, J.L., Höning, K., Ebert, T.S., and Hornung, V. (2016). CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. Nat Commun *7*, 12338. 10.1038/ncomms12338.
Schmidt, T.G., Batz, L., Bonet, L., Carl, U., Holzapfel, G., Kiem, K., Matulewicz, K., Niermeier, D., Schuchardt, I., and Stanar, K. (2013). Development of the Twin-Strep-tag® and its application for purification of recombinant proteins from cell culture supernatants. Protein Expr Purif *92*, 54-61. 10.1016/j.pep.2013.08.021.
Schödel, J., Grampp, S., Maher, E.R., Moch, H., Ratcliffe, P.J., Russo, P., and Mole, D.R. (2016). Hypoxia, Hypoxia-inducible Transcription Factors, and Renal Cancer. Eur Urol *69*, 646-657. 10.1016/j.eururo.2015.08.007.
Schödel, J., Mole, D.R., and Ratcliffe, P.J. (2013). Pan-genomic binding of hypoxia-inducible transcription factors. Biol Chem *394*, 507-517. 10.1515/hsz-2012-0351.

Schödel, J., Oikonomopoulos, S., Ragoussis, J., Pugh, C.W., Ratcliffe, P.J., and Mole, D.R. (2011). High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. Blood *117*, e207-217. 10.1182/blood-2010-10-314427.
Schofield, C.J., and Ratcliffe, P.J. (2004). Oxygen sensing by HIF hydroxylases. Nat Rev Mol Cell Biol *5*, 343-354. 10.1038/nrm1366.
Scholz, C.C., Cavadas, M.A., Tambuwala, M.M., Hams, E., Rodríguez, J., von Kriegsheim, A., Cotter, P., Bruning, U., Fallon, P.G., Cheong, A., et al. (2013). Regulation of IL-1β-induced NF-κB by hydroxylases links key hypoxic and inflammatory signaling pathways. Proc Natl Acad Sci U S A *110*, 18490-18495. 10.1073/pnas.1309718110.
Schöneich, C. (2000). Mechanisms of metal-catalyzed oxidation of histidine to 2-oxo-histidine in peptides and proteins. J Pharm Biomed Anal *21*, 1093-1097. 10.1016/s0731-7085(99)00182-x.
Scott, D., Oldham, N.J., Strachan, J., Searle, M.S., and Layfield, R. (2015). Ubiquitin-binding domains: mechanisms of ubiquitin recognition and use as tools to investigate ubiquitin-modified proteomes. Proteomics *15*, 844-861. 10.1002/pmic.201400341.
Segura, I., Lange, C., Knevels, E., Moskalyuk, A., Pulizzi, R., Eelen, G., Chaze, T., Tudor, C., Boulegue, C., Holt, M., et al. (2016). The Oxygen Sensor PHD2 Controls Dendritic Spines and Synapses via Modification of Filamin A. Cell reports *14*, 2653-2667. 10.1016/j.celrep.2016.02.047.
Semenza, G.L. (1999). Regulation of mammalian O2 homeostasis by hypoxia-inducible factor 1. Annu Rev Cell Dev Biol *15*, 551-578. 10.1146/annurev.cellbio.15.1.551.
Semenza, G.L., and Wang, G.L. (1992). 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. Mol Cell Biol *12*, 5447-5454. 10.1128/mcb.12.12.5447-5454.1992.
Shao, C., Zhang, Y., and Sun, W. (2014). Statistical characterization of HCD fragmentation patterns of tryptic peptides on an LTQ Orbitrap Velos mass spectrometer. Journal of Proteomics *109*, 26-37. https://doi.org/10.1016/j.jprot.2014.06.012.
Shoulders, M.D., and Raines, R.T. (2009). Collagen structure and stability. Annu Rev Biochem *78*, 929-958. 10.1146/annurev.biochem.77.032207.120833.
Silva, A.M.N., Vitorino, R., Domingues, M.R.M., Spickett, C.M., and Domingues, P. (2013). Post-translational modifications and mass spectrometry detection. Free Radic Biol Med *65*, 925-941. 10.1016/j.freeradbiomed.2013.08.184.
Simpson, P.D., Eipper, B.A., Katz, M.J., Gandara, L., Wappner, P., Fischer, R., Hodson, E.J., Ratcliffe, P.J., and Masson, N. (2015). Striking Oxygen Sensitivity of the Peptidylglycine α-Amidating Monooxygenase (PAM) in Neuroendocrine Cells. The Journal of biological chemistry *290*, 24891-24901. 10.1074/jbc.M115.667246.
Singleton, R.S., Liu-Yi, P., Formenti, F., Ge, W., Sekirnik, R., Fischer, R., Adam, J., Pollard, P.J., Wolf, A., Thalhammer, A., et al. (2014). OGFOD1 catalyzes prolyl hydroxylation of RPS23 and is involved in translation control and stress granule formation. Proc Natl Acad Sci U S A *111*, 4031-4036. 10.1073/pnas.1314482111.
Singleton, R.S., Trudgian, D.C., Fischer, R., Kessler, B.M., Ratcliffe, P.J., and Cockman, M.E. (2011). Quantitative mass spectrometry reveals dynamics of factor-inhibiting hypoxia-inducible factor-catalyzed hydroxylation. The Journal of biological chemistry *286*, 33784-33794. 10.1074/jbc.M111.262808.

Stadtman, E.R., and Levine, R.L. (2003). Free radical-mediated oxidation of free amino acids and amino acid residues in proteins. Amino Acids *25*, 207-218. 10.1007/s00726-003-0011-2.
Stebbins, C.E., Kaelin, W.G., Jr., and Pavletich, N.P. (1999). Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. Science *284*, 455-461. 10.1126/science.284.5413.455.
Steen, H., and Mann, M. (2001). Similarity between condensed phase and gas phase chemistry: Fragmentation of peptides containing oxidized cysteine residues and its implications for proteomics. J. Am. Soc. Mass Spectrom. *12*, 228-232. 10.1021/jasms.8b01580.
Steen, H., and Mann, M. (2004). The abc's (and xyz's) of peptide sequencing. Nature Reviews Molecular Cell Biology *5*, 699-711. 10.1038/nrm1468.
Steinhoff, A., Pientka, F.K., Möckel, S., Kettelhake, A., Hartmann, E., Köhler, M., and Depping, R. (2009). Cellular oxygen sensing: Importins and exportins are mediators of intracellular localisation of prolyl-4-hydroxylases PHD1 and PHD2. Biochem Biophys Res Commun *387*, 705-711. 10.1016/j.bbrc.2009.07.090.
Stockhammer, A., Benz, L.S., Freund, C., Kuropka, B., and Bottanelli, F. (2021). When less is more - Endogenous tagging with TurboID increases the sensitivity of proximity labelling-based experiments. bioRxiv, 2021.2011.2019.469212. 10.1101/2021.11.19.469212.
Stoehr, A., Yang, Y., Patel, S., Evangelista, A.M., Aponte, A., Wang, G., Liu, P., Boylston, J., Kloner, P.H., Lin, Y., et al. (2016). Prolyl hydroxylation regulates protein degradation, synthesis, and splicing in human induced pluripotent stem cell-derived cardiomyocytes. Cardiovasc Res *110*, 346-358. 10.1093/cvr/cvw081.
Tabaro, F., Minervini, G., Sundus, F., Quaglia, F., Leonardi, E., Piovesan, D., and Tosatto, S.C. (2016). VHLdb: A database of von Hippel-Lindau protein interactors and mutations. Sci Rep *6*, 31128. 10.1038/srep31128.
Tabb, D.L., Smith, L.L., Breci, L.A., Wysocki, V.H., Lin, D., and Yates, J.R., 3rd (2003). Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides. Anal Chem *75*, 1155-1163. 10.1021/ac026122m.
Takahashi, N., Kuwaki, T., Kiyonaka, S., Numata, T., Kozai, D., Mizuno, Y., Yamamoto, S., Naito, S., Knevels, E., Carmeliet, P., et al. (2011). TRPA1 underlies a sensing mechanism for O2. Nat Chem Biol *7*, 701-711. 10.1038/nchembio.640.
Tanimoto, K., Makino, Y., Pereira, T., and Poellinger, L. (2000). Mechanism of regulation of the hypoxia-inducible factor-1 alpha by the von Hippel-Lindau tumor suppressor protein. Embo j *19*, 4298-4309. 10.1093/emboj/19.16.4298.
Tannous, A., Pisoni, G.B., Hebert, D.N., and Molinari, M. (2015). N-linked sugar-regulated protein folding and quality control in the ER. Semin Cell Dev Biol *41*, 79-89. 10.1016/j.semcdb.2014.12.001.
Tarade, D., and Ohh, M. (2018). The HIF and other quandaries in VHL disease. Oncogene *37*, 139-147. 10.1038/onc.2017.338.
Tello, D., Balsa, E., Acosta-Iborra, B., Fuertes-Yebra, E., Elorza, A., Ordóñez, Á., Corral-Escariz, M., Soro, I., López-Bernardo, E., Perales-Clemente, E., et al. (2011). Induction of the mitochondrial NDUFA4L2 protein by HIF-1α decreases oxygen consumption by inhibiting Complex I activity. Cell Metab *14*, 768-779. 10.1016/j.cmet.2011.10.008.
Teng-Umnuay, P., van der Wel, H., and West, C.M. (1999). Identification of a UDP-GlcNAc:Skp1-hydroxyproline GlcNAc-transferase in the cytoplasm of Dictyostelium. The Journal of biological chemistry *274*, 36392-36402. 10.1074/jbc.274.51.36392.

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., and Hamon, C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem *75*, 1895-1904. 10.1021/ac0262560.
Tian, Y.M., Yeoh, K.K., Lee, M.K., Eriksson, T., Kessler, B.M., Kramer, H.B., Edelmann, M.J., Willam, C., Pugh, C.W., Schofield, C.J., and Ratcliffe, P.J. (2011). Differential sensitivity of hypoxia inducible factor hydroxylation sites to hypoxia and hydroxylase inhibitors. The Journal of biological chemistry *286*, 13041-13051. 10.1074/jbc.M110.211110.
Tran, N.H., Rahman, M.Z., He, L., Xin, L., Shan, B., and Li, M. (2016). Complete De Novo Assembly of Monoclonal Antibody Sequences. Sci Rep *6*, 31730. 10.1038/srep31730.
Tran, N.H., Zhang, X., Xin, L., Shan, B., and Li, M. (2017). De novo peptide sequencing by deep learning. Proc Natl Acad Sci U S A *114*, 8247-8252. 10.1073/pnas.1705691114.
Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419. 10.1126/science.1260419.
Uhlmann, T., Geoghegan, V.L., Thomas, B., Ridlova, G., Trudgian, D.C., and Acuto, O. (2012). A method for large-scale identification of protein arginine methylation. Molecular & cellular proteomics : MCP *11*, 1489-1499. 10.1074/mcp.M112.020743.
Ullah, K., Rosendahl, A.H., Izzi, V., Bergmann, U., Pihlajaniemi, T., Mäki, J.M., and Myllyharju, J. (2017). Hypoxia-inducible factor prolyl-4-hydroxylase-1 is a convergent point in the reciprocal negative regulation of NF-κB and p53 signaling pathways. Sci Rep *7*, 17220. 10.1038/s41598-017-17376-0.
van Bentum, M., and Selbach, M. (2021). An Introduction to Advanced Targeted Acquisition Methods. Molecular & cellular proteomics : MCP *20*, 100165. 10.1016/j.mcpro.2021.100165.
van der Wal, L., Bezstarosti, K., Sap, K.A., Dekkers, D.H.W., Rijkers, E., Mientjes, E., Elgersma, Y., and Demmers, J.A.A. (2018). Improvement of ubiquitylation site detection by Orbitrap mass spectrometry. J Proteomics *172*, 49-56. 10.1016/j.jprot.2017.10.014.
van Huizen, N.A., Burgers, P.C., Saintmont, F., Brocorens, P., Gerbaux, P., Stingl, C., Dekker, L.J.M., Ijzermans, J.N.M., and Luider, T.M. (2019). Identification of 4-Hydroxyproline at the Xaa Position in Collagen by Mass Spectrometry. Journal of proteome research *18*, 2045-2051. 10.1021/acs.jproteome.8b00930.
Verrastro, I., Pasha, S., Jensen, K.T., Pitt, A.R., and Spickett, C.M. (2015). Mass spectrometry-based methods for identifying oxidized proteins in disease: advances and challenges. Biomolecules *5*, 378-411. 10.3390/biom5020378.
Vizcaíno, J.A., Deutsch, E.W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., Dianes, J.A., Sun, Z., Farrah, T., Bandeira, N., et al. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat. Biotechnol. *32*, 223-226. 10.1038/nbt.2839.
Wang, G.L., Jiang, B.H., Rue, E.A., and Semenza, G.L. (1995). Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. Proc Natl Acad Sci U S A *92*, 5510-5514. 10.1073/pnas.92.12.5510.
Wang, J., Zhang, W., Ji, W., Liu, X., Ouyang, G., and Xiao, W. (2014). The von hippel-lindau protein suppresses androgen receptor activity. Mol Endocrinol *28*, 239-248. 10.1210/me.2013-1258.

Wang, X., Hu, J., Fang, Y., Fu, Y., Liu, B., Zhang, C., Feng, S., and Lu, X. (2022). Multi-Omics Profiling to Assess Signaling Changes upon VHL Restoration and Identify Putative VHL Substrates in Clear Cell Renal Cell Carcinoma Cell Lines. Cells *11*. 10.3390/cells11030472.
Weits, D.A., Giuntoli, B., Kosmacz, M., Parlanti, S., Hubberten, H.M., Riegler, H., Hoefgen, R., Perata, P., van Dongen, J.T., and Licausi, F. (2014). Plant cysteine oxidases control the oxygen-dependent branch of the N-end-rule pathway. Nat Commun *5*, 3425. 10.1038/ncomms4425.
Wenger, R.H., and Kurtz, A. (2011). Erythropoietin. Compr Physiol *1*, 1759-1794. 10.1002/cphy.c100075.
White, M.D., Klecker, M., Hopkinson, R.J., Weits, D.A., Mueller, C., Naumann, C., O'Neill, R., Wickens, J., Yang, J., Brooks-Bartlett, J.C., et al. (2017). Plant cysteine oxidases are dioxygenases that directly enable arginyl transferase-catalysed arginylation of N-end rule targets. Nat Commun *8*, 14690. 10.1038/ncomms14690.
Wierenga, A.T., Vellenga, E., and Schuringa, J.J. (2014). Convergence of hypoxia and TGFβ pathways on cell cycle regulation in human hematopoietic stem/progenitor cells. PLoS One *9*, e93494. 10.1371/journal.pone.0093494.
Wysocki, V.H., Tsaprailis, G., Smith, L.L., and Breci, L.A. (2000). Mobile and localized protons: a framework for understanding peptide dissociation. J Mass Spectrom *35*, 1399-1406. 10.1002/1096-9888(200012)35:12<1399::Aid-jms86>3.0.Co;2-r.
Xia, X., Lemieux, M.E., Li, W., Carroll, J.S., Brown, M., Liu, X.S., and Kung, A.L. (2009). Integrative analysis of HIF binding and transactivation reveals its role in maintaining histone methylation homeostasis. Proc Natl Acad Sci U S A *106*, 4260-4265. 10.1073/pnas.0810067106.
Xiao, Z., Chang, J.G., Hendriks, I.A., Sigurðsson, J.O., Olsen, J.V., and Vertegaal, A.C. (2015). System-wide Analysis of SUMOylation Dynamics in Response to Replication Stress Reveals Novel Small Ubiquitin-like Modified Target Proteins and Acceptor Lysines Relevant for Genome Stability. Molecular & cellular proteomics : MCP *14*, 1419-1434. 10.1074/mcp.O114.044792.
Xie, L., Pi, X., Mishra, A., Fong, G., Peng, J., and Patterson, C. (2012). PHD3-dependent hydroxylation of HCLK2 promotes the DNA damage response. J Clin Invest *122*, 2827-2836. 10.1172/jci62374.
Xie, L., Pi, X., Townley-Tilson, W.H., Li, N., Wehrens, X.H., Entman, M.L., Taffet, G.E., Mishra, A., Peng, J., Schisler, J.C., et al. (2015). PHD2/3-dependent hydroxylation tunes cardiac response to β-adrenergic stress via phospholamban. J Clin Invest *125*, 2759-2771. 10.1172/jci80369.
Xie, L., Xiao, K., Whalen, E.J., Forrester, M.T., Freeman, R.S., Fong, G., Gygi, S.P., Lefkowitz, R.J., and Stamler, J.S. (2009). Oxygen-regulated beta(2)-adrenergic receptor hydroxylation by EGLN3 and ubiquitylation by pVHL. Sci Signal *2*, ra33. 10.1126/scisignal.2000444.
Xu, G., and Chance, M.R. (2007). Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. Chem Rev *107*, 3514-3543. 10.1021/cr0682047.
Xu, G., Paige, J.S., and Jaffrey, S.R. (2010). Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. Nat Biotechnol *28*, 868-873. 10.1038/nbt.1654.
Xu, W., Zhou, B., Zhao, X., Zhu, L., Xu, J., Jiang, Z., Chen, D., Wei, Q., Han, M., Feng, L., et al. (2018). KDM5B demethylates H3K4 to recruit XRCC1 and promote chemoresistance. Int J Biol Sci *14*, 1122-1132. 10.7150/ijbs.25881.

Yan, B., Huo, Z., Liu, Y., Lin, X., Li, J., Peng, L., Zhao, H., Zhou, Z.N., Liang, X., Liu, Y., et al. (2011a). Prolyl hydroxylase 2: a novel regulator of β2 -adrenoceptor internalization. J Cell Mol Med *15*, 2712-2722. 10.1111/j.1582-4934.2011.01268.x.
Yan, B., Jiao, S., Zhang, H.S., Lv, D.D., Xue, J., Fan, L., Wu, G.H., and Fang, J. (2011b). Prolyl hydroxylase domain protein 3 targets Pax2 for destruction. Biochem Biophys Res Commun *409*, 315-320. 10.1016/j.bbrc.2011.05.012.
Yasumoto, K., Kowata, Y., Yoshida, A., Torii, S., and Sogawa, K. (2009). Role of the intracellular localization of HIF-prolyl hydroxylases. Biochim Biophys Acta *1793*, 792-797. 10.1016/j.bbamcr.2009.01.014.
Yates, J.R., Ruse, C.I., and Nakorchevsky, A. (2009). Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. Annual Review of Biomedical Engineering *11*, 49-79. 10.1146/annurev-bioeng-061008-124934.
Yeh, T.L., Leissing, T.M., Abboud, M.I., Thinnes, C.C., Atasoylu, O., Holt-Martyn, J.P., Zhang, D., Tumber, A., Lippl, K., Lohans, C.T., et al. (2017). Molecular and cellular mechanisms of HIF prolyl hydroxylase inhibitors in clinical trials. Chem Sci *8*, 7651-7668. 10.1039/c7sc02103h.
Yeo, K.S., Tan, M.C., Lim, Y.-Y., and Ea, C.-K. (2017). JMJD8 is a novel endoplasmic reticulum protein with a JmjC domain. Scientific reports *7*, 15407-15407. 10.1038/s41598-017-15676-z.
Yoneyama, T., Ohtsuki, S., Ono, M., Ohmine, K., Uchida, Y., Yamada, T., Tachikawa, M., and Terasaki, T. (2013). Quantitative targeted absolute proteomics-based large-scale quantification of proline-hydroxylated α-fibrinogen in plasma for pancreatic cancer diagnosis. Journal of proteome research *12*, 753-762. 10.1021/pr3008144.
Yoon, H., Spinelli, J.B., Zaganjor, E., Wong, S.J., German, N.J., Randall, E.C., Dean, A., Clermont, A., Paulo, J.A., Garcia, D., et al. (2020). PHD3 Loss Promotes Exercise Capacity and Fat Oxidation in Skeletal Muscle. Cell Metab *32*, 215-228.e217. 10.1016/j.cmet.2020.06.017.
Yu, F., White, S.B., Zhao, Q., and Lee, F.S. (2001). HIF-1alpha binding to VHL is regulated by stimulus-sensitive proline hydroxylation. Proc Natl Acad Sci U S A *98*, 9630-9635. 10.1073/pnas.181341498.
Zhang, J., Wu, T., Simon, J., Takada, M., Saito, R., Fan, C., Liu, X.D., Jonasch, E., Xie, L., Chen, X., et al. (2018). VHL substrate transcription factor ZHX2 as an oncogenic driver in clear cell renal cell carcinoma. Science *361*, 290-295. 10.1126/science.aap8411.
Zhang, J., and Zhang, Q. (2018). VHL and Hypoxia Signaling: Beyond HIF in Cancer. Biomedicines *6*. 10.3390/biomedicines6010035.
Zhang, W.M., Kapyla, J., Puranen, J.S., Knight, C.G., Tiger, C.F., Pentikainen, O.T., Johnson, M.S., Farndale, R.W., Heino, J., and Gullberg, D. (2003). alpha 11beta 1 integrin recognizes the GFOGER sequence in interstitial collagens. The Journal of biological chemistry *278*, 7270-7277. 10.1074/jbc.M210313200.
Zhang, Z., Wu, S., Stenoien, D.L., and Paša-Tolić, L. (2014). High-Throughput Proteomics. Annual Review of Analytical Chemistry *7*, 427-454. 10.1146/annurev-anchem-071213-020216.
Zhao, Y., and Jensen, O.N. (2009). Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques. Proteomics *9*, 4632-4641. 10.1002/pmic.200900398.
Zhao, Z., Su, Z., Liang, P., Liu, D., Yang, S., Wu, Y., Ma, L., Feng, J., Zhang, X., Wu, C., et al. (2020). USP38 Couples Histone Ubiquitination and Methylation via KDM5B to Resolve Inflammation. Adv Sci (Weinh) *7*, 2002680. 10.1002/advs.202002680.

Zheng, X., Zhai, B., Koivunen, P., Shin, S.J., Lu, G., Liu, J., Geisen, C., Chakraborty, A.A., Moslehi, J.J., Smalley, D.M., et al. (2014). Prolyl hydroxylation by EglN2 destabilizes FOXO3a by blocking its interaction with the USP9x deubiquitinase. Genes Dev *28*, 1429-1444. 10.1101/gad.242131.114.
Zhou, B., Zhu, Y., Xu, W., Zhou, Q., Tan, L., Zhu, L., Chen, H., Feng, L., Hou, T., Wang, X., et al. (2021). Hypoxia Stimulates SUMOylation-Dependent Stabilization of KDM5B. Front Cell Dev Biol *9*, 741736. 10.3389/fcell.2021.741736.
Zhou, T., Erber, L., Liu, B., Gao, Y., Ruan, H.-B., and Chen, Y. (2016). Proteomic analysis reveals diverse proline hydroxylation-mediated oxygen-sensing cellular pathways in cancer cells. Oncotarget *7*, 79154-79169. 10.18632/oncotarget.12632.
Zi, Z., Yuan, L., Ding, Q., Kim, C., Wang, X.-d., Zhang, X., and Yu, Y. (2021). Hydroxylation of Actin Impairs Cell Motility Through Blocking its' Histidine Methylation. bioRxiv, 2021.2008.2026.456853. 10.1101/2021.08.26.456853.
Zolg, D.P., Wilhelm, M., Schmidt, T., Médard, G., Zerweck, J., Knaute, T., Wenschuh, H., Reimer, U., Schnatbaum, K., and Kuster, B. (2018). ProteomeTools: Systematic Characterization of 21 Post-translational Protein Modifications by Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) Using Synthetic Peptides. Molecular & cellular proteomics : MCP *17*, 1850-1863. 10.1074/mcp.TIR118.000783.
Zurlo, G., Guo, J., Takada, M., Wei, W., and Zhang, Q. (2016). New Insights into Protein Hydroxylation and Its Important Role in Human Diseases. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer *1866*, 208-220. https://doi.org/10.1016/j.bbcan.2016.09.004.
Zurlo, G., Liu, X., Takada, M., Fan, C., Simon, J.M., Ptacek, T.S., Rodriguez, J., von Kriegsheim, A., Liu, J., Locasale, J.W., et al. (2019). Prolyl hydroxylase substrate adenylosuccinate lyase is an oncogenic driver in triple negative breast cancer. Nat Commun *10*, 5177. 10.1038/s41467-019-13168-4.