UNIVERSITY OF LIVERPOOL

Identification and characterisation of the post-translational modifications that regulate the Hypoxia Inducible Factors, HIF-1α and HIF-2α.

Thesis submitted in accordance with the requirements of the
University of Liverpool
For the degree of Doctor in Philosophy by

Leonard Daly

September 2019

I. Acknowledgements

Reflecting on the past four years, this PhD has allowed me to grow both as a scientist and a person. The continual rollercoaster of emotions has certainly been a challenging experience, but one that has been made easier by the people I have met along the way. I would like to say thankyou to everyone who I have had the pleasure of working with on this journey.

I am indebted to my supervisors Violaine & Claire for their copious support, knowledge and expert guidance; without you both I wouldn't be the person I am today, I will always be grateful. I would also like to thank Philip Brownridge, a fountain of knowledge who always had a solution for my problems. A huge thank you goes to every member of Labs A & B past and present for being by my side through the last four years.

On a more personal note, I would like to thank my family, friends, pets and girlfriend for entertaining my answer to "how is work?", I know you may not always understand my answer, but your support and encouragement has been endearing throughout this journey.

To whoever has the 'pleasure' to read this thesis, I hope you enjoy it as much as I have.

II. Abstract

Eukaryotic organisms, including human, require molecular oxygen (O_2) to survive. During periods of low O_2 availability (hypoxia), a family of protein transcription factors become stabilised (Hypoxia Inducible Factors, HIF) and allow the adaption to hypoxia by regulating gene expression. Hypoxic adaption is required for cellular survival and is considered a hallmark of cancer. HIF is a heterodimeric protein consisting of a stable beta subunit (HIF-1β) and an O_2 labile alpha subunits (HIF-1α or HIF-2α). The two HIFα isoforms share ~50% sequence homology, yet have different target genes, O_2 sensitivity and sub-nuclear localisation. The O_2-dependent stability of HIFα subunits is due to an O_2 dependent, proline hydroxylation post translational modification (PTM) resulting in degradation. The current understanding of post-translational regulation beyond the O_2 dependent hydroxylation are poorly understood. Previous published studies aiming to identify HIF PTMs did it using a targeted/biased approach and failed to compare isoforms. Therefore, the primary aim of this work was to expand the understanding of the regulatory network of HIFα proteins, by obtaining an unbiased identification of PTMs and binding partners using a proteomics approach. We have performed an in-depth analysis of PTMs across ~90% of the total protein sequence for HIF-1α and HIF-2α in response to hypoxia, with a specific focus on phosphorylation. In total, ~50 different PTMs were confidently identified (~25 of which were phosphorylation) for each HIFα proteins, with the majority of these PTM sites being novel. Identified PTM sites were investigated through a combination of hypoxia regulation, evolutionary analysis, domain localisation and crystal structure modelling to identify potentially interesting sites to prioritise functional characterisation. This led to the discovery of HIF-1α Serine 31 phosphorylation, a previously superficially investigated site, as a potentially important mechanism to fully abolish HIF-1α-mediated transcription by preventing its binding to DNA. In addition, HIFα binding partners in response to hypoxia were identified. We demonstrated that many more proteins interact with HIFα proteins than currently known. The binding partner profiles were hypoxia-dependent, especially for HIF-2α which had >10 fold more binding partners confidently identified in hypoxia than normoxia. Combined with Gene ontology (GO) analysis, the binding partners identified strongly suggest a role for HIFα with mitochondria. Whilst we have discovered many novel data, this project has opened many avenues for further investigation. Overall, it is clear that the current understanding of HIF mediated hypoxia signalling is incomplete, and that the signalling pathways at play are orders of magnitude more complex than the current understanding.

III. Contents

IV. Table of figures

V. Table of tables

VI. Abbreviations

2D	2 dimensional
3D	3 dimensional
ACN	Acetonitrile
Ambic	Ammonium bicarbonate
APS	ammonium persulfate
ATP	Adenosine triphosphate
bHLH	basic helix-loop-helix
BSA	Bovine serum albumin
$CaCl_2$	Calcium chloride
CBP	CREB binding protein
CID	Collision induced dissociation
CMV	Cytomegalovirus
CO_2	Carbon dioxide
$CoCl_2$	Cobalt chloride
Co-IP	Co-immunoprecipitation
CTAD	Carboxy terminal transactivation domain
Da	Dalton
DDA	Data dependent acquisition
DFX	Deferasirox
DMEM	Dulbecco's modified eagles medium
DMOG	Dimethyloxalylglycine
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGTA	Egtazic acid
ER	Endoplasmic reticulum
ETD	Electron transfer dissociation
EThcD	Electron transfer, higher energy collisional, dissociation
FDR	False discovery rate
FIH	Factor inhibiting HIF
GSK3β	Glycogen synthase kinase 3β
H_2O	Water
HA	Hemagglutinin
HCD	Higher energy collisional dissociation
HIF	Hypoxia inducible factor
HRE	Hypoxia responsive element
HTP	High throughput
IAA	Iodoacetamide
IMAC	Immobilised metal ion affinity chromatography
IP	Immunoprecipitation
LB	Luria broth
LC	Liquid chromatography

LC-MS/MS	Liquid chromatography tandem mass spectrometry
LFQ	Label free quantification
LTP	Low throughput
m/z	mass to charge ratio
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulphate
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NCE	Normalised collision energy
NP-40	Nonidet-P40
NTAD	Amino terminal transactivation domain
O_2	Oxygen
ODDD	Oxygen dependent degradation domain
PAS	Per/ARNT/Sim domain
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PD	Proteome discoverer
PEG	Polyethylene glycol
PEI	Polyethylenimine
PHD	Prolyl hydroxylase
PKA	Protein kinase A
ppm	Parts per million
PSM	Peptide spectral match
qRT-PCR	Quantitative real time polymerase chain reaction
SCX	Strong cation exchange
SDM	Site directed mutagenesis
SDS	Sodium dodecyl sulphate
SRM	Selective reaction monitoring
TAE	Tris-acetate Ethylenediaminetetraacetic acid
TBS	Tris buffered saline
TBST	Tris buffered saline plus tween
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
TiO_2	Titanium dioxide
VHL	Von Hippel lindau E3 ubiquitin ligase

VIII. Amino acid codes

Amino acid	Three letter code	One letter code
Alanine	ala	A
Arginine	arg	R
Asparagine	asn	N
Aspartic acid	asp	D
Cysteine	cys	C
Glutamic acid	glu	E
Glutamine	gln	Q
Glycine	gly	G
Histidine	his	H
Isoleucine	ile	I
Leucine	leu	L
Lysine	lys	K
Methionine	met	M
Phenylalanine	phe	F
Proline	pro	P
Serine	ser	S
Threonine	thr	T
Tryptophan	trp	W
Tyrosine	tyr	Y
Valine	val	V

1. Chapter 1: Introduction

1.1. Cellular adaptation and the requirement of cell signalling

All living organisms, either single or multi-cellular, need to be able adapt to their changing environment in order to survive. Because various environmental signals (stimuli) are possible, organisms require mechanisms to sense specific environmental stimuli and transduce the signal to promote the appropriate adaptive response. Over a prolonged time period, the adaption process to the environment drives the evolution of species, highlighting its significance. For multicellular organisms, adaption at the cellular level permits cells to function as independent cell types to form highly complex systems.

An extracellular signal needs be transduced intracellularly to promote an adaptive response, known as cell signalling. The adaptive response is ultimately achieved by a change in the gene expression programme. The steps required involve protein complexes and a range of chemical reactions to convey the message to the DNA binding proteins. The mechanisms evolved by cells to identify a specific stimuli and regulate protein interactions, for the appropriate response, are highly varied. A generalised principle is that the stimuli is detected by specialised receptor proteins localised on cellular membranes and result in protein structural changes, permitting the interaction with other target proteins (known as effector proteins) which may then lead to gene expression regulation. However, certain stimuli can freely diffuse into cells allowing intracellular protein receptors to induce a signalling response, such as Estrogen signalling (Björnström *et al.*, 2005).

One important aspect of cell signalling is the ability to switch off the adaptive response once the cell has responded to the stimulus, or once the stimulus is no longer present, to prevent excessive response. Hence, effector proteins generally result in the expression of inhibitor proteins that inactivate the cellular signalling pathways, as part of a negative feedback loop system. A very common mechanism to regulate effector proteins is the cyclic protein accumulation/degradation in response to stimuli. Besides being highly metabolically inefficient, this mechanism can result in delays between stimulus detection and gene expression changes. Therefore, cells have evolved mechanisms to rapidly modify an already translated protein to swiftly alter its structural and functional properties, known as post translation modification (PTM). PTMs can be reversible and irreversible modifications. Thus, PTMs permit cell signalling without the accumulation/degradation cycling, providing a means for much more rapid responses.

1.1.1. Post translational modification (PTM)

PTM is a process that results in the covalent modification of a specific amino acid residue within the protein sequence. Thus, PTM alters the modified residues biochemical properties by size, and potentially charge, which can act as a switch to alter protein binding partners and function; thereby enabling signal transduction. As singular entities, PTMs are highly varied with >400 different types of PTM recorded in the UniProt database (The UniProt Consortium, 2017), highlighting the range of different strategies that have evolved to regulate protein function. The same PTM on different proteins does not necessarily result in the same functional outcome, nor does the same PTM of different sites within a given protein. Furthermore, identical residues can be modified by different PTMs, depending on the signal, and result in different regulatory roles. PTMs do not generally occur as single events but rather are highly abundant in 'decorating' the protein at sub-stoichiometric levels below total protein. Thus, PTMs result in the simultaneous fine tuning of signalling pathways and allow for a single protein to participate in multiple different regulatory pathways at once (Mann *et al.*, 2003). Therefore, a protein can be viewed as separate 'proteoforms', each having a different PTM map and function. Because of the large number of PTMs possible, only a few relevant PTMs to my work are discussed below.

1.1.1.1. Phosphorylation

Phosphorylation is a reversible PTM that involves the functional attachment of a negatively charged phosphate group to a residue, performed by a kinase and hydrolysed by a phosphatase (Figure 1.1). Phosphorylation is highly abundant within cells with >500 different human kinases identified, equating to ~2% of the proteins encoded by the human genome (Manning *et al.*, 2002), that are regulated by different cell signals and/or target different sequences for PTM within a protein. For scale, it is predicted that at any one moment approximately a third of all proteins in the cellular protein complement are phosphorylated (Cohen, 2001 & Olsen *et al.*, 2006). Phosphorylation canonically occurs on Serine (S), Threonine (T) and Tyrosine (Y) residues at a ratio of ~86% : 12% : 2% (Olsen *et al.*, 2006), however more recent studies identify that Histidine (H), Arginine (R), Lysine (K), Aspartic acid (D), Glutamic acid (E) and Cysteine (C) are also extensively phosphorylated in humans (Hardman *et al.*, 2019).

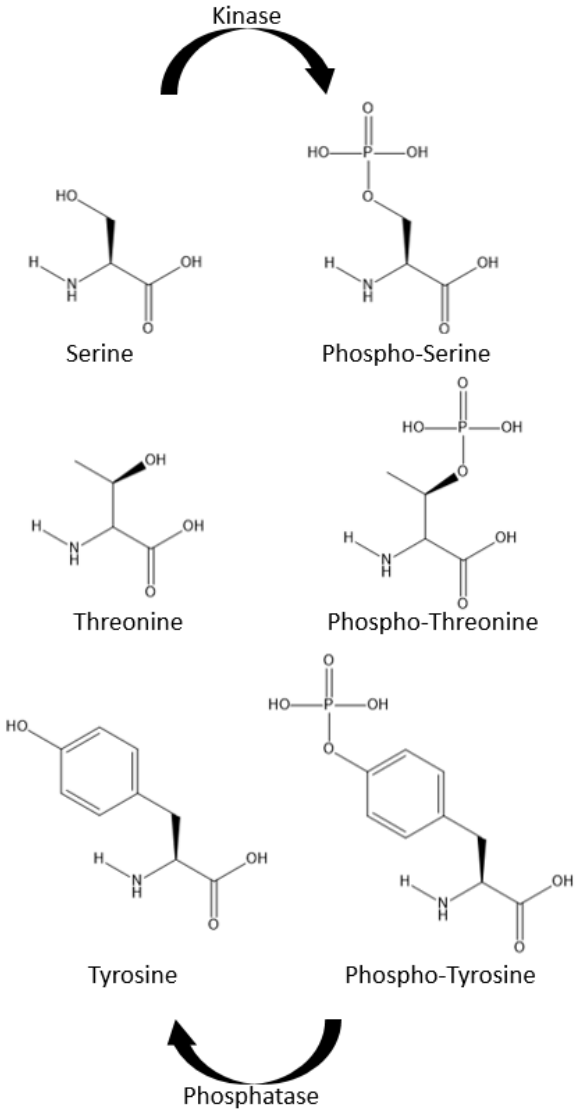


Figure 1.1: The structure of non-phosphorylated and phosphorylated Serine, Threonine and Tyrosine residues.

The relatively large and negatively charged phosphate group has been shown to have various effects to a modified protein, either by protein structural changes or acting as a docking site (to promote) or blocking agent (to prevent) additional binding partners from associating. Functionally, phosphorylation has been shown to regulate protein function by many mechanisms including the aforementioned binding partner role, altering the sub-cellular localisation of a protein and/or stability roles (reviewed by Hunter *et al.*, 1992). With such a vast a role in cellular signal transduction, aberrant phosphorylation/phosphatase expression (or function) is linked to many diseases and represents one of the largest markets for drug discovery (Cohen, 2001 & Cohen, 2002).

1.1.1.1. Lysine PTMs

Lysine is a positively charged amino acid that has been identified to undergo multiple PTM states including acetylation, methylation, ubiquitination and SUMOylation (Figure 1.2).

Many more lysine PTMs exist and have functionally important roles, however only the aforementioned PTMs are described here. Lysine acetylation, performed by acetyl-transferases and removed by de-acetylases, results in the addition of an acetyl group to the positively charged amino-group side chain of lysine and resulting in charge neutralisation. Acetylation has been shown to have various roles in protein regulation, similar to phosphorylation. Although acetylation has been identified as an essential signalling mechanism on a large proportion of human proteins, and at multiple different sites, the number of acetyl-transferases in the human genome (~20, Drazic *et al.*, 2016) is much smaller than the number of kinases identified. Dysregulation of acetylation is linked with multiple diseases (Drazic *et al.*, 2016 & Timmermann *et al.*, 2001).

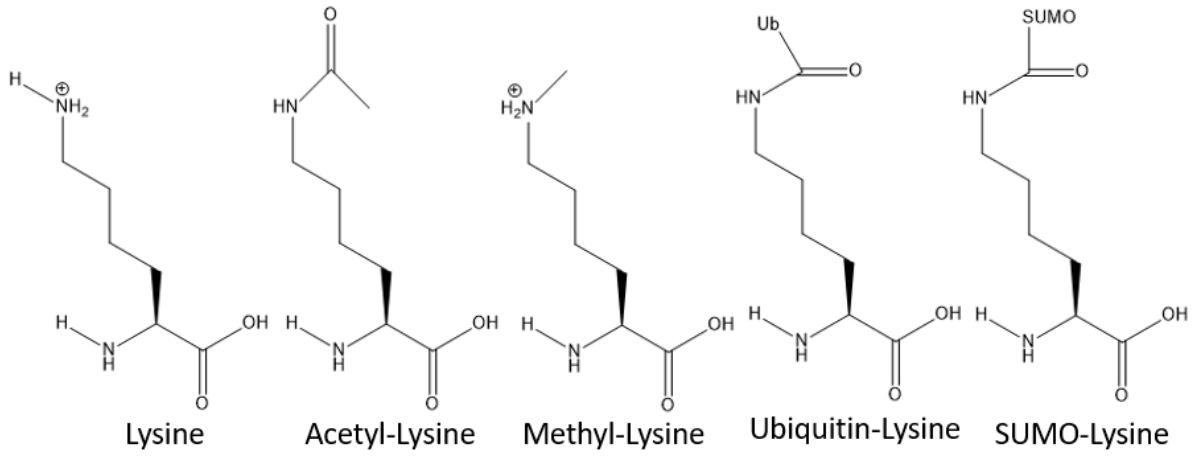


Figure 1.2: The structure of a selection of lysine PTMs.

Most of the investigation into the role of lysine methylation has been conducted on DNA bound proteins such as histones. These studies show that lysine can be mono-, di- and tri-methylated, each known to have different regulatory roles. Unlike acetylation, methylation does not result in charge neutralisation (Figure 1.2), hence is suggested to play a role primarily acting as docking/blocking site to change protein binding partners (Lanouette *et al.*, 2014 & Blanc *et al.*, 2017). Recent studies demonstrated the key role of methylation in cellular signalling pathways, as, for example, for the Epidermal growth factor receptor (EGFR) pathway (Hsu *et al.*, 2011), and for regulating transcription factor function by binding partners and/or stability means (Huang *et al.*, 2007 & Lee *et al.*, 2017). Because of the relatively recent discovery of their role beyond histone proteins, few methylated proteins have been identified, however the role of methylation in cancer progression is well known (Lanouette *et al.*, 2014).

Ubiquitin (Ub in Figure 1.2) is an example of a PTM that involves the covalent addition of a small (~8.5 kDa) protein tag to lysine residues. Ubiquitination is a 3-stage process involving a ubiquitin activating (E1), ubiquitin conjugating (E2) and ubiquitin ligase (E3) enzyme, thus providing the mechanism to target multiple different lysine residues (Komander *et al.*, 2012). Ubiquitination is a highly complex PTM, the ubiquitin molecule itself contains 7 different lysine residues which can undergo sequential ubiquitination PTMs leading to the formation of highly branched, poly-ubiquitin chains (Komander *et al.*, 2012). Ubiquitination is a reversible PTM by the function of de-ubiquitinase enzymes. There are ~80 encoded de-ubiquitinases in the human genome, highlighting the complexity of ubiquitin mediated signalling (Komander *et al.*, 2009). Poly-ubiquitination is generally associated with a catabolic role, increasing the rate that tagged proteins are degraded by the 26S proteasome. However, the different branching patterns of poly-ubiquitination are suggested to have different roles, and have been shown to stabilise proteins too (Komander *et al.*, 2012 & Sun *et al.*, 2004).

Similar to ubiquitin, Small Ubiquitin related Modifier (SUMO, Figure 1.2) is a 12 kDa protein tag that involves a 3-stage activation process for lysine modification (SUMOylation), and is reversible by SUMO-specific proteases (Flotho *et al.*, 2013). SUMOylation can result in complex SUMO chains, similar to ubiquitin (Flotho *et al.*, 2013). Humans encode 4 different SUMO isoforms (SUMO-1/2/3/4), where the SUMO-1 has poor sequence similarity to the highly conserved SUMO-2/3/4 isoforms (Flotho *et al.*, 2013). SUMOylation has varied roles dependent on the target and location (similar to phosphorylation and acetylation), but generally alters binding partners by acting as a docking or blocking site.

1.1.1.2. Redox sensitive PTMs

General cellular processes such as mitochondrial function result in the production of highly reactive free radical species, known as reactive oxygen and nitrogen species (ROS/RNS). ROS/RNS can non enzymatically react with multiple amino acids including methionine, phenylalanine, tryptophan, histidine and cysteine (Sharma *et al.*, 2010). Cysteine is particularly prone to ROS/RNS reaction by containing a free thiol group and results in multiple different reversible and irreversible PTM states known to regulate cellular signalling (Hess *et al.*, 2005 & Chung *et al.*, 2013).

Under severe oxidative stress, cysteine can undergo irreversible oxidation into 2+ and 3+ oxidative states, sulfinic acid and sulfonic acid respectively (Figure 1.3). Although not well

studied, these irreversible oxidative cysteine states are known to be important PTMs for modulating protein structure (Vivancos *et al.*, 2005, Blackinton *et al.*, 2009 & Fujiwara *et al.*, 2007). Additionally, enzymes have also been identified that utilise O_2 as a cofactor to hydroxylate specific residues including lysine, tryptophan, asparagine and proline (Figure 1.3). The role of hydroxylation is context specific, playing a major role in collagen structure but having different roles intracellularly on different proteins, from degradation to altering binding partners (Tak *et al.*, 2019 & Zurlo *et al.*, 2016). Recent studies suggest that intracellular proline hydroxylation is highly specific to a single family of proteins that are known to be stability regulated by O_2 tension (Cockman *et al.*, 2019).

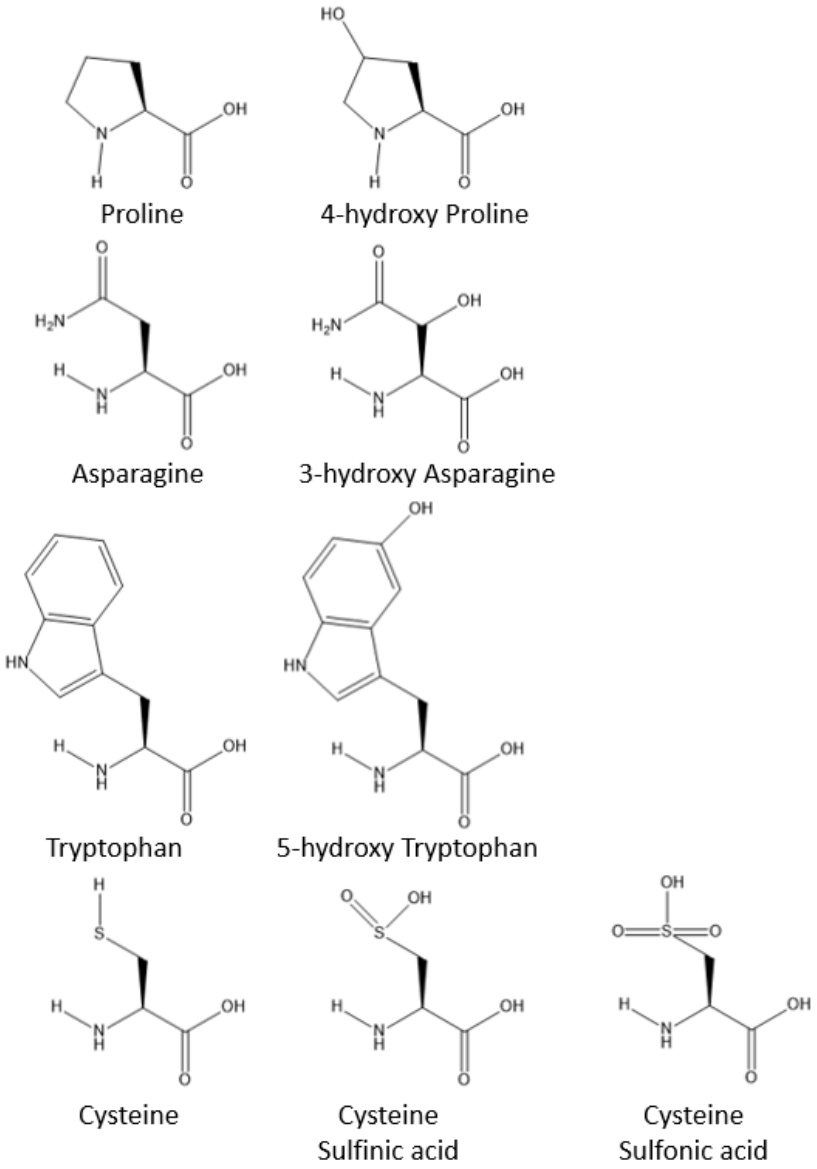


Figure 1.3: The structure of residues that can undergo redox sensitive PTMs.
Hydroxylation of proline, asparagine and tryptophan are enzyme mediate and are reversible. Cysteine oxidation states are ROS mediated and irreversible.

1.2. Oxygen-dependent signalling

For eukaryotic organisms, one example of an environmental stimuli that cells monitor is the availability of molecular oxygen (O_2). Eukaryotic organisms, including humans, are known as obligate aerobes and require O_2 to survive. O_2 is used in a process called oxidative phosphorylation, where O_2 is reduced into water (H_2O) to yield high quantities of ATP (Adenosine triphosphate), a chemical energy storage molecule. ATP is ubiquitously used by all organisms and cells to provide an energy source to promote chemically unfavourable reactions, including some PTMs. Although O_2 is essential for life, both over oxygenation (hyperoxia) and under oxygenation (hypoxia) can result in cellular stress, and are potentially lethal for organisms. Thus, cells have evolved highly specific O_2 sensing mechanisms to stringently monitor available O_2 and allow the adaption to changes in environmental O_2 tension. Therefore, beyond its role in metabolism, O_2 is a signalling molecule. Low O_2 levels in the cells surroundings (hypoxia) trigger a range of intracellular events, known as hypoxia signalling. This signalling pathway is the focus of this thesis.

1.3. Hypoxia

The definition of hypoxia is when the O_2 demand by a cell outweighs the available O_2 supply to that cell. Simply, if the O_2 supply to a tissue or cell is insufficient for regular metabolic activity then they are deemed hypoxic. Table 1.1 contains a summary of organs and required O_2 tension for normal function.

Table 1.1: Examples of O_2 tension in regular organs and associated tumours.
Median values of O_2 tension shown, data collected from Muz et al., 2015 & Carreau et al., 2011.

Tissue/Organ	Oxygen tension (Median % O_2)	Tumour oxygen tension (Median % O_2)
Inspired Air	20	X
Brain	4.6	1.7
Breast	8.5	1.5
Cervix	5.5	1.2
Heart	1.9	X
Kidney	9.5	1.3
Liver	5.5	0.8
Lung	5.6	2.2
Pancreas	7.5	0.3
Intestines	7.5	X
Rectal mucosa	3.9	1.8

In response to hypoxia, cells trigger pathways to ultimately promote the restoration of normal O_2 tension and O_2 independent energy metabolism, increasing cell survival rates

during periods of prolonged hypoxia. For example, hypoxia promotes the upregulation of genes encoding proteins that are involved with iron metabolism (Transferrin, Rolfs *et al.*, 1997), erythropoiesis (EPO, Wang *et al.*, 1993), angiogenesis (VEGF, Forsythe *et al.*, 1996), and the anaerobic stages of energy metabolism- glycolysis (Chen *et al.*, 2001 & Marsin *et al.*, 2002). Thus, pathways aim to increase the available O_2 saturation per unit of blood and permit increased rates of O_2 independent energy metabolism during hypoxia, however inefficiently.

Physiologically, hypoxia can be separated into two categories: global and localised hypoxia. Mild global hypoxia is an essential factor for correct embryonic development (where O_2 availability is limited by placental diffusion, (Dunwoodie, 2009), and adaptions to high altitudes (where the partial pressure of O_2 in the air is limited, Frisancho, 2013). While pathologies can induce global hypoxia, such as anaemia where there is a lack of haemoglobin containing iron to carry O_2, pathologies generally result in localised hypoxia by blocking the delivery of O_2 to tissues/cells, for example in heart attacks and strokes where blockages in blood vessels directly disrupts the supply of oxygenated blood to tissues and cells.

1.4. Hypoxia and cancer

Due to the rapid replication rate of tumours/cancers, the cell masses created are usually large in size and have a poor vasculature network, resulting in localised hypoxic cores within the tumour microenvironment (Figure 1.4). Table 1.1 has examples of the median O_2 tension of different tumour cell types compared to their normal cellular state. Since the first draft of “The Hallmarks of Cancer” in 2000, the increased rate of angiogenesis was a known characteristic that promotes cancer survival (Hanahan *et al.*, 2000). In 2011, an updated draft of The Hallmarks of Cancer was published which included the upregulation of glycolytic pathways (Hanahan *et al.*, 2011). Significantly, both of these pathways related to tumour survival are canonically hypoxia upregulated pathways, highlighting the potential importance of hypoxia in cancer.

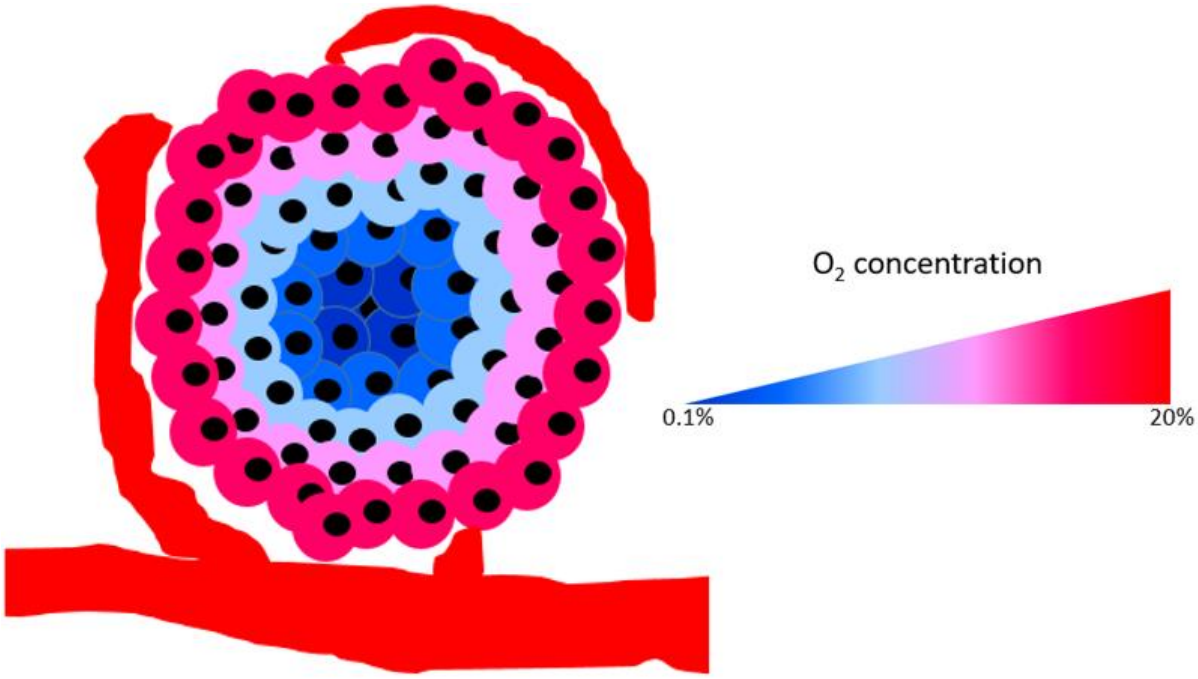


Figure 1.4: Hypoxia and the tumour microenvironment.
A large tumour mass combined with poor vasculature can cause hypoxic tumour cores, as low as 0.1% O_2.

Infact, it has been shown that an extended, or unregulated, hypoxic response can result in multiple of the hallmarks of cancer becoming upregulated (Figure 1.5), including: 1) Activating invasion and metastasis, through the upregulation of extracellular metabolism proteins such as matrix metalloproteases (MMPs, Ben-Yosef *et al.*, 2002) and uPA/uPAR (Urokinase Plasminogen Activator and receptor, Gupta *et al.*, 2011). 2) Sustaining proliferative signalling, through regulating autocrine signalling proteins such as Transforming Growth Factor –α (TGFα, Krishnamachary *et al.*, 2003) and Insulin-like Growth Factor 2 (IGF2, Feldser *et al.*, 1999). 3) Resisting cell death, through various mechanisms including the reduced expression of proteins involved with apoptosis pathways (Erler *et al.*, 2004) and upregulation of Multi-Drug Resistance (MDR) proteins; resulting chemo-resistance (Comerford *et al.*, 2002). Thus tumour/cancer hypoxia is a poor prognostic marker for cancer patient survival.

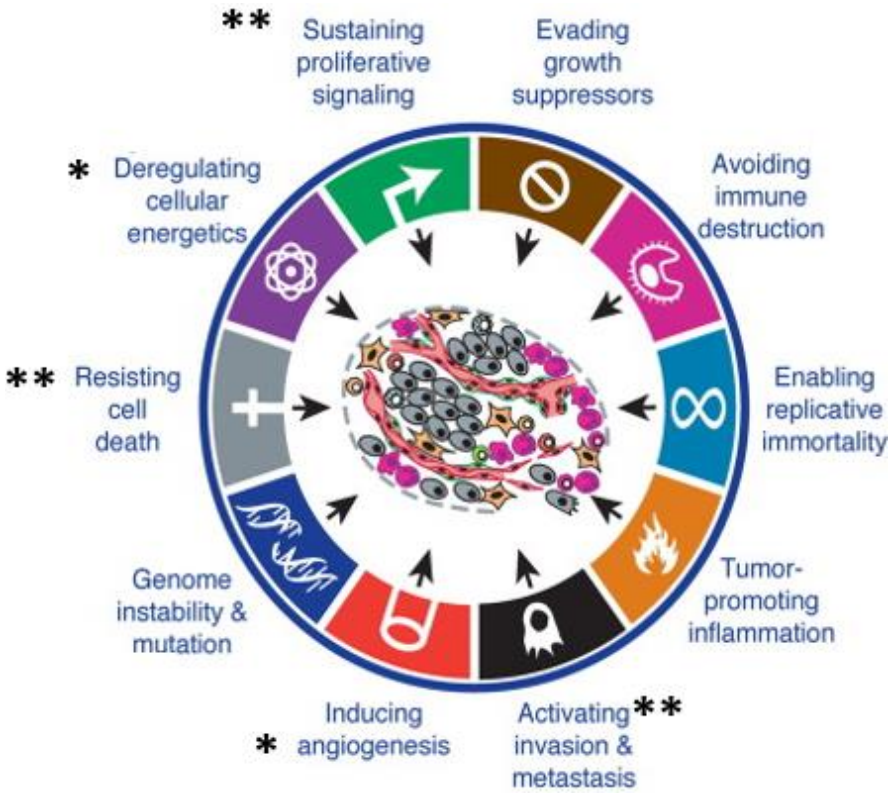


Figure 1.5: The hypoxia regulated hallmarks of cancer survival and progression.
Each section of the wheel contains a hallmark trait which promotes cancer survival and progression. *indicates primary targets of hypoxic regulation, **indicates secondary targets of extended hypoxia regulation. Modified from Hanahan *et al.*, 2011.

1.5. Hypoxia Inducible Factor- The master regulator of the hypoxic response

In 1995, the Greg Semenza lab discovered an O_2 dependent, labile protein transcription factor was responsible for the hypoxic induced expression of erythropoietin (EPO), and subsequent increased rate of angiogenesis, termed the Hypoxia Inducible Factor (HIF, Wang *et al.*, 1995). Wang and co-workers were able to identify that the active HIF transcription factor was a basic Helix-Loop-Helix (bHLH), Per-Arnt-Sim (PAS) domain containing, heterodimeric complex consisting of the O_2 dependently degraded protein (termed HIF-1α) and an O_2 insensitive protein (termed HIF-1β, also known as Aryl hydrocarbon Nuclear Translocator (ARNT)) (Wang *et al.*, 1995 & Wang *et al.*, 1995). Thus, in response to hypoxia (limited O_2 availability) HIF-1α is no longer degraded, permitting the formation of the active HIFαβ dimer that can regulate gene expression.

1.1.2. HIFα isoforms

Further characterisation of the HIF-1α protein (Pugh *et al.*, 1997 & Jiang *et al.*, 1997) identified three additional functional domains, and their minimalistic sequences: the

Oxygen Dependent Degradation domain (ODDD), the Amino-terminal Transactivation domain (NTAD) and Carboxyl-terminal Transactivation domain (CTAD), with the 2 TAD domains having distinctly different functions (schematically depicted in Figure 1.6).

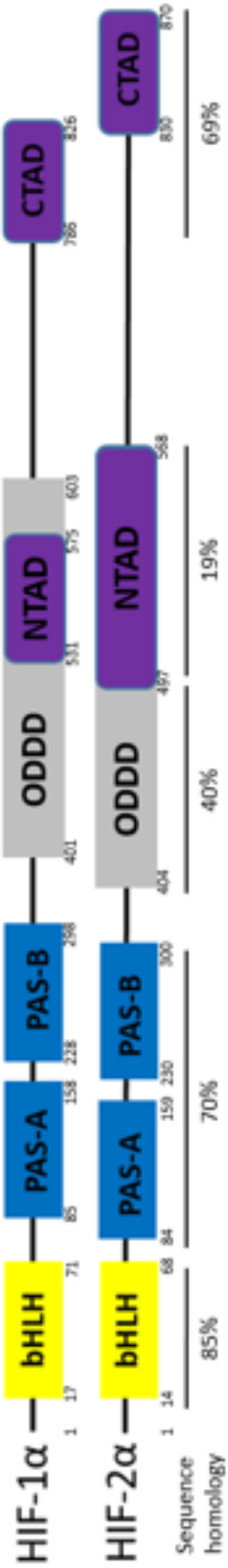


Figure 1.6: Schematic diagram of the HIF-1α and HIF-2α domain structure.
Both HIF-1α and HIF-2α contain a basic helix-loop-helix domain (bHLH), PAS domains, an oxygen dependent degradation domain (ODDD) and N-terminal and C-terminal transactivation domains (NTAD and CTAD). Domain numbering from Pugh *et al.*, 1997 and Jiang *et al.*, 1997. Sequence homology between major domains is detailed underneath the domain schematic (Tian *et al.*, 1997).

At the same time, a second hypoxia inducible transcription factor, with high sequence homology to HIF-1α, was simultaneously identified by four separate laboratories, thus the multiple names associated to it: Endothelial PAS domain protein -1 (EPAS1, Tian *et al.*, 1997), HIF-1α Like Factor (HLF, Ema *et al.*, 1997), HIF Related Factor (HRF, Flamme *et al.*, 1997) and Member Of PAS Superfamily 2 (MOPS2, Hogenesch *et al.*, 1997). Later studies identified that this second hypoxia responsive protein, though slightly larger than HIF-1α, could be functionally compartmentalised into bHLH, PAS, ODDD, NTAD and CTAD domains, identically to HIF-1α (Figure 1.6); thus this protein was termed HIF-2α (Wiesener *et al.*, 1998 and O'Rourke *et al.*, 1999).

A third HIFα gene locus exists (HIF-3α), which exhibits high rates of multiple splicing, giving rise to over six separate protein isoforms that generally contain an ODDD and NTAD in combination with different domains (Maynard *et al.*, 2003 & Gu *et al.*, 1998). The most predominant HIF-3α isoform is a 307 amino acid protein with extremely high sequence homology to the bHLH and PAS domains of HIF-1α and HIF-2α and functions in a dominant negative mechanism to inhibit HIFα signalling, termed Inhibitory PAS domain protein (IPAS, Makino *et al.*, 2001 & Makino *et al.*, 2002). Thus HIF-3α was not a target for this study, which solely focuses on HIF-1α and HIF-2α.

1.1.3. HIF-1α versus HIF-2α

Although HIF-1α and HIF-2α are expressed almost ubiquitously by cells *in vivo* (Stroka *et al.*, 2001, Wiesener *et al.*, 2003 & Rosenberger *et al.*, 2002) and have an overall high sequence homology of ~50% (Tian *et al.*, 1997), their functional characteristics and downstream roles are different. For example, their target genes regulated, their O_2 sensitivity and their sub-nuclear localisation are different between the two HIFα isoforms (discussed below).

1.1.3.1. Target genes

Many studies have determined the promoter region where active HIF dimers bind, irrespective of HIFα isoform, as 5'-RCGTG-3, termed the Hypoxia responsive element (HRE, Wenger *et al.*, 2005 & Schödel *et al.*, 2011). The HRE is highly abundant within the human genome, however HIF-1α and HIF-2α occupy <1% (~500) of the total HREs available in response to hypoxia (Mole *et al.*, 2009). Significantly, <20% of HREs identified can be occupied by both HIFα isoforms, thus many HREs have HIF-1α or HIF-2α exclusivity, suggesting differential roles for HIF-1α and HIF-2α in response to hypoxia (Mole *et al.*, 2009). This conclusion is further supported by additional genome wide Chromatin

Immunoprecipitation assays (Schödel *et al.*, 2011 & Smythies *et al.*, 2019), microarray analysis (Hu *et al.*, 2003) and biochemical approaches (Raval *et al.*, 2005 & Lau *et al.*, 2007) which all identified HIF-1α and HIF-2α specific target gene regulation. For example, HIF-1α specifically regulates genes encoding enzymes of the glycolytic pathway (Hu *et al.*, 2003), while HIF-2α specifically regulates the Transforming growth factor alpha gene (TGFα, Raval *et al.*, 2005), yet both HIFα isoforms regulate genes encoding proteins involved with angiogenesis (Hu *et al.*, 2003). Recently, the gene specificity of HIFα isoforms has been shown to be due to specific binding locations of HIFα in respect to the hypoxia regulated gene, with HIF-1α predominately binding at HREs in close proximity to gene promoters while HIF-2α binds predominately at distant gene enhancer HREs (Smythies *et al.*, 2019).

1.1.3.2. Stability

Since its discovery, HIF-1α protein stability has been known to be O_2 dependent, only stabilising at 1% O_2 (Wang *et al.*, 1995). HIF-1 α is rapidly degraded upon reoxygenation, with a half-life of ~5 min (Moroz *et al.*, 2009). Although many published data show a similar O_2 dependent stability profile for HIF-2α, there is growing evidence that HIF-2α stability is less sensitive to O_2 tensions than HIF-1α. Many published data show that HIF-2α is strongly stabilised at 2-5% O_2 (Holmquist-Mengelbier *et al.*, 2006 & Nilsson *et al.*, 2005) and more recent data report that HIF-2α is actually stable at atmospheric O_2 (21%), with only marginal increase in stability at 1% O_2 (Hara *et al.*, 1999, Fujita *et al.*, 2012, Uchida *et al.*, 2004, Xie *et al.*, 2018, Bagnall *et al.*, 2014 & Taylor *et al.*, 2016). As suggested by these publications, HIF-2α protein stability is probably cell-line dependent. This is supported by Bracken *et al.*, 2006, who demonstrated that HIF- 1α and HIF-2α protein stability and transcriptional function vary dramatically by cell line.

1.1.3.3. Nuclear localisation

HIF-1α appears as a homogenous distribution throughout the nucleus, while HIF-2α localises to distinct punctate during hypoxia (Hara *et al.*, 1999). Later studies even suggest that HIF-2α is physically trapped within these fine punctate, potentially by nuclear tethering (Taylor *et al.*, 2016, Figure 1.7).

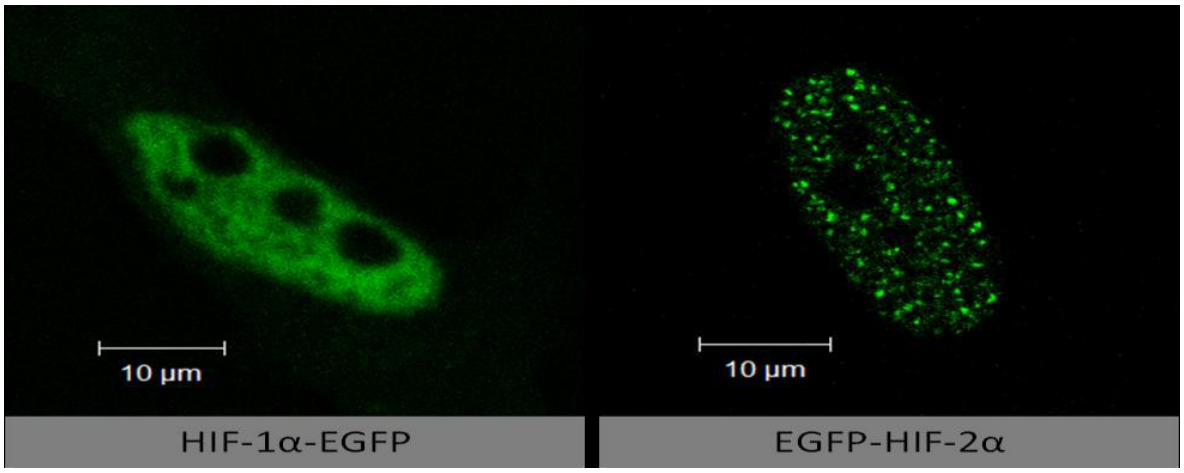


Figure 1.7: The nuclear localisation of HIF-1α and HIF-2α tagged with EGFP.
HIF-1α homogenously distributed throughout the nucleus, while HIF-2α is non-homogenously distributed as intense punctate (taken from Taylor *et al.*, 2016).

Although the exact role played by sub-nuclear localisation in regulation/differentiating HIF-1α and HIF-2α is not fully determined, a recent study has shown that transcription factors can form concentrated hubs via, which have the potential to phase-separate at higher concentrations (Chong *et al.*, 2018). In this study the authors postulated that the weak, dynamic, and transient contacts between transcription factors within the hubs, play a role in disease-causing dysregulation of gene expression. In the HIF context, they might underpin some target gene expression regulatory differences.

1.1.4. HIFα domain function

As stated, both HIF-1α and HIF-2α can be compartmentalised into five distinct functional domains: bHLH, PAS, ODDD, NTAD and CTAD (Figure 1.6). Sequence alignment comparisons between HIF-1α and HIF-2α (Tian *et al.*, 1997) identified that the bHLH and PAS domains share the highest degree of sequence homology of all domains at 85% and 70% respectively. This is unsurprising considering their importance for DNA binding and HIF-1β binding. Tian *et al.*, 1997 also identified that the CTAD also has significant sequence homology at ~70%, while the ODDD and NTAD have minimal sequence homology (~40% and ~20% respectively, Figure 1.6). Later studies characterising the domains of HIFα proteins identified that the CTAD is required for binding to the histone acetyltransferase proteins CREB binding protein (CBP) and p300, the CBP/p300 complex (Ema *et al.*, 1999). The HIF-p300/CBP binding axis is well known to be essential for the HIF dependent transcriptional response to hypoxia, likely by chromosomal remodelling (Ema *et al.*, 1999 & Kallio *et al.*, 1998). The ODDD between HIFα isoforms shares ~40% sequence homology, however it appears to have a similar role in the O_2 dependent degradation of both HIF-1α

and HIF-2α (discussed further in the O_2 dependent regulation section 1.6). The NTAD shares the least sequence homology between HIFα isoforms and was initially thought to explain the gene targeting differences observed by HIF-1α and HIF-2α (Hu *et al.*, 2003). In 2007, the NTAD was proved to control gene specific targeting of HIFα isoforms. Using mutational analysis (where the HIF-2α NTAD was inserted into HIF-1α, and vice versa), it was demonstrated that mutant HIF-1α could express HIF-2α specific genes, and vice versa (Hu *et al.*, 2007 & Lau *et al.*, 2007). It was also found that secondary, isoform specific, cofactor binding partners bound within the NTAD and were responsible for isoform specific gene regulation; including ETS like protein 1 (ELK) for HIF-2α (Hu *et al.*, 2007).

1.6. O_2 dependent regulation

The O_2-dependent regulation mechanism is shown in Figure 1.8. A class of enzymes, the Prolyl Hydroxylases (PHDs), encoded by the EGLN genes, utilise molecular O_2 and iron as cofactors to result in the proline hydroxylation of HIFα subunits (Epstein *et al.*, 2001 & Bruick *et al.*, 2001). Proline hydroxylation occurs within the ODDD at the specific residues: HIF-1α P402/P564 and HIF-2α P405/531 (Jaakkola *et al.*, 2001, Ivan *et al.*, 2001 & Masson *et al.*, 2001). HIFα Proline hydroxylation acts as a recognition site for binding of the Von Hippel-Lindau E3 ubiquitin ligase complex (pVHL), resulting in the poly-ubiquitination of HIFα and subsequent protein degradation through the 26S proteasome (Ivan *et al.*, 2001, Jaakkola *et al.*, 2001 & Masson *et al.*, 2001).

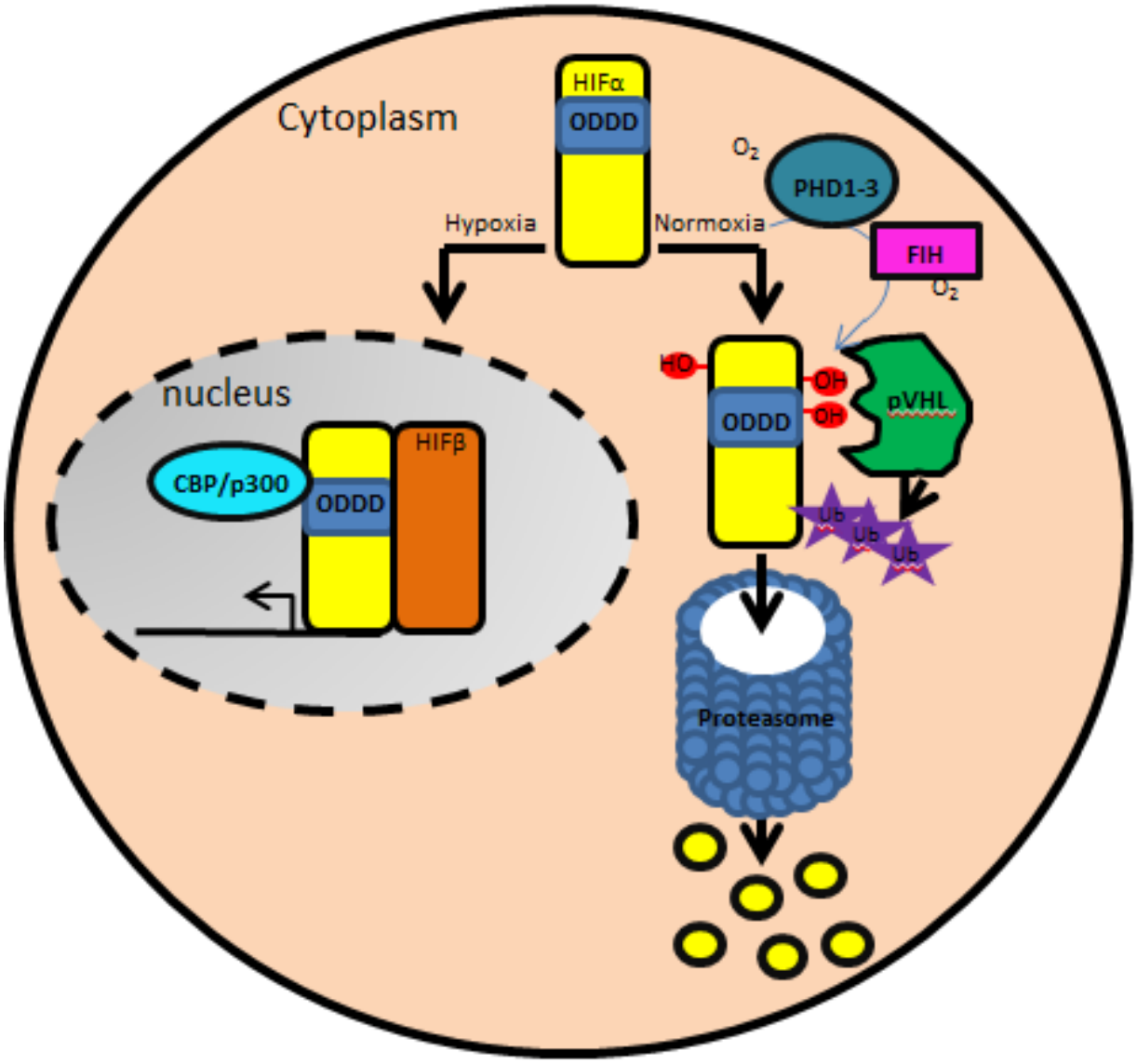


Figure 1.8: The canonical oxygen dependent regulation pathway of HIFα during hypoxia and normoxia. In normoxia (right hand side), available O_2 is used as a cofactor for PHDs to hydroxylate HIFα proline residues (HIF-1α P402/P564 and HIF-2α P405/P531), resulting in the recognition and poly-ubiquitination by VHL and subsequent degradation by the 26S proteasome. FIH also hydroxylates HIFα on an Asparagine residue (HIF-1α N803 and HIF-2α N847) causing the inhibition of essential interactions with CBP/p300. During hypoxia (left hand side), a lack of O_2 inhibits both PHDs and FIH, thus resulting in the accumulation and translocation of HIFα into the nucleus where dimerization with HIF-1β and complexing with the essential co-factors CBP/p300 occur; resulting in transcriptional regulation.

As shown in Figure 1.8, a second O_2 dependent regulatory pathway has also been discovered. Factor Inhibiting HIF-1 (FIH-1) is an asparaginyl hydroxylase protein, using molecular O_2 to hydroxylate a single asparagine residue within the CTAD of both HIF-1α (N803) and HIF-2α (N847). FIH-1 maintains function at much lower O_2 tensions than PHDs, as low as 0.5% O_2, and is responsible for abolishing transcriptional activity independently of protein stability by blocking the essential cofactor interactions with CBP/p300 (Lando *et al.*, 2002, Lando *et al.*, 2002 & Mahon *et al.*, 2001).

Overall, in normoxia HIFα proteins are rapidly degraded and transcriptionally inactivated by hydroxylation through the function of PHDs and FIH-1, both requiring O_2 as a cofactor. Thus, when there is a lack of O_2, the activity of PHDs and FIH-1 become greatly reduced; allowing the accumulation and translocation of HIFα into the nucleus, where the active HIF transcription factor forms and regulates gene expression.

1.7. O_2 independent regulation

Many PTMs do not require O_2 directly as a co-factor, as is the case for hydroxylation. For example, phosphorylation, acetylation, methylation, ubiquitination and SUMOylation are all O_2 independent. Thus, many PTMs can regulate protein function independently of O_2 tension. These PTMs have all been shown to regulate HIFα proteins in hypoxia, or hypoxia mimicking conditions, by 3 mechanisms: 1) PTM occurs directly on the HIFα proteins (discussed below). 2) PTM occurs on a secondary binding partner, thus indirectly regulating HIFα proteins (Kewley *et al.*, 2005). 3) PTM occurs on proteins that do not interact with HIFα, yet result in the elevated translation rate of HIFα proteins (Page *et al.*, 2002).

Because of the important association of HIF with cancer progression, many different pathways have been identified as essential regulators of the HIFα proteins, such as the MAPK cascade and PI3K-mTOR pathway. However, importantly, the exact mechanisms by which these pathways regulate HIF is unclear, with many pathways being found to regulate HIFα proteins both directly and indirectly (Reviewed by: Kietzmann *et al.*, 2016, Dengler *et al.*, 2014, Lee *et al.*, 2004 & Semenza, 2003). Additionally, many of the studies that show direct regulation of HIFα by PTMs fail to identify the exact site of modification, an essential aspect of functional outcomes, but rather narrow PTM location to a specific domain. Therefore, only PTMs that have been localised to a specific residues of HIF-1α or HIF-2α are discussed here. A summary of the modifications previously described is depicted in Figure 1.9 and listed in Table 1.2 & Table 1.3 for HIF-1α and HIF-2α respectively.

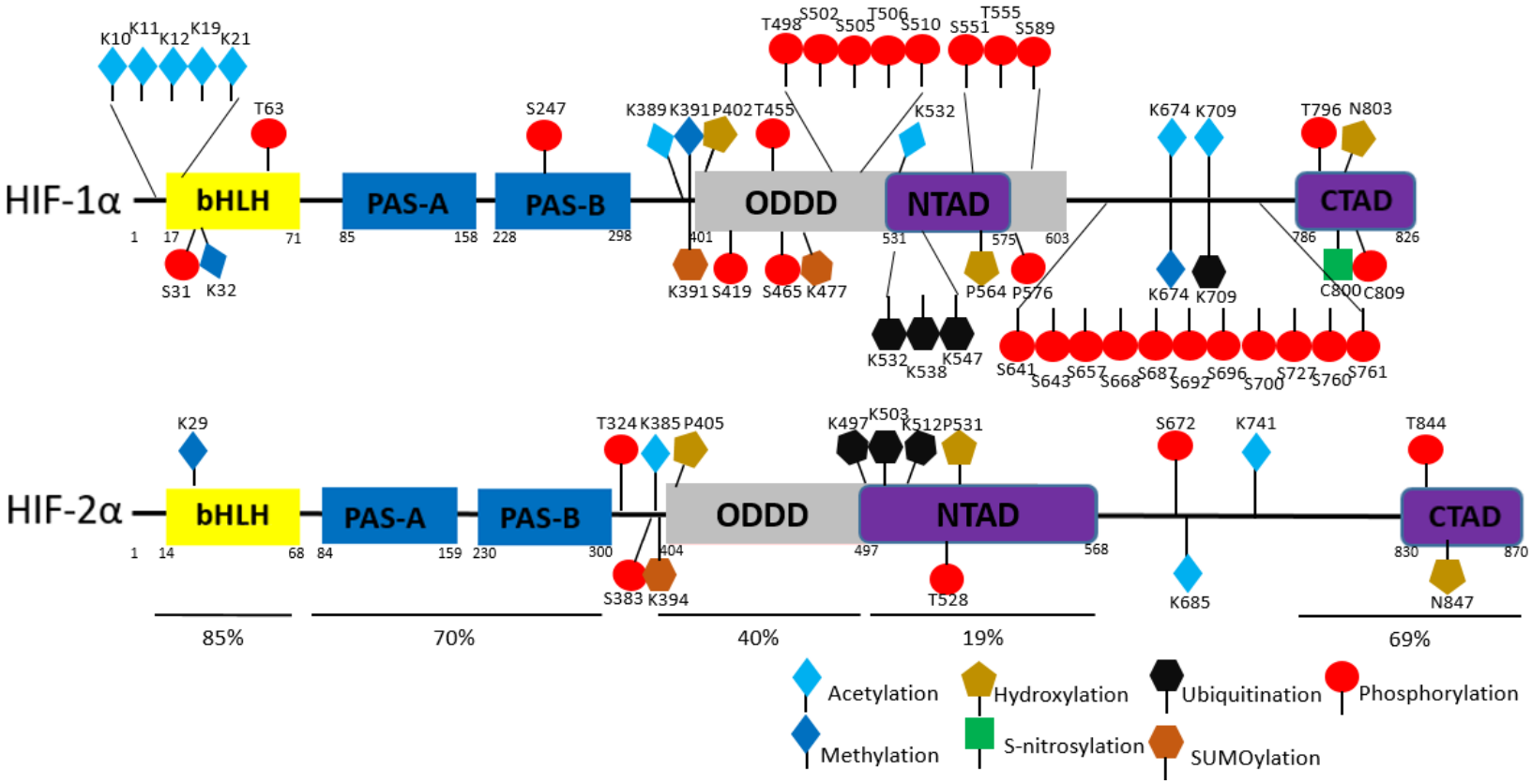


Figure 1.9: Schematic depiction of all known (to date), site specific PTMs that regulate HIF-1α or HIF-2α.
PTMs include type and position within the protein sequence. Sequence similarity is depicted below

Table 1.2: Post translational modifications of HIF-1α.

Site uses the 1 letter amino acid code to identify amino acid modified, numbering system from human protein unless stated. Table includes the PTM type, modifying enzyme and effect (if known). – indicates that the authors did not provide this information. Table also includes the techniques used to identify PTM site and functional effects of PTM. References for sites are included.

Site	Modification	Enzyme	Effect	Technique used for PTM localisation	Technique used for PTM characterisation	Reference
K10/11/12/19/21	Acetylation	-	Degradation	bHLH random mutagenesis	Full length, multi-site acetylation-null mutation	Geng *et al.* , 2011
S31	Phosphorylation	PKA	-	*In vitro* assay & recombinant fragments	Fragment, phopsho-null mutation	Bullen *et al.* , 2016
K32	Methylation	SET7/9	Inactivation	*In vitro* assay & recombinant fragments	Methyl-transferase overexpression, endogenous	Liu *et al.* , 2015
K32	Methylation	SET7/9	Degradation	Endogenous IP	Mehtyl-transferase overexpression, endogenous	Kim *et al.* , 2016
T63	Phosphorylation	PKA	Stabilisation	*In vitro* assay & recombinant fragments	Fragment, phopsho-null mutation	Bullen *et al.* , 2016
S247	Phosphorylation	CK1δ	Inhibition	*In vitro* assay & recombinant fragments	Fragment, phopsho-mimetic mutation	Kalousi *et al.* , 2010
K389	Acetylation	PCAF	Transactivation	*In vitro* assay & recombinant fragments	Deacetylase overexpression, endogenous	Xenaki *et al.* , 2008
K391	Methylation	Set9	Degradation	*In vitro* assay & recombinant full length	Full length, methylation-null mutation	Lee *et al.* , 2017
K391	SUMOlyation	Ubc9/RanBP2	Stabilisation	*In vitro* assay & recombinant full length	Fragment, SUMOylation-null mutation	Bae *et al.* , 2004
K391	SUMOlyation	Ubc9/RanBP2	Inhibition	*In vitro* assay & *In vitro* fragment	Full length, SUMOylation-null mutation	Berta *et al.* , 2007
K391	SUMOlyation	Ubc9/RanBP2	Degradation	Motif analysis	Full length, SUMOylation-null mutation	Cheng *et al.* , 2007 & Kang et al., 2010
S419 (*C. elegans*)	Phosphorylation	AMPK	Degradation	*In vitro* assay & recombinant fragments	AMPK gene deletion	Hwang *et al.* , 2014
T455	Phosphorylation	PKA	-	*In vitro* assay & recombinant fragments	Fragment, phopsho-null mutation	Bullen *et al.* , 2016
S465	Phosphorylation	PKA	-	*In vitro* assay & recombinant fragments	Fragment, phopsho-null mutation	Bullen *et al.* , 2016
K477	SUMOylation-1/2	Ubc9/RanBP2	Stabilisation	*In vitro* assay & recombinant full length	Fragment, SUMOylation-null mutation	Bae *et al.* , 2004
K477	SUMOylation-1/2	Ubc9/RanBP2	Inhibition	*In vitro* assay & *In vitro* fragment	Full length, SUMOylation-null mutation	Berta *et al.* , 2007
K477	SUMOylation-1/2	Ubc9/RanBP2	Degradation	Motif analysis	Full length, SUMOylation-null mutation	Cheng *et al.* , 2007 & Kang et al., 2010
T498/S502/S505/T506/S510	Phosphorylation	GSK3β	Degradation	Motif analysis	Full length, multi-site phospho-null mutation	Cassavaugh *et al.* , 2011
K532	Acetylation	ARD1	Degradation	*In vitro* assay & recombinant fragments	Fragment, acetylation-null mutation	Jeong *et al.* , 2002
K532/538/547	Ubiquitination	VHL	Degradation	ODDD random mutagenesis	Fragment, ubiquitination-null mutation	Tanimoto *et al.* , 2000
K532/538/547	Ubiquitination	VHL	Degradation	ODDD random mutagenesis	Fragment, ubiquitination-null mutation	Paltoglou & Roberts, 2007
S551/T555/S589	Phosphorylation	GSK3β	Degradation	Motif analysis	Full length, multi-site phospho-null mutation	Flügel et al., 2007 & Cassavaugh et al., 2011
S576	phosphorylation	PLK3	Degradation	*In vitro* assay & recombinant fragments	Full length, phospho-null mutation	Xu *et al.* , 2010
S641	Phosphorylation	ERK1/2	Nuclear localisation	*In vitro* assay & recombinant full length	Full length, phospho -null & -mimetic mutation	Mylonis *et al.* , 2006 & Mylonis *et al.* , 2008
S643	Phosphorylation	ERK1/2	Nuclear localisation	*In vitro* assay & recombinant full length	Full length, phospho -null & -mimetic mutation	Mylonis *et al.* , 2006 & Mylonis *et al.* , 2008
S657	phosphorylation	PLK3	Degradation	*In vitro* assay & recombinant fragments	Full length, phospho-null mutation	Xu *et al.* , 2010
S668	Phosphorylation	CDK1	Stabilisation	*In vitro* assay & recombinant fragments	Full length, phospho -null & -mimetic mutation	Warfel *et al.* , 2013
K674	Acetylation	PCAF	Transactivation	*In vitro* assay & recombinant fragments	Full length, acetylation-null mutation	Lim *et al.* , 2010
K674	Methylation	G9a/GLP	Inhibition	*In vitro* assay & recombinant fragments	Full length, methylation-null mutation	Bao *et al.* , 2018
S687	Phosphorylation	CDK5	Stabilisation	*In vitro* assay & recombinant fragments	Full length, phospho -null & -mimetic mutation	Herzog *et al.* , 2016
S692	Phosphorylation	PKA	Stabilisation	*In vitro* assay & recombinant fragments	Fragment, phopsho-null mutation	Bullen *et al.* , 2016
S696	Phosphorylation	ATM	Stabilisation	*In vitro* assay & recombinant fragments	Full length, phospho-null mutation	Cam *et al.* , 2010
S700	Phosphorylation	PKA	-	*In vitro* assay & recombinant fragments	Fragment, phopsho-null mutation	Bullen *et al.* , 2016
K709	Acetylation	p300	Stabilisation	Co-overexpression, full length IP	Full length, Non-conservative acetylation null mutation	Geng *et al.* , 2012
S727	Phosphorylation	PKA	-	*In vitro* assay & recombinant fragments	Fragment, phopsho-null mutation	Bullen *et al.* , 2016
S760	Phosphorylation	PKA	-	*In vitro* assay & recombinant fragments	Fragment, phopsho-null mutation	Bullen *et al.* , 2016
S761	Phosphorylation	PKA	-	*In vitro* assay & recombinant fragments	Fragment, phopsho-null mutation	Bullen *et al.* , 2016
C800	S-nitrosylation	-	Transactivation	CTAD random mutagenesis	recombinant p300 fragment, *In vitro* IP, HIF-1α fragment S-nitrosylation-null mutation	Yasinska & Sumbayev, 2003
C800	S-nitrosylation	-	Inhibition	CTAD random mutagenesis	recombinant p300 fragment, *In vitro* binding assay, HIF-1α S-nitrosylated & S-nitrosylation-null mutation fragments	Cho *et al.* , 2007
S809	Phosphorylation	PKA	-	*In vitro* assay & recombinant fragments	Fragment, phopsho-null mutation	Bullen *et al.* , 2016

Table 1.3: Post translational modifications of HIF-2α.

Site uses the 1 letter amino acid code to identify amino acid modified, numbering system from human protein. Table include the PTM type, modifying enzyme and effect (if known). – indicates that the authors did not provide this information. Table also includes the techniques used to identify PTM site and functional effects of PTM. References for sites are included.

Site	Modification	Enzyme	Effect	Technique used for PTM localisation	Technique used for PTM characterisation	Reference
K29	Methylation	SET7/9	Degradation	*In vitro* assay & recombinant fragments	Methyl-transferase overexpression, endogenous	Liu *et al.*, 2015
T324	Phosphorylation	PKD1	Transactivation	*In vitro* assay & recombinant fragments	Fragment, phospho -null & -mimetic mutation	To *et al.*, 2006
S383	Phosphorylation	CK1δ	Nuclear localisation	*In vitro* assay & recombinant fragments	Full length, phospho-null mutation	Pangou *et al.*, 2016
K385/685/741	Acetylation	-	Inhibition	Full length, O_2 insensitive IP	Full length, multi-site acetylation-null mutation	Dioum *et al.*, 2009
K394	SUMOylation	SUMO-2	Degradation	Motif analysis	Full length, SUMOtylation-null mutation	van Hagen *et al.*, 2010
K497/503/512	Ubiquitination	VHL	Degradation	HIF-1α alignment	Fragment, ubiquitination-null mutation	Paltoglou & Roberts, 2007
T528	Phosphorylation	CK1δ	Nuclear localisation	*In vitro* assay & recombinant fragments	Full length, phospho-null mutation	Pangou *et al.*, 2016
S672	Phosphorylation	ERK1/2	Nuclear localisation	*In vitro* assay & recombinant fragments	Full length, phospho-mimetic mutation	Gkotinakou *et al.*, 2019
T844	Phosphorylation	-	Transactivation	CTAD random mutagenesis	*In vitro* translated, full length, phospho-null mutation	Gradin *et al.*, 2002

It is clear from Figure 1.9 and Table 1.2 & Table 1.3 that many more PTMs have been identified of HIF-1α than HIF-2α, possibly because of its earlier discovery. The functional consequences of these PTMs are discussed below.

1.1.5. HIFα phosphorylation

Phosphorylation has been shown to regulate both HIFα isoforms, in terms of protein stability, transactivation and nuclear localisation. A total of 28 different phosphorylation sites have been identified for HIF-1α, although not all are considered to have functional effects, and 5 different sites for HIF-2α.

1.1.5.1. Stability

The phosphorylation status of HIF-1α has been shown to have large effects on protein stability, both positively and negatively, although no phosphorylation mediated stability roles have yet been defined for HIF-2α. Two separate studies have identified Glycogen Synthase Kinase 3β (GSK3β) phosphorylation at HIF-1α sites: T498, S502, S505, T506, S510, S551, T555 and S589, resulting in protein degradation (Flügel *et al.*, 2007 & Cassavaugh *et al.*, 2011). Infact, mutation of these GSK3β phosphorylation sites to phospho-null alanine residues increased protein stability >10 fold in normoxia (Flügel *et al.*, 2007 & Cassavaugh *et al.*, 2011), highlighting the importance that PTMs can play. Polo-like Kinase 3 (Plk3) has also been demonstrated to have a role in the degradation of HIF-1α, by phosphorylation at S576 and S657. In *C. elegans,* the redox sensitive AMP-activated protein kinase (AMPK) was shown to phosphorylate HIF-1α S419 leading to its degradation (Hwang *et al.*, 2014). However, human cell line studies conflict this and suggest AMPK induces HIF-1α protein stability, mediated through JNK signalling pathways (Jung *et al.*, 2008).

Phosphorylation at different sites of HIF-1α has been demonstrated to have stabilising roles. Bullen *et al.*, 2016 identified that Protein Kinase A (PKA) could multiply phosphorylate HIF-1α at 10 different sites *in vitro,* but found that only T63 and S692 phosphorylation increased protein stability. A role for Cyclin Dependent Kinases (CDK) in promoting HIF-1α stability has also been identified. CDK1 was shown to phosphorylate HIF-1α at S668, and CDK5 at S687, both resulting in increased protein stability. Cam *et al.*, 2010 also report that Ataxia telangiectasia Mutated (ATM) can phosphorylate HIF-1α at S696 to result in increased protein stability.

1.1.5.2. Transactivation

There is a single reported case of phosphorylation-induced global transcriptional inactivation for HIF-1α and HIF-2α, S247 (Kalousi *et al.*, 2010) and T844 (Gradin *et al.*, 2002) respectively. Casein Kinase 1δ (CK1δ) was shown to phosphorylate HIF-1α at S247, preventing the formation of HIFαβ dimers and thus inhibiting transcriptional roles without affecting protein stability (Kalousi *et al.*, 2010). Gradin *et al.*, 2002 did not identify a kinase that causes the phosphorylation, but rather mutated serine and threonine residues within the CTAD to determine a potential role for phosphorylation. However, phospho-null mutants of T844 to alanine resulted in the inability to bind p300/CBP thereby inhibiting transcription.

Phosphorylation has also been demonstrated to have a role in differentiating HIF-1α and HIF-2α binding partners for gene specificity (To *et al.*, 2006). To *et al.*, 2006 identified that HIF-2α T324 can be phosphorylated by Protein Kinase D1 (PKD1), which prevents the SP1 transcription factor from binding. HIF-1α contains a sequence variation that prevents PKD1 phosphorylation, thus SP1 can bind and mediate a specific set of gene targets.

1.1.5.3. Nuclear localisation

Phosphorylation has been shown to regulate the nuclear localisation of both HIF-1α and HIF-2α proteins, without affecting protein stability. Extracellular signal Regulated protein kinases 1 & 2 (ERK1/2) have been demonstrated to phosphorylate both HIF-1α (at S641 and S643) and HIF-2α (at S672) (Mylonis *et al.*, 2006, Mylonis *et al.*, 2008 & Gkotinakou *et al.*, 2019). Phosphorylation at these sites was shown to promote nuclear accumulation, while phospho-null mutations were exclusively cytoplasmic, this was identified as the result of preventing interactions with the nuclear exporter protein CRM-1. Additionally, CK1δ phosphorylation of HIF-2α S383 and T528 results in nuclear accumulation, similarly by blocking CRM-1 mediated nuclear export (Pangou *et al.*, 2016).

1.1.6. HIFα acetylation

Acetylation has been shown to regulate both HIFα isoforms, regulating protein stability and transactivation. Although acetylation status of HIFα is much less studied compared to phosphorylation, current theories suggest a functional role in terms of stability and transactivation.

1.1.6.1. Stability

The first reported acetyl-transferase to have function in regulating HIF-1α was Arrest Defective Protein 1 (ARD-1, an acetyl transferase protein), which resulted in the acetylation of K532 (Jeong *et al.*, 2002). It was demonstrated that K532 acetylation was essential for the O_2 dependent, VHL mediated degradation post proline hydroxylation, and that acetylation-null mutations to arginine residues were stable in normoxia (Jeong *et al.*, 2002 & Lee *et al.*, 2010). However, two separate studies (albeit in different cell lines, Murray-Rust *et al.*, 2006 & Arnesen *et al.*, 2005) have failed to identify ARD-1 mediated HIF-1α acetylation. This therefore could suggest that acetylation mediated degradation may be cell line dependent.

Geng *et al.*, 2012 identified that p300 dependent acetylation of HIF-1α at K709 results in increased protein stability. This study also identified reduced rates of ubiquitination by less-conservative acetylation-null K709 mutations to alanine, suggesting potential PTM competition where acetylation prevents ubiquitination mediated protein degradation. A previous study by the same group used multi-site acetylation-null mutations of the lysine residues in the bHLH domain (K10, K11, K12, K19 and K21) to demonstrate acetylation induced HIF-1α degradation (Geng *et al.*, 2011). The methodology employed by Geng *et al.*, 2011 did not identify an acetyl-transferase protein responsible for the modification nor the role of individual sites, hence it is possible only 1 site requires acetylation for degradation. HIF-2α has not yet been identified to be regulated at the protein stability level by acetylation.

1.1.6.2. Transactivation

Acetylation has also been reported to affect HIF-1α and HIF-2α transactivation, by both global and specific gene targeting mechanisms. p300/CBP associated factor (PCAF) was demonstrated to acetylate HIF-1α at 2 separate sites to regulate transactivation. Lim *et al.*, 2010 identified K674 as an acetylation site that results in the global upregulation of HIF-1α target genes, however did not identify an explanation for this observation. While Xenaki *et al.*, 2008 demonstrated that PCAF acetylation at K389 resulted in the differential regulation of specific HIF-1α target genes, such as the upregulation of Carbonic anhydrase-IX and downregulation of Lactate dehydrogenase, highlighting a gene specific targeting mechanism of acetylation. A single publication has shown the role of acetylation on HIF-2α activity. Dioum *et al.*, 2009 identified the acetylation sites K385, K685 and K741 on O_2 stable HIF-2α (P405A/P531A), and used multi-site acetylation-null mutation analysis to

demonstrate HIF-2α acetylation increased transcriptional function without affecting protein stability.

1.1.7. HIFα methylation

Most research on methylation as a PTM has been conducted on DNA bound proteins such as histones, where it is known to regulate secondary protein binding by either acting as a docking site (to promote) or blocking agent (to prevent) protein interactions (reviewed by Lanouette *et al.*, 2014 & Blanc *et al.*, 2017). More recently, methylation has been identified as a key PTM in regulating protein function in signal transduction pathways, including the EGFR – ERK pathway (Hsu *et al.*, 2011), and for regulating transcription factor function, such as p53 (Huang *et al.*, 2007). Methylation has also been shown to regulate both protein stability and transactivation of HIF-1α and HIF-2α.

1.1.7.1. Stability

Methylation has been suggested to be an important PTM for regulating HIF-1α protein stability. Lee *et al.*, 2017 demonstrated that SET9 could methylate HIF-1α K391 and was required for the efficient normoxic proline hydroxylation by PHDs, with methylation null mutations to arginine being stable in normoxia. Thus, Lee *et al.*, 2017 suggest that K391 methylation may be an important pre-requisite for the O_2-dependent PHD mediated degradation pathway.

HIF-1α methylation at K32 has also been identified by 2 separate groups, however have conflicting evidence regarding the functional outcomes (Kim *et al.*, 2016 & Liu *et al.*, 2015). Kim *et al.*, 2016 shows that the overexpression of a methyl-transferase protein (SET7/9) results in the increased degradation rate of HIF-1α, while Liu *et al.*, 2015 use the same assay to report transcriptional inhibition independent of HIF-1α protein degradation. Liu *et al.*, 2015 also identified that the respective HIF-2α site (K29) is also a methylation site with identical functional outcomes.

1.1.7.2. Transactivation

Bao *et al.*, 2018 identified K674 as a target for mono- and di- methylation by Histone-lysine N-methyltransferase EHMT2 (G9a) and G9a Like protein (GLP). K674 methylation was demonstrated to inhibit HIF-1α dependent transcription, independent of protein stability. This observation is particularly interesting considering acetylation at K674 was shown to

have identical regulatory roles (Lim *et al.*, 2010), potentially suggesting a mechanism of regulation that is more dependent on charge neutralisation of K674 than the PTM type.

1.1.8. HIFα SUMOylation

The role of SUMOylation in HIF-1α dependent signalling is controversial, multiple studies all identify the identical SUMOylation sites yet report a different functional outcome by PTM. Initially, Bae *et al.*, 2004 identified HIF-1α K391 and K477 as SUMOylation sites for the SUMO-ligase Ubc9, and demonstrated that SUMOylation induced protein stability. However, a second independent group identified PIASy as the SUMO-ligase of HIF-1α K391 and K477 sites, and demonstrated SUMOylation induced degradation (Cheng *et al.*, 2007 & Kang *et al.*, 2010). While a third independent group identify RanBP2/NUP538 as the SUMO-ligase responsible for modifying HIF-1α K391 and K477 sites and could not identify a protein stability role of SUMOylation, yet showed transcriptional inactivation by SUMOylation. Again, the conflicting reports for SUMOylation of the same sites, might be due to the different cell lines used, but also to the use of different SUMO-ligases and HIF-1α recombinant fragments. A single publication has reported SUMOylation of HIF-2α at K394 (van Hagen *et al.*, 2010). van Hagen *et al.*, 2010 identified SUMOylation motifs with HIF-2α and used SUMOylation-null arginine mutations to show that K394 SUMOylation induced protein degradation.

1.1.9. HIFα Ubiquitination

As described earlier, the canonical pathway of O_2 dependent degradation involves the hydroxylation of proline residues followed by poly-ubiquitination, resulting in rapid proteasomal degradation (Ivan *et al.*, 2001, Jaakkola *et al.*, 2001, Masson *et al.*, 2001). Tanimoto *et al.*, 2000 sequentially mutated all K residues within the HIF-1α ODDD to identify K532 as the major site responsible for O_2 dependent ubiquitination and degradation. Using similar techniques, Paltoglou *et al.*, 2007 identified that HIF-1α K532, K538 and K542 (and respective HIF-2α sites K497, K503 and K512) could also be ubiquitinated in normoxic conditions. Because my work focuses on the role of specific sites of PTM, only ubiquitination that has been associated to specific HIFα residues are discussed. However it is important to mention that the ubiquitination status of HIFα is a highly studied field and has been shown to be a much more complex regulatory system than only proline hydroxylation based degradation by involvement of different ubiquitin ligases and de-ubiquitinating enzymes (reviewed by Schober *et al.*, 2016). One interesting

example of ubiquitination is a possible role in switching between HIF-1α to HIF-2α dependent signalling in prolonged hypoxia. Koh *et al.*, 2008 identified a novel E3 ubiquitin ligase termed Hypoxia Associated Factor (HAF) which specifically poly-ubiquitinates HIF-1α only, resulting in the rapid degradation by the 26S proteasome. Later studies from the same group found that HAF could bind to HIF-2α at the CTAD, promoting transcriptional function of HIF-2α target genes by promoting DNA binding at HREs (Koh *et al.*, 2011 & Koh *et al.*, 2014).

1.1.10. HIFα nitrosylation

All modifications discussed so far were enzyme driven. However, PTMs can occur through non-enzyme mediated reactions. For example, the production of reactive oxygen, or nitrogen, species (ROS/RNS) can react with free thiol groups of cysteine residues, resulting in multiple different cysteine PTM modifications that are both reversible and irreversible (Reviewed by Hess *et al.*, 2005 & Chung *et al.*, 2013). ROS/RNS production occurs endogenously through general mitochondrial function and cell signalling pathways, and is greatly increased in response to hypoxia (Chandel *et al.*, 1998 & Bell *et al.*, 2007). Nitric Oxide (NO) is a RNS that has been shown to modify HIF-1α C800, with contradictory reports regarding its function. Initially, Yasinska *et al.*, 2003 has shown that S-nitrosylation of C800, by chemical or signalling meditated strategies, resulted in a HIF-1α CTAD mutated fragment (N803A, to prevent asparagine hydroxylation) binding strongly to p300/CBP. Whereas, a later study by Cho *et al.*, 2007 used a non-mutated variant of the same CTAD fragment to show that S-nitrosylation-null mutants to alanine bound stronger to p300/CBP, the opposite effect of the previous study.

1.8. Limitations of previous studies

HIF-1α, and to a lesser extent HIF-2α, have been shown to be regulated by various PTMs, impacting either protein stability, transactivation or nuclear localisation. An important consideration is that many of these studies relied on the use of recombinant fragment based approaches coupled with *in vitro* assays to identify PTMs (Table 1.2 & Table 1.3). These strategies however have multiple intrinsic issues: 1) recombinant proteins, particularly small fragments, may not fold correctly. 2) A reported PTM site discovered on a small fragment, may be buried within the full length protein, and hence physically blocked by internal stoichiometric mechanisms. 3) Proteins tend to form large protein complexes for regulation, a fragment may lack many protein interactions in neighbouring regions of

the protein which could block the reported PTM site 4) The ratios used by *in vitro* assays might force interactions, resulting in modification that would not otherwise be observed *in cellulo*. Overall, the false discovery (and false negative) rate of these techniques can be potentially quite high and may not reflect true *in vivo* regulatory mechanisms.

Additionally, for hypoxic conditions, studies generally used hypoxia mimicking drugs such as DMOG (Dimethyloxalylglycine), DFO (Deferoxamine), DFX (Deferasirox) and $CoCl_2$ (cobalt chloride), which are PHD inhibitors (directly, by iron chelation or iron replacement). Therefore, hypoxia mimicking drugs do not necessarily reflect the true effects observed by a lack of O_2. For example, DFO based iron chelation, used to increase HIFα stability, has additionally been shown to induce cell cycle arrest at the G1/S boundary by reducing the protein levels of cyclin D1 and p21; by inhibiting their mRNA cytoplasmic translocation and increased non-ubiquitin dependent proteasomal degradation (Fu *et al.*, 2007 & Nurtjahja-Tjendraputra *et al.*, 2007).

Furthermore, the use of multiple different cancer cell lines, many with important regulatory proteins and modifying proteins deleted or mutated, may explain some of the observed discrepancies (Bracken *et al.*, 2006). This is particularly evident for HIF-1α K532 acetylation (Jeong *et al.*, 2002, Lee *et al.*, 2010, Arnesen *et al.*, 2005 & Murray-Rust *et al.*, 2006) and HIF-1α K391 and K477 SUMOylation (Bae *et al.*, 2004, Berta *et al.*, 2007, Cheng *et al.*, 2007 & Kang *et al.*, 2010).

Finally, all these studies were performed in a targeted manner, aimed at identifying a single PTM through *in vitro* assays, motif analysis and PTM-null mutations. Thus, these studies lack any perspective into how multiple PTMs may affect overall protein function inside cells, a potentially important aspect for fine tuning regulatory responses (Mann *et al.*, 2003 & Lanucara *et al.*, 2013). Unbiased PTM discovery is therefore critical to understand how the PTM map of HIF-1α and HIF-2α changes in response to true O_2 deprivation. Such investigation is essential for the discovery of novel regulatory mechanisms and to unravel the signalling strategies that coordinate the hypoxic response. One method to explore the PTM modification status of proteins in an unbiased manner is to use a proteomics approach.

1.9. Proteomics

The term proteome was used originally to describe all expressed proteins within a cell at a given time. However, it has evolved to also include all proteoforms (differentially PTM forms of the same protein), a reflection that the functional aspect of proteins are orders of magnitude more complicated than simply the complement of expressed genes (Tyers *et al.*, 2003), as exemplified with HIFα PTMs discussed above. Many more PTMs exist than have been discussed above in the context of HIFα, including glycosylation, sulphation and acylation (Mann *et al.*, 2003). Additionally, PTMs can occur in a combinatorial fashion which may regulate protein function differently than either PTM alone (Mann *et al.*, 2003). As many PTMs are reversible, often occurring at sub-stoichiometric levels compared to total protein, this allows the simultaneous co-existence of multiple functionally distinct subsets of the same protein within a cell (Mann *et al.*, 2002); hence can pose a challenge for proteomics analysis. With the advancements in technology over the past 25 years, mass spectrometry (MS) has become an essential component of proteomics analysis. MS can be used not only for protein identification and to determine PTM status of proteins, but also to quantify protein levels, sub-cellular protein localisation, protein interactions and protein structures (Reviewed by Han *et al.*, 2008, Aebersold *et al.*, 2016, Mallick *et al.*, 2010, Schneider *et al.*, 2018 & Lanucara *et al.*, 2014).

1.1.11. Bottom-Up Proteomics

Although MS-based proteomics analysis can be performed on full length proteins (Top-Down proteomics) there are many technical challenges to this, including: 1) A lack of fractionation techniques to decrease the complexity of samples. 2) The inefficiency of protein fragmentation in a gaseous phase, required to identify the protein. 3) Difficulty in the ionisation of proteins (Lanucara *et al.*, 2013 & Zhang *et al.*, 2013). Peptides, made from the proteolysis of proteins, are much smaller and circumvent the problems associated with analysis of intact proteins. Thus a 'Bottom-Up' approach is common, where proteins undergo proteolysis and protein identity is inferred from peptide sequence. Bottom-Up proteomics approaches are also referred to as shotgun proteomics due to its similarity to shotgun genomic sequencing (Lanucara *et al.*, 2013 & Zhang *et al.*, 2013).

For a typical Bottom-Up approach, proteins are extracted from the system of interest, followed by denaturation, reduction of disulphide bonds and alkylation of cysteine residues to prevent reformation of disulphide bonds, before proteolysis. Proteolysis most commonly

uses Trypsin as a protease because it has a highly specific cleavage pattern at the C-terminus of Arginine (R) and Lysine (K) residues (unless the K/R +1 position is proline (P)). Electrospray ionisation (ESI, section 1.1.15) of tryptic digested products thus typically results in doubly protonated peptide ions, which is beneficial in terms of analysing large peptides with mass analysers (section 1.1.16) of smaller *m/z* ratios (Paizs *et al.*, 2005, Zhang *et al.*, 2013 & Michalski *et al.*, 2012).

However, proteolysis exponentially increases sample complexity. Thus, for highly complex samples there can be competitive/preferential ionisation of peptides combined with an inability for the MS to analyse all peptides available at any one time. Hence, low abundance proteins, and different PTM proteoforms, are challenging to analyse without prior fractionation techniques, purification or enrichment strategies being additionally employed (Mann *et al.*, 2002). Consequently, peptides are separated by Reverse-Phase chromatography prior to MS analysis. Immunoprecipitation (IP) of target proteins or PTM enrichment can also aid in analysis.

1.1.12. IP-coupled Proteomics

A cells proteins do not function as separate entities but rather form highly complex, multi-protein complexes that result in protein function (Alberts, 1998), for example the HIF-1α – HIF-1β complex (Wang *et al.*, 1995). Additionally, for a PTM to occur a modifying enzyme must come into contact with the target protein, even if transiently. Thus, protein binding partners result in an additional layer of complexity when attempting to understand cellular regulatory mechanisms, thus is an important aspect to investigate.

A method to investigate binding partners, and simultaneously solve the problem of potential low abundance proteins, is to specifically purify the protein of interest prior to proteomics analysis. The most common method for protein purification is to use an antibody to the target protein, that is bound to a solid phase, thus purifying the protein of interest and, under the correct conditions, any interacting proteins at a specific time (Co-IP) away from the whole cell proteome (Dunham *et al.*, 2012). Thus Co-IP reduces sample complexity, increases the relative abundance of the target protein and allows identification of target protein interactors (Mallick *et al.*, 2010). Although it is possible to identify protein interactors post IP through alternative techniques to proteomics, such as western blotting, these are largely Low Throughput (LTP) and are inherently biased by the requirement of prior knowledge of potential binding partners to investigate. Proteomics analysis provides a

High Throughput (HTP), unbiased identification approach, significantly improving understanding of protein interaction networks (Dunham *et al.*, 2012 & Mallick *et al.*, 2010).

1.1.13. Phospho-Proteomics

As mentioned, the proteome is highly fluidic in nature, in part due to the dynamic nature of PTMs (Mann *et al.*, 2002). Due to the sub-stoichiometric levels of PTMs, and the fact that certain regions of the protein will not be analysable by MS following tryptic digest, it can be difficult to identify and locate the site of a PTM using standard Bottom-Up proteomics approaches, even following IP. Thus, enrichment strategies have been developed that can enrich for a specific PTM in the background of unmodified peptides. Many enrichment strategies for phosphorylation have been developed, including antibody-based techniques and targeting the charge introduced by phosphorylation. IP using anti phospho-tyrosine antibodies is fairly efficient. However phospho -serine and –threonine antibodies are highly dependent on the sequence surrounding the PTM site (Fíla *et al.*, 2012), thus are not particularly useful for the vast majority of the phospho-proteome.

Alternatively, using metal cations bound to a solid matrix, known as Immobilised Metal Ion Affinity Chromatography (IMAC), results in the enrichment of peptides that contain the negatively charged phosphate moiety. Thus, IMAC reduces sample complexity for more in depth analysis of phospho-peptides (Mann *et al.*, 2002 & Rainer *et al.*, 2015). The most commonly used phospho-peptide enrichment method is by Titanium dioxide (TiO_2).

1.1.14. Liquid Chromatography Mass Spectrometry

As mentioned, MS has become an integral part of proteomics, however sample complexity is often too high for in-depth MS analysis alone. To deconvolute sample complexity, a prior Liquid Chromatography (LC) separation step is commonly coupled prior to MS analysis, where peptides are bound to a C18 column and eluted using an increasing acetonitrile concentration gradient, separating peptides based on their hydrophobicity. The LC system is coupled in-line with the ESI source of the mass spectrometer, thus as peptides are eluted off the LC column, they directly enter the mass spectrometer for analysis. The reduced complexity at any given time thus dramatically increases the depth of coverage of MS-based proteomics analysis.

In order to determine peptide primary sequence for protein identification, tandem MS (MS/MS) is generally performed. At any given time, the *m/z* ratio is recorded for all peptide

ions eluting from the LC column in an MS1 scan. Subsequently, the most intense, or desired, ions are sequentially isolated and fragmented. Fragment ions are then analysed by a second round of MS, and the *m/z* ratios of the product ions identified, generating an MS2 spectrum. To prevent the same precursor *m/z* ion from MS1 being selected for multiple rounds of fragmentation and MS2 analysis, MS/MS is generally performed in a data dependent-acquisition mode (DDA) combined with an exclusion time window. Thereby, a 'Top X' approach can be adopted to obtain tandem mass spectra analysis of a defined number of most intense precursor ions from the MS1 scan.

A typical mass spectrometer consists of three essential components: an ionisation source, a mass analyser and a detector, although the exact components will vary depending on the type of instrument. For LC-MS/MS, the ionisation source is coupled to the LC elution and ionises peptides into a gaseous phase to enter the mass spectrometer. A mass analyser manipulates ions according to *m/z* ratios and a detector records the signal of ions when ejected from the mass analyser. Mass spectra are created from a MS scan by plotting detected *m/z* ratio against relative intensity of all ions in each respective MS scan. Tandem mass spectra can thus be used to determine the primary sequence of a peptide, which is searched against a database of all potential proteins in a sample. In this manner, a tandem mass spectrum can be used to identify a peptide and the gene product it was derived from.

1.10. Thermo Orbitrap Fusion Tribrid

The continued development of MS techniques has led to the generation of multiple mass analysers, and peptide ion fragmentation techniques. The most advanced mass spectrometers combine these technologies into hybrid systems to provide extensive flexibility for MS/MS analysis. The Thermo Orbitrap Fusion Tribrid mass spectrometer is an example of this, combining a quadrupole, linear ion trap and Orbitrap mass analyser (Figure 1.10, Senko *et al.*, 2013). Utilising multiple mass analysers, it is possible to perform mass analysis (MS1, MS2 or MS^n) in either the Orbitrap or the Iontrap, thus maximising efficiency. Additionally, this allows MS/MS analysis to be performed in a time-dependent mode, rather than a Top X mode, where a maximal (undefined) number of precursor ions can be analysed in the Orbitrap while product ions are detected in the Iontrap (Senko *et al.*, 2013). The Thermo Orbitrap Fusion Tribrid mass spectrometer also incorporates multiple different fragmentation strategies (see section 1.1.17), thus making its use in proteomics highly advantageous for both in-depth protein identification and phospho-proteome analysis.

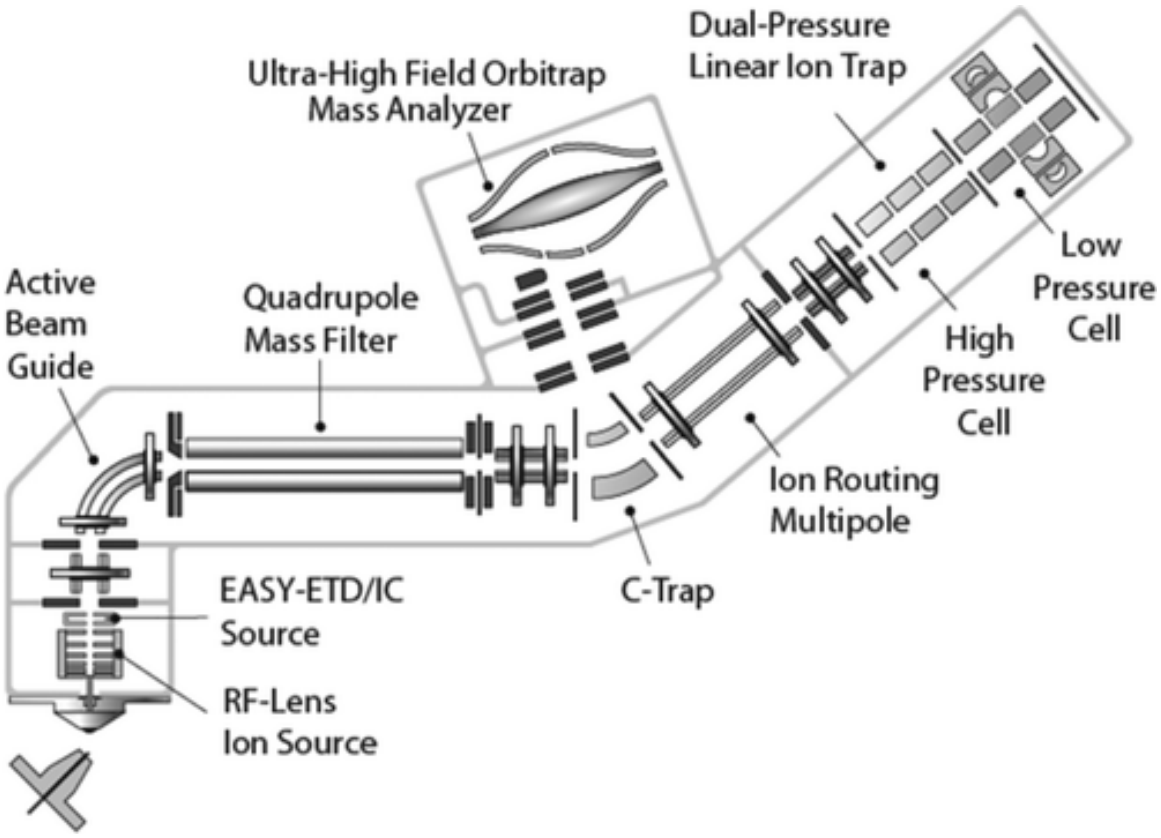


Figure 1.10: Schematic view of the Thermo Orbitrap Fusion Tribrid Mass Spectrometer.
All major components of the MS are labelled, including: ion guides, mass analysers, ETD source and electrospray ionisation source. Taken from Senko *et al.*, 2013.

1.1.15. Electrospray Ionisation

To analyse large biomolecules (peptides), 'soft' ionisation mechanisms are required to ensure that ions produced are not simultaneously fragmented. Electrospray ionisation (ESI) is the most commonly used 'soft' ionisation technique used for proteomics analysis. During ESI, a sample in a volatile solvent (acetonitrile) is passed through a high voltage capillary, generating a highly charged droplet cloud, of equal charge to the capillary (Fenn *et al.*, 1989). Charged droplets are heated to high temperatures (~300 ^{o}C) as they enter a high-to-low pressure gradient, resulting in the desolvation of droplets and creating multiply charged gaseous ions, which can be analysed by MS (Fenn *et al.*, 1989 & Kebarle *et al.*, 1993). ESI produces ions in the form of $[Mass+nH]^{n+}$ where n = the number of protons added.

1.1.16. Mass analysers

Upon entering the mass spectrometer, ions are guided into a mass analyser. Mass analysers use a variable magnetic and/or electric field to eject ions and allow all ions of a single *m/z* ratio to be detected. There are multiple types of mass analyser, each with inherent properties to analyse different *m/z* ranges, sensitivity, resolution and mass accuracy (Haag,

2016 & Yates *et al.*, 2009). The Thermo OrbiTrap Fusion Tribrid primarily consists of two main mass analysers, an Iontrap and an Orbitrap, with the quadrupole typically being used in a *m/z* filtering mode only.

1.1.16.1. Iontrap

Iontrap mass analysers rely on a high frequency, oscillating electric field to spatially capture all ions within electrodes. The application of a varying radio frequency (RF) voltage to the capturing electrodes results in the selective resonance of a particular *m/z* ion, resulting in its ejection from the Iontrap and detection by the detector; thus ejection/detection is dependent upon *m/z* ratio (Stafford *et al.*, 1984). Two main types of Iontrap mass analysers have been developed: 3D (also known as Paul's) Iontrap and 2D linear Iontrap (Figure 1.11).

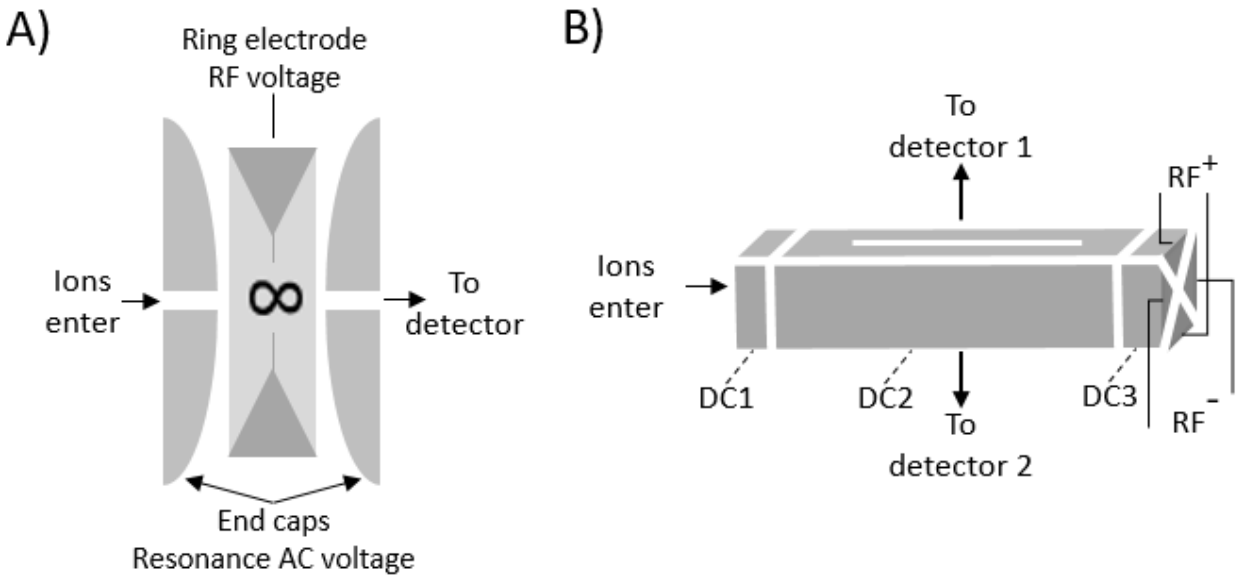


Figure 1.11: Schematic view of Iontrap mass analysers.
A) 3D (Paul) Iontrap, consisting of two hyperbolic end cap electrodes (at the same AC voltage) surrounding a ring electrode (exhibits the varying RF frequency) to trap ions in a figure of 8. B) 2D Linear Iontrap, consisting of four parallel, oppositely charged, hyperbolic pole electrodes (that exhibit varying RF frequencies) with capping electrodes (exhibit DC voltages) to trap ions within an internal cavity cloud. The end caps and central trap can each experience different DC voltages. Detectors are perpendicular to the length of the Iontrap.

A 3D Iontrap consists of a ring electrode surrounded by two hyperbolic electrode plates (Figure 1.11 A). All ions that enter into the 3D Iontrap are trapped spatially in a 3D figure-of-eight trajectory, and experience resonance ejection from the ion cloud by changing the RF voltage applied to the central ring electrode (Stafford *et al.*, 1984 & Williams *et al.*, 1994). A 2D linear Iontrap applies the same basic principles as a 3D Iontrap, however has a different structural layout. A 2D linear Iontrap consists of four parallel hyperbolic pole electrodes, of opposite changes, surrounded by end cap electrodes, thus trapping ions linearly (Figure 1.11 B). The RF voltage is applied to the end caps of the 2D linear Iontrap, resulting in the

selective resonance ejection of a particular *m/z* ratio, perpendicularly to the Iontrap direction, enabling detection (Schwartz *et al.*, 2002). Because a 2D linear Iontrap has a large internal capacity between electrodes, compared to other mass analysers, it allows for the large accumulation of ions before space-charge ion effects occur, thus providing the ability to achieve high sensitivity. However, Iontraps lack mass accuracy and resolving power compared to more advanced mass analysers such as the Orbitrap (Douglas *et al.*, 2005)

1.1.16.2. Orbitrap

The Orbitrap is another type of ion-trapping mass analyser that was used in this project. An Orbitrap consists of a spindle-like pole electrode encapsulated by concave electrodes to form a barrel-like structure (Figure 1.12, Makarov, 2000). A linear electric field is applied between electrodes which causes ions to oscillate bi-spatially, orbitally around and axially along the inner spindle-like electrode, with frequency of oscillations being independent of each other (Makarov, 2000). The frequency at which a particular ion axially oscillates is directly proportional to the *m/z* ratio of the ion. When a mixture of ions is detected, a complex, repeating wave is generated over time. Fourier Transformation (FT) can then be used to convert the complex repeating wave into its component wave frequencies, which are proportionate to the *m/z* ratios and their respective intensities, to generate a mass spectrum (Scigelova *et al.*, 2011). As frequency of oscillation can be detected much more accurately than time, Orbitrap mass analysers provide very high mass accuracy and resolution. However, they are more prone to space-charge effects, resulting in decreased sensitivity (Scigelova *et al.*, 2011 & Makarov, 2000).

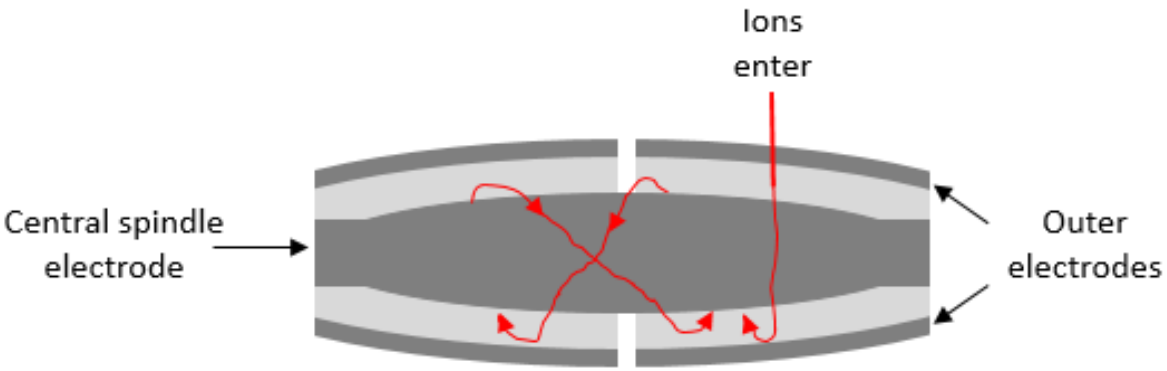


Figure 1.12: Schematic view of an Orbitrap mass analyser.
Consisting of two concave electrodes forming a barrel around a central spindle electrode. Ions are injected in time interval 'packets' through a whole in an outer electrode and ions oscillate both orbitally around and axially along the central spindle.

1.1.17. Peptide fragmentation

Peptide sequence determination requires peptide ion fragmentation and MS2 (MS/MS) analysis. For peptide analysis, multiple fragmentation methods have been developed which primarily result in fragmentation along the peptide backbone (Figure 1.13 , Hunt *et al.*, 1986, Johnson *et al.*, 1988, Zubarev *et al.*, 1998, Syka *et al.*, 2004 & Frese *et al.*, 2012).

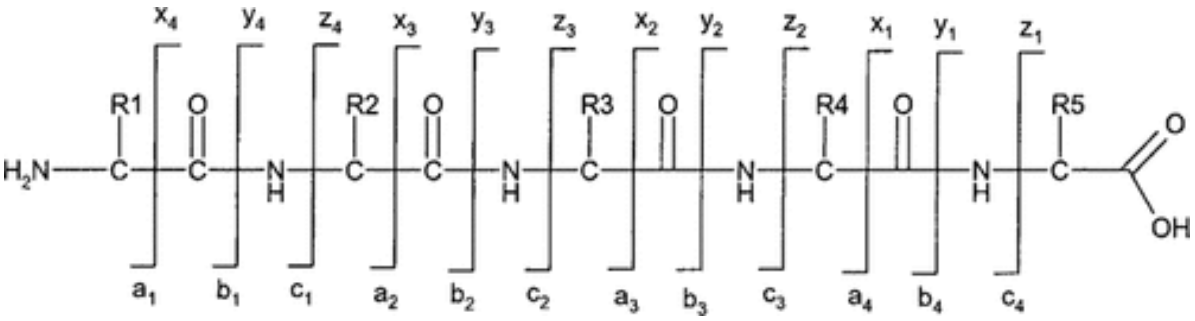


Figure 1.13: Nomenclature of peptide fragment ions.
a, b and c ions are different fragmentations of the peptide backbone, numbered sequentially from the amino-terminus of peptides. x, y and z ions are respective counterparts, of a, b, c ions, and labelled from the carboxy-terminus of peptides. R groups are of unknown amino acid side chains. CID/HCD predominantly results in b, y ions, ECD/ETD predominantly results in c, z ions and EThcD results in all four ion types. Taken from Paizs et al., 2005.

1.1.17.1. Collisional-induced dissociation

The most commonly used fragmentation technique is collision-induced dissociation (CID). CID involves selectively exciting a specific *m/z* ratio and allowing it to collide with an inert gas to promote fragmentation. Generally, CID fragmentation occurs along the amide bond, producing a heterogeneous mixture of b and y ions, depending on how the charge localises post fragmentation (Figure 1.13). b –ions are derived from N-terminal peptide ion fragments, and y –ions are derived from C-terminal peptide ion fragments (Johnson *et al.*, 1988). By deducing the exact mass of b/y ions, it is possible to determine the primary sequence of the peptide by calculating mass changes between identified *m/z* ratio peaks at MS2 (Johnson *et al.*, 1988 & Hunt *et al.*, 1986).

There are 2 general types of CID, resonance CID (performed in Iontraps, Johnson *et al.*, 1988) and beam type CID, also known as higher energy collisional dissociation (HCD, performed in a specialised collision cell, Olsen *et al.*, 2007). Resonance CID, involves the selective resonance excitation of a single *m/z* ratio (by changing the RF voltage), leading to their increased collision rate with the inert gas (commonly helium) and fragmentation (Johnson *et al.*, 1988). Thus, upon a single fragmentation event, a different *m/z* ratio is generated and the ions are no longer excited, hence MS2 spectra can be information limited. HCD involves filtering a specific *m/z* ratio into a specialised collision cell which has a variable RF voltage applied to accelerate ions into a high pressure of inert gas (commonly

nitrogen) to promote fragmentation (Olsen *et al.*, 2007). Because HCD does not require resonance excitation, fragment ions undergo secondary fragmentation events to provide greater depth of b/y ion production, allowing the more efficient determination of peptide primary sequence (Olsen *et al.*, 2007). However, a particular problem of CID is the susceptibility of phosphorylation to both CID and HCD methods (Zhang *et al.*, 2009).

1.1.17.2. Electron mediated dissociation

Another commonly used fragmentation technique is through the absorption of thermal electrons by multiply protonated, charged peptide ions. Electron Transfer Dissociation (ETD) utilises a chemical anion vehicle to introduce electrons efficiently to peptide ions in a gaseous state (Syka *et al.*, 2004). The absorption of electrons by ions results in the spontaneous fragmentation of electron hypervalent species, primarily at the N-Cα bond; producing a heterogeneous mixture of c and z ions (Figure 1.13, McLafferty *et al.*, 2001 & Zubarev *et al.*, 1998). c –ions are derived from N-terminal peptide ion fragments, and z – ions are derived from C-terminal peptide ion fragments (Zubarev *et al.*, 1998). In a similar manner to CID/HCD, deducing the exact mass of c/z ions allows for the determination of peptide primary sequence. ETD results in peptide fragmentation in a non-energetic reaction, thus preserving PTM localisation data. However these fragmentation strategies are limited to multiply charged species (3+), due to the electron-induced charge reduction and inability to fragment singly charged (1+) ions (Sobott *et al.*, 2009).

1.1.17.3. Dual fragmentation

More recently, a combined dual fragmentation approach has been developed which utilises both HCD and ETD fragmentation strategies (EThcD, Frese *et al.*, 2012). EThcD fragmentation creates a highly heterogeneous mixture of b, y, c and z ions in order to obtain more in-depth data to determine primary sequence while maintaining an increased chance that the PTM is not lost (Frese *et al.*, 2012 & Frese *et al.*, 2013).

1.11. Data analysis

In order to identify peptides, and thus proteins, from MS/MS analysis, search engines can be used to search obtained spectra against a database consisting of all potential proteins in the sample, for example the human protein complement. Three search engines are relevant to this project: MASCOT (Perkins *et al.*, 1999), PEAKS DB (Zhang *et al.*, 2012) and Andromeda (Cox *et al.*, 2011). Search engines use complex algorithms to match MS2 spectra (also known as peptide spectral matches (PSMs)) against the database to determine

primary peptide sequence and infer protein identity. As a peptide can be unique to a single protein or found within multiple proteins, inferred protein identities are scored based on the total number of PSMs that match to a specific protein.

To estimate the false-positive rate of PSMs randomly matching a target database, all PSMs are subject to a second-round search against a fictional decoy database of equal size, which is usually the primary database with the sequence of proteins reversed. Comparison of the total number of matching PSMs from the desired and decoy databases allows the determination of a false discovery rate (FDR). Thus, FDR reflects the probability that any given PSM that matches the desired database is due to chance. Hence, a FDR cut-off can be applied, generally at 1%, to statistically control the number of false positive protein identifications per sample.

For PTM discovery, data analysis needs to account for the known mass change introduced at the peptide level in the MS1 spectrum and the amino acid level in the MS2 spectrum, to determine the PTM site (Perkins *et al.*, 1999). However, as stated, fragmentation techniques can result in the neutral loss of PTMs and difficulty in localising its site (Mann *et al.*, 2002), essential information to characterise the biological function of a PTM. Therefore, common proteomics practise is to provide a site localisation confidence score of identified PTMs, determined through bioinformatics approaches. phosphoRS (ptmRS, Taus *et al.*, 2011) is a tool that analyses specific product ions to predict PTM site localisation, and has been used here with MASCOT as part of the Proteome Discoverer pipeline. Similar approaches are applied for PEAKS PTM and Ascore (part of the PEAKS pipeline, Han *et al.*, 2011), and Andromeda with PTM score (part of the MaxQuant pipeline, Cox *et al.*, 2011).

1.1.18. Quantitative proteomics

In response to a specific stimulus, PTMs and protein interactions can change, thus altering the functions of the protein of interest. It is therefore relevant to define quantitatively the changes, in order to further understand the proteome. However, MS analysis is not inherently quantitative due to differences in ionisation efficiency of peptides (Bantscheff *et al.*, 2007).Therefore, multiple strategies have been developed to quantify proteins, either in relative terms (fold change comparisons between treatments) or absolute terms (the number of copies a cell has of a single protein). Relative quantification can be further divided into two distinct techniques: label-free or labelled approaches (Bantscheff *et al.*,

2007 & Bantscheff *et al.*, 2012). Here, relative quantification was performed using a label-free approach, described below.

1.1.18.1. Label free quantification

Label-free quantification is a less accurate approach compared to label based techniques as the samples are prepared and analysed independently, resulting in increased experimental and analytical variation (Bantscheff *et al.*, 2007 & Bantscheff *et al.*, 2012). Whilst highly flexible and theoretically able to compare an unlimited number of samples, label free quantification requires extensive normalisation techniques to compensate for experimental error. Commonly used normalisations include LC elution time correction and signal intensity correction against house-keeping proteins, both of which are assumed to be near identical for different samples (Gillet *et al.*, 2016 & Välikangas *et al.*, 2018). Thus, although not the most powerful technique for quantification, label free techniques provide a rapid and low-cost strategy to simultaneously extract semi-quantitative information from proteomics datasets.

Two strategies have been developed to infer protein abundance from peptides: Spectral counting and peptide intensity averaging. Spectral counting relies on the fact that for a given protein, a greater number of PSMs will be identified for a protein of greater abundance, when normalising for protein length (Washburn *et al.*, 2001, Arike *et al.*, 2014, Old *et al.*, 2005 & Bantscheff *et al.*, 2007). However, spectral counting approaches are relatively controversial due not measuring any physical traits of identified peptides and is further complicated by the inclusion of dynamic exclusion windows which prevent the same *m/z* ratio being sent for MS2 analysis multiple times, thus limiting the number of PSMs (Old *et al.*, 2005).

Alternatively, a more computationally intensive, intensity-based quantification is possible. For each *m/z* peak identified at MS1, the intensity is integrated against the timescale of the LC chromatogram to determine peptide intensity as a measure of ion area; accounting for physiochemical properties of peptides and different elution width windows. Each peptide identified for a protein has the intensity areas averaged to infer protein intensity, which is then compared between experimental conditions (Bondarenko *et al.*, 2002, Old *et al.*, 2005 & Bantscheff *et al.*, 2007).

1.1.18.2. Label based quantification

Label-based quantification approaches are generally considered more accurate because they allow the mixing of samples at an early stage in sample preparation so that each sample experiences identical experimental and analytical variation (Bantscheff *et al.*, 2007 & Bantscheff *et al.*, 2012). Labelling approaches require the covalent incorporation of a known mass difference, either by *in cellulo* metabolic C13 labelling (SILAC, Ong *et al.*, 2002) or post-digest isobaric tagging of peptides (TMT, Thompson *et al.*, 2003 or iTRAQ, Ross *et al.*, 2004). However, these experiments are expensive and complicated by the need to optimally achieve near 100% labelling efficiency. They also require samples to be prepared simultaneously, which was not feasible with the experiments performed here (Chapter 3).

1.1.19. Software tools

Various software packages are available to analyse LC-MS/MS data, which can be open source or commercial. Generally, the ideal software tool is dependent on the required analysis and is influenced by the experimental design and MS instrumentation used. In this thesis, Proteome Discoverer was used for protein identification and phosphorylation analysis, through the MASCOT and ptmRS pipeline described above. PEAKS PTM was used for searching of all known PTMs, both biological and artefactual, in an 'open' PTM search, and MaxQuant was used for label free, intensity-based quantification.

1.1.19.1. PEAKS PTM

PEAKS is a software package that allows identification of proteins without a database present, or has a poor quality database, through the *de novo* sequencing of PSMs to directly determine peptide primary sequence (Ma *et al.*, 2003 & Zhang *et al.*, 2012). PEAKS PTM then creates a smaller secondary database of all proteins identified from the initial *de novo* peptide primary sequence. The concatenated secondary database is then used in a database search fashion for all unidentified *m/z* ratios with variable modifications of all known PTMs in the Unimod database (>300, Han *et al.*, 2011 & Creasy *et al.*, 2004). PEAKS PTM uses a novel FDR calculation where the initial large database is used as reversed decoy database to search against in an open PTM fashion. The identifications from the larger reversed database are then compared against the identifications in the smaller correct orientation database to determine the FDR in a more strigent manner, reducing potential bias for underestimating FDR (Zhang *et al.*, 2012 & Bern *et al.*, 2009). As such PEAKS was used for open PTM searching of all LC-MSMS data.

1.1.19.2. MaxQuant

MaxQuant is an open source software package with high flexibility for the quantification of both labelled and label free LC-MS/MS data by spectral matching and ion intensity methods, discussed above (Cox *et al.*, 2014). MaxQuant uses its in-built search engine, Andromeda (Cox *et al.*, 2011). As discussed, quantification approaches generally rely on the ability for large, global normalisation techniques against 'housekeeping proteins' assumed to be at identical levels independently of treatment (Välikangas *et al.*, 2018). However, hypoxia is known to induce large effects on the level of expression for many housekeeping genes (Caradec *et al.*, 2010), and thus potentially protein level, and may not be suitable for these type of normalisation strategies.

Critically, MaxQuant utilises a process termed delayed normalisation, in which protein intensities are initially inferred from peptide area intensities (Cox *et al.*, 2014). Identified proteins with minimal differences in protein intensity between experimental conditions are then used to normalise data, independently of being considered a housekeeping protein or not (Cox *et al.*, 2011 & Cox *et al.*, 2014). Thus, MaxQuant software is a more suitable package for the label free quantification of the LC-MS/MS analysis data obtained in this work. Perseus is a complementary, open source, software designed to aid in the extraction and interpretation of required information from MaxQuant, having various built in statistical and visualisation tools (Tyanova *et al.*, 2016).

1.12. Research Aims:

The aim of this thesis was discover putative PTMs and binding partners that may regulate full length HIF-1α and HIF-2α proteins *in cellulo,* and in response to hypoxia. The results will aid to further understand the cellular mechanisms used to regulate HIF-1α and HIF-2α, and how their dysregulation may lead to pathologies. Using proteomics techniques, I identified a large number of novel PTMs and binding partners that occur in both O_2-dependent and -independent manners. Their significance were further investigated by initial functional characterisation.

2. Chapter 2: Materials and Methods

2.1. Chemicals and reagents

Tissue culture reagents were purchased from Gibco. Powdered chemical reagents and custom DNA primers were purchased from Sigma-Aldrich, unless otherwise stated. Mass spectrometry (MS) solvents were purchased from Thermofisher, HPLC grade. All Eppendorf tubes used are Ultra-High recovery Eppendorf tubes (STARLAB).

2.2. Cell culture, transfection and treatment

2.2.1. Cell passaging

HeLa cells (ECACC catalogue #: 93021013) were grown in DMEM media supplemented with 10% (v/v) Foetal calf serum, 1% (v/v) Non-Essential Amino Acids and 1% (v/v) Penicillin/streptomycin and incubated at 37 ^{o}C, 5% CO_2 , 21% O_2. Cells were passaged at 80-90% confluency. Passaging involved removing the media, washing cells in PBS and incubating with 1X Trypsin/EDTA for 5 min. Detached cells were re-suspended in growth media, cell count obtained using a TC20- Automate cell Counter (BioRad) and seeded at a density of $1x10^5$ cells/mL (~$1.35x10^5$ cells/cm^2).

2.2.2. PEI 40K MAX linear stock solution

PEI 40K MAX linear powder (Polysciences #24765) was resuspended to make a 1% (w/v) solution in sterile PBS by continual stirring for 18 hr at 4 ^{o}C. The stock solution was adjusted to pH 7.5 by addition of 5 M NaOH before sterile filtering through a 0.22 µm filter, aliquoting and storage at 4 ^{o}C. Referred to as PEI in this thesis.

2.2.3. Transient transfection

Transient transfection was conducted 24 hr prior to experimental use. During transfection optimisation, various reagents were tested, following manufacturers guidelines (Table 2.1). Additionally, the recommended ratio for volume of transfection reagent (µL):quantity of DNA (µg), was changed 0.5 and 2 fold, in order to identify if a better ratio existed. The final transfection conditions used throughout were as follows: unsupplemented DMEM media was used to dilute DNA to a final DNA concentration of 10 ng/µL. For MS experiments (High HIFα expression level), the total quantity of DNA was at a 1:1 ratio of HA-Clover-HIFα:empty vector (pcDNA3(-)). For biochemical and functional assays (low HIFα expression level), the DNA was at a 1:19 ratio of HA-Clover-HIFα:empty vector. To this, a ratio of 4 µL PEI:1 µg of DNA was added. Transfection mixes were vortexed and left to stand at room temperature for 30 min. The volume of transfection mix added to cultured cells is equivalent to 5% of

the total cell culture volume (example volumes in Table 2.2). Cells were incubated for 18 hr before media was replaced and cells treated. All plasmids used are listed in Table 2.3.

Table 2.1: Transfection reagents tested during optimisation and recommended conditions.
*Recommended to determine empirically for different plasmids.

Transfection reagent	Mechanism of action	Manufacturer	Catalogue no:	Recommended ratio (Reagent : DNA)	Quantity DNA / 8.8 cm^2 plate (µg)
FuGENE 6	Polymer	Promega	E2691	3:1	1
FuGENE HD	Polymer	Promega	E2311	3:1	1
Turbofectin 8.0	Liposomal	ORIGENE	TF81001	3:1	1
Lipofectamine 3000	Liposomal	Invitrogen	L3000001	2:1	2.5
PEI 40K MAX, linear	Polymer	Polyplus	24765-1	4:1	NA*
JetPrime	Polymer	Polyplus	114-01	2:1	1.5
Viromer RED	Virus- like	ORIGENE	TT100302S	2:1	1.6

Table 2.2: Transfection conditions for a 10 cm (56.7 cm^2) plate.
Table includes required volumes (µl) of reagents and quantities (µg) of DNA for the High and low expression levels of HA-Clover-HIFα and empty vector.

Expression level:	High	Low
Reagent	Quantity	Quantity
Serum free medium	600 µl	600 µl
PEI	24 µl	24 µl
HA-Clover-HIFα DNA	3 µg	300 ng
pcDNA3(-) DNA	3 µg	5.7 µg

Table 2.3: Plasmids used and their availability.

Plasmid	Source
HA-Clover-HIF-1α	Cloned here
HA-Clover-HIF-2α	Cloned here
HA-Clover only	Cloned here
HA-HIF-1α	Gift from Prof Rocha
GFP-HIF-1α-DM	Previously cloned in the See lab
GFP-HIF-2α-DM	Previously cloned in the See lab
pcDNA3(-)	Invitrogen (V79520)
HaloTag-HIF-2α	Kazusa DNA Research Institute (pFN21AB4384)
HRE-Luciferase	Addgene (26731)
pcDNA-mRUBY	Addgene (40260)

2.2.4. Hypoxic incubation

Hypoxic incubation was at 1% O_2, performed in a Don Whitley H35 Hypoxystation. To ensure correct O_2 tensions, the hypoxic chamber was tested monthly using a Microx 4 fibre Optic Oxygen Meter (PreSens). Cells, PBS and lysis buffer were incubated in hypoxia for 4 hr, unless otherwise stated.

2.3. Cell lysis and protein extraction

2.3.1. Cell lysis and protein extraction

Post treatment, cells were washed in PBS before adding 11.4 µL/cm^2 (equivalent to 600 µL / 10 cm plate) lysis buffer (50 mM Tris pH 8.0, 120 mM NaCl, 5 mM EDTA, 0.5% (v/v) NP-40, 1X EDTA-free cOmplete protease inhibitor (Roche) and 1X phosSTOP (Roche), unless otherwise specified). Cells were scraped and supernatant collected into Ultra-High recovery Eppendorf tubes (STARLAB) before removal from the hypoxic chamber, in case of hypoxic incubation. Normoxic treated cells were lysed identically. Lysates were rotated end-over-end for 30 min at 4 °C, unless otherwise specified, before centrifugation at 10,000 *g* for 10 min at 4 °C. Cleared supernatant was collected into fresh Ultra-High recovery Eppendorf tubes.

2.3.2. Protein concentration determination

Protein quantification was performed using the Pierce BCA Protein assay Kit (ThermoFisher) following the manufacturer's recommended protocol, using a serial dilution of bovine serum albumin (BSA) of: 1.5, 1.0, 0.5, 0.25 mg/mL, as the reference standard.

2.4. HaloTag visualisation

2.4.1. *In vivo* labelling

The required volume of HaloTag- Oregon Green (Promega, G2801), of 0.25 µL/cm^2 of culture plate, was diluted 200 fold in unsupplemented DMEM media. An equivalent volume of media was removed from cells and replaced by the diluted HaloTag- Oregon Green solution. Cells were returned to incubation for 15 min and then washed 2 X with equal volumes of unsupplemented DMEM, before adding growth media. Cells were incubated for 30 min and the growth media replaced before being taken for microscopy or lysed for SDS-PAGE analysis.

2.4.2. *In vitro* labelling

Cells were lysed as described above. HaloTag- Oregon Green (Promega, G2801) was diluted 200 fold by direct addition into cell lysate. This was rotated end-over-end for 30 min at room temperature and used directly for SDS-PAGE analysis.

2.5. Immunoprecipitation (IP)

2.5.1. Endogenous protein IP

Post hypoxic incubation, $5x10^6$ cells (1x 10 cm plate) were lysed as previously described (600 µL lysis buffer) using the HA-Tag IP optimised lysis buffer (25 mM Tris pH 8.0, 350 mM NaCl, 1 mM EDTA, 1 mM EGTA, 0.5% (v/v) Triton X-100) or antibody manufacturer recommended lysis buffer (50 mM Tris pH 8.0, 120 mM NaCl, 1 mM EDTA, 0.2% (v/v) NP-40), both supplemented with EDTA free cOmplete protease inhibitors (Roche) and phosSTOP phosphatase inhibitors (Roche). IP used the Pierce CrossLink Magnetic IP/CO-IP Kit (88805) following manufacturers recommended protocol, omitting the antibody-bead crosslinking stages. Briefly, using a magnetic stand for all bead collection steps, 25 µL of protein A/G beads were washed in the provided lysis buffer (25 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% (v/v) NP-40, 5% (v/v) glycerol) before 2 µg of antibody in 100 µL 1X coupling buffer (provided) was added and left to shake at 600 rpm for 15 min at room temperature, to pre-bind antibodies to beads. Antibody-bead complexes were washed in supplied lysis buffer and 1 mg of cleared protein lysate added and left to rotate end-over-end for 18 hr at 4 ^oC. IPs were washed 3 X in lysis buffer and eluted in 25 µL of 2 X Laemmli buffer with boiling at 95 ^oC for 10 min.

2.5.2. HaloTag IP

All components required for IP are part of the HaloTag Mammalian Pull-Down System (G6504) kit, except the wash buffer (TBS + 0.05% (v/v) NP-40). The manufacturers recommended protocol for IP was followed, alongside 2 published protocols (Ohana *et al.*, 2011 & Daniels *et al.*, 2014), all scaled to work with $5x10^6$ cells (1 X 10 cm plate). All methods collect cells by scraping in PBS followed by centrifugation at 2,000 *g* for 10 min and lysed in 50 µL/mL of cell culture medium using the supplied mammalian lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% (v/v) Triton X-100, 0.1% (v/v) sodium deoxycholate & 1 X Protease inhibitor cocktail (G6521)). Physical perturbations to aid cell lysis were as follows: 1) Manufacturer recommend: passing cell lysate through a 25- gauge needle, 2) Ohana *et al.*, 2011: Sonication max power 10 s on, 10 s off (Cavitek 27 L professional ultrasonic cleaner, Allendale-ultrasonics) 3) Daniels *et al.*, 2014: -80 ^oC for 1 hr freeze-thaw cycle with dounce glass homogenizer (2.0 mL size). Cell lysates were then cleared by centrifugation at 10,000 *g* for 10 min at 4 ^oC and collected into fresh tubes before following the manufacturer recommend IP protocol. Briefly, all centrifugation steps were performed at 2000 *g* for 3 min, 100 µL of HaloLink resin was washed 3 X in the supplied lysis buffer before

adding 1 mg of protein lysate and rotating end-over-end for 2 hr at room temperature. Bead-protein complexes were washed 3 X in the supplied lysis buffer and 2 X in wash buffer. Bound protein was eluted with 30 U of TEV protease (V6101) according to manufacturer's recommended protocol.

2.5.3. GFP-Traps IP for Mass spectrometry analysis

A total of ~$1.5x10^8$ cells (30 X 10 cm plates) were lysed as previously described in section: 2.3.1, but collected into a 50 mL falcon tube (~18 mL lysis buffer). Cell lysate (20 mg) was diluted in dilution buffer (lysis buffer without NP-40) to a final concentration of 0.2% NP-40, a 2.5 fold dilution. Preclearing was performed with bab-20 (CHROMOTEK) beads at a volume equivalent to 1:200 of bab-20 beads (µL):diluted lysate (µL) (~300 µL of bab-20 bead suspension). Unless stated, all centrifugation steps were performed at 3,000 *g* for 2 min. Before addition to diluted lysate, bab-20 beads were equilibrated by washing 3 X in 5 volumes worth of dilution buffer, compared to volume of bead suspension. To preclear, the diluted lysate - bab-20 beads were rotated end-over-end for 1 hr at room temperature and centrifuged at 5,000 *g* for 5 min, supernatant was collected into fresh tubes. IP was performed with GFP-TRAP_A (CHROMOTEK) beads at a volume equivalent to 1:800 of GFP-TRAP_A (µL):diluted lysate (µL) (~75 µL of GFP-TRAP_A bead suspension). Before addition to diluted lysate, GFP-TRAP_A beads were equilibrated as stated for bab-20 beads. To IP, the diluted lysate – GFP-TRAP_A beads were rotated end-over-end for 18 hr at 4 ^oC. IP complexes were collected and washed sequentially 3 X as described for bead equilibration and 2 X in 25 mM AmBic (ammonium bicarbonate in HPLC grade H_2O). For all subsequent steps, HPLC grade solvents were used. An equal volume of 1% (w/v) Rapigest SF (WATERS) in 25 mM AmBic, to the initial volume of GFP-TRAP_A bead suspension, was added to washed beads and boiled at 95 ^oC for 15 min with 5 s vortexing every 2.5 min. Rapigest SF was diluted to a final concentration of 0.06% (v/v) in 25 mM AmBic and centrifuged 10000 *g* for 5 min and eluted supernatant collected into Ultra high recovery Eppendorf tubes (STARLAB).

2.6. Gel based analysis

2.6.1. SDS-PAGE sample preparation

Cleared protein lysates (30 µg) were mixed with 5X Laemmli buffer (250 mM Tris-HCl pH 6.8, 30% (v/v) glycerol, 10% (w/v) SDS, 500 mM DTT, 0.05% (w/v) bromophenol blue) and heated at 95 ^{o}C for 10 min. Samples were allowed to cool before loading onto a SDS-PAGE gel.

2.6.2. SDS-PAGE

Gels were made to obtain 3 mL a 4% stacking gel (4% (v/v) acrylamide, 125 mM Tris-HCl pH 6.8, 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.1% (v/v) TEMED) and 5 mL of a 7.5% resolving gel (7.5% (v/v) acrylamide, 375 mM Tris-HCl pH 8.8, 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.1% (v/v) TEMED). For IP efficiency determination, an equal volume of protein sample was loaded per lane. For specific protein quantification, equal total protein (20 µg) of protein samples was loaded. For western blotting, all gels were loaded with 5 µL of Color prestained Protein Standard, Broad Range (10-250 kDa, New England Biolabs, P7719). For Coomassie staining, all gels were loaded with 5 µL of unstained SDS-PAGE standard, broad range (BIORAD, 161-0317). Electrophoretic separation was performed at 200 V until the bromophenol blue dye front reached the end of the gel (~45 min) in electrophoresis running buffer (25 mM Tris pH ~8.3, 192 mM Glycine, 1% (w/v) SDS).

2.6.3. Coomassie staining

Colloidal coomassie stain was prepared as described by Candiano *et al.*, 2004. Briefly colloidal coomassie consists of: 0.12% (w/v) Coomassie G250 dye, 10% (w/v) ammonium sulphate, 10% (v/v) phosphoric acid and 20% (v/v) methanol. Gels were immersed in colloidal coomassie solution and allowed to stain for 18 hr at room temperature on an orbital shaker. Gels were destained in milliQ H_2O on an orbital shaker at room temperature, changing the water frequently, until the gel was clear. Images were captured using an Epson Scanner.

2.6.4. Western blotting

Protein transfer onto nitrocellulose membrane (0.2 µm, BIORAD) was performed in transfer buffer (electrophoresis buffer + 10% (v/v) ethanol) for 2 hr at 300 mA, 4 ^{o}C. Membranes were blocked in a 5% (w/v) skimmed milk solution (Marvel) dissolved in TBST (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween-20) for 1 hr at room temperature on an orbital shaker. Membranes were washed 3 X 10 min in TBST with shaking before addition of

primary antibody for overnight incubation at 4 °C with orbital shaking. After primary incubation, membranes were washed 3 X 10 min in TBST and incubated with the secondary antibody for 1 hr at room temperature with orbital shaking. All antibodies used are listed in Table 2.4, and were diluted in 5% (w/v) BSA in TBST. After secondary antibody incubation, membranes were washed 3 X 10 min in TBST and detected using ECL reagents: ECL clarity (BIORAD) or, the more sensitive, Amersham ECL select (GE Lifesciences). Images were captured using a Syngene gel imaging G-Box using the GeneSnap image acquisition software. Densitometry of western blots was performed using Fiji (Schindelin *et al.*, 2012).

Table 2.4: Summary of antibodies used for western blotting and IP.

Primary Antibodies					
Target	Company	Catalogoue number	Dilution factor	Monoclonal vs Polyclonal	Host species
HIF-1α	Proteintech	20960-1-AP	1:1000	Polycloncal	Rabbit
	Bethyl labratories	a700-001	1:1000	Monoclonal	Rabbit
	Novus biologicals	NB100-105	1:1000	Monoclonal	Mouse
	BD biosciences	BD610959	1:500	Monoclonal	Mouse
HIF-2α	Proteintech	26422-1-AP	1:500	Polycloncal	Rabbit
	Bethyl labratories	a700-002	1:1000	Monoclonal	Rabbit
	Bethyl labratories	a700-003	1:1000	Monoclonal	Rabbit
	Abcam	Ab199	1:1000	Polycloncal	Rabbit
GFP	Roche	11814460001	1:1000	2x Monoclonal	Mouse

Secondary Antibodies					
Target	Company	Catalogoue number	Dilution factor	Monoclonal vs Polyclonal	Host species
Mouse	Abcam	ab6808	1:5000	Polycloncal	Sheep
Rabbit	Cell signalling	7074s	1:3000	Not specified	Goat

2.7. Bacterial expression

2.7.1. LB (Luria-Bertani) broth

A 2.5% (w/v) solution was made using LB broth powder (Miller, Merck Millipore) in MilliQ H_2O, as per manufacturer's recommendations. LB broth was autoclaved once dissolved and allowed to cool to room temperature before use.

2.7.2. LB agar and antibiotic selection

A 3.7% (w/v) solution was made using LB agar powder (Miller, Merck Millipore) in MilliQ H_2O, as per manufacturer's recommendations. LB agar was autoclaved once dissolved and allowed to cool to 'touch hot' temperature before adding required antibiotic selection and pouring plates under flame. Plates were allowed to cool and either used instantly or stored at 4 °C. Final concentrations of 50 µg/mL kanamycin or 100 µg/mL ampicillin were used.

2.7.3. Generation of heat-shock competent cells

Subcloning Efficiency DH5α competent cells (Thermofisher #18265017) were streaked onto agar plates without selection and grown for 18 hr at 37 °C. A single colony was picked and grown in 100 mL LB broth, without antibiotic selection, at 37 °C in an orbital shaker to an OD_{600} of 0.6. Bacterial suspension was centrifuged at 1500 *g* for 10 min at 4 °C and the supernatant discarded. Maintaining all steps and buffers on ice, the cell pellet was resuspended in 50 mL of 50 mM autoclaved $CaCl_2$ (calcium chloride) and incubated on ice for 30 min. Centrifugation was repeated and cells resuspended in 1 mL of 100 mM autoclaved $CaCl_2$ and incubated at 4 °C for 18 hr. 100% glycerol was added to a final concentration of 20% (v/v) (250 µL), and cells aliquoted and stored at -80 °C, minimum of 48 hr before use.

2.7.4. Heat shock transformation

For cloning purposes STELLAR competent cells were used (TakaraBio), for plasmid propagation purposes DH5α heat shock competent cells were used (in-house generated, section: 2.7.3). Bacteria were defrosted from -80 °C on ice for 30 min. Unless otherwise stated, 5 ng plasmid DNA was added to 50 µL bacterial cells and left on ice for further 30 min in polypropylene tubes (Fisher scientific #10384641). Using a water bath, tubes were heated to 42 °C for 45 s before returning to ice for 2 min. SOC media (450 µL, ThermoFisher #15544034) was added and cells incubated at 37 °C on an orbital shaker for 1 hr. Bacteria were plated, under flame, at 100 µL and 400 µL onto relevant antibiotic selection plates and grown at 37 °C for 18 hr.

2.8. DNA based and cloning techniques

2.8.1. Plasmid visualisation and cloning design

SnapGene Viewer was used to visualise plasmids maps and sequences, identify restriction sites and aid in the design primers (SnapGene software (from GSL Biotech; available at snapgene.com)).

2.8.2. Restriction digestion

All restriction enzymes were purchased from New England Biolabs (NEB) in their HF (High Fidelity) format. Digestion was performed using the CutSmart buffer provided at 37 °C for 1 hr. For testing purposes, 250 ng of DNA was digested with 0.25 µL of enzyme(s) in a total

volume of 25 µL. For DNA extraction and further cloning purposes, 1 µg of DNA was digested with 1 µL of enzyme(s), in a total volume of 25 µL.

2.8.3. Primer design and generation

The sequences of all primers used are available in Table 2.5 to Table 2.8. All primers were synthesised by Sigma-Aldrich as standard oligonucleotides in a desalted format. Primers were resuspended to 100 µM (information supplied by manufacturer) in MilliQ water and vigorously vortexed before use, stored at -20 °C. Primers were diluted to 5 µM stocks for use.

Table 2.5: Sequencing primers.
Includes a description of which primers were used for each construct. Sequencing of full length HIFα genes required multiple primers due to the length of the genes, only 1 primer used per sequencing reaction. The location of the primer binding within the respective gene, and the stop codons, respective to promoter and 5' -> 3' sequence are included.

Plasmid	Name	Position (bp from promoter)	Direction	Sequence (5' → 3')
All	CMV promoter	0	Forward	AACGGGACTTTCCAAAATG
All	Clover	767	Forward	GCAGAACACCCCCATCGG
HA-Clover	Clover-STOP	942	X	X
HA-Clover-HIF-1α	HIF-1α (1)	1128	Forward	CTTACCATCAGCTATTTGCG
HA-Clover-HIF-1α	HIF-1α (2)	1708	Forward	GTGATGAAAGAATTACCG
HA-Clover-HIF-1α	HIF-1α (3)	2378	Forward	AGAACCAAATCCAGAGTCAC
HA-Clover-HIF-1α	HIF-1α (4)	2904	Forward	GCCACATCATCACCATATAG
HA-Clover-HIF-1α	HIF-1α STOP	3426	X	X
HA-Clover-HIF-2α	HIF-2α (1)	1202	Forward	CAACTTGTACCTGAAAG
HA-Clover-HIF-2α	HIF-2α (2)	1691	Forward	CATGAAGTTCACCTACTG
HA-Clover-HIF-2α	HIF-2α (3)	2394	Forward	GAAGACTATTACACATC
HA-Clover-HIF-2α	HIF-2α (4)	3135	Forward	TCACATTTGATGTGGAAAC
HA-Clover-HIF-2α	HIF-2α STOP	3552	X	X

Table 2.6: In-Fusion cloning primers.
Includes a description of how the primer was used, 5' -> 3' sequence and predicted melting temperature (Tm) for the gene specific portion of primer. For sequence, uppercase letters are vector specific and bases inserted to correct reading frame, lower case letters are gene specific.

Name	Sequence (5' → 3')	Tm (°C)
Clover into HA-HIF-1α (forward)	ACTACGCCTCCCTCGGATCCatggtgagcaagggcgag	61.7
Clover into HA-HIF-1α (reverse)	CCATGGTGGCGGATCCcttgtacagctcgtccat	55.7
HIF-2α into HA-Clover (forward)	TGGCATGGACGAGCTGTACAAGGGATCCatgacagctgacaaggagaagaaaag	62.7
HIF-2α into HA-Clover (reverse)	CGCCAGTGTGATGGATATCtcaggtggcctggtccag	63.2

Table 2.7: Site Directed mutagenesis primers.
Includes plasmid for mutation, position of amino acid mutation site with from and to mutations, 5'-> 3' sequence and predicted melting temperature. For sequence, uppercase letters match plasmid sequence, lower case letters are sites of mutation.

Plasmid	Mutation	Sequence (5' → 3')	Tm (°C)
Both	Flanking	GATCCGCCACAACGTTGAG	~62
Both	Flanking	AGACAATGCGATGCAATTTCC	
HA-Clover-HIF-1α	S31A	CAGATTCTTTAtcTCGCCGAG	~61
HA-Clover-HIF-1α	S31D	CAGATTCTTTAccTCGCCGAG	
HA-Clover-HIF-2α	S345A	CTCAATCTCAgcCAGGACGTAG	~62
HA-Clover-HIF-2α	S345D	CTCAATCTCAtcCAGGACGTAG	
HA-Clover-HIF-2α	T406A	CTCCTGGGGcGGGAGCC	~63
HA-Clover-HIF-2α	T406E	GTCTCCTGGttcGGGAGCCAG	
HA-Clover-HIF-2α	T528A	CTTGGAGgCACTGGCACC	~60
HA-Clover-HIF-2α	T528E	CTTGGAGgaACTGGCACC	
HA-Clover-HIF-2α	S581A	CCCGCACgcTCCCTTCC	~61
HA-Clover-HIF-2α	S581D	CCCGCACgaTCCCTTCC	
HA-Clover-HIF-1α	S786A	CTGGGGCAAgCAATGGATGAAAG	~61
HA-Clover-HIF-1α	S786D	CTGGGGCAAgatATGGATGAAAG	

Table 2.8: Quantitative Real Time PCR (qRT-PRC) primers.
Includes the gene name that primers were against, direction of primer and 5' -> 3' sequence.

Gene	Direction	Sequence (5' → 3')
CYCLOA	Forward	GCTTTGGGTCCGGAATGG
	Reverse	GTTGTCCACAGTCAGCAAT
EGLN1	Forward	TTTTCTGGTCTGACCGTCGC
	Reverse	CCTCACACCTTTTTCACCTGT
EGLN3	Forward	AGATCGTAGGAACCCACACG
	Reverse	TTCTGCCCTTTCTTCAGCAT
GLUT-1	Forward	GAACTCTTCAGCCAGGGTCC
	Reverse	ACCACACAGTTGCTCCACAT
VEGF	Forward	CTCCATGCCAAGTGGTC
	Reverse	GCAGTAGCTGCGCTGATAGA

2.8.4. Polymerase Chain Reaction (PCR)

The KOD polymerase kit (Merck Millipore), a high fidelity, proof-reading polymerase, was used throughout to reduce mutation risk. PCR reactions were made using the supplied reagents and as described by manufacturer. For a 50 µL reaction mix: 5 µL 10X KOD polymerase buffer, 5 µL mixed dNTPs (2 mM each dNTP), 3 µL $MgSO_4$ (25 mM), 3 µL of each diluted primer stock (5 µM), 20 ng plasmid and milliQ H_2O to make the final volume. For new primers, a 50 µL reaction was split equally into 10 tubes to test a gradient of

temperatures at Stage 2 (b). Once the optimal temperature is determined a larger scale PCR reaction was set-up for extraction purposes. PCR reaction conditions in Table 2.9.

Table 2.9 PCR reaction conditions.
A gradient of 55-70 °C is used for new primers to determine optimal temperature. X = 1 min/kBp of amplified gene.

	# cycles	Temp (°C)	Time (min)
Stage 1	1	95	2
Stage 2	30	95	0.5
		55-70	0.5
		72	X
Stage 3	1	72	10
Stage 4	1	4	Indefinitely

2.8.5. Agarose gel electrophoresis

A 1% (w/v) high melting temperature agarose gel (Bioline) was made by diluting agarose in TAE buffer (40 mM Tris-acetate pH ~8.5, 1 mM EDTA) and boiled until bubbling, ~2 min/100mL, in a microwave. Molten agar was left to cool until 'touch hot' before a final concentration of 5% (v/v) Midori Green advanced DNA stain (Nippon Genetics) was added, mixed and poured into the gel casting tray (~75 mL). DNA samples were mixed with homemade DNA loading dye (5X stock) to a 1X concentration (10% (v/v) glycerol, 3.3 mM Tris pH 8.0, 10 mM EDTA, 0.01% (w/v) Bromophenol Blue). Samples ran alongside 10 µL of HyperLadder 1KB+ (Bioline). Samples were run at 100 V for 30 min. For extraction, gels were visualised using a UV box. For imaging, gels were visualised and imaged using a Syngene gel imaging G-Box using the GeneSnap image acquisition software.

2.8.6. In-gel DNA extraction

DNA fragments were excised from agarose gels using a scalpel and purified using the E.Z.N.A in gel extraction kit (Omega bio-tek) following adjusted manufacturers recommendations. Briefly, the agarose slice is melted, bound to a DNA binding membrane and washed before elution. Adjustments include: extending washing conditions to 4 X 5 min, and, after the max speed drying spin step, a 10 min incubation at 70 °C with the lid open to further dry spin the column membrane before elution. Elution was done in TAE buffer, elution volume was dependent on downstream purposes: for In-Fusion Cloning: 20 µL, for mutagenic PCR and ligation: 50 µL.

2.8.7. In-Fusion Cloning

In-Fusion Cloning (Clontech) technology was used to create all HA-Clover based plasmids (HA-Clover only, HA-Clover-HIF-1α and HA-Clover-HIF-2α) following manufacturers recommended protocols. Briefly, destination vector was linearised with restriction enzymes, PCR was performed on the insert of interest using In-Fusion cloning compatible primers: Primers as described in Table 2.6 (with a 5' 15 bp overhang complementary to the destination plasmid). Bases were inserted as necessary to correct the reading frame and

$$\text{Insert mass (g)} = \frac{\text{ratio of insert:vector} \times \text{mass of vector (g)} \times \text{insert length (bp)}}{\text{vector length (bp)}}$$

recover restriction digest sites. Both linearised plasmid and amplified inserts were purified using in gel extraction (section 2.8.6). In-Fusion reaction mixes were performed in a total volume of 20 µL with a 3:1 molar ratio of insert:vector using the following equation:
For transformation, 5 µL of the infusion reaction mix was used to with Stellar competent cells (Clontech), (conditions described in 2.10.4).

2.8.8. Ligation

All ligation reactions were performed using T4 DNA ligase and supplied buffer reaction (1X concentration: 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ 1 mM ATP, 10 mM DTT) following the manufacturer's recommended protocol (New England Biolabs, #M0202S). Briefly, 50 ng of digested linearized vector backbone was mixed with digested insert a 1:3 molar ratio of vector:insert, calculated using the above equation, with 2 µL of T4 DNA ligase and made to total volume of 20 µL with milliQ H_2O. Ligation was performed at 37 ^oC for 30 min, followed by heat inactivation at 65 ^oC for 15 min. For bacterial heat shock transformation, 5 µL of ligation reaction was transformed into STELLAR competent cells (Clontech).

2.8.9. Plasmid amplification

For cloned and mutagenic plasmids, 5 colonies were picked from antibiotic selection plates and added into separate 50 mL falcon tubes containing 5 mL LB broth, maintaining antibiotic selection, and grown at 37 ^oC shaking on an orbital shaker for 8 hr. 0.5 mL cell suspension was mixed with 0.5 mL 80% (v/v) sterile glycerol to create glycerol stocks and froze at -80 ^oC. The remaining cell suspension was used for DNA extraction using the GeneJET plasmid miniprep kit (ThermoFisher). Briefly, cells were pelleted, lysed, neutralised to precipitate protein and cleared by centrifugation (13,000 *g* for 5 min) before applying to a DNA binding membrane for washing and elution using centrifugation (13,000 *g* for 1 min). The same adaptions as described for in gel extraction (section 2.8.6) are used. Plasmids

were sequenced (see LightRun DNA sequencing section 2.8.10). For a plasmid containing the correct gene, 100 µL of the glycerol stock was added into 200 mL LB broth, with antibiotic selection, and grown for 18 hr at 37 ^{o}C in an orbital shaker. Plasmid was extracted from bacteria using the PureLink HiPure Plasmid Maxiprep Kit (Invitrogen), following the manufacturer's protocol. Principles are identical to miniprep, but include an additional isopropanol-ethanol precipitation step to further purify the DNA.

2.8.10. LightRun DNA sequencing

All sequencing was performed with GATC Biotech (Eurofins genomics) using the LightRun sequencing option. Sequencing required, 100 ng of purified plasmid mixed with 5 µL of a single diluted sequencing stock (5 µM), made to a total of 10 µL in H_2O. Sequencing primers listed in Table 2.5.

2.9. Cloning HA-Clover plasmids

A schematic view of cloning protocol is presented in Figure 3.10.

2.9.1. HA-Clover-HIF-1α

HA-HIF-1α plasmid (a gift from Prof Sonia Rocha) was linearised with BamHI. The Clover gene was PCR amplified using the stated primers (Table 2.6) from a Clover-only plasmid (Addgene #40259). The Clover stop codon was removed and both BamHI sites were restored (either end of sticky ends) during primer design, to maintain the reading frame and allow easy removal of the Clover gene should other tags be required in the future. The final plasmid map is depicted in (Figure 2.1).

Figure 2.1: Plasmid map of cloned HA-Clover-HIF-1α.
Viewed and annotated in Snapgene viewer (SnapGene software (from GSL Biotech; available at snapgene.com))

2.9.2. HA-Clover only

The HA-HIF-1α plasmid was digested with both BamHI and EcoRV to remove the HIF-1α gene. The clover gene was amplified with new stated primers (Table 2.6) from the Clover-only plasmid. Primer design restored the N-terminal Clover BamHI and C-terminal EcoRV sites, and a stop codon C-terminally adjacent to the EcoRV site. This was done for future cloning with easy production of N-terminal and C-terminal clover constructs by linearization with EcoRV or BamHI respectively. The final plasmid map is depicted in Figure 2.2.

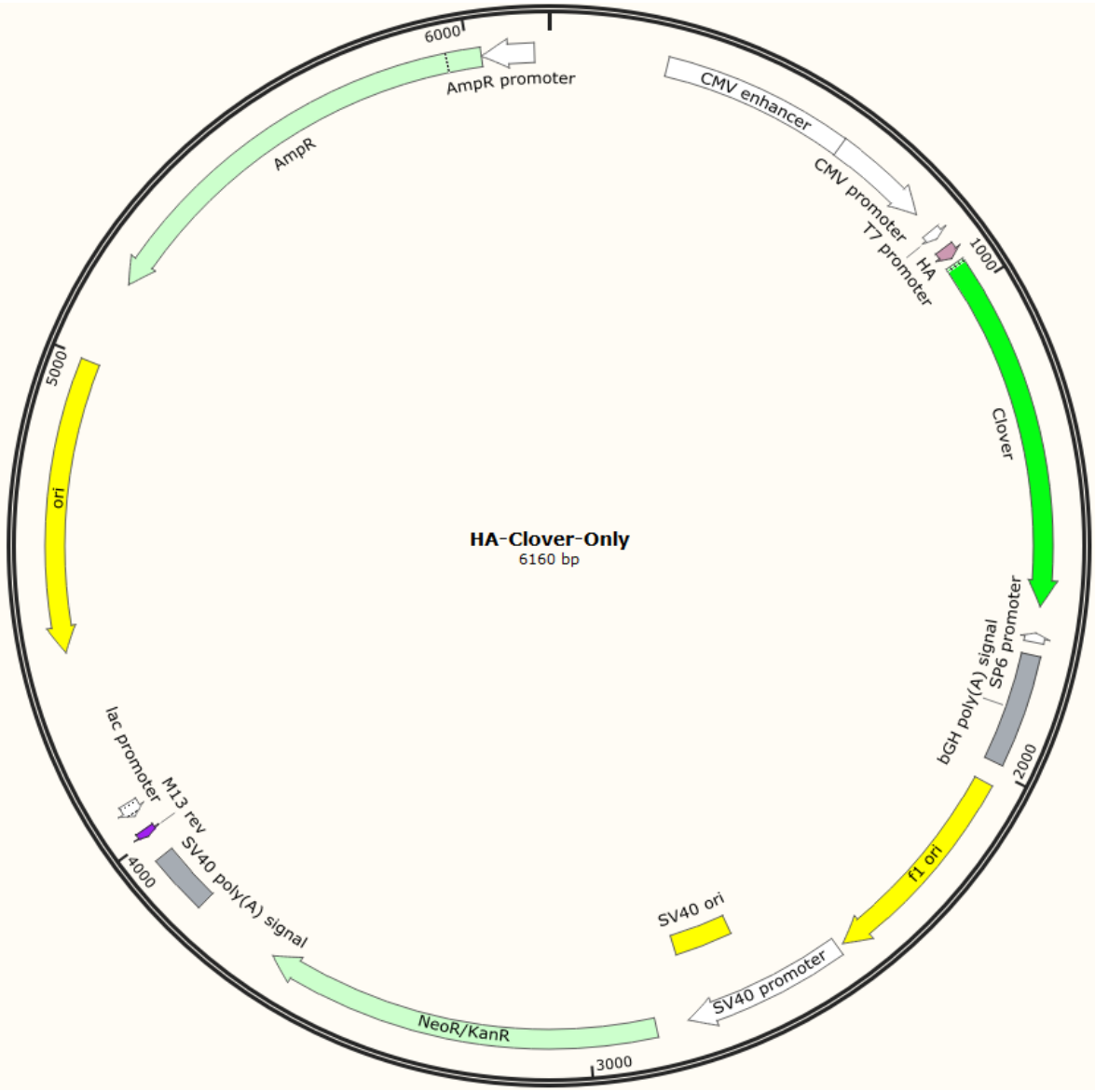


Figure 2.2: Plasmid map of cloned HA-Clover only.
Viewed and annotated in Snapgene viewer (SnapGene software (from GSL Biotech; available at snapgene.com)).

2.9.3. HA-Clover-HIF-2α

The HA-Clover only plasmid was linearised with EcoRV and the HIF-2α gene PCR amplified with stated primers (Table 2.6) from a HaloTag-HIF-2α plasmid available in the Sée lab. The C-terminal EcoRV site was restored in primer design. The final plasmid map is depicted in Figure 2.3.

Figure 2.3: Plasmid map of cloned HA-Clover-HIF-2α.
Viewed and annotated in Snapgene viewer (SnapGene software (from GSL Biotech; available at snapgene.com))

2.10. Site directed mutagenesis (SDM) - MEGAprimer

SDM was used to create point mutations within the HIFα gene. MEGAprimer protocol is depicted diagrammatically in Figure 2.4. The PCR protocol described (2.8.4) was used for SDM. Flanking primers were designed within the vector backbone to create a full-length mutant HIFα gene, irrespective of HIF-1α or HIF-2α gene, thus allowing the same flanking primers to be used for SDM of either the HIF-1α or HIF-2α genes. Mutagenic primers (7 bp either side of mutated codon) were designed so that their orientation would create the smallest PCR fragment in a first round of PCR (all primers listed in Table 2.7). First round

PCR used the mutagenic primer and a single flanking primer to create the MEGAprimer. MEGAprimer was extracted following the in-gel extraction protocol described (section 2.8.6). Second round PCR followed an adapted version of the described PCR protocol (2.8.4), using: 250 ng MEGAprimer and 50 ng plasmid. The second flanking primer was used with the MEGAprimer to amplify the full length gene. The full length gene was in-gel extracted and digested with the restriction enzymes: BsrGI and NotI. HA-Clover was digested identically and both subjected to in-gel extraction (2.8.6). Vector backbone and insert were ligated together (see ligation section: 2.8.8) and transformed by heat shock (2.7.4) into STELLAR competent cells (Takarabio).

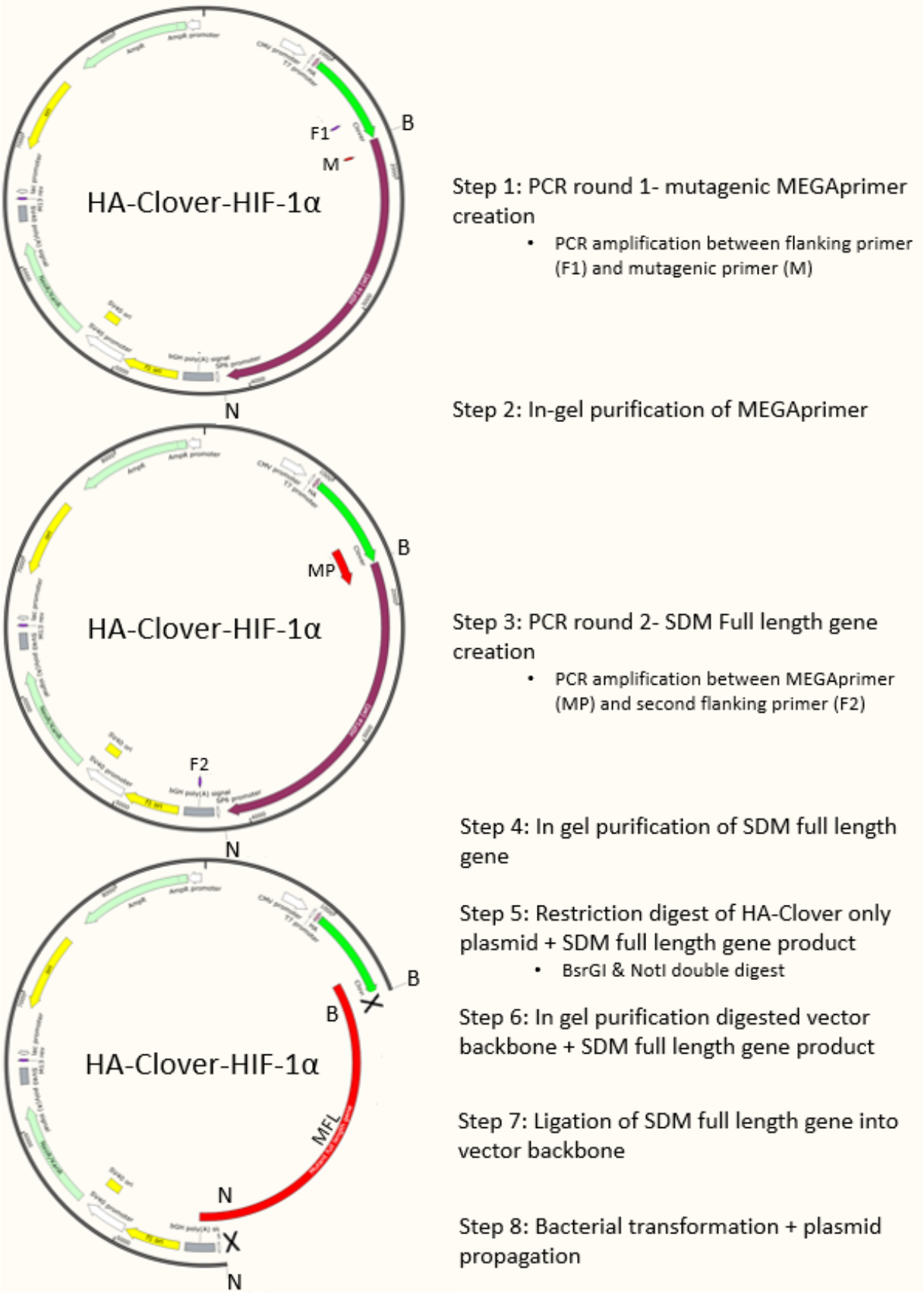


Figure 2.4: Schematic depiction of the MEGAprimer protocol used for SDM of S31 of HIF-1α.
Includes a description of each step performed throughout the creation of SDM plasmid. F1 and F2 = flanking primer locations, M = mutagenic primer location, MP = MEGAprimer location, MFL = mutant full-length gene. B = BsrGI and N = NotI restriction enzyme sites. X = ligation site. Plasmid viewed and automatically annotated using Snapgene Viewer (SnapGene software (from GSL Biotech; available at snapgene.com)).

2.11. Sample preparation for mass spectrometry (MS)

HPLC grade solvents and highest purity powdered chemical reagents were used throughout.

2.11.1. Reduction and alkylation

A Nanodrop-2000 was used to determine protein concentration of eluted material at a wavelength of 205 nm. DTT (Dithiothreitol), dissolved in 25 mM AmBic, was added to a final concentration of 3.33 mM and heated at 60 ^{o}C for 10 min. The sample was cooled to room temperature and IAA (Iodoacetamide), dissolved in 25 mM AmBIC, was added to a final concentration of 9.5 mM and incubated at room temperature in dark for 30 min. Remaining IAA was quenched by the addition of DTT to a final concentration of 7 mM DTT.

2.11.2. Proteolytic digestion

Reduced and alkylated samples were split equally into three Ultra low bind tubes for digestion by either: 10:1 (w/w) Trypsin Gold (Promega), 7.5:1 (w/w) Chymotrypsin (Promega) or 5:1 (w/w) Elastase (Promega) (w/w: total eluted protein : enzyme). Total eluted protein determined from the nanodrop 2000 protein concentration reading in 2.11.1. Samples were incubated at manufacturers recommended temperatures for 18 hr with 600 rpm shaking on a Thermomixer (Eppendorf). Post digestion, trifluoroacetic acid (TFA) and acetonitrile (ACN) was added to a final concentration of 1.5% (v/v) and 3% (v/v), respectively, and incubated at 37 ^{o}C with 600 rpm shaking for 2 hr before incubating on ice for 2 hr. The sample was centrifuged at 13000 *g* for 15 min at 4 ^{o}C and the clear supernatant collected. For mass spectrometry analysis of binding partners, 5% (20 µL) of the supernatant was removed, the remaining 95% was dried to completion under cooled vacuum centrifugation (Centrifuge: UNIVAPO – 150 ECH, Cooling unit: UNICRYO MC2L -60 ^{o}C, Vacuum pump: UNIVAC DQ4).

2.11.3. Strong Cation Exchange (SCX)

Strong cation exchange (SCX) was used for removal of PEG contamination from peptide samples prior to TiO_2 phospho-peptide enrichment. Dried peptides were dissolved in 200 µL of 1.5% (v/v) TFA in H_2O and sonicated at maximum power for 10 min (Cavitek 27 L professional ultrasonic cleaner, Allendale-ultrasonics). SCX stage tips were prepared by packing 5 discs of SCX membrane (Empore™ Supelco 47 mm Cation Exchange disc #2251)

into a 200 µL pipette tip. Tips were equilibrated by the sequential washing of 2X 200 µL of each: acetone, methanol, H_2O, 5% (v/v) ammonium hydroxide (in H_2O) and H_2O. All centrifugation steps were performed at 4000 *g* for 4 min, or until all liquid had passed through the stage tip. Peptide samples were passed through the equilibrated tip 2 X and washed 5 X in 50 µL 1.5% (v/v) TFA in H_2O before eluting in 3 fractions of 100 µL of 5% (v/v) ammonium hydroxide (in H_2O). Elutions were combined and dried to completion under cooled vacuum centrifugation (described in section: 2.11.2).

2.11.4. Titanium dioxide (TiO_2) phospho-peptide enrichment

Dried peptides were dissolved in loading buffer (80% (v/v) ACN, 5% (v/v) TFA, 1 M glycolic acid in H_2O) to a concentration of 25 ng/µl and sonicated as stated for SCX (section 2.11.3). Concentration was determined using the Nanodrop settings described in 2.11.1. A ratio of 100:1 (w/w) TiO_2 resin (GL Sciences) to peptide was added and mixed at 1400 rpm on a thermomixer (Eppendorf) at room temperature for 20 min with intermittent vortexing every 5 min for 5 s. Sequential wash steps equivalent to 1.2:1 (w/v) TiO_2 resin to wash buffer were performed subsequently with loading buffer, wash buffer 1 (80% (v/v) ACN, 1% (v/v) TFA in H_2O) and wash buffer 2 (10% (v/v) ACN, 0.2% (v/v) TFA in H_2O) with 1400 rpm shaking for 10 min. All centrifugation steps were at 2000 *g* for 1 min and supernatant was removed between washes. Peptides were eluted by sequential 1% (v/v) and 5% (v/v) ammonium hydroxide elutions, equivalent to 1.5:1 (w/v) TiO_2 resin to elution buffer, and combined before drying to completion by vacuum centrifugation (as described in 2.11.2). Dried peptides were resuspended in 20 µL of 3% (v/v) ACN, 1% (v/v) TFA (in H_2O) and sonicated as stated in 2.11.2. Samples were centrifuged at 13000 *g* for 15 min at 4 ^{o}C and 18 µL taken for LC-MS/MS analysis.

2.12. Orbitrap Fusion Tribrid Mass spectrometer

2.12.1. Liquid chromatography peptide separation

Peptides were separated by reverse-phase HPLC using an UltiMate 3000 nano system (Dionex) coupled in-line to an Orbitrap Fusion mass spectrometer (ThermoScientific). Peptides were loaded onto a trapping column (PepMap100, C18, 300 µm x 5 mm) in MS loading buffer (3% (v/v) ACN, 0.1% (v/v) TFA) at a flow rate of 9 µL/min for seven minutes. Peptides were then resolved at a flow rate of 0.3 µL/min on an analytical column (Easy-

Spray C18 75 µm x 500 mm, 2 µm bead diameter column) over a 60 minute gradient of 3% buffer A (0.1% (v/v) Formic acid in H_2O):97% buffer B (80% (v/v) ACN, 0.1% (v/v) Formic acid in H_2O) to 20% buffer A:80% buffer B, with a final 5 min 100% B wash.

2.12.2. High-Low MS/MS method, binding partner identification

For unenriched samples, MS1 spectra were acquired in the Orbitrap (60K resolution at a *m/z* 200) over a *m/z* range of 350-2000, AGC target = $2E^5$, maximum injection time = 100 ms. MS2 data were acquired in a data dependent acquisition (DDA) mode, using a 'top speed' method with a cycle time of 3 s. HCD fragmentation was set to 32% normalised collision energy (NCE) in the Iontrap. Iontrap MS2 settings were as follows: rapid mode (15K resolution at *m/z* 200), maximum injection time = 50 ms, fragmentation intensity threshold = $5E^4$ for 2+ to 5+ charge states. A dynamic exclusion window of 60 s was applied at a 0.5 Da mass tolerance.

2.12.3. HTP HCD and EThcD method development

The High-High method by Ferries *et al.*, 2017 was used as an initial starting point to optimise MS data acquisition for the IP and TiO_2 enriched samples. Briefly, the Orbitrap was used for MS1 and MS2 scans (60K and 30K resolution at 200 *m/z* respectively) in a DDA mode, 32% NCE HCD only fragmentation, AGC target = $2E^5$, fragmentation intensity threshold = $5E^4$, maximum injection time = 100 ms. For EThcD methods, the instrument parameter settings were kept the same. ETD reaction time was calibrated relative to the ion charge state, using an angiotensin standard. The HCD NCE was changed between 22%-18% in sequential LC-MS/MS experiments.

2.12.4. High-High MS/MS method, phospho-peptide identification

Final method for TiO_2 enriched phospho-peptide samples, MS1 spectra were acquired in the Orbitrap (60K resolution at a *m/z* 200) over a *m/z* range of 350-2000, AGC target = $5E^5$ ions, maximum injection time = 250 ms. MS2 data was acquired in a DDA mode, using a top speed method with a cycle time of 3 s. HCD fragmentation was set to 32% NCE in the Orbitrap. Orbitrap MS2 settings were: 30K resolution at *m/z* 200, maximum injection time = 250 ms, fragmentation intensity threshold = $2E^4$ for 2+ to 5+ charge states and 2 µscans performed. A dynamic exclusion window of 5 s was applied at a 10 ppm mass tolerance.

2.13. Mass spectrometry Data analysis

2.13.1. Proteome Discoverer (PD)

Proteome Discoverer v1.4 was used to process LC-MS/MS data from the Orbitrap Fusion Tribrid for binding partner identification and phospho-peptide identification and site localisation. All database searches were performed using the MASCOT search engine against the UniProt Human reviewed database (updated August 2018); instrument type = ESI-FTICR, fixed modifications = cysteine carbamidomethylation; variable modifications = methionine oxidation. For TiO_2 enriched samples, phosphorylation of S/T/Y was also included as a variable modification. For samples digested with trypsin (K/R -not P) a maximum of 2 miscleaves allowed. For samples digested with chymotrypsin (F/Y/L/W/M) a maximum of 4 miscleaves allowed. For samples digested with elastase (A/V/S/G/L/I) a maximum of 8 miscleaves allowed. For the High-Low MS/MS methods (binding partner identification), mass tolerances were MS1 = 10 ppm and MS2 = 0.5 Da. For High-High MS/MS methods (phospho-peptide detection and PTM localisation), mass tolerances were MS1 = 10 ppm and MS2 = 0.01 Da. The ptmRS mode, in PD, was used to localise the site of phosphorylation on the peptides. All identified peptides were filtered to a 1% False Discovery Rate (FDR) at the Peptide Spectral Match (PSM) level.

2.13.2. MaxQuant and Perseus

MaxQuant (1.6.7.0, Cox *et al.*, 2011 & Cox *et al.*, 2014) was used for label free quantification (LFQ) of high-low data for binding partner analysis. Perseus (1.6.7.0, Tyanova *et al.*, 2016) was used to extract and visualise data from the MaxQuant output. All data were searched against the UniProt Human Reviewed database (August 2019) and the in-built contaminants database using the Andromeda search engine. Miscleaves for the different enzymes were set as described previously for Proteome Discoverer (section 2.13.1). For intensity calculations, split peaks was disabled and peptides were selected if they were 7 amino acids or longer. Constant modification of cysteine carbamidomethylation and variable modifications of methionine oxidation and N-term acetylation were used. Default instrument parameter settings were used for an Orbitrap - Iontrap system. The match between runs option was enabled with a time window of 10 min. MS1 and MS2 scans were set to a 10 ppm and 0.5 Da tolerances respectively, peptides were filtered to a 1% FDR at the PSM level. Perseus was used to group identified peptides by unique protein identifier and average peptide intensity to estimate protein abundance. Missing values were imputed with a normalised distribution below the lowest intensity

peptides identified. A 2-sample t-test was performed between the two O_2 conditions, s0 = 0.1. Data was imported into R for use with a custom R script (provided by Dr Amy Campbell, part of the Eyers group) that colours significant identifications (P value <0.05) red, labels points with Gene name (from UniProt) if P value <0.01, and scales point size depending on the sum of PSMs between all replicates and conditions.

2.14. Final method for HA-Clover immunoprecipitation and mass spectrometry analysis

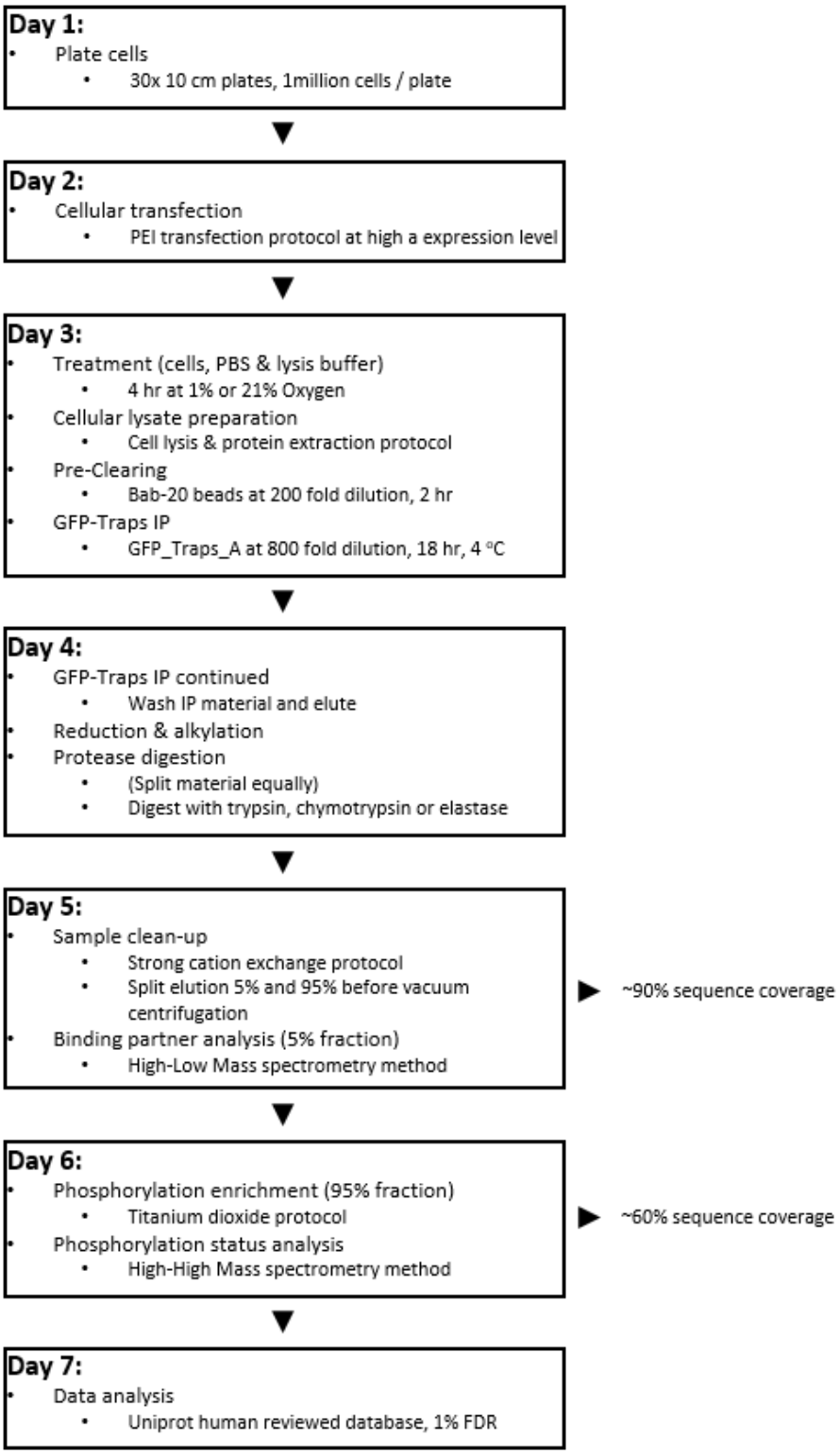


Figure 2.5: Flow diagram of the final methodology applied for the immunoprecipitation of HA-Clover tagged HIFα proteins

Figure includes: Day each step was performed, which protocol was used (indented bullet points) and approximate sequence coverage obtained at a 1% FDR for both HIF-1α and HIF-2α proteins.

2.15. Bioinformatics analysis

2.15.1. Phylogeny

Phylogenetic analysis was performed using the human HIF-1α (Q16665) and HIF-2α (Q99814) sequences following instructions from Hall, 2013. Briefly, protein sequences were BLAST searched for all homologous sequences and a manual filter of 50% sequence homology applied. All partial and 'unknown protein' labelled sequences were removed. If multiple isoforms existed for a single species, reciprocal blast searches were performed to identify those sequences with most similarity to the human equivalent, and the others removed. Genus-species names were converted into common names using the Taxize plugin for R (using the Global Names Resolver (GNR)). Phylogeny analysis was performed in MEGA7 using 500 bootstrap replicates.

2.15.2. DAVID

Gene ontology functional annotation was performed using DAVID (Database for Annotation, Visualisation and Integrated Discovery, version 6.8, Dennis *et al.*, 2003). Molecular function and Biological process annotations were kept, filtering out all other annotations, and P-values adjusted using the Benjamini-Hochberg method; a stringent P-value correction that multiplies the initial P-value by the total number of annotations divided by the rank of a given annotation. Data visualisation was performed using a custom R script (provided by Dr Amy Campbell).

2.16. Biochemical assays

2.16.1. Real Time Quantitative PCR (RT-qPCR)

Per sample, $4x10^5$ cells (1 X 6 cm plate) were plated, transfected as stated (Section 2.2.3) and incubated for a further 24 hr at 21% or 1% O_2. Cells were washed in PBS, lysed (400 µL of the supplied lysis buffer (Roche)) and RNA extracted using the HiPure RNA isolation kit (Roche), following the manufacturer's recommended protocol. RNA (1 µg), determined from a Nanodrop 2000 reading, was converted into cDNA using the SuperScript Vilo reverse transcription master mix (Invitrogen), following the manufacturer's recommended protocol. Briefly, all components were mixed and incubated at 25 ^oC for 10 min, followed by 1 hr at 42 ^oC and terminated at 85 ^oC for 5 min. cDNA was diluted 20 fold in RNAse/DNAse free H_2O before use. RT-qPCR was performed using a LightCycler480 using white bottom, 96 -well plates (Roche). Master mixes were prepared so that a single well contained a total of

20 µL, consisting of: 10 µL SYBR Green PCR mix (Roche), 2 µL cDNA (or H_2O for negative controls), 1 µL of each forward and reverse diluted stock primers (5 µM) for VEGF/GLUT-1 in H_2O. For EGLN1/EGLN3 4 µL cDNA and 3 µL of each forward and reverse diluted stock primers (5 µM) were used. All primers are listed in Table 2.8. Three technical replicates were performed per sample, per run. Three biological repeats were performed. Real Time qPCR cycling parameters are detailed in Table 2.10. Data analysis was performed using the LightCycler480 SW1.8 software, using the relative quantification method against the house keeping gene CYCLOA.

Table 2.10: Cycling parameters used for RT-qPCR.

Step	Temperature (°C)	Time (s)	Number of Cycles
Pre-incubation	50	120	1
	95	600	
Amplification	95	10	40
	60	15	
	72	15	
Melt Curve	95	5	1 / amplification cycle
	65	60	
	97	Continuous	
Cooling	40	10	1 / amplification cycle

2.16.2. Luciferase assay

Per sample, a total of $2x10^5$ cells per dish/well (1 X 35 mm dish / 6 well plate) were grown and transfected as described (2.2.3), using 50% of the total DNA quantity for transfection as a HRE-Luciferase plasmid (Addgene # 26731). Cells were washed in PBS and lysed in 35 µL/cm^2 of plate surface area (200 µL/ well) using luciferase lysis buffer (25 mM Tris-phosphate pH 7.5, 15% (v/v) glycerol, 1% (w/v) BSA, 8 mM $MgCl_2$, 0.1 mM EDTA, 2 mM DTT, 1% (v/v) Triton-X 100) and shaken for 10 min on an orbital shaker. Cell lysate was split into three separate wells of 80 µL in a white walled and bottomed 96 well plate (Greiner), for triplicate technical replicates. A 2.5 fold excess (200 µL) of luciferase working solution was added to each well (500 µM Luciferin (Abcam, #ab145164), 5 µM ATP (Sigma-Aldrich, #FLAAS-1VL) in lysis buffer). A BMG Labtech FLUOstar Omega plate reader was used to shake the plate at 200 rpm for 5s before standing in the dark for 5 min and an endpoint luminometry readings taken. Readings were taken twice to ensure that the luciferase reaction was at a steady state.

2.16.3. CHIP

Cells (~5x10^6, 1x 15 cm dish) were transfected following the high expression level transfection (as described in section 2.2.3). CHIP protocol was followed from Batie *et al.*, 2019. Cells were subjected to incubation at normoxia or 1% O_2 for 4 hr, before washing in PBS and crosslinking in 1% formaldehyde for 10 min, while maintaining O_2 incubation conditions. Excess formaldehyde was quenched using 125 mM glycine for 5 min before taking cells to the bench. Cells were washed 2 X in PBS before lysing in 400 µL CHIP lysis buffer (50 mM Tris pH 8.1, 1% (w/v) SDS, 10 mM EDTA and 1X EDTA free cOmplete protease inhibitors (Roche)) per 15 cm dish (2.75 µL/cm^2). Cells were scraped for collection and left on ice for 10 min before sonication, 8 cycles of 15 s on 30s off at 50% amplification on ice (Sonics Vibra-Cell # VCX130). Samples were cleared by centrifugation at 13000 *g*, 4 °C for 10 min and supernatant collected. Lysate (100 µL) was diluted 10 fold in dilution buffer (20 mM Tris pH 8.1, 1% (v/v) Triton-X100, 2 mM EDTA, 150 mM NaCl) and precleared using 2 µg sheered salmon sperm DNA, 20 µL G-Sepharose 50% bead slurry (Generon) and end-over-end rotation for 2 hr at room temperature. IP was performed by addition of a final concentration of 0.1% (v/v) Brij-35, 2 µg of the anti-GFP antibody (Roche, #11814460001) and end-over-end rotation at 4 °C for 18 hr. Anti-Mouse IgG was used as a negative control for IP (I5381, Sigma). Antibody complexes were captured using 30 µL G-Sepharose 50% bead slurry (Generon) in 2 µg sheered salmon sperm DNA with end-over-end rotation at 4 °C for 1 hr. Captured complexes were sequentially washed for 5 min each at 4 °C with end-over-end rotation in wash buffer 1 (20 mM Tri pH 8.1, 0.1% (w/v) SDS, 1% (v/v) Triton X-100, 2 mM EDTA and 150 mM NaCl), wash Buffer 2 (wash buffer 1 + 500 mM NaCl), and Wash Buffer 3 (10 mM Tris pH 8.1, 250 mM Lithium Chloride, 1% (v/v) NP-40, 1% (w/v) sodium deoxycholate, 1 mM EDTA), 2 X TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) before elution with 120 µL elution buffer (1% (w/v) SDS, 100 mM sodium-bicarbonate). Crosslinks were reversed by addition of NaCl to a final concentration of 200 mM and incubated at 65 °C for 18 hr with 300 rpm shaking on a Thermomixer. For proteolytic digestion, Tris-HCl pH 6.5 and EDTA were added to a final concentration of 40 mM and 10 mM respectively. Proteinase K (20 µg) was added and incubated at 45 °C for 1 hr. DNA was purified using the PCR product purification kit (NBS biologicals) following the manufacturer's recommended protocol. Purified DNA (3 µL) was used for RT-qPCR analysis, as described (section: 2.15.1).

2.16.4. Microscopy

Samples were grown and transfected, as stated, using 35 mm glass bottom dishes (Greiner). Imaging was performed on a LSM780 Zeiss Microscope equipped with GFP (488 nm) and RFP (561 nm) filters, 20X and 63X objectives (numerical apertures of 0.75 and 1.4 respectively) and a temperature (37 ^{o}C), O_2 (21%) and CO_2 (5%) tension control unit.

3. Chapter 3: Development of a mass spectrometry compatible immunoprecipitation protocol of HIFα

3.1. Introduction:

As discussed in the introduction, HIF-1α and HIF-2α have a high sequence homology of ~50% but, despite this homology, have different characteristics, including: sub-nuclear localisation, oxygen sensitivity and target genes. Regulatory mechanisms of the HIFα subunits have been investigated through targeted approaches, yet lack a discovery style experiment to identify all PTMs in an unbiased manner. Current published data investigating PTM regulation of HIFα subunits have used recombinant fragment-based approaches coupled with *in vitro* assays to identify and characterise PTMs. However, there are intrinsic issues with using these types of approaches to define PTMs and their roles, including incorrect folding of recombinant proteins, particularly for fragments, and thus there is a high possibility that the target proteins may not interact with endogenous modifying proteins. Hence, the false discovery (and false negative) rate of these techniques can be quite high. It is essential to investigate the regulatory roles of PTMs and of interactions with binding partners on full length HIFα expressed in cells, in order to understand the isoform specific characteristics observed in cells.

To achieve this, we aimed to adopt a high throughput (HTP) mass spectrometry (MS) analysis strategy for the in-depth characterisation of HIF-1α and HIF-2α PTMs and their binding partners in cells cultured at different O_2 levels. The initial objective was to design an immunoprecipitation (IP) protocol for full length HIF-1α and HIF-2α proteins from human HeLa cell lines that is compatible with such MS approaches.

3.2. Aims:

The aim of this chapter was to optimise a strategy for the specific IP of full-length HIF-1α and HIF-2α proteins from the widely used human HeLa cell line model. Due to the low expression levels of endogenous HIFα and the relatively high amount of protein required for discovery proteomics and comprehensive PTM analysis, this ultimately required optimisation of an exogenous protein expression system, utilising a tagged protein for IP.

3.3. Endogenous HIF-1α and HIF-2α immunoprecipitation:

3.3.1. Antibody selection

For greatest biological relevance (minimal artefactual data) in identified PTMs and binding partners, the direct IP of endogenous cellular HIF-1α and HIF-2α proteins is required to ensure that the physiological signalling pathways are maintained. However, this relies on the availability of antibodies that are isoform specific and that recognise proteins in a folded state. HIF-1α and HIF-2α have a high sequence homology and their secondary structures are unknown; thus, specific antibody selection can be problematic. For IP, monoclonal and polyclonal antibodies have intrinsic advantages and disadvantages. For example, monoclonal antibodies have little batch-to-batch variation due to their robust production techniques, however, although a single antibody may increase the isoform specificity, it simultaneously increases the chance of epitope masking due to altered protein folding or differential protein binding and/or PTMs within the epitope. These could potentially result in IP variability/bias or even failure (Wardle *et al.*, 2015). On the other hand, polyclonal antibodies detect multiple epitopes, limiting the risk of epitope masking but also may lack isoform specificity, and have a larger risk of batch-to-batch variation (Wardle *et al.*, 2015). The most important consideration for IP-based PTM mapping is the risk of enrichment of proteins in an unmodified state, or conversely in a particular modified state, associated with monoclonal antibody-based IPs. Thus, monoclonal antibody-based IP can potentially introduce a bias in identified PTMs and binding partners, if the pathways are mutually exclusive, a problem circumvented by polyclonal antibodies and detection of multiple epitopes.

A review of the literature identified 4 commonly used antibodies suitable for IP for each isoform, which were purchased and tested. A summary of the antibodies evaluated is provided in Table 2.4. Interestingly, there is a lack of evidence by manufacturers, or in publications, that unequivocally show isoform specificity of antibodies. Considering that HIF-1α and HIF-2α have near identical masses from western blotting (~120 kDa) it is impossible to distinguish between them without tagged-overexpression (resulting in a molecular weight shift) or silencing/knockdown controls.

Specificity testing was conducted using the exogenous expression of a GFP tagged- HIF-1α or HIF-2α protein, using O_2 stable GFP-HIF-1α-DM and GFP-HIF-2α-DM constructs, available

in the Sée lab. The DM constructs have been mutated to remove the O_2 sensitive proline residues (P402A/P564A and P405A/P531A for HIF-1α and HIF-2α respectively), thus eliminating their O_2 dependent regulation for maximal stability, and abundance, regardless of environmental O_2 levels. These GFP-tagged HIFα proteins have an increased apparent molecular weight to ~150 kDa by western blot, thus independent exogenous expression can be used for cross reactivity testing by western blot (Figure 3.1).

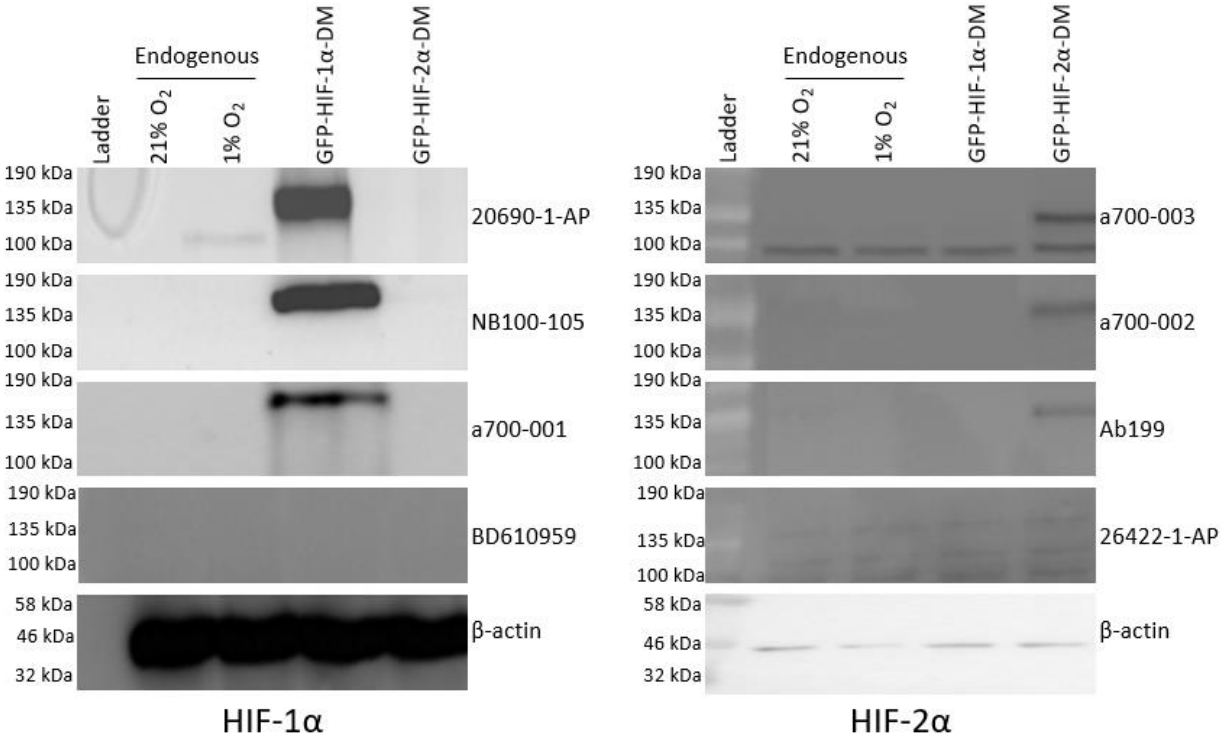


Figure 3.1: Sensitivity and specificity testing of 8 different commercially available HIF-1α and HIF-2α antibodies.
Untransfected HeLa cells were incubated for 4 hr at 21% or 1% O_2 as indicated. Transiently transfected GFP-HIFα-DM isoforms were left at 21% O_2 for 24 hr post transfection, before lysing all conditions simultaneously. Western blotting antibody dilutions are stated in Table 2.4. β-actin was used as a loading control.

Except for BD610959 and 26422-1-AP, every antibody tested could specifically and adequately detect the exogenously expressed GFP-HIF-1α-DM or GFP-HIF-2α-DM proteins. However, only 20960-1-AP (Proteintech) and a700-003 (Bethyl laboratories) could detect the endogenous HIF-1α and HIF-2α proteins respectively, thus these antibodies were used for IP optimisation. These antibodies are used throughout this thesis for western blotting and referred to as HIF-1α or HIF-2α primary antibodies.

3.3.2. Antibody IP optimisation

When using IP of bait proteins to identify binding partners it is important to consider and minimise the co-purification of non-specific 'background' proteins. These contaminants are

generally environmentally introduced, inadvertent from sample preparation, or of high cellular abundance proteins, such as serum albumin or myosin, and are predicted to account for up to 95% of all IP coupled mass spectrometry identified proteins (Trinkle-Mulcahy, 2012). Hence it is essential to optimise IP conditions to limit non-specific interactions, while maintaining weak or transiently interacting proteins of interest.

A protocol for the IP of HA-tagged proteins using a commercially available anti-HA antibody cross-linked to magnetic beads (Pierce) had previously been optimised in the Eyers lab. Therefore, this protocol was used as a starting point for the IP of endogenous HIFα proteins, substituting the anti-HA antibody for the anti-HIF-1α or anti-HIF-2α antibodies. Although this protocol uses a cross-linking strategy, for initial optimisations this step was not included. This protocol uses a relatively stringent lysis and wash buffer to minimise non-specific interactions, however is above the manufacturers recommended concentrations, therefore a second lysis buffer recommended by the antibody manufacturers was also tested (Figure 3.2).

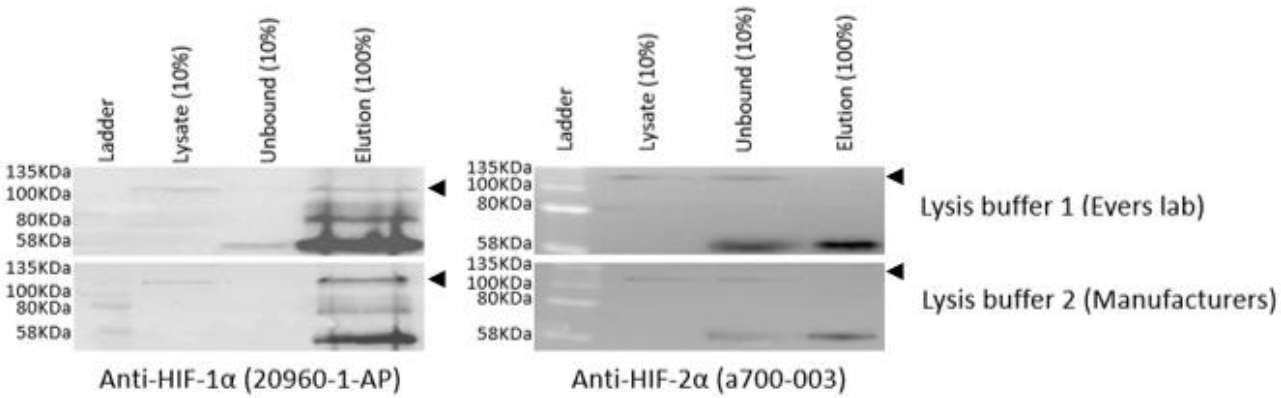


Figure 3.2: IP optimisation for endogenous HIF-1α and HIF-2α proteins.
Western blot analysis using either the HIF-1α or the HIF-2α antibodies of total lysate (10% volume), unbound protein material (10% volume) and eluted fraction (100% volume) following IP using either the Eyers lab protocol, or manufacturer recommended conditions. HeLa cells were incubated at 1% O_2 for 4 hr before lysis in the relevant lysis buffers. Arrows indicate bands of the correct molecular mass for HIF-1α and HIF-2α. Lysis buffer 1 composition: 25 mM Tris pH 8.0, 350 mM NaCl, 1 mM EDTA, 1 mM EGTA, 0.5% Triton X-100. Lysis buffer 2 composition: 50 mM Tris pH 8.0, 120 mM NaCl, 1 mM EDTA, 0.2% NP-40. Both lysis buffers were supplemented with EDTA-free cOmplete protease inhibitors and phosSTOP phosphatase inhibitors (Roche).

The initial immunoprecipitation protocol from the Eyers lab was successful in purifying ~40% of total endogenous HIF-1α protein. IP efficiency was increased by ~2 fold when using the manufacturer recommended lysis/wash buffer (Figure 3.2), determined by densitometry (data not shown). However, neither method tested was successful for IP of HIF-2α, shown primarily by a lack of a band at the correct molecular weight in the elution lane, and secondarily by equally intense bands between the lysate and unbound lanes.

Combined, these data suggest a potential problem with antibody-epitope recognition or antibody-bead association, rather than loss caused by wash steps or failure to elute. An intense band of ~55 kDa is seen in the elution lane of all conditions; presumably the antibody heavy chain given that the same antibody is used for IP and western blotting. This suggests that the anti HIF-2α antibody can efficiently bind to the magnetic beads, thus the problem rather lies in epitope recognition. To confirm this, the antibody-bead binding efficiency was tested by following the IP protocol for both anti-HIF-1α and anti-HIF-2α antibodies without cellular lysate. 'Mock' IPs were then analysed by SDS-PAGE and stained with Colloidal Coomassie stain to highlight all proteins in the sample (Figure 3.3).

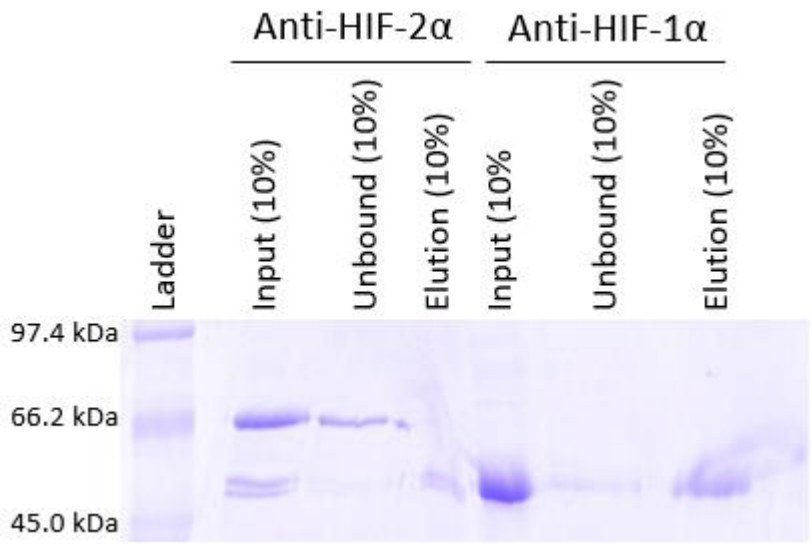


Figure 3.3: Antibody bead binding efficiency.
Colloidal coomassie stained SDS-PAGE gel of lysate free, mock IPs using the anti-HIF-1α and anti-HIF-2α antibodies**.** IPs were performed following the optimised protocol from the Eyers lab without cross-linking and using the manufacturer recommended lysis/wash buffer. Lysis performed in 5X Laemmli's buffer. 10% of input, unbound and elution samples were loaded and stained with Colloidal Coomassie stain.

For the anti-HIF-1α antibody, where IP was successful (Figure 3.2), there was a single band of ~55 kDa in the input material. As expected for successful IP, without cross-linking, this band is significantly reduced and then recovered by bead binding and elution stages respectively (Figure 3.3). A faint band in the unbound lane can be explained by the high quantity of antibody used potentially being above the maximum binding capacity of the beads. For the anti-HIF-2α antibody, 3 bands were detected in the input lane; a prominent ~66 kDa band and two equally intense bands of ~55 kDa. The manufacturer (Bethyl Laboratories) states this antibody is supplied in a 0.1% BSA solution (~66 kDa), which aligns to the 66 kDa BSA standard of the molecular weight ladder (BIORAD Broad MW range marker unstained, catalogue #: 1610317), thus confirming the identity of the heavier band. This suggests that the ~55 kDa bands are the antibody chains. The ~55 kDa bands have an

identical band pattern to the anti-HIF-1α antibody positive control (Figure 3.3), thus this suggests that the anti-HIF-2α antibody successfully binds to, and elutes from, the magnetic beads. Hence, the likely explanation for IP failure is that the HIF-2α antibody cannot identify the epitope; either by protein folding or protein-complex formation that masks the epitope, as hypothesised previously, but works for western blotting because of protein denaturation.

Overall, endogenous HIF-1α could be successfully and efficiently pulled-down using the manufacturer recommended buffers with the method developed in the Eyers lab. However, due to the high amount of protein necessary for discovery PTM analysis (see Figure 3.15 (A)) and without a successful method to IP endogenous HIF-2α, it was decided to explore tag-based exogenous expression strategies.

3.4. Exogenous expression of tagged- HIF-1α and HIF-2α for immunoprecipitation:

3.4.1. Exogenous expression optimisation:

There are many options for the expression of tagged fusion proteins in cell lines including genomic knock-in technology (CRISPR/Cas9), bacterial artificial chromosomes (BACs) and plasmids, each with inherent advantages and disadvantages. Although technically challenging and costly, the development of CRISPR/Cas9 has revolutionised human cell genetic manipulation. This technology can insert genes/tags within the genome at specific sites, thereby maintaining endogenous protein levels and improving biological reliability for pathway interrogation (Cong *et al.*, 2013 & Hsu *et al.*, 2014). BACs are large supercoiled segments of DNA that allow expression of a target gene of interest under its endogenous promoter, following expression into relevant cells. BACs are typically ~200 kbp in size and contain large segments of upstream and downstream neighbouring DNA to the endogenous gene of interest. Putting the gene into its natural context increases the likelihood that endogenous gene expression regulation may be maintained, but is not 100% successful. As an added complication, BACs require extensive genetic manipulation for tag insertion and screening of clones to evaluate copy number, and thus BAC-mediated protein expression levels, in comparison with endogenous protein (Shizuya *et al.*, 2001). In contrast, plasmids are small (usually <10 kbp) circular pieces of DNA which contain all the essential signalling motifs to promote self-replication and transcription of the tagged-gene of interest independently to the host genome (reviewed by del Solar *et al.*, 1998). Consequently,

plasmids do not contain endogenous transcriptional promoters or regulatory pathways and typically result in high expression of the desired protein, either continuously or following induction. Their easy genetic manipulation, ability for large scale fusion-protein production and ease for transient transfection/stable cell line selection have popularised the use of plasmids exponentially.

Without prior knowledge of which tags will work for successful HIFα IP, combined with the difficulty of creating knock-in/BACs systems and the requirement for large quantities of protein post IP (which would not be available with endogenous expression levels (Figure 3.15 (A)), a transient plasmid based exogenous expression system was selected for optimisation for the pull-down and PTM analysis experiments.

Plasmid DNA introduction into cells involves getting the hydrophilic DNA through the hydrophobic phospholipid bilayer. Three main strategies have been developed for DNA transfection into cells: virus-mediated, chemical reagents and physical perturbations, each having different degrees of cytotoxicity, efficiency and reproducibility (reviewed by Kim *et al.*, 2010). The most common, and simplest, technique for DNA transfection is the use of commercially available chemical reagents, known as transfection reagents. These are designed to facilitate DNA endocytosis and subsequent ejection inside the cytoplasm. As high quantities of tagged-HIFα proteins are required for IP and subsequent comprehensive PTM analysis by MS, it was important to optimise transfection efficiency. Several available transfection reagents were tested (Table 2.1). Optimisation studies aimed to obtain the greatest number of transfected cells, with a reasonable level of expression, while maintaining nuclear localisation, cell viability and avoiding cellular stress (membrane blebbing).

For initial optimisations, a luciferase reporter construct under the control of a strong viral CMV promoter (CMV-Luc) was transfected into HeLa cells, following the manufacturers recommended protocols for each reagent. Thus, luciferase signal reflects the total transfection rate of cells. Additionally, the volume of each transfection reagent was altered 0.5X and 2X, maintaining the quantity of plasmid DNA, to identify the best DNA:reagent ratio for transfection efficiency (Figure 3.4 (A)).

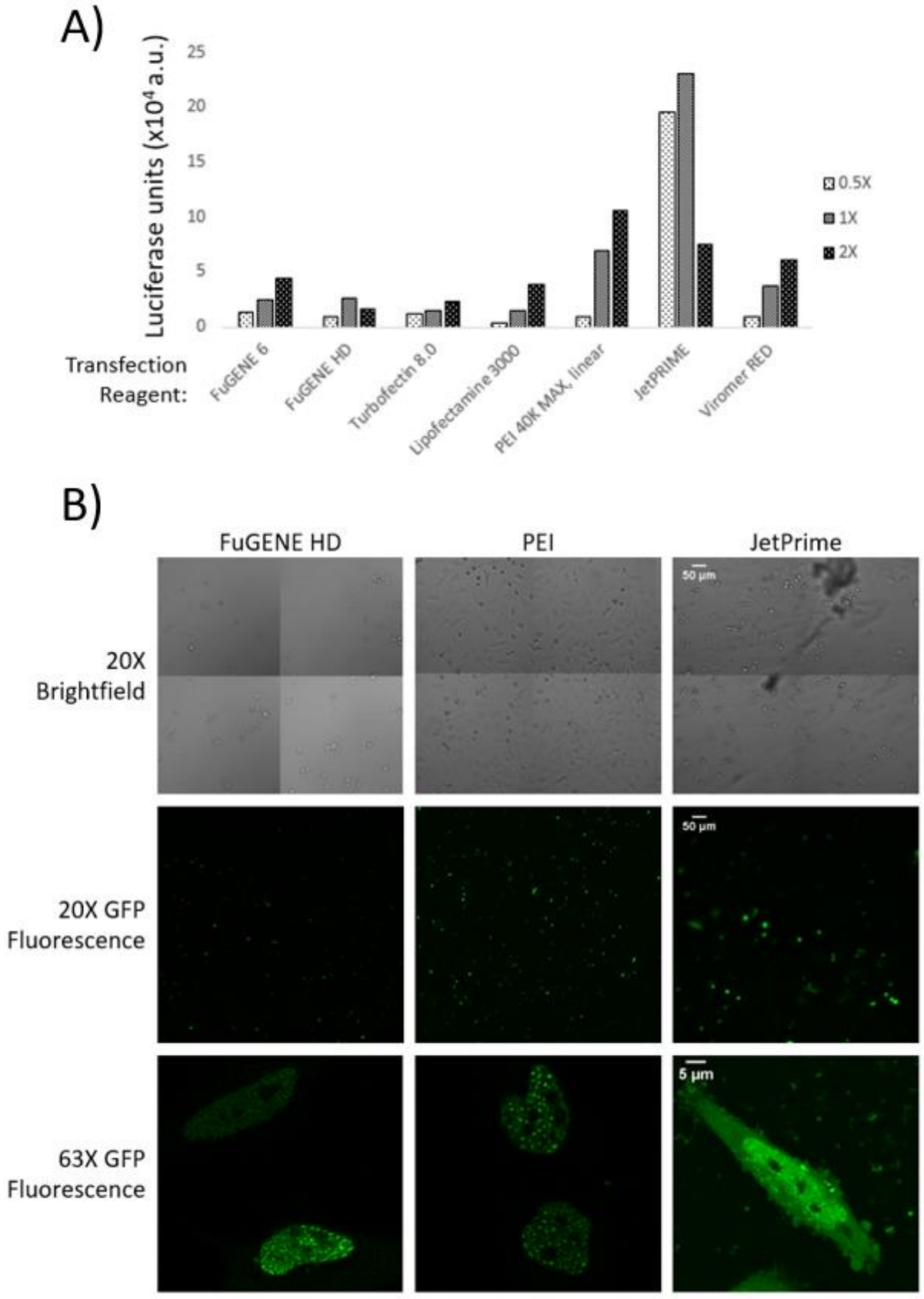


Figure 3.4: Transfection efficiency optimisation by luciferase assay and fluorescence imaging.
A) Comparison of transfection efficiencies between 7 different transfection reagents, measured as a luciferase response from a CMV-Luciferase plasmid in HeLa cells. B) Microscopy analysis of FuGENE HD, PEI and JetPrime using the optimised DNA:reagent ratios from (A). Overexpression of GFP-HIF-2α-DM was used for fluorescence imaging at 20X magnification, with a 2x2 grid of images at brightfield and 488 nm fluorescence excitation, and single images at 63X magnification using a Zeiss LSM 780 microscope.

These studies showed that PEI 40K MAX linear (referred to as PEI from now) and JetPrime had the greatest transfection efficiencies according to CMV-luciferase response, considerably outperforming the more common transfection reagents by 2-5 fold respectively (Figure 3.4 (A)). Although luciferase assays provide a high throughput method for determining transfection efficiency, they cannot distinguish between an increased number of transfected cells or dramatic luciferase overexpression in fewer cells. Therefore the two best transfection reagents (PEI and JetPrime) and FuGENE HD, used as a benchmark as it was previously optimised within the Sée lab, were further assessed for exogenous protein expression by confocal fluorescence microscopy (Figure 3.4 (B)).

GFP-HIF-2α-DM expression was used because of the highly distinct, nuclear only, fine punctate localisation of the HIF-2α protein (Hara et al., 1999 & Taylor et al., 2016). Strong overexpression of GFP-HIF-2α-DM has been shown to lead to aggregation and cytoplasmic expression; thereby allowing an assessment of aberrant expression. The 20X magnification images (Figure 3.4 (B)) clearly confirm the better transfection efficiency of PEI and JetPrime compared to FuGENE HD, measured by the number of fluorescent cells. Brightfield images of the same region shows considerably more cell death (rounded cells) with JetPrime and FuGENE HD transfections compared to PEI transfection. At 63X magnification, which permits visualisation of nuclear localisation, it is apparent that JetPrime causes significant cellular stress with membrane blebbing and cytoplasmic expression of GFP-HIF-2α-DM. Conversely, FuGENE HD and PEI transfections have the expected specific nuclear punctate localisation (Hara et al., 1999 & Taylor et al., 2016). Thus, PEI was selected as the transfection reagent for optimising tagged-HIFα IPs, at a 4 µL:1 µg transfection reagent:DNA ratio per 200,000 cells in a 35 mm dish (8.8 cm^2).

3.4.2. Tag selection

Exogenous expression of an IP tagged- fusion protein can have many advantages over target-antibody based IP approaches, including the circumnavigation of problems associated with the use of polyclonal and monoclonal antibodies discussed previously (section 3.3.1). A multitude of tags have been developed for IP, each with intrinsic advantages such as enhancing solubility or fluorescence for ease of visualisation (reviewed by Terpe, 2003). Within the Sée lab, there has been some success in using the HaloTag and GFP-Traps technology for purification of non-HIFα proteins. N-terminal tagged HaloTag-HIF-2α and GFP-HIFα-DM constructs (N-terminal and C-terminal tagged), were readily available

in the Sée lab and were therefore used as initial starting points for optimisation. IP kits against these tags are commercially available, HaloLink resin (Promega) and GFP-Traps (CHROMOTEK), and have been successfully used by many groups for IP and MS analysis (Daniels *et al.*, 2014, Ohana *et al.*, 2011, Trinkle-Mulcahy *et al.*, 2008, Lipinszki *et al.*, 2014 & Smits *et al.*, 2013).

3.4.3. HaloTag IP optimisation

A problem associated with all antibody-based IP protocols is that the IP efficiency reaches a saturation equilibration, thus resulting in efficiencies of 60-90%. This is dependent on factors such as antibody strength/dissociation constant (Kd) and steric hindrance of neighbouring antibodies by target protein binding (Encell *et al.*, 2012). This effect can be overcome by using an excess of antibody and capturing beads, but simultaneously results in the increased risk of contaminant binding partner identification. Thus, for low expression level proteins, such as transcription factors, large numbers of cells may be required. HaloTag was developed to bypass this equilibrium effect by covalent capture of HaloTagged- fusion proteins (Los *et al.*, 2008). HaloTag is a haloalkane dehalogenase protein that contains a mutated aspartate catalytic triad. This allows the activation and nucleophilic attack of haloalkanes to form a covalently bound intermediate but lacks the ability to hydrolyse the bond to release free enzyme (shown schematically in Figure 3.5). Thus HaloLink resin, coated with haloalkanes, covalently captures HaloTag- fusion proteins. For elution, a highly specific protease sequence can be incorporated into a linker sequence between the HaloTag and protein of interest to recover the protein and binding partners, away from the HaloTag and non-specific contaminants (Los *et al.*, 2008, Encell *et al.*, 2012 & Daniels *et al.*, 2014).

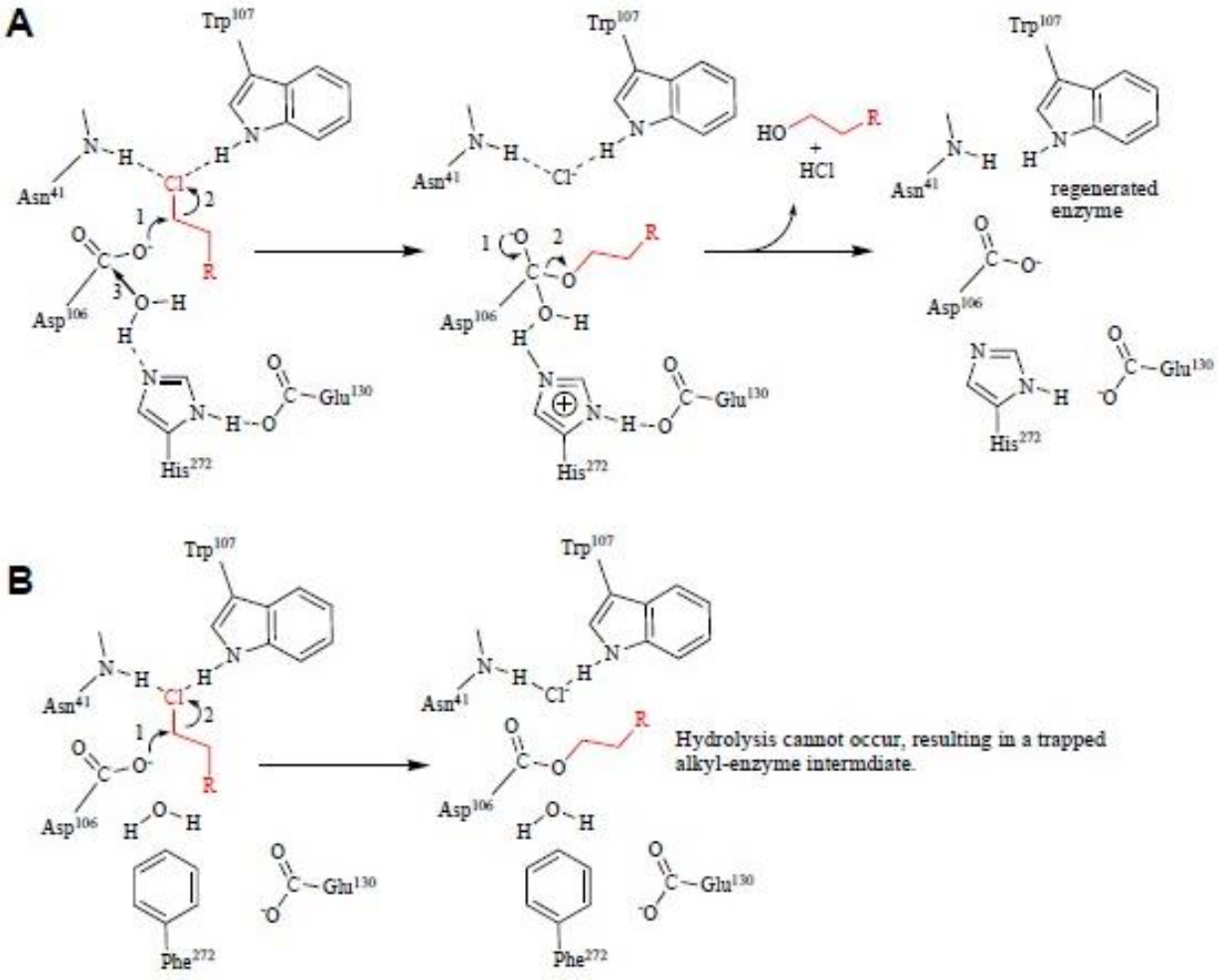


Figure 3.5: Schematic view of the catalytic mechanism of the Haloalkane dehalogenase protein versus the HaloTag technology.
Critical residues required for function are shown with hydrogen bonding (dotted lines) and movement of electrons (arrows). Red is the haloalkane molecule that is coupled to R, for example beads. A) Endogenous haloalkane dehalogenase, Asp106 acts as a nucleophile to attack haloalkanes to form the covalently bonded alkane-enzyme intermediate. Glu130 promotes the general base nature of His272 resulting in the hydrolysis of the alkane-enzyme bond, releasing free enzyme. B) HaloTag technology, mutation of His272 to Phe results in the inability to hydrolyse the covalently bound intermediate; thus covalently trapping HaloTag (Taken from Encell *et al.*, 2012). Asp = aspartic acid, His = histidine, Glu = glutamic acid, Trp = tryptophan, Phe = phenylalanine

Initially we needed to determine whether the exogenous expression of HaloTag-HIF-2α maintained the physiological properties of HIF, such as nuclear localisation. A benefit of the HaloTag is the ability for specific labelling and direct visualisation, using a cell permeable fluorescent dye (Los *et al.*, 2008), such as HaloTag- Oregon Green (Promega). HaloTag-HIF-2α was transfected into HeLa cells and visualised by fluorescent confocal microscopy (Figure 3.6). Inclusion of an untransfected control highlights that the labelling was specific to HaloTag expressing cells. At 10X magnification a poor transfection efficiency was observed at ~20% (Figure 3.6), however this could be artificially under-represented due to a potentially inefficient labelling step. Importantly, at 100X magnification nuclear localisation is punctate, as expected (Figure 3.6). Thus, IP optimisation was conducted.

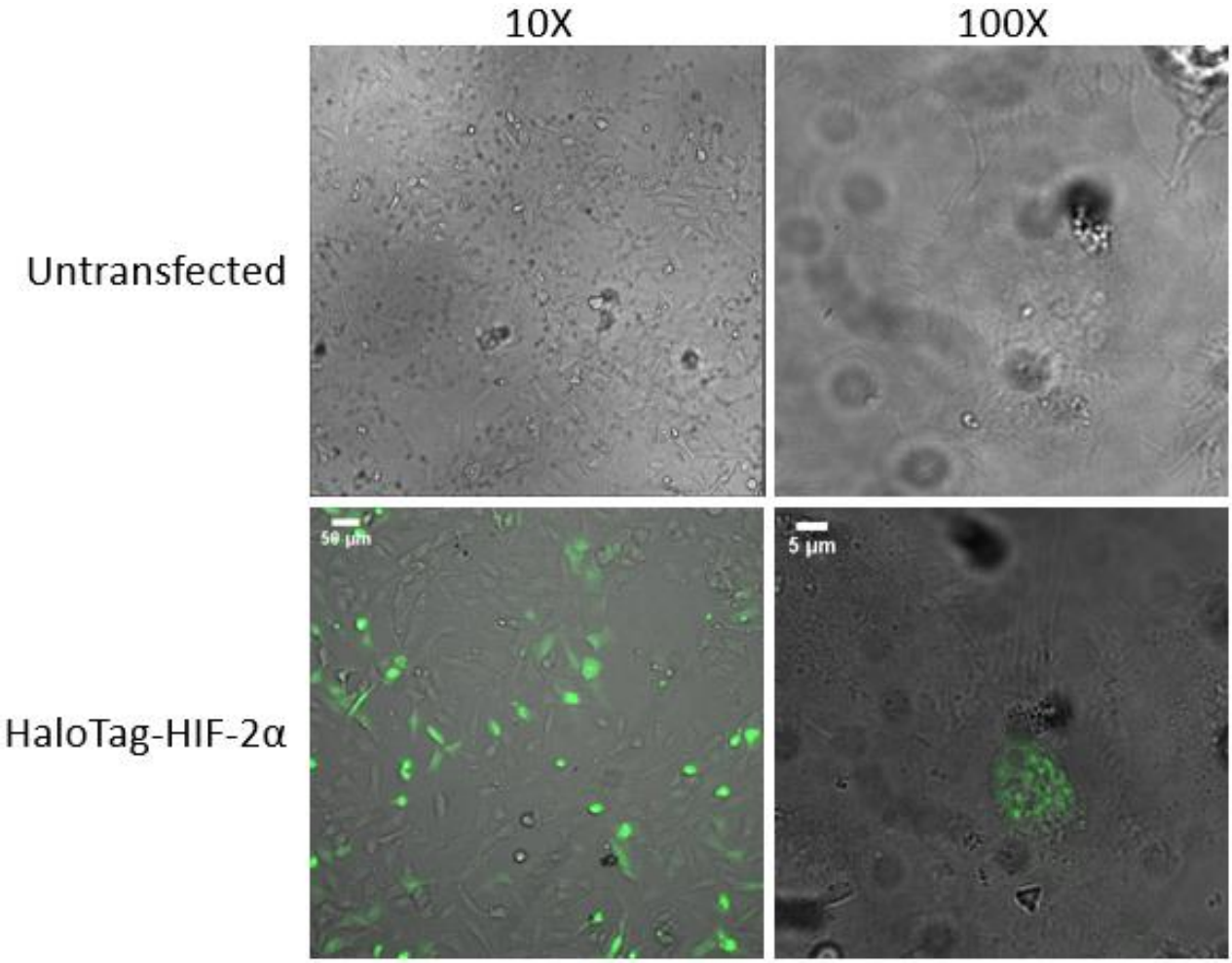


Figure 3.6: Transfection efficiency and nuclear localisation of HaloTag-HIF-2α.
Untransfected or HaloTag-HIF-2α transfected HeLa cells, using the PEI protocol as described and imaged following HaloTag- Oregon Green labelling, following the manufacturer's protocol. Single images were acquired using brightfield and at 488 nm fluorescence excitation at 10X and 100X magnification with a Zeiss LSM 780 microscope.

To obtain enough HaloTag-HIFα protein for LC-MS/MS analysis it was essential to ensure that transfection efficiency scaled proportionally with plate size. Due to the covalent nature of the HaloTag, either fluorescent labelling or IP can be performed per experiment. Hence, cells were grown in small dishes (as optimised) or large 10 cm dishes and transfected, scaling for plate surface area (~6 fold). Cells from both plate sizes were lysed, as for IP, and an aliquot of lysate was fluorescently labelled *in vitro*. Subsequent SDS-PAGE and fluorescent gel scanning allowed me to assess the protein expression levels of HaloTag-HIF-2α in comparison to a standard that was fluorescently labelled in live cells (Figure 3.7).

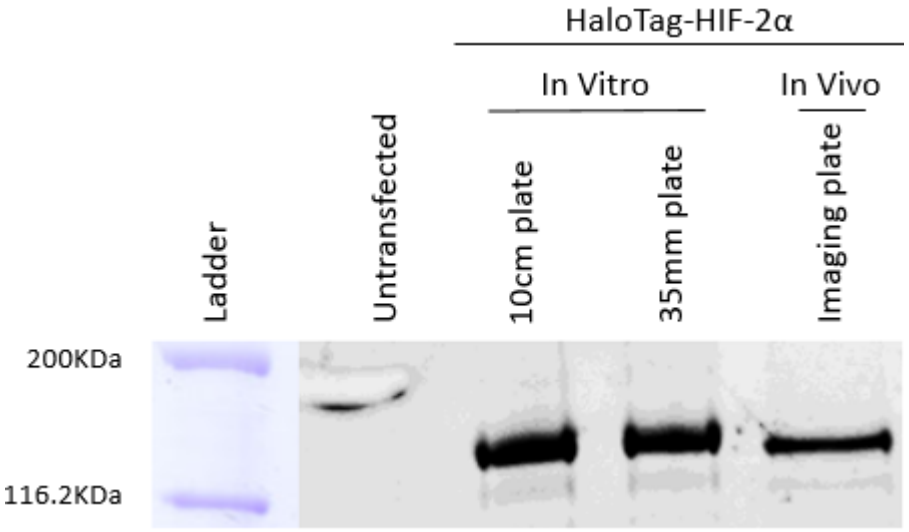


Figure 3.7: Transfection efficiency & scalability of HaloTag-HIF-2α exogenous expression.
A composite image of HaloTag- Oregon Green labelled untransfected or HaloTag-HIF-2α expressing HeLa cell lysates, grown as stated. Labelling was performed as indicated and lysis performed following manufacturer's IP protocol. Labelled lysates were analysed by SDS-PAGE and fluorescently imaged on an ImageQuant LAS4000 system, using the Cy2 filter, and then stained colloidal coomassie.

Similar to in culture labelling, *in vitro* labelling is specific to HaloTag expressing cells (Figure 3.7). Comparing identical plate sizes of the in culture labelled standard to the *in vitro* labelled lysate shows an ~1.5 fold increase in fluorescent intensity, thus ~50% transfection efficiency (determined by densitometry, data not shown). This apparent increase is likely the result of a more efficient labelling technique *in vitro* than in culture, confirming the potential under-estimation of transfection efficiency when live cell imaging, as discussed earlier. Similar band intensities were observed between small (35 mm) and large (10 cm) cell culture plates, showing that the transfection efficiency was unaffected by plate size. Therefore, large plates were prepared for initial IP optimisations. Protocols selected as an initial basis included the manufacturer's recommended protocol and two published protocols from Ohana *et al.*, 2011 and Daniels *et al.*, 2014. These protocols are referred to as Method 1, Method 2 and Method 3 respectively from now on.

All methods evaluated used the manufacturer recommended lysis buffer, short bead binding times and TEV based elution. However, physical perturbations to aid cell lysis differed: method 1 uses a 25-gauge needle, method 2 uses sonication and method 3 uses -80 °C freeze-thaw cycles. All cell lysis methods are followed by centrifugation to remove cellular debris. Post lysis, all samples were analysed through *In vitro* lysate labelling and SDS-PAGE analysis. A lack of a fluorescent band was observed for each tested method, suggesting a failure in transfection. However, an identically prepared plate labelled in culture and analysed by fluorescence confocal microscopy showed adequate levels of

transfection (data not shown). Combined, these data suggest a potential loss of the HaloTag-HIF-2α protein during lysate preparation, which was not observed previously. Differences between the previously used *in vitro* labelling protocol and the IP protocol is the inclusion of a centrifugation step to remove insoluble cellular debris in the latter. Thus, to check the effects of centrifugation, lysate and pellets were analysed by western blot (Figure 3.8).

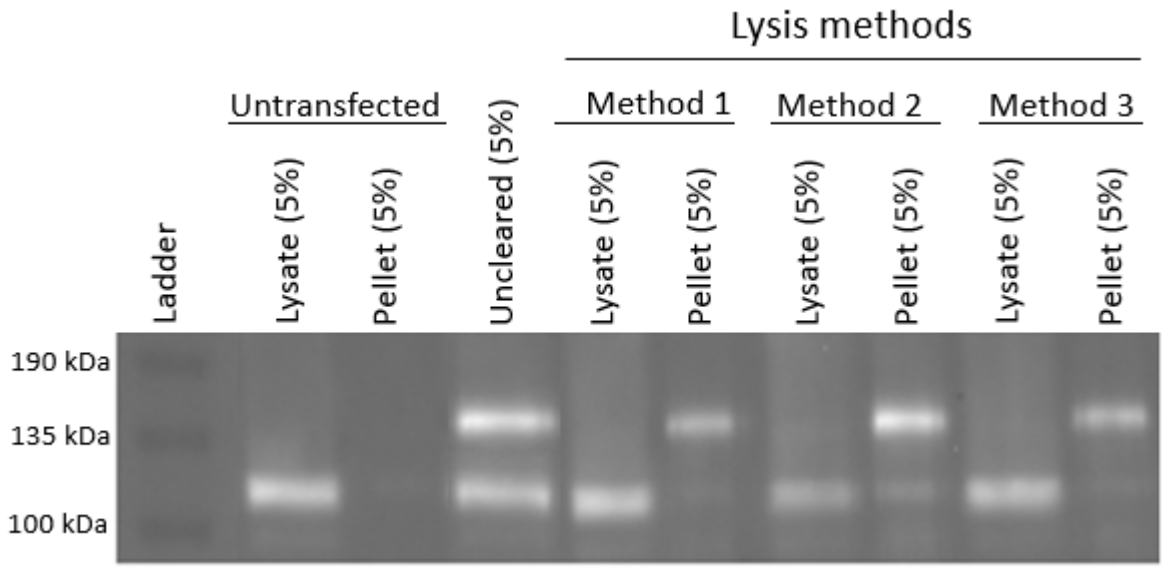


Figure 3.8: Analysis of HaloTag-HIF-2α solubility when lysed under different conditions.
Western blot analysis of untransfected and HaloTag-HIF-2α transfected HeLa cells, probed with an anti-HIF-2α primary antibody. Each method uses a lysis buffer of: 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT, 0.5 mM EDTA and 0.005% NP-40. Method 1: manufacturer recommended, lysis by passing through a 25-gauge needle. Method 2: Ohana *et al.*, 2011, lysis by sonication, maximum power for 10 s on 10 s off for 2 min. Method 3: Daniels *et al.*, 2014, lysis by a 1 hr -80 ^{o}C Freeze-Thaw cycle. Post lysis, lysates were centrifuged 10 min at 13,000 *g* and cleared supernatant collected. Cellular debris pellets were resuspended in Laemmli's buffer. Method 1 was used for the untransfected and uncleared lysates, but without the centrifugation step for the latter.

Using an anti-HIF2α antibody provides the advantage of simultaneous investigation of the endogenous and exogenously expressed HIF-2α proteins. An untransfected control shows that lysis is efficient and the endogenous HIF-2α protein is soluble, with a single band of ~120 kDa in the cleared lysate lane only (Figure 3.8). For the transfected samples, the uncleared lysate shows a doublet band pattern equating to endogenous HIF-2α and the HaloTag-HIF-2α protein, at ~150 kDa. All three tested lysis methods result in the removal of the HaloTag-HIF-2α protein, shown by a single 150 kDa band in the resuspended cellular debris lanes only. The solubility of endogenous HIF-2α protein was unaffected by the exogenous expression of HaloTag-HIF-2α protein, yet the HaloTag-HIF-2α protein was clearly lost from the lysate fraction post centrifugation. Without centrifugal clearing, or the ability to pre-clear samples, the ratio of non-specific interactors will inevitably increase

dramatically, which may mask the signal of true binding partners. Thus, HaloTag based IP was abandoned and GFP tagged IP approaches were explored.

3.4.4. GFP-Trap IP optimisation

GFP-Traps (CHROMOTEK) are commercially available, pre-conjugated antibody-beads designed for the IP of GFP-tagged proteins. Unlike traditional antibody IP approaches, GFP-Traps utilise the development of nanobodies, which are comparable to a single heavy chain of antibodies but retain epitope recognition. Due to their more simplistic nature, nanobodies can be created with much greater affinities; usually with dissociation constants (Kd) in the pM range (reviewed by Muyldermans, 2013). GFP-Traps are claimed to have a Kd of <1 pM and are available conjugated to different matrices; agarose coupled are used throughout these studies.

GFP-HIF-2α-DM was previously used during transfection optimisation (Figure 3.4 (B)), as such localisation studies were not repeated. Previous data from the Sée lab, and manufacturer's statements, suggests orientation of the GFP-tag in relation to the protein of interest can be essential for successful IP. Thus, both GFP-HIFα-DM and HIFα-DM-GFP were tested for IP efficiency (Figure 3.9 (A)). By comparing elution efficiency it was clear that N-terminally tagged GFP-HIFα-DM constructs are required for IP; with ~90% efficiency compared to an undetectable band for C-terminal HIFα-DM-GFP constructs (Figure 3.9 (A)). C-terminal HIFα-DM-GFP constructs appear to bind GFP-Traps, because of the lack of a band in unbound lanes, but lack an elution band. This result is unlikely to be due to failure to elute, because of the use of 5X Laemmli's buffer (10% (w/v) SDS). Alternatively, a more likely scenario is that weak binding interactions between GFP traps and C-terminal HIFα-DM-GFP may result in the dissociation of the HIFα-DM-GFP protein from the GFP-Traps during washing.

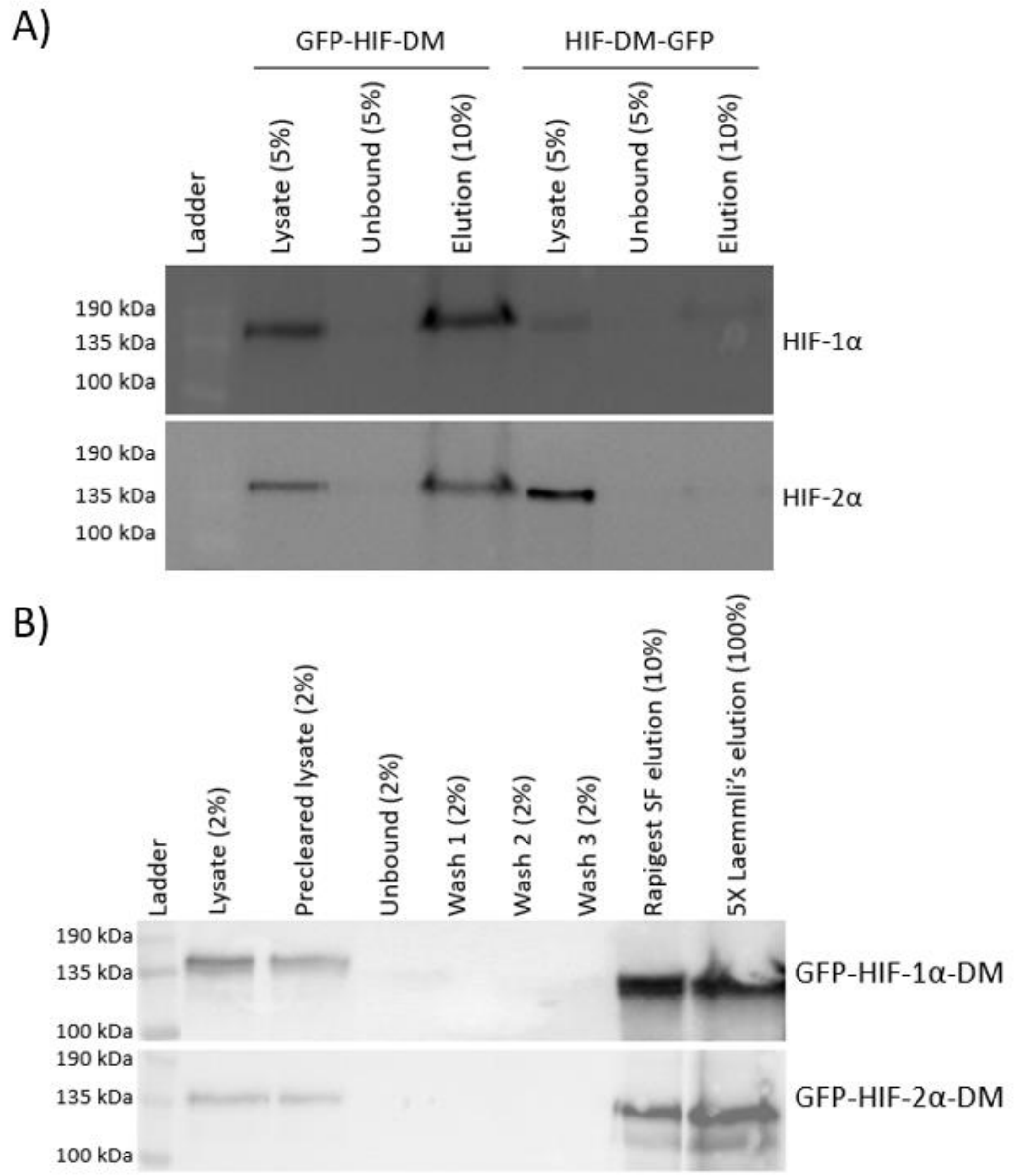


Figure 3.9: The development of a mass spectrometry compatible GFP-Trap IP protocol.
HeLa cells were grown in 21% O_2 and transfected with the indicated plasmid. Cells were lysed 24 hr post transfection in 50 mM Tris pH 8.0, 120 mM NaCl, 1 mM EDTA, 0.5% NP-40 supplemented with EDTA free cOmplete protease inhibitors (Roche) and phosSTOP (Roche). Lysate was diluted to 0.2% NP-40 in an identical lysis buffer lacking NP-40. IP was performed according to the manufacturer's protocol. Western blots were performed with a primary anti-GFP antibody. A) Efficiency of IP using N-terminal and C-terminal GFP tagged HIFα-DM constructs, elution with 5X Laemmli's buffer. B) Development of a GFP-Trap IP protocol compatible with high throughput mass spectrometry analysis. Changes include a preclearing step with bab-20 beads, HPLC grade buffers, addition of 2 X 25 mM Ambic washes, and elution with 1% RapiGest SF in 25 mM Ambic pH 8.0. A secondary elution in 5X Laemmli's was performed subsequently.

N-terminal GFP-HIFα-DM was efficiently immunoprecipitated for both HIFα proteins, therefore further optimisations focused on modifications to create a MS compatible protocol and limit non-specific binding partners (Figure 3.9 (B)). To this end, all buffers were made with HPLC grade components, a preclearing step and elution in a MS-compatible buffer were all employed. bab-20 beads (CHROMOTEK) have an identical composition of

matrix compared to the GFP-Traps, but lack the nanobody conjugation. Thus, incubation of bab-20 beads with cleared lysate, prior to IP, can remove non-specific binding partners resulting from matrix association (pre-clear). RapiGest SF (Waters) is an acid labile, mild detergent that can be degraded into an insoluble product (and a volatile soluble component) for removal by centrifugation (Yu *et al.*, 2003). Thus, RapiGest SF elution allows the recovery of proteins and HTP MS analysis without the need for sample clean-up steps.

Western blot analysis of the modified protocol (Figure 3.9 (B)) shows that pre-clearing does not affect the levels of available GFP-HIFα-DM protein, shown by near identical band intensities of lysates pre- and post- preclearing lanes. Importantly, RapiGest SF elution is highly efficient, resulting in ~80% IP efficiency of both GFP-HIF-1α-DM and GFP-HIF-2α-DM proteins (Figure 3.9 (B)). Although a prominent band exists with the secondary elution, it is crucial to note that 100% of this elution was loaded for analysis, 10 fold greater than the RapiGest SF elution. Hence, the secondary elution is artificially abundant for direct comparisons. Once the IP protocol was developed, WildType (WT) GFP-HIFα constructs were required for biological relevance of IP experiments. Previous WT HIFα plasmids obtained by the Sée lab were found to have multiple point mutations that result in amino acid substitutions, therefore new WT HIFα constructs were obtained and cloned to contain the GFP tag.

3.5. WildType HA-Clover-HIFα Cloning:

A WT HA-HIF-1α plasmid was obtained (a gift from the Prof Rocha lab) and a GFP isoform (Clover) was cloned into the plasmid for GFP-Trap IP, and allow further downstream microscopy analysis. Although a HA-tag is present in this construct there is no guarantee that the previously optimised HA IP protocol within the Eyers lab would be successful. Clover is a GFP isoform with greater dynamic range and photostability, thus is preferable for advanced microscopy techniques (Lam *et al.*, 2012). GFP-Traps are stated to be as efficient for IP when using Clover. To evaluate this, a Clover- only plasmid was immunoprecipitated prior to cloning. Western blot analysis showed an efficiency of ~80% for Clover IP (data not shown). Thus, insertion of the clover gene into the HA-HIF-1α plasmid was conducted (HA-Clover-HIF-1α), and subsequent cloning of HA-Clover-HIF-2α and HA-Clover only plasmids performed, using the In-Fusion cloning (TakaraBio) technology. All plasmid cloning is schematically depicted in Figure 3.10.

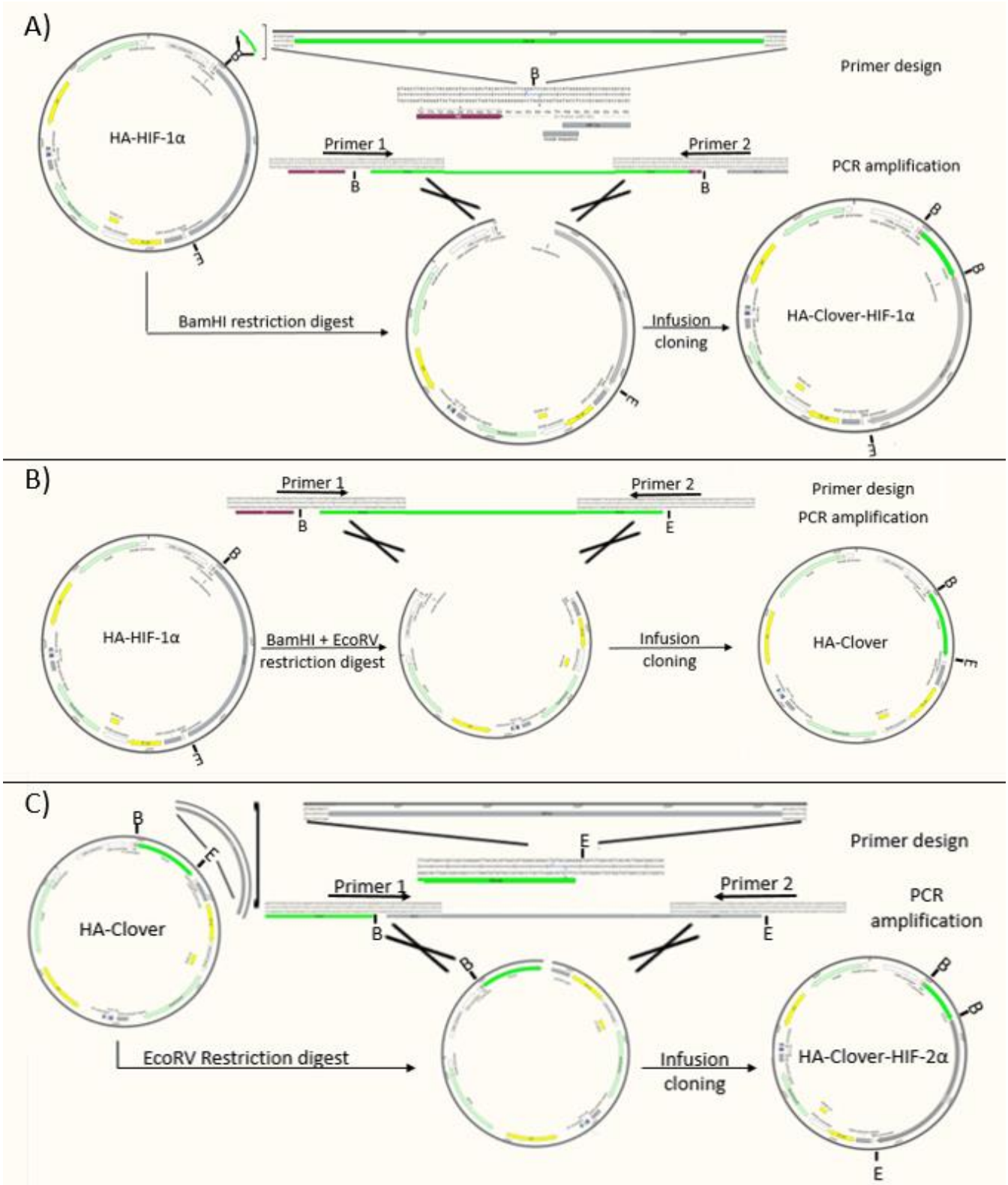


Figure 3.10: Schematic view of the In-Fusion cloning strategy used for the creation of HA-Clover-HIFα and HA-Clover plasmids.
Green = Clover sequence, Grey = HIFα sequences, yellow/pale green = regulatory domains of plasmid. Restriction enzyme sites (B=BamHI, E=EcoRV) and primer designs for each plasmid are included A) HA-Clover-HIF-1α from HA-HIF-1α. B) HA-Clover from HA-HIF-1α. C) HA-Clover-HIF-2α from HA-Clover. Plasmid maps viewed in SnapGene software (SnapGene software (from GSL Biotech; available at snapgene.com)).

Initially it was presumed that the HA-tag would have minimal effect on IP efficiency of HA-Clover-HIFα, due to its short length of 9 amino acids. Therefore, once cloned, HA-Clover-HIF-1α was tested for IP efficiency with GFP-Traps; confirming this assumption with an efficiency of ~85% (Figure 3.11). Thus, cloning of HA-Clover and HA-Clover-HIF-2α was performed (Figure 3.10 (B) and (C) respectively). The HA-Clover only construct was used as a negative control to allow contaminant binding partner identification and background subtraction of non-specific binding partners associated with the HA-Clover tag.

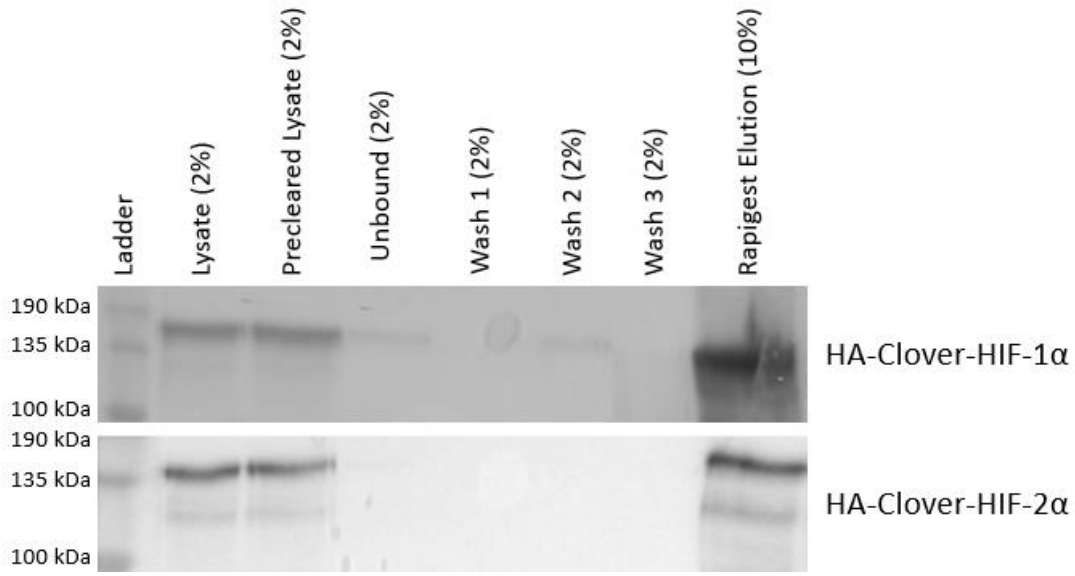


Figure 3.11: Determination of the suitability of the HA-Clover tag for IP.
Western blot analysis of HeLa cells transfected with the indicated plasmid 24 hr prior to 4 hr incubation at 1% O_2. The mass spectrometry compatible GFP-Trap IP protocol was followed and probed with an anti-GFP primary antibody.

All plasmids were validated by DNA sequencing using GATC Biotech, sequencing primers listed in Table 2.5. As before for HA-Clover-HIF-1α, HA-Clover-HIF-2α was immunoprecipitated and western blotting performed to ensure successful IP with the HA-Clover tagged construct (Figure 3.11). This shows a similar IP efficiency of ~85% is obtained. To obtain more biologically relevant data, I undertook final optimisation steps to test the feasibility of low level exogenous expression systems for in-depth PTM analysis.

3.6. Final optimisations:

3.6.1. Exogenous expression levels

The newly cloned HA-Clover-HIFα plasmids are under the control of different regulatory elements (in the plasmid backbone) compared to the previously used, and optimised, GFP-HIFα-DM plasmids. Thus, identical transfection conditions may result in different expression

levels. In order to optimise the exogenous expression levels of HIFα, maintaining physiological O_2 dependent protein stability and transcriptional activity, a reporter assay using a luciferase plasmid under the control of the Hypoxia Responsive Element (HRE) was used (HRE-Luciferase). Hence, luciferase signal is representative of the level of active HIF transcription. The optimised PEI transfection protocol was used and the amount of HA-Clover-HIF-1α was sequentially altered (Figure 3.12). HA-Clover-HIF-1α exogenous expression was selected for optimisations due to the endogenous protein being completely O_2 dependently degraded, thus aberrant exogenous expression at 21% O_2 is easy to identify. To maintain the overall quantity of DNA in transfection mixes, the addition of a transcriptionally inactive empty vector was used, pcDNA3.1(-).

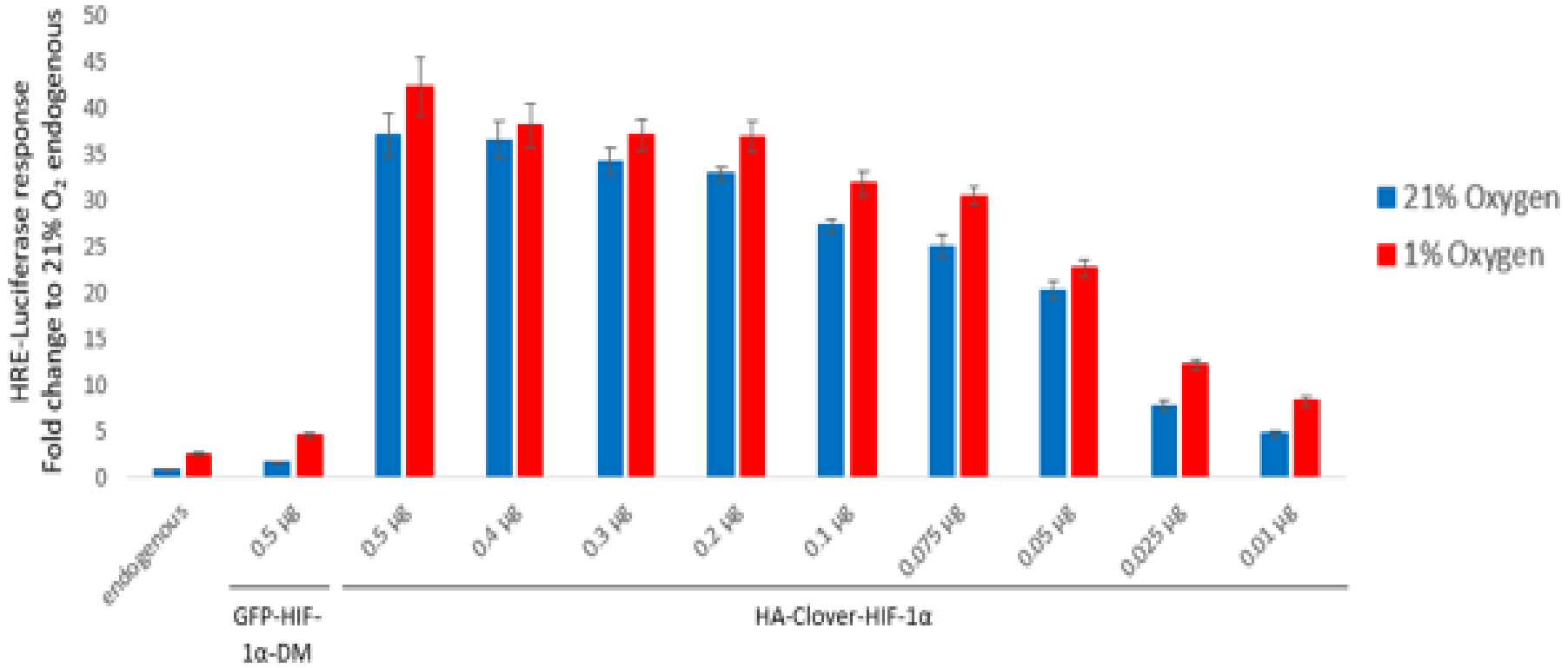


Figure 3.12: Expression levels of HA-Clover-HIF-1α, determined by HRE-Luciferase assay.
The transcriptional activity of HA-Clover-HIF-1α transfection at a range of 0.5 µg – 0.01 µg plasmid DNA in HeLa cells. Cells incubated at 21% O_2 for 24 hr post transfection followed by incubation as indicated for an additional 6 hr. Optimised GFP-HIF-1α-DM transfection and HRE-Luciferase only controls included. Data shown as fold change to the HRE-luciferase only (endogenous) 21% O_2 sample. Data from 3 biological replicates and standard error plotted.

The optimised expression levels of GFP-HIF-1α-DM result in <2 fold greater transcriptional activity compared to endogenous HIF-1α (Figure 3.12). An O_2 dependent induction of transcription of fully stable HIF-1α is unsurprising considering the second O_2 dependent regulatory mechanism involving FIH-1 (Lando *et al.*, 2002), see introduction (1.6). A dramatic increase in transcriptional output was observed (~35 fold) for identical transfection conditions using the HA-Clover-HIF-1α plasmid. By using decreased amounts of the HA-Clover-HIF-1α plasmid (as low as 10 ng (0.01 μg)) an O_2 dependent transcriptional activity could be observed, at a reasonable level compared to endogenous conditions. To highlight the transcriptional output of the cloned HA-Clover-HIF-1α, at 10 ng of plasmid there was an observable ~2 fold increase in HRE-luciferase signal compared to the optimised (0.5 μg (50 X more DNA)) GFP-HIF-1α-DM conditions.

There are two possible explanations for these observations: 1) induction of expression from the HA-Clover-HIF-1α plasmid is orders of magnitudes higher than the GFP-HIF-1α-DM plasmid, therefore although HA-Clover-HIF-1α is O_2 dependently degraded, the quantity of HA-Clover-HIF-1α protein outweighs the system flux, and thus accumulates in normoxia. 2) The GFP-HIF-1α-DM protein is highly inefficient at initiating transciption. Possible explanations for the latter include: a short linker sequence resulting in steric hindrance with the GFP-tag, or the known point mutations result in amino acid substitutions that effect HIF-1α transactivation, although being fully stable. Sequencing the full length HIF-1α gene, from the GFP-HIF-1α-DM plasmid, identified six amino acid substitution mutations, besides the known proline hydroxylation sites: 13N insertion, K14R, V87M, D544G, K548R and T627I. Although removing the prospect of direct comparisons between constructs, the identification of these mutations may be helpful in the future when identifying PTMs to characterise.

Considering in-depth discovery PTM analysis requires a large quantity of protein, we decided to benchmark a low expression level (fully O_2 dependently degraded) against a high expression level (maximal expression without observable effects on nuclear localisation, cell stress and cell viability) to determine the scale required for MS analysis, and feasibility of using a low expression model. Therefore, similar to Figure 3.12, sequentially decreasing levels of HA-Clover-HIF-1α plasmid was transfected into cells and investigated using microscopy and western blotting techniques.

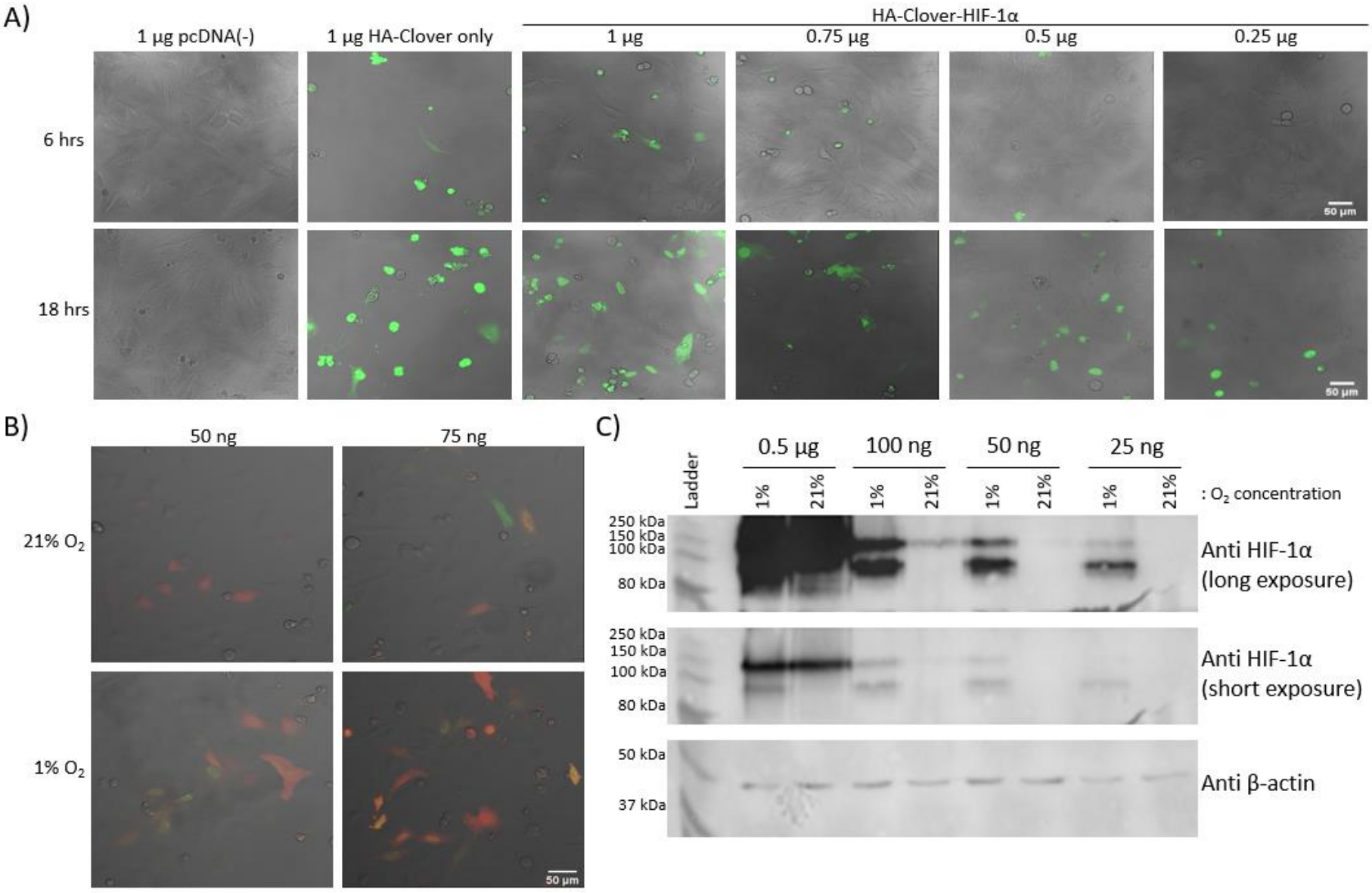


Figure 3.13: Creation of low and high expression HA-Clover-HIF-1α models.
HeLa cells transfected using the optimised PEI:DNA ratio protocol varying the quantity of HA-Clover-HIF-1α plasmid as indicated, maintaining the DNA ratio using a pcDNA3.1(-) empty vector. A) Fluorescence confocal microscopy analysis of transfections incubated with cells for different lengths of time, media changes performed at stated times. Cells incubated at 21% O_2 until 24 hrs post initial transfection, brightfield and 488 nm fluorescence images taken using a Zeiss LSM780. B) Live cell fluorescence confocal microscopy analysis of transfections incubated 18 hrs before media change. Brightfield, 488 nm and 561 nm wavelength fluorescence images taken at 21% O_2 and again 8 hrs after 1% O_2 incubation. Cells co-transfected with pcDNA-mRUBY. C) Western blot analysis of the high expression model and range of low expression models determined from (B). Cells incubated at stated O_2 concentrations 8 hrs prior to lysis in the GFP-Trap lysis buffer. Probed with a anti-HIF-1α antibody or anti-β-actin loading control.

Complementary to luciferase data, transfection of 1 µg HA-Clover-HIF-1α resulted in vast protein expression at 21% O_2. Additionally microscopy analysis identified that 1 µg of HA-Clover-HIF-1α resulted in significant cell death rates of ~70% (Figure 3.13 (A)). This result was mimicked with the HA-Clover only plasmid, however no significant cell death was observed with the empty vector control. Therefore, combined, these data suggest that the HA-Clover tagged plasmids are causing lethality, likely due to their high expression levels. To reduce expression levels, we tried lowering the length of timings that transfection mixes were incubated with cells and/or reducing the quantity of HA-Clover-HIF-1α plasmid DNA in the transfection mix (Figure 3.13 (A)).

Reduction of the transfection time to 6 hr resulted in poor transfection efficiencies (<10%) at any DNA quantity, however maintained a high cell death rate in fluorescent cells (Figure 3.13 (A)). Reduction of HA-Clover-HIF-1α plasmid maintained high transfection efficiences whilst reducing the rate of cell death. At 0.5 µg HA-Clover-HIF-1α (per 200,000 cells in a 35 mm plate), cell death rates were comparable to the empty vector controls (<10%) and maintained nuclear only expression. Thus 0.5 µg equivalents for transfection were selected as the high expression level (Figure 3.13 (A)).

To create the low expression level system, which is fully O_2 dependently degraded, clover fluorescence was to be visualised in an O_2 dependent manner. Thus, fluorescence imaging was performed pre- and post- 1% O_2 incubation. Thus, an absence of fluorescence at 21% O_2 could be explained by a failure to transfect, or an expression level too low for detection. Therefore, a different wavelength fluorescent protein encoding plasmid (pcDNA3-mRUBY2) was included as an internal transfection control, in order to monitor transfection efficiency (Figure 3.13 (B)).

A time point of 8 hr of 1% O_2 incubation was used for microscopy analysis, as this is when maximal endogenous HIF-1α protein accumulation occurs (Bagnall *et al.*, 2014). At 50 ng (0.05 µg) DNA (per 200,000 cells in a 35 mm plate) clover fluroescence was found to be O_2 sensitive, and maintaining a transfection efficiency of ~50% (Figure 3.13 (B)). Greater or reduced DNA concentrations resulted in either abnormal expression at 21% O_2 or undetectable expression levels respectively (Figure 3.13 (B)). Thus 50 ng equivalents were selected as the low expression level.

To validate microscopy data, the same transfection condtions were analysed by western blotting (Figure 3.13 (C)). Probing with anti-HIF-1α shows that the HA-Clover-HIF-1α protein (~150 kDa) at the low expression level had an identical banding pattern to endogenous HIF-1α protein (~120 kDa); having complete O_2 dependent degradation (Figure 3.13 (C)). Supporting microscopy data, the high expression level of HA-Clover-HIF-1α was not O_2 sensitive, shown by equally intense bands at both O_2 tensions (Figure 3.13 (C)). Comparisons between low and high expression models finds an ~20 fold increase in exogenous expression in the latter (determined by densitometry analysis, data not shown).

Interestingly, the large overexpression of HA-Clover-HIF-1α does not result in the dysregulation of the O_2 dependent degradation pathways for endogenous HIF-1α, which maintained 21% O_2 degradation (Figure 3.13 (C)). It could be argued that the large overexpression of HA-Clover-HIF-1α in 21% O_2 may be the result of overloading the endogenous system flux, thus maintaining endogenous HIF-1α regulation is surprising.

3.6.2. Mass spectrometry optimisations

As mentioned, discovery proteomics requires a considerable quantity of protein to be purified for analysis. Therefore, with the development of high expression and low expression models, we decided to benchmark both expression levels for LC-MS/MS analysis, post Titanium Dioxide (TiO_2) phospho-peptide enrichment (TitanSphere, GL Sciences). This will aid in the determination of whether scaling the low expression (more biologically relevent) model is feasible (Figure 3.14). Figure 3.14 (A) shows western blot analysis of 600 µg lysate (1x 10 cm plate) IPs from both the high and low expression levels. For IP, lysate requires a 2.5 fold dilution; likely explaining the lack of a band for endogenous HIF-1α or HA-Clover-HIF-1α in the low expression model whole cell lysate. Importantly, Figure 3.14 (C) shows that the high expression level model obtains ~20 fold more HA-Clover-HIF-1α protein post IP than compared to the low expression model (in agreement with Figure 3.13 (C)). Hence, 20 X 10 cm plates would be required for similar levels of HA-Clover-HIF-1α post IP.

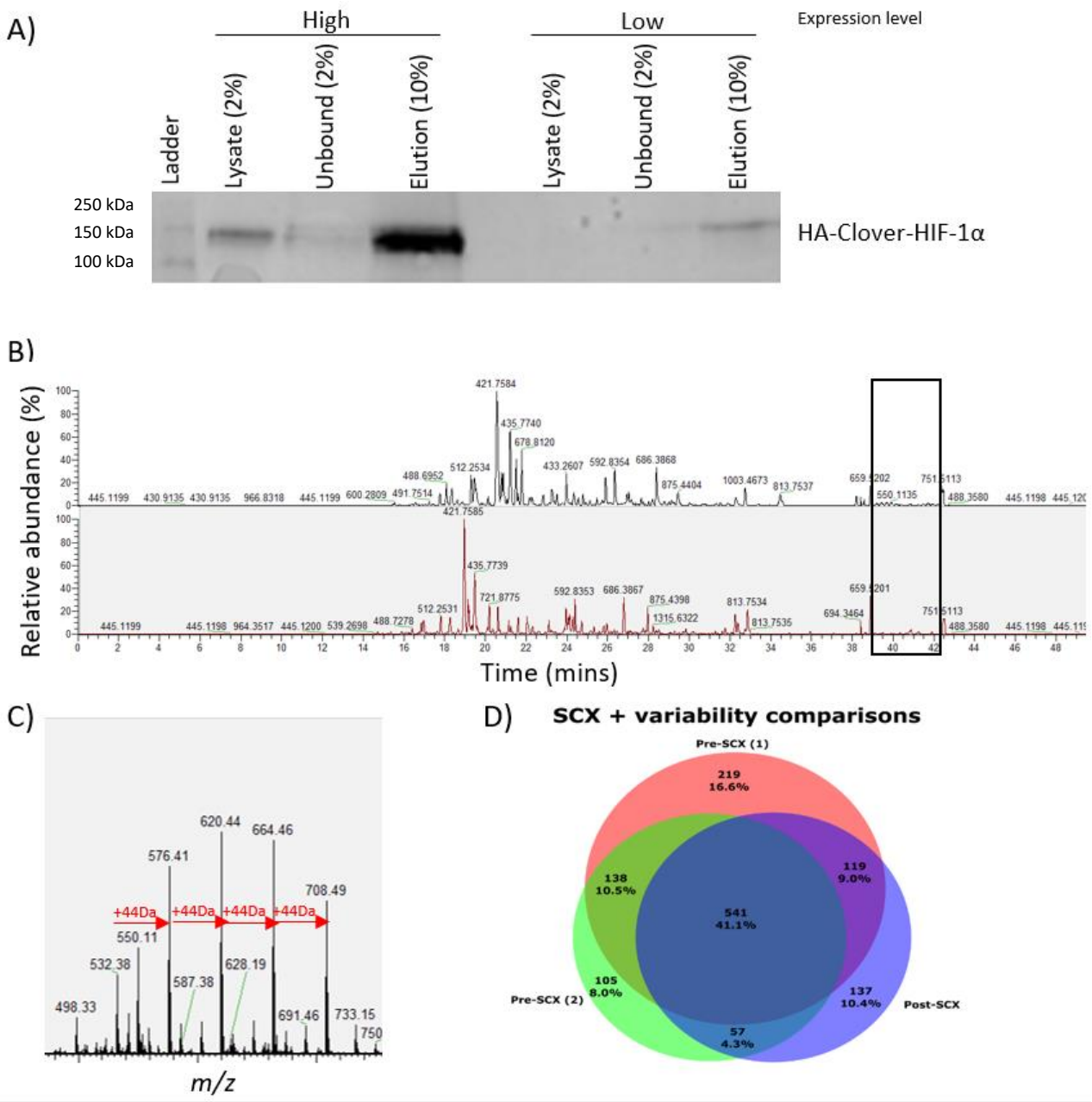


Figure 3.14: IP model comparison and assessment of the variability introduced by SCX based removal of PEG. A) Western blot analysis of 600 µg IPs of high and low HA-Clover-HIF-1α expression models in HeLa cells incubated with 4 hr 1% O_2. B) BPC's of trypsin digested IPs pre- (white) and post-SCX (grey). C) Enlarged boxed region of pre-SCX BPC of (B), highlighting the +44 Da repeating unit (red arrows). D) Venn diagram of proteins identified at 1% FDR confidence from identical injections pre-SCX at time point 0 (1), 24 hr later (2) and post SCX. Includes the number of proteins and equivalent percentage for each overlap.

Trypsin digested IPs were split 5% for LC-MS/MS analysis (to analyse sequence coverage and binding partners, Figure 3.14 (B)) and 95% for TiO_2 enrichment (for phospho-peptide identification and site localisation), following the HCD Orbitrap only method from Ferries *et al.*, 2017. Digest injections found a repeating unit contamination of +44 Da at ~5% relative intensity, likely to be Polyethylene Glycol (PEG) (Figure 3.14 (B) and Figure 3.14 (C)). TiO_2 enrichment for phospho-peptides resulted in the co-enrichment of the PEG contamination,

to much greater levels than any phospho-peptides (data not shown). A 'mock' control IP, using identical buffers without cellular lysate, lacked this PEG contamination (data not shown). Thus, the likely source of contamination is during lysate preparatory stages in cell culture plates.

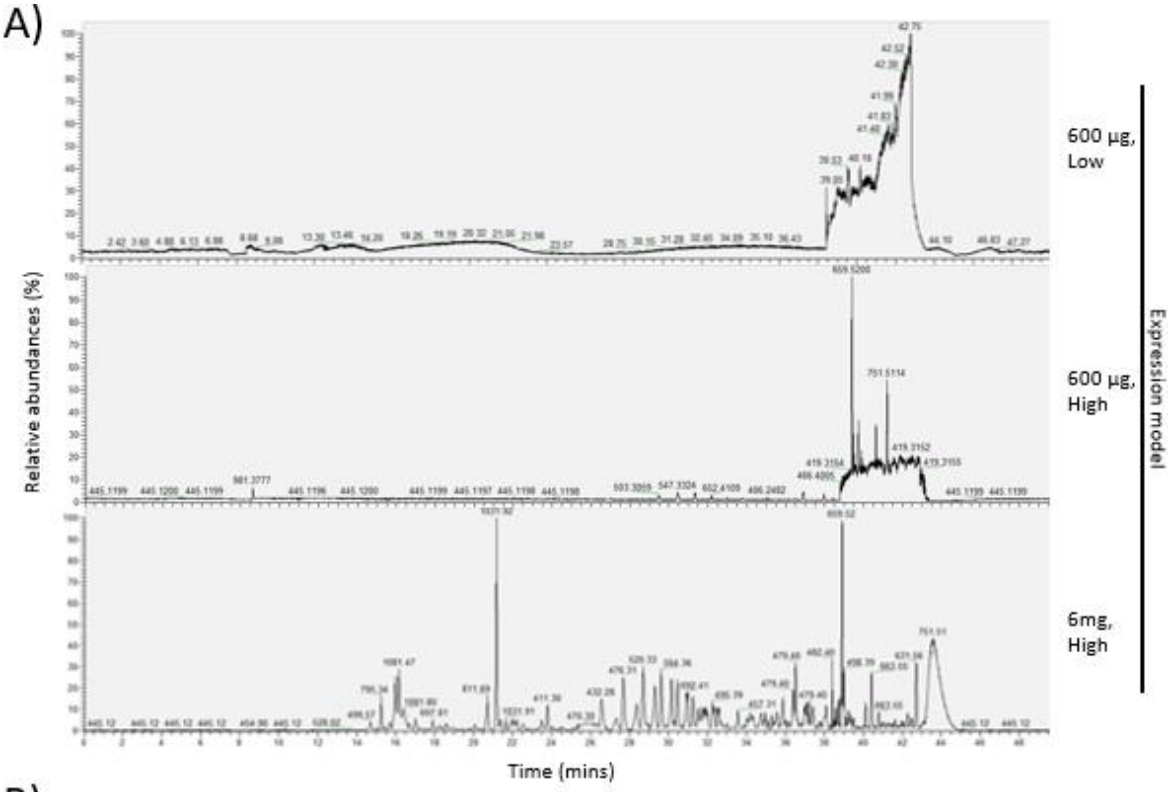
With the requirement of using a hypoxic chamber to generate hypoxic samples, there is not an easy and reliable method to incorporate a trypsinisation protocol during cellular lysis. Therefore, a strategy to remove PEG post IP and digestion was adopted, using a Strong Cation Exchange (SCX) stage-tip protocol. LC-MS/MS analysis of digests post SCX shows the PEG contaminant was successfully removed, however the Base Peak Chromatogram (BPC) was simultaneously affected (Figure 3.14 (B)). It is likely that SCX may cause the loss of peptides in a stoichiometric fashion. However, SCX incorporation results in the addition of 24 hr of sample preparation time prior to LC-MS/MS analysis, hence BPC variations could be the result of analytical variance. To investigate the variability induced by SCX, a large IP was performed and split equally for analysis in a pre-SCX state, at the time points of 0 hr and 24 hr, whilst SCX was performed and analysed alongside the 24 hr time point.

Data was analysed through a Proteome Discoverer coupled MASCOT pipeline and filtered to a 1% False Discovery rate (FDR), matching protein identifications were determined using the BioVenn tool (Hulsen *et al.*, 2008) (Figure 3.14 (D)). Figure 3.14 (D) found that the 2 pre-SCX samples had ~65% of all identified proteins matched, hence a baseline of ~35% variability. The majority of this variance is accounted for by substantially more protein identifications at the initial 0 hr time point analysis, likely reflecting peptide instability and precipitation over time. Figure 3.14 (D) also shows that post-SCX identifications matched ~80% of the identified proteins that were identified in the 24 hr pre-SCX time point control. Additionally, ~70% of all proteins identified in both pre-SCX time points were also identified post-SCX, thus SCX inclusion does not appear to introduce any further variation into LC-MS/MS based protein identification Figure 3.14 (D).

IPs (600 µg) of high and low expression level lysates were repeated, with the incorporation of the SCX PEG removal and TiO_2 enrichment protocols. BPC's of TiO_2 enriched IPs show very small peaks in the high expression level samples only (Figure 3.15 (A)). This intensity is far below the fragmentation intensity threshold required for discovery analysis, hence no peptide identifications were obtained. Considering that a ~20 fold increase in HA-Clover-HIF-1α is obtained by high expression level IP (Figure 3.14 (A)), the low expression model

would be insufficient for discovery PTM analysis without very large experiments, therefore this route was abandoned.

A)



B)

Peptide	modifications	score	E-value
SPNVLSVALsQR	S10(Phospho)	59	2.50E-05
QSsPEPNsPSEYcFYVDSDmVNEFK	S3(Phospho); S8(Phospho); C13(Carbamidomethyl); M20(Oxidation)	48	2.90E-04
LEPNPESLELSFTmPQIQDQTPsPSDGSTR	M14(Oxidation); S23(Phospho)	46	5.60E-04
QSsPEPNSPSEYcFYVDSDmVNEFKLELVEK	S3(Phospho); C13(Carbamidomethyl); M20(Oxidation)	39	2.60E-03
SHPRsPNVLSVALSQR	S5(Phospho)	23	1.00E-01
ILIAsPsPTHIHK	S5(Phospho); S7(Phospho)	30	2.10E-02
ILIASPsPTHIHK	S7(Phospho)	45	7.10E-04
ETTSATSsPYRDTQSR	S8(Phospho)	32	1.10E-02
QSSPEPNsPSEYcFYVDSDmVNEFK	S8(Phospho); C13(Carbamidomethyl); M20(Oxidation)	67	3.70E-06
LQNINLAmsPLPTAETPKPLR	M8(Oxidation); S9(Phospho)	48	3.10E-04
LQNINLAmsPLPTAEtPKPLR	M8(Oxidation); S9(Phospho); T16(Phospho)	25	6.60E-02
ETTsAtSSPYRDTQSR	S4(Phospho); T6(Phospho)	25	6.50E-02
VESEDtSSLFDK	T6(Phospho)	56	5.00E-05
SSADPALNQEVALK		35	5.60E-03
SIYEYYHALDSDHLTK		28	3.40E-02
LLGQSmDESGLPQLTSYDcEVNAPIQGSR	M6(Oxidation); C19(Carbamidomethyl)	24	8.80E-02
THHDmFTK	M5(Oxidation)	21	1.50E-01

Figure 3.15: TiO_2 enrichment phospho-peptide detection and scaling.
A) BPC's of HA-Clover-HIF-1α IPs post TiO_2 enrichment. IPs of model and quantity of protein as stated from HeLa cells incubated with 4 hr 1% O_2. B) Table of identified HIF-1α peptides at a 1% FDR confidence. Data from 6 mg High expression model IP. Analysis through Proteome Discoverer 1.4 using the MASCOT search engine. Fixed modifications = Carbamidomethlation of Cysteine. Variable modifications = phosphorylation of Serine, Threonine or Tyrosine and oxidation of methionine. Table includes peptide sequence (lower case letter symbolises modified residue), modifications and position, peptide MASCOT score and E-value.

A 10 fold increase of initial high expression lysate (10 X 10 cm plates, 6 mg lysate) was tested for phospho-peptide detection post TiO_2 enrichment. The BPC showed numerous intense peaks (Figure 3.15 (A)). Using the described Proteome Discoverer data analysis pipeline, with variable modifications of phospho- Serine, Threonine and Tyrosine, 17 peptides were confidently identified at a 1% FDR, 13 of which were phosphorylated (Figure 3.15 (B)).

3.6.3. Bead ratio and elution optimisations

Currently, the manufacturer recommend bead:lysate ratio was used (25 µl bead slurry:1 X 10 cm plate (600 µg lysate)). However, because a 10 fold increase to the recommended amount of lysate was required for LC-MS/MS analysis, this would result in a large quantity of GFP-Traps used per experiment. Besides the associated cost, an excess of GFP-Traps provides an increased opportunity for 'noise' proteins to bind the bead resin. Therefore, identical lysate was split for multiple IPs using varying densities of GFP-Traps (Figure 3.16 (A)). It has to be noted that the band intensity in the unbound lane is ~2 fold greater when using 10 fold less GFP-Traps (2.5 µl / 600 µg lysate). However, RapiGest SF elution resulted in equal quantities of HA-Clover-HIF-2α obtained. Therefore this bead:lysate ratio was incorporated into the protocol.

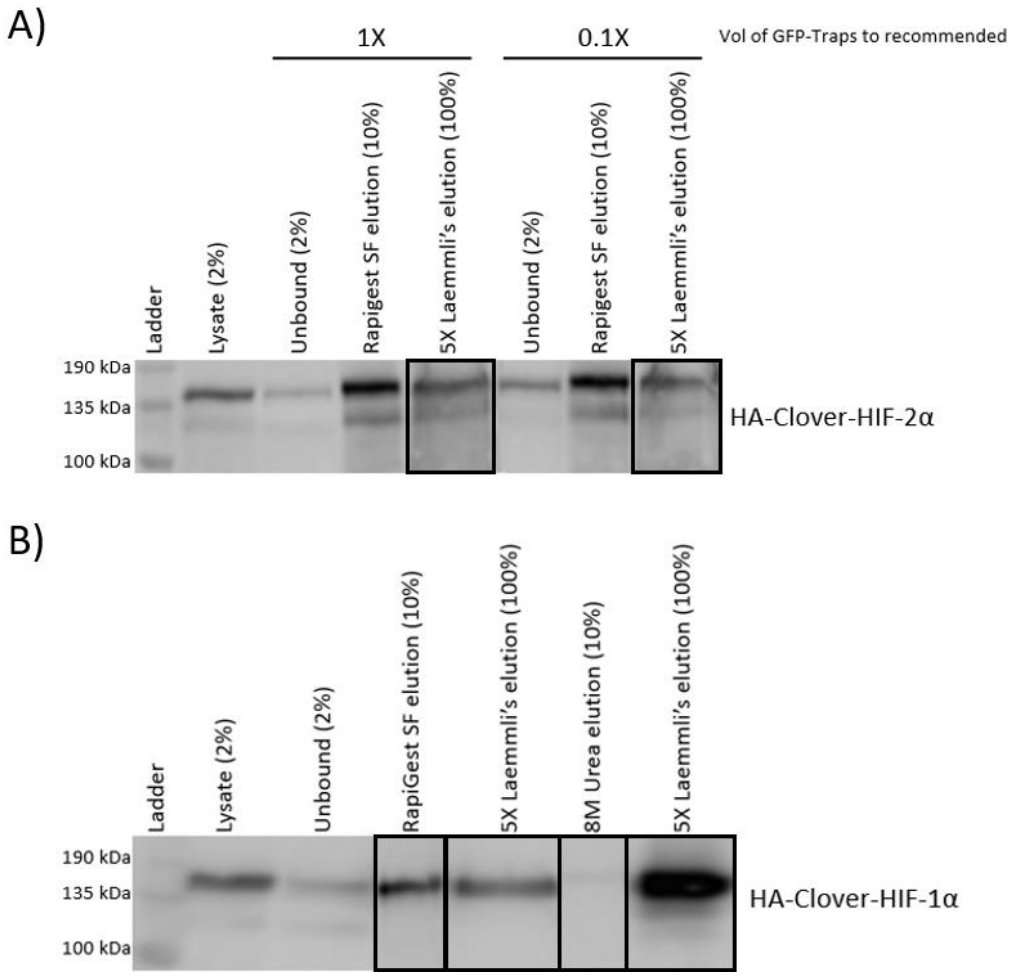


Figure 3.16: Bead density and non-canonical phosphorylation compatibility optimisations.
High expression model HA-Clover-HIFα (as stated) HeLa cell lysate (600 µg) incubated at 1% O_2 for 4 hr prior to IP. A) GFP-Trap bead density optimisation, stated fold change of GFP-Traps (and buffer volumes) used to manufacturer's recommendation, probed with a primary anti-HIF-2α antibody. B) Non-canonical phosphorylation compatibility, identical IPs eluted as stated, probed with a primary anti-HIF-1α antibody. Boxed regions are to highlight that a lane from a single gel has been moved post-processing.

Within the Eyers lab there is particular interest in non-canonical phosphorylation. Histidine, Lysine, Arginine, Aspartic acid, Glutamic acid and Cyesteine can all be phosphorylated, as well as the traditional Serine, Threonine and Tyrosine residues (Besant *et al.*, 2009, Attwood *et al.*, 2011 & Hardman *et al.*, 2017). However, non-canonical phosphorylation is heat and acid labile, therefore nor the GFP-Trap IP protocol, SCX clean-up, TiO_2 phospho-peptide enrichment or traditional mass spectrometry methodolgy is applicable for such investigations. A non-canonical phosphorylation pipeline has been developed in the Eyers lab, using HeLa cells and 8 M urea (Hardman *et al.*, 2017). Therefore, identical IPs were performed and eluted by RapiGest SF (optimised here) or an equal volume of 8 M urea, in order to determine the feasibility of adopting the non-canonical phosphorylation pipeline (Figure 3.16 (B)). Western blot analysis found that the urea elution resulted in the lack of a band, whereas a secondary elution in 5X Laemmli's buffer resulted in a very significant band

(Figure 3.16 (B)). Combined, this suggests that 8 M urea is not a strong enough elution condition to dissociate HA-Clover-HIF-1α from GFP-Traps. Thus, the non-canonical phosphorylation analysis was abandoned, and we decided to focus on traditional phosphorylation sites.

3.7. Discussion:

At the time of writing, this chapter describes the first protocol able to IP full length HIF-1α and HIF-2α from mammalian cell lines with sufficient abundance for in-depth discovery PTM analysis. Initially, target antibody-based IPs were investigated for the analysis of endogenous HIFα protein regulation. Although several antibodies were tested for sensitivity, selectivity and IP ability, they all failed the requirements. Combined with the apparent level of expression required for discovery PTM analysis, exogenous overexpression of IP-tag based HIFα was investigated. The HaloTag tag was found to make the HIF-2α protein insoluble, whereas GFP tagging was successfully used for IP with ~85% efficiency. The use of HA-Clover-HIFα proteins not only provides a means of IP but also allows for live-cell microscopy techniques, hence having the advantage of not requiring immunostaining protocols for visualisation.

Although a large exogenous overexpression is required for the initial identification of PTMs, future experiments could adopt a Selective Reaction Monitoring (SRM) MS approach to specifically search for *m/z* ratios of interest. SRM can provide in excess of 100 fold increases to sensitivity, feasibly allowing endogenous protein investigations, however requires pre-knowledge of known modified peptides and respective *m/z* ratios for selective MS filtering and fragmentation (reviewed by Lange *et al.*, 2008). Therefore, this approach is not possible currently. Thus, future experiments could use SRM to validate PTMs identified here using the low expression model or endogenous HIFα proteins in the future.

An interesting observation is that the large exogenous overexpression of WT HA-Clover-HIF-1α (>20 fold greater than endogenous HIF-1α) resulted in the exogenous protein expression at 21% O_2 without affecting endogenous HIF-1α O_2 dependent stability, which maintained complete O_2 dependent degradation. Several possibilities could explain this, including: 1) The HA-Clover tag interferes with the PHD-VHL targeting for degradation, possibly by steric hindrance and blocking modifying proteins from binding as efficiently. 2) A different, and prominent, unknown degradation pathway exists for HIF-1α, which is disrupted by N-terminal tagging. A potential example could be the N-end rule; endogenous HIF-1α protein

has a glutamic acid residue at its secondary position, known to be destabilising in mammals, while HA-Clover-HIF-1α protein has an alanine residue, known to be stabilising in mammals (Varshavsky, 2011 & Tasaki *et al.*, 2007). Therefore, the O_2 dependent stability observed for endogenous HIF-1α, although there is a large excess of HA-Clover-HIF-1α that accumulates in 21 O_2, may not be due to the PHD-VHL pathway but rather the result of the N-end rule pathway, which is more efficient in degrading the endogenous protein.

Another interesting observation is the comparison of WT HA-Clover-HIF-1α transcriptional output versus GFP-HIF-1α-DM, with 50 fold less DNA of HA-Clover-HIF-1α resulting in a greater HIF dependent luciferase signal than GFP-HIF-1α-DM. We never investigated protein expression levels, thus the WT HA-Clover-HIF-1α protein expression level could be several fold higher than GFP-HIF-1α-DM, although maintains the ability for O_2 dependent degradation at low expression levels. Further investigations should determine the protein expression level differences between plasmids, western blotting of samples will provide an answer to this. However, assuming similar protein expression levels, then the difference in transcriptional activity is due to a biochemical reason. The identification of 6 separate mutations within the GFP-HIF-1α-DM protein is of potential interest as these may block the efficient transactivation of the exogenous protein. These identified mutations may guide functional characterisation studies, if PTMs are identified in close proximity.

We attempted to adopt a non-canonical phospho-proteomics pipeline to investigate the complete phosphorylation status of HIF-1α and HIF-2α, however the GFP-Traps technology bound the HA-Clover tag too strongly to elute in current compatible methods. Future studies could use different strategies to investigate non-canonical phosphorylation, such as: anti-GFP antibody-based IPs, which are weaker than GFP-Traps and therefore may elute in 8 M urea. Or eluting in a detergent based buffer and performing a buffer exchange into MS compatible buffers for digestion.

We have been able to show that TiO_2 based phospho-peptide is possible for in-depth LC-MS/MS analysis, however this does not reflect the full PTM status of HIFα proteins; with many other PTMs known to regulate HIF-1α and HIF-2α (see introduction section 1.7). Additionally, TiO_2 enrichment is known to be inefficient for tyrosine phosphorylation identification (Lombardi *et al.*, 2015). Therefore, future studies could look to adopt the protocol described here for IP followed by secondary IP enrichment using antibodies against specific PTMs, for example anti- lysine acetylation antibodies.

4. Chapter 4: HIFα PTMs and binding partners

4.1. Introduction

As discussed in the introduction, cell signalling strategies have been vital in the development and wide evolutionary variation of multi-cellular organisms. However, for a cell to elicit a response, a ligand needs to transduce a signal through the cell membrane to induce its downstream effect, such as gene transcription. The evolution of reversible modifications of proteins, known as post translational modifications (PTMs) has allowed for rapid activation and termination of effector proteins to finely regulate a response. Many PTMs exist and regulate each protein differently, however phosphorylation is considerably one of the largest mechanism utilised by cells with >500 known kinases in humans (Manning *et al.*, 2002). It is also predicted that a third of all proteins expressed at any time are phosphorylated (Olsen *et al.*, 2006).

To further complicate PTM regulation, the same type of PTM within a given protein can have functionally distinct effects depending on the residue that is modified. For example, HIF-1α phosphorylation at S692 results in protein stabilisation (Bullen *et al.*, 2016), whilst phosphorylation on S657 results in protein degradation (Xu *et al.*, 2010) and phosphorylation on S641/S643 promotes nuclear localisation (Mylonis *et al.*, 2006 & Mylonis *et al.*, 2008). PTMs are generally sub-stoichiometric, thus allowing diverse and independent regulatory mechanisms and pathways to co-exist by functionally altering subsets of the protein (Mann *et al.*, 2002). Thus, the involvement of PTMs results in the proteome being exponentially more complex and dynamic than a collection of expressed gene products.

There has not yet been a study that defines the PTM status of full length HIFα proteins occurring within cells or in response to hypoxia. As a consequence the effects of hypoxia on the HIF-1α and HIF-2α PTM network have not been elucidated either. This is particularly relevant for HIF-2α, where O_2 dependent protein stability appears to play a lesser role in functional regulation, compared to HIF-1α (see introduction 1.1.3).

The lack of *in cellulo* data for HIFα PTMs can be explained by several technical challenges, including:

1) Endogenous HIFα proteins are of extremely low abundances within cells, 0.123 ppm and 1.53 ppm for HIF-1α and HIF-2α respectively (data obtained from PAXdb, Wang *et al.*, 2012). For comparison the p53 protein (gene: TP53), an essential

transcription factor where dysregulation of PTM status is well-known for cancer progression (reviewed by: Meek *et al.*, 2009), has an abundance of 22.2 ppm; ~200-20 fold greater than the HIFα isoforms respectively (data obtained from PAXdb, Wang *et al.*, 2012). Combined with poor HIFα antibodies that are not suitable for IP (discussed in chapter 3, section 3.3), the ability to purify endogenous HIFα in sufficient quantities for discovery mass spectrometry analysis is difficult. Given the stoichiometric ratio of PTMs, this reduces the abundance of the different PTM states (proteoforms) further, to levels potentially below the limit of detection for discovery MS experiments. Thus PTM identification and correct site localisation are challenging (Mann *et al.*, 2002). I solved these challenges in Chapter 3, by using HA-Clover-HIFα constructs overexpression and IP against the Clover tag.

2) The selection of PTMs to investigate. From previous studies (Table 1.2 & Table 1.3), HIF-1α has been shown to be modified by multiple different types of PTMs, including: hydroxylation (Jaakkola *et al.*, 2001) phosphorylation (Kalousi *et al.*, 2010), acetylation (Jeong *et al.*, 2002), methylation (Lee *et al.*, 2017), ubiquitination (Tanimoto *et al.*, 2000), SUMOylation (Bae *et al.*, 2004) and S- nitrosylation (Yasinska *et al.*, 2003). Because the stoichiometric ratio of PTMs is undetermined, it is possible that PTM containing peptides may be below the limit of detection of the analytical system and/or masked by unmodified peptides. Thus, in a paradoxical fashion, discovery analysis attempting to identify all PTMs can lead to the lack of confident data. Strategies exist to enrich a specific PTM, in the background of unmodified/different PTM peptides, allowing the in-depth analysis of a specific PTM per LC-MS/MS experiment. Enrichment strategies can be antibody IP based or affinity based. The former is useful for specific modifications such as ubiquitination (Udeshi *et al.*, 2013). However, phospho- serine and threonine antibodies are generally poor and highly dependent on the sequence environment of the PTM (Fíla *et al.*, 2012). Therefore, it is preferential to investigate phosphorylation through Immobilised Metal ion Affinity Chromatography (IMAC). As discussed in the introduction, IMAC exploits the charge specific properties of phosphorylation to bind and purify phospho-peptides from the background of unmodified peptides (Mann *et al.*, 2002 & Rainer *et al.*, 2015). In chapter 3, we demonstrated that it was possible to use TiO_2, the most commonly used IMAC resin, for phosphorylation enrichment of HIF-1α peptides. Thus, phosphorylation will be the main focus for PTM identification in this work.

3) The poor sequence coverage observed by LC-MS/MS analysis of HIFα digestion. A traditional proteomics workflow utilises trypsin as a proteolytic enzyme to create analysable peptides from proteins. Trypsin is preferentially used due to its specific cleavage pattern at the C-terminus of Lysine (K) and Arginine (R) residues, thus resulting in 2+ ion peptides (on average) which aid in peptide flyability and fragmentation (Paizs *et al.*, 2005, Zhang *et al.*, 2013 & Michalski *et al.*, 2012). However, the peptides analysed are thus dependent on the distribution of K and R residues throughout a protein of interest. Hence tryptic digestion can be disadvantageous because it can result in peptides too large or too small for LC-MS/MS analysis, determined by sequence (Tsiatsiani *et al.*, 2015 & Mallick *et al.*, 2006). As a consequence, trypsin digestion alone can result in large sections of proteins that are unanalysed, resulting in poor protein sequence coverage and concomitant loss of PTMs being identified. By utilising a combination of proteolytic strategies, each generating their own identifiable peptides, it is possible to increase protein coverage detected by mass spectrometry analysis; providing a more in depth view of the PTM status of proteins (reviewed by Meyer *et al.*, 2011).
4) The labile nature of the phosphate moiety and difficulty in PTM localisation. Although phosphorylation is a covalent modification, the phospho-bond is much more susceptible to collisional induced fragmentation techniques (CID), which are required for amide bond fragmentation to obtain the peptide primary sequence and to localise the PTM. Thus, CID and, more pertinently, higher energy collision dissociation (HCD) can result in the neutral loss of the phosphate moiety and reduced ability to localise the PTM (Lehmann *et al.*, 2007). Software has been developed to statistically infer phosphorylation localisation post neutral loss (Taus *et al.*, 2011). Alternatively, additional fragmentation methods are available such as Electron Transfer Dissociation (ETD). ETD induces fragmentation specifically at the N-Cα bond in a non-vibrational fragmentation method, bypassing the neutral loss effect of phosphorylation (Syka et al., 2004). However, ETD fragmentation is highly inefficient for low charge state peptide ions (<3+) and the prominent tryptic ions are 2+. Hence a combinatorial method utilising both ETD and HCD fragmentations (EThcD) was developed (Frese *et al.*, 2012 & Frese *et al.*, 2013). In theory, by creating mixed ETD/HCD fragmentation spectra it maximises fragmentation for determining peptide primary sequence (HCD) while increasing the likelihood that localisation data may be maintained (ETD) (Frese *et al.*, 2012 & Frese *et al.*, 2013).

5) Biological interpretation. The development of large datasets can make it difficult to identify interesting aspects to initially focus on for further characterisation. Beyond defining presence and absence of PTMs between biological conditions, MS can be used quantitatively to investigate how the level of a PTM changes depending on a specific treatment (Thompson *et al.*, 2003). The more accurate labelling strategies require the different samples to be prepared simultaneously to remove LC-MS/MS variation, a prospect not possible here due to the experimental size (Chapter 3, section 3.6.2). However, label free quantification can be performed, which experiences greater experimental error but allows the theoretical comparison of an unlimited number of samples. Additionally, investigating the evolutionary history of a PTM site can be insightful as highly conserved sites are likely to have essential functional importance (Bui *et al.*, 2016, Beltrao *et al.*, 2013 & Capra *et al.*, 2007). Label free quantification and evolutionary analysis strategies were used here.

An additional aspect of IP-coupled MS analysis is the Co-IP of binding partners. Although impossible to determine whether a Co-IP'd protein is a direct or an indirect interactor of the target protein, binding partners can be essential for function, for example the HIFα – HIF-1β interaction (Wang *et al.*, 1995). Additionally, for a PTM to occur the modifying enzyme must come into contact with the protein of interest, although may not be stably bound. Thus, using a High Throughput (HTP) mass spectrometry approach to discover the binding partners that are Co-IP'd with HIF-1α and HIF-2α may not only unlock the signalling mechanisms differentiating the two isoforms but also potentially identify any modifying enzymes which may cause a given PTM. The latter can be fed back into pathway analysis to understand the global regulatory pathways acting on the HIFα proteins.

4.2. Aims:

The aim of this chapter was to use the developed GFP-Traps IP protocol for HA-Clover-HIFα proteins in order to identify: 1) The PTMs that occur in response to hypoxia for full length HIF-1α and HIF-2α *in cellulo*. A specific focus on phosphorylation will be applied, for the technical reasons mentioned above, however all PTMs were investigated. 2) The binding partners of HIFα isoforms in response to hypoxia. As mentioned above, several technical challenges needed to be overcome for an in-depth analysis of both HIFα proteins. This included:

1) Improving sequence coverage identified by LC-MS/MS analysis. Traditional tryptic digests resulted in <35% HIFα protein coverage, thus digestion with multiple proteases was investigated.
2) The attainment of maximal and confidently localised PTM data, by benchmarking different MS/MS fragmentation methods.
3) Biological Interpretation of identified PTMs. Upon PTM data acquisition, a quintuple approach to biological interpretation was applied. Domain, treatment, isoform sequence comparison, phylogenetic/evolutionary and cancer database screening analyses were performed to aid in the identification of potentially more important sites of regulation. Where applicable, PTM sites were analysed using crystal structure modelling to predict the impact of a modification through close proximity intra- and inter- molecular interactions.

Additionally, using the binding partner information obtained, gene ontology analysis and label free quantification approaches were performed to identify how binding partners acting on HIFα change between O_2 tensions. I used these data to determine specifically enriched pathways, and identify the upstream enzymes potentially responsible of the PTMs identified, which may have a regulatory role in the HIF signalling network.

4.3. Improving sequence coverage:

In order to define an in-depth PTM map it is important that most of the protein sequence of interest can be analysed, at the peptide level, using MS methods. Therefore, the developed GFP-Trap IP coupled MS protocol was followed and the digested eluent analysed using a high scan frequency Orbitrap-Iontrap method. Trypsin digestion resulted in <35% of the HIF-1α protein sequence being identified at a 1% FDR (Figure 4.1), hence two-thirds of the protein, and potentially PTMs occurring, were not detectable. To improve the depth of PTM map obtainable, the IP process was repeated multiple times and digested with different proteases and analysed identically to trypsin. Proteases tested included: Chymotrypsin, Elastase, Asp-N, Glu-C and Lys-C (Figure 4.1, not all data shown).

Chymotrypsin (cleaves at tyrosine (Y), phenylalanine (F), tryptophan (W), leucine (L) and methionine (M)) and particularly Elastase (cleaves at alanine (A), glycine (G), valine (V), serine (S), leucine (L) and isoleucine (I)) are prone to not cleaving at every possible site, known as miscleaves. Miscleave frequency needs to be accounted for at the time of data

interrogation because increasing the number of permitted miscleaves allows for more peptide identifications, however simultaneously increasing the chance that a given fragmentation spectrum (PSM) can match the reversed sequence database, used to determine FDR, by chance. Thus a balance exists when filtering to a 1% FDR, allowing too few potential miscleaves results in not all fragmented peptides being identified, and allowing too many increases the false positive rate and subsequent removal of correctly matched data by FDR filtering. To calculate the optimal parameters for data analysis, the same data file was analysed repeatedly, sequentially changing the permitted number of miscleaves per peptide, and recording the number of peptides and total number of PSMs identified at 1% FDR filtering (Table 4.1). This allowed me to evaluate the miscleave propensity for each protease and determine optimal parameters to assess obtainable protein coverage.

Table 4.1: Determination of the best miscleave parameters for maximal confidently identified peptides for Elastase and Chymotrypsin.
Table includes stated number of miscleaves post 1% FDR filtering, Protein confidence as determined by MASCOT score, number of unique peptides and the total number of PSMs identified for HIF-1α. LC-MS/MS was acquired following digestion at pre- (using a High-Low high scan frequency Orbitrap-Iontrap MS/MS method) and post-TiO_2 enrichment (using a High-High low scan frequency Orbitrap-Orbitrap method), digest peptides and TiO_2 enriched peptides respectively, # PSMs is summed number of PSMs from both runs that were confidently identified.

Elastase				
# Miscleaves allowed	MASCOT Score	# Digest peptides	# TiO_2 enriched peptides	# PSMs
9	475.36	8	4	15
8	1267.93	31	21	90
7	1109.61	28	18	73
6	972.11	24	14	63
5	589.86	20	11	50
4	436.76	16	8	33

Chymotrypsin				
# Miscleaves allowed	MASCOT Score	# Digest peptides	# TiO_2 enriched peptides	# PSMs
5	2296	61	25	174
4	2296	61	25	174
3	2151	58	23	155
2	2051	56	23	152

Table 4.1 shows that, as expected, as the number of miscleaves decreases, the number of PSMs and peptides identified generally do as well. At a 1% FDR cut-off, elastase with 9 miscleaves identified 15 PSMs, however, successive decreases to 8 and 7 miscleaves found that 90 and 73 PSMs were maintained respectively. Hence 8 miscleaves was selected for elastase digestion samples, equating to ~65% protein sequence coverage (Figure 4.1). At a 1% cut-off, chymotrypsin identified equal numbers of PSMs for both 5 and 4 miscleaves (174). Thus, to reduce the potential for increased false positive rates, 4 miscleaves was used to analyse chymotrypsin digest samples; equating to ~45% protein sequence coverage (Figure 4.1).

>Trypsin

MAYPYDVPDYASLGSMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTFGYGVACFSRYPDHMKQHDFF
KSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNT
PIGDGPVLLPDNHYLSHQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSATMEGAGGANDKKKISSERRKEKSRDAARSRRSKESEVFYELAHQLP
LPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDDMKAQMNCFYLKALDGFVMVLTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEM
REMLTHRNGLVKKGKEQNTQRSFFLRMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYDTNSNQPQCGYKKPPMTCLVLICEPIPHPSNIEIPLDSKTFLSRH
SLDMKFSYCDERITELMGYEPEELLGRSIYEYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKNSQPQCIVCVNYVVSGIIQ
HDLIFSLQQTECVLKPVESSDMKMTQLFTKVESEDTSSLFDKLKKEPDALTLLAPAAGDTIISLDFGSNDTETDDQQLEEVPLYNDVMLPSPNEKLQNINLAM
SPLPTAETPKPLRSSADPALNQEVALKLEPNPESLELSFTMPQIQDQTPSPSDGSTRQSSPEPNSPSEYCFYVDSDMVNEFKLELVEKLFAEDTEAKNPFSTQ
DTDLDLEMLAPYIPMDDDFQLRSFDQLSPLESSSASPESASPQSTVTVFQQTQIQEPTANATTTTATTDELKTVTKDRMEDIKILIASPSPTHIHKETTSATSSP
YRDTQSRTASPNRAGKGVIEQTEKSHPRSPNVLSVALSQRTTVPEEELNPKILALQNAQRKRKMEHDGSLFQAVGIGTLLQQPDDHAATTSLSWKRVKGCK
SSEQNGMEQKTIILIPSDLACRLLGQSMDESGLPQLTSYDCEVNAPIQGSRNLLQGEELLRALDQVN

>Chymotrypsin

MAYPYDVPDYASLGSMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTFGYGVACFSRYPDHMKQHDFF
KSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNT
PIGDGPVLLPDNHYLSHQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSATMEGAGGANDKKKISSERRKEKSRDAARSRRSKESEVFYELAHQLP
LPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDDMKAQMNCFYLKALDGFVMVLTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEM
REMLTHRNGLVKKGKEQNTQRSFFLRMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYDTNSNQPQCGYKKPPMTCLVLICEPIPHPSNIEIPLDSKTFLSRH
SLDMKFSYCDERITELMGYEPEELLGRSIYEYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKNSQPQCIVCVNYVVSGIIQ
HDLIFSLQQTECVLKPVESSDMKMTQLFTKVESEDTSSLFDKLKKEPDALTLLAPAAGDTIISLDFGSNDTETDDQQLEEVPLYNDVMLPSPNEKLQNINLAM
SPLPTAETPKPLRSSADPALNQEVALKLEPNPESLELSFTMPQIQDQTPSPSDGSTRQSSPEPNSPSEYCFYVDSDMVNEFKLELVEKLFAEDTEAKNPFSTQ
DTDLDLEMLAPYIPMDDDFQLRSFDQLSPLESSSASPESASPQSTVTVFQQTQIQEPTANATTTTATTDELKTVTKDRMEDIKILIASPSPTHIHKETTSATSSP
YRDTQSRTASPNRAGKGVIEQTEKSHPRSPNVLSVALSQRTTVPEEELNPKILALQNAQRKRKMEHDGSLFQAVGIGTLLQQPDDHAATTSLSWKRVKGCK
SSEQNGMEQKTIILIPSDLACRLLGQSMDESGLPQLTSYDCEVNAPIQGSRNLLQGEELLRALDQVN

>Elastase

MAYPYDVPDYASLGSMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTFGYGVACFSRYPDHMKQHDFF
KSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNT
PIGDGPVLLPDNHYLSHQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSATMEGAGGANDKKKISSERRKEKSRDAARSRRSKESEVFYELAHQLP
LPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDDMKAQMNCFYLKALDGFVMVLTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEM
REMLTHRNGLVKKGKEQNTQRSFFLRMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYDTNSNQPQCGYKKPPMTCLVLICEPIPHPSNIEIPLDSKTFLSRH
SLDMKFSYCDERITELMGYEPEELLGRSIYEYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKNSQPQCIVCVNYVVSGIIQ
HDLIFSLQQTECVLKPVESSDMKMTQLFTKVESEDTSSLFDKLKKEPDALTLLAPAAGDTIISLDFGSNDTETDDQQLEEVPLYNDVMLPSPNEKLQNINLAM
SPLPTAETPKPLRSSADPALNQEVALKLEPNPESLELSFTMPQIQDQTPSPSDGSTRQSSPEPNSPSEYCFYVDSDMVNEFKLELVEKLFAEDTEAKNPFSTQ
DTDLDLEMLAPYIPMDDDFQLRSFDQLSPLESSSASPESASPQSTVTVFQQTQIQEPTANATTTTATTDELKTVTKDRMEDIKILIASPSPTHIHKETTSATSSP
YRDTQSRTASPNRAGKGVIEQTEKSHPRSPNVLSVALSQRTTVPEEELNPKILALQNAQRKRKMEHDGSLFQAVGIGTLLQQPDDHAATTSLSWKRVKGCK
SSEQNGMEQKTIILIPSDLACRLLGQSMDESGLPQLTSYDCEVNAPIQGSRNLLQGEELLRALDQVN

>Overlay

MAYPYDVPDYASLGSMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTFGYGVACFSRYPDHMKQHDFF
KSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNT
PIGDGPVLLPDNHYLSHQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSATMEGAGGANDKKKISSERRKEKSRDAARSRRSKESEVFYELAHQLP
LPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDDMKAQMNCFYLKALDGFVMVLTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEM
REMLTHRNGLVKKGKEQNTQRSFFLRMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYDTNSNQPQCGYKKPPMTCLVLICEPIPHPSNIEIPLDSKTFLSRH
SLDMKFSYCDERITELMGYEPEELLGRSIYEYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQATVIYNTKNSQPQCIVCVNYVVSGIIQ
HDLIFSLQQTECVLKPVESSDMKMTQLFTKVESEDTSSLFDKLKKEPDALTLLAPAAGDTIISLDFGSNDTETDDQQLEEVPLYNDVMLPSPNEKLQNINLAM
SPLPTAETPKPLRSSADPALNQEVALKLEPNPESLELSFTMPQIQDQTPSPSDGSTRQSSPEPNSPSEYCFYVDSDMVNEFKLELVEKLFAEDTEAKNPFSTQ
DTDLDLEMLAPYIPMDDDFQLRSFDQLSPLESSSASPESASPQSTVTVFQQTQIQEPTANATTTTATTDELKTVTKDRMEDIKILIASPSPTHIHKETTSATSSP
YRDTQSRTASPNRAGKGVIEQTEKSHPRSPNVLSVALSQRTTVPEEELNPKILALQNAQRKRKMEHDGSLFQAVGIGTLLQQPDDHAATTSLSWKRVKGCK
SSEQNGMEQKTIILIPSDLACRLLGQSMDESGLPQLTSYDCEVNAPIQGSRNLLQGEELLRALDQVN

Figure 4.1: Identified sequence map of HIF-1α.
HA-Clover-HIF-1α protein sequence, with peptide identifications at a 1% FDR highlighted. Data from the digest stage LC-MS/MS analysis using a High-Low MS/MS method. Bold and underlined = HA-Clover tag sequence, Pink = Trypsin digested (2 miscleaves), Green = Chymotrypsin digested (4 miscleaves), Grey = Elastase digested (8 miscleaves). An overall coverage map is displayed to determine complete protein coverage, where a hierarchy for highlighting was applied of Trypsin, Chymotrypsin, and Elastase.

As Figure 4.1 shows, combining data obtained from trypsin, chymotrypsin and elastase digestions equates to >90% of the HIF-1α protein sequence confidently identified at a 1%

FDR. The majority of the missing sequence is located at the HIF-1α N-terminus which is K/R rich, thus producing small peptides that are difficult to multiply charge, hence having poor fragmentation efficiency. Significantly, for phosphorylation analysis, >95% of the total S/T/Y residues were identified, allowing the creation of an in-depth phosphorylation map. Although additional proteases were investigated (Asp-N & Glu-C), no further protein coverage was obtained by their addition into the protocol, as such were not used. This experiment was repeated with HA-Clover-HIF-2α, using the same digest miscleave parameters, similarly obtaining >90% sequence coverage (data not shown).

4.4. MS/MS method development:

As discussed in the introduction, different methods for peptide fragmentation exist, namely HCD, ETD & EThcD. Many benchmark studies comparing these fragmentation methods have been performed for HTP phosphorylation analysis of trypsin digested human proteome, yet leading to different conclusions. Frese *et al.*, 2012 showed that EThcD resulted in an improved peptide confidence score in 72% of all peptides when compared to HCD alone, translating to a marginal increase in correct phosphorylation localisation to 97%, from 95%. However, due to the increased time required for a dual fragmentation approach, EThcD resulted in 11% less total number of PSMs, thus a trade-off exists in HTP ability versus peptide confidence. This study was performed on the human proteome, thus is much more complex than my samples that are post IP (of a low copy number protein) and TiO_2 enrichment. As a result, the reduced PSM count of EThcD may be inconsequential for my study, if providing more confidently localised PTMs. However, a more recent study from the Eyers lab found that an Orbitrap only, HCD only fragmentation strategy combined with ptmRS analysis outcompeted EThcD in the total number of PSMs and correct phosphorylation site localisation (Ferries *et al.*, 2017). Therefore, it was necessary to determine whether an HCD only or EThcD strategy provided the most confident data for my samples.

The MS/MS methods from Ferries *et al.*, 2017 (Orbitrap only, HCD collision energy (NCE) at 32% and Orbitrap only EThcD, ETD calibrated time and HCD collision energy at 25%) were tested. Alongside these, a series of successive changes to the HCD collision energy of EThcD methods (25%-18%) were tested to potentially increase PTM localisation confidence, by providing less collisional energy. Additionally, it was theorised that a second HCD only method, designed to increase sensitivity and fragmentation time, could be beneficial within

this study (due to the vastly reduced complexity of samples post –IP and –TiO_2 phospho-peptide enrichment, compared to the human proteome). Changes to the Ferries *et al.*, 2017 method included: intensity required for peptide fragmentation (lowered to $2E^4$ ions, from $2E^5$), targeted peptide ion count before fragmentation (AGC target, increased to $5E^5$, from $2E^5$), maximum injection time (time spent collecting ions before fragmentation, if AGC target is not reached, increased to 250 ms from 100 ms) and the inclusion of 2 μ-scans (a technique that fragments a particular *m/z* ion twice and additively combines spectra to improve the signal to noise ratio). Each method was tested on the identical, post TiO_2 enrichment, HA-Clover-HIF-2α IP sample. The number of peptides identified, ptmRS score, peptide confidence MASCOT score and E-value were recorded for benchmarking (Table 4.2).

Table 4.2: MS/MS method development for most confidently identified and localised phospho-peptides.
A large scale HA-Clover-HIF-2α IP and TiO_2 enrichment injected identically and analysed using different MS/MS methods, as stated. MS/MS parameters for phospho-peptide analysis from Ferries *et al.*, 2017 was used initially and modified as stated. Alterations for 'Modified' include decreased minimum intensity = $4e^4$, increased Maximum injection time = 250 ms, AGC target = $5e^5$ and microscans = 2. % refers to normalised HCD energy. Data analysed using Proteome Discoverer with the MASCOT search engine using the human proteome database, with ptmRS enabled. ptmRS score, MASCOT score and E-value were all recorded for benchmarking.

	Modified			Ferries et al., 2017 Method			EThcD (25%)			EThcD (22%)			EThcD (20%)			EThcD (18%)		
Peptide	ptmRS (best)	MASCOT Score	E-value	ptmRS (best)	MASCOT Score	E-value	ptmRS (best)	MASCOT Score	E-value	ptmRS (best)	MASCOT Score	E-value	ptmRS (best)	MASCOT Score	E-value	ptmRS (best)	MASCOT Score	E-value
HLPLPQPPSAIsPGENSK	S12(Phospho): 100	32	1.40E-02				S12(Phospho): 99.69	52	1.30E-04	S9(Phospho): 50; S12(Phospho): 50	63	1.00E-05	S12(Phospho): 99.66	44	8.80E-04	S12(Phospho): 99.65	58	3.20E-05
GPDVLsPAmVALSNK	S6(Phospho): 100	81	1.50E-07	S6(Phospho): 100	59	2.30E-05										S6(Phospho): 50; S13(Phospho): 50	16	5.80E-01
LLSsVcSENESEAEADQQmDNLYLK	S4(Phospho): 98.91	97	7.50E-09															
GGscPLmPDKPLSANVPNDK	S3(Phospho): 100	45	9.30E-04															
LLSsVcSENEsEAEADQQmDNLYLK	S4(Phospho): 99.37; S11(Phospho): 99.37	72	1.80E-06	S4(Phospho): 99.35; S7(Phospho): 50.11; S11(Phospho): 50.11	41	2.10E-03	S3(Phospho): 50; S4(Phospho): 50; S7(Phospho): 50; S11(Phospho): 50	32	1.70E-02									
LLGPsFESYLLPELTR	S5(Phospho): 49.9; S8(Phospho): 49.9	29	2.90E-02															
LLSSVcSENESEAEADQQmDNLYLK		72	2.20E-06		46	9.10E-04		90	4.00E-08		62	2.10E-05		74	1.50E-06		28	5.90E-02
FQQQLESK		48	1.90E-04		43	1.10E-03		26	2.90E-02									
KTEPEHRPmSSIFFDAGSK		41	2.50E-03											27	6.70E-02		25	9.60E-02
GQVVSGQYR		39	2.60E-03					42	9.00E-04		41	1.00E-03		31	1.70E-02		13	1.10E+00
SHQNLcTK		36	5.10E-03											28	3.30E-02		18	2.90E-01
HLPLPQPPSAISPGENSK		32	1.20E-02															
GGScPLmPDKPLSANVPNDK		31	2.40E-02															
SAYEFYHALDSENmTK		30	2.50E-02											36	6.30E-03		15	8.10E-01
FTYcDDR		27	4.40E-02															
TEFLGAAPLGPPVsPPHVSTFK				S14(Phospho): 100	44	8.90E-04												
LLSSVcSENEsEAEADQQmDNLYLK				S11(Phospho): 99.61	76	9.40E-07												
LLSSVcsENESEAEADQQmDNLYLK							S7(Phospho): 99.45	49	3.90E-04	S7(Phospho): 98.72	60	3.30E-05	S7(Phospho): 49.59; S11(Phospho): 49.59	39	4.70E-03	S7(Phospho): 49.66; S11(Phospho): 49.66	64	1.50E-05
TEFLGAAPLGPPVsPPHVSTFK																S14(Phospho): 97.31	10	2.20E+00

As reported by Ferries *et al.*, 2017, we found that the HCD-Orbitrap only method outperformed any of the tested EThcD methods in terms of the number of confidently localised phospho-peptides. We observed that many of the peptides identified in HCD methods are of 2+ charge state, thus explaining their lack of identification in EThcD methods (Sobott *et al.*, 2009). The newly created high sensitivity, HCD only, Orbitrap only method resulted in the most phospho-peptides identified at a 1% FDR. Peptides identified were also at a greater confidence using MASCOT Score and E-value statistics and, significantly, resulted in greater ptmRS scores for identical peptides identified by different MS/MS methods. For example the LLSsVcSENEsEAEADQQmDNLYLK doubly phosphorylated peptide had both sites of phosphorylation identified to a confidence of >99.0, while the Ferries *et al.*, 2017 method could not distinguish between S7 or S11 as a phosphorylation site, both at a 50/50 chance. Thus, for this study, all phosphorylation analysis was performed using the newly developed HCD only MS/MS method.

4.5. Biological interpretation of MS data- Phylogeny analysis:

With the identification of potentially many PTM sites, attempting to prioritise sites for further investigation is a non-trivial task. Using a combined approach of MS data analysis and evolutionary history analysis of a particular site, it can potentially provide insight into the functional importance of a modification. Evolutionary conserved sites have more chance to be associated with essential functionality, while a highly mutated site could reflect poor functional importance (Beltrao *et al.*, 2013 & Capra *et al.*, 2007). However, the mutations that arise during evolution can also provide insight. For example, if a phosphorylated serine residue is mutated to a negatively charged aspartic acid residue it may reflect a state of permanent phosphorylation, conversely mutation to an uncharged, unmodifiable residue, such as alanine, could reflect a state of permanent unphosphorylation (Pham *et al.*, 2008, diagrammatically depicted in Figure 5.1). Therefore, combining the evolution of a PTM site with species knowledge, of where a mutation occurred (for example Carp species known for extreme hypoxia tolerance (Nilsson *et al.*, 2004)), may lead to the identification of potential sites of importance.

Following the guidelines from Hall, 2013, all protein matches to either HIF-1α or HIF-2α were obtained by BLAST searching. This was reduced to a non-redundant database by

reciprocal BLAST searching and a phylogenetic tree created for both HIF-1α (data not shown) and HIF-2α (Figure 4.2). The multiple sequence alignments were reoriented to match the order of the evolutionary tree, thus allowing the evolution of a specific site to be followed with ease. For this analysis, only vertebrate species were used for two reasons: 1) HIF-2α has only evolved in species with extensive oxygen delivery systems (Graham *et al.*, 2017). 2) Invertebrate HIF-1α homologs are vastly different to vertebrate homologs. For example, *Drosophila melanogaster* HIF-1α homolog, sima, is >1500 residues long (approximately twice the size), contains only one proline (P) hydroxylation site, lacks a definable CTAD and has primary functions in tracheal proliferation (Gorr *et al.*, 2004). Therefore, it is reasonable to assume the regulatory pathways at play may be very different.

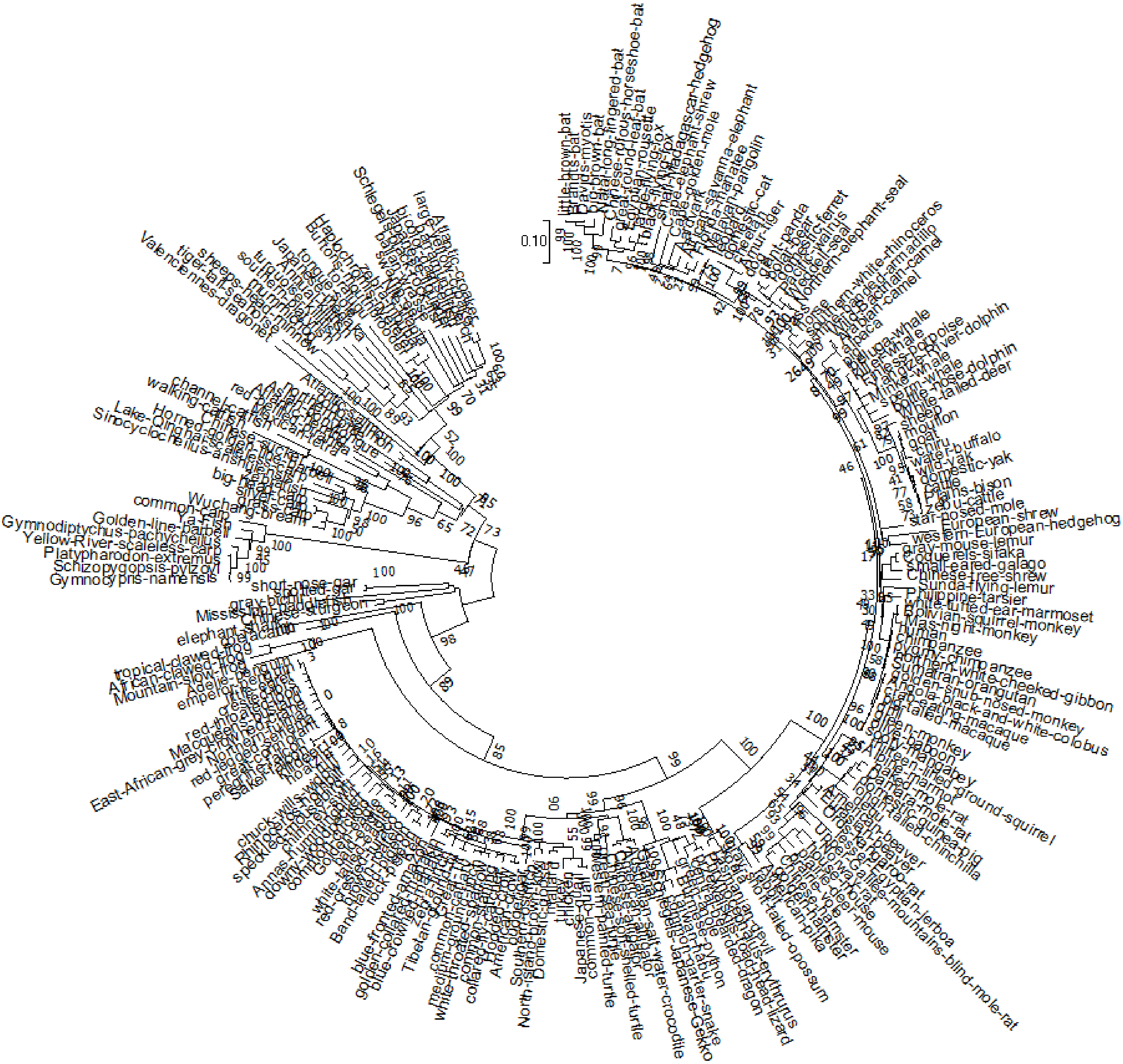


Figure 4.2: Molecular Phylogenetic analysis of HIF-2α protein sequences by Maximum Likelihood method. Evolutionary history was inferred by use of maximum Likelihood based on the JTT matrix model (Jones *et al.*, 1992). The highest log likelihood (-47644.35) tree is shown. Initial tree(s) for the heuristic search were obtained using Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and selecting the superior log likelihood topology. A discrete Gamma distribution was used to model evolutionary rate among sites (5 categories (+*G*, parameter = 0.9749)). The rate variation model allowed for some sites to be evolutionarily invariable ([+*I*], 8.63% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 227 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 730 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016). 500 bootstrap replicates were used and values represent the confidence that a branch point exists. Common names, from the R Taxize plugin, are used for naming purposes.

4.6. Mass spectrometry sample preparation:

Following the optimised GFP-Trap IP protocol on exogenously expressed HA-Clover-HIFα, IPs were performed in duplicate at both 21% and 1% O_2 concentrations. After SCX clean-up, 5% of the sample was analysed on a 1 hr LC gradient using a High-Low MS/MS method for maximal protein identifications, facilitating binding partner identification. The remaining 95% was TiO_2 enriched and 100% analysed using the optimised HCD only High-High Orbitrap only MS/MS method (section 4.4). All data were analysed using Proteome Discover, searching against the human non-redundant database and filtered to 1% FDR, equating to ~90% and ~60% sequence coverage post IP and TiO_2 enrichment respectively. The next sections discuss the PTMs and binding partners identified.

4.7. Phosphorylation data:

Overall, ~25 phosphorylation sites were identified for both of HIF-1α and HIF-2α at a confidence of 1% FDR and ptmRS >99.0 (Figure 4.3, supporting peptide data in appendix 1), showing these proteins undergo extensive PTM. Comparing the phosphorylation sites identified to previously published data (Table 1.2 & Table 1.3) and HTP data recorded in PhosphoSitePlus (Hornbeck *et al.*, 2015 & Hornbeck *et al.*, 2012), showed that the vast majority of the identified sites here are novel, with only 7 sites for HIF-1α and 2 sites for HIF-2α previously characterised. Considering the wealth of data published for HIF-1α phosphorylation, ~40 sites when including HTP identified sites, it is surprising that so few were identified in our data. However it is important to consider that the majority of proteomics studies will not be under hypoxic conditions, as well as the inefficiencies with trypsin digestion (Figure 4.1), therefore will artificially select against HIFα phospho-peptide detection. It can also be argued that the lack of previously identified phospho-sites (with functional characterisation) could be a reflection of the cell type specific regulation or a combination of previous studies not using true O_2 deprivation hypoxia or the potentially high false positive rates of *in vitro* assays.

For discussion, phosphorylation sites have been grouped together below based on their previous discovery, O_2 dependence and domain localisation. The groups have been selected by observed characteristics and discussed in an order based on their potential interest, with a grouping order of: 1) previously characterised, 2) localisation in the ODDD, 3) localisation in the NTAD, 4) O_2 dependency, 5) proline hydroxylation site proximity 6) localisation in the

bHLH and PAS-A/B domains, 7) clustering and 8) localisation in the inhibitory domain. If a particular PTM is applicable to multiple groups it is discussed in the primary group only, unless otherwise stated. However the site is present in the respective table containing their evolutionary history and isoform comparisons at the start of each group (Table 4.3-Table 4.19).

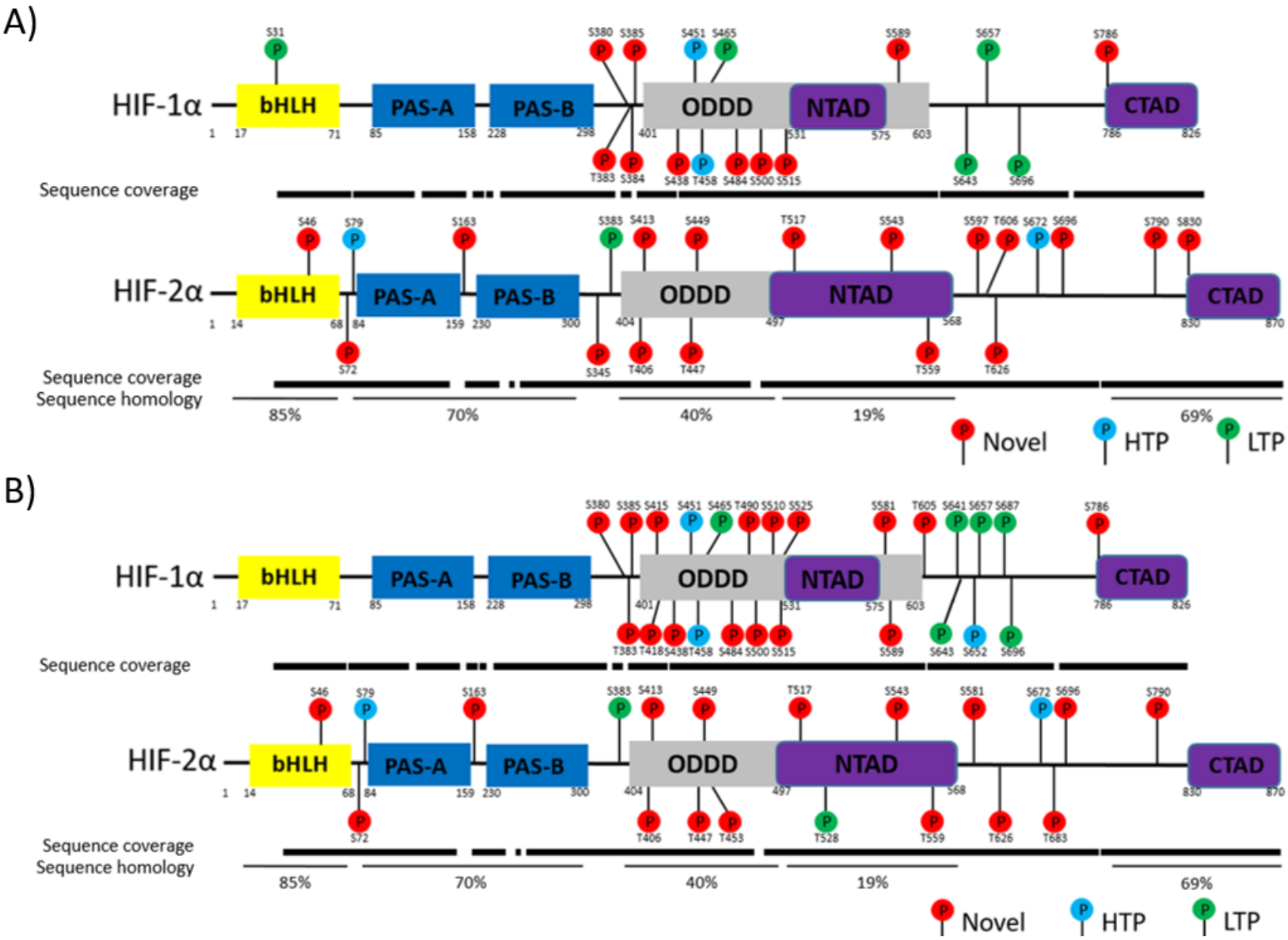


Figure 4.3: Schematic view of the phosphorylation map of HIF-1α and HIF-2α.
Domains of HIFα proteins are highlighted and include canonical domain numbering. Confidently identified phosphorylation sites (1% FDR, ptmRS score >99.0) are mapped to the protein and coloured dependent on whether they are novel (red), have been seen by previous HTP mass spectrometry studies (blue) or have undergone functional characterisation (green). The sequence coverage seen from mass spectrometry analysis (from Figure 4.1) is depicted below the protein schematic along with the sequence homology between isoform domains. A) Phosphorylation sites identified in 1% O_2, B) Phosphorylation sites identified in 21% O_2.

4.7.1. Previously characterised:

In total 7 phosphorylation sites, out of 22, have been previously characterised for HIF-1α (S31, S465, S641, S643, S657, S687 and S696) and 3, out of 5, have been previously characterised for HIF-2α (S383, T528 and S830), listed in Table 4.3.

Table 4.3: Characteristics of HIF-1α and HIF-2α phosphorylation sites identified.
Table includes position of PTM, how many studies have identified the respective site with function and reference, evolutionary history where variations are highlighted by single letter amino acid code, the O_2 tension a PTM was identified in and comparisons to HIF-2α sequence. NA = not applicable, X = residue was not phosphorylated.

HIF-1α					
Site	Characterisation	Conservation	Observed in: 21%, 1% or Both?	HIF-2α alignment	phosphorylated HIF-2α?
31 S	1 LTP PKA- no investigation Bullen et al,. 2016	Fish (except Gar) = G	1%	S	X
465 S	2 LTP + 2 HTP PKA- no investigation Bullen et al,. 2016 + So et al,. 2016	Fish (post Gar) = deleted region Birds = N	Both	H	NA
641 S	3 LTP + 2 HTP MAPK- Nuclear localisation Mylonis et al,. 2006 + 2008 + Han et al., 2016	Ungulates/Whales/Dolphins = F Birds = P Fish (in clades) = A/V/deleted region	21%	D	NA
643 S	3 LTP + 6 HTP MAPK- Nuclear localisation Mylonis et al,. 2006 + 2008 + Han et al., 2016	Lizards/Snakes = A/I	Both	G	NA
657 s	1LTP + 2 HTP Plk3- Degradation Xu et al., 2010	Conserved	Both	T	21% + 1% (626)
687 S	1 LTP + 5 HTP (687) CDK5- stabilising Herzog et al., 2016	Frogs/Reptiles = T Birds (in clades) = A/V Fish (in clades) = T/S/A/deleted region	21%	S	21% + 1% (672)
696 S	1 LTP + 1 HTP (696) ATM- stabilising Cam et al., 2010	Birds/Frogs = N Lizards/Snakes = G Fish (in clades) = N/K	Both	T	X

HIF-2α					
Site	Characterisation	Conservation	Observed in: 21%, 1% or Both?	HIF-1α alignment	phosphorylated HIF-1α?
383 S	1 LTP Inhibits nuclear export Pangou et al. (2016)	Birds/Reptiles = T Fish (single clade) = G/A	Both	S	21% + 1% (380)
528 T	1 LTP Inhibits nuclear export Pangou et al. (2016)	Conserved	21%	M	NA
830 S	1 LTP (mouse) Transcriptional activation Gradin et al., 2002	Conserved	1%	S	21% + 1% (786 S)

4.7.1.1. Serine 31, a potential DNA binding role

Utilising a recombinant fragment of HIF-1α in an *in vitro* PKA assay, Bullen *et al.*, 2016 identified multiple sites of phosphorylation (see introduction, 1.7); including S31 and S465. Based on the lack of observable effects on protein stability, when expressing the HIF-1α fragments in cells which contained a phospho-null mutation (serine to alanine), both sites were deemed not to have a biological function and not investigated further. Importantly, only investigating a phospho-null mutation removes the effect that phosphorylation may have through the introduction of charge. Similarly, it is possible that phosphorylation may not have a resultant effect on protein stability but rather transactivation. Thus the

evolutionary history of both S31 and S465 poses a potentially interesting view, particularly for S31 which resides in the bHLH domain, an almost 100% conserved domain in all species of HIF-1α (Figure 4.4 (A)) and HIF-2α. Interestingly, the residue at position 31 is a non-phosphorylatable glycine residue in all bony fish species (*Osteichthyes*). To investigate this site further, I performed Pymol modelling on the partial HIF-1α-HIF-1β-DNA crystal structure (PDB: 4ZPR, Wu *et al.*, 2015), using the PyTMs plugin to *in silico* phosphorylate S31 (Warnecke *et al.*, 2014, Figure 4.4 (B)).

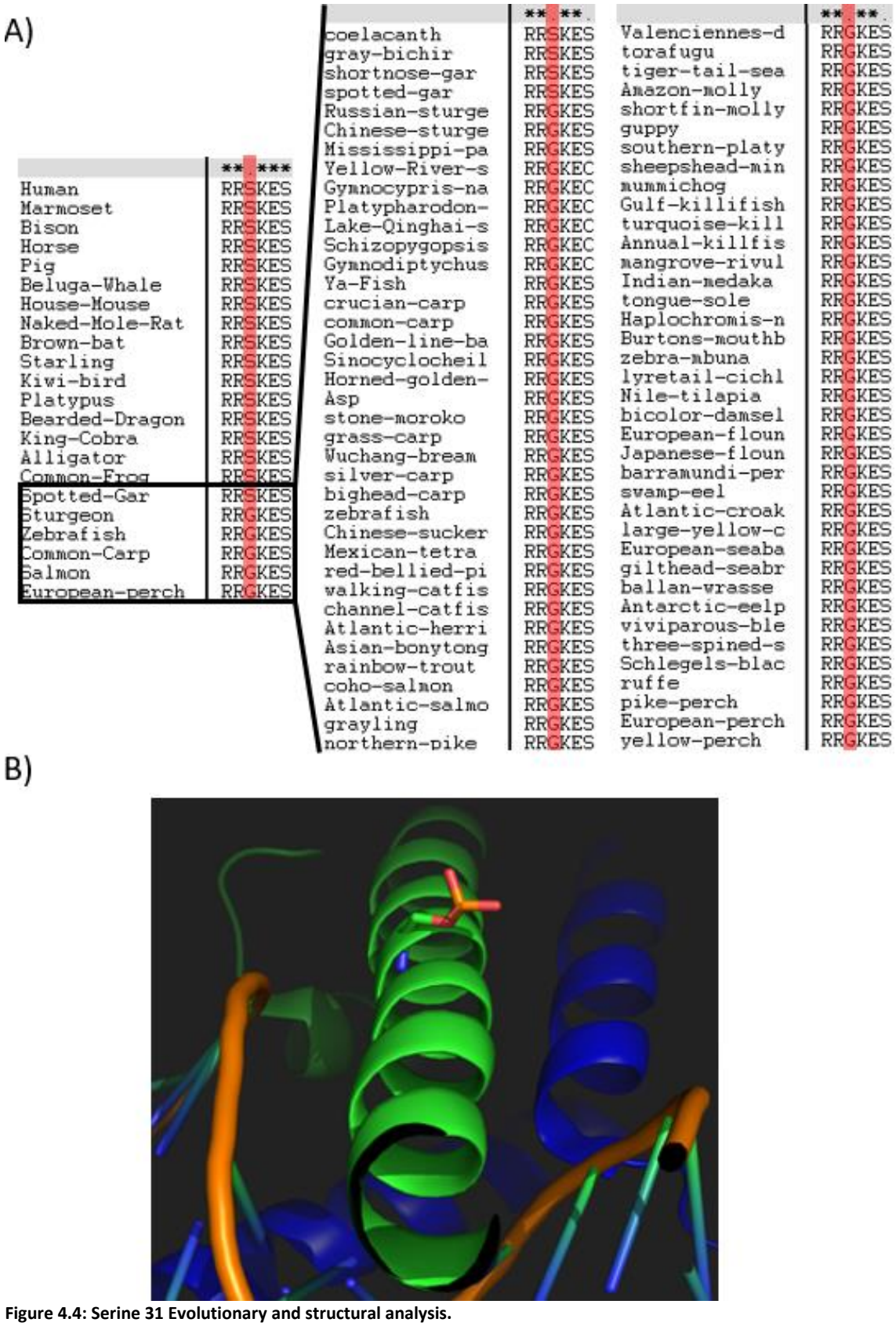


Figure 4.4: Serine 31 Evolutionary and structural analysis.
A) Multiple sequence alignment comparisons of residues 29-34, arrow identifies S31. Selection of model organisms from a variety of phylogenetic families within the full dataset are shown, boxed region shows the Bony fish species (Osteichthyes), which is expanded to include all Bony fish species within the dataset. B) Crystal structure of phosphorylated S31. PyTMs plugin (Warnecke et al., 2014) was used to *in silico* phosphorylate S31 in PDB file 4ZPR (Wu et al., 2015). Green = HIF-1α, Blue = HIF-1β, Orange = DNA backbone.

Figure 4.4 (B) shows that S31 phosphorylation is in close proximity to the DNA backbone. It could be hypothesised that the negatively charged phosphate group may repel the negatively charged DNA backbone, thus resulting in transcriptional inhibition induced by phosphorylation, due to decreased ability to bind DNA. Hence, Bullen *et al.*, 2016 would not have discovered the role of S31 phosphorylation, based only on protein stability measurements and phospho-null mutations. Interestingly, in our dataset and published data, there are no reports of phosphorylation at the respective HIF-2α serine residue (S28), thus identifying a potential HIF-1α specific pathway. Sequence alignment comparisons between isoforms finds the only amino acid different in the surrounding region is HIF-1α S28 to HIF-2α C25, identifying a potentially important site for motif analysis.

4.7.1.2. Nuclear localisation

HIF-1α phosphorylation at S641 and/or S643, by ERK1/2, have been shown to be responsible for nuclear accumulation, by preventing its nuclear export (Mylonis *et al.*, 2006 & Mylonis *et al.*, 2008). In their study, MS analysis was inconclusive to determine which serine residue, if not both, were phosphorylated. However phospho –mimetic (nuclear accumulation) and –null (nuclear exclusion) mutations of either site resulted in the same phenotypic effect. Data presented here provide strong evidence that both phosphorylation sites co-exist as a doubly phosphorylated peptide (appendix 1). Interestingly, S643 phosphorylation was detected as a singly phosphorylated peptide while S641 was only identified in the presence of S643, as a doubly phosphorylated peptide. This suggests a possible priming, or synergistic, mechanism where S643 phosphorylation is essential for the secondary phosphorylation at S641 (as seen with GSK3β signalling (Beurel *et al.*, 2015)). In agreement with this hypothesis, S643 is highly evolutionarily conserved while S641 is poorly conserved, with variation primarily to non-phosphorylatable residues. Hence, these sites may work synergistically or have slightly different roles not yet discovered.

HIF-2α phosphorylation at S383 and T528, by CK1δ, have previously been shown to be responsible for nuclear accumulation, in an identical mechanism to HIF-1α S641/643 (Pangou *et al.*, 2016). Interestingly, both S383 and T528 are highly conserved across vertebrate species, with 100% conservation of T528, suggesting an important role. Sequence comparisons identifies that HIF-2α S383 aligns with the novel S380 phosphorylation site of HIF-1α. The surrounding sequences of these phosphorylation sites are poorly conserved between isoforms. Interestingly, HIF-2α T528 phosphorylation is O_2 dependent, only present at 21% O_2, and is within very close proximity to the proline

hydroxylation site P531. Thus, T528 may have underlying roles not identified by Pangou *et al.*, 2016, discussed further in the Proline hydroxylation proximity section (4.7.5).

4.7.1.3. Protein stability

The remaining HIF-1α phosphorylation sites (S657, S687 and S696) have all been shown to regulate protein stability. S657, phosphorylated by Plk3, has been shown to promote HIF-1α degradation (Xu *et al.*, 2010), while both S687, phosphorylated by CDK5 (Herzog *et al.*, 2016) and S696, phosphorylated by ATM (Cam *et al.*, 2010) stabilise HIF-1α. S657 is a 100% conserved site and sequence alignment comparisons finds that it aligns to the, novel T626 phosphorylated residue of HIF-2α, which too is highly conserved. However, sequences surrounding S657 of HIF-1α and T626 of HIF-2α are poorly conserved between isoforms.

HIF-1α S687 phosphorylation by CDK5 suggests a role in cell cycle progression, thus evolutionary analysis identifying that a serine residue only exists in primate species is intriguing. The ability for phosphorylation is maintained in other mammals by variaton to a threonine residue, except for Ungulate species which instead have a non-phosphorylatable alanine residue. All other vertebrate species contain a non-phosphorylatable residue, with no distinct residue selection. Interestingly, HIF-1α S687 aligns to the novel phosphorylation site S672 of HIF-2α, which has an identical evolutionary pattern to the former. However, the surrounding sequence is poorly conserved, this may provide an explanation to how HIF-1α S687 phosphorylation only occurs in 21% O_2 while HIF-2α S672 is O_2 independent. S696 of HIF-1α is evolutionarily poorly conserved, and although aligns to a threonine residue of HIF-2α the latter is not phosphorylated.

4.7.1.4. Transcriptional effects

Of all phosphorylation sites that have previously been characterised and identified in this study, only HIF-2α S830 has been shown to effect transcription. S830 phosphorylation resulted in an increase of transcriptional output by ~2 fold, without affecting protein stability, although this was not investigated further (Gradin *et al.*, 2002). S830 resides in the CTAD, involved with p300/CBP binding interactions, which is a 100% conserved domain among all vertebrate species analysed. Sequence comparisons between isoforms shows that HIF-2α aligns to the novel phosphorylation site of HIF-1α: S786, which too is 100% conserved. The surrounding regions of respective sites are also highly conserved, both between isoforms and evolutionarily, thus suggesting a potentially essential function shared between isoforms, such as regulating p300/CBP association. Interestingly, HIF-2α

S830 phosphorylation was only detected in 1% O_2 while HIF-1α S786 phosphorylation was detected O_2 independently.

All phosphorylation sites discussed below are novel and therefore have no functional assessment in the narrative.

4.7.2. ODDD hyperphosphorylation

The ODDD was named due to containing the oxygen sensitive proline resides, which upon hydroxylation leads to degradation (see introduction 1.6). However, a striking feature in our data is the hyperphosphorylation of the HIF-1α ODDD, especially compared to the HIF-2α ODDD. Indeed, there was 15 and 5 phosphorylation sites identified respectively (Figure 4.5), listed in Table 4.4 and Table 4.5 respectively.

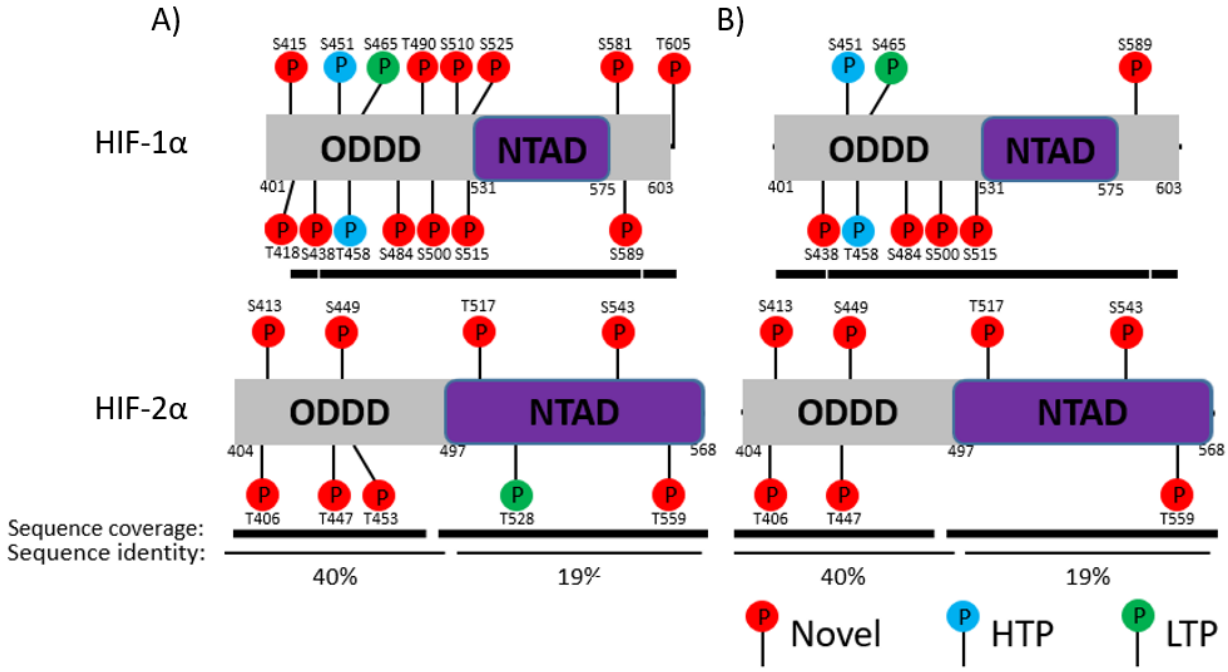


Figure 4.5: Schematic view of the ODDD and NTAD phosphorylation map of HIF-1α and HIF-2α.
Stated domains of HIFα proteins are highlighted, with canonical domain numbering. Confidently identified phosphorylation sites (1% FDR, ptmRS score >99.0) are mapped to the protein and coloured dependent on whether they are novel (red), have been seen by previous HTP mass spectrometry studies (blue) or have previously undergone functional characterisation (green). The sequence coverage seen from mass spectrometry analysis (from Figure 4.1, bold line) and the sequence homology between isoform domains (fine line and score) is depicted below the protein schematic. A) Phosphorylation sites identified in 1% O_2, B) Phosphorylation sites identified in 21% O_2.

Table 4.4: Characteristics of HIF-1α phosphorylation sites identified within the ODDD.
Table includes position of PTM, how many studies have identified the respective site with function and reference, evolutionary history where variations are highlighted by single letter amino acid code, the O_2 tension a PTM was identified in and comparisons to HIF-2α sequence. NA = not applicable, X = residue was not phosphorylated.

Site	Novel	Conservation	Observed in: 21%, 1% or Both?	HIF-2α alignment	phosphorylated HIF-2α?
415 S	Y	Sharks = P Fish (in clades) = K/N/R/C	21%	NA	NA
418 T	Y	Only in Mammals Deleted region in all other species	21%	NA	NA
438 S	Y	Conserved	Both	NA	NA
451 S	5 HTP	Conserved	Both	NA	NA
458 T	1 HTP	Frogs = A/P Most fish = deleted Remaining fish = P	Both	W	NA
465 S	2 LTP + 2 HTP PKA- no investigation Bullen et al,. 2016 + So et al,. 2016	Deleted in all fish post Gar Birds = N	Both	H	NA
484 S	Y	Monotreme/Marsupials/Repltiles/Gar/Birds = P Fish (Carp) = D Remaining Fish = deleted region	Both	NA	NA
490 T	Y	Reptiles = S Fish (Carp) = A Remaining Fish = E	21%	T	X
500 S	Y	Fish (in clades) = D/A/T Lizards/Snakes = T	Both	T	X
510 S	Y	Fish (single clade) = P	21%	C	NA
515 S	Y	Fish (single clade) = N/G	Both	S	X
525 S	Y	Monotreme/Marsupials/Repltiles/Birds = N	21%	X	NA
581 S	Y	Fish (except Gar) = D/E Turtles = P	21%	C	NA
589 S	Y	Birds/Crocodiles = G Fish (in clades) = Q/K/S Whales/Dolphins = T	Both	E	NA
605 T	Y	Only in Mammals/Birds/Crocodiles Deleted region in all other species	21%	T	X

Interestingly, 7 (out of 15) of the HIF-1α phosphorylation sites within the ODDD are O_2 dependent, being observed in my investigations only in 21% O_2: S415, T418, T490, S510, S525, S581 and T605. Surprisingly, no phosphorylation sites within the ODDD were identified in 1% O_2 only. Thus, it is likely that the O_2 dependent regulation of HIF-1α protein stability, through the ODDD, is more complex than initially thought and may be influenced by phosphorylation status. It could be hypothesised that the hyperphosphorylation status of the ODDD in 21% O_2 may enhance the efficiency of HIF-1α degradation, potentially by promoting PHD interactions and subsequent degradation. Interestingly none of these phosphorylation sites are highly conserved, with approximately half of these residues present in mammalian species only. Furthermore, the majority of residue variation at these phosphorylation sites were to non-phosphorylatable residues, or to regions of the protein that are missing, suggesting a potentially important regulatory mechanism that has evolved specifically in mammals. Similarly, since HIF-1α and HIF-2α share ~40% sequence homology in the ODDD, it is surprising that none of the phosphorylation sites identified in HIF-1α mapped to phosphorylation sites of HIF-2α. If these HIF-1α phosphorylation sites have a role in protein degradation, it may partially explain why HIF-2α regulation is less dependent

on O_2 dependent degradation compared to HIF-1α. Intriguingly, HIF-1α phosphorylation sites S510 and S581 both align to Cysteine (C) residues of HIF-2α, which can experience redox sensitive PTMs and may have unique roles (reviewed by: Chung *et al.*, 2013). Although there is currently no evidence for redox-mediated modification of these C residues.

We observed that the O_2 dependent phosphorylation sites: T490, S510 and S525 are interspersed with O_2 independent phosphorylation sites: S484, S500 and S515. This region is the most densely phosphorylated portion of either HIF-1α or HIF-2α, and has an approximate repeating unit of a multiple of four, potentitally suggesting this region could be α-helical in structure (Pauling *et al.*, 1951). Theoretically, this could allow an O_2 independent phosphorylation face of the α-helix to be used in essential functions, such as correct protein folding, while allowing an O_2 dependent phosphorylation face to have secondary functions, such as protein degradation. Without a crystal structure it is impossible to validate this theory. Therefore we adopted an *ab initio* (sequence inferred) modelling approach of a 60 residue fragment encompassing these 6 phosphorylation sites (A475-L535), using the Robetta online server (Song *et al.*, 2013 & Raman *et al.*, 2009). The most likely model is shown in Figure 4.6 and has been phosphorylated *in silico* at all 6 identified sites. Importantly, the homology aspect of modelling was very poor, at a score of 0.05 (between 0 (no crystal structures for modelling) – 1 (complete crystal structure)), hence the model is highly reliant on sequence based protein structure prediction, which although has limitations on large proteins can be useful for smaller proteins (Lee *et al.*, 2017).

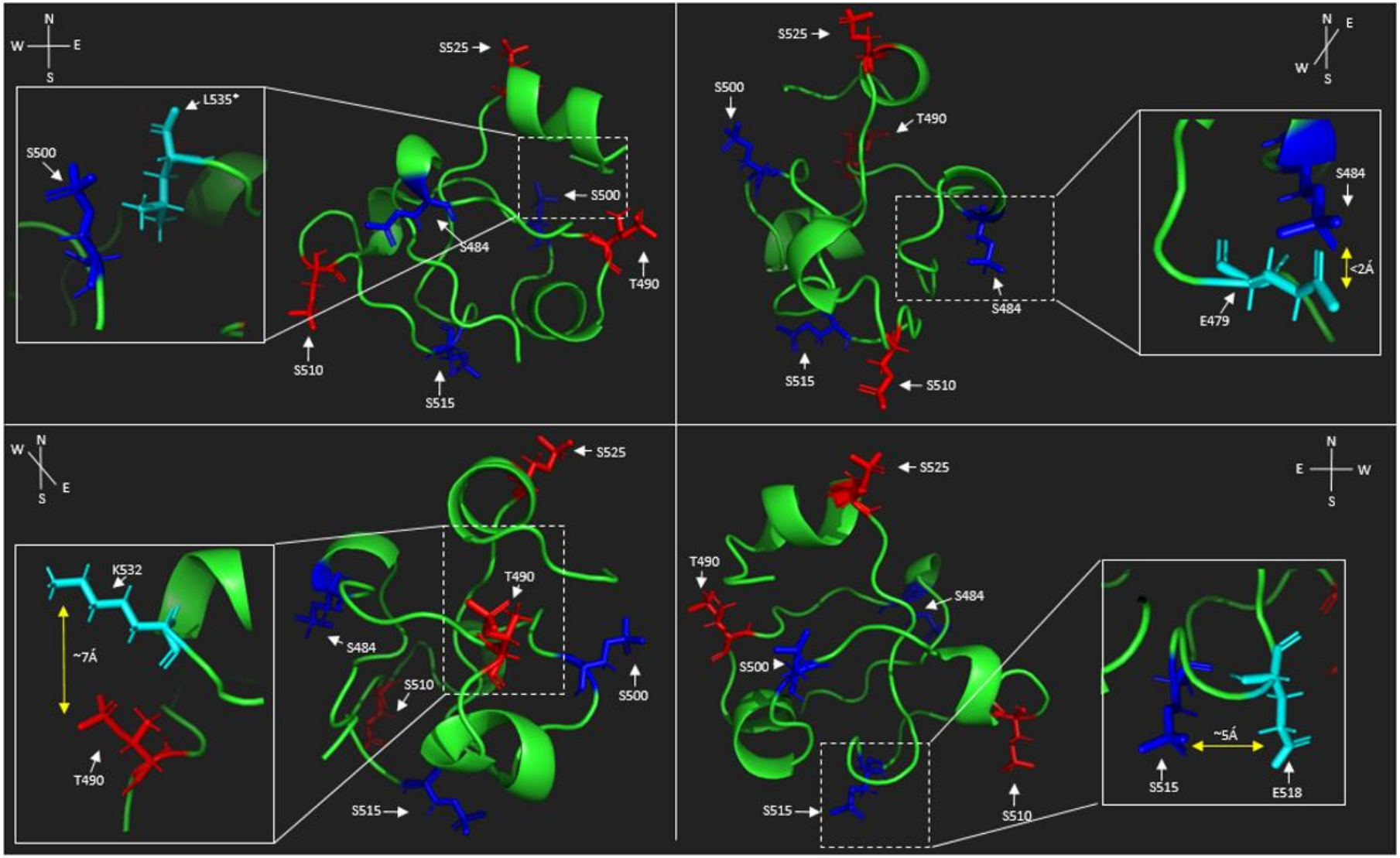


Figure 4.6: Combined homology and *ab initio* modelling of the high phosphorylation density region of the HIF-1α ODDD.
A residues from A475-L535 were analysed through the Robetta online server tool (Song *et al.*, 2013 & Raman *et al.*, 2009). The model of highest significance was taken for Pymol modelling and *in silico* phosphorylated at identified sites using the Pytms plugin (Warnecke,. et al. 2014) O_2 insensitive phospho-sites (S484/S500/S515) coloured blue, O_2 sensitive sites(only present in 21% O_2) (T490/S510/S525) coloured red. Dotted boxed regions highlight where phospho-sites are in close proximity to other residues within the model, expanded in solid boxed regions. Close proximity residues coloured cyan. Distances between these residues were calculated using the Pymol distance tool. All sites are labelled using the human HIF-1α protein sequence. Orientation of model is shown by compass directions. * indicates last residue in model

Figure 4.6 shows that although there are α-helical regions present, the initial hypothesis of a large α-helix is incorrect. However, it is clear that the O_2 independent phosphorylation sites are in very close proximity to negatively charged glutamic acid residues: S484-E479 is <2 Á and S515-E518 is <5 Á, thus are likely to experience strong repulsive interactions (Barlow *et al.*, 1983). We notice that the last O_2 dependent site (S500) is ~7 Á in distance from Leucine (L) 535, the last residue that was used in this model. Investigating the next amino acids identifies E537, therefore it is possible this glutamic acid residue is within closer proximity to the S500 phosphorylation site than the modelled leucine residue; resulting in charge repulsion. Therefore, although not α-helical as initially hypothesised, the O_2 independent phosphorylation sites are likely to have a dramatic role on protein folding. Additionally, all O_2 dependent phosphorylation sites (T490, S510 and S25) are modelled to face externally, thus potentially acting as docking sites for binding partners. However, without knowing the stoichiometry of how these phosphorylation sites occur, it is possible that all occur simultaneously. Thus, the O_2 independent phosphorylation sites, facing negatively charge glutamic acid residues, may result in large structural rearrangements, hence the O_2 dependent sites will likely be in different positions to as modelled here.

Evolutionary analysis of the modelled region shows that it is well conserved in the majority of vertebrate species analysed, except for Bony fish which show high rates of variation within this region; interesting considering they experience temporal hypoxia much more frequently than most species. Interestingly, sequence comparison to HIF-2α shows that this region is poorly conserved between isoforms, but is rich in serine/threonine residues; containing a total of 13 sites. Unfortunately, large parts of this region were not detectable by the MS/MS methods used, thus we were unable to evaluate the phosphorylation status of this region.

Remaining HIF-1α phosphorylation sites within the ODDD (S438, S451, T458 and S589) are all O_2 independent. Although all novel in terms of LTP studies, S451 phosphorylation has been observed in 5 separate HTP MS studies, thus supporting its existence. Interestingly, S451 is a 100% conserved site among species and lies within a highly conserved region between species, suggesting a functionally important role. Sequence comparisons with HIF-2α show that this region is missing and therefore could be hypothesised to have a role in differentiating HIF-1α from HIF-2α. S438 and T458 are both highly conserved, with the former 100% conserved in all species analysed. S589 is much less conserved with non-

phosphorylatable residues of glycine and lysine identified in bird species and bony fish species respectively. The identification that a negatively charged phosphorylation site aligns to a positively charged lysine residue is particularly interesting, and may suggest a potential charge importance role. For example, phosphorylation may inhibit a binding partner through charge repulsion, while a lysine residue can promote it by strengthening electrostatic interactions. Sequence comparisons to HIF-2α identifies that a negatively charged glutamic acid residue aligns to this phosphorylation site, thus could support this hypothesis and suggest a permanent role for HIF-2α in signal transduction and temporal role for HIF-1α.

Table 4.5: Characteristics of HIF-2α phosphorylation sites identified within the ODDD.
Table includes position of PTM, how many studies have identified the respective site with function and reference, evolutionary history where variation are highlighted by single letter amino acid code, the O_2 tension a PTM was identified in and comparisons to HIF-1α sequence. NA = not applicable, X = residue was not phosphorylated.

Site	Novel	Conservation	Observed in: 21%, 1% or Both?	HIF-1α alignment	phosphorylated HIF-1α?
406 T	Y	Fish (single clade) = M	Both	A	NA
413 S	Y	Fish (single clade) = M	Both	S	X
447 T	Y	Only in Primates/Cats/Ungulates Frogs/Marsupials/Fish/Reptiles = S Fish (single clade) = P All other species = A	Both	D	NA
449 S	Y	All birds = G Fish (in clades) = G/R	Both	A	NA
453 S	Y	All birds/reptiles = T Fish (in clades) = S/ N/K	21%	E	NA

Comparison of Table 4.4 and Table 4.5 highlights how different the phosphorylation status of the ODDD is between HIF-1α and HIF-2α. Interestingly, T406 and S413 are virtually conserved in all vertebrate species analysed and are in close proximity to the proline hydroxylation site P405, discussed further in the proline hydroxylation proximity section (4.7.5).

The remaining phospho-sites T447, S449 and S453 form a tight cluster, however phosphorylated T477 is identified as an exclusive site while phosphorylated S449 and S453 are identified as singly and doubly phosphorylated peptides (Table 4.6). It may be hypothesised that the T447 and S449/S453 phosphorylation sites could have different roles, with the latter two phosphorylation sites potentially having synergistic effects. However, this would require further investigation to support such statements. Evolutionary analysis reveals that all three sites are poorly conserved, with no species other than mammals having all three residues.

Table 4.6: Identifying peptides and confidence of T447, S449 and S453 phospho-sites.
Includes the peptide sequence, Uppercase letter = unmodified, Lowercase letter = phosphorylation site, the number of PSMs that identified a particular peptide, the ptmRS score, peptide confidence score and E-value from the best PSM.

Site	Peptide	# PSMs	PTMrs best	Score	E-value
447 T	ATELRSHStQSEAGSLPAF	12	T9(1xPhospho): 99.27	36	5.60E-03
449 S	HSTQsEAGSLPA	3	S5(1xPhospho): 100	58	3.40E-05
453 S	RSHSTQSEAGsLPAF	2	S11(1xPhospho): 100	51	1.70E-04
449 S + 453 S	RSHSTQsEAGsLPAF	4	S7(1xPhospho): 99.78; S11(1xPhospho): 100	45	6.70E-04

Overall, the functional characterisation of the identified phosphorylation status of the HIFα isoforms ODDD will be an interesting avenue to explore in the future, potentially providing an insight for better understanding of the fine tuning of HIFα protein stability at 21% O_2.

4.7.3. N-Terminal Transactivation Domain (NTAD) hypophosphorylation:

The NTAD is the fundamental region involved with determining gene specificity of the HIFα isoforms, and has the least sequence similarity of the canonical domains (Hu *et al.*, 2007 & Tian *et al.*, 1997). We therefore hypothesised that the NTAD may experience heavy PTM regulation in order to alter binding partners and gene specificity. Thus the complete absence of identified phosphorylation sites within the HIF-1α NTAD is surprising (Figure 4.7). This observation is supported by published data too, where a distinct lack of phosphorylation is identified within the NTAD of HIF-1α (data from PhosphoSitePlus). In fact, only one publication has identified phosphorylation within the HIF-1α NTAD at S551 and T555, both resulting in protein degradation; thus may explain why we were unable to identify these sites (Flügel *et al.*, 2007 & Flügel *et al.*, 2012).

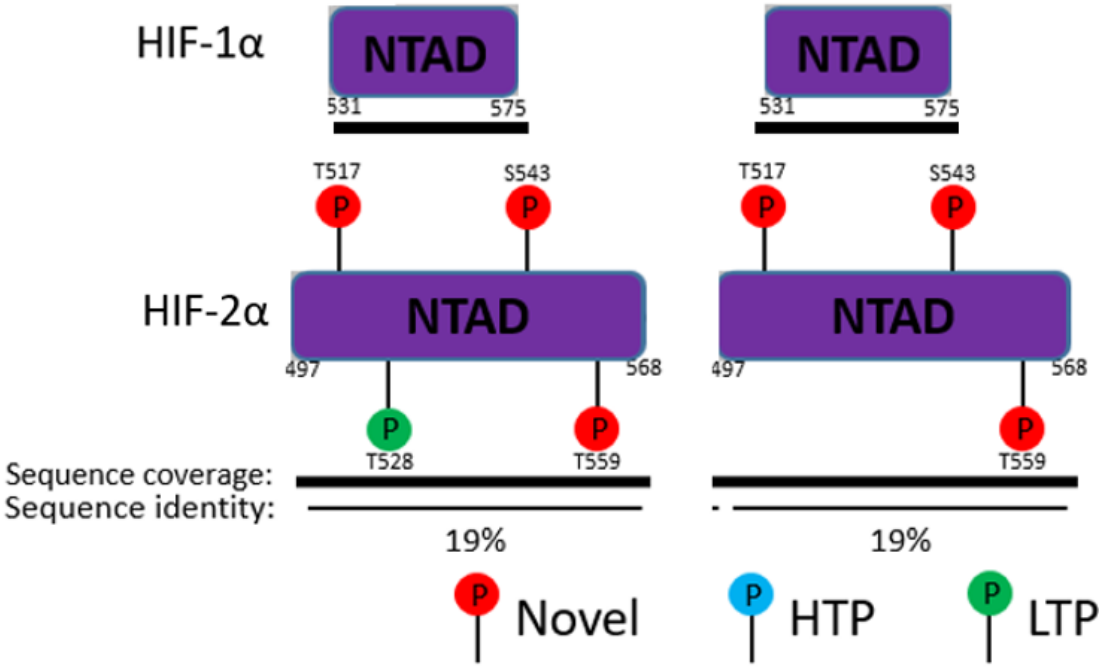


Figure 4.7: Schematic view of the NTAD phosphorylation map of HIF-1α and HIF-2α.
Confidently identified phosphorylation sites (1% FDR, ptmRS score >99.0) are mapped to the protein and coloured dependent on whether they are novel (red), have been seen by previous HTP mass spectrometry studies (blue) or whether they have undergone previous characterisation (green). The sequence coverage seen from mass spectrometry analysis (from Figure 4.1) is depicted below the protein schematic (bold line) along with the sequence homology between isoform domains (fine line and score). A) Phosphorylation sites identified in 1% O_2, B) Phosphorylation sites identified in 21% O_2.

Table 4.7: Characteristics of HIF-2α phosphorylation sites identified within the NTAD.
Table includes position of PTM, how many studies have identified the respective site with function and reference, evolutionary history where variation are highlighted by single letter amino acid code, the O_2 tension at which the PTM was identified in and comparisons to HIF-1α sequence. NA = not applicable, X = residue was not phosphorylated.

Site	Novel	Conservation	Observed in: 21%, 1% or Both?	HIF-1α alignment	phosphorylated HIF-1α?
517 T	Y	Birds/Reptiles = S Fish (in clades) = S/Q	Both	T	X
528 T	1 LTP CK1δ- Inhibits nuclear export Pangou et al. (2016)	Conserved	21%	M	NA
543 S	Y	Fish (in clades) = H/N	Both	D	NA
559 T	Y	Ungulates = A	Both	D	NA

Figure 4.7 and Table 4.7 show that HIF-2α is phosphorylated at 4 positions within the NTAD: T517, S543, T559 and T528, the latter discussed in the proline hydroxylation proximity section (4.7.5). All 4 sites are highly conserved across vertebrate species, although the NTAD itself is poorly conserved. Despite the poor sequence homology between HIF-1α and HIF-2α, it is interesting that the HIF-2α S543 and T559 phospho-sites both align to negatively charged aspartic acid residues of HIF-1α.

Overall, it will be interesting to investigate whether the HIFα isoforms undergo novel PTMs of different types in the NTAD domain, such as acetylation or ubiquitination which have previously been published (Jeong *et al.*, 2002 & Paltoglou *et al.*, 2007). Alternatively, it will be interesting to investigate whether the NTAD undergoes PTM regulation in different cell lines, which may explain differences seen (Bracken *et al.*, 2006), or whether the sequence differences alone are responsible for the specificity of HIFα gene targeting.

4.7.4. O_2 dependent phosphorylation:

Although the process of phosphorylation does not directly require O_2 as a cofactor, wit is known that O_2 tension can regulate kinase (or phosphatase) function by PTM (Shao *et al.*, 2014), therefore it can also be hypothesised that signalling pathways in response to hypoxia may change. Therefore the identification of O_2 dependent phosphorylation sites is not surprising, with a total of 12 and 8 phosphorylation sites only identified in an O_2 dependent manner for HIF-1α (Table 4.8) and HIF-2α (Table 4.9) respectively.

Table 4.8: Characteristics of HIF-1α phosphorylation sites identified that are O_2 dependent.
Table includes position of PTM, how many studies have identified the respective site with function and reference, evolutionary history where variation are highlighted by single letter amino acid code, the O_2 tension a PTM was identified in and comparisons to HIF-2α sequence. NA = not applicable, X = residue was not phosphorylated.

Site	Novel	Conservation	Observed in: 21%, 1% or Both?	HIF-2α alignment	phosphorylated HIF-2α?
31 S	1 LTP PKA- no investigation Bullen et al,. 2016	Fish (except Gar) = G	1%	S	X
384 S	Y	Birds/Reptiles/Sharks/Sturgeon = N Fish (in clades) = L/P/S/D/A	1%	N	NA
415 S	Y	Sharks = P Fish (in clades) = K/N/R/C	21%	NA	NA
418 T	Y	Only in Mammals Deleted region in all other species	21%	NA	NA
490 T	Y	Reptiles = S Fish (Carp) = A Remaining Fish = E	21%	T	X
510 S	Y	Fish (single clade) = P	21%	C	NA
525 S	Y	Monotreme/Marsupials/Repltiles/Birds = N	21%	X	NA
581 S	Y	Fish (except Gar) = D/E Turtles = P	21%	C	NA
605 T	Y	Only in Mammals/Birds/Crocodiles Deleted region in all other species	21%	T	X
641 S	3 LTP + 2 HTP MAPK- Nuclear localisation Mylonis et al,. 2006 + 2008 + Han et al., 2016	Birds = P Fish (in clades) = A/V/deleted region	21%	D	NA
652 T	1 HTP	Ungulates = A/P Fish (in clades) = C/H	21%	C	NA
687 S	1 LTP + 5 HTP CDK5- Protein stability Herzog et al., 2016	Frogs/Reptiles = T Birds (in clades) = A/V Fish (in clades) = T/S/A/deleted region	21%	S	21% + 1% (672)

Interestingly, only two phosphorylation sites on HIF-1α were identified in 1% O_2 only: S31 and S384, the former discussed previously in previously characterised section (4.7.1.1) and the latter in the negatively charged cluster section (4.7.7). The remaining ten phospho-sites that were differentially regulated by O_2 tension were only identified in 21% O_2: S415, T418, T490, S510, S525, S581, T605, S641, T652 and S687. Of these, only T652 has not been discussed above, phospho-sites S415-T605 were discussed in the ODDD section (4.7.2) and S641 & S687 were discussed in the previously characterised section (4.7.1.2 and 4.7.1.3 respectively). T652 has been identified previously in a single HTP study and was only detected as a doubly phosphorylated peptide in this study with S657. As we identify S657 as a singly phosphorylated peptide it suggests that a possible priming mechanism may exist, in which S657 is initially phosphorylated and can result in secondary phosphorylation at S652, possibly affecting the functional degradation role of S657 (Xu *et al.*, 2010).

Table 4.9: Characteristics of HIF-2α phosphorylation sites identified that are O_2 dependent.
Table includes position of PTM, how many studies have identified the respective site with function and reference, evolutionary history where variations are highlighted by single letter amino acid code, the O_2 tension a PTM was identified in and comparisons to HIF-1α sequence. NA = not applicable, X = residue was not phosphorylated.

Site	Novel	Conservation	Observed in: 21%, 1% or Both?	HIF-1α alignment	phosphorylated HIF-1α?
345 S	Y	Conserved	1%	S	X
453 S	Y	Birds/Reptiles = T Fish (single clade) = N/K	21%	E	NA
528 T	1 LTP Inhibits nuclear export Pangou et al. (2016)	Conserved	21%	M	NA
581 S	Y	Fish (in clades) = G/deleted region Bats = G Reptiles/Birds = N	21%	T	X
597 T	Y	Whales/Dolphins/Crocodiles= I Birds (in clades) = I/M Remaining Reptiles = M Fish (in clades) = S/K/R	1%	E	NA
606 S	Y	Snakes/Lizards = L Whales/Dolphins/Ungulates = F Fish (in clades) = S/A/P	1%	I	NA
683 S	Y	Birds = P Fish (single clade) = F/P Marsupials = L	21%	R	NA
830 S	1 LTP (mouse) Transcriptional activation Gradin et al., 2002	Conserved	1%	S	21% + 1% (786 S)

HIF-2α phosphorylation status appears to be more O_2 dependent, with 4 sites specifically identified in 21% O_2 (S453, T528, S581 and S683) and in 1% O_2 (S345, T597, S606 and S830), 3 of these sites are discussed elsewhere: S453 in the ODDD section (4.7.2), T528 in the proline hydroxylation proximity section (4.7.5), and S830 in the previously characterised section (4.7.1.4). Both remaining 21% O_2 only phospho-sites (S581 and S683) lie within an

undefined domain of HIF-2α that is poorly conserved among all species investigated. Similarly, there is poor sequence homology between HIFα isoforms within this region, thus it is difficult to determine their importance or their potential role beyond differentiateing HIFα isoforms. Of the remaining 1% O_2 only phospho-sites, T597 and S606 are both within the same undefined domain of HIF-2α, as S581 and S683, and are equally poorly conserved. Based on my observations of numerous O_2 dependent phosphorylation sites, it is possible that this undefined domain could have an important role in HIF-2α O_2 dependent regulation.

The last HIF-2α 1% O_2 only phosphorylation site (S345) is within the PAS-A/B domains and is highly conserved among all vertebrate species analysed. S345 has been crystallised in the HIF-2α-HIF-1β crystal structure complex (PDB: 4ZP4, Wu *et al.*, 2015), therefore I could model the S345 phospho-site (Figure 4.8). Figure 4.8 (A) shows that S345 phosphorylation introduces the negative charge within 2 Å from a negatively charged glutamic acid residue (E348, cyan), thus likely resulting in structural rearrangements through charge repulsion. Figure 4.8 (B) shows that this region of HIF-2α (green) is encapsulated by HIF-1β (blue). Although the amino acids 347-359 of HIF-1β failed to crystallise, it is clear that to connect the two fragments of HIF-1β that did crystallise then the unstructured region would have to pass over the HIF-2α S345 containing domain. Therefore, it is plausible that if phosphorylation at S345 results in a structural rearrangement, then the HIF-1β binding domain would be disrupted too; potentially regulating HIF-2α dependent transcription by regulating dimerization in an O_2 dependent, phosphorylation dependent manner.

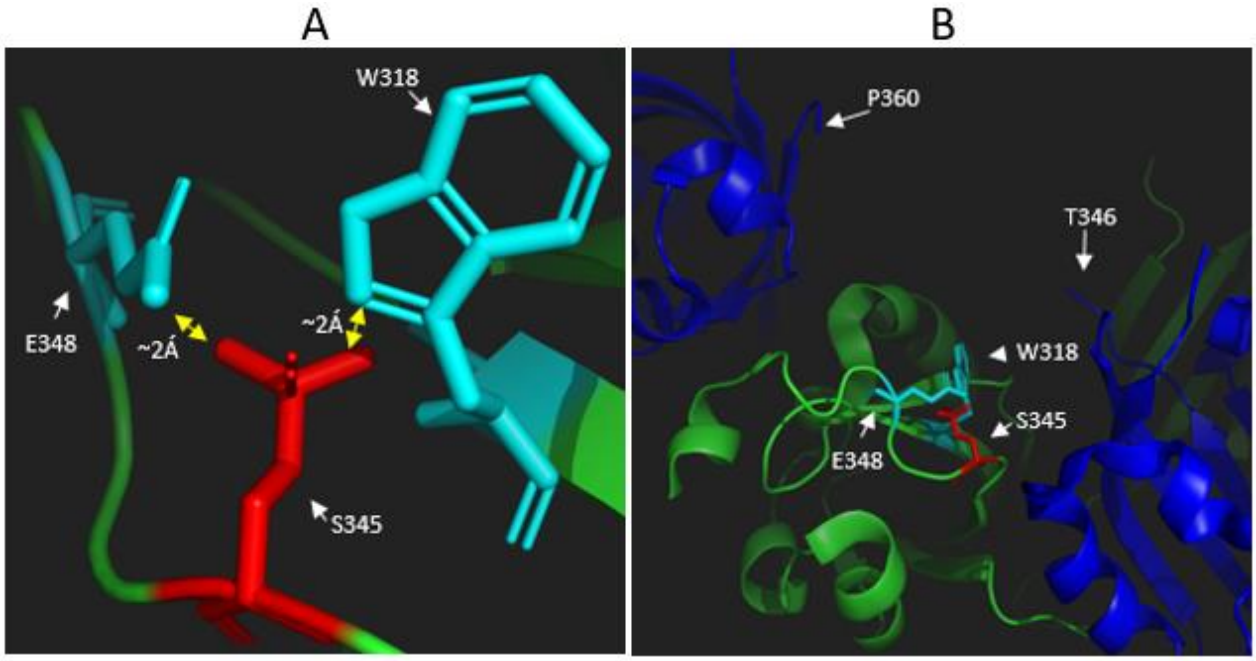


Figure 4.8: Phosphorylated S345 of HIF-2α and potential resultant affects.
Crystal structure (PDB: 4ZP4, Wu *et al.*, 2015), modelled in Pymol and phosphorylated using the PyTMs plugin (Warnecke *et al.*, 2014). Green: HIF-2α, Blue: HIF-1β, Red: phospho-S345, Cyan: possible S345 interacting residues; W318 & E348. Distances measured using Pymol Distance wizard. A) Close proximity of phospho-S345 to W318 & E348. B) Zoomed out image to highlight possible role in HIF-1β binding. Residues between T346-P360 did not crystallise (marked on the structure to highlight the start and end points of the missing region).

4.7.5. Proline hydroxylation proximity

The canonical proline hydroxylation sites are P402/P564 for HIF-1α and P405/P531 for HIF-2α. These proline hydroxylation sites reside in highly conserved regions between isoforms (Figure 4.9). From domain analysis, the LAP (leucine, alanine, proline-hydroxylated) motif is essential for proline hydroxylation, observed at all sites for both isoforms. We found that a third site of the LAP motif exists within HIF-2α specifically (P576), which resides in a region of very poor sequence homology to HIF-1α (Figure 4.9). Evolutionary analysis shows that this third LAP motif only exists in mammalian species and thus may suggest an important, isoform specific, evolved function.

Figure 4.9: Hydroxylated Proline residues and close proximity phosphorylation sites observed with HIF-2α. Shows the aligned HIFα sequences from MUSCLE multiple sequence alignment and viewed in clustal X. * indicate identical residues at a position, : identifies a strong conservation mutation, . identifies a weak conservation mutation and blank spaces indicate no conservation mutation (determined by clustal matrix). Red boxes highlight the LAP motif for Proline hydroxylation and arrows identify sites of HIF-2α specific phosphorylation.

Interestingly, as Figure 4.9 shows, although the canonical proline hydroxylation sites are highly conserved between HIFα isoforms, HIF-2α is specifically phosphorylated within close proximity to all hydroxylated Proline residues (arrowed): P405/T406, P531/T528 and the novel P576/S581. Additionally, HIF-1α has non-phosphorylatable residues at each respective site. Considering the increased O_2 dependent stability of HIF-2α compared to HIF-1α, it is possible that these phosphorylation sites (T406, T528 and S581) could block, or reduce the affinity of, PHD binding; thereby preventing the O_2 dependent degradation. Supporting this hypothesis, all three phosphorylation sites are identified at 21% O_2 and, for the canonical proline hydroxylation sites, are highly conserved in all species (Table 4.10).

Table 4.10: Characteristics of HIF-2α phosphorylation sites identified that are within close proximity to Proline hydroxylation sites.
Table includes position of PTM, how many studies have identified the respective site with function and reference, evolutionary history where variation are highlighted by single letter amino acid code, the O_2 tension a PTM was identified in and comparisons to HIF-1α sequence. NA = not applicable, X = residue was not phosphorylated.

Site	Novel	Conservation	Observed in: 21%, 1% or Both?	HIF-1α alignment	phosphorylated HIF-1α?
406 T	Y	Fish (single clade) = M	Both	A	NA
528 T	1 LTP CK1δ- Inhibits nuclear export Pangou et al. (2016)	Conserved	21%	M	NA
581 S	Y	Fish (in clades) = G/deleted region Bats = G Reptiles/Birds = N	21%	T	X

Evolutionary analysis of all species identifies that S581 is less conserved, however, since this LAP motif only exists in mammalian species then the conservation of S581 in these species could support this theory. Overall, HIF-2α phosphorylation in close proximity to proline hydroxylation sites may stabilise the protein in an O_2 independent manner, a mechanism

that is not present in HIF-1α; potentially explaining the 21% O_2 HIF-2α protein stability observed.

4.7.6. DNA binding and HIF-1β interaction domains:

The βHLH and PAS-A/B domains are the most highly conserved domains between HIF-1α and HIF-2α, with 85% and 75% sequence homology respectively. Therefore, it was surprising to find that 5 of the 6 phosphorylation sites identified in this region align to non-phosphorylatable residues between isoforms, and that the remaining HIF-1α S31 phospho-site aligns to the non-phosphorylated HIF-2α S28, in this study (Figure 4.10). Phosphorylation sites are: HIF-1α S31, discussed in the previously characterised section (4.7.1.1), and HIF-2α sites S46, S72, S79, S163 and S345, the latter discussed in the O_2 dependent phosphorylation section (4.7.4, Table 4.11).

Figure 4.10: Phosphorylated residues within the βHLH domain of HIF-1α and HIF-2α.
Shows the aligned HIFα sequences from MUSCLE multiple sequence alignment and viewed in clustal X. * indicate identical residues at a position, : identifies a strong conservation mutation, . identifies a weak conservation mutation and blank spaces indicate no conservation mutation (determined by clustal matrix). Red boxes highlight the phosphorylation site and arrows identify sites position of phosphorylation, number based from isoform that is modified.

Table 4.11: Characteristics of HIF-2α phosphorylation sites identified within the bHLH and PAS-A/B domains.
Table includes position of PTM, how many studies have identified the respective site with function and reference, evolutionary history where variation are highlighted by single letter amino acid code, the O_2 tension at which a PTM was identified in and comparisons to HIF-2α sequence. NA = not applicable, X = residue was not phosphorylated.

Site	Novel	Conservation	Observed in: 21%, 1% or Both?	HIF-1α alignment	phosphorylated HIF-1α?
46 S	Y	Birds/Reptiles/Gar/Shark/Marsupials = N	Both	N	NA
72 S	Y	Fish (single clade) = T	Both	A	NA
79 S	1 HTP	Birds/Crocdiles = N Lizards/Snakes = K/R Fish (in clades) = T/E/G/A/deleted region	Both	I	NA
163 S	Y	Frogs/Fish (Gar/Carp) = N	Both	N	NA
345 S	Y	Conserved	1%	S	X

Evolutionary analysis shows that the HIF-2α bHLH domain is virtually 100% conserved among all species. However, S46 only exists in mammals and bony fish, with remaining species having the non-phosphorylatable residue asparagine. Sequence alignments to HIF-1α also identifies an asparagine residue at this position. Since asparagine is present in all HIF-1α species and non-mammalian species of HIF-2α, it could highlight a HIF-2α specific function that has evolved in mammalian species only. HIF-2α S163 is better evolutionarily

conserved than S46, however is similarly a non-phosphorylatable residue in some species and in HIF-1α.

Interestingly, HIF-2α S72 and S79 are located within a 16 residue region of the bHLH domain which is the only part of this domain that is not highly conserved between HIFα isoforms (Figure 4.10). HIFα isoform sequence alignment shows that HIF-2α phospho-sites S72 and S79 both align to non-phosphorylatable residues of alanine and isoleucine in HIF-1α, respectively. HIF-2α S72 and S79 were both identified as singly phosphorylated peptides and as a doubly phosphorylated peptide (data in appendix 1) and are O_2 independent. Therefore, it may be hypothesised that these sites may have roles in differentiating HIFα isoform functions through DNA or HIF-1β binding mechanisms.

Interestingly, S72 is evolutionarily conserved, while S79 is only present in mammalian species, with non-phosphorylatable residues identified in all other species (Table 4.11). Hence, a previous HTP study providing supportive evidence for S79 phosphorylation makes this site of potential interest. The crystal structure of the bHLH and PAS-A/B domains (PDB 4ZP4, Wu *et al.*, 2015), did not crystallise efficiently within this region, with multiple large fragments missing which encompass the phosphorylation sites S72, S79 and S163 (Figure 4.11).

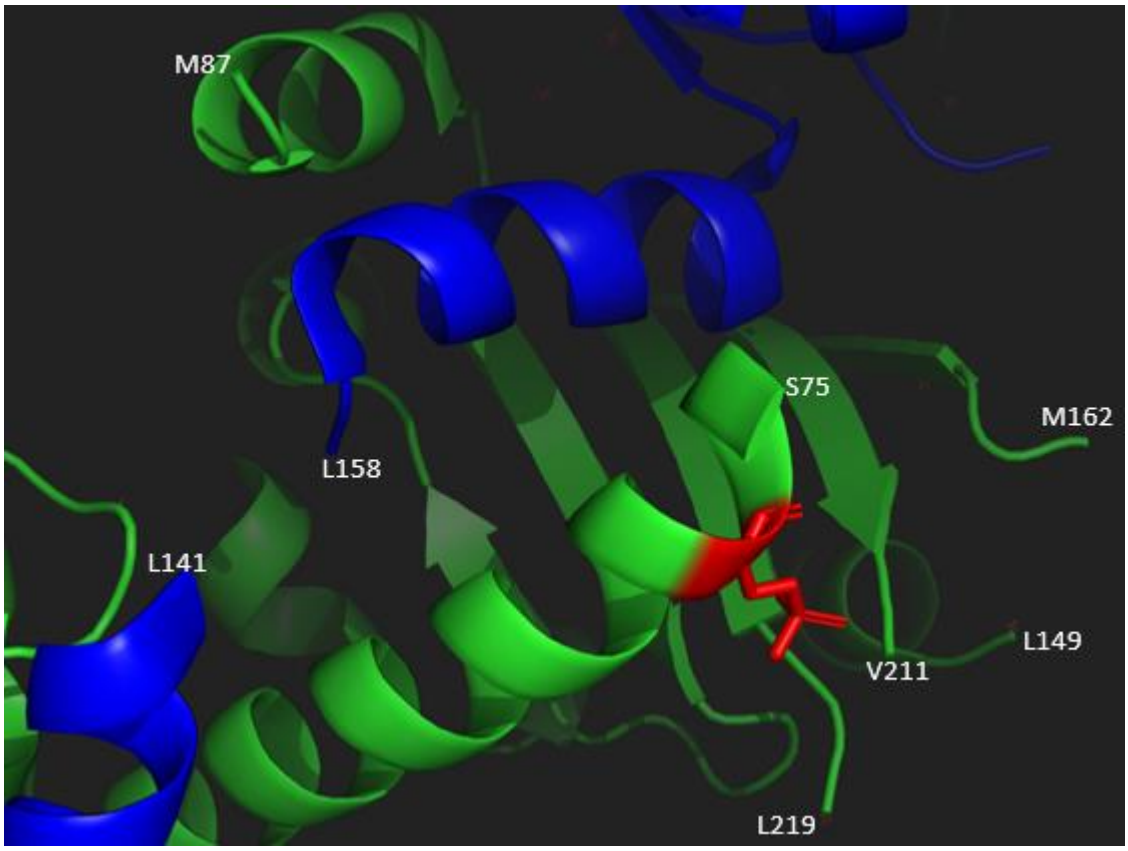


Figure 4.11: Phosphorylated S72 of HIF-2α and potential resultant effects.
Crystal structure (PDB: 4ZP4, Wu *et al.*, 2015), modelled in Pymol and phosphorylated using the PyTMs plugin (Warnecke *et al.*, 2014). Green: HIF-2α, Blue: HIF-1β, Red: phospho-S72. Residues that failed to crystallise are highlight by white text identifying the positions of last crystallised residues.

As Figure 4.11 shows, HIF-2α (green) regions between S75-M87, L149-M162 and V211-L219, and HIF-1β (blue) region between L141-L158, are all missing. It is clear that this region is important for HIF-2α - HIF-1β binding interactions. For HIF-2α S75-M87 to connect, the missing sequence must directly pass over, or under, the HIF-1β α-helix. Interestingly, the lower face of the HIF-1β α-helix contains negatively charged glutamic acid residues while the upper face contains positively charged lysine residues. Thus, depending on how the missing sequence between S75-M87 folds, it is possible that phosphorylation at S79 could result in repulsive or attractive forces respectively. S72 appears to face into a missing region of HIF-2α (V211-L219), which contains both negatively charged glutamic acid residues and positively charged lysine residues. Thus, similar to S79, S72 phosphorylation may result in repulsive or attractive forces respectively. Hence it is possible that S72 and S79 phosphorylation could either result in structural rearrangements, by charge repulsion, and result in the loss of HIF-1β binding, or alternatively stabilise the HIF-2α - HIF-1β dimer, by charge attraction. Without the full crystal structure, it is impossible to determine either possibility.

4.7.7. HIF-1α Negatively charged cluster:

From our data we have identified a highly phosphorylated domain of HIF-1α, between residues 380-385: S380, T383, S384 and S385. Sequence analysis identifies that this phosphorylation cluster contains multiple negatively charged residues as well, possibly creating a negatively charged pocket (378 V-E-Sp-E-D-Tp-Sp-Sp-L 387). However, the available crystal structure (4ZPR) does not cover these regions to further validate this. Mass spectrometry analysis confidently identifies each phospho-site with >10 PSMs and ptmRS scores of >99.5 for each, thus I have confidence in their identifications (Table 4.12). Interestingly, these phosphorylation sites show a degree of mutual exclusivity, both S380 and S384 only appear as singly phosphorylated peptides (mutually exclusive), while T383 and S385 appear as singly- and doubly- phospho-peptides (Table 4.12).

Table 4.12: Identifying peptides and confidence of S380, T383, S384 and S385 HIF-1α phospho-sites. Includes the peptide sequence, Uppercase letter = unmodified, Lowercase letter = phosphorylation site, the number of PSMs that identified a particular peptide, the ptmRS score, peptide confidence score and E-value from the best PSM.

Site	Peptide	# PSMs	PTMrs best	Score	E-value
380 S	TKVEsEDTSSLF	21	S10(1xPhospho): 100	66	7.70E-06
383 T	VESEDtSSLFDKLK	15	T6(1xPhospho): 100	69	2.80E-06
384 S	VESEDTsSLFDKLKK	11	S7(1xPhospho): 99.68	45	8.80E-04
385 S	TKVESEDTSsLFDKL	21	S10(1xPhospho): 100	61	1.50E-05
383 T +385 S	VESEDtSsLFDKLK	12	T6(1xPhospho): 99.76; S8(1xPhospho): 99.76	56	5.00E-05

Table 4.13: Characteristics of HIF-1α phosphorylation sites identified within the negatively charged cluster. Table includes position of PTM, how many studies have identified the respective site with function and reference, evolutionary history where variation are highlighted by single letter amino acid code, the O_2 tension a PTM was identified in and comparisons to HIF-2α sequence. NA = not applicable, X = residue was not phosphorylated.

Site	Novel	Conservation	Observed in: 21%, 1% or Both?	HIF-2α alignment	phosphorylated HIF-2α?
380 S	Y	Birds + Reptiles = L Crocodiles = V Fish (in clades) = A/V/L Gar/Frogs/Sharks =D/E	Both	S	21% + 1% (383)
383 T	Y	Reptiles = S Fish, in sub clades = L/P/Q/S/C	Both	S	X
384 S	Y	All birds/reptiles/sharks/ sturgeon = N Fish, in clades = L/P/S/D/A	1%	N	NA
385 S	Y	Mice/Voles/Hamster = C Sharks = N Fish cyprinids clade = D/E All other Fish = N/C	Both	F	NA

All singly phosphorylated sites within this cluster are O_2 independent, except for S384 which was only identified in 1% O_2 (Table 4.12). The doubly phosphorylated (T383 and

S385) peptide only identified in 21% O_2 (Table 4.12). Overall, this could suggest two possibilities: 1) the different phosphorylation sites may have different roles, or 2) the introduction of a negative charge by phosphorylation is required for functional outcomes, irrespective of location. Evolutionary analysis finds this region to be conserved within phylogenetic vertebrate families, but poorly between families; especially for Bony fish which have large insertions. However, it is interesting that both mutually exclusive phosphorylation sites (S380 and S384) are exclusive to mammalian species, with non-phosphorylatable residues in all other species (Table 4.13). Hence, evolutionary analysis could suggest that possibility (1) is more likely and that S380/S384 may allow for different regulatory pathways that have evolved specifically in mammalian species. Sequence alignment to HIF-2α shows this HIF-1α cluster and surrounding regions are poorly conserved.

4.7.8. Inhibitory domain:

The inhibitory domain was originally termed through deletion construct analysis attempting to identify the main functional domains of HIF-1α (Jiang *et al.*, 1997). Jiang *et al.*, 1997 found that the HIF-1α CTAD was of residues 786-826, while a construct from residues 576-826 had no function; hence the term “inhibitory domain” was coined for residues 576-785. From previously published data (PhosphoSitePlus), the inhibitory domain is the most PTM dense containing domain of HIF-1α; with 27 separate modifications identified from LTP and HTP studies. Here, we show that the inhibitory domain is a heavily phosphorylated domain for HIF-1α (Table 4.14) and the most densely phosphorylated domain of HIF-2α (Table 4.15).

Table 4.14: Characteristics of HIF-1α phosphorylation sites identified within the inhibitory domain.
Table includes position of PTM, how many studies have identified the respective site with function and reference, evolutionary history where variation are highlighted by single letter amino acid code, the O_2 tension at which a PTM was identified in and comparisons to HIF-2α sequence. NA = not applicable, X = residue was not phosphorylated.

Site	Novel	Conservation	Observed in: 21%, 1% or Both?	HIF-2α alignment	phosphorylated HIF-2α?
605 T	Y	Only in Mammals/Birds/Crocodiles Deleted region in all other species	21%	T	X
641 S	3 LTP + 2 HTP MAPK- Nuclear localisation Mylonis et al,. 2006 + 2008 + Han et al., 2016	Birds = P Fish (in clades) = A/V/deleted region	21%	D	NA
643 S	3 LTP + 6 HTP MAPK- Nuclear localisation Mylonis et al,. 2006 + 2008 + Han et al., 2016	Lizards/Snakes = A/I	Both	G	NA
652 T	1 HTP	Ungulates = A/P Fish (in clades) = C/H	21%	C	NA
657 s	1LTP + 2 HTP Plk3- Degradation Xu et al., 2010	Conserved	Both	T	21% + 1% (626)
687 S	1 LTP + 5 HTP CDK5- stabilising Herzog et al., 2016	Frogs/Reptiles = T Birds (in clades) = A/V Fish (in clades) = T/S/A/deleted region	21%	S	21% + 1% (672)
696 S	1 LTP + 1 HTP (696) ATM- downregulation of mTORC1 pathway Cam et al., 2010	Birds/Frogs = N Lizards/Snakes = G Fish (in clades) = N/K	Both	T	X

Table 4.15: Characteristics of HIF-2α phosphorylation sites identified within the inhibitory domain.
Table includes position of PTM, how many studies have identified the respective site with function and reference, evolutionary history where variation are highlighted by single letter amino acid code, the O_2 tension at which a PTM was identified in and comparisons to HIF-1α sequence. NA = not applicable, X = residue was not phosphorylated.

Site	Novel	Conservation	Observed in: 21%, 1% or Both?	HIF-1α alignment	phosphorylated HIF-1α?
581 S	Y	Fish (in clades) = G/deleted region Bats = G Reptiles/Birds = N	Both	T	X
597 T	Y	Whales/Dolphins/Crocodiles= I Birds (in clades) = I/M Remaining Reptiles = M Fish (in clades) = S/K/R	X	E	NA
606 S	Y	Snakes/Lizards = L Whales/Dolphins/Ungulates = F Fish (in clades) = S/A/P	X	I	NA
626 T	Y	Cfish (single clade) = N	Both	S	21% + 1% (657)
672 S	1 HTP	Only Primates = S Ungulates = A Marsupials = P Remaining Mammals = T Frogs/Birds/Reptiles = H Fish (in clades) = F/P/C/deleted region	Both	S	21% only (687)
683 S	Y	Birds = P Fish (single clade) = F/P Marsupials = L	X	R	NA
696 S	Y	Frogs = H Birds/Reptiles/Marsupials = N	Both	N	NA
790 S	Y	Fish (in clades) = G/N/I Snakes/Lizards = N	Both	NA	NA

Interestingly, sequence alignments between HIFα isoforms shows that the HIF-2α inhibitory domain has a large (~60 residue) insertion within the inhibitory domain, and in aligned regions, has very poor sequence homology between HIFα isoforms (Figure 4.12). However,

as Figure 4.12 shows, two sites of phosphorylation between HIFα isoforms do align (green), HIF-1α/HIF-2α numbering: S657/T625 and S687/S672, which could reflect an important shared function. The fact that many of the identified phosphorylation sites within the inhibitory domain are O_2 dependent, for both HIFα isoforms, suggests that this domain could be an important domain for O_2 dependent functions and differentiating HIFα isoform functions.

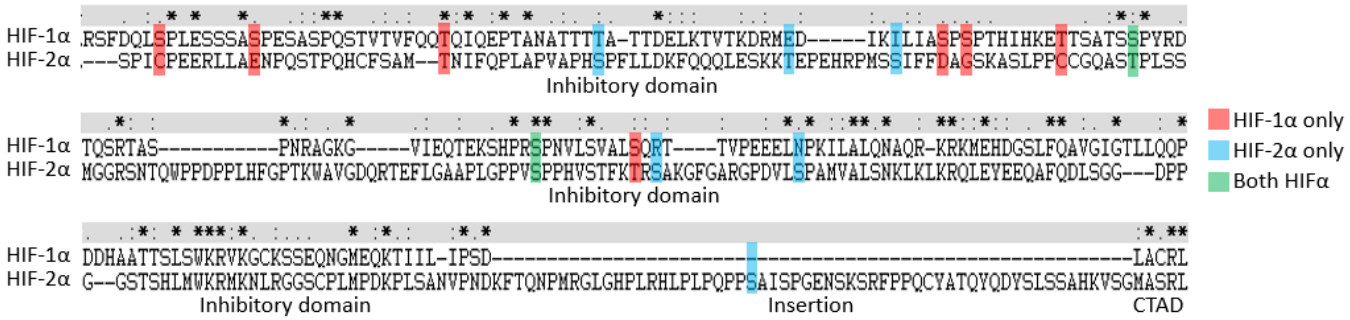


Figure 4.12: Phosphorylated residues within the inhibitory domains of HIF-1α and HIF-2α.
Shows the aligned HIFα sequences from MUSCLE multiple sequence alignment and viewed in clustal X. * indicate identical residues at a position, : identifies a strong conservation mutation, . identifies a weak conservation mutation and blank spaces indicate no conservation mutation (determined by clustal matrix). Red boxes highlight HIF-1α specific phosphorylation, Blue boxes highlight HIF-2α specific phosphorylation and Green boxes highlight phosphorylation of the same aligned residue between HIFα isoforms.

Because many of the identified phosphorylation sites within the inhibitory domain have previously been characterised or are O_2 dependent, they have been discussed in relevant sections (4.7.1 and 4.7.4). Therefore only HIF-2α S696 and S790 are discussed here. HIFα isoform sequence alignments show that HIF-2α S696 is within the poorly conserved region between both isoforms, aligning to a non-phosphorylatable asparagine residue, while S790 is within the large insertion of HIF-2α (Figure 4.12). Interestingly, this entire region is poorly evolutionarily conserved among species and could reflect an evolutionary 'hot-pocket' for evolving novel regulatory mechanisms for HIFα isoforms.

4.8. Search for other PTMs

Since a dual mass spectrometry approach was used, for identifying binding partners and phosphorylation sites (post TiO_2 enrichment), the former will contain every HIFα peptide present and thus all PTMs that have a role in regulating HIFα proteins. As discussed, un-enriched, PTMs are likely to be at a very low level in comparison to unmodified peptides and therefore it is very likely that not every PTM peptide will be detected by our MS analysis. However, some PTMs may be at a high enough level to be detected. Additionally, it is possible that, TiO_2 enriched phospho-peptides may contain additional types of PTM

that would not have been identified previously because they were not considered in the data analysis parameters.

PEAKS is a software tool that can be used to perform an 'open PTM' search for over 300 different PTMs (that are both biological and artefactual in nature), and is known to outperform similar software when filtering to a 1% FDR (Zhang *et al.*, 2012, Han *et al.*, 2011, Ma *et al.*, 2003 & Creasy *et al.*, 2004). Therefore all digested and TiO_2 enriched LC-MS/MS data were additionally analysed using PEAKS, employing the optimised miscleave parameters (described in section 4.3). Searches were performed twice, where cysteine carbamidomethylation was set as a fixed and variable modification, in order to identify potential redox based cysteine PTMs; potentially interesting considering the O_2 sensitive nature of HIFα proteins. Modifications confidently identified at 1% FDR and 5% minimal ion intensity (a measure of localisation score) are depicted in Figure 4.13 and labelled in Table 4.16 to Table 4.19. Interestingly, the open PTM search clearly identifies considerably more PTMs in 21% O_2 than 1% O_2, in agreement with my phosphorylation data. Each PTM type is discussed as a separate modification below. No phosphorylation data is mentioned here as no additional sites were identified by PEAKS.

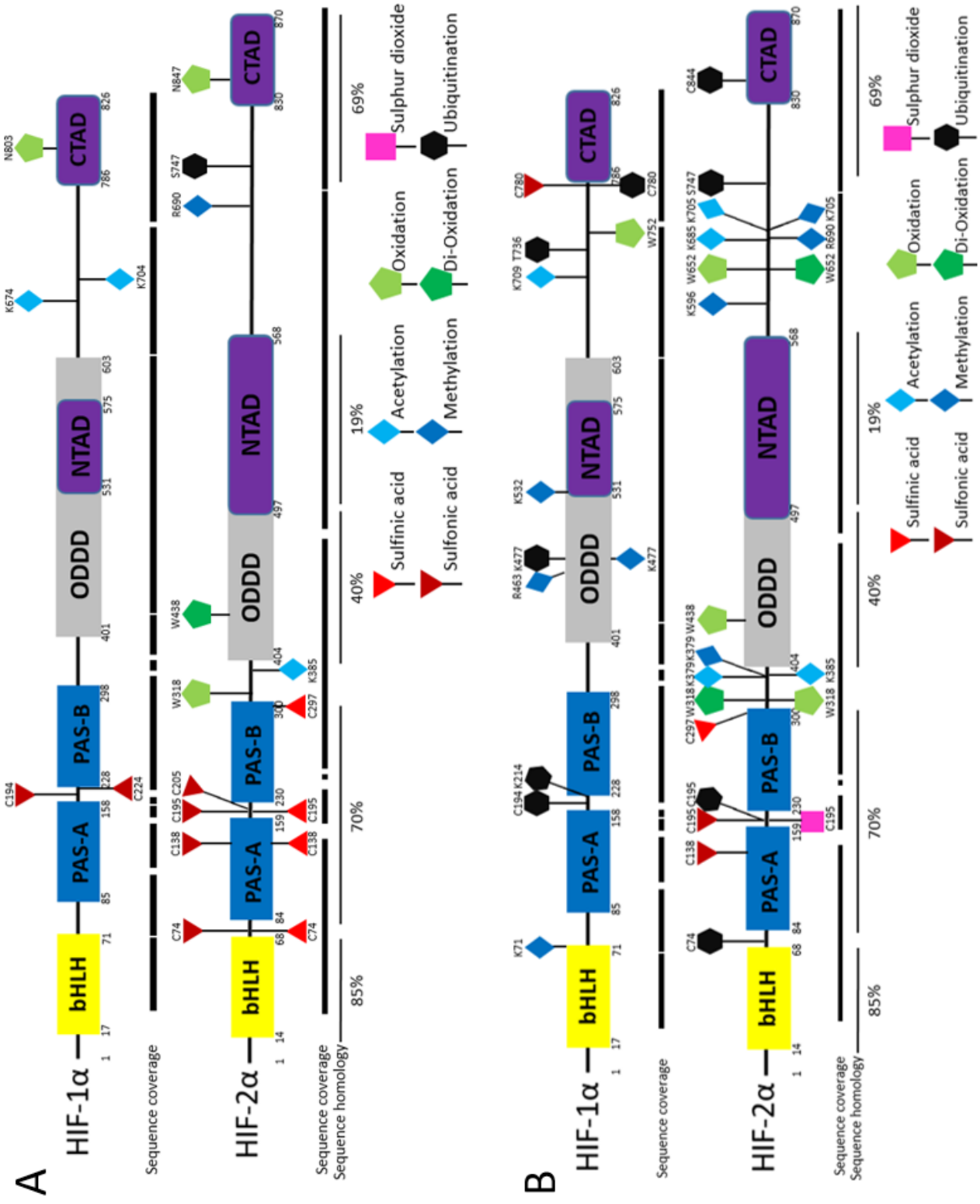


Figure 4.13: Schematic view of the PTM status of HIF-1α and HIF-2α, without phosphorylation. Domains of HIFα proteins are highlighted, with canonical domain numbering. Confidently identified PTMs (1% FDR, 5% minimal ion intensity) are mapped to the protein and coloured dependent on modification. The sequence coverage seen from mass spectrometry analysis (from Figure 4.1) is depicted below the protein schematic along with the sequence homology between isoform domains. A) PTMs identified in 1% O_2, B) PTMs identified in 21% O_2.

4.8.1. Oxidative modifications

Redox sensitive PTMs are well known regulators of both HIFα protein stability, through proline hydroxylation (Jaakkola *et al.*, 2001, Ivan *et al.*, 2001 & Bruick *et al.*, 2001), and transcriptional activity, through asparagine hydroxylation (Lando *et al.*, 2002, Lando *et al.*,

2002 & Mahon *et al.*, 2001). Considering that proline hydroxylation results in the rapid degradation of HIFα, it is not surprising that the canonical sites of proline hydroxylation were not identified (HIF-1α: P405 and P531 and HIF-2α: P405 and P531 (Jaakkola *et al.*, 2001, Ivan *et al.*, 2001 & Masson *et al.*, 2001). Therefore it was incredibly interesting to identify novel HIF-2α proline hydroxylation sites (P137 and P206) at a 5% FDR (removed at 1% (data not shown)), that are specific to 1% O_2 conditions. Therefore it is possible these sites are real but poorly fragment or, equally valid, they are matching by chance; hence should be validated before investigating further. These novel hydroxylation sites do not contain the LAP motifs required for the EGLN PHD function, either suggesting the action of different proline hydroxylases, discussed further in the known binding partners section (4.13.1.4), or that these sites may be false positive by the reduced stringency filter.

It is known that FIH-1 maintains enzymatic activity in O_2 tensions as low as 0.5% O_2 (Lando *et al.*, 2002), thus it was not surprising to identify N803 and N847 hydroxylation (HIF-1α and HIF-2α respectively) in 1% O_2. It is likely that protein molecules that have been proline hydroxylated in 21% O_2 are also aspargine hydroxylated, thus explaining the lack of identification for aspargine hydroxylation at 21% (due to it being rapidly degraded).

Oxidation is also possible on the following amino acids: methionine, phenylalanine, tryptophan, histidine and cysteine, by either enzymatic reactions or interaction with Reactive oxygen species (ROS) (Sharma *et al.*, 2010). Although methionine oxidation can have biological roles in disease and ageing (Mohsenin *et al.*, 1989 & Hoshi *et al.*, 2001), it is generally accepted that standard MS sample preparation techniques can artefactually increase methionine oxidation rates (Ghesquière *et al.*, 2014), thus it was disregarded as a PTM here due to the lack of adequate controls. Neither phenylalanine or histidine oxidation was identified here.

Although tryptophan oxidation is also potentially artifactual, it requires much stronger oxidising conditions and is seldom seen due to proteomics based sample preparation. During cellular lysis, tryptophan oxidation can occur artifactually but is generally identified in a very specific manner that is enriched to mitochondrial proteins, and is at a much reduced frequency than methionine oxidiation (Perdivara *et al.*, 2010 & Taylor *et al.*, 2003). Additionally, there are various pathologies associated with the dysregulation of these enzymes (Ehrenshaft *et al.*, 2015).

Similarly, cysteine residues can undergo multiple, reversible oxidation states by reaction with both reactive oxygen and nitrogen Species (ROS/RNS) at the free thiol group, if in a non-disulfide bonded state (reviewed by Chung *et al.*, 2013). Under severe oxidative stress, cysteine residues can be irreversibly oxidised into 2+ and 3+ oxidative states: sulfinic acid and sulfonic acid respectively. Published studies have found an important regulatory role for irreversibly oxidised cysteine residues (Vivancos *et al.*, 2005, Blackinton *et al.*, 2009 & Fujiwara *et al.*, 2007). Considering ROS/RNS accumulate in hypoxia (Chandel *et al.*, 1998 & Bell *et al.*, 2007), it is likely that hypoxia could ellict an oxidative based PTM response in HIF signalling. Thus both tryptophan and cysteine oxidative PTMs were considered as true PTMs rather than artefactual, and are discussed below. All oxidiative PTMs discussed are located within Table 4.16.

Table 4.16: Characteristics of Oxidative based PTMs of HIF-1α and HIF-2α.
Table includes position and type of PTM, whether the PTM is novel or previously identified (with references), the O_2 tension a PTM was observed in and whether the site is observed as a different PTM site in published data or in data presented here. Table also includes the evolutionary conservation, sequence alignment to the other HIFα isoform and whether the other isoform aligned residue was identified as a site of PTM. NA = not applicable, X = residue was not modified.

HIF-1α								
Site	Modification	Novel	Observed in: 21%, 1% or Both?	Other modification at site: published	Other modification at site: Identified here (Observed in)	Conservation	HIF-2α alignment	HIF-2α modification? (Observed in)
194 C	Sulfonic Acid	Y	1%	X	Ubiquitination (21%)	Conserved	C (195)	Sulfonic acid (Both) Sulfinic acid (1%) Ubiquitination (21%)
224 C	Sulfonic Acid	Y	1%	X	X	Conserved	C	X
W752	Oxidation	Y	21%	X	X	Fish = G/deleted region	W	X
C780	Sulfonic Acid	Y	21%	X	X	Only in mammals All other species = S	S	X

HIF-2α								
Site	Modification	Novel	Observed in: 21%, 1% or Both?	Other modification at site: published	Other modification at site: Identified here (Observed in)	Conservation	HIF-1α alignment (position)	HIF-1α modification? (Observed in)
C74	Sulfinic acid	Y	1%	X	Ubiquitination (21%)	Fish (in clades) = C/R/Y/W	NA	NA
C74	sulfonic acid	Y	1%	X	Ubiquitination (21%)	Fish (in clades) = C/R/Y/W	NA	NA
C138	sulfonic acid	Y	Both	X	X	Conserved	C	X
C138	Sulfinic acid	Y	1%	X	X	Conserved	C	X
C195	sulfonic acid	Y	Both	X	Ubiquitination (21%) Sulfur dioxide (21%) Sulfinic acid (1%)	Conserved	C (194)	Sulfonic acid (1% only) Ubiquitination (21% only)
C195	Sulfinic acid	Y	1%	X	Ubiquitination (21%) Sulfur dioxide (21%) Sulfonic acid (both)	Conserved	C (194)	Sulfonic acid (1% only) Ubiquitination (21% only)
C195	Sulfur Dioxide	Y	21%	X	Ubiquitination (21%) Sulfinic acid (1%) Sulfonic acid (both)	Conserved	C (194)	Sulfonic acid (1% only) Ubiquitination (21% only)
C205	sulfonic acid	Y	1%	X	X	Fish (single clade) = R	NA	NA
C297	Sulfinic acid	Y	Both	X	X	Conserved	F	NA
W318	Oxidation	Y	Both	X	Dihydroxy (21%)	Conserved	W	X
W318	Di-oxidation	Y	21%	X	Oxidation (both)	Conserved	W	X
W438	Oxidation	Y	21%	X	Di-oxidation (1%)	Conserved	T (458)	Phosphorylation (1% + 21%)
W438	Di-oxidation	Y	1%	X	Oxidation (21%)	Conserved	T (458)	Phosphorylation (1% + 21%)
W652	Oxidation	Y	21%	X	Di-oxidation (21%)	Fish (in clades) = A/I	A	NA
W652	Di-oxidation	Y	21%	X	Oxidation (21%)	Fish (in clades) = A/I	A	NA
N847	Oxidation	2 LTP- Inhibition Lando et al., 2002 & 2002	1%	X	X	Conserved	N (803)	Oxidation (1% only)

4.8.1.1. Tryptophan oxidation

HIF-1α has a single site of tryptophan oxidation (W752), specific to 21% O_2. W752 is within a highly conserved region and is also a highly conserved site, with only a single clade of Bony Fish having a glycine residue. The functional importance of this variation is hard to predict; although glycine can not undergo an oxidative modification, the size and hydrophobicity differences between tryptophan and glycine are likely to have a larger structural effect.

HIF-2α has 3 tryptophan oxidation sites (W318, W438 and W652) that are identified in both mono-oxidised and di-oxidised states. All 3 tryptophan sites are highly conserved from evolutionary analysis. Interestingly, the oxygenation state of both W318 and W438 are O_2 dependent; W318 is mono-oxidised in both O_2 tensions but only di-oxidised in 21% O_2, whilst W438 is mono-oxidised only in 1% O_2 and di-oxidised only in 21% O_2. Similarly, W652 is both mono- and di- oxidised only in 21% O_2, suggesting that the sites are all independently regulated. Of particular interest, W438 resides within the ODDD, which is a highly conserved domain between isoforms. Therefore it is surprising to identify HIF-2α W438 aligns to the phosphorylated HIF-1α T458 residue, with a single HTP study providing supportive evidence of its existence. These residues and PTMs have highly differing characteristics, suggesting possible functional differences.

4.8.1.2. Cysteine oxidation

Considering the accumilation of ROS/RNS in hypoxia (Chandel *et al.*, 1998 & Bell *et al.*, 2007), it is not surprising to identify that the number of both sulfinic acid and sulfonic acid irreversible cysteine oxidative PTMs increases by hypoxic incubation, with 7 cysteines modified in 1% O_2 compared to 4 in 21% O_2. As the majority of cysteine oxidation sites identified here are found in both sulfinic and sulfonic acid forms, it is unknown whether they may have separate functions or function interchangably. All oxidised cysteine residues detected at 21% O_2 were also detected at 1% O_2, except for C780 in HIF-1α, suggesting that some sites are more readily oxidized than others. Besides C780, all oxidised cysteines are localised in the DNA binding (bHLH) and HIF-1β binding (PAS-A/B) domains. Interestingly, although the bHLH and PAS-A/B domains share the most sequence homology between HIFα isoforms, HIF-2α is much more extensively oxidised than HIF-1α and in a 1% O_2 only dependent manner. Considering the near 100% conservation of all cysteine residues within these domains, this could suggest an important function that may differentiate HIFα isoforms.

4.8.2. Ubiquitination

Ubiquitination is a 8.5 kDa protein tag which is typically covalently attached to lysine residues, through a 3-stage activation process. Mono-ubiquitination and poly-ubiquitination chains can occur through multiple linkage strategies of monomers. Different linkage strategies have been shown to have different cellular outcomes. However, poly-ubiquitination generally results in tagged-protein degradation through the proteasome (reviewed by Komander *et al.*, 2012). Traditional shotgun- based proteomics (protein digestion) result in the concomitant digestion of ubiquitin, thus consequently it is impossible to distinguish between mono- or poly- ubiquitination. However, ubiqutination at a specific site is detectable by a 'GG-tag' that remains attached to the lysine residue post digestion (reviewed by Kirkpatrick *et al.*, 2005).

Ubiquitination is an important HIFα PTM, having roles in regulating protein stability, and potentially switching to a HIF-2α dependent signalling state in prolonged hypoxic conditions (Koh *et al.*, 2008, Koh *et al.*, 2011, Koh *et al.*, 2014, Paltoglou *et al.*, 2007, Tanimoto *et al.*, 2000 & Maxwell *et al.*, 1999, reviewed by Schober *et al.*, 2016). Recently, ubiquitination has been shown to be involved in a non-canonical activation pathway that results in the ubiquitination of cysteine, serine, threonine and tyrosine residues through thiol-ester and hydroxy-ester bonding respectively. These non-canonical ubiquitination generally result in protein degradation (Reviewed by: McDowell *et al.*, 2013). All ubiquitination sites identified here are discussed below, and are listed in Table 4.17.

Table 4.17: Characteristics Ubiquitination PTMs of HIF-1α and HIF-2α

Table includes position and type of PTM, whether the PTM is novel or previously identified (with references), the O_2 tension a PTM was observed in and whether the site is observed as a different PTM site in published data or in data presented here. Table also includes the evolutionary conservation, sequence alignment to the other HIFα isoform and whether the other isoform aligned residue was identified as a site of PTM. NA = not applicable, X = residue was not modified.

HIF-1α								
Site	Modification	Novel	Observed in: 21%, 1% or Both?	Other modification at site: published	Other modification at site: Identified here (Observed in)	Conservation	HIF-2α alignment	HIF-2α modification? (Observed in)
C194	Ubiquitination	Y	21%	X	Sulfonic acid	Conserved	C (195)	Sulfonic acid (Both) Sulfinic acid (1% only) Ubiquitination (21% only)
K214	Ubiquitination	Y	21%	X	X	Fish = E	E	NA
K477	Ubiquitination	1 HTP	21%	1 HTP- Ubiquitination 3 LTP- Sumoylation, Stabilising / inhibition Bae et al., 2004, Berta et al., 2007 & Cheng et al., 2007	Methylation (21%)	Fish = deleted region	NA	NA
T736	Ubiquitination	Y	21%	X	X	Birds = S Fish (in clades) = A/S/T	NA	NA
C780	Ubiquitination	Y	21%	X	X	Only in mammals All other species = S	S	X

HIF-2α								
Site	Modification	Novel	Observed in: 21%, 1% or Both?	Other modification at site: published	Other modification at site: Identified here (Observed in)	Conservation	HIF-1α alignment	HIF-1α modification? (Observed in)
C74	ubiquitination	Y	21%	X	Sulfinic acid (1%) Sulfonic acid (1%)	Fish (in clades) = C/R/Y/W	NA	NA
C195	ubiquitination	Y	21%	X	Sulfonic acid (Both) Sulfur dioxide (21%) Sulfinic acid (1%)	Conserved	C (194)	Sulfonic acid (1% only) Ubiquitination (21% only)
S747	ubiquitination	Y	Both	X	X	Only in primates + Fish Ungulates/Rodents = T Remaining Mammals/Reptiles/Birds = N	E	NA
C844	ubiquitination	Y	21%	X	X	Conserved	C	NA

Two sites of lysine ubiquitination were detected in total, both in HIF-1α at K214 and K477. Considering the rapid degradatory nature of ubiquitination, it is possible that this PTM is not easily identifiable without proteasomal inhibition. Interestingly, K214 is highly evolutionarily conserved, except in Bony fish where a negatively charged glutamic acid residue aligns. Not only does a glutamic acid residue remove the ability for ubiquitination, lysine and glutamic acid are oppositely charged residues, thus may reflect a dual role of K214 where there is a charge importance. K477 ubiquitination has a HTP study for supportive evidence, and was also identified as a site of methylation in this study. K477 has also been identified in multiple LTP as a site of SUMOylation, with conflicting results in terms of protein stability and transactivation roles (Cheng *et al.*, 2007, Bae *et al.*, 2004 & Berta *et al.*, 2007).

Taking into account the degradatory nature of non-canonical ubiquitination sites, it was not surprising to detect non-canonical ubiquitination predominantly in 21% O_2, with only a single site identified in 1% O_2 (HIF-2α at S747). We identified two hydroxy-ester ubiquitination sites: HIF-1α at T736 and HIF-2α at S747, these do not align between HIFα isoforms and have fairly poor evolutionary conservation.

The remaining ubiquitination sites are all cysteine linked. Commonly used reduction strategies for mass spectrometry sample preparation have been shown to result in the removal of cysteine ubiquitination (Williams *et al.*, 2007), which could explain the lack of previous identification within the proteome. It is observed that every cysteine ubiquitination site is neighboured by a serine/threonine residue, which could have resulted in mis-annotation of PTM localisation if the peptide was poorly fragmentated. However, setting the data analysis for fixed cysteine carbamidomethylation (chemically induced by sample preparation), in an attempt to force localisation onto the neighbouring serine/threonine residue, resulted in the PSM no longer being identified. This suggests that either: cysteine is infact ubiquitinated and is more stable than initially thought, or serine/threonine ubiquitination blocks the ability to chemically modify the cysteine residues. Interestingly, many of the cysteine ubiquitination sites detected are also sites of oxidative PTM in 1% O_2 (HIF-1α: C194, HIF-2α: C74 and C195), which could suggest a bi-functional role for these cysteine residues. For example, 21% O_2 ubiquitination promotes degradation whereas oxidative PTMs in 1% O_2 could promote accumulation.

4.8.3. Acetylation

Lysine acetylation involves the addition of an acetyl group to the positively charged amino group of lysine residues, resulting in charge neutralisation. Both HIF-1α and HIF-2α have been shown to be regulated by acetylation, having both protein stability and transcriptional function roles (Lim *et al.*, 2010 & Dioum *et al.*, 2009). Six separate acetylation sites were identified in this dataset, HIF-1α: K674 and K709 and HIF-2α: K379, K385, K685 and K705, with two of these being novel: HIF-2α K379 and K705. All data listed within Table 4.18.

Table 4.18: Characteristics Acetylation PTMs of HIF-1α and HIF-2α

Table includes Table includes position and type of PTM, whether the PTM is novel or previously identified (with references), the O_2 tension a PTM was observed in and whether the site is observed as a different PTM site in published data or in data presented here. Table also includes the evolutionary conservation, sequence alignment to the other HIFα isoform and whether the other isoform aligned residue was identified as a site of PTM.

HIF-1α								
Site	Modification	Novel	Observed in: 21%, 1% or Both?	Other modification at site: published	Other modification at site: Identified here (Observed in)	Conservation	HIF-2α alignment	HIF-2α modification? (Observed in)
K674	Acetylation	2 LTP Promotes CBP/p300 binding Lim et al., 2010 & Geng et al., 2012	1%	2HTP- Ubiquitination	X	Fish (in clades) = T/P	V	NA
K709	Acetylation	2 LTP Protein stabilising Geng et al., 2012 & Seo et al., 2015	Both	4HTP- Ubiquitination	X	Fish (single clade) = R	A	NA
HIF-2α								
Site	Modification	Novel	Observed in: 21%, 1% or Both?	Other modification at site: published	Other modification at site: Identified here (Observed in)	Conservation	HIF-1α alignment	HIF-1α modification? (Observed in)
K379	Acetylation	Y	21%	X	Methylation (21%)	Rodents/Bats = E/D Sharks/Reptiles/Birds = G Fish (in clades) = A/V	NA	NA
K385	Acetylation	1 LTP- Inhibition Dioum et al., 2009	Both	2 HTP- ubiqutination	X	Fish (in clades) = A/P	D	X
K685	Acetylation	1 LTP- Inhibition Dioum et al., 2009	21%	1 HTP- ubiqutination	X	Only in mammals Birds = D Reptiles/Fish = E	NA	NA
K705	Acetylation	Y	21%	X	Methylation (21%)	Mammals only Fish (in clades) = N/S Birds/Reptiles = S	A	NA

HIF-1α K674 acetylation, by PCAF, was shown to promote transcriptional function by enhancing p300/CBP-HIF-1α interactions (Lim *et al.*, 2010). Interestingly, HIF-1α K674 has been identified by two separate published HTP studies as a site of ubiquitination. Thus HIF-

1α K674 might have a dual role, influenced by PTMs, in a mechanism where acetylation prevents protein degradation by blocking ubiquitination at this site. Suporting this theory, HIF-1α K674 is virtually 100% evolutionarily conserved and acetylation is only identified in 1% O_2. Hypoxia dependent acetylation could thus prevent normoxic ubiquitination and degradation, an avenue not explored in the previous publication.

HIF-1α K709 acetylation was shown to directly promote protein stability (Geng *et al.*, 2012). Interestingly, 4 separate HTP studies have identified K709 as a site of ubiquitination, therefore it can be argued that the 2 PTMs may compete for HIF-1α K709 modification to regulate protein stability, as discussed for HIF-1α K674. HIF-1α K709 resides within an evolutionarily highly conserved region of ~20 residues, however in all Bony fish there is an arginine residue. This variation corresponds to what is used for biochemical mutational analysis for acetylation, where lysine residues are mutated to arginine for acetylation-null investigations because it maintains the positive charge without the ability to be acetylated (Gorsky *et al.*, 2016).

Dioum *et al.*, 2009 showed that HIF-2α was acetylated on K385, K685 (both identified in this study) and K741, and that a triple mutant reduced transcriptional function without affecting protein stability. However the effect of individual sites was not explored. Both HIF-2α K385 and K685 have been identified as sites of ubiquitination by HTP studies, which could suggest a protein stability role, as discussed for HIF-1α K674, which was not observed by Dioum *et al.*, 2009. However, it is possible that the use of different cell lines and treatments before analysis could impact HIF-2α stability, with different regulatory profiles for different cell lines (Bracken *et al.*, 2006). Interestingly, HIF-2α K685 is only present in mammalian species, with all other species having variation to aspartic acid or glutamic acid mutations, which are oppositely charged to lysine and may therefore suggest an importance of charge.

Both HIF-2α K379 and K705 acetylation sites are novel and are specific to HIF-2α, aligning respectively to a missing region and an alanine residue within HIF-1α. Interestingly, evolutionary analysis identifies that HIF-2α K379 has specifically evolved in primate and ungulate species only, with remaining mammals having a negatively charged glutamic acid residue. All other vertebrate species investigated have a charge neutral amino acid. Overall this could suggest a recently developed functional trait that is charge dependent. Similarly, HIF-2α K705 is only present in mammalian species, with all other species having a serine residue, which hypothetically can undergo phosphorylation to introduce a negative charge.

4.8.4. Methylation

Lysine and arginine methylation have classically been associated with histone modification and gene expression regulation. Methylation has also been identified as an important PTM in signal transduction pathways, including the EGFR – ERK pathway (Hsu *et al.*, 2011), and for regulation of protein function such as p53, where mono-methylation inhibits function and di-methylation activates function through co-factor binding (Huang *et al.*, 2007), reviewed by Biggar *et al.*, 2014. Infact, HIF-1α and HIF-2α have been shown to be methylated at K32 and K29, respectively, but has conflicting reported functional effects with reports showing inhibition of transcriptional function without affecting their stability (Liu *et al.*, 2015) and methylation induced degradation (Kim *et al.*, 2016). Additionally, HIF-1α methylation at K391 was found to promote PHD-VHL dependent degradation (Lee *et al.*, 2017), showing the potential importance of this modification in hypoxia signalling. None of these sites were identified in this study, however a total of 4 novel HIF-1α and HIF-2α methylation sites were identified, listed in Table 4.19.

Table 4.19: Characteristics Methylation PTMs of HIF-1α and HIF-2α

Table includes position and type of PTM, whether the PTM is novel or previously identified (with references), the O_2 tension at which a PTM was observed in and whether the site is observed as a different PTM site in published data or in data presented here. Table also includes the evolutionary conservation, sequence alignment to the other HIFα isoform and whether the other isoform aligned residue was identified as a site of PTM. NA = not applicable, X = residue not modified

HIF-1α								
Site	Modification	Novel	Observed in: 21%, 1% or Both?	Other modification at site: published	Other modification at site: Identified here (Observed in)	Conservation	HIF-2α alignment	HIF-2α modification? (Observed in)
K71	Methylation	Y	21%	2 HTP- Ubiquitination	X	Conserved	K	X
R463	Methylation	Y	21%	X	X	Snakes/Sharks = C Fish (in clades) = T/S/D/N/H	R	X
K477	Methylation	Y	21%	1 HTP- Ubiquitination 3 LTP- Sumoylation, Stabilising / Inhibition Bae et al., 2004, Berta et al., 2007 & Cheng et al., 2007	Ubiquitination (21%)	Fish = Deleted region	NA	NA
K532	Methylation	Y	21%	2 LTP- Acetylation, Degradation Jeong et al., 2002 & Lee et al., 2010 2 LTP- no modification/affect Murray-Rust et al., 2006 & Arnesen et al., 2005 2 LTP- Ubiquitination, Degradation Tanimoto et al., 2000 & Paltoglou et al., 2007	X	Conserved	K	X

HIF-2α								
Site	Modification	Novel	Observed in: 21%, 1% or Both?	Other modification at site: published	Other modification at site: Identified here (Observed in)	Conservation	HIF-1α alignment	HIF-1α modification? (Observed in)
K379	Methylation	Y	21%	X	Acetylation (21%)	Rodents/Bats = E/D Sharks/Reptiles/Birds = G Fish (in clades) = A/V	NA	NA
K596	Methylation	Y	21%	X	X	Reptiles/Birds = N Birds (single clade) = K Fish (Carp) = D	M	NA
R690	Methylation	Y	Both	X	X	Only Primates/Marsupials/Rodents = R Remaining mammals = Q Fish (in clades) = V/N/A	V	NA
K705	Methylation	Y	21%	X	Acetylation (21%)	Mammals only Fish (in clades) = N/S Birds/Reptiles = S	A	NA

Our data show that all HIF-1α methylation sites are novel and are present at 21% O_2 only, suggesting a potential O_2 dependent regulatory role. Interestingly, all 3 HIF-1α lysine methylation sites (K71, K477 and K532) have been identified as sites for other PTMs, suggesting a potential dual function (as discussed previously in the acetylation section (4.8.3)). HIF-1α K71 was detected as a site of ubiquitination in two separate HTP studies while both K477 and K532 have been extensively investigated in LTP studies as SUMOylation and acetylation/ubiquitination sites. K477 was also identified as a site of ubiquitination in our data, therefore was discussed previsouly in section 4.8.2.

LTP studies have shown HIF-1α K532 is ubiquitinated, post PHD Proline hydroxylation, resulting in degradation (Tanimoto *et al.*, 2000 & Paltoglou *et al.*, 2007). Additionally, HIF-1α K532 was shown to be a site of acetylation, by ARD-1 (Arrest Defective protein -1), triggering degradation (Jeong *et al.*, 2002 & Lee *et al.*, 2010). However, the role of HIF-1α K532 has conflicting evidence, with studies showing that acetylation does not occur (Murray-Rust *et al.*, 2006 & Arnesen *et al.*, 2005). Considering the range of reported PTMs (ubiquitination, acetylation, and methylation from this study), combined with 100% conservation of HIF-1α K532, it is highly likely that K532 has essential regulatory roles which are not yet fully understood. The final HIF-1α methylation site is a novel arginine residue (R463), which is poorly conserved.

All HIF-2α methylation sites detected in this study (K379, K596, R690 and K705) are novel sites. Similarly to HIF-1α, lysine methylation sites were only present at 21% O_2, R690 methylation was identified in both O_2 tensions. Evolution analysis shows that HIF-2α R690 only exists in primates, marsupials and rodents, with remaining mammalian species having a Glutamine variation. This variation is used as an acetylation-mimetic mutation to maintain side chain structure but removing the charge (Gorsky *et al.*, 2016). Thus, it is interesting as methylation does not result in charge neutralisation, like acetylation. Interestingly, both K379 and K705 are sites of acetlyation, and were therefore discussed above.

4.8.5. SUMOylation

Although SUMOylation has been shown as an important PTM that regulates HIF-1α (Bae *et al.*, 2004, Berta *et al.*, 2007 and Cheng *et al.*, 2007) and HIF-2α (van Hagen *et al.*, 2010) function, no sites were detected within this study. It is possible that the cell line and O_2 treatments used did not result in SUMOylation. However, a more plausible explanation is the difficulty in identifying SUMOylation PTMs using MS without targeted approaches

(Hendriks *et al.*, 2016, Hendriks *et al.*, 2018 & Tammsalu *et al.*, 2015). SUMO is a protein tag, similar to ubiqutin, however it is poorly digested with trypsin and results in a 30 amino acid tag attached to the peptide; too large for standard proteomics analysis.

4.9. COSMIC database

With the advancements of genome sequencing technology, high quality genome sequences can be produced and computationally aligned/annotated relatively quickly. This has lead to a wealth of data where different cancer samples have been sequenced to identify mutational patterns in genes of interest, that may result in amino acid mutation (missense mutation). The COSMIC database (Catalogue Of Somatic Mutations In Cancer) is the largest repository for such data with over a million partial and >28,000 full genome tumour sequence entries (Bamford *et al.*, 2004 & Forbes *et al.*, 2016). The COSMIC database was analysed to identify mutations in the HIF-1α and HIF-2α (EPAS-1) genes that result in missense mutations, with the aim to identify potential hotspots of mutation that align to PTMs identified in this study; potentially pointing to important functional sites worthy of prioritising for further investigation.

A total of 50,243 and 48,762 unique tumour samples had been sequenced containing the HIF-1α and HIF-2α genes respectively. Considering the importantance of hypoxia, and active HIF, in tumour survival and progression, this number of tumours lead to a surprisinly low total mutation count of 261 and 357 for HIF-1α and HIF-2α respectively (~0.5% of all tumours). For comparison p53 (TP53 gene) has been sequenced in 167,617 unique tumours and was found mutated in 41,588 tumours (~25% of all samples), equating to a mutation frequency 50 fold greater in p53 than either HIFα genes.

Since the specific aim of this analysis was to identify site-specific mutational hotspots for potential PTM effects, only missense mutations (amino acid substitution) were investigated, hence synonymous, nonsense, frameshift insertions/deletion and in-frame insertion/deletion mutations were removed from the dataset. This lowered the mutation count to 190 for HIF-1α and 270 for HIF-2α. The COSMIC database stores mutations of the same site to different DNA bases as separate records, thus a highly mutated site to no specific residue would not be identified as significantly mutated. Therefore all mutations for

a single site were cumulated together and plotted against amino acid position for respective protein (Figure 4.14).

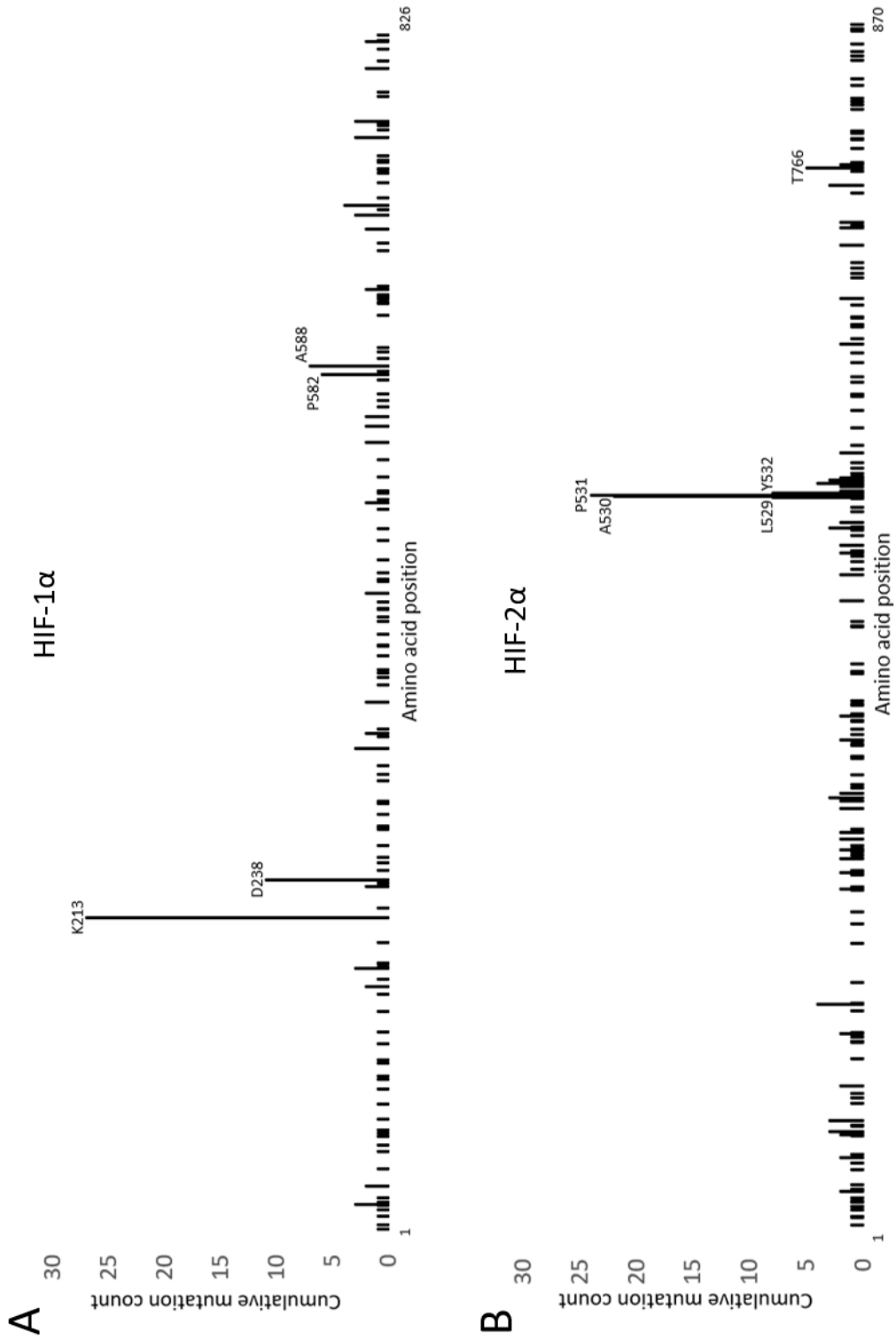


Figure 4.14: Missense mutational hotspot mapping of HIF-1α and HIF-2α.
Data collected from COSMIC datasbase from 50,243 and 48,762 unique tumour samples sequenced containing the HIF-1α and HIF-2α genes respectively. Cumulative missense mutation count is plotted against protein length and sites with a count of 5+ are labelled. A) HIF-1α, total of 190 mutations. B) HIF-2α, total of 270 mutations.

Figure 4.14 shows that both HIF-1α and HIF-2α have a mutational hotspot with ~15% of all HIF-1α mutations located at K213 + D238, and ~23% of all HIF-2α mutations between positions 529-532. Interestingly, none of the PTM sites identified in our study, or in published studies, are identified as mutational hotspots in cancer samples. The two enriched HIF-1α mutation sites (K213 and D238) are relatively close together and oppositely charged, therefore these sites may interact with each other to form an ion pair; hence mutation to either can give the same phenotypic effect. HIF-1α K213 and D238 were modelled using the crystal structure (PDB: 4ZPR, Wu *et al.*, 2015, Figure 4.15). Figure 4.15 shows that K213 is situated within a missing region (199-218), however D238 is spatially distant from C218 (the closest amino acid in the crystal structure to K213). An alternatively theory that Figure 4.15 could suggest is that HIF-1α K213 and D238 are involved with HIF-1β binding regulation, both being in close proximity to a HIF-1β α-helix. Interestingly, the HIF-2α mutational hotspot localises to the prolyl-hydroxylation site (P531) and consensus 'LAP' motif seen in all known proline hydroxylation sites (Figure 4.14), suggesting that the loss of degradation for HIF-2α is an important cancer survival mechanism.

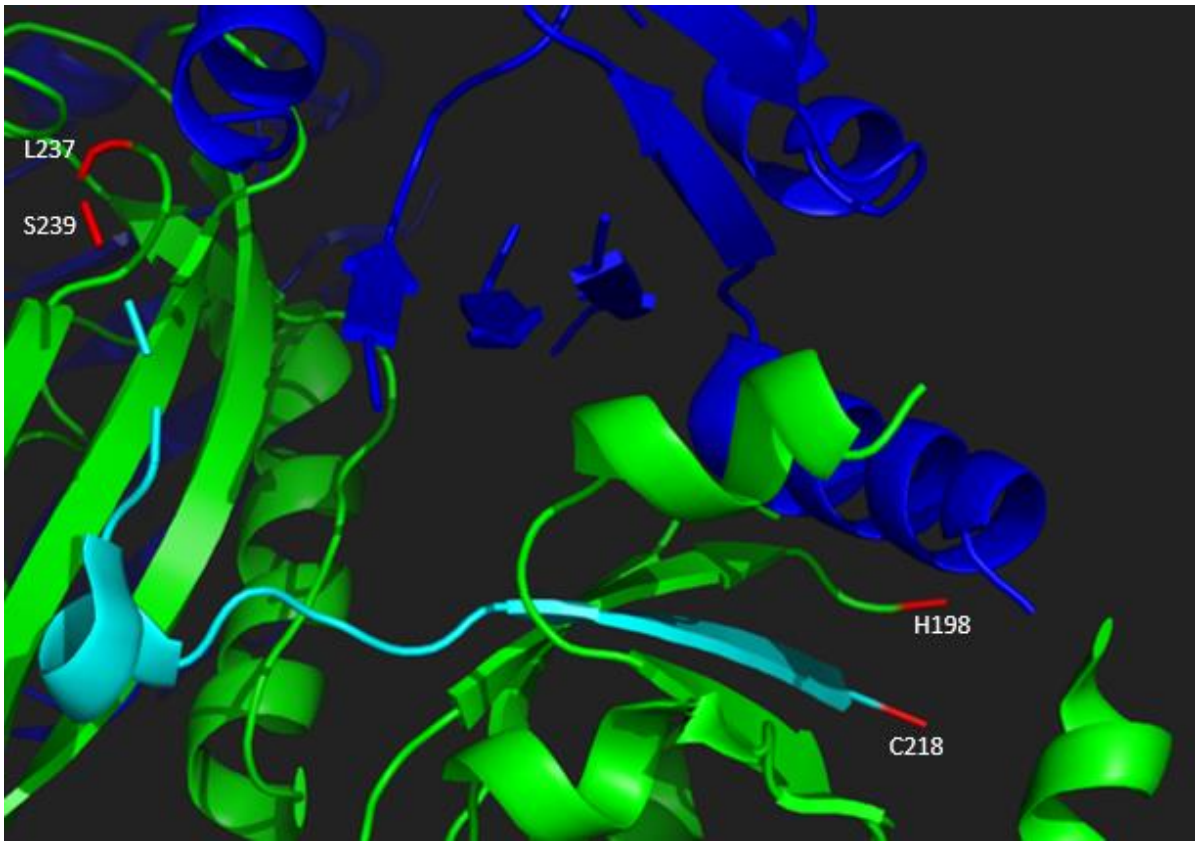


Figure 4.15: Crystal structure mapping to investigate K213 and D238 cancer mutations of HIF-1α.
PDB: 4zpr, green = HIF-1α, blue = HIF-1β, red = surrounding amino acids of the deletion encompassing K213 and D238, cyan = amino acids connecting K213 and D238. Numbering following crystal structure numbers, from *Mus musculus*.

4.10. Binding partners

MS analysis of the eluted material, before TiO_2 phospho-peptide enrichment, allowed for total protein analysis to identify Co-IP'd HIF-1α and HIF-2α binding partners. A confidently identified peptide count per protein filter was not applied, however identified proteins were only kept in the dataset if identified in both replicates. Finally, using the HA-Clover only control we created background subtracted lists by the removal of non-specific binding partners. All binding partners from each IP and replicate filtering are listed in Appendix 2.

4.10.1. Replicate variability

Although significant optimisation was performed, including lysis/wash buffer and length of washes (Chapter 3), there was relatively low reproducibility in protein identifications (50-60%) between replicate IPs performed under the same conditions (Figure 4.16). The low identity reproducibility is also observed using the HA-Clover only negative control, suggesting an analytical repeatibility issue rather than a specific HIFα characteristic. It is clear that the HA-Clover only tag, and/or GFP-Traps resin, have their own specific binding partners, 'background' proteins that are removed from the HIFα list to identify 'real' binding partners. Nevertheless, the number of unspecific 'noise' proteins identified is significantly lower than the number of proteins identified by either HIFα isoform (Figure 4.16).

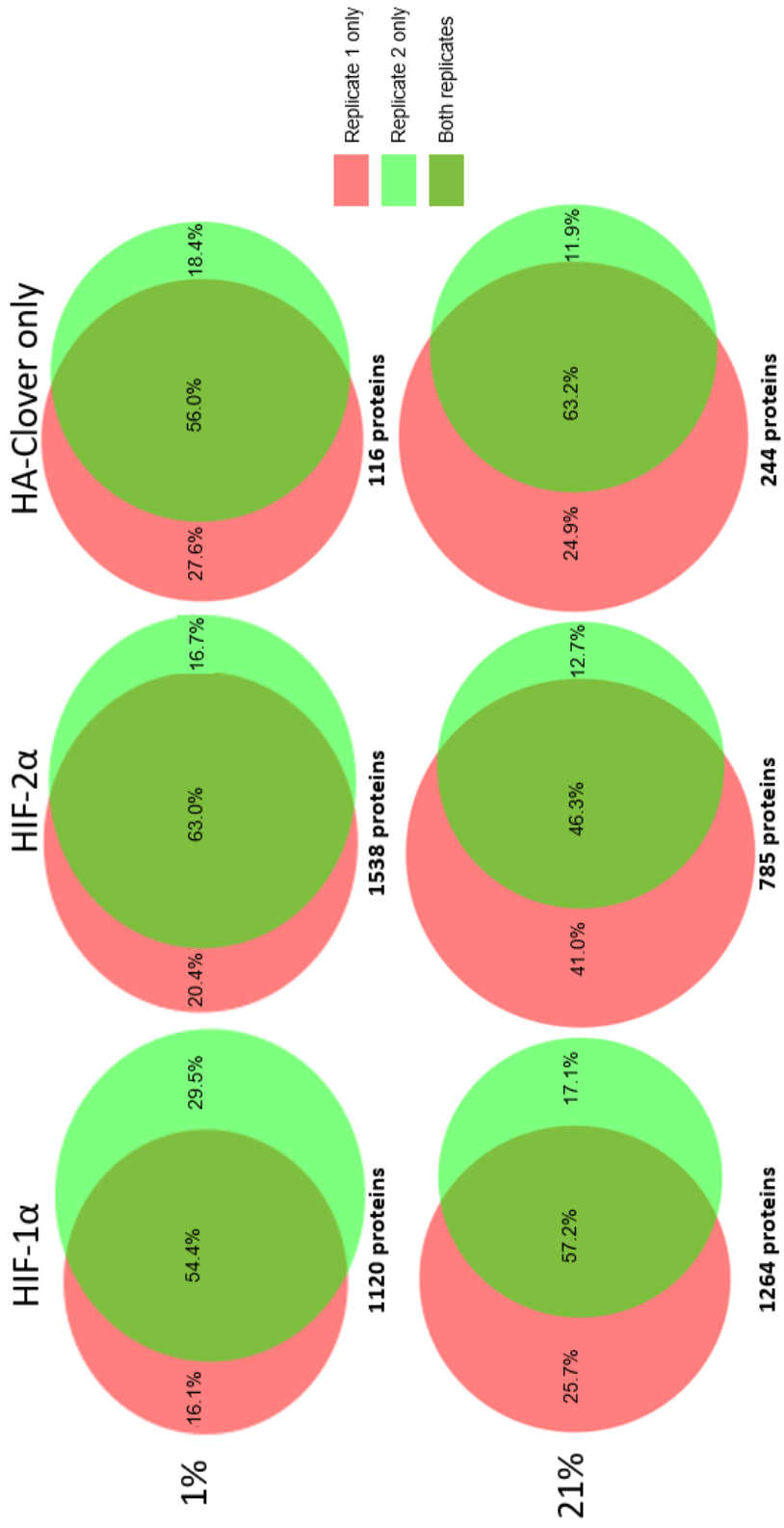


Figure 4.16: Replicate variability of identified proteins post IP of HA-Clover-HIF-1α, HA-Clover-HIF-2α and HA-Clover only.
Eluted material (5%, 15 µL) of post IP and digest was analysed by a high scan frequency High-Low LC-MS/MS method on a 1 hr gradient, data analysed using Proteome Discoverer using the MASCOT search engine and filtered to 1% FDR. Venn diagrams created using the BioVenn tool (Hulsen *et al.*, 2008). Protein count indicates the total number of different proteins identified from summation of replicates. % values reflect the number of binding partners identified per section, relative to the total protein count. Red = seen in replicate 1 only, light green = seen in replicate 2 only, dark green = seen in both replicates. Circles are scaled respectively to the number of binding partners identified per section.

4.10.2. O_2 dependent binding partners

Proteins that were confidently identified in both replicates were kept for further evaluation (Dark green of Figure 4.16), while proteins identified in only one replicate were filtered out.

For background subtraction, the same filtering was applied to HA-Clover only proteins identified and any proteins that were identified in the same O_2 tension experiments for both HA-Clover only and HA-Clover-HIFα IPs were removed. Post background subtraction, HIF-1α had a total of 610 and 571 protein identifications for 1% and 21% O_2 respectively, HIF-2α had a total of 900 and 270 protein identifications for 1% and 21% O_2 respectively (Figure 4.17).

Figure 4.17: O2 dependent binding partners for HIF-1α and HIF-2α.
Venn diagrams of proteins identified in 21% O_2 versus 1% O_2 for HIF-1α and HIF-2α post background subtraction. Red = proteins specific to 21% O_2, light green = proteins specific to 1% O_2, dark green = seen in both O_2 tensions. % values reflect the number of binding partners identified per section, relative to the total protein count. Circles are scaled respectively to the number of binding partners identified per section. Venn diagrams created using the BioVenn tool (Hulsen *et al.*, 2008).

As might have been expected, the HIFα binding partners change significantly as a result of different O_2 tensions. Interestingly, whilst similar total numbers of protein identifications

were identified as HIF-1α binding partners (571 vs 610 for 21% and 1% O_2 respectively), only ~34% of proteins are identified in both conditions; thus ~66% of the proteins identified are O_2 dependent. More striking is the HIF-2α binding partner data. As previously discussed, it could be hypothesised that PTMs and binding partners may play a greater role in regulating HIF-2α as its stability is less dependent on O_2 tension than HIF-1α. Data here indeed suggest this, with >700 proteins identified specifically in 1% O_2 compared to 73 at 21% O_2. This ~10 fold increase in the size of the interacting network induced by hypoxia is interesting (especially when compared to HIF-1α) and will require additional invesitgation.

4.10.3. Gene Ontology (GO) analysis

Gene ontology (GO) enrichment analysis was performed to identify which cellular pathways interact with HIFα isoforms, in an O_2 dependent and independent manner (Huang *et al.*, 2007 & Huang *et al.*, 2008). For GO annotation enrichment analysis, the DAVID (Database for Annotation, Visualisation and Integrated Discovery) tool was used (Dennis *et al.*, 2003). Data was stringently filtered for GO-terms of Biological process and Molecular function annotations at a Benjamini-Hochberg corrected P-value <0.05 only (Figure 4.18 to **Figure 4.21: GO Enrichment Analysis of HIF-2α interactors in either O2 tension.**Figure 4.21), all GO enrichment data in appendix 3). These filtering criteria were applied to provide a broad overview of the processes in which HIFα may be involved with and of the specific pathways that may regulate HIFα function.

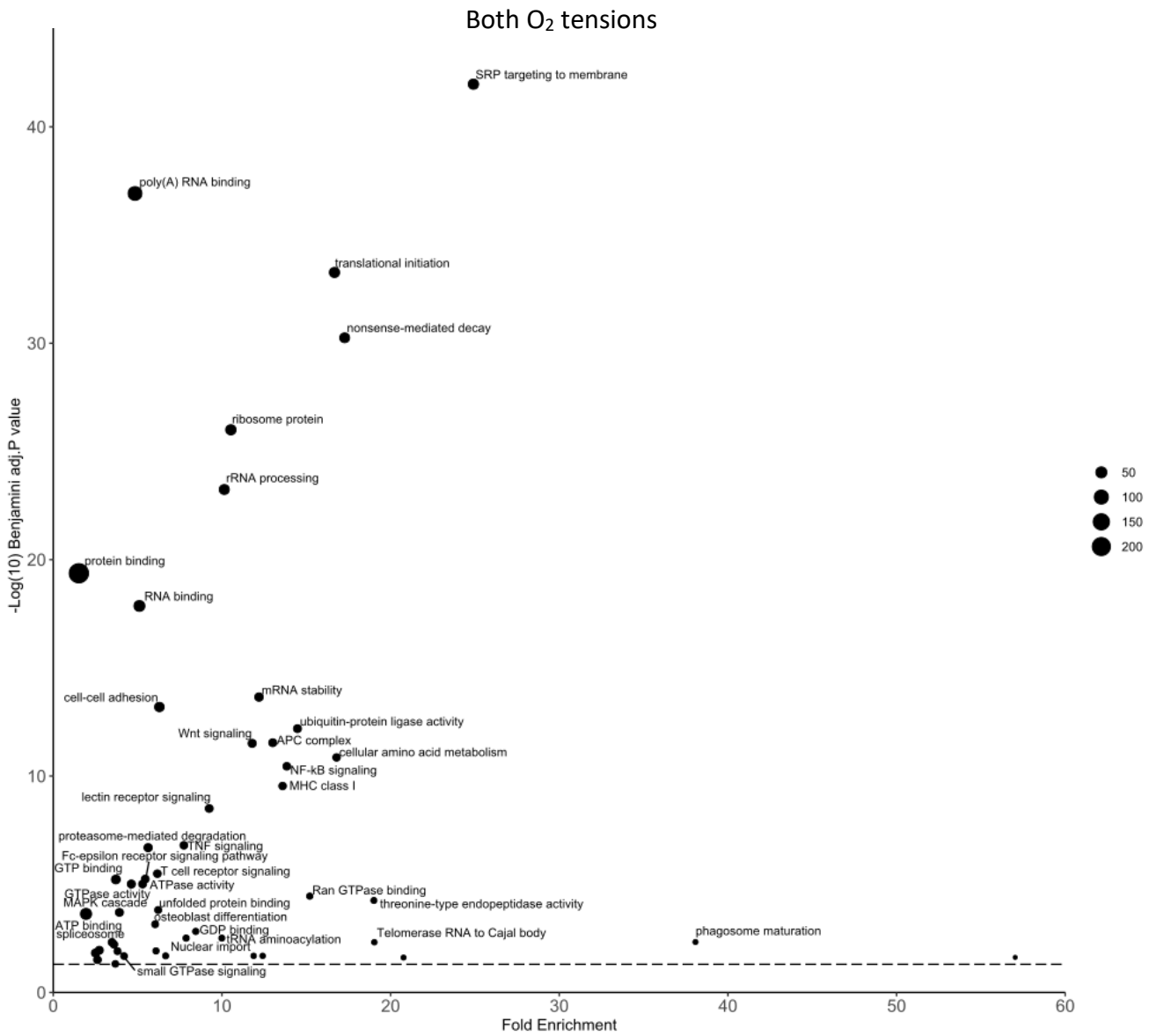


Figure 4.18: GO Enrichment Analysis of HIF-1α interactors identified in both O_2 tensions.
GO enrichment analysis was performed in DAVID using the background subtracted list of binding partners identified in both O_2 tensions (427 proteins). A Fishers exact test is performed for P-values and were adjusted for multiple hypothesis testing using the Benjamini-Hochberg correction and are plotted against fold enrichment. The dotted horizontal line represents an adjusted P-value of 0.05. Points are only plotted if the adjusted P-value is <0.05 and are labelled if the adjusted P-value is <0.01. Points are labelled with respective GO annotation and are scaled according to the number of proteins within an annotation.

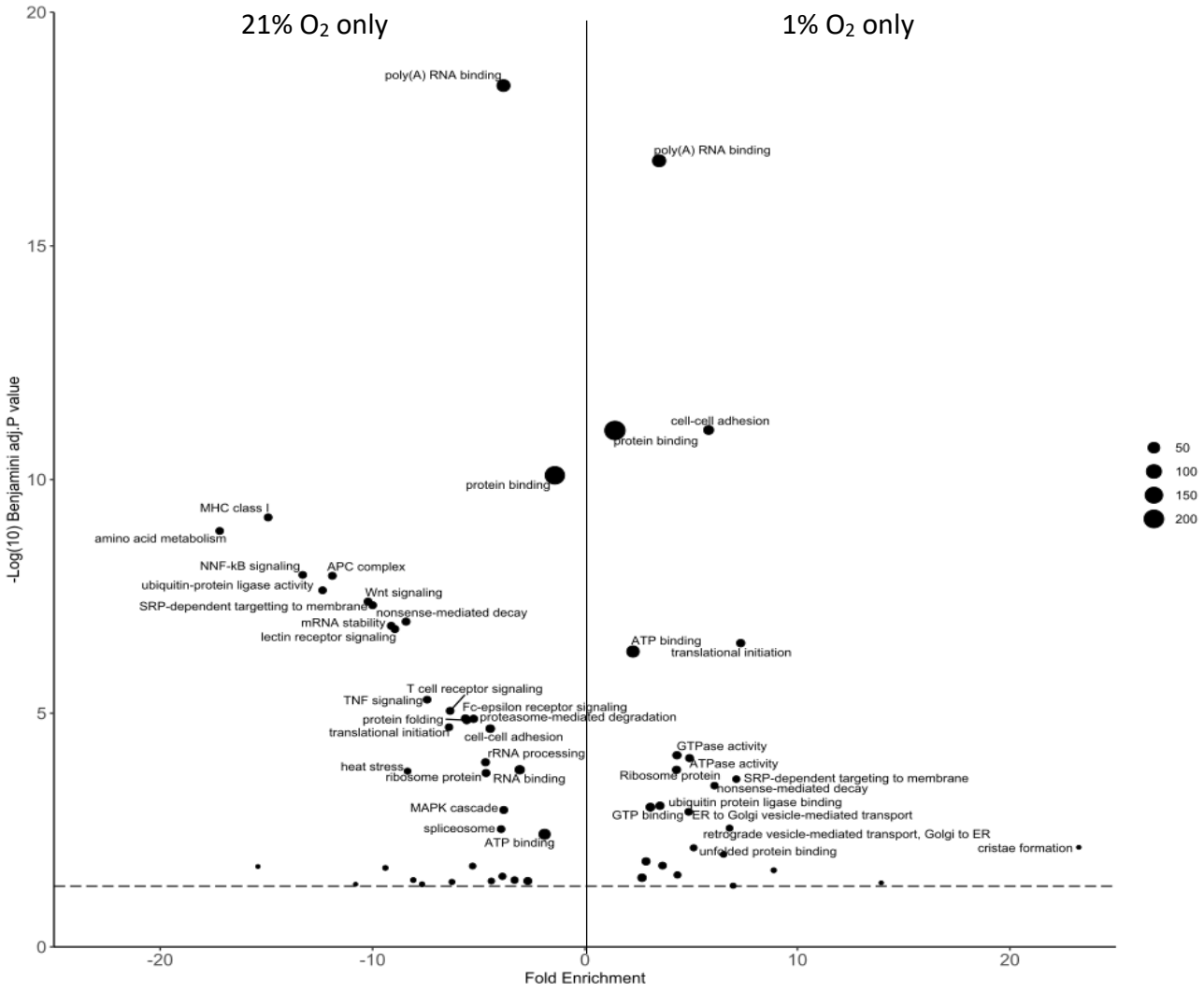


Figure 4.19: GO Enrichment Analysis of HIF-1α interactors identified in either O_2 tension.
GO enrichment analysis was performed in DAVID using the lists of background subtracted binding partners identified in 21% O_2 only (310 proteins), 1% O_2 only (235 proteins. A Fishers exact test is performed for P-values and were adjusted for multiple hypothesis testing using the Benjamini-Hochberg correction and are plotted against fold enrichment. The dotted horizontal line represents an adjusted P-value of 0.05. Points are only plotted if the adjusted P-value is <0.05 and are labelled if the adjusted P-value is <0.01. Points are labelled with respective GO annotation and are scaled according to the number of proteins within an annotation. Proteins identified in 21% O_2 only (left hand side) and 1% O_2 only (right hand side).

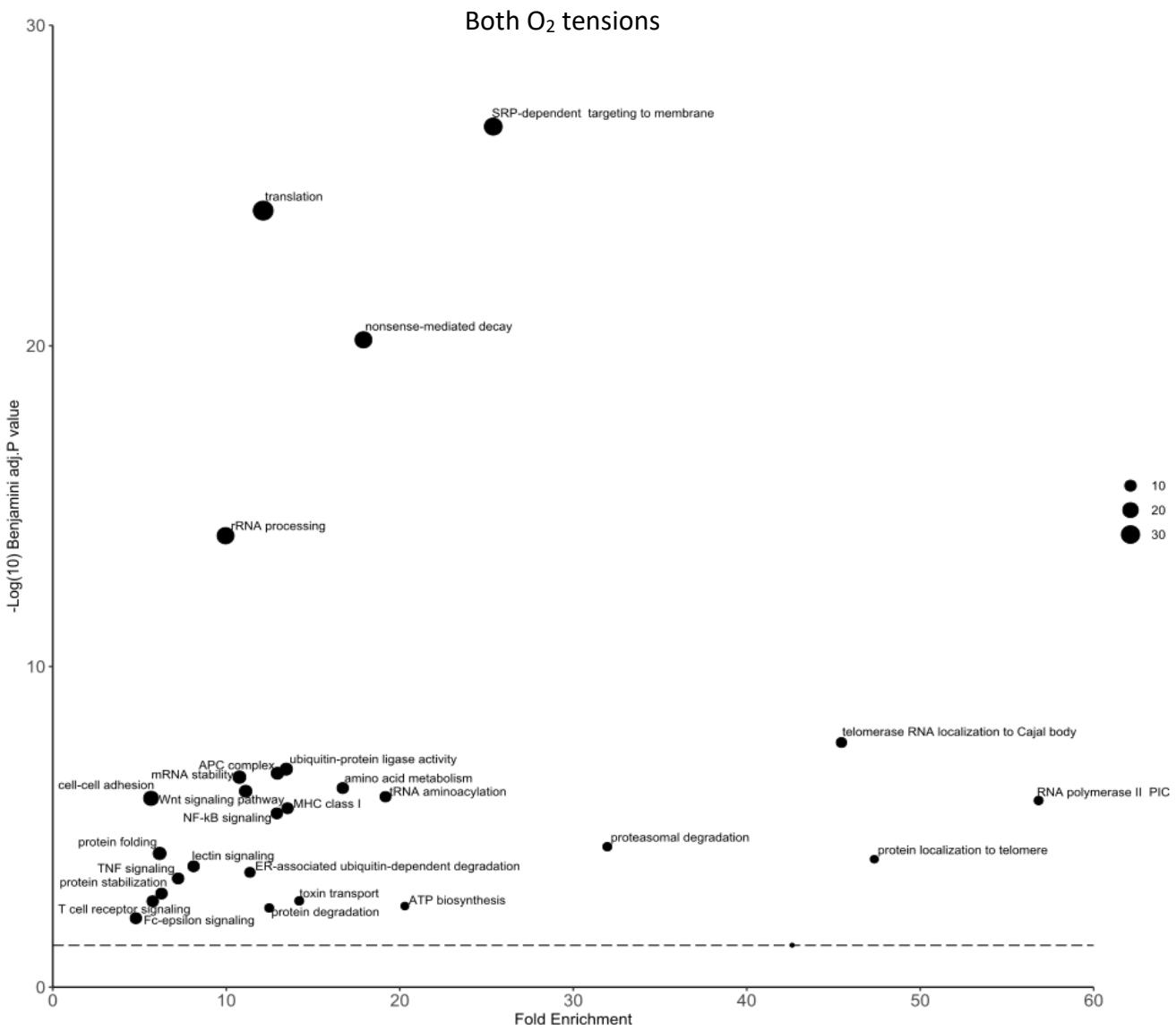


Figure 4.20: GO Enrichment Analysis of HIF-2α interactors in both O_2 tensions.
GO enrichment analysis was performed in DAVID using the lists of background subtracted binding partners identified in both O_2 tensions (292proteins). A Fishers exact test is performed for P-values and were adjusted for multiple hypothesis testing using the Benjamini-Hochberg correction and are plotted against fold enrichment. The dotted horizontal line represents an adjusted P-value of 0.05. Points are only plotted if the adjusted P-value is <0.05 and are labelled if the adjusted P-value is <0.01. Points are labelled with respective GO annotation and are scaled according to the number of proteins within an annotation.

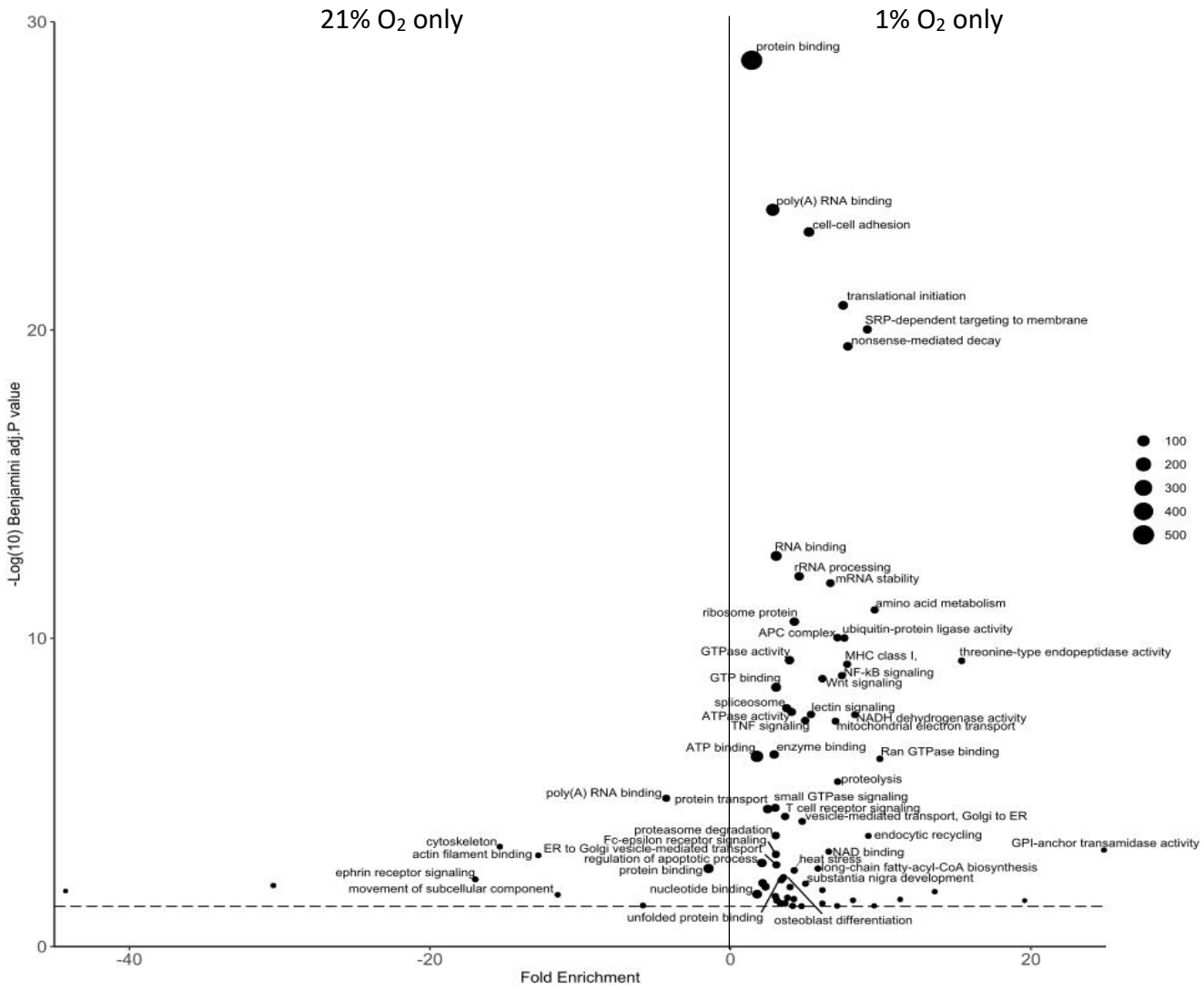


Figure 4.21: GO Enrichment Analysis of HIF-2α interactors in either O_2 tension.
GO enrichment analysis was performed in DAVID using the lists of background subtracted binding partners identified in 21% O_2 only (70 proteins), 1% O_2 only (677 proteins. A Fishers exact test is performed for P-values and were adjusted for multiple hypothesis testing using the Benjamini-Hochberg correction and are plotted against fold enrichment. The dotted horizontal line represents an adjusted P-value of 0.05. Points are only plotted if the adjusted P-value is <0.05 and are labelled if the adjusted P-value is <0.01. Points are labelled with respective GO annotation and are scaled according to the number of proteins within an annotation. Proteins identified in 21% O_2 only (left hand side) and 1% O_2 only (right hand side).

Although GO enrichment analysis was performed on binding partners specifically identified in 1% or 21% O_2 only, many of the significantly enriched GO annotations are found in both O_2 tensions (Figure 4.18 and Figure 4.20) and the separate O_2 tensions (Figure 4.19 and Figure 4.21), for example protein binding and poly(A) RNA binding. The likely explanation for this is the broad classification system used by DAVID, thus resulting in many different proteins grouped under the same GO annotation. An example of which is proteasomal proteins, significantly enriched GO annotations of: proteasomal degradation, protein binding, NF-kB signalling, WNT signalling, MAPK cascade, ATPase activity (and more) all

consist primarily of the proteasomal proteins with a single differentiating protein to classify them as independent groups; NEMO in NF-kB signalling or MAP2K3 in MAPK cascade. Therefore, although this analysis may provide some insight into potential functions, further more in-depth analysis is needed in respect of actual proteins identified before conclusions can be made. Hence only a few GO annotations will be discussed here. Due to issues of bioinformatics complexity and time constraints, HIF-1α versus HIF-2α comparisons were not performed.

For both HIFα isoforms, the amount of ribosomal proteins identified post background subtraction is interesting; with >50 different proteins associated to the ribosome identified. Although ribosomal proteins are generally considered background 'noise' of the IP, the lack of these in the HA-Clover only IPs in comparison to HIFα IPs would suggest otherwise. It is undeterminable from this dataset whether HIFα interacts with these proteins either directly or indirectly.

Interestingly, the ATPase activity/ATP binding groups conisist of many parts of the mitochondrial electron transport chain, including the majority of the ATP-synthase pump. A search of the data for mitochondrial proteins (DAVID GO-annotation) highlights a significant number of proteins IP'd that are endogenously located within the mitochondria at >100 in total (~10% of total proteins identified) for both HIFα isoform, and O_2 independently binding. This number of internally localised mitochondrial proteins identified would suggest that the HIFα proteins may have a fraction of total protein localised into the mitochondria. This hypothesis is also supported by some published studies which used biochemical and proteomics approaches to show a sub-population of HIF-1α localised into the mitochondria (Briston *et al.*, 2011, Concolino *et al.*, 2018 & Thomas *et al.*, 2019). Although these papers only mention HIF-1α, they do not investigate HIF-2α. Therefore it is possible from the data presented here that HIF-2α also localises to mitochondria. Because of these identifications, we decided to search the mitochondrial DNA (UniProt accession: NC_012920, Andrews *et al.*, 1999) for HREs to suggest whether HIF may have a role in regulating mitochondrial gene expression. Interestingly, we identified 20 putative HRE promoter sequences ((A/G)CGTG, Schödel *et al.*, 2011), although further investigation is needed to understand whether these may be promoters and regulate gene expression, or are intronic in nature.

For HIF-1α, the GO annotations identified are marginally different between protein identifications of 21% only, 1% only and both O_2 conditions. The potentially more interesting GO annotations, such MAPK cascade and NF-kB singalling that are only identified in 21% O_2, primarily consist of the identical proteins with a single differentiating protein (Protesomal proteins with either MAP2K3 or NEMO respectively). Therefore GO annotations need further investigation before drawing conclusions, as discussed above.

Significantly more enriched GO annotations were identified for HIF-2α at 1% O_2, compared to 21% O_2, as would be expected given the ~10 fold increase in unique protein identifications (73 compared to >700 respectively). Surprisingly, HIF-2α in 21% O_2 had cytoskeletal proteins specifically enriched, with no equivalent GO annotations in 1% O_2 or any O_2 tension of HIF-1α. The major constituents of the cytoskeletal protein GO annotation are: proteins of the Arp2/3 complex, Talin and Tubulin isoforms.

Interestingly, enzymes essential for GPI (glycosylphosphatidylinositol) anchoring are specifically enriched by HIF-2α at 1% O_2 only: GPAA1, PIGK, PIGS, PIGT & PIGU, all of which are located within the Endoplasmic Reticulum (ER) (GPI anchor function and localisation of binding partners discovered by: Ohishi *et al.*, 2000 & Ohishi *et al.*, 2001). Predicition tools (PredGPI) suggest HIF-2α does not have any sites for GPI-anchoring (Pierleoni *et al.*, 2008). This therefore suggests that, upon hypoxic incubation, HIF-2α may become localised into the ER. In addition, our data finds that the GO annotation of: 'ER to Golgi vesicle mediated transport' is also significantly enriched, specifically in 1% O_2 and for HIF-2α only. This could reflect a potential role of HIF-2α in extracellular vesicle packaging. Aga *et al.*, 2014 has shown that HIF-1α can be packaged into extracellular vesicles, which then influence surrounding cells to promote cancer progression. However HIF-2α has not been investigated for similar function.

Another interesting aspect is the enrichment of telomerase influencing proteins identified as HIF-2α specific binding partners; with both 'telomerase RNA localization to Cajal bodies' and 'protein localisation to teleomere' significantly enriched in both O_2 tensions. Due to the nuclear speckle localisation of telomeres (Abreu *et al.*, 2011 & Lee *et al.*, 2015), it is feasible that HIF-2α interaction with proteins associated with telomeres could partially explain the speckled nuclear localisation obsered of HIF-2α.

Investigating the proteins which constitute these telomere associated GO annotations show they consist of identical proteins: RUVB like AAA ATPase's and 6 Chaperonin containing TCP1 subunits (2, 3, 5, 6, 7, 8). The latter are components of the TRiC complex which has been shown to regulate PHD3 specifically (Masson *et al.*, 2004). However, PHD3 was not identified as a HIF-2α binding partner here, thus the PHD3-HIF-2α association cannot be blamed for the HIF-2α-TRiC identification.

4.11. Label Free Quantification analysis of binding partners

Due to experimental size and setup, label based quantification was not appropiate. Therefore a label free approach was used to quantify binding partner fold changes induced by O_2 tension. As discussed in the introduction (Section 1.1.18.1) Label Free Quantification (LFQ) is less accurate than labelling approaches as samples are prepared and analysed independently of each other, hence experimental error and variation between runs can impact the precision and accuracy of protein quantification (Bantscheff *et al.*, 2012 & Bantscheff *et al.*, 2007). Therefore LFQ approaches are reliant on normalisation against 'housekeeping proteins' (Välikangas *et al.*, 2018). However, hypoxia induces large effects in gene expression of many house keeping genes (Caradec *et al.*, 2010), that are used as housekeeping proteins, combined with IP specifically enriching a subset proteins from the proteome in an unknown manner (and thus may include common housekeeping proteins), these approaches are not necessarily adequate here.

MaxQuant is an open source software that allows the LFQ analysis of such data. Critically, MaxQuant determines the summed protein intensities initially to identify proteins that minimally change between conditions, which are used to normalise data to in a process termed delayed normalisation (Cox *et al.*, 2014 & Cox *et al.*, 2011). Perseus, also open source software, has been designed to aid the extraction of required information from MaxQuant outputs and aid in visualisation with a range of tools (Tyanova *et al.*, 2016).

LFQ was based on a minimum of 1 peptide identified at 1% FDR, calculating the area under the peak. Peptides were grouped by their unique protein identifier and were only retained in the data if they were observed in both replicates of either O_2 tension. Peptide intensities were averaged to infer protein intensity and normalised to the fold change of HIFα proteins, as small abundance differences (<1.2 fold) were found between O_2 tensions. Thus, the data contains all binding partner interactions, as filtered previously when using

Proteome Discoverer, with calculated protein intensities that are normalised to HIFα levels, available in Appendix 4.

Importantly, MaxQuant matches peaks between conditions even if one condition lacks an identified MS/MS spectra, due to too low intensities. Missing values (undetected *m/z* ratios) are imputed with a fictional number, from a normally distributed dataset that is below the lowest intensity recorded from the MS/MS data. Theoretically this should result in statistically valid differences, however if intensities are variable and of low intensity it is possible that imputed values lead to the loss of significance. Hence data shown from the Proteome Discoverer pipeline (Figure 4.17), that requires MS/MS data, have many more proteins that are O_2 dependently regulated, whereas LFQ data (Figure 4.22 & Figure 4.23) has statistically significant fold change differences of data that includes proteins identified in both O_2 tensions. For example, it is feasible to have a 10 fold change in protein intensity, identified by LFQ, which was initially identified in both O_2 conditions.

Figure 4.22: LFQ analysis of HIF-1α binding partners in response to O_2 tension.
LFQ was performed using MaxQuant, based on a minimum of 1 peptide identified at 1% FDR in both replicates of either O_2 tension, with match between runs applied. Peptide intensity was measured by the area under the peak and all peptide intensities averaged for protein intensity. Missing intensities were imputed from a normal distribution. Protein intensities were normalised to HIF-1α protein intensity. P-values were determined using a Students 2-sample t-test. Plot is of $-Log_{10}$(P-Value) against Log_2(fold change), where positive values are enriched in 1% O_2 and negative values enriched in 21% O_2. Dotted lines are at a P-value <0.05 and 2- Fold enrichment. Points coloured red are significant at P-value <0.05. Labelled points are significant at P-Value <0.01 and have a fold change of >2. Point size is scaled to the number of peptides identifying a protein.

Figure 4.23: LFQ analysis of HIF-2α binding partners in response to O_2 tension.
LFQ was performed using MaxQuant, based on a minimum of 1 peptide identified at 1% FDR in both replicates of either O_2 tension, with match between runs applied. Peptide intensity was measured by the area under the peak and all peptide intensities averaged for protein intensity. Missing intensities were imputed from a normal distribution. Protein intensities were normalised to HIF-2α protein intensity. P-values were determined using a Students 2-sample t-test. Plot is of $-Log_{10}$(P-Value) against Log_2(fold change), where positive values are enriched in 1% O_2 and negative values enriched in 21% O_2. Dotted lines are at a P-value <0.05 and 2- Fold enrichment. Points coloured red are significant at P-value <0.05. Labelled points are significant at P-Value <0.01 and have a fold change of >2. Point size is scaled to the number of peptides identifying a protein.

As suggested, the number of binding partners that are significantly differentially IP'd is lower than the Proteome Discoverer pipeline to identify O_2 dependent binding partners. For example, Proteome Discoverer identified >300 binding partners specific to HIF-1α in 1% O_2

only, compared to <150 identified with LFQ at a P-value <0.05. Interestingly, LFQ shows that some binding partners do have very dramatic fold changes induced by O_2 tension. For example the fold change of greatest statistical significance for HIF-1α (~80 fold, P-value = ~$1x10^{-7}$) is ANKRD34B. Greater fold changes do exist, for example HSPA5 (also known as BiP) with >1000 fold change between O_2 tensions, however investigating the data shows both 21% O_2 replicates resulted in no identified peptides and are thus has imputed values. Therefore, HSPA5 is a true hypoxia regulated binding partner, interesting considering its ER specific localisation.

Brief investigation of this LFQ data identifies interesting fold changes from a PTM standpoint. VCPIP is a deubiquitinating enzyme (DUB), known to be important in mitosis and golgi assembly (Wang *et al.*, 2004), which bound HIF-1α 5 fold greater in 21% O_2 than 1%; a surprising observation considering the role of ubiquitination and degradation. HIF-2α binding partner shows an inverted ubiquitination protein associated binding partner pattern, with the significantly increased intensities of ubiqutinating enzymes LTN1 and HUWE1 in 1% O_2, by ~10- and ~ 4- fold respectively. Therefore, these data suggest very different roles of ubiquitination in HIFα isoform regulation.

4.12. Protein kinase binding partners of HIFα

As we were particularly interested in phosphorylation as a PTM, we searched the DAVID data output by kinases. The kinase data was then filtered down for protein kinases specifically (Table 4.20), as the datset contained nucleotide, lipid and sugar kinases as well as kinase interacting proteins, such as NEMO. A total of 14 and 19 proteins kinases were identified for HIF-1α and HIF-2α respectively. This is perhaps not surprising given the transient nature of kinase – substrate interactions.

Table 4.20: HIF-1α and HIF-2α interacting protein kinases and O_2 tensions observed in.
All binding partners that were observed in 1% O_2, 21% O_2 or both O_2 tensions were analysed through DAVID and GO annotations for kinases kept. Non-protein kinases were filtered out of the data.

HIF-1α

1% only		
Accession	Description (Gene name)	Type (S/T/Y)
P30530	AXL receptor tyrosine kinase(AXL)	Y
Q13418	Integrin linked kinase(ILK)	S/T
P51617	Interleukin 1 receptor associated kinase 1(IRAK1)	S/T
P42345	Mechanistic target of rapamycin(MTOR)	S/T
P36507	Mitogen-activated protein kinase kinase 2(MAP2K2)	T/Y
O95747	Oxidative stress responsive 1(OXSR1)	S/T

21% only		
O14965	Aurora kinase A(AURKA)	S/T
P49841	Glycogen synthase kinase 3 beta(GSK3B)	S/T
P46734	Mitogen-activated protein kinase kinase 3(MAP2K3)	T/Y
O75962	Trio Rho guanine nucleotide exchange factor(TRIO)	S/T
Q9Y3F4	Serine/threonine kinase receptor associated protein(STRAP)	S/T

BOTH O_2 tensions		
P10398	A-Raf proto-oncogene, serine/threonine kinase(ARAF)	S/T
P06493	Cyclin dependent kinase 1(CDK1)	S/T
P78527	Protein kinase, DNA-activated, catalytic polypeptide(PRKDC)	S/T

HIF-2α

1% only		
Accession	Description (Gene name)	Type (S/T/Y)
P10398	A-Raf proto-oncogene, serine/threonine kinase(ARAF)	S/T
P30530	AXL receptor tyrosine kinase(AXL)	Y
P08581	MET proto-oncogene, receptor tyrosine kinase(MET)	Y
Q8TDX7	NIMA related kinase 7(NEK7)	S/T
O14976	Cyclin G associated kinase(GAK)	S/T
P06493	Cyclin dependent kinase 1(CDK1)	S/T
P00533	Epidermal growth factor receptor(EGFR)	Y
O14920	Inhibitor of nuclear factor kappa-B kinase subunit bet	S/T
Q13418	Integrin linked kinase(ILK)	S/T
P51617	Interleukin 1 receptor associated kinase 1(IRAK1)	S/T
P42345	Mechanistic target of rapamycin(MTOR)	S/T
P36507	Mitogen-activated protein kinase kinase 2(MAP2K2)	T/Y
P46734	Mitogen-activated protein kinase kinase 3(MAP2K3)	T/Y
O95747	Oxidative stress responsive 1(OXSR1)	S/T
P15735	Phosphorylase kinase catalytic subunit gamma 2(PHKG2)	S/T
Q13131	Protein kinase AMP-activated catalytic subunit alpha 1(PRKAA1)	S/T
Q9Y3F4	Serine/threonine kinase receptor associated protein(STRAP)	S/T

NO 21% only kinases identified

BOTH O_2 tensions		
P78527	Protein kinase, DNA-activated, catalytic polypeptide(PRKDC)	S/T
P53350	Serine/threonine-protein kinase PLK1	S/T

Table 4.20 clearly shows that the HIFα binding kinome is influenced by O_2 tension, with the majority of kinases only identified in a specific O_2 tension. This is partly unexpected considering the phopshorylation maps of both HIFα isoforms (Figure 4.3) did not differ greatly between O_2 tensions. However, as discussed, it is likely that some kinases that phosphorylate HIFα have not been identified due to their transient nature. Importantly, it is impossible to decipher whether the identified kinases are directly binding to HIFα proteins or are indirectly immunoprecipitated through a secondary binding partner.

The identification of tyrosine specific kinases is of interest since no tyrosine phosphorylation sites were identified from our mass spectrometry analysis. It is possible that the level of tyrosine phosphorylation is below the limit of detection for the methodology or, alternatively, an indication that HIFα is localised with these kinases in a 'primed' state for the correct signal to enable a rapid response. Additionally, TiO_2 is known to have limitations in terms of phospho-tyrosine identification, therefore a more targeted approach would be useful for these investigations using phospho-tyrosine specific antibodies for secondary IPs (Lombardi *et al.*, 2015). Whilst some kinases identified have previously been shown to phosphorylate the HIFα proteins, including: MTOR (Land *et al.*, 2007), GSK3β (Flügel *et al.*, 2007) and CDK1 (Warfel *et al.*, 2013), the majority of the kinases identified are novel interactors of HIF-1α and HIF-2α.

All of the HIF-1α 1% O_2 only identified kinase interactors were also identified as HIF-2α 1% O_2 only binding partners. Of the 21% O_2 only HIF-1α kinases, both STRAP and MAP2K3 were also identified as HIF-2α binding partners, however only in 1% O_2. AURKA and GSK3β were the only kinases specific to HIF-1α, both identified in 21% O_2. Although conflicting, a transcriptional relationship has been shown between HIF-1α and AURKA, yet no interaction or phosphorylation has been found (Cui *et al.*, 2013 & Fanale *et al.*, 2013). GSK3β has been shown to multiply phosphorylate HIF-1α within the ODDD, ultimately leading to protein degradation, independently of proline hydroxylation and O_2 tension (Flügel *et al.*, 2007, Flügel *et al.*, 2012, Cassavaugh *et al.*, 2011 & Mottet *et al.*, 2003), therefore its HIF-1α only identification could potentially explain some of the stability differences between HIF-1α and HIF-2α.

The HIF-2α specific kinase interactors identified are all novel. Interestingly, PLK1 was identified in this study, with the PLK3 isoform previously shown to phosphorylate HIF-1α at S657, although HIF-2α was not investigated in this published study (Xu *et al.*, 2010). Here

we show that, using HIFα isoform sequence alignments, the known PLK3 dependent HIF-1α S657 phosphorlyation site aligns to the novel HIF-2α T626 site, therefore could suggest the phosphorylation site for the PLK1 kinase identified.

MET (Rankin *et al.*, 2014 & Vanichapol *et al.*, 2015), NEK7 (Korgaonkar *et al.*, 2008), EGFR (Franovic *et al.*, 2009 & Alam *et al.*, 2016) and GAK (a CDK5 associated protein, Herzog *et al.*, 2016) are all kinases essential for mitotic cell cycle progression that have been linked as target genes of the hypoxic response as part of a feedback loop. None of these kinases have been implicated in HIFα regulation, rather identifying expression level changes of stated kinases in response to hypoxia.

4.13. Known binding partners

The current main regulatory pathways for HIFα regulation is the O_2 dependent degradation and inactivation pathway and transcriptional complex formation, consisting of 7 proteins: PHD1-3 (EGLN1-3), FIH-1, HIF-1β and p300/CBP (Ivan *et al.*, 2001, Epstein *et al.*, 2001, Bruick *et al.*, 2001, Jaakkola *et al.*, 2001, Semenza *et al.*, 1992, Wang *et al.*, 1995, Lando *et al.*, 2002, Lando *et al.*, 2002, Masson *et al.*, 2003, Jiang *et al.*, 1996, Arany *et al.*, 1996 & Kallio *et al.*, 1998)). Therefore we searched our data to investigate how these interactions change in response to hypoxia, simultaneously validating our findings. This is of particular interest as the experimental design used here overexpresses WT HIFα, rather than traditionally used O_2 insensitive proline mutations to prevent the hydroxylation-degradation pathway.

4.13.1.1. HIF-1β

For HIF-1α, HIF-1β (ARNT) was found as a binding partner in both O_2 tensions. LFQ however identifies, counter intuitatively, that HIF-1β was enriched >6 fold more in 21% O_2 than 1% O_2, and is significant to a P-value <0.05 (hence not labelled in Figure 4.22). For HIF-2α, HIF-1β was found to be 1% O_2 specific using the Proteome Discoverer analysis pipeline, yet MaxQuant showed a non-significant no fold change. The latter is likely due to variation in the data at 21% O_2, with one replicate showing imputed data values (no peptides identified, hence removed as an interactor in Proteome Discoverer data) while the second replicate shows minimal differences to 1% O_2 intensities. Thus the nature of O_2 sensitive binding is undeterminable without further replicates.

4.13.1.2. FIH-1

FIH-1 was not identified, in any condition and for either HIFα isoform, although the 'open' PTM search identified HIFα asparagine hydroxylation. This could suggest highly transient or weak interactions of HIFα - FIH-1.

4.13.1.3. p300/CBP

p300/CBP was not identified in any HIF-1α sample. For HIF-2α, p300/CBP showed different results, depending on the data analysis platform used (similar to HIF-1β). Proteome discoverer failed to identify p300/CBP in 21% O_2, thereby classifying it as 1% O_2 specific protein. However, MaxQuant detects >2 fold enrichment of p300/CBP in 21% O_2 with a significant P-value (<0.01). In these circumstances, it is of greater significance to believe Proteome Discoverer data because they require MS/MS spectra and determination of primary peptide sequence to identify a peptide, where MaxQuant can, infrequently, mistake non-fragmented identical *m/z* ratios at ther MS1 stage as the same peptide, which may not be the case.

4.13.1.4. PHDs

The PHDs that have been shown to regulate both HIFα isoforms are translated from the EGLN genes. For HIF-1α EGLN1 (PHD2) was identified as a binding partner in both O_2 tensions. LFQ found that there was infact a 4 fold increase of EGLN1 in 21% compared to 1% O_2, as may be expected when O_2 is a limiting factor. Interestingly, for HIF-2α EGLN1 was identified as a binding partner specific to 1% O_2, through Proteome Discoverer, and supported by LFQ identifying >3 fold more EGLN1 in 1% O_2 than 21% O_2. The functional importance of this is unknown, however would suggest that in early hypoxia HIF-2α may experience an increased rate of proline hydroxylation. HIF-1α IPs did not identify any other interacting EGLN proteins, likely due to their involvement with rapid HIF degradation. Similarly, EGLN2 (PHD1) was not detected by HIF-2α IPs. EGLN3 (PHD3) was identified using MaxQuant, however there was no significant quantitative fold change dependent on O_2 tension observed.

Interestingly, searching for prolyl hydroxylases using DAVID identified multiple other proline hydroxylases as novel interactors of HIFα proteins, including: Prolyl 4-hydroxylases 1, and 2 (procollagen-proline, 2-oxoglutarate-4-dioxygenases), Protein Disulfide isomerase (Prolyl 4-hydroxylase subunit beta) and Prolyl 3-hydroxylases 1, 2 and 3 (Leprecans), referred to by their gene names (P4HA1, P4HA2, P4HB, P3H1, P3H2 and P3H3) from now. Although P4HB is a protein disulfide isomerase, and not a proline hydroxylase, it is an

essential component of the P4HA1/2 complexes which result in proline hydroxylation. Thus, P4HB acts as supporting evidence for the correct identification of these additional proline hydroxylases.

P4HA1/2 are localised to the endoplasmic reticulum and hydroxylate proline residues at position 4 of the proline side chain (Pihlajaniemi *et al.*, 1991 & Myllyharju, 2003). Their primarily known for their function in hydroxylating collagen proteins. However, P4HA1/2 have been shown to hydroxylate other proteins at conserved A-P-G motifs too, such as prion proteins and Argonaute 2 (Gill *et al.*, 2000 & Qi *et al.*, 2008), although this motif is not present in either HIFα isoform. HIF-1α has been shown to increase the expression of the P4HA1/2 genes in response to hypoxia, which in turn can act as a positive feedback loop to promote HIF-1α stability, by limiting the availability α-ketoglutarate, an essential cofactor for the EGLN proteins (Epstein *et al.*, 2001, Gilkes *et al.*, 2013 & Xiong *et al.*, 2018). We found that HIF-1α had ~6 fold more P4HA2 enriched in 1% O_2 than 21% O_2, (P-value <0.05), while PHB had no fold change and P4HA1 was not identified. For HIF-2α a similar pattern was observed with ~4 fold more P4HA2 enriched in 1% O_2, while both P4HA1 and P4HB were found to have no fold change induced by O_2 tension.

P3H1/2/3 hydroxylate proline residues at position 3 of the side chain of Proline residues. P3H1 is localised within the ER and basement membrane of cells, while P3H2/3 are both localised within the golgi. P3Hs are much less studied in comparison to the PHDs and P4HAs, however are known to be highly specific in their hydroxylation of collagen isoforms (Vranka *et al.*, 2004, Hudson *et al.*, 2013 & reviewed by Gorres *et al.*, 2010).

4.14. Discussion

The data presented in this chapter are the first known discovery proteomics analysis to map O_2 dependent PTMs and binding partners of full length HIFα isoforms from human cells. Previous studies have used targeted proteomics approachs and/or recombinant or overexpressed fragments of the HIFα proteins to identidy single PTMs from *in vitro* assays. Besides the very low abundance of HIFα proteins in human cells, their primary sequence makes mass spectrometry analysis, by canonical tryptic digestion, data limiting; with maximal sequence coverage of ~35%. We were able to achieve >90% sequence coverage of the HIFα proteins by combining data from different protease digests, a process not commonly applied to proteomics due to the cleavage sites of different proteases increasing

the chance for 1+ charge state peptide ions, which are suboptimal for MS/MS based primary sequence determination.

Although an open PTM search was performed to identify any PTM, the specific focus of this study was on phosphorylation, primarily due to the availablity of phospho-peptide enrichment strategies from the background of all unmodified/alternative PTM peptides and the known roles of phosphorylation in regulating HIFα proteins. All phosphorylation data were obtained using an optimised high sensitivity, high resolution MS/MS method to maximise peptide fragmentation and phosphorylation site localisation. Phosphorylation localisation from MS/MS data was analysed using a statistical software tool (ptmRS), to predict the confidence of site localisation. In total, ~50 different PTMs were identified on both HIF-1α and HIF-2α, with ~25 of these being phosphorylation at different sites. It is important to consider that the level of overexpression required for this in-depth PTM analysis may result in artifactual PTMs being identified. Whilst this strategy maintains the PHD-VHL pathway, although the rate of HIFα synthesis outweighs the degradation rate, therefore could be argued more biologically relevant than proline hydroxylation mutants that are traditionally used for O_2 independent protein expression.

Combining identified PTMs with phylogeny analysis, I set out to identify potentially functionally important PTM sites; highly conserved regions are thought to be more likely to have important regulatory roles. Similarly, sites of phopshorylation that show variation specifically to negatively charged residues could reflect an important role that has evolved to be signalling activated, where residues that are uncharged and non-phosphorylatable could reflect the loss of a signalling pathway.

However, with previous knowledge, the phylogeny analysis can also be hypothesis driven. For example species that are hypoxia tolerant vs sensitive may have sequence variation to aid in adaptions. This is supported by recent studies by anthropologists, where humans that have been living at high altitudes for many generations (Tibetan, Andean) have a significant enrichment of mutations in HIF-2α. However these were largely within intronic regions (Simonson *et al.*, 2010, Yi *et al.*, 2010 & Beall *et al.*, 2010). Interestingly, our cancer genomics database mining also found that the HIFα genes were very rarely mutated (both in exonic and intronic regions), although HIFα expression levels and downstream targets were generally increased. This suggests that regulatory proteins of HIFα are more prevalent to mutation and in turn alter HIFα function, as known for the EGLN genes (Simonson *et al.*,

2010 & Bigham *et al.*, 2009). This study also compared HIF-1α and HIF-2α PTMs to potentially identify sites/patterns, which could aid in explaining the functional differences between isoforms.

4.14.1. PTM data

It is clear from Table 1.2 & Table 1.3 that the majority of published PTMs of HIFα proteins, including phosphorylation, were not identified in this study. There are numerous potential reasons for this: 1) past studies generally using *in vitro* assays and recombinant HIFα fragment based approaches which are inherently suspectible to artefacts. 2) HIFα regulation, including O_2 dependent regulation, is highly dependent on cell type (Bracken *et al.*, 2006), thus the use of a single cell line, single hypoxic O_2 tension and single incubation time could all affect identified PTMs. 3) Only phosphorylation was analysed in depth. Other PTMs that may occur at low frequencies may not be identified if they are below the limit of detection in these analyses. Interestingly, the main functional domains that share minimal sequence homology, thus can be argued to differentiate, HIF-1α and HIF-2α have distinctly different PTM profiles.

4.14.1.1. ODDD

It is obvious from Figure 4.5 that the ODDD of HIF-1α is hyperphosphorylated, with a total of 15 different phosphorylation sites identified. Interestingly, 7 of these phosphorylations are O_2 dependent and only occur in 21% O_2, particularly curious considering the process of phosphorylation does not require O_2 as a cofactor, like hydroxylases. Additionally, considering there are a total of 12 O_2 dependent phosphorylation events for HIF-1α, this means that 67% of the O_2 dependent phosphorylation sites identified are located within the ODDD. Overall, this may reflect on three possibilies: 1) the role of phosphorylation in the canonical PHD degradation pathway, where phosphorylation may promote PHD association and degradation. 2) The complexity of alternative degradation pathways at play, or 3) kinases that are O_2 dependently regulated have function in the PTM of HIF, for example AMPK kinase (Shao *et al.*, 2014 & Hwang *et al.*, 2014).

As a crystal structure of the ODDD is unavailable, sequence based *ab initio* modelling was performed on the region encompassing the densest, and O_2 dependent, phosphoylation region (475-535). This revealed that all O_2 independent sites are internally facing and in such close proximity to negatively charged glutamic acid residues (<5 Å) that space-charge effects will likely lead to structural rearrangements. Modelling also highlighted that O_2

dependent phosphorylations are all externally facing, which could potentially act as docking sites for binding partners and secondary functions, such as promoting degradation. Unfortunately, our data could not decipher whether these PTMs co-exist or are mutually exclusive, therefore it is possible that O_2 dependent PTMs are mutually exclusive to O_2 independent. Hence, it will be interesting to investigate how mutational analysis will affect HIF-1α function. However, this may require mutation of multiple sites for functional effects to be observed. To limit the number of mutations required, a 'Middle-Down' MS approach could be adopted, where HIF-1α is minimally digested, in order to determine the co-existance of phosphorylation sites on larger peptides (~10 kDa).

Conversely, HIF-2α has few phosphorylation sites within the ODDD, a total of 5, with only one O_2 dependent site. Therefore, if phosphorylation is linked to protein degradation, it is conceivable that the differences in protein stability between HIF-1α and HIF-2α at 21% O_2 could be explained by the different phosphorylation status of the HIFα ODDD.

4.14.1.2. N-Terminal Transactivation Domain

Before obtaining PTM data, I hypothesised that the NTAD would undergo extensive PTM changes due to its role in coordinating HIFα specific gene targeting. However, the opposite was observed, where a distinct lack of all PTMs was found in this domain, even though ~100% of the NTAD was observed by MS for both HIFα isoforms. Infact, HIF-1α only had 1 PTM identified: K532 methylation, and HIF-2α had 4 phosphorylations identified: T517, T528, S543 and T559, making the NTAD the least dense PTM containing domain of both HIFα isoforms. Since the NTAD is the region of least sequence homology between HIFα isoforms, it is likely sequence/structural differences within this domain explains the differential gene targeting seen by HIFα isoforms, rather than PTM status. However, future more in-depth analyses of different PTMs may highlight a significant role of PTMs in the NTAD, as PTMs could be at low stoichiometric levels, considering the multiple different PTMs of K532 acetylation, ubiquitination and methylation (identified here).

4.14.1.3. Proline hydroxylation sites

Potentially the most interesting observation from the PTM discovery data is the identification of HIF-2α phosphorylation sites in 21% O_2 that neighbour all of proline hydroxylation sites, including a potentially novel site (P576). The proline hydroxylation sites are within highly conserved domains between HIFα isoforms, and consists of an LAP motif. However, HIF-2α has single amino acid variations in both of the canonical proline

hydroxylation sites that are phosphorylated: (HIF-1α/HIF-2α), P402-A403/P405-T406 and P564-M561/P531-T528. Therefore it is feasible that the size and charge differences of HIF-2α phosphorylated threonine residues, to HIF-1α alanine and methionine residues, could explain the 21% O_2 protein stability observed by HIF-2α; due to blocking of PHD binding, thus removing the hydorxylation and degradation pathway. This would be fairly easy to investigate through mutational analysis and western blotting techniques. However, as mentioned, shotgun proteomics approaches mean it is undeterminable whether each site co-exists or are mutually exclusive, therefore further investigation is required such as by 'Middle-down' approaches.

4.14.2. Binding partner analysis

The binding partner discovery work presented here will be highly insightful for a targeted/hypothesis driven approach. The discovery analysis performed by this study has generated a huge quantity of novel data for potential binding partners that may regulate HIFα isoforms both O_2 dependently and independently. However, IP makes it is impossible to determine whether these identified binding partners are the result of direct or indirect interactions to the HIFα protein. Thus, without a specific characteristic/hypothesis to filter identified proteins, data analysis is difficult to draw meaningful conclusions from. Future studies investigating binding partners could easily adopt a cross linking and MS analysis strategy to identify peptides that crosslink, and thus must interact (Sutherland *et al.*, 2008), to determine direct and indirect binding partners. Additionally, the data presented here could be used in network mapping to investigate how proteins interact with each other through known interactions (Szklarczyk *et al.*, 2015, Szklarczyk *et al.*, 2017 & Shannon *et al.*, 2003). Global categorisation, through GO annotation enrichment, and O_2 depenecy of binding partners was investigated in this study.

4.14.2.1. GO annotation enrichment

Although it is clear that O_2 tension has a significant effect on the binding partner profile of both HIFα isoforms, with ~66% of all HIF-1α proteins and a >10 fold increase of HIF-2α proteins, GO annotation enrichment analysis was relatively unuseful. The GO annotations categorise proteins broadly, thus a single protein is part of multiple GO terms. Hence if a specfic protein complex is significantly enriched it can result in multiple different GO annotations that consist primarily of identical proteins, as observed here with proteasomal proteins. Combined with the strict filtering applied, the vast majority of GO annotations that were plotted for Figure 4.18 to Figure 4.21 are identical, yet categorically contain

different proteins due to analysis of 21% or 1% O_2 only binding partners. Hence, although useful for categorising proteins based on function/pathway, manually mining all GO annotations (both significant and non-significant) for specific queries will be of greater use, as done for kinases and mitochondrial proteins here.

4.14.2.2. O_2 dependent protein binding

Two separate approaches were applied to identify binding partner differences induced by O_2 tension: an MS/MS driven identification and filtering approach (Proteome Discoverer) and a label free quantification (LFQ) approach (MaxQuant). Both strategies have their own advantages and disadvantages, with Proteome discoverer providing greater confidence in O_2 tension classification, by requiring MS/MS and determination of primary sequence, while MaxQuant is able to analyse the overlap group (both O_2 tensions) for statistically valid fold changes. Therefore, further analysis of data can be done in combination of both O_2 dependent outputs and GO annotations when extracting information. This identified some interesting observations.

4.14.2.3. Mitochondrial proteins

It is clear from our data that there is significant enrichment of mitochondrial proteins for both HIFα isoforms. It could be argued that upon organelle lysis mitochondrial proteins are Co-IP'd in a HIFα specific manner, hence not removed by background subtraction although endogenously do not interact, thus are artefactual in nature. However, recent publications have shown that HIF-1α may have a sub-population of total protein localised into these sub-cellular compartments, although functional aspects are unknown. Therefore, in combination of data provided here, it will be interesting to see how the field develops. Since many transporter proteins into mitochondria are enriched in 1% O_2, it could be theorised that HIFα is transported into the mitochondria to upregulate genes of the oxidative phosphorylation pathway in an attempt to maximise the efficiency of the limited available O_2.

4.14.2.4. Kinases

As a specific focus of this chapter was on phosphorylation, we investigated bound protein kinases. However, as stated previously, it is unknown whether these directly bind HIFα or are secondary indirect binding partners. Due to the transient nature of kinase function, it is likely that many kinases that phosphoylate HIFα will not have been identified. The majority of kinases identified are novel HIFα interactors and are generally cell cycle regulators,

suggesting an important role of HIF in correct cell cycle progression. It is known that HIF is important for cell cycle progression (Goda *et al.*, 2003, Koshiji *et al.*, 2004, Kaidi *et al.*, 2007, & Hubbi *et al.*, 2013), however the exact mechanism of regulation is unknown. Therefore, the identification of many cell cycle dependent kinases could aid in unravelling this system.

4.14.2.5. PHDs

Considering the primary regulatory pathway of HIFα proteins, through proline hydroxylation, it is highly interesting that we identify novel interactions with additional proline hydroxylase enzymes (P4HA1, P4HA2, P4HB, P3H1, P3H2 & P3H3). The biological significance of this is unknown, especially considering they are ER/golgi specific localised, however the identification of novel proline hydroxylation events in HIF-2α (P137 and P206) could suggest a functional importance.

4.14.3. Concluding remarks

Overall, this chapter contains a significant quantity of novel data for HIF-1α and HIF-2α O_2 dependent and independent regulation, through PTM and binding partner based regulatory pathways. However, it is important to consider that the HIF overexpression used might have resulted in some artefactual data. We strongly argue that this dataset is the most biologically relevant yet, compared to published studies which use fragment based approaches, *in vitro* assays or HIFα proline mutants that are oxygen insensitive, bypassing the main regulatory pathway. With the identification of these vast numbers of PTMs and binding partners future studies will be needed to validate the findings on endogenous HIFα proteins (if suitable antibodies are available), potentially in combination with targeted MS techniques (that can result in 100 fold increases to sensitivity). Future studies could also investigate how PTM status may change as a function of the severity of the hypoxic conditions within the same cell line, or different cell lines, to investigate the full complexity of HIFα regulation.

5. Chapter 5: PTM functional characterisation

5.1. Introduction:

Following my discovery of multiple new PTMs, the next step is to functionally characterise some of these potentially interesting PTMs, determined from the analyses in Chapter 4, to investigate their roles as HIF regulators. Since there was a specific focus on phosphorylation, and robust statistical analysis was performed to aid in the confident localisation of this PTM, sites of phosphorylation were chosen for initial characterisation experiments.

Because phosphorylation is generally dynamic and the regulators that promote/inhibit phosphorylation of these identified sites are currently unknown, it is difficult to investigate the role of the PTM of interest endogenously. Additionally, phosphorylation is likely to occur at a stoichiometric ratio below the levels of total protein (Mann *et al.*, 2002), thus complicating the functional analysis and compromising the ability to draw conclusions. Site directed mutagenesis (SDM) is a method used to introduce a site-specific mutation into DNA that can result in an amino acid change. Using this approach, it is possible to mutate a PTM site into residues that mimic or prevent phosphorylation. Figure 5.1 shows an example of serine phosphorylation, which introduces a negative charge, hence mutation to aspartic acid (a permanently negatively charged amino acid) can mimic the functional characteristics of the protein in a completely phosphorylated state. Conversely, mutation of the serine to an alanine removes the potential for phosphorylation at a specific site, hence allowing investigations of the protein in a completely unphosphorylated state.

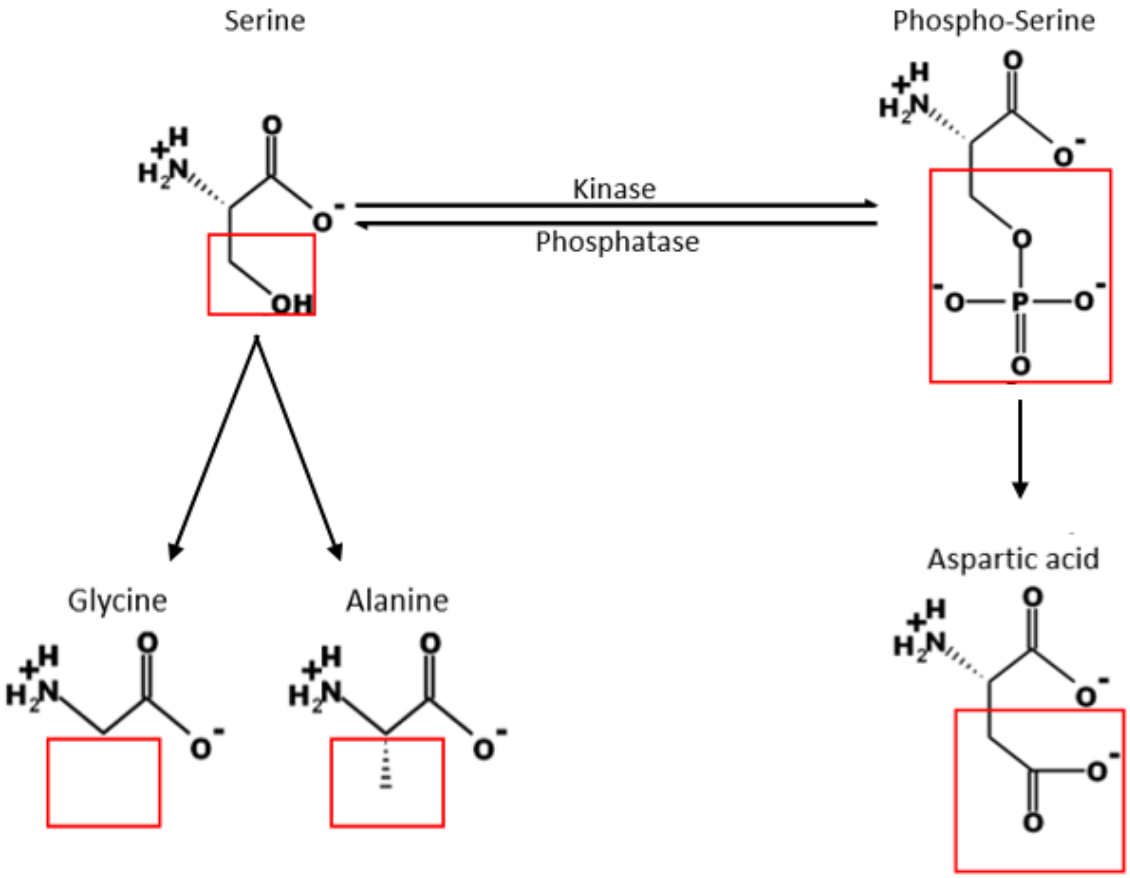


Figure 5.1: Schematic view of Serine phosphorylation and phospho –mimetic and –null mutations.
All amino acids are labelled and boxed region highlights the functional group of amino acids.

5.2. Aims:

The aims of this chapter was to functionally characterise a selection of phosphorylation sites identified from Chapter 4. A site-directed mutagenesis (SDM) protocol had to be implemented, which could reliably mutate single sites while limiting the risk of random mutations through PCR amplification. Once the SDM protocol was optimised, mutational strategy used the traditional phospho -mimetic (serine->aspartic acid, threonine->glutamic acid) and -null (serine/threonine->alanine) mutations, unless there was a guided mutation inferred from evolutionary analysis (Chapter 4). Analogous mutational analysis of HIFα is a common strategy to investigate regulatory mechanisms (Table 1.2 & Table 1.3, example Mylonis *et al.*, 2008). Once mutants were created, an initial functional screening of mutants was performed through a HIF-dependent luciferase-based transcriptional assay. We further characterised, Serine-31 (S31) mutations by a range of biochemical techniques to identify their regulatory roles.

5.3. Site Directed Mutagenesis (SDM) optimisation:

SDM is a method used to mutate specific DNA bases, which can result in an amino acid substitution in the protein sequence (missense mutation). The most common SDM technique uses overlapping primers that contain the mutation, thus creating full-length mutant plasmid in a single PCR reaction (Quikchange, Agilent, Liu *et al.*, 2008). However, even the highest fidelity polymerases can result in random mutagenesis. A random mutation rate of ~0.1% is known for the KOD polymerase used here (McInerney *et al.*, 2014 & Manufacturers notes). Thus, PCR amplification of the whole HA-Clover-HIFα plasmids (~9 kBp) could potentially result in 9 random mutations per SDM experiment. These mutations are incorporated randomly within the plasmid, therefore it is likely that different random mutations will occur for each different SDM plasmid. Problematically, the random mutagenesis could occur within the plasmid regulatory regions, such as the strong CMV promoter where single base pair mutations can result in large gene expression level changes (Alper *et al.*, 2005). Thus, random mutagenesis could significantly affect protein expression levels, resulting in the false conclusion that the mutation of interest has a protein stability role. Hence, this SDM strategy would require sequencing of the whole plasmid to ensure that only the desired mutation is present.

MEGAprimer is a SDM technique that limits the PCR reaction length to the gene of interest only, depicted in Figure 2.4. Initially, a primer in the flanking gene region and a mutagenic primer are used in a PCR reaction to create a mutated DNA fragment (the MEGAprimer). The mutated MEGAprimer is then used with a second flanking primer to amplify the full length gene, which is then reintegrated into the vector backbone by restriction digest and ligation (Sarkar *et al.*, 1990). We created flanking primers that allow SDM independently of HA-Clover-HIFα gene, but rather dependent on the mutagenic primer created. Because the vector backbone comes from the WT digested plasmid there is no risk of random mutagenesis between SDM plasmids outside of the PCR fragment, hence sequencing is only required of the HIFα gene (~2.5 kBp).

Using the optimised SDM protocol, I made phospho –mimetic and –null mutations for HIF-1α: S31 and S786, and HIF-2α: S345, T528 and S581. For serine and threonine phosphorylation, mimetic mutations to aspartic acid and glutamic acid were used respectively, due to the extra carbon in the side chain of threonine and glutamic acid to

mimic space-charge effects. For all phosphorylation null mutations, except S31, Alanine was used as a non-charged nor modifiable analogue. In the case of S31, evolutionary analysis identified that all bony fish (*Osteichthyes*) had a non-phosphorylatable glycine residue at position 31 (Chapter 4, section 4.7.1.1), thus serine 31 was mutated to glycine. Once cloned, all plasmids were sequenced through the Clover and HIFα genes to ensure that only the mutation of interest was present.

5.4. HIF dependent luciferase assay

Due to the number of mutants and conditions to be tested (2 different mutants for all 5 sites selected, plus WT and endogenous controls at 1% O_2 and 21% O_2) we initially used a HTP HIF dependent luciferase assay to measure transcriptional function. As HIFα proteins are canonically regulated by stability, it is important to note that this assay cannot provide information on whether the transcriptional effects observed are due to alteration of transcriptional activity, transactivation or protein stability. The assay uses a plasmid encoding the luciferase gene under the regulation of three HRE repeats (HRE-Luciferase, Coulet *et al.*, 2003). Therefore, the luciferase gene is only transcribed and translated into protein if active HIF dimers are present. Thus, luminometry readings of the mutants, in comparison to the WT counterpart, allows an initial assessment of the functional effects induced by a mutation.

Cells were transfected as described for HRE-luciferase assays using the 'low' expression level model designed to maintain O_2 dependent regulation (Chapter 3). As phosphorylation sites under investigation were identified at a 4 hr time point in either 1% or 21% O_2, these conditions were repeated for initial screening (Figure 5.2).

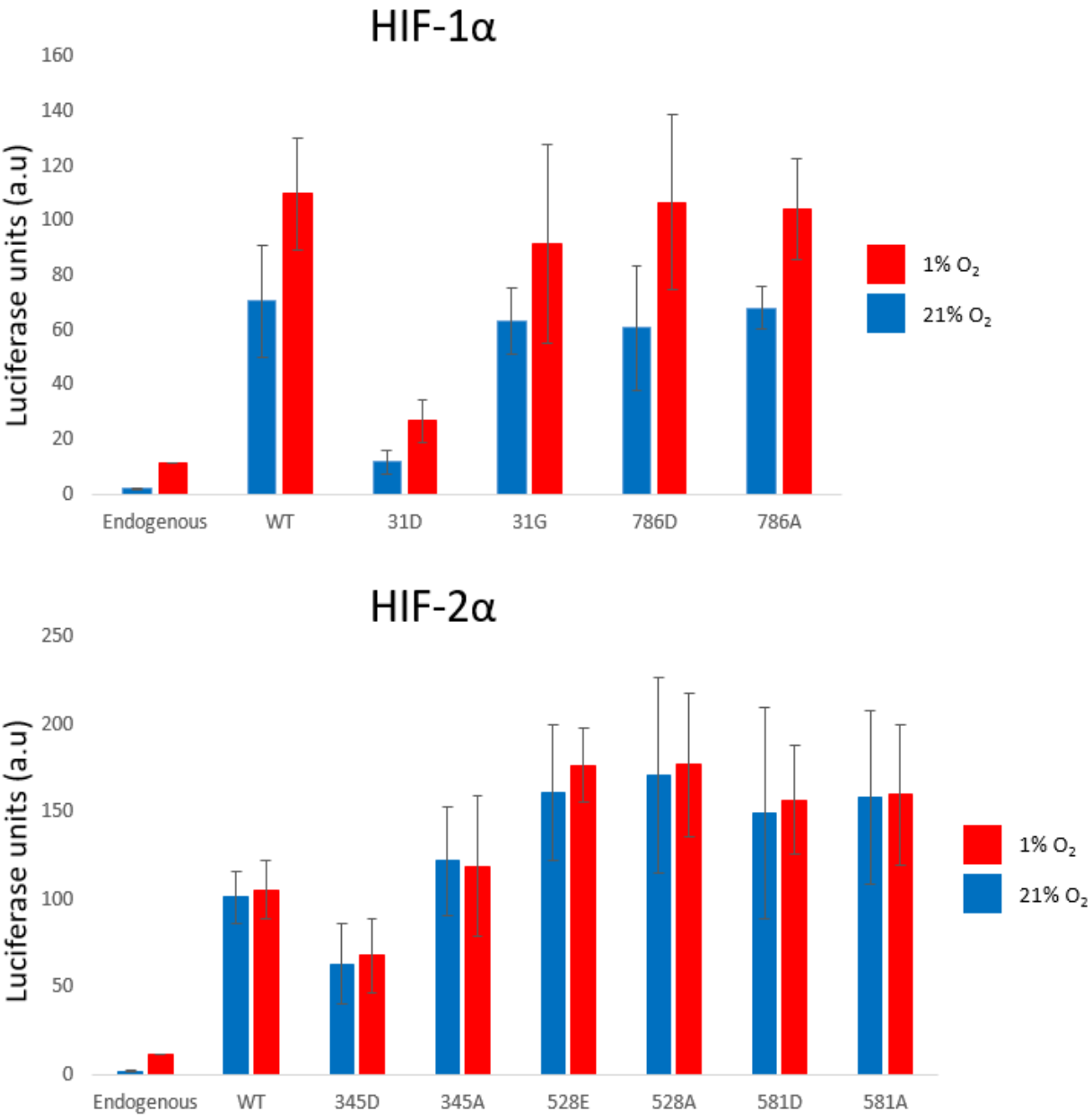


Figure 5.2: Luciferase transcriptional assay of HIFα SDM mutants.
Cells were co-transfected with 0.5 µg HRE-Luciferase plasmid (Addgene #26731) alongside 50 ng of the stated HA-Clover-HIFα plasmid (low expression level model) per 200,000 cells in a 35 mm plate. Cells were incubated with transfection mix for 18 hr before media change and incubating in 1% or 21% O_2 for 4 hr. Cells were then lysed and luminometry readings taken using a BMG Labtech FLUOstar Omega plate reader. Readings were normalised against luciferase working solution only readings. Data present for n=2 biological replicates +/- standard deviation.

It is important to consider that for endogenous readings it is impossible to distinguish the relative effects of HIF-1α and HIF-2α without using knockdown controls. These readings were thus primarily used to ensure the assay and hypoxic incubation were working correctly. A clear observation from Figure 5.2 is that hypoxia induced an ~1.5 fold transcriptional increase of all HA-Clover-HIF-1α plasmids, whilst on HA-Clover-HIF-2α plasmids, hypoxia had a very mild transcriptional induction. For comparison, endogenous

HIF had an ~5 fold luciferase signal increase induced at 1% O_2. The weak hypoxic induction of HA-Clover-HIF-2α plasmids could be a result of the low exogenous expression model saturating HRE-luciferase binding sites (similar to HA-Clover-HIF-1α at a high expression level Figure 3.12). However, this explanation is unlikely since WT readings for HA-Clover-HIF-1α and HA-Clover-HIF-2α are <1.1 fold different, and a ~1.8 fold greater response is observed with HA-Clover-HIF-2α T528 mutants. Alternatively, this could suggest that the lack of hypoxic induction in early hypoxia could reflect HIF-2α specific characteristics, such as HIF-2α being known to become active in prolonged hypoxia (as reviewed by Koh *et al.*, 2012).

Because this experiment was for preliminary screening to identify a specific mutant set to further focus on, only 2 repeats were performed, thus statistical analysis could not be done. From Figure 5.2, the HIF-1α S786 mutations had very little effect on transcriptional function, mimicking the WT control within a 1.15 fold difference at both O_2 tensions. Conversely, the S31D phospho-mimetic mutation of HIF-1α dramatically reduced the transcriptional activity by ~5 fold, compared to the WT HIF-1α protein. However, the S31G phospho-null mutation is <1.2 –fold different to the WT HIF-1α controls. These data suggest that S31 phosphorylation inhibits the transcriptional activity of HIF-1α. Based on structural analysis (Figure 4.4), we hypothesised that this transcriptional inhibition could be due to a reduced ability of HIF-1α to bind DNA in a phosphorylated state, likely due to charge repulsion with the DNA backbone.

A similar, but less pronounced, decrease in luciferase signal was observed with the HIF-2α S345D phospho-mimetic mutation compared to the WT HIF-2α (Figure 5.2). An ~40% decrease in luciferase signal was observed for S345D in both O_2 conditions (Figure 5.2), suggesting a possible analogous role of phosphorylation at HIF-2α S345 to HIF-1α S31 in causing the decrease in transcriptional activity. However, based on structural analysis (Figure 4.8) we hypothesised that HIF-2α S345 phosphorylation might regulate transcriptional activity by the reduced ability to bind HIF-1β.

Considering that the T528 and S581 mutations were selected based on their close proximity to the known proline hydroxylation site (P531) and a potentially novel hydroxylation site (P576), it is interesting that the mutation of these sites to either their phospho-mimetic (glutamic acid and aspartic acid respectively) or phospho-null (alanine) mutations both resulted in an increase in transcriptional activity by ~50%. Our initial hypothesis was that

phosphorylation at a site in such close proximity to the proline hydroxylation sites may block the function of PHD enzymes, thus increasing protein stability and luciferase signal. However, these data could suggest a dual role for the hydroxyl group of serine/threonine residues, for example, the unmodified hydroxyl group of the serine/threonine could be required for effective targeting of PHDs and subsequent proline hydroxylation. Hence, mutation to the phospho-null alanine residue also inhibits this degradation of HIF-2α, thus increasing luciferase signal. Where phosphorylation (aspartic acid/glutamic acid) prevents PHD binding, thus also increasing luciferase signal

5.5. Exploring the mechanism of HIF-1α S31 phosphorylation-induced transcriptional inhibition

As stated, the luciferase assay, does not allow the determination of whether the observed results are because of direct transcriptional effects, or through alteration of protein stability or change in binding partners/transactivation. Therefore, we first determined whether S31 mutations affected protein stability Figure 5.3 (A)). Figure 5.3 (A) shows that there is no significant difference in protein stability between WT, S31G and S31D HA-Clover-HIF-1α. Normalisation against the β-actin control shows that the S31D and S31D mutations are ~1.4 fold and ~1.2 fold more intense than the WT protein (based on densitometry analysis, data not shown). Hence the transcriptional inhibition of S31D is not due to a decrease in protein level.

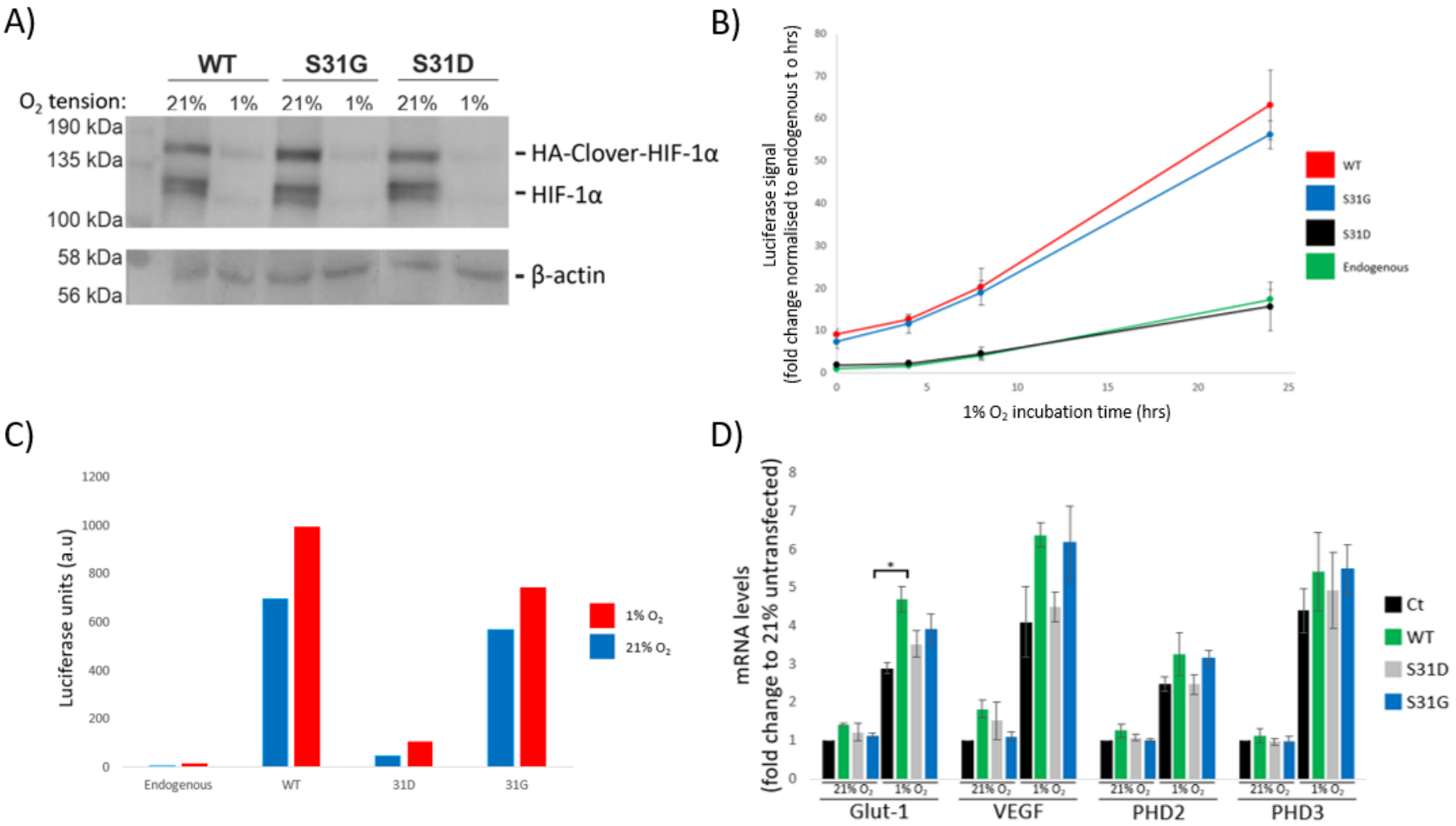


Figure 5.3: Biochemical assessment of WT HA-Clover-HIF-1α and S31 phospho –mimetic (S31D) and -null (S31G) mutations.
A) Western blot analysis of cells transfected with the stated plasmid with the low expression method and incubated for 4 hrs at indicated O_2 tensions. Western blots were probed with anti-HIF-1α and β-actin antibodies, transfected HA-Clover-HIF-1α and endogenous HIF-1α are labelled, image representative of n=2 biological replicates. B) Time course luciferase activity of cells transfected with indicated plasmid using the low expression method and incubated for 4, 8 or 24 hrs in 1% O_2. All samples were lysed simultaneously. Data from n=3 biological replicates C) Luciferase assay of cells transfected with stated plasmid using the high expression method and incubated for 4 hrs at indicated O_2 tension, normalised to luciferase working solution only readings. D) RT-qPCR of 4 known hypoxia regulated target genes (GLUT-1, PHD2, VEGF and PHD3). Cells were transfected at the low expression model level with stated plasmid and incubated in stated O_2 tension for 24 hrs before lysing and RNA extraction. Data measured using a LightCycler480 mRNA levels are normalised to cyclophillinA mRNA, +/- standard error. Data presented of n=3 biological replicates. A 2-way ANOVA analysis, with paired-comparisons, was performed. * = P-value <0.05.

To check if transcriptional efficiency is chronically reduced by S31D mutation, rather than a transient effect, two strategies were employed: 1) A low expression model time course over 24 hr of 1% O_2 incubation (Figure 5.3 (B)) and 2) a large overexpression using the high expression model (Figure 5.3 (C)). Both Figure 5.3 (B) and (C) show conclusively that S31D dramatically reduces the luciferase signal, to levels comparative to the non-transfected endogenous control, while the S31G mimics the WT overexpression profile. For comparison, after 24 hr incubation at 1% O_2 using the low expression levels, the HIF-1α WT and S31G proteins were ~5 fold greater, and significantly different, to the S31D mutation. The S31D mutant does not significantly differ from the endogenous control. Using the high expression level, the HIF-1α WT and S31G proteins resulted in an >50 fold increase in luciferase response compared to the S31D. To ensure the S31D mutation had a physiological effect on endogenous target gene expression, Real-Time quantitative PCR (RT-qPCR) was performed on the following hypoxia inducible genes: GLUT-1, PHD2 (EGLN1), VEGF and PHD3 (EGLN3) (Figure 5.3 (D)). Figure 5.3 (D) shows that the HIF-1α S31D mutation results in a statistically significant reduction in expression of both GLUT-1 and VEGF in comparison to the WT control (P-value <0.05), to levels that are comparable of the non-transfected control. However, no significant difference was observed in the mRNA levels of PHD2 and PHD3 with HIF-1α S31D compared to the WT control.

Overall, the data presented here suggest that the phosphorylation of HIF-1α at position S31, inferred from S31D mutation, completely inhibits the transcriptional function of HIF-1α in a protein stability independent manner. However, the potential mechanism for the transcriptional inhibition is unknown and could be explained by: 1) the impairment of binding to known essential cofactors, such as HIF-1β. 2) The impairment of binding to a novel binding partner. 3) Differential localisation, such as nuclear exclusion, or 4) direct DNA binding inhibition through charge repulsion. These scenarios are explored below.

5.5.1. Investigating S31 binding partners

To investigate how the binding partners were altered by S31 phosphorylation (exploring both scenarios (1) and (2) above), the WT, S31D or S31G plasmids were expressed at the low expression level and immunoprecipitated for analysis of their binding partners(Figure 5.4), as in Chapter 4.

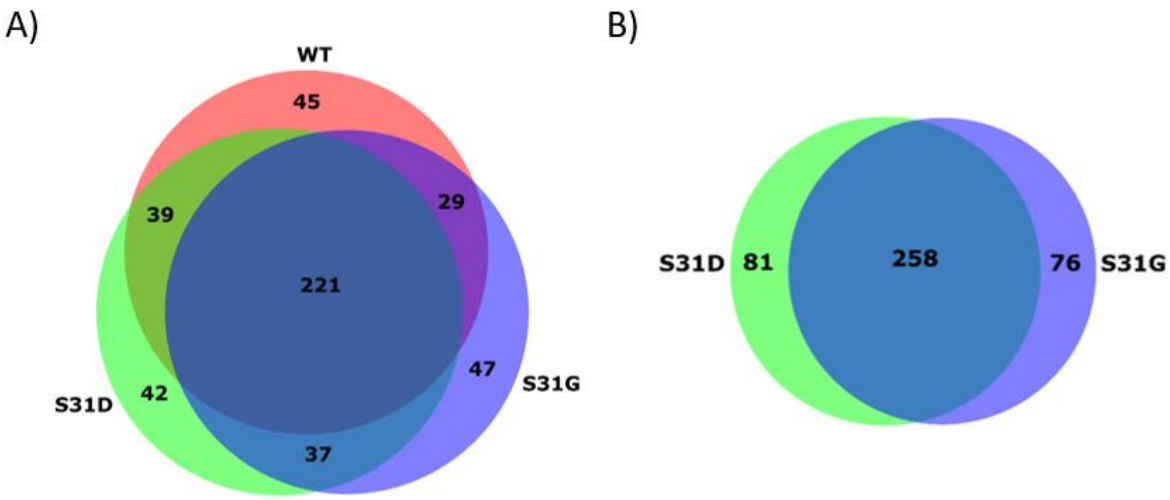


Figure 5.4: Visual depiction of binding partner differences between HA-Clover-HIF-1α WT, S31D and S31G. IPs were performed using HeLa cells transfected with the low level expression model of indicated plasmid. Data represents binding partners only identified in both replicates post background subtraction. Numbers reflect total number of proteins in each segment, circles and overlaps scaled respective to number. A) Venn diagram of WT, S31D and S31G binding partners. B) Venn diagram of S31D and S31G binding partners.

As Figure 5.4 (A) shows, ~70% of the total number of proteins identified were common between the WT, S31D and S31G HIF-1α proteins. A list of protein identifications for each HIF-1α protein is provided in appendix 5. Importantly, HIF-1β was identified as a binding partner for all HIF-1α mutants. Similar to the experiments presented in chapter 4, FIH-1 and CBP/p300 were not identified in any sample. Thus, it does not appear that the reduced transcriptional function of S31D is the result of failing to form the active transcription complex.

Therefore, the data was analysed for potentially novel binding partner changes induced by phosphorylation. It is highly likely that the expressed WT HIF-1α protein will be in an unknown state of equilibrium between S31 phosphorylated and unphosphorylated protein forms. Thus, if S31 phosphorylation is responsible for promoting, or inhibiting, the interaction with an essential unknown binding partner, the WT dataset would preclude its identification. Hence a focused analysis was performed, comparing HIF-1α S31D and S31G binding partners only (Figure 5.4 (B)). Theoretically, direct comparison of the S31D and S31G will identify binding partner differences arising as a function of phosphorylation, allowing me to hypothesise the functional role played by phosphorylation. A list of all mutant specific proteins is presented in Table 5.1.

Table 5.1: Binding partner differences between S31D and S31G.
Identifications were present in two biological replicates post background subtraction, using HeLa cells transfected with indicated plasmids at low expression levels upon 4 hr incubation at 1% O_2.

S31D			
Accession	Gene name	Accession	Gene name
P62820	RAB1A	A8MWD9	SNRPGP15
P05023	ATP1A1	P11166	SLC2A1
P07900	HSP90AA1	P14324	FDPS
Q6KB66	KRT80	Q9P258	RCC2
O00148	DDX39A	Q9UJY1	HSPB8
P33316	DUT	Q9H2P0	ADNP
P01891	HLA-A	Q08211	DHX9
Q96P70	IPO9	P62195	PSMC5
P50454	SERPINH1	Q96BM9	ARL8A
P06858	LPL	Q99714	HSD17B10
P07737	PFN1	P62888	RPL30
P17844	DDX5	Q96HS1	PGAM5
P15924	DSP	Q14558	PRPSAP1
P20340	RAB6A	Q9HA64	FN3KRP
P39019	RPS19	Q9UBB4	ATXN10
Q9H9B4	SFXN1	P31153	MAT2A
O60701	UGDH	Q16795	NDUFA9
P30041	PRDX6	Q6NXT2	H3F3C
Q8NC51	SERBP1	Q8TEQ6	GEMIN5
P55072	VCP	P60842	EIF4A1
O00425	IGF2BP3	Q04637	EIF4G1
P04083	ANXA1	Q969S3	ZNF622
O75369	FLNB	P12004	PCNA
P63104	YWHAZ	P00403	MT-CO2
P27824	CANX	Q92499	DDX1
P61289	PSME3	Q96AE4	FUBP1
P27105	STOM	Q7Z2W4	ZC3HAV1
O43896	KIF1C	Q99805	TM9SF2
Q9H3K6	BOLA2	P46977	STT3A
P07237	P4HB	Q9NX63	CHCHD3
Q9BWF3	RBM4	Q9UBV8	PEF1
P35268	RPL22	P51991	HNRNPA3
P22061	PCMT1	Q9UI30	TRMT112
Q14498	RBM39	P40938	RFC3
P46459	NSF	P55884	EIF3B
O75643	SNRNP200	P62633	CNBP
O43684	BUB3	Q92973	TNPO1
Q08554	DSC1	Q14697	GANAB
P21912	SDHB	Q9Y277	VDAC3
O96008	TOMM40	Q13123	IK

S31G			
Accession	Gene name	Accession	Gene name
P04259	KRT6B	P29692	EEF1D
Q13509	TUBB3	P48047	ATP5PO
P52597	HNRNPF	P41252	IARS
P08134	RHOC	O60814	HIST1H2BK
Q9H0U4	RAB1B	Q86VI3	IQGAP3
Q13637	RAB32	P43243	MATR3
Q9Y305	ACOT9	Q9Y2L1	DIS3
Q14257	RCN2	P48556	PSMD8
O75396	SEC22B	P24666	ACP1
P61026	RAB10	O14949	UQCRQ
Q96TA2	YME1L1	P61254	RPL26
P53621	COPA	P56385	ATP5ME
Q9BWM7	SFXN3	P61009	SPCS3
Q9UBX3	SLC25A10	P47756	CAPZB
P51148	RAB5C	Q6UXV4	APOOL
P08865	RPSA	P62191	PSMC1
O94905	ERLIN2	P01111	NRAS
P83731	RPL24	P53007	SLC25A1
P67809	YBX1	Q96CS3	FAF2
Q63ZY3	KANK2	P14625	HSP90B1
P23258	TUBG1	P62913	RPL11
P63151	PPP2R2A	P62070	RRAS2
P25398	RPS12	P63173	RPL38
P11717	IGF2R	P26640	VARS
Q8N726	CDKN2A	P51665	PSMD7
P50914	RPL14	P25789	PSMA4
Q99436	PSMB7	Q13895	BYSL
O60762	DPM1	P48507	GCLM
P53985	SLC16A1	Q969Z3	MARC2
Q15233	NONO	Q9NUQ8	ABCF3
P60953	CDC42	Q04941	PLP2
P05388	RPLP0	O95299	NDUFA10
P62917	RPL8	O15260	SURF4
Q96HC4	PDLIM5	O43837	IDH3B
P35030	PRSS3	Q92616	GCN1
Q04837	SSBP1		
P62318	SNRPD3		
Q96PK6	RBM14		
Q27J81	INF2		
Q15392	DHCR24		

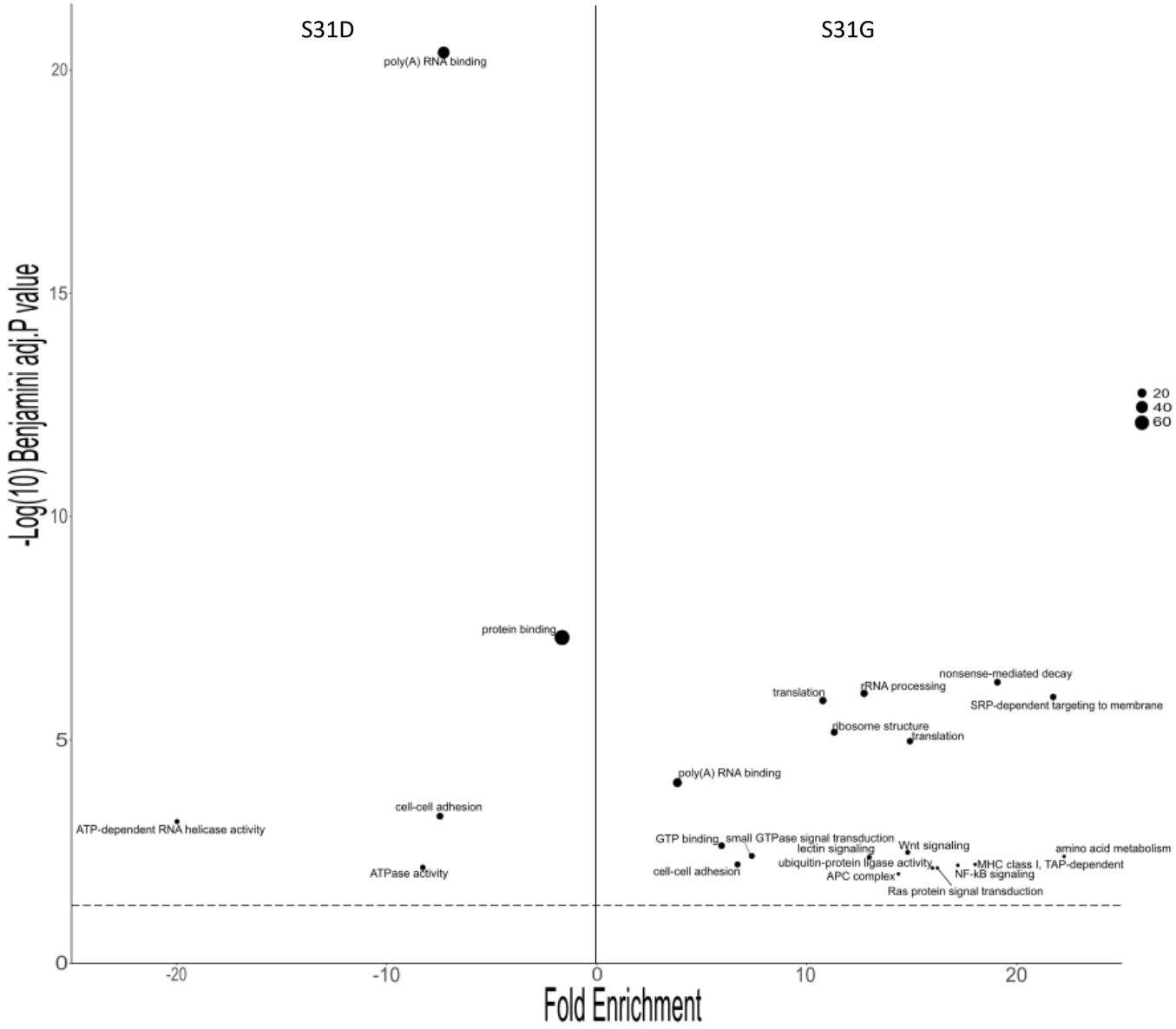


Figure 5.5: GO pathway enrichment analysis of S31D vs S31G.
Identities from Table 5.1 used in DAVID GO annotation analysis, Biological process and Molecular function annotations filtered for only. A Fishers exact test was performed, plotted annotations are of a Benjamini-Hochberg adjusted P-Value < 0.01. Left hand side = S31D only proteins identified, right hand side = S31G only proteins identified.

Table 5.1 contains >150 different proteins. To identify functional protein categories and signalling pathways associated with these unique binding partners, the data were analysed through Gene Ontology enrichment analysis using DAVID (Figure 5.5, Dennis *et al.*, 2003), as performed in chapter 4. Figure 5.5 shows that more functional categories were significantly enriched for in S31G IPs. However further investigation found that multiple annotations consisted of the same proteins and were categorically labelled as different GO terms, similar to previously discussed in Chapter 4. Interestingly, S31G has three significantly identified annotations for GTP related proteins (GTP binding, small GTPase signal transduction and Ras protein signal transduction) equating to 10 different proteins

(Table 5.2). These GO annotations, and related proteins, were not identified in S31D even when relaxing stringency of confidence. Thus, this data suggests that unphosphorylated S31 (S31G) is associated with a GTP signalling cascade that is essential to trigger HIF-1α dependent transcription, and is blocked by S31 phosphorylation (S31D).

Table 5.2: Identified GTP binding proteins that bind specifically to S31G, and not S31D.
Table includes the UniProt accession code, Gene Name and description of protein.

Accession	Gene Name	Description
P61026	RAB10	Ras-related protein Rab-10
Q9H0U4	RAB1B	Ras-related protein Rab-1B
Q13637	RAB32	Ras-related protein Rab-32
P51148	RAB5C	Ras-related protein Rab-5C
P60953	CDC42	Cell division control protein 42 homolog
P01111	NRAS	GTPase NRas (Transforming protein N-Ras)
P08134	RHOC	Rho-related GTP-binding protein RhoC
P62070	RRAS2	Ras-related protein R-Ras2
Q13509	TUBB3	Tubulin beta-3 chain
P23258	TUBG1	Tubulin gamma-1 chain

The downstream regulation post Ras signalling (reviewed by: Downward, 2003) includes many protein kinases known to influence HIF-1α activity, including ERKs (Minet *et al.*, 2000), GSK3β (Flügel *et al.*, 2012) and mTOR (Land *et al.*, 2007). Therefore, it is possible that a Ras-dependent pathway promotes HIF-1α dependent transcription through secondary phosphorylation sites. This possibility was not explored due to the poor sequence coverage by MS analysis in the neighbouring region of HIF-1α, as discussed in Chapter 4. Additionally, it is possible that the Ras-dependent signalling pathway that interacts with HIF-1α is indirect and results in the modification or transactivation of a novel HIF-1α binding partner, in turn promoting HIF signalling; these options remain to be explored.

Although the binding partners comparison, between S31G and S31D mutants, revealed some interesting potential upstream signalling mechanisms, it did not provide a definitive explanation for the functional mechanism underpinning the transcriptional inhibition induced by a phospho-mimetic mutant of S31 phosphorylation (S31D). The role of S31 phosphorylation in HIF-1α nuclear localisation and direct DNA binding were further explored.

5.5.2. S31 phosphorylation and Nuclear localisation

The third hypothesis was explored by confocal microscopy to visualise the correct nuclear localisation (Figure 5.6 (A)). For nuclear localisation studies, the high expression level of HA-Clover-HIFα was used for WT, S31D and S31G plasmids. Figure 5.6 (A) shows that, whilst highly overexpressed, the WT HIF-1α protein remained exclusively expressed in the nucleus as a homogenous distribution. The same observation was observed for both HIF-1α S31D and S31G mutants, thus mis-localisation of the S31D phospho-mimetic mutant is not responsible for the transcriptional inhibition observed, under these conditions.

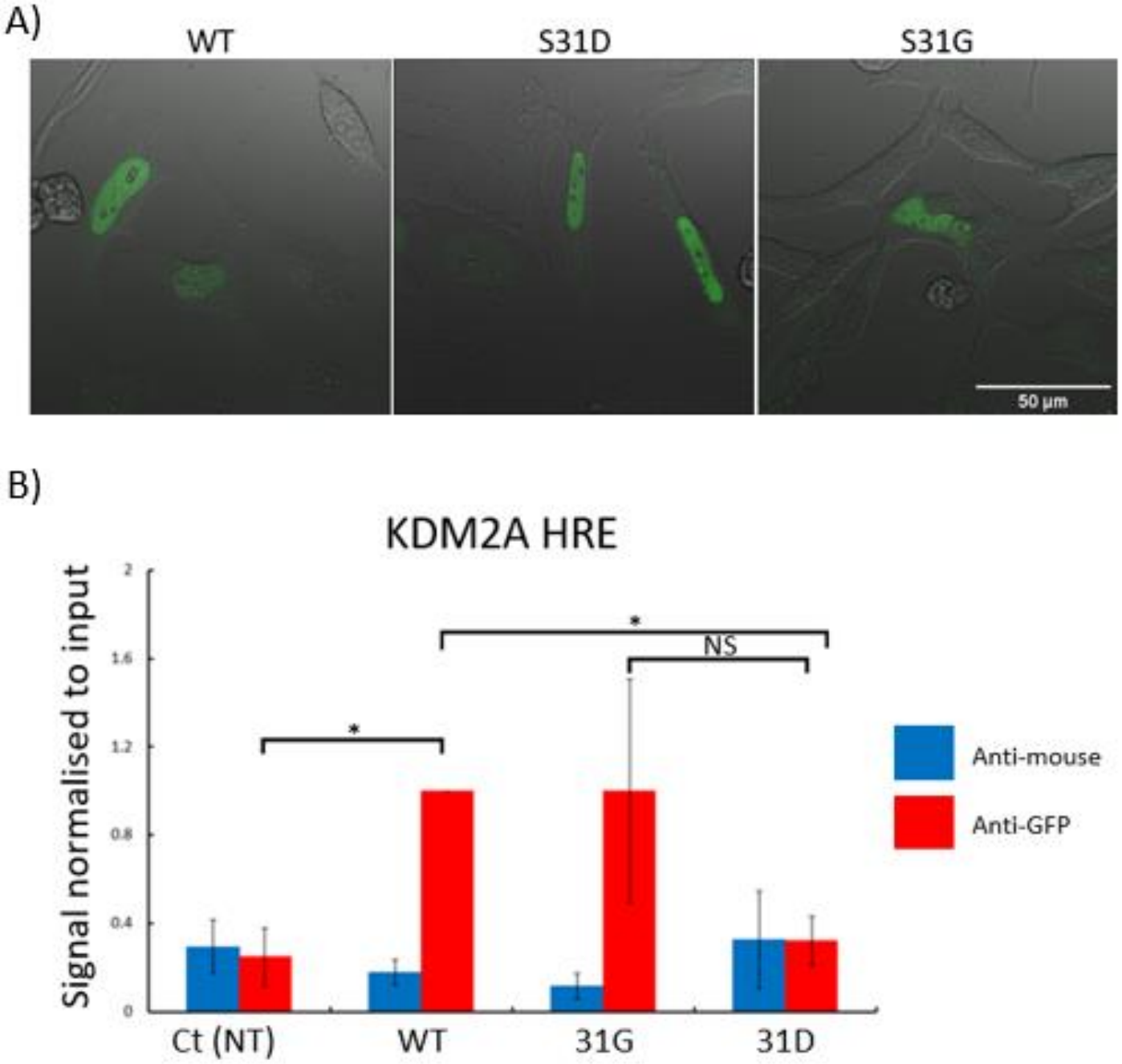


Figure 5.6: Nuclear localisation and DNA binding efficiency characteristics of HIF-1α S31 WT, S31D and S31G. HeLa cells transfected with either WT, S31G or S31D HA-Clover-HIF-1α plasmids at the high expression level. A) Nuclear localisation analysis, Microscopy analysis at 63X magnification with images taken at brightfield and 488 nm fluorescence excitation using a Zeiss LSM 780 microscope. B) DNA binding efficiency, determined through CHIP and RT-qPCR analysis of the KDM2A HRE promoter. Ct(NT) = Negative control using the HA-Clover only plasmid. A Student t-test was performed, * = P-value < 0.01, NS = not significant. Data of n=3 biological replicates +/- standard error. Performed by Dr Michael Batie, from the Prof Rocha group.

5.5.3. Direct DNA binding inhibition

The ability of WT, S31D and S31G HA-Clover-HIF-1α proteins to bind DNA was measured by Chromatin Immunoprecipitation (CHIP) (Figure 5.6 (B)), experiments performed by Dr Michael Batie, from the Prof Rocha group). The high expression level of HA-Clover-HIF-1α was used to exacerbate functional differences observed, because it is known that the S31D

mutant protein maintains its transcriptional inhibition at high expression levels (Figure 5.3 (C)).

Analysis of the KDMA2 HRE promoter was used as a well-known HIF-1α regulated promoter (Batie *et al.*, 2017), which works well for HIF-CHIP experiments. CHIP analysis (Figure 5.6 (B)) shows that the WT HA-Clover-HIF-1α resulted in a significant ~4 fold greater DNA binding signal at the KDMA2 promoter than compared to the HA-Clover only (Ct(NT)), and anti-mouse, off-target antibody negative controls ($P<0.01$). The S31G mutant shows a similar DNA binding pattern as the WT, consistent with the luminometry and qPCR experiments, but exhibited large variability (Figure 5.6 (B)). The S31D mutation resulted in almost identical DNA binding signal to the negative control, and was significantly different to the WT plasmid ($P<0.01$). These data are consistent with the luciferase reporter assays and RT-qPCR data. Additionally, the DNA binding with HIF-1α S31D was not significantly different to either of the negative controls performed (HA-Clover only and anti-mouse IPs), highlighting how little DNA was bound to HIF-1α S31D.

Overall, these data suggest that phosphorylation at S31, as determined using the S31D phospho-mimetic mutation. It was determined that S31D prevents active HIF dimers from binding to DNA, and therefore inhibits transcriptional activation, without affecting HIF-1α protein stability, the HIF-1α - HIF-1β binding interactions or nuclear localisation. Combined with the identification that the S31D mutant lacked GTPase proteins, it is possible that an unknown GTP dependent pathway has a role in promoting the association of HIF – DNA through an unknown mechanism, either directly or indirectly.

5.6. Discussion

In this chapter we used a SDM protocol to specifically create phospho -mimetic and –null mutants of identified phosphorylation sites from Chapter 4. Of the mutants generated, the HIF-1α S31 phospho-mimetic mutation (S31D) exhibited the greatest disruption to a luciferase based transcriptional assay, and thus was further characterised. S31 has previously been identified as a phosphorylation site for PKA when using recombinant HIF-1α fragments in an *in vitro* kinase assay (Bullen *et al.*, 2016). Using a S31 phospho-null mutation of the fragment, to an alanine residue, Bullen *et al.*, 2016 determined that there were no effects on protein stability when expressed in human cancer cells lines, and this mutant was not investigated further. This is in agreement with our data which show that

the stability of full-length HIF-1α is not affected by the S31G phospho-null mutation. However, data presented in Chapter 4 suggests a potential role in DNA binding, thus regulating transcription in a manner that is independent of protein stability, this was not investigated by Bullen *et al.*, 2016.

Interestingly, the phospho-mimetic aspartic acid mutation was completely transcriptionally inactive while the phospho-null glycine mutations resulted in generally marginal sub-level expression in all luciferase and RT-qPCR experiments, but exhibited large variations in measurement. A potential explanation for variability could be that a mutation to a glycine residue creates a consensus sequence for neighbouring PTM sites that experience PTM in a cell cycle dependent manner. Hence, the variation in cell cycle stage of cultured cells could explain the increased variability observed. This could be investigated by the use of cell cycle stalling agents, such as hydroxyurea, nocodazole, mimosine or thymidine blocking (Jackman *et al.*, 2001, Bostock *et al.*, 1971 & Koc *et al.*, 2004), prior to hypoxic incubation and gene transcription.

Investigating a potential cause of transcriptional inhibition by HIF-1α S31D by binding partner analysis identified that 10 separate GTPase proteins were specifically purified with the phospho-null HIF-1α S31G mutant, suggesting a phosphorylation induced lack of GTPase binding. From data presented here, there are two equally valid mechanisms that could explain the GTPase and DNA binding inefficiency correlation: 1) GTPases are essential for loading of active HIF onto the DNA, through unknown mechanisms. Consequently phosphorylation at HIF-1α S31 prevents GTPase binding and subsequent DNA binding. 2) Active HIF associates with the DNA together with GTPase proteins stabilising/activating transcription. The reduced DNA binding resulting from HIF-1α S31 phosphorylation thus simultaneously decreases GTPase protein binding.

There is emerging evidence that a class of GTPase proteins (Rho GTPases) have a regulatory role in gene expression by promoting secondary interactions within the nucleus (Phuyal *et al.*, 2019). Cdc42 is a rho GTPase protein identified here as a binding partner for the non-phosphorylatable HIF-1α S31G mutant. Cdc42 is a well characterised protein, known to have multiple roles as an activator of various cell signalling pathways involved with cellular polarity (Melendez *et al.*, 2011 & Etienne-Manneville, 2004). Thus, it is possible that Cdc42 activates HIF-1α in a S31 non-phosphorylated state (S31G) for DNA binding by direct

mechanisms or indirectly, through a secondary protein interactors becoming activated (such as a kinase). Such a role in hypoxia based signalling has not yet been identified.

The major constituents of the GTPases identified to bind S31G and WT HIF-1α are the Ras-related Rab proteins. Rab proteins are known for their intracellular organisation roles, targeting proteins to specific membrane bound organelles (Hutagalung *et al.*, 2011 & Wandinger-Ness *et al.*, 2014). Rab32, Rab1 and Rab10 are involved with mitochondrial fission and ER vesicle formation respectively, an interesting observation due to the large quantity of mitochondrial and ER specific proteins identified in Chapter 4 (>10% of all proteins). Rab5 has roles in nuclear-cytoplasmic shuttling of effector proteins (such as PI3K) or sequestration of proteins to regulate gene expression, rather than direct protein binding (Schwartz *et al.*, 2007 & Wandinger-Ness *et al.*, 2014). Since Rab proteins can simultaneously act as scaffolds (Wandinger-Ness *et al.*, 2014), and the fact that it is impossible to determine whether binding partners identified by Co-IP are direct or indirect, it is possible that a novel HIF binding partner may utilise Rab proteins for nuclear shuttling and activation. Thus, if the novel protein was blocked from interacting with HIF-1α by S31 phosphorylation it could explain the loss of transcriptional function and the Rab protein binding. It would be interesting to investigate how Rab protein known-down experiments could influence the hypoxic response.

Although the HIF-1α S786 mutations seemed to have no overall effect on transcription, it is possible that the luciferase assay failed to identify a transcriptional effect due to the nature of the experiment. HIF-1α S786 is situated within the CTAD which is essential for p300/CBP binding, and is functionally involved with chromatin remodelling for gene expression (reviewed by Chan *et al.*, 2001). Therefore, it is possible that if p300/CBP binding is inhibited that endogenous gene expression regulation may be affected, while extra-nuclear plasmid DNA (not included in the chromatin and devoid of histone proteins) could remain unaffected. Thus, analysis of the S786 mutants through RT-qPCR should be performed before disregarding functional transcriptional outcomes.

The HIF-2α S345 was selected because crystal modelling predicted a potential role in HIF-1β binding. HIF-2α S345 phospho-null mutation (S345A) had no effect on transcriptional response compared to WT, while a phospho-mimetic (S345D) mutation resulted in a large decrease in luciferase response. This could suggest that HIF-2α S345 phosphorylation weakens the HIF-1β binding strength, as predicted from modelling, thus limiting the

transcriptional efficiency of the HIF-2α protein by failure to form the active dimer. This could be investigated relatively simply through IP and western blotting for HIF-1β, however due to time restraints was not investigated further here.

It is interesting that mutation to phospho –null, or –mimetic mutations of either HIF-2α T528 or S581 phosphorylation sites, which neighbour LAP proline hydroxylation sites, resulted in ~1.5 fold increase in transcription. One hypothesised explanation is that the hydroxyl group of the threonine/serine residue in an unmodified form is an essential component of a motif to target a secondary modification that results in the reduced transcriptional efficiency, through either protein stability or transactivation. Thus, mutation to a charged (glutamic acid/aspartic acid) or hydrophobic (alanine) residue abolishes the inactivating secondary PTM, hence stabilising HIF-2α transcriptional signal independently of physiochemical properties introduced by phosphorylation. With the creation of HIF-2α T406 phospho -mimetic (T406E) and -null (T406A) mutants, neighbouring the remaining known proline hydroxylation site, it will be interesting to investigate whether similar observations are identified to T528 and S581 mutations. We hypothesise that the increased HIF-2α transcription observed by these mutations, are due to protein stability, therefore multi-site mutational analysis may have synergistic effects. Thus, it is possible that the phosphorylation pattern of the HIF-2α T406/T528/S581 sites in 21% O_2 (Chapter 4) could explain the observed 21% O_2 stability of HIF-2α, while HIF-1α is rapidly degraded.

Overall, using S31 as an example, this chapter highlights how a single PTM can dramatically alter the functional status of the HIFα proteins. In the context of drug development, HIF targeting in tumour progression is a major research interest with current approaches using HTP screening techniques to identify novel drugs (reviewed by Semenza, 2003 & Masoud *et al.*, 2015). Problematically, the functional aspect of drug-based inhibition is lost and can lead to drugs with various side effects. A more targeted approach is preferable but requires mechanistic understanding of essential signalling pathways that regulate the hypoxic response, so that drugs can be designed to bind to specific regions of the protein, or the regulatory enzyme for specific functions and/or modifications to be altered. Thus with ~100 different novel PTMs identified, it is not only exciting for the field of hypoxia regulation, but also in the future investigation of therapeutics.

6. Chapter 6: Final discussion

6.1. Final discussion

This thesis describes the development and utilisation of tools to investigate the regulation of HIFα proteins by hypoxia in an unbiased manner. I was able to identify an in-depth phosphorylation map (~25 phosphorylation sites) for each of the HIF-1α and HIF-2α proteins, with the majority of phosphorylation sites identified being novel. Furthermore, an open PTM search (all possible PTMs) identified a potential further 25 PTMs per HIFα protein (additional to phosphorylation) including acetylation, methylation and irreversible cysteine oxidative modification. Combined with domain and evolutionary analysis of ~250 different vertebrate species, we were able to identify several interesting aspects from isoform differences and evolutionary conservation/variation, which we used to guide the investigation of potentially interesting phosphorylation sites (discussed in Chapter 4). Using mutational analysis, I determined that HIF-1α S31 phosphorylation inhibits transcriptional activity, independent of protein stability mechanisms, but rather through preventing the ability to bind to DNA. Additionally, I was able to identify many more binding partners of HIFα proteins than previously defined and evaluated their O_2 dependent binding by label-free quantification proteomics approaches. I used GO annotation enrichment analysis to identify O_2 dependent cellular processes interacting with HIFα proteins to further understand the cellular processes that regulate these. There are several points to consider to guide future experimentation, discussed below.

6.1.1. Novelty of the approach

We have overexpressed full length HA-Clover tagged proteins to investigate the HIFα PTM status and binding partners. Such a strategy has the obvious drawback that HIFα expression is under a strong, exogenous CMV promoter and thus lacks endogenous DNA based regulatory mechanisms. Additionally, it is well-documented that purification tags, particularly larger tags such as GFP/Clover (~27 kDa, approximately 25% of the HIFα protein size), can influence protein function/localisation (Weill *et al.*, 2019). However, it is important to consider that countless studies have utilised GFP tagging of proteins to investigate protein function by proteomics and microscopy techniques (Trinkle-Mulcahy *et al.*, 2008, Lipinszki *et al.*, 2014, Kubitscheck *et al.*, 2000 & Soboleski *et al.*, 2005), including for HIFα protein investigation (Jiang *et al.*, 2001, Mylonis *et al.*, 2006, Mylonis *et al.*, 2008, Kalousi *et al.*, 2010, Bagnall *et al.*, 2014 & Taylor *et al.*, 2016). These studies still constituted an ideal starting point to elucidate cellular functional mechanisms of HIF signalling.

To ensure physiological relevance of the experimental system, we have used microscopy and transcriptional assays to confirm that HA-Clover-HIFα proteins had the correct nuclear localisation and were functionally active as transcription factors. Thus, in comparison to previous studies using recombinant fragment-based *in vitro* assays to determine PTMs in a targeted approach (Table 1.2 & Table 1.3), my methodology provided a much more biologically relevant representation of the HIFα signalling systems.

One important focus of this thesis was to identify molecular features that could explain the observed differences between HIFα isoforms, such as sub-nuclear localisation and O_2 dependent stability. A discovery proteomics approach was adopted, primarily focusing on phosphorylation because of the availability of enrichment tools. However, from this study and other published studies, it is clear that HIFα proteins are subjected to multiple other types of PTM. Antibody based enrichment strategies are available for other PTMs. Whilst they are less easy to adopt than the TiO_2 phosphorylation enrichment used in this study, they permit the enrichment of other PTMs, and thus could be used to provide a full in-depth PTM map of HIFα proteins. This will allow the field to fully decipher the regulatory pathways of HIFα.

In total, ~50 PTMs have been identified in this study for both HIF-1α and HIF-2α, which is of a similar order to the number of PTMs identified for endogenous p53. p53 has 86 different characterised PTMs and ~20 other PTMs have been identified in HTP MS studies (data obtained from PhosphoSitePlus (Hornbeck *et al.*, 2015 & Hornbeck *et al.*, 2012)). Therefore, although validation of the HIFα PTMs on endogenous proteins will be vital, this study provides an essential starting point to further understand HIFα regulation and its implication in signalling networks.

Interestingly, few of the previously identified HIFα PTMs were identified in our dataset, thus posing an intriguing question for the hypoxia field: has the previously used methodology generated high rates of false positive data for both PTM identification and binding partner discovery? Two arguments could provide an explanation to the apparent discrepancy: 1) cell line specific regulatory mechanisms, and 2) the 'hypoxia' conditions used. Bracken *et al.*, 2006 have shown that the protein expression levels and transcriptional profiles of HIFα isoforms vary considerably between cell lines, O_2 tension used and whether hypoxia mimicking agents are used (specifically the iron chelator DFX). Hence the multiple different cell lines used, combined with different hypoxia conditions (many publications

using different hypoxia mimicking drugs, at different concentrations and for different incubation times), could explain some of the differences observed. Therefore, further experimentation following this study should: 1) validate the data in a range of cell lines and conditions, 2) investigate the dynamics of PTMs and binding partners, and 3) functionally characterise the hypoxia regulatory network.

6.1.1.1. Validation of data

Due to the discovery nature of these experiments, and the enormous quantity of cells required to obtain sufficient amount of HIFα proteins for robust MS analysis of the complete protein sequence, combined with a lack of good antibodies for IP, analysis of endogenous proteins would have been infeasible. However, now that the PTMs have been identified it will be possible to determine the exact *m/z* ratio of specific modified peptides and use a targeted MS approach (selective reaction monitoring (SRM)), increasing sensitivity up to 100 fold (Lange *et al.*, 2008). Thus, theoretically, endogenous HIFα could be used to validate identified PTMs. Furthermore, an SRM pipeline could simultaneously be easily applied to previously published PTM sites by the *in silico* determination of proteolytic peptides and respective *m/z* ratios. However, for endogenous protein IP this would require the development of better antibodies, particularly for HIF-2α. Alternatively, CRISPR/Cas9 technology could be used to insert a tag of interest into the genome, therefore maintaining endogenous expression mechanisms.

6.1.1.2. Expanding the regulatory network

It will be of interest to repeat the discovery proteomics aspect in different cell lines to investigate whether different cell lines have different HIFα signalling pathways. This could be further extended into investigation of hypoxia time courses and severity of hypoxia treatment. For example, prolonged (chronic) hypoxia is known to downregulate HIF-1α and promote HIF-2α signalling (Koh *et al.*, 2012 & Uchida *et al.*, 2004). Hypoxia severity is an interesting question, because cultured cells incubated under prolonged physiological levels of O_2 (3-5%) are known to stabilise HIF-1α (Carrera *et al.*, 2014). Therefore, it may be possible to compare regulatory mechanisms from a more endogenous regulation system to hypoxic cancers with 0.5-1.5% O_2 (Table 1.1). Additionally, it will be interesting to investigate O_2 independent mechanisms of HIFα stabilisation, such as in response to Insulin-like Growth factor signalling (IGF2, Feldser *et al.*, 1999), Epidermal growth factor receptor signalling (EGFR, Peng *et al.*, 2006) or inflammation (Palazon *et al.*, 2014), which all have varying downstream effects.

6.1.2. HIF-1α versus HIF-2α

From the PTM data obtained here, there are several noticeable patterns of interest (discussed in Chapter 4). The most striking observation is the possible role of phosphorylation in the O_2 dependent degradation for both HIFα isoforms. It is clear that the HIF-1α ODDD is hyperphosphorylated, particularly in 21% O_2 (4.7.2). Considering emerging evidence of the requirement of other PTMs for the efficient proline hydroxylation and degradation (Lee *et al.*, 2017), it is possible that phosphorylation contributes to the O_2 dependent degradation of HIF-1α. Conversely, we predict that the phosphorylation status of HIF-2α may be, in some part, responsible for its O_2 dependent stability. We find that for HIF-2α, there is a phosphorylation site adjacent to both canonical proline hydroxylation sites, and a novel third LAP motif. We further showed that mutation of any of these sites resulted in ~1.5 fold increase in transcriptional output. However, this requires further investigation, and evaluation of protein stability.

It is important to consider that utilising a 'Bottom-Up' proteomics approach makes it is impossible to understand the co-regulation of PTM sites, unless identified on the same proteolytically generated peptide. Therefore a 'Middle-Down' proteomics approach could be adopted, where a minimally cleaving protease is used to produce large peptides of ~10 kDa in size (Cristobal *et al.*, 2017). This will allow the determination of combinations of PTMs before investigating PTMs using multi-site mutational analysis. This is particularly important when considering the potential roles of HIF-1α ODDD hyperphosphorylation, where multiple sites may co-exist and provide a degree of synergism between PTMs, therefore complicating the identification of functional roles by single site mutational analysis. However, multi-site mutational analysis of so many sites will likely result in off target effects, as structural and polarity changes in amino acids could have additive effects. Therefore, it will be important to investigate the co-existence of PTMs in a combinatorial fashion to define the PTM pattern and identify the mutants required for multi-site analysis.

To investigate the binding partners we decided not to include benzonase, a DNA digesting enzyme, because DNA and long non-coding RNAs (lncRNA) are known to regulate binding partner interactions (reviewed by Siggers *et al.*, 2014 & Schmitz *et al.*, 2016). Recent data identifies that a lncRNA has an important role in stabilising HIF-1α protein, by preventing PHD2 binding (Chen *et al.*, 2019). However, there is a possibility that chromatin DNA bound to HIF will Co-IP, resulting in the identification of proteins which do not interact with active

HIF but rather are associated with DNA. Therefore, the IP experiments could be repeated in the presence of benzonase to identify protein interactions that do not require DNA scaffolding mechanisms. Although this study is the first discovery style experiment for HIF binding partners, it was reliant on overexpression with a large IP tag. Therefore, it could be argued that the rate of false positive binding partners identified could be high, even though measures were taken for background subtraction. We were able to use the low level expression (near endogenous levels) to detect binding partners of mutants (5.5.1). Therefore, with optimisation and scaling, it will be easily possible to IP endogenous HIF-1α (and HIF-2α if an adequate antibody is found) for the detection of protein interactors of endogenous proteins. If further experiments are performed to identify endogenous HIFα binding proteins, they should strongly consider a reversible crosslinking IP strategy (Smith *et al.*, 2011), because many protein interactors will be stimuli specific and highly transient, particularly when considering modifying enzymes. Such a strategy, will allow to obtain a more in-depth interactor map.

The fact that >100 proteins (~10% of all proteins) identified for both HIFα isoforms were mitochondrial in nature is interesting. The validity of this discovery is supported by a number of published studies demonstrating a sub-population of HIF-1α localised in mitochondria (Briston *et al.*, 2011, Concolino *et al.*, 2018 & Thomas *et al.*, 2019). Additionally, it has been shown that mitochondria have a role in the stabilisation of HIFα proteins through the development of ROS, and may act as a secondary O_2 sensing mechanism to PHDs (Guzy *et al.*, 2005 & Chandel *et al.*, 1998). Overall, this could raise an interesting question: does HIF have a role in regulating mitochondrial DNA? Indeed, when I searched the mitochondrial DNA (UniProt accession: NC_012920, Andrews *et al.*, 1999) we identified 20 putative HRE promoter sequences ((A/G)CGTG, Schödel *et al.*, 2011). Using an annotated gene map of mitochondrial DNA will provide initial insight to whether these identified HREs are in proximity to functional genes, and may be HIF regulated, or whether they are intronic in nature. This will be an interesting avenue to pursue. It is likely that fluorescent tagging for live cell microscopy, previously used to investigate sub-cellular localisation, will struggle to identify HIFα in mitochondria as they may likely be below the limit of detection. Therefore, a proteomics approach that employs sub-cellular fractionation to purify mitochondria before lysis, potentially coupled with an SRM targeted approach, could be used for validation.

6.1.3. Evolutionary analysis

Our evolutionary analysis was performed on HIFα sequences from vertebrates only, for reasons explained in Chapter 4, including the fact that invertebrate species do not have specialised O_2 delivery systems. However, more advanced invertebrate species do infact have complex O_2 delivery systems including dense vasculature networks and haemoglobin based O_2 transport systems, for example in crab and squid species (Birk *et al.*, 2018 & Alter *et al.*, 2015). Although these 'HIF-1α' proteins are larger than human HIFα (~1000 vs ~850 residues) they are much smaller than insect HIFα proteins (~1500) and could reflect the evolution of developing specialist O_2 delivery systems. A more directed approach to investigating HIF-1α evolution could potentially be adopted by combining a knowledge based, selective inclusion of invertebrate species, especially considering that very early invertebrate jellyfish species have HIF-1α homologs (Wang *et al.*, 2014). Additionally, the bHLH and PAS domains are common domains in many proteins, therefore it would be of interest to investigate how the respective domains have evolved between proteins to identify potential regulatory differences and PTM sites of interest.

From the sequence alignments performed here, it is clear that both HIFα isoforms are highly evolutionarily conserved, highlighting their functional importance. However, the inhibitory domain is poorly conserved between isoforms and among vertebrate species, generally only conserved within families. Interestingly, Bony fish species show extensive variation in this region, with all species, except for hypoxia tolerant Carp, missing the first ~50 residues, which is serine/threonine rich in human with 15 different residues. Unfortunately, from our study this serine/threonine rich region was unanalysable by MS, even using multiple different proteases, therefore it would be of interest to investigate this further using different proteolytic strategies to identify if phosphorylation has a significant role here. The inhibitory domain could possibly be viewed as an evolutionary 'hot pocket' for mutations, to both distinguish between HIFα isoforms and provide slightly altered function between phylogenetic families.

6.1.4. Outlook and future perspective

Whilst we have discovered many novel PTMs and binding partners in this work, it has posed many new questions, including: 1) The functional meaning of the PTMs and interactions identified. 2) Whether the PTM status may be transient/dynamic and its robustness as a result of different hypoxic incubations/treatments and cell lines. 3) The potential role of HIFα in mitochondria, and 4) The accuracy and relevance of previously used methodologies

to study HIF PTM status and regulatory roles. Overall it appears that the canonical proline hydroxylation pathway is not the only pathway to act on HIFα. However, unravelling the functional outcomes post PTM will require extensive further work. The results of further characterising HIFα regulation/signalling might make it possible to target hypoxic cells in tumours in a selective manner, by inhibiting (or promoting) modifying enzymes. If this can be achieved, it would prevent the blanket inhibition of HIF signalling, which is likely to have various side effects for hypoxic adaption elsewhere in the body.

Finally, our data suggest that the signalling pathways regulating HIF function may be orders of magnitude more complex than the current understanding, opening up multiple avenues for further work to fully understand the cellular adaption to hypoxia.

7. References

Abreu, E., R. M. Terns and M. P. Terns (2011). "Visualization of human telomerase localization by fluorescence microscopy techniques." Methods Mol Biol **735**: 125-137.

Aebersold, R. and M. Mann (2016). "Mass-spectrometric exploration of proteome structure and function." Nature **537**: 347.

Aga, M., G. L. Bentz, S. Raffa, M. R. Torrisi, S. Kondo, N. Wakisaka, T. Yoshizaki, J. S. Pagano and J. Shackelford (2014). "Exosomal HIF1alpha supports invasive potential of nasopharyngeal carcinoma-associated LMP1-positive exosomes." Oncogene **33**(37): 4613-4622.

Alam, M. W., C. U. Persson, S. Reinbothe, J. U. Kazi, L. Rönnstrand, C. Wigerup, H. J. Ditzel, A. E. Lykkesfeldt, S. Påhlman and A. Jögi (2016). "HIF2α contributes to antiestrogen resistance via positive bilateral crosstalk with EGFR in breast cancer cells." Oncotarget **7**(10): 11238-11250.

Alberts, B. (1998). "The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists." cell **92**(3): 291-294.

Alper, H., C. Fischer, E. Nevoigt and G. Stephanopoulos (2005). "Tuning genetic control through promoter engineering." Proceedings of the National Academy of Sciences of the United States of America **102**(36): 12678-12683.

Alter, K., K. Paschke, P. Gebauer, J.-P. Cumillaf and H.-O. Pörtner (2015). "Differential physiological responses to oxygen availability in early life stages of decapods developing in distinct environments." Marine Biology **162**(5): 1111-1124.

Andrews, R. M., I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull and N. Howell (1999). "Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA." Nat Genet **23**(2): 147.

Arany, Z., L. E. Huang, R. Eckner, S. Bhattacharya, C. Jiang, M. A. Goldberg, H. F. Bunn and D. M. Livingston (1996). "An essential role for p300/CBP in the cellular response to hypoxia." Proceedings of the National Academy of Sciences **93**(23): 12969-12973.

Arike, L. and L. Peil (2014). Spectral Counting Label-Free Proteomics. Shotgun Proteomics: Methods and Protocols. D. Martins-de-Souza. New York, NY, Springer New York**:** 213-222.

Arnesen, T., X. Kong, R. Evjenth, D. Gromyko, J. E. Varhaug, Z. Lin, N. Sang, J. Caro and J. R. Lillehaug (2005). "Interaction between HIF-1α (ODD) and hARD1 does not induce acetylation and destabilization of HIF-1α." Febs Letters **579**(28): 6428-6432.

Attwood, P. V., P. G. Besant and M. J. Piggott (2011). "Focus on phosphoaspartate and phosphoglutamate." Amino Acids **40**(4): 1035-1051.

Bae, S. H., J. W. Jeong, J. A. Park, S. H. Kim, M. K. Bae, S. J. Choi and K. W. Kim (2004). "Sumoylation increases HIF-1 stability and its transcriptional activity." Biochemical and Biophysical Research Communications **324**(1): 394-400.

Bagnall, J., J. Leedale, S. E. Taylor, D. G. Spiller, M. R. H. White, K. J. Sharkey, R. N. Bearon and V. See (2014). "Tight Control of Hypoxia-inducible Factor-alpha Transient Dynamics Is Essential for Cell Survival in Hypoxia." Journal of Biological Chemistry **289**(9): 5549-5564.

Bamford, S., E. Dawson, S. Forbes, J. Clements, R. Pettett, A. Dogan, A. Flanagan, J. Teague, P. A. Futreal, M. R. Stratton and R. Wooster (2004). "The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website." British Journal of Cancer **91**(2): 355-358.

Bantscheff, M., S. Lemeer, M. M. Savitski and B. Kuster (2012). "Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present." Anal Bioanal Chem **404**(4): 939-965.

Bantscheff, M., M. Schirle, G. Sweetman, J. Rick and B. Kuster (2007). "Quantitative mass spectrometry in proteomics: a critical review." Anal Bioanal Chem **389**(4): 1017-1031.

Bantscheff, M., M. Schirle, G. Sweetman, J. Rick and B. Kuster (2007). "Quantitative mass spectrometry in proteomics: a critical review." Analytical and bioanalytical chemistry **389**(4): 1017-1031.

Bao, L., Y. Chen, H. T. Lai, S. Y. Wu, J. E. Wang, K. J. Hatanpaa, J. M. Raisanen, M. Fontenot, B. Lega, C. M. Chiang, G. L. Semenza, Y. Wang and W. Luo (2018). "Methylation of hypoxia-inducible factor (HIF)-1alpha by G9a/GLP inhibits HIF-1 transcriptional activity and cell migration." Nucleic Acids Res **46**(13): 6576-6591.

Barlow, D. J. and J. M. Thornton (1983). "Ion-pairs in proteins." Journal of Molecular Biology **168**(4): 867-885.

Batie, M., J. Druker, L. D'Ignazio and S. Rocha (2017). "KDM2 Family Members are Regulated by HIF-1 in Hypoxia." Cells **6**(1): 8.

Batie, M., J. Frost, M. Frost, J. W. Wilson, P. Schofield and S. Rocha (2019). "Hypoxia induces rapid changes to histone methylation and reprograms chromatin." Science **363**(6432): 1222-1226.

Beall, C. M., G. L. Cavalleri, L. Deng, R. C. Elston, Y. Gao, J. Knight, C. Li, J. C. Li, Y. Liang, M. McCormack, H. E. Montgomery, H. Pan, P. A. Robbins, K. V. Shianna, S. C. Tam, N. Tsering, K. R. Veeramah, W. Wang, P. Wangdui, M. E. Weale, Y. Xu, Z. Xu, L. Yang, M. J. Zaman, C. Zeng, L. Zhang, X. Zhang, P. Zhaxi and Y. T. Zheng (2010). "Natural selection on EPAS1 (HIF2α) associated with low hemoglobin concentration in Tibetan highlanders." Proceedings of the National Academy of Sciences **107**(25): 11459-11464.

Bell, E. L., T. A. Klimova, J. Eisenbart, P. T. Schumacker and N. S. Chandel (2007). "Mitochondrial Reactive Oxygen Species Trigger Hypoxia-Inducible Factor-Dependent Extension of the Replicative Life Span during Hypoxia." Molecular and cellular biology **27**(16): 5737-5745.

Beltrao, P., P. Bork, N. J. Krogan and V. van Noort (2013). "Evolution and functional cross-talk of protein post-translational modifications." Molecular systems biology **9**: 714-714.

Ben-Yosef, Y., N. Lahat, S. Shapiro, H. Bitterman and A. Miller (2002). "Regulation of endothelial matrix metalloproteinase-2 by hypoxia/reoxygenation." Circ Res **90**(7): 784-791.

Bern, M., B. S. Phinney and D. Goldberg (2009). "Reanalysis of Tyrannosaurus rex Mass Spectra." Journal of Proteome Research **8**(9): 4328-4332.

Berta, M. A., N. Mazure, M. Hattab, J. Pouysségur and M. C. Brahimi-Horn (2007). "SUMOylation of hypoxia-inducible factor-1α reduces its transcriptional activity." Biochemical and Biophysical Research Communications **360**(3): 646-652.

Besant, P. G., P. V. Attwood and M. J. Piggott (2009). "Focus on Phosphoarginine and Phospholysine." Current Protein & Peptide Science **10**(6): 536-550.

Beurel, E., S. F. Grieco and R. S. Jope (2015). "Glycogen synthase kinase-3 (GSK3): regulation, actions, and diseases." Pharmacology & therapeutics **148**: 114-131.

Biggar, K. K. and S. S. C. Li (2014). "Non-histone protein methylation as a regulator of cellular signalling and function." Nature reviews Molecular cell biology **16**: 5.

Bigham, A. W., X. Mao, R. Mei, T. Brutsaert, M. J. Wilson, C. G. Julian, E. J. Parra, J. M. Akey, L. G. Moore and M. D. Shriver (2009). "Identifying positive selection candidate loci for high-altitude adaptation in Andean populations." Hum Genomics **4**(2): 79-90.

Birk, M. A., A. K. Dymowska and B. A. Seibel (2018). "Do squid breathe through their skin?" The Journal of Experimental Biology **221**(19): jeb185553.

Björnström, L. and M. Sjöberg (2005). "Mechanisms of Estrogen Receptor Signaling: Convergence of Genomic and Nongenomic Actions on Target Genes." Molecular Endocrinology **19**(4): 833-842.

Blackinton, J., M. Lakshminarasimhan, K. J. Thomas, R. Ahmad, E. Greggio, A. S. Raza, M. R. Cookson and M. A. Wilson (2009). "Formation of a stabilized cysteine sulfinic acid is critical for the mitochondrial function of the parkinsonism protein DJ-1." J Biol Chem **284**(10): 6476-6485.

Blanc, R. S. and S. Richard (2017). "Arginine Methylation: The Coming of Age." Molecular cell **65**(1): 8-24.

Bondarenko, P. V., D. Chelius and T. A. Shaler (2002). "Identification and Relative Quantitation of Protein Mixtures by Enzymatic Digestion Followed by Capillary Reversed-Phase Liquid Chromatography−Tandem Mass Spectrometry." Analytical Chemistry **74**(18): 4741-4749.

Bostock, C. J., D. M. Prescott and J. B. Kirkpatrick (1971). "An evaluation of the double thymidine block for synchronizing mammalian cells at the G1-S border." Exp Cell Res **68**(1): 163-168.

Bracken, C. P., A. O. Fedele, S. Linke, W. Balrak, K. Lisy, M. L. Whitelaw and D. J. Peet (2006). "Cell-specific regulation of hypoxia-inducible factor (HIF)-1α and HIF-2α stabilization and transactivation in a graded oxygen environment." Journal of Biological Chemistry **281**(32): 22575-22585.

Briston, T., J. Yang and M. Ashcroft (2011). "HIF-1α localization with mitochondria: a new role for an old favorite?" Cell cycle (Georgetown, Tex.) **10**(23): 4170-4171.

Bruick, R. K. and S. L. McKnight (2001). "A Conserved Family of Prolyl-4-Hydroxylases That Modify HIF." Science **294**(5545): 1337-1340.

Bui, V. M., S. L. Weng, C. T. Lu, T. H. Chang, J. T. Weng and T. Y. Lee (2016). "SOHSite: incorporating evolutionary information and physicochemical properties to identify protein S-sulfenylation sites." BMC Genomics **17 Suppl 1**: 9.

Bullen, J. W., I. Tchernyshyov, R. J. Holewinski, L. DeVine, F. Wu, V. Venkatraman, D. L. Kass, R. N. Cole, J. Van Eyk and G. L. Semenza (2016). "Protein kinase A-dependent phosphorylation stimulates the transcriptional activity of hypoxia-inducible factor 1." Science signaling **9**(430): ra56.

Cam, H., J. B. Easton, A. High and P. J. Houghton (2010). "mTORC1 signaling under hypoxic conditions is controlled by ATM-dependent phosphorylation of HIF-1α." Molecular cell **40**(4): 509-520.

Candiano, G., M. Bruschi, L. Musante, L. Santucci, G. M. Ghiggeri, B. Carnemolla, P. Orecchia, L. Zardi and P. G. Righetti (2004). "Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis." Electrophoresis **25**(9): 1327-1333.

Capra, J. A. and M. Singh (2007). "Predicting functionally important residues from sequence conservation." Bioinformatics **23**(15): 1875-1882.

Caradec, J., N. Sirab, C. Keumeugni, S. Moutereau, M. Chimingqi, C. Matar, D. Revaud, M. Bah, P. Manivet, M. Conti and S. Loric (2010). "'Desperate house genes': the dramatic example of hypoxia." British Journal of Cancer **102**(6): 1037-1043.

Carreau, A., B. El Hafny-Rahbi, A. Matejuk, C. Grillon and C. Kieda (2011). "Why is the partial oxygen pressure of human tissues a crucial parameter? Small molecules and hypoxia." Journal of cellular and molecular medicine **15**(6): 1239-1253.

Carrera, S., J. Senra, M. I. Acosta, M. Althubiti, E. M. Hammond, P. J. de Verdier and S. Macip (2014). "The Role of the HIF-1α Transcription Factor in Increased Cell Division at Physiological Oxygen Tensions." PLoS One **9**(5): e97938.

Cassavaugh, J. M., S. A. Hale, T. L. Wellman, A. K. Howe, C. Wong and K. M. Lounsbury (2011). "Negative regulation of HIF-1α by an FBW7-mediated degradation pathway during hypoxia." Journal of cellular biochemistry **112**(12): 3882-3890.

Chan, H. M. and N. B. La Thangue (2001). "p300/CBP proteins: HATs for transcriptional bridges and scaffolds." Journal of Cell Science **114**(13): 2363-2373.

Chandel, N. S., E. Maltepe, E. Goldwasser, C. E. Mathieu, M. C. Simon and P. T. Schumacker (1998). "Mitochondrial reactive oxygen species trigger hypoxia-induced transcription." Proceedings of the National Academy of Sciences **95**(20): 11715-11720.

Chen, C., N. Pore, A. Behrooz, F. Ismail-Beigi and A. Maity (2001). "Regulation of glut1 mRNA by hypoxia-inducible factor-1. Interaction between H-ras and hypoxia." J Biol Chem **276**(12): 9519-9525.

Chen, F., J. Chen, L. Yang, J. Liu, X. Zhang, Y. Zhang, Q. Tu, D. Yin, D. Lin, P.-P. Wong, D. Huang, Y. Xing, J. Zhao, M. Li, Q. Liu, F. Su, S. Su and E. Song (2019). "Extracellular vesicle-packaged HIF-1α-stabilizing lncRNA from tumour-associated macrophages regulates aerobic glycolysis of breast cancer cells." Nature Cell Biology **21**(4): 498-510.

Cheng, J., X. Kang, S. Zhang and E. T. Yeh (2007). "SUMO-specific protease 1 is essential for stabilization of HIF1α during hypoxia." cell **131**(3): 584-595.

Cho, H. J., D. R. Ahn, H. Park and E. G. Yang (2007). "Modulation of p300 binding by posttranslational modifications of the C-terminal activation domain of hypoxia-inducible factor-1 alpha." Febs Letters **581**(8): 1542-1548.

Chong, S., C. Dugast-Darzacq, Z. Liu, P. Dong, G. M. Dailey, C. Cattoglio, A. Heckert, S. Banala, L. Lavis, X. Darzacq and R. Tjian (2018). "Imaging dynamic and selective low-complexity domain interactions that control gene transcription." Science **361**(6400): eaar2555.

Chung, H. S., S.-B. Wang, V. Venkatraman, C. I. Murray and J. E. Van Eyk (2013). "Cysteine Oxidative Posttranslational Modifications." Circulation research **112**(2): 382-392.

Chung, H. S., S. B. Wang, V. Venkatraman, C. I. Murray and J. E. Van Eyk (2013). "Cysteine oxidative posttranslational modifications: emerging regulation in the cardiovascular system." Circ Res **112**(2): 382-392.

Cockman, M. E., K. Lippl, Y. M. Tian, H. B. Pegg, W. D. J. Figg, M. I. Abboud, R. Heilig, R. Fischer, J. Myllyharju, C. J. Schofield and P. J. Ratcliffe (2019). "Lack of activity of recombinant HIF prolyl hydroxylases (PHDs) on reported non-HIF substrates." Elife **8**.

Cohen, P. (2001). "The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture." Eur J Biochem **268**(19): 5001-5010.

Cohen, P. (2002). "Protein kinases--the major drug targets of the twenty-first century?" Nat Rev Drug Discov **1**(4): 309-315.

Comerford, K. M., T. J. Wallace, J. Karhausen, N. A. Louis, M. C. Montalto and S. P. Colgan (2002). "Hypoxia-inducible factor-1-dependent regulation of the multidrug resistance (MDR1) gene." Cancer Res **62**(12): 3387-3394.

Concolino, A., E. Olivo, L. Tammè, C. V. Fiumara, M. T. De Angelis, B. Quaresima, V. Agosti, F. S. Costanzo, G. Cuda and D. Scumaci (2018). "Proteomics Analysis to Assess the Role of Mitochondria in BRCA1-Mediated Breast Tumorigenesis." Proteomes **6**(2): 16.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang (2013). "Multiplex Genome Engineering Using CRISPR/Cas Systems." Science **339**(6121): 819-823.

Coulet, F., S. Nadaud, M. Agrapart and F. Soubrier (2003). "Identification of hypoxia-response element in the human endothelial nitric-oxide synthase gene promoter." J Biol Chem **278**(47): 46230-46240.

Cox, J., M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj and M. Mann (2014). "Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ." Molecular & cellular proteomics : MCP **13**(9): 2513-2526.

Cox, J., N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen and M. Mann (2011). "Andromeda: a peptide search engine integrated into the MaxQuant environment." J Proteome Res **10**(4): 1794-1805.

Creasy, D. M. and J. S. Cottrell (2004). "Unimod: Protein modifications for mass spectrometry." PROTEOMICS **4**(6): 1534-1536.

Cristobal, A., F. Marino, H. Post, H. W. P. van den Toorn, S. Mohammed and A. J. R. Heck (2017). "Toward an Optimized Workflow for Middle-Down Proteomics." Analytical Chemistry **89**(6): 3318-3325.

Cui, S.-Y., J.-Y. Huang, Y.-T. Chen, H.-Z. Song, G.-C. Huang, W. De, R. Wang and L.-B. Chen (2013). "The role of Aurora A in hypoxia-inducible factor 1α-promoting malignant phenotypes of hepatocelluar carcinoma." Cell cycle (Georgetown, Tex.) **12**(17): 2849-2866.

Daniels, D. L., J. Mendez, H. Benink, A. Niles, N. Murphy, M. Ford, R. Jones, R. Amunugama, D. Allen and M. Urh (2014). "Discovering Protein Interactions and Characterizing Protein Function Using HaloTag Technology." Jove-Journal of Visualized Experiments(89): 9.

Daniels, D. L., J. Mendez, H. Benink, A. Niles, N. Murphy, M. Ford, R. Jones, R. Amunugama, D. Allen and M. Urh (2014). "Discovering protein interactions and characterizing protein function using HaloTag technology." J Vis Exp(89).

del Solar, G., R. Giraldo, M. J. Ruiz-Echevarría, M. Espinosa and R. Díaz-Orejas (1998). "Replication and Control of Circular Bacterial Plasmids." Microbiology and Molecular Biology Reviews **62**(2): 434-464.

Dengler, V. L., M. Galbraith and J. M. Espinosa (2014). "Transcriptional regulation by hypoxia inducible factors." Critical reviews in biochemistry and molecular biology **49**(1): 1-15.

Dennis, G., Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane and R. A. Lempicki (2003). "DAVID: Database for Annotation, Visualization, and Integrated Discovery." Genome biology **4**(9): R60-R60.

Dioum, E. M., R. Chen, M. S. Alexander, Q. Zhang, R. T. Hogg, R. D. Gerard and J. A. Garcia (2009). "Regulation of hypoxia-inducible factor 2α signaling by the stress-responsive deacetylase sirtuin 1." Science **324**(5932): 1289-1293.

Douglas, D. J., A. J. Frank and D. Mao (2005). "Linear ion traps in mass spectrometry." Mass Spectrometry Reviews **24**(1): 1-29.

Downward, J. (2003). "Targeting RAS signalling pathways in cancer therapy." Nature Reviews Cancer **3**(1): 11-22.

Drazic, A., L. M. Myklebust, R. Ree and T. Arnesen (2016). "The world of protein acetylation." Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1864**(10): 1372-1401.

Dunham, W. H., M. Mullin and A.-C. Gingras (2012). "Affinity-purification coupled to mass spectrometry: Basic principles and strategies." PROTEOMICS **12**(10): 1576-1590.

Dunwoodie, S. L. (2009). "The Role of Hypoxia in Development of the Mammalian Embryo." Developmental Cell **17**(6): 755-773.

Ehrenshaft, M., L. J. Deterding and R. P. Mason (2015). "Tripping up Trp: Modification of protein tryptophan residues by reactive oxygen species, modes of detection, and biological consequences." Free Radic Biol Med **89**: 220-228.

Ema, M., K. Hirota, J. Mimura, H. Abe, J. Yodoi, K. Sogawa, L. Poellinger and Y. Fujii-Kuriyama (1999). "Molecular mechanisms of transcription activation by HLF and HIF1alpha in response to hypoxia: their stabilization and redox signal-induced interaction with CBP/p300." Embo j **18**(7): 1905-1914.

Ema, M., S. Taya, N. Yokotani, K. Sogawa, Y. Matsuda and Y. Fujii-Kuriyama (1997). "A novel bHLH-PAS factor with close sequence similarity to hypoxia-inducible factor 1alpha regulates the VEGF expression and is potentially involved in lung and vascular development." Proc Natl Acad Sci U S A **94**(9): 4273-4278.

Encell, L. P., R. Friedman Ohana, K. Zimmerman, P. Otto, G. Vidugiris, M. G. Wood, G. V. Los, M. G. McDougall, C. Zimprich, N. Karassina, R. D. Learish, R. Hurst, J. Hartnett, S. Wheeler, P. Stecha, J. English, K. Zhao, J. Mendez, H. A. Benink, N. Murphy, D. L. Daniels, M. R. Slater, M. Urh, A. Darzins, D. H. Klaubert, R. F. Bulleit and K. V. Wood (2012). "Development of a dehalogenase-based protein fusion tag capable of rapid, selective and covalent attachment to customizable ligands." Current chemical genomics **6**: 55-71.

Epstein, A. C. R., J. M. Gleadle, L. A. McNeill, K. S. Hewitson, J. O'Rourke, D. R. Mole, M. Mukherji, E. Metzen, M. I. Wilson, A. Dhanda, Y.-M. Tian, N. Masson, D. L. Hamilton, P. Jaakkola, R. Barstead, J. Hodgkin, P. H. Maxwell, C. W. Pugh, C. J. Schofield and P. J. Ratcliffe (2001). "C. elegans EGL-9 and Mammalian Homologs Define a Family of Dioxygenases that Regulate HIF by Prolyl Hydroxylation." cell **107**(1): 43-54.

Erler, J. T., C. J. Cawthorne, K. J. Williams, M. Koritzinsky, B. G. Wouters, C. Wilson, C. Miller, C. Demonacos, I. J. Stratford and C. Dive (2004). "Hypoxia-mediated down-regulation of Bid

and Bax in tumors occurs via hypoxia-inducible factor 1-dependent and -independent mechanisms and contributes to drug resistance." Mol Cell Biol **24**(7): 2875-2889.

Etienne-Manneville, S. (2004). "Cdc42 - the centre of polarity." Journal of Cell Science **117**(8): 1291-1300.

Fanale, D., V. Bazan, L. R. Corsini, S. Caruso, L. Insalaco, M. Castiglia, G. Cicero, G. Bronte and A. Russo (2013). "HIF-1 is involved in the negative regulation of AURKA expression in breast cancer cell lines under hypoxic conditions." Breast Cancer Research and Treatment **140**(3): 505-517.

Feldser, D., F. Agani, N. V. Iyer, B. Pak, G. Ferreira and G. L. Semenza (1999). "Reciprocal positive regulation of hypoxia-inducible factor 1alpha and insulin-like growth factor 2." Cancer Res **59**(16): 3915-3918.

Fenn, J., M. Mann, C. Meng, S. Wong and C. Whitehouse (1989). "Electrospray ionization for mass spectrometry of large biomolecules." Science **246**(4926): 64-71.

Ferries, S., S. Perkins, P. J. Brownridge, A. Campbell, P. A. Eyers, A. R. Jones and C. E. Eyers (2017). "Evaluation of Parameters for Confident Phosphorylation Site Localization Using an Orbitrap Fusion Tribrid Mass Spectrometer." Journal of Proteome Research **16**(9): 3448-3459.

Fíla, J. and D. Honys (2012). "Enrichment techniques employed in phosphoproteomics." Amino Acids **43**(3): 1025-1047.

Flamme, I., T. Frohlich, M. von Reutern, A. Kappel, A. Damert and W. Risau (1997). "HRF, a putative basic helix-loop-helix-PAS-domain transcription factor is closely related to hypoxia-inducible factor-1 alpha and developmentally expressed in blood vessels." Mech Dev **63**(1): 51-60.

Flotho, A. and F. Melchior (2013). "Sumoylation: a regulatory protein modification in health and disease." Annual review of biochemistry **82**: 357-385.

Flügel, D., A. Görlach and T. Kietzmann (2012). "GSK-3β regulates cell growth, migration, and angiogenesis via Fbw7 and USP28-dependent degradation of HIF-1α." Blood **119**(5): 1292-1301.

Flügel, D., A. Görlach, C. Michiels and T. Kietzmann (2007). "Glycogen synthase kinase 3 phosphorylates hypoxia-inducible factor 1α and mediates its destabilization in a VHL-independent manner." Molecular and cellular biology **27**(9): 3253-3265.

Forbes, S. A., D. Beare, H. Boutselakis, S. Bamford, N. Bindal, J. Tate, C. G. Cole, S. Ward, E. Dawson, L. Ponting, R. Stefancsik, B. Harsha, C. Y. Kok, M. Jia, H. Jubb, Z. Sondka, S. Thompson, T. De and P. J. Campbell (2016). "COSMIC: somatic cancer genetics at high-resolution." Nucleic acids research **45**(D1): D777-D783.

Forsythe, J. A., B. H. Jiang, N. V. Iyer, F. Agani, S. W. Leung, R. D. Koos and G. L. Semenza (1996). "Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1." Mol Cell Biol **16**(9): 4604-4613.

Franovic, A., C. E. Holterman, J. Payette and S. Lee (2009). "Human cancers converge at the HIF-2alpha oncogenic axis." Proc Natl Acad Sci U S A **106**(50): 21306-21311.

Frese, C. K., A. F. Altelaar, H. van den Toorn, D. Nolting, J. Griep-Raming, A. J. Heck and S. Mohammed (2012). "Toward full peptide sequence coverage by dual fragmentation

combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry." Anal Chem **84**(22): 9668-9673.

Frese, C. K., H. Zhou, T. Taus, A. F. M. Altelaar, K. Mechtler, A. J. R. Heck and S. Mohammed (2013). "Unambiguous Phosphosite Localization using Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)." Journal of Proteome Research **12**(3): 1520-1525.

Frisancho, A. R. (2013). "Developmental Functional Adaptation to High Altitude: Review." American Journal of Human Biology **25**(2): 151-168.

Fu, D. and D. R. Richardson (2007). "Iron chelation and regulation of the cell cycle: 2 mechanisms of posttranscriptional regulation of the universal cyclin-dependent kinase inhibitor p21CIP1/WAF1 by iron depletion." Blood **110**(2): 752-761.

Fujita, N., K. Chiba, I. M. Shapiro and M. V. Risbud (2012). "HIF-1α and HIF-2α degradation is differentially regulated in nucleus pulposus cells of the intervertebral disc." Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research **27**(2): 401-412.

Fujiwara, N., M. Nakano, S. Kato, D. Yoshihara, T. Ookawara, H. Eguchi, N. Taniguchi and K. Suzuki (2007). "Oxidative modification to cysteine sulfonic acid of Cys111 in human copper-zinc superoxide dismutase." J Biol Chem **282**(49): 35933-35944.

Geng, H., C. T. Harvey, J. Pittsenbarger, Q. Liu, T. M. Beer, C. H. Xue and D. Z. Qian (2011). "HDAC4 Protein Regulates HIF1 alpha Protein Lysine Acetylation and Cancer Cell Response to Hypoxia." Journal of Biological Chemistry **286**(44): 38095-38102.

Geng, H., Q. Liu, C. Xue, L. L. David, T. M. Beer, G. V. Thomas, M.-S. Dai and D. Z. Qian (2012). "HIF1α protein stability is increased by acetylation at lysine 709." Journal of Biological Chemistry **287**(42): 35496-35505.

Ghesquière, B. and K. Gevaert (2014). "Proteomics methods to study methionine oxidation." Mass Spectrometry Reviews **33**(2): 147-156.

Gilkes, D. M., S. Bajpai, P. Chaturvedi, D. Wirtz and G. L. Semenza (2013). "Hypoxia-inducible factor 1 (HIF-1) promotes extracellular matrix remodeling under hypoxic conditions by inducing P4HA1, P4HA2, and PLOD2 expression in fibroblasts." J Biol Chem **288**(15): 10819-10829.

Gill, A. C., M. A. Ritchie, L. G. Hunt, S. E. Steane, K. G. Davies, S. P. Bocking, A. G. Rhie, A. D. Bennett and J. Hope (2000). "Post-translational hydroxylation at the N-terminus of the prion protein reveals presence of PPII structure in vivo." The EMBO journal **19**(20): 5324-5331.

Gillet, L. C., A. Leitner and R. Aebersold (2016). "Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing." Annual Review of Analytical Chemistry **9**(1): 449-472.

Gkotinakou, I.-M., C. Befani, G. Simos and P. Liakos (2019). "ERK1/2 phosphorylates HIF-2α and regulates its activity by controlling its CRM1-dependent nuclear shuttling." Journal of Cell Science **132**(7): jcs225698.

Goda, N., H. E. Ryan, B. Khadivi, W. McNulty, R. C. Rickert and R. S. Johnson (2003). "Hypoxia-inducible factor 1alpha is essential for cell cycle arrest during hypoxia." Mol Cell Biol **23**(1): 359-369.

Gorr, T. A., T. Tomita, P. Wappner and H. F. Bunn (2004). "Regulation of Drosophila hypoxia-inducible factor (HIF) activity in SL2 cells: identification of a hypoxia-induced variant isoform of the HIFalpha homolog gene similar." J Biol Chem **279**(34): 36048-36058.

Gorres, K. L. and R. T. Raines (2010). "Prolyl 4-hydroxylase." Critical reviews in biochemistry and molecular biology **45**(2): 106-124.

Gorsky, M. K., S. Burnouf, J. Dols, E. Mandelkow and L. Partridge (2016). "Acetylation mimic of lysine 280 exacerbates human Tau neurotoxicity in vivo." Scientific Reports **6**: 22685.

Gradin, K., C. Takasaki, Y. Fujii-Kuriyama and K. Sogawa (2002). "The transcriptional activation function of the HIF-like factor requires phosphorylation at a conserved threonine." Journal of Biological Chemistry **277**(26): 23508-23514.

Graham, A. M. and J. S. Presnell (2017). "Hypoxia Inducible Factor (HIF) transcription factor family expansion, diversification, divergence and selection in eukaryotes." PLoS One **12**(6): e0179545-e0179545.

Gu, Y. Z., S. M. Moran, J. B. Hogenesch, L. Wartman and C. A. Bradfield (1998). "Molecular characterization and chromosomal localization of a third alpha-class hypoxia inducible factor subunit, HIF3alpha." Gene Expr **7**(3): 205-213.

Gupta, R., C. Chetty, P. Bhoopathi, S. Lakka, S. Mohanam, J. S. Rao and D. E. Dinh (2011). "Downregulation of uPA/uPAR inhibits intermittent hypoxia-induced epithelial-mesenchymal transition (EMT) in DAOY and D283 medulloblastoma cells." Int J Oncol **38**(3): 733-744.

Guzy, R. D., B. Hoyos, E. Robin, H. Chen, L. Liu, K. D. Mansfield, M. C. Simon, U. Hammerling and P. T. Schumacker (2005). "Mitochondrial complex III is required for hypoxia-induced ROS production and cellular oxygen sensing." Cell Metabolism **1**(6): 401-408.

Haag, A. M. (2016). Mass Analyzers and Mass Spectrometers. Modern Proteomics – Sample Preparation, Analysis and Practical Applications. H. Mirzaei and M. Carrasco. Cham, Springer International Publishing**:** 157-169.

Hall, B. G. (2013). "Building Phylogenetic Trees from Molecular Data with MEGA." Molecular Biology and Evolution **30**(5): 1229-1235.

Han, X., A. Aslanian and J. R. Yates, 3rd (2008). "Mass spectrometry for proteomics." Current opinion in chemical biology **12**(5): 483-490.

Han, X., L. He, L. Xin, B. Shan and B. Ma (2011). "PeaksPTM: Mass spectrometry-based identification of peptides with unspecified modifications." J Proteome Res **10**(7): 2930-2936.

Han, X., L. He, L. Xin, B. Shan and B. Ma (2011). "PeaksPTM: Mass Spectrometry-Based Identification of Peptides with Unspecified Modifications." Journal of Proteome Research **10**(7): 2930-2936.

Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." cell **100**(1): 57-70.

Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of cancer: the next generation." cell **144**(5): 646-674.

Hara, S., C. Kobayashi and N. Imura (1999). "Nuclear localization of hypoxia-inducible factor-2alpha in bovine arterial endothelial cells." Molecular cell biology research communications : MCBRC **2**(2): 119-123.

Hardman, G., S. Perkins, P. J. Brownridge, C. J. Clarke, D. P. Byrne, A. E. Campbell, A. Kalyuzhnyy, A. Myall, P. A. Eyers, A. R. Jones and C. E. Eyers (2019). "Strong anion exchange-mediated phosphoproteomics reveals extensive human non-canonical phosphorylation." Embo j: e100847.

Hardman, G., S. Perkins, Z. Ruan, N. Kannan, P. Brownridge, D. P. Byrne, P. A. Eyers, A. R. Jones and C. E. Eyers (2017). "Extensive non-canonical phosphorylation in human cells revealed using strong-anion exchange-mediated phosphoproteomics." bioRxiv: 202820.

Hendriks, I. A., D. Lyon, D. Su, N. H. Skotte, J. A. Daniel, L. J. Jensen and M. L. Nielsen (2018). "Site-specific characterization of endogenous SUMOylation across species and organs." Nature communications **9**(1): 2456.

Hendriks, I. A. and A. C. O. Vertegaal (2016). "A comprehensive compilation of SUMO proteomics." Nature reviews Molecular cell biology **17**: 581.

Herzog, J., S. M. Ehrlich, L. Pfitzer, J. Liebl, T. Fröhlich, G. J. Arnold, W. Mikulits, C. Haider, A. M. Vollmar and S. Zahler (2016). "Cyclin-dependent kinase 5 stabilizes hypoxia-inducible factor-1α: a novel approach for inhibiting angiogenesis in hepatocellular carcinoma." Oncotarget **7**(19): 27108.

Hess, D. T., A. Matsumoto, S.-O. Kim, H. E. Marshall and J. S. Stamler (2005). "Protein S-nitrosylation: purview and parameters." Nature reviews Molecular cell biology **6**(2): 150-166.

Hogenesch, J. B., W. K. Chan, V. H. Jackiw, R. C. Brown, Y. Z. Gu, M. Pray-Grant, G. H. Perdew and C. A. Bradfield (1997). "Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the dioxin signaling pathway." J Biol Chem **272**(13): 8581-8593.

Holmquist-Mengelbier, L., E. Fredlund, T. Lofstedt, R. Noguera, S. Navarro, H. Nilsson, A. Pietras, J. Vallon-Christersson, A. Borg, K. Gradin, L. Poellinger and S. Pahlman (2006). "Recruitment of HIF-1alpha and HIF-2alpha to common target genes is differentially regulated in neuroblastoma: HIF-2alpha promotes an aggressive phenotype." Cancer Cell **10**(5): 413-423.

Hornbeck, P. V., J. M. Kornhauser, S. Tkachev, B. Zhang, E. Skrzypek, B. Murray, V. Latham and M. Sullivan (2012). "PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse." Nucleic acids research **40**(Database issue): D261-D270.

Hornbeck, P. V., B. Zhang, B. Murray, J. M. Kornhauser, V. Latham and E. Skrzypek (2015). "PhosphoSitePlus, 2014: mutations, PTMs and recalibrations." Nucleic acids research **43**(Database issue): D512-D520.

Hoshi, T. and S. Heinemann (2001). "Regulation of cell function by methionine oxidation and reduction." J Physiol **531**(Pt 1): 1-11.

Hsu, J.-M., C.-T. Chen, C.-K. Chou, H.-P. Kuo, L.-Y. Li, C.-Y. Lin, H.-J. Lee, Y.-N. Wang, M. Liu, H.-W. Liao, B. Shi, C.-C. Lai, M. T. Bedford, C.-H. Tsai and M.-C. Hung (2011). "Crosstalk between Arg 1175 methylation and Tyr 1173 phosphorylation negatively modulates EGFR-mediated ERK activation." Nature Cell Biology **13**: 174.

Hsu, P. D., E. S. Lander and F. Zhang (2014). "Development and applications of CRISPR-Cas9 for genome engineering." cell **157**(6): 1262-1278.

Hu, C.-J., A. Sataur, L. Wang, H. Chen and M. C. Simon (2007). "The N-terminal transactivation domain confers target gene specificity of hypoxia-inducible factors HIF-1α and HIF-2α." Molecular Biology of the Cell **18**(11): 4528-4542.

Hu, C.-J., L.-Y. Wang, L. A. Chodosh, B. Keith and M. C. Simon (2003). "Differential roles of hypoxia-inducible factor 1alpha (HIF-1alpha) and HIF-2alpha in hypoxic gene regulation." Molecular and cellular biology **23**(24): 9361-9374.

Hu, C.-J., L.-Y. Wang, L. A. Chodosh, B. Keith and M. C. Simon (2003). "Differential Roles of Hypoxia-Inducible Factor 1α (HIF-1α) and HIF-2α in Hypoxic Gene Regulation." Molecular and cellular biology **23**(24): 9361-9374.

Hu, C. J., A. Sataur, L. Y. Wang, H. Q. Chen and M. C. Simon (2007). "The N-terminal Transactivation domain confers target gene specificity of hypoxia-inducible factors HIF-1 alpha and HIF-2 alpha." Molecular Biology of the Cell **18**(11): 4528-4542.

Huang, D. W., B. T. Sherman and R. A. Lempicki (2008). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc **4**: 44.

Huang, D. W., B. T. Sherman, Q. Tan, J. R. Collins, W. G. Alvord, J. Roayaei, R. Stephens, M. W. Baseler, H. C. Lane and R. A. Lempicki (2007). "The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists." Genome biology **8**(9): R183-R183.

Huang, J., R. Sengupta, A. B. Espejo, M. G. Lee, J. A. Dorsey, M. Richter, S. Opravil, R. Shiekhattar, M. T. Bedford, T. Jenuwein and S. L. Berger (2007). "p53 is regulated by the lysine demethylase LSD1." Nature **449**(7158): 105-108.

Hubbi, M. E., Kshitiz, D. M. Gilkes, S. Rey, C. C. Wong, W. Luo, D. H. Kim, C. V. Dang, A. Levchenko and G. L. Semenza (2013). "A nontranscriptional role for HIF-1alpha as a direct inhibitor of DNA replication." Sci Signal **6**(262): ra10.

Hudson, D. M. and D. R. Eyre (2013). "Collagen prolyl 3-hydroxylation: a major role for a minor post-translational modification?" Connect Tissue Res **54**(4-5): 245-251.

Hulsen, T., J. de Vlieg and W. Alkema (2008). "BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams." BMC Genomics **9**(1): 488.

Hunt, D. F., J. R. Yates, 3rd, J. Shabanowitz, S. Winston and C. R. Hauer (1986). "Protein sequencing by tandem mass spectrometry." Proceedings of the National Academy of Sciences of the United States of America **83**(17): 6233-6237.

Hunter, T. and M. Karin (1992). "The regulation of transcription by phosphorylation." cell **70**(3): 375-387.

Hutagalung, A. H. and P. J. Novick (2011). "Role of Rab GTPases in membrane traffic and cell physiology." Physiological reviews **91**(1): 119-149.

Hwang, A. B., E.-A. Ryu, M. Artan, H.-W. Chang, M. H. Kabir, H.-J. Nam, D. Lee, J.-S. Yang, S. Kim and W. B. Mair (2014). "Feedback regulation via AMPK and HIF-1 mediates ROS-dependent longevity in Caenorhabditis elegans." Proceedings of the National Academy of Sciences **111**(42): E4458-E4467.

Ivan, M., K. Kondo, H. F. Yang, W. Kim, J. Valiando, M. Ohh, A. Salic, J. M. Asara, W. S. Lane and W. G. Kaelin (2001). "HIF alpha targeted for VHL-mediated destruction by proline hydroxylation: Implications for O-2 sensing." Science **292**(5516): 464-468.

Jaakkola, P., D. R. Mole, Y.-M. Tian, M. I. Wilson, J. Gielbert, S. J. Gaskell, A. von Kriegsheim, H. F. Hebestreit, M. Mukherji and C. J. Schofield (2001). "Targeting of HIF-α to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation." Science **292**(5516): 468-472.

Jackman, J. and P. M. O'Connor (2001). "Methods for synchronizing cells at specific stages of the cell cycle." Curr Protoc Cell Biol **Chapter 8**: Unit 8.3.

Jeong, J.-W., M.-K. Bae, M.-Y. Ahn, S.-H. Kim, T.-K. Sohn, M.-H. Bae, M.-A. Yoo, E. J. Song, K.-J. Lee and K.-W. Kim (2002). "Regulation and destabilization of HIF-1α by ARD1-mediated acetylation." cell **111**(5): 709-720.

Jeong, J. W., M. K. Bae, M. Y. Ahn, S. H. Kim, T. K. Sohn, M. H. Bae, M. A. Yoo, E. J. Song, K. J. Lee and K. W. Kim (2002). "Regulation and destabilization of HIF-1 alpha by ARD1-mediated acetylation." Cell **111**(5): 709-720.

Jiang, B. H., E. Rue, G. L. Wang, R. Roe and G. L. Semenza (1996). "Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor 1." J Biol Chem **271**(30): 17771-17778.

Jiang, B. H., J. Z. Zheng, S. W. Leung, R. Roe and G. L. Semenza (1997). "Transactivation and inhibitory domains of hypoxia-inducible factor 1alpha. Modulation of transcriptional activity by oxygen tension." J Biol Chem **272**(31): 19253-19260.

Jiang, H., R. Guo and J. A. Powell-Coffman (2001). "The Caenorhabditis elegans hif-1 gene encodes a bHLH-PAS protein that is required for adaptation to hypoxia." Proceedings of the National Academy of Sciences of the United States of America **98**(14): 7916-7921.

Johnson, R. S., S. A. Martin and K. Biemann (1988). "Collision-induced fragmentation of (M + H)+ ions of peptides. Side chain specific sequence ions." International Journal of Mass Spectrometry and Ion Processes **86**: 137-154.

Jones, D. T., W. R. Taylor and J. M. Thornton (1992). "The rapid generation of mutation data matrices from protein sequences." Comput Appl Biosci **8**(3): 275-282.

Jung, S.-N., W. K. Yang, J. Kim, H. S. Kim, E. J. Kim, H. Yun, H. Park, S. S. Kim, W. Choe and I. Kang (2008). "Reactive oxygen species stabilize hypoxia-inducible factor-1 alpha protein and stimulate transcriptional activity via AMP-activated protein kinase in DU145 human prostate cancer cells." Carcinogenesis **29**(4): 713-721.

Kaidi, A., A. C. Williams and C. Paraskeva (2007). "Interaction between beta-catenin and HIF-1 promotes cellular adaptation to hypoxia." Nat Cell Biol **9**(2): 210-217.

Kallio, P. J., K. Okamoto, S. O'Brien, P. Carrero, Y. Makino, H. Tanaka and L. Poellinger (1998). "Signal transduction in hypoxic cells: inducible nuclear translocation and recruitment of theCBP/p300 coactivator by the hypoxia-induciblefactor-1α." The EMBO journal **17**(22): 6573-6586.

Kalousi, A., I. Mylonis, A. S. Politou, G. Chachami, E. Paraskeva and G. Simos (2010). "Casein kinase 1 regulates human hypoxia-inducible factor HIF-1." Journal of Cell Science **123**(17): 2976-2986.

Kang, X., J. Li, Y. Zou, J. Yi, H. Zhang, M. Cao, E. Yeh and J. Cheng (2010). "PIASy stimulates HIF1α SUMOylation and negatively regulates HIF1α activity in response to hypoxia." Oncogene **29**(41): 5568-5578.

Kebarle, P. and L. Tang (1993). "From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry." Analytical Chemistry **65**(22): 972A-986A.

Kewley, R. J. and M. L. Whitelaw (2005). "Phosphorylation inhibits DNA-binding of alternatively spliced aryl hydrocarbon receptor nuclear translocator." Biochem Biophys Res Commun **338**(1): 660-667.

Kietzmann, T., D. Mennerich and E. Y. Dimova (2016). "Hypoxia-Inducible Factors (HIFs) and Phosphorylation: Impact on Stability, Localization, and Transactivity." Frontiers in cell and developmental biology **4**: 11-11.

Kim, T. K. and J. H. Eberwine (2010). "Mammalian cell transfection: the present and the future." Analytical and bioanalytical chemistry **397**(8): 3173-3178.

Kim, Y., H. J. Nam, J. Lee, D. Y. Park, C. Kim, Y. S. Yu, D. Kim, S. W. Park, J. Bhin and D. Hwang (2016). "Methylation-dependent regulation of HIF-1α stability restricts retinal and tumour angiogenesis." Nature communications **7**.

Kirkpatrick, D. S., C. Denison and S. P. Gygi (2005). "Weighing in on ubiquitin: the expanding role of mass-spectrometry-based proteomics." Nature Cell Biology **7**(8): 750-757.

Koc, A., L. J. Wheeler, C. K. Mathews and G. F. Merrill (2004). "Hydroxyurea arrests DNA replication by a mechanism that preserves basal dNTP pools." J Biol Chem **279**(1): 223-230.

Koh, M. Y., B. G. Darnay and G. Powis (2008). "Hypoxia-associated factor, a novel E3-ubiquitin ligase, binds and ubiquitinates hypoxia-inducible factor 1α, leading to its oxygen-independent degradation." Molecular and cellular biology **28**(23): 7081-7095.

Koh, M. Y., R. Lemos, X. Liu and G. Powis (2011). "The hypoxia-associated factor switches cells from HIF-1α-to HIF-2α-dependent signaling promoting stem cell characteristics, aggressive tumor growth and invasion." Cancer research **71**(11): 4015-4027.

Koh, M. Y., V. Nguyen, R. Lemos, B. G. Darnay, G. M. Kiriakova, M. F. Abdelmelek, T. H. Ho, J. A. Karam, F. A. Monzon and E. Jonasch (2014). "Hypoxia-induced SUMOylation of E3 ligase HAF determines specific activation of HIF-2 in clear cell renal cell carcinoma." Cancer research: canres. 2190.2013.

Koh, M. Y. and G. Powis (2012). "Passing the baton: the HIF switch." Trends Biochem Sci **37**(9): 364-372.

Komander, D., M. J. Clague and S. Urbé (2009). "Breaking the chains: structure and function of the deubiquitinases." Nature reviews Molecular cell biology **10**(8): 550-563.

Komander, D. and M. Rape (2012). "The Ubiquitin Code." Annual review of biochemistry **81**(1): 203-229.

Korgaonkar, S. N., X. Feng, M. D. Ross, T.-c. Lu, V. D'Agati, R. Iyengar, P. E. Klotman and J. C. He (2008). "HIV-1 Upregulates VEGF in Podocytes." Journal of the American Society of Nephrology **19**(5): 877-883.

Koshiji, M., Y. Kageyama, E. A. Pete, I. Horikawa, J. C. Barrett and L. E. Huang (2004). "HIF-1alpha induces cell cycle arrest by functionally counteracting Myc." Embo j **23**(9): 1949-1956.

Krishnamachary, B., S. Berg-Dixon, B. Kelly, F. Agani, D. Feldser, G. Ferreira, N. Iyer, J. LaRusch, B. Pak, P. Taghavi and G. L. Semenza (2003). "Regulation of colon carcinoma cell invasion by hypoxia-inducible factor 1." Cancer Res **63**(5): 1138-1143.

Kubitscheck, U., O. Kückmann, T. Kues and R. Peters (2000). "Imaging and Tracking of Single GFP Molecules in Solution." Biophysical Journal **78**(4): 2170-2179.

Kumar, S., G. Stecher and K. Tamura (2016). "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets." Molecular Biology and Evolution **33**(7): 1870-1874.

Lam, A. J., F. St-Pierre, Y. Gong, J. D. Marshall, P. J. Cranfill, M. A. Baird, M. R. McKeown, J. Wiedenmann, M. W. Davidson, M. J. Schnitzer, R. Y. Tsien and M. Z. Lin (2012). "Improving FRET dynamic range with bright green and red fluorescent proteins." Nature methods **9**(10): 1005-1012.

Land, S. C. and A. R. Tee (2007). "Hypoxia-inducible factor 1alpha is regulated by the mammalian target of rapamycin (mTOR) via an mTOR signaling motif." J Biol Chem **282**(28): 20534-20543.

Lando, D., D. J. Peet, J. J. Gorman, D. A. Whelan, M. L. Whitelaw and R. K. Bruick (2002). "FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor." Genes & development **16**(12): 1466-1471.

Lando, D., D. J. Peet, D. A. Whelan, J. J. Gorman and M. L. Whitelaw (2002). "Asparagine hydroxylation of the HIF transactivation domain: A hypoxic switch." Science **295**(5556): 858-861.

Lange, V., P. Picotti, B. Domon and R. Aebersold (2008). "Selected reaction monitoring for quantitative proteomics: a tutorial." Molecular systems biology **4**: 222-222.

Lanouette, S., V. Mongeon, D. Figeys and J. F. Couture (2014). "The functional diversity of protein lysine methylation." Molecular systems biology **10**(4): 724.

Lanucara, F. and C. E. Eyers (2013). "Top-down mass spectrometry for the analysis of combinatorial post-translational modifications." Mass Spectrometry Reviews **32**(1): 27-42.

Lanucara, F., S. W. Holman, C. J. Gray and C. E. Eyers (2014). "The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics." Nature Chemistry **6**: 281.

Lau, K. W., Y. M. Tian, R. R. Raval, P. J. Ratcliffe and C. W. Pugh (2007). "Target gene selectivity of hypoxia-inducible factor-alpha in renal cancer cells is conveyed by post-DNA-binding mechanisms." British Journal of Cancer **96**(8): 1284-1292.

Lee, J., P. L. Freddolino and Y. Zhang (2017). Ab Initio Protein Structure Prediction. From Protein Structure to Function with Bioinformatics. D. J. Rigden. Dordrecht, Springer Netherlands**:** 3-35.

Lee, J. H., S. A. Jeong, P. Khadka, J. Hong and I. K. Chung (2015). "Involvement of SRSF11 in cell cycle-specific recruitment of telomerase to telomeres at nuclear speckles." Nucleic Acids Res **43**(17): 8435-8451.

Lee, J. W., S. H. Bae, J. W. Jeong, S. H. Kim and K. W. Kim (2004). "Hypoxia-inducible factor (HIF-1)alpha: its protein stability and biological functions." Exp Mol Med **36**(1): 1-12.

Lee, J. Y., J. H. Park, H. J. Choi, H. Y. Won, H. S. Joo, D. H. Shin, M. K. Park, B. Han, K. P. Kim, T. J. Lee, C. M. Croce and G. Kong (2017). "LSD1 demethylates HIF1alpha to inhibit hydroxylation and ubiquitin-mediated degradation in tumor angiogenesis." Oncogene **36**(39): 5512-5521.

Lee, M.-N., S.-N. Lee, S.-H. Kim, B. Kim, B.-K. Jung, J. H. Seo, J.-H. Park, J.-H. Choi, S. H. Yim, M.-R. Lee, J.-G. Park, J.-Y. Yoo, J. H. Kim, S.-T. Lee, H.-M. Kim, S. Ryeom, K.-W. Kim and G. T. Oh (2010). "Roles of Arrest-Defective Protein 1225 and Hypoxia-Inducible Factor 1α in Tumor Growth and Metastasis." JNCI: Journal of the National Cancer Institute **102**(6): 426-442.

Lehmann, W. D., R. Krüger, M. Salek, C.-W. Hung, F. Wolschin and W. Weckwerth (2007). "Neutral Loss-Based Phosphopeptide Recognition: A Collection of Caveats." Journal of Proteome Research **6**(7): 2866-2873.

Lim, J. H., Y. M. Lee, Y. S. Chun, J. Chen, J. E. Kim and J. W. Park (2010). "Sirtuin 1 Modulates Cellular Responses to Hypoxia by Deacetylating Hypoxia-Inducible Factor 1 alpha." Molecular cell **38**(6): 864-878.

Lipinszki, Z., P. Wang, R. Grant, C. Lindon, N. S. Dzhindzhev, P. P. D'Avino, M. R. Przewloka, D. M. Glover and V. Archambault (2014). "Affinity purification of protein complexes from Drosophila embryos in cell cycle studies." Methods Mol Biol **1170**: 571-588.

Liu, H. and J. H. Naismith (2008). "An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol." BMC Biotechnology **8**(1): 91.

Liu, X., Z. Chen, C. Xu, X. Leng, H. Cao, G. Ouyang and W. Xiao (2015). "Repression of hypoxia-inducible factor α signaling by Set7-mediated methylation." Nucleic acids research **43**(10): 5081-5098.

Lombardi, B., N. Rendell, M. Edwards, M. Katan and J. G. Zimmermann (2015). "Evaluation of phosphopeptide enrichment strategies for quantitative TMT analysis of complex network dynamics in cancer-associated cell signalling." EuPA Open Proteomics **6**: 10-15.

Los, G. V., L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit and K. V. Wood (2008). "HaloTag: a novel protein labeling technology for cell imaging and protein analysis." ACS Chem Biol **3**(6): 373-382.

Ma, B., K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby and G. Lajoie (2003). "PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry." Rapid Communications in Mass Spectrometry **17**(20): 2337-2342.

Mahon, P. C., K. Hirota and G. L. Semenza (2001). "FIH-1: a novel protein that interacts with HIF-1α and VHL to mediate repression of HIF-1 transcriptional activity." Genes & development **15**(20): 2675-2686.

Makarov, A. (2000). "Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis." Analytical Chemistry **72**(6): 1156-1162.

Makino, Y., R. Cao, K. Svensson, G. Bertilsson, M. Asman, H. Tanaka, Y. Cao, A. Berkenstam and L. Poellinger (2001). "Inhibitory PAS domain protein is a negative regulator of hypoxia-inducible gene expression." Nature **414**(6863): 550-554.

Makino, Y., A. Kanopka, W. J. Wilson, H. Tanaka and L. Poellinger (2002). "Inhibitory PAS domain protein (IPAS) is a hypoxia-inducible splicing variant of the hypoxia-inducible factor-3 alpha locus." Journal of Biological Chemistry **277**(36): 32405-32408.

Mallick, P. and B. Kuster (2010). "Proteomics: a pragmatic perspective." Nature Biotechnology **28**(7): 695-709.

Mallick, P., M. Schirle, S. S. Chen, M. R. Flory, H. Lee, D. Martin, J. Ranish, B. Raught, R. Schmitt, T. Werner, B. Kuster and R. Aebersold (2006). "Computational prediction of proteotypic peptides for quantitative proteomics." Nature Biotechnology **25**: 125.

Mann, M. and O. N. Jensen (2003). "Proteomic analysis of post-translational modifications." Nature Biotechnology **21**(3): 255-261.

Mann, M., S.-E. Ong, M. Grønborg, H. Steen, O. N. Jensen and A. Pandey (2002). "Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome." Trends in Biotechnology **20**(6): 261-268.

Manning, G., D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsanam (2002). "The protein kinase complement of the human genome." Science **298**(5600): 1912-1934.

Marsin, A. S., C. Bouzin, L. Bertrand and L. Hue (2002). "The stimulation of glycolysis by hypoxia in activated monocytes is mediated by AMP-activated protein kinase and inducible 6-phosphofructo-2-kinase." J Biol Chem **277**(34): 30778-30783.

Masoud, G. N. and W. Li (2015). "HIF-1α pathway: role, regulation and intervention for cancer therapy." Acta Pharmaceutica Sinica. B **5**(5): 378-389.

Masson, N., R. J. Appelhoff, J. R. Tuckerman, Y.-m. Tian, H. Demol, M. Puype, J. Vandekerckhove, P. J. Ratcliffe and C. W. Pugh (2004). "The HIF prolyl hydroxylase PHD3 is a potential substrate of the TRiC chaperonin." Febs Letters **570**(1): 166-170.

Masson, N. and P. J. Ratcliffe (2003). "HIF prolyl and asparaginyl hydroxylases in the biological response to intracellular O-2 levels." Journal of Cell Science **116**(15): 3041-3049.

Masson, N., C. Willam, P. H. Maxwell, C. W. Pugh and P. J. Ratcliffe (2001). "Independent function of two destruction domains in hypoxia-inducible factor-α chains activated by prolyl hydroxylation." The EMBO journal **20**(18): 5197-5206.

Maxwell, P. H., M. S. Wiesener, G.-W. Chang, S. C. Clifford, E. C. Vaux, M. E. Cockman, C. C. Wykoff, C. W. Pugh, E. R. Maher and P. J. Ratcliffe (1999). "The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis." Nature **399**(6733): 271-275.

Maynard, M. A., H. Qi, J. Chung, E. H. Lee, Y. Kondo, S. Hara, R. C. Conaway, J. W. Conaway and M. Ohh (2003). "Multiple splice variants of the human HIF-3 alpha locus are targets of the von Hippel-Lindau E3 ubiquitin ligase complex." J Biol Chem **278**(13): 11032-11040.

McDowell, G. S. and A. Philpott (2013). "Non-canonical ubiquitylation: Mechanisms and consequences." The International Journal of Biochemistry & Cell Biology **45**(8): 1833-1842.

McInerney, P., P. Adams and M. Z. Hadi (2014). "Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase." Molecular Biology International **2014**: 8.

McLafferty, F. W., D. M. Horn, K. Breuker, Y. Ge, M. A. Lewis, B. Cerda, R. A. Zubarev and B. K. Carpenter (2001). "Electron capture dissociation of gaseous multiply charged ions by fourier-transform ion cyclotron resonance." Journal of the American Society for Mass Spectrometry **12**(3): 245-249.

Meek, D. W. and C. W. Anderson (2009). "Posttranslational modification of p53: cooperative integrators of function." Cold Spring Harb Perspect Biol **1**(6): a000950.

Melendez, J., M. Grogg and Y. Zheng (2011). "Signaling role of Cdc42 in regulating mammalian physiology." J Biol Chem **286**(4): 2375-2381.

Meyer, B., D. G. Papasotiriou and M. Karas (2011). "100% protein sequence coverage: a modern form of surrealism in proteomics." Amino Acids **41**(2): 291-310.

Michalski, A., E. Damoc, O. Lange, E. Denisov, D. Nolting, M. Müller, R. Viner, J. Schwartz, P. Remes, M. Belford, J.-J. Dunyach, J. Cox, S. Horning, M. Mann and A. Makarov (2012). "Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes." Molecular & cellular proteomics : MCP **11**(3): O111.013698-O013111.013698.

Minet, E., T. Arnould, G. Michel, I. Roland, D. Mottet, M. Raes, J. Remacle and C. Michiels (2000). "ERK activation upon hypoxia: involvement in HIF-1 activation." FEBS Lett **468**(1): 53-58.

Mohsenin, V. and J. L. Gee (1989). "Oxidation of alpha 1-protease inhibitor: role of lipid peroxidation products." Journal of Applied Physiology **66**(5): 2211-2215.

Mole, D. R., C. Blancher, R. R. Copley, P. J. Pollard, J. M. Gleadle, J. Ragoussis and P. J. Ratcliffe (2009). "Genome-wide Association of Hypoxia-inducible Factor (HIF)-1 alpha and HIF-2 alpha DNA Binding with Expression Profiling of Hypoxia-inducible Transcripts." Journal of Biological Chemistry **284**(25): 16767-16775.

Moroz, E., S. Carlin, K. Dyomina, S. Burke, H. T. Thaler, R. Blasberg and I. Serganova (2009). "Real-time imaging of HIF-1α stabilization and degradation." PLoS One **4**(4): e5077.

Mottet, D., V. Dumont, Y. Deccache, C. Demazy, N. Ninane, M. Raes and C. Michiels (2003). "Regulation of hypoxia-inducible factor-1alpha protein level during hypoxic conditions by the phosphatidylinositol 3-kinase/Akt/glycogen synthase kinase 3beta pathway in HepG2 cells." J Biol Chem **278**(33): 31277-31285.

Murray-Rust, T. A., N. J. Oldham, K. S. Hewitson and C. J. Schofield (2006). "Purified recombinant hARD1 does not catalyse acetylation of Lys532 of HIF-1α fragments in vitro." Febs Letters **580**(8): 1911-1918.

Muyldermans, S. (2013). "Nanobodies: Natural Single-Domain Antibodies." Annual review of biochemistry **82**(1): 775-797.

Muz, B., P. de la Puente, F. Azab and A. K. Azab (2015). "The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy." Hypoxia (Auckland, N.Z.) **3**: 83-92.

Myllyharju, J. (2003). "Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis." Matrix Biology **22**(1): 15-24.

Mylonis, I., G. Chachami, E. Paraskeva and G. Simos (2008). "Atypical CRM1-dependent nuclear export signal mediates regulation of hypoxia-inducible factor-1α by MAPK." Journal of Biological Chemistry **283**(41): 27620-27627.

Mylonis, I., G. Chachami, M. Samiotaki, G. Panayotou, E. Paraskeva, A. Kalousi, E. Georgatsou, S. Bonanou and G. Simos (2006). "Identification of MAPK phosphorylation sites and their role in the localization and activity of hypoxia-inducible factor-1 alpha." Journal of Biological Chemistry **281**(44): 33095-33106.

Nilsson, G. E. and G. M. C. Renshaw (2004). "Hypoxic survival strategies in two fishes: extreme anoxia tolerance in the North European crucian carp and natural hypoxic preconditioning in a coral-reef shark." Journal of Experimental Biology **207**(18): 3131-3139.

Nilsson, H., A. Jogi, S. Beckman, A. L. Harris, L. Poellinger and S. Pahlman (2005). "HIF-2alpha expression in human fetal paraganglia and neuroblastoma: relation to sympathetic differentiation, glucose deficiency, and hypoxia." Exp Cell Res **303**(2): 447-456.

Nurtjahja-Tjendraputra, E., D. Fu, J. M. Phang and D. R. Richardson (2007). "Iron chelation regulates cyclin D1 expression via the proteasome: a link to iron deficiency-mediated growth suppression." Blood **109**(9): 4045-4054.

O'Rourke, J. F., Y.-M. Tian, P. J. Ratcliffe and C. W. Pugh (1999). "Oxygen-regulated and Transactivating Domains in Endothelial PAS Protein 1: Comparison with Hypoxia-inducible Factor-1α." Journal of Biological Chemistry **274**(4): 2060-2071.

Ohana, R. F., R. Hurst, J. Vidugiriene, M. R. Slater, K. V. Wood and M. Urh (2011). "HaloTag-based purification of functional human kinases from mammalian cells." Protein Expr Purif **76**(2): 154-164.

Ohishi, K., N. Inoue and T. Kinoshita (2001). "PIG-S and PIG-T, essential for GPI anchor attachment to proteins, form a complex with GAA1 and GPI8." The EMBO journal **20**(15): 4088-4098.

Ohishi, K., N. Inoue, Y. Maeda, J. Takeda, H. Riezman and T. Kinoshita (2000). "Gaa1p and gpi8p are components of a glycosylphosphatidylinositol (GPI) transamidase that mediates attachment of GPI to proteins." Molecular Biology of the Cell **11**(5): 1523-1533.

Old, W. M., K. Meyer-Arendt, L. Aveline-Wolf, K. G. Pierce, A. Mendoza, J. R. Sevinsky, K. A. Resing and N. G. Ahn (2005). "Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics." Molecular & Cellular Proteomics **4**(10): 1487-1502.

Olsen, J. V., B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen and M. Mann (2006). "Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks." cell **127**(3): 635-648.

Olsen, J. V., B. Macek, O. Lange, A. Makarov, S. Horning and M. Mann (2007). "Higher-energy C-trap dissociation for peptide modification analysis." Nature methods **4**: 709.

Ong, S.-E., B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey and M. Mann (2002). "Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics." Molecular & Cellular Proteomics **1**(5): 376-386.

Page, E. L., G. A. Robitaille, J. Pouyssegur and D. E. Richard (2002). "Induction of hypoxia-inducible factor-1alpha by transcriptional and translational mechanisms." J Biol Chem **277**(50): 48403-48409.

Paizs, B. and S. Suhai (2005). "Fragmentation pathways of protonated peptides." Mass Spectrometry Reviews **24**(4): 508-548.

Palazon, A., A. W. Goldrath, V. Nizet and R. S. Johnson (2014). "HIF transcription factors, inflammation, and immunity." Immunity **41**(4): 518-528.

Paltoglou, S. and B. Roberts (2007). "HIF-1α and EPAS ubiquitination mediated by the VHL tumour suppressor involves flexibility in the ubiquitination mechanism, similar to other RING E3 ligases." Oncogene **26**(4): 604-609.

Pangou, E., C. Befani, I. Mylonis, M. Samiotaki, G. Panayotou, G. Simos and P. Liakos (2016). "HIF-2α phosphorylation by CK1δ promotes erythropoietin secretion in liver cancer cells under hypoxia." J Cell Sci **129**(22): 4213-4226.

Pauling, L., R. B. Corey and H. R. Branson (1951). "The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain." Proceedings of the National Academy of Sciences **37**(4): 205-211.

Peng, X. H., P. Karna, Z. Cao, B. H. Jiang, M. Zhou and L. Yang (2006). "Cross-talk between epidermal growth factor receptor and hypoxia-inducible factor-1alpha signal pathways increases resistance to apoptosis by up-regulating survivin gene expression." J Biol Chem **281**(36): 25903-25914.

Perdivara, I., L. J. Deterding, M. Przybylski and K. B. Tomer (2010). "Mass spectrometric identification of oxidative modifications of tryptophan residues in proteins: chemical artifact or post-translational modification?" J Am Soc Mass Spectrom **21**(7): 1114-1117.

Perkins, D. N., D. J. Pappin, D. M. Creasy and J. S. Cottrell (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data." Electrophoresis **20**(18): 3551-3567.

Pham, P., M. B. Smolka, P. Calabrese, A. Landolph, K. Zhang, H. Zhou and M. F. Goodman (2008). "Impact of phosphorylation and phosphorylation-null mutants on the activity and deamination specificity of activation-induced cytidine deaminase." J Biol Chem **283**(25): 17428-17439.

Phuyal, S. and H. Farhan (2019). "Multifaceted Rho GTPase Signaling at the Endomembranes." Frontiers in cell and developmental biology **7**(127).

Pierleoni, A., P. L. Martelli and R. Casadio (2008). "PredGPI: a GPI-anchor predictor." BMC Bioinformatics **9**: 392-392.

Pihlajaniemi, T., R. Myllyla and K. I. Kivirikko (1991). "Prolyl 4-hydroxylase and its role in collagen synthesis." J Hepatol **13 Suppl 3**: S2-7.

Pugh, C. W., J. F. O'Rourke, M. Nagao, J. M. Gleadle and P. J. Ratcliffe (1997). "Activation of hypoxia-inducible factor-1; definition of regulatory domains within the alpha subunit." J Biol Chem **272**(17): 11205-11214.

Qi, H. H., P. P. Ongusaha, J. Myllyharju, D. Cheng, O. Pakkanen, Y. Shi, S. W. Lee, J. Peng and Y. Shi (2008). "Prolyl 4-hydroxylation regulates Argonaute 2 stability." Nature **455**(7211): 421-424.

Rainer, M. and G. K. Bonn (2015). "Enrichment of phosphorylated peptides and proteins by selective precipitation methods." Bioanalysis **7**(2): 243-252.

Raman, S., R. Vernon, J. Thompson, M. Tyka, R. Sadreyev, J. Pei, D. Kim, E. Kellogg, F. DiMaio, O. Lange, L. Kinch, W. Sheffler, B. H. Kim, R. Das, N. V. Grishin and D. Baker (2009). "Structure prediction for CASP8 with all-atom refinement using Rosetta." Proteins **77 Suppl 9**: 89-99.

Rankin, E. B., K. C. Fuh, L. Castellini, K. Viswanathan, E. C. Finger, A. N. Diep, E. L. LaGory, M. S. Kariolis, A. Chan, D. Lindgren, H. Axelson, Y. R. Miao, A. J. Krieg and A. J. Giaccia (2014). "Direct regulation of GAS6/AXL signaling by HIF promotes renal metastasis through SRC and MET." Proceedings of the National Academy of Sciences **111**(37): 13373-13378.

Raval, R. R., K. W. Lau, M. G. B. Tran, H. M. Sowter, S. J. Mandriota, J.-L. Li, C. W. Pugh, P. H. Maxwell, A. L. Harris and P. J. Ratcliffe (2005). "Contrasting Properties of Hypoxia-Inducible Factor 1 (HIF-1) and HIF-2 in von Hippel-Lindau-Associated Renal Cell Carcinoma." Molecular and cellular biology **25**(13): 5675-5686.

Rolfs, A., I. Kvietikova, M. Gassmann and R. H. Wenger (1997). "Oxygen-regulated transferrin expression is mediated by hypoxia-inducible factor-1." J Biol Chem **272**(32): 20055-20062.

Rosenberger, C., S. Mandriota, J. S. Jurgensen, M. S. Wiesener, J. H. Horstrup, U. Frei, P. J. Ratcliffe, P. H. Maxwell, S. Bachmann and K. U. Eckardt (2002). "Expression of hypoxia-inducible factor-1alpha and -2alpha in hypoxic and ischemic rat kidneys." J Am Soc Nephrol **13**(7): 1721-1732.

Ross, P. L., Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson and D. J. Pappin (2004). "Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents." Molecular & cellular proteomics : MCP **3**(12): 1154-1169.

Sarkar, G. and S. S. Sommer (1990). "The "megaprimer" method of site-directed mutagenesis." Biotechniques **8**(4): 404-407.

Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak and A. Cardona (2012). "Fiji: an open-source platform for biological-image analysis." Nat Methods **9**(7): 676-682.

Schmitz, S. U., P. Grote and B. G. Herrmann (2016). "Mechanisms of long noncoding RNA function in development and disease." Cellular and molecular life sciences : CMLS **73**(13): 2491-2509.

Schneider, M., A. Belsom and J. Rappsilber (2018). "Protein Tertiary Structure by Crosslinking/Mass Spectrometry." Trends Biochem Sci **43**(3): 157-169.

Schober, A. S. and E. Berra (2016). "DUBs, New Members in the Hypoxia Signaling clUb." Frontiers in Oncology **6**(53).

Schödel, J., S. Oikonomopoulos, J. Ragoussis, C. W. Pugh, P. J. Ratcliffe and D. R. Mole (2011). "High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq." Blood **117**(23): e207-e217.

Schwartz, J. C., M. W. Senko and J. E. P. Syka (2002). "A two-dimensional quadrupole ion trap mass spectrometer." Journal of the American Society for Mass Spectrometry **13**(6): 659-669.

Schwartz, S. L., C. Cao, O. Pylypenko, A. Rak and A. Wandinger-Ness (2007). "Rab GTPases at a glance." Journal of Cell Science **120**(22): 3905-3910.

Scigelova, M., M. Hornshaw, A. Giannakopulos and A. Makarov (2011). "Fourier transform mass spectrometry." Molecular & cellular proteomics : MCP **10**(7): M111.009431-M009111.009431.

Semenza, G. L. (2003). "Targeting HIF-1 for cancer therapy." Nat Rev Cancer **3**(10): 721-732.

Semenza, G. L. (2003). "Targeting HIF-1 for cancer therapy." Nature Reviews Cancer **3**(10): 721-732.

Semenza, G. L. and G. L. Wang (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." Molecular and cellular biology **12**(12): 5447-5454.

Senko, M. W., P. M. Remes, J. D. Canterbury, R. Mathur, Q. Song, S. M. Eliuk, C. Mullen, L. Earley, M. Hardman, J. D. Blethrow, H. Bui, A. Specht, O. Lange, E. Denisov, A. Makarov, S. Horning and V. Zabrouskov (2013). "Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates." Analytical Chemistry **85**(24): 11710-11714.

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome research **13**(11): 2498-2504.

Shao, D., S.-I. Oka, T. Liu, P. Zhai, T. Ago, S. Sciarretta, H. Li and J. Sadoshima (2014). "A redox-dependent mechanism for regulation of AMPK activation by Thioredoxin1 during energy starvation." Cell Metabolism **19**(2): 232-245.

Sharma, V. K. and N. J. D. Graham (2010). "Oxidation of Amino Acids, Peptides and Proteins by Ozone: A Review." Ozone: Science & Engineering **32**(2): 81-90.

Shizuya, H. and H. Kouros-Mehr (2001). "The development and applications of the bacterial artificial chromosome cloning system." Keio J Med **50**(1): 26-30.

Siggers, T. and R. Gordan (2014). "Protein-DNA binding: complexities and multi-protein codes." Nucleic Acids Res **42**(4): 2099-2111.

Simonson, T. S., Y. Yang, C. D. Huff, H. Yun, G. Qin, D. J. Witherspoon, Z. Bai, F. R. Lorenzo, J. Xing, L. B. Jorde, J. T. Prchal and R. Ge (2010). "Genetic evidence for high-altitude adaptation in Tibet." Science **329**(5987): 72-75.

Smith, A. L., D. B. Friedman, H. Yu, R. H. Carnahan and A. B. Reynolds (2011). "ReCLIP (Reversible Cross-Link Immuno-Precipitation): An Efficient Method for Interrogation of Labile Protein Complexes." PLoS One **6**(1): e16206.

Smits, A. H., P. W. Jansen, I. Poser, A. A. Hyman and M. Vermeulen (2013). "Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics." Nucleic Acids Res **41**(1): e28.

Smythies, J. A., M. Sun, N. Masson, R. Salama, P. D. Simpson, E. Murray, V. Neumann, M. E. Cockman, H. Choudhry, P. J. Ratcliffe and D. R. Mole (2019). "Inherent DNA-binding specificities of the HIF-1α and HIF-2α transcription factors in chromatin." EMBO reports **20**(1): e46401.

Soboleski, M. R., J. Oaks and W. P. Halford (2005). "Green fluorescent protein is a quantitative reporter of gene expression in individual eukaryotic cells." Faseb j **19**(3): 440-442.

Sobott, F., S. J. Watt, J. Smith, M. J. Edelmann, H. B. Kramer and B. M. Kessler (2009). "Comparison of CID Versus ETD Based MS/MS Fragmentation for the Analysis of Protein Ubiquitination." Journal of the American Society for Mass Spectrometry **20**(9): 1652-1659.

Song, Y., F. DiMaio, R. Y. Wang, D. Kim, C. Miles, T. Brunette, J. Thompson and D. Baker (2013). "High-resolution comparative modeling with RosettaCM." Structure **21**(10): 1735-1742.

Stafford, G. C., P. E. Kelley, J. E. P. Syka, W. E. Reynolds and J. F. J. Todd (1984). "Recent improvements in and analytical applications of advanced ion trap technology." International Journal of Mass Spectrometry and Ion Processes **60**(1): 85-98.

Stroka, D. M., T. Burkhardt, I. Desbaillets, R. H. Wenger, D. A. Neil, C. Bauer, M. Gassmann and D. Candinas (2001). "HIF-1 is expressed in normoxic tissue and displays an organ-specific regulation under systemic hypoxia." Faseb j **15**(13): 2445-2453.

Sun, L. and Z. J. Chen (2004). "The novel functions of ubiquitination in signaling." Current Opinion in Cell Biology **16**(2): 119-126.

Sutherland, B. W., J. Toews and J. Kast (2008). "Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions." J Mass Spectrom **43**(6): 699-715.

Syka, J. E. P., J. J. Coon, M. J. Schroeder, J. Shabanowitz and D. F. Hunt (2004). "Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry." Proceedings of the National Academy of Sciences of the United States of America **101**(26): 9528-9533.

Szklarczyk, D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen and C. von Mering (2015). "STRING v10: protein-protein interaction networks, integrated over the tree of life." Nucleic acids research **43**(Database issue): D447-D452.

Szklarczyk, D., J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, L. J. Jensen and C. von Mering (2017). "The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible." Nucleic acids research **45**(D1): D362-D368.

Tak, I.-u.-R., F. Ali, J. S. Dar, A. R. Magray, B. A. Ganai and M. Z. Chishti (2019). Chapter 1 - Posttranslational Modifications of Proteins and Their Role in Biological Processes and Associated Diseases. Protein Modificomics. T. A. Dar and L. R. Singh, Academic Press**:** 1-35.

Tammsalu, T., I. Matic, E. G. Jaffray, A. F. Ibrahim, M. H. Tatham and R. T. Hay (2015). "Proteome-wide identification of SUMO modification sites by mass spectrometry." Nat Protoc **10**(9): 1374-1388.

Tanimoto, K., Y. Makino, T. Pereira and L. Poellinger (2000). "Mechanism of regulation of the hypoxia-inducible factor-l alpha by the von Hippel-Lindau tumor suppressor protein." Embo Journal **19**(16): 4298-4309.

Tanimoto, K., Y. Makino, T. Pereira and L. Poellinger (2000). "Mechanism of regulation of the hypoxia-inducible factor-1α by the von Hippel-Lindau tumor suppressor protein." The EMBO journal **19**(16): 4298-4309.

Tasaki, T. and Y. T. Kwon (2007). "The mammalian N-end rule pathway: new insights into its components and physiological roles." Trends Biochem Sci **32**(11): 520-528.

Taus, T., T. Kocher, P. Pichler, C. Paschke, A. Schmidt, C. Henrich and K. Mechtler (2011). "Universal and confident phosphorylation site localization using phosphoRS." J Proteome Res **10**(12): 5354-5362.

Taylor, S., J. Bagnall, D. Mason, R. Levy, D. Fernig and V. See (2016). "Differential sub-nuclear distribution of hypoxia-inducible factors (HIF)-1 and-2 alpha impacts on their stability and mobility." Open biology **6**(9): 160195.

Taylor, S. W., E. Fahy, J. Murray, R. A. Capaldi and S. S. Ghosh (2003). "Oxidative post-translational modification of tryptophan residues in cardiac mitochondrial proteins." J Biol Chem **278**(22): 19587-19590.

Terpe, K. (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems." Appl Microbiol Biotechnol **60**(5): 523-533.

The UniProt Consortium (2017). "UniProt: the universal protein knowledgebase." Nucleic acids research **45**(D1): D158-D169.

Thomas, L. W. and M. Ashcroft (2019). "Exploring the molecular interface between hypoxia-inducible factor signalling and mitochondria." Cellular and molecular life sciences : CMLS **76**(9): 1759-1777.

Thompson, A., J. Schäfer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann and C. Hamon (2003). "Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS." Analytical Chemistry **75**(8): 1895-1904.

Tian, H., S. L. McKnight and D. W. Russell (1997). "Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells." Genes Dev **11**(1): 72-82.

Timmermann, S., H. Lehrmann, A. Polesskaya and A. Harel-Bellan (2001). "Histone acetylation and disease." Cellular and molecular life sciences : CMLS **58**(5-6): 728-736.

To, K. K. W., O. A. Sedelnikova, M. Samons, W. M. Bonner and L. E. Huang (2006). "The phosphorylation status of PAS-B distinguishes HIF-1α from HIF-2α in NBS1 repression." The EMBO journal **25**(20): 4784-4794.

Trinkle-Mulcahy, L. (2012). "Resolving protein interactions and complexes by affinity purification followed by label-based quantitative mass spectrometry." PROTEOMICS **12**(10): 1623-1638.

Trinkle-Mulcahy, L., S. Boulon, Y. W. Lam, R. Urcia, F. M. Boisvert, F. Vandermoere, N. A. Morrice, S. Swift, U. Rothbauer, H. Leonhardt and A. Lamond (2008). "Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes." J Cell Biol **183**(2): 223-239.

Tsiatsiani, L. and A. J. R. Heck (2015). "Proteomics beyond trypsin." The FEBS Journal **282**(14): 2612-2626.

Tyanova, S., T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann and J. Cox (2016). "The Perseus computational platform for comprehensive analysis of (prote)omics data." Nat Methods **13**(9): 731-740.

Tyers, M. and M. Mann (2003). "From genomics to proteomics." Nature **422**(6928): 193-197.

Uchida, T., F. Rossignol, M. A. Matthay, R. Mounier, S. Couette, E. Clottes and C. Clerici (2004). "Prolonged hypoxia differentially regulates hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha expression in lung epithelial cells: implication of natural antisense HIF-1alpha." J Biol Chem **279**(15): 14871-14878.

Udeshi, N. D., P. Mertins, T. Svinkina and S. A. Carr (2013). "Large-scale identification of ubiquitination sites by mass spectrometry." Nat Protoc **8**(10): 1950-1960.

Välikangas, T., T. Suomi and L. L. Elo (2018). "A systematic evaluation of normalization methods in quantitative label-free proteomics." Briefings in bioinformatics **19**(1): 1-11.

van Hagen, M., R. M. Overmeer, S. S. Abolvardi and A. C. O. Vertegaal (2010). "RNF4 and VHL regulate the proteasomal degradation of SUMO-conjugated Hypoxia-Inducible Factor-2α." Nucleic acids research **38**(6): 1922-1931.

Vanichapol, T., K. Leelawat and S. Hongeng (2015). "Hypoxia enhances cholangiocarcinoma invasion through activation of hepatocyte growth factor receptor and the extracellular signal-regulated kinase signaling pathway." Molecular medicine reports **12**(3): 3265-3272.

Varshavsky, A. (2011). "The N-end rule pathway and regulation by proteolysis." Protein science : a publication of the Protein Society **20**(8): 1298-1345.

Vivancos, A. P., E. A. Castillo, B. Biteau, C. Nicot, J. Ayté, M. B. Toledano and E. Hidalgo (2005). "A cysteine-sulfinic acid in peroxiredoxin regulates H₂O₂-sensing by the antioxidant Pap1 pathway." Proceedings of the National Academy of Sciences of the United States of America **102**(25): 8875-8880.

Vranka, J. A., L. Y. Sakai and H. P. Bachinger (2004). "Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes." J Biol Chem **279**(22): 23615-23621.

Wandinger-Ness, A. and M. Zerial (2014). "Rab proteins and the compartmentalization of the endosomal system." Cold Spring Harb Perspect Biol **6**(11): a022616-a022616.

Wang, G., Z. Yu, Y. Zhen, T. Mi, Y. Shi, J. Wang, M. Wang and S. Sun (2014). "Molecular characterisation, evolution and expression of hypoxia-inducible factor in Aurelia sp.1." PLoS One **9**(6): e100057-e100057.

Wang, G. L., B. H. Jiang, E. A. Rue and G. L. Semenza (1995). "Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension." Proceedings of the National Academy of Sciences of the United States of America **92**(12): 5510-5514.

Wang, G. L. and G. L. Semenza (1993). "Desferrioxamine induces erythropoietin gene expression and hypoxia-inducible factor 1 DNA-binding activity: implications for models of hypoxia signal transduction." Blood **82**(12): 3610-3615.

Wang, G. L. and G. L. Semenza (1995). "Purification and characterization of hypoxia-inducible factor 1." J Biol Chem **270**(3): 1230-1237.

Wang, M., M. Weiss, M. Simonovic, G. Haertinger, S. P. Schrimpf, M. O. Hengartner and C. von Mering (2012). "PaxDb, a database of protein abundance averages across all three domains of life." Molecular & cellular proteomics : MCP **11**(8): 492-500.

Wang, Y., A. Satoh, G. Warren and H. H. Meyer (2004). "VCIP135 acts as a deubiquitinating enzyme during p97-p47-mediated reassembly of mitotic Golgi fragments." J Cell Biol **164**(7): 973-978.

Wardle, F. C. and H. Tan (2015). "A ChIP on the shoulder? Chromatin immunoprecipitation and validation strategies for ChIP antibodies." F1000Research **4**: 235-235.

Warfel, N. A., N. G. Dolloff, D. T. Dicker, J. Malysz and W. S. El-Deiry (2013). "CDK1 stabilizes HIF-1α via direct phosphorylation of Ser668 to promote tumor growth." Cell cycle **12**(23): 3689-3701.

Warnecke, A., T. Sandalova, A. Achour and R. A. Harris (2014). "PyTMs: a useful PyMOL plugin for modeling common post-translational modifications." BMC Bioinformatics **15**(1): 370.

Washburn, M. P., D. Wolters and J. R. Yates (2001). "Large-scale analysis of the yeast proteome by multidimensional protein identification technology." Nature Biotechnology **19**(3): 242-247.

Weill, U., G. Krieger, Z. Avihou, R. Milo, M. Schuldiner and D. Davidi (2019). "Assessment of GFP Tag Position on Protein Localization and Growth Fitness in Yeast." Journal of Molecular Biology **431**(3): 636-641.

Wenger, R. H., D. P. Stiehl and G. Camenisch (2005). "Integration of Oxygen Signaling at the Consensus HRE." Science's STKE **2005**(306): re12-re12.

Wiesener, M. S., J. S. Jurgensen, C. Rosenberger, C. K. Scholze, J. H. Horstrup, C. Warnecke, S. Mandriota, I. Bechmann, U. A. Frei, C. W. Pugh, P. J. Ratcliffe, S. Bachmann, P. H. Maxwell and K. U. Eckardt (2003). "Widespread hypoxia-inducible expression of HIF-2alpha in distinct cell populations of different organs." Faseb j **17**(2): 271-273.

Wiesener, M. S., H. Turley, W. E. Allen, C. Willam, K. U. Eckardt, K. L. Talks, S. M. Wood, K. C. Gatter, A. L. Harris, C. W. Pugh, P. J. Ratcliffe and P. H. Maxwell (1998). "Induction of endothelial PAS domain protein-1 by hypoxia: characterization and comparison with hypoxia-inducible factor-1alpha." Blood **92**(7): 2260-2268.

Williams, C., M. van den Berg, R. R. Sprenger and B. Distel (2007). "A conserved cysteine is essential for Pex4p-dependent ubiquitination of the peroxisomal import receptor Pex5p." J Biol Chem **282**(31): 22534-22543.

Williams, J. D., K. A. Cox, R. G. Cooks, S. A. McLuckey, K. J. Hart and D. E. Goeringer (1994). "Resonance Ejection Ion Trap Mass Spectrometry and Nonlinear Field Contributions: The Effect of Scan Direction on Mass Resolution." Analytical Chemistry **66**(5): 725-729.

Wu, D., N. Potluri, J. Lu, Y. Kim and F. Rastinejad (2015). "Structural integration in hypoxia-inducible factors." Nature **524**(7565): 303-308.

Xenaki, G., T. Ontikatze, R. Rajendran, I. J. Stratford, C. Dive, M. Krstic-Demonacos and C. Demonacos (2008). "PCAF is an HIF-1α cofactor that regulates p53 transcriptional activity in hypoxia." Oncogene **27**(44): 5785-5796.

Xie, L., A. Yin, A. S. Nichenko, A. M. Beedle, J. A. Call and H. Yin (2018). "Transient HIF2A inhibition promotes satellite cell proliferation and muscle regeneration." The Journal of clinical investigation **128**(6): 2339-2355.

Xiong, G., R. L. Stewart, J. Chen, T. Gao, T. L. Scott, L. M. Samayoa, K. O'Connor, A. N. Lane and R. Xu (2018). "Collagen prolyl 4-hydroxylase 1 is essential for HIF-1α stabilization and TNBC chemoresistance." Nature communications **9**(1): 4456.

Xu, D., Y. Yao, L. Lu, M. Costa and W. Dai (2010). "Plk3 functions as an essential component of the hypoxia regulatory pathway by direct phosphorylation of HIF-1α." Journal of Biological Chemistry **285**(50): 38944-38950.

Yasinska, I. M. and V. V. Sumbayev (2003). "S-nitrosation of Cys-800 of HIF-1α protein activates its interaction with p300 and stimulates its transcriptional activity." Febs Letters **549**(1-3): 105-109.

Yates, J. R., C. I. Ruse and A. Nakorchevsky (2009). "Proteomics by Mass Spectrometry: Approaches, Advances, and Applications." Annual Review of Biomedical Engineering **11**(1): 49-79.

Yi, X., Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. P. Cuo, J. E. Pool, X. Xu, H. Jiang, N. Vinckenbosch, T. S. Korneliussen, H. Zheng, T. Liu, W. He, K. Li, R. Luo, X. Nie, H. Wu, M. Zhao, H. Cao, J. Zou, Y. Shan, S. Li, Q. Yang, Asan, P. Ni, G. Tian, J. Xu, X. Liu, T. Jiang, R. Wu, G. Zhou, M. Tang, J. Qin, T. Wang, S. Feng, G. Li, Huasang, J. Luosang, W. Wang, F. Chen, Y. Wang, X. Zheng, Z. Li, Z. Bianba, G. Yang, X. Wang, S. Tang, G. Gao, Y. Chen, Z. Luo, L.

Gusang, Z. Cao, Q. Zhang, W. Ouyang, X. Ren, H. Liang, H. Zheng, Y. Huang, J. Li, L. Bolund, K. Kristiansen, Y. Li, Y. Zhang, X. Zhang, R. Li, S. Li, H. Yang, R. Nielsen, J. Wang and J. Wang (2010). "Sequencing of 50 human exomes reveals adaptation to high altitude." Science (New York, N.Y.) **329**(5987): 75-78.

Yu, Y.-Q., M. Gilar, P. J. Lee, E. S. P. Bouvier and J. C. Gebler (2003). "Enzyme-Friendly, Mass Spectrometry-Compatible Surfactant for In-Solution Enzymatic Digestion of Proteins." Analytical Chemistry **75**(21): 6023-6028.

Zhang, J., L. Xin, B. Shan, W. Chen, M. Xie, D. Yuen, W. Zhang, Z. Zhang, G. A. Lajoie and B. Ma (2012). "PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification." Molecular & cellular proteomics : MCP **11**(4): M111.010587.

Zhang, J., L. Xin, B. Shan, W. Chen, M. Xie, D. Yuen, W. Zhang, Z. Zhang, G. A. Lajoie and B. Ma (2012). "PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification." Molecular & cellular proteomics : MCP **11**(4): M111.010587-M010111.010587.

Zhang, Y., S. B. Ficarro, S. Li and J. A. Marto (2009). "Optimized Orbitrap HCD for Quantitative Analysis of Phosphopeptides." Journal of the American Society for Mass Spectrometry **20**(8): 1425-1434.

Zhang, Y., B. R. Fonslow, B. Shan, M.-C. Baek and J. R. Yates, 3rd (2013). "Protein analysis by shotgun/bottom-up proteomics." Chemical reviews **113**(4): 2343-2394.

Zubarev, R. A., N. L. Kelleher and F. W. McLafferty (1998). "Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process." Journal of the American Chemical Society **120**(13): 3265-3266.

Zurlo, G., J. Guo, M. Takada, W. Wei and Q. Zhang (2016). "New Insights into Protein Hydroxylation and Its Important Role in Human Diseases." Biochimica et biophysica acta **1866**(2): 208-220.