Investigating the Role of Cysteamine Dioxygenase in Oxygen Sensing



Much ADO about O_2 Sensing

A thesis submitted to the Board of the Mathematical, Physical, and Life Sciences Division in partial fulfilment of the degree of Doctor of Philosophy

Karen Heathcote
Wolfson College
2023

Dedicated to Andrew Heathcote.

Acknowledgements

My first, and biggest, thanks has to go to Emily Flashman for being the best primary supervisor anyone could wish for. Her support and guidance over the past four years have been invaluable. I am really going to miss working with her.

I am also incredibly grateful for all of the advice and input I've received from Peter Ratcliffe as my supervisor. His keen insight has been useful in many different areas and aspects of this PhD. It is impossible to mention Peter without also thanking Catherine King, who keeps the whole Ratcliffe group functioning. My thanks must also go to Jacob Bush, my industrial supervisor at GSK, for all of his help.

A big thank you to all members of the Flashman group, past and present. It has been a pleasure to share a lab with you all. Thank you also to the Ratcliffe group for welcoming me so readily when I started my cellular work. I'd also like to thank the members of the Schofield and Kawamura groups who I've shared labs and equipment with over the years.

In the lab, I need to thank Grace Roper, for showing me how to make and purify peptides, and Leah Taylor-Kearney, Emily Flashman, and Dona Gunawardana for teaching me all things protein production. Thank you to Dona, also, for her help with protein crystallography, along with Anka Lucic and Mike McDonough. Thank you to Tony Tumber for his unending and invaluable assistance with the RapidFire mass spectrometer and LCMS system, to John Walsby-Tickle and Elisabete Pires for theirs with the Xevo, and to Nader Amin for his with NMR. A massive thank you to Tom Keeley for showing me how to work with human cells. Thank you, also, to Garrett Morris for showing me how to dock peptides.

I'd like to thank Tom Keeley and Norma Masson for all their help and advice regarding ADO. Thank you to Mark White, who started the biochemical ADO work and provided me with a great foundation of resources and protocols to start this project with.

Thank you to my collaborators: Thomas Arnesen and Matti Myllykoski at the University of Bergen, and Colin Levy at the University of Manchester.

I am very grateful to have had Akane Kawamura as a programme director for the GSK-Crick Chemical Biology CDT. Her enthusiasm for our projects never waned and I always knew I could count on her support. A big thank you is also due to Nina Stamps and Katherine Baysan for their help with all the various administrative issues that have come up over the years.

Thank you to my small army of proof-readers: Nyasha, Issy, Anna, Tom, Salma, Jess, Rebecca, Dona, Ollie, and M.

Finally, thank you to my family for all of their support, to my fiancé Pete for deeds too many to mention, and to my dog and WFH buddy Copper, who often helped to keep me writing by sticking his head in my chair legs and trapping me at my desk (see image below).



Declaration

The work described in this thesis is entirely my own, except where I have directly acknowledged help from a named person or have referenced a published source.

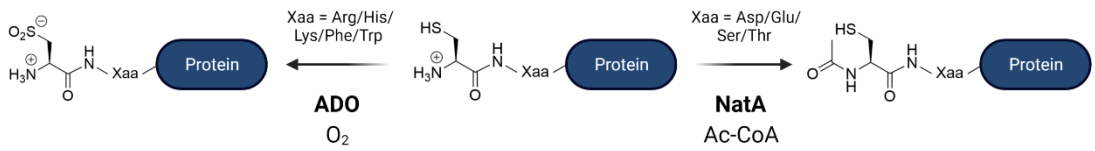
Abstract

Many organisms have developed systems that allow them to sense and respond to changes in O_2 levels, and so maintain O_2 homeostasis. Recently, a new O_2-sensing system was identified in animals, which involves a thiol dioxygenase (TDO) enzyme, cysteamine dioxygenase (2-aminoethanethiol dioxygenase or ADO)[1]. In normoxia, ADO catalyses the oxidation of N-terminal cysteine (Nt-Cys) on target proteins to cysteine sulfinic acid, which triggers protein degradation via the Arg/N-degron pathway[1,2]. ADO activity is inhibited by hypoxia, so its target proteins are stabilised. In this manner, ADO confers O_2-dependent stability on its protein substrates, similar to the action of the related plant cysteine oxidases (PCOs)[1,3,4].

Work conducted in this Thesis aimed to increase understanding of ADO and its role in the hypoxic response. Only three protein substrates of ADO were identified in the initial study (regulators of G-protein signalling RGS4/5 and a chemokine IL32)[1]. In this Thesis, broad *in vitro* screening of ADO activity towards additional Nt-Cys-bearing sequences revealed that it is highly substrate specific, preferentially catalysing oxidation of peptides with basic or aromatic residues after the Nt-Cys. Interestingly, this preference was also observed when ADO was tested with PCO substrates, although catalytic activity towards these was significantly lower than towards ADO's known human substrates.

ADO activity towards peptides proved to be broadly a good indicator of ADO activity as manifest in cells with N-terminal substrate sequence fusion protein reporter assays. However, two sequences appeared to not be ADO substrates in this cell assay, contrary to *in vitro* data with corresponding peptides. Furthermore, inconsistencies with cellular assays examining ADO activity towards full-length putative substrates highlighted that other factors, perhaps including protein folding, are important to consider when using *in vitro* data to predict *in cellula* regulation by ADO. No novel physiological protein substrates were identified in this Thesis. However, the results presented greatly increase understanding of ADO activity and will be useful for prediction of novel substrates in the future.

The unexpected *in cellula* results noted above (i.e., the lack of regulation *in cellula* of some sequences despite positive *in vitro* data) led to investigation of Nt-Cys acetylation and its influence on ADO activity, through a collaboration with Thomas Arnesen and colleagues at the University of Bergen. Nt-Cys acetylation was, for the first time, demonstrated to be catalysed by N-terminal acetyltransferase A (NatA). Notably, NatA activity was also shown to be substrate dependent, with a preference contrasting to that of ADO. NatA preferentially catalyses acetylation of peptides with acidic or polar residues after the Nt-Cys. This suggests that ADO and NatA have evolved to modify distinct pools of Nt-Cys proteins. Nt-Cys acetylation was also shown to abrogate ADO activity (and vice versa for oxidation and NatA activity), raising the possibility that some Nt-Cys sequences could be shielded from ADO-dependent degradation by co-translational acetylation.



ADO and NatA preferentially catalyse modification of distinct pools of Nt-Cys-bearing proteins, determined by the physicochemical properties of the residue following the Nt-Cys.

Furthermore, peptide and small molecules were characterised as ADO inhibitors, with substrate-based Nt-selenocysteine peptides identified as the best candidates for specific and potent chemical probes. Finally, investigation of ADO's structure using an *in silico*[4] model and published crystal structures[5,6] identified features within the active site that may contribute to substrate binding and explain some of the experimental observations made in this Thesis.

Overall, the results presented in this Thesis contribute to understanding of ADO's role in O_2-dependent protein signalling, including the extent and determinants of its protein substrates, its interaction with other N-terminal modifications, and its structure-function relationship. Knowledge of ADO's activity on novel protein substrates, and inhibitors thereof, can help to characterise the hypoxic response and could be important in disease conditions, such as ischaemia and cancer[4].

Abbreviations

Ac	acetyl
ADO	cysteamine dioxygenase (also known as 2-aminoethanethiol dioxygenase)
ANOVA	analysis of variance
AP2/ERF	Apetala2/ethylene responsive factor
AS	'anti'-substrate
At	*Arabidopsis thaliana*
ATE	arginyl transferase
ATP	adenosine triphosphate
Aq	*Amphimedon queenslandica*
Bis-Tris	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl) methane
Bis-Tris propane	1,3-Bis[tris(hydroxymethyl)methylamino]propane
BSA	bovine serum albumin
Bt	*Bos taurus*
C	Celsius
cDNA	complementary DNA
CDO	cysteine dioxygenase
CHCA	α-cyano-4-hydroxycinnamic acid
CoA	coenzyme A
CSA	cysteine sulfinic acid
CV	column volumes
Da	dalton (mass unit)
DCM	dichloromethane
DETA NONOate	2,2′-(hydroxynitrosohydrazono)bis-ethanimine
DFO	deferoxamine
DFOR	dual-fluorescence oxygen reporter
DIC	*N,N'*-diisopropylcarbodiimide
DMEM	Dulbecco's Modified Eagle Medium
DMF	dimethyl formaldehyde
DMOG	dimethyloxallyl glycine
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid

DOPE	Discrete Optimised Protein Energy
Dp44Mt	di-2-pyridylketone 4,4-dimethyl-3-thiosemicarbazone
Dr	*Danio rerio*
DSBH	double stranded β helix
DTT	1,4-dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EGFP	enhanced GFP
ERF-VII	Group VII ethylene response factor
ESI-MS	electrospray ionisation MS
FBS	fetal bovine serum
FIH	factor inhibiting HIF
Fmoc	fluorenlymethyloxycarbonyl
FPLC	fast protein liquid chromatography
gDNA	genomic DNA
GFP	green fluorescent protein
Hcy	homocysteine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIF	hypoxia inducible factor
HPLC	high-performance liquid chromatography
HRE	hypoxia response element
Hs	*Homo sapiens*
IC_{50}	half maximal inhibitory concentration
IL32	interleukin-32
IMAC	immobilised metal affinity chromatography
iMet	initiator methionine
IPTG	isopropyl β-D-1-thiogalactopyranoside
k_{cat}	turnover number
K_M	Michaelis constant
KO	knock-out
LB	Luria Bertani broth
LC	liquid chromatography
Lm	*Leishmania major*

MAP	methionine aminopeptidase
MALDI-TOF	matrix-assisted laser desorption ionisation time of flight
MBHA	4-methylbenzhydrylamine
MC-initiating	Met-Cys-initiating
MeCN	acetonitrile
Mm	*Mus musculus*
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MW	molecular weight
MWCO	molecular weight cut-off
m/z	mass to charge ratio
NAT	N-terminal acetyltransferase
NatA	N-terminal acetyltransferase A
NCO	N-terminal cysteine oxidase
ND	not determined
NMR	nuclear magnetic resonance
Nt	N-terminal
NTA	N-terminal acetylation
OD	optical density
OE	over-expressing
OS	'optimum' substrate
PAGE	polyacrylamide gel electrophoresis
PCO	plant cysteine oxidase
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PHD	prolyl hydroxylase
PRT6	proteolysis 6
P2A	porcine teschovirus-1 2A
Q-ToF	quadrupole time of flight
RAP	related to AP-2
Rn	*Rattus norvegicus*
RNA	ribonucleic acid

RGS	regulator of G-protein signalling
ROS	reactive oxygen species
r.t	room temperature
sd	standard deviation
SDS	sodium dodecyl sulfate
Sec	selenocysteine
SEC	size exclusion chromatography
sem	standard error of means
SOC	super optimal broth with catabolite repression
Sp	*Spizellomyces punctatus*
SPPS	solid phase peptide synthesis
TCEP	tris(2-carboxyethyl)phosphine
TDO	thiol dioxygenase
TFA	trifluoroacetic acid
TIPS	Triisopropylsilane
TLC	thin layer chromatography
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
UPLC	ultra-high-performance liquid chromatography
UV	ultraviolet
v	velocity
V	Volts
V_{max}	maximum velocity
WT	wild type
(w/v)	weight to volume ratio
(v/v)	volume to volume ratio
2OG	2-oxoglutatarate
2 TY	2-tryptone yeast extract

Common one- and three- letter amino acid abbreviations are used throughout

Contents

Acknowledgements ..... **i**
Declaration ..... **iii**
Abstract ..... **iv**
Abbreviations ..... **v**
1 Introduction **1**
1.1 Oxygen sensing in animals 1
1.1.1 Oxygen homeostasis in animals 1
1.1.2 HIF transcription factors and PHD enzymes 2
1.1.3 Transcriptional effects of HIF 4
1.2 Oxygen sensing in plants 5
1.2.1 Oxygen homeostasis in plants 5
1.2.2 ERF-VIIs and PCO enzymes 6
1.2.3 Effects of ERF-VIIs 9
1.3 Protein N-termini 9
1.3.1 Common N-terminal modifications 10
1.3.2 N-degron pathways 13
1.4 Thiol dioxygenases 16
1.4.1 Small molecule thiol dioxygenase enzymes 17
1.4.2 N-terminal cysteine oxidases 19
1.5 Cysteamine dioxygenase 21
1.6 Thesis Aims 28
2 Investigation of *Hs*ADO Substrates *in vitro* **30**
2.1 Introduction 30
2.2 Chapter Aims 31
2.3 Preparation of materials for *in vitro* experiments 31
2.3.1 Expression and purification of recombinant *Hs*ADO 31
2.3.2 Preparation of peptides for *in vitro* assays 32
2.4 Recombinant *Hs*ADO catalyses oxidation of cysteamine 33
2.5 Recombinant *Hs*ADO catalyses oxidation of RGS4 peptide 35
2.6 Investigation of NCO activity with RGS4 peptides 37
2.6.1 Methionine cleavage is necessary for Cys2 oxidation 37
2.6.2 *Hs*ADO does not catalyse oxidation of internal cysteines 38
2.6.3 *Hs*ADO can catalyse oxidation of 14- to 7-mer peptides 39
2.7 *Hs*ADO is selective amongst Nt-Cys peptides *in vitro* 41
2.7.1 *Hs*ADO activity *in vitro* varies with Nt-Cys peptide sequence 41

2.7.2 'Optimum' and 'anti'-substrate peptides demonstrate *Hs*ADO substrate preferences 47
2.8 Characterisation of *Hs*ADO activity with PCO substrates 52
2.8.1 *Hs*ADO can catalyse oxidation of PCO substrate peptides 53
2.8.2 Steady-state kinetic analysis of *At*RAP2.12$_{2-15}$ with *Hs*ADO 54
2.9 Discussion 57
3 Investigation of *Hs*ADO Substrates in Human Cells 59
3.1 Introduction 59
3.2 Chapter Aims 59
3.3 The dual-fluorescence oxygen reporter (DFOR) assay 60
3.4 Preparation of materials for DFOR cellular assay 62
3.4.1 Selection of Nt-sequences from peptide activity assays 62
3.4.2 Plasmid construction 64
3.4.3 Generation of SH-SY5Y cells stably expressing DFOR constructs 65
3.5 *Hs*ADO is selective amongst *Nt*-Cys sequences in DFOR assay 66
3.5.1 Regulation of *Hs*ADO activity in DFOR assay with 2,2-dipyridyl 66
3.5.2 Regulation of *Hs*ADO activity in the DFOR assay with hypoxia 71
3.6 Cellular assays with full-length putative substrate proteins 76
3.7 Discussion 80
4 The Interplay of Nt-Cys Oxidation and Acetylation 82
4.1 Introduction 82
4.2 Chapter Aims 84
4.3 NatA is selective amongst Nt-Cys peptides *in vitro* 84
4.3.1 NatA activity *in vitro* varies with Nt-Cys peptide sequence 85
4.3.2 Acetylation occurs on the Nt-Cys of JUNB$_{2-15}$ 90
4.3.3 NatA activity with MTPN$_{2-15}$ varies with identity of residue 3 92
4.3.4 NatA has distinct substrate preferences to ADO 92
4.4 Nt-Cys acetylation abrogates ADO-catalysed oxidation 95
4.5 Nt-Cys oxidation abrogates NatA-catalysed acetylation 97
4.6 Discussion 98
5 Peptide and Small Molecule Inhibitors of *Hs*ADO 101
5.1 Introduction 101
5.2 Chapter Aims 102
5.3 Peptide inhibitors 102
5.3.1 Nt-Sec peptides inhibit *Hs*ADO *in vitro* 104
5.3.2 Nt-Sec peptide inhibitors have similar potencies 106
5.4 Small molecule inhibitors 109

5.4.1 2,2-dipyridyl and DP44mT inhibit *Hs*ADO *in vitro* 110
5.4.2 DP44mT is more potent than 2,2-dipyridyl *in vitro* 112
5.5 Investigation of modes of *Hs*ADO inhibition 113
5.5.1 KH1 mode of inhibition 115
5.5.2 2,2-dipyridyl mode of inhibition 117
5.6 Discussion 119
6 Structural Investigation of ADO 122
6.1 Introduction 122
6.2 Chapter Aims 125
6.3 Crystallisation trials with *Hs*ADO 125
6.4 *In silico* modelling of *Hs*ADO structure 128
6.5 Co-crystallisation trials with *Hs*ADO and KH1 peptide inhibitor 134
6.6 AutoDock4 studies with *Mm*ADO 136
6.6.1 Docking results with RGS4 peptide 137
6.6.2 Docking results with IL32 peptide 139
6.6.3 Docking results with RAP2.12 peptide 144
6.7 Discussion 145
7 Conclusions 152
8 Materials and Methods 165
8.1 Biochemical *in vitro* Experiments 165
8.1.1 Recombinant protein production 165
8.1.2 Recombinant protein purification 166
8.1.3 SDS-PAGE gel analysis 168
8.1.4 Fe quantification of recombinant protein 168
8.1.5 Peptide synthesis 170
8.1.6 Thin-layer chromatography assays 173
8.1.7 Peptide oxidation assays 174
8.1.8 Peptide acetylation assays 178
8.1.9 Mass spectrometry analysis 181
8.1.10 Protein crystallography 182
8.2 Experiments in Mammalian Cells 184
8.2.1 Preparation of plasmids for DFOR assays 184
8.2.2 General cell culture procedures 187
8.2.3 Production of stable SH-SY5Y cell lines for DFOR assay 187
8.2.4 Dual-fluorescence oxygen reporter assay 189
8.2.5 Analysis of DNA from SH-SY5Y cell lines 190
8.2.6 Assay of full-length protein stability in mammalian cells 191

8.3 Computational Experiments 193
8.3.1 Comparative protein modelling 193
8.3.2 Peptide docking experiments 194
8.3.3 Cavity prediction 196
References 197
Appendix 1: Sequence of recombinant *N*-His$_6$-tagged *Hs*ADO and representative UV traces from purification 210
Appendix 2: Peptides purchased for biochemical experiments 211
Appendix 3: MALDI spectra for peptides synthesised in this Thesis 213
Appendix 4: *Hs*ADO *in vitro* Nt-Cys peptide screening data 220
Appendix 5: Statistical analysis of OS peptides 1 minute assay data 221
Appendix 6: Statistical analysis of AS peptide 1 minute assay data 223
Appendix 7: Primers 225
Appendix 8: pcDNA3-GFP-P2A-mCherry plasmid map and sequence 227
Appendix 9: Combined NatA and *Hs*ADO *in vitro* activity data 228
Appendix 10: Deconvoluted mass spectra of oxidised peptides used in Chapter 4 229
Appendix 11: Raw data from mode of inhibition assays 231
Appendix 12: Crystallisation screening trays 233
Appendix 13: MODELLER scripts and .ali files 235
Appendix 14: 1,10-phenanthroline contamination during ADO purification 240
Appendix 15: ADO + KH1 co-crystallisation trial microscope pictures 242
Appendix 16: CavityPlus analysis of *Mm*ADO 244
Appendix 17: Potential O_2 access tunnel in ADO structures 245

1 Introduction

1.1 Oxygen sensing in animals

Oxygen is essential to all animals, with only one exception (a parasite identified in 2020 that does not require oxygen for cellular respiration)[7]. In this Section, the need for oxygen homeostasis in animals will be discussed and a key transcriptional regulation system in the hypoxic response will be introduced.

1.1.1 Oxygen homeostasis in animals

Animals are defined as multicellular, eukaryotic organisms and are sometimes referred to as metazoans. The first metazoan organisms are believed to have evolved around 0.5 billion years ago, and their evolution was likely dependent on oxidative phosphorylation[8]. Oxidative phosphorylation is the process by which metazoans release energy from fuel to survive. In this process, molecular oxygen (O_2) is utilised as the final electron acceptor in the mitochondrial electron chain, and is therefore essential for metazoans to produce energy. The reduction of O_2 during oxidative phosphorylation can lead to the release of reactive oxygen species (ROS), such as superoxide or peroxide radicals. ROS can damage cellular proteins, lipids, and other important molecules, potentially causing cell dysfunction and death[9]. Excess ROS may be generated when O_2 levels rise or fall sharply[8,10]. Furthermore, low O_2 concentrations lead to reduced oxidative phosphorylation, and therefore reduced energy production, which impacts a range of cellular processes. Therefore, the process of oxidative phosphorylation and energy production in animals appears to be optimised to function within a narrow O_2 range[8].

The level of O_2 in the atmosphere on Earth is usually around 21 %. This O_2 level is generally referred to as normoxia when studying animals and animal cells. O_2 levels can be reduced environmentally or locally within the body, known as hypoxia. To maintain O_2 levels in the

physiological range, i.e., to maintain O_2 homeostasis, animals must be able to sense and respond to changes in O_2 availability. The carotid body has been identified as a chemoreceptor that senses changes in arterial oxygen, inducing a rapid physiological response to hypoxia[11]. Over a longer timescale, at the transcriptional level, the response to hypoxia is controlled by the hypoxia inducible factor (HIF) family of transcription factors[12,13].

1.1.2 HIF transcription factors and PHD enzymes

HIF-1 was first discovered in 1992 as a protein that bound to the hypoxia-inducible enhancer of the human erythropoietin (EPO) gene[12]. EPO is a hormone that regulates red blood cell production, and thus O_2 delivery within the body. It was later found that this O_2-sensing mechanism was widespread across mammalian cells, and HIF transcription factors were found to direct a range of processes involved in the hypoxic response[13–15].

HIF transcription factors are heterodimeric proteins, composed of an unstable HIFα protein and a stable HIFβ protein[13,16]. Metazoans possess three different HIFα proteins: HIF1α, HIF2α, and HIF3α. The stability of HIFα proteins is controlled by O_2-sensitive prolyl hydroxylase (PHD) enzymes, which catalyse hydroxylation of a HIFα proline residue[17,18]. Hydroxylated HIFα proteins can be bound by von Hippel-Lindau (VHL) protein and subsequently targeted for ubiquitination and proteasomal degradation[19]. In hypoxia, PHD activity is inhibited, so non-hydroxylated HIFα proteins accumulate and dimerise with HIF1β to form transcriptionally active HIF. The formation of transcriptionally active HIF is also regulated by factor inhibiting HIF (FIH) enzyme. FIH catalyses hydroxylation of an asparagine residue in HIFα proteins, which prevents binding of coactivators p300/CREB-binding protein[20–22]. As with PHD-catalysed prolyl hydroxylation, FIH-catalysed asparaginyl hydroxylation is inhibited by hypoxia. Therefore, in normoxic conditions HIF transcriptional activity is inhibited, due to the action of PHD and FIH

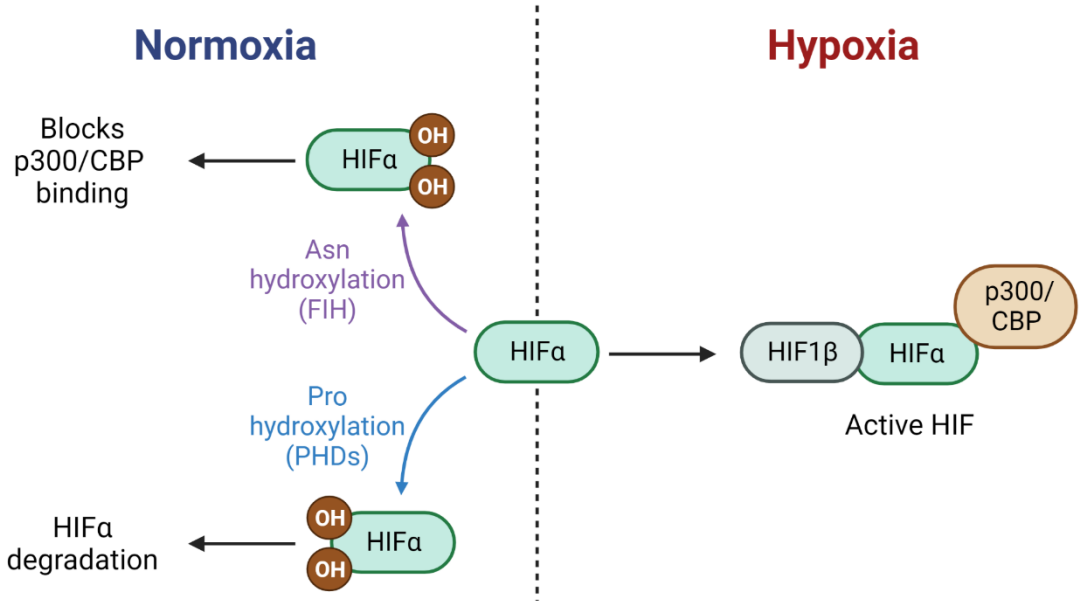


Figure 1.1: Fates of HIFα in normoxia and hypoxia
Post-translational modifications of HIFα, catalysed by prolyl hydroxylase (PHD) enzymes (proline hydroxylation) and by factor inhibiting HIF (FIH) enzyme (asparagine hydroxylation), lead to inactivation of HIF via HIFα degradation and prevention of cofactor recruitment, respectively. FIH and PHD enzymes are inhibited by hypoxia, so transcriptionally active HIF can form.
Created with Biorender.com

enzymes, while in hypoxia transcriptionally active HIF can form and regulate the hypoxic response (Figure 1.1).

PHD and FIH enzymes are iron(II)-dependent, non-heme, 2-oxoglutatarate (2OG) dependent oxygenases. Three HIF prolyl hydroxylases have been identified: PHD1, PHD2, and PHD3[23,24]. As members of the 2OG-oxygenase family, structures of FIH and the PHDs feature a characteristic double stranded β helix (DSBH) that supports an iron-binding site consisting of a facial triad of two His residues and one Asp or Glu residue (Figure 1.2)[25–27]. In addition to iron(II) and 2OG, these enzymes also require an ascorbate cofactor for activity[28,29]. FIH has also been shown to catalyse hydroxylation of asparagine residues in ankyrin repeat domain (ARD)-containing proteins[30].

The oxygen sensitivity of FIH and PHD enzymes has been investigated both *in vivo* and *in vitro*. In cells, PHD activity has been shown to be more sensitive to changes in O_2 levels than FIH activity[31]. Studies have shown that PHD2 reacts slowly with O_2; this is thought to contribute to its function as an O_2 sensor[29,32]. Steady state kinetic analyses have found that

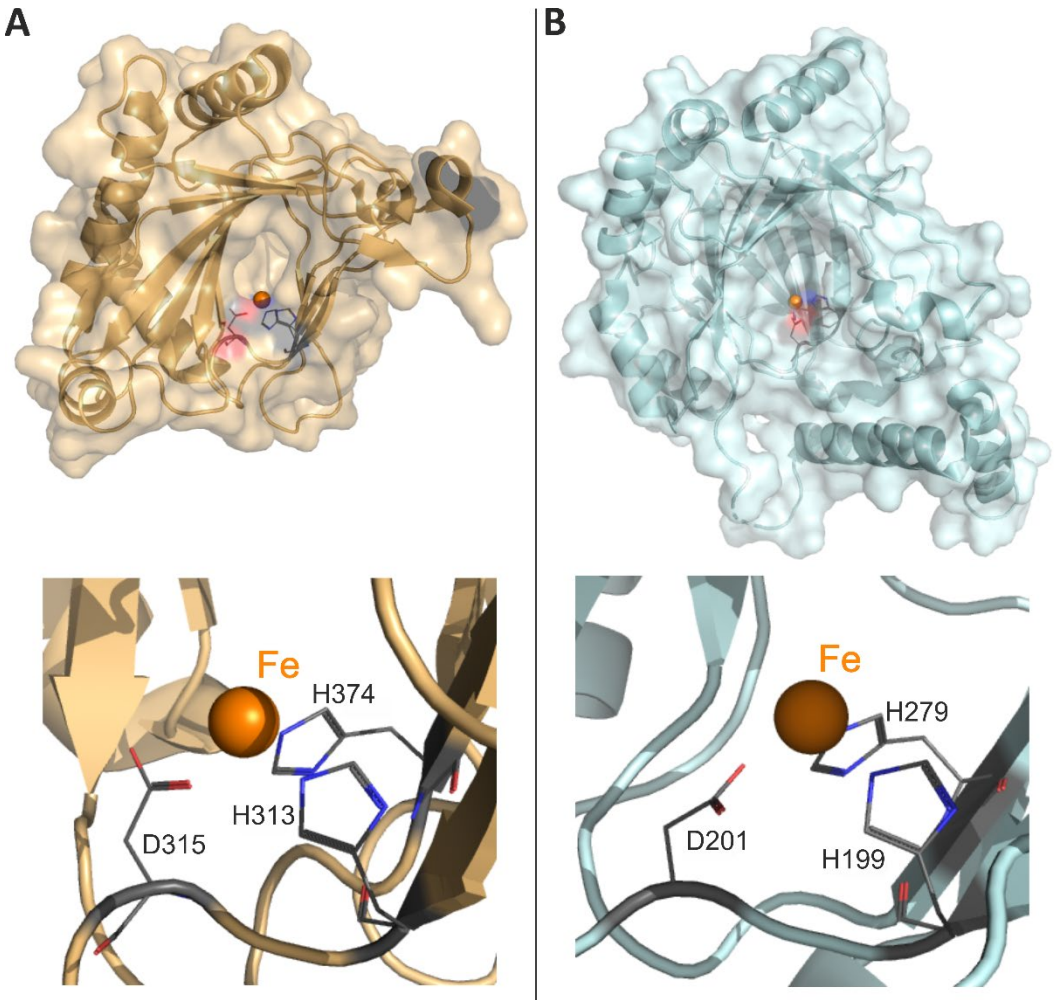


Figure 1.2: Structures of PHD2 and FIH enzymes
A: Crystal structure of catalytic domain of PHD2 (PDB: 2G19)
B: Crystal structure of FIH (PDB: 1H2L)
Active site Fe shown in orange; Fe-binding facial triad residues shown as grey lines.

FIH has a lower $K_M^{app}(O_2)$ than PHD2, and that FIH O_2 sensitivity varies with substrate[33–35]. Two studies have found that the $K_M^{app}(O_2)$ values of FIH are lower with ARD-containing substrates than with HIF substrates[33,35]. Tarhonskaya *et al.* determined $K_M^{app}(O_2)$ values of 110 µM (approx. 8 % O_2) for FIH and of 410 to > 450 µM (> 30 % O_2) for PHD2, both with HIFα substrates[35]. These kinetic properties reflect the ability of PHD and FIH enzymes to act as oxygen sensors[35].

1.1.3 Transcriptional effects of HIF

Active HIF induces a transcriptional response by binding to hypoxia response elements (HRE) on DNA. This activates transcription of over 1500 different target genes[11]. Some genes are targeted specifically by HIF1α, some specifically by HIF2α, and others by both[13].

HIF3α is less well understood. Multiple variants of HIF3α have been identified that appear to have distinct functions, one of which has been shown to be the inhibition of HIF1/2α activity by competing with HIF1β for binding[36–38]. Targets of HIF have been identified as being involved in a wide range of processes, including energy metabolism, hormonal regulation, angiogenic signalling, vasomotor regulation, and cell migration[15]. Regulation of the hypoxic response via the HIF pathway extends beyond direct transcriptional activation: HIF has been shown to interact with other signal transduction and transcriptional systems including the Notch and Wnt pathways[13,39,40]. Overall, since its discovery the HIF pathway has been shown to play a widespread and significant role in the hypoxic response. This was recognised in 2019 when Kaelin, Ratcliffe, and Semenza were jointly awarded the Nobel Prize for Physiology and Medicine[41].

1.2 Oxygen sensing in plants

Oxygen is also essential to plants. In this Section the need for oxygen homeostasis in plants will be discussed. This Section will also introduce an O_2-sensing system involved in the plant hypoxic response that is reminiscent of the HIF pathway described above (Section 1.1).

1.2.1 Oxygen homeostasis in plants

The evolution of photosynthesising organisms on Earth has been linked to the increased availability of O_2 in the atmosphere[42]. Endosymbiosis of these first photosynthesising cyanobacteria led to the evolution of plants, which, like animals, are aerobic organisms. Plants require molecular O_2 for various biochemical pathways, including oxidative phosphorylation (as discussed above in Section 1.1.1) and the production of hormones such as ethylene[43] and abscisic acid[44]. Reactive oxygen species (ROS) that form in plants can cause oxidative damage to DNA and lipids, but can also be involved in development and signalling[45–47]. Unlike animals, plants do not have a system for active delivery of O_2 to

tissues so are reliant on passive transport processes like diffusion, and are also able to produce O_2 by photosynthesis[48]. O_2 levels within plants vary between and within tissues, where they can drop close to zero[48]. Environmental factors, such as submergence, can also lead to changes in plant O_2 levels.

When the O_2 level is reduced in plant tissues, there can be metabolic (e.g., switching from aerobic to anaerobic respiration) and morphological (e.g., rapid stem elongation to escape submergence) effects[48]. In order to control these responses, plants require the ability to sense and respond to changes in O_2 levels. As in animals, O_2 homeostasis in plants involves transcriptional regulation; some genes are upregulated in response to low O_2, while other transcripts are ROS-responsive[48]. In vascular plants, the hypoxic response is regulated by group VII of the ethylene response factor (ERF-VIIs) family of transcription factors[49–51].

1.2.2 ERF-VIIs and PCO enzymes

ERF-VII transcription factors are part of the ERF family, which in turn belongs to the APETALA2/Ethylene Responsive Factor (AP2/ERF) superfamily[51–54]. ERF proteins are characterised by the presence of a single AP2/ERF domain that is involved in DNA binding[52] and was first identified in tobacco plants in 1995[55]. In addition to the AP2/ERF domain, ERF-VIIs also contain an N-terminal (Nt) domain (MCGGAII/L), which is conserved among flowering plant species[52,53]. The ERF-VIIs have been most thoroughly studied in *Arabidopsis thaliana*, where two subfamilies have been identified, related to AP-2 (RAP-type) and hypoxia responsive ERF (HRE-type)[51]. The Nt-domain shared by ERF-VIIs allows their stability to be regulated by plant cysteine oxidase (PCO) enzymes in an O_2-dependent manner via the N-degron pathway[3,56–58].

PCO enzymes catalyse oxidation of the N-terminal cysteine (Nt-Cys) that is revealed when the initiator Met is removed co-translationally by the action of methionine amino

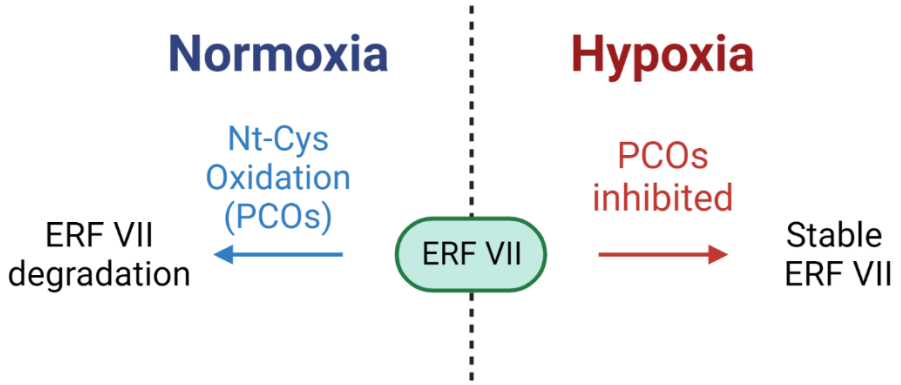


Figure 1.3: Fates of ERF-VIIs in normoxia and hypoxia
Nt-Cys oxidation of ERF-VIIs, catalysed by plant cysteine oxidase (PCO) enzymes, leads to degradation of ERF-VIIs via the Arg/N-degron pathway. PCO enzymes are inhibited by hypoxia, so ERF-VIIs are stabilised.
Created with Biorender.com

peptidases[3,58–60]. The PCOs were first identified in *Arabidopsis thaliana* in 2014 by Weits *et al.*; in this study five PCOs were identified (*At*PCO1-5)[3]. The Nt-Cys is oxidised to cysteine sulfinic acid (CSA), allowing the N-termini of *At*ERF-VIIs to be recognised by arginyl tRNA transferase (ATE) enzymes, which catalyse Nt-arginylation[58]. Nt-arginylated *At*ERF-VIIs can then be ubiquitinated by an E3 ligase proteolysis 6 (PRT6)[61], and subsequently undergo proteasomal degradation. This pathway is part of the Arg/N-degron pathway[2]. N-degron (previously referred to as N-end rule) pathways will be discussed in more detail below (Section 1.3.2). PCO enzymes require O_2 as a co-substrate, so are inhibited by hypoxia[3,58,62]. Thus, in hypoxia ERF-VIIs are stabilised and can regulate the transcriptional response in plants (Figure 1.3). Degradation of ERF-VIIs *in vivo* has also been demonstrated to require nitric oxide (NO)[63]. However, *in vitro* studies have found that NO is not required for PCO activity and the precise role of NO in the Arg/N-degron pathway remains to be determined[3,58,62,64]. Interestingly, PCOs capable of catalysing oxidation of the Nt-Cys of *At*ERF-VIIs *in vitro* have been identified in early land plants and algae, suggesting that this pathway could have a role in O_2 homeostasis throughout the plant kingdom[65].

PCOs are iron(II)-dependent thiol dioxygenase enzymes. Thiol dioxygenase enzymes will be discussed in detail in Section 1.4. PCOs are members of the cupin superfamily and their structures feature a characteristic double stranded β helix (DSBH) that supports an iron-binding site consisting of a facial triad of three His residues (Figure 1.4)[66,67]. In *Arabidopsis*, five ERF-VII substrates have been identified: RAP2.12, RAP2.2, RAP2.3, HRE1, and HRE2. *At*PCOs have also been shown to catalyse Nt-Cys oxidation of non-ERF-VII substrates, polycomb group protein vernalisation 2 (VRN2)[68] and LITTLE ZIPPER 2 (ZPR2)[69].

Catalytic activity has been observed to vary between *At*PCOs 1 – 5; *At*PCO4 and *At*PCO5 were more catalytically efficient *in vitro* with *At*RAP2.12/2.2 than *At*PCO1, *At*PCO2, and *At*PCO3[62]. This study also characterised the O_2-sensitivity of the *At*PCOs *in vitro,* and found that this too varied between enzymes. $K_M^{app}(O_2)$ values for *At*PCO2 and *AtP*CO5 (7 and 5 %, respectively) were lower than those for *At*PCO1 and *AtP*CO4 (16 and 17 %, respectively), suggesting that *At*PCO2 and *AtP*CO5 have a higher affinity for O_2[62]. These values were comparable to the O_2 concentrations measured within different plant tissues[48], so demonstrated that PCOs could be physiologically relevant O_2 sensors[4,62].

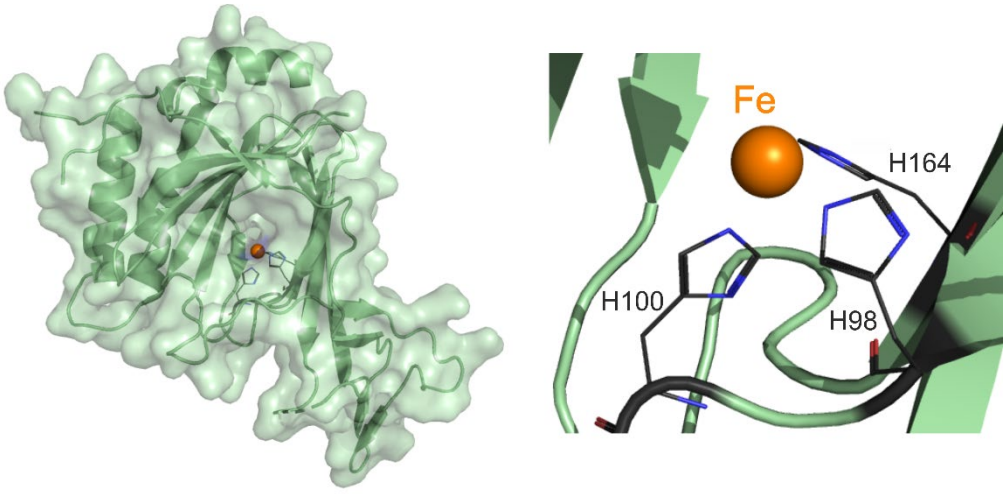


Figure 1.4: Structure of *At*PCO4
A: Crystal structure of *At*PCO4 (PDB: 6S7E). Active site Fe shown in orange; Fe-binding facial triad residues shown as grey lines with N atoms in blue and O atoms in red.

1.2.3 Effects of ERF-VIIs

The ERF-VIIs have been shown to have an important role in controlling the response of plants to flooding and low O_2[53]. The five *At*ERF-VIIs (Section 1.2.2) bind to hypoxia response promoter elements (HRPEs) present in anaerobic genes[51,70]. HRE1, HRE2, and RAP2.2 have been linked to upregulation of hypoxic-response gene alcohol dehydrogenase 1 (ADH1)[71–73]. *At*PCO1 and *At*PCO2 are upregulated in hypoxia, likely regulated by *At*RAP2.12[3,57].

There are two mechanisms by which plants respond to flooding: quiescence (e.g., reduced respiration, inhibited flowering) and escape (e.g., rapid stem growth, vascular changes)[74]. In rice, different ERF-VIIs promote different survival mechanisms: SUB1A promotes quiescence, while SNORKEL1 and SNORKEL2 regulate escape[53,75,76]. Notably, SUB1A is not regulated by the N-degron pathway, despite possessing an Nt-Cys; this has been attributed to an interaction between its N- and C-termini that shields the Nt-Cys from oxidation[77]. The increased stability of SUB1A could contribute to the improved flood tolerance of rice varieties carrying this gene[53].

Overall, ERF-VIIs have been shown to activate transcription of a variety of genes involved in the plant hypoxic response[51,53]. ERF-VIIs have also been linked to the plant response to ROS and nitric oxide (NO). While there is still much to investigate about the ERF-VIIs it is clear that they have an important role in O_2 homeostasis in plants.

1.3 Protein N-termini

The PCO enzymes introduced in the previous section catalyse oxidation of the N-terminal cysteine of their target proteins. The O_2-sensing pathway involving PCOs and their substrates is an example of how the nature of a protein's N-terminus can have implications for its physiological role. This Section will describe some common Nt-modifications that

occur co- and post-translationally on proteins. N-degron pathways, which were introduced briefly above, will also be discussed.

1.3.1 Common N-terminal modifications

The majority of the proteins expressed in eukaryotes will be modified at their N-terminus[78]. These modifications can occur as the nascent protein emerges from the ribosome (co-translationally), or later in the life of the protein (post-translationally). A variety of different enzymes catalyse these modifications.

One of the most common modifications is initiator methionine excision, which is catalysed by methionine aminopeptidases (MAPs)[79]. Met cleavage occurs co-translationally on nascent polypeptides once they reach 20 – 40 amino acids in length[78,80–83]. MAPs will only catalyse Met cleavage when the initiator Met is followed by an Ala, Cys, Gly, Pro, Ser, Thr, or Val residue, and the amino acid at the third position can also influence MAP selectivity[78,79]. Initiator Met cleavage is a process that is highly conserved in all organisms, and over 50 % of proteins are estimated to undergo this modification[59,84]. The physiological function of Met cleavage is not fully understood; however, it has been shown to be involved in controlling protein stability[59].

Another Nt-modification that is common in eukaryotes is Nt-acetylation (NTA)[85]. NTA involves the transfer of an acetyl group from acetyl coenzyme A (CoA) to the Nt-amine of a protein, and is catalysed by N-terminal acetyltransferases (NATs) (Figure 1.5). The majority (80 – 90 %) of soluble proteins in plants and animals are believed to be N-terminally acetylated[85–89]. There is a variety of different NATs with different substrate specificities, mostly determined by the identity of the first two amino acids at the N-terminus of a protein[85,90]. Eight are currently known in eukaryotes: these are NatA-H.

NatA is responsible for acetylation of approximately 40 % of the human proteome and catalyses NTA of small amino acids after the initiator methionine has been removed[87,90]. NatB catalyses acetylation of Met-Asn/Asp/Gln/Glu-initiating N-termini[85,91]. This accounts for approximately 20 % of the Nt-acetylome in humans. NatC, NatE, and NatF also catalyse acetylation of N-termini that have retained their Nt-Met, and are believed to account for around 20 % of NTA in humans[85]. Their substrate specificities are less well defined, although NatF is specific for membrane proteins[88,92]. NatD catalyses acetylation of Ser-initiating N-termini, and has been found to be specific for histones H2A and H4[85,93,94]. NatH has also been found to have a specific set of substrates, catalysing acetylation of mature acidic actin N-termini possessing several Asp or Glu residues[95–97]. Finally, NatG is

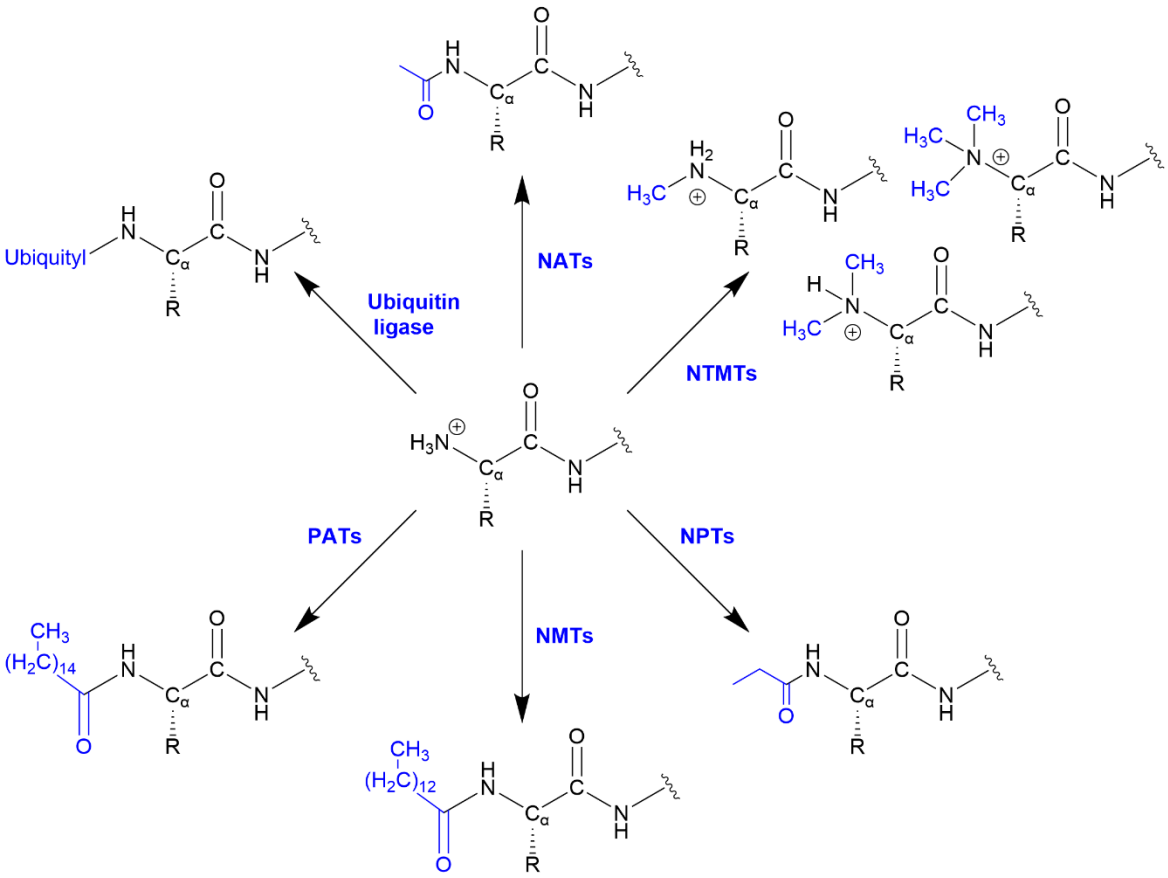


Figure 1.5: Protein Nt-modifications and the enzymes responsible
Examples of some modifications that can occur on the Nt-amine of proteins and the enzymes that catalyse the reactions.
NATs = N-terminal acetyltransferases; NTMTs = N-terminal methyltransferases; NPTs = N-terminal propionyltransferases; NMTs = N-terminal myristoyltransferases; PATs = palmitoylacyltransferases.
Adapted from Varland *et al.* (2015)[78].

responsible for NTA of Met/Ala/Ser/Thr-initiating proteins in chloroplasts[98]. These NAT substrate specificities are not necessarily absolute; specificity can be affected by factors such as downstream residues and secondary structure[78,99]. Within the NatA substrate pool, Ser-initiating N-termini are almost always acetylated, while acetylation of Gly- and Pro-initiating N-termini is much rarer[85,87,100,101].

NTA mostly occurs co-translationally on nascent polypeptides when they reach approximately 25-50 amino acids long[78,101,102]. Some examples of post-translational NTA have been identified, however, including by NatG and NatH[97,98]. Thus far, no N-terminal deacetylases have been found, so NTA is generally assumed to be irreversible[85]. NTA can have a diverse range of effects on proteins. This Nt-modification has been shown to mediate protein-protein interactions and protein complex formation[85,103–105]. NTA can also have a role in subcellular protein localisation and in protein aggregation and folding[85]. In 2010, the Ac/N-degron pathway was discovered, linking NTA to protein degradation[106]. Recently, NTA has also been linked to protein stabilisation in both plants and animals[107–109].

Additional Nt-modifications have been identified on proteins, including propionylation[110], myristoylation[111], and palmitoylation[112,113] (Figure 1.5). Some NATs have been shown to also catalyse Nt-propionylation using propionyl CoA as a propionyl donor[114]. Nt-propionylation is thought to occur both co- and post-translationally, but is much less abundant than NTA[78]. The functional significance of Nt-propionylation has not yet been established[78]. Nt-myristoylation is the transfer of a myristoyl group ($C_{14}OH_{25}$) from myristoyl-CoA to an Nt-glycine, catalysed by N-terminal myristoyltransferases (NMTs). It was initially identified as a co-translational modification, but has since been shown to also occur post-translationally[78]. Met cleavage is required prior to Nt-myristoylation to allow NMTs to recognise the Nt-Gly[111,115]. About 0.5 % of cellular proteins are predicted to be Nt-myristoylated[78,116,117]. Significantly less is known about Nt-palmitoylation than

Nt-myristoylation[78]. Palmitoylation mostly occurs on internal cysteines[118]. However, some instances of palmitoylation have been observed on N-termini[113,119], including on Sonic Hedgehog (SHH) signalling protein[112]. Nt-myristoylation and Nt-palmitoylation are both examples of protein fatty acylation, which is generally associated with targeting proteins to lipid membranes and promoting membrane binding (due to increased hydrophobicity)[78]. Both modifications have been linked to cell signalling and signal transduction[78,119,120].

Other modifications that can occur on the Nt-amine include methylation[121], catalysed by N-terminal methyltransferases (NTMTs), and ubiquitination[122], catalysed by E3 ligases (Figure 1.5). There are also modifications that can occur on other functional groups of the Nt- residues. For example, Nt-Cys oxidation catalysed by plant cysteine oxidases occurs on the thiol group of the Cys side chain[58].

Overall, there is a broad range of different modifications that can occur on a protein's N-terminus, which can occur co- or post-translationally, and may require the action of other N-terminal modifying enzymes (e.g., MAPs) to take place.

1.3.2 N-degron pathways

As mentioned previously, the nature of a protein's N-terminus can have implications for its stability. PCO-catalysed Nt-Cys oxidation triggers degradation of proteins via the Arg/N-degron pathway[3] and NAT-catalysed acetylation can trigger degradation of proteins in the Ac/N-degron pathway[106]. N-degron pathways, previously referred to as N-end rule pathways, define the *in vivo* half-life of proteins by their N-termini[2]. In an N-degron pathway, a destabilising Nt-residue is recognised by a protein (e.g., E3 ubiquitin ligase), which is referred to as an N-recognin. In eukaryotes, the destabilising Nt-residue is accompanied by an internal lysine in the substrate protein that can be polyubiquitinated, thus enabling proteolytic degradation by the 26S proteasome[2,123]. In eukaryotes the

N-degron pathways are referred to as the Ac/Arg/Pro/N-degron pathways. The Pro/N-degron pathway involves the recognition of a Pro residue at the N-terminus[124]. There is also the fMet/N-degron pathway, in which formylation of an Nt-Met residue can trigger protein degradation[125]. The Arg/N-degron pathway is of the greatest relevance to this Thesis, and will therefore be described in the most detail.

In the eukaryotic Arg/N-degron pathway, there are primary, secondary, and tertiary destabilising Nt-residues (Figure 1.6). Tertiary destabilising residues are Gln, Asn, and Cys, which are converted into secondary destabilising residues by the action of different enzymes[2]. Gln (Q) and Asn (N) can be deamidated by specific Nt-amidohydrolases, Nt^{Q}-amidase (NTAQ1)[126] and Nt^{N}-amidase (NTAN1)[127], to generate Nt-Glu and Nt-Asp. Cys can be oxidised to cysteine sulfinic acid, a reaction that can be catalysed by PCO enzymes[3,58], and also by cysteamine dioxygenase (ADO)[1]. Oxidation of Nt-Cys *in vivo* has also been shown to affected by NO[128]. Nt-Cys sulfinic acid, Nt-Glu, and Nt-Asp can be

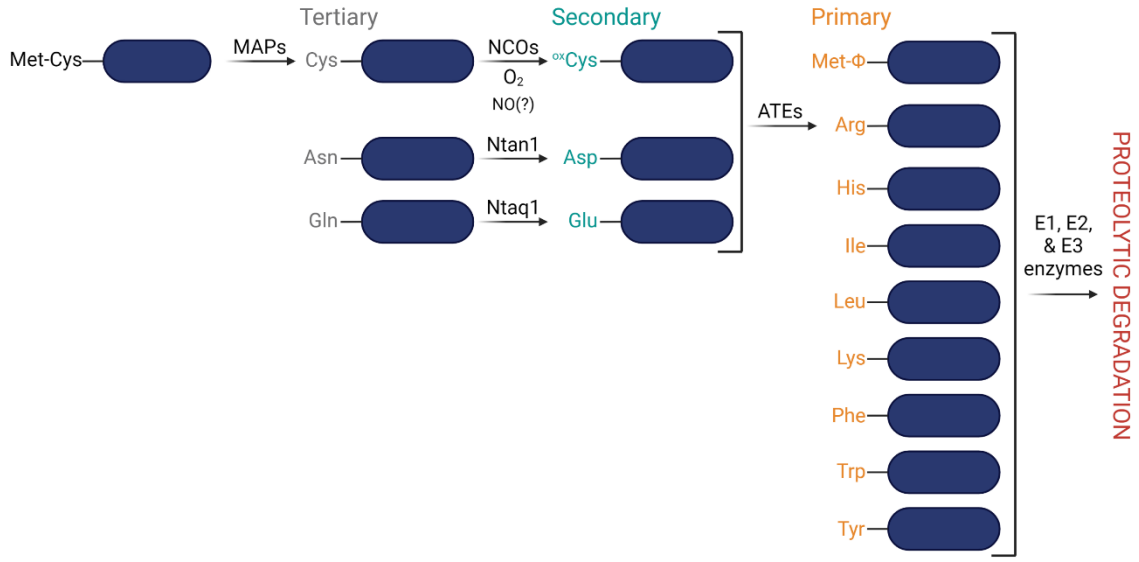


Figure 1.6: Eukaryotic Arg/N-degron pathway
Methionine aminopeptidases can catalyse iMet cleavage to reveal Nt-Cys. Nt-Cys/Asn/Gln are tertiary destabilising residues. Nt-Cys oxidases (NCOs), such as PCOs and ADO, can catalyse Nt-Cys oxidation to cysteine sulfinic acid (^{ox}Cys); NTAN1 and NTAQ1 catalyse deamidation of Nt-Asn and Nt-Gln, respectively. Nt-^{ox}Cys/Asp/Glu are secondary destabilising residues, which can undergo arginyl-tRNA transferase (ATE)-catalysed arginylation to produce Nt-Arg. Primary destabilising residues can be recognised by Arg/N-recognins (E3 ligases), ubiquitinated by the action of E1, E2, and E3 enzymes, and subsequently degraded by the ubiquitin-proteasome system.
Created using Biorender.com

recognised by arginyl-tRNA transferases (ATEs), which catalyse Nt-arginylation[56,128–130]. There is also evidence that ATEs are able to recognise Nt-Cys sulfonic acid, which is presumed to be formed non-enzymatically[58,128,131]. Nt-Arg is a primary destabilising residue that can be recognised by E3 ligase N-recognins[2]. In mammals, these E3 ligases include UBR1, 2, 4, and 5[132], and in plants, PRT6[61]. Within the Arg/N-degron pathway Nt-Lys/His/Leu/Phe/Tyr/Trp/Ile can also be recognised by the Arg/N-recognins, as can Nt-Met if followed by a bulky hydrophobic residue[2]. E3 ligases catalyse ubiquitination of an internal lysine in the substrate protein[133]; this process also requires the action of E2 ubiquitin-conjugating enzymes and E1 ubiquitin-activating enzymes[134]. The 26S proteasome can bind ubiquitinated proteins, which are then unfolded and broken down into short polypeptides[135]. Ubiquitination is a reversible process; deubiquitylases (DUBs) can catalyse deubiquitination and edit poly-ubiquitin chains[2,136]. DUBs play an important role in maintaining protein homeostasis and releasing ubiquitin[136].

The process of Nt-arginylation has been the subject of a significant volume of research over the years. In 2018, a study was published examining the substrate specificity of mammalian ATE enzymes, which found that different isoforms of mouse ATE1 targeted different substrate pools and that arginylation could depend on the physicochemical environment around the Nt-residue[137]. The authors of this paper were able to generate an algorithm that enabled prediction of arginylation of Nt-Asp/Glu-bearing proteins. More recently, human ATE1 has been identified as part of a targeting complex, consisting of five enzymes of the human Arg/N-degron pathway by Oh *et al*[138]. This complex was formed by interactions between NTAQ1, NTAN1, ATE1, E3 ligase UBR1, and E2 enzyme UBE2. Oh *et al.* proposed that this complex was involved in substrate channelling in the Arg/N-degron pathway. In 2022, two crystal structures of yeast ATE1 enzymes (*Kluyveromyces lactis* ATE1 and *Saccharomyces cerevisiae* ATE1) were published, which provided much needed

insight on ATE1's mechanism and substrate binding sites[139,140]. Kim *et al.* identified potential binding sites in *Kl*ATE1, including two pockets, one positively and one negatively charged, which the authors proposed bind to the negatively charged N-terminal Asp/Glu or oxidised Cys substrate protein residues and to the positively charged Arg moiety of the Arg-tRNA^{Arg} co-substrate, respectively[139]. Van *et al.* used their structural findings with *Sc*ATE1 to propose a mechanism for arginylation, which was supported by Kim *et al.*'s findings with *Kl*ATE1[139,140].

The Ac/N-degron pathway was discovered in 2010[106] and is less well understood than the Arg/N-degron pathway. This pathway was first identified in yeast, where acetylated Nt-Met, Ala, Val, Ser, Thr, and Cys residues were found to be recognised by Ac/N-degrons[106,141]. There is also evidence that the Ac/N-degron pathway exists in mammals[142] and could exist in plants[64,143]. In 2017, the Pro/N-degron pathway was uncovered; in yeast, Gid4 was identified as a Pro/N-recognin that could bind to Pro at the N-terminus or position two in the presence of distinct sequence motifs[124]. The most recently identified eukaryotic N-degron pathway involves the recognition of Nt-formyl methionine, and was also found in yeast[125].

Overall, N-degron pathways have been found to be essential for protein homeostasis across eukaryotes. These pathways involve a variety of different enzymes, and to date all 20 amino acids have been implicated as N-degrons or pro-N-degrons[2].

1.4 Thiol dioxygenases

Thiol dioxygenase enzymes were briefly introduced above in the context of the PCO (plant cysteine oxidase) enzymes. This section will discuss thiol dioxygenase enzymes, their structure, and function in greater detail.

Thiol dioxygenase (TDO) enzymes are part of the cupin superfamily of proteins, meaning that they share two cupin sequence motifs and a distinct double stranded β helix (DSBH) within their protein structure[144,145]. These enzymes are iron-dependent and have been identified in eukaryotes and prokaryotes with a range of different functions[4].

1.4.1 Small molecule thiol dioxygenase enzymes

The first TDO to be characterised was cysteine dioxygenase (CDO), which catalyses the oxidation of L-cysteine to L-cysteine sulfinic acid (Figure 1.7)[146,147]. This reaction is part of cysteine homeostasis and the taurine biosynthesis pathway[148,149]. Cysteine can be toxic at high levels and has been linked to diseases such as Alzheimer's and Parkinson's[150]. Therefore, cysteine homeostasis is an essential process. Taurine synthesis is also essential; taurine deficiency is associated with heart dysfunction, brain development abnormalities, and other pathologies[149,151,152]. Expression of mammalian CDO has been found to be tissue specific and has been found mostly in adipose tissues, the liver, and the lungs[153,154].

CDO's sequence was found to contain the two conserved cupin motifs, and crystal structures obtained for mouse, rat, and human CDO revealed that its structure contained the characteristic DSBH[155–157] (Figure 1.8). Interestingly, the iron coordination site in CDO was observed to differ from that of other cupin proteins[144]. Cupin proteins typically coordinate

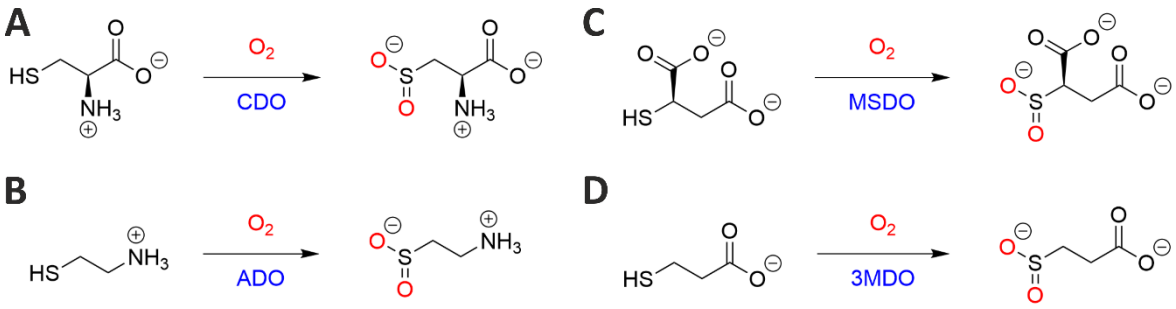


Figure 1.7: Reactions catalysed by small molecule TDO enzymes
A: Oxidation of cysteine to cysteine sulfinic acid, catalysed by cysteine dioxygenase (CDO)
B: Oxidation of cysteamine to hypotaurine, catalysed by cysteamine dioxygenase, also known as 2-aminoethanethiol dioxygenase, (ADO)
C: Oxidation of mercaptosuccinate to sulfinosuccinate, catalysed by mercaptosuccinate dioxygenase (MSDO)
D: Oxidation of 3-mercaptopropionic acid to 3-sulfiniopropionic acid, catalysed by 3-mercaptopropionic acid dioxygenase (3MDO)

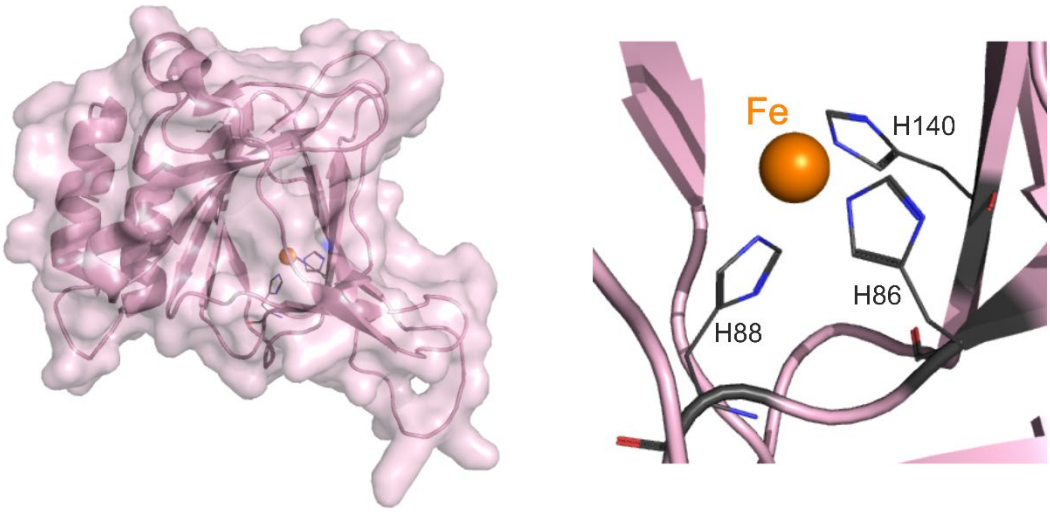


Figure 1.8: Structure of *Hs*CDO
Crystal structure of *Hs*CDO (PDB: 2IC1). Active site Fe shown in orange; Fe-binding facial triad residues shown as grey lines.

their active site metal with a 3-His-1-Glu or 2-His-1-Glu/Asp motif (as observed in 2OG-oxygenases)[25]. CDO's structure, however, features a 3-His facial triad, which coordinates the active site iron (Figure 1.8)[155–157]. This iron binding motif was also observed in the PCOs[66,67] (discussed in Section 1.2 and 1.4.2). Other interesting structural features of CDO include a conserved Ser-His-Tyr motif (Ser153, His155, and Tyr157 in *Hs*CDO) and Arg residue (Arg60 in *Hs*CDO), both found in the active site. The active site Arg is believed to form a salt bridge with the carboxylate group of the cysteine substrate[158] and a hydrogen bonding network forms across the Ser-His-Tyr motif. Tyr157 in the Ser-His-Tyr active site motif has been shown to form a thioether bond with Cys93 (*Hs*CDO numbering)[155–157]. Formation of this Cys-Tyr crosslink has been linked to increased CDO activity and has been shown to require both substrate and substrate turnover[157,159–163]. The CDO Cys-Tyr cofactor is unusual compared to amino acid cofactors in other enzymes, and is an interesting mechanism of regulating protein function[144].

Only one other mammalian thiol dioxygenase enzyme has been identified to date: cysteamine dioxygenase (2-aminoethanethiol dioxygenase, ADO). This enzyme is the primary subject of this Thesis and will therefore be discussed in detail below (Section 1.5).

Briefly, the existence of ADO was first demonstrated in the 1960s[164,165], and the enzyme was first characterised in 2007 by Dominy *et al.*[166]. In this study, it was shown that ADO catalyses the oxidation of cysteamine to hypotaurine (Figure 1.7), which is also a reaction involved in taurine biosynthesis. Expression of ADO has been shown to occur in a wider range of tissues than CDO, with higher expression identified in the brain and muscle[166,167].

Small molecule TDO enzymes have also been found in bacteria; these include bacterial CDOs[168], 3-mercaptopropionic acid dioxygenase (3MDO)[169], and mercaptosuccinate dioxygenase (MSDO)[170]. 3MDO catalyses the oxidation of 3-mercaptopropionic acid and MSDO the oxidation of mercaptosuccinate (Figure 1.7). Bacterial TDOs can be categorised as either "Arg-type" or "Gln-type", which refers to whether the active site Arg60 (*Hs*CDO) is conserved or substituted for a Gln[161]. In CDO Arg60 is associated with substrate binding via interaction with the carboxylate group. MSDO and 3MDO have both been characterised as "Gln-Type" TDOs; however, Arg residues have been identified in both enzymes (in addition to the Gln residue), which likely have a role in substrate binding (Arg66 in *Vp*MSDO and Arg168 in *Av/Pa*3MDO)[4,171,172]. The Cys-Tyr crosslink observed in mammalian CDO does not appear to be conserved in bacterial TDOs[4].

1.4.2 N-terminal cysteine oxidases

In 2014, TDO enzymes were identified in plants that catalysed oxidation of protein substrates rather than small molecules: the plant cysteine oxidases (PCOs)[3] (introduced in Section 1.2.2). These enzymes catalyse the addition of two oxygen atoms to the Nt-Cys of

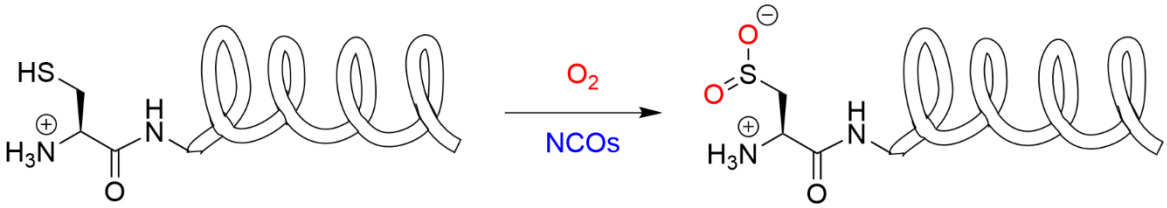


Figure 1.9: Reaction catalysed by N-terminal cysteine oxidase enzymes
Oxidation of a protein's N-terminal cysteine to cysteine sulfinic acid, catalysed by N-terminal cysteine oxidases (NCOs)

a target protein to produce Nt-cysteine sulfinic acid, and have been termed N-terminal cysteine oxidases (NCOs)[4] (Figure 1.9). As mentioned previously (Section 1.2.2), this reaction is a precursor to ATE-catalysed arginylation[58] and subsequent degradation of the target protein via the Arg/N-degron pathway (Section 1.3.2). In 2019, Masson *et al.* demonstrated that cysteamine dioxygenase (ADO) is also an NCO, and is involved in the Arg/N-degron pathway in humans[1] (Section 1.5). This made ADO the first, and thus far only, TDO found to be able to catalyse oxidation of both small molecule and protein substrates.

Structures of NCO enzymes were first reported for the PCOs. As in CDO, the PCO enzymes have been characterised as cupin proteins, and the DSBH core motif has been observed in structures of *At*PCO2, 4, and 5[66,67] (Figure 1.4). The iron binding site in PCOs also resembles that of CDO, with the active site metal coordinated by a 3-His facial triad. However, the Ser-His-Tyr motif found in the active site of CDO is not conserved in the PCOs, where it is replaced by an Ile-Asp-Leu motif. The role of this motif is not yet fully understood, but mutation of the negatively charged Asp residue to an Asn in *At*PCO4 reduced both enzyme activity and iron occupancy[66]. Furthermore, although potential crosslink residues have been identified in the PCOs, a Cys-Tyr crosslink (as seen in CDO), has not been observed[4,66]. The PCOs also lack an Arg residue in their active site that would correspond to Arg60 in *Hs*CDO. In *Hs*CDO, this Arg interacts with the carboxylate group of its substrate; PCO substrates lack a carboxylate group, so the absence of this Arg is not surprising[66]. Another key difference that has been noted between the structures of PCOs and small molecule TDOs is the positioning of a C-terminal loop, which creates an open tunnel-like cavity in the PCOs but sits across the active site in small molecule TDOs and hinders access to the active site[4]. Finally, a potential substrate binding loop was identified

near the entrance to the active site in PCOs that has not been observed in structures of small molecule TDOs[4,66].

Over the years, two separate classes of TDOs have therefore emerged; those capable of accepting small molecule substrates and those that accept protein substrates[4]. The plant NCOs, the PCOs, have been characterised as O_2-sensing and help to regulate the hypoxic response in plants. A role in O_2 homeostasis has not yet been identified for small molecule TDOs such as CDO, 3MDO, and MSDO, although their O_2 kinetics are yet to be determined as they have been for the PCOs[4,62]. Cysteamine dioxygenase (ADO), however, appears to belong to both classes of TDO, catalysing oxidation of both cysteamine and Nt-Cys residues on target proteins[1,4,166].

1.5 Cysteamine dioxygenase

Far less is known about cysteamine dioxygenase (2-aminoethanethiol dioxygenase, ADO) than CDO, its fellow mammalian TDO. It was not biochemically characterised until 2007[166], and its role as an NCO was only published in 2019[1]; it is this function of ADO that is the primary subject of investigation in this Thesis. In this Section, ADO's role as a small molecule TDO and NCO enzyme will be reviewed.

Since the 1960s, researchers have postulated the existence of an enzyme capable of oxidising cysteamine to hypotaurine[164,165,173]. However, attempts to purify this enzyme from different tissues resulted in conflicting reports regarding its molecular weight, activity, and cofactors[144,165,166,174,175]. In 2006, Coloso *et al*. demonstrated that many tissues in rats and mice were capable of converting cysteamine to hypotaurine[176]. Then, in 2007, Dominy *et al.* identified a hypothetical cupin protein that shared low, but significant, sequence identity with *Mm*CDO[166]. This protein (accession Q6PDY2) was identified as *Mm*ADO, and characterised as a TDO with specificity for cysteamine[166].

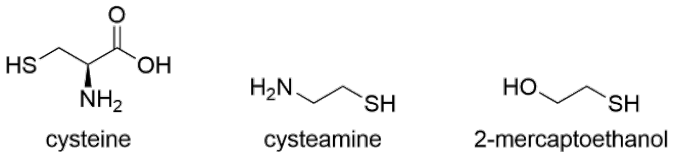


Figure 1.10: Small molecule thiols
Structures of cysteine, cysteamine, and 2-mercaptoethanol

Dominy *et al.* were able to recombinantly express *Mm*ADO and demonstrate that it could catalyse oxidation of cysteamine to hypotaurine[166] (Figure 1.7). Kinetic parameters for ADO with cysteamine were determined (K_M 3.8 mM; k_{cat} 1.6 s^{-1}), and found to be similar to those for CDO with cysteine[166,168,177]. ADO was shown to not catalyse oxidation of cysteine to cysteine sulfinic acid, and cysteine was found to be a weak inhibitor of its cysteamine oxidising activity[166]. 2-mercaptoethanol, whose structure is similar to that of cysteamine (Figure 1.10), was also found to be a weak inhibitor of ADO. Furthermore, Dominy *et al.* demonstrated that iron was required for catalytic activity of ADO, finding 96 % iron occupancy. When a highly conserved His residue (His95), believed to be involved in metal coordination based on comparison to other cupin proteins, was mutated to Ala, both iron occupancy and catalytic activity of ADO were greatly reduced. Dominy *et al*. predicted that iron coordination involved three His residues, as seen in other TDOs[166]. Finally, this study found that ADO was widely expressed, with the highest levels found in the brain and muscle. The authors suggested that this meant that the cysteamine to hypotaurine to taurine pathway may be more important than previously thought; this pathway had been regarded as minor compared to the cysteine to cysteine sulfinic acid to taurine pathway that involves CDO[166].

In 2018, Wang *et al.* investigated whether the Cys-Tyr crosslink observed in CDO (Section 1.4.1) was also present in ADO. Neither of the Cys-Tyr crosslink residues in CDO (Cys93, Tyr157) are conserved in ADO, but the existence of a crosslink had been postulated based on similarities between the two enzymes[144,178]. Using tandem mass spectrometry (MS/MS), Wang *et al.* were able to identify a small amount of crosslinked wild-type (WT)

*Hs*ADO; further analysis suggested that the crosslink formed between Cys220 and Tyr222[178]. A mutant of ADO was then produced, incorporating a 3,5-difluorotyrosine (F_2-Tyr) in place of Tyr222. This substitution led to a greater mass difference between the crosslinked and non-crosslinked forms, as a fluorine (19 Da), rather than a hydrogen (1 Da), was lost during crosslink formation. Mass spectrometry with the F_2-Tyr222 ADO mutant confirmed the existence of a crosslink between Cys220 and Tyr222 (at the *ortho*-position). Wang *et al.* also used ^{19}F NMR spectroscopy to probe the crosslink in the F_2-Tyr222 ADO mutant[178]. Fluoride release, as would be expected from crosslink formation, was detected when the ADO mutant protein was incubated with cysteamine overnight in aerobic conditions, but not when incubated in anaerobic conditions. This suggested that cofactor formation was promoted in the presence of ADO's substrates cysteamine and oxygen[178]. The authors also noted that Cys220 and Tyr222 appeared to be close enough to facilitate crosslink formation in a predicted model of ADO's structure. Mutation of Tyr222 to Ala or Phe led to a reduction in ADO activity with cysteamine (to approximately 23 % relative to WT), as did mutation of another Tyr to Ala (Tyr87Ala), although Tyr87 mutation produced a small reduction in activity (to approximately 51 % relative to WT)[178]. These results suggested that the Cys-Tyr cofactor might be important for ADO catalytic activity with cysteamine.

Following the discovery of the PCOs, Masson *et al*. noted that CDO and ADO shared similarity with the PCOs and tested both mammalian enzymes as potential N-terminal cysteine oxidases (NCOs)[1]. Regulators of G-protein signalling RGS4 and RGS5 had already been identified as targets of the Cys/Arg N-degron pathway in humans[179] (Section 1.3.2), so these were chosen as potential substrates for this study. Masson *et al*. demonstrated that ADO, but not CDO, was required for O_2-dependent degradation of RGS4 and RGS5 in human cells (Figure 1.11), and that this degradation also required active ATE1[1]. These

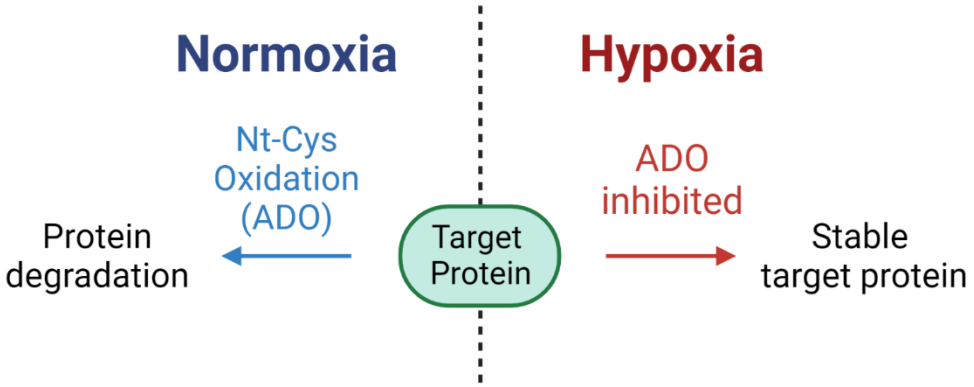


Figure 1.11: Fates of ADO protein substrates in normoxia and hypoxia
Nt-Cys oxidation of target proteins, catalysed by cysteamine dioxygenase (ADO), leads to protein degradation via the Arg/N-degron pathway. ADO is inhibited by hypoxia, so proteins are stabilised. Created with Biorender.com

results suggested that ADO was acting upstream of ATE1 in the N-degron pathway[1]. ADO was also shown to compensate for loss of PCO activity in *Arabidopsis thaliana* plants. Biochemical assays with RGS4/5 N-terminal peptides demonstrated that ADO-catalysed dioxygenation utilised oxygen atoms from molecular O_2, rather than solvent. The modification was confirmed to occur on the Nt-Cys using MS/MS.

Degradation of RGS proteins had previously been reported to involve NO[128,180], so Masson *et al.* investigated the influence of an NO donor (DETA NONOate) on RGS4 stability in wild-type and ADO-deficient cells[1]. The presence of NO reduced accumulation of RGS4 protein during hypoxia in wild-type, but not ADO-deficient, cells[1]. *In vitro* experiments in this study demonstrated that ADO-catalysed dioxygenation does not require NO[1]. The role of NO in the N-degron pathway, and its interaction with ADO, remains an interesting avenue of investigation.

Kinetic parameters for ADO with RGS4/5 peptides were determined (K_M 71 – 123 μM; k_{cat} 17 – 20 s^{-1}); these suggested that ADO has higher affinity and turnover with peptide substrates than with its small molecule substrate cysteamine[1,4]. This is exemplified by comparing the catalytic efficiency ($\frac{k_{cat}}{K_M}$) of ADO with different substrates: 4×10^2 M^{-1} s^{-1} for cysteamine and approximately 2×10^5 M^{-1} s^{-1} with RGS4/5 peptides[1,4,166]. These

kinetic data suggest that the primary biological function of ADO may be Nt-Cys oxidation, although its small molecule thiol dioxygenase activity is likely still relevant[4]. The O_2-sensitivity of ADO was also measured *in vitro*. Masson *et al*. calculated an apparent K_MO_2 of around 50 % O_2 with respect to RGS4 and RGS5 oxidation[1]. This was comparable to that calculated for HIF prolyl hydroxylase enzyme PHD2 (> 450 μM or > 30 %)[35], demonstrating that ADO is a physiologically relevant O_2-sensor[1]. Cysteamine and cysteine were examined as inhibitors of ADO's NCO activity and found to be weak inhibitors (IC_{50}, 37 and 13 mM, respectively)[1].

Masson *et al.* also characterised a chemokine interleukin-32 (IL32) as a substrate of ADO both *in vitro* and *in cellula*[1]. Eight other peptides representing Nt-Cys-bearing human proteins were also screened *in vitro*; ADO was found to have much lower catalytic activity towards these peptides than RGS5 and IL32 peptides, demonstrating that ADO-catalysed Nt-Cys oxidation is not universal[1]. Interestingly, the three protein substrates identified of ADO, RGS4/5 and IL32, were also all reported to be transcriptional targets of HIF[181–183]. Masson *et al.* suggest that the ADO and HIF pathways could interact in specific cellular settings to coordinate the physiological response to hypoxia, with ADO able to operate over a shorter timescale due to its direct effect on protein stability (compared to the transcriptional response induced by HIF)[1].

The study by Masson *et al.* described above was published at the commencement of this project, and was the basis for many of the research questions posed by this Thesis (Section 1.6). The findings on ADO described below have been published since commencement of the research project described in this Thesis.

In 2020, Fernandez *et al.* and Wang *et al.* both performed spectroscopic studies with ADO to characterise its iron centre[184,185]. Fernandez *et al.* used magnetic circular dichroism and

electron paramagnetic resonance to examine ADO iron coordination and binding of various small molecule substrate analogues[184]. Wang *et al.* used electron paramagnetic resonance (EPR), Mössbauer, and UV-visible spectroscopies to investigate the binding of both cysteamine and RGS5. These studies demonstrated that both cysteamine and RGS5 peptide bind to the active site iron in a monodentate manner via their free thiols[184,185]. This is different to the manner in which cysteine coordinates to iron in the active site of CDO; the thiol and amine groups of cysteine have both been observed to coordinate to iron in CDO in a bidentate manner[186,187]. ADO was expected to bind iron via a 3-His facial triad, based on its sequence similarity with other TDOs and the activity of His to Ala mutants[1,166]. Fernandez *et al.*'s results provided further support for this hypothesis[184]. This study also demonstrated that free cysteine could coordinate to iron in ADO, which aligns with the observation that cysteine can (albeit weakly) inhibit ADO activity[184]. Results obtained by Wang *et al*. suggested that substrate binding could reduce ADO's active site metal, reactivating the enzyme if the active site became oxidised[185].

In 2021, crystal structures of mouse and human ADO were published by Fernandez *et al.*[6] and Wang *et al*, respectively[5]. These structures will be discussed in detail in Chapter 6 (page 128). Briefly, both structures confirmed that ADO contains a DSBH motif and a 3-His iron binding site (Figure 1.12), as observed in structures of other TDOs[5,6]. The structures of ADO[5,6] bore a close resemblance to those of the PCOs[66,67] (the only other NCOs with structures reported to date); ADO's active site features a Leu-Asp-Phe motif, similar to the Ile-Asp-Leu motif in PCOs. ADO's structures also contain a potential substrate binding (or hairpin) loop at the entrance to the active site and a wide open active site cavity, as seen in the PCOs[4–6,66,67]. Notably, neither of the ADO structures featured the predicted Cys-Tyr crosslink, although the authors noted that this was not surprising under the conditions in

which the crystals were obtained[5,6]. Fernandez *et al.* also proposed that a small tunnel on the backside of ADO could be involved in O_2 delivery to the active site[6].

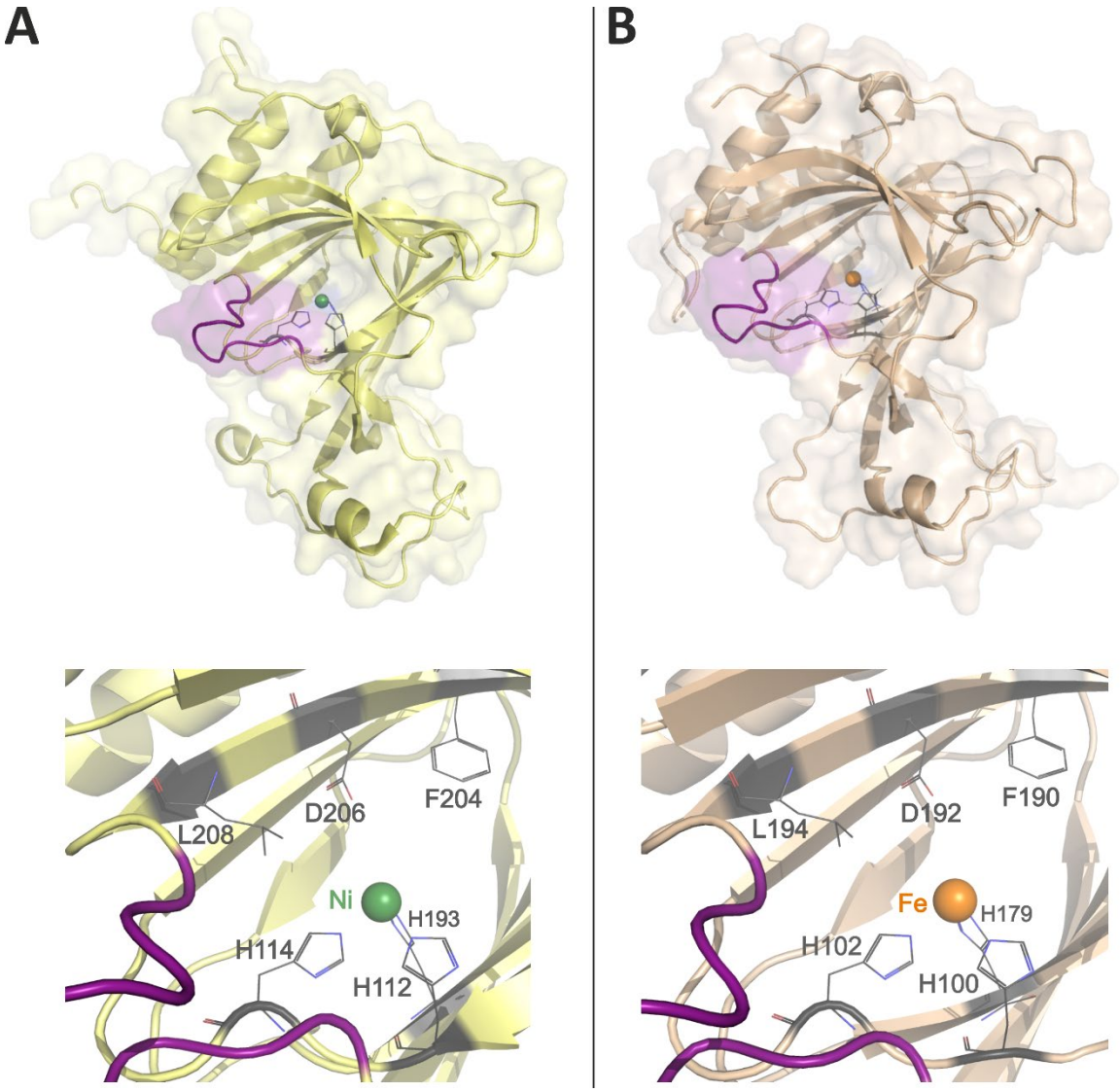


Figure 1.12: Crystal structures of ADO
A: Wang *et al.*'s crystal structure of Ni-bound *Hs*ADO (PDB: 7REI); showing hairpin loop (cartoon, purple), active site Ni (sphere, green), 3 Fe-binding His residues and Phe-Asp-Leu catalytic triad (lines)
B: Fernandez *et al.*'s crystal structure of Fe-bound *Mm*ADO (PDB: 7LVZ; chain D); showing hairpin loop (cartoon, purple), active site Fe (sphere, orange), 3 Fe-binding His residues and Phe-Asp-Leu catalytic triad (lines)

Overall, ADO represents an interesting subset of TDOs, capable of catalysing oxidation of both small molecules and proteins. Of particular interest is its role in O_2-dependent protein degradation and therefore its potential role in the physiological response to hypoxia.

1.6 Thesis Aims

This Thesis aims to address some of the research questions that remain open regarding ADO, especially in regard to its recently uncovered role as an NCO.

Masson *et al*. have shown that ADO is an O_2-sensitive enzyme that catalyses the oxidation of Nt-Cys on target proteins and initiates their degradation via the N-degron pathway[1]. Three novel protein substrates were characterised in this study: RGS4, RGS5, and IL32. However, there are many other Nt-Cys-bearing proteins with the potential to be substrates of ADO. In order to better understand ADO's role in the physiological hypoxic response, it is necessary to establish the extent of its role in O_2-dependent protein signalling. To this end, Chapters 2 and 3 address the research question "What are the extent and determinants of ADO protein substrates?", *in vitro* and *in cellula*, respectively. Chapter 3 contains work carried out in collaboration with Dr. Thomas Keeley (Nuffield Department of Medicine).

As discussed above (Section 1.3.1), there are a variety of different modifications that could also occur on the Nt-Cys of a protein. It is possible that other modifications could influence ADO activity, and by extension the O_2-dependent stability of a protein. Chapter 4 therefore investigates one of the most common modifications, N-terminal acetylation, and its influence on ADO activity, *in vitro* in the context of Nt-Cys. This Chapter contains work carried out in collaboration with the Arnesen group at the University of Bergen.

Research to date has demonstrated that ADO is a fascinating enzyme to study, both in the context of O_2-sensing and taurine biosynthesis. It is also possible that ADO could be a useful therapeutic target as its protein substrates so far identified have been shown to be involved in the cardiovascular system, angiogenesis, and inflammation[188–192]. There is, therefore, a need to develop potent specific inhibitors of ADO for use as chemical probes or therapeutic candidates. This need is addressed in Chapter 5, which investigates peptides and small molecules as inhibitors of ADO.

Finally, prior to 2021, no crystal structures of ADO had been published. Crystal structures are useful for investigating enzyme mechanisms and substrate binding, so one of the original aims of this Thesis was to obtain a crystal structure of ADO. Chapter 6 describes work towards this goal, and towards elucidation of the structural features of ADO that are involved in substrate binding.

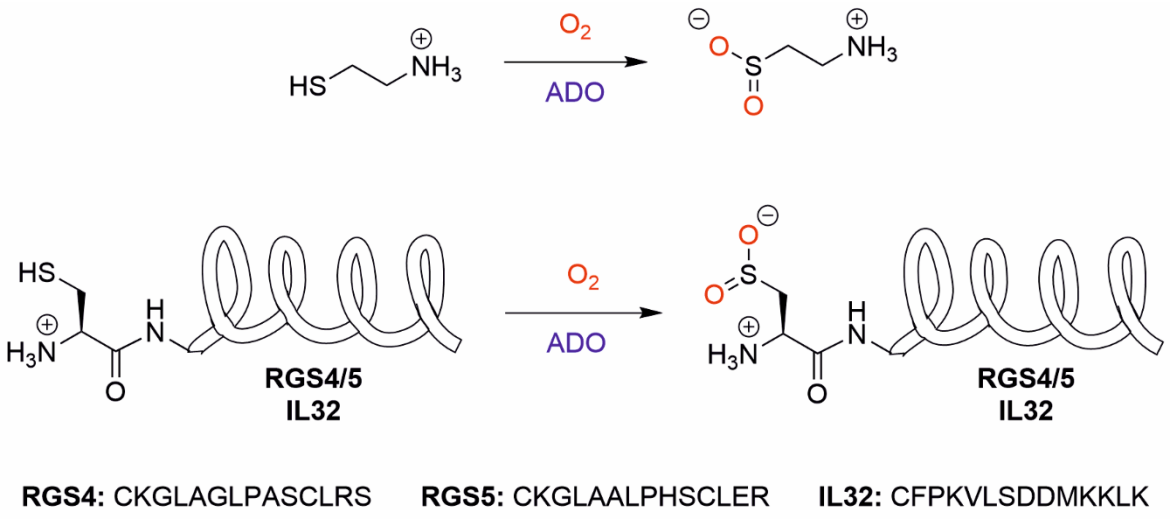
In summary, this Thesis aims to address the following research questions:

- How many other Nt-Cys proteins are substrates of ADO, and can a consensus sequence be determined that will aid future identification of novel protein substrates? (Chapters 2 and 3)
- Can other N-terminal modifications influence ADO catalytic activity, and by extension protein stability? (Chapter 4)
- Can specific and potent inhibitors be found for ADO? (Chapter 5)
- Can ADO's structure provide information on substrate binding and enzyme function? (Chapter 6)

2 Investigation of *Hs*ADO Substrates *in vitro*

2.1 Introduction

Cysteamine dioxygenase (2-aminoethanethiol dioxygenase, ADO) was first characterised as a small molecule thiol dioxygenase (TDO), oxidising cysteamine to hypotaurine (Scheme 2.1 **A**) as part of the taurine biosynthesis pathway[166]. Later, ADO was found to have another role oxidising the N-terminal cysteine (Nt-Cys) of target proteins to cysteine sulfinic acid (Scheme 2.1 **B**) in an O_2-dependent manner as part of the Arg/N-degron pathway[1]. Three protein substrates of *Hs*ADO were characterised *in vitro* and *in cellula*: regulators of G-protein signalling (RGS4 (accession P49798) and RSG5 (accession O15539)) and a chemokine, interleukin-32 (IL32 (accession P24001)) (Scheme 2.1 **B**). These are not the only sequences in the human proteome with the Nt-Cys motif. In October 2022, there were over 400 potential Met-Cys-initiating (MC-initiating) sequences in the human proteome according to UniProt[193]. These represent a broad range of potential ADO substrates. This raises the question of whether ADO is able to regulate the stability of other proteins, and if so whether these proteins have a role in the hypoxic response.



Scheme 2.1: Reactions catalysed by cysteamine dioxygenase (ADO)
A: Oxidation of cysteamine to hypotaurine.
B: Oxidation of N-terminal cysteine to cysteine sulfinic acid; the first 14 amino acids of Met-excised N-termini of RGS4, RGS5, and IL32 proteins are shown below reaction scheme.

2.2 Chapter Aims

In this Chapter, biochemical experiments with recombinant *N*-His_6-tagged *Hs*ADO were used to investigate various substrates of ADO *in vitro*. First, recombinant *Hs*ADO prepared in this Thesis was confirmed to catalyse oxidation of known substrates, cysteamine and an RGS4 peptide. Then, different peptides based on the Nt-sequence of RGS4 were used to investigate: whether Nt-methionine cleavage is necessary for ADO-catalysed Nt-Cys oxidation; if ADO is able to catalyse oxidation of internal Cys residues; and the length of peptide required for ADO binding.

This Chapter next addressed the question of whether other protein substrates of ADO could be identified in the human proteome. Peptides representing the Met-excised N-terminus of different MC-initiating sequences were screened against ADO in initial activity assays. The peptides screened varied in the identity of amino acids after the Nt-Cys, and so these activity assays also provided information of ADO's substrate preferences. These preferences were validated using 'optimum' substrate (OS) and 'anti'-substrate (AS) peptides. Finally, substrates of related plant cysteine oxidase (PCO) enzymes were investigated as potential ADO substrates.

2.3 Preparation of materials for *in vitro* experiments

2.3.1 Expression and purification of recombinant *Hs*ADO

All *in vitro* assays used recombinant human *N*-His_6-tagged $ADO_{1–270}$ protein (*Hs*ADO) (Appendix 1), which was overexpressed in Rosetta (DE3) *E. coli* cells and purified from lysate by fast protein liquid chromatography (FPLC) (Chapter 8, page 165). *Hs*ADO protein was first purified by immobilised metal affinity chromatography (IMAC) on a 5 mL HisTrap™ Ni(II) affinity chromatography column, eluting with increasing concentrations of imidazole. Protein fractions were analysed by ultraviolet (UV) trace at A_{280nm} (Appendix 1) and sodium dodecyl sulfate - polyacrylamide gel electrophoresis

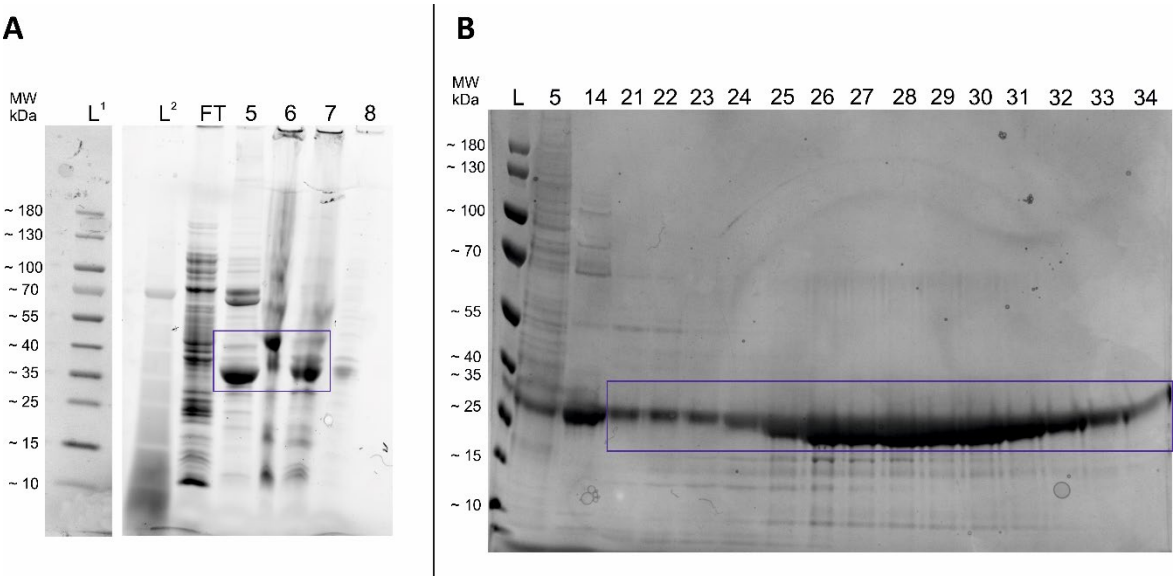


Figure 2.1: SDS-PAGE gels from IMAC and SEC purification of recombinant *Hs*ADO
A: SDS-PAGE gel from IMAC. L^1 – ladder cropped from image of same gel imaged under different conditions, L^2 - ladder imaged using same conditions as remainder of gel image, FT – flow-through, numbers – collected fractions. Fractions 5-7 were selected for further purification.
B: SDS-PAGE gel from SEC elution. L – ladder, numbers – collected fractions. Fraction 21 – 34 were pooled, concentrated and frozen at -80 °C for use in *in vitro* experiments.
Expected MW of *N*-His_6-tagged $ADO_{1–270}$ protein (*Hs*ADO) is 32 kDa.

(SDS-PAGE) (Figure 2.1) to select fractions of sufficient purity, which were pooled and concentrated using 10,000 Da MWCO ultracentrifugation device. An aliquot of IMAC purified protein was flash frozen in liquid N_2 and retained for later comparison against other protein batches (Section 2.5). *Hs*ADO protein was further purified by size exclusion chromatography (SEC) on a Superdex™ 75 pg column, eluting fractions in 50 mM Tris, 400 mM NaCl, pH 7.5. Fractions of sufficient purity were selected through analysis of UV trace (A_{280nm}) and SDS-PAGE (Figure 2.1). Selected fractions were pooled and concentrated as before, then aliquoted and flash frozen in liquid N_2 and stored at −80 °C.

2.3.2 Preparation of peptides for *in vitro* assays

Peptides used for *in vitro* assays were either purchased from GL Biochem or GenScript (Appendix 2), or synthesised by solid-phase peptide synthesis (SPPS) (Table 2.1), unless otherwise stated. All peptides were obtained with a C-terminal amide group. Peptides in Table 2.1 were synthesised using a Liberty Blue™ automated microwave peptide synthesiser (CEM) by standard Fmoc-protected SPPS (Chapter 8, page 170). Peptides were

subsequently purified by high-performance liquid chromatography (HPLC) and analysed by MALDI-TOF mass spectrometry to confirm peptide mass and purity (Appendix 3).

Table 2.1: Peptides prepared by SPPS for *in vitro* substrate investigations

Peptide Name	Sequence	Molecular Weight / Da
$RGS4_{2-15}$[C3]	KCGLAGLPASCLRS-NH_2	1375.67
$RGS4_{2-15}$[C5]	KGLCAGLPASCLRS-NH_2	1375.67
$RGS4_{2-15}$[C7]	KGLAGCLPASCLRS-NH_2	1375.67
$RGS4_{2-12}$	CKGLAGLPASC-NH_2	1018.26
$RGS4_{2-10}$	CKGLAGLPA-NH_2	828.04
$RGS4_{2-8}$	CKGLAGL-NH_2	659.85
$RGS4_{2-6}$	CKGLA-NH_2	489.64
OS1	CRGRRKKATTKAWR-NH_2	1717.09
OS2	CRGGRKKATTKAWR-NH_2	1617.95
OS3	CRGGFKKATTKAWR-NH_2	1608.94
OS4	CRGGFMKATTKAWR-NH_2	1611.95
OS5	CRGGFMKAPTKAWR-NH_2	1607.97
$ZPR2_{2-15}$	CLTTSEPPFPDTDT-NH_2	1522.64

2.4 Recombinant *Hs*ADO catalyses oxidation of cysteamine

The first substrate identified of *Hs*ADO was cysteamine[166] (Scheme 2.1 **A**). Thin layer chromatography (TLC) was used to confirm that recombinant *Hs*ADO prepared above (Section 2.3.1) could oxidise cysteamine to hypotaurine. The protocol for the TLC assay was adapted from that used by Fernandez *et al.*[184] (Chapter 8, page 173). Cysteamine (20 mM) was incubated with recombinant *Hs*ADO (20 μM) at 37 °C for 30 minutes in assay buffer (50 mM Bis-Tris Propane, 50 mM NaCl, pH 7.5). The reaction was stopped by heating at 95 °C for 2 minutes to denature the enzyme, then the precipitated protein was pelleted by centrifugation. The supernatant was spotted onto a silica TLC plate alongside cysteamine (20 mM) and hypotaurine (20 mM) control solutions in assay buffer, and placed in a beaker containing 20:20:60 (v/v) water/acetic acid/1-butanol running solution until the

solvent front had progressed sufficiently up the plate (> 1 hour). The plate was removed and developed in a ninhydrin solution (1.5 % w/v) in acetic acid (3 % v/v) in ethanol, then heated to activate the stain.

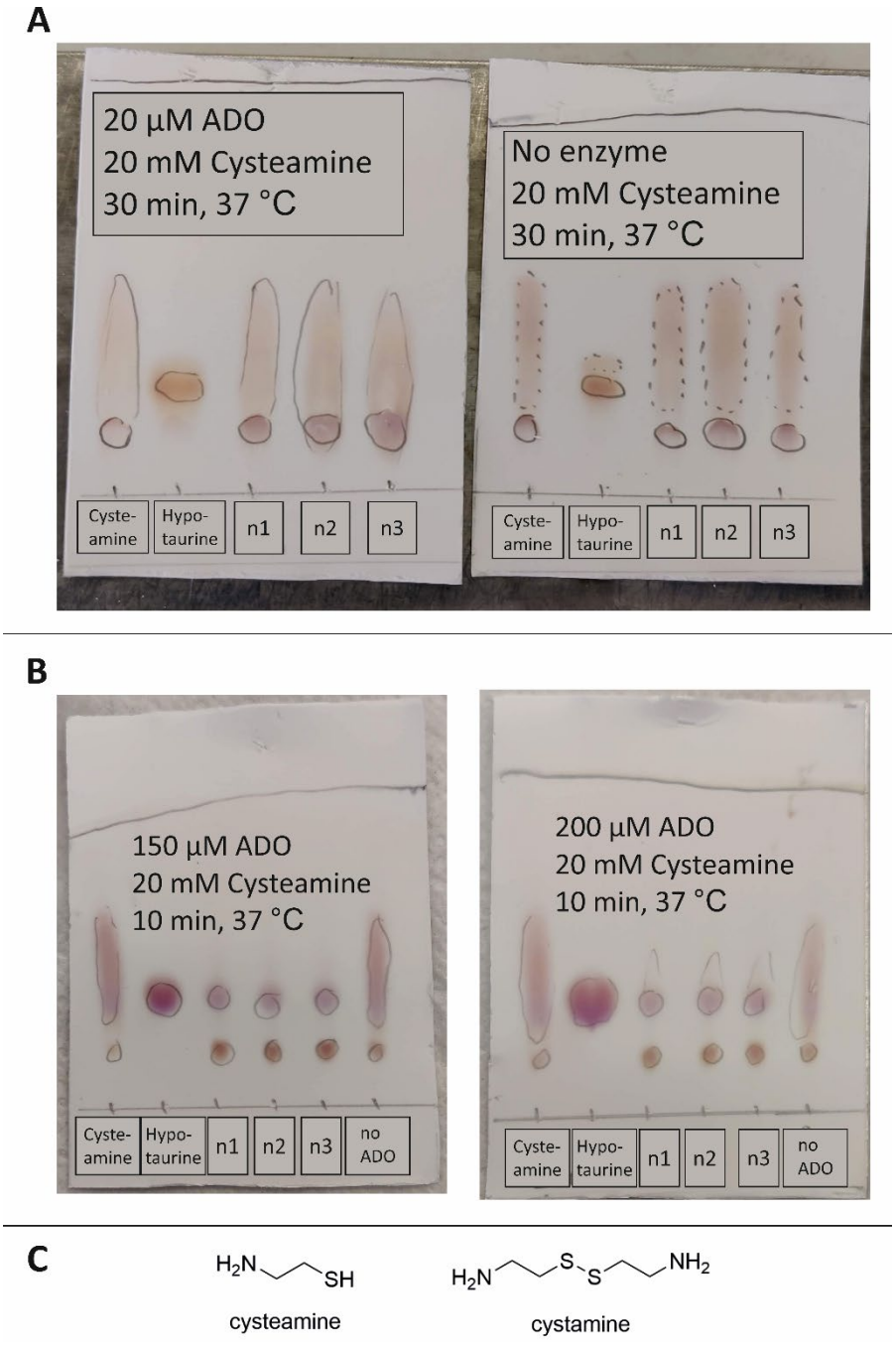


Figure 2.2: TLC analysis of *Hs*ADO-catalysed cysteamine oxidation
Analysis of samples from reaction of recombinant HsADO with cysteamine at 37 °C. Reactions were stopped by heat-denaturation and spotted onto TLC plate. Reactions were carried out in triplicate. 20 mM solutions of cysteamine and hypotaurine were spotted as controls. Upper broad spot for cysteamine is assumed to be monomeric cysteamine, while lower spot is assumed to be dimeric cystamine.
A: Reaction of 20 µM ADO with 20 mM cysteamine for 30 minutes.
B: Reaction of 200 µM or 150 µM ADO with 20 mM cysteamine for 10 minutes.
C: Structures of cysteamine and cystamine

No hypotaurine formation was observed by TLC under initial reaction conditions (20 mM cysteamine, 20 μM *Hs*ADO) (Figure 2.2 **A**). Therefore, the reaction was repeated with the concentration of *Hs*ADO increased to 150 μM and 200 μM. Under these conditions strong spots for hypotaurine were observed in all replicates after 10 minutes (Figure 2.2 **B**). In all TLC analyses, two spots were observed for cysteamine: a broad upper spot and a lower spot; this was also observed by Fernandez *et al.*[184]. The lower spot was postulated to be dimeric cystamine and the upper monomeric cysteamine (Figure 2.2 **C**). The broad nature of the upper spot is likely due to interaction of the free thiol with the silica in the TLC plate. The intensity of the lower cystamine spot did not vary visibly between cysteamine controls, enzyme-catalysed reactions, and no enzyme controls. In contrast, only a faint cysteamine spot could be seen in reactions with 150 μM and 200 μM enzyme (Figure 2.2 **B**), which suggested that almost all cysteamine had been consumed under these conditions. Thus, recombinant *Hs*ADO prepared above (Section 2.3.1) was confirmed to be an active small molecule TDO enzyme, capable of oxidising cysteamine to hypotaurine.

2.5 Recombinant *Hs*ADO catalyses oxidation of RGS4 peptide

In 2019, ADO was shown to be an N-terminal cysteine oxidase (NCO), in addition to its small molecule TDO activity[1]. In this paper, a peptide representing the Met-excised N-terminus of RGS4 ($RGS4_{2\text{-}15}$) was shown to be a substrate of *Hs*ADO *in vitro*, so this peptide was used to confirm NCO activity of recombinant *Hs*ADO prepared above (Section 2.3.1). $RGS4_{2\text{-}15}$ (20 μM) was incubated with recombinant *Hs*ADO (0.05 μM) at 37 °C for 3 minutes in assay buffer (50 mM Bis-Tris Propane, 50 mM NaCl, pH 7.5) supplemented with cofactors (TCEP (5 mM), iron(II) (20 μM), and L-ascorbate (1 mM)). Samples were taken from the reaction mixture at specific time intervals and quenched in 1 % (v/v) formic acid (1:10 v/v) (Chapter 8, page 174). Analysis of samples was carried out by high throughput mass spectrometry (MS) using a RapidFire RF360 sampling robot

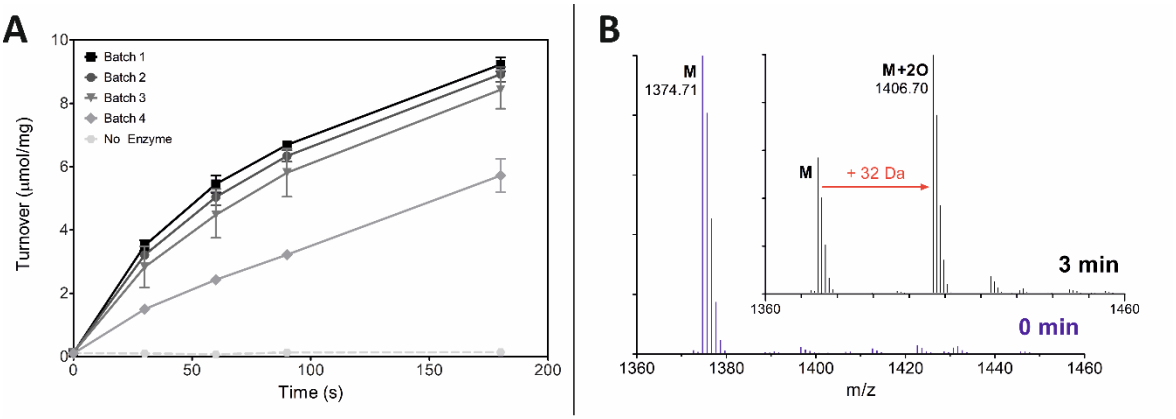


Figure 2.3: *Hs*ADO-catalysed oxidation of RGS4$_{2-15}$ peptide
Oxidation of RGS4$_{2-15}$ peptide incubated with *Hs*ADO at 37 °C for 3 minutes (20 µM peptide, 0.05 µM ADO).
A: Turnover of RGS4$_{2-15}$ peptide incubated with different batches of recombinant *Hs*ADO: Batch 1 = prepared by Dr. Mark White (IMAC purification only); Batch 2 = prepared in this Thesis (IMAC purification only); Batch 3 = prepared in this Thesis (IMAC and SEC purification): Batch 4 = protein from Batch 1 after one freeze-thaw cycle.
B: Mass spectra of RGS4$_{2-15}$ peptide before and after incubation with *Hs*ADO (Batch 3 (**A**)). Deconvoluted spectra were produced with Agilent MassHunter Qualitative Analysis software.

connected to a 6530 Accurate-Mass Q-ToF mass spectrometer (Agilent) operated in positive ion mode (Chapter 8, page 181). Spectra were analysed using MassHunter Qualitative Analysis software (Agilent) and the integrated area under peaks of interest were extracted using RapidFire Integrator (Agilent). Oxidation of RGS4$_{2-15}$ peptide was evidenced by a mass shift of +32 Da (Figure 2.3 **B**), corresponding to the addition of two oxygen atoms to the sulfur atom of the Nt-Cys.

The *Hs*ADO protein used by Dr. Mark White (former Flashman Group member) in previous work[1] was purified by IMAC only; therefore *Hs*ADO protein purified by IMAC only was compared with protein purified by both IMAC and SEC (Section 2.3.1). Comparison of the turnover of the RGS4$_{2-15}$ peptide following incubation with different batches of *Hs*ADO protein showed that there was not a notable difference in activity of *Hs*ADO protein purified by IMAC only or by IMAC and SEC, and that protein prepared above (Section 2.3.1) had comparable activity to protein expressed and purified by Dr. White (Figure 2.3 **A**). Recombinant *Hs*ADO protein that had been subjected to an additional freeze-thaw cycle (Batch 4, Figure 2.3 **A**) catalysed oxidation of RGS4$_{2-15}$ peptide more slowly than all other

batches tested. Overall, these results demonstrated that recombinant *Hs*ADO prepared in this Thesis had similar NCO activity to enzyme prepared by Dr. White and could therefore be used to investigate oxidation of peptide substrates.

2.6 Investigation of NCO activity with RGS4 peptides

2.6.1 Methionine cleavage is necessary for Cys2 oxidation

In eukaryotes, initiation of protein translation requires binding of initiator Met-tRNA to the initiator codon AUG[194,195]. Therefore, nascent eukaryotic proteins emerge from the ribosome bearing an Nt-Met residue. Commonly, this initiator methionine (iMet) residue is removed co-translationally by methionine amino peptidases (MAPs)[59,79]. To confirm that this methionine excision is necessary for *Hs*ADO to catalyse oxidation of a cysteine at

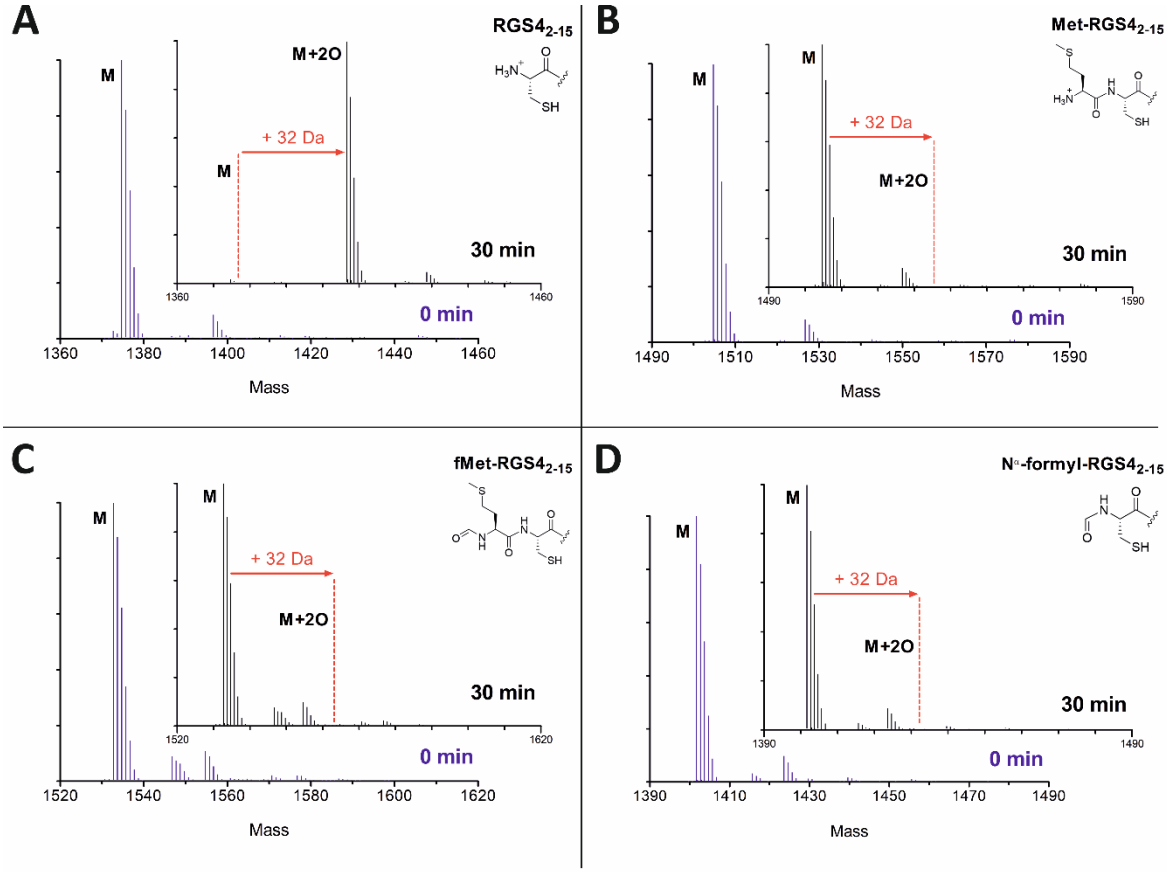


Figure 2.4: *Hs*ADO-catalysed oxidation of Met-Cys RGS4 peptides
Mass spectra of RGS4 peptides before and after incubation with *Hs*ADO at 37 °C for 30 minutes (100 µM peptide, 0.1 µM ADO). Deconvoluted spectra were produced with Agilent MassHunter Qualitative Analysis software. N-terminal structures of peptides are shown in top right of each spectra.
A: RGS4$_{2-15}$ **B:** RGS4$_{1-15}$ (Met-RGS4$_{2-15}$) **C:** N$^{\alpha}$-formyl-RGS4$_{1-15}$ (fMet-RGS4$_{2-15}$)
D: N$^{\alpha}$-formyl-RGS4$_{2-15}$ (fCys-RGS4$_{3-15}$)
RGS4$_{2-15}$ sequence: CKGLAGLPASCLRS

residue 2 (Cys2), a peptide representing the non-Met-excised N-terminus of RGS4 (Met-RGS4$_{2-15}$) was incubated with *Hs*ADO at 37 °C for 30 minutes (100 µM peptide, 0.1 µM *Hs*ADO) using the same assay conditions as above (Section 2.5). Oxidation of peptides was monitored using high throughput MS as described above (Section 2.5). After 30 minutes, mass spectra of RGS4$_{2-15}$ and Met-RGS4$_{2-15}$ showed that oxidation (evidenced by a +32 Da mass increase) only occurred in the absence of the Nt-Met (Figure 2.4 **A** and **B**). Proteins synthesised in mitochondria initiate with a formyl methionine (fMet) residue instead of methionine, so an RGS4 peptide bearing an Nt-fMet residue (fMet-RGS4$_{2-15}$) was also tested (Figure 2.4 **C**). This peptide was also not oxidised after 30 minutes incubation with *Hs*ADO. It is possible that the formyl group could migrate onto the Nt-Cys[196]; an Nt-formyl cysteine RGS4 peptide (fCys-RGS4$_{3-15}$) was also found to not be oxidised when incubated with *Hs*ADO (Figure 2.4 **D**). These results demonstrated that *Hs*ADO can only catalyse oxidation of Cys2 if it is exposed following removal of an Nt-Met (or fMet) residue. Migration of a formyl group onto the Nt-Cys also prevents *Hs*ADO-catalysed oxidation.

2.6.2 *Hs*ADO does not catalyse oxidation of internal cysteines

After establishing that Nt-Met excision was necessary for *Hs*ADO activity, it was next confirmed that *Hs*ADO could not catalyse oxidation of internal cysteines in a known substrate sequence (RGS4). Three peptides were synthesised (Section 2.3.2) in which the Nt-Cys (Cys2) was moved along the Met-excised Nt-sequence of RGS4 (CKGLAGLPASCLRS) to positions 3, 5, and 7: RGS4$_{2-15}$[C3] (KCGLAGLPASCLRS); RGS4$_{2-15}$[C5] (KGLCAGLPASCLRS); RGS4$_{2-15}$[C7] (KGLAGCLPASCLRS). These peptides were incubated with *Hs*ADO at 37 °C for 1 minute using the same assay conditions as above (Section 2.5). Oxidation of peptides was monitored using high throughput MS as described above (Section 2.5). The characteristic +32 Da mass shift indicating oxidation

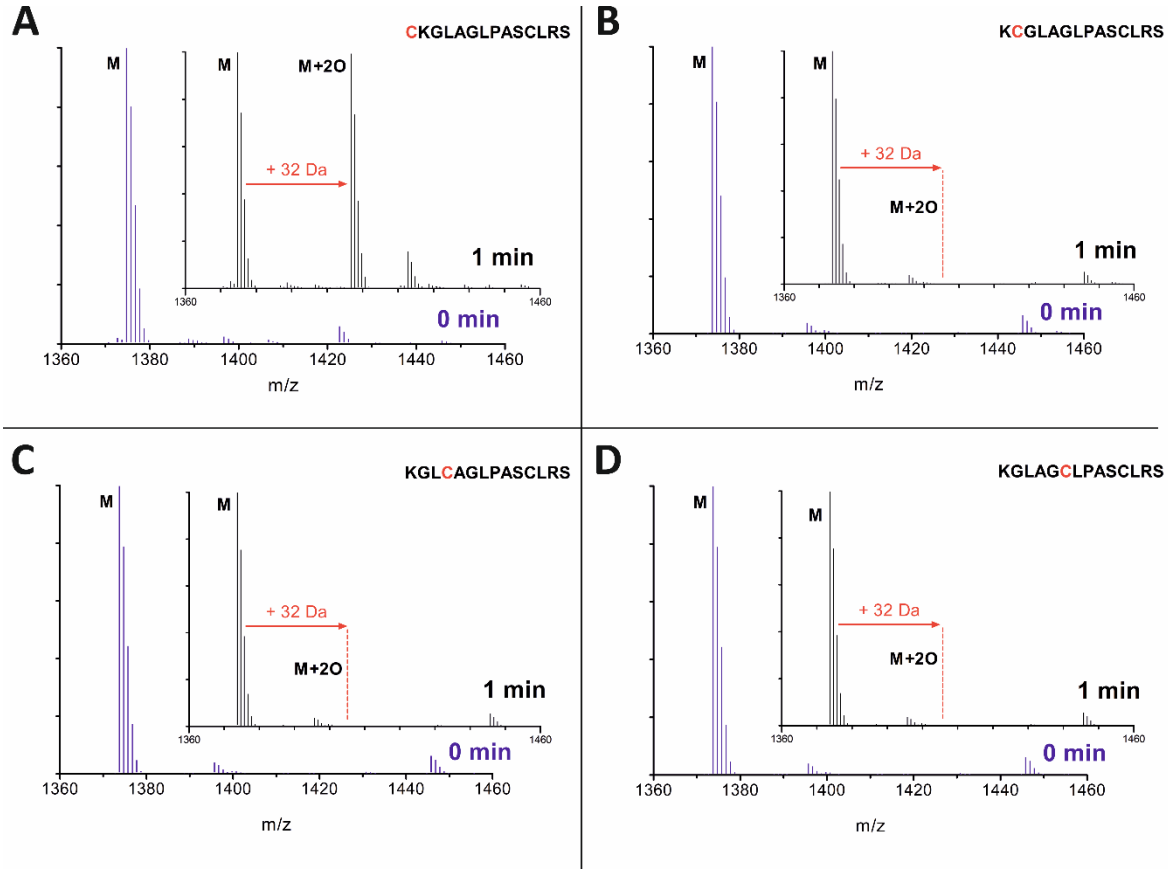


Figure 2.5: *Hs*ADO-catalysed oxidation of RGS4 C3/C5/C7 mutant peptides
Mass spectra of $RGS4_{2-15}$ peptides before and after incubation with *Hs*ADO at 37 °C for 1 minute (100 µM peptide, 0.1 µM ADO). Deconvoluted spectra were produced with Agilent MassHunter Qualitative Analysis software.
A: wild-type RGS4 sequence
B: RGS4 sequence with Nt-Cys moved to position 3 ($RGS4_{2-15}$[C3])
C: RGS4 sequence with Nt-Cys moved to position 5 ($RGS4_{2-15}$[C5])
D: RGS4 sequence with Nt-Cys moved to position 7 ($RGS4_{2-15}$[C7])

was only observed in the mass spectra of the wild-type $RGS4_{2-15}$ peptide (Figure 2.5). Previously, Masson *et al.* had used tandem mass spectrometry (MS/MS) to demonstrate that oxidation of RGS4 and RGS5 peptides occurred on the Nt-Cys[1], but had not tested the effect of moving the Cys along the sequence as described here. The data presented here demonstrated that *Hs*ADO can only catalyse oxidation of Nt-Cys under these conditions.

2.6.3 *Hs*ADO can catalyse oxidation of 14- to 7-mer peptides

Experiments were next performed to investigate the requirements for efficient binding of the RGS4 substrate to the *Hs*ADO active site. It is not yet known how *Hs*ADO binds peptide and protein substrates; at the beginning of this project, ADO's role in Nt-Cys oxidation had

not yet been published and no crystal structures of ADO had been obtained. Spectroscopic data has subsequently been published suggesting that RGS5 peptide binds to the active site iron in *Hs*ADO via the thiol group in a monodentate manner[185]. Two structures of ADO were then published in 2021[5,6], however no structures have been published with substrate bound. Computational studies by Fernandez *et al*. with an RGS4/5 tetramer (CKGL) suggested that the Nt-amine can form a hydrogen bond to an Asp residue in the active site (*Mm*ADO Aps192), and that this peptide is able to interact with a larger number of residues in the active site cavity than the small molecule substrate cysteamine[6].

In order to investigate how much of the N-terminus of a protein is required for ADO to bind to and catalyse oxidation, peptides representing the first five (residues 2-6), seven (residues 2-8), nine (residues 2-10), and eleven (residues 2-12) amino acids of the Met-excised N-terminus of RGS4 were synthesised to test alongside the $RGS4_{2-15}$ peptide previously used in *in vitro* experiments (Section 2.3.2). Peptides were incubated with *Hs*ADO for 1 minute at 37 °C, and their oxidation monitored by high throughput MS as described above (Section 2.5). Unfortunately, $RGS4_{2-6}$ peptide could not be detected by high throughput MS, likely due to its small size. All other RGS4 peptides were oxidised in this assay (Figure 2.6). The initial activity of *Hs*ADO was lower with the shorter peptides compared to $RGS4_{2-15}$, although this decrease was only statistically significant for $RGS4_{2-12}$. These data provided evidence that *Hs*ADO is able to bind as few as seven amino acids in its active site. This is perhaps not surprising as *Hs*ADO is known to bind to and catalyse oxidation of the small molecule cysteamine; however *Hs*ADO is unable to catalyse oxidation of free cysteine[166]. It is possible that shorter peptides could also bind to *Hs*ADO, as suggested by Fernandez *et al.*'s computational studies[6], but it was not possible to test this experimentally using established assays. As *Hs*ADO was the most active with the 14-mer $RGS4_{2-15}$, peptides of this length continued to be used for biochemical experiments.

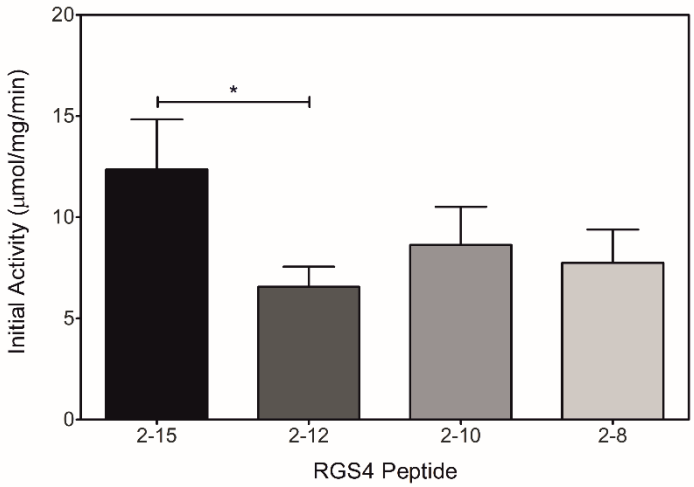


Figure 2.6: *Hs*ADO -catalysed oxidation of RGS4 7-, 9-, 11- and 14-mer peptides
Initial activity of *Hs*ADO with different RGS4 peptides (100 µM peptide, 0.1 µM HsADO, 37 °C). Reactions were carried out in triplicate. Error bars show standard deviation. Statistical analysis carried out using GraphPad Prism v5 1-way ANOVA with Tukey's multiple comparison post-test. Only significant p value summaries (< 0.05) are shown.

2.7 *Hs*ADO is selective amongst Nt-Cys peptides *in vitro*

Three protein substrates of *Hs*ADO have been characterised so far: RGS4, RGS5, and IL32[1]. There are many other MC-initiating sequences in the human proteome. These represent a range of potential substrates of *Hs*ADO, whose stability could be regulated in an O_2-dependent manner. This Section describes experiments that investigated which of these could be *Hs*ADO substrates at the peptide level.

2.7.1 *Hs*ADO activity *in vitro* varies with Nt-Cys peptide sequence

In order to probe the extent of *Hs*ADO's role in O_2-dependent protein signalling a variety of MC-initiating sequences were selected from the human proteome (Figure 2.7). An initial selection was made by Dr. Thomas Keeley and Dr. Norma Masson (Ratcliffe Group, Nuffield Department of Medicine), which provided a good distribution of different amino acids at the position after the Nt-Cys to help determine any substrate specificity requirements of ADO (Figure 2.7 **A**). To demonstrate the distribution of different amino acids at residue 3 in the sequences selected (Figure 2.7 **A**), the amino acid composition at

this position in the selected proteins was compared to the amino acid composition at all positions in all proteins in the UniProt Swiss-Prot database in 2022[197] (Figure 2.7 **B**). While some amino acids were over-represented in the initial selection (e.g., Cys, Trp), and others were under-represented (e.g., Ala, Ile), overall, there was a good distribution of different amino acids after the Nt-Cys. Only three amino acids were not represented at residue 3 in the selected sequences (Gln, Met, and Tyr) and many of the other amino acids were similarly represented at residue 3 in the selected sequences as they were in the Swiss-Prot database (e.g. Asn, Glu, Lys, Thr, and Val). Cysteine had the second lowest abundance compared to all amino acids in the Swiss-Prot database (1.38 %), so it is not surprising that not all amino acids could be well represented at residue 3 in the initial selection of Nt-Cys sequences.

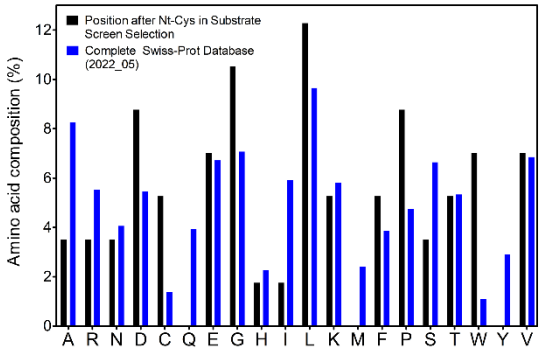
In addition to this initial selection, some sequences were selected from the human proteome whose N-terminal amino acids resembled those of known substrates RGS4/5 and IL32 (Figure 2.7 **C** (blue)). Specifically, these sequences all possessed a basic or aromatic residue at the position after the Nt-Cys. Three other sequences were selected for testing based on negative results in cellular assays with *Hs*ADO performed by Dr. Keeley (Figure 2.7 **C** (orange)): SHHN, AAAS, and LYPLA2 (unpublished data, Ratcliffe Group). In contrast, NT5E was selected for testing based on promising initial results from Dr. Keeley's cellular assays, although results from these assays were inconsistent. GJC3 was selected following identification as a potential substrate of ADO in mice by Dr. Masson (unpublished data, Ratcliffe Group) (Figure 2.7 **C** (grey)). Finally, another Nt-Cys bearing regulator of G-protein signalling, RGS16, was included for screening. As with RGS4 and RGS5, RGS16 had previously been shown to be a substrate of the Arg/N-degron pathway[128,198], but had not yet been characterised as a substrate of ADO.

Peptides representing the first fourteen amino acids of the Met-excised N-termini of Nt-Cys sequences (Figure 2.7 **A** and **C**) were incubated with *Hs*ADO for 1 minute at 37 °C and analysed by high throughput MS, as described above (Section 2.5). Due to solubility issues

A

Accession	Gene	Peptide Sequence													
P49798	RGS4	C	K	G	L	A	G	L	P	A	S	C	L	R	S
O15539	RGS5	C	K	G	L	A	A	L	P	H	S	C	L	E	R
P24001	IL32	C	F	P	K	V	L	S	D	D	M	K	K	L	K
Q8TE56	ADAMTS17	C	D	G	A	L	L	P	P	L	V	L	P	V	L
Q60I27	ALS2CL	C	N	P	E	E	A	A	L	L	R	L	E	E	V
D6R9V6	ARHGAP22	C	V	P	P	H	R	K	D	G	S	I	P	C	P
F5H1R4	ARHGAP45	C	I	C	G	T	A	H	P	V	L	D	E	G	P
B4E3H0	ARL16	C	L	L	L	G	A	T	G	V	G	K	T	L	L
P08243	ASNS	C	G	I	W	A	L	F	G	S	D	D	C	L	S
Q5JW98	CALHM4	C	P	T	L	N	N	I	V	S	S	L	Q	R	N
A0A087WTJ6	CDX1	C	W	T	R	I	R	P	G	T	P	S	S	P	G
A0A0C4DGJ7	CHURC1	C	G	D	C	V	E	K	E	Y	P	N	R	G	N
C9K0K2	CPSF4	C	P	F	R	H	I	S	G	E	K	T	V	V	C
H7BXY3	DHX30	C	V	N	P	T	P	G	G	T	I	S	R	A	S
Q99766	DMAC2CL	C	C	A	V	S	E	Q	R	L	T	C	A	D	Q
A0A1B0GTX1	DNMBP	C	E	T	Q	K	A	R	K	E	R	E	V	E	E
F6S220	DOCK2	C	V	C	A	E	D	G	K	T	L	P	V	A	V
A2A369	DOCK8	C	G	E	D	A	S	N	A	M	P	V	R	R	E
P63167	DYNLL1	C	D	R	K	A	V	I	K	N	A	D	M	S	E
Q5THR3	EFCAB6	C	K	M	A	I	I	P	D	W	L	R	S	H	P
Q8TBG4	ETNPPL	C	E	L	Y	S	K	R	D	T	L	G	L	R	K
S4R381	FGFR2	C	L	N	K	T	K	Q	L	F	G	V	A	K	R
Q9BZD3	GCOM2	C	L	L	P	R	G	F	E	P	Q	A	P	E	D
O94808	GFPT2	C	G	I	F	A	Y	M	N	Y	R	V	P	R	T
Q99680	GPR22	C	F	S	P	I	L	E	I	N	M	Q	S	E	S
P07203	GPX1	C	A	A	R	L	A	A	A	A	A	A	A	Q	S
K7EN03	iASPP-SV*	C	W	L	Y	A	L	E	V	V	E	R	E	C	W
P46695	IER3/IEX1	C	H	S	R	S	C	H	P	T	M	T	I	L	Q
P29459	IL12A/NKSF1	C	P	A	R	S	L	L	L	V	A	T	L	V	L
B4DNV3	IL6	C	E	S	S	K	E	A	L	A	E	N	N	L	N
A0A0A0MTP6	INPPL1	C	T	R	I	A	F	C	L	M	E	K	I	S	W
P17275	JUNB	C	T	K	M	E	Q	P	F	Y	H	D	D	S	Y
A0A087WZG8	LIPI	C	W	V	R	S	D	N	K	R	P	C	L	E	F
P58546	MTPN	C	D	K	E	F	M	W	A	L	K	N	G	D	L
B4DVL2	NAAA	C	D	F	M	N	L	S	L	A	D	C	L	L	V
A0A0B4J216	NADSYN1	C	C	Q	V	C	E	A	V	R	S	G	N	E	E
A0A0B4J2B3	NBPF14	C	R	D	G	Y	E	P	C	V	N	E	G	M	C
Q7Z6G3	NECAB2	C	E	R	A	A	R	L	C	R	A	G	A	H	R
Q7Z3S9	NOTCH2NLA	C	V	T	Y	H	N	G	T	G	Y	C	K	C	P
Q8NGY3	OR6K3	C	W	T	M	P	S	P	F	T	G	S	S	T	R
H3BT22	ORC6	C	L	D	L	A	A	S	W	M	K	C	P	L	D
O15547	P2RX6	C	P	Q	L	A	G	A	G	S	M	G	S	P	G
F6U1T9	PPP3R1	C	S	H	F	D	A	D	E	I	K	R	L	G	K
B7Z3L0	RAB37	C	C	K	V	M	L	L	G	D	T	G	V	G	K
E9PGC0	RASA1	C	L	V	K	S	W	F	L	S	I	R	S	R	W
Q8IXN7	RIMKLA	C	S	Q	L	W	F	L	T	D	R	R	I	R	E
H3BM89	RPL4	C	R	G	G	R	M	F	A	P	T	K	T	W	R
Q5QGT7	RTP2	C	T	S	L	T	T	C	E	W	K	K	V	F	Y
E9PDM8	SEC24D	C	P	F	M	Q	F	I	E	G	G	R	R	Y	Q
Q15465	SHHC	C	F	P	G	S	A	T	V	H	L	E	Q	G	G
Q15465	SHHN	C	G	P	G	R	G	F	G	K	R	R	H	P	K
B0QZ35	SIRT1	C	L	C	S	G	R	K	T	I	L	E	I	Y	P
F5H013	SNRPG	C	L	F	M	D	K	K	L	S	L	K	L	N	G
Q9GZT4	SRR	C	A	Q	Y	C	I	S	F	A	D	V	E	K	A
Q6ZMT1-2	STAC2	C	K	V	S	V	H	L	W	C	S	E	E	I	
E9PBG3	TRIM36	C	D	L	C	K	P	P	P	Q	E	S	T	K	S
B2RNG4	TRIM6-TRIM34	C	G	S	E	R	I	L	Q	A	G	N	I	L	E
Q9Y277	VDAC3	C	N	T	P	T	Y	C	D	L	G	K	A	A	K
Q9P243-2	ZFAT	C	K	C	C	N	L	F	S	P	N	Q	S	E	L
Q504Y0	ZIP12	C	F	R	T	K	L	S	V	S	W	V	P	L	F

B



C

Accession	Gene	Peptide Sequence													
P49798	RGS4	C	K	G	L	A	G	L	P	A	S	C	L	R	S
O15539	RGS5	C	K	G	L	A	A	L	P	H	S	C	L	E	R
O15492	RGS16	C	R	T	L	A	A	F	P	T	T	C	L	E	R
P24001	IL32	C	F	P	K	V	L	S	D	D	M	K	K	L	K
Q8N6D5	ANKRD29	C	R	M	S	F	K	K	E	T	P	L	A	N	A
P19022	CDH2	C	R	I	A	G	A	L	R	T	L	L	P	L	L
Q8TCI5	PIFO	C	F	S	R	A	D	A	A	D	N	Y	P	F	G
Q92537	SUSD6	C	H	G	R	I	A	P	K	S	T	S	V	F	A
Q9H0V1	TMEM168	C	K	S	L	R	Y	C	F	S	H	C	L	Y	L
Q13077	TRAF1	C	Y	R	A	P	C	S	E	S	Q	E	E	L	A
A6NNY8	USP27X	C	K	D	Y	V	Y	D	K	D	I	E	Q	I	A
P46695	IER3/IEX1	C	H	S	R	S	C	H	P	T	M	T	I	L	Q
Q15465	SHHC	C	F	P	G	S	A	T	V	H	L	E	Q	G	G
Q15465	SHHN	C	G	P	G	R	G	F	G	K	R	R	H	P	K
Q9NRG9	AAAS	C	S	L	G	L	F	P	P	P	P	P	R	G	Q
O95372	LYPLA2	C	G	N	T	M	S	V	P	L	L	T	D	A	A
P21589	NT5E	C	P	R	A	A	R	A	P	A	T	L	L	L	A
Q8NFK1	GJC3	C	G	R	F	L	R	R	L	L	A	E	E	S	R

Figure 2.7: Nt-Cys sequences selected for *in vitro* screening
A: List of MC-initiating sequences selected from the human proteome for screening as ADO substrates. Accession of protein, gene name and sequence of peptide used for screening are included. Sequences selected to provide variety at the position after the Nt-Cys. Gene names of known substrates are highlighted in green. Gene names also featuring in **C** are coloured accordingly. *iASPP-SV is a splice variant of iASPP protein encoded by PPP1R13L gene.
B: Amino acid composition (%) of residues highlighted yellow in **A** (position after Nt-Cys) compared to the amino acid composition (%) of all proteins in the UniProt Swiss-Prot database (release 2022_05).
C: Additional MC-initiating sequences selected from human proteome for investigation, based on similarity to known substrates (blue shading), negative results in cellular assays (orange shading), or of other interest (grey shading). Residue after Nt-Cys colour coded according to chemical properties: basic (H/K/R) = red; aromatic (F/W/Y) = orange; polar (S/T) = purple
Known substrates (RGS4/5, IL32) are shaded green.

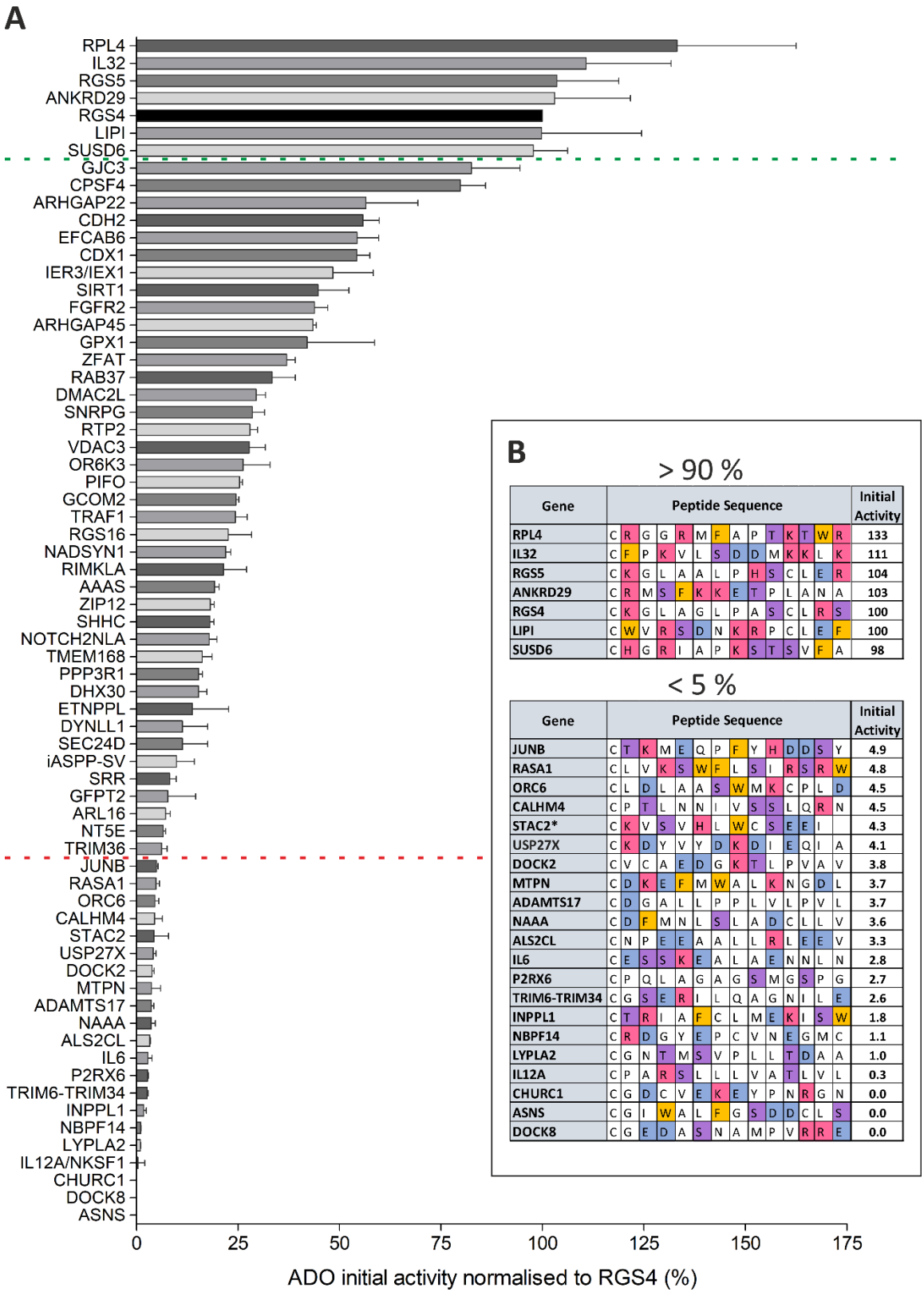


Figure 2.8: *Hs*ADO initial activity with different Nt-Cys peptides

A: *Hs*ADO-catalysed turnover of peptides representing N-termini of potential *Hs*ADO substrates (listed by gene name) after 1 minute (100 µM peptide, 0.1 µM HsADO, 37 °C); initial activity was normalised to turnover of $RGS4_{2\text{-}15}$ positive control peptide. Error bars show standard deviation, reactions were carried out in triplicate.

B: N-terminal sequences of peptides with which *Hs*ADO was observed to have > 90 % or < 5 % initial activity (µmol mg^{-1} min^{-1}) relative to $RGS4_{2\text{-}15}$ positive control peptide. Amino acids are colour coded according to chemical properties: basic (H/K/R) = red; acidic (D/E) = blue; aromatic (F/W) = orange; polar (S/T) = purple.

during synthesis one sequence, STAC2, was tested using a 13-mer peptide instead of a 14-mer. Initial activity of *Hs*ADO with these peptides was normalised to turnover of $RGS4_{2\text{-}15}$ peptide in the same assay. Four peptides could not be detected by high throughput MS at sufficient levels for analysis: $DNMBP_{2\text{-}15}$; $GPR22_{2\text{-}15}$; $NECAB2_{2\text{-}15}$; $SHHN_{2\text{-}15}$. Detection levels of these peptides remained low when analysis was attempted by ultrahigh-performance liquid chromatography (UPLC) MS (Chapter 8, page 181), so the activity of *Hs*ADO with these peptides was not assessed.

Turnover of peptides by *Hs*ADO after 1 minute varied from 0 to > 130 % relative to $RGS4_{2\text{-}15}$ (Figure 2.8). The most rapidly oxidised peptide represented the N-terminus of a computationally mapped isoform of RPL4 (Accession: H3BM89), which has an Arg at the

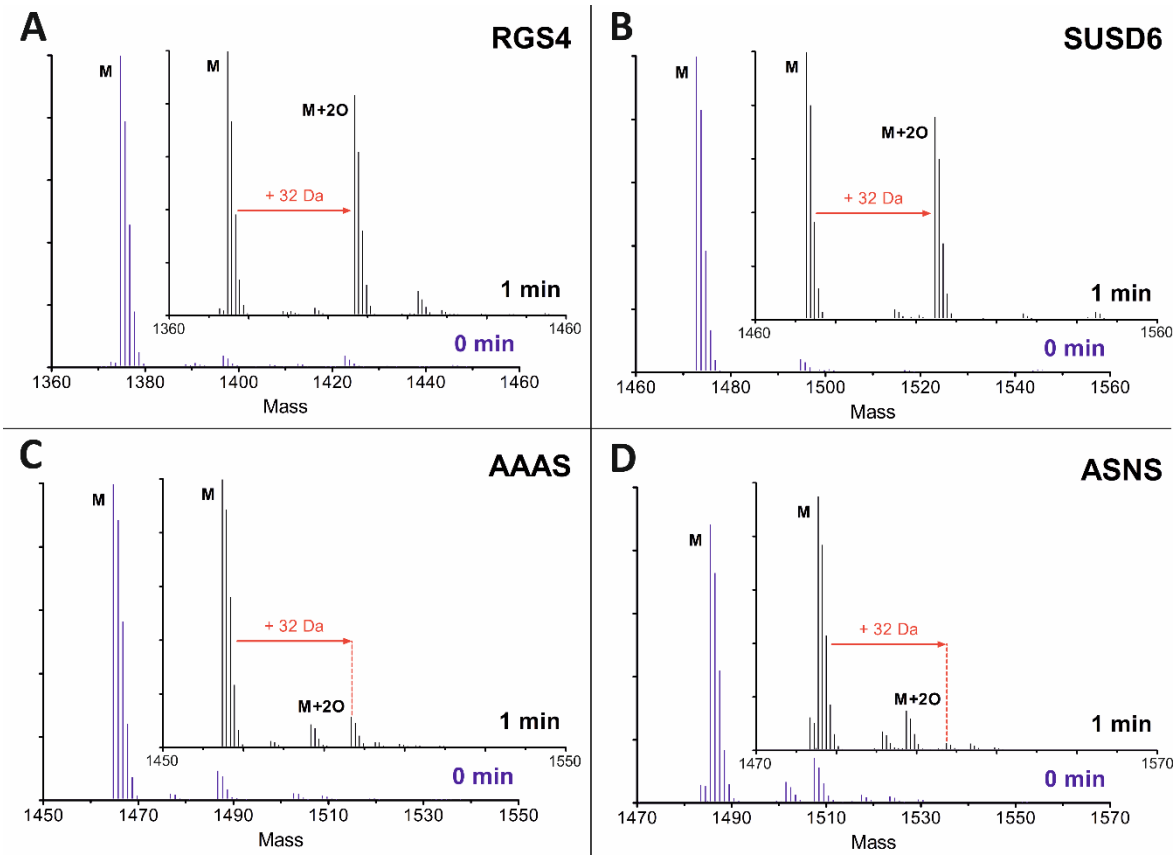


Figure 2.9: Mass spectra of representative Nt-Cys peptides after incubation with *Hs*ADO
Mass spectra of Nt-Cys peptides from substrate screening before and after incubation with *Hs*ADO at 37 °C for 1 minute (100 µM peptide, 0.1 µM ADO). Deconvoluted spectra were produced with Agilent MassHunter Qualitative Analysis software.
A: $RGS4_{2\text{-}15}$ **B:** $SUSD6_{2\text{-}15}$ **C:** $AAAS_{2\text{-}15}$ **D:** $ASNS_{2\text{-}15}$

position after the Nt-Cys. From the positive bias selection (Figure 2.7 **C**), only ANKRD29$_{2\text{-}15}$ (Figure 2.8) and SUSD6$_{2\text{-}15}$ (Figure 2.8, Figure 2.9 **B**) had > 90 % initial *Hs*ADO activity relative to RGS4$_{2\text{-}15}$. Other sequences from this selection had between 56 % and 4 % initial *Hs*ADO activity relative to RGS4$_{2\text{-}15}$; USP27X$_{2\text{-}15}$ had the lowest initial *Hs*ADO activity (4.1 % relative to RGS4$_{2\text{-}15}$) of this selection. This sequence has a Lys after the Nt-Cys, directly followed by an Asp. AAAS$_{2\text{-}15}$ and LYPLA2$_{2\text{-}15}$ peptides were not expected to be substrates based on Dr. Keeley's negative cellular assay results. While LYPLA2$_{2\text{-}15}$ had very low initial *Hs*ADO activity (0.96 % relative to RGS4$_{2\text{-}15}$ (Figure 2.8)), *Hs*ADO had moderate initial activity towards AAAS$_{2\text{-}15}$ (19.24 % relative to RGS4$_{2\text{-}15}$ (Figure 2.8, Figure 2.9 **C**)). Initial activity assay results for NT5E$_{2\text{-}15}$ (Figure 2.8) also did not correlate well with initial cellular assay results that indicated this protein might be a *Hs*ADO substrate (unpublished data, Dr. Keeley). Initial *Hs*ADO activity with NT5E$_{2\text{-}15}$ was 6.6 % relative to RGS4$_{2\text{-}15}$. Initial cellular assay results by Dr. Keeley were not repeatable, with subsequent assays suggesting that NT5E was not an endogenous ADO substrate, so this *in vitro* result was less surprising. The potential substrate identified by Dr. Masson in mice, GJC3$_{2\text{-}15}$, was a good *in vitro* substrate, with 83 % turnover after 1 minute, relative to RGS4$_{2\text{-}15}$. Some peptides were not detectably oxidised after incubation with *Hs*ADO for 1 minute: ASNS$_{2\text{-}15}$ (Figure 2.9 **D**), DOCK8$_{2\text{-}15}$, and CHURC1$_{2\text{-}15}$ (Figure 2.8).

Substrate preferences of *Hs*ADO began to emerge upon inspection of peptides towards which *Hs*ADO had particularly high (> 90 % relative to RGS4$_{2\text{-}15}$) or low (< 5 % relative to RGS4$_{2\text{-}15}$) initial activity (Figure 2.8 **B**). In this Chapter, and throughout this Thesis, amino acids are colour-coded in order to visualise physicochemical properties of side chains: basic (H/K/R) = red; acidic (D/E) = blue; aromatic (F/W) = orange; polar (S/T) = purple. A Two Sample Logo[199] was used to inspect the sequences. Logo plots were



Figure 2.10: Two Sample Logo of sequences in Figure 2.8 B
Two Sample Logo (t-test) produced using sequences of peptides towards which *Hs*ADO had > 90 % initial activity relative to $RGS4_{2-15}$ as the positive sample and sequences of peptides with low initial activity (< 5 % relative to $RGS4_{2-15}$) as the negative sample (residue shown if $p < 0.5$).

developed as a method for visualising sequence patterns in datasets[200,201]. The Two Sample Logo, developed by Vacic *et al*., provides a visual representation of the differences between two data sets: a positive and a negative (background)[199]. Here, sequences with > 90 % initial activity relative to $RGS4_{2-15}$ were used as the positive set, and sequences with < 5 % as the negative set (Figure 2.10). A two sample t-test was used to determine which residues in the sequences in the positive set were enriched or depleted compared to the negative (background) set.

All peptides towards which *Hs*ADO had high initial activity had a basic (red) or aromatic (orange) residue after the Nt-Cys. In the Two Sample Logo basic residues were enriched across the sequence, especially at position 3. Meanwhile, sequences with which *Hs*ADO had low initial activity had a prevalence of acidic residues (blue), which were also depleted in the Two Sample Logo. Residues with a polar (purple) side chain were also depleted (to a lesser degree) at positions 3, 4, and 7. Based on these data, *Hs*ADO appeared to have a strong preference for a basic or aromatic residue after the Nt-Cys, with a broader preference for basic over acidic residues within the Nt-sequence.

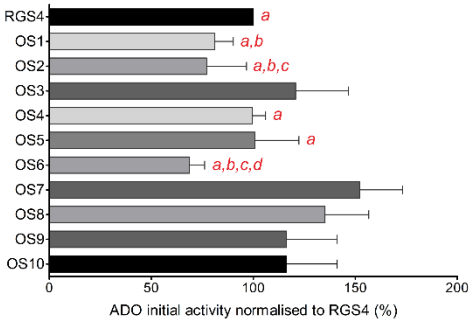
2.7.2 'Optimum' and 'anti'-substrate peptides demonstrate *Hs*ADO substrate preferences

To investigate the emerging substrate preferences of *Hs*ADO, ten 'optimum' substrate (OS) and twelve 'anti'-substrate (AS) peptides were designed and tested in *in vitro* assays (Figure 2.11). Peptides were incubated with *Hs*ADO for up to 30 minutes using the assay

conditions described above (Section 2.5). For peptides OS3-5 and AS1-12 oxidation was monitored by high throughput MS as described above (Section 2.5). Peptides OS1, OS2, and OS6-10 could not be detected at sufficient levels for analysis by high throughput MS, likely due to their high basicity causing low retention times. Oxidation of these peptides was instead monitored using an Agilent 1200 Infinity B LCMS connected to an Agilent 6530

A

Name	Peptide Sequence														Initial Activity
Residue	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	
IL32	C	F	P	K	V	L	S	D	D	M	K	K	L	K	111
RGS5	C	K	G	L	A	A	L	P	H	S	C	L	E	R	104
RGS4	C	K	G	L	A	G	L	P	A	S	C	L	R	S	100
OS1	C	R	G	R	R	K	K	A	T	T	K	A	W	R	81
OS2	C	R	G	G	R	K	K	A	T	T	K	A	W	R	77
OS3	C	R	G	G	F	K	K	A	T	T	K	A	W	R	121
OS4	C	R	G	G	F	M	K	A	T	T	K	A	W	R	100
OS5	C	R	G	G	F	M	K	A	P	T	K	A	W	R	101
OS6	C	K	G	R	R	M	K	A	T	T	K	A	W	R	69
OS7	C	F	G	R	R	M	K	A	T	T	K	A	W	R	152
OS8	C	R	M	R	R	M	K	A	T	T	K	A	W	R	135
OS9	C	R	P	R	R	M	K	A	T	T	K	A	W	R	116
OS10	C	R	G	R	R	M	K	A	T	T	K	A	W	R	113



B

Name	Peptide Sequence														Initial Activity
Residue	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	
RGS4	C	K	G	L	A	G	L	P	A	S	C	L	R	S	100.0
ASNS	C	G	I	W	A	L	F	G	S	D	D	C	L	S	0.0
AS1	C	G	E	L	A	M	K	A	T	T	K	A	W	R	2.9
AS2	C	G	S	L	A	M	K	A	T	T	K	A	W	R	31.6
AS3	C	D	E	L	A	M	K	A	T	T	K	A	W	R	2.7
AS4	C	D	S	L	A	M	K	A	T	T	K	A	W	R	3.2
AS5	C	T	E	L	A	M	K	A	T	T	K	A	W	R	13.6
AS6	C	T	S	L	A	M	K	A	T	T	K	A	W	R	30.5
AS7	C	G	E	L	A	M	P	A	S	E	E	S	L	S	0.0
AS8	C	G	S	L	A	M	P	A	S	E	E	S	L	S	0.0
AS9	C	D	E	L	A	M	P	A	S	E	E	S	L	S	0.0
AS10	C	D	S	L	A	M	P	A	S	E	E	S	L	S	0.0
AS11	C	T	E	L	A	M	P	A	S	E	E	S	L	S	0.0
AS12	C	T	S	L	A	M	P	A	S	E	E	S	L	S	2.9

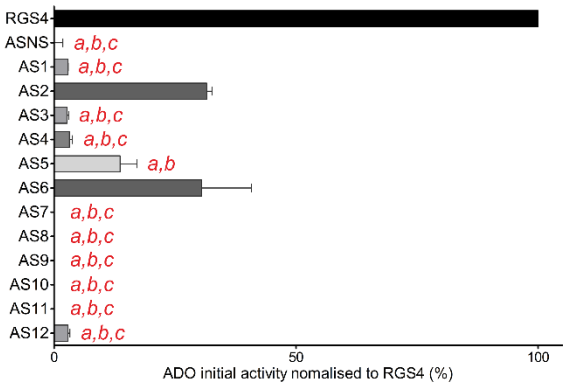


Figure 2.11: Initial activity of *Hs*ADO towards Optimum Substrate (OS) and Anti-Substrate (AS) peptides
A: Sequences of *Hs*ADO optimum substrate (OS) peptides and initial activity of *Hs*ADO with OS peptides. Significant differences ($p < 0.05$) from 1-way ANOVA results are summarised on the graph: a = lower than OS7; b = lower than OS8; c = lower than OS3; d = lower than OS9 and OS10. Full 1-way ANOVA results are available in Appendix 5.
B: Sequences of *Hs*ADO anti-substrate (AS) peptides and initial activity of *Hs*ADO with AS peptides. ASNS was included as a negative control for ADO activity. Significant differences ($p < 0.05$) from 1-way ANOVA results are summarised on the graph: a = lower than AS2; b = lower than AS6; c = lower than AS5; ASNS and AS1-12 all lower than RGS4. Full 1-way ANOVA results are available in Appendix 6.

Initial activity is turnover of peptide after 1-minute incubation with *Hs*ADO normalised to that of $RGS4_{2\text{-}15}$ positive control peptide. Error bars show standard deviation, reactions were carried out in triplicate. Amino acids are colour coded according to chemical properties: basic (H/K/R) = red; acidic (D/E) = blue; aromatic (F/W) = orange; polar (S/T) = purple.

Accurate-Mass Q-ToF mass spectrometer (Chapter 8, page 181). Spectra were analysed and the integrated area under peaks extracted using MassHunter Qualitative Analysis software (Agilent). $RGS4_{2-15}$ peptide was included in all assays as a positive control, and $ASNS_{2-15}$ was included in assays with AS peptides as a negative control.

*Hs*ADO had > 68 % initial activity relative to $RGS4_{2-15}$ towards all OS peptides (Figure 2.11 **A**). OS7 had the highest initial *Hs*ADO activity of all peptides analysed (152 % compared to $RGS4_{2-15}$). This sequence was the only OS peptide with a Phe at position 3 (after the Nt-Cys). As in IL32, this Phe residue at position 3 was followed by a basic residue at position 5. This motif of an aromatic residue at position 3 followed by a basic residue at position 5 was also seen in LIPI, a sequence towards which *Hs*ADO had the same initial activity as RGS4 (Figure 2.8 **B**). This suggested that this motif may be particularly favourable for *Hs*ADO binding.

Initial *Hs*ADO activity was significantly higher towards OS7 than OS6, whose sequence varied from that of OS7 only at position 3 (Lys instead of Phe). OS6 also had significantly lower *Hs*ADO activity than OS10, which shared the same sequence except for an Arg at position 3. These data suggested that it was less favourable for *Hs*ADO binding to have a Lys at position 3, compared to Phe or Arg. There was not significant variation in initial *Hs*ADO activity towards OS8, OS9, and OS10. These sequences varied only at position 4; these data suggested that *Hs*ADO does not have a preference between Met, Pro, or Gly at this position. Initial *Hs*ADO activity was significantly lower towards OS4 and OS5 than OS7. Comparison of these peptide sequences suggested that this difference could be due to the combined effect of having an Arg at position 3 instead of Phe and the lack of basic residues at positions 5 and 6. OS1 and OS2 had significantly lower initial *Hs*ADO activity than OS8, with which OS1 had sequence identity except for at position 7. *Hs*ADO activity was also significantly lower towards OS2 than OS3, with which OS2 had sequence identity

except for at position 6. OS1 and OS2 were also the only OS peptides to have three or more basic residues in a row, which the observations above suggested could be less favourable for *Hs*ADO binding. Overall, these data demonstrated that it was possible to design good *Hs*ADO substrate peptides using *in vitro* screening data. These data also confirmed that *Hs*ADO prefers substrates with an aromatic or basic residue after the Nt-Cys and highly basic N-termini, although data for OS1 and OS2 suggested that there is a limit to how concentrated the basic residues could be.

In contrast, all AS peptides were oxidised significantly less than $RGS4_{2\text{-}15}$ after 1 minute incubation with *Hs*ADO (Figure 2.11, Figure 2.12). All AS peptides shared sequence identity at positions 5-7 (LAM). AS1-6 then shared sequence identity at positions 8-15 (KATTKAWR); these peptides also shared sequence identity at these positions with all OS peptides other than OS5. AS7-12 shared a different sequence at positions 8-15 (PASEESLS). While AS1-6 were all oxidised to some degree after 1 minute, AS7-11 were not oxidised at all and AS12 was oxidised significantly less than AS6 (with which it shared sequence identity at positions 2-7). Turnover of AS7-11 remained low after 30 minutes incubation with *Hs*ADO (Figure 2.12). These data showed that residues later in the Nt-sequence can have a profound impact on *Hs*ADO activity. AS7-12 had an 'EE' motif at positions 11 and 12, which was reminiscent of the 'DD' motif at these positions in $ASNS_{2\text{-}15}$. The Nt-sequence of ASNS does not have any acidic or polar residues earlier in the sequence, which screening data suggested is unfavourable for *Hs*ADO. It is proposed, instead, that this 'DD' motif later in the sequence is responsible for *Hs*ADO's low activity towards the ASNS peptide. Other poor *Hs*ADO substrates identified in *in vitro* screening (Figure 2.8 **B**) also have a 'DD' or 'EE' motif after position 11: $JUNB_{2\text{-}15}$, $STAC2_{2\text{-}14}$, and $ALS2CL_{2\text{-}15}$. A potential substrate binding loop rich in acidic residues has been identified in *in silico* and crystal structures of *Hs*ADO[4–6] which may provide an explanation for this observation; this

will be discussed in detail in Chapter 6 (page 128). Interestingly, one of the known *in cellula* *Hs*ADO substrates, IL32, also has a 'DD' motif in its sequence at positions 9 and 10. This suggests that it is possible to overcome the effect of this unfavourable double acidic residue motif if other regions of the peptide bind strongly enough. Residues 3 and 5 in IL32 are

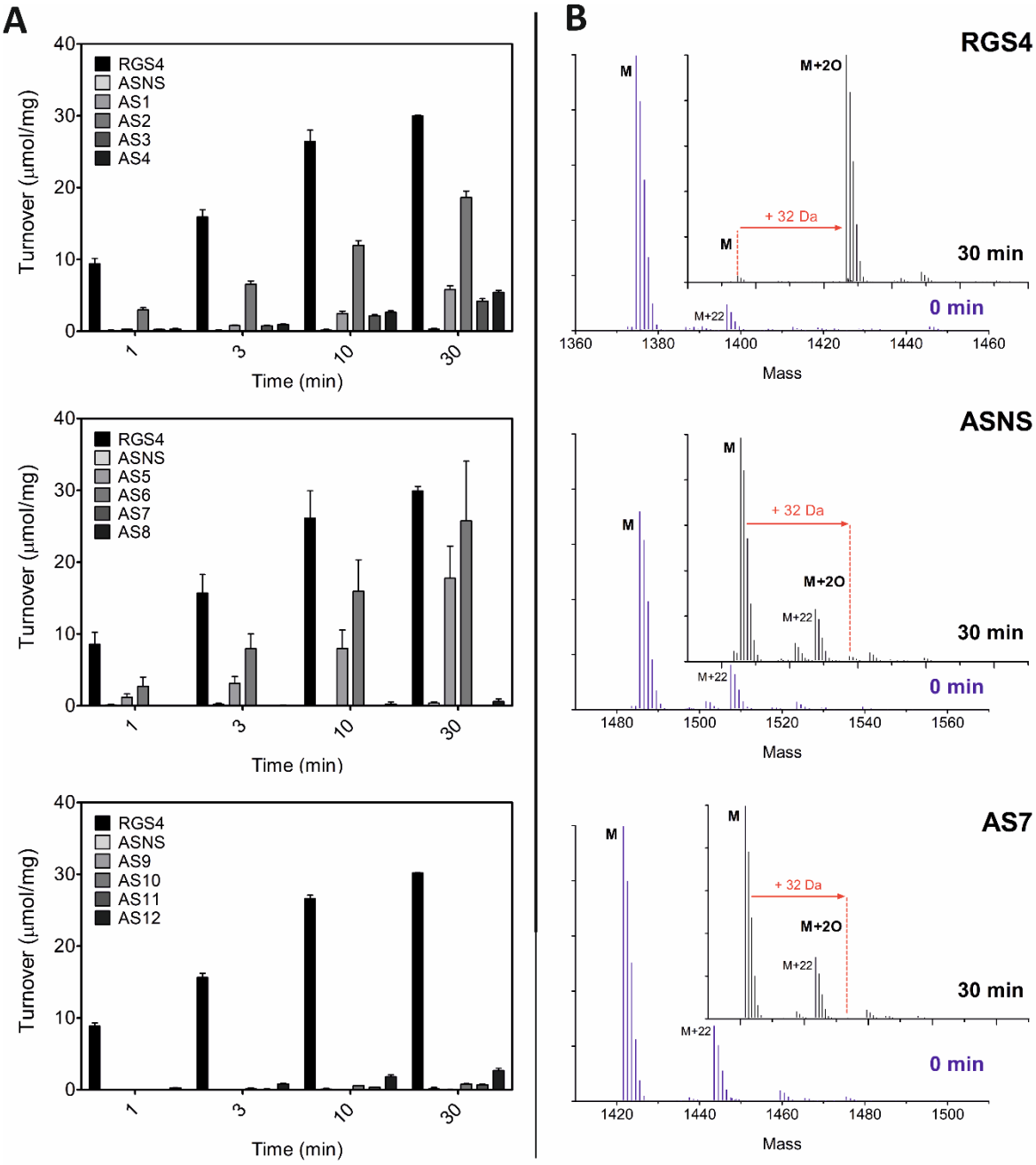


Figure 2.12: Time-courses and mass spectra of AS peptides
A: Oxidation of anti-substrate (AS) peptides in presence of ADO over 30 minutes. ASNS included as a negative control; RGS4 as a positive control. Error bars show standard deviation, reactions carried out in triplicate.
B: Deconvoluted mass spectra of $RGS4_{2\text{-}15}$, $ASNS_{2\text{-}15}$ and AS7 peptides before and after incubation with ADO. Deconvoluted spectra were produced with Agilent MassHunter Qualitative Analysis software.

thought to be particularly favourable for binding (as discussed above) and the 'DD' motif is also closely followed by a 'KK' motif (Figure 2.8 **B**), which could interact favourably with the acidic residues in the proposed substrate binding loop.

Within peptides AS1-6, *Hs*ADO had significantly higher activity towards AS2, AS5, and AS6 than towards AS1, AS3, and AS4. Initial *Hs*ADO activity was also significantly higher towards both AS2 and AS6 than AS5. AS2 and AS6 were the only AS peptides not to have any acidic residues at all, which supported the hypothesis that acidic residues are unfavourable for *Hs*ADO binding. AS5 had significantly higher initial *Hs*ADO activity than AS4. AS5 had Thr-Glu (polar-acidic) at positions 3 and 4, while AS4 had Asp-Ser (acidic-polar). This suggested that an acidic residue at position 3 may be more unfavourable than at position 4. Overall, activity data with AS peptides confirmed that acidic residues are unfavourable for *Hs*ADO binding, both early (positions 3 and 4) and later (positions 11 and 12) in the Nt-sequence.

2.8 Characterisation of *Hs*ADO activity with PCO substrates

ADO was first identified as a potential Nt-Cys oxidase (NCO) based on its similarity to plant cysteine oxidase (PCO) enzymes, which catalyse oxidation of the Nt-Cys of group VII ethylene response factors (ERF-VIIs)[1,3,58]. Masson *et al.* demonstrated that the stability of RAP2.12, a PCO substrate in *Arabidopsis thaliana,* was dependent on iron and O_2 availability in human cells when incorporated into a GFP reporter protein, in a manner that was dependent on the presence of a Cys at the N-terminus[1]. Human ADO was also able to compensate for loss of PCOs *in planta*. These observations suggested that it would be interesting to characterise RAP2.12, and other PCO substrates, as ADO substrates *in vitro*.

2.8.1 *Hs*ADO can catalyse oxidation of PCO substrate peptides

PCO substrates selected for investigation included four ERF-VIIs: RAP2.12, RAP2.3, HRE1, and HRE2[51,62]. In 2019, protein LITTLE ZIPPER2 (ZPR2) was also identified as a substrate of *At*PCOs[69], so was included in *in vitro* investigations in this Chapter. Sequences of four additional potential PCO substrates (MJH1/2/5/6) were kindly provided by Prof. Michael Holdsworth (University of Nottingham).

Table 2.2: PCO substrate peptides

Peptide Name	Sequence
RAP2.12$_{2\text{-}15}$	CGGAIISDFIPPPR-NH_2
RAP2.3$_{2\text{-}15}$	CGGAIISDYAPLVT-NH_2
HRE1$_{2\text{-}15}$	CGGAVISDYIAPEK-NH_2
HRE2$_{2\text{-}15}$	CGGAIISDFIWSKS-NH_2
ZPR2$_{2\text{-}15}$	CLTTSEPPFPDTDT-NH_2
MJH1	CKKFKIFMRKLQIC-NH_2
MJH2	CRSPDFERYVEKES-NH_2
MJH5	CSSKLKNLTQENIS-NH_2
MJH6	CHVGYALIELLENV-NH_2

Peptides representing the Met-excised N-terminus of the nine different PCO substrates (Table 2.2) were incubated with recombinant *Hs*ADO for 1 minute at 37 °C using the assay conditions described above (Section 2.5), and oxidation was monitored by high throughput MS as described above (Section 2.5). The sequences of the PCO substrate peptides are presented in Table 2.2, with acidic residues in blue text and basic residues in red text. *Hs*ADO initial activity was significantly lower towards PCO substrate peptides than the human substrate RGS4$_{2\text{-}15}$ (Figure 2.13). Initial activity was particularly low towards ZPR2$_{2\text{-}15}$ (0.57 μmol mg^{-1} min^{-1}), RAP2.3$_{2\text{-}15}$ (0.58 μmol mg^{-1} min^{-1}), and HRE1 (1.22 μmol mg^{-1} min^{-1}). Of the PCO substrate peptides tested, ZPR2$_{2\text{-}15}$ had both the

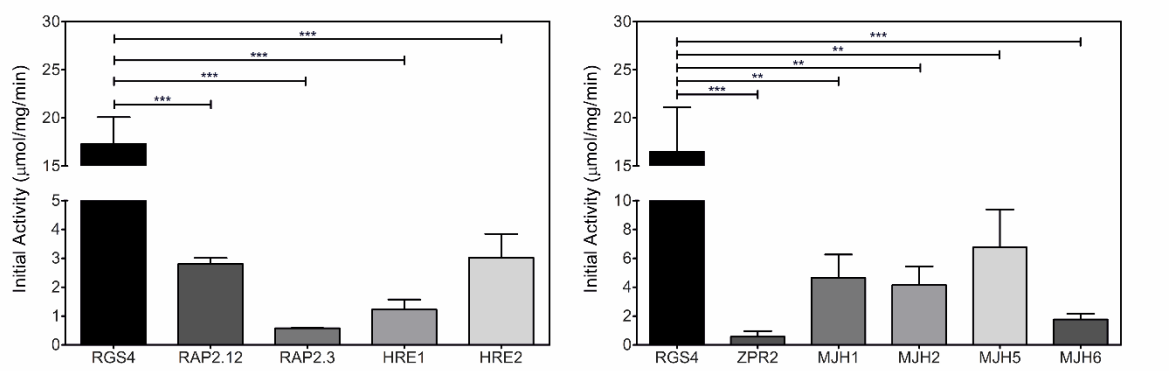


Figure 2.13: Initial activity of ADO with PCO substrate peptides.
Initial activity is turnover of peptide after 1-minute incubation with *Hs*ADO. Error bars show standard deviation, reactions were carried out in triplicate. Statistical analysis carried out using GraphPad Prism v5 1-way ANOVA with Tukey's multiple comparison post-test. Only significant *p* value summaries (< 0.05) are shown: * ($p \leq 0.05$); ** ($p \leq 0.01$); *** ($p \leq 0.001$):

lowest initial activity and the most acidic residues in its sequence (three). *Hs*ADO activity was generally higher towards the MJH peptides (> 4.14 μmol/mg/min), which were also the peptides with the most basic residues. The exception was MJH6 (1.76 μmol/mg/min); this peptide had only one basic residue (His3) and also had two acidic Glu residues (at positions 10 and 13). These data aligned with previous observations made regarding *Hs*ADO's *in vitro* substrate preferences (Section 2.7).

2.8.2 Steady-state kinetic analysis of *At*RAP2.12$_{2\text{-}15}$ with *Hs*ADO

To further investigate ADO's apparent low affinity for PCO substrates, steady-state kinetic analysis was carried out with *Hs*ADO and *At*RAP2.12$_{2\text{-}15}$ peptide. Michaelis-Menten kinetics[202,203] can be used to provide information on the substrate specificity of an enzyme. In steady-state model kinetics, the enzyme reaction is defined as:

$$E + S \underset{k_r}{\overset{k_f}{\rightleftharpoons}} ES \xrightarrow{k_{cat}} E + P$$

(Equation 2.1)

The model makes the following assumptions:

1 – No product reverts to substrate.

2 – The concentration of substrate is much higher than that of the product ([S] >> [P]) and that of the enzyme ([S] >> [E]), i.e., steady state conditions.

3 – The concentration of enzyme-substrate complex ([ES]) is constant.

Under these assumptions, the Michaelis-Menten equation can be derived, where reaction velocity (v) is defined in terms of maximum velocity (V_{max}) and the Michaelis-Menten constant (K_M):

$$v = \frac{V_{max}[S]}{K_M + [S]}$$

(Equation 2.2)

where

$$K_M = \frac{k_r + k_{cat}}{k_f}$$

(Equation 2.3)

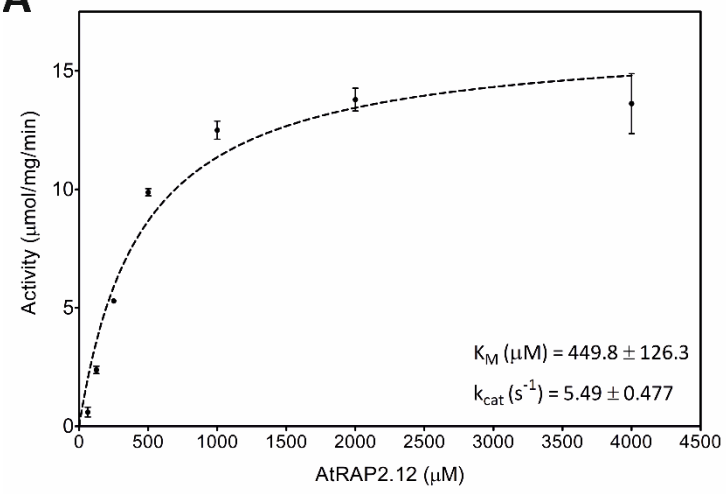
The Michaelis-Menten constant (K_M), is the substrate concentration at which velocity is half V_{max}, and is indicative of enzyme affinity for a particular substrate. The catalytic constant (k_{cat}), also referred to as turnover number, is a measure of how active an enzyme is once substrate is bound.

The K_M and k_{cat} were determined for *At*RAP2.12$_{2\text{-}15}$ with *Hs*ADO (Figure 2.14 **A**) (Chapter 8, page 174). *At*RAP2.12$_{2\text{-}15}$ peptide, at concentrations between 15.625 and 4000.0 µM, was incubated with *Hs*ADO (0.05 µM) for 1 minute and 20 seconds using the assay conditions described above (Section 2.5). The reaction was sampled every 20 seconds and analysed by high throughput MS as described previously (Section 2.5). These conditions were selected to ensure steady state conditions. Peptide turnover (µmol mg^{-1}) was calculated at each time point and used to calculate reaction velocity as specific activity (µmol mg^{-1} min^{-1}). The iron(II) content of the recombinant enzyme used was quantified

(Chapter 8, page 168) and found to be 73 % of that predicted by 1:1 stoichiometry of metal and enzyme; this was taken as the percentage of active enzyme so peptide turnover was adjusted by a factor of 0.73. Specific activity was plotted against substrate concentration, and K_M and k_{cat} were determined using GraphPad Prism.

The K_M for *At*RAP2.12$_{2\text{-}15}$ with *Hs*ADO was determined to be approximately 450 µM (Figure 2.14 **A**), which was higher than the K_M values for the known human protein substrates of *Hs*ADO[1] (Figure 2.14 **B**). RGS4, RGS5, and IL32 14-mer peptides all had K_M values between 33 and 123 µM. However, the K_M value for *At*RAP2.12 was still much lower than the K_M value for cysteamine with *Hs*ADO (3.8 mM)[166]. This suggested that although *Hs*ADO has higher affinity for human peptide substrates than plant peptide substrates, overall *Hs*ADO has a higher affinity for peptide over small molecule substrates. Interestingly the K_M value for *At*RAP2.12$_{2\text{-}15}$ with *Hs*ADO was similar to the K_M value for *At*RAP2.12$_{2\text{-}15}$ with *At*PCO3 (560 µM)[62]. The K_M values for *At*RAP2.12$_{2\text{-}15}$ with *At*PCO1-5 varied between 90 and 690 µM[62]; *Hs*ADO therefore has comparable affinity with *At*RAP2.12 as the *At*PCOs.

A



B

	RGS4	RGS5	IL32
K_m (uM)	123 ± 20.76	71.51 ± 9.55	33.6 ± 3.42
K_{cat} (s^{-1})	20.13 ± 1.93	16.92 ± 1.1	10.3 ± 0.39
K_{cat}/K_m (M^{-1} s^{-1})	1.64E+05	2.37E+05	3.07E+05
***At* RAP2.12**		**Cysteamine**	
K_m (uM)	449.8 ± 126.3	K_m (uM)	3800
K_{cat} (s^{-1})	5.49 ± 0.48	K_{cat} (s^{-1})	1.6
K_{cat}/K_m (M^{-1} s^{-1})	1.22E+04	K_{cat}/K_m (M^{-1} s^{-1})	4.21E+02

Figure 2.14: Steady-state kinetic analysis of *At*RAP2.12 with *Hs*ADO
A: Kinetic analysis of *At*RAP2.12$_{2\text{-}15}$ peptide with *Hs*ADO. K_M and k_{cat} were determined using GraphPad Prism Michaelis-Menten and k_{cat} non-linear fit models.
B: Table comparing kinetic data for different substrates with *Hs*ADO, showing Michaelis-Menten constant (K_m), turnover number (k_{cat}), and catalytic efficiency (k_{cat}/K_m). Blue: literature values for RGS4$_{2\text{-}15}$, RGS5$_{2\text{-}15}$, and IL32$_{2\text{-}15}$ peptides; green: values for *At*RAP2.12$_{2\text{-}15}$ (A); yellow: literature values for cysteamine.

Catalytic efficiency can be measured by calculating $\frac{k_{cat}}{K_M}$. These values for peptide and small molecule substrates further demonstrate *Hs*ADO's substrate preferences (Figure 2.14 **B**). Catalytic efficiency for *At*RAP2.12 was an order of magnitude lower than for RGS4, RGS5, and IL32; catalytic efficiency for cysteamine was then two orders of magnitude lower than for *At*RAP2.12. This is evidence that the primary role of *Hs*ADO might be Nt-Cys oxidation, rather than small molecule thiol dioxygenation.

2.9 Discussion

In this Chapter, recombinant *Hs*ADO was produced (Section 2.3) and shown to oxidise both cysteamine (Section 2.4) and an RGS4 peptide (Section 2.5). This Chapter then described the use of RGS4 peptides to further investigate the NCO activity of *Hs*ADO (Section 2.6). These experiments demonstrated that methionine excision was necessary for *Hs*ADO to catalyse cysteine oxidation, and that *Hs*ADO could only catalyse oxidation of Nt-Cys, not internal Cys residues. As few as seven amino acids of the Met-excised N-terminus of RGS4 were shown to be sufficient for *Hs*ADO binding and activity. Computational studies by Fernandez *at al*. suggest that ADO could bind shorter peptides, but it was not possible to test this under the conditions used for this experiment.

As only three protein substrates of *Hs*ADO had been characterised before the start of this project, one of the key aims of this Chapter was to determine if other human Nt-Cys proteins could be substrates of *Hs*ADO. This Chapter described *in vitro* screening of over 70 different peptides representing the Met-excised N-terminus of Met-Cys-initiating sequences from the human proteome (Section 2.7). *Hs*ADO was found to be highly selective amongst Nt-Cys peptides. Due to the varied nature of the amino acid after the Nt-Cys in the sequences screened, these data provided information of the substrate preferences of *Hs*ADO. Sequences with a basic or aromatic residue after the Nt-Cys made the best *Hs*ADO

substrates, while sequences rich in acidic residues were poor substrates. These preferences were confirmed using 'optimum' substrate (OS) and 'anti'-substrate (AS) peptides (Section 2.7). These experiments highlighted the effect that residues later in the Nt-sequence can have on *Hs*ADO binding and activity, with a double acidic residue motif after residue 11 being particularly unfavourable. Data for OS and AS peptides further demonstrated *Hs*ADO's preference for basic over acidic residues at the N-terminus, and for a basic or aromatic residue after the Nt-Cys. The link between these observations and the structure of ADO will be discussed in detail in Chapter 6 (page 136). *Hs*ADO's substrate preferences is further investigated in Chapter 3 (page 59) using *in cellula* experiments with human cells.

Finally, this Chapter described the characterisation of substrates of the related Plant Cysteine Oxidase (PCO) enzymes as substrates of *Hs*ADO (Section 2.8). *Hs*ADO was previously shown to compensate for PCO activity *in planta*[1], so it was not surprising that *Hs*ADO was able to catalyse oxidation of the N-terminus of nine different PCO substrate peptides. However, the turnover of all these peptides was significantly lower than the turnover of the human substrate $RGS4_{2\text{-}15}$ peptide. Within these results, *Hs*ADO activity was higher towards peptides that had more basic residues within their sequence, and lower towards those with a higher prevalence of acidic residues, as expected from *in vitro* data with human substrate peptides (Section 2.7). Steady state kinetic analysis of *At*RAP2.12 with *Hs*ADO showed that *Hs*ADO has lower affinity with *At*RAP2.12 peptide than with RGS4, RGS5, and IL32 peptides (K_M 450 μM versus 33 – 123 μM). Interestingly, *Hs*ADO affinity with *At*RAP2.12 was comparable to the affinity of the *At*PCO1-5 enzymes with *At*RAP2.12[62] (K_M 90 – 690 μM). Catalytic efficiency and affinity were highest for *Hs*ADO with RGS4, RGS5, and IL32, than with *At*RAP2.12, and were lowest for cysteamine. These data provide supporting evidence that Nt-Cys oxidation could be the primary role of *Hs*ADO, rather than oxidation of cysteamine to hypotaurine.

3 Investigation of *Hs*ADO Substrates in Human Cells

3.1 Introduction

In Chapter 2 (page 41), *in vitro* experiments demonstrated that *Hs*ADO is highly substrate specific. *Hs*ADO was shown to preferentially catalyse oxidation of peptides with a basic or aromatic residue after the N-terminal cysteine (Nt-Cys), and, more generally, to preferentially catalyse oxidation of 14-mer Nt-Cys-bearing peptides rich in basic residues over those rich in acidic residues. However, prior to the start of this project, only three protein substrates of *Hs*ADO had been characterised *in cellula*: RGS4, RGS5, and IL32[1]. In order to identify novel protein substrates of *Hs*ADO, and therefore determine the scope of its role in O_2-dependent protein degradation, it is important to establish whether *in vitro* substrate preferences also apply *in cellula*.

3.2 Chapter Aims

In this Chapter, a dual-fluorescence oxygen reporter (DFOR) cellular assay, developed by Smith and Keeley[204], was used to investigate the regulation of N-terminal (Nt) sequences by *Hs*ADO *in cellula.* These Nt-sequences were selected using results from *in vitro* peptide activity assays (Chapter 2, page 41). In the DFOR assay (described in detail below), protein levels could be monitored by fluorescence, and Nt-sequences investigated as *Hs*ADO substrates *in cellula* without the need for extraction of protein from cells or immunoblotting.

This Chapter will also discuss results obtained by Dr. Thomas Keeley (Nuffield Department of Medicine) investigating the regulation of full-length putative protein substrates by *Hs*ADO in human cells.

3.3 The dual-fluorescence oxygen reporter (DFOR) assay

In the dual-fluorescence oxygen reporter (DFOR) assay (Figure 3.1)[204], constructs were designed such that the N-terminus of a protein of interest (MCX_n) was fused to the N-terminus of enhanced green fluorescent protein[205] (EGFP). The initiator methionine (M) of these sequences is expected to be removed co-translationally by methionine aminopeptidases (MAPs)[79], to produce CX_n-EGFP proteins. CX_n sequences were selected using results from *in vitro* peptide activity assays (Chapter 2, page 41), and are given in Table 3.1 (Section 3.4). If the CX_n sequence added was a target of *Hs*ADO *in cellula* then its location at the N-terminus of EGFP was hypothesised to render the resulting fusion protein an *Hs*ADO substrate, that might manifest degradation by the Arg/N-degron

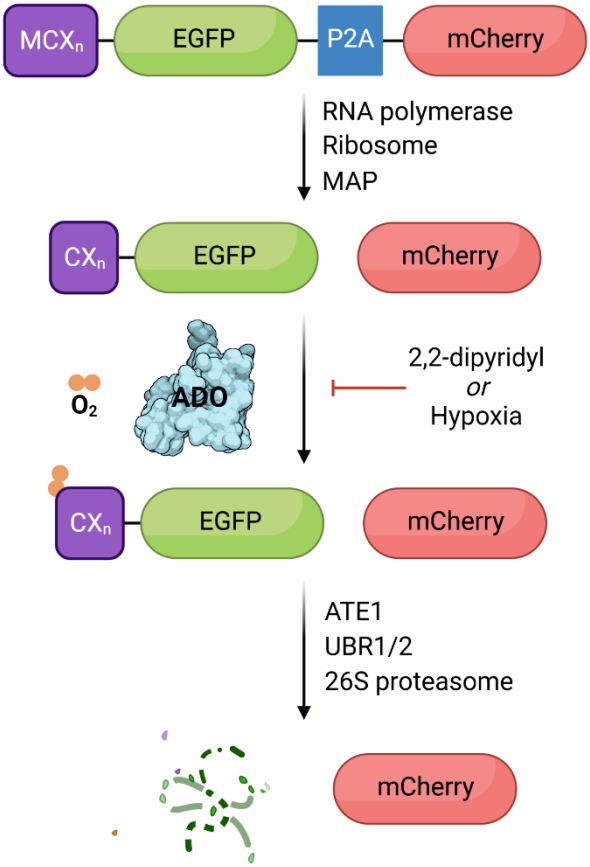


Figure 3.1: Dual-Fluorescence Oxygen Reporter (DFOR) assay
MCX_n-EGFP-mCherry plasmids are transfected into cells. EGFP and mCherry DNA sequences are separated by a self-cleaving P2A sequence; CXn-EGFP and mCherry protein levels should therefore be the same following protein translation and Met-excision by methionine aminopeptidases (MAP). If the CX_n N-terminal sequence is an ADO substrate the Nt-Cys will be oxidised and the CX_n-EGFP protein should be degraded via the Arg/N-degron pathway; EGFP fluorescence will be reduced relative to mCherry fluorescence so CX_n-EGFP levels can be monitored by the ratio of EGFP/mCherry fluorescence. Created using Biorender.com

pathway[1,2]. This would result in a decrease in EGFP fluorescence in the presence of active ADO. The constructs used for DFOR assays also included DNA encoding for mCherry (a red fluorescent protein), which was separated from the DNA encoding for EGFP by a self-cleaving P2A (porcine teschovirus-1 2A) sequence (Figure 3.1). '2A' sequences are viral peptides that induce co-translational 'cleavage' by causing a bond-skipping event in the 2A peptide sequence[206–208]. Therefore, cells transfected with the DFOR plasmids (MCX_n-EGFP:P2A:mCherry) should produce CX_n-EGFP and mCherry proteins at the same level. The mCherry protein was used as a reference protein: any factors other than N-terminal targeting affecting CX_n-EGFP levels, e.g., alterations in transcription or translation, would also affect mCherry levels.[204]

Fluorescence measurements were then used to monitor the levels of both EGFP and mCherry proteins in cells. It had previously been established that hypoxia and 2,2-dipyridyl (an iron chelator) could inhibit *Hs*ADO activity *in cellula*[1]. It was therefore predicted that changes in *Hs*ADO-dependent degradation of CX_n-EGFP could be induced by treating cells with 2,2-dipyridyl or by subjecting cells to hypoxia (1 % O_2)[204]. These changes could be measured by comparing the EGFP/mCherry ratio in cells treated with 2,2-dipyridyl with cells treated with a vehicle control (DMSO), or the ratio in cells incubated in hypoxia versus those incubated in normoxia (~20 % O_2). An increase in the EGFP/mCherry ratio for MCX_n-EGFP:P2A:mCherry expressing cells treated with 2,2-dipyridyl versus DMSO, or hypoxia versus normoxia, should indicate that the CX_n Nt-sequence was an *Hs*ADO substrate in the Arg/N-degron pathway *in cellula*.

The DFOR system provided a more rapid procedure than the alternative of harvesting protein from cells and measuring protein levels by immunoblotting. Fluorescence levels could be measured in minutes using a plate reader (such as a FLUOstar® Omega plate reader (BMG LABTECH)), and multiple Nt-sequences could be analysed simultaneously.

Immunoblotting, however, requires specific antibodies for each protein of interest and multiple incubations of an hour or longer. One advantage of protein extraction and immunoblotting is that levels of native protein can be assessed in cells. However, not all candidate ADO substrates are expressed in all cells, so comparisons of ADO's targeting efficiency would be complicated by the use of different cellular settings. Another favourable feature of the DFOR system is the fact that the same type of cells could be used to compare all sequences, so the cellular environment should be broadly the same, other than the presence of specific DFOR constructs.

3.4 Preparation of materials for DFOR cellular assay

3.4.1 Selection of Nt-sequences from peptide activity assays

Although the DFOR assay provided a more rapid method of *in cellula* analysis than protein extraction and immunoblotting, due to the cloning and cell culture requirements involved in establishing the DFOR assay (Section 3.4.2-3) it was not feasible to test all of the sequences that were screened against *Hs*ADO *in vitro* (Chapter 2, page 41). Therefore, representative candidates were selected from peptide activity assay results (Table 3.1). RGS4 was selected for use as a positive control sequence, while ASNS was selected as a negative control. Two Nt-sequences were selected whose representative peptides had high initial *Hs*ADO activity: ANKRD29 and SUSD6. A further seven Nt-sequences were selected, whose representative peptides had between 42.05 % and 0.96 % initial *Hs*ADO activity *in vitro* relative to $RGS4_{2\text{-}15}$ peptide. All sequences selected were from proteins predicted to be expressed in cells based on RNA expression databases[209,210], and were therefore likely to be biologically relevant. STAC2 was an exception to this, as the MC-initiating variant used in this Thesis is a computationally mapped isoform predicted from alternative splicing by UniProt[193].

Name	Sequence of Peptide Tested *in vitro*														Initial *HsADO* Activity *in vitro* (% RGS4)	MCX_n
OS7	C	F	G	R	R	M	K	A	T	T	K	A	W	R	152.2	MCFGRRMKATTKAWR
ANKRD29	C	R	M	S	F	K	K	E	T	P	L	A	N	A	103.0	MCRMSFKKETPLANA
RGS4	C	K	G	L	A	G	L	P	A	S	C	L	R	S	100.0	MCKGLAGLPASCLRS
SUSD6	C	H	G	R	I	A	P	K	S	T	S	V	F	A	97.7	MCHGRIAPKSTSVFA
GJC3	C	G	R	F	L	R	R	L	L	A	E	E	S	R	82.5	MCGRFLRRLLAEESR
GPX1	C	A	A	R	L	A	A	A	A	A	A	A	Q	S	42.0	MCAARLAAAAAAAQS
AAAS	C	S	L	G	L	F	P	P	P	P	P	R	G	Q	19.2	MCSLGLFPPPPPRGQ
***At* RAP2.12**	C	G	G	A	I	I	S	D	F	I	P	P	P	R	16.4	MCGGAIISDFIPPPR
GFPT2	C	G	I	F	A	Y	M	N	Y	R	V	P	R	T	7.8	MCGIFAYMNYRVPRT
***Mm* Tex19.1**	C	P	P	V	S	V	R	H	G	A	R	G	M	S	7.3	MCPPVSVRHGARGMS
JUNB	C	T	K	M	E	Q	P	F	Y	H	D	D	S	Y	4.9	MCTKMEQPFYHDDSY
STAC2	C	K	V	S	V	H	L	W	C	S	E	E	I		4.3	MCKVSVHLWCSEEI
MTPN	C	D	K	E	F	M	W	A	L	K	N	G	D	L	3.7	MCDKEFMWALKNGDL
LYPLA2	C	G	N	T	M	S	V	P	L	L	T	D	A	A	1.0	MCGNTMSVPLLTDAA
ASNS	C	G	I	W	A	L	F	G	S	D	D	C	L	S	0.0	MCGIWALFGSDDCLS

Table 3.1: Sequences selected for investigation in the DFOR assay
Name: Names of sequences selected for investigation in DFOR assay: referring to human (*Homo sapiens*) proteins with exception of: RAP2.12 (*Arabidopsis thaliana*); Tex19.1 (*Mus musculus* (mouse)); OS7 (optimum substrate peptide tested in Chapter 2 (page 47)).
Sequences of Peptides Tested *in vitro*: Amino acid sequences of corresponding peptides tested in *in vitro* assays in Chapter 2 (page 41). Amino acids are colour coded according to their chemical properties: basic (H/K/R) = red; acidic (D/E) = blue; aromatic (F/W) = orange; polar (S/T) = purple.
Initial *Hs*ADO Activity *in vitro* (% RGS4): Initial *Hs*ADO activity data obtained for peptides, normalised to initial activity of *Hs*ADO with RGS4 peptide (Chapter 2, page 41).
MCX_n: Amino acid sequence for N-terminus of protein (or OS7 peptide) appended to N-terminus of EGFP for DFOR assay (MCX_n).

In addition to these human sequences, a substrate of the related Plant Cysteine Oxidase (PCO) enzymes was chosen for investigation: *At*RAP2.12[3,58]. *Hs*ADO has previously been shown to compensate for loss of PCO activity *in planta*[1], and *At*RAP2.12$_{2\text{-}15}$ has been characterised as a moderately good substrate of *Hs*ADO *in vitro* (Chapter 2, page 52). The stability of an *At*RAP2.12$_{1\text{-}50}$-GFP reporter was also regulated by iron and oxygen in human cells, in a manner that was dependent on the presence of Cys2, and not on the presence of an active HIF pathway[1]. The N-terminus of *At*RAP2.12 was therefore expected to behave as an *Hs*ADO substrate in DFOR assays. In Chapter 2 (page 47), ten optimum substrate (OS) sequences were designed and tested to validate *Hs*ADO's *in vitro* substrate preferences. OS7, the most rapidly oxidised of these peptides was also selected for the

DFOR assay to investigate whether an artificial Nt-sequence could be an *Hs*ADO substrate *in cellula*. Finally, a mouse protein, Tex19.1, was also included. *Mm*Tex19.1 has been shown to interact with and be stabilised by E3 ligase UBR2; in Ubr2-deficient cells Tex19.1 is depleted[211], possibly via the Arg/N-degron pathway. *Mm*Tex19.1 may therefore be an ADO substrate; although *Hs*ADO had reasonably low initial activity with the *Mm*Tex19.1$_{2\text{-}15}$ peptide *in vitro* (Table 3.1).

This selection allowed the relationship between *in vitro* and *in cellula* activity to be investigated. It was predicted that the N-termini of proteins corresponding to peptides towards which *Hs*ADO had high *in vitro* activity (e.g., ANKRD29, SUSD6) would confer *Hs*ADO-dependent stability on the EGFP fusion protein, while those with lower activity *in vitro* (e.g., ASNS, LYPLA2) would not. The N-termini of the proteins chosen for DFOR assays also provided variety in amino acid sequence, enabling *Hs*ADO's substrate preferences to be probed in human cells.

3.4.2 Plasmid construction

MCX$_n$-EGFP:P2A:mCherry plasmids were generated for selected sequences (Table 3.1) by PCR and restriction enzyme digestion and ligation (Chapter 8, page 184). For all proteins other than STAC2, MCX$_n$ corresponded to the first 15 amino acids of the nascent N-terminus; as a 13-mer peptide of STAC2 was used in *in vitro* activity assays, the first 14 amino acids were used. For OS7, an initiator methionine was added to the peptide sequence. The initiator methionine in all these sequences was predicted to be removed co-translationally to reveal the Nt-Cys when the proteins were expressed in human cells[79].

PCR reactions were carried out using EGFP template DNA, reverse primer (R_GFP_XbaI (Appendix 7)), and the relevant forward primer (e.g., F_HindIII_RGS4[15] (Appendix 7)). PCR products were purified using QIAquick® PCR Purification Kit (Qiagen), and

subsequently digested using FastDigest *HindIII* and *XbaI* restriction enzymes (Thermo Scientific™) to produce MCX_n-EGFP DNA inserts. A linearised vector was produced by digestion with the same restriction enzymes and dephosphorylation with calf intestinal phosphatase of pcDNA3.1-GFP:P2A:mCherry template plasmid (kindly provided by Dr. Keeley (Appendix 8))[204]. Digested DNA inserts and linearised vector were separated from reaction mixtures by electrophoresis on a 1 % agarose gel and extracted using the QIAquick® Gel Extraction Kit (Qiagen), then inserts and vector were ligated using T4 DNA Ligase (Thermo Scientific™) according to the manufacturer's protocol. Ligation products were transformed into XL10-Gold® Ultracompetent Cells (Agilent) according to the manufacturer's protocol. Following incubation at 37 °C overnight, up to three colonies from each transformation were selected and grown individually in 5 mL LB media (supplemented with ampicillin (100 μg mL^{-1})) at 37 °C overnight. MCX_n-EGFP:P2A:mCherry plasmid DNA was purified using the QIAprep® Spin Miniprep Kit (Qiagen); sequences were confirmed by Sanger Sequencing (Source BioScience).

3.4.3 Generation of SH-SY5Y cells stably expressing DFOR constructs

In order to generate pools of cells stably expressing MCX_n-EGFP:P2A:mCherry, DFOR plasmids produced above (Section 3.4.2) were transfected into SH-SY5Y cells using GeneJuice® (Novagen) or Lipofectamine™ 3000 (Invitrogen™) according to the manufacturer's protocol. SH-SY5Y is a human-derived neuroblastoma cell line that has previously been used to study *Hs*ADO activity *in cellula*[1]. The transfected SH-SY5Y cells were cultured in DMEM-F12 growth medium (Gibco™), supplemented with fetal bovine serum (10 %), L-Glutamine (2 mM), penicillin (100 U mL^{-1}), and streptomycin (10 μg mL^{-1}); cells were maintained at 37 °C under an atmosphere of 5 % CO_2 in air. Following transfection, cells were treated with G418 antibiotic (0.5 mg mL^{-1}, Santa Cruz Biotechnology) for 2-4 weeks to generate a polyclonal population, i.e., a population of cells

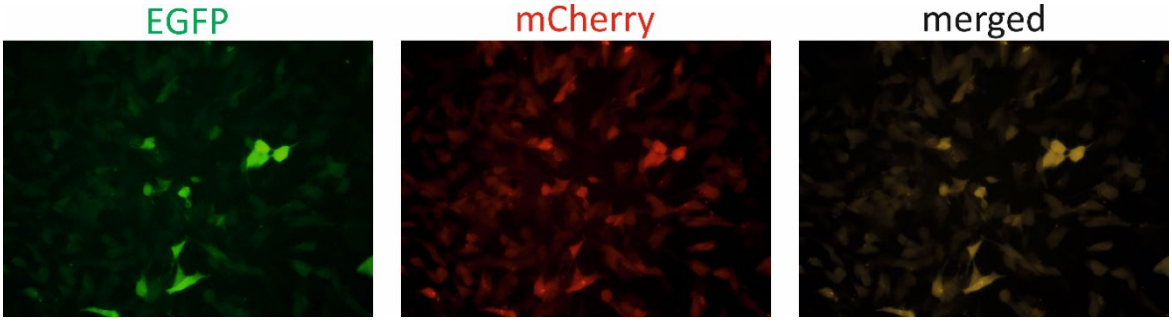


Figure 3.2 SH-SY5Y cells expressing RGS4$_{1\text{-}15}$-EGFP-mCherry
SH-SY5Y cells stably expressing the RGS4$_{1\text{-}15}$-EGFP-mCherry plasmid. Cells were imaged using an Olympus CKX41 microscope and a CoolLED pE300-W light source. EGFP was imaged under blue light ($488_{ex}/520_{em}$), and mCherry was imaged under green light ($550_{ex}/590_{em}$). Images were taken in black and white using QIClick™ CoolLED digital camera and colour filters were added with ImageJ.

that may express the DFOR construct at different levels (Chapter 8, page 187). Cells expressing DFOR plasmids successfully were verified by mCherry fluorescence (Figure 3.2).

3.5 *Hs*ADO is selective amongst *Nt*-Cys sequences in DFOR assay

3.5.1 Regulation of *Hs*ADO activity in DFOR assay with 2,2-dipyridyl

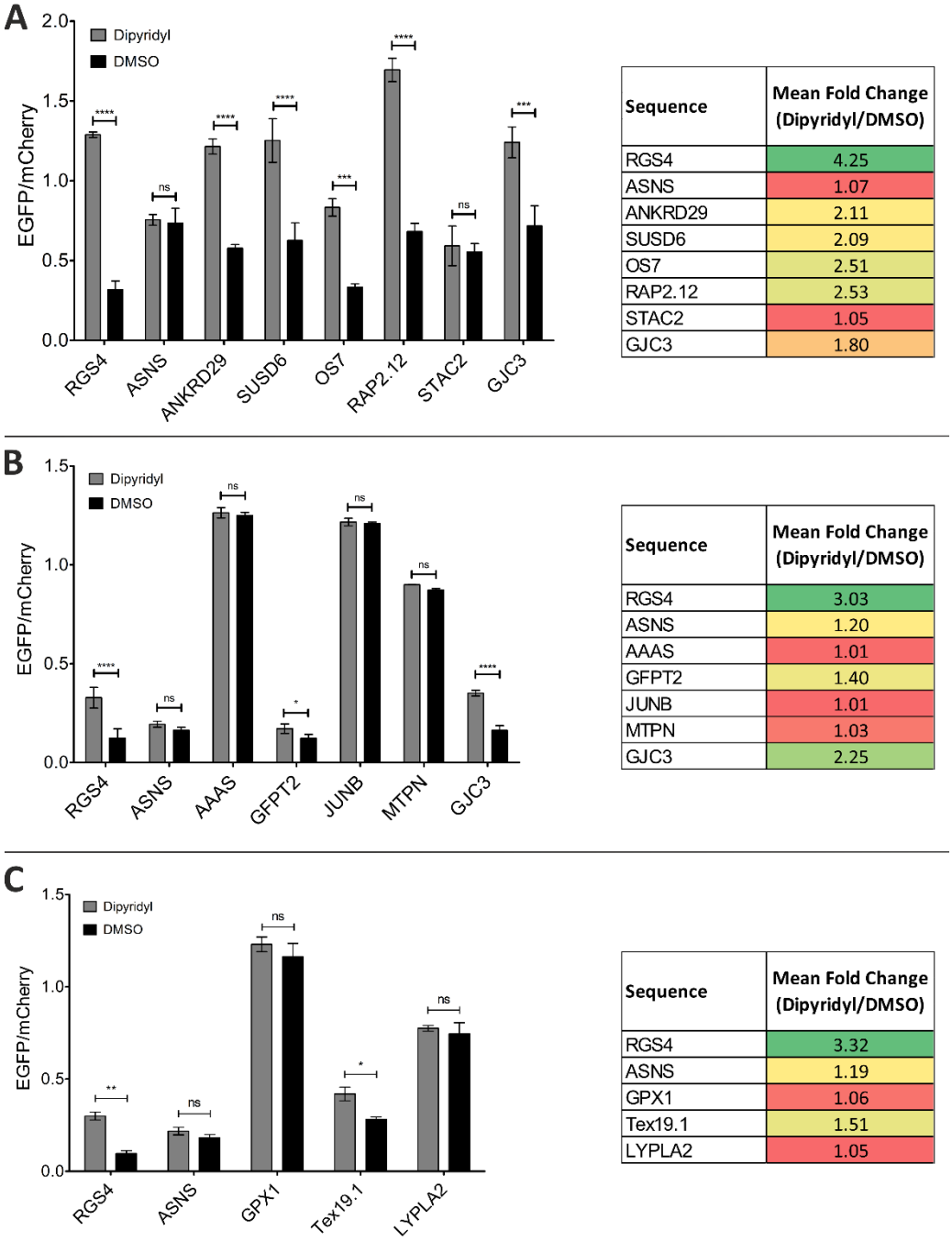
Nt-sequences were first investigated *in cellula* using 2,2-dipyridyl treatment to inhibit *Hs*ADO activity (Chapter 8, page 189). SH-SY5Y cells stably expressing MCX$_n$-EGFP:P2A:mCherry constructs (Table 3.1) were seeded in a 96 well plate, allowing six wells per construct. Cells were allowed to reach 80 % confluency and then growth medium was replaced with DMEM without phenol red (supplemented with fetal bovine serum, L-Glutamine, penicillin, and streptomycin as before (Section 3.4.3)). Phenol red was excluded from this growth medium as it would interfere with fluorescence readings. For three wells per construct, growth media was treated with 2,2-dipyridyl (200 μM); the remaining three wells were treated with an equivalent volume of vehicle control (DMSO, 0.1%). Cells were incubated in normoxia for 24 hours, then EGFP and mCherry fluorescence were measured by multichromatic fluorescence imaging using a FLUOstar® Omega plate reader (BMG LABTECH); gain was adjusted to prevent saturation of mCherry signal and kept consistent across biological repeats. The EGFP fluorescence was measured

at 485 nm excitation, 520 nm emission and mCherry at 544 nm excitation, 590 nm emission. The EGFP/mCherry ratio was calculated using blank-adjusted values. At least three biological replicates were performed for each N-terminus of interest.

Due to the plate-based nature of the assay, it was only possible to investigate up to eight sequences at a time. Each plate included cells expressing $RGS4_{1-15}$-EGFP and $ASNS_{1-15}$-EGFP proteins as positive and negative controls for *Hs*ADO activity respectively. Dr. Keeley had previously conducted DFOR assays using $RGS4_{1-11}$-EGFP. Therefore, cells expressing $RGS4_{1-15}$-EGFP could also be used as a control for the experimental technique as they were expected to behave the same as cells expressing $RGS4_{1-11}$-EGFP based on *in vitro* results with peptides of varying length (Chapter 2, page 37) and previous *in cellula* results[204].

In the first DFOR assay of this kind (Figure 3.3 **A**), the sequences with which *Hs*ADO had the highest *in vitro* initial activity were investigated, along with *At*RAP2.12 and STAC2. 2,2-Dipyridyl treatment, which was expected to result in *Hs*ADO inhibition[1] (Chapter 5, page 109), led to a mean 4.25-fold increase in EGFP/mCherry ratio for cells expressing EGFP conjugated with the positive control (RGS4 Nt-sequence). EGFP/mCherry ratio was also significantly higher (1.80- to 2.53-fold increase) following 2,2-dipyridyl treatment for cells expressing EGFP conjugated with ANKRD29, SUSD6, GJC3, OS7, and *At*RAP2.12 Nt-sequences (Figure 3.3 **A**). These results indicated that these N-termini might all be substrates of *Hs*ADO *in cellula* and aligned with *in vitro* activity results. Also in line with *in vitro* activity data, neither ASNS nor STAC2 N-termini were able to confer 2,2-dipyridyl-dependent stability on EGFP; no significant change in EGFP/mCherry ratio was observed in cells expressing EGFP conjugated with these N-termini.

In the next DFOR assay using 2,2-dipyridyl treatment (Figure 3.3 **B**), sequences were investigated towards which *Hs*ADO had moderate (19.2 % relative to $RGS4_{2-15}$) to low (< 5 % relative to $RGS4_{2-15}$) *in vitro* initial activity. In addition, investigation of the GJC3



Sequence	Mean Fold Change (Dipyridyl/DMSO)
RGS4	4.25
ASNS	1.07
ANKRD29	2.11
SUSD6	2.09
OS7	2.51
RAP2.12	2.53
STAC2	1.05
GJC3	1.80

Sequence	Mean Fold Change (Dipyridyl/DMSO)
RGS4	3.03
ASNS	1.20
AAAS	1.01
GFPT2	1.40
JUNB	1.01
MTPN	1.03
GJC3	2.25

Sequence	Mean Fold Change (Dipyridyl/DMSO)
RGS4	3.32
ASNS	1.19
GPX1	1.06
Tex19.1	1.51
LYPLA2	1.05

Figure 3.3: DFOR assays inhibiting *Hs*ADO with dipyridyl
SH-SY5Y cell lines stably expressing MCX_{13}-EGFP:P2A:mCherry were treated for 20 hours with 2,2-dipyridyl (200 µM) or DMSO vehicle control. EGFP/mCherry ratio calculated using fluorescence measurements. Higher EGFP/mCherry ratio indicates stabilisation of EGFP protein; increased stability after dipyridyl treatment consistent with reduced ADO activity. Error bars show SEM. Statistical significance determined with 2-way ANOVA with Holm-Sidak multiple comparisons post-hoc test: ns ($p > 0.05$); * ($p \leq 0.05$); ** ($p \leq 0.01$); *** ($p \leq 0.001$): **** ($p \leq 0.0001$). Data was collected across three biological repeats. Fold change between EGFP/mCherry ratios for dipyridyl and DMSO treated cells calculated for each biological replicate and averaged across replicates.

N-terminal sequence was repeated as this protein showed the most promise as a potential novel substrate of *Hs*ADO (based on unpublished data obtained by Dr. Norma Masson (Nuffield Department of Medicine) from mice). Interestingly, in this assay, three sequences stabilised EGFP in a non-iron-dependent manner, to the extent that the gain had to be adjusted in fluorescence measurements to prevent saturation of the EGFP signal (Figure 3.3 **B**). This was observed when the N-termini of AAAS, JUNB, and MTPN were added to the N-terminus of EGFP. As with ASNS (and STAC2 in the previous assay (Figure 3.3 **A**)), no significant change in EGFP/mCherry ratio was observed in cells expressing these DFOR constructs following 2,2-dipyridyl treatment. This suggested that these N-termini were not substrates of *Hs*ADO *in cellula*, despite *Hs*ADO having similar initial *in vitro* activity towards $AAAS_{2\text{-}15}$ peptide as *At*RAP2.$12_{2\text{-}15}$. Results obtained in this assay for the GJC3 sequence (Figure 3.3 **B**) aligned with results obtained in the previous assay (Figure 3.3 **A**), with a 2.25-fold increase in EGFP/mCherry ratio observed after 2,2-dipyridyl treatment for cells expressing $GJC3_{1\text{-}15}$-EGFP protein. A slight increase in EGFP/mCherry was measured after 2,2-dipyridyl treatment for cells expressing $GFPT2_{1\text{-}15}$-EGFP (1.40-fold, Figure 3.3 **B**); *in vitro Hs*ADO activity with $GFPT2_{2\text{-}15}$ peptide was 7.8 % relative to $RGS4_{2\text{-}15}$, which is between that for *At*RAP2.$12_{2\text{-}15}$ (an *Hs*ADO substrate *in cellula*, Figure 3.3 **A**) and $JUNB_{2\text{-}15}$ (not an *Hs*ADO substrate *in cellula*, Figure 3.3 **B**). This suggested that GFPT2's N-terminus is a substrate of *Hs*ADO *in vitro* and *in cellula*, but is only weakly regulated.

The third DFOR 2,2-dipyridyl assay assessed the remaining N-termini: GPX1; LYPLA2; *Mm*Tex19.1 (Figure 3.3 **C**). Of these sequences, only *Mm*Tex19.1 was able to confer any 2,2-dipyridyl-dependent stability onto EGFP. A 1.51-fold increase in EGFP/mCherry ratio was observed after 2,2-dipyridyl treatment in cells expressing *Mm*Tex19.$1_{1\text{-}15}$-EGFP (Figure 3.3 **C**). This increase was similar to that observed with $GFPT2_{1\text{-}15}$-EGFP

(Figure 3.3 **B**); *in vitro Hs*ADO activity was similar towards *Mm*Tex19.1$_{2\text{-}15}$ and GFPT2$_{2\text{-}15}$ peptides (7.3 % and 7.8 % relative to RGS4$_{2\text{-}15}$, respectively). These data suggested that weaker *in vitro* substrate sequences may also be regulated less strongly by *Hs*ADO *in cellula.* An unexpected result was that no significant change in EGFP/mCherry was observed for cells expressing GPX1$_{1\text{-}15}$-EGFP when treated with 2,2-dipyridyl or DMSO (Figure 3.3 C). The GPX1$_{2\text{-}15}$ peptide had 42 % initial *Hs*ADO activity *in vitro* relative to RGS4$_{2\text{-}15}$, so this Nt-sequence was expected to be an *Hs*ADO substrate *in cellula.* EGFP fluorescence was also very high for both 2,2-dipyridyl and DMSO treated GPX1$_{1\text{-}15}$-EGFP expressing cells, suggesting that the GPX1 N-terminus stabilised EGFP in a non-*Hs*ADO-dependent manner.

Overall, results from DFOR assays in which *Hs*ADO activity was regulated by 2,2-dipyridyl (Figure 3.3) aligned well with results from *in vitro* activity assays (Table 3.1; Chapter 2, page 41). Addition of highly active *in vitro Hs*ADO substrate sequences (e.g., ANKRD29 and GJC3) conferred *Hs*ADO-dependent stability onto EGFP, while addition of inactive *in vitro* substrate sequences (e.g., ASNS and LYPLA2) did not. Some Nt-sequences that produced weaker peptide substrates of *Hs*ADO *in vitro* (GFPT2 and *Mm*Tex19.1), also seemed to be regulated to a lesser degree in the DFOR assay.

However, some key differences were observed. The N-termini of AAAS and GPX1 did not confer *Hs*ADO-dependent stability onto EGFP, suggesting that these are not *Hs*ADO substrates *in cellula* despite *Hs*ADO having higher *in vitro* activity towards their representative peptides than the *At*RAP2.12 peptide, which did behave as an *Hs*ADO substrate *in cellula* (Figure 3.3). A possible explanation for this is that another Nt-modification occurs on these sequences, which may be able to block *Hs*ADO activity; this will be discussed in Chapter 4 (page 82). It is also possible that these N-termini are substrates of *Hs*ADO *in cellula*, but are not substrates of other components of the

Arg/N-degron pathway (e.g., ATE1). The Arg/N-recognin of the Arg/N-degron pathway E3 ligase requires an internal lysine for ubiquitylation[2]. Although there is not a lysine within the first 14 amino acids of the Met-excised N-terminus of either AAAS or GPX1, there is a lysine three amino acids into the sequence of EGFP, as well as multiple other lysines later in EGFP's sequence. Thus, lack of an internal lysine is likely not responsible for these N-termini not destabilising EGFP in the DFOR system. Another interesting observation was the 2,2-dipyridyl-independent increase in EGFP stability observed with $AAAS_{1-15}$-, $JUNB_{1-15}$-, $MTPN_{1-15}$-, $GPX1_{1-15}$-, and $LYPLA2_{1-15}$-EGFP. This suggested that these N-termini could undergo a stabilising modification *in cellula*.

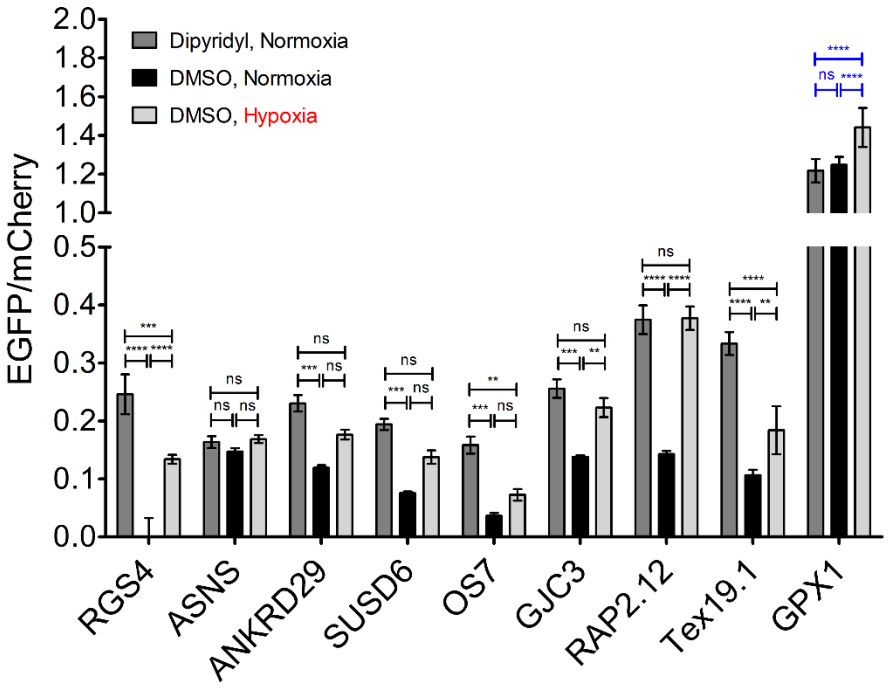
3.5.2 Regulation of *Hs*ADO activity in the DFOR assay with hypoxia

*Hs*ADO activity had previously been shown to be O_2-dependent *in cellula* with three protein substrates (RGS4/5 and IL32)[1]. Therefore, the effect of hypoxia on *Hs*ADO activity with potential substrate N-termini was also investigated in DFOR assays. N-termini with convincing evidence of being *Hs*ADO substrates in 2,2-dipyridyl DFOR assays (Figure 3.3) were chosen for hypoxia DFOR assays: ANKRD29; SUSD6; OS7; GJC3; *At*RAP2.12. Once again, RGS4 and ASNS sequences were used as positive and negative controls, respectively. Finally, *Mm*Tex19.1 and GPX1 were selected examples of a weak *in vitro* substrate that was regulated in initial DFOR assays and a good *in vitro* substrate that was not (Figure 3.3 **C**).

SH-SY5Y cells stably expressing chosen MCX_n-EGFP:P2A:mCherry were seeded into two 96 well plates (Chapter 8, page 189). After cells reached 80 % confluency, the medium was changed as described above (Section 3.5.1). In the first plate, cells were treated with either 2,2-dipyridyl or DMSO vehicle control as before (Section 3.5.1). In the second plate, cells were treated with DMSO vehicle control only. The first plate was incubated in normoxia as before (Section 3.5.1) for 20 hours. The second plate was incubated at 37 °C in hypoxia within an atmosphere-regulated workstation set at 1 % O_2: 5 % CO_2: balance N_2 (Invivo$_2$

400, Baker-Ruskinn Technologies) for 20 hours. After incubation, EGFP and mCherry fluorescence were measured by multichromatic fluorescence imaging as before (Section 3.5.1). Three biological replicates were performed.

$RGS4_{1\text{-}15}$-EGFP and $ASNS_{1\text{-}15}$-EGFP controls behaved as expected in this assay (Figure 3.4). A significant increase in EGFP/mCherry ratio was observed for cells expressing $RGS4_{1\text{-}15}$-EGFP between both 2,2-dipyridyl and DMSO treatment (7.69-fold),



Sequence	Mean Fold Change (Dipyridyl/DMSO)	Mean Fold Change (Hypoxia/Normoxia)
RGS4	7.69	3.99
ASNS	1.11	1.14
ANKRD29	1.94	1.49
SUSD6	2.57	1.84
OS7	4.46	2.12
GJC3	1.85	1.62
RAP2.12	2.62	2.66
Tex19.1	3.22	1.72
GPX1	0.98	1.16

Figure 3.4: DFOR assays inhibiting *Hs*ADO with hypoxia
SH-SY5Y cell lines stably expressing MCX_n-EGFP:P2A:mCherry were treated for 20 hours with 2,2-dipyridyl (200 µM) or DMSO vehicle control and incubated in normoxia or hypoxia (1 % O_2). EGFP/mCherry ratio was calculated using fluorescence measurements. Higher EGFP/mCherry ratio indicates stabilisation of EGFP protein; increased stability after 2,2-dipyridyl or hypoxia treatment is consistent with reduced ADO activity. Error bars show SEM. Statistical significance determined with Two-way ANOVA with Holm-Sidak multiple comparisons post-hoc test: ns ($p > 0.05$); * ($p \leq 0.05$); ** ($p \leq 0.01$); *** ($p \leq 0.001$): **** ($p \leq 0.0001$). Statistical analysis excluding GPX1 data is shown in black and analysis including GPX1 data is shown in blue. Data was collected across three biological repeats. Fold change between EGFP/mCherry ratios for 2,2-dipyridyl/DMSO and hypoxia/normoxia treated cells was calculated for each biological replicate and averaged across replicates.

and between hypoxic and normoxic incubation (3.99-fold). Meanwhile no significant differences were observed for $ASNS_{1-15}$-EGFP-expressing cells subjected to different conditions. Once again, EGFP fluorescence was high for all cells expressing GPX_{1-15}-EGFP (EGFP/mCherry ratio > 1.0) compared to other cells (EGFP/mCherry ratio < 0.5). This appeared to skew grouped statistical analysis; therefore, analysis was conducted for this assay both with and without data for the GPX1 sequence (Figure 3.4). Statistical analysis suggested that a significant increase in EGFP/mCherry ratio was observed between $GPX1_{1-15}$-EGFP-expressing cells incubated in hypoxia versus normoxia. However, this increase was only 1.16-fold (similar to that for ASNS (1.14-fold), so the GPX1 N-terminus was not thought to have conferred O_2-dependent stability onto EGFP.

Conjunction of all other N-termini to EGFP lead to an increase in EGFP/mCherry ratio of 1.49-fold or greater following 2,2-dipyridyl or hypoxia treatment. For cells expressing *At*RAP2.12_{1-15}-EGFP, 2,2-dipyridyl treatment and hypoxia led to a similar level of EGFP stabilisation (2.62-fold vs 2.66-fold increase in EGFP/mCherry ratio). However, in other constructs the stabilising effect of hypoxia on MCX_n-EGFP was smaller than that of 2,2-dipyridyl. The effect was most notable for cells expressing OS7-EGFP; a 4.46-fold increase in EGFP/mCherry ratio was observed when these cells were treated with 2,2-dipyridyl versus DMSO, while only a 2.12-fold increase was seen between hypoxia and normoxia. The concentration of 2,2-dipyridyl used in DFOR assays in this Chapter was twice that previously used by Dr. Keeley[204]. So, it was not unexpected that 2,2-dipyridyl treatment produced more significant inhibition of *Hs*ADO *in cellula* than incubation in 1 % O_2, where some O_2 was still available for *Hs*ADO-catalysed oxidation. Overall, the data from the hypoxia DFOR assay (Figure 3.4) agreed with the results from initial assays with 2,2-dipyridyl (Figure 3.3), indicating that the Nt-sequences of ANKRD29, SUSD6, GJC3, OS7, *At*RAP2.12, and *Mm*Tex19.1 were *Hs*ADO substrates *in cellula.*

All DFOR plasmids were sequenced prior to transfection. However, in light of some unexpected results (e.g., the lack of regulation of the GPX1 Nt-sequence) and the potential identification of novel ADO substrate sequences *in cellula,* a PCR screen was conducted as a secondary verification method for the constructs used in the hypoxia DFOR assays. Furthermore, it was useful to confirm that the O_2-dependence of the CX_n-EGFP fluorescence was due to the nature of the CX_n N-terminus, as maturation of the GFP fluorophore has been shown to require oxygen[212]. In practice, 1 % O_2 should still be sufficient for GFP fluorescence[204,213].

The PCR screen was designed to confirm the presence of the correct DFOR constructs in the cell lines used for the hypoxia DFOR assays. Forward primers were designed for this screen, which would only bind to DNA corresponding to the Nt-sequence appended to EGFP (F_HindIII_RGS4[15]_NoGFP, etc., (Appendix 7)). These primers did not contain any DNA complementary to the N-terminus of EGFP itself; only the reverse primer would bind to the EGFP DNA sequence. Genomic DNA (gDNA) was extracted from cells transfected with MCX_n-EGFP:P2A:mCherry constructs corresponding to the sequences examined using hypoxia (Chapter 8, page 189). PCR reactions were carried out mixing gDNA from each cell line with the reverse primer (R_GFP_Xba1 (Appendix 7)) and each different forward primer (Appendix 7) (Figure 3.5 **A**). DNA would only be amplified in these reactions if the sequence appended to EGFP in the DFOR plasmids matched that of the forward primer added. PCR reactions were analysed by gel electrophoresis (Figure 3.5 **B**).

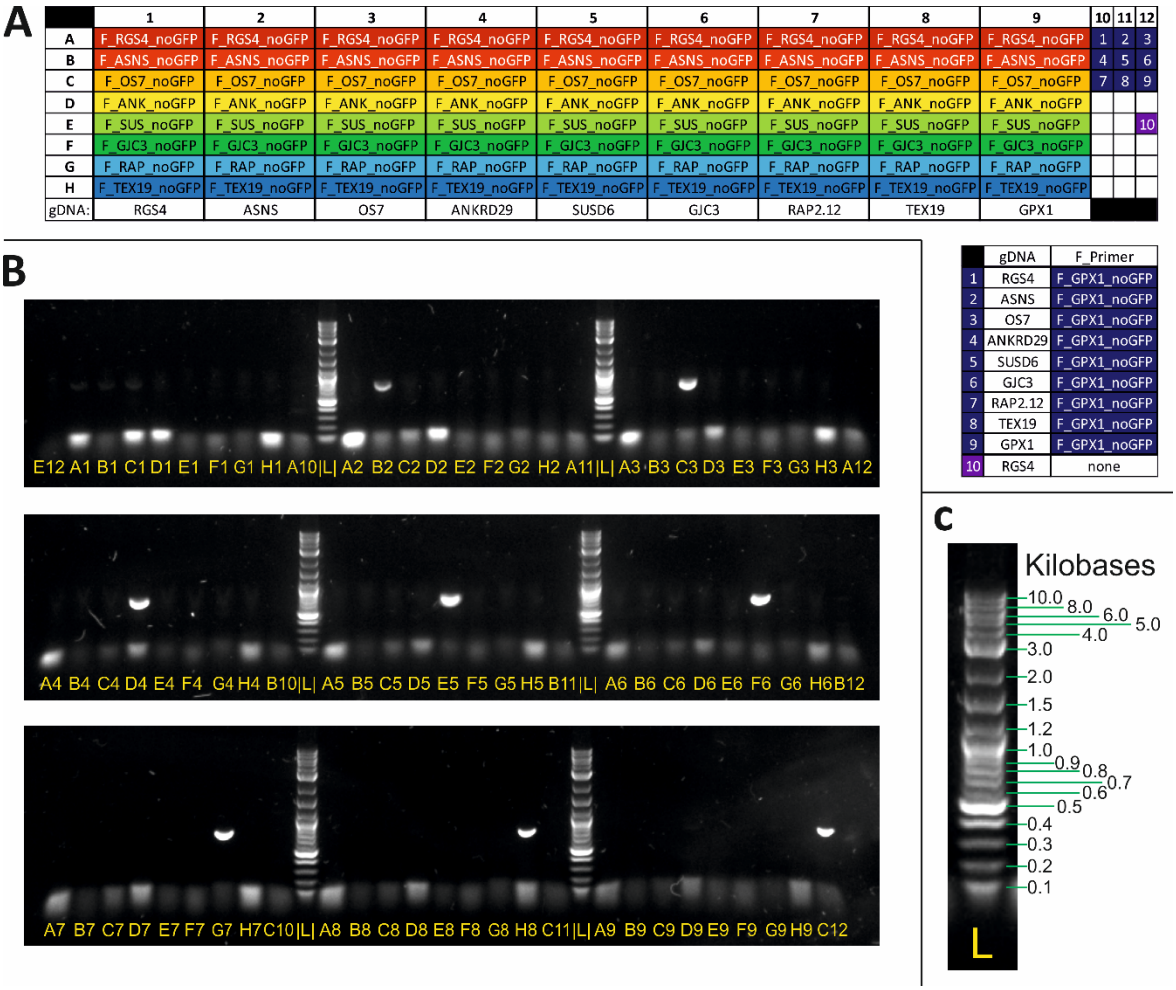


Figure 3.5: PCR analysis of gDNA extracted from SH-SY5Y cells used in the DFOR hypoxia assay
A: Plate map describing combinations of genomic DNA (gDNA) and forward primers used in the PCR reactions. gDNA: gDNA extracted from cells stably expressing DFOR plasmids, e.g., 'RGS4' gDNA extracted from cells expressing RGS4$_{1\text{-}15}$-EGFP:P2A:mCherry plasmid.
B: Gel electrophoresis analysis of PCR reactions described in **A**. A gel band for plasmid DNA was only observed when gDNA combined with the corresponding forward primer (e.g., ASNS gDNA and F_ASNS_noGFP). gDNA for RGS4 was lost by experimental error during extraction.
C: Expanded view of 1 kb plus DNA ladder (New England BioLabs) used as reference.

Experimental error led to loss of gDNA from cells transfected with RGS4$_{1\text{-}15}$-EGFP:P2A:mCherry during extraction so only primer bands were observed by gel electrophoresis for these PCR reactions. Bands at the expected length were observed for all other combinations of gDNA with the correct forward primer (e.g., gDNA extracted from cells transfected with ASNS$_{\text{-}15}$-EGFP:P2A:mCherry (ASNS gDNA) and F_ASNS_noGFP). This confirmed that cells used in the hypoxia DFOR assay (Figure 3.4) contained the correct constructs.

Although gDNA was not successfully extracted from cells transfected with $RGS4_{1-15}$-EGFP:P2A:mCherry, comparison of the behaviour of these cells with cells previously used by Dr. Keeley[204] (expressing $RGS4_{1-11}$-EGFP:P2A:mCherry) suggested that these cells were also expressing the correct construct. Similar changes in GFP and mCherry levels were observed for both sets of cells following treatment with 2,2-dipyridyl and hypoxia. This, combined with sequencing results obtained prior to transfection was taken as sufficient evidence that the cells used in this Chapter contained the $RGS4_{1-15}$-EGFP:P2A:mCherry construct.

Overall, DFOR assays with both 2,2-dipyridyl (Figure 3.3) and hypoxia (Figure 3.4) demonstrated that *Hs*ADO was selective amongst Nt-Cys sequences *in cellula*, and that *in vitro* peptide activity data was generally a good indicator of whether an Nt-sequence would be a substrate in DFOR assays. Exceptions to this included GPX1 and AAAS N-termini, which were moderately good *Hs*ADO substrates in peptide activity assays but were not regulated by *Hs*ADO in DFOR assays. A surprisingly low level of *in vitro Hs*ADO activity with a representative peptide (> 7 % relative to RGS4) proved to be sufficient for an N-terminal sequence to confer *Hs*ADO substrate properties onto EGFP (for example, *Mm*Tex19.1).

3.6 Cellular assays with full-length putative substrate proteins

Alongside *in vitro* experiments with potential *Hs*ADO substrates described in Chapter 2 (page 41) and the DFOR assays described above (Section 3.5), Dr. Keeley (NDM) performed experiments in human-derived cells to investigate the behaviour of potential *Hs*ADO substrates as full-length proteins. Experimental details and results described in this Section were kindly provided by Dr. Keeley.

To examine the effect of *Hs*ADO on endogenous levels of putative *Hs*ADO substrate proteins, wild-type (WT), ADO knock-out (KO), and ADO over-expressing (OE) SH-SY5Y cells were exposed to normoxia or hypoxia (1 % O_2) for 16 hours (Figure 3.6 **A**). Protein was then extracted from cells and examined by immunoblotting. Levels of RGS4 and RGS5 proteins (known *Hs*ADO substrates[1]) were used as positive controls. RGS4 and RGS5 were stabilised by hypoxia in WT cells; in ADO KO cells these proteins were stabilised independent of hypoxia and in ADO OE cells these proteins were degraded in both normoxia

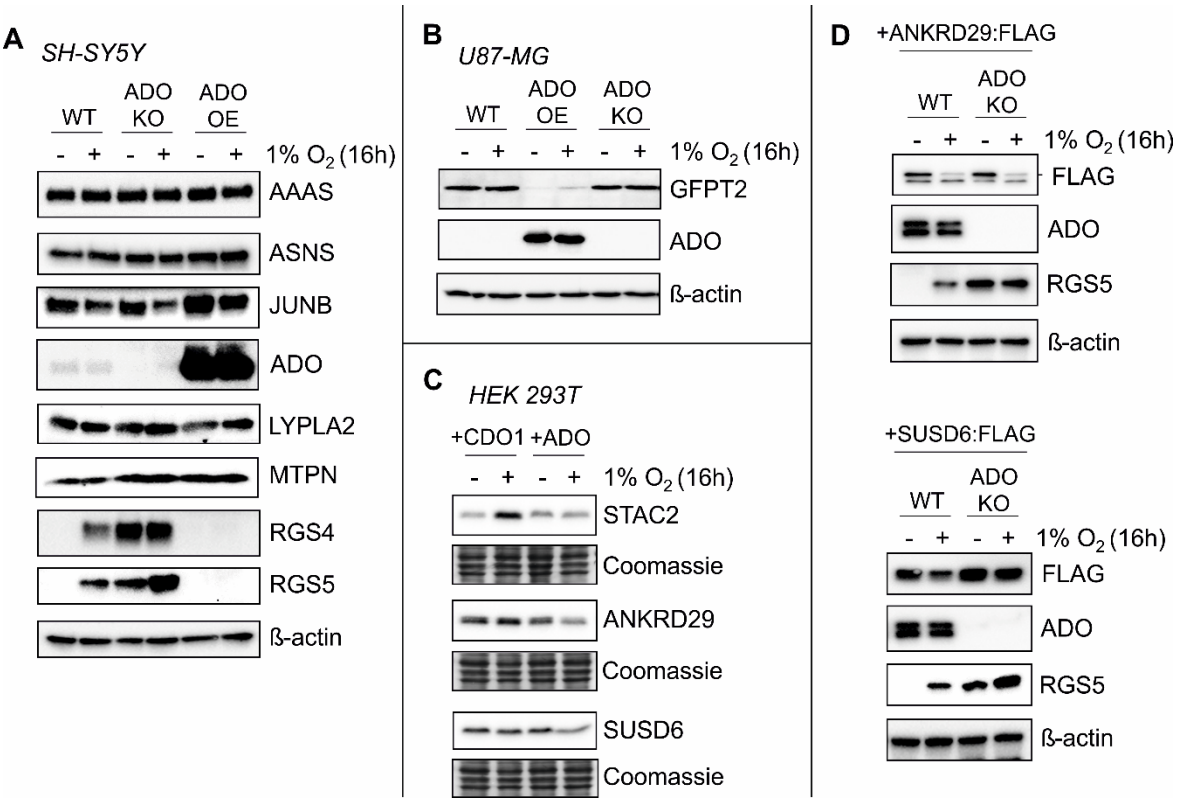


Figure 3.6: Regulation of putative ADO substrates in human cells
Blot images and experiment details kindly provided by Dr. Keeley. All blots are representative of at least 3 independent experiments.
A: wild-type (WT), ADO knock-out (KO), or ADO overexpressing (OE) SH-SY5Y cells were exposed to hypoxia (1 % O_2) for 16 hours. Protein samples were collected and analysed by Western Blot to examine levels of potential ADO substrate proteins. RGS4 and RGS5 were used as positive controls.
B: WT, ADO KO, or ADO OE U87-MG cells were exposed to hypoxia as before (**A**). Protein samples were collected and analysed by Western Blot to examine levels of GFPT2.
C: HEK293T cells were transiently co-transfected with cDNAs encoding for STAC2, ANKRD29, or SUSD6, alongside either CDO1 or ADO. Co-transfection with ADO mimics ADO overexpression, while co-transfection with CDO1 was used as a control. Cells were exposed to hypoxia as before (**A**) and analysed by Western Blot.
D: WT or ADO KO SH-SY5Y cells stably expressing ANKRD29:FLAG, or SUSD6:FLAG were subjected to hypoxia as before (**A**) and analysed by Western Blot using an anti-FLAG antibody. RGS5 levels were analysed as a positive control.

and hypoxia. Levels of all other proteins analysed in this assay did not vary substantially with hypoxia or ADO levels, indicating that these proteins (AAAS, ASNS, JUNB, LYPLA2, and MTPN) were not substrates of *Hs*ADO *in cellula* (Figure 3.6 **A**). These data correlated with data from the DFOR assays, which also suggested that the N-termini of these proteins were not *Hs*ADO substrates *in cellula* (Figure 3.3). GFPT2 was not detectable in SH-SY5Y cells, so U87-MG cells were used to investigate this protein as an ADO substrate *in cellula*. WT, ADO OE and ADO KO U87-MG cells were also exposed to hypoxia as described above, and protein extraction and immunoblotting were used to examine levels of endogenous GFPT2. Interestingly, GFPT2 protein levels were consistent across WT and ADO KO cells exposed to normoxia or hypoxia, but were depleted in normoxic ADO OE cells, an effect that was reversed when these cells were subjected to hypoxia. This suggested that GFPT2 can be regulated by *Hs*ADO in cells, but only when *Hs*ADO is present at higher than WT levels, which is consistent with *in vitro* (Chapter 2, page 41) and DFOR assay data (Figure 3.3 **B**) that suggested that the GFPT2 N-terminus was a weak substrate of *Hs*ADO.

Endogenous levels of ANKRD29, SUSD6, and STAC2 could not be detected in cells using antibodies available at the time, so cDNA encoding for these proteins were transiently co-transfected into HEK 293T cells, alongside either ADO or CDO1 cDNA (Figure 3.6 **C**). Co-transfection alongside ADO would mimic the ADO over-expression model in SH-SY5Y (Figure 3.6 **A**) and U87-MG (Figure 3.6 **B**) cells, repressing the effect of hypoxia on protein stability for *Hs*ADO substrates. Meanwhile, co-transfection with CDO1 cDNA would mimic the wild-type model, so protein stability should be O_2-dependent for *Hs*ADO substrates, while also acting as a transfection control. In this assay neither ANKRD92 nor SUSD6 behaved as would be expected of *Hs*ADO substrates, while STAC2 stability was O_2-dependent in the CDO1 co-transfected cells. This did not align with observations from *in vitro* (Chapter 2, page 41) and DFOR experiments (Section 3.5), which suggested the

opposite: that ANKRD29 and SUSD6 should be *Hs*ADO substrates, and not STAC2. It is proposed that the Nt-sequence of STAC2 is unfavourable for *Hs*ADO binding, but that there is another binding site on the full-length STAC2 protein, which is able to compensate (discussed further below (Section 3.7)).

Transient transfection often leads to very high expression of target proteins, so it was possible that the lack of regulation of ANKRD29 and SUSD6 in HEK293T cells (Figure 3.6 **C**) was due to saturation of the endogenous *Hs*ADO pathway. Therefore, ANKRD29 and SUSD6 proteins were further investigated *in cellula* using SH-SY5Y cells stably expressing C-terminally FLAG-tagged proteins, in which stable integration of the cDNA into the host genome generally results in a lower level of expression than transient transfection (Figure 3.6 **D**). In this experiment, levels of ANKRD29:FLAG and SUSD6:FLAG did not vary in an ADO- or O_2-dependent manner. Cellular data therefore indicated that ANKRD29 and SUSD6 proteins were not *in cellula* substrates of *Hs*ADO (Figure 3.6 **C** and **D**). A possible explanation for the lack of *in cellula Hs*ADO activity with ANKRD29 and SUSD6 proteins, despite positive *in vitro* (Chapter 2, page 41) and DFOR assay (Figure 3.3 and Figure 3.4) data, lies in protein folding. In rice, there is an ERF-VII transcription factor that is not regulated by the N-degron pathway, despite having an Nt-Cys like other ERF-VIIs (e.g., *At*RAP2.12): SUB1A-1[56]. In 2019, it was shown that this lack of regulation was due to the C-terminus of SUB1A-1 interacting with its N-terminus, leaving the N-terminus unavailable for modification[77]. It is possible that something similar occurs in full-length ANKRD29 and SUSD6 proteins, protecting their N-termini from *Hs*ADO-catalysed Cys oxidation. In the DFOR assay, previous experiments had suggested that the N-terminus was likely to be accessible[204], so this phenomenon was expected not a primary concern.

3.7 Discussion

In this Chapter, Nt-Cys-bearing sequences were investigated as *Hs*ADO substrates in cellular assays. These experiments built upon results obtained in *in vitro* peptide activity assays with recombinant *Hs*ADO (Chapter 2, page 41). This was first achieved using a dual-fluorescence oxygen reporter (DFOR) assay, in which an Nt-sequence of interest was appended to EGFP, allowing *Hs*ADO-dependent stability of these N-termini to be assessed using fluorescence measurements (Section 3.3). This method was used to test a selection of N-termini towards which *Hs*ADO had varied levels of activity in peptide activity assays *in vitro* (Section 3.4.1). Results from the DFOR assays (Section 3.5) were mostly in keeping with the *in vitro* data, and demonstrated that *Hs*ADO was selective amongst Nt-Cys sequences *in cellula*. N-termini of proteins whose representative peptides were good *in vitro* substrates (> 80 % initial activity relative to RGS4, e.g., ANRKD29 and SUSD6) acted as *Hs*ADO substrates in DFOR assays, while those whose peptides were poor *in vitro* substrates (< 5 % initial activity relative to RGS4, e.g., JUNB and LYPLA2) did not. However, DFOR data for two Nt-sequences did not align with peptide activity data: AAAS and GPX1, which were not observed to be *Hs*ADO substrates in the DFOR assay, despite *Hs*ADO having higher initial activity towards their representative peptides *in vitro* than towards other sequences that were found to be *Hs*ADO substrates *in cellula* (e.g., *At*RAP2.12). An explanation proposed for this phenomenon was that another Nt-modification may be influencing *Hs*ADO activity; a hypothesis that will be explored in Chapter 4 (page 82).

This Chapter also discussed *in cellula* experiments performed by Dr. Keeley investigating *Hs*ADO activity with full-length putative substrate proteins (Section 3.6). Data from these assays revealed further discrepancies between *in vitro* and cellular experiments. The full-length STAC2 protein behaved as an *Hs*ADO substrate when transiently transfected

into HEK 293T cells (Figure 3.6 **C**). This sequence was not expected to be an *Hs*ADO substrate based on DFOR and *in vitro* peptide activity assay data; the STAC2 N-terminus also includes an 'EE' motif after position 11, which was shown to be particular unfavourable for *Hs*ADO activity in experiments with 'anti-substrate' peptides (Chapter 2, page 47). It is hypothesised that the N-terminus of STAC2 is unfavourable for *Hs*ADO binding, but that there could another binding site present in the full-length protein that might be able to overcome this. Although this is an interesting phenomenon, investigation of STAC2 as a new protein substrate of ADO was not pursued as the MC-initiating variant used in this Thesis is an isoform predicted from alternative splicing and not the canonical isoform.

Finally, two proteins that were expected to be *Hs*ADO substrates, based on high activity of *Hs*ADO with their representative peptides *in vitro* (Chapter 2, page 41) and regulation of their Nt-sequences in the DFOR assays (Section 3.5), did not behave as *Hs*ADO substrates in two separate cellular contexts (Section 3.6): ANKRD29 and SUSD6. One possibility is that this lack of regulation of the full-length proteins by *Hs*ADO *in cellula* could be due to their N-termini being protected from modification through interaction with other parts of the protein when fully folded. This has been previously observed in SUB1A-1 protein in rice, which is protected from PCO-catalysed oxidation through interaction of its N- and C-termini[77].

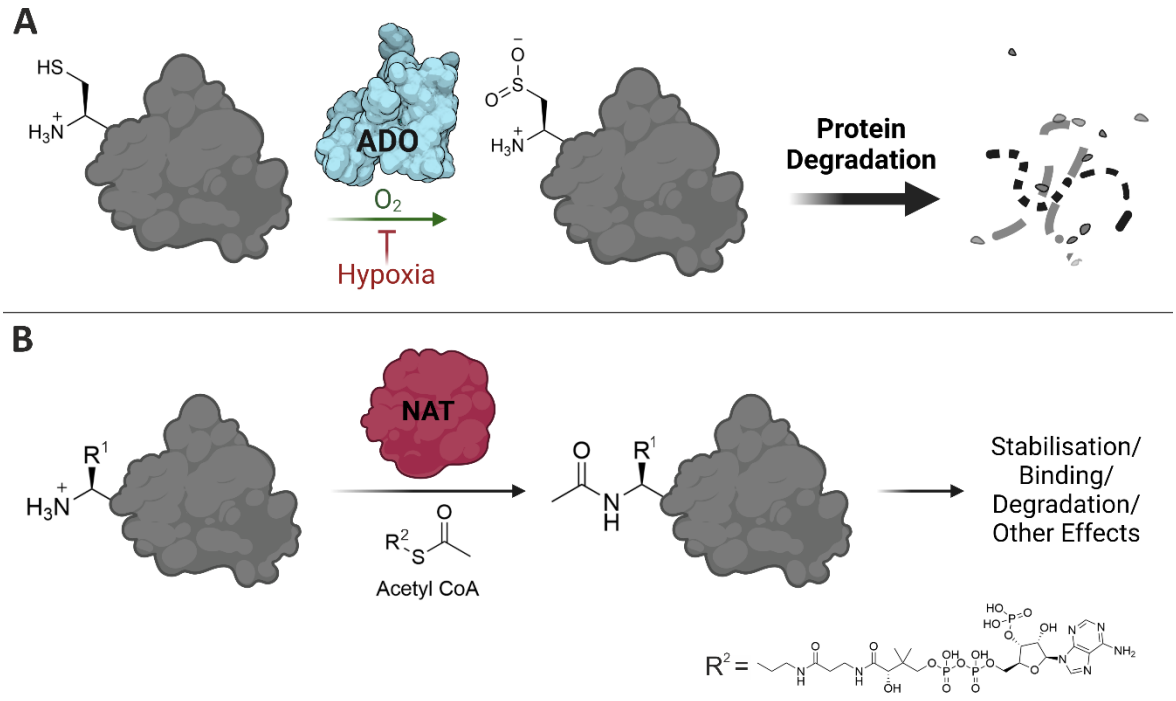
Overall, the experiments discussed in this Chapter demonstrated that *Hs*ADO is selective amongst Nt-Cys sequences *in cellula*, as well as *in vitro* (Chapter 2, page 41). However, these data also serve as a reminder that there are many other factors to consider when using *in vitro* data to make predictions about *in cellula* behaviour.

4 The Interplay of Nt-Cys Oxidation and Acetylation

4.1 Introduction

In Chapter 3 (page 80), differences were noted between *in vitro* and *in cellula* data obtained with potential substrates of *Hs*ADO. Notably, there were some sequences that were expected to be regulated by *Hs*ADO based on *in vitro* data, which did not behave like substrates in human cells: AAAS and GPX1. It was suggested that this could be due to another N-terminal (Nt) modification occurring on these sequences and influencing *Hs*ADO activity *in cellula*. There are a variety of different modifications that can occur on a protein's N-terminus[78] (Chapter 1, page 9). Many Nt-modifications occur on the Nt-amine, such as acetylation, propionylation, and myristoylation[78]. Nt-modifications can occur both co-translationally and post-translationally, and can occur sequentially. For example, in Chapter 2 (page 37) it was demonstrated *in vitro* that methionine excision[79] is necessary to reveal the Nt-Cys prior to *Hs*ADO-catalysed oxidation of the thiol group (Scheme 4.1 **A**).

One of the most common Nt-modifications is Nt-acetylation (NTA), with approximately 80 % of the human proteome estimated to be N-terminally acetylated[85] (Chapter 1, page 9). NTA is the transfer of an acetyl group from acetyl coenzyme A (CoA) to the Nt-amine of a protein, and is catalysed by N-terminal acetyltransferases (NATs) (Scheme 4.1 **B**). Thus far no N-terminal deacetylases have been identified, so NTA is generally assumed to be irreversible. Most known NTA occurs co-translationally, and can lead to a variety of different effects at the protein level, including localisation and degradation[85]. In eukaryotes, there are eight known NATs, which have specificity towards different proteins[85]. NatA catalyses acetylation of protein N-termini with a relatively small N-terminal amino acid (A/S/T/V/C/G) after the initial methionine is cleaved, although the level of acetylation varies within this pool of substrates[87]. NatB is specific for N-termini that initiate with MD/ME/MN/MQ[91], while NatC, NatE, and NatF can catalyse acetylation of protein termini



Scheme 4.1: ADO and NAT catalysed reactions on N-termini
A: Cysteamine dioxygenase (ADO)-catalysed oxidation of Nt-Cys, leading to degradation of target proteins
B: N-terminal acetyltransferase (NAT)-catalysed acetylation of Nt-residues, leading to a variety of different effects. (R^1 = amino acid side chain).

starting with ML/MI/MF/MY/MK, MS/MT/MA/MV/ML/MI/MF/MY/MK, and ML/MI/MF/MY/MK, respectively[100,214,215], with NatF specifically catalysing acetylation of membrane proteins[88]. NatD has been shown to be specific for histones H2A and H4[93,94]. In plants, NatG is able to catalyse acetylation of proteins with M/A/S/T N-termini[98]. Finally, in animals, acetylation of mature acidic actin N-termini, generated by post-translational cleavage[95,96], can be catalysed by NatH[97].

NatA is the NAT believed to be responsible for catalysis of acetylation of Nt-Cys, along with other small amino acids such as serine and threonine[85]. In a recent study, NatA was linked to protein stabilisation in animals; Mueller *et al.* demonstrated that NTA of some NatA substrates was protective against inhibitor of apoptosis protein (IAP) family E3 ubiquitin ligases, which are linked to protein degradation[109]. Furthermore, depletion of NatA activity in Arabidopsis has been shown to increase protein turnover[108]. It was hypothesised

that the lack of *Hs*ADO-mediated degradation observed with some sequences in Chapter 3 (page 66) could be due to stabilisation caused by acetylation.

4.2 Chapter Aims

In this Chapter, NatA activity with Nt-Cys sequences was investigated *in vitro*, through a collaboration with Prof. Thomas Arnesen and colleagues at the University of Bergen. Initial NatA activity screens were performed by Dr. Matti Myllykoski (University of Bergen) with a variety of different Nt-Cys peptides to determine whether human NatA could catalyse acetylation of Nt-Cys and to investigate its substrate preferences. To further validate these data, NatA activity with the most active substrate sequences identified in these screens was then assessed by high throughput MS, including assessment of the influence of peptide length on *in vitro* NatA activity. Tandem mass spectrometry (MS/MS) was used to confirm that NatA-catalysed acetylation occurred on the Nt-Cys residue. NatA substrate preferences were further investigated using variants of a substrate peptide, in which the residue after the Nt-Cys was mutated to different amino acids. Lastly, this Chapter addressed whether Nt-Cys acetylation and oxidation by NatA and ADO were mutually exclusive, using acetylated and oxidised versions of substrate peptides in *in vitro* activity assays.

4.3 NatA is selective amongst Nt-Cys peptides *in vitro*

NatA has been assumed to be responsible for acetylation of Nt-Cys, along with other small Nt-residues[85]. However, there is little evidence for NatA-catalysed acetylation of Nt-Cys. Proteins with Nt-Cys are generally poorly represented in proteomic datasets investigating the Nt-acetylome[87,88,91,97,100,214–218], which has made investigation of Nt-Cys acetylation challenging. It was therefore necessary to establish that NatA catalysed acetylation of Nt-Cys *in vitro* and to probe its substrate preferences regarding residues beyond the Nt-Cys.

4.3.1 NatA activity *in vitro* varies with Nt-Cys peptide sequence

In order to investigate NatA activity with different Nt-Cys proteins, thirty MC-initiating proteins from the human proteome were selected, which had previously been screened against *Hs*ADO *in vitro* (Chapter 2, page 41). A further four Met-Cys-initiating proteins were selected by Prof. Arnesen and Dr. Myllykoski (University of Bergen) due to their previous detection in the Nt-acetylated state in cells[91,215]: myosin light polypeptide 6 (MYL6; accession P60660); protein arginine *N*-methyltransferase 3 (PRMT3; accession O60678); endonuclease III-like protein 1 (NTHL1; accession P78549); uncharacterized protein FAM241A (FAM241A; accession Q8N8J7). Two more sequences were chosen: actin alpha 2 (ACTA2; accession P62736), which has a highly acidic N-terminus that would be expected to be a poor *Hs*ADO substrate based on *in vitro* data (Chapter 2, page 41); and mouse receptor-interacting serine/threonine-protein kinase 1 (*Mm*Ripk1; accession Q60855), a mouse protein in which an Nt-Cys is revealed by caspase cleavage after Asp325[219,220]. Finally, the most rapidly oxidised peptide from *in vitro* *Hs*ADO assays, OS7, (Chapter 2, page 47) was included to assess whether a synthetic 'optimum' *Hs*ADO substrate could be acetylated by NatA.

NatA activity with these Nt-Cys sequences was investigated *in vitro* by Dr. Myllykoski using an established radioactivity-based assay[221]. Recombinant NatA protein used for activity assays was a 1:1 complex of human $NAA10_{1-160}$ and *N*-His_6-tagged $NAA15_{1-866}$. Potential substrate peptides were incubated with recombinant NatA protein and a 1:3 mix of [^{14}C]-labelled and unlabelled acetyl CoA in assay buffer (50 mM Na-phosphate pH 7.0, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 2 mg ml^{-1} BSA) for 5 minutes at 37 °C with shaking, using a protocol adapted from Drazic and Arnesen[221] (Chapter 8, page 178). The reaction was transferred to phosphocellulose P81 filter discs to quench the reaction and to extract peptides for analysis. Unincorporated [^{14}C]-acetyl CoA was removed from P81 filter

discs by washing, and the amount of [^{14}C]-acetyl incorporated into peptides was determined by measuring their radioactivity. Initial NatA activity with each peptide was calculated as the level of [^{14}C]-acetyl incorporation after 5 minutes, normalised to [^{14}C]-acetyl incorporation into a HMGA1 positive control peptide[87] (Figure 4.1). Background acetylation of Cys-containing peptides was also observed, and was therefore measured using 'no enzyme' controls and subtracted from the acetylation measured in the presence of NatA.

Radioactivity-based NatA activity assays were initially attempted using 14-mer peptides as in ADO activity assays (Chapter 2, page 41). However, these peptides were found to be incompatible with the assay set-up, which was attributed to insufficient positive charge for binding to the filter discs (as described in the assay procedure above). To resolve this issue, peptides were instead purchased featuring seven unique amino acids at the N-terminus (CX_6), followed by residues 8 - 24 (RWGRPVGRRRRPVRVYP) from the HMGA1 positive control peptide to facilitate binding of the peptides to the P81 filter paper. The unique amino acids in the different peptides tested corresponded to the first seven amino acids of the Met-excised N-terminus of the protein of interest, the caspase-cleaved C-terminal protein fragment in the case of Ripk1, or the peptide sequence of OS7.

Initial NatA activity with Nt-Cys peptides ranged from 0 to 120 % relative to the HMGA1 positive control peptide (Figure 4.1). Good *in vitro* peptide substrates of NatA included those representing proteins that had previously been detected in the Nt-acetylated state in cells by proteomic studies: MYL6[97]; PRMT3[109,215]; NTHL1[91]; FAM241[215]; MTPN[100,109]; AAAS[109,215]; and JUNB[109,215]. The best NatA substrate identified in Dr. Myllykoski's screening was a peptide featuring the Met-excised N-terminus of DYNLL1. Notably, NatA initial activity was very low with known *Hs*ADO protein substrates[1], RGS4 and IL32 (0.13 and 0.48 % relative to HMGA1, respectively). All but one of the Nt-Cys peptides with which NatA had > 10 % activity relative to HMGA1 had an acidic (D, E) or polar (S, T)

N-terminus	Name	Peptide residues 1-7	NatA activity normalised to SESS
SE	HMGA1	SESSSKS	100.0
CD	MYL6	CDFTEDQ	6.7
	MTPN	CDKEFMW	22.6
	DYNLL1	CDRKAVI	119.5
CE	ACTA2	CEEEDST	3.4
	ETNPPL	CELYSKR	3.3
	IL6	CESSKEA	8.1
CS	FAM241A	CSAGELL	5.5
	PRMT3	CSLASGA	32.5
	AAAS	CSLGLFP	23.3
	NTHL1	CSPQESG	7.6
	RIMKLA	CSQLWFL	6.0
CT	JUNB	CTKMEQP	22.2
CH	SUSD6	CHGRIAP	0.1
	IER3/IEX1	CHSRSCH	10.2
CK	USP27X	CKDYVYD	0.0
	RGS4	CKGLAGL	0.1
	STAC2	CKVSVHL	0.2
CR	CDH2	CRIAGAL	0.0
	ANKRD29	CRMSFKK	0.0
	RGS16	CRTLAAF	0.5
CF	OS7	CFGRRMK	0.0
	SHHC	CFPGSAT	0.1
	IL32	CFPKVLS	0.5
	ZIP12	CFRTKLS	0.0
CA	GPX1	CAARLAA	1.0
CG	CHURC1	CGDCVEK	0.2
	GFPT2	CGIFAYM	0.9
	ASNS	CGIWALF	0.0
	LYPLA2	CGNTMSV	0.2
CN	ALS2CL	CNPEEAA	0.3
	VDAC3	CNTPTYC	0.7
CP	P2RX6	CPQLAGA	0.2
	CD73	CPRAARA	0.3
	CAHLM4	CPTLNNI	0.0
CV	Ripk1*	CVPLPPS	0.0
CY	TRAF1	CYRAPCS	0.6

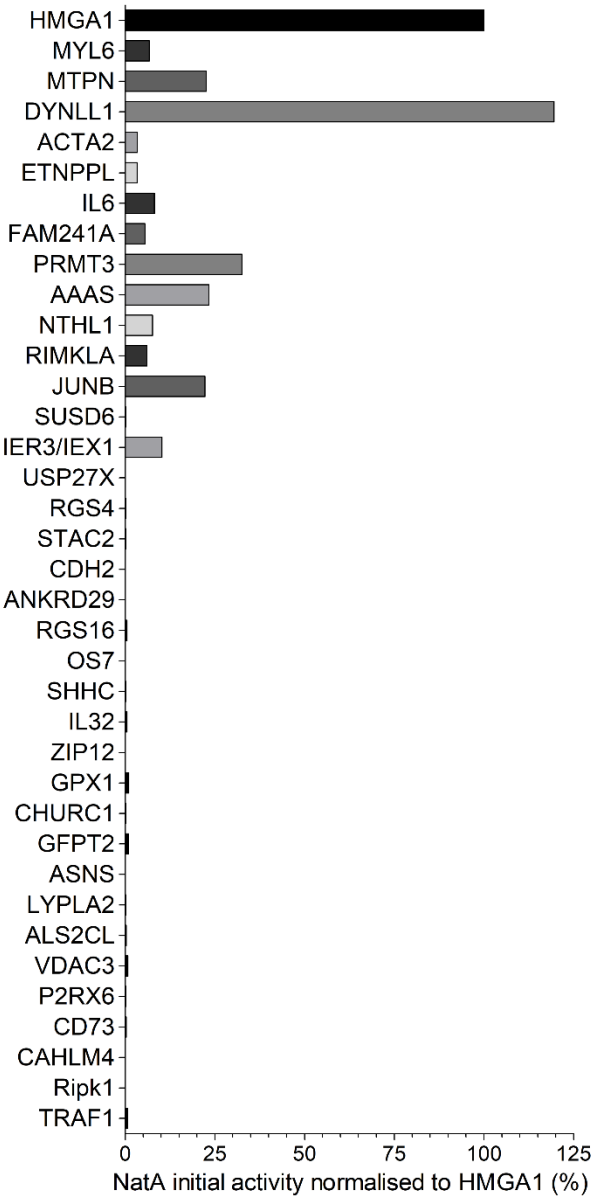


Figure 4.1: NatA-catalysed acetylation of Nt-Cys peptides
NatA-catalysed turnover of Nt-Cys peptides of interest after 5 minutes (200 µM peptide, 0.3 µM NatA, 37 °C); initial activity was normalised to turnover of HMGA1 positive control peptide. Peptides took the form [1-7]-RWGRPVGRRRRPVRVYP, where residues 1-7 represented the first 7 amino acids of the Met-excised N-terminus of the protein (or of OS7). Data kindly provided by Dr. Matti Myllykoski (University of Bergen).
*Nt-Cys of Ripk1 revealed following caspase 8 cleavage at residue 326.

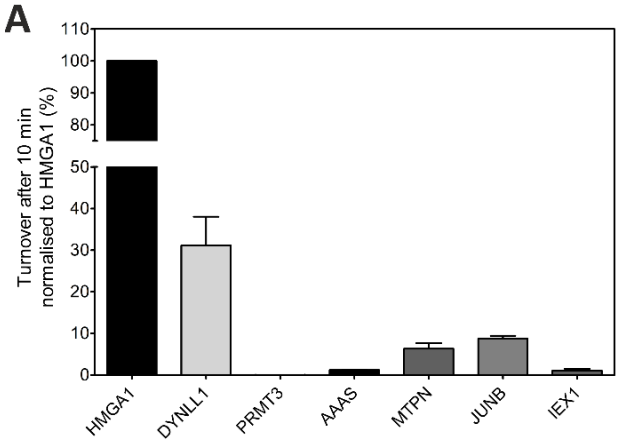
amino acid after the Nt-Cys. The exception was IEX1, whose representative peptide had a basic His residue after the Nt-Cys, and 10.2 % initial NatA activity relative to HMGA1.

Mass spectrometry experiments were next conducted in Oxford to examine NatA activity towards 14-mer peptides, as used to investigate *Hs*ADO activity *in vitro* (Chapter 2,

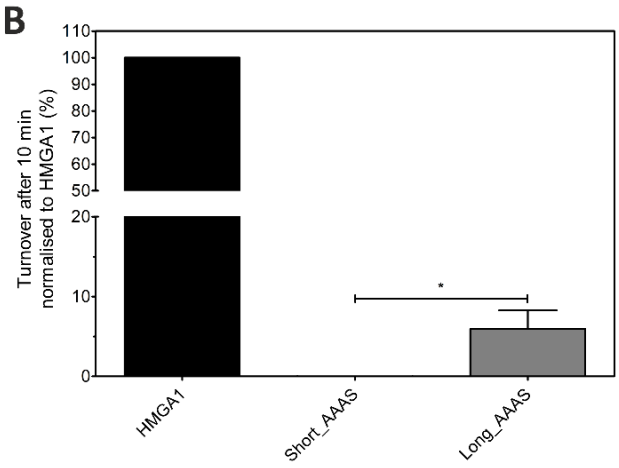
page 41). Human NatA used for these experiments was kindly donated by Dr. Myllykoski. Peptides representing the Met-excised N-termini of the best NatA substrates (> 10 % activity relative to HMGA1) identified by Dr. Myllykoski (Figure 4.2 **A**) were incubated with NatA (0.1 μM) and acetyl CoA (200 μM) at 37 °C for 10 minutes in 50 mM HEPES, 100 mM NaCl, 1 mM EDTA, pH 7.5 supplemented with 1 mM TCEP (Chapter 8, page 178). The reaction was quenched in 1 % (v/v) formic acid (1:10 v/v) and analysed by high throughput MS using a RapidFire RF360 sampling robot connected to a 6530 Accurate-Mass Q-ToF mass spectrometer (Agilent) operated in positive ion mode. Spectra were analysed using MassHunter Qualitative Analysis software (Agilent) and the integrated area under peaks of interest were extracted using RapidFire Integrator (Agilent) (Chapter 8, page 181). Acetylation was evidenced by a mass shift of +42 Da, and was normalised to acetylation of HMGA1 control peptide (kindly provided by Dr. Myllykoski). Some acetylation occurred on peptides incubated with acetyl CoA only (no enzyme controls); therefore, acetylation of peptides in no enzyme control reactions was subtracted from acetylation of peptides incubated with NatA to calculate NatA-catalysed acetylation. The peptides used in mass spectrometry activity assays did not contain the positive C-terminus (RWGRPVGRRRRPVRVYP) that was included in peptides used in radioactivity-based activity assays above, as it was not necessary to facilitate binding to P81 filter paper in this protocol[221].

Activity of NatA with 14-mer peptides observed in MS assays (Figure 4.2 **A**) was much lower than that observed by Dr. Myllykoski in [^{14}C] radioactivity assays (Figure 4.1). For example, 31 % turnover relative to HMGA1 was observed for DYNLL1 peptide, compared with 120 % initial NatA activity relative to HMGA1 measured by Dr. Myllykoski. It was postulated that these differences could be due to the shorter length of peptides used in the MS assays. Thus, activity of NatA with a 14-mer $AAAS_{2\text{-}15}$ peptide (Short_AAAS) was

compared with that of the AAAS peptide used by Dr. Myllykoski (Long_AAAS) in the MS assay described above (Figure 4.2 **B**). While negligible turnover of Short_AAAS was detected after 10 minutes incubation with NatA, 5.94 % turnover relative to HMGA1 was measured for Long_AAAS. This was still lower than the initial NatA activity detected by Dr. Myllykoski with the same Long_AAAS peptide (23.34 % relative to HMGA1); however, this result did demonstrate that peptide length could affect NatA activity. There is also the possibility that peptide charge could be a factor, as the Long_AAAS peptide included the positive C-terminal region needed to facilitate binding to filter paper in the



Name	Peptide Sequence
SESS	SESSSKSRWGRPVGRRRRPVRVYP
DYNLL1	CDRKAVIKNADMSE
PRMT3	CSLASGATGGRGAV
AAAS	CSLGLFPPPPPRGQ
MTPN	CDKEFMWALKNGDL
JUNB	CTKMEQPFYHDDSY
IEX1	CHSRSCHPTMTILQ



Name	Peptide Sequence
SESS	SESSSKS RWGRPVGRRRRPVRVYP
Short_AAAS	CSLGLFPPPPPRGQ
Long_AAAS	CSLGLFP RWGRPVGRRRRPVRVYP

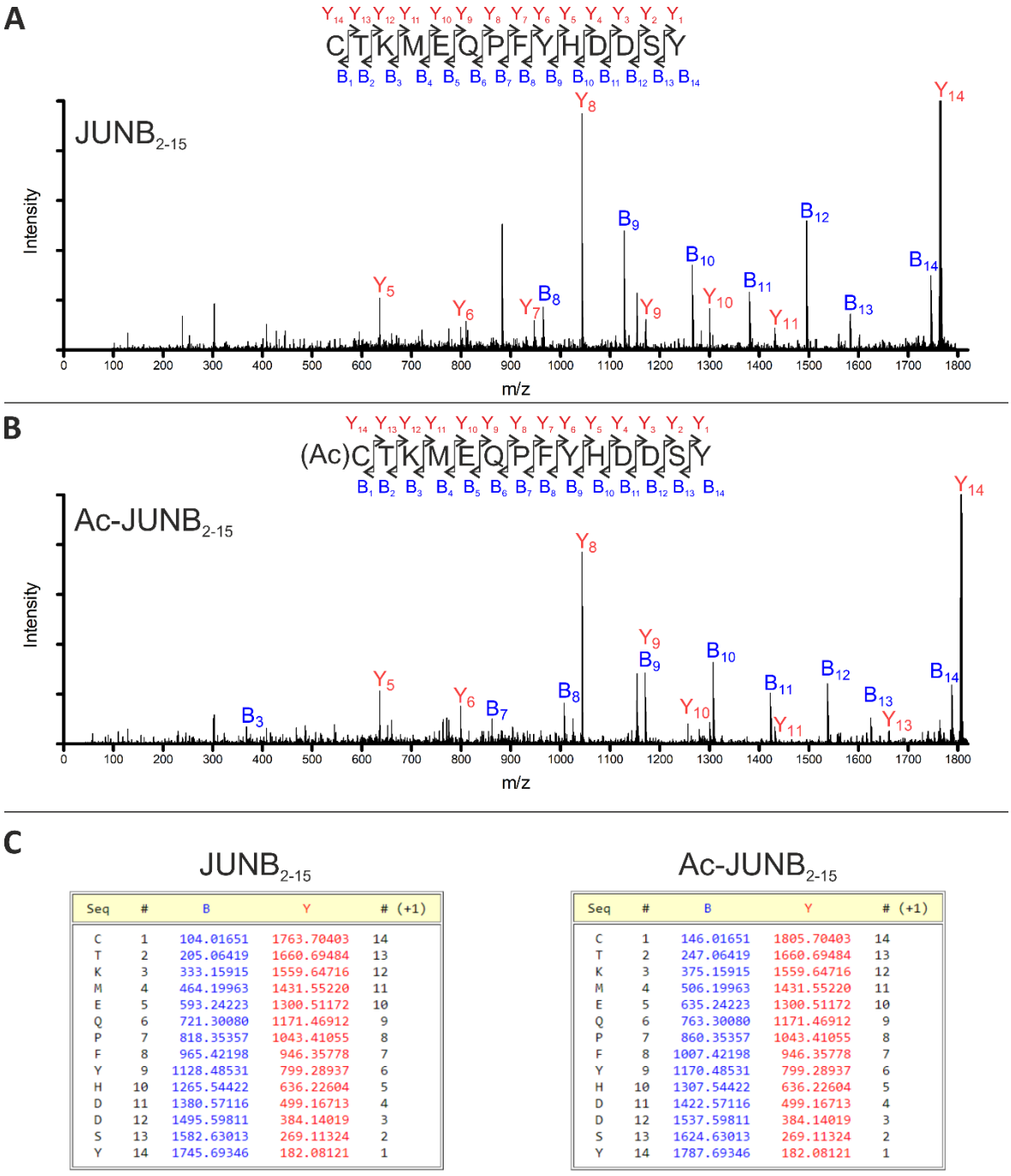
Figure 4.2: NatA-catalysed acetylation of Nt-Cys peptides in MS assays
A: Turnover of 14-mer peptides representing the Met-excised N-termini of the 6 best NatA substrates identified by Dr. Matti Myllykoski (University of Bergen).
B: Turnover of 14-mer AAAS peptide (Short_AAAS), compared to turnover of 24-mer AAAS (Long_AAAS) peptide used by Dr. Matti Myllykoski (University of Bergen). Statistical significance between turnover of Long_AAAS and Short_AAAS determined using unpaired t-test: * ($p < 0.05$).
Turnover of peptides in **A** and **B** was measured by high throughput MS after 10 minutes incubation with NatA at 37 °C (100 µM peptide, 0.1 µM NatA), normalised to turnover of HMGA1 positive control peptide. Error bars show standard deviation. Reactions were carried out in triplicate. Peptide sequences are provided in table next to each graph.

radioactivity assay protocol. Previous studies have shown that co-translational NTA occurs when the nascent chain reaches 25 or more amino acids in length[102,222,223]. It is possible then, that NatA binding is reduced when peptide length is too short.

4.3.2 Acetylation occurs on the Nt-Cys of JUNB$_{2\text{-}15}$

Experiments were next conducted to confirm that the NatA-catalysed acetylation observed *in vitro* was occurring on the N-terminus rather than elsewhere on the peptide. JUNB$_{2\text{-}15}$ was selected as a peptide with which a sufficient level of acetylation was observed to allow detection of acetylated and non-acetylated peptide using tandem mass spectrometry (MS/MS). In MS/MS, peptide ions at specific masses can be targeted for fragmentation, which enables peptides to be inspected one amino acid at a time by mass spectrometry. JUNB$_{2\text{-}15}$ peptide (100 µM) was incubated with NatA (0.1 µM (kindly provided by Dr. Myllykoski)) and acetyl CoA (200 µM) at 37 °C for 60 minutes in 50 mM HEPES, 100 mM NaCl, 1 mM EDTA, pH 7.5 supplemented with 1 mM TCEP (Chapter 8, page 178). The reaction was quenched in 1 % (v/v) formic acid (1:10 v/v) and analysed by MS/MS using an Acquity UPLC system coupled to a Xevo G2-XS Q-ToF mass spectrometer (Waters) operated in positive electrospray mode (Chapter 8, page 181). Comparison of the spectra obtained for JUNB$_{2\text{-}15}$ (Figure 4.3 **A**) and acetylated JUNB$_{2\text{-}15}$ (Ac-JUNB$_{2\text{-}15}$; Figure 4.3 **B**) with the predicted fragment ions for each peptide (Figure 4.3 **C**) confirmed that acetylation was occurring on the Nt-residue. Peptide fragmentation leads to formation of *B* and *Y* ions, resulting from cleavage of the peptide bond between amino acids; B_n ions are formed when the charge is retained on the N-terminal fragment, and Y_n ions are formed when charge is retained on the C-terminal fragment[224,225]. In the MS/MS spectra for Ac-JUNB (Figure 4.3 **B**), the Y_{13} ion, corresponding to loss of the Nt-residue only, was observed at the mass expected if the acetyl modification was present on the Nt-Cys (1661 m/z). Other *B* and *Y* ions observed were consistent with those predicted

for the respective peptides. These MS/MS spectra, together with the activity assay data above (Section 4.3.1), are the first direct evidence of human NatA catalysing Nt-Cys acetylation *in vitro*.



C

$JUNB_{2\text{-}15}$

Seq	#	B	Y	# (+1)
C	1	104.01651	1763.70403	14
T	2	205.06419	1660.69484	13
K	3	333.15915	1559.64716	12
M	4	464.19963	1431.55220	11
E	5	593.24223	1300.51172	10
Q	6	721.30080	1171.46912	9
P	7	818.35357	1043.41055	8
F	8	965.42198	946.35778	7
Y	9	1128.48531	799.28937	6
H	10	1265.54422	636.22604	5
D	11	1380.57116	499.16713	4
D	12	1495.59811	384.14019	3
S	13	1582.63013	269.11324	2
Y	14	1745.69346	182.08121	1

$Ac\text{-}JUNB_{2\text{-}15}$

Seq	#	B	Y	# (+1)
C	1	146.01651	1805.70403	14
T	2	247.06419	1660.69484	13
K	3	375.15915	1559.64716	12
M	4	506.19963	1431.55220	11
E	5	635.24223	1300.51172	10
Q	6	763.30080	1171.46912	9
P	7	860.35357	1043.41055	8
F	8	1007.42198	946.35778	7
Y	9	1170.48531	799.28937	6
H	10	1307.54422	636.22604	5
D	11	1422.57116	499.16713	4
D	12	1537.59811	384.14019	3
S	13	1624.63013	269.11324	2
Y	14	1787.69346	182.08121	1

Figure 4.3: MS/MS of $JUNB_{2\text{-}15}$ peptide after incubation with NatA
A and **B**: MS/MS spectra of sample taken from reaction of $JUNB_{2\text{-}15}$ peptide (100 µM) with NatA (0.1 µM) at 37 °C after 1 hour. Sample was diluted in 1 % (v/v) formic acid (1:10 v/v) and spectra were produced using Xevo G2-XS Q-ToF mass spectrometer, targeting masses of 1763 (**A**) and 1805 (**B**) for fragmentation (collision energy 55-75).
C: Fragment ions for $JUNB_{2\text{-}15}$ and $Ac\text{-}JUNB_{2\text{-}15}$ peptides, calculated using http://db.systemsbiology.net/proteomicsToolkit/FragIonServlet.html

4.3.3 NatA activity with MTPN$_{2\text{-}15}$ varies with identity of residue 3

Initial NatA activity assays with 24-mer peptides (Dr. Myllykoski) suggested that NatA preferred sequences with an acidic or polar residue after the Nt-Cys (Figure 4.1). These substrate preferences were further investigated using one of the best NatA substrate peptides: MTPN. Five variants of the MTPN peptide were investigated by Dr. Myllykoski using the [^{14}C]-radioactivity assay described above (Section 4.3.1). In these variants, the residue after the Nt-Cys (Asp3) was mutated to either another acidic (Glu, E), a polar (Ser, S), a basic (Arg, R), or a hydrophobic (Gly, G or Leu, L) amino acid (Figure 4.4). NatA activity was only retained when Asp3 was mutated to an acidic (D3E) or polar (D3S) amino acid. NatA activity was almost completely lost when Asp3 was mutated to Arg (D3R), Gly (D3G), or Leu (D3L). These data confirmed that NatA has a preference for acidic or polar residues after the Nt-Cys.

Name	CX_6	NatA activity normalised to SESS (%)
MTPN	CDKEFMW	22.6
MPTN-D3E	CEKEFMW	2.4
MTPN-D3S	CSKEFMW	5.1
MTPN-D3R	CRKEFMW	0.5
MTPN-D3L	CLKEFMW	0.3
MTPN-D3G	CGKEFMW	0.2

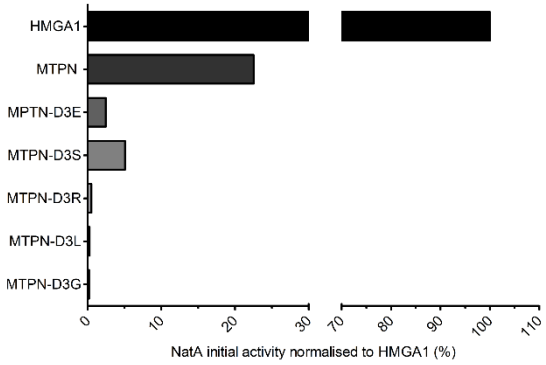


Figure 4.4: NatA-catalysed acetylation of MTPN D3X peptides
NatA-catalysed turnover of MTPN D3X peptides after 5 minutes (200 µM peptide, 0.3 µM NatA, 37 °C); initial activity was normalised to turnover of HMGA1 positive control peptide. Data kindly provided by Dr. Matti Myllykoski (University of Bergen). MTPN-D3X peptide N-termini shown on left where full peptide sequence is [CX_6]-RWGRPVGRRRRPVRVYP

4.3.4 NatA has distinct substrate preferences to ADO

Having established that human NatA was able to acetylate Nt-Cys sequences *in vitro*, *Hs*NatA activity was next compared with *Hs*ADO activity data for the same sequences, to determine whether ADO could also catalyse oxidation of these NatA targets. For the majority of the sequences tested in the NatA activity assays described above, *Hs*ADO

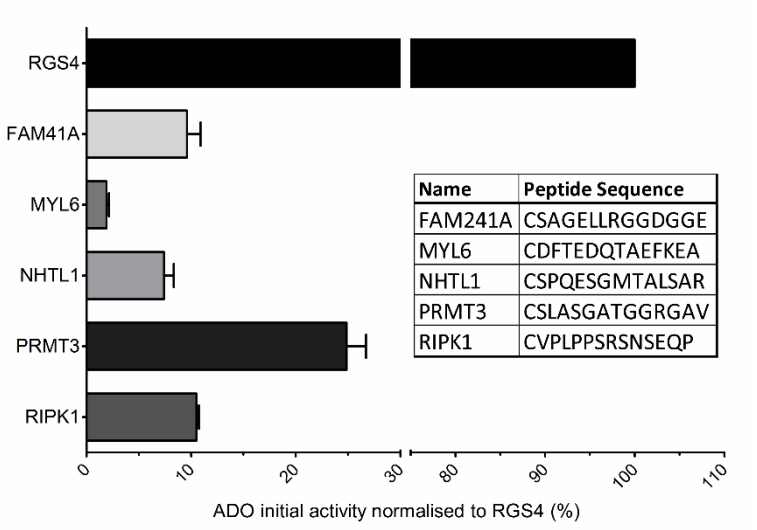


Figure 4.5: ADO-catalysed oxidation of additional Nt-Cys peptides from NatA activity assays
ADO-catalysed turnover of peptides representing N-termini of potential ADO substrates (listed by gene name) after 1 minute (100 μM peptide, 0.1 μM ADO, 37 °C); initial activity was normalised to turnover of $RGS4_{2\text{-}15}$ positive control peptide. Error bars show standard deviation, reactions were carried out in triplicate.

activity data had previously been obtained in Chapter 2 (page 41). However, ADO activity data had not been assessed towards FAM241A, MYL6, NHTL1, RIPK1, PRMT3, and ACTA2 sequences. 14-mer peptides representing the Met-excised N-termini (or caspase-cleaved C-terminal protein fragment of $Ripk1^{219,220}$) of these sequences were purchased from GL Biochem. Unfortunately, a 14-mer peptide representing the N-terminus of ACTA2 could not be synthesised due to lack of solubility. ADO initial activity with these peptides was then measured using the established assay (Chapter 8, page 174) (Figure 4.5). Initial ADO activity with these peptides was < 30 % relative to $RGS4_{2\text{-}15}$, and was lowest for $MYL6_{2\text{-}15}$. Interestingly, PRMT3, whose representative peptide was one of the best NatA substrates identified by Dr. Myllykoski in initial assays, produced the best ADO substrate peptide in this assay (25 % initial ADO activity relative to RGS4).

To compare the relative activities of NatA and ADO with Nt-Cys sequences, initial NatA activity data (Dr. Myllykoski) was plotted against initial ADO activity data (Chapter 2, page 41) (Figure 4.5). This comparison (Figure 4.6) demonstrated that sequences towards which NatA had higher initial activity (> 20 % relative to HMGA1) had low initial ADO

activity (< 25 % relative to RGS4), while the best ADO substrates (> 95 % initial activity relative to RGS4) had little to no initial NatA activity (< 0.5 % relative to HMGA1). Notably, AAAS was one of the best NatA substrates identified in peptide activity assays (Figure 4.1, Figure 4.6) and was also one of the sequences that was not regulated by ADO *in cellula* (Chapter 3, page 66) despite being an ADO substrate *in vitro* (Chapter 2, page 41). This suggested that NatA-catalysed acetylation of AAAS could be linked to the lack of ADO regulation observed with this sequence *in cellula*. However, GPX1, the other sequence that was an ADO substrate *in vitro* but not *in cellula*, was a poor NatA substrate *in vitro* (1.0 % initial activity relative to HMGA1; Figure 4.1).

These data demonstrated that NatA and ADO are both selective amongst Nt-Cys sequences *in vitro*, with distinct substrate preference based on the physicochemical properties of the

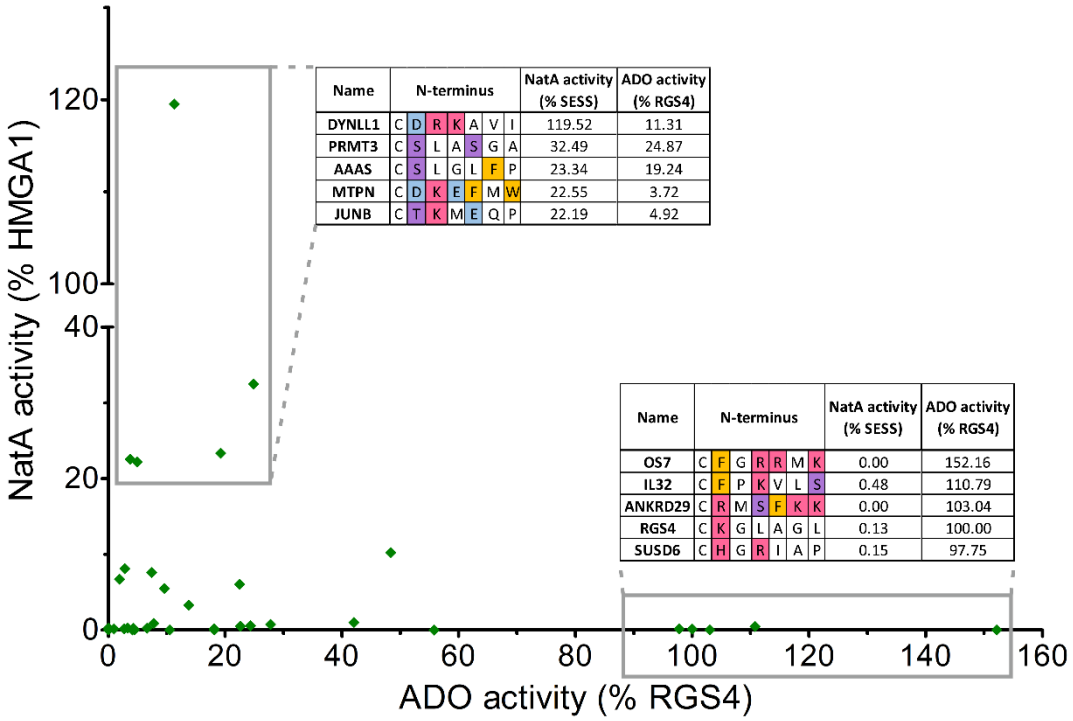


Figure 4.6: ADO initial activity compared to NatA initial activity towards Nt-Cys sequences
Initial NatA activity data for Nt-Cys sequences (Figure 4.1) plotted against initial ADO activity data for same Nt-Cys sequences (Chapter 2 (page 41); Figure 4.5). Names, N-termini (residues 2-8), and activity data are shown for 5 best NatA and ADO substrates. Amino acids are colour coded according to chemical properties: basic (H/K/R) = red; acidic (D/E) = blue; aromatic (F/W) = orange; polar (S/T) = purple. Full combined *in vitro* activity data is available in Appendix 9.

amino acid(s) following the Nt-Cys. NatA displayed a preference for sequences with polar or acidic residues after the Nt-Cys, while ADO displayed a preference for sequences with aromatic or basic residues after the Nt-Cys.

4.4 Nt-Cys acetylation abrogates ADO-catalysed oxidation

NatA-catalysed Nt-acetylation and ADO-catalysed Nt-Cys oxidation are not necessarily mutually exclusive; acetylation occurs on the Nt-amine, while oxidation occurs on the thiol side chain[1,85]. It was next decided to determine whether an ADO substrate, once acetylated, could still be oxidised (Section 4.4), and whether a NatA substrate, once oxidised, could still be acetylated (Section 4.5). As NTA by NatA occurs co-translationally[85], it was first investigated whether ADO could oxidise an Nt-Cys that had already been acetylated.

N-terminally acetylated versions of some of the best *in vitro* ADO substrate peptides (Chapter 2, page 41) were synthesised by solid-phase peptide synthesis (SPPS), using acetic anhydride to acetylate the Nt-amine prior to deprotection and cleavage of the synthesised peptides (Chapter 8, page 170). Peptides (Ac-RGS4$_{2\text{-}15}$, Ac-IL32$_{2\text{-}15}$, Ac-ANKRD29$_{2\text{-}15}$, and Ac-RPL4$_{2\text{-}15}$) were purified by HPLC, their purity was checked by MALDI MS (Appendix 3) and concentration was quantified by NMR. Additional N-terminally acetylated versions of some NatA substrate peptides (Section 4.3) were purchased from GL Biochem (Ac-AAAS$_{2\text{-}15}$, Ac-JUNB$_{2\text{-}15}$, Ac-MTPN$_{2\text{-}15}$).

Acetylated and non-acetylated 14-mer peptides (100 μM) were incubated with ADO (0.1 μM) at 37 °C for 30 minutes (Figure 4.7) using established assay conditions (Chapter 8, page 174). Reactions were quenched in 1 % (v/v) formic acid (1:10 v/v) and analysed by high throughput MS as described previously. After 30 minutes incubation with ADO, RGS4$_{2\text{-}15}$ and IL32$_{2\text{-}15}$ peptides were 100 % oxidised (Figure 4.7). AAAS$_{2\text{-}15}$, JUNB$_{2\text{-}15}$, and MPTN$_{2\text{-}15}$ peptides were also partially oxidised (70 %, 16 %, and 7 % respectively).

Ac-RGS4$_{2\text{-}15}$ peptide was < 1.5 % oxidised after 30 minutes incubation with ADO, Ac-MTPN$_{2\text{-}15}$ was < 0.9 % oxidised, and no oxidation was detected of the other acetylated peptides (Figure 4.7). These data demonstrated that acetylation of the Nt-amine abrogates ADO-catalysed oxidation of good, moderate, and poor ADO substrate peptides. The lack of oxidation of the Ac-AAAS$_{2\text{-}15}$ peptide provided evidence to support the hypothesis that acetylation of the Nt-Cys could be responsible for the AAAS sequence not behaving as an ADO substrate *in cellula* (Chapter 3, page 66).

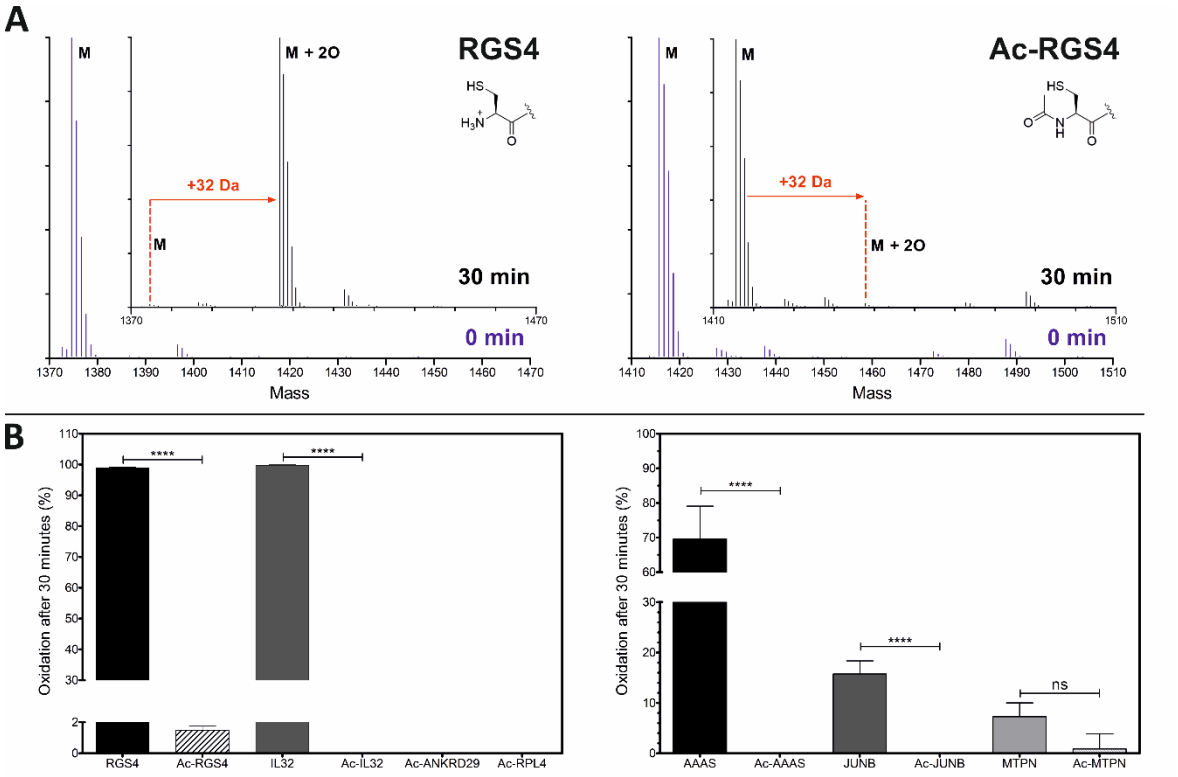


Figure 4.7: ADO-catalysed oxidation of Nt-acetylated peptides
A: Mass spectra of non Nt-acetylated (RGS4; M = 1374 Da) and Nt-acetylated (Ac-RGS4; M = 1416) RGS4$_{2\text{-}15}$ peptides before and after incubation with ADO at 37 °C for 30 minutes (100 µM peptide, 0.1 µM ADO). Deconvoluted spectra were produced with Agilent MassHunter Qualitative Analysis software. Structure of N-terminal cysteine in unoxidised peptide (**M**) shown in top right of spectra.
B: Oxidation of non Nt-acetylated and Nt-acetylated peptides representing the first 14 amino acids of the Met-excised N-termini of proteins of interest after incubation with ADO at 37 °C for 30 minutes (100 µM peptide, 0.1 µM ADO). Error bars show standard deviation, reactions were carried out in triplicate. Statistical significance determined using Two-way ANOVA with Bonferroni's multiple comparison test: ns ($p > 0.05$); * ($p \leq 0.05$); ** ($p \leq 0.01$); *** ($p \leq 0.001$): **** ($p \leq 0.0001$).

4.5 Nt-Cys oxidation abrogates NatA-catalysed acetylation

Having shown that NTA could significantly reduce, and in some cases completely prevent, ADO-catalysed oxidation, it was next investigated whether the opposite was also true, i.e., would oxidation of the Nt-Cys thiol block NatA-catalysed acetylation of the Nt-amine.

Oxidised versions of 14-mer NatA substrate peptides (Ox-DYNLL1$_{2-15}$, Ox-JUNB$_{2-15}$, and Ox-MTPN$_{2-15}$) were produced enzymatically using recombinant ADO. All peptides selected for this investigation had previously been shown to be acetylated in the presence of NatA in the MS assay (Figure 4.2). DYNLL1$_{2-15}$, JUNB$_{2-15}$, and MTPN$_{2-15}$ peptides (1 mM) were incubated with ADO (1 µM (DYNLL1, JUNB) or 10 µM (MTPN)) for 1 hour at 37 °C using established ADO assay conditions (Chapter 8, page 170). ADO was extracted from the reaction mix using a 10,000 Da MWCO ultracentrifugation device and oxidised peptides were obtained from the flow-through. Oxidation was confirmed by high throughput MS as described previously (Appendix 10). Only 80 % oxidation was observed for DYNLL1$_{2-15}$, but due to limited peptide stocks the oxidation reaction could not be repeated at sufficient scale. As > 16 % acetylation of unoxidised DYNLL1$_{2-15}$ was observed previously after 10 minutes incubation with NatA (Figure 4.2; Appendix 10), this level of oxidation was considered sufficient for the investigation provided a long enough timescale was used.

Oxidised and unoxidised peptides (100 µM) were incubated with NatA (0.15 µM; kindly provided by Dr. Myllykoski) and acetyl CoA (100 µM) at 37 °C for 90 minutes in assay buffer (50 mM Na-phosphate, 100 mM NaCl, pH 7.0) supplemented with TCEP (1 mM) (Chapter 8, page 178). Peptides bearing Nt-CysO$_2$ (Ox- peptides) were all significantly less acetylated than their unoxidised counterparts (Figure 4.8). These data suggested that oxidation of the Nt-Cys thiol group abrogates NatA-catalysed acetylation.

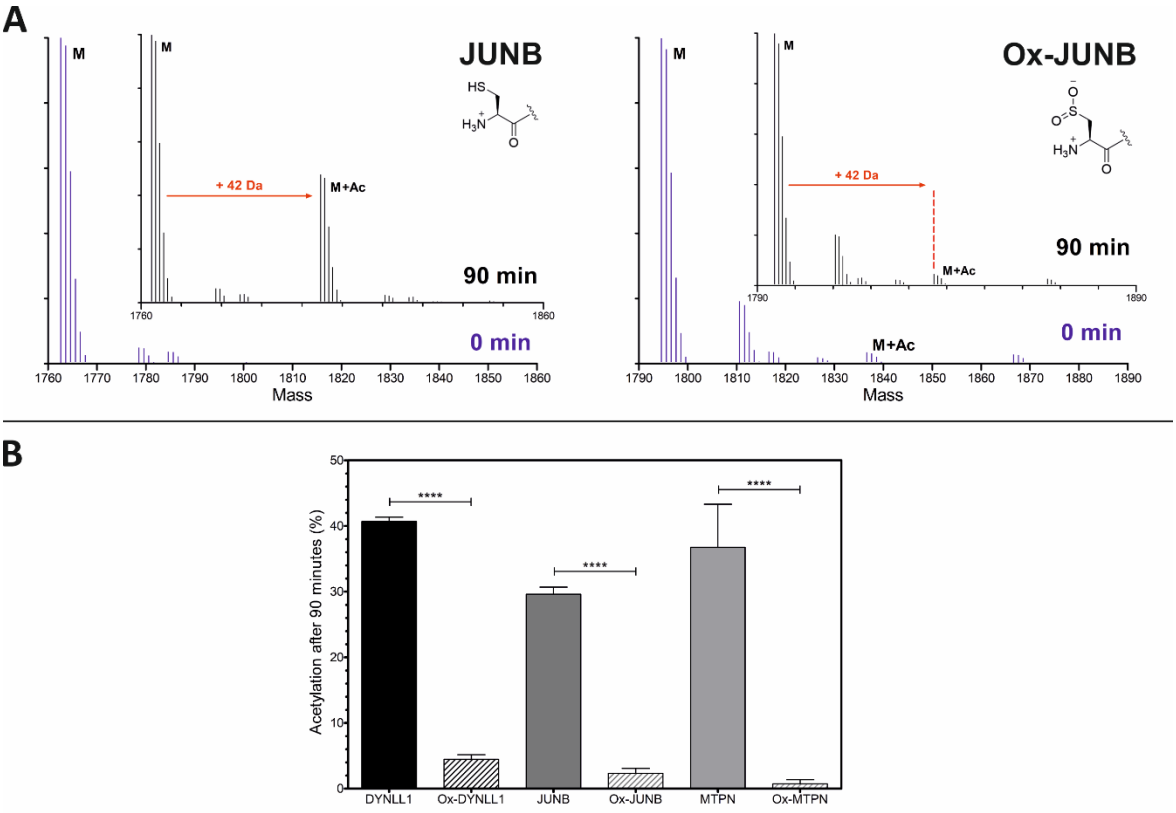


Figure 4.8: NatA-catalysed acetylation of Nt-oxidised peptides
A: Mass spectra of $JUNB_{2\text{-}15}$ peptides with Nt-Cys (JUNB; M = 1763) and Nt-$CysO_2$ (Ox-JUNB; M = 1795) before and after incubation with NatA at 37 °C for 90 minutes (100 µM peptide, 0.15 µM NatA). Deconvoluted spectra were produced with Agilent MassHunter Qualitative Analysis software. Structure of N-terminal cysteine in unacetylated peptide (**M**) shown in top right of spectra.
B: Acetylation of unoxidised and oxidised peptides representing the first 14 amino acids of the Met-excised N-termini of proteins of interest after incubation with ADO at 37 °C for 30 minutes (100 µM peptide, 0.15 µM NatA). Error bars show standard deviation, reactions were carried out in triplicate. Statistical significance determined using 2-way ANOVA with Bonferroni multiple comparison post-test: ns ($p > 0.05$); * ($p \leq 0.05$); ** ($p \leq 0.01$); *** ($p \leq 0.001$): **** ($p \leq 0.0001$).

4.6 Discussion

In this Chapter, the link between NTA and ADO-catalysed Nt-Cys oxidation was investigated using biochemical experiments with Nt-Cys substrate peptides *in vitro*. Results obtained in this Chapter and by Dr. Matti Myllykoski (University of Bergen) demonstrated that human NatA catalysed acetylation of the Nt-Cys of peptides *in vitro* (Section 4.3). NatA was shown to preferentially catalyse acetylation of peptides with an acidic (Asp/Glu) or polar (Ser/Thr) amino acid after the Nt-Cys. This substrate preference was distinct to that of ADO, which has been shown to preferentially catalyse oxidation of peptides with a basic (Arg/His/Lys) or aromatic (Phe/Trp) amino acid after the Nt-Cys.

Notably, AAAS was shown to be one of the best Nt-Cys substrates of NatA *in vitro*. This sequence was of interest due to its lack of regulation by ADO *in cellula* (Chapter 3, page 66), despite *in vitro* data that suggested AAAS should be an ADO substrate (Chapter 2, page 41). However, NatA had low *in vitro* activity with GPX1 peptide, another sequence that was unexpectedly not a substrate of ADO *in cellula*. Addition of the N-termini of AAAS and GPX1 to the N-terminus of EGFP stabilised the resulting fusion proteins compared to the other constructs in a non-ADO-dependent manner in the DFOR assay (Chapter 3, page 66); this was also observed with JUNB, MPTN, and LYPLA2. Three of these five sequences (AAAS, JUNB, and MTPN) were among the best NatA substrates identified *in vitro* (Section 4.3). This suggested that NatA-catalysed acetylation could be a stabilising modification on Nt-Cys bearing proteins. NatA has recently been linked to stabilisation of protein substrates of IAP family E3 ubiquitin ligases by Mueller *et al.*[109]. GPX1 and LYPLA2 N-termini were not good NatA substrates *in vitro*; the stabilising effect of these sequences observed in the DFOR assay could be due to the action of a different Nt-modifying enzyme.

It was postulated that NatA-catalysed acetylation protected AAAS from ADO-catalysed Nt-Cys oxidation *in cellula*. This hypothesis was supported by data that demonstrated that Nt-acetylation was able to block ADO-catalysed oxidation *in vitro* (Section 4.4). Interestingly, Nt-Cys oxidation also abrogated NatA-catalysed acetylation *in vitro* (Section 4.5). NatA-catalysed acetylation is believed to occur co-translationally and to be irreversible[85]. It has not yet been shown exactly when ADO-catalysed oxidation occurs. The most likely *in cellula* application of the above observations is that NTA shields some proteins from ADO-catalysed oxidation, which prevents their degradation via the N-degron pathway. This could affect proteins such as AAAS and PRMT3, whose representative peptides have been shown to be moderate ADO substrates *in vitro*. Further experiments with

NatA and ADO *in cellula* are required to test this hypothesis. These could include stability assays using ADO and NatA knockout cells under normoxic and hypoxic conditions and Nt-proteomics with potential ADO and NatA substrates. Overall, the results presented in this Chapter demonstrated how different Nt-modifications can interact with each other, and the potential impact this can have on protein stability.

5 Peptide and Small Molecule Inhibitors of *Hs*ADO

5.1 Introduction

It is often useful to develop potent and selective inhibitors of an enzyme, for use as chemical probes or as leads for potential therapeutic candidates. There are few known inhibitors of ADO, which is perhaps not surprising given the enzyme was not characterised until 2007[166], and its role as an N-terminal cysteine oxidase (NCO) was not published until 2019[1]. In Chapter 3 (page 66), and in a previous study[1], a non-specific iron chelator, 2,2-dipyridyl, was used to inhibit *Hs*ADO activity in cells. Furthermore, Masson *et al.* also demonstrated that both cysteamine and cysteine were weak inhibitors of *Hs*ADO's NCO activity *in vitro* (with IC_{50} values of 37 and 13 mM, respectively). However, thus far no inhibitors of ADO have been fully characterised.

In recent years, more and more researchers have been exploring the use of peptides as inhibitors over small molecules, especially when targeting protein-protein interactions[226,227]. The reaction being targeted for inhibition in this Chapter involves interaction between ADO and the N-terminus of a target protein (or peptide)[1]. Therefore, peptide inhibitors were considered likely to bind strongly and specifically to the active site of ADO, and thus inhibit its NCO activity. Peptides have many advantages over small molecules: peptides are generally more specific and less toxic. However, many peptides have poor bioavailability, with weak membrane permeability and low *in cellula* stability[226], so small molecules continue to have their use as inhibitors. In some cases of enzyme inhibition small molecule mimics of co-substrates can be used as inhibitors. For example, the PHD enzymes, which regulate the stability of HIF transcription factors[15], utilise 2-oxoglutarate (2OG) as a co-substrate, and are inhibited by 2OG mimics such as N-oxalylglycine (NOG) and its cell-penetrating derivative dimethyl oxalylglycine (DMOG)[228,229]. ADO, however, does not have a small molecule co-substrate, and its small molecule substrate cysteamine is a poor

inhibitor of its NCO activity[1]. ADO has previously been inhibited by an iron chelator (as mentioned above), so small molecule iron chelators are an avenue worth exploring to develop potent *in cellula* inhibitors.

5.2 Chapter Aims

In this Chapter, candidate peptide inhibitors based on the N-terminal (Nt) sequence of known *Hs*ADO substrates (RGS4/5 and IL32)[1] were screened against recombinant *Hs*ADO *in vitro*. Peptides which were found to be inhibitory in initial assays were further characterised to determine their IC_{50} values, and thus assess their potency. Results from experiments with optimal substrate peptides in Chapter 2 (page 47) were used to optimise peptide inhibitors identified in initial screening experiments. Four small molecules were also tested as *Hs*ADO inhibitors *in vitro*, including known *in cellula* inhibitor 2,2-dipyridyl. IC_{50} values were determined for the two of these small molecules, which were selected based on initial assay results.

Finally, experiments were performed to investigate the mode of inhibition of one peptide inhibitor (KH1) and one small molecule inhibitor (2,2-dipyridyl). Although these experiments were not conclusive they provided preliminary data, which gave an indication of the type of inhibition observed in each case. These results could be used to plan future experiments to fully characterise the mode of inhibition of the different inhibitors.

5.3 Peptide inhibitors

Prior to the commencement of this project a selection of potential peptide inhibitors was designed based on the Nt-sequences of three HsADO substrates: RGS4, RGS5, and IL32[1] (Figure 5.1). These were mostly shortened versions (4- or 7-mer) of the 14-mer substrate peptides used in biochemical experiments, with the Nt-Cys substituted with a different proteinogenic or non-proteinogenic amino acid residue. Within the first seven amino acids

of their Met-excised N-termini, RGS4 and RGS5 vary at only one position, residue 8. In RGS4 this is a glycine, while in RGS5 it is an alanine. In the candidate peptide inhibitors based on RGS4/5 a leucine residue was used at this position. Histidine, selenocysteine (Sec),

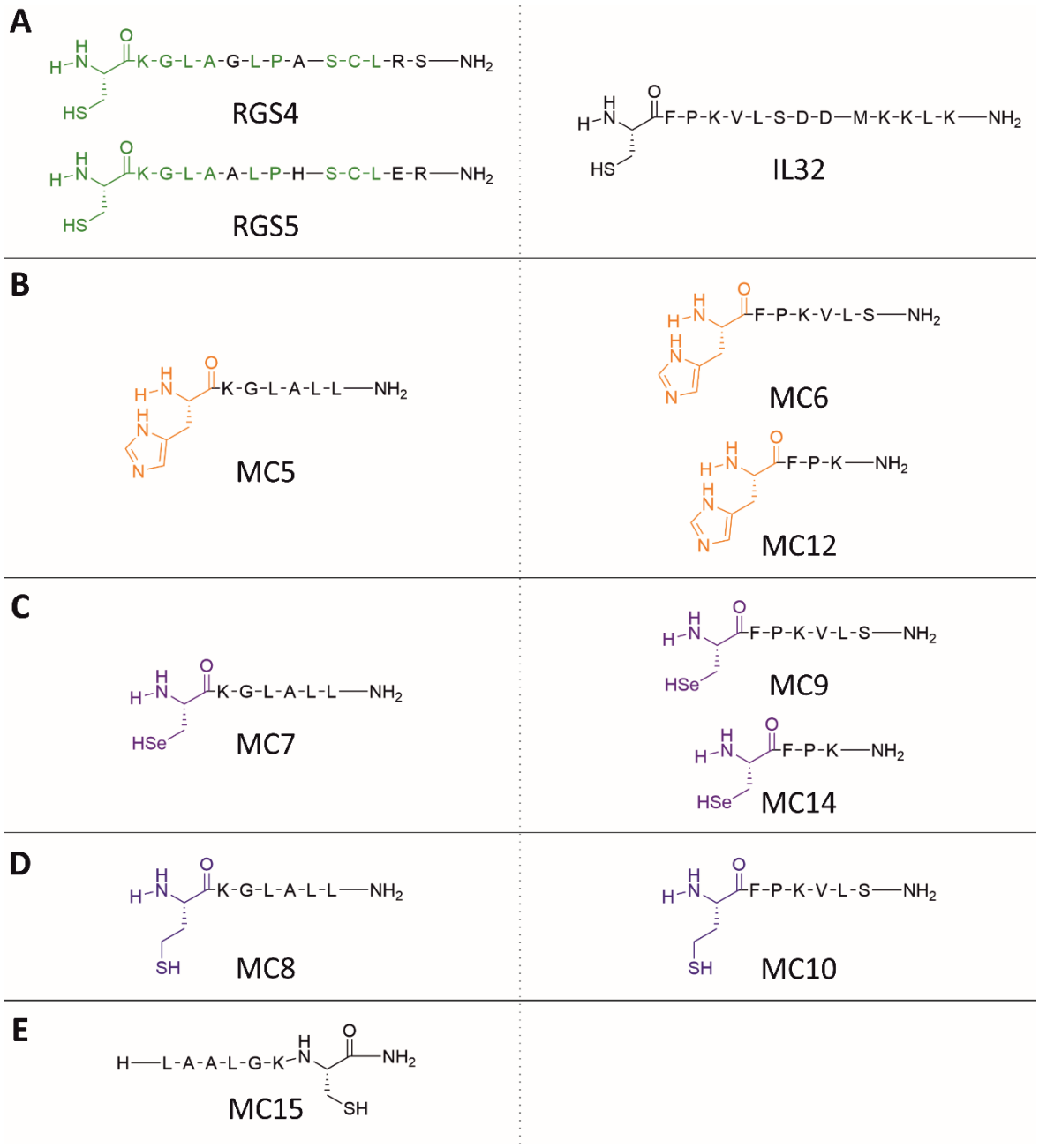


Figure 5.1: Candidate peptide inhibitors selected for screening against *Hs*ADO
A: Sequences of 14-mer peptides representing the Met-excised N-termini of known *Hs*ADO substrates: RGS4, RGS5, and IL32. Residues shared between RGS4 and RGS5 are coloured green.
B: Sequences of candidate inhibitors based on substrate sequences (**A**) with Nt-Cys replaced with Nt-His
C: Sequences of candidate inhibitors based on substrate sequences (**A**) with Nt-Cys replaced with Nt-Sec (selenocysteine)
D: Sequences of candidate inhibitors based on substrate sequences (**A**) with Nt-Cys replaced with Nt-Hcy (homocysteine)
E: Sequence candidate inhibitor based on reversed sequence of $RGS5_{2\text{-}8}$ (**A**)

and homocysteine (Hcy) were used to replace the Nt-Cys in the candidate peptide inhibitors (Figure 5.1 **B**, **C**, and **D**). In the crystal structure of the related enzyme PCO4, the His-tag from a neighbouring PCO4 molecule was observed to bind in the active site; this suggested that an Nt-His would be able to bind to the active site iron in *Hs*ADO. Hcy and Sec both contain a thiol or selenol moiety, so may be able to mimic the Nt-Cys and bind in the active site. An additional candidate peptide inhibitor was designed, in which the $RGS5_{2-8}$ sequence was reversed to place the Cys at the C-terminus instead of the N-terminus (Figure 5.1 **E**). All candidate peptide inhibitors were synthesised by the Peptide Chemistry facility at the Francis Crick Institute. It was thought that altering the sequence slightly could convert a substrate peptide into an inhibitor. Substrate-based peptide inhibitors were expected to still be able to bind to *Hs*ADO, while altering the sequence would prevent *Hs*ADO-catalysed oxidation.

5.3.1 Nt-Sec peptides inhibit *Hs*ADO *in vitro*

Candidate peptide inhibitors (Figure 5.1) were screened against *Hs*ADO *in vitro* (Figure 5.2). Substrate peptides, $RGS4_{2-15}$ or $IL32_{2-15}$, (20 μM) were incubated with recombinant *Hs*ADO (0.05 μM) at 37 °C for 10 minutes in the presence of different peptide inhibitors (200 μM) using standard *Hs*ADO assay buffer and cofactors (Chapter 8, page 174). Recombinant *Hs*ADO used in these initial screening assays was expressed and purified by Dr. Mark White, and had been subjected to one freeze-thaw cycle. Therefore, *Hs*ADO activity was reduced compared to previous *in vitro* assays (Chapter 2, page 35), but was still sufficient for relative assessment of inhibitors. Reactions were quenched in 1 % (v/v) formic acid (1:10 v/v), and analysed by high throughput MS as described previously.

Peptides in which the Nt-Cys had been replaced with His (MC5/6/12) or Hcy (MC8/10) did not inhibit *Hs*ADO activity with either $RGS4_{2\text{-}15}$ or $IL32_{2\text{-}15}$ peptide substrates (Figure 5.2 **A**); nor did MC15, the peptide with a Ct-Cys. However, peptides with an Nt-Sec (MC7/9/14) were able to significantly inhibit *Hs*ADO activity with both substrates (Figure 5.2 **B**). *Hs*ADO activity with $IL32_{2\text{-}15}$ in the presence of MC9 could not be assessed due to overlap of MC9 peaks with peaks for oxidised $IL32_{2\text{-}15}$ in the mass spectra. Analysis of the mass spectra of MC7 and MC9 incubated with *Hs*ADO and $RGS4_{2\text{-}15}$ for 10 minutes revealed that Nt-Sec peptides were not oxidised under the assay conditions used

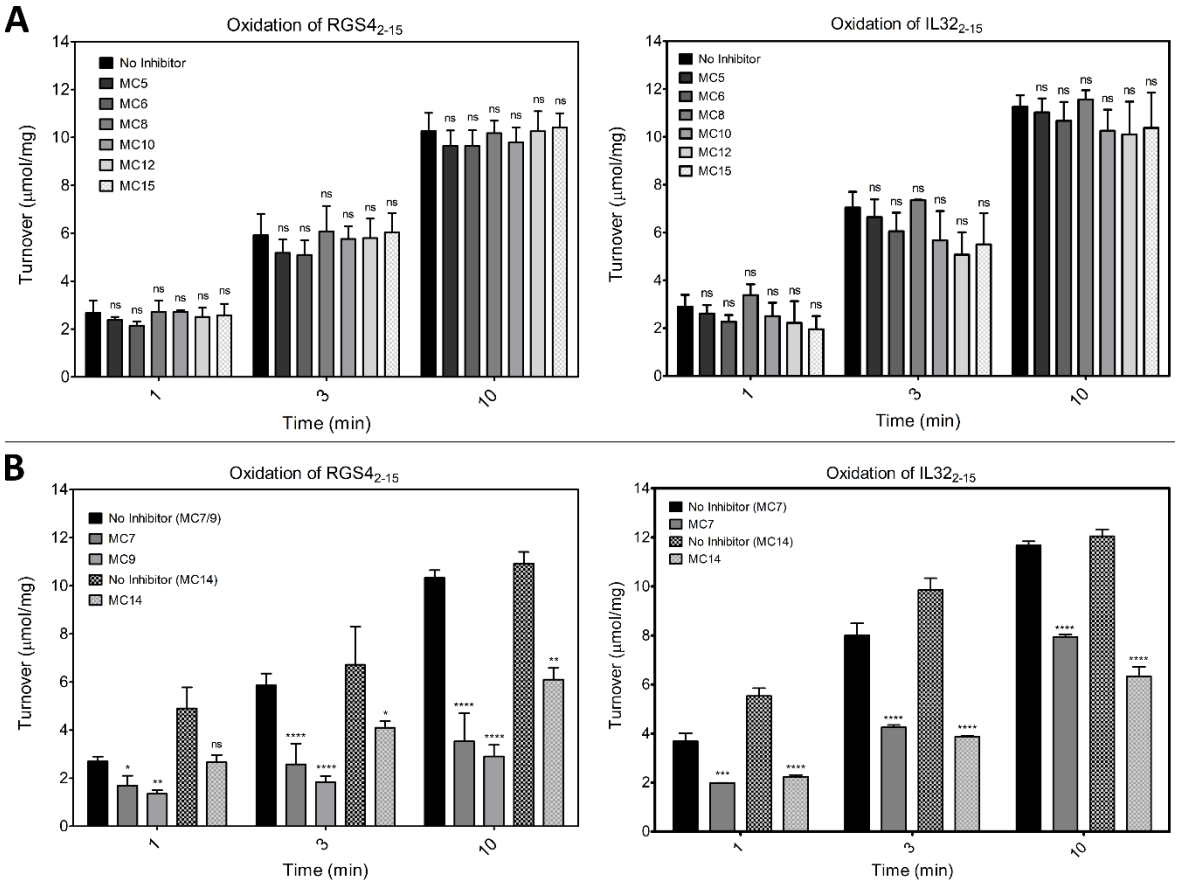


Figure 5.2: *Hs*ADO-catalysed turnover of 14-mer substrate peptides in the presence of potential peptidomimetic inhibitors

$RGS4_{2\text{-}15}$ and $IL32_{2\text{-}15}$ peptides (20 µM) were incubated with and without potential inhibitors (200 µM) in the presence of *Hs*ADO (0.05 µM) at 37 °C for 10 minutes.

A: MC5, MC6, MC8, MC10, MC12, and MC15

B: MC7, MC9, and MC14

Reactions were carried out in triplicate. Error bars show standard deviation. Statistical significance was determined using 2-way ANOVA with Bonferroni multiple comparisons post-test (comparing turnover in the presence of each inhibitor to the relevant no inhibitor control): ns ($p > 0.05$); * ($p \leq 0.05$); ** ($p \leq 0.01$); *** ($p \leq 0.001$): **** ($p \leq 0.0001$).

For sequences of substrate peptides and inhibitors see Figure 5.1

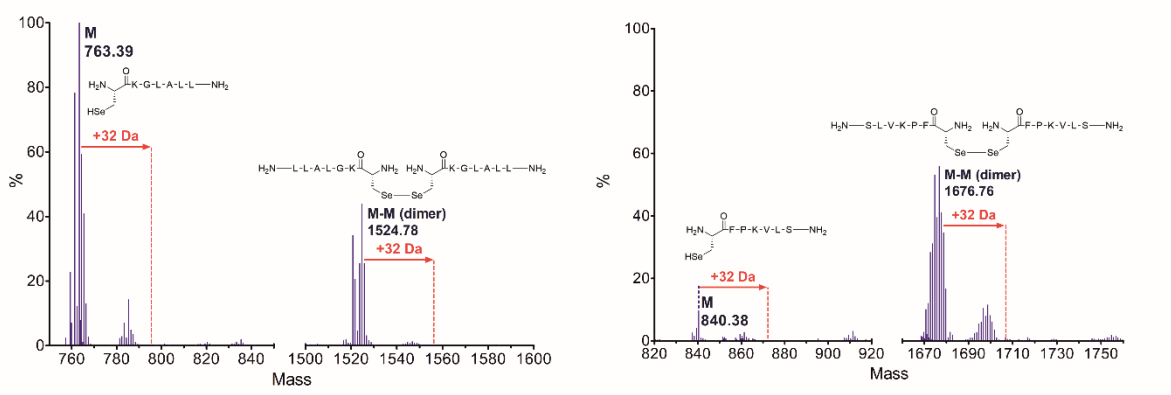


Figure 5.3: Spectra showing no oxidation of inhibitory peptides
Deconvoluted mass spectra of MC7 and MC9 peptides (200 µM) after incubation with *Hs*ADO (0.05 µM) at 37 °C for 10 minutes (in the presence of $RGS4_{2\text{-}15}$ (20 µM)). MC7: MW = 762.87 Da. MC9: MW = 838.92 Da. Presence of a +32 Da peak in the spectra would be indicative of oxidation of peptide. Levels of dimer in deconvoluted spectra may not be completely accurate due to overlap of peaks for single peptide and dimer in raw spectra (e.g., for MC7 single peptide $[M^+] = 764.39$, and for dimer $[M^{2+}] = 763.39$). Spectra are representative of at least 3 repeats.

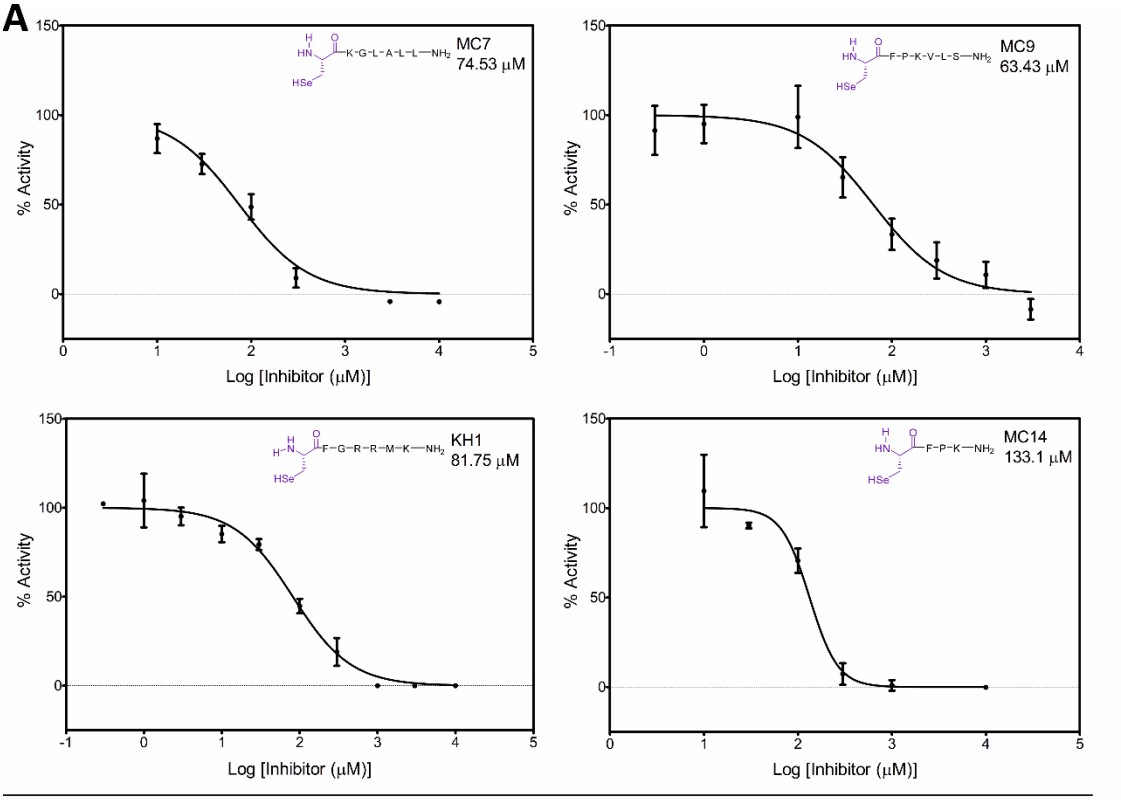
(Figure 5.3). However, significant levels of dimerised peptide were observed for both MC7 and MC9. This dimerisation most likely occurred via a diselenide bond between Nt-Sec residues, despite the presence of a reducing agent (TCEP) in the assay. It appeared that while the concentration of TCEP used in biochemical *Hs*ADO assays (5 mM) was sufficient to reduce disulfide bonds between substrate peptide molecules, it was not enough to reduce the diselenide bonds between inhibitor peptide molecules. It was not possible to accurately quantify the levels of dimer using the established assay due to overlap of peaks in mass spectra of the single peptide and dimer. Future work could involve a more detailed investigation to determine whether dimer formation and inhibitor activity are linked.

5.3.2 Nt-Sec peptide inhibitors have similar potencies

In order to assess the potency of the Nt-Sec peptide inhibitors identified in the initial screening assay (Figure 5.2 **B**), *in vitro* experiments were performed to determine the IC_{50} values of MC7, MC9, and MC14 (Chapter 8, page 174). The IC_{50} value of an inhibitor is the concentration needed to inhibit enzyme activity by 50 %. It was theorised that the potency of the peptide inhibitor could be linked to *in vitro* activity of *Hs*ADO with the substrate peptide used to design the inhibitor, i.e., that a better substrate would produce a better

inhibitor when its Nt-Cys is replaced with Nt-Sec. To test this hypothesis another Nt-Sec peptide inhibitor was designed using the first seven amino acids of the most rapidly oxidised optimum substrate peptide, OS7 (Chapter 2, page 47): Sec-FGRRMK (Figure 5.4 **A**). This peptide (henceforth KH1), was synthesised using a Liberty Blue™ automated microwave peptide synthesiser (CEM) by standard Fmoc-protected SPPS (Chapter 8, page 170). Selenocysteine was incorporated using Fmoc-Sec(pMeOBzl)-OH (Sigma Aldrich), according to protocol kindly provided by Stefania Federico (Francis Crick Institute). KH1 peptide was purified by HPLC, then purity was checked by MALDI MS (Appendix 10) and peptide stock concentration was quantified by NMR. The recombinant *Hs*ADO used for IC_{50} determination was from the same batch as for the initial screening assays (Section 5.3.1), but had not been subjected to a freeze-thaw cycle and was therefore more active (Chapter 2, page 35). To calculate the IC_{50} values for Nt-Sec peptides, $RGS4_{2-15}$ substrate peptide (20 μM) was incubated with recombinant *Hs*ADO (0.05 μM) at 37 °C for 1 minute in the presence of different concentrations of peptide inhibitor (0.3 - 1000 μM) using standard *Hs*ADO assay buffer and cofactors (Chapter 8, page 174). Reactions were quenched in 1 % (v/v) formic acid as before (1:10 v/v), and analysed by ultrahigh-performance liquid chromatography (UPLC) MS using an Acquity UPLC system coupled to a Xevo G2-XS Q-ToF mass spectrometer (Waters) operated in positive electrospray mode (Chapter 8, page 181). Turnover of $RGS4_{2-15}$ peptide was quantified using integrated area under peaks for modified and unmodified peptide, which was extracted from chromatograms using MassLynx V4.1 (Waters).

All 7-mer Nt-Sec peptide inhibitors (MC7, MC9, and KH1) were found to have IC_{50} values of < 100 μM (Figure 5.4), ranging between 63 and 82 μM. Confidence intervals calculated for these IC_{50} values indicated that they were within error of each other, and so not significantly different (Figure 5.4 **B**). The IC_{50} value for MC14, the 4-mer Nt-Sec peptide,



B

Peptide	Sequence							IC_{50} (µM)	95 % Confidence Interval - IC_{50} (µM)
MC9	Sec	F	P	K	V	L	S	63	43.39 to 92.74
MC7	Sec	K	G	L	A	L	L	75	55.43 to 100.2
KH1	Sec	F	G	R	R	M	K	82	68.04 to 98.23
MC14	Sec	F	P	K				133	104.8 to 169.1

Figure 5.4: IC_{50} analysis of Nt-Sec peptide inhibitors
A: IC_{50} analysis of Nt-Sec peptide inhibitors using 20 µM $RGS4_{2-15}$ substrate and 0.05 µM *Hs*ADO. Peptide names, sequences and IC_{50} values given in top right of each graph.
B: IC_{50} values and 95 % confidence interval values for each peptide inhibitor
IC_{50} analysis was performed using GraphPad Prism 5: log(inhibitor) vs. normalized response -- Variable slope

was slightly higher than for the 7-mers: 133 µM. This suggested that peptide length, rather than peptide sequence, had the greater influence on inhibitor potency. However, it is predicted that an Nt-Sec peptide based on the sequence of a poor substrate (e.g., ASNS (Chapter 2, page 41)) would not inhibit, as it is expected that this peptide would not have as many favourable binding interactions with *Hs*ADO. Future work could involve testing of this hypothesis and of Nt-Sec peptides at different lengths to determine the optimum peptide length for inhibition.

5.4 Small molecule inhibitors

Although peptide-based inhibitors tend to have better potency and lower toxicity *in cellula*, small molecule inhibitors generally have greater bioavailability[226,227]. Small molecules are more cell permeable than peptides due to their reduced molecular size, allowing them to more easily target intracellular molecules[230]. This also means small molecule drugs are more likely to be orally administered, which is safer for patients. It was therefore useful to investigate potential small molecule inhibitors of *Hs*ADO. *Hs*ADO activity had previously been shown to be iron-dependent[166], and to be inhibited by a non-specific iron chelator *in cellula* (2,2-dipyridyl)[1]. Thus, in this Section, three iron-chelating small molecules were assessed as *Hs*ADO inhibitors (Figure 5.5 **A**, **B**, and **D**): 2,2-dipyridyl; DP44mT; and DFO. As mentioned, 2,2-dipyridyl was known to inhibit *Hs*ADO *in cellula*, but its *in vitro* potency had not yet been assessed. DP44mT is a 2,2-dipyridyl derived iron chelator that was designed by Yuan *et al.*, and shown to be a more active iron chelator than 2,2-dipyridyl and other 2,2-dipyridyl derivatives[231]. DP44mT was also shown to be a more potent chelator than DFO (deferoxamine)[231], which did not inhibit *Hs*ADO *in cellula* when used in cellular assays by Masson *et al.*[1]. So, it was expected that 2,2-dipyridyl and DP44mT would inhibit *Hs*ADO *in vitro*, but that DFO would not. DP44mT was also predicted to be the most potent *Hs*ADO inhibitor. In addition to the iron chelators, DMOG, a nonspecific inhibitor of

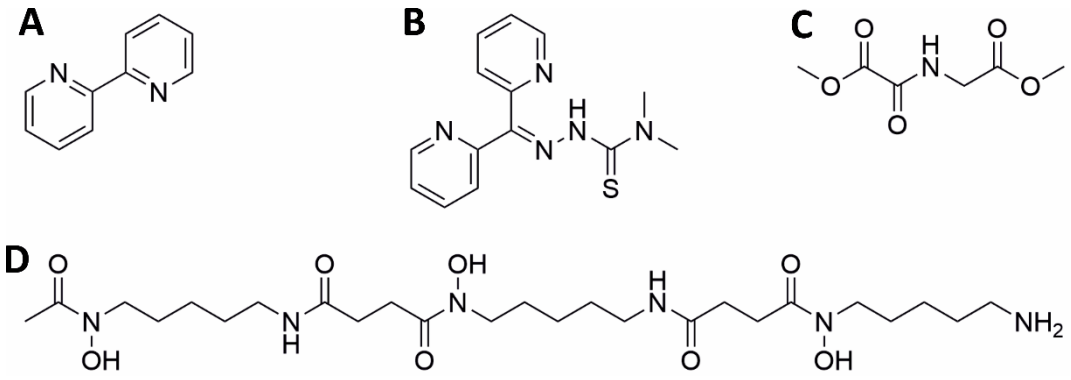


Figure 5.5: Structures of small molecules screened against *Hs*ADO
A: 2,2-dipyridyl **B**: 2,2-dipyridylketone-4,4,-dimethyl-3-thiosemicarbazone (DP44mT)
C: dimethyloxalylglycine (DMOG) **D**: Deferoxamine (DFO)

2OG-dependent dioxygenases, was included in initial small molecule inhibition assays (Figure 5.5 **C**). 2OG-dependent dioxygenases are a diverse range of iron-binding enzymes, including the prolyl hydroxylase (PHD) enzymes that regulated the stability of hypoxia inducible factor transcription factors as part of the hypoxia response[232]. DMOG had previously been shown not to inhibit *Hs*ADO *in cellula*[1], so was predicted to also not inhibit *in vitro.*

5.4.1 2,2-dipyridyl and DP44mT inhibit *Hs*ADO *in vitro*

Initial inhibitor screening assays with small molecules were performed using the same assay protocol and recombinant *Hs*ADO enzyme as described above (Section 5.3.1), with the exception that no iron(II) was added as a cofactor (Chapter 8, page 174). Three of the four molecules tested were iron chelators; addition of iron(II) to the assay would have sequestered these small molecules and prevented their interaction with the enzyme. *Hs*ADO activity was assessed with $RGS4_{2\text{-}15}$ peptide substrate using high throughput MS as before. DP44mT was expected to be significantly more potent than 2,2-dipyridyl, so was used at a lower concentration in all assays (10 × less than 2,2-dipyridyl). As 2,2-dipyridyl and DP44mT stocks required DMSO for solubility, a suitable concentration of DMSO was used as a 'no inhibitor' control for addition of these small molecules (Figure 5.6 **A**, **C**, and **D**). Assay buffer was used as the control for addition of DFO and DMOG. At concentrations of 200 μM inhibitor (20 μM for DP44mT), 20 μM substrate, and 0.05 μM *Hs*ADO, none of the small molecules significantly inhibited *Hs*ADO activity (Figure 5.6 **A**).

It was expected that 2,2-dipyridyl would inhibit *Hs*ADO in *vitro*, based on *in cellula* experiments (Chapter 3, page 66; Masson *et al.*[1]). Therefore, the assay was repeated increasing the concentration of inhibitor to 1 mM for 2,2-dipyridyl, DFO, and DMOG, and to 100 μM for DP44mT. Under these conditions DFO and DMOG still did not inhibit ADO (Figure 5.6 **B**). To account for the increased volume of 2,2-dipyridyl and DP44mT stocks

added to achieve final assay concentrations of 1 and 0.1 mM, an increased volume of DMSO was also added; this increased the final assay DMSO concentration to 10 % (v/v). This level of DMSO (10 % v/v) was found to inhibit *Hs*ADO activity *in vitro* (Figure 5.6 **C**). However, at 2 % (v/v) DMSO, *Hs*ADO activity was still comparable to activity in the 'no inhibitor' (or positive controls) where assay buffer was added (Figure 5.6 **C**). Based on these observations, in future experiments the DMSO concentration was kept below 2 %. When the assay with 2,2-dipyridyl and DP44mT was repeated using more concentrated inhibitor stocks and a DMSO concentration of 1 % (v/v), both small molecules significantly inhibited

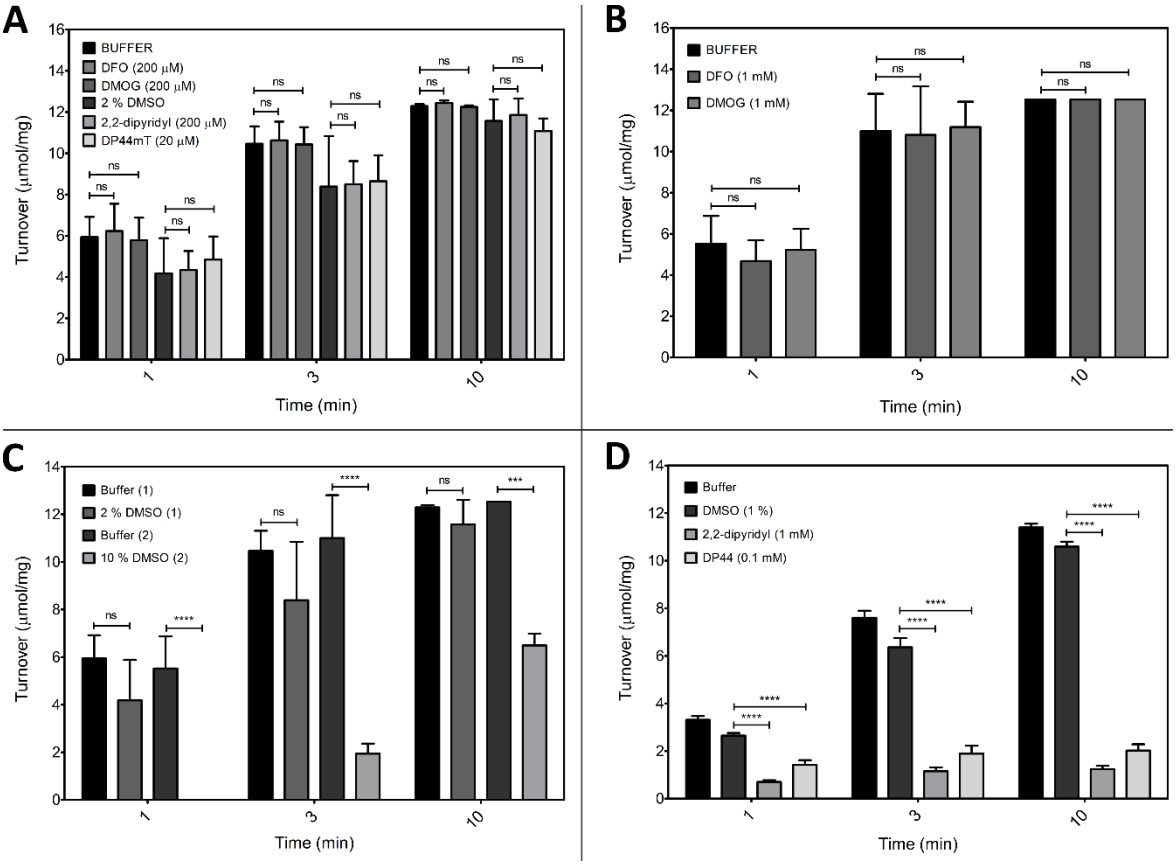


Figure 5.6: *Hs*ADO-catalysed turnover of RGS4$_{2\text{-}15}$ substrate peptide in the presence of potential small molecule inhibitors

Turnover of RGS4$_{2\text{-}15}$ (20 μM) incubated with *Hs*ADO (0.05 μM) at 37 °C for 10 minutes:

A: In presence of DFO, DMOG, or 2,2-dipyridyl (200 μM), or DP44mT (20 μM). 2 % DMSO used as a control for addition of 2,2-dipyridyl and DP44mT, and 'BUFFER' as a control for addition of DFO and DMOG.

B: In presence of DFO or DMOG (1 mM)

C: In presence of varying concentrations of DMSO.

D: In presence of 2,2-dipyridyl (1mM) or DP44mT (100 μM). 1 % DMSO used as a control.

Reactions were carried out in triplicate. Error bars show standard deviation. Statistical significance determined using 2-way ANOVA with Bonferroni multiple comparisons post-test (comparing turnover in the presence of each inhibitor to the relevant no inhibitor control): ns ($p > 0.05$); * ($p \leq 0.05$); ** ($p \leq 0.01$); *** ($p \leq 0.001$): **** ($p \leq 0.0001$).

*Hs*ADO activity (Figure 5.6 **D**). As predicted, 100 µM DP44mT was able to produce a similar level of *Hs*ADO inhibition as 1 mM 2,2-dipyridyl.

5.4.2 DP44mT is more potent than 2,2-dipyridyl *in vitro*

The potency of DP44mT and 2,2-dipyridyl inhibitors *in vitro* was assessed through IC_{50} determination as before (Section 5.3.2). Briefly, $RGS4_{2-15}$ substrate peptide (20 µM) was incubated with recombinant *Hs*ADO (0.05 µM) at 37 °C for 1 minute in the presence of different concentrations of peptide inhibitor (1 – 3000 µM 2,2-dipyridyl; 0.1 – 1000 µM DP44mT) using standard assay buffer and cofactors (Chapter 8, page 174). As above (Section 5.4.1), iron(II) cofactor was excluded from the assays. Reactions were quenched in 1 % (v/v) formic acid as before (1:10 v/v), and were analysed by high throughput MS. DP44mT was found to have similar potency to the Nt-Sec peptide inhibitors, which had IC_{50} values between 63 and 133 µM (Figure 5.4); the IC_{50} value for DP44mT was 125 µM (Figure 5.7). As predicted, 2,2-dipyridyl was a less potent inhibitor, with an IC_{50} of 522 µM (Figure 5.7).

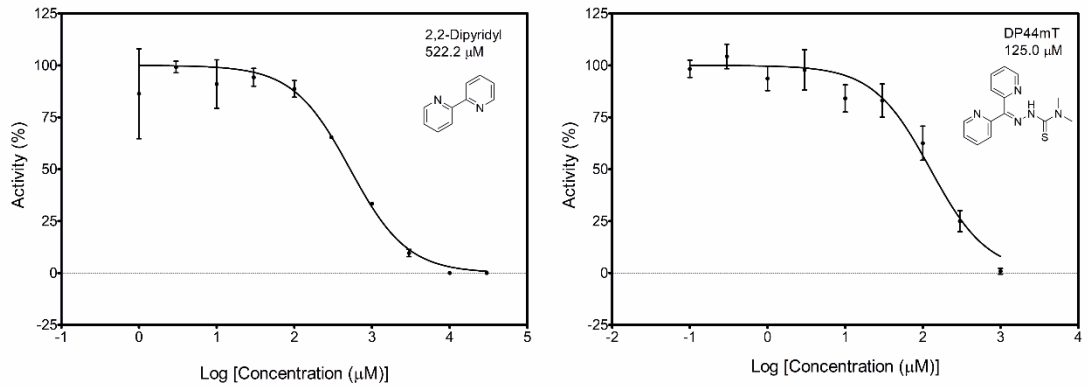


Figure 5.7: IC_{50} analysis of 2,2-dipyridyl and DP44mT
IC_{50} analysis of small molecule inhibitors using 20 µM $RGS4_{2-15}$ substrate and 0.05 µM *Hs*ADO. Small molecule names, structures, and IC_{50} values are shown in the top right of each graph. IC_{50} analysis was performed using GraphPad Prism 5: log(inhibitor) vs. normalized response - variable slope. 95 % confidence intervals (µM): 2,2-dipyridyl = 405.1 to 673.1; DP44mT = 99.45 to 157.0.

5.5 Investigation of modes of *Hs*ADO inhibition

Having established that both peptides and small molecules could inhibit *Hs*ADO *in vitro* with a similar level of potency (Figure 5.5 and Figure 5.7), experiments were performed to investigate whether these inhibitors acted via different modes of inhibition. Reversible enzyme inhibition can be described as either competitive, non-competitive, or uncompetitive[233]. Kinetic experiments can be performed that examine the effect of inhibitor concentration on Michaelis-Menten parameters[202,203] (discussed in Chapter 2, page 54): comparing inhibitor concentration with maximum rate (V_{max}) and Michaelis-Menten constant (K_M) can be used to determine whether an inhibitor's mode of action is competitive, non-competitive, or uncompetitive[233,234].

In competitive inhibition, substrate and inhibitor compete for binding at the same site; in this mode of inhibition V_{max} is unaffected by increased inhibitor concentration, while the apparent K_M is increased. In non-competitive inhibition, inhibitor and substrate can bind to the enzyme simultaneously at different sites; inhibitor can therefore bind to either free enzyme or the enzyme-substrate complex. In this mode of inhibition V_{max} is reduced by increased inhibitor concentration and the apparent K_M is unchanged. In uncompetitive inhibition, inhibitor can only bind to the enzyme-substrate complex; in this mode both V_{max} and apparent K_M are reduced by increased inhibitor concentration.

The effect of different modes of inhibition on enzyme reaction kinetics can be visualised using double reciprocal, or Lineweaver-Burk, plots (Figure 5.8). According to the Michaelis-Menten equation (discussed in Chapter 2, page 54), where reaction velocity (v) is defined in terms of maximum velocity (V_{max}) and the Michaelis-Menten constant (K_M):

$$v = \frac{V_{max}[S]}{K_M + [S]}$$

(Equation 2.2)

The double-reciprocal gives:

$$\frac{1}{v} = \frac{K_M}{V_{max}} \times \frac{1}{[S]} + \frac{1}{V_{max}}$$

(Equation 5.1)

Therefore, in a Lineweaver-Burk plot, where $\frac{1}{[S]}$ (or 1/S) is plotted against $\frac{1}{v}$ (or 1/V), the slope gives $\frac{K_M}{V_{max}}$, the y-intercept gives $\frac{1}{V_{max}}$, and the x-intercept gives $-\frac{1}{K_M}$. The effects of each type of inhibitor on K_M and V_{max} discussed above thus lead to changes in the slope and intercepts in Lineweaver-Burk plots for enzyme-catalysed reactions (Figure 5.8).

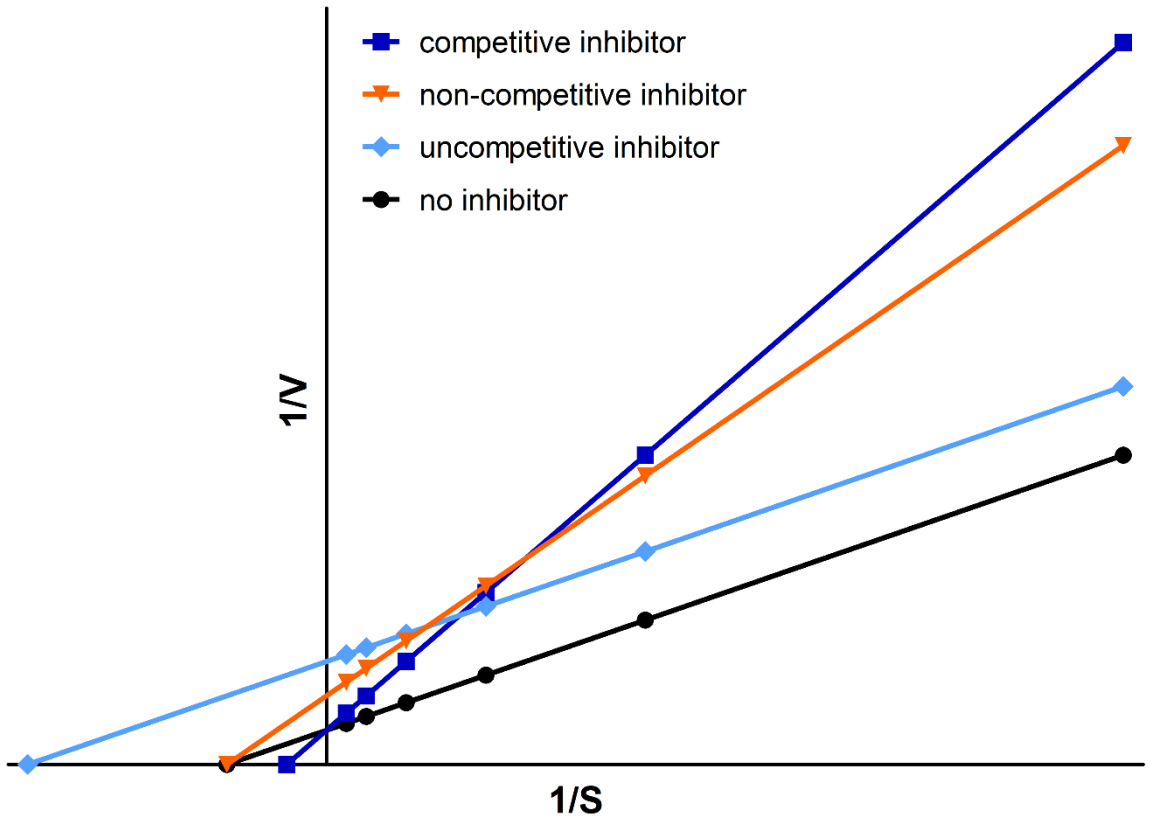


Figure 5.8: Example Lineweaver-Burk plots for different modes of inhibition
Lineweaver-Burk (or double reciprocal) plots representing the effect of adding different types of inhibitors to an enzyme-catalysed reaction.

5.5.1 KH1 mode of inhibition

In order to estimate the mode of inhibition of the Nt-Sec peptide inhibitors, the K_M and V_{max} were determined for $RGS4_{2-15}$ with *Hs*ADO in the presence of increasing concentrations of KH1 inhibitor (from no inhibitor to > 2 × IC_{50}). $RGS4_{2-15}$ peptide, at concentrations between 25 μM and 1 mM, was incubated with *Hs*ADO (0.05 μM) for 1 minute using the established assay conditions (Chapter 8, page 174). This experiment was repeated for each concentration of KH1. The reaction was sampled every 20 seconds and analysed by high throughput MS as described previously. Peptide turnover (μmol mg^{-1}) was calculated at each time point and used to calculate reaction velocity as specific activity (μmol mg^{-1} min^{-1}). When KH1 inhibitor was included in the reactions, the reaction velocity was not linear over all timepoints, therefore only data for between 0 and 40 seconds was used for determination of kinetic parameters. Turnover of $RGS4_{2-15}$ peptide also varied significantly between replicates, which led to exclusion of some data points and large standard deviations (Appendix 11 (A)). The results presented here should therefore be considered as preliminary data that can give an initial indication of mode of inhibition and provide information for planning of future experiments. Analysis could not be repeated in this Thesis due to limited peptide stocks.

Lineweaver-Burk plots were generated for each inhibitor concentration using linear regression to analyse 1/S against 1/V (Figure 5.9 **A**), and K_M and V_{max} were plotted against inhibitor concentration (Figure 5.9 **B**), in order to visualise the influence of KH1 inhibitor on the *Hs*ADO-catalysed reaction. Preliminary results suggested that KH1 may have different modes of inhibition at different concentrations. At lower concentrations (< 50 μM KH1), both K_M and V_{max} appeared to be lower, as would be expected for an uncompetitive inhibitor. The changes seen in the Lineweaver-Burk plots also resembled those of an uncompetitive inhibitor (Figure 5.8). At the highest concentration

(200 µM KH1), K_M was higher than without inhibitor, while the value for V_{max} was within error of that obtained without inhibitor (20.48 ± 0.52 µmol mg^{-1} min^{-1} versus 19.76 ± 0.25 µmol mg^{-1} min^{-1}). These data, together with the appearance of the Lineweaver-Burk plots, indicated that KH1 was acting as a competitive inhibitor at 200 µM.

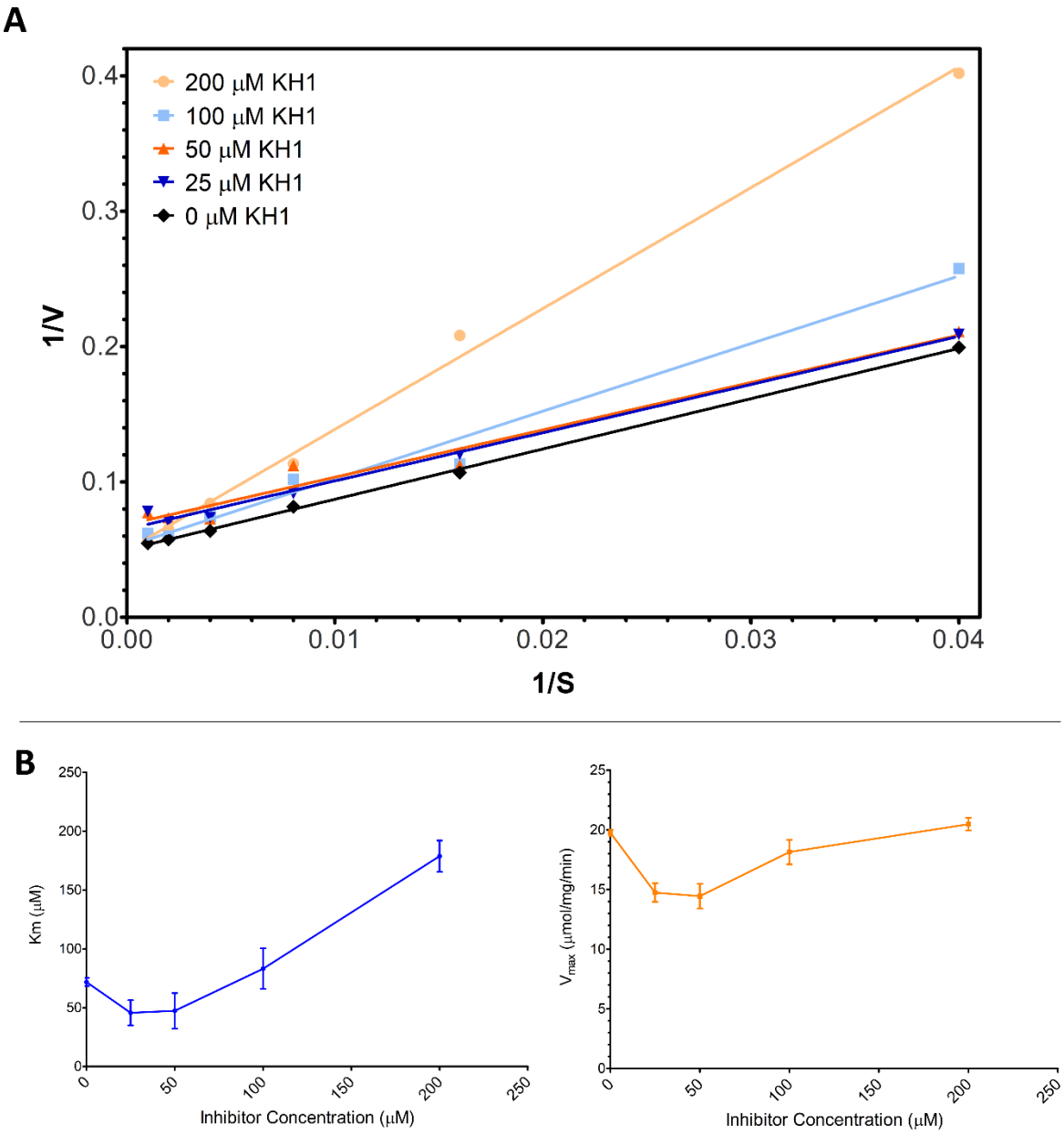


Figure 5.9: Examining mode of inhibition of KH1 peptide inhibitor
A: Lineweaver-Burk plots of 1/V (V = reaction velocity in µmol/mg/min) against 1/S (S = [RGS4$_{2\text{-}15}$] in µM) for addition of increasing concentrations of KH1 inhibitor to reaction of *Hs*ADO (0.05 µM) with RGS4$_{2\text{-}15}$ at 37 °C. Lines show linear regression of data.
B: Kinetic parameters K_M and V_{max} calculated using GraphPad Prism Michaelis-Menten analysis of reaction of *Hs*ADO (0.05 µM) with RGS4$_{2\text{-}15}$ at 37 °C with increasing KH1 inhibitor concentration.

As the KH1 peptide was designed using a *Hs*ADO substrate sequence, the most likely location of KH1 binding was considered to be the active site. Therefore, KH1 would be expected to act via competitive inhibition, as suggested by the differences in kinetic parameters observed between the experiments with no inhibitor and 200 μM KH1. However, the amount of error observed in the raw data mean it is not possible to draw firm conclusions from these experiments. Repetition is required and conditions should be adjusted such that steady state conditions are ensured across all timepoints before KH1's mode of inhibition can be defined.

5.5.2 2,2-dipyridyl mode of inhibition

Similar experiments as described above (Section 5.5.1) were performed to estimate the mode of inhibition of 2,2-dipyridyl. 2,2-dipyridyl was selected for these experiments over the more potent small molecule inhibitor DP44mT due to relative purchasing costs. As described above the K_M and V_{max} were determined for $RGS4_{2\text{-}15}$ with *Hs*ADO in the presence of increasing concentrations of 2,2-dipyridyl inhibitor (from no inhibitor to approximately $2 \times IC_{50}$); an equivalent volume of DMSO was added to 'no inhibitor' reactions as the 2,2-dipyridyl stocks were in DMSO (Chapter 8, page 174). In these experiments, reaction velocity was linear across all timepoints within the 1 minute timescale, although large standard deviations were still observed for some data points (Appendix 11 (B)).

Lineweaver-Burk plots were generated for each inhibitor concentration using linear regression to plot lines of 1/S against 1/V (Figure 5.10 **A**), and K_M and V_{max} plotted against inhibitor concentration (Figure 5.10 **B**), in order to visualise the influence of 2,2-dipyridyl on the *Hs*ADO-catalysed reaction. Linear regressions in the Lineweaver-Burk plots did not fit the data as closely as hoped, and large errors were obtained when calculating K_M and V_{max} using GraphPad Prism Michaelis-Menten analysis. This experiment should therefore be repeated to confirm the observations made here; this was not possible within this Thesis due

to time constraints. Preliminary data shown here suggested that 2,2-dipyridyl could be a competitive inhibitor; K_M increased steadily with increasing inhibitor, while V_{max} was within error of the no inhibitor value for most concentrations of inhibitor. This would fit the hypothesis that 2,2-dipyridyl is able to bind to iron in the active site of *Hs*ADO, where spectroscopic experiments by Wang *et al.*[185] suggest that it would compete for binding with the thiol group of the Nt-Cys of the substrate.

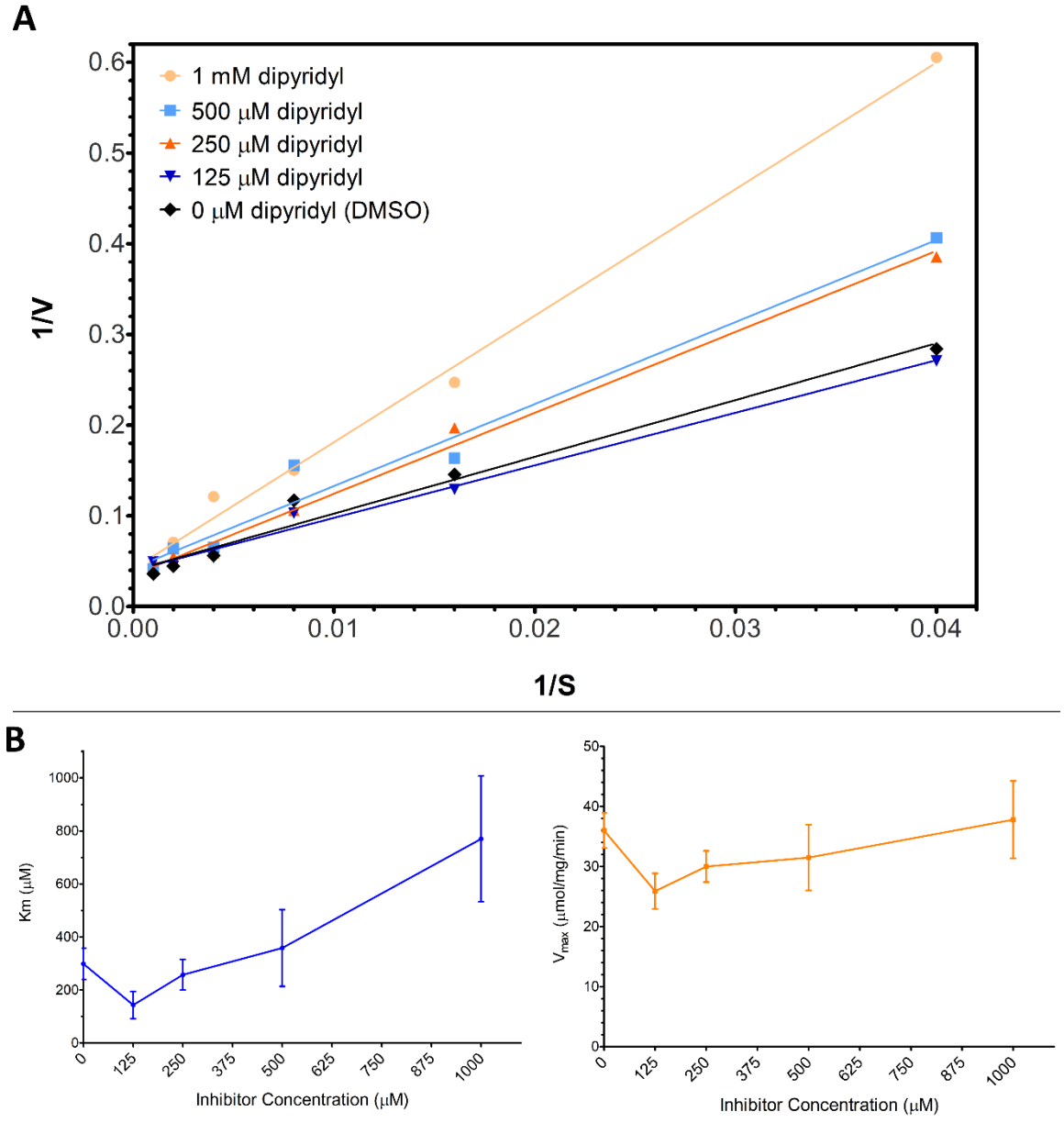


Figure 5.10: Examining mode of inhibition of 2,2-dipyridyl small molecule inhibitor
A: Lineweaver-Burk plots of 1/V (V = reaction velocity in µmol/mg/min) against 1/S (S = [$RGS4_{2-15}$] in µM) for addition of increasing concentrations of 2,2-dipyridyl inhibitor to reaction of *Hs*ADO (0.05 µM) with $RGS4_{2-15}$ at 37 °C. Lines show linear regression of data.
B: Kinetic parameters K_M and V_{max} calculated using GraphPad Prism Michaelis-Menten analysis of reaction of *Hs*ADO (0.05 µM) with $RGS4_{2-15}$ at 37 °C with increasing 2,2-dipyridyl inhibitor concentration.

5.6 Discussion

In this Chapter, a selection of substrate-like peptides was screened as inhibitors of *Hs*ADO *in vitro* (Section 5.3). It was found that substituting the Nt-Cys with an Nt-Sec in a 4- or 7-mer peptide based on the Nt-sequence of RGS4/5, IL32, or OS7 produced an *Hs*ADO inhibitor with an IC_{50} value of around 100 μM. However, replacing the Nt-Cys with an Nt-His or Nt-Hcy, or reversing the sequence to generate a Ct-Cys, in substrate-like peptides did not produce inhibitory peptides in the conditions tested. Interestingly, Nt-Sec peptides were not oxidised in the presence of *Hs*ADO, but did form dimers via a suspected diselenide bond under standard assay conditions. It was not determined whether this dimerisation was linked to inhibitor potency, as accurate quantification of dimer levels was not possible using the established assay. However, future work on these Nt-Sec peptides should include investigation of this phenomenon. The *in cellula* activity of these peptide inhibitors should also be assessed.

Preliminary kinetic experiments with KH1 suggested that at concentrations above its IC_{50} value KH1 acted as a competitive inhibitor (Section 5.5). However, at lower concentrations there was some evidence that KH1 behaved like an uncompetitive inhibitor. There was significant variability within the raw data used to produce Lineweaver-Burk plots and determine kinetic parameters (Appendix 11). Therefore, it was not possible to define KH1's mode of inhibition; further experiments and optimisation of conditions are required.

Four small molecules were also tested as potential *Hs*ADO inhibitors *in vitro* (Section 5.4). Three were iron chelators: DFO; 2,2-dipyridyl; and DP44mT. Of these three iron chelators, only 2,2-dipyridyl and DP44mT were able to inhibit *Hs*ADO *in vitro*. DFO, and another small molecule, DMOG (a 2OG-dioxygenase inhibitor), did not inhibit *Hs*ADO even when used at a 50:1 inhibitor:substrate ratio. DFO, 2,2-dipyridyl, and DP44mT all have different modes of iron binding. DFO binds iron in a hexadentate manner and has a strong affinity

for Fe(III), while 2,2-dipyridyl and DP44mT have softer donor atoms so should bind more strongly to Fe(II), which has a lower charge density[235]. Furthermore, 2,2-dipyridyl binds in a bidentate manner, and DP44mT in a tridentate manner[236]. In the active site of ADO, iron is bound to three histidine residues, leaving three available binding sites in an octahedral coordination (to which waters are bound in the crystal structures of *Hs*ADO and *Mm*ADO)[5,6]. It is suggested that 2,2-dipyridyl and DP44mT were able to bind to these available sites on the active site iron, but that DFO was not able to outcompete the His triad to access the three additional sites that it would require to bind iron in a hexadentate fashion. Data from initial mode of inhibition experiments with 2,2-dipyridyl supported this theory (Section 5.5); 2,2-dipyridyl behaved like a competitive inhibitor, which would bind *Hs*ADO at the same site as the substrate, i.e., at the active site iron. As with the peptide mode of inhibition experiments, repetition and optimisation are required to confirm this. It is predicted that DP44mT would have the same mode of inhibition as 2,2-dipyridyl, although experiments would be required to validate this prediction; these would be best performed following optimisation of experiment conditions using 2,2-dipyridyl.

Overall, results from this Chapter provided *in vitro* data that supported *in cellula* results obtained with 2,2-dipyridyl, DFO, and DMOG. Novel peptide inhibitors of *Hs*ADO were also discovered, which are expected to be more specific for *Hs*ADO than the iron chelating small molecule inhibitors. The bioavailability of these peptides is yet to be assessed, and future experiments should address this. However, if these peptides are found to have, or can be optimised to have, good bioavailability they could prove useful as chemical probes *in cellula* or potential therapeutic candidates. Substitution of unnatural amino acids (such as N-methylated or D-amino acids) into peptide inhibitors can help to improve membrane permeability[237–239] and should be investigated if further development of these peptide inhibitors is conducted. Inhibitors could also be useful in crystal trials as they could stabilise

*Hs*ADO; in the case of the peptide inhibitors co-crystallisation trials could produce a peptide-bound structure that would provide information on the manner in which *Hs*ADO binds peptide substrates. Experiments of this kind were trialled in Chapter 6 (page 134).

6 Structural Investigation of ADO

6.1 Introduction

Although the enzyme cysteamine dioxygenase (ADO) has been postulated since the 1960s[165], and was first characterised in 2007[166], at the commencement of this project no crystal structures had been published. In contrast, many structures had been published of cysteine dioxygenase (CDO), the only other mammalian thiol dioxygenase[156,157,159,162,187,240–244]. Furthermore, in 2019 researchers in the Flashman group obtained crystal structures of two related plant cysteine oxidase enzymes (*At*PCO4 and *At*PCO5), which were published in 2020[66], closely followed by structures of *At*PCO2/4/5 published by Chen *et al.*[67].

ADO, CDO, and the PCOs are members of the cupin superfamily of proteins[245–247]; cupin proteins share common structural features, including a double stranded β helix (DSBH) and an active site within the cupin fold, as seen in structures of *At*PCO4 (PDB: 6S7E)[66] and *Hs*CDO (PDB: 2IC1)[157] (Figure 6.1). CDO and PCO proteins share a common 3-His iron-binding motif that differs from that of other cupin proteins (Figure 6.1 **C-D**)[4,247]. However, key differences exist between the structures of mammalian CDOs and *Arabidopsis thaliana* PCOs, which may reflect their roles of oxidising small molecule thiols and protein N-terminal cysteines, respectively[4]. The active site cavity of CDO is less open than that of the PCOs, with a C-terminal loop sitting across the entrance to the cavity; in the PCOs the C-terminal loop is positioned to create a wide tunnel, which potentially allows access of polypeptide substrates to the active site (Figure 6.1 **A-B**). Additionally, in PCOs a potential substrate binding (hairpin) loop has been identified near the opening to the active site cavity (Figure 6.1 **A**) that has not been observed in CDO structures[4,66]. Finally, a Ser-His-Tyr (S-H-Y) motif adjacent to the active site iron is present in CDO and other small

molecule thiol dioxygenases, but is not conserved in the *At*PCOs, where it is replaced with an Ile-Asp-Leu (I-D-L) motif (Figure 6.1 **C-D**).

ADO has been shown to be capable of acting as both a small molecule thiol dioxygenase (smTDO) and an N-terminal cysteine oxidase (NCO)[1,4,166]. An alignment[248,249] of *Hs*ADO

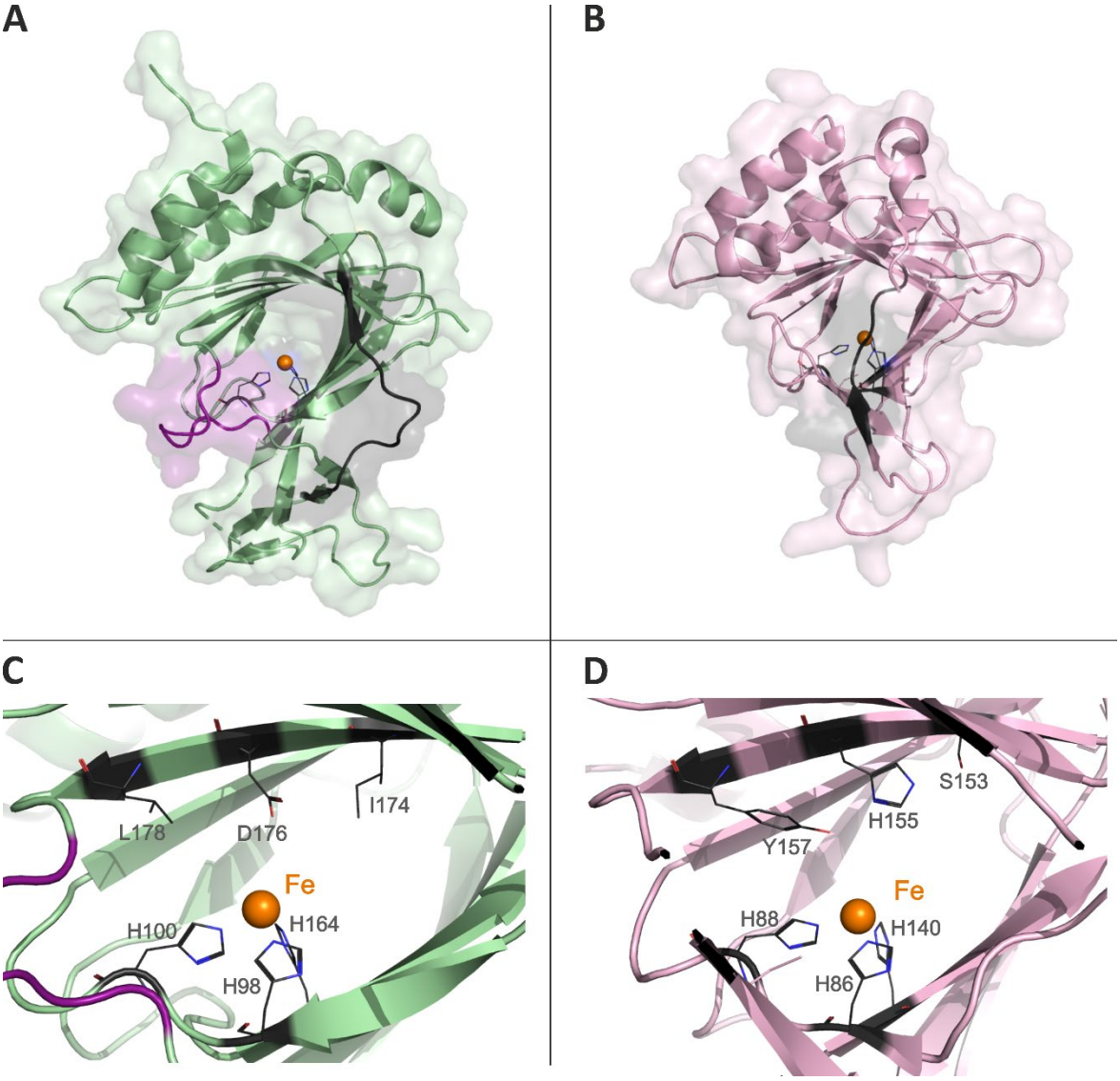


Figure 6.1: Views of crystal structures of *At*PCO4 and CDO

A: Crystal structure of *At*PCO4 (PDB: 6S7E), showing core DSBH motif flanked by loops and helices; C-terminal loop (cartoon, gray), potential substrate binding loop (cartoon, purple), 3 active site His residues (lines, gray), and coordinating active site Fe (sphere, orange)
B: Crystal structure of *Hs*CDO (PDB: 2IC1), showing core DSBH motif flanked by loops and helices; C-terminal loop (cartoon, gray), 3 active site His residues (lines, gray), and coordinating active site Fe (sphere, orange)
C: Close-up view of *At*PCO4 (**A**) active site showing Ile-Asp-Leu motif (lines, gray), 3 Fe-binding His residues (lines, gray), and coordinating active site Fe (sphere, orange)
D: Close-up view of *Hs*CDO (**B**) active site showing Ser-His-Tyr motif (lines, gray), 3 Fe-binding His residues (lines, gray), and coordinating active site Fe (sphere, orange)

with *At*PCO4 and *Hs*CDO shows that the three iron-binding His residues are conserved across all three enzymes, but that the CDO S-H-Y motif residue is replaced by a Phe-Asp-Leu (F-D-L) motif in *Hs*ADO, which more closely resembles the I-D-L motif in *At*PCO4 (Figure 6.2). *Hs*ADO also shares 26.6 % sequence identity (39.2 % similarity) with *At*PCO4, compared to 15.8 % sequence identity (26.5 % similarity) with *Hs*CDO. Therefore, based on function and sequence similarities it was expected that the structure of ADO would more closely resemble that of the *At*PCOs than the other mammalian thiol dioxygenase CDO. Two structures of ADO were published during the course of this Thesis[5,6], which will be discussed in this Chapter.

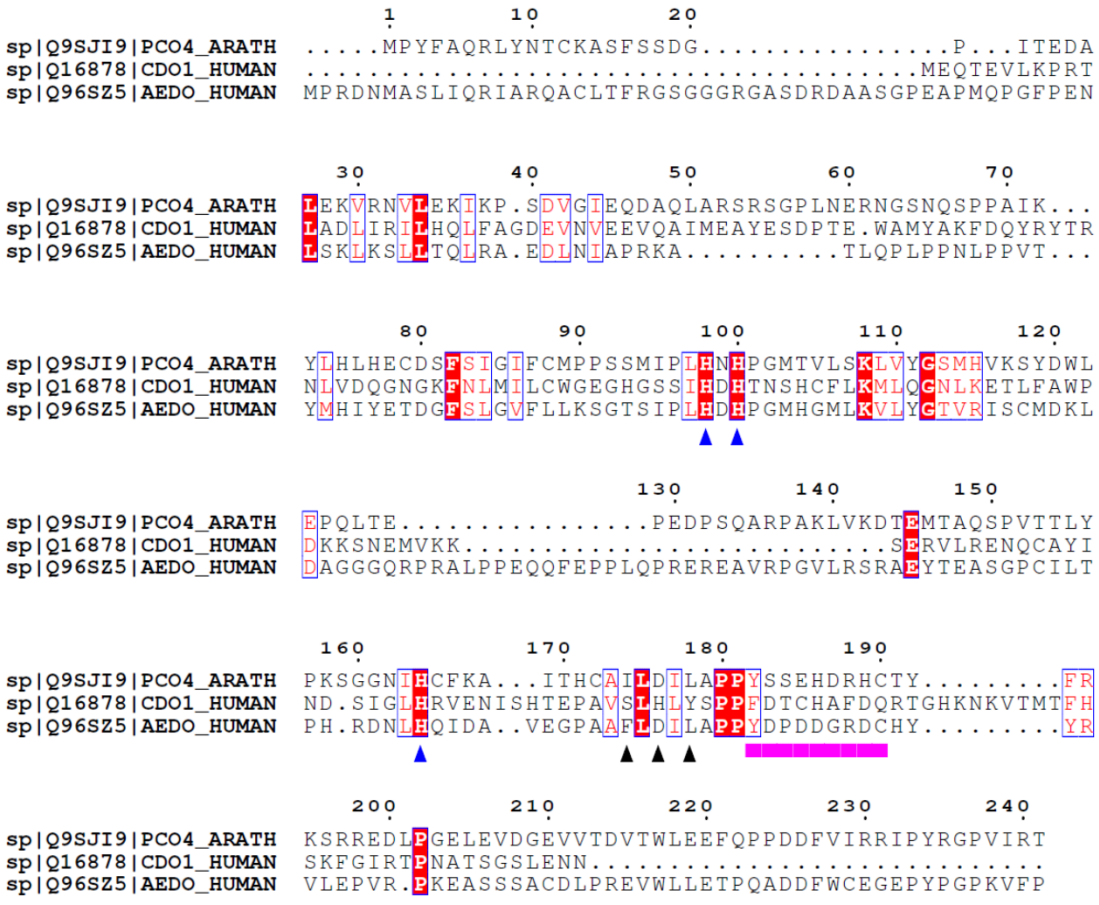


Figure 6.2: Alignment of *At*PCO4, *Hs*CDO, and *Hs*ADO protein sequences
CLUSTAL O(1.2.4) multiple sequence alignment of *At*PCO4 (Q9SJI9), *Hs*CDO (Q16878), and *Hs*ADO (Q96SZ5) rendered with ESPript 3.0 server. Fe-binding His residues marked with blue triangles, I-D-L motif (*At*PCO4) residues marked with black triangle, and potential substrate binding loop (PCO4, Figure 6.1) marked with magenta bar.

6.2 Chapter Aims

Work described in this Chapter originally aimed to generate crystals of *Hs*ADO in order to determine its structure. Alongside crystallisation trials, an *in silico* structure of *Hs*ADO was produced with MODELLER[250] using the structure of *At*PCO4[66] as a template. Crystal trials were ultimately unsuccessful, but this Chapter will discuss the *in silico* structure, the two crystal structures of ADO (published in 2021[5,6]), and the key features in these structures that may link to the function of ADO. The published structures were of ADO without bound substrate, therefore an ADO:substrate structure remained an objective. Towards this, a peptide inhibitor of *Hs*ADO, KH1 (Chapter 5, page 106) was used in co-crystallisation trials to attempt to obtain a crystal structure of *Hs*ADO with a substrate-like molecule bound in the active site. These trials were supplemented with docking studies, performed using AutoDock4[251], to investigate the *in silico* binding modes of different substrate peptides in ADO's active site.

6.3 Crystallisation trials with *Hs*ADO

Given sequence similarities between *Hs*ADO and the *At*PCOs, initial attempts to crystallise *Hs*ADO used conditions similar to those used for *At*PCO crystallisation[66]. Recombinant *N*-His_6-tagged *Hs*ADO was expressed and purified as described previously (Chapter 2, page 31). Initial crystallisation trials used protein expressed and purified by Prof. Emily Flashman, which was aliquoted and flash frozen in liquid N_2 at a concentration of 16 mg mL^{-1}; aliquots were defrosted on ice prior to addition to crystallisation screens. Initial crystallisation screens were conducted aerobically at room temperature without the addition of iron or substrate. The lack of substrate should have prevented any turnover, even in the presence of oxygen, which should improve protein homogeneity and aid crystal formation. A sitting drop vapour diffusion method was used in 24 well format, trialling a protein/precipitant solution ratio of 1:1 in each well (Chapter 8, page 182). Precipitant

solution compositions in initial screens (Table 6.1; Appendix 12) were selected from conditions used by Dr. Mark White to obtain crystals of *At*PCOs[66]. Fine precipitate was observed in multiple drops following addition of protein to precipitant solution in 24 Well Screens 1 and 2 (Table 6.1). Crystallisation screens were checked under a LEICA M205 C stereo microscope after five days and then every two to four days for two weeks. Although large square precipitates were observed in some wells, and small needles were observed in others, these did not develop into promising protein crystals. No potential protein crystals were observed for any other conditions trialled.

Table 6.1: *Hs*ADO crystallisation screening conditions

Screen	Conditions
24 Well – 1	12 – 22 % (w/v) PEG 3350, 0.1 - 0.4 M ammonium nitrate
24 Well – 2	10 – 20 % (w/v) PEG 4000, 0.1 – 0.4 M sodium acetate trihydrate, 0.1 M Tris pH 8.5
24 Well – 3	16 – 26 % (w/v) PEG 3350. 0.1 – 0.4 M sodium fluoride

As *Hs*ADO activity is inhibited by hypoxia[1], crystallisation trials were repeated under anaerobic conditions (Chapter 8, page 182); it was expected that this would stabilise ADO and increase the likelihood of obtaining protein crystals. In the anaerobic trials, it was decided to include a substrate-like peptide; the lack of oxygen would prevent turnover, so these trials presented an opportunity to seek a crystal structure of substrate-bound ADO that would provide information on substrate binding in the active site. Stock solutions of chemicals required for crystallisation screens (Table 6.1) were purged with nitrogen then transferred to an anaerobic glove box (Belle Technology), where they were left to equilibrate overnight. All subsequent steps for the initial anaerobic crystallisation trials were performed in the anaerobic chamber at room temperature. Stock solutions were used to prepare precipitant solutions in 24 well sitting drop plates according to the crystallisation screening

conditions described in Table 6.1 (and Appendix 12). The same batch of *Hs*ADO protein was used as above. In these initial anaerobic crystallisation trials, ADO was mixed with an $RGS4_{2-15}$ peptide bearing an Nt-^{D}Cys prior to addition of protein to the crystallisation screens. It was predicted that substitution of ^{L}Cys with ^{D}Cys in the substrate peptide would help to prevent turnover and improve protein homogeneity for crystal formation. The ^{D}Cys-containing $RGS4_{2-15}$ peptide was prepared by solid phase peptide synthesis (SPPS) using a Liberty Blue™ peptide synthesiser, purified by HPLC, and checked by MALDI MS (Chapter 8; Appendix 3). ^{D}Cys-$RGS4_{3-15}$ peptide (1 μL, 50 mM) was added to *Hs*ADO protein (30 μL, 16.3 mg mL^{-1}), producing a solution of 2.5 mM peptide and 15.8 mg mL^{-1} (~ 494 μM) protein, which was used in crystallisation trials at a protein/precipitant solution ratio of 1:1. It was hoped that the ^{D}Cys-$RGS4_{3-15}$ peptide would co-crystallise with *Hs*ADO to allow determination of a substrate-bound protein structure. Fine precipitate was observed in multiple drops following addition of protein to precipitant solution in all three screens. Crystallisation screens were checked under a Wild M3C stereo microscope after three days and then at least once a week. Solutions in some drops were observed to contain fine precipitate, but no potential protein crystals were observed for any conditions trialled.

As initial crystallisation trials did not produce any potential crystals or promising conditions it was decided to attempt broad anaerobic crystallisation screens with freshly purified protein +/- substrate. Recombinant *N*-His_6-tagged *Hs*ADO was expressed and purified as described previously; *Hs*ADO was concentrated to 20.1 mg mL^{-1}, stored in ice and delivered to Manchester, where Dr. Colin Levy (University of Manchester) used a Mosquito® Crystal robot (SPT Labtech) housed in an anaerobic glove box to perform broad screens with JCSG Plus™, PACT Premier™, The BCS Screen, SG1™ Screen, Morpheus®, and LMB Crystallization Screen (Molecular Dimensions). Screens were performed using 96 well sitting drop plates (SWISSSCI) and a protein/precipitant solution ratio of 1:1.

Anaerobic crystallisation trials were repeated with and without inclusion of $RGS5_{2-15}$ substrate peptide at a peptide/protein ratio of 2.5:1. Crystallisation screens were checked regularly by Dr. Levy for four weeks, after which point the plates were almost completely dehydrated and no protein crystals had been observed.

6.4 *In silico* modelling of *Hs*ADO structure

Alongside attempts to obtain a crystal structure of *Hs*ADO experimentally, comparative protein structure modelling was used to produce an *in silico* structure[250,252,253]. This was achieved using MODELLER[252] (Chapter 8, page 193); source files and python scripts used are provided in Appendix 13 . The crystal structure of *At*PCO4[66] was selected as a template for ADO structure modelling due to sequence and function similarities[4] (Section 6.1).

First, MODELLER was used to align the sequence of *Hs*ADO (accession Q96SZ5) with the PDB file containing the structure of *At*PCO4 (PDB: 6S7E[66]). In order to ensure that the active site iron was included in the *Hs*ADO model the alignment script (align2d.py) provided[252,254] was modified to include heteroatoms (Appendix 13). Five models of *Hs*ADO's structure were produced using the model-ligand.py script (Appendix 13), which also assessed the Discrete Optimised Protein Energy (DOPE)[255] and GA_{341} score[256] of each model. All five models produced had a GA_{341} score of 1. GA_{341} score is a fold assessment function restricted to values between 0 and 1 that depends on percentage sequence identity in the target-template alignment, model compactness, and the z-score of the model[257]; a GA_{341} score > 0.7 indicates an 'accurate' model[256]. As all models had the same GA_{341} score, DOPE was used to select the best model. A lower DOPE score indicates a more accurate and complete model[255], therefore the model with the lowest DOPE score was taken forward (Figure 6.3 **A**; Appendix 13).

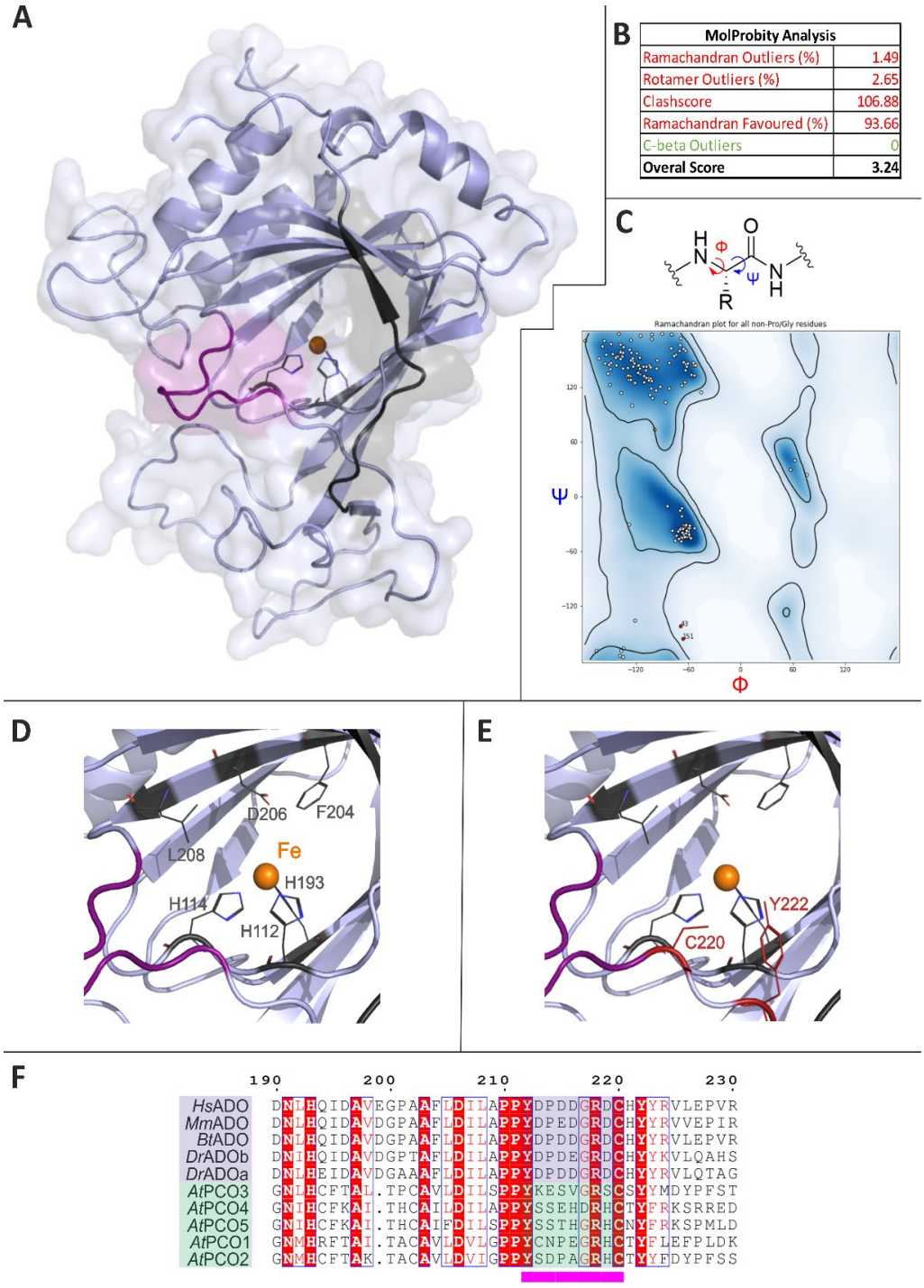


Figure 6.3: *in silico* structure of *Hs*ADO produced with MODELLER
A: *in silico* structure of *Hs*ADO produced with MODELLER using *At*PCO4 (PDB: 6S7E) as a template. Potential substrate binding (hairpin) loop shown in purple; active site iron shown in orange (sphere); iron-binding His residues (H112, H114, H193) shown as lines.
B: MolProbity analysis of *in silico* structure (**A**).
C: Ramachandran plot for all non Pro/Gly residues in *in silico* structure (**A**).
D: Close-up view of *Hs*ADO *in silico* structure (**A**) active site showing Phe-Asp-Leu motif (lines, gray), 3 Fe-binding His residues (lines, gray), and coordinating active site Fe (sphere, orange)
E: Alternative view of active site (**D**) also showing potential C220-Y222 cross link residues (lines, red)
F: CLUSTAL O(1.2.4) multiple sequence alignment of hairpin loop (magenta line) region of NCO sequences rendered with ESPript 3.0 server: *Homo sapiens* (human) ADO; *Mus musculus* (mouse) ADO; *Bos taurus* (bovine) ADO; *Danio rerio* (zebrafish) ADO a and b; *Arabidopsis thaliana* (mouse-ear cress) PCO1-5.

The *in silico* model selected following DOPE measurement was subsequently assessed using MolProbity[258] (Figure 6.3 **B** – **C**). The overall MolProbity score is representative of the experimental resolution expected for a model of the quality observed, so a lower score was better. Other measures that were assessed included 'clashscore', which measured the steric clash overlaps ≥ 0.4 Å per thousand atoms (a lower score is more favourable), and Ramachandran and rotamer criteria. The MolProbity clashscore was particularly high for the *in silico* model produced using the *At*PCO4 template (106.88); however, other measures were close to being within the acceptable ranges for an experimental crystal structures (> 98 % Ramachandran favoured, < 1 % Rotamer outliers, 0 C-beta outliers). It was not surprising that that the *in silico* structure had less favourable MolProbity statistics than an experimentally determined structure, especially given *Hs*ADO shared only 26.6 % sequence identity with the template *At*PCO4. Loop refinement was attempted to improve the quality of the *in silico* model but was not successful (Chapter 8, page 193). The overall quality of the *in silico* model was considered to be sufficient for initial investigation of *Hs*ADO's structural features.

As expected, the *in silico* structure of *Hs*ADO (Figure 6.3 **A**, **D-E**) featured a double stranded β helix (DSBH), characteristic of cupin family proteins[259,260], and a 3-His iron-binding site as seen in the structures of mammalian CDO and *At*PCOs (Figure 6.1). Furthermore, the Phe-Asp-Leu (F-D-L) motif in *Hs*ADO was positioned in the active site cavity similarly to the S-H-Y and I-D-L motifs in CDO and PCOs respectively (Figure 6.1, Figure 6.3). In CDO, a Cys-Tyr crosslink has been observed that has been linked to increased catalytic activity[156,157,159–163,172,241,261]. In 2018, Wang *et al.* detected evidence of a similar crosslink in ADO between Cys220 and Tyr222 (Figure 6.3 **E**); unlike in CDO, crosslink formation in ADO did not appear to induce a large structural change[178]. In the *in silico Hs*ADO structure proposed in this Chapter, the two potential crosslink residues

were located near the entrance to the active site, with the Cys sulfur atom 4.0 Å from the *ortho* Tyr carbon. This distance is comparable to the distance observed between the two equivalent atoms in the structure of non-crosslinked *Rn*CDO[160]. The *in silico* structure produced in this Chapter therefore suggested that the Cys220-Tyr222 crosslink is feasible.

The *Hs*ADO *in silico* structure shared similarities with the structure of *At*PCO4[66]; the placement of the C-terminal loop created a more open active site cavity compared to that of CDO and a hairpin loop was observed near at the opening of the active site cavity[4]. In *Hs*ADO the hairpin loop, predicted to be involved in substrate binding[4,66], is rich in acidic residues (Figure 6.3 **F**). This aligns with the observation made in Chapter 2 (page 41) that *Hs*ADO preferentially catalyses oxidation of 14-mer peptides rich in basic residues over those containing more acidic residues. Interestingly, this hairpin loop is more conserved in ADO proteins across different species than it is within the five *At*PCOs (Figure 6.3 **F**).

In 2021, two experimentally determined crystal structures of ADO were published: one of nickel-bound *Hs*ADO (PDB: 7REI)[5] and one of iron-bound *Mm*ADO (PDB: 7LVZ)[6]. These structures both bore close resemblance to the *in silico Hs*ADO structure produced with MODELLER and *At*PCO4 template (Figure 6.4). Wang *et al.* (PDB: 7REI) found that initial crystallisation attempts with wild-type ADO were unsuccessful and it was necessary to mutate two surface cysteines (C18 and C239) to serine in order to prevent protein dimerisation via disulfide bonds and obtain homogenous protein suitable for crystallization. In this study, the active site iron was replaced with nickel, and crystals were obtained under aerobic conditions using 67 mg mL^{-1} protein mixed 1:1 (v/v) with a crystallization buffer of 0.1 M BisTris-HCl (pH 5.5), 0.2 M Li_2SO_4, and 20% (w/v) PEG3350 at 289 K (~16 °C). Crystals formed after one week and were prepared for X-ray diffraction after two weeks, suggesting that timescale was not the issue in initial crystallization trials described in this Chapter (Section 6.3). Notably, this protein concentration was much higher than the

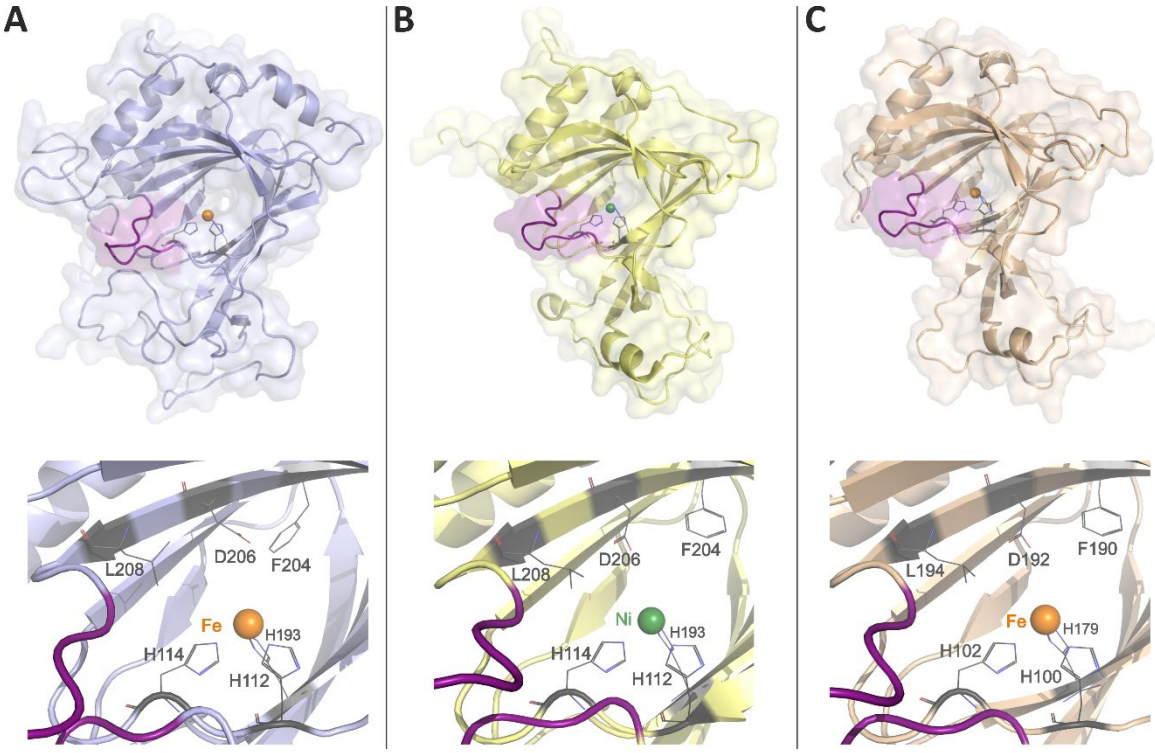


Figure 6.4: Comparison of *in silico* and crystal structures of ADO
A: *in silico* structure of *Hs*ADO (Figure 6.3); showing hairpin loop (cartoon, purple), active site Fe (sphere, orange), 3 Fe-binding His residues and Phe-Asp-Leu motif (lines)
B: Wang *et al.*'s crystal structure of Ni-bound *Hs*ADO (PDB: 7REI); showing hairpin loop (cartoon, purple), active site Ni (sphere, green), 3 Fe-binding His residues and Phe-Asp-Leu motif (lines)
C: Fernandez *et al.*'s crystal structure of Fe-bound *Mm*ADO (PDB: 7LVZ; chain D); showing hairpin loop (cartoon, purple), active site Fe (sphere, orange), 3 Fe-binding His residues and Phe-Asp-Leu motif (lines)

concentration used in *Hs*ADO crystallization trials performed in this Chapter (Section 6.3). The protein used in this Chapter was also wild-type *Hs*ADO, without surface cysteine mutation as described by Wang *et al.* However, it is suspected that the primary reason ADO crystals were not observed in this Chapter was low protein concentration rather than inhomogeneity. Wang *et al.* observed a shoulder in the chromatogram obtained during size exclusion chromatography (SEC) purification of wild-type *Hs*ADO[5], which was not observed during SEC purification of *Hs*ADO protein used in this Thesis (Appendix 1). This suggests that protein used in this Chapter (Section 6.3) was homogenous, rather than a mix of dimer and single protein molecules that would negatively impact crystallisation. Fernandez *et al.* also did not describe a need to mutate surface cysteines[6], although C18 and C239 are both conserved in *Mm*ADO. Fernandez *et al.* (PDB: 7LVZ) did not provide

information on protein concentration; in this study, crystals of iron-bound *Mm*ADO were obtained under anaerobic conditions with a crystallization buffer of 15% (w/v) PEG 3350, 300 mM $MgCl_2$, and 0.1 M Tris-HCl (pH 8.0) at 293 K (~20 °C).

Both crystal structures of ADO featured the characteristic DSBH cupin fold, with the active site metal bound by 3 His residues[5,6]. Wang *et al.* and Fernandez *et al.* both observed octahedral coordination of the active site metal, with water molecules bound to the three remaining metal binding sites. Wang *et al.* noted that the structure of *Hs*ADO was very flexible, with 62 % of the protein consisting of loops or disordered regions, which they proposed helps ADO to bind substrates of varying sizes[5]. The predicted Cys-Tyr crosslink[178] was not observed in either crystal structure, which was not unexpected as crosslink formation is believed to require iron and oxygen[5,6]. Wang *et al.* suggested that crosslink formation could be facilitated by movement of the hairpin loop closer to the metal centre and could tune *Hs*ADO catalytic activity to favour small molecule over Nt-Cys oxidation[5]. Fernandez *et al.* noted that a Tyr on the other end of the hairpin loop (Y198 in *Mm*ADO) was the same distance from the predicted crosslink Cys (C206 in *Mm*ADO) as the other predicted crosslink residue (Y208 in *Mm*ADO). This other Tyr is conserved in *Hs*ADO (Y212). In this paper, Fernandez *et al.* suggest that Tyr198 could be more important than Tyr208 for substrate binding and enzyme activity. Interestingly, Fernandez *et al.* also identified a smaller tunnel on the opposite side of the ADO protein to the broad active site tunnel, which they suggested could be how O_2 enters the active site. In their structure they identified two cysteines (C120 and C169) that line this potential O_2 access tunnel; they proposed that these could form a disulfide bond 'gate' to control O_2 access to the active site in response to changes in the redox environment.

Overall, *in silico* and experimentally determined crystal structures of ADO shared more similarity with structures of the *At*PCOs than with those of mammalian CDOs[4–6,66,262].

Structures of ADO also contained features that could be linked to enzyme function and substrate binding; the relationship between structure and function will be discussed further below (Sections 6.6 and 6.7).

6.5 Co-crystallisation trials with *Hs*ADO and KH1 peptide inhibitor

Although crystal structures of ADO contained features that provided indicators of how *Hs*ADO could bind to both small molecule and protein substrates[4–6], it would be particularly useful to obtain a structure of ADO protein with a substrate bound in the active site. Spectroscopic studies by Wang *et al.* have suggested that the active site iron in *Hs*ADO binds to the Nt-Cys of substrate proteins in a monodentate manner via the thiol group[185]. Furthermore, Fernandez *et al.* conducted computational studies using $RGS5_{2\text{-}5}$ that demonstrated that the Nt-amine of the peptide could interact with the Asp residue in the Phe-Asp-Leu motif. In order to further investigate ADO substrate binding, it was decided to use a substrate-like peptide inhibitor in crystallographic studies.

The substrate-like peptide inhibitor selected for co-crystallisation trials was KH1 (Sec-FGRRMK). The sequence of this peptide inhibitor was based on the most rapidly oxidised optimum substrate peptide (Chapter 2, page 47), and KH1 was found to have an IC_{50} of approximately 82 μM *in vitro* (Chapter 5, page 106). It was predicted that this peptide would bind to *Hs*ADO in a similar manner to a productive substrate, while also stabilising *Hs*ADO. Crystallisation trials with KH1 and *Hs*ADO were conducted aerobically at room temperature (Chapter 8, page 182).

Recombinant *N*-His_6-tagged *Hs*ADO protein was expressed and purified as described previously (Chapter 2, page 31). *Hs*ADO was concentrated to 81.3 mg mL^{-1} and kept on ice. During concentration it was observed that the *Hs*ADO solution was pink, which was likely due to phenanthroline contamination during SEC (Appendix 14). It was necessary to

determine whether this pink form of ADO showed reduced activity; it was seen to be as active as a control sample of wild-type *Hs*ADO (Appendix 14), so was still used for crystallisation trials and assumed that KH1 peptide would be able to outcompete any bound contaminant. However, the protein solution was supplemented with iron(II) to account for any iron(II) sequestered by phenanthroline in the solution. *Hs*ADO (81.3 mg mL^{-1}, 235 µL) was mixed with KH1 peptide (100 mM, 30 µL), Fe_2SO_4 (500 mM in 20 mM HCl, 2.4 µL) and SEC buffer ((400 mM NaCl, 50 mM Tris, pH 7.5), 32.6 µL) to produce a solution of 63.7 mg mL^{-1} *Hs*ADO and 10 mM KH1 (1:5 protein/inhibitor). This *Hs*ADO/KH1 mix was used in crystallisation trials at a protein/precipitant solution ratio of 1:1. Two 6 well screens were conducted using a 24 well hanging drop plate; conditions in these screens were selected based on crystallisation conditions used by Wang *et al.*[5] and Fernandez *et al.*[6] to obtain crystals of ADO previously (Table 6.2). In addition, four broad screens were conducted using a Mosquito® Xstal3 robot (SPT Labtech) to prepare 96 well sitting drop plates (Table 6.2).

Table 6.2: KH1 co-crystallisation screening conditions

Screen	Conditions
6 Well - 1	0.1 M Bis-Tris HCl, 15 – 25 % (w/v) PEG 3350, 0.1 – 0.3 M lithium sulfate
6 Well - 2	0.1 M Tris (pH 8.0), 10 – 20 % (w/v) PEG 3350, 0.2 – 0.4 magnesium chloride
Broad Screens	JCSG Plus™; PACT Premier™; Morpheus® (Molecular Dimensions). Crystal Screen™ (Hampton Research).

Crystallisation screens were checked under a LEICA M205 C stereo microscope after six days, then every two days for a week, and then once a week for three more weeks. Precipitate

was visible in various conditions across all screens (Appendix 15); however, no potential protein crystals were observed for any conditions trialled.

6.6 AutoDock4 studies with *Mm*ADO

ADO substrate binding was also investigated *in silico* using protein structure visualisation and modelling software packages, PyMOL and Autodock4[251]. Of the two published crystal structures of ADO, the structure of *Mm*ADO[6] (PDB: 7LVZ) was selected for use in docking studies as it contained iron in the active site, while the structure of *Hs*ADO[5] (PDB: 7REI) contained nickel. *Hs*ADO and *Mm*ADO share 85.6 % sequence identity and their structures are very similar (Figure 6.4), so it was expected that there would not be significant differences in the manner in which these proteins bound to substrates. The structure of *Mm*ADO published by Fernandez *et al.* consists of four chains of four independent copies of ADO; these chains were inspected in PyMOL and chain D was selected for use in docking studies as it was the most complete. As mentioned previously, the active site iron in this structure was octahedrally coordinated to three His residues and three water molecules (Figure 6.5). The three water molecules were inspected as potential substrate binding sites;

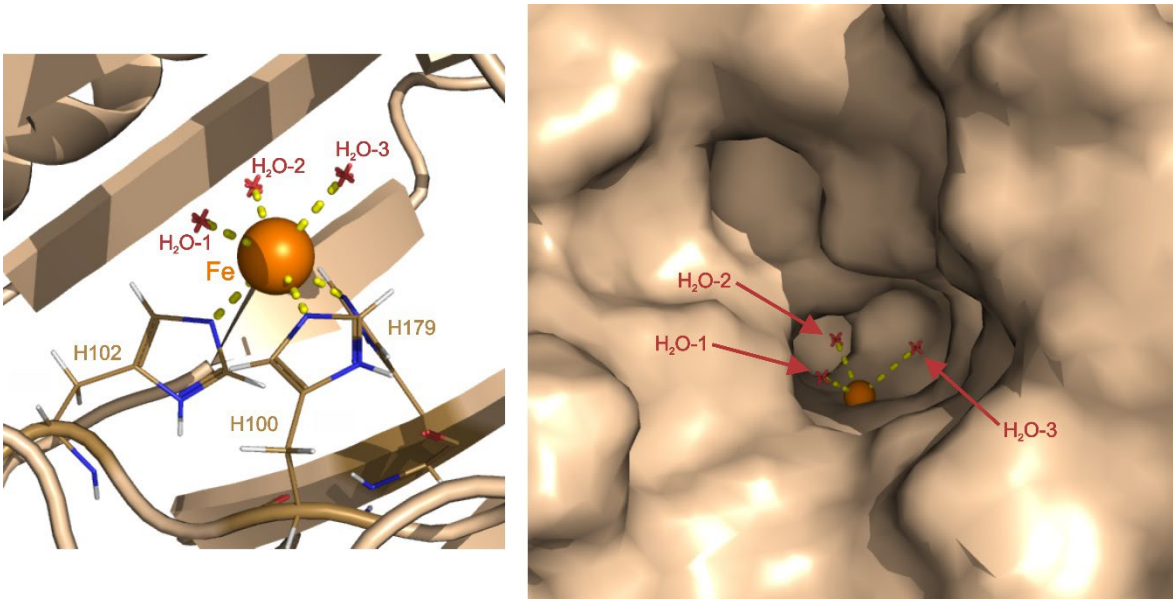


Figure 6.5: Potential binding sites on *Hs*ADO active site iron
Close-up view of active site iron in *Mm*ADO crystal structure (PDB: 7LVZ, chain D), showing cartoon (left) and surface (right). 3 Fe-binding His residues (lines, sand) and 3 Fe-binding water molecules (labelled red crosses). H_2O-3 identified as most likely site for binding thiol of peptide/protein substrate.

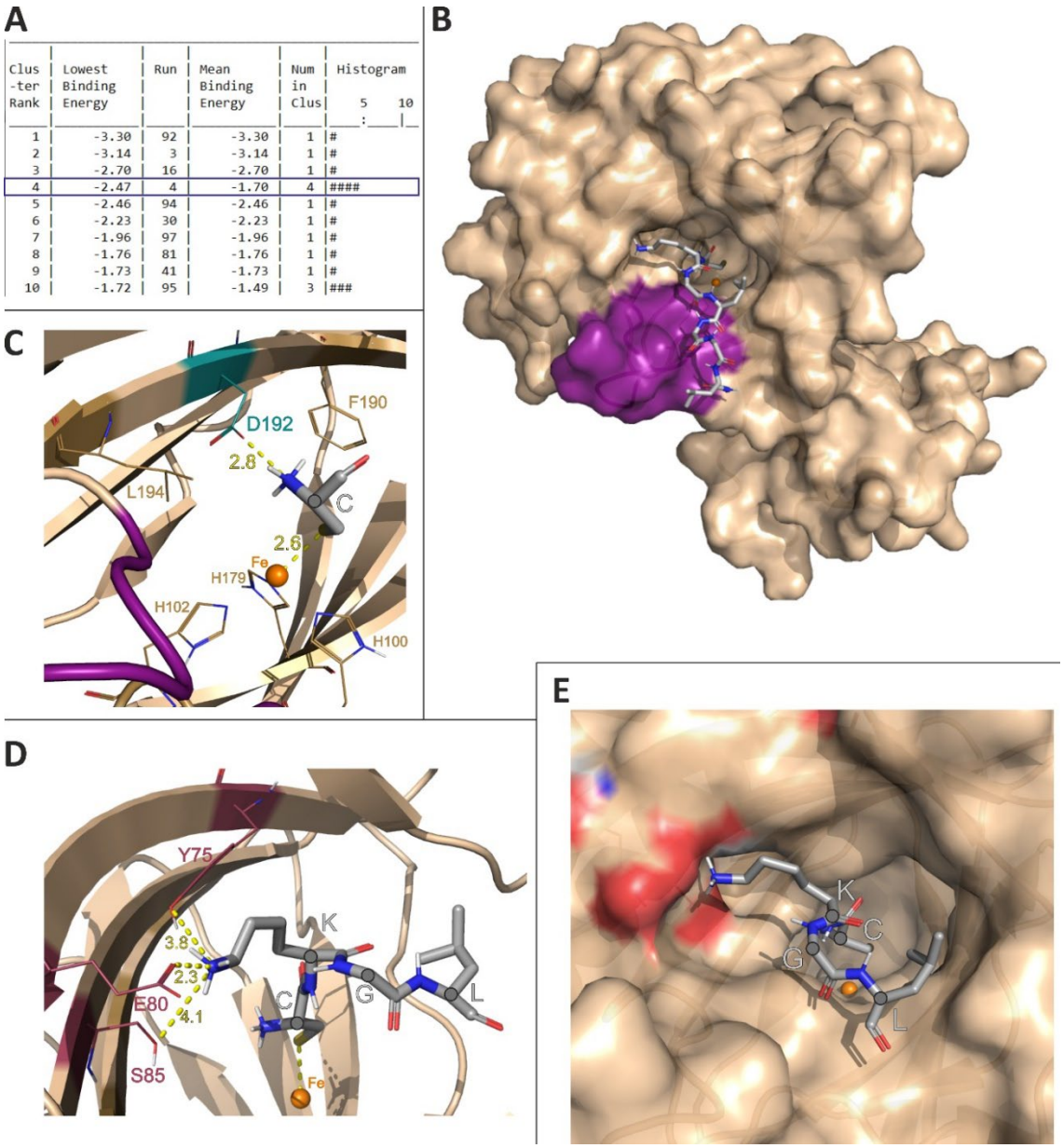
one water molecule (Figure 6.5, H_2O-2) pointed away from the main active site cavity towards the predicted O_2 access tunnel[6], while the other two water molecules were located closer to the main active site cavity (Figure 6.5, H_2O-1 and -3). H_2O-3 was selected as the most likely site for thiol binding as access to this site via the main active site cavity appeared to be the least obstructed by the rest of the protein structure. PyMOL was then used to generate structures of *Mm*ADO with *Hs*RGS4$_{2\text{-}8}$ (CKGLAGL), *Hs*IL32$_{2\text{-}8}$ (CFPKVLS), and *At*RAP2.12$_{2\text{-}8}$ (CGGAIIS) bound to the active site iron via the Nt-Cys thiol sulfur atom (as observed in spectroscopic studies by Wang *et al.*[185]).

In AutoDock4, the iron-peptide residues were set as the flexible residues, while the remaining residues of *Mm*ADO were set to be rigid and the ligand was set to be a water molecule (Chapter 8, page 193). For each complex, 100 docking experiments were performed using the Lamarckian genetic algorithm with the default parameters with a maximum of 2.5 million energy evaluations applied for each experiment. This set-up allowed the lowest energy conformations to be searched for with the three different peptides bound to the active site of *Mm*ADO.

6.6.1 Docking results with RGS4 peptide

AutoDock4 docking experiments with RGS4$_{2\text{-}8}$ and *Mm*ADO (PDB: 7LVZ_D) produced 100 different conformations, whose binding energies were evaluated and which were sorted into clusters of similar conformations. 'Run 4' was selected as this conformation had the fourth lowest binding energy and the largest cluster of the top ten conformations (Figure 6.6 **A-B**). The RGS4$_{2\text{-}8}$ peptide in this conformation was well positioned to potentially bind to residues in the hairpin loop region of *Mm*ADO (Figure 6.6 **B**). However, in these docking experiments the hairpin loop region was kept rigid, while the peptide residues remained flexible. In reality it is likely that the hairpin loop residues would be able to move to interact more favourably with Nt-residues of a substrate protein or peptide.

Although these docking results likely do not provide an accurate representation of substrate binding to the hairpin loop, they do help to validate the suggestion that this loop is involved



Clus-ter Rank	Lowest Binding Energy	Run	Mean Binding Energy	Num in Clus	Histogram 5 10
1	-3.30	92	-3.30	1	#
2	-3.14	3	-3.14	1	#
3	-2.70	16	-2.70	1	#
4	-2.47	4	-1.70	4	####
5	-2.46	94	-2.46	1	#
6	-2.23	30	-2.23	1	#
7	-1.96	97	-1.96	1	#
8	-1.76	81	-1.76	1	#
9	-1.73	41	-1.73	1	#
10	-1.72	95	-1.49	3	###

Figure 6.6: AutoDock4 results with RGS4$_{2-8}$ peptide and *Mm*ADO (PDB: 7LVZ_D)
A: Clustering histogram from AutoDock4 run (Lamarckian GA, 100 runs, maximum of 2.5 million energy evaluations per run) with RGS4$_{2-8}$ peptide and *Mm*ADO (PDB: 7LVZ, chain D). Binding energy calculated in kcal mol^{-1}. 'Run 4' selected.
B: RGS4$_{2-8}$ peptide (sticks, grey) docked into *Mm*ADO (**A** – 'Run 4'). Hairpin loop residues (Y198-C206) coloured purple.
C: Close-up view of Nt-Cys of RGS4$_{2-8}$ docked into *Mm*ADO (**B**), showing active site Fe (sphere, orange), 3 Fe-binding His residues (lines), Phe-Asp-Leu motif (lines), and Nt-Cys of RGS4$_{2-8}$ (sticks, grey). Nt-amine within H-bonding distance of Asp192 (lines, teal). Distances between Nt-amine and Asp192, and Nt-thiol and Fe, are given in angstrom in yellow.
D: View showing Lys3 (of RGS4$_{2-8}$) in 'polar pocket', consisting of Tyr75, Glu80, and Ser85 residues (lines, raspberry). Distances with side chain amine of Lys3 and polar pocket residues given in angstrom in yellow. First 4 residues of RGS4$_{2-8}$ peptide shown (sticks, grey).
E: Close-up view of first 4 residues of RGS4$_{2-8}$ peptide (sticks, grey) docked into surface representation of *Mm*ADO (**B**); polar pocket residues are coloured by element (O = red, N = blue, C = wheat).

in substrate binding[4,66].

In the chosen conformation, the amine of the Nt-Cys was orientated towards the Asp residue of the Phe-Asp-Leu motif (Figure 6.6 **C**), as observed in computational studies by Fernandez *et al.*[6]. In Chapter 4 (page 95), *in vitro* experiments demonstrated that acetylation of this amine significantly reduced (and in many cases eliminated all) *Hs*ADO-catalysed oxidation of the Nt-Cys of peptide substrates. In combination, these *in vitro* and *in silico* results indicate that the Nt-amine is important for ADO substrate binding, and that this interaction requires an active site Asp residue. This Asp residue is highly conserved across NCOs, but not in small molecule TDOs[4]. Notably, mutation of this residue in *At*PCO4 (Asp176) to an Asn, thus removing the negative charged side chain, resulted in almost complete loss of catalytic activity and significantly reduced iron incorporation, relative to the wild-type enzyme[66].

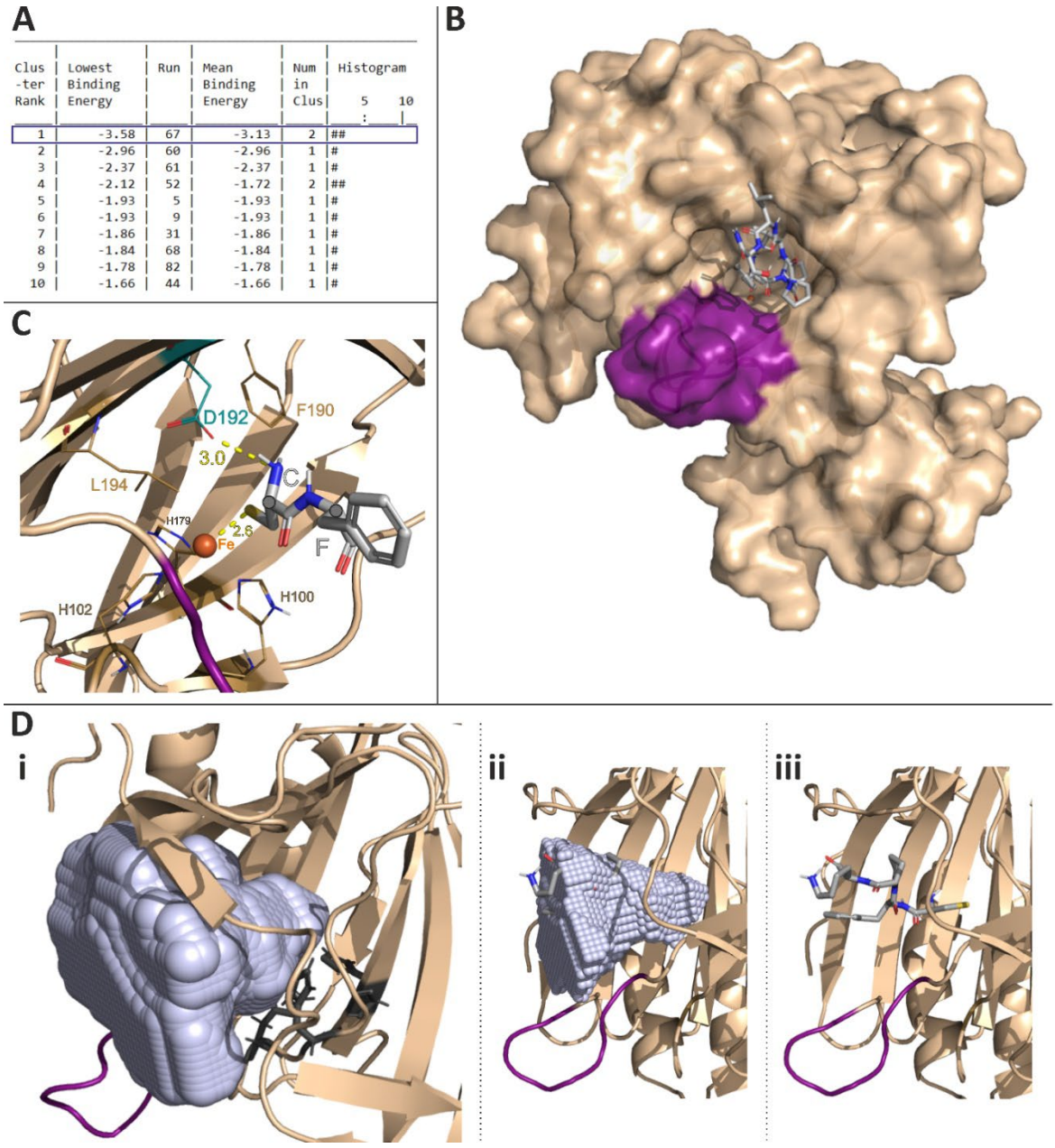
Also in this conformation, Lys3 of $RGS4_{2-8}$ was observed to be orientated towards a 'polar pocket', consisting of residues Tyr75, Glu80 and Ser85 (Figure 6.6 **D-E**). The side chain amine of RGS4 Lys3 was within hydrogen bonding range of all three ADO 'polar pocket' residues. These residues are conserved in *Hs*ADO (Tyr87, Glu92, Ser97). In Chapter 2 (page 41), the highest *in vitro Hs*ADO initial activity was observed with peptides that had a basic or aromatic residue at the position after the Nt-Cys. This docking result suggests that a basic residue at position 3 in a substrate peptide or protein could be able to interact with the ADO 'polar pocket' residues and improve substrate binding to the active site.

6.6.2 Docking results with IL32 peptide

As before (Section 6.6.1) AutoDock4 docking experiments with $IL32_{2-8}$ and *Mm*ADO (PDB: 7LVZ_D) produced 100 different conformations. Less clustering was observed with these experiments than in experiments with RGS4; in this case 'Run 67' was selected as this

conformation had the lowest binding energy (Figure 6.7 **A-B**). More intra-peptide interactions were observed in $IL32_{2-8}$ (CFPKVLS) docking results than in $RGS4_{2-8}$ (CKGLAGL) results; this was likely due to the presence of more amino acids with polar or basic side chains. It is difficult to predict if this observation is representative of behaviour of the N-terminus of full-length IL32 protein, as without a crystal structure it is not known how much of the N-terminus may be flexible.

In conformation 67, the Nt-amine was once again observed in an orientation that would allow H-bond formation with Asp192 (Figure 6.7 **C**). There also appeared to be sufficient space in the active site cavity to accommodate the bulky Phe residue after the Nt-Cys (Figure 6.7 **D**). Protein cavity detection in *Mm*ADO using the CavityPlus server[263] predicted that the active site cavity in *Mm*ADO has a volume of approximately 790 $Å^3$ (Figure 6.7 **D**, Appendix 16). Visual inspection of CavityPlus results and AutoDock4 results with $IL32_{2-8}$ suggested that residue 5 partially protrudes from the main active site cavity (Figure 6.7 **D ii-iii**). The volumes of Cys, Phe, and Pro (the three IL32 residues predicted to be fully within the cavity) are believed to be approximately 190, 100, and 120 $Å^3$, respectively[264]. This is a total volume of 410 $Å^3$. Therefore, even accounting for the partial volume of residue 5 (Lys, approximately 170 $Å^3$), there would be expected to be spare volume within the 790 $Å^3$ active site cavity, enabling *Mm*ADO to easily accommodate the bulky aromatic residue at position 3 in the substrate.



Clus-ter Rank	Lowest Binding Energy	Run	Mean Binding Energy	Num in Clus	Histogram 5 10 :
1	-3.58	67	-3.13	2	##
2	-2.96	60	-2.96	1	#
3	-2.37	61	-2.37	1	#
4	-2.12	52	-1.72	2	##
5	-1.93	5	-1.93	1	#
6	-1.93	9	-1.93	1	#
7	-1.86	31	-1.86	1	#
8	-1.84	68	-1.84	1	#
9	-1.78	82	-1.78	1	#
10	-1.66	44	-1.66	1	#

Figure 6.7: AutoDock4 results with IL32$_{2-8}$ peptide and *Mm*ADO (PDB: 7LVZ_D)
A: Clustering histogram from AutoDock4 run (Lamarckian GA, 100 runs, maximum of 2.5 million energy evaluations per run) with IL32$_{2-8}$ peptide and *Mm*ADO (PDB: 7LVZ, chain D). Binding energy calculated in kcal mol^{-1}. 'Run 67' selected.
B: IL32$_{2-8}$ peptide (sticks, grey) docked into *Mm*ADO (**A – '**Run 67'). Hairpin loop residues (Y198-C206) coloured purple.
C: Close-up view of Nt-Cys-Phe of IL32$_{2-8}$ docked into *Mm*ADO (**B**), showing active site Fe (sphere, orange), 3 Fe-binding His residues (lines), Phe-Asp-Leu motif (lines), and first two residues of IL32$_{2-8}$ (sticks, grey). Nt-amine within H-bonding distance of Asp192 (lines, teal). Distances between Nt-amine and Asp192, and Nt-thiol and Fe, are given in angstrom in yellow.
D: CavityPlus results (cavity 1). **i** – vacant space in cavity 1 as predicted by CavityPlus analysis of *Mm*ADO (PDB: 7LVZ, chain D), hairpin loop in purple, iron-binding His residues in dark grey. **ii** – vacant space in cavity 1 (**i**) overlaid on AutoDock4 result with IL32$_{2-8}$ peptide (sticks, light grey) and *Mm*ADO (**A – '**Run 67'), showing residues 2 – 5 of IL32 (CFPK). **iii** – AutoDock4 result with IL32$_{2-8}$ peptide (sticks, light grey), as seen in **ii** with view of cavity 1 removed.

Furthermore, Tyr198 was identified as a residue in the *Mm*ADO active site that was located close to the IL32 Phe3 residue in this conformation. Rotation around the Cα – Cβ bond of Tyr198 placed the phenyl ring close enough to the Phe3 phenyl ring to allow π-π bonding (Figure 6.8 **A**). It is unclear how likely this orientation of Tyr198 is to occur, however there are five other aromatic residues within the active site cavity of *Mm*ADO: Tyr75; Phe89; Phe190; Tyr208; Phe242 (Figure 6.8 **B**). The prevalence of aromatic residues in the active site cavity represents a high potential for π-π interactions between substrate and enzyme, and may help to explain the observation from Chapter 2 (page 41) that many of the best

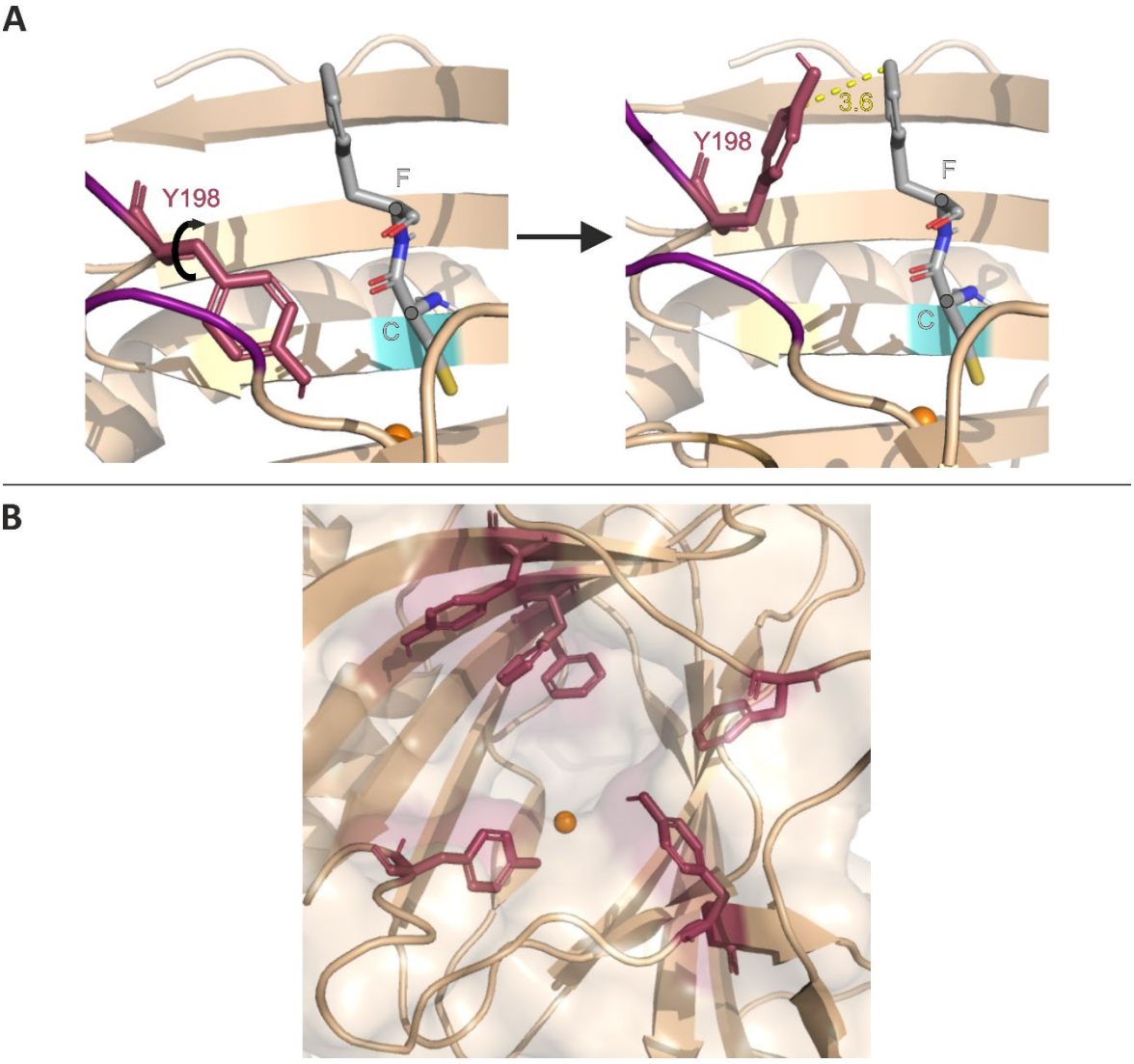


Figure 6.8: Interaction of IL32$_{2\text{-}8}$ with aromatic residues in *Mm*ADO active site
A: View of IL32$_{2\text{-}8}$-*Mm*ADO (Figure 6.7 **B**) showing how rotation around C^{α}-C^{β} bond brings phenyl ring of Tyr198 close enough to Phe3 of IL32$_{2\text{-}8}$ for π-π interaction.
B: Active site cavity of *Mm*ADO (PDB: 7LVZ, chain D) showing aromatic residues (sticks, raspberry) within 4 Å of IL32$_{2\text{-}8}$ peptide in AutoDock4 result (Figure 6.7).

in vitro peptide substrates of *Hs*ADO feature an aromatic residue after the Nt-Cys. All six active site aromatic residues are conserved in *Hs*ADO, and many are also conserved in *At*PCO4/5, while only Tyr198 (*Mm*ADO numbering) is complemented by an aromatic residue (Phe) at the equivalent position in *Hs*CDO (Figure 6.9). This alignment also shows that the residues that form the polar pocket noted above (Section 6.6.1) are conserved in *At*PCO4/5, as well as between *Hs*ADO and *Mm*ADO, but are not present in *Hs*CDO. In *At*PCO1-3 the polar pocket is also largely conserved: the Tyr-Glu-Ser motif observed in ADO is replaced by a Tyr-Gln-Ser, Tyr-Ala-Ser, and Phe-Glu-Thr in *At*PCOs 1,2, and 3, respectively.

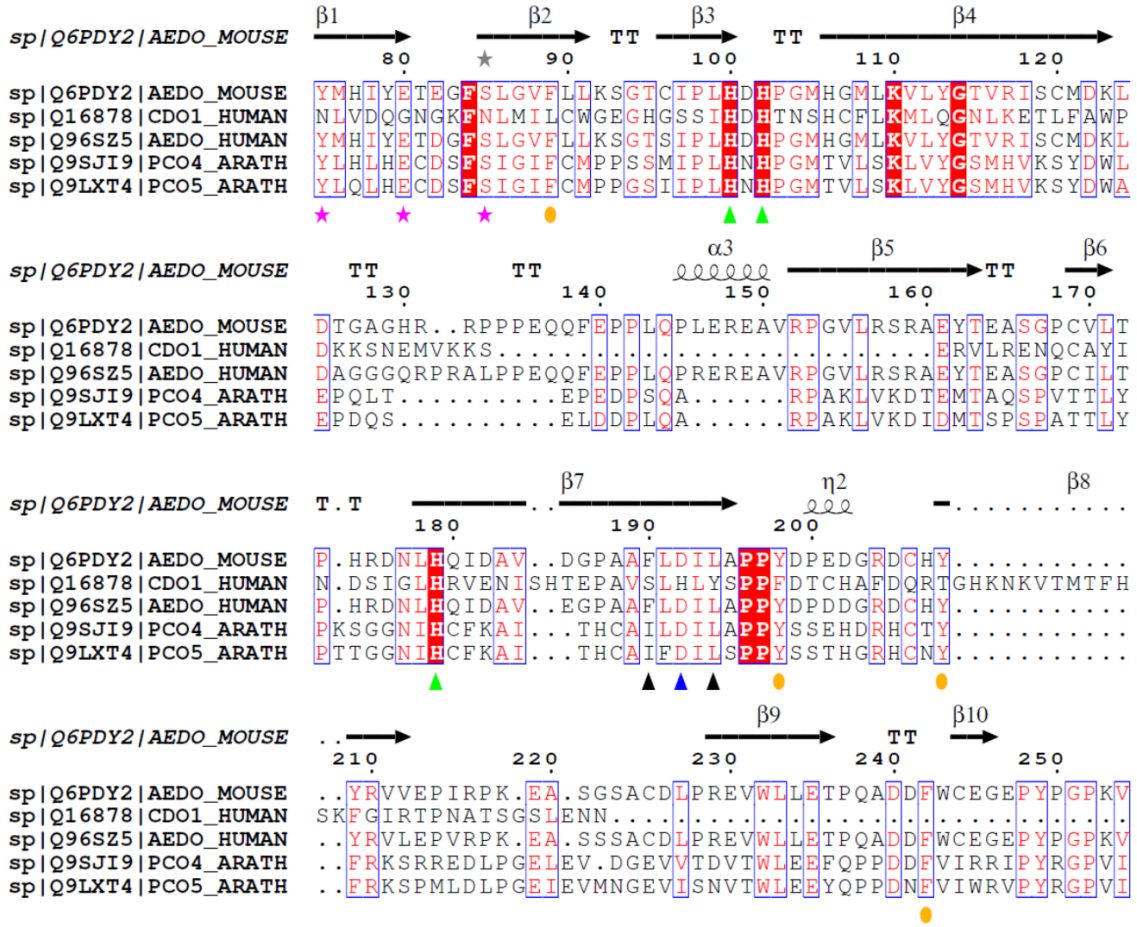


Figure 6.9: Conservation of key ADO active site residues in CDO and PCOs
CLUSTAL O(1.2.4) multiple sequence alignment of *Mm*ADO (Q6PDY2), *Hs*CDO (Q16878), *Hs*ADO (Q96SZ5), *At*PCO4 (Q9SJI9), and *At*PCO5 (Q9LXT4) rendered with ESPript 3.0 server. Fe-binding His residues marked with green triangles, Phe-Asp-Leu motif (ADO) marked with black/blue triangles (Asp192 (*Mm*ADO numbering) marked with blue triangle), polar pocket (identified in Fig. 6.6) residues marked with pink stars, and aromatic active site residues (Fig. 6.8) marked with orange circles (with the exception of Tyr75 and Phe190 (*Mm*ADO numbering), which are part of the polar pocket and Phe-Asp-Leu motif (ADO) respectively).

6.6.3 Docking results with RAP2.12 peptide

The final docking experiments of this Chapter were conducted with the PCO substrate, *At*RAP2.12$_{2\text{-}8}$. *Hs*ADO has previously been shown to accept *At*RAP2.12 as a substrate both *in vitro* (Chapter 2, page 52) and *in cellula*[1] (Chapter 3, page 66). However, catalytic efficiency of *Hs*ADO with *At*RAP2.12$_{2\text{-}15}$ peptide (1.2×10^4 M^{-1} s^{-1}) was lower than with RGS4$_{2\text{-}15}$, RGS5$_{2\text{-}15}$, and IL32$_{2\text{-}15}$ peptides ($1.6 - 3.1 \times 10^5$ M^{-1} s^{-1}). Therefore, docking experiments were performed to investigate whether there was a structural explanation for ADO's apparent reduced affinity for *At*RAP2.12 *in vitro* compared to the human substrates. From the 100 docking results 'Run 5' was selected; this conformation had a calculated binding energy of -5.05 kcal mol^{-1}, which was at least 1.35 kcal mol^{-1} lower than all other conformations (Figure 6.10 **A**-**B**).

As with IL32$_{2\text{-}8}$ (Section 6.6.2), some intra-peptide bonding was observed and the peptide structure was more 'curled up' than in docking experiments with RGS4$_{2\text{-}8}$. Residues later in the sequence of *At*RAP2.12$_{2\text{-}8}$ did appear to be orientated towards the *Mm*ADO hairpin loop (Figure 6.10 **B**), suggesting that there could be interaction between the peptide and the loop. As observed in previous docking results (Section 6.6.1-2), and by Fernandez *et al.*[6], the Nt-amine group was observed to be within H-bonding distance of Asp192 (Figure 6.10 **C**). However, no other potential interactions were identified between *At*RAP2.12$_{2\text{-}8}$ and *Mm*ADO active site residues. The first three RAP2.12 residues (GGA) after the Nt-Cys are small hydrophobic residues, unlike the equivalent residues in RGS4 (KGL) and IL32 (FPK). Docking experiments with RGS4$_{2\text{-}8}$ and IL32$_{2\text{-}8}$ peptides suggested that these three residues could be important for improving substrate binding in ADO's active site. Therefore, this docking result suggested that while the N-terminus of *At*RAP2.12 could be easily accommodated in the active site (Figure 6.10 **D**), it likely binds less strongly than the mammalian substrates. This could begin to explain the reduced catalytic efficiency *in vitro*.

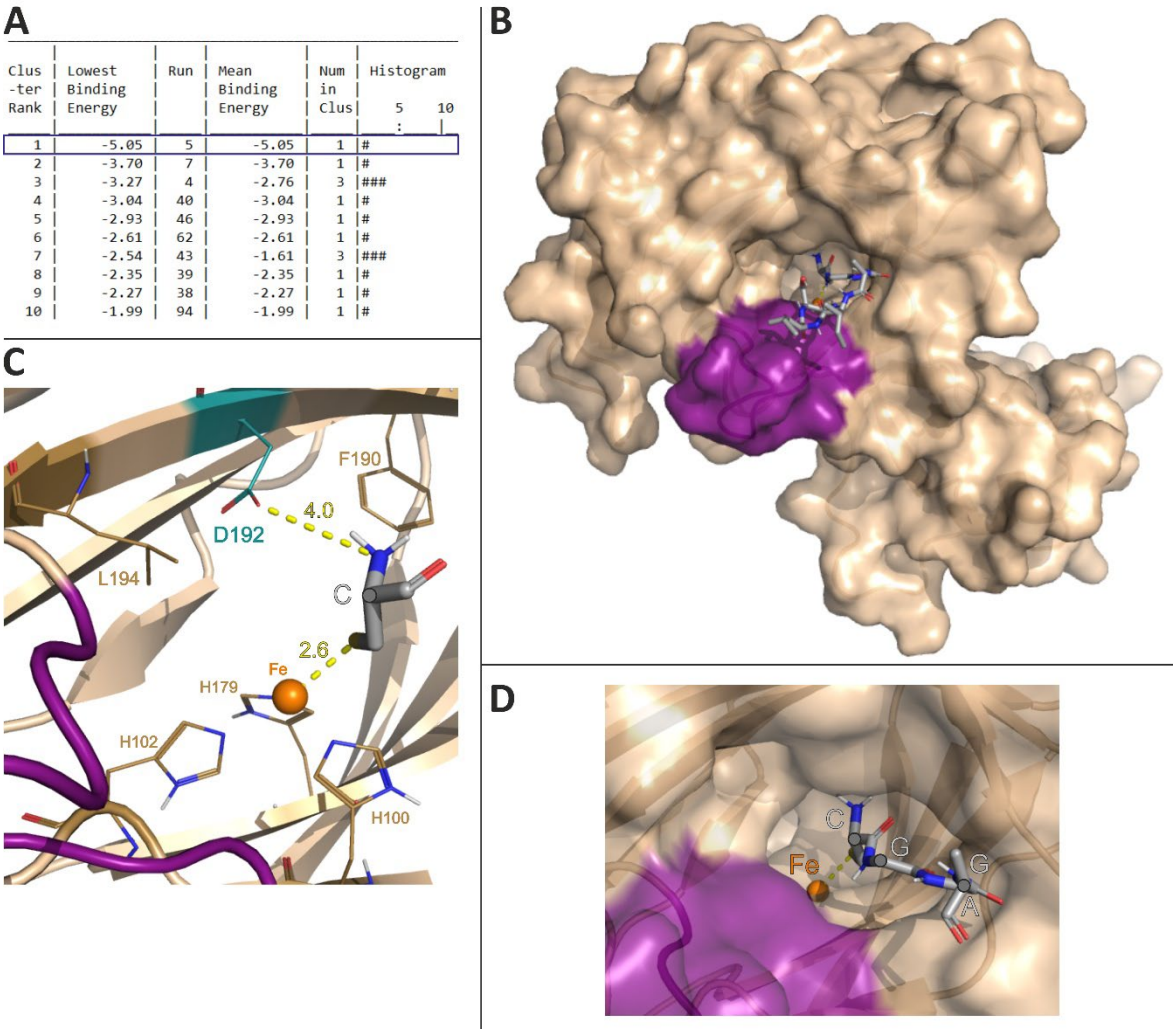


Figure 6.10: AutoDock4 results with *At*RAP2.12$_{2\text{-}8}$ peptide and *Mm*ADO (PDB: 7LVZ_D)
A: Clustering histogram from AutoDock4 run (Lamarckian GA, 100 runs, maximum of 2.5 million energy evaluations per run) with *At*RAP2.12$_{2\text{-}8}$ peptide and *Mm*ADO (PDB: 7LVZ, chain D). Binding energy calculated in kcal mol^{-1}. Run 5 selected.
B: *At*RAP2.12$_{2\text{-}8}$ peptide (sticks, grey) docked into *Mm*ADO (**A** – run 5). Hairpin loop residues (Y198-C206) coloured purple.
C: Close-up view of Nt-Cys of *At*RAP2.12$_{2\text{-}8}$ docked into *Mm*ADO (**B**), showing active site Fe (sphere, orange), 3 Fe-binding His residues (lines), Phe-Asp-Leu motif (lines), and Nt-Cys of *At*RAP2.12$_{2\text{-}8}$ (sticks, grey). Nt-amine within H-bonding distance of Asp192 (lines, teal). Distances between Nt-amine and Asp192, and Nt-thiol and Fe, are given in angstrom in yellow.
D: Close-up view of first 4 residues of *At*RAP2.12$_{2\text{-}8}$ docked into active site cavity of *Mm*ADO (**B**).

6.7 Discussion

In this Chapter, crystallisation trials were performed with *Hs*ADO protein under both aerobic and anaerobic conditions, and with and without substrate-like peptides (Section 6.3). Unfortunately, these trials were unsuccessful; however, they provided information that could be useful for planning future trials as some crystallisation conditions could now be eliminated. An *in silico* structure of *Hs*ADO was proposed using

MODELLER[252] and the structure of *At*PCO4[66] as a template (Section 6.4). This *in silico* structure bore a close resemblance to the two experimentally determined crystal structures of ADO published by Wang *et al.*[5] and Fernandez *et al.*[6] in 2021. As expected from sequence similarity and function, structures of ADO were more similar to those of the *Arabidopsis thaliana* PCOs, than to those of the other mammalian thiol dioxygenase CDO.

ADO structures featured the double stranded β helix (DSBH) that is common to proteins in the cupin superfamily[259,260]. In these structures, the active site metal was coordinated by three His residues, as seen in structures of mammalian CDO[156,157,241] and *At*PCOs[66,67]. The active site cavity in ADO structures was more open than in CDO and a hairpin loop was observed near the entrance to the active site, as seen in structures of the *At*PCOs. The hairpin loop is proposed to be involved in substrate binding[66] and has not been observed in CDO structures[4]. The sequence of the hairpin loop in ADO is rich in acidic residues. This is particularly interesting when considered alongside the results from *in vitro* peptide activity assays with 'anti-substrate' sequences (Chapter 2, page 47). These results demonstrated that the C-terminal half of 14-mer peptides could have a significant influence on ADO activity, and identified a DD/EE motif as a potential feature of poor ADO substrate peptides. It seems possible that this motif could interact unfavourably with the acidic residues in the hairpin loop and negatively impact peptide binding to ADO.

Furthermore, a Cys-Tyr crosslink, previously observed in ADO by mass spectrometry and by NMR in the presence of cysteamine and oxygen[178], was not seen in either published crystal structure. This does not necessarily disprove the existence of the crosslink as it would not be expected to form in the absence of small molecule substrate, oxygen, or iron[5]. A similar Cys-Tyr crosslink has been observed in CDO structures and has been linked to increased catalytic activity[156,157,159–163,172,241,261]. Given this, it seems most likely that the crosslink would be linked to ADO's activity towards the small molecule cysteamine, if it is

present, rather than ADO's NCO activity. Interestingly, Fernandez *et al.* noted the presence of a smaller tunnel accessing the active site of *Mm*ADO from the opposite side to the main active site cavity, which they proposed could be the route by which O_2 reaches the active site[6]. This tunnel was also observed in the experimental[5] and *in silico*[4] (Section 6.4) structures of *Hs*ADO (Appendix 17). Fernandez *et al.* proposed that O_2 access via this tunnel could be controlled by two cysteine residues in response to changes in redox environment[6].

Crystallisation trials with substrate-like peptide inhibitor KH1 were performed to investigate substrate binding to ADO, using broad screens and conditions informed by Wang *et al.*[5] and Fernandez *et al.*[6] (Section 6.5). These trials were once again unsuccessful, but could be used to help plan future crystallisation experiments; these could include repetition of co-crystallisation broad screens in anaerobic conditions or with active site iron replaced with nickel to assist with protein stabilisation.

In the absence of crystal structures to show ADO/substrate complexes, docking experiments were conducted using AutoDock4[251] to investigate the binding of three peptide substrates (*Hs*RGS4$_{2\text{-}8}$, *Hs*IL32$_{2\text{-}8}$, and *At*RAP2.12$_{2\text{-}8}$) to the structure of iron-bound *Mm*ADO[6] (Section 6.6). Docking results for all three peptides demonstrated that the Nt-amine had the potential to interact with the Asp residue in the Phe-Asp-Leu motif. This had previously been observed in computational studies by Fernandez *et al.*[6] with RGS5$_{2\text{-}5}$. This Asp residue is conserved in *Hs*ADO and *At*PCOs, but not in CDO[66]. In addition, mutation of this residue in *At*PCO4 significantly reduced catalytic activity and iron incorporation relative to the wild-type enzyme[66]. The importance of this interaction is further highlighted by results from *in vitro* experiments with ADO and acetylated peptide substrates, which showed that acetylation of the Nt-amine abrogated ADO-catalysed oxidation (Chapter 4, page 95). Acetylation of the Nt-amine would neutralise its charge and prevent hydrogen bonding between the substrate amine and the carboxylate of the active site Asp. It would also increase

the size of the Nt-amine, potentially introducing steric clash that would disfavour substrate binding.

Additional structural features of ADO were identified that could potentially improve binding of certain peptides into the active site. A polar pocket was identified in docking experiments with $RGS4_{2-8}$ that was positioned to interact with a basic residue after the Nt-Cys in a substrate (Section 6.6.1). The residues in this pocket (*Mm*ADO Tyr75, Glu80 and Ser85) are conserved in *Hs*ADO, *At*PCO4 and *At*PCO5, but not in CDO. In substrate screening assays with Nt-Cys peptides based on sequences from the human proteome (Chapter 2, page 41), five of the seven peptides with which *Hs*ADO had > 95 % initial activity relative to $RGS4_{2-15}$ featured an Arg, Lys, or His residue after the Nt-Cys. Subsequent experiments with 'optimum substrate' peptides confirmed that a basic residue after the Nt-Cys was favourable for *Hs*ADO activity. Docking results with $RGS4_{2-8}$ described in this Chapter (Section 6.6.1) suggest that the higher initial activity of *Hs*ADO with these peptides could be due to improved binding, facilitated by H-bonding interactions between the basic residue and the polar pocket residues. In these docking results, Glu80 (an acidic residue) is the closest *Mm*ADO residue to the basic Lys residue that follows the Nt-Cys in $RGS4_{2-8}$; the distance between the Lys amine nitrogen and a carboxylate oxygen in Glu80 was 2.3 Å. In Chapter 2 (page 41) and Chapter 3 (page 66), sequences with an acidic residue after the Nt-Cys were observed to be poor *Hs*ADO substrates (e.g., MTPN, 'anti-substrate' peptides AS3/4/9/10). The presence of the acidic side chain of Glu80 (*Mm*ADO numbering) in the polar pocket could help to explain why ADO does not readily accept substrates with an acidic residue after the Nt-Cys as the negatively charged carboxylate groups could repel each other.

In docking experiments with $IL32_{2-8}$ the possibility for a π-π binding interaction was noted between *Mm*ADO Tyr198 and IL32 Phe3, and five other aromatic residues were identified

in the active site that were highly conserved in *Hs*ADO, *At*PCO4 and *At*PCO5, but not in CDO (Section 6.6.2). In Chapter 2 (page 41), some of the peptides with which *Hs*ADO had the highest *in vitro* activity had an aromatic residue after the Nt-Cys: IL32, LIPI, and OS7. The OS7 sequence also behaved as a good *Hs*ADO substrate in the cell reporter assay (Chapter 3, page 66). The prevalence of aromatic residues in the active site cavity of ADO suggests that there is potential for π-π interactions to stabilise binding of substrates with an aromatic residue after the Nt-Cys.

In contrast to docking results with $RGS4_{2-8}$ and $Il32_{2-8}$, no additional potential binding interactions were identified between *At*$RAP2.12_{2-8}$ and *Mm*ADO. This could help to explain why *Hs*ADO had lower catalytic efficiency *in vitro* with *At*$RAP2.12_{2-15}$ peptide than with $RGS4_{2-15}$, $RGS5_{2-15}$, and $IL32_{2-15}$ peptides (Chapter 2, page 54). No potential negative interactions were observed either, suggesting that the *At*RAP2.12 N-terminus may be readily accommodated by *Hs*ADO, but binding cannot be stabilised by H-bonding or π-π interactions. In light of this, it is interesting that many of the potential substrate binding residues are conserved in the *At*PCOs. ADO catalytic efficiency towards *At*$RAP2.12_{2-15}$ peptide (1.22×10^4 M^{-1} s^{-1}) is comparable with that previously determined for the *At*PCOs (8.39×10^3 - 4.94×10^4 M^{-1} s^{-1})[62]. This suggests that the potential substrate binding residues conserved in the PCOs may have a similar effect on catalytic activity as they appear to in ADO, and raises questions about the evolution of these enzymes and their substrates.

ADO-like proteins have been identified in eukaryotic organisms, including sponge, protists, and fungi[50,265]. A selection of these proteins was investigated *in vitro* by Thomas Bancroft (Flashman Group); Mr. Bancroft measured the NCO activity of ADO-like proteins from *Danio rerio* (zebrafish), *Amphimedon queenslandica* (sponge), *Spizellomyces puncatus* (fungus), and *Leishmania major* (protist). All ADO-like proteins tested (*Dr*ADOa, *Aq*ADO, *Sp*ADO, and *Lm*ADO) were able to catalyse oxidation of $RGS5_{2-15}$ peptide, although the

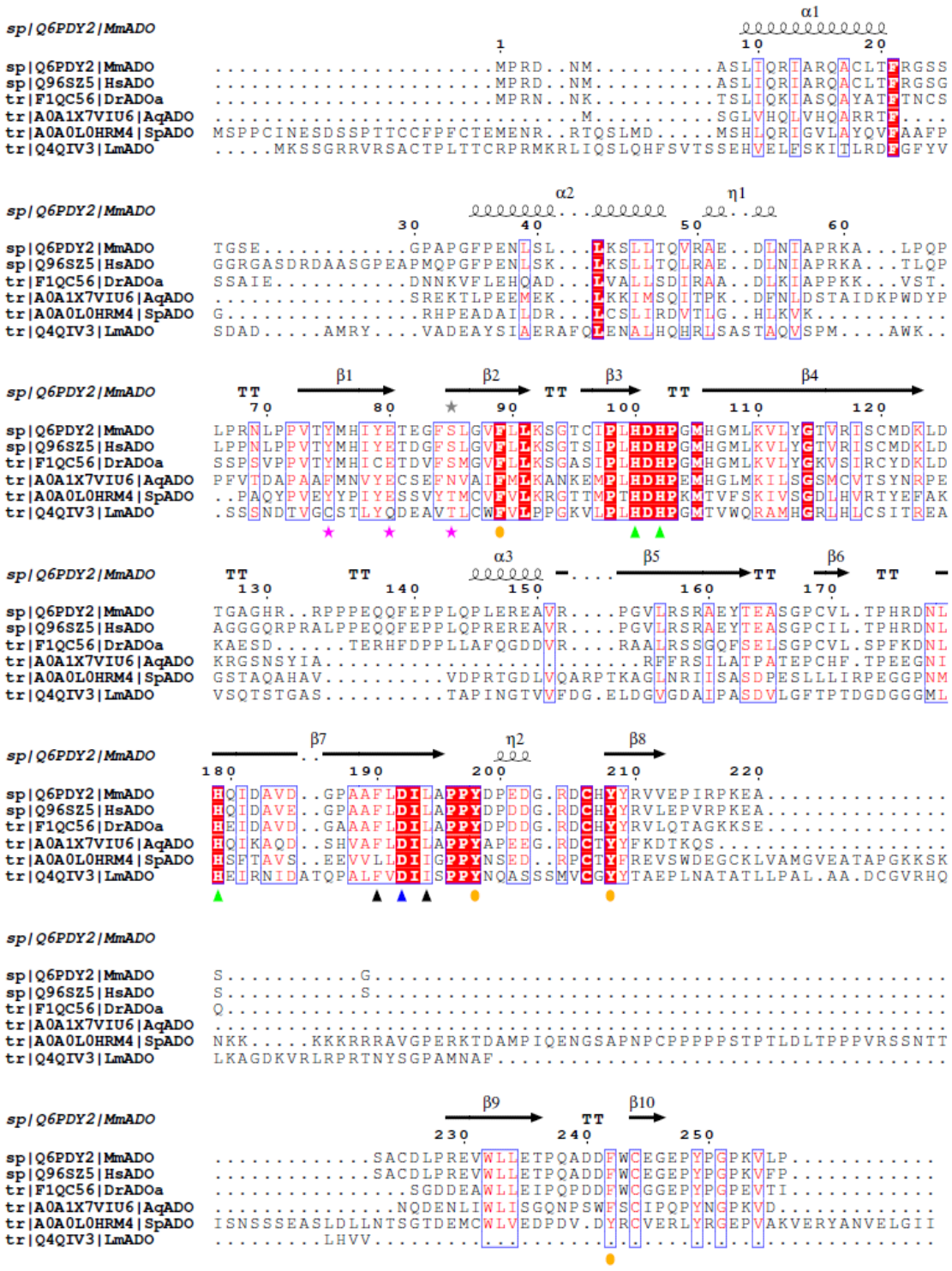


Figure 6.11: Alignment of ADO-like proteins from organisms across different eukaryotic kingdoms
CLUSTAL O(1.2.4) multiple sequence alignment of eukaryotic ADO-like proteins rendered with ESPript 3.0 server. Fe-binding His residues marked with green triangles, Phe-Asp-Leu motif (ADO) marked with black or blue triangles (Asp192 (*Mm*ADO numbering) marked with blue triangle), polar pocket (identified in Fig. 6.6) residues marked with pink stars, and aromatic active site residues (Fig. 6.8) marked with orange circles (with the exception of Tyr75 and Phe190 (*Mm*ADO numbering), which are part of the polar pocket and Phe-Asp-Leu motif (ADO) respectively).
*Mm*ADO = mouse ADO; *Hs*ADO = human ADO; *Dr*ADOa = zebrafish ADO a; *Aq*ADO = ADO-like protein in sea sponge (*Amphimedon queenslandica)*; *Sp*ADO = ADO-like protein in fungus (*Spizellomyces puncatus*); *Lm*ADO = ADO-like protein in protist (*Leishmania major*).

fungal protein, *Sp*ADO, had much lower activity than the other proteins. An alignment of these proteins with mouse and human ADO (Figure 6.11) shows that many of the potential substrate-binding residues identified in this Chapter are well conserved. At least one of the three residues in the proposed polar pocket (*Mm*ADO Tyr85, Glu80, Ser85) is present in each of the ADO-like proteins, and many of the aromatic residues observed in the active site of *Mm*ADO are also highly conserved (Section 6.6.2). This provides evidence that these residues could be important for NCO functionality. Furthermore, the Phe-Asp-Leu mammalian ADO active site motif is present in all except *Sp*ADO and *Lm*ADO, which possess a Leu-Asp-Ile and Phe-Asp-Ile motif, respectively. The residue predicted to bind to the Nt-amine of a substrate, Asp192 (*Mm*ADO numbering), is conserved in all proteins, further highlighting its importance to the function of NCO proteins.

Overall, in this Chapter, *in silico* modelling of *Hs*ADO identified structural features shared with *At*PCO enzymes that could be linked to their function as NCOs[4]. These structural features were also seen in published crystal structures of *Hs*ADO[5] and *Mm*ADO[6]. Additional features were identified through docking studies with *Hs*RGS4$_{2\text{-}8}$, *Hs*IL32$_{2\text{-}8}$, and *At*RAP2.12$_{2\text{-}8}$ peptides that provided clues to the manner in which ADO binds substrate peptides and proteins. These docking results also provided potential explanations for some *in vitro* results in Chapters 2 and 4 (pages 41 and 95).

7 Conclusions

Cysteamine dioxygenase (2-aminoethanethiol dioxygenase, ADO) is a thiol dioxygenase enzyme from the cupin family of proteins that has recently been linked to O_2-dependent protein stability, and therefore the hypoxic response, in humans[1]. ADO was first characterised as a small molecule thiol dioxygenase, converting cysteamine to hypotaurine as part of the taurine biosynthesis pathway[166]. However, in 2019, ADO was also shown to catalyse oxidation of the N-terminal cysteine (Nt-Cys) of target proteins, which facilitated their degradation via the Arg/N-degron pathway[1,2]. This activity was inhibited by hypoxia, meaning that the stability of ADO's substrate proteins was O_2-dependent. Three substrates of ADO were characterised in this context: regulators of G-protein signalling, RGS4 and RGS5, and a chemokine, interleukin-32, IL32. The O_2-sensitivity of ADO was found to be similar to that of the prolyl hydroxylase (PHD) enzymes[1]; PHDs regulate the stability of hypoxia inducible factor (HIF) transcription factors, which have an important role in the human hypoxic response[15,35]. Furthermore, ADO is closely related to plant cysteine oxidase (PCO) enzymes, which are also thiol dioxygenases (TDOs). The PCOs are involved in the plant hypoxic response by catalysing oxidation of the Nt-Cys of Group VII ethylene response factors (ERF-VIIs) in an O_2-dependent manner, leading to degradation of ERF-VIIs in the Arg/N-degron pathway[3,56,58,62].

The aim of the work described in this Thesis was to investigate the role of ADO in O_2 homeostasis. At the commencement of this research project, ADO's role as an O_2-sensing N-terminal cysteine oxidase (NCO) had only recently been discovered, and just three protein substrates had been identified. Therefore, a key objective of this Thesis was to identify other potential protein substrates of ADO and to investigate characteristics of these that would facilitate further substrate searching within the human proteome. This would aid elucidation of the extent of ADO's role in O_2-dependent protein degradation. Additional

aims of this Thesis included screening for inhibitors of ADO, which could be useful as chemical probes or as potential therapeutic candidates, and investigation of ADO's structure, which could increase understanding of ADO's function and mechanism.

In Chapter 2 (page 30), mass spectrometry peptide activity assays were used to investigate ADO's substrate preferences *in vitro*. Removal of the initiator methionine to expose the Nt-Cys was demonstrated to be necessary for ADO-catalysed oxidation to occur. Thus, a protein must be a substrate of methionine aminopeptidase in order to be modified by ADO. In this Chapter, it was also shown that ADO would not catalyse oxidation of internal cysteines, confirming that ADO is specific for Nt-Cys. ADO was also found to accept peptides as short as seven amino acids long, which provided an initial indication that ADO may only require a small portion of a protein's N-terminus to facilitate substrate binding.

A wide range of Met-Cys-initiating proteins were next screened as ADO substrates *in vitro* using peptides representing the first 13 - 14 amino acids of their Met-excised N-terminus. Initial ADO activity towards these peptides was demonstrated to be highly substrate dependent over a one minute timescale, with no evidence of oxidation being detected for some peptides, while others were oxidised more rapidly than the positive control $RGS4_{2-15}$ peptide. This showed that ADO is selective amongst Nt-Cys sequences *in vitro*. The highest initial ADO activity *in vitro* was measured towards peptides that had a basic (H/K/R) or aromatic (F/W) residue at the position after the Nt-Cys. In contrast, peptides towards which ADO had lower initial activity had a higher prevalence of acidic amino acids in their sequences. These data suggested that ADO selectivity is determined by the physicochemical properties of the N-terminus, and that ADO preferentially catalyses oxidation of peptides with basic or aromatic residues after the Nt-Cys. These substrate preferences were verified using 'optimum substrate' (OS) and 'anti-substrate' (AS) peptides. Experiments with these peptides also identified a double acidic residue motif that, when present in the latter half of

the 14-mer peptides, was linked to low ADO activity. This, and other *in vitro* substrates preferences observed in this Chapter, were particularly interesting when considered within the context of ADO's structure (discussed below).

ADO had previously been shown to compensate for loss of PCO activity in *Arabidopsis thaliana* plants and to regulate the stability of *At*RAP2.12 (a PCO substrate) in human cells[1]. Therefore, ADO activity with *At*PCO substrate peptides was investigated *in vitro*. Chapter 2 (page 52) described peptide activity assays with nine different *At*PCO substrate sequences. Oxidation was detected on all *At*PCO substrate peptides following incubation with ADO for one minute; within these results a higher level of oxidation was measured on peptides that contained the most basic residues in their sequences. This provided additional validation of the substrate preferences of ADO observed with sequences from the human proteome. Notably, ADO initial activity was significantly lower towards all *At*PCO substrate peptides than its known substrate from humans (the positive control peptide, $RGS4_{2\text{-}15}$). Kinetic analysis of ADO with *At*$RAP2.12_{2\text{-}15}$ peptide found that ADO had lower catalytic efficiency with *At*$RAP2.12_{2\text{-}15}$ (1.22×10^4 M^{-1} s^{-1}) than with all known human substrate peptides ($1.64 - 3.07 \times 10^5$ M^{-1} s^{-1})[1]. However, catalytic efficiency with all peptide substrates was at least two orders of magnitude higher than that measured by Dominy *et al.* for ADO with its small molecule substrate, cysteamine[166]. This suggested that ADO's primary role is likely to be as an NCO, rather than as a small molecule TDO. As a weak inhibitor of NCO activity[1], cysteamine may be able to compete with protein substrates when the cellular concentration is high enough. Therefore, ADO's small molecule role is likely still biologically relevant, particularly in cases where cysteamine may reach cytotoxic levels ($10^{-4} - 10^{-3}$M)[259].

In Chapter 3 (page 59), *in cellula* experiments were described that investigated the regulation of putative ADO substrates in human cells. First, a representative selection of

N-termini from peptide activity assays (Chapter 2, page 41) were studied using a dual-fluorescence oxygen reporter (DFOR) assay (developed by Dr. Thomas Keeley (Nuffield Department of Medicine))[204]. In this assay the first 15 amino acids of N-termini of interest were fused to the N-terminus of EGFP; if the N-terminus added was then a substrate of ADO in the N-degron pathway, the stability of EGFP would become ADO dependent and could be altered using 2,2-dipyridyl or hypoxia to inhibit ADO activity. The stability of EGFP could be monitored by fluorescence, which was measured relative to fluorescence of an mCherry control that was expressed at the same level as EGFP in cells transfected with DFOR constructs. Fifteen different Nt-sequences were investigated in this context: *in vitro* ADO activity towards representative peptides of these sequences varied from 0 to 152 % relative to $RGS4_{2\text{-}15}$. This selection therefore allowed the relationship between *in vitro* and *in cellula* ADO activity to be probed. Sequences chosen also had varying physicochemical properties, so the substrate preferences identified in Chapter 2 could be tested *in cellula*. One of the sequences included in the DFOR assay was not selected from peptide activity assay results obtained in Chapter 2 (page 41): *Mm*Tex19.1. This protein was of interest as it had previously been shown to interact with and be stabilised by E3 ligase UBR2[211], and was a candidate ADO substrate.

DFOR assay results correlated well with *in vitro* peptide activity data, with some notable exceptions. Addition of the N-termini of RGS4, ANKRD29, SUSD6, *At*RAP2.12, and OS7 (the sequence of the best optimum substrate peptide) all conferred both 2,2-dipyridyl and hypoxia dependent stability onto EGFP. This suggested that these Nt-sequences were substrates of ADO *in cellula*. Representative peptides of these sequences all had > 16 % initial ADO activity *in vitro* relative to $RGS4_{2\text{-}15}$. The N-termini of GFPT2 and *Mm*Tex19.1, whose representative peptides had around 7 % initial ADO activity *in vitro* relative to $RGS4_{2\text{-}15}$, had a similar but reduced effect on EGFP stability, indicating that these

sequences could also be ADO substrates. However, two Nt-sequences whose representative peptides also had > 16 % initial ADO activity *in vitro* relative to $RGS4_{2-15}$ did not behave as ADO substrates in this assay: AAAS and GPX1. It was postulated that this could be due to another Nt-modification potentially preventing ADO-catalysed oxidation. Another possibility was that these sequences were ADO substrates, but were not substrates of other components of the N-degron pathway (e.g., ATE1). Chapter 3 (page 66) described another interesting observation from DFOR data. The addition of some N-termini (AAAS, JUNB, GPX1, LYPLA2, and MTPN) appeared to confer increased stability on the EGFP that was not 2,2-dipyridyl dependent. A possible explanation for this is that a stabilising modification occurred on these N-termini.

Chapter 3 (page 76) also described experiments conducted by Dr. Keeley (NDM) investigating the regulation of full-length putative ADO substrate proteins in human cells. Notably, these results demonstrated that the stability of GFPT2 was regulated by ADO only when ADO was over-expressed *in cellula*. This aligned with peptide activity assay and DFOR results, which indicated that the N-terminus of GFPT2 was a weak substrate *in vitro* and *in cellula.* Interestingly, the stability of two proteins that were expected to be good ADO substrates based on peptide activity and DFOR results, ANKRD29 and SUSD6, did not appear to be regulated by ADO *in cellula*. It was theorised that this could be a consequence of protein folding; their N-termini could be shielded from modification through interaction with other parts of the protein. This phenomenon had been previously observed in rice, where an Nt-Cys-initiating protein SUB1A-1 is protected from PCO-catalysed oxidation by interaction of its N- and C-termini[77].

Unexpectedly, the stability of STAC2 protein was ADO-dependent *in cellula*, in contrast to peptide activity and DFOR assay results. It was hypothesised that while the N-terminus of STAC2 is unfavourable for ADO binding, there could be another binding site on the

full-length protein that may be able to compensate. The isoform of STAC2 used in this Thesis was not the canonical isoform, but is instead predicted by alternative splicing. The canonical isoform of STAC2 does not have an Nt-Cys, therefore this protein has not been pursued further as a potential new substrate of ADO. Importantly, the experiments described in Chapter 3 (page 66) demonstrated that there are factors beyond *in vitro* peptide activity that must be considered when attempting to identify new protein substrates of ADO *in cellula.*

One of the factors postulated to be influencing ADO activity *in cellula* was the presence of another Nt-modification. Acetylation of the Nt-amine is one of the most common Nt-modifications of human proteins[85]. Nt-Cys acetylation had been postulated to be catalysed by N-terminal acetyltransferase A (NatA) in humans. However, definitive evidence of this was lacking in the literature. Chapter 4 (page 82) described experiments conducted *in vitro* to investigate NatA activity with Nt-Cys peptides, in collaboration with Prof. Thomas Arnesen and Dr. Matti Myllykoski (University of Bergen). Radioactivity-based peptide activity assay results (kindly provided by Dr. Myllykoski) demonstrated that NatA could catalyse acetylation of Nt-Cys peptides *in vitro*, but that NatA was selective amongst Nt-Cys sequences, preferentially catalysing acetylation of peptides with an acidic or polar residue after the Nt-Cys. This substrate preference was distinct from that observed with ADO (Chapter 2, page 41), suggesting that these two Nt-modifying enzymes may have evolved to act on separate pools of Nt-Cys-bearing proteins. Tandem mass spectrometry (conducted in Oxford with NatA kindly provided by Dr. Myllykoski) was used to verify that NatA-catalysed acetylation occurred on the Nt-Cys of a $JUNB_{2\text{-}15}$ peptide. These results provided the first direct evidence that NatA catalysed Nt-Cys acetylation *in vitro*.

Interestingly, AAAS was among the sequences identified as the best NatA substrates *in vitro*, as well as MTPN and JUNB. These N-termini were observed to stabilise EGFP in the DFOR assay independently of 2,2-dipyridyl treatment (Chapter 3, page 66). This is an initial indication that acetylation of these N-termini could possibly be a stabilising modification. NatA has recently been linked to stabilisation of protein substrates of IAP family E3 ubiquitin ligases[109], so this would not be unprecedented. AAAS was one of the N-termini expected to be a substrate of ADO in the DFOR assay based on peptide activity data, so it was hypothesised that Nt-acetylation could protect against ADO-catalysed oxidation. This hypothesis was supported by mass spectrometry assays with Nt-acetylated peptides, which demonstrated that ADO activity was significantly reduced or ablated with acetylated versions of ADO substrate peptides (Chapter 4, page 95). Similar assays with Nt-oxidised peptides and NatA showed that NatA activity was also significantly reduced with oxidised versions of substrate peptides (Chapter 4, page 97). These data suggested that Nt-Cys acetylation and oxidation could be mutually exclusive modifications, and raised the possibility that co-translational acetylation could shield some Nt-Cys proteins from ADO-catalysed oxidation.

GPX1 was another sequence that was predicted to be an ADO substrate based on peptide activity data, but whose N-terminus did not confer 2,2-dipyridyl-dependent stability onto EGFP in the DFOR assay (suggesting that it was not an *in cellula* substrate of ADO in the Arg/N-degron pathway). This sequence was not among those identified as a NatA substrate *in vitro*, however. There are a variety of other Nt-modifications that could also influence ADO activity *in cellula*. Activity data for Nt-formylated (Chapter 2, page 37) and Nt-acetylated (Chapter 4, page 95) peptides *in vitro* suggest that modifications to the Nt-amine are likely to prevent ADO activity. Furthermore, Nt-fatty acid modifications such as palmitoylation can promote membrane binding[78], which could also limit ADO activity.

Future work could explore the effect of other Nt-modifications on ADO activity, which could uncover an explanation for GPX1's lack of regulation *in cellula*. It is also worth noting that cellular assays used in this Thesis were only able to identify ADO substrates that were also substrates of the other components of the N-degron pathway. It is formally possible that GPX1 and other *in vitro* peptide substrates undergo Nt-Cys oxidation *in cellula* but are not degraded; this would imply an alternative signalling role for ADO that would be an interesting future avenue of investigation. Nt-proteomic studies would be useful to identify Nt-Cys oxidation of GPX1 and other proteins that are not degraded *in cellula* (for example ANKRD29 and SUSD6).

Chapter 5 (page 101) described *in vitro* experiments investigating small molecule and peptide inhibitors of ADO. Small molecules and peptides have different advantages and disadvantages as inhibitors. Peptides are likely to be more specific and potent due to their larger potential surface area for target binding, while small molecules are more bioavailable as their small size renders them more cell permeable. Therefore, both avenues were pursued in this Thesis. A selection of substrate-based peptides was screened against ADO *in vitro*. Peptides based on the N-termini of RGS4/5 and IL32, and on the best optimum substrate peptide OS7, were found to be inhibitory when the Nt-Cys was replaced with an Nt-selenocysteine (Nt-Sec), with *in vitro* IC_{50} values of 63 - 133 μM. Dimerisation of these peptides was observed, likely via a diselenide bond. However, the influence of this dimerisation on inhibitor activity could not be investigated using established assay conditions. Preliminary results from a mode of inhibition experiment suggested that these peptides could be competitive inhibitors. Importantly, these Nt-Sec peptides represent useful candidates for ADO-specific chemical probes.

Two small molecule iron chelators, 2,2-dipyridyl and a 2,2-dipyridyl derivative DP44mT, were characterised as ADO inhibitors *in vitro*. 2,2-dipyridyl had previously been shown to

inhibit ADO activity *in cellula*[1], so was expected to also inhibit *in vitro*. DP44mT is a more potent iron chelator than 2,2-dipyridyl[231], so was also predicted to be a more potent ADO inhibitor *in vitro*. Mass spectrometry assays demonstrated that both small molecules inhibited ADO, with IC_{50} values of 522 and 125 μM for 2,2-dipyridyl and DP44mT, respectively. An initial mode of inhibition experiment with 2,2-dipyridyl suggested that these small molecules could also be competitive inhibitors of ADO, which would indicate that they bind in the active site of ADO. As these molecules are iron chelators it seems most likely that they would bind to the active site iron. These small molecule inhibitors are likely to be less useful as specific ADO inhibitors as they could influence activity of other iron-dependent enzymes. However, they can be useful when used in conjunction with suitable controls.

The final aim addressed by this Thesis was investigation of ADO's structure. At the commencement of this project no structures of ADO had been published. Therefore, crystallisation trials were conducted to attempt to resolve the structure of ADO. These trials are described in Chapter 6 (page 125). Trials were conducted in aerobic and anaerobic conditions, and in the absence and presence of substrate-like peptides. No potential protein crystals were observed in the conditions trialled. However, in 2021 two crystal structures were published[5,6]. Details published by Wang *et al*. regarding their successful crystallisation conditions[5] suggested that protein concentration may have been too low in the trials described in Chapter 6.

Chapter 6 (page 128) also described production of an *in silico* model of ADO's structure by comparative protein modelling. A structure of *At*PCO4 obtained by White *et al.*[66] was used as a template for protein modelling, which was performed using MODELLER[252]. The *in silico* structure obtained resembled the experimentally determined crystal structures published in 2021[4–6]. These structures revealed that ADO contained the characteristic cupin

double stranded β helix (DSBH) and 3-His iron-binding site, as expected from sequence comparison with other thiol dioxygenases[4–6]. ADO was also observed to have a hairpin loop near the entrance to the active site. This hairpin loop had also been seen in structures of PCOs and had been proposed to be involved in substrate binding[66]. In ADO, the hairpin loop is rich in acidic residues, which correlates with the observation that peptides and N-termini rich in acidic residues appear to be poor substrates of ADO. Fernandez *et al.* also noted the existence of a small tunnel on the opposite side of ADO to the main active site cavity, which was also observed in other *in silico* and experimental structures[4–6]. The authors proposed that this tunnel could be involved in O_2 delivery to the active site, with two cysteine residues providing 'gated' access in response to changes in redox environment[6]. The DFOR assay used in Chapter 3 (page 60) could be a useful setting to interrogate this, for example using live fluorescence measurements in a plate reader equipped with temperature and atmosphere control, as described by Smith and Keeley[204].

Finally, Chapter 6 (page 134) addressed the need to understand ADO substrate binding through experimental and *in silico* studies. Further crystallisation trials were conducted using a peptide inhibitor characterised in Chapter 5 (page 106). It was hoped that this inhibitor would bind in a similar manner to substrate, while also stabilising ADO to increase the likelihood of crystal formation. Unfortunately, these trials did not yield any promising protein crystals, despite using an increased protein concentration (in line with Wang *et al.*'s crystallisation conditions[5]). These experiments did, however, provide information that could be used to inform future crystallisation trials.

Docking experiments were conducted with selected peptide substrates: human ADO substrates $RGS4_{2\text{-}8}$ and $IL32_{2\text{-}8}$, and a PCO substrate *At*$RAP2.12_{2\text{-}8}$. These *in silico* experiments identified residues within ADO's active site that could be involved in substrate binding. Computational studies previously conducted by Fernandez *et al.* had noted a

potential interaction between the amine group of a peptide substrate and an Asp residue in ADO's active site[6]. AutoDock4 experiments described in Chapter 6 (page 136) corroborated this finding. These results aligned with the observation that acetylation of the Nt-amine prevents oxidation (Chapter 4, page 95); acetylation would prevent H-bonding between the Asp carboxylate and the Nt-amine, disfavouring substrate binding and therefore turnover. Furthermore, a 'polar pocket' was observed in ADO's active site, which was well positioned to interact with basic residues after the Nt-Cys of a substrate peptide. This polar pocket was highly conserved in different NCO sequences, and could help to explain why sequences with basic residues after the Nt-Cys are good ADO substrates.

Overall, the work described in this Thesis represents a significant contribution to the field of ADO research, and, more broadly, to research on O_2-sensing and thiol dioxygenases. ADO has been demonstrated to be highly substrate specific, with implications for its role in O_2-dependent protein stability. Interestingly, ADO substrate preferences were found to depend on physicochemical properties of residues downstream of the Nt-Cys, and to be distinct to those of another Nt-Cys modifying enzyme, NatA. This is not a unique phenomenon; amino acids downstream of the N-terminus have been shown to influence the selectivity of other Nt-modifying enzymes, such as methionine aminopeptidases[79], arginyl transferases[137], and other NATs[99]. In addition, novel peptide inhibitors were identified, which have the potential to be developed as chemical probes for future ADO research. Finally, investigation of ADO's structure provided justification for some of the results obtained in Chapters 2 – 4, and demonstrated how ADO's structure is tailored to its function as an NCO.

Future work could include continued pursuit of a substrate-bound structure of ADO, potentially utilising the Nt-selenocysteine peptide inhibitors characterised in this Thesis. A substrate-bound structure would confirm the accuracy of the predictions made in this Thesis

regarding potential binding residues in ADO's active site, and would provide useful information for elucidation of ADO's mechanism and for inhibitor design. It would also be of interest to characterise ADO's O_2-sensitivity with additional substrate peptides to further investigate its role as on O_2-sensor in protein degradation. The O_2-sensitivity of another O_2-sensing enzyme, FIH, has been shown to be substrate-dependent[33,35]. It would be interesting to know if ADO's role in the hypoxic response may be tailored towards different proteins. Nitric oxide has previously been linked to degradation of proteins via the Arg/N-degron pathway[63,128,180], and Masson *et al.* reported that the presence of an NO donor reduced hypoxic stabilisation of RGS4 protein in wild-type, but not ADO KO, cells[1]. Therefore, it would be valuable to biochemically characterise the effect of NO on ADO activity in hypoxic and normoxic conditions.

Furthermore, little is known of ADO's mechanism, or of the mechanism of thiol dioxygenases in general. This could be addressed using stopped-flow and quench flow experiments to identify reactive iron-oxygen species and the rate-limiting step in ADO catalysis. Knowledge of enzyme mechanism can be particularly useful for inhibitor design using the structures of reaction intermediates. As noted previously, potent and specific inhibitors of ADO would be useful as therapeutic candidates and chemical probes to investigate ADO's role in the hypoxic response further.

There is still much to be uncovered about ADO and its role as an O_2-sensor in animals. Its regulation of known substrates, RGS4/5 and IL32, could be important in the physiological response to hypoxia, and there may be other novel substrates that could also have significant roles[1]. O_2-dependent protein signalling by ADO can facilitate a rapid response to changes in O_2 levels, which takes place over a shorter timescale than the transcriptional response controlled by the HIF transcription factors and could have a key role in acute hypoxic disease states[1,4]. Results in this Thesis will help direct novel substrate identification efforts

towards likely candidates. Furthermore, as NCOs appear to be conserved among eukaryotes, knowledge of ADO could inform research into O_2-sensing systems in other organisms[4,49,50,265]. It will be interesting to probe the relationship between Nt-Cys oxidation and acetylation in plants, as this may dictate which Nt-Cys proteins are hypoxically regulated. The structural work presented here will help define NCO/substrate interactions that may be useful across broader research areas. In summary, the results presented here have begun to answer some of the research questions surrounding ADO and will be useful for future studies of O_2 homeostasis and NCOs.

8 Materials and Methods

Graphs were produced and statistical analyses performed using GraphPad Prism v5 or v8. Experimental procedures were conducted at r.t., unless otherwise stated.

8.1 Biochemical *in vitro* Experiments

8.1.1 Recombinant protein production

LB agar and 2 TY media used during recombinant protein production were prepared according to recipes in Table 8.1. After mixing the components, Milli-Q water was added up to the required volume and the mixture was autoclaved. SOC medium was purchased from Sigma Aldrich. Kanamycin was used at a final concentration of 40 μg/mL.

Table 8.1: Media recipes for recombinant protein production

	LB agar (per 100 mL)	2 TY media (per 1 L)
Tryptone	1.0 g	16.0 g
Yeast extract	0.5 g	5.0 g
NaCl	1.0	10.0 g
KCl	-	-
NaOH (1 M)	-	-
Agar	1.5 g	-

8.1.1.1 Transformation of E. coli with ADO plasmid

The plasmid for expression of recombinant *N*-His_6-tagged $ADO_{1–270}$ protein (*Hs*ADO) in *E. coli* was purchased from GenScript using the pET-28a(+) vector. 35 μL of Rosetta™ (DE3) competent *E. coli* cells (Merck) were mixed with 50 ng of *Hs*ADO plasmid and incubated on ice for 20 min. The mixture was heat shocked at 42 °C for 45 s then left on ice for 2 min. Cells were then added to SOC media (200 μL) and incubated at 37 °C for approximately 1 hour before being plated on an LB agar plate containing kanamycin. The plate was allowed to dry for 10 min then incubated at 37 °C for 18 hours. The plate was subsequently stored at 4 °C.

8.1.1.2 Protein expression in E. coli

In preparation for large-scale protein production, a single colony was selected from the transformation plate and used to inoculate 100 mL of 2 TY medium containing kanamycin. The mixture was incubated at 37 ℃ overnight with shaking (180 rpm). 12 × 600 mL flasks of 2 TY medium containing kanamycin were each inoculated with overnight culture (6 mL) and incubated at 37 ℃ with shaking (180 rpm). Cultures were grown to an $O.D._{600}$ of 0.6 - 0.8 and induced with IPTG (0.5 M, 600 µL per flask). Cultures were then incubated at 18 ℃ overnight with shaking (180 rpm), before harvesting by centrifugation (5000 rpm (~10,000 g), 10 min). The harvested cell pellet was frozen at −80 ℃.

8.1.2 Recombinant protein purification

Proteins were purified using an AKTA™ Fast Protein Liquid Chromatography (FPLC) system (Cytiva™ (formerly GE Healthcare)) at 4 ℃.

Buffers used for protein purification by FPLC were prepared in Milli-Q water according to recipes in Table 8.2 and adjusted to pH 7.5. Buffers were then filtered using 0.22 µm filters and degassed prior to use.

Table 8.2: Buffer recipes for FPLC protein purification

	Immobilized Metal Affinity Chromatography (IMAC)		**Size exclusion chromatography (SEC)**
	Binding Buffer (A)	**Elution Buffer (B)**	**SEC Buffer (C)**
Tris	50 mM	50 mM	50 mM
NaCl	400 mM	400 mM	400 mM
Imidazole	20 mM	1 M	-

8.1.2.1 Cell lysate preparation

0.5 – 1 tablet of cOmplete™ EDTA-free protease inhibitor cocktail (Sigma Aldrich), 20 µL DNase I (Thermo Fisher Scientific), and 100 µL DNase buffer (Thermo Fisher Scientific) were added to Buffer A (50 mL). Approximately 10 g of cell pellet was

suspended in this solution by gentle stirring at 4 °C. Following suspension, cells were lysed by sonication (5 cycles of 9.9 s on, 9.9 s off for 40 s, with stirring on ice). The lysed cells were centrifuged (48000 g, 30 min, 4 °C) and the supernatant was collected and filtered using a 0.45 µm syringe filter (Millipore).

8.1.2.2 Immobilized Metal Affinity Chromatography

The cleared lysate was loaded onto an equilibrated 5 mL HisTrap™ Ni(II) affinity chromatography column (Cytiva™) at a flow rate of 1 mL min^{-1} and washed with Buffer A until a steady UV absorbance baseline was reached. Protein was eluted in a gradient of 0 – 100 % (v/v) Buffer B in Buffer A (2 % B for 2 CV, then 2 – 100 % B over 2 CV, then 100 % B for 6 CV), and collected in 5 mL fractions. Fractions containing protein were identified through inspection of the UV trace and SDS-PAGE gel (as described below). These fractions were pooled for further purification.

8.1.2.3 Size exclusion chromatography

The pooled fractions from immobilised metal affinity chromatography (IMAC) were concentrated to approximately 3 mL using a 10,000 MWCO ultracentrifugation device (Millipore), and desalted using a PD-10 Sephadex G-25 Desalting Column (Cytiva™), buffer exchanging to Buffer C. A HiLoad™ 26/600 Superdex™ 75 pg column (Cytiva™) was equilibrated with Buffer C and loaded with desalted protein. Protein was eluted in Buffer C in 2 mL fractions. Fractions containing pure protein were determined by analysis of the UV trace and SDS-PAGE gel (as described below). These fractions were pooled, concentrated and aliquoted, before being flash-frozen in liquid N_2. Protein aliquots were stored at −80 °C. Protein concentration was determined using A_{280nm} measurements taken with using an LVis plate on a CLARIOstar® (BMG LABTECH).

8.1.3 SDS-PAGE gel analysis

SDS-PAGE gel analysis was carried out using either Mini-PROTEAN® TGX™ precast stain-free gel (Bio-Rad) or NuPAGE® 4 – 15 % Bis-Tris gel (Invitrogen™). A PageRuler™ Prestained Protein Ladder (ThermoFisher Scientific) was used on all gels as a molecular weight marker. Gels were run at 200 V for approximately 90 min (or until the dye front reached the bottom of the gel).

8.1.3.1 Mini-PROTEAN® TGX™ precast stain-free gel analysis

TGS running buffer (10 ×) was prepared in Milli-Q water using the components in Table 8.3, then diluted with Milli-Q water (1:10 v/v). Samples were mixed 1:1 with sample buffer (95 % (v/v) Laemmli sample buffer (Bio-Rad), 5 % (v/v) β-mercaptoethanol) and heated at 70 ℃ for 10 min before loading onto gel.

Table 8.3: TGS running buffer (10 ×)

Tris	Glycine	SDS
250 mM	1.92 M	1 % (w/v)

8.1.3.2 NuPAGE® 4-15 % Bis-Tris gel analysis

NuPAGE® MOPS SDS running buffer (ThermoFisher Scientific) (20 ×) was diluted with Milli-Q water (1:20 v/v). Samples were mixed 1:1 with sample buffer (95 % (v/v) NuPAGE® LDS sample buffer, 5 % (v/v) β-mercaptoethanol) and heated at 70 ℃ for 10 min before loading onto gel.

8.1.4 Fe quantification of recombinant protein

A stock solution of iron(II) ($(NH_4)_2SO_4$ $FeSO_4.6H_2O$) was prepared at 250 µM in 0.5 M HCl. This stock was diluted in 0.5 M HCl to prepare stock solutions at 25 µM intervals between 225 and 25 µM. Solutions of saturated ammonium acetate,

100 mM sodium metabisulfite, and 100 mM bathophenanthroline disulfonic acid were prepared in Milli-Q water.

The protein sample was prepared for analysis by adding 1.25 µL of 12 M HCl to 28.75 µL of 202.4 µM recombinant *Hs*ADO. This resulted in a 193.97 µM solution of ADO in 0.5 M HCl. The resulting mixture was heated at 80 °C for 10 minutes then centrifuged at 15000 rpm (16602 g) for 3 minutes to pellet precipitated protein. The supernatant was recovered, then centrifuged again (15000 rpm (16602 g), 3 min) to ensure that all protein was removed from the sample.

A master mix was prepared with 1350 µL of Milli-Q water, 500 µL sat. ammonium acetate, 125 µL of 100 mM sodium metabisulfite and 25 µL of 100 mM bathophenanthroline disulfonic acid. 80 µL of master mix was mixed with 30 µL of each of the iron(II) stock solutions (0 – 250 µM in 0.5 M HCl) and the protein sample solution. The reactions were incubated at r.t. for 10 minutes before transferring 3 x 30 µL aliquots from each reaction

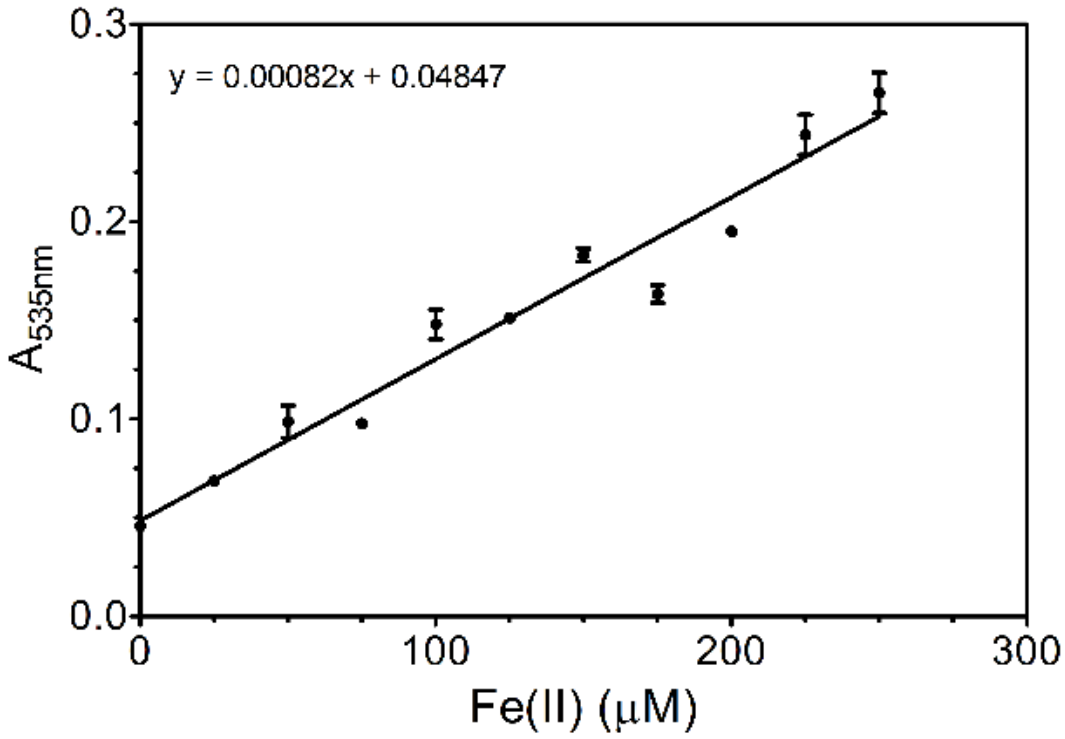


Figure 8.1: Iron(II) quantification graph
A_{535nm} readings from iron(II) plotted against stock concentration (µM), showing equation of line of linear regression in top left of graph.

into a clear 96 well plate (Greiner Bio-One). Absorbance was measured at 535 nm using a CLARIOstar® (BMG LABTECH). A_{535nm} readings from iron(II) stocks were plotted against concentration (μM) and a line of linear regression was fitted (Figure 8.1). The equation of the line of linear regression was used to calculated the concentration of iron(II) in the protein sample solution from its A_{535nm} reading.

8.1.5 Peptide synthesis

8.1.5.1 Solid phase peptide synthesis

Peptides were synthesised on 0.1 mmol or 0.05 mmol scale using a Liberty Blue™ automated microwave peptide synthesiser (CEM) by standard Fmoc-protected solid phase peptide synthesis (SPPS). Fmoc Rink Amide MBHA resin was used as the solid support. Fmoc-protected amino acids were dissolved in DMF at a stock concentration of 0.2 M. Oxyma Pure and *N,N'*-Diisopropylcarbodiimide (DIC) were used as coupling reagents. Synthesised peptides were all prepared with C-terminal amides.

8.1.5.2 Acetylation of peptide N-terminus

When acetylation of the N-terminus of the peptide was required, excess DMF was removed after SPPS to leave protected peptide attached to resin in approximately 5 mL DMF. Acetic anhydride (100 μL) was added and the mixture was incubated at r.t for 3 hours, mixing by vortex every 30 min. Peptides were then deprotected and cleaved from resin as described below.

8.1.5.3 Peptide deprotection and cleavage from resin

Protected peptide on resin was drained under vacuum and washed with DMF, followed by washing with DCM. The resin was then dried under vacuum. Approximately 5 mL of cleavage cocktail (92.5 % (v/v) trifluoroacetic acid (TFA), 2.5 % (v/v) triisopropylsilane (TIPS), 2.5 % (v/v) water, 2.5 % (v/v) 1,3- dimethoxybenzene) was added to the dry resin

and the resulting mixture was incubated for 3 hours at r.t, with occasional gentle shaking. The resin was then drained and the filtrate collected into 50 mL falcon tubes containing diethyl ether (approximately 40 mL), which had been cooled on ice. The resulting suspension was kept at −20 °C for at least one hour to aid precipitation of peptide. Solid peptide was collected by centrifugation (4700 rpm, 10–20 min, 4 °C). The peptide pellet was then redissolved in a mixture of acetonitrile (MeCN) and water, and lyophilised. Lyophilised crude peptide was stored at −20 °C.

8.1.5.4 Selenocysteine incorporation

Selenocysteine was incorporated into KH1 peptide using Fmoc-Sec(pMeOBzl)-OH (Sigma Aldrich) according to protocol kindly provided by Stefania Federico (Francis Crick Institute). For SPPS on the Liberty Blue™ peptide synthesiser, Fmoc-Sec(pMeOBzl)-OH was dissolved in DMF at a stock concentration of 0.2 M and used alongside other amino acid stocks. After SPPS, the protected peptide on resin was drained and washed as described above. Approximately 5 mL of cleavage cocktail (89 % (v/v) TFA, 5 % (v/v) DCM, 5 % (v/v) water, 1 % (v/v) TIPS) was added to the dry resin and the resulting mixture was incubated for 3 hours at r.t, with occasional gentle shaking. The resin was drained and the cleaved peptide was collected by cold ether precipitation and centrifugation as described above. 1 mL of TFA and 1 mL of bromotrimethylsilane were added to dried peptide and the resulting mixture was stirred at r.t. overnight under an N_2 atmosphere. Fully deprotected peptide was then collected by cold ether precipitation and centrifugation as described above. The peptide pellet was then redissolved in a mixture of acetonitrile (MeCN) and water, and lyophilised. Lyophilised crude peptide was stored at −20 °C.

8.1.5.5 Peptide purification by high-performance liquid chromatography

Lyophilised crude peptide was dissolved in the minimum volume of MeCN and Milli-Q water mixture. These solutions contained the minimum % (v/v) MeCN at which the

peptide remained soluble. The peptide solution was filtered using a 0.45 µm syringe filter prior to injection into high-performance liquid chromatography (HPLC) system.

Peptides were purified using either a JASCO or Shimadzu HPLC system with a Vydac 218GCC1520 C18 column (Grace Division Discovery Sciences™). MeCN and water used in HPLC purification were supplemented with TFA (to a final concentration of 0.1 % (v/v)) and degassed before use. The peptide was purified by HPLC using a suitable gradient of MeCN in water, determined by performing an initial small-scale purification. Peptide fractions of sufficient purity for use were identified using the UV trace and MALDI-TOF analysis using an autoflex™ speed MALDI-TOF mass spectrometer (Bruker). These fractions were pooled and lyophilised to yield solid pure peptide.

8.1.5.6 Quantification by NMR

Purified peptides were quantified using ^{1}H NMR spectroscopy on the AVB 500 (Bruker) with 1D-^{1}H-nuclear Overhauser effect spectroscopy (1D-NOESY) solvent suppression. Samples were prepared by adding 5 µL of 10 or 50 mM peptide and 5 µL of 58 mM (10 mg mL^{-1}) 3-(trimethylsilyl)-propanoic acid (TSP) were added to 500 µL D_2O. NMR peaks for methyl groups on aliphatic residues were used to determine concentration of the peptide relative to the TSP standard (^{1}H chemical shift = 0 ppm). Spectra were generated in TopSpin 3.2 and integrated manually in MestReNova.

8.1.5.7 Nt-Cys oxidation using recombinant HsADO

When oxidation of the N-terminal cysteine of a peptide was required, peptide (1 mM) was incubated with recombinant *Hs*ADO (1 or 10 µM) in assay buffer (50 mM Bis-Tris Propane, 50 mM NaCl, pH 7.5) supplemented with TCEP (5 mM), Fe_2SO_4 (20 µM), and L-ascorbate (1 mM) at 37 °C for 60 minutes. ADO was removed from the reaction mixture using a 10,000 Da MWCO ultracentrifugation device. Flow-through contained oxidised

peptide (1 mM), so was retained and stored at -20 °C. Oxidation level was verified using high throughput mass spectrometry (MS).

8.1.6 Thin-layer chromatography assays

The thin-layer chromatography (TLC) assay protocol used to assess the small molecule thiol dioxygenase activity of recombinant *Hs*ADO was adapted from a protocol used by Fernandez *et al.*[184].

Ninhydrin stain solution was prepared by dissolving 0.75 g of ninhydrin in 500 mL of a solution of 3 % (v/v) acetic acid in ethanol. The resulting solution was wrapped in foil to protect it from the light. TLC running buffer was prepared by mixing 1:1:3 (v/v) water, acetic acid, and 1-butanol. Standard solutions of cysteamine and hypotaurine were prepared to a final concentration of 20 mM in assay buffer (50 mM Bis-Tris Propane, 50 mM NaCl, pH 7.5).

For the assay, cysteamine (40 mM in assay buffer, 3 × 55 µL) and recombinant *Hs*ADO (40, 300, or 400 µM in SEC buffer (Buffer C from Table 8.2), 3 × 50 µL) were incubated separately in 0.6 mL tubes at 37 °C for 5 minutes. 50 µL of cysteamine solution was added to 50 µL of ADO solution and the resulting reaction mixture was incubated at 37 °C for 10 – 30 minutes, as required. The reaction mixture was then heated at 95 °C for 2 minutes to denature the enzyme and stop the reaction. Precipitated protein was removed by centrifugation (14000 rpm (14462 g), 5 minutes) and supernatant was spotted onto a TLC plate (Supelco® TLC Silica gel 60, aluminium support), alongside standard solutions of cysteamine and hypotaurine. The TLC plate was placed in a beaker containing running buffer, and the beaker was covered with a foil lid. After approximately 1 hour (when the solvent front almost reached the top of the TLC plate), the TLC plate was removed from the beaker and developed with the ninhydrin stain solution, which was heat activated using a

heat gun. Spots associated with cysteamine and hypotaurine in the reaction mixture were identified by comparison with the standards. Two spots were observed for the cysteamine standard, as also seen by Fernandez *et al.*[184], which was postulated to be due to cystamine formation.

8.1.7 Peptide oxidation assays

Oxidation assays with peptides were performed in the following final reaction conditions unless otherwise stated:

Assay Buffer: 50 mM Bis-Tris Propane, 50 mM NaCl, pH 7.5

Cofactors: 5 mM TCEP, 20 μM iron(II), 1 mM L-ascorbate

Di-ammonium iron(II) sulfate hexahydrate, $(NH_4)_2SO_4$ $FeSO_4.6H_2O$, was used as the source of iron(II) in all experiments. This was dissolved in 20 mM HCl at a concentration of 100 mM, followed by dilution to 10 mM in Milli-Q water before dilution in assay buffer to a stock concentration of 2 mM. Stock solutions of other cofactors were prepared in assay buffer at the following concentrations: 500 mM TCEP; 100 mM ascorbate. Cofactor stocks were all prepared fresh on the day of the assay.

Enzyme aliquots were thawed on ice and diluted to an appropriate stock concentration in SEC buffer (Buffer C from Table 8.2). A reduction in enzyme activity was observed if enzyme was diluted in assay buffer too early; the most consistent assay results were obtained when enzyme was diluted in assay buffer just before starting each replicate of an assay. Peptides stocks were prepared at a concentration of 10 mM in Milli-Q water, then diluted to 1 mM in assay buffer. Reaction volume was 50 μL, unless otherwise stated.

8.1.7.1 ADO activity assays

In activity assays with recombinant *Hs*ADO, final reaction concentrations were either 20 μM peptide and 0.05 μM ADO or 100 μM peptide and 0.1 μM ADO.

For assays analysing ADO activity towards a single substrate peptide, a master mix of substrate peptide and cofactors was prepared in assay buffer (all components at 2 × final assay concentrations). Enzyme was diluted to 2 × final assay concentration in assay buffer. Master mix (0.5 × reaction volume) and enzyme (0.5 × reaction volume) were incubated separately in a 96 well plate at 37 °C for 5 min. A '0 minute' timepoint was taken just before the end of the 5 minute pre-incubation period by adding a sample of master mix and a sample of enzyme mix to 1 % (v/v) formic acid (Optima™ LC/MS grade, Fisher Chemical™). The reaction was initiated by addition of enzyme to master mix. For 'no enzyme' controls, assay buffer was incubated at 37 °C for 5 min and added to master mix to 'initiate' the reaction. At requisite time points, a sample of reaction mixture was taken and quenched with 1 % (v/v) formic acid. Samples were stored at −20 °C until needed for analysis. All reactions were performed in triplicate.

For assays comparing ADO activity towards different peptides, a solution of cofactors (12.5 mM TCEP, 50 μM iron(II), 2.5 mM L-ascorbate) was prepared in assay buffer. This cofactor solution was added to the 96 well plate, then putative substrate peptides were added to the relevant wells of the plate to create a master mix for each substrate peptide (all components at 2 × final assay concentrations). The assay was then performed as described above, with a 'no enzyme' control included for each substrate.

Time points taken for each assay are specified in main Thesis text and figures. Assays were analysed by high throughput MS, ultrahigh-performance liquid chromatography (UPLC) MS, or liquid chromatography MS (LCMS) (Section 8.1.9).

8.1.7.2 Steady state kinetics assays

Steady state kinetics analysis of *At*RAP2.12$_{2\text{-}15}$ peptide was performed using a final ADO concentration of 0.05 μM. Stock solutions of *At*RAP2.12$_{2\text{-}15}$ peptide were prepared in assay

buffer at concentrations of: 40 mM, 20 mM, 10 mM, 5 mM, 2.5 mM, 1.25 mM, 0.625 mM, 0.3125 mM, and 0.15625 mM. Final assay concentrations of *At*RAP2.12$_{2\text{-}15}$ peptide for steady state kinetics analysis were: 4 mM, 2 mM, 1 mM, 500 µM, 250 µM, 125 µM, 62.5 µM, 31.25 µM, and 15.625 µM.

As in the activity assays, a solution of cofactors in assay buffer was added to a 96 well plate (20 µL per well). 5 µL of each peptide stock was added to the relevant wells in the 96 well plate to generate a master mix for each required final concentration of peptide. The assay was then performed as described above. A 'no enzyme' control was included, with a final peptide assay concentration of 125 µM.

Time points were taken at 0, 20, 40, 60, and 80 seconds. Assays were analysed by high throughput MS.

8.1.7.3 Inhibition assays

ADO inhibition assays were performed using final reaction concentrations of 20 µM peptide and 0.05 µM recombinant *Hs*ADO. Iron(II) cofactor was excluded from assays testing inhibition by small molecules as all but one of the small molecules used were iron chelators. Inhibitor stock concentrations were prepared in assay buffer or DMSO (2,2-dipyridyl and DP44mT).

A solution of substrate peptide and cofactors was prepared in assay buffer and added to a 96 well plate. Inhibitor stock solution was added to the relevant wells in the 96 well plate to create a master mix of inhibitor, substrate, and cofactors for each inhibitor (all components at 2 × final assay concentrations). For positive controls, addition of inhibitor to master mix was replaced with addition of either assay buffer or DMSO, depending on the solvent used in the inhibitor stock. The assay was then performed as described above. A 'no

enzyme' control was included for each inhibitor to examine the interaction of inhibitor and substrate peptides.

Time points were taken at 0, 1, 3, and 10 minutes. Assays were analysed by high throughput MS.

8.1.7.4 IC_{50} determination

Assays to determine IC_{50} were performed as described above for inhibitor assays. Instead of including different inhibitors in the different wells of the 96 well assay plate, master mixes were prepared to produce varying final inhibitor concentrations (as specified in Chapter 5).

Time points were taken at 0 and 1 minute. Assays were analysed by high throughput MS (small molecule inhibitors) or UPLC MS (peptide inhibitors).

8.1.7.5 Mode of inhibition experiments

To determine the mode of inhibition for ADO inhibitors, steady state kinetic analysis was performed for $RGS4_{2-15}$ peptide in the presence of different concentrations of inhibitor. The final concentration of recombinant *Hs*ADO in these assays was 0.05 μM. Final $RGS4_{2-15}$ concentrations used were: 1000 μM, 500 μM, 250 μM, 125 μM, 62.5 μM, 25 μM. Final KH1 concentrations were: 200 μM, 100 μM, 50 μM, 25 μM, 0 μM (assay buffer). Final 2,2-dipyridyl concentrations were: 1 mM, 0.5 mM, 0.25 mM, 0.125 mM, 0 mM (DMSO). 10 × stocks were prepared for each concentration of substrate and inhibitor.

For each inhibitor, at each concentration, a solution of inhibitor and cofactors was prepared in assay buffer, which was added to the 96 well assay plate. As in steady state kinetic analysis with $RAP2.12_{2-15}$, 5 μL of each peptide stock was added to the relevant wells in the 96 well plate to generate a master mix for each required final concentration of peptide (all components at 2 × final assay concentrations). The assay was then performed as described

above. A ‘no enzyme’ control was included, with a final substrate peptide assay concentration of 125 μM.

Time points were taken at 0, 20, 40, and 60 seconds. Assays were analysed by high throughput MS.

8.1.8 Peptide acetylation assays

Recombinant NatA protein used in acetylation assays was composed of human NAA10 residues 1-160 and *N*-His_6-tagged full-length NAA15.

NAA10 (1-160):
MNIRNARPEDLMNMQHCNLLCLPENYQMKYYFYHGLSWPQLSYIAEDENGKIVGYVLAKM
EEDPDDVPHGHITSLAVKRSHRRLGLAQKLMDQASRAMIENFNAKYVSLHVRKSNRAALH
LYSNTLNFQISEVEPKYYADGEDAYAMKRDLTQMADELRR

N-His_6-tagged NAA15:
HHHHHHMPAVSLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNPKFAEHGETLAMKGL
TLNCLGKKEEAYELVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKCYRNALKWDKDNL
QILRDLSLLQIQMRDLEGYRETRYQLLQLRPAQRASWIGYAIAYHLLEDYEMAAKILEEF
RKTQQTSPDKVDYEYSELLLYQNQVLREAGLYREALEHLCTYEKQICDKLAVEETKGELL
LQLCRLEDAADVYRGLQERNPENWAYYKGLEKALKPANMLERLKIYEEAWTKYPRGLVPR
RLPLNFLSGEKFKECLDKFLRMNFSKGCPPVFNTLRSLYKDKEKVAIIEELVVGYETSLK
SCRLFNPNDDGKEEPPTTLLWVQYYLAQHYDKIGQPSIALEYINTAIESTPTLIELFLVK
AKIYKHAGNIKEAARWMDEAQALDTADRFINSKCAKYMLKANLIKEAEEMCSKFTREGTS
AVENLNEMQCMWFQTECAQAYKAMNKFGEALKKCHEIERHFIEITDDQFDFHTYCMRKIT
LRSYVDLLKLEDVLRQHPFYFKAARIAIEIYLKLHDNPLTDENKEHEADTANMSDKELKK
LRNKQRRAQKKAQIEEEKKNAEKEKQQRNQKKKKDDDDEEIGGPKEELIPEKLAKVETPL
EEAIKFLTPLKNLVKNKIETHLFAFEIYFRKEKFLLMLQSVKRAFAIDSSHPWLHECMIR
LFNTAVCESKDLSDTVRTVLKQEMNRLFGATNPKNFNETFLKRNSDSLPHRLSAAKMVYY
LDPSSQKRAIELATTLDESLTNRNLQTCMEVLEALYDGSLGDCKEAAEIYRANCHKLFPY
ALAFMPPGYEEDMKITVNGDSSAEAEELANEI

8.1.8.1 Acetylation assays conducted in Bergen

Radioactivity-based acetylation assays with NatA *in vitro* were performed by Dr. Matti Myllykoski at the University of Bergen (Arnesen Lab). The protocol used was adapted from that published by Drazic *et al*.[221].

Peptides for acetylation assays performed in Bergen were purchased from Invitrogen. Protocol was adapted from that used by Drazic *et al*.[221] in the following ways: reaction time

was shortened to 5 minutes, enzyme concentration was increased to 0.3 μM, and reaction buffer was changed to 50 mM Na-phosphate pH 7.0, 100 mM NaCl, 1 mM EDTA, 1 mM DTT and 2 mg/ml BSA. Otherwise, assay were performed as described by Drazic *et al*.[221].

8.1.8.2 Acetylation assays conducted in Oxford

MS-based acetylation assays with NatA *in vitro* were performed in Oxford using recombinant NatA protein kindly provided by Dr. Myllykoski. Acetylation assays were analysed by high throughput MS (Section 8.1.9).

1 mM peptide stocks were prepared in Milli-Q water. 20 mM acetyl CoA and 100 mM TCEP stocks were prepared in water. NatA aliquots were thawed on ice.

Initial MS-based acetylation assays used an assay buffer of 50 mM HEPES, 100 mM NaCl, 1 mM EDTA, pH 7.5 supplemented with 1 mM TCEP. Final reaction concentrations were 100 μM substrate peptide, 0.1 μM NatA, and 200 μM acetyl CoA.

For these assays, a master mix of 2.5 mM TCEP and 500 μM acetyl CoA was prepared in acetylation assay buffer and added to a 96 well assay plate (20 μL per well). Master mix, peptide (1 mM), and enzyme (0.2 μM) were incubated separately 37 °C for 5 minutes. It was important to incubate the peptide separately from acetyl CoA to reduce background acetylation of cysteines. Just before the end of the 5 minute pre-incubation period 5 μL peptide was added to the master mix, and '0 minute' timepoint was quickly taken by adding a 2.5 μL sample of peptide in master mix and a 2.5 μL sample of enzyme mix to 45 μL of 1 % (v/v) formic acid. The reaction was then initiated by adding 22.5 μL of enzyme to the solution of peptide and master mix. At requisite time points, a 5 μL sample of reaction mixture was taken and quenched in 45 μL of 1 % (v/v) formic acid. Samples were stored at −20 °C until needed for analysis. All reactions were performed in triplicate.

MS-based acetylation assays investigating the influence of Nt-Cys oxidation on NatA activity used an assay buffer of 50 mM Na-phosphate, 100 mM NaCl, pH 7.0, supplemented with 1 mM TCEP. Final reaction concentrations were 100 μM substrate peptide, 0.15 μM NatA, and 100 μM acetyl CoA.

For these assays, a solution of TCEP in acetylation assay buffer was added to a 96 well assay plate. Peptide was then added to the relevant wells in the 96 well assay plate to generate a master mix of 2.5 mM TCEP and 250 μM peptide. Master mix, acetyl CoA (1 mM), and enzyme (0.3 μM) were incubated separately 37 °C for 5 minutes. Just before the end of the 5 minute pre-incubation period, 5 μL of acetyl CoA was added to the master mix, and a '0 minute' timepoint was quickly taken by adding a 2.5 μL sample of master mix (with acetyl CoA) and a 2.5 μL sample of enzyme mix to 45 μL of 1 % (v/v) formic acid. The reaction was then initiated by adding 22.5 μL enzyme to the solution of peptide and master mix. At requisite time points, a 5 μL sample of reaction mixture was taken and quenched in 45 μL of 1 % (v/v) formic acid. Samples were stored at −20 °C until needed for analysis. All reactions were performed in triplicate.

8.1.8.3 Preparation of MS/MS sample

A sample containing acetylated $JUNB_{2-15}$ peptide was prepared by incubating $JUNB_{2-15}$ peptide (100 μM) with NatA (0.1 μM) and acetyl CoA (200 μM) in 50 mM HEPES, 100 mM NaCl, 1 mM EDTA, pH 7.5 supplemented with 1 mM TCEP at 37 °C for 60 minutes. The reaction was quenched in 1 % (v/v) formic acid to produce a solution containing 10 μM peptide, which was analysed by UPLC MS/MS.

8.1.9 Mass spectrometry analysis

8.1.9.1 High throughput MS

High throughput MS was performed using a RapidFire 365 high-throughput MS System (Agilent) and 6530 Accurate-Mass Quadrupole time-of-flight (Q-TOF) mass spectrometer (Agilent) operated in the positive ion mode. Peptide samples for high throughput MS were prepared in 1 % (v/v) formic acid. Samples were aspirated under vacuum and loaded onto a C4 solid phase extraction (SPE) cartridge where they were washed with 0.1% (v/v) formic acid in LCMS grade water at a flow rate of 1.5 mL min^{-1} and eluted with a solution of 85% (v/v) acetonitrile, 15% (v/v) water, 0.1% (v/v) formic acid. Samples were loaded onto the mass spectrometer at 1.25 mL min^{-1}. Mass spectra were assessed in Qualitative Analysis software B.07.00, and peptide turnover was quantified using integrated area under the peaks for modified and unmodified peptide, extracted using RapidFire Integrator software.

8.1.9.2 UPLC MS

In activity assays, some peptides ($DNMBP_{2\text{-}15}$; $GPR22_{2\text{-}15}$; $NECAB2_{2\text{-}15}$; $SHHC_{2\text{-}15}$) could not be detected using high-throughput MS, so analysis of assays with these peptides was attempted by ultrahigh-performance liquid chromatography (UPLC) MS using an Acquity UPLC system (Waters) coupled to a Xevo G2-S or G2-XS Q-TOF mass spectrometer (Waters) operated in positive electrospray mode. UPLC MS was also used for analysis of IC_{50} assays with peptide inhibitors to ensure separation of substrate and inhibitor peaks. Instrument parameters, data acquisition, and data processing were controlled by MassLynx 4.1 with source conditions adjusted to maximize sensitivity and minimize fragmentation. Samples were injected onto a Chromolith Performance RP-18e 100 2-mm column (Merck) heated to 40 °C, and eluted at 0.3 mL min^{-1} using a gradient of 95 % (v/v) Milli-Q water supplemented with 0.1 % (v/v) formic acid to 95 % (v/v) acetonitrile. Mass spectra was analysed using MassLynx 4.1; peptide turnover

was calculated from the integrated area under the peaks for modified and unmodified peptide.

8.1.9.3 UPLC MS/MS

MS/MS of $JUNB_{2\text{-}15}$ peptide was performed using UPLC MS as described above. Fragmentation spectra of substrate and product peptide ions were obtained using a targeted approach with a collision-induced dissociation energy ramp of 55 to 75 eV. Analysis was carried out with the same source settings, flow rate, and column elution conditions as above.

8.1.9.4 LCMS

Due to their high basicity, peptides OS1, OS2, and OS6-10 could not be analysed by high throughput MS or UPLC MS. Assays with these peptides were analysed using a 1200 Infinity B LCMS (Agilent) connected to a 6530 Accurate-Mass Q-ToF mass spectrometer (Agilent) operated in positive ion mode. Samples were injected onto a ProSwift™ RP-4H analytical 1 × 50 mm column (Thermo Scientific™) and eluted at 0.2 mL min^{-1} using a gradient of 95 % (v/v) LCMS grade water supplemented with 0.1% (v/v) formic acid to 100 % (v/v) LCMS grade acetonitrile. Spectra were analysed and peptide turnover calculated from the integrated area under peaks, extracted using MassHunter Qualitative Analysis software (Agilent).

8.1.10 Protein crystallography

24 well sitting drop plates for initial crystallisation trials (described in Section 6.3) were prepared according to the plate maps in Appendix 12. For initial aerobic trials, a 30 µL aliquot of 16.3 mg mL^{-1} recombinant *Hs*ADO (prepared by Prof. Emily Flashman as described above) was defrosted on ice. 2 µL of reservoir solution from each well was mixed with 2 µL of ADO in the 24 well sitting drop plates. The plates were stored aerobically at

r.t., and were checked periodically for potential crystals under a M205 C stereo microscope (Leica).

For initial anaerobic trials, stock solutions for reservoir solutions in 24 well sitting drop plates were purged with nitrogen and transferred to an anaerobic glove box (Belle Technology), where they were left to equilibrate overnight. 24 well sitting drop plates for trials were then prepared in the glove box according to the plate maps in Appendix 12. An aliquot of 30 µL aliquot of 16.3 mg mL^{-1} recombinant *Hs*ADO and a solution of 50 mM ^{D}Cys-RGS4$_{3-15}$ peptide were transferred to the glove box and allowed to equilibrate for at least 10 minutes. 1.5 µL of ^{D}Cys-RGS4$_{3-15}$ peptide was added to the aliquot of ADO. Prior to addition to crystal trays, solutions containing ADO protein were kept cold using a cooling block that had been stored at -20 °C. 2 µL of reservoir solution from each well was mixed with 2 µL of the solution of peptide and ADO in the 24 well sitting drop plates. The plates were stored anaerobically in the glove box at r.t. and checked periodically for potential crystals under a WILD M3C stereo microscope (Leica).

Anaerobic broad screens were performed in Manchester by Dr. Colin Levy. Recombinant *Hs*ADO was expressed and purified as described above. ADO was concentrated to 20.1 mg mL^{-1} using a 10,000 MWCO ultracentrifugation device. The resulting ADO solution was stored on ice and delivered to Manchester from Oxford directly after purification, along with a solution of 50 mM RGS5$_{2-15}$ peptide in SEC buffer. In Manchester, Dr. Levy used a Mosquito® Crystal robot (SPT Labtech) housed in an anaerobic glove box to perform broad screens with JCSG Plus™, PACT Premier™, The BCS Screen, SG1™ Screen, Morpheus®, and LMB Crystallization Screen (Molecular Dimensions). Screens were performed using 96 well sitting drop plates and a protein/precipitant solution ratio of 1:1. Anaerobic crystallisation trials were repeated with and without inclusion of RGS5$_{2-15}$

substrate peptide at a peptide/protein ratio of 2.5:1. The screens were monitored by Dr. Levy for potential protein crystals.

Aerobic co-crystallisation trials (Section 6.5) were conducted using freshly purified recombinant *Hs*ADO and KH1 peptide. Stocks of ADO and KH1 peptide were prepared in SEC buffer (Table 8.2). ADO (81.3 mg mL^{-1}, 235 µL) was mixed with KH1 peptide (100 mM, 30 µL), iron(II) (500 mM in 20 mM HCl, 2.4 µL), and SEC buffer (Table 8.2) on ice to produce a solution of 63.7 mg mL^{-1} ADO and 10 mM KH1 (1:5 protein/inhibitor). Two 6 well screens were prepared in a 24 well hanging drop plate according to the plate maps in Appendix 12. 96 well broad screens were prepared using a Mosquito® Xstal3 robot (SPT Labtech). In all screens a 1:1 ratio of protein solution to precipitant solution was used. Trays were stored aerobically at r.t. and checked periodically for potential crystals under a M205 C stereo microscope (Leica).

8.2 Experiments in Mammalian Cells

8.2.1 Preparation of plasmids for DFOR assays

Plasmids for dual-fluorescence oxygen reporter (DFOR) assays were generated by PCR and restriction enzyme cloning using the primers listed in Appendix 7 and a pcDNA3.1-GFP:P2A:mCherry template plasmid as described by Smith and Keeley[204].

8.2.1.1 Addition of sequence of interest to EGFP DNA template by PCR

The PCR reaction was prepared to a final volume of 50 µL in 0.6 mL PCR tubes according to Table 8.4, using stocks prepared in sterile dH_2O.

Table 8.4: PCR reaction constituents:

Component	Volume (µL)	Final Concentration
pcDNA3.1 + -EGFP template	1	0.1 – 10 ng
Forward primer	2.5	1 µM
Reverse primer	2.5	1 µM
Phusion High Fidelity master mix (NEB)	25	1 ×
Sterile dH_2O	19	-

PCR tubes were transferred to a thermocycler and the PCR reaction was initiated using the following conditions:

1 cycle: 98 °C for 30 s

30 cycles: 98 °C for 10 s, 55 °C for 15 s, 68 °C for 30 s

1 cycle: 72 °C for 300 s

1 cycle: 4 °C for ∞ (until ready for next step)

PCR products were purified using QIAquick® PCR Purification Kit (Qiagen), eluting in a final volume of 50 µL sterile dH_2O.

8.2.1.2 Digestion of PCR products

Restriction digest reactions were prepared with 50 µL of PCR product, 6 µL of 10 × FastDigest buffer (Thermo Scientific™), and 2 µL each of FastDigest *HindIII* and *XbaI* restriction enzymes (Thermo Scientific™). Digest reactions were incubated at 37 °C for 45 – 60 minutes and quenched by reducing the temperature to 4 °C.

Linearised vector was produced by preparing a restriction digest reaction, replacing the 50 µL of PCR product with 30 µL of sterile dH_2O and 20 µL of pcDNA3-GFP-P2A-mCherry plasmid provided by Dr. Thomas Keeley (Ratcliffe Group, Nuffield Department of Medicine) (Appendix 8). After the digest reaction, linearised vector was dephosphorylated by mixing 53 µL of linearised plasmid with 6 µL of 10 × CutSmart®

buffer (NEB) and 6 µL of 1 U μL^{-1} Quick CIP (NEB), and incubating the resulting mixture at 37 °C for 30 minutes.

Quenched digestion reactions and dephosphorylated and linearised vector were mixed with 6 × SDS-loading dye and run on a 1 % (w/v) agarose gel at 100 V for approximately 30 minutes, with a pre-stained DNA ladder for reference. Digested DNA inserts and linearised vector were excised from gel and purified using the QIAquick® Gel Extraction Kit (Qiagen). Concentrations of purified fragments and linear plasmids were determined using a Nanodrop spectrophotometer.

8.2.1.3 Ligation of DNA inserts and linearised vector

Ligation reactions were prepared in fresh 0.6 mL tubes according to Table 8.5.

The ligation reactions were incubated at r.t. for 1 hour. Reactions were stopped by placing tubes on ice.

Table 8.5: Ligation reaction constituents:

Component	Volume (µL)	Final Concentration
Linear plasmid	-	50 ng
DNA fragment	-	~ 100 ng
T4 ligase (Thermo Scientific)	1	1 U
T4 ligase buffer (10×)	2	1 ×
Sterile dH_2O	up to 20	-

8.2.1.4 Transformation of E. coli with DFOR plasmids

Ligation reactions were transformed into XL10-Gold® Ultracompetent Cells (Agilent) according to the manufacturer's protocol, then plated onto agar plates containing ampicillin selection antibiotic. Plates were incubated at 37 °C overnight. Colonies were selected and used to inoculate minipreps (5 mL LB media containing 100 µg mL^{-1} ampicillin), which were incubated in a shaking incubator at 37 °C overnight. Plasmids were then purified from

minipreps using the QIAprep® Spin Miniprep Kit (Qiagen) and sequences were confirmed by Sanger Sequencing (Source BioScience).

8.2.2 General cell culture procedures

All cell culture procedures were carried out using sterile technique in a laminar flow cabinet. All pipette tips and microfuge tubes used were sterile cell culture grade. All equipment and consumables were sprayed with 70 % (v/v) ethanol before being introduced into the cabinet. Solutions were pre-warmed in a 37 °C water bath prior to use. Sterile phosphate buffered saline (PBS) was used in all wash steps in cell culture experiments, unless otherwise stated. All mammalian cells were incubated at 5 % CO_2, 37 °C. SH-SY5Y cells were cultured in sterile DMEM-F12 growth medium (Gibco™), supplemented with 10 % (v/v) fetal bovine serum (FBS), 2 mM L-Glutamine, 100 U mL^{-1} penicillin, and 10 µg mL^{-1} streptomycin, unless otherwise stated.

8.2.3 Production of stable SH-SY5Y cell lines for DFOR assay

8.2.3.1 Splitting cells for subculture

SH-SY5Y cell culture was allowed to grow to ~80 % confluency. Growth medium was aspirated and cells were washed to remove excess medium. Cells were detached from the culture dish using an appropriate volume (e.g., 1 mL for a 10 cm culture dish) of Trypsin-EDTA solution (Sigma Aldrich) and incubation at 37 °C for 2 – 5 minutes. When cells were observed to be sufficiently detached, the trypsin solution was neutralised by adding fresh medium (e.g., 9 mL for a 10 cm culture dish). Cell solution was mixed until a homogenous single-cell suspension was achieved. Cells were diluted with fresh medium and distributed into new cell culture dishes or different wells in a new cell culture plate.

8.2.3.2 Transfection of DFOR plasmids into SH-SY5Y cells

Cells were grown to ~70 % confluency in a 6 well cell culture plate (each well containing 2 mL of medium). Plasmids were transfected into SH-SY5Y cells using GeneJuice® (Novagen) or Lipofectamine™ 3000 (Invitrogen™) according to the manufacturer's protocol.

8.2.3.3 Selection of transfected cells

Approximately 24 hours after transfection, 10 µL of 100 mg mL^{-1} G418 antibiotic (Santa Cruz Biotechnology) was added to each well of the 6 well cell culture plate. Cells were grown in medium containing G418 (0.5 mg mL^{-1}) for 2 – 4 weeks, splitting cells at a ratio of 1:4 when they reached > 90 % confluency, until > 95 % of cells could be seen expressing EGFP and mCherry when imaged using a CKX41 microscope (Olympus) and pE300-W light source (CoolLED). EGFP was imaged under blue light ($488_{ex}/520_{em}$), and mCherry was imaged under green light ($550_{ex}/590_{em}$). After selection, a polyclonal population of cells remained, all stably expressing the reporter gene. Following selection cells were returned to medium without G418 antibiotic.

8.2.3.4 Freezing aliquots of cells

Cells were grown to > 80 % confluency in a 10 cm culture dish. Medium was removed from cells, which were washed and detached from the cell culture dish as described above. Cells were resuspended in growth medium as before, then transferred to 15 mL falcon tubes and pelleted by centrifugation in a benchtop centrifuge (3 min, 1200 rpm). Medium was aspirated and cell pellet was resuspended in 1 mL freezing medium (DMEM (Sigma Aldrich) supplemented with 15 % DMSO and 20 % FBS) and transferred to a labelled 2 mL cryovial on dry ice. Frozen cell aliquots were stored at -80 °C.

8.2.3.5 Thawing aliquots of cells

Frozen cell aliquots in 2 mL cryovials were defrosted in a 37 °C water bath. Defrosted cells were quickly added to a 15 mL falcon tube containing 9 mL growth medium and pelleted by centrifugation in a benchtop centrifuge (3 min, 1200 rpm). Medium was aspirated and cell pellets were resuspended in 10 mL fresh medium, then transferred to 10 cm culture dishes and incubated at 5 % CO_2, 37 °C.

8.2.4 Dual-fluorescence oxygen reporter assay

Initial assays with dipyridyl

Cells were split into individual wells of a 96 well cell culture plate with black walls and a clear base (diluting to 20 % (v/v) in fresh medium from a 10 cm cell culture dish at > 90 % confluency). Three wells were allocated for each cell line per treatment condition. Cells were allowed to grow to 80 % confluency, then growth medium was aspirated. Cells were washed with PBS, then DMEM without phenol red (Gibco™) containing either DMSO (20 µL per 10 mL medium) or 2,2-dipyridyl (20 µL of 100 mM solution in DMSO per 10 mL medium) was added. The 96 well plate was then incubated at 5 % CO_2, 37 °C for 24 hours. EGFP and mCherry fluorescence were measured by multichromatic fluorescence imaging using a FLUOstar® Omega plate reader (BMG LABTECH); gain was adjusted to prevent saturation of mCherry signal and kept consistent across biological repeats. The EGFP fluorescence was measured at 485 nm excitation, 520 nm emission and mCherry at 544 nm excitation, 590 nm emission. The EGFP/mCherry ratio was calculated using blank-adjusted values. At least three biological replicates were performed for each cell line.

Assays with hypoxia

Cells were split into two 96 well cell culture plates (black walls, clear base) and allowed to grow to 80 % confluency as described above. Growth medium was then replaced with DMEM without phenol red (Gibco™) containing either DMSO (20 µL per 10 mL medium)

or 2,2-dipyridyl (20 µL of 100 mM solution in DMSO per 10 mL medium). In one plate, cells were treated with DMSO and 2,2-dipyridyl, and in the other, with only DMSO. The plate containing both DMSO- and 2,2-dipyridyl-treated cells was incubated at 5 % CO_2, 37 °C for 20 hours. The other plate was incubated in hypoxia within an atmosphere-regulated workstation set at 1 % O_2: 5 % CO_2: balance N_2 (Invivo$_2$ 400, Baker-Ruskinn Technologies), 37 °C for 20 hours. EGFP and mCherry fluorescence were then measured as above.

8.2.5 Analysis of DNA from SH-SY5Y cell lines

Genomic DNA extraction

Cells were grown to ~80 % confluency in a 12 well cell culture plate. Medium was aspirated from cells, which were then washed and detached from the cell culture plate as above. Cells were resuspended in 1 mL growth medium and transferred to 1.5 mL tubes, then pelleted by centrifugation in a benchtop centrifuge (3 min, 2000 rpm). Cell were lysed in 400 µL of lysis buffer containing Proteinase K (100 mM Tris, 5 mM EDTA, 200 mM NaCl, 0.2 % SDS, pH 8 + 10 µL mL^{-1} Proteinase K (100 µg mL^{-1})) at 55 °C overnight. 400 µL of ice-cold isopropanol was added to lysed cells to precipitate gDNA, and the resulting mixture was vortexed for approximately 10 seconds. Precipitated gDNA in 1.5 mL tubes was pelleted by centrifugation in a benchtop centrifuge (5 min, 13000 rpm). Supernatant was removed and gDNA pellet was washed with 70 % ethanol (2 × 800 µL). Any remaining ethanol solution was allowed to evaporate until the gDNA pellet stopped appearing fully opaque, but before it became fully transparent. gDNA pellets were resuspended in 400 µL of Tris-EDTA (TE) buffer (Thermo Scientific) at 55 °C for 2 hours. Resulting gDNA solutions were stored at -20 °C.

Experimental error: 1 mL of lysis buffer was accidentally added to sample taken from cells expressing RGS4$_{1\text{-}15}$-EGFP-P2A-mCherry plasmid. Therefore, lysed cells had to be split

into two aliquots for subsequent treatment and gDNA pellets were lost during ethanol washes. Each 'pellet' was resuspended in 200 µL TE buffer and combined in one tube for PCR screen.

PCR screen

Nine different gDNA samples were analysed in the PCR screen. A master mix (MM1) was prepared with 67.5 µL of DMSO, 112.5 µL of R_GFP_XbaI primer (Appendix 7), 900 µL of Phusion High Fidelity master mix (NEB), and 157.5 µL of sterile dH_2O. 137.5 µL of MM1 was added to 50 µL of each gDNA solution to produce gDNA master mixes (gDNA MMs). To prepare individual PCR reactions, 1.25 µL of "F_HindIII_XXX[15]_NoGFP" forward primer (Appendix 7) was added to 18.75 µL of gDNA MM. PCR reactions were prepared in a 96 well PCR plate, which was transferred to a thermocycler and the PCR reactions initiated using the following conditions:

1 cycle: 98 °C for 180 s

30 cycles: 98 °C for 30 s, 63 °C for 30 s, 72 °C for 60 s

1 cycle: 72 °C for 300 s

1 cycle: 8 °C for ∞ (until ready for next step)

4 µL of 6 × purple gel loading dye (NEB) was added to each PCR reaction and 15 µL of the resulting mixture was run on 1 % agarose gel at 90 V for 30 minutes, with a pre-stained DNA ladder for reference. Gels were imaged using a Bio-Rad ChemiDoc system.

8.2.6 Assay of full-length protein stability in mammalian cells

Assays assessing the stability of full-length putative ADO substrate proteins in human cells were conducted by Dr. Thomas Keeley (NDM) as described below.

Protein stability was assessed in parental SH-SY5Y or U87-MG cells. ADO knock-out (KO) cells were generated using CRISPR-Cas9[1]. ADO over-expressing (OE) cells were generated

by infecting ADO KO cells with pRRL-FLAG:ADO-IRES-GFP lentivirus. WT, ADO KO, or ADO OE cells were seeded in 6 well cell culture plates and allowed to reach 80 % confluency before normoxic or hypoxic incubation (as above) for 16 hours. Protein expression in different cell lines was compared by immunoblotting

Full-length STAC2, SUSD6, and ANKRD29 were not detectable in any of the above cell lines. Therefore, HEK293T cells were transiently transfected with 1 μg of pcDNA3.1-STAC2:FLAG, -SUSD6:FLAG or -ANKRD29:FLAG plasmids (GenScript) alongside pRRL-FLAG-ADO or -CDO1 (as a control) using polyethyleneimine. 24 hours later, cells were split into two 6 well cell culture plates and allowed to adhere before incubation in normoxia or hypoxia for 16 hours. In parallel, WT or ADO KO SH-SY5Y cells were transfected with pcDNA3.1-SUSD6:FLAG or -ANKRD29:FLAG plasmids and polyclonal stably expressing populations generated as described above for DFOR plasmids. Protein expression in different cell lines was compared by immunoblotting

Immunoblotting was performed as described by Masson *et al.*[1], using the following antibodies: AAAS (sc-374073), ASNS (sc-365809), LYLPLA2 (sc-390546), MTPN (sc-166072), RGS5 (sc-514184), and horseradish peroxidase-conjugated FLAG (M2, A8592) from Santa Cruz Biotech, JunB (3753) and RGS4 (15129) from Cell Signalling Technologies, GFPT2 (NBP1-56688) and SUSD6 (H00009766-B01P) from Novus Biotech, ANKRD29 (23999-1-AP) from Proteintech, and horseradish peroxidase-conjugated β-actin (ab49900) from Abcam. HRP-conjugated secondary antibodies were supplied by DAKO.

8.3 Computational Experiments

8.3.1 Comparative protein modelling

Comparative protein modelling of *Hs*ADO was performed using MODELLER software (version 9.23)[250,252]. *At*PCO4 (PDB: 6S7E) was used as the template structure. The sequence for *Hs*ADO was obtained from UniProt[193]:

```
MPRDNMASLIQRIARQACLTFRGSGGGRGASDRDAASGPEAPMQPGFPENLSKLKSLLTQ
LRAEDLNIAPRKATLQPLPPNLPPVTYMHIYETDGFSLGVFLLKSGTSIPLHDHPGMHGM
LKVLYGTVRISCMDKLDAGGGQRPRALPPEQQFEPPLQPREREAVRPGVLRSRAEYTEAS
GPCILTPHRDNLHQIDAVEGPAAFLDILAPPYDPDDGRDCHYYRVLEPVRPKEASSSACD
LPREVWLLETPQADDFWCEGEPYPGPKVFP*
```

This sequence was used to produce the ADO.ali file in Appendix 13. All scripts and .ali files used for comparative protein modelling in MODELLER can be found in Appendix 13.

MODELLER was used to align *Hs*ADO with *At*PCO4, using the "align2d.py" script and the "ADO.ali" and "6s7e.pdb" files to produce the "ADO-6s7e.ali" file. The "align2d.py" script was downloaded from the MODELLER online tutorial[254], with a line added to ensure inclusion of the Fe atom (env.io.hetatm = True). The "model-ligand.py" script was then used to produce five different models, using the "ADO-6s7e.ali" file. The output from this script included an assessment of the DOPE and GA_{341} scores, as described in Chapter 6 (page 128), which can be found in Appendix 13. A DOPE profile was generated for the best model produced (chosen by comparison of model DOPE scores) using the "evaluate_model.py", "evaluate_template.py", and "plot_profiles.py" scripts (Figure 8.2). This model was also assessed using MolProbity (downloaded as part of Phenix1.18-3861 software).

The DOPE profile (Figure 8.2) was used to select loops in the *Hs*ADO model to refine using the "loop_refine.py". Models produced by the "loop_refine.py" script were assessed using the "model_energies.py" script, and the model with the lowest DOPE score from each refinement was assessed as described above. Refinement was attempted for loops 22 to 46,

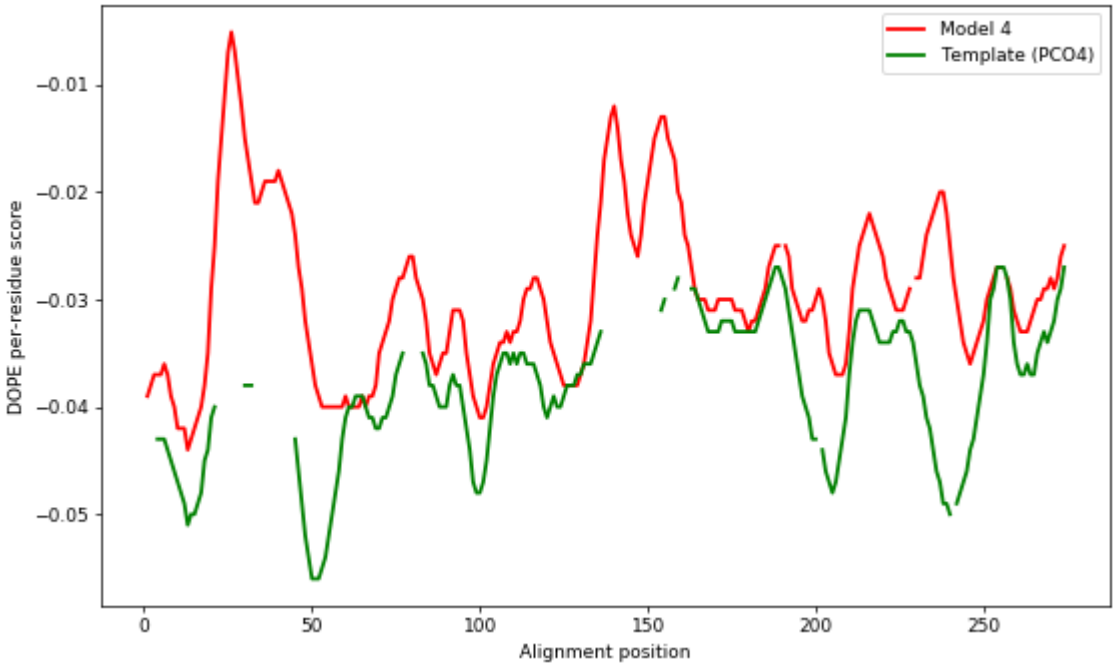


Figure 8.2: DOPE profile for *in silico* model and *At*PCO4 template
Plot of DOPE scores (per residue) for best model produced by "model-ligand.py" script (Model 4) and the *At*PCO4 structure (PDB: 6S7E) template (Template (PCO4)).

135 to 164, and 232 to 242 from the best model produced by the "model-ligand.py". However, none of these refinements resulted in an improved model according to DOPE scores and MolProbity analysis. Therefore, the original best model produced by the "model-ligand.py" script was used as the final *in silico* model.

8.3.2 Peptide docking experiments

PyMOL (version 2.3.0a0 Open-Source)[266] was used to assess the four chains in the structure of *Mm*ADO published by Fernandez *et al.* (PDB: 7LVZ)[6]. Chain D was selected as it appeared to be the most complete, i.e., included the largest portion of the protein's sequence. This chain was exported to a new .pdb file (7lvzD.pdb) for use in docking experiments.

The waters bound to the active site iron in 7lvzD.pdb were inspected as potential substrate binding sites as described in Chapter 6 (page 136). Water molecule 491 was selected as the most likely binding site. A Cys residue was fabricated in PyMOL and aligned with water molecule 491, using pair fitting. Water molecule 491 was then removed, and a bond was created between the S atom of the new Cys residue and the Fe atom in the protein active

site. Next, the chosen substrate peptide (*Hs*RGS4$_{2-8}$ (CKGLAGL), *Hs*IL32$_{2-8}$ (CFPKVLS), or *At*RAP2.12$_{2-8}$ (CGGAIIS)) was fabricated in PyMOL and its structure was minimised using the sculpt function. The Nt-Cys of the peptide was aligned to the Cys residue bound to the Fe atom using pair fitting and the peptide was rotated around the C_α-C_β bond of the Nt-Cys until the peptide was in a sensible orientation within the protein structure. The original Cys residue was then removed and a bond was created between the S atom of the peptide Nt-Cys and the Fe atom. The charge on the S atom of the peptide Nt-Cys was adjusted to remove the residual H atom, and the C-terminal residue of the peptide was adjusted to have a C-terminal amide moiety. The structure of peptide-bound ADO was minimised using the sculpt function, and chain and residue IDs were assigned to the peptide. The Fe atom and peptide residues were extracted together as a new .pdb file (named "Fe_[peptide]_FlexRes"). All water molecules not bound to the active site iron were removed (leaving two waters), and the remaining residues of the protein (excluding the Fe atom and the peptide residues) were extracted to a new .pdb file (named "7lvzD_rigid"). Finally, one of the water molecules was exported as a .pdb file (HOH.pdb).

Docking experiments were performed with AutoDock4 (version 4.2.6) using the AutoDockTools (version 1.5.7) graphical user interface[251].

In AutoDockTools, "HOH.pdb" and "7lvzD_rigid.pdb" were converted to .pdbqt files. "Fe_[peptide]_FlexRes.pdb" was loaded as a ligand file and the Fe atom was set as the root for the torsion tree description of flexibility. If necessary, the number of rotatable bonds in the "Fe_[peptide]_FlexRes" structure was adjusted to be within the software limit of 32. "Fe_[peptide]_FlexRes" was then saved as a .pdbqt file; in this file "BEGIN_RES" was added between the last "REMARK…" line and the "ROOT…" line, and "END_RES" was added at the end of the file before the "TORSDOF" line. The grid parameters for Autogrid were set by using the grid box to choose an appropriate docking area and confirming that

all required atoms were included in the map types, then saved as a .gpf file. Docking parameters for Autodock were set as follows: "7lvzD_rigid.pdbqt" was set as the "Rigid Filename"; "Fe_[peptide]_FlexRes.pdbqt" was set as the "Flexible Residues Filename"; the ligand was set as "HOH.pdbqt". Within the docking parameters, the search parameters were set for the Genetic Algorithm (GA), with 10 runs and 250,000 maximum evaluations for initial trial experiments, and 100 runs and 2,500,000 maximum evaluations for the full experiments. Docking parameters were saved as a .dpf file using the Lamarckian GA (4.2) output. Autogrid was run in AutoDockTools using the .gpf file created above, followed by Autodock using the .dpf file created above. Results from Autodock runs could be viewed in AutoDockTools and exported as .pdbqt files for visualisation in PyMOL.

8.3.3 Cavity prediction

Protein cavities in *Mm*ADO were predicted using the CavityPlus server[263]. A .pdb file containing chain D of Fernandez *et al.*'s *Mm*ADO structure (PDB: 7LVZ)[6] was uploaded to the CavityPlus server[267]. Cavity prediction was performed using the default parameters, selecting "Try to rescue large cavities" and "Apply soft separation" (Figure 8.3). Results were viewed on the CavityPlus webpage, and downloaded for visualisation in PyMOL.

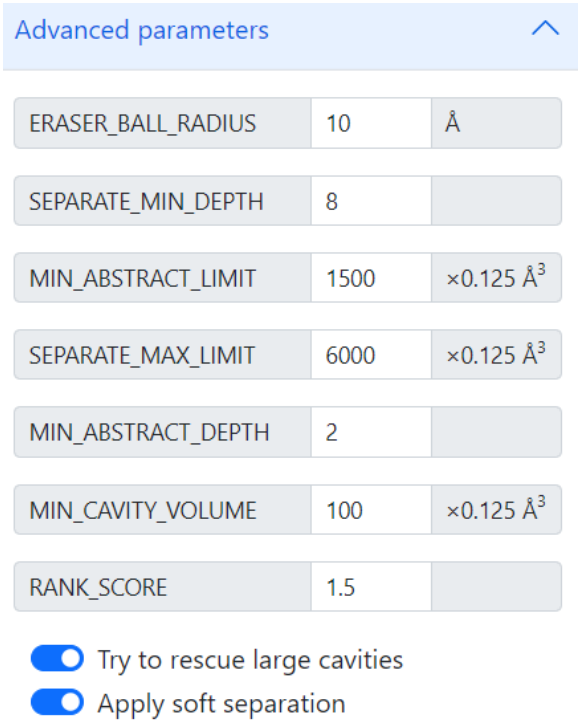


Figure 8.3: CavityPlus parameters

References

1. Masson, N. *et al.* Conserved N-terminal cysteine dioxygenases transduce responses to hypoxia in animals and plants. *Science* **365**, 65–69 (2019).
2. Varshavsky, A. N-degron and C-degron pathways of protein degradation. *Proc. Natl. Acad. Sci.* **116**, 358–366 (2019).
3. Weits, D. A. *et al.* Plant cysteine oxidases control the oxygen-dependent branch of the N-end-rule pathway. *Nat. Commun.* **5**, 3425 (2014).
4. Gunawardana, D. M., Heathcote, K. C. & Flashman, E. Emerging roles for thiol dioxygenases as oxygen sensors. *FEBS J.* **289**, 5426–5439 (2022).
5. Wang, Y., Shin, I., Li, J. & Liu, A. Crystal structure of human cysteamine dioxygenase provides a structural rationale for its function as an oxygen sensor. *J. Biol. Chem.* **297**, 101176 (2021).
6. Fernandez, R. L. *et al.* The Crystal Structure of Cysteamine Dioxygenase Reveals the Origin of the Large Substrate Scope of This Vital Mammalian Enzyme. *Biochemistry* **60**, 3728–3737 (2021).
7. Yahalomi, D. *et al.* A cnidarian parasite of salmon (Myxozoa: *Henneguya*) lacks a mitochondrial genome. *Proc. Natl. Acad. Sci.* **117**, 5358–5363 (2020).
8. Semenza, G. L. Life with Oxygen. *Science* **318**, 62–64 (2007).
9. Semenza, G. L. Oxygen-dependent regulation of mitochondrial respiration by hypoxia-inducible factor 1. *Biochem. J.* **405**, 1–9 (2007).
10. Guzy, R. D. & Schumacker, P. T. Oxygen sensing by mitochondria at complex III: the paradox of increased reactive oxygen species during hypoxia. *Exp. Physiol.* **91**, 807–819 (2006).
11. Prabhakar, N. R. & Semenza, G. L. Oxygen Sensing and Homeostasis. *Physiology* **30**, 340–348 (2015).
12. Semenza, G. L. & Wang, G. L. A nuclear factor induced by hypoxia via de novo protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation. *Mol. Cell. Biol.* **12**, 5447–54 (1992).
13. Kaelin, W. G. & Ratcliffe, P. J. Oxygen Sensing by Metazoans: The Central Role of the HIF Hydroxylase Pathway. *Mol. Cell* **30**, 393–402 (2008).
14. Maxwell, P. H., Pugh, C. W. & Ratcliffe, P. J. Inducible operation of the erythropoietin 3' enhancer in multiple cell lines: evidence for a widespread oxygen-sensing mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 2423 (1993).
15. Schofield, C. J. & Ratcliffe, P. J. Oxygen sensing by HIF hydroxylases. *Nat. Rev. Mol. Cell Biol.* **5**, 343–354 (2004).
16. Wang, G. L., Jiang, B. H., Rue, E. A. & Semenza, G. L. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5510–5514 (1995).
17. Ivan, M. *et al.* HIFα Targeted for VHL-Mediated Destruction by Proline Hydroxylation: Implications for O 2 Sensing. *Science* **292**, 464–468 (2001).
18. Jaakkola, P. *et al.* Targeting of HIF-α to the von Hippel-Lindau Ubiquitylation Complex by O 2 -Regulated Prolyl Hydroxylation. *Science* **292**, 468–472 (2001).
19. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–79 (1998).
20. Lando, D. *et al.* FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev.* **16**, 1466–1471 (2002).
21. Sang, N., Fang, J., Srinivas, V., Leshchinsky, I. & Caro, J. Carboxyl-Terminal Transactivation Activity of Hypoxia-Inducible Factor 1α Is Governed by a von

Hippel-Lindau Protein-Independent, Hydroxylation-Regulated Association with p300/CBP. *Mol. Cell. Biol.* **22**, 2984–2992 (2002).
22. Hewitson, K. S. *et al.* Hypoxia-inducible Factor (HIF) Asparagine Hydroxylase Is Identical to Factor Inhibiting HIF (FIH) and Is Related to the Cupin Structural Family. *J. Biol. Chem.* **277**, 26351–26355 (2002).
23. Epstein, A. C. R. *et al. C. elegans* EGL-9 and Mammalian Homologs Define a Family of Dioxygenases that Regulate HIF by Prolyl Hydroxylation. *Cell* **107**, 43–54 (2001).
24. Bruick, R. K. & McKnight, S. L. A Conserved Family of Prolyl-4-Hydroxylases That Modify HIF. *Science* **294**, 1337–1340 (2001).
25. Clifton, I. J. *et al.* Structural studies on 2-oxoglutarate oxygenases and related double-stranded β-helix fold proteins. *J. Inorg. Biochem.* **100**, 644–669 (2006).
26. Elkins, J. M. *et al.* Structure of Factor-inhibiting Hypoxia-inducible Factor (HIF) Reveals Mechanism of Oxidative Modification of HIF-1α. *J. Biol. Chem.* **278**, 1802–1806 (2003).
27. McDonough, M. A. *et al.* Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc. Natl. Acad. Sci.* **103**, 9814–9819 (2006).
28. Knowles, H. J., Raval, R. R., Harris, A. L. & Ratcliffe, P. J. Effect of ascorbate on the activity of hypoxia-inducible factor in cancer cells. *Cancer Res.* **63**, 1764–1768 (2003).
29. Flashman, E. *et al.* Evidence for the Slow Reaction of Hypoxia-Inducible Factor Prolyl Hydroxylase 2 with Oxygen. *FEBS J.* **277**, 4089 (2010).
30. Cockman, M. E. *et al.* Posttranslational hydroxylation of ankyrin repeats in IκB proteins by the hypoxia-inducible factor (HIF) asparaginyl hydroxylase, factor inhibiting HIF (FIH). *Proc. Natl. Acad. Sci. U. S. A.* **103**, 14767–14772 (2006).
31. Tian, Y.-M. *et al.* Differential Sensitivity of Hypoxia Inducible Factor Hydroxylation Sites to Hypoxia and Hydroxylase Inhibitors. *J. Biol. Chem.* **286**, 13041–13051 (2011).
32. Tarhonskaya, H. *et al.* Investigating the contribution of the active site environment to the slow reaction of hypoxia-inducible factor prolyl hydroxylase domain 2 with oxygen. *Biochem. J.* **463**, 363–372 (2014).
33. Wilkins, S. E. *et al.* Differences in hydroxylation and binding of Notch and HIF-1α demonstrate substrate selectivity for factor inhibiting HIF-1 (FIH-1). *Int. J. Biochem. Cell Biol.* **41**, 1563–1571 (2009).
34. Koivunen, P., Hirsilä, M., Günzler, V., Kivirikko, K. I. & Myllyharju, J. Catalytic Properties of the Asparaginyl Hydroxylase (FIH) in the Oxygen Sensing Pathway Are Distinct from Those of Its Prolyl 4-Hydroxylases. *J. Biol. Chem.* **279**, 9899–9904 (2004).
35. Tarhonskaya, H. *et al.* Kinetic Investigations of the Role of Factor Inhibiting Hypoxia-inducible Factor (FIH) as an Oxygen Sensor. *J. Biol. Chem.* **290**, 19726–19742 (2015).
36. Duan, C. Hypoxia-inducible factor 3 biology: Complexities and emerging themes. *Am. J. Physiol. - Cell Physiol.* **310**, C260–C269 (2016).
37. Yang, S. L., Wu, C., Xiong, Z. F. & Fang, X. Progress on hypoxia-inducible factor-3: Its structure, gene regulation and biological function (Review). *Mol. Med. Rep.* **12**, 2411–2416 (2015).
38. Makino, Y. *et al.* Inhibitory PAS domain protein is a negative regulator of hypoxia-inducible gene expression. *Nature* **414**, 550–554 (2001).
39. Kaidi, A., Williams, A. C. & Paraskeva, C. Interaction between β-catenin and HIF-1 promotes cellular adaptation to hypoxia. *Nat. Cell Biol.* **9**, (2007).

40. Gustafsson, M. V. *et al.* Hypoxia requires Notch signaling to maintain the undifferentiated cell state. *Dev. Cell* **9**, 617–628 (2005).
41. The Nobel Prize in Physiology or Medicine 2019 - Press release. https://www.nobelprize.org/prizes/medicine/2019/press-release/.
42. Fischer, W. W., Hemp, J. & Johnson, J. E. Evolution of Oxygenic Photosynthesis. *Annu. Rev. Earth Planet. Sci.* **44**, 647–683 (2016).
43. Wang, K. L. C., Li, H. & Ecker, J. R. Ethylene Biosynthesis and Signaling Networks. *Plant Cell* **14**, S131–S151 (2002).
44. Schwartz, S. H., Tan, B. C., Gage, D. A., Zeevaart, J. A. D. & McCarty, D. R. Specific Oxidative Cleavage of Carotenoids by VP14 of Maize. *Science* **276**, 1872–1874 (1997).
45. Tsukagoshi, H., Busch, W. & Benfey, P. N. Transcriptional Regulation of ROS Controls Transition from Proliferation to Differentiation in the Root. *Cell* **143**, 606–616 (2010).
46. Apel, K. & Hirt, H. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.* **55**, 373–99 (2004).
47. Gilroy, S. *et al.* A tidal wave of signals: calcium and ROS at the forefront of rapid systemic signaling. *Trends Plant Sci.* **19**, 623–630 (2014).
48. van Dongen, J. T. & Licausi, F. Oxygen Sensing and Signaling. *Annu. Rev. Plant Biol.* **66**, 345–367 (2015).
49. Hammarlund, E. U., Flashman, E., Mohlin, S. & Licausi, F. Oxygen-sensing mechanisms across eukaryotic kingdoms and their roles in complex multicellularity. *Science* **370**, eaba3512 (2020).
50. Holdsworth, M. J. & Gibbs, D. J. Comparative Biology of Oxygen Sensing in Plants and Animals. *Curr. Biol.* **30**, R362–R369 (2020).
51. Giuntoli, B. & Perata, P. Group VII Ethylene Response Factors in *Arabidopsis*: Regulation and Physiological Roles. *Plant Physiol.* **176**, 1143–1155 (2018).
52. Nakano, T., Suzuki, K., Fujimura, T. & Shinshi, H. Genome-Wide Analysis of the ERF Gene Family in *Arabidopsis* and Rice. *Plant Physiol.* **140**, 411–432 (2006).
53. Gibbs, D. J. *et al.* Group VII Ethylene Response Factors Coordinate Oxygen and Nitric Oxide Signal Transduction and Stress Responses in Plants. *Plant Physiol.* **169**, 23–31 (2015).
54. Riechmann, J. L. *et al. Arabidopsis* Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes. *Science* **290**, 2105–2110 (2000).
55. Ohme-Takagi, M. & Shinshi, H. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *Plant Cell* **7**, 173 (1995).
56. Gibbs, D. J. *et al.* Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. *Nature* **479**, 415–418 (2011).
57. Licausi, F. *et al.* Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. *Nature* **479**, 419–422 (2011).
58. White, M. D. *et al.* Plant cysteine oxidases are dioxygenases that directly enable arginyl transferase-catalysed arginylation of N-end rule targets. *Nat. Commun.* **8**, 14690 (2017).
59. Giglione, C., Boularot, A. & Meinnel, T. Protein N-terminal methionine excision. *Cell. Mol. Life Sci.* **61**, 1455–1474 (2004).
60. Ross, S., Giglione, C., Pierre, M., Espagne, C. & Meinnel, T. Functional and Developmental Impact of Cytosolic Protein N-Terminal Methionine Excision in Arabidopsis. *Plant Physiol.* **137**, 623–637 (2005).
61. Garzón, M. *et al.* PRT6/At5g02310 encodes an *Arabidopsis* ubiquitin ligase of the N-end rule pathway with arginine specificity and is not the CER3 locus. *FEBS Lett.*

581, 3189–96 (2007).
62. White, M. D., Kamps, J. J. A. G., East, S., Taylor Kearney, L. J. & Flashman, E. The plant cysteine oxidases from *Arabidopsis thaliana* are kinetically tailored to act as oxygen sensors. *J. Biol. Chem.* **293**, 11786–11795 (2018).
63. Gibbs, D. J. *et al.* Nitric Oxide Sensing in Plants Is Mediated by Proteolytic Control of Group VII ERF Transcription Factors. *Mol. Cell* **53**, 369–379 (2014).
64. Holdsworth, M. J., Vicente, J., Sharma, G., Abbas, M. & Zubrycka, A. The plant N-degron pathways of ubiquitin-mediated proteolysis. *J. Integr. Plant Biol.* **62**, 70–89 (2020).
65. Taylor-Kearney, L. J. *et al.* Plant Cysteine Oxidase Oxygen-Sensing Function Is Conserved in Early Land Plants and Algae. *ACS Bio Med Chem Au* **2**, 521–528 (2022).
66. White, M. D. *et al.* Structures of *Arabidopsis thaliana* oxygen-sensing plant cysteine oxidases 4 and 5 enable targeted manipulation of their activity. *Proc. Natl. Acad. Sci.* **117**, 23140–23147 (2020).
67. Chen, Z. *et al.* Molecular basis for cysteine oxidation by plant cysteine oxidases from *Arabidopsis thaliana*. *J. Struct. Biol.* **213**, 107663 (2021).
68. Gibbs, D. J. *et al.* Oxygen-dependent proteolysis regulates the stability of angiosperm polycomb repressive complex 2 subunit VERNALIZATION 2. **9**, (2018).
69. Weits, D. A. *et al.* An apical hypoxic niche sets the pace of shoot meristem activity. *Nature* **569**, 714–717 (2019).
70. Gasch, P. *et al.* Redundant ERF-VII Transcription Factors Bind to an Evolutionarily Conserved cis-Motif to Regulate Hypoxia-Responsive Gene Expression in Arabidopsis. *Plant Cell* **28**, 160–80 (2016).
71. Hess, N., Klode, M., Anders, M. & Sauter, M. The hypoxia responsive transcription factor genes ERF71/HRE2 and ERF73/HRE1 of *Arabidopsis* are differentially regulated by ethylene. *Physiol. Plant.* **143**, 41–49 (2011).
72. Licausi, F. *et al.* HRE1 and HRE2, two hypoxia-inducible ethylene response factors, affect anaerobic responses in *Arabidopsis thaliana*. *Plant J.* **62**, 302–315 (2010).
73. Hinz, M. *et al. Arabidopsis* RAP2.2 : An Ethylene Response Transcription Factor That Is Important for Hypoxia Survival. *Plant Physiol.* **153**, 757–772 (2010).
74. Voesenek, L. A. C. J. & Bailey-Serres, J. Flood adaptive traits and processes: an overview. *New Phytol.* **206**, 57–73 (2015).
75. Hattori, Y. *et al.* The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. *Nature* **460**, 1026–1030 (2009).
76. Xu, K. *et al.* Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* **442**, 705–708 (2006).
77. Lin, C.-C. *et al.* Regulatory cascade involving transcriptional and N-end rule pathways in rice under submergence. *Proc. Natl. Acad. Sci.* **116**, 3300–3309 (2019).
78. Varland, S., Osberg, C. & Arnesen, T. N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. *Proteomics* **15**, 2385–2401 (2015).
79. Sherman, F., Stewart, J. W. & Tsunasawa, S. Methionine or not methionine at the beginning of a protein. *BioEssays* **3**, 27–31 (1985).
80. Kerwar, S. S., Weissbach, H. & Glenner, G. G. An aminopeptidase activity associated with brain ribosomes. *Arch. Biochem. Biophys.* **143**, 336–7 (1971).
81. Jackson, R. & Hunter, T. Role of methionine in the initiation of haemoglobin synthesis. *Nature* **227**, 672–6 (1970).
82. Palmiter, R. D., Gagnon, J. & Walsh, K. A. Ovalbumin: a secreted protein without a transient hydrophobic leader sequence. *Proc. Natl. Acad. Sci.* **75**, 94–98 (1978).

83. Arfin, S. M. & Bradshaw, R. A. Cotranslational processing and protein turnover in eukaryotic cells. *Biochemistry* **27**, 7979–7984 (1988).
84. Bradshaw, R. A., Brickey, W. W. & Walker, K. W. N-Terminal processing: the methionine aminopeptidase and Nα-acetyl transferase families. *Trends Biochem. Sci.* **23**, 263–267 (1998).
85. Aksnes, H., Ree, R. & Arnesen, T. Co-translational, Post-translational, and Non-catalytic Roles of N-Terminal Acetyltransferases. *Mol. Cell* **73**, 1097–1114 (2019).
86. Linster, E. & Wirtz, M. N-terminal acetylation: an essential protein modification emerges as an important regulator of stress responses. *J. Exp. Bot.* **69**, 4555–4568 (2018).
87. Arnesen, T. *et al.* Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc. Natl. Acad. Sci.* **106**, 8157–8162 (2009).
88. Aksnes, H. *et al.* An Organellar Nα-Acetyltransferase, Naa60, Acetylates Cytosolic N Termini of Transmembrane Proteins and Maintains Golgi Integrity. *Cell Rep.* **10**, 1362–1374 (2015).
89. Bienvenut, W. V. *et al.* Comparative Large Scale Characterization of Plant versus Mammal Proteins Reveals Similar and Idiosyncratic N-α-Acetylation Features. *Mol. Cell. Proteomics* **11**, M111.015131 (2012).
90. Polevoda, B., Norbeck, J., Takakura, H., Blomberg, A. & Sherman, F. Identification and specificities of N-terminal acetyltransferases from Saccharomyces cerevisiae. *EMBO J.* **18**, 6155–68 (1999).
91. Van Damme, P. *et al.* N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 12449–54 (2012).
92. Aksnes, H., Drazic, A., Marie, M. & Arnesen, T. First Things First: Vital Protein Marks by N-Terminal Acetyltransferases. *Trends Biochem. Sci.* **41**, 746–760 (2016).
93. Hole, K. *et al.* The Human N-Alpha-Acetyltransferase 40 (hNaa40p/hNatD) Is Conserved from Yeast and N-Terminally Acetylates Histones H2A and H4. *PLoS One* **6**, e24713 (2011).
94. Song, O. K., Wang, X., Waterborg, J. H. & Sternglanz, R. An Nα-acetyltransferase responsible for acetylation of the N-terminal residues of histones H4 and H2A. *J. Biol. Chem.* **278**, 38109–38112 (2003).
95. Redman, K. & Rubenstein, P. A. NH2-terminal processing of *Dictyostelium discoideum* actin in vitro. *J. Biol. Chem.* **256**, 13226–13229 (1981).
96. Rubenstein, P. A. & Martin, D. J. NH2-terminal processing of actin in mouse L-cells in vivo. *J. Biol. Chem.* **258**, 3961–3966 (1983).
97. Drazic, A. *et al.* NAA80 is actin's N-terminal acetyltransferase and regulates cytoskeleton assembly and cell motility. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 4399–4404 (2018).
98. Dinh, T. V. *et al.* Molecular identification and functional characterization of the first Nα-acetyltransferase in plastids by global acetylome profiling. *Proteomics* **15**, 2426–2435 (2015).
99. Polevoda, B. & Sherman, F. N-terminal Acetyltransferases and Sequence Requirements for N-terminal Acetylation of Eukaryotic Proteins. *J. Mol. Biol.* **325**, 595–622 (2003).
100. Van Damme, P. *et al.* NatF Contributes to an Evolutionary Shift in Protein N-Terminal Acetylation and Is Important for Normal Chromosome Segregation. *PLoS Genet.* **7**, e1002169 (2011).
101. Strous, G. J. A. M., Westreenen, H., Bloemendal, H., van Westreenen, H. &

Bloemendal, H. Synthesis of lens protein in vitro. N-terminal acetylation of alpha-crystallin. *Eur. J. Biochem.* **38**, 79–85 (1973).
102. Strous, G. J., Berns, A. J. & Bloemendal, H. N-terminal acetylation of the nascent chains of alpha-crystallin. *Biochem. Biophys. Res. Commun.* **58**, 876–84 (1974).
103. Scott, D. C., Monda, J. K., Bennett, E. J., Harper, J. W. & Schulman, B. A. N-terminal acetylation acts as an avidity enhancer within an interconnected multiprotein complex. *Science* **334**, 674–678 (2011).
104. Monda, J. K. *et al.* Structural Conservation of Distinctive N-terminal Acetylation-Dependent Interactions across a Family of Mammalian NEDD8 Ligation Enzymes. *Structure* **21**, 42–53 (2013).
105. Scott, D. C. *et al.* Blocking an N-terminal acetylation–dependent protein interaction inhibits an E3 ligase. *Nat. Chem. Biol.* **13**, 850–857 (2017).
106. Hwang, C.-S., Shemorry, A. & Varshavsky, A. N-Terminal Acetylation of Cellular Proteins Creates Specific Degradation Signals. *Science* **327**, 973–977 (2010).
107. Gibbs, D. J., Bailey, M. & Etherington, R. D. A stable start: cotranslational Nt-acetylation promotes proteome stability across kingdoms. *Trends Cell Biol.* **32**, 374–376 (2022).
108. Linster, E. *et al.* Cotranslational N-degron masking by acetylation promotes proteome stability in plants. *Nat. Commun.* **13**, 810 (2022).
109. Mueller, F. *et al.* Overlap of NatA and IAP substrates implicates N-terminal acetylation in protein stabilization. *Sci. Adv.* **7**, (2021).
110. Dormeyer, W., Mohammed, S., Breukelen, B. van, Krijgsveld, J. & Heck, A. J. R. Targeted Analysis of Protein Termini. *J. Proteome Res.* **6**, 4634–4645 (2007).
111. Martin, D. D. O., Beauchamp, E. & Berthiaume, L. G. Post-translational myristoylation: Fat matters in cellular life and death. *Biochimie* **93**, 18–31 (2011).
112. Pepinsky, R. B. *et al.* Identification of a Palmitic Acid-modified Form of Human Sonic hedgehog. *J. Biol. Chem.* **273**, 14037–14045 (1998).
113. Kleuss, C. & Krause, E. Galpha(s) is palmitoylated at the N-terminal glycine. *EMBO J.* **22**, 826–32 (2003).
114. Foyn, H. *et al.* Protein N-terminal Acetyltransferases Act as N-terminal Propionyltransferases In Vitro and In Vivo. *Mol. Cell. Proteomics* **12**, 42–54 (2013).
115. Towler, D. A., Eubanks, S. R., Towery, D. S., Adams, S. P. & Glaser, L. Amino-terminal processing of proteins by N-myristoylation. Substrate specificity of N-myristoyl transferase. *J. Biol. Chem.* **262**, 1030–1036 (1987).
116. Podell, S. & Gribskov, M. Predicting N-terminal myristoylation sites in plant proteins. *BMC Genomics* **5**, 37 (2004).
117. Maurer-Stroh, S., Eisenhaber, B. & Eisenhaber, F. N-terminal N -myristoylation of proteins: prediction of substrate proteins from amino acid sequence 1 1Edited by J. Thornton. *J. Mol. Biol.* **317**, 541–557 (2002).
118. Linder, M. E. *et al.* Lipid modifications of G proteins: alpha subunits are palmitoylated. *Proc. Natl. Acad. Sci.* **90**, 3675–3679 (1993).
119. Resh, M. D. Palmitoylation of Ligands, Receptors, and Intracellular Signaling Molecules. *Sci. STKE* **2006**, re14 (2006).
120. Hayashi, N. & Titani, K. N-myristoylated proteins, key components in intracellular signal transduction systems enabling rapid and flexible cell responses. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* **86**, 494–508 (2010).
121. Stock, A., Clarke, S., Clarke, C. & Stock, J. N-terminal methylation of proteins: Structure, function and specificity. *FEBS Lett.* **220**, 8–14 (1987).
122. Ciechanover, A. & Ben-Saadon, R. N-terminal ubiquitination: more protein substrates join in. *Trends Cell Biol.* **14**, 103–6 (2004).

123. Hershko, A., Ciechanover, A. & Varshavsky, A. The ubiquitin system. *Nat. Med.* **6**, 1073–1081 (2000).
124. Chen, S.-J., Wu, X., Wadas, B., Oh, J.-H. & Varshavsky, A. An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes. *Science* **355**, (2017).
125. Kim, J.-M. *et al.* Formyl-methionine as an N-degron of a eukaryotic N-end rule pathway. *Science* **362**, eaat0174 (2018).
126. Wang, H., Piatkov, K. I., Brower, C. S. & Varshavsky, A. Glutamine-Specific N-Terminal Amidase, a Component of the N-End Rule Pathway. *Mol. Cell* **34**, 686–695 (2009).
127. Grigoryev, S. *et al.* A Mouse Amidase Specific for N-terminal Asparagine. *J. Biol. Chem.* **271**, 28521–28532 (1996).
128. Hu, R.-G. *et al.* The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. *Nature* **437**, 981–986 (2005).
129. Soffer, R. L. Enzymatic Modification of Proteins: V. PROTEIN ACCEPTOR SPECIFICITY IN THE ARGININE-TRANSFER REACTION. *J. Biol. Chem.* **246**, 1602–1606 (1971).
130. Gonda, D. K. *et al.* Universality and Structure of the N-end Rule. *J. Biol. Chem.* **264**, 16700–16712 (1989).
131. Kwon, Y. T. An Essential Role of N-Terminal Arginylation in Cardiovascular Development. *Science* **297**, 96–99 (2002).
132. Kim, J. G. *et al.* Signaling Pathways Regulated by UBR Box-Containing E3 Ligases. *Int. J. Mol. Sci.* **22**, 8323 (2021).
133. Yang, Q., Zhao, J., Chen, D. & Wang, Y. E3 ubiquitin ligases: styles, structures and functions. *Mol. Biomed.* **2**, 23 (2021).
134. Scheffner, M., Nuber, U. & Huibregtse, J. M. Protein ubiquitination involving an E1–E2–E3 enzyme ubiquitin thioester cascade. *Nature* **373**, 81–83 (1995).
135. Bard, J. A. M. *et al.* Structure and Function of the 26S Proteasome. *Annu. Rev. Biochem.* **87**, 697–724 (2018).
136. Clague, M. J., Urbé, S. & Komander, D. Breaking the chains: deubiquitylating enzyme specificity begets function. *Nat. Rev. Mol. Cell Biol. 2019 206* **20**, 338–352 (2019).
137. Wang, J. *et al.* Target site specificity and in vivo complexity of the mammalian arginylome. *Sci. Rep.* **8**, 16177 (2018).
138. Oh, J.-H., Hyun, J.-Y., Chen, S.-J. & Varshavsky, A. Five enzymes of the Arg/N-degron pathway form a targeting complex: The concept of superchanneling. *Proc. Natl. Acad. Sci.* **117**, 10778–10788 (2020).
139. Kim, B. H. *et al.* Crystal structure of the Ate1 arginyl-tRNA-protein transferase and arginylation of N-degron substrates. *Proc. Natl. Acad. Sci.* **119**, e2209597119 (2022).
140. Van, V., Ejimogu, N.-E., Bui, T. S. & Smith, A. T. The Structure of *Saccharomyces cerevisiae* Arginyltransferase 1 (ATE1). *J. Mol. Biol.* **434**, 167816 (2022).
141. Kim, H.-K. *et al.* The N-Terminal Methionine of Cellular Proteins as a Degradation Signal. *Cell* **156**, 158–169 (2014).
142. Park, S.-E. *et al.* Control of mammalian G protein signaling by N-terminal acetylation and the N-end rule pathway. *Science* **347**, 1249–1252 (2015).
143. Xu, F. *et al.* Two N-Terminal Acetyltransferases Antagonistically Regulate the Stability of a Nod-Like Receptor in Arabidopsis. *Plant Cell* **27**, 1547–1562 (2015).
144. Stipanuk, M. H., Simmons, C. R., Andrew Karplus, P. & Dominy, J. E. Thiol dioxygenases: unique families of cupin proteins. *Amino Acids* **41**, 91–102 (2011).
145. Dunwell, J. M., Khuri, S. & Gane, P. J. Microbial relatives of the seed storage proteins of higher plants: conservation of structure and diversification of function

during evolution of the cupin superfamily. *Microbiol. Mol. Biol. Rev.* **64**, 153–179 (2000).
146. Lombardini, J. B., Singer, T. P. & Boyer, P. D. Cystein oxygenase. II. Studies on the mechanism of the reaction with 18oxygen. *J. Biol. Chem.* **244**, 1172–1175 (1969).
147. Sakakibara, S. *et al.* Purification and some properties of rat liver cysteine oxidase (cysteine dioxygenase). **422**, 273–279 (1976).
148. Ueki, I. & Stipanuk, M. H. Enzymes of the Taurine Biosynthetic Pathway Are Expressed in Rat Mammary Gland ,. *J. Nutr.* **137**, 1887–1894 (2007).
149. Vitvitsky, V., Garg, S. K. & Banerjee, R. Taurine biosynthesis by neurons and astrocytes. *J. Biol. Chem.* **286**, 32002–32010 (2011).
150. Heafield, M. T. *et al.* Plasma cysteine and sulphate levels in patients with motor neurone, Parkinson's and Alzheimer's disease. *Neurosci. Lett.* **110**, 216–220 (1990).
151. Sturman, J. A. Taurine in development. *Physiol. Rev.* **73**, 119–147 (1993).
152. Huxtable, R. J. Taurine in the central nervous system and the mammalian actions of taurine. *Prog. Neurobiol.* **32**, 471–533 (1989).
153. Stipanuk, M. H., Ueki, I., Dominy Jr, J. E., Simmons, C. R. & Hirschberger, L. L. Cysteine dioxygenase: a robust system for regulation of cellular cysteine levels. *Amino Acids* **37**, 55–63 (2009).
154. Ueki, I. *et al.* Knockout of the murine cysteine dioxygenase gene results in severe impairment in ability to synthesize taurine and an increased catabolism of cysteine to hydrogen sulfide. *Am. J. Physiol. Metab.* **301**, E668–E684 (2011).
155. Simmons, C. R. *et al.* Crystal structure of mammalian cysteine dioxygenase. A novel mononuclear iron center for cysteine thiol oxidation. *J. Biol. Chem.* **281**, 18723–33 (2006).
156. McCoy, J. G. *et al.* Structure and mechanism of mouse cysteine dioxygenase. *Proc. Natl. Acad. Sci.* **103**, 3084–3089 (2006).
157. Ye, S. *et al.* An Insight into the Mechanism of Human Cysteine Dioxygenase. *J. Biol. Chem.* **282**, 3391–3402 (2007).
158. Aloi, S., Davies, C. G., Karplus, P. A., Wilbanks, S. M. & Jameson, G. N. L. Substrate Specificity in Thiol Dioxygenases. *Biochemistry* **58**, 2398–2407 (2019).
159. Driggers, C. M. *et al.* Structure-Based Insights into the Role of the Cys–Tyr Crosslink and Inhibitor Recognition by Mammalian Cysteine Dioxygenase. *J. Mol. Biol.* **428**, 3999–4012 (2016).
160. Forbes, D. L., Meneely, K. M., Chilton, A. S., Lamb, A. L. & Ellis, H. R. The 3-His Metal Coordination Site Promotes the Coupling of Oxygen Activation to Cysteine Oxidation in Cysteine Dioxygenase. *Biochemistry* **59**, 2022–2031 (2020).
161. Driggers, C. M., Hartman, S. J. & Karplus, P. A. Structures of Arg- and Gln-type bacterial cysteine dioxygenase homologs. *Protein Sci.* **24**, 154–161 (2015).
162. Li, J., Koto, T., Davis, I. & Liu, A. Probing the Cys-Tyr Cofactor Biogenesis in Cysteine Dioxygenase by the Genetic Incorporation of Fluorotyrosine. *Biochemistry* **58**, 2218–2227 (2019).
163. Dominy, J. E. *et al.* Synthesis of Amino Acid Cofactor in Cysteine Dioxygenase Is Regulated by Substrate and Represents a Novel Post-translational Regulation of Activity. *J. Biol. Chem.* **283**, 12188–12201 (2008).
164. Cavallini, D., de Marco, C., Scandurra, R., Dupré, S. & Graziani, M. T. The Enzymatic Oxidation of Cysteamine to Hypotaurine. *J. Biol. Chem.* **241**, 3189–3196 (1966).
165. Cavallini, D., Scandurra, R. & De Marco, C. The Enzymatic Oxidation of Cysteamine to Hypotaurine in the Presence of Sulfide. *J. Biol. Chem.* **238**, 2999–3005 (1963).
166. Dominy, J. E. *et al.* Discovery and Characterization of a Second Mammalian Thiol

Dioxygenase, Cysteamine Dioxygenase. *J. Biol. Chem.* **282**, 25189–25198 (2007).
167. Ueki, I. & Stipanuk, M. H. 3T3-L1 Adipocytes and Rat Adipose Tissue Have a High Capacity for Taurine Synthesis by the Cysteine Dioxygenase/Cysteinesulfinate Decarboxylase and Cysteamine Dioxygenase Pathways ,. *J. Nutr.* **139**, 207–214 (2009).
168. Dominy, J. E., Simmons, C. R., Karplus, P. A., Gehring, A. M. & Stipanuk, M. H. Identification and Characterization of Bacterial Cysteine Dioxygenases: a New Route of Cysteine Degradation for Eubacteria. *J. Bacteriol.* **188**, 5561–5569 (2006).
169. Bruland, N., Wu, J. H., Steinbu, A., Wu¨bbeler, J. H. & Steinbu¨chel, A. 3-Mercaptopropionate Dioxygenase, a Cysteine Dioxygenase Homologue, Catalyzes the Initial Step of 3-Mercaptopropionate Catabolism in the 3,3-Thiodipropionic Acid-degrading Bacterium *Variovorax paradoxus*. *J. Biol. Chem.* **284**, 660–672 (2009).
170. Brandt, U., Schürmann, M. & Steinbüchel, A. Mercaptosuccinate Dioxygenase, a Cysteine Dioxygenase Homologue, from *Variovorax paradoxus* Strain B4 Is the Key Enzyme of Mercaptosuccinate Degradation. *J. Biol. Chem.* **289**, 30800–30809 (2014).
171. Brandt, U., Galant, G., Meinert-Berning, C. & Steinbüchel, A. Functional analysis of active amino acid residues of the mercaptosuccinate dioxygenase of *Variovorax paradoxus* B4. *Enzyme Microb. Technol.* **120**, 61–68 (2019).
172. Li, W., Blaesi, E. J., Pecore, M. D., Crowell, J. K. & Pierce, B. S. Second-sphere interactions between the C93-Y157 cross-link and the substrate-bound Fe site influence the O_2 coupling efficiency in mouse cysteine dioxygenase. *Biochemistry* **52**, 9104–19 (2013).
173. Cavallini, D., De Marco, C. & Mondovi, B. Detection and Distribution of Enzymes for oxidizing Thiocysteamine. *Nature* **192**, 557–558 (1961).
174. Rotilio, G., Federici, G., Calabrese, L., Costa, M. & Cavallini, D. An Electron Paramagnetic Resonance Study of the Nonheme Iron of Cysteamine Oxygenase. *J. Biol. Chem.* **245**, 6235–6236 (1970).
175. Richerson, R. B. & Ziegler, D. M. Cysteamine dioxygenase. in *Methods in Enzymology* vol. 143 410–415 (Academic Press, 1987).
176. Coloso, R. M., Hirschberger, L. L., Dominy, J. E., Lee, J.-I. & Stipanuk, M. H. Cysteamine Dioxygenase: Evidence for the Physiological Conversion of Cysteamine to Hypotaurine in Rat and Mouse Tissues. in *Taurine 6* vol. 583 25–36 (Springer US, 2006).
177. Simmons, C. R., Hirschberger, L. L., Machi, M. S. & Stipanuk, M. H. Expression, purification, and kinetic characterization of recombinant rat cysteine dioxygenase, a non-heme metalloenzyme necessary for regulation of cellular cysteine levels. *Protein Expr. Purif.* **47**, 74–81 (2006).
178. Wang, Y. *et al.* Cofactor Biogenesis in Cysteamine Dioxygenase: C−F Bond Cleavage with Genetically Incorporated Unnatural Tyrosine. *Angew. Chemie Int. Ed.* **57**, 8149–8153 (2018).
179. Lee, M. J. *et al.* RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. *Proc. Natl. Acad. Sci.* **102**, 15030–15035 (2005).
180. Jaba, I. M. *et al.* NO triggers RGS4 degradation to coordinate angiogenesis and cardiomyocyte growth. *J. Clin. Invest.* **123**, 1718–1731 (2013).
181. Olechnowicz, S. W. Z., Fedele, A. O. & Peet, D. J. Hypoxic Induction of the Regulator of G-Protein Signalling 4 Gene Is Mediated by the Hypoxia-Inducible Factor Pathway. *PLoS One* **7**, e44564 (2012).
182. Jin, Y. *et al.* RGS5, a hypoxia-inducible apoptotic stimulator in endothelial cells. *J.*

Biol. Chem. **284**, 23436–23443 (2009).
183. Zahoor, M. *et al.* Hypoxia promotes IL-32 expression in myeloma cells, and high expression is associated with poor survival and bone loss. *Blood Adv.* **1**, 2656–2666 (2017).
184. Fernandez, R. L., Dillon, S. L., Stipanuk, M. H., Fox, B. G. & Brunold, T. C. Spectroscopic Investigation of Cysteamine Dioxygenase. *Biochemistry* **59**, 2450–2458 (2020).
185. Wang, Y. *et al.* Characterization of the nonheme iron center of cysteamine dioxygenase and its interaction with substrates. *J. Biol. Chem.* **295**, 11789–11802 (2020).
186. Simmons, C. R. *et al.* A putative Fe2+-bound persulfenate intermediate in cysteine dioxygenase. *Biochemistry* **47**, 11390–11392 (2008).
187. Driggers, C. M. *et al.* Cysteine Dioxygenase Structures from pH4 to 9: Consistent Cys-Persulfenate Formation at Intermediate pH and a Cys-Bound Enzyme at Higher pH. *J. Mol. Biol.* **425**, 3121–3136 (2013).
188. Kwon, Y. T. *et al.* An essential role of N-terminal arginylation in cardiovascular development. *Science* **297**, 96–99 (2002).
189. Arnold, C. *et al.* Hypertension-evoked RhoA activity in vascular smooth muscle cells requires RGS5. *FASEB J.* **32**, 2021–2035 (2018).
190. Hamzah, J. *et al.* Vascular normalization in Rgs5-deficient tumours promotes immune destruction. *Nature* **453**, 410–414 (2008).
191. Nold-Petry, C. A. *et al.* IL-32 Promotes Angiogenesis. *J. Immunol.* **192**, 589–602 (2014).
192. Sloot, Y. J. E., Smit, J. W., Joosten, L. A. B. & Netea-Maier, R. T. Insights into the role of IL-32 in cancer. *Semin. Immunol.* **38**, 24–32 (2018).
193. Bateman, A. *et al.* UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
194. Zou, K., Ouyang, Q., Li, H. & Zheng, J. A global characterization of the translational and transcriptional programs induced by methionine restriction through ribosome profiling and RNA-seq. *BMC Genomics* **18**, 1–12 (2017).
195. Cigan, A. M., Feng, L. & Donahue, T. F. tRNAimet Functions in Directing the Scanning Ribosome to the Start Site of Translation. *Science* **242**, 93–97 (1988).
196. Wu, X., Ma, Z., Feng, T. & Zhu, C. Radical-mediated rearrangements: past, present, and future. *Chem. Soc. Rev.* **50**, 11577–11613 (2021).
197. UniProtKB/Swiss-Prot Release 2022_05 statistics. https://web.expasy.org/docs/relnotes/relstat.html.
198. Davydov, I. V. & Varshavsky, A. RGS4 Is Arginylated and Degraded by the N-end Rule Pathway in Vitro. *J. Biol. Chem.* **275**, 22931–22941 (2000).
199. Vacic, V., Iakoucheva, L. M. & Radivojac, P. Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. *Bioinformatics* **22**, 1536–1537 (2006).
200. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990).
201. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188–1190 (2004).
202. Cornish-Bowden, A. The origins of enzyme kinetics. *FEBS Lett.* **587**, 2725–2730 (2013).
203. Johnson, K. A. & Goody, R. S. The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry* **50**, 8264–8269 (2011).
204. Smith, E. & Keeley, T. P. Monitoring ADO dependent proteolysis in cells using

fluorescent reporter proteins. in *Methods in Enzymology* (Academic Press, 2023). doi:10.1016/bs.mie.2023.02.004.
205. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
206. Liu, Z. *et al.* Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci. Rep.* **7**, 2193 (2017).
207. Daniels, R. W., Rossano, A. J., Macleod, G. T. & Ganetzky, B. Expression of Multiple Transgenes from a Single Construct Using Viral 2A Peptides in Drosophila. *PLoS One* **9**, e100637 (2014).
208. Ahier, A. & Jarriault, S. Simultaneous expression of multiple proteins under a single promoter in *Caenorhabditis elegans* via a versatile 2A-based toolkit. *Genetics* **196**, 605–613 (2014).
209. The Human Protein Atlas. https://www.proteinatlas.org/.
210. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, (2015).
211. Yang, F. *et al.* The Ubiquitin Ligase Ubr2, a Recognition E3 Component of the N-End Rule Pathway, Stabilizes Tex19.1 during Spermatogenesis. *PLoS One* **5**, e14017 (2010).
212. Heim, R., Prasher, D. C. & Tsien, R. Y. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci.* **91**, 12501–12504 (1994).
213. Coralli, C., Cemazar, M., Kanthou, C., Tozer, G. M. & Dachs, G. U. Limitations of the Reporter Green Fluorescent Protein under Simulated Tumor Conditions 1. *CANCER Res.* **61**, 4784–4790 (2001).
214. Van Damme, P., Hole, K., Gevaert, K. & Arnesen, T. N-terminal acetylome analysis reveals the specificity of Naa50 (Nat5) and suggests a kinetic competition between N-terminal acetyltransferases and methionine aminopeptidases. *Proteomics* **15**, 2436–2446 (2015).
215. Van Damme, P. *et al.* A Role for Human N-alpha Acetyltransferase 30 (Naa30) in Maintaining Mitochondrial Integrity. *Mol. Cell. Proteomics* **15**, 3361–3372 (2016).
216. Myklebust, L. M. *et al.* Biochemical and cellular analysis of Ogden syndrome reveals downstream Nt-acetylation defects. *Hum. Mol. Genet.* **24**, 1956–76 (2015).
217. Varland, S. *et al.* N-terminal Acetylation Levels Are Maintained During Acetyl-CoA Deficiency in *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* **17**, 2309–2323 (2018).
218. Van Damme, P., Støve, S. I., Glomnes, N., Gevaert, K. & Arnesen, T. A *Saccharomyces cerevisiae* Model Reveals In Vivo Functional Impairment of the Ogden Syndrome N-Terminal Acetyltransferase NAA10 Ser37Pro Mutant. *Mol. Cell. Proteomics* **13**, 2031–2041 (2014).
219. Lin, Y., Devin, A., Rodriguez, Y. & Liu, Z. G. Cleavage of the death domain kinase RIP by Caspase-8 prompts TNF-induced apoptosis. *Genes Dev.* **13**, 2514–2526 (1999).
220. Newton, K. *et al.* Cleavage of RIPK1 by caspase-8 is crucial for limiting apoptosis and necroptosis. *Nat. 2019 5747778* **574**, 428–431 (2019).
221. Drazic, A. & Arnesen, T. [14C]-Acetyl-Coenzyme A-Based In Vitro N-Terminal Acetylation Assay. in *Methods in molecular biology (Clifton, N.J.)* vol. 1574 1–8 (Humana Press Inc., 2017).
222. Driessen, H. P. C., De Jong, W. W., Tesser, G. I. & Bloemendal, H. The Mechanism of N-Terminal Acetylation of Protein. *Crit. Rev. Biochem.* **18**, 281–325 (1985).
223. Pestana, A. & Pitot, H. C. Acetylation of nascent polypeptide chains on rat liver polyribosomes in vivo and in vitro. *Biochemistry* **14**, 1404–1412 (1975).

224. Roepstorff, P. & Fohlman, J. Letter to the editors. *Biol. Mass Spectrom.* **11**, 601–601 (1984).
225. Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T. & Watson, J. T. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.* **59**, 2621–2625 (1987).
226. Lau, J. L. & Dunn, M. K. Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorg. Med. Chem.* **26**, 2700–2707 (2018).
227. Otvos, L. & Wade, J. D. Current challenges in peptide-based drug discovery. *Front. Chem.* **2**, 62 (2014).
228. Mole, D. R. *et al.* 2-oxoglutarate analogue inhibitors of HIF prolyl hydroxylase. *Bioorg. Med. Chem. Lett.* **13**, 2677–80 (2003).
229. Rose, N. R., McDonough, M. A., King, O. N. F., Kawamura, A. & Schofield, C. J. Inhibition of 2-oxoglutarate dependent oxygenases. *Chem. Soc. Rev.* **40**, 4364 (2011).
230. Wang, L. *et al.* Therapeutic peptides: current applications and future directions. *Signal Transduct. Target. Ther. 2022 71* **7**, 1–27 (2022).
231. Yuan, J., Lovejoy, D. B. & Richardson, D. R. Novel di-2-pyridyl–derived iron chelators with marked and selective antitumor activity: in vitro and in vivo assessment. *Blood* **104**, 1450–1458 (2004).
232. Schofield, C. & Hausinger, R. *2-Oxoglutarate-Dependent Oxygenases*. (Royal Society of Chemistry, 2015).
233. Cornish-Bowden, A. Reversible Inhibition and Activation. in *Fundamentals of Enzyme Kinetics* 133–165 (Wiley-Blackwell, 2012).
234. Berg, J. M., Tymoczko, J. L., Gatto, G. J. J. & Stryer, L. Enzymes: Basic Concepts and Kinetics. in *Biochemistry* 535–623 (2019).
235. Liu, Z. D. & Hider, R. C. Design of iron chelators with therapeutic application. *Coord. Chem. Rev.* **232**, 151–171 (2002).
236. Yu, Y. *et al.* Thiosemicarbazones from the Old to New: Iron Chelators That Are More Than Just Ribonucleotide Reductase Inhibitors. *J. Med. Chem.* **52**, 5271–5294 (2009).
237. Lamers, C. Overcoming the shortcomings of peptide-based therapeutics. *Futur. Drug Discov.* **4**, (2022).
238. Bruno, B. J., Miller, G. D. & Lim, C. S. Basics and recent advances in peptide and protein drugdelivery. *Ther. Deliv.* **4**, 1443 (2013).
239. Renukuntla, J., Vadlapudi, A. D., Patel, A., Boddu, S. H. S. & Mitra, A. K. Approaches for enhancing oral bioavailability of peptides and proteins. *Int. J. Pharm.* **447**, 75–93 (2013).
240. Simmons, C. R. *et al.* A Putative Fe 2+ -Bound Persulfenate Intermediate in Cysteine Dioxygenase. *Biochemistry* **47**, 11390–11392 (2008).
241. Simmons, C. R. *et al.* Crystal structure of mammalian cysteine dioxygenase. A novel mononuclear iron center for cysteine thiol oxidation. *J. Biol. Chem.* **281**, 18723–18733 (2006).
242. Davies, C. G., Fellner, M., Tchesnokov, E. P., Wilbanks, S. M. & Jameson, G. N. L. The Cys-Tyr Cross-Link of Cysteine Dioxygenase Changes the Optimal pH of the Reaction without a Structural Change. *Biochemistry* **53**, 7961–7968 (2014).
243. Fellner, M. *et al.* Influence of cysteine 164 on active site structure in rat cysteine dioxygenase. *J. Biol. Inorg. Chem.* **21**, 501–510 (2016).
244. Li, J. *et al.* Cleavage of a carbon–fluorine bond by an engineered cysteine dioxygenase. *Nat. Chem. Biol.* **14**, 853–860 (2018).

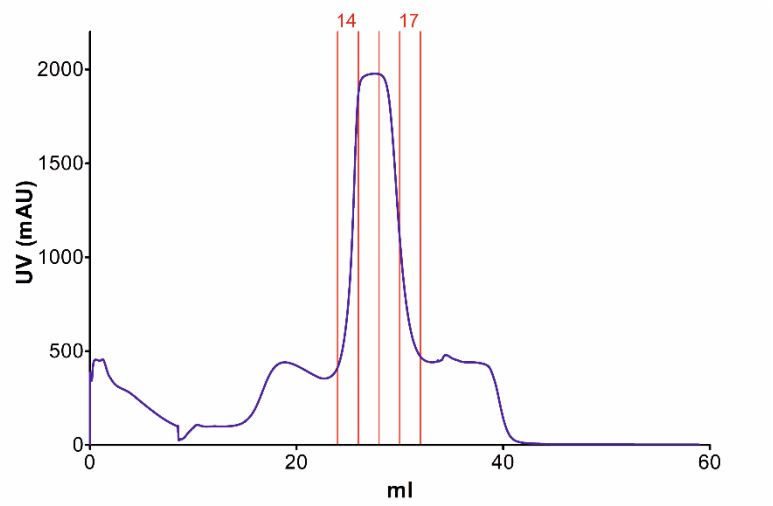
245. Dunwell, J. M. Cupins: A New Superfamily of Functionally Diverse Proteins that Include Germins and Plant Storage Proteins. *Biotechnol. Genet. Eng. Rev.* **15**, 1–32 (1998).
246. Dunwell, J. M., Purvis, A. & Khuri, S. Cupins: The most functionally diverse protein superfamily? *Phytochemistry* **65**, 7–17 (2004).
247. Stipanuk, M. H., Simmons, C. R., Andrew Karplus, P. & Dominy, J. E. Thiol dioxygenases: unique families of cupin proteins. *Amino Acids* **41**, 91–102 (2011).
248. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
249. Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276-9 (2022).
250. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
251. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
252. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinforma.* **54**, 5.6.1-5.6.37 (2016).
253. Fiser, A., Do, R. K. G. & Šali, A. Modeling of loops in protein structures. *Protein Sci.* **9**, 1753–1773 (2000).
254. MODELLER Tutorial. https://salilab.org/modeller/tutorial/.
255. Shen, M. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–24 (2006).
256. Melo, F. & Sali, A. Fold assessment for comparative protein structure modeling. *Protein Sci.* **16**, 2412 (2007).
257. Melo, F., Sa´nchez, R., Sa´nchez, S. & Sali, A. Statistical potentials for fold assessment. (2002) doi:10.1110/ps.22802.
258. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
259. Sarkar, B., Kulharia, M. & Mantha, A. K. Understanding human thiol dioxygenase enzymes: structure to function, and biology to pathology. *Int. J. Exp. Pathol.* **98**, 52–66 (2017).
260. Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R. & Goodenough, P. W. Evolution of functional diversity in the cupin superfamily. *Trends in Biochemical Sciences* vol. 26 740–746 (2001).
261. Njeri, C. W. & Ellis, H. R. Shifting redox states of the iron center partitions CDO between crosslink formation or cysteine oxidation. *Arch. Biochem. Biophys.* **558**, 61–69 (2014).
262. Guo, Q. *et al.* Molecular basis for cysteine oxidation by Plant Cysteine Oxidases from Arabidopsis thaliana *bioRxiv* 2020.09.15.298976 (2020) doi:10.1101/2020.09.15.298976.
263. Xu, Y. *et al.* CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. *Nucleic Acids Res.* **46**, W374–W379 (2018).
264. Harpaz, Y., Gerstein, M. & Chothia, C. Volume changes on protein folding. *Structure* **2**, 641–649 (1994).
265. Weits, D. A. *et al.* Acquisition of hypoxia inducibility by oxygen sensing N-terminal cysteine oxidase in spermatophytes. *Plant. Cell Environ.* **46**, 322–338 (2023).
266. Schrödinger LLC. The PyMOL Molecular Graphics System Version 2.3.0 (Open Source).
267. CavityPlus computation. http://162.105.160.28:2017/cavityplus#/computation.

Appendix 1: Sequence of recombinant *N*-His$_6$-tagged *Hs*ADO and representative UV traces from purification

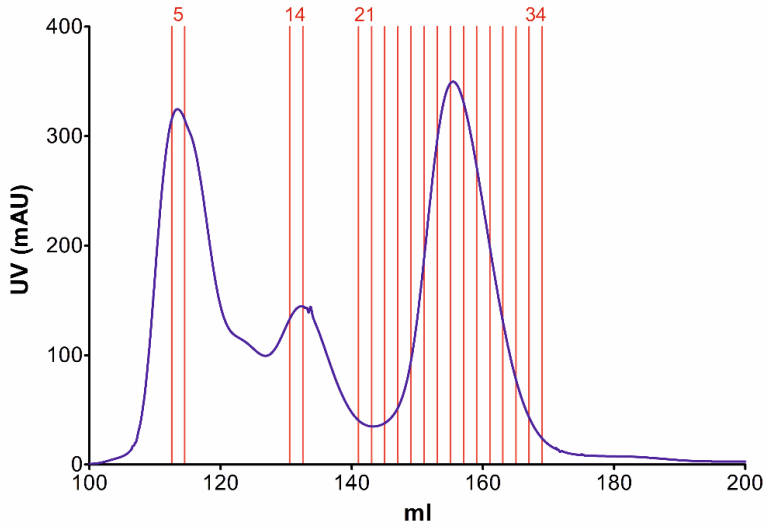
Amino acid sequence, molecular weight, and extinction coefficient of recombinant *N*-His$_6$-tagged *Hs*ADO protein:

Molecular Weight	**Sequence**
31914.19	MGSSHHHHHHSSGLVPRGSHMPRDNMASLIQRIARQACLTFRGSGGGRG ASDRDAASGPEAPMQPGFPENLSKLKSLLTQLRAEDLNIAPRKATLQPLPP NLPPVTYMHIYETDGFSLGVFLLKSGTSIPLHDHPGMHGMLKVLYGTVRI SCMDKLDAGGGQRPRALPPEQQFEPPLQPREREAVRPGVLRSRAEYTEAS GPCILTPHRDNLHQIDAVEGPAAFLDILAPPYDPDDGRDCHYYRVLEPVRP KEASSSACDLPREVWLLETPQADDFWCEGEPYPGPKVFP*
Molar extinction coefficient	
22920	

AKTA UV trace from IMAC purification (fractions X-Y collected):



AKTA UV trace from SEC purification (fractions 21-34 collected):



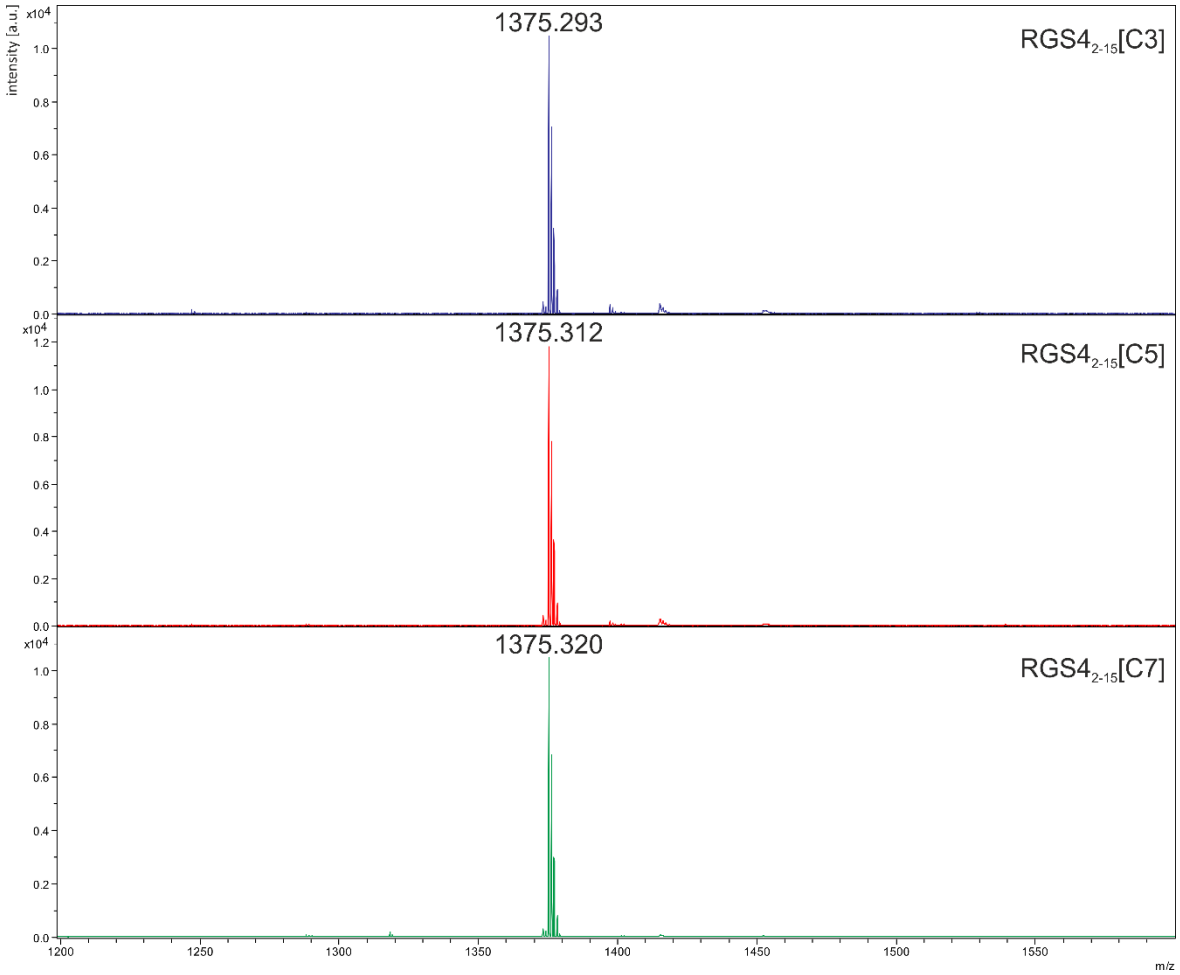
Fractions analysed by SDS-PAGE gels (Figure 2.1) are shown in red.

Appendix 2: Peptides purchased for biochemical experiments

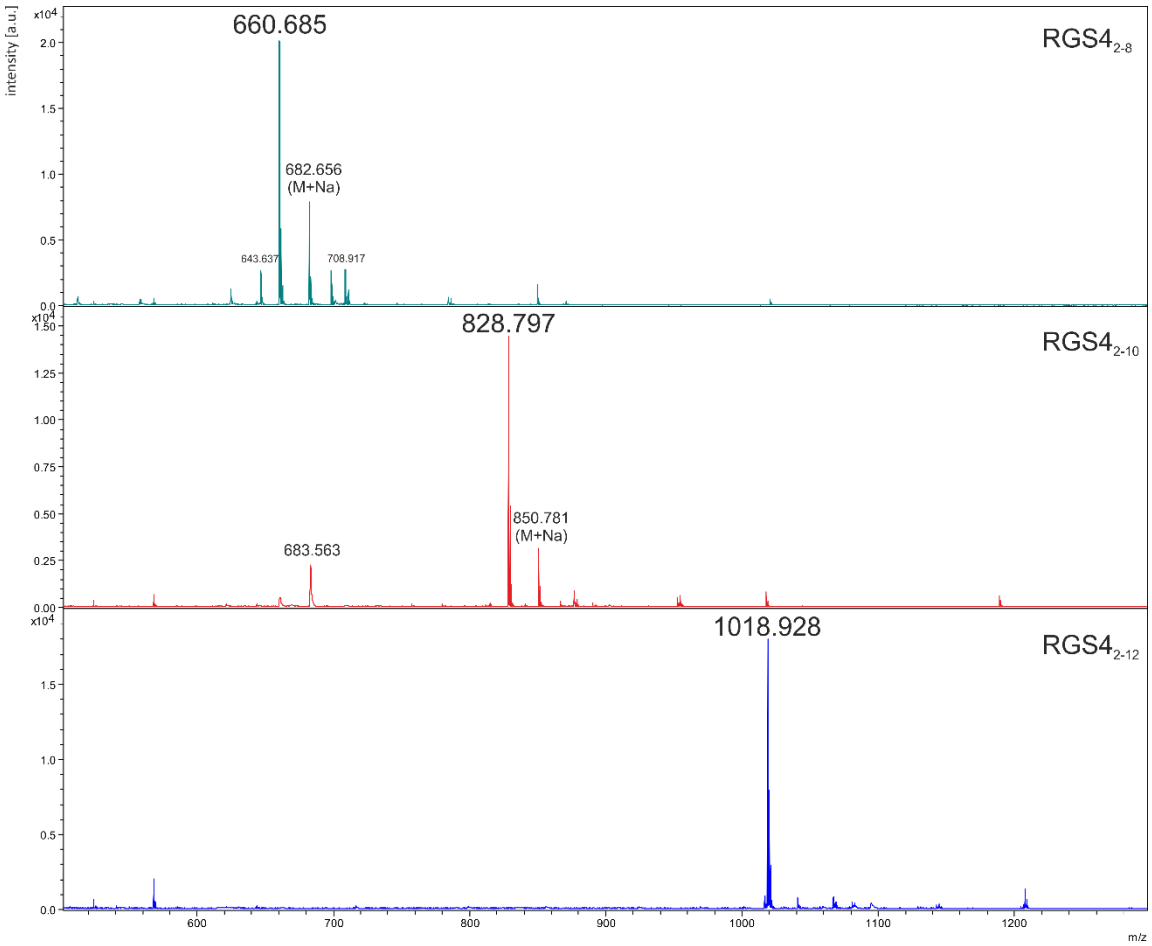
Peptide Name	Sequence	MW	Purchased from
AAAS	CSLGLFPPPPPRGQ-NH_2	1465.74	GLBiochem
Ac-AAAS	Ac-CSLGLFPPPPPRGQ-NH_2	1507.78	GLBiochem
Ac-JUNB	Ac-CTKMEQPFYHDDSY-NH_2	1805.96	GLBiochem
Ac-MTPN	Ac-CDKEFMWALKNGDL-NH_2	1711.97	GLBiochem
ADAMTS17	CDGALLPPLVLPVL-NH_2	1419.79	GLBiochem
ALS2CL	CNPEEAALLRLEEV-NH_2	1585.80	GLBiochem
ANKRD29	CRMSFKKETPLANA-NH_2	1595.90	GLBiochem
ARHGAP22	CVPPHRKDGSIPCP-NH_2	1505.78	GLBiochem
ARHGAP45	CICGTAHPVLDEGP-NH_2	1411.61	GLBiochem
ARL16	CLLLGATGVGKTLL-NH_2	1358.71	GenScript
AS1	CGELAMKATTKAWR-NH2	1564.87	GLBiochem
AS10	CDSLAMPASEESLS-NH2	1438.58	GLBiochem
AS11	CTELAMPASEESLS-NH2	1466.63	GLBiochem
AS12	CTSLAMPASEESLS-NH2	1424.59	GLBiochem
AS2	CGSLAMKATTKAWR-NH2	1522.83	GLBiochem
AS3	CDELAMKATTKAWR-NH2	1622.91	GLBiochem
AS4	CDSLAMKATTKAWR-NH2	1580.87	GLBiochem
AS5	CTELAMKATTKAWR-NH2	1608.92	GLBiochem
AS6	CTSLAMKATTKAWR-NH2	1566.88	GLBiochem
AS7	CGELAMPASEESLS-NH2	1422.58	GLBiochem
AS8	CGSLAMPASEESLS-NH2	1380.54	GLBiochem
AS9	CDELAMPASEESLS-NH2	1480.61	GLBiochem
ASNS	CGIWALFGSDDCLS-NH_2	1486.68	GenScript
CALHM4	CPTLNNIVSSLQRN-NH_2	1558.78	GLBiochem
NT5E (CD73)	CPRAARAPATLLLA-NH_2	1423.75	Francis Crick Institute
CDH2	CRIAGALRTLLPLL-NH_2	1509.92	GLBiochem
CDX1	CWTRIRPGTPSSPG-NH_2	1514.73	GLBiochem
CHURC1	CGDCVEKEYPNRGN-NH_2	1583.71	GLBiochem
CPSF4	CPFRHISGEKTVVC-NH_2	1575.87	GLBiochem
DHX30	CVNPTPGGTISRAS-NH_2	1359.52	GLBiochem
DMAC2CL	CCAVSEQRLTCADQ-NH_2	1526.72	GLBiochem
DNMBP	CETQKARKEREVEE-NH_2	1734.91	GLBiochem
DOCK2	CVCAEDGKTLPVAV-NH_2	1404.66	GLBiochem
DOCK8	CGEDASNAMPVRRE-NH_2	1534.69	GLBiochem
DYNLL1	CDRKAVIKNADMSE-NH_2	1579.81	GenScript
EFCAB6	CKMAIIPDWLRSHP-NH_2	1667.03	GenScript
ETNPPL	CELYSKRDTLGLRK-NH_2	1681.98	GLBiochem
FAM241A	CSAGELLRGGDGGE-NH_2	1319.41	GLBiochem
FGFR2	CLNKTKQLFGVAKR-NH_2	1604.91	GLBiochem
GCOM2	CLLPRGFEPQAPED-NH_2	1570.74	GLBiochem
GFPT2	CGIFAYMNYRVPRT-NH_2	1689.81	GenScript
GPR22	CFSPILEINMQSES-NH_2	1596.71	GLBiochem
GPX1	CAARLAAAAAAAQS-NH_2	1244.63	GenScript
iASPP-SV	CWLYALEVVERECW-NH_2	1797.82	GLBiochem
IER3/IEX1	CHSRSCHPTMTILQ-NH_2	1612.72	GLBiochem
IL12A NKSF1	CPARSLLLVATLVL-NH_2	1467.88	GLBiochem
IL32	CFPKVLSDDMKKLK-NH_2	1650.88	GLBiochem & GenScript
IL6	CESSKEALAENNLN-NH_2	1520.67	GLBiochem
INPPL1	CTRIAFCLMEKISW-NH_2	1699.82	GLBiochem
JUNB	CTKMEQPFYHDDSY-NH_2	1762.69	GLBiochem
LIPI	CWVRSDNKRPCLEF-NH_2	1751.82	GLBiochem
LYPLA2	CGNTMSVPLLTDAA-NH_2	1391.64	GLBiochem

Peptide Name	Sequence	MW	Purchased from
MTPN	CDKEFMWALKNGDL-NH_2	1668.76	GLBiochem
MYL6	CDFTEDQTAEFKEA-NH_2	1632.72	GLBiochem
NAAA	CDFMNLSLADCLLV-NH_2	1555.70	GLBiochem
NADSYN1	CCQVCEAVRSGNEE-NH_2	1525.59	GLBiochem
NBPF14	CRDGYEPCVNEGMC-NH_2	1574.56	GLBiochem
NHTL1	CSPQESGMTALSAR-NH_2	1436.62	GLBiochem
NECAB2	CERAARLCRAGAHR-NH_2	1569.83	GLBiochem
NOTCH2NLA	CVTYHNGTGYCKCP-NH_2	1545.77	GLBiochem
OR6K3	CWTMPSPFTGSSTR-NH_2	1557.76	GLBiochem
ORC6	CLDLAASWMKCPLD-NH_2	1565.88	GenScript
OS10	CRGRRMKATTKAWR-NH_2	1720.10	GLBiochem
OS6	CKGRRMKATTKAWR-NH_2	1692.09	GLBiochem
OS7	CFGRRMKATTKAWR-NH_2	1711.09	GLBiochem
OS8	CRMRRMKATTKAWR-NH_2	1794.24	GLBiochem
OS9	CRPRRMKATTKAWR-NH_2	1760.17	GLBiochem
P2RX6	CPQLAGAGSMGSPG-NH_2	1232.39	GLBiochem
PIFO	CFSRADAADNYPFG-NH_2	1533.64	GLBiochem
PPP3R1	CSHFDADEIKRLGK-NH_2	1618.83	GLBiochem
PRMT3	CSLASGATGGRGAV-NH_2	1205.36	GLBiochem
RAB37	CCKVMLLGDTGVGK-NH_2	1423.77	GLBiochem
RASA1	CLVKSWFLSIRSRW-NH_2	1781.16	GLBiochem
RGS16	CRTLAAFPTTCLER-NH_2	1580.89	GLBiochem
RGS4	CKGLAGLPASCLRS-NH_2	1375.67	GLBiochem
RGS5	CKGLAALPHSCLER-NH_2	1497.80	GLBiochem & GenScript
RIMKLA	CSQLWFLTDRRIRE-NH_2	1823.11	GLBiochem
RIPK1	CVPLPPSRSNSEQP-NH_2	1509.7	GLBiochem
RPL4	CRGGRMFAPTKTWR-NH_2	1666.99	GenScript
RTP2	CTSLTTCEWKKVFY-NH_2	1709.01	GLBiochem
SEC24D	CPFMQFIEGGRRYQ-NH_2	1732.01	GLBiochem
SHH-C	CFPGSATVHLEQGG-NH_2	1402.55	GLBiochem
SHH-N	CGPGRGFGKRRHPK-NH2	1552.83	GLBiochem
SIRT1	CLCSGRKTILEIYP-NH_2	1595.94	GLBiochem
SNRPG	CLFMDKKLSLKLNG-NH_2	1610.01	GLBiochem
SRR	CAQYCISFADVEKA-NH_2	1547.76	GenScript
STAC2	CKVSVHLWCSEEI-NH_2	1531.81	GenScript
SUSD6	CHGRIAPKSTSVFA-NH_2	1473.72	GLBiochem
TMEM168	CKSLRYCFSHCLYL-NH_2	1736.10	GLBiochem
TRAF1	CYRAPCSESQEELA-NH_2	1585.73	GLBiochem
TRIM36	CDLCKPPPQESTKS-NH_2	1532.75	GLBiochem
TRIM6-TRIM34	CGSERILQAGNILE-NH_2	1502.71	GLBiochem
USP27X	CKDYVYDKDIEQIA-NH_2	1702.90	GLBiochem
VDAC3	CNTPTYCDLGKAAK-NH_2	1484.71	GenScript
ZFAT	CKCCNLFSPNQSEL-NH_2	1585.83	GLBiochem
ZIP12	CFRTKLSVSWVPLF-NH_2	1683.05	GLBiochem
At RAP2.12	CGGAIISDFIPPPR-NH_2	1442.7	GLBiochem
At RAP 2.3	CGGAIISDYAPLVT-NH_2	1379.6	GLBiochem
At HRE1	CGGAVISDYIAPEK-NH_2	1422.6	GLBiochem
At HRE2	CGGAIISDFIWSKS-NH_2	1483.7	GLBiochem
MJH1	CKKFKIFMRKLQIC-NH_2	1786.3	GLBiochem
MJH2	CRSPDFERYVEKES-NH_2	1744.9	GLBiochem
MJH5	CSSKLKNLTQENIS-NH_2	1564.8	GLBiochem
MJH6	CHVGYALIELLENV-NH_2	1572.8	GLBiochem
Mm Tex19.1	CPPVSVRHGARGMS-NH_2	1452.7	GLBiochem

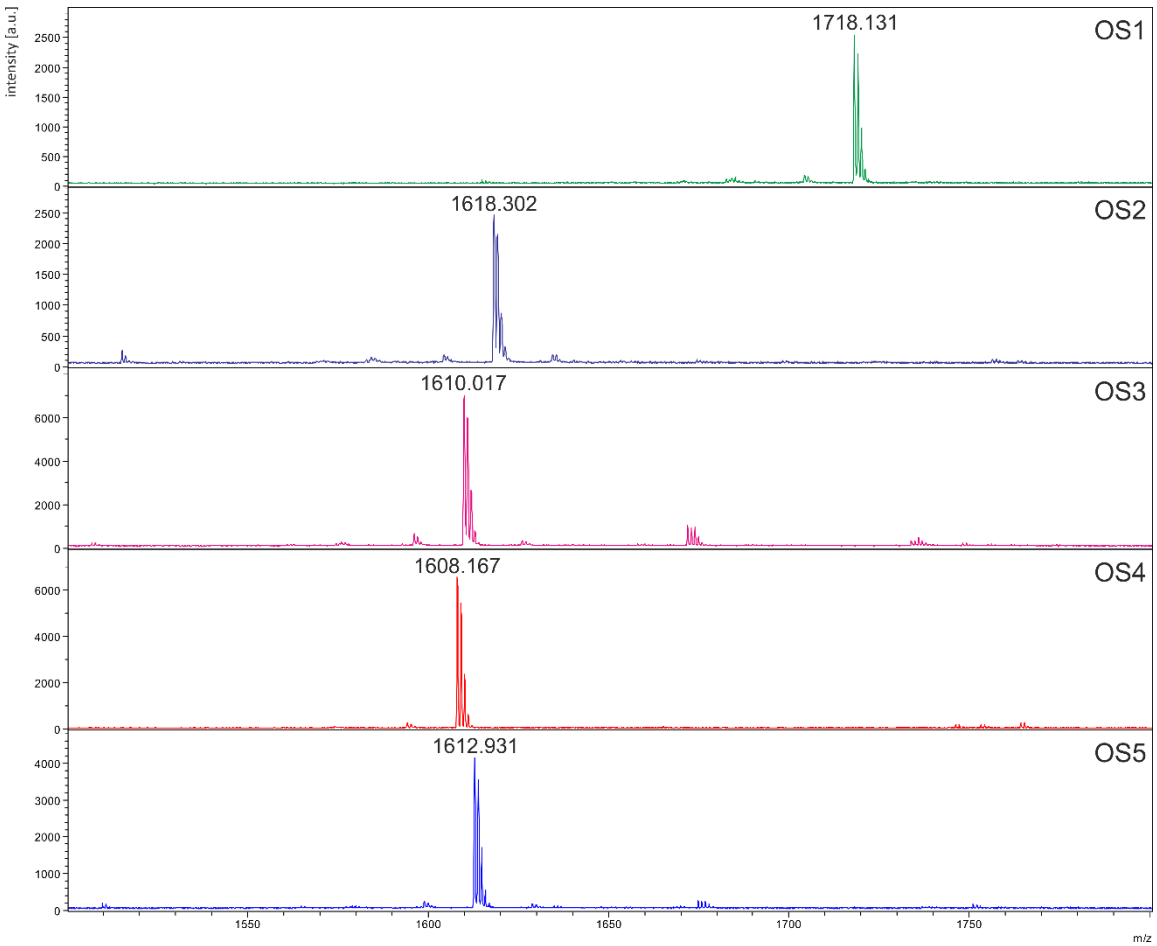
Appendix 3: MALDI spectra for peptides synthesised in this Thesis



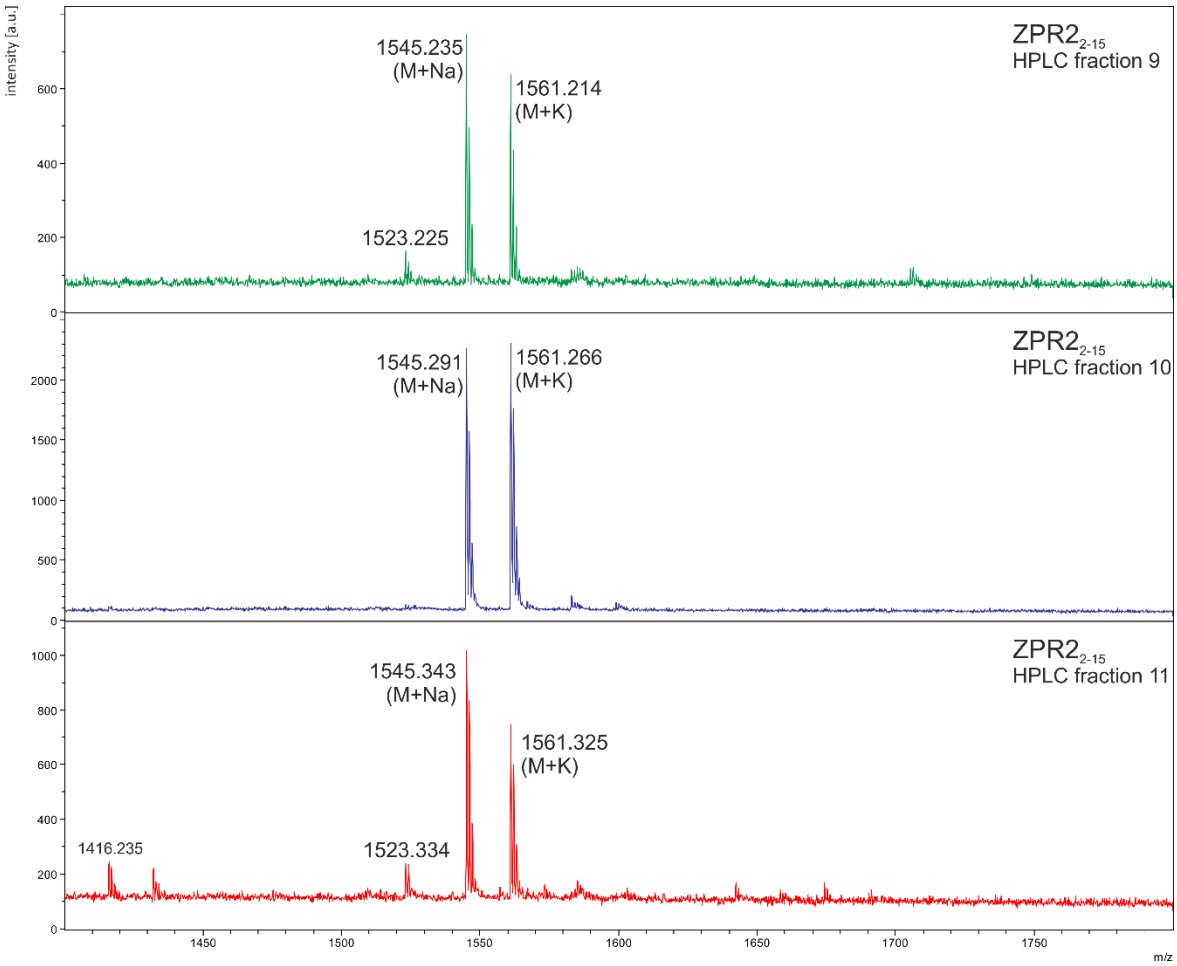
MALDI spectra of $RGS4_{2\text{-}15}$[C3,C5,C7] peptides after HPLC purification; MW: 1375.67



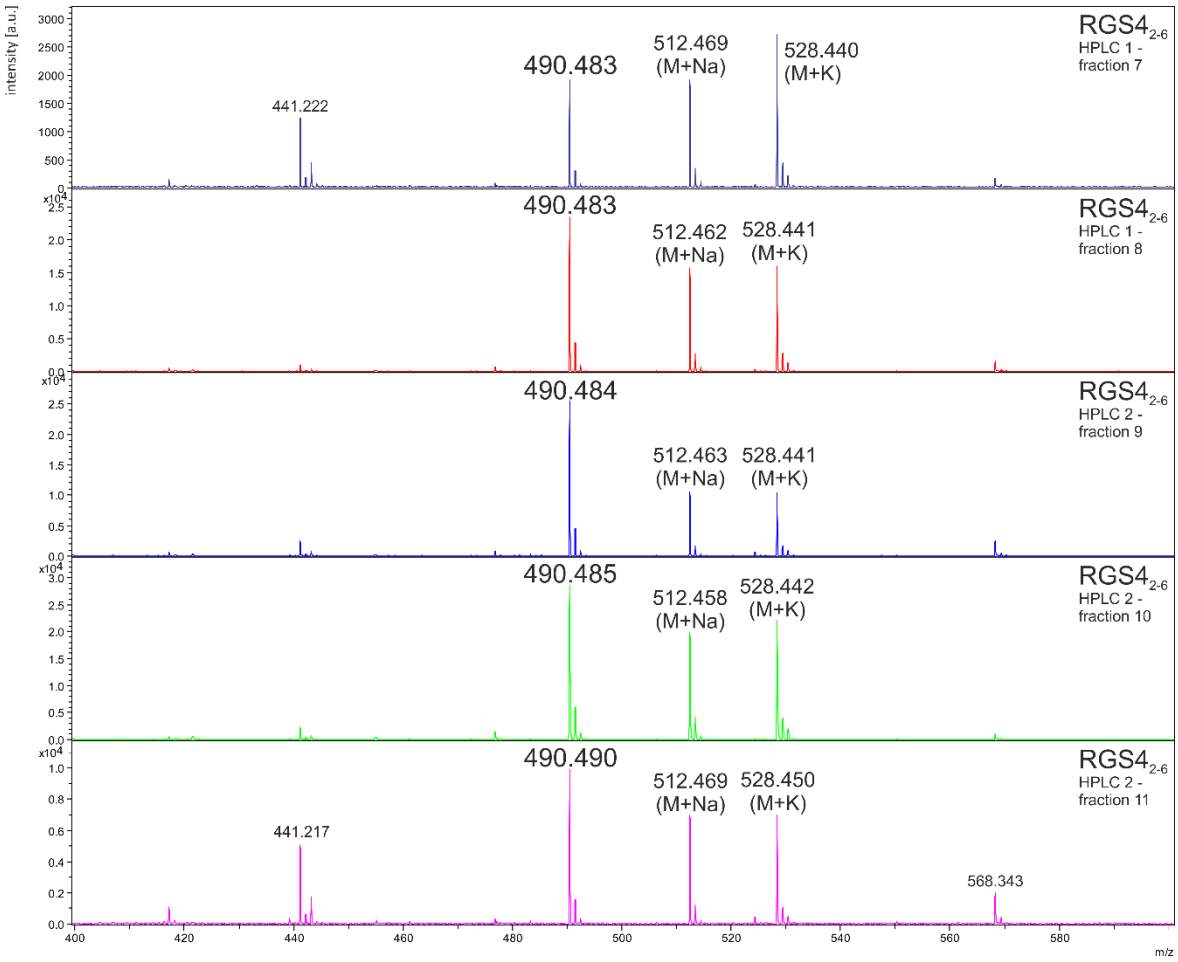
MALDI spectra of RGS4$_{2-8,10,12}$ peptides after HPLC purification; MW: 659.85; 828.04; 1018.26



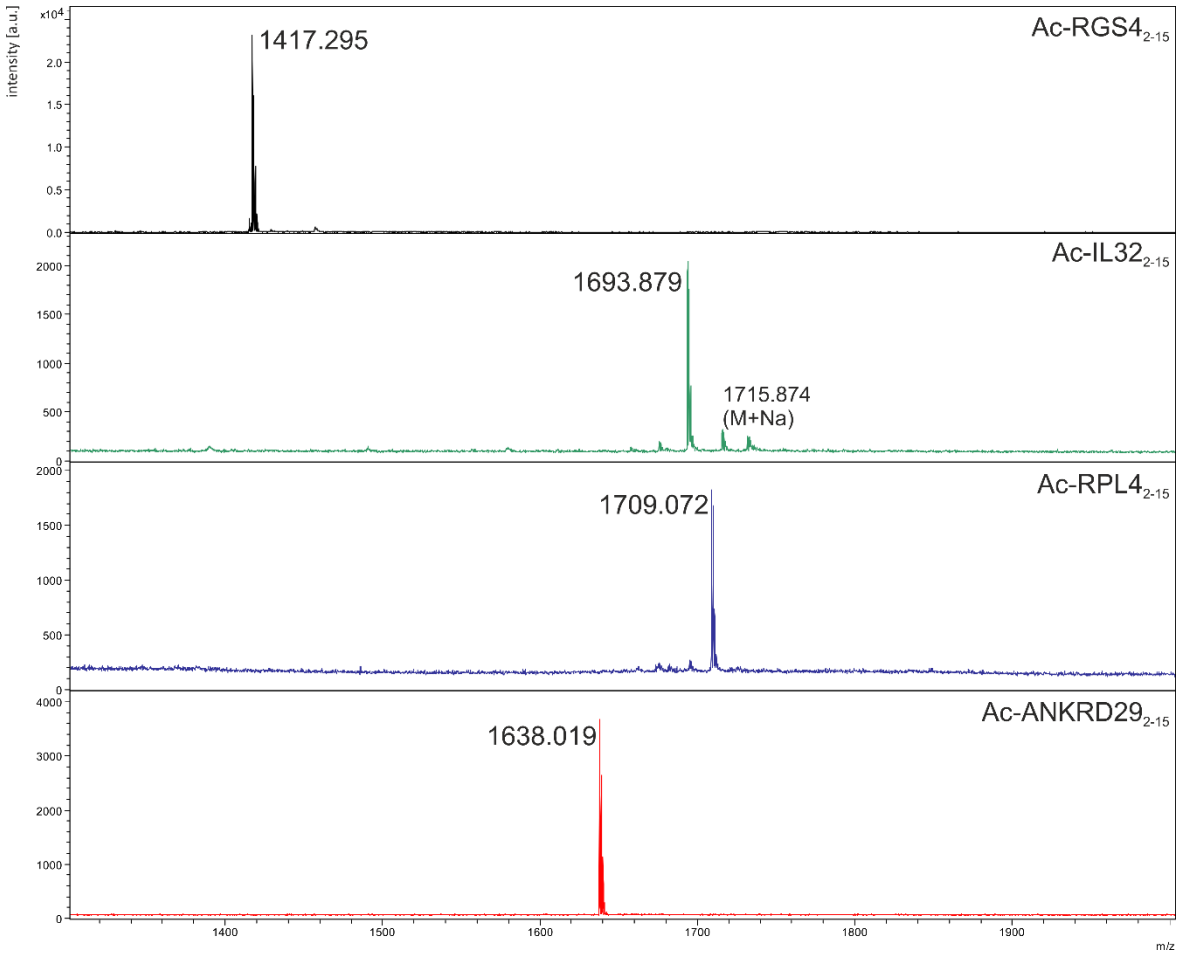
MALDI spectra of OS1-5 peptides after HPLC purification; MW: 1717.09; 1617.95; 1608.94; 1611.95; 1607.97



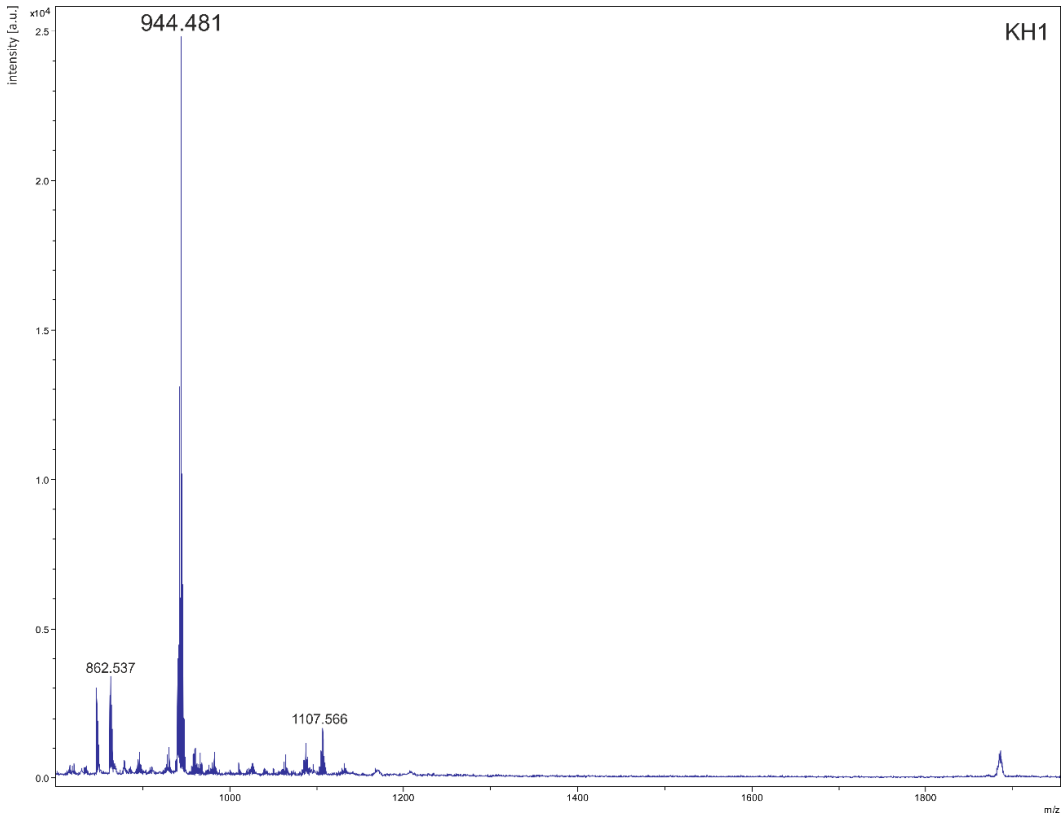
MALDI spectra of $ZPR2_{2\text{-}15}$ HPLC fractions, which were combined for use in assays; MW: 1522.64



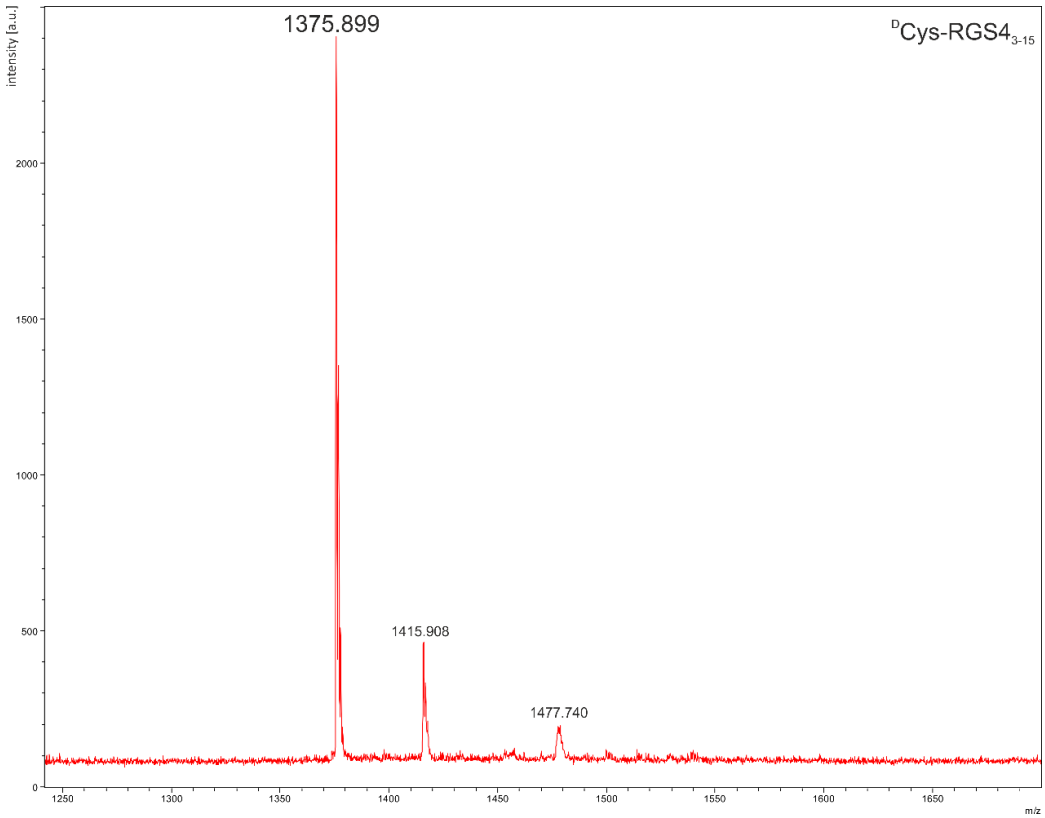
MALDI spectra of $RGS4_{2-6}$ HPLC fractions from two runs, which were combined for use in assays; MW: 489.64



MALDI spectra of Ac-RGS4$_{2\text{-}15}$, Ac-IL32$_{2\text{-}15}$, Ac-RPL4$_{2\text{-}15}$, Ac-ANKRD29$_{2\text{-}15}$; MW: 1416.73; 1693.10; 1708.04; 1636.95



MALDI spectra of KH1; MW: 943.06



MALDI spectra of ^{D}Cys-RGS4$_{3\text{-}15;}$ MW: 1375.67

Appendix 4: *Hs*ADO *in vitro* Nt-Cys peptide screening data

Accession	Gene	Peptide Sequence														Initial ADO Activity *in vitro* (% RGS4)
H3BM89	**RPL4**	C	R	G	G	R	M	F	A	P	T	K	T	W	R	133.13
P24001	**IL32**	C	F	P	K	V	L	S	D	D	M	K	K	L	K	110.79
O15539	**RGS5**	C	K	G	L	A	A	L	P	H	S	C	L	E	R	103.60
Q8N6D5	**ANKRD29**	C	R	M	S	F	K	K	E	T	P	L	A	N	A	103.04
P49798	**RGS4**	C	K	G	L	A	G	L	P	A	S	C	L	R	S	100.00
A0A087WZG8	**LIPI**	C	W	V	R	S	D	N	K	R	P	C	L	E	F	99.80
Q92537	**SUSD6**	C	H	G	R	I	A	P	K	S	T	S	V	F	A	97.75
Q8NFK1	**GJC3**	C	G	R	F	L	R	R	L	L	A	E	E	S	R	82.51
B7Z7B0	**CPSF4**	C	P	F	R	H	I	S	G	E	K	T	V	V	C	79.76
D6R9V6	**ARHGAP22**	C	V	P	P	H	R	K	D	G	S	I	P	C	P	56.56
P19022	**CDH2**	C	R	I	A	G	A	L	R	T	L	L	P	L	L	55.80
Q5THR3	**EFCAB6**	C	K	M	A	I	I	P	D	W	L	R	S	H	P	54.35
A0A087WTJ6	**CDX1**	C	W	T	R	I	R	P	G	T	P	S	S	P	G	54.29
P46695	**IER3/IEX1**	C	H	S	R	S	C	H	P	T	M	T	I	L	Q	48.36
B0QZ35	**SIRT1**	C	L	C	S	G	R	K	T	I	L	E	I	Y	P	44.76
S4R381	**FGFR2**	C	L	N	K	T	K	Q	L	F	G	V	A	K	R	43.84
F5H1R4	**ARHGAP45**	C	I	C	G	T	A	H	P	V	L	D	E	G	P	43.50
P07203	**GPX1**	C	A	A	R	L	A	A	A	A	A	A	A	Q	S	42.05
E9PBN4	**ZFAT**	C	K	C	C	N	L	F	S	P	N	Q	S	E	L	37.01
B7Z3L0	**RAB37**	C	C	K	V	M	L	L	G	D	T	G	V	G	K	33.39
Q99766	**DMAC2L**	C	C	A	V	S	E	Q	R	L	T	C	A	D	Q	29.50
F5H013	**SNRPG**	C	L	F	M	D	K	K	L	S	L	K	L	N	G	28.55
Q5QGT7	**RTP2**	C	T	S	L	T	T	C	E	W	K	K	V	F	Y	27.94
Q9Y277	**VDAC3**	C	N	T	P	T	Y	C	D	L	G	K	A	A	K	27.79
Q8NGY3	**OR6K3**	C	W	T	M	P	S	P	F	T	G	S	S	T	R	26.28
Q8TCI5	**PIFO**	C	F	S	R	A	D	A	A	D	N	Y	P	F	G	25.41
Q9BZD3	**GCOM2**	C	L	L	P	R	G	F	E	P	Q	A	P	E	D	24.51
Q13077	**TRAF1**	C	Y	R	A	P	C	S	E	S	Q	E	E	L	A	24.36
O15492	**RGS16**	C	R	T	L	A	A	F	P	T	T	C	L	E	R	22.61
A0A0B4J216	**NADSYN1**	C	C	Q	V	C	E	A	V	R	S	G	N	E	E	21.98
Q8IXN7	**RIMKLA**	C	S	Q	L	W	F	L	T	D	R	R	I	R	E	21.44
Q9NRG9	**AAAS**	C	S	L	G	L	F	P	P	P	P	P	R	G	Q	19.24
Q504Y0	**ZIP12**	C	F	R	T	K	L	S	V	S	W	V	P	L	F	18.14
Q15465	**SHHC**	C	F	P	G	S	A	T	V	H	L	E	Q	G	G	18.11
Q7Z3S9	**NOTCH2NLA**	C	V	T	Y	H	N	G	T	G	Y	C	K	C	P	17.93
Q9H0V1	**TMEM168**	C	K	S	L	R	Y	C	F	S	H	C	L	Y	L	16.20
F6U1T9	**PPP3R1**	C	S	H	F	D	A	D	E	I	K	R	L	G	K	15.34
H7BXY3	**DHX30**	C	V	N	P	T	P	G	G	T	I	S	R	A	S	15.30
Q8TBG4	**ETNPPL**	C	E	L	Y	S	K	R	D	T	L	G	L	R	K	13.77
P63167	**DYNLL1**	C	D	R	K	A	V	I	K	N	A	D	M	S	E	11.31
E9PDM8	**SEC24D**	C	P	F	M	Q	F	I	E	G	G	R	R	Y	Q	10.45
K7EN03	**iASPP-SV**	C	W	L	Y	A	L	E	V	V	E	R	E	C	W	9.84
Q9GZT4	**SRR**	C	A	Q	Y	C	I	S	F	A	D	V	E	K	A	8.18
O94808	**GFPT2**	C	G	I	F	A	Y	M	N	Y	R	V	P	R	T	7.78
B4E3H0	**ARL16**	C	L	L	L	G	A	T	G	V	G	K	T	L	L	7.17
P21589	**NT5E**	C	P	R	A	A	R	A	P	A	T	L	L	L	A	6.62
E9PBG3	**TRIM36**	C	D	L	C	K	P	P	P	Q	E	S	T	K	S	6.21
P17275	**JUNB**	C	T	K	M	E	Q	P	F	Y	H	D	D	S	Y	4.92
E9PGC0	**RASA1**	C	L	V	K	S	W	F	L	S	I	R	S	R	W	4.84
H3BT22	**ORC6**	C	L	D	L	A	A	S	W	M	K	C	P	L	D	4.54
Q5JW98	**CALHM4**	C	P	T	L	N	N	I	V	S	S	L	Q	R	N	4.47
Q6ZMT1	**STAC2**	C	K	V	S	V	H	L	W	C	S	E	E	I		4.31
A6NNY8	**USP27X**	C	K	D	Y	V	Y	D	K	D	I	E	Q	I	A	4.12
Q92608-2	**DOCK2**	C	V	C	A	E	D	G	K	T	L	P	V	A	V	3.85
P58546	**MTPN**	C	D	K	E	F	M	W	A	L	K	N	G	D	L	3.72
Q8TE56	**ADAMTS17**	C	D	G	A	L	L	P	P	L	V	L	P	V	L	3.67
B4DVL2	**NAAA**	C	D	F	M	N	L	S	L	A	D	C	L	L	V	3.60
Q60I27	**ALS2CL**	C	N	P	E	E	A	A	L	L	R	L	E	E	V	3.30
B4DNV3	**IL6**	C	E	S	S	K	E	A	L	A	E	N	N	L	N	2.82
O15547	**P2RX6**	C	P	Q	L	A	G	A	G	S	M	G	S	P	G	2.74
B2RNG4	**TRIM6-TRIM34**	C	G	S	E	R	I	L	Q	A	G	N	I	L	E	2.64
AAA96658	**INPPL1**	C	T	R	I	A	F	C	L	M	E	K	I	S	W	1.80
A0A0B4J2B3	**NBPF14**	C	R	D	G	Y	E	P	C	V	N	E	G	M	C	1.07
O95372	**LYPLA2**	C	G	N	T	M	S	V	P	L	L	T	D	A	A	0.96
P29459	**IL12A NKSF1**	C	P	A	R	S	L	L	L	V	A	T	L	V	L	0.34
Q8WUH1	**ASNS**	C	G	I	W	A	L	F	G	S	D	D	C	L	S	0.00
P08243	**CHURC1**	C	G	D	C	V	E	K	E	Y	P	N	R	G	N	0.00
KAI4006482	**DOCK8**	C	G	E	D	A	S	N	A	M	P	V	R	R	E	0.00

Appendix 5: Statistical analysis of OS peptides 1 minute assay data

One-ANOVA followed by Tukey's Multiple Comparison Test

One-way analysis of variance

P value < 0.0001 P value summary ****

Are means significantly different? (P < 0.05) Yes

Number of groups 11 F 7.582 R square 0.6661

ANOVA Table	SS	df	MS
Treatment (between columns)	25457	10	2546
Residual (within columns)	12760	38	335.8
Total	38217	48	

Tukey's Multiple Comparison Test	Mean Diff.	q	Significant? P < 0.05?	Summary	95% CI of diff
OS2 vs OS3	-43.63	5.019	Yes	*	-85.72 to -1.534
OS3 vs OS6	52.05	5.26	Yes	*	4.129 to 99.98
OS5 vs OS7	-51.37	5.606	Yes	*	-95.73 to -6.997
OS6 vs OS9	-47.55	5.025	Yes	*	-93.37 to -1.723
OS6 vs OS10	-47.55	5.025	Yes	*	-93.37 to -1.723
OS1 vs OS8	-53.97	5.891	Yes	**	-98.34 to -9.603
OS2 vs OS8	-57.98	6.128	Yes	**	-103.8 to -12.16
OS4 vs OS7	-52.59	5.74	Yes	**	-96.96 to -8.226
OS6 vs OS8	-66.41	6.277	Yes	**	-117.6 to -15.18
RGS4 vs OS7	-52.16	6.236	Yes	**	-92.66 to -11.65
OS1 vs OS7	-70.99	8.488	Yes	***	-111.5 to -30.49
OS2 vs OS7	-75	8.629	Yes	***	-117.1 to -32.91
OS6 vs OS7	-83.43	8.43	Yes	***	-131.4 to -35.50
OS1 vs OS2	4.012	0.5114	No	ns	-33.98 to 42.01
OS1 vs OS3	-39.61	4.736	No	ns	-80.12 to 0.8899
OS1 vs OS4	-18.39	2.199	No	ns	-58.90 to 22.11
OS1 vs OS5	-19.62	2.346	No	ns	-60.13 to 20.88
OS1 vs OS6	12.44	1.358	No	ns	-31.93 to 56.81
OS1 vs OS9	-35.11	4.474	No	ns	-73.10 to 2.888
OS1 vs OS10	-35.11	4.474	No	ns	-73.10 to 2.888
OS2 vs OS4	-22.41	2.578	No	ns	-64.50 to 19.69
OS2 vs OS5	-23.64	2.719	No	ns	-65.73 to 18.46
OS2 vs OS6	8.427	0.8906	No	ns	-37.40 to 54.25
OS2 vs OS9	-39.12	4.774	No	ns	-78.80 to 0.5654
OS2 vs OS10	-39.12	4.774	No	ns	-78.80 to 0.5654
OS3 vs OS4	21.22	2.316	No	ns	-23.15 to 65.59
OS3 vs OS5	19.99	2.182	No	ns	-24.38 to 64.36
OS3 vs OS7	-31.38	3.425	No	ns	-75.74 to 12.99
OS3 vs OS8	-14.36	1.451	No	ns	-62.28 to 33.56
OS3 vs OS9	4.506	0.5184	No	ns	-37.59 to 46.60
OS3 vs OS10	4.506	0.5184	No	ns	-37.59 to 46.60
OS4 vs OS5	-1.229	0.1342	No	ns	-45.60 to 43.14
OS4 vs OS6	30.83	3.116	No	ns	-17.09 to 78.76

Tukey's Multiple Comparison Test	Mean Diff.	q	Significant? P < 0.05?	Summary	95% CI of diff
OS4 vs OS8	-35.58	3.595	No	ns	-83.50 to 12.35
OS4 vs OS9	-16.71	1.923	No	ns	-58.80 to 25.38
OS4 vs OS10	-16.71	1.923	No	ns	-58.80 to 25.38
OS5 vs OS6	32.06	3.24	No	ns	-15.86 to 79.99
OS5 vs OS8	-34.35	3.471	No	ns	-82.27 to 13.58
OS5 vs OS9	-15.48	1.781	No	ns	-57.58 to 26.61
OS5 vs OS10	-15.48	1.781	No	ns	-57.58 to 26.61
OS7 vs OS8	17.02	1.72	No	ns	-30.91 to 64.94
OS7 vs OS9	35.88	4.128	No	ns	-6.210 to 77.97
OS7 vs OS10	35.88	4.128	No	ns	-6.210 to 77.97
OS8 vs OS9	18.87	1.994	No	ns	-26.96 to 64.69
OS8 vs OS10	18.87	1.994	No	ns	-26.96 to 64.69
OS9 vs OS10	0	0	No	ns	-39.68 to 39.68
RGS4 vs OS1	18.83	2.518	No	ns	-17.39 to 55.06
RGS4 vs OS2	22.85	2.912	No	ns	-15.15 to 60.84
RGS4 vs OS3	-20.78	2.484	No	ns	-61.28 to 19.72
RGS4 vs OS4	0.4395	0.05255	No	ns	-40.06 to 40.94
RGS4 vs OS5	-0.7896	0.09441	No	ns	-41.29 to 39.71
RGS4 vs OS6	31.27	3.413	No	ns	-13.10 to 75.64
RGS4 vs OS8	-35.14	3.835	No	ns	-79.51 to 9.230
RGS4 vs OS9	-16.27	2.074	No	ns	-54.27 to 21.72
RGS4 vs OS10	-16.27	2.074	No	ns	-54.27 to 21.72

Appendix 6: Statistical analysis of AS peptide 1 minute assay data

One-Way ANOVA followed by Tukey's Multiple Comparison Test

One-way analysis of variance:

P value< 0.0001 P value summary ****

Are means significantly different? (P < 0.05)Yes

Number of groups 14 F 827.3 R square 0.9963

ANOVA Table	SS	df	MS
Treatment (between columns)	71433	13	5495
Residual (within columns)	265.7	40	6.642
Total	71699	53	

Tukey's Multiple Comparison Test	Mean Diff.	q	Significant? P < 0.05?	Summary	95% CI of diff
AS1 vs AS2	-28.68	19.27	Yes	***	-36.19 to -21.18
AS1 vs AS5	-10.76	7.229	Yes	***	-18.26 to -3.251
AS1 vs AS6	-27.62	18.56	Yes	***	-35.13 to -20.12
AS2 vs AS10	31.56	21.21	Yes	***	24.05 to 39.06
AS2 vs AS11	31.56	21.21	Yes	***	24.05 to 39.06
AS2 vs AS12	28.66	19.26	Yes	***	21.16 to 36.17
AS2 vs AS3	28.89	19.41	Yes	***	21.38 to 36.39
AS2 vs AS4	28.39	19.08	Yes	***	20.88 to 35.89
AS2 vs AS5	17.92	12.05	Yes	***	10.42 to 25.43
AS2 vs AS7	31.56	21.21	Yes	***	24.05 to 39.06
AS2 vs AS8	31.56	21.21	Yes	***	24.05 to 39.06
AS2 vs AS9	31.56	21.21	Yes	***	24.05 to 39.06
AS3 vs AS5	-10.96	7.368	Yes	***	-18.47 to -3.458
AS3 vs AS6	-27.83	18.7	Yes	***	-35.33 to -20.32
AS4 vs AS5	-10.46	7.031	Yes	***	-17.97 to -2.957
AS4 vs AS6	-27.33	18.37	Yes	***	-34.83 to -19.82
AS5 vs AS10	13.63	9.161	Yes	***	6.127 to 21.14
AS5 vs AS11	13.63	9.161	Yes	***	6.127 to 21.14
AS5 vs AS12	10.74	7.217	Yes	***	3.234 to 18.24
AS5 vs AS6	-16.87	11.33	Yes	***	-24.37 to -9.360
AS5 vs AS7	13.63	9.161	Yes	***	6.127 to 21.14
AS5 vs AS8	13.63	9.161	Yes	***	6.127 to 21.14
AS5 vs AS9	13.63	9.161	Yes	***	6.127 to 21.14
AS6 vs AS10	30.5	20.5	Yes	***	22.99 to 38.00
AS6 vs AS11	30.5	20.5	Yes	***	22.99 to 38.00
AS6 vs AS12	27.6	18.55	Yes	***	20.10 to 35.11
AS6 vs AS7	30.5	20.5	Yes	***	22.99 to 38.00
AS6 vs AS8	30.5	20.5	Yes	***	22.99 to 38.00
AS6 vs AS9	30.5	20.5	Yes	***	22.99 to 38.00
ASNS vs AS2	-31.55	25.97	Yes	***	-37.67 to -25.42
ASNS vs AS5	-13.62	11.21	Yes	***	-19.75 to -7.494
ASNS vs AS6	-30.49	25.09	Yes	***	-36.62 to -24.36
RGS4 vs AS1	97.12	79.94	Yes	***	91.00 to 103.3
RGS4 vs AS10	100	82.31	Yes	***	93.87 to 106.1

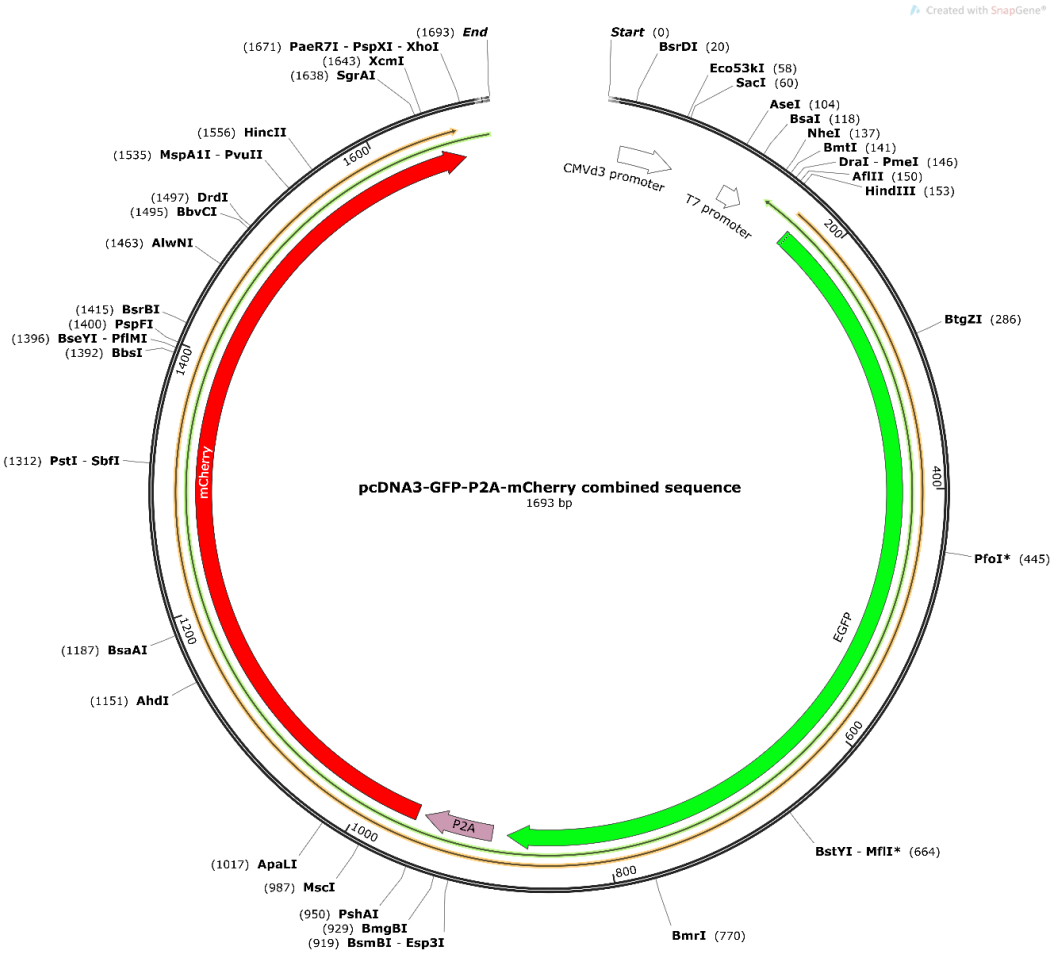
Tukey's Multiple Comparison Test	Mean Diff.	q	Significant? P < 0.05?	Summary	95% CI of diff
RGS4 vs AS11	100	82.31	Yes	***	93.87 to 106.1
RGS4 vs AS12	97.11	79.93	Yes	***	90.98 to 103.2
RGS4 vs AS2	68.44	56.34	Yes	***	62.32 to 74.57
RGS4 vs AS3	97.33	80.11	Yes	***	91.20 to 103.5
RGS4 vs AS4	96.83	79.7	Yes	***	90.70 to 103.0
RGS4 vs AS5	86.37	71.09	Yes	***	80.24 to 92.50
RGS4 vs AS6	69.5	57.21	Yes	***	63.37 to 75.63
RGS4 vs AS7	100	82.31	Yes	***	93.87 to 106.1
RGS4 vs AS8	100	82.31	Yes	***	93.87 to 106.1
RGS4 vs AS9	100	82.31	Yes	***	93.87 to 106.1
RGS4 vs ASNS	99.99	116.4	Yes	***	95.66 to 104.3
AS1 vs AS10	2.876	1.933	No	ns	-4.630 to 10.38
AS1 vs AS11	2.876	1.933	No	ns	-4.630 to 10.38
AS1 vs AS12	-0.01732	0.01164	No	ns	-7.523 to 7.488
AS1 vs AS3	0.2064	0.1387	No	ns	-7.299 to 7.712
AS1 vs AS4	-0.2946	0.198	No	ns	-7.800 to 7.211
AS1 vs AS7	2.876	1.933	No	ns	-4.630 to 10.38
AS1 vs AS8	2.876	1.933	No	ns	-4.630 to 10.38
AS1 vs AS9	2.876	1.933	No	ns	-4.630 to 10.38
AS10 vs AS11	0	0	No	ns	-7.505 to 7.505
AS10 vs AS12	-2.893	1.944	No	ns	-10.40 to 4.613
AS11 vs AS12	-2.893	1.944	No	ns	-10.40 to 4.613
AS2 vs AS6	1.059	0.7117	No	ns	-6.446 to 8.564
AS3 vs AS10	2.669	1.794	No	ns	-4.836 to 10.17
AS3 vs AS11	2.669	1.794	No	ns	-4.836 to 10.17
AS3 vs AS12	-0.2237	0.1504	No	ns	-7.729 to 7.282
AS3 vs AS4	-0.501	0.3367	No	ns	-8.006 to 7.004
AS3 vs AS7	2.669	1.794	No	ns	-4.836 to 10.17
AS3 vs AS8	2.669	1.794	No	ns	-4.836 to 10.17
AS3 vs AS9	2.669	1.794	No	ns	-4.836 to 10.17
AS4 vs AS10	3.17	2.13	No	ns	-4.335 to 10.68
AS4 vs AS11	3.17	2.13	No	ns	-4.335 to 10.68
AS4 vs AS12	0.2773	0.1863	No	ns	-7.228 to 7.783
AS4 vs AS7	3.17	2.13	No	ns	-4.335 to 10.68
AS4 vs AS8	3.17	2.13	No	ns	-4.335 to 10.68
AS4 vs AS9	3.17	2.13	No	ns	-4.335 to 10.68
AS7 vs AS10	0	0	No	ns	-7.505 to 7.505
AS7 vs AS11	0	0	No	ns	-7.505 to 7.505
AS7 vs AS12	-2.893	1.944	No	ns	-10.40 to 4.613
AS7 vs AS8	0	0	No	ns	-7.505 to 7.505
AS7 vs AS9	0	0	No	ns	-7.505 to 7.505
AS8 vs AS10	0	0	No	ns	-7.505 to 7.505
AS8 vs AS11	0	0	No	ns	-7.505 to 7.505
AS8 vs AS12	-2.893	1.944	No	ns	-10.40 to 4.613
AS8 vs AS9	0	0	No	ns	-7.505 to 7.505
AS9 vs AS10	0	0	No	ns	-7.505 to 7.505
AS9 vs AS11	0	0	No	ns	-7.505 to 7.505
AS9 vs AS12	-2.893	1.944	No	ns	-10.40 to 4.613
ASNS vs AS1	-2.866	2.359	No	ns	-8.994 to 3.263
ASNS vs AS10	0.01	0.008231	No	ns	-6.118 to 6.138
ASNS vs AS11	0.01	0.008231	No	ns	-6.118 to 6.138
ASNS vs AS12	-2.883	2.373	No	ns	-9.011 to 3.245
ASNS vs AS3	-2.659	2.189	No	ns	-8.787 to 3.469
ASNS vs AS4	-3.16	2.601	No	ns	-9.288 to 2.968
ASNS vs AS7	0.01	0.008231	No	ns	-6.118 to 6.138
ASNS vs AS8	0.01	0.008231	No	ns	-6.118 to 6.138
ASNS vs AS9	0.01	0.008231	No	ns	-6.118 to 6.138

Appendix 7: Primers

Primer Name	DNA Sequence
F_HindIII_RGS4[15]	ATTAAGCTTGCCACCATGTGCAAAGGCCTGGCAG GCCTGCCAGCAAGCTGCCTGAGGAGCGTGAGCA AGGGCGAGGAGCTGTT
F_HindIII_AAAS[15]	ATTAAGCTTGCCACCATGTGCAGCCTGGGCCTGT TCCCACCACCACCACCAAGGGGCCAAGTGAGCA AGGGCGAGGAGCTGTT
F_HindIII_ANKRD29[15]	ATTAAGCTTGCCACCATGTGCAGGATGAGCTTCA AGAAGGAGACCCCCCTGGCCAACGCCGTGAGCA AGGGCGAGGAGCTGTT
F_HindIII_ASNS[15]	ATTAAGCTTGCCACCATGTGCGGAATCTGGGCCC TGTTCGGCAGCGACGACTGCCTGAGCGTGAGCA AGGGCGAGGAGCTGTT
F_HindIII_At RAP 2.12[15]	ATTAAGCTTGCCACCATGTGCGGAGGAGCAATCA TCAGCGACTTCATCCCCCCCCCCAGGGTGAGCAA GGGCGAGGAGCTGTT
F_HindIII_GFPT2[15]	ATTAAGCTTGCCACCATGTGCGGCATCTTCGCCT ACATGAACTACAGGGTGCCCAGGACCGTGAGCA AGGGCGAGGAGCTGTT
F_HindIII_GPX1[15]	ATTAAGCTTGCCACCATGTGCGCAGCAAGGCTGG CAGCAGCAGCAGCAGCAGCACAGAGCGTGAGCA AGGGCGAGGAGCTGTT
F_HindIII_JUNB[15]	ATTAAGCTTGCCACCATGTGCACCAAGATGGAGC AGCCCTTCTACCACGACGACAGCTACGTGAGCAA GGGCGAGGAGCTGTT
F_HindIII_LYPLA2[15]	ATTAAGCTTGCCACCATGTGCGGCAACACCATGA GCGTGCCCCTGCTGACCGACGCAGCAGTGAGCA AGGGCGAGGAGCTGTT
F_HindIII_MTPN[15]	ATTAAGCTTGCCACCATGTGCGACAAGGAGTTCA TGTGGGCCCTGAAGAACGGCGACCTGGTGAGCA AGGGCGAGGAGCTGTT
F_HindIII_OS7[15]	ATTAAGCTTGCCACCATGTGCTTCGGCAGGAGGA TGAAGGCCACCACCAAGGCCTGGAGGGTGAGCA AGGGCGAGGAGCTGTT
F_HindIII_STAC2[14]	ATTAAGCTTGCCACCATGTGCAAGGTGAGCGTGC ACCTGTGGTGCAGCGAGGAGATCGTGAGCAAGG GCGAGGAGCTGTT
F_HindIII_SUSD6[15]	ATTAAGCTTGCCACCATGTGCCACGGCAGGATCG CACCCAAGAGCACCAGCGTGTTCGCAGTGAGCA AGGGCGAGGAGCTGTT
F_HindIII_TEX19[15]	ATTAAGCTTGCCACCATGTGCCCACCAGTGAGCG TGAGGCACGGAGCAAGGGGAATGAGCGTGAGCA AGGGCGAGGAGCTGTT
F_HindIII_GJC3[15]	ATTAAGCTTGCCACCATGTGCGGAAGGTTCCTGA GGAGGCTGCTGGCCGAGGAGAGCAGGGTGAGCA AGGGCGAGGAGCTGTT
R_GFP_XbaI	GCCTCTAGAGCCCTTGTACAGCTCGTCCA
F_HindIII_RGS4[15]_NoGFP	ATTAAGCTTGCCACCATGTGCAAAGGCCTGGCAG GCCTGCCAGCAAGCTGCCTGAGGAGC
F_HindIII_ANKRD29[15]_NoGFP	ATTAAGCTTGCCACCATGTGCAGGATGAGCTTCA AGAAGGAGACCCCCCTGGCCAACGCC
F_HindIII_ASNS[15]_NoGFP	ATTAAGCTTGCCACCATGTGCGGAATCTGGGCCC TGTTCGGCAGCGACGACTGCCTGAGC
F_HindIII_At RAP 2.12[15]_NoGFP	ATTAAGCTTGCCACCATGTGCGGAGGAGCAATCA TCAGCGACTTCATCCCCCCCCCCAGG

Primer Name	DNA Sequence
F_HindIII_GPX1[15]_NoGFP	ATTAAGCTTGCCACCATGTGCGCAGCAAGGCTGGCAGCAGCAGCAGCAGCAGCACAGAGC
F_HindIII_OS7[15]_NoGFP	ATTAAGCTTGCCACCATGTGCTTCGGCAGGAGGATGAAGGCCACCACCAAGGCCTGGAGG
F_HindIII_SUSD6[15]_NoGFP	ATTAAGCTTGCCACCATGTGCCACGGCAGGATCGCACCCAAGAGCACCAGCGTGTTCGCA
F_HindIII_TEX19[15]_NoGFP	ATTAAGCTTGCCACCATGTGCCCACCAGTGAGCGTGAGGCACGGAGCAAGGGGAATGAGC
F_HindIII_GJC3[15]_NoGFP	ATTAAGCTTGCCACCATGTGCGGAAGGTTCCTGAGGAGGCTGCTGGCCGAGGAGAGCAGG

Appendix 8: pcDNA3-GFP-P2A-mCherry plasmid map and sequence



>pcDNA3-GFP-P2A-mCherry sequence (1693 bp)

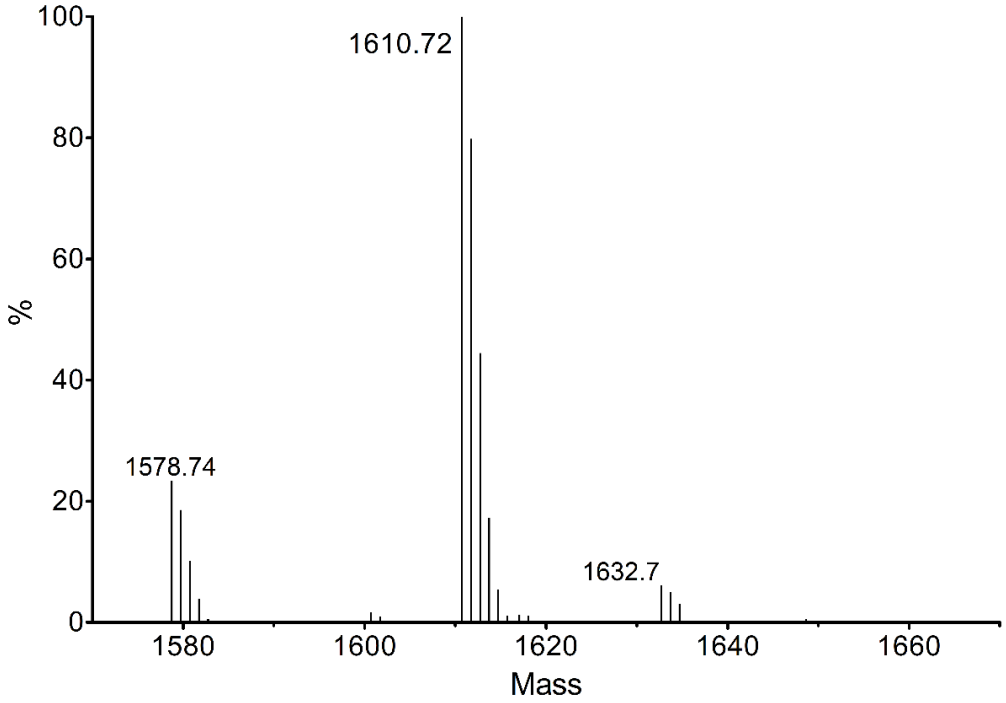
NNNNNNNNTGACGCAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAA
CTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAG
CGTTTAAACTTAAGCTTGCGGCCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC
CTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGA
CCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAG
TCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCC
GCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGA
GGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGAC
AAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTC
GCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGA
GCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGAC
CGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGCTCTAGAGGCGCTACTAACTTCAGCCTG
CTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCTGGGTCCATGGTGAGCAAGGGCGAGGAGGAT
AACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGT
TCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCA
AGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTG
AAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGAT
GAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTAC
AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGG
AGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCT
GAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCC
CGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAG
TACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAAGCCCTCGAGTCTA
GANNNNNCNNNNN

Appendix 9: Combined NatA and *Hs*ADO *in vitro* activity data

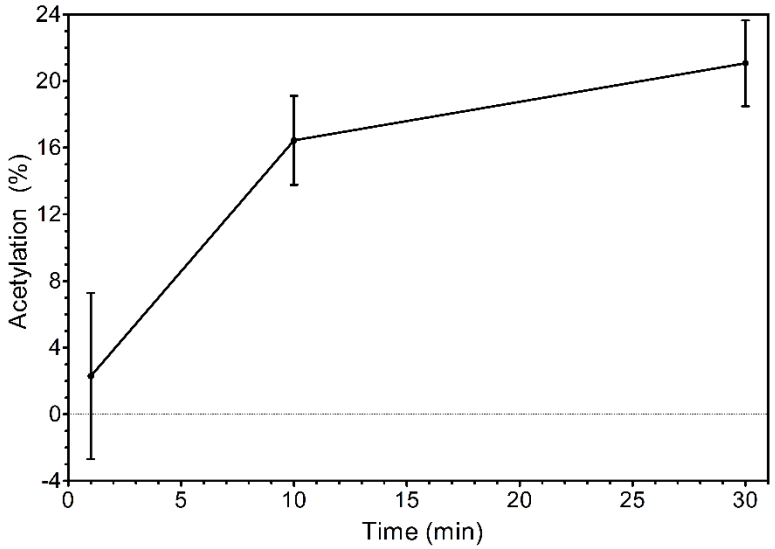
Peptide	Nt-Sequence (residues 2-8)							NatA 5 min Acetylation normalised to HMGA1 (%)	ADO 1 min Oxidation normalised to RGS4 (%)
DYNLL1	C	D	R	K	A	V	I	119.52	11.31
HMGA1	S	E	S	S	S	K	S	100	
PRMT3	C	S	L	A	S	G	A	32.49	24.87
AAAS	C	S	L	G	L	F	P	23.34	19.24
MTPN	C	D	K	E	F	M	W	22.55	3.72
JUNB	C	T	K	M	E	Q	P	22.19	4.92
IER3/IEX1	C	H	S	R	S	C	H	10.23	48.36
IL6	C	E	S	S	K	E	A	8.11	2.82
NHTL1	C	S	P	Q	E	S	G	7.59	7.41
MYL6	C	D	F	T	E	D	Q	6.71	1.91
RIMKLA	C	S	Q	L	W	F	L	6.04	22.50
FAM241A	C	S	A	G	E	L	L	5.50	9.59
ACTA2	C	E	E	E	D	S	T	3.38	
ETNPPL	C	E	L	Y	S	K	R	3.30	13.77
GPX1	C	A	A	R	L	A	A	1.01	42.05
GFPT2	C	G	I	F	A	Y	M	0.88	7.78
VDAC3	C	N	T	P	T	Y	C	0.74	27.79
TRAF1	C	Y	R	A	P	C	S	0.57	24.36
RGS16	C	R	T	L	A	A	F	0.49	22.61
IL32	C	F	P	K	V	L	S	0.48	110.79
NT5E	C	P	R	A	A	R	A	0.26	6.62
ALS2CL	C	N	P	E	E	A	A	0.26	3.30
CHURC1	C	G	D	C	V	E	K	0.25	0.00
STAC2	C	K	V	S	V	H	L	0.22	4.31
LYPLA2	C	G	N	T	M	S	V	0.17	0.96
P2RX6	C	P	Q	L	A	G	A	0.16	2.74
SUSD6	C	H	G	R	I	A	P	0.15	97.75
SHHC	C	F	P	G	S	A	T	0.15	18.11
RGS4	C	K	G	L	A	G	L	0.13	100.00
USP27X	C	K	D	Y	V	Y	D	0.04	4.12
CAHLM4	C	P	T	L	N	N	I	0.03	4.47
ZIP12	C	F	R	T	K	L	S	0.00	18.14
ASNS	C	G	I	W	A	L	F	0.00	0.00
Ripk1	C	V	P	L	P	P	S	0.00	10.50
ANKRD29	C	R	M	S	F	K	K	0.00	103.04
CDH2	C	R	I	A	G	A	L	0.00	55.80
OS7	C	F	G	R	R	M	K	0.00	152.16

Appendix 10: Deconvoluted mass spectra of oxidised peptides used in Chapter 4

Ox-DYNLL1$_{2-15}$

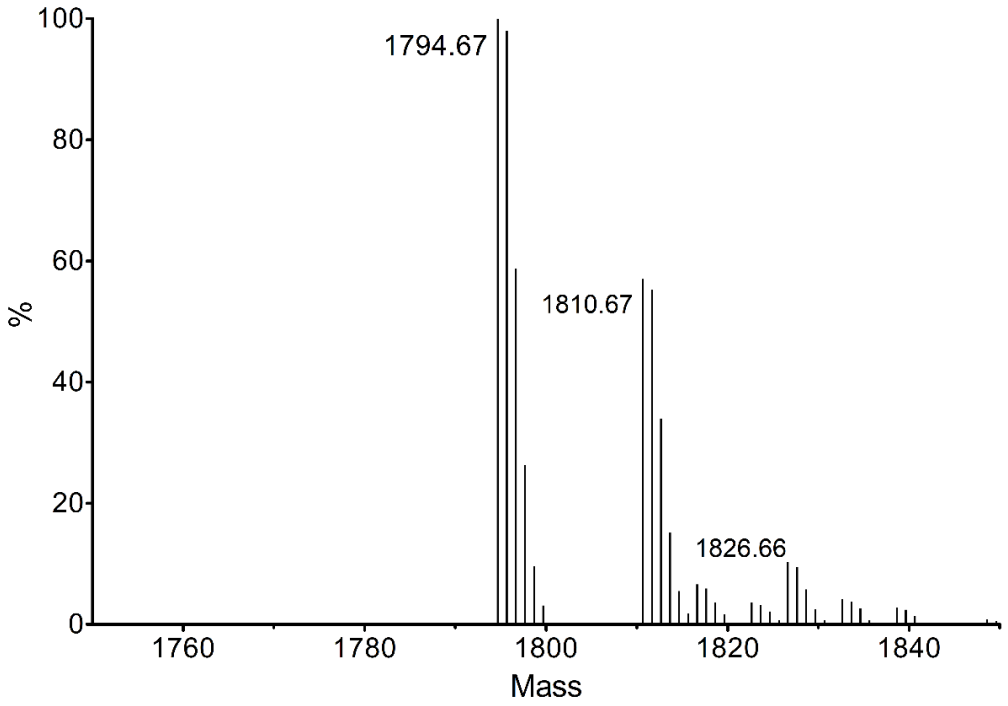


Deconvoluted mass spectra of DYNLL1$_{2-15}$ peptide (1 mM) after 1 hour incubation with *Hs*ADO (1 µM) at 37 °C. DYNLL1$_{2-15}$ MW = 1579.81; Ox-DYNLL1$_{2-15}$ MW = 1611.80



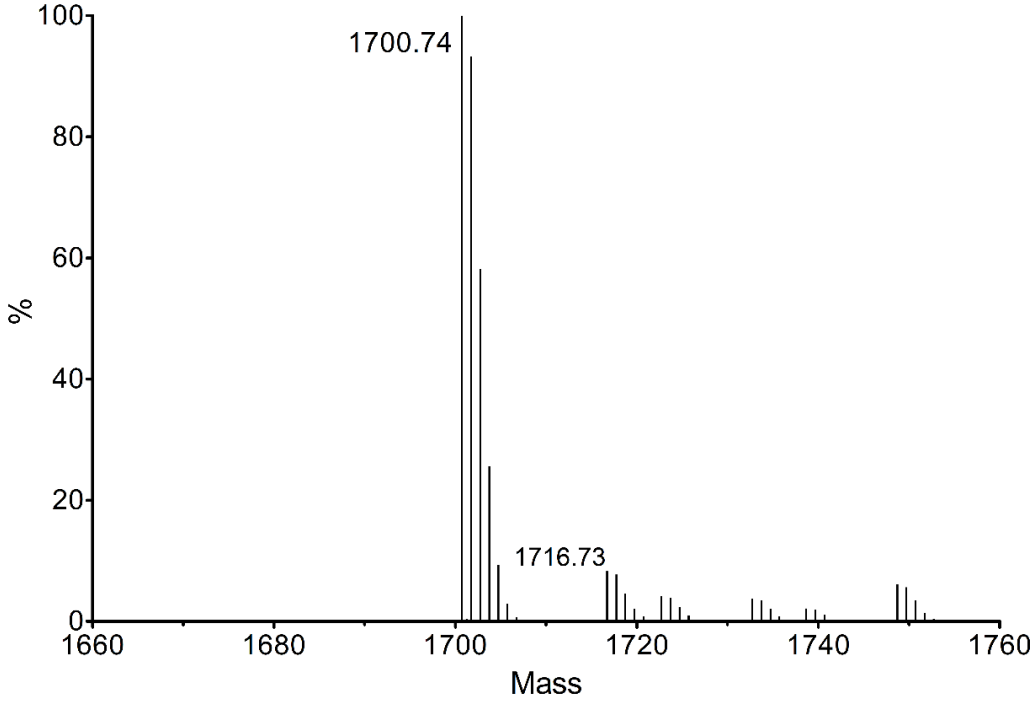
DYNLL1$_{2-15}$ peptide (100 µM) incubated with NatA (0.1 µM (kindly provided by Dr. Myllykoski)) and acetyl CoA (200 µM) at 37 °C for 30 minutes in 50 mM HEPES, 100 mM NaCl, 1 mM EDTA supplemented with 1 mM TCEP (Chapter 8).

Ox-JUNB$_{2-15}$



Deconvoluted mass spectra of JUNB$_{2-15}$ peptide (1 mM) after 1 hour incubation with *Hs*ADO (1 μM) at 37 °C. JUNB$_{2-15}$ MW = 1763.92; Ox-MTPN$_{2-15}$ MW = 1795.91

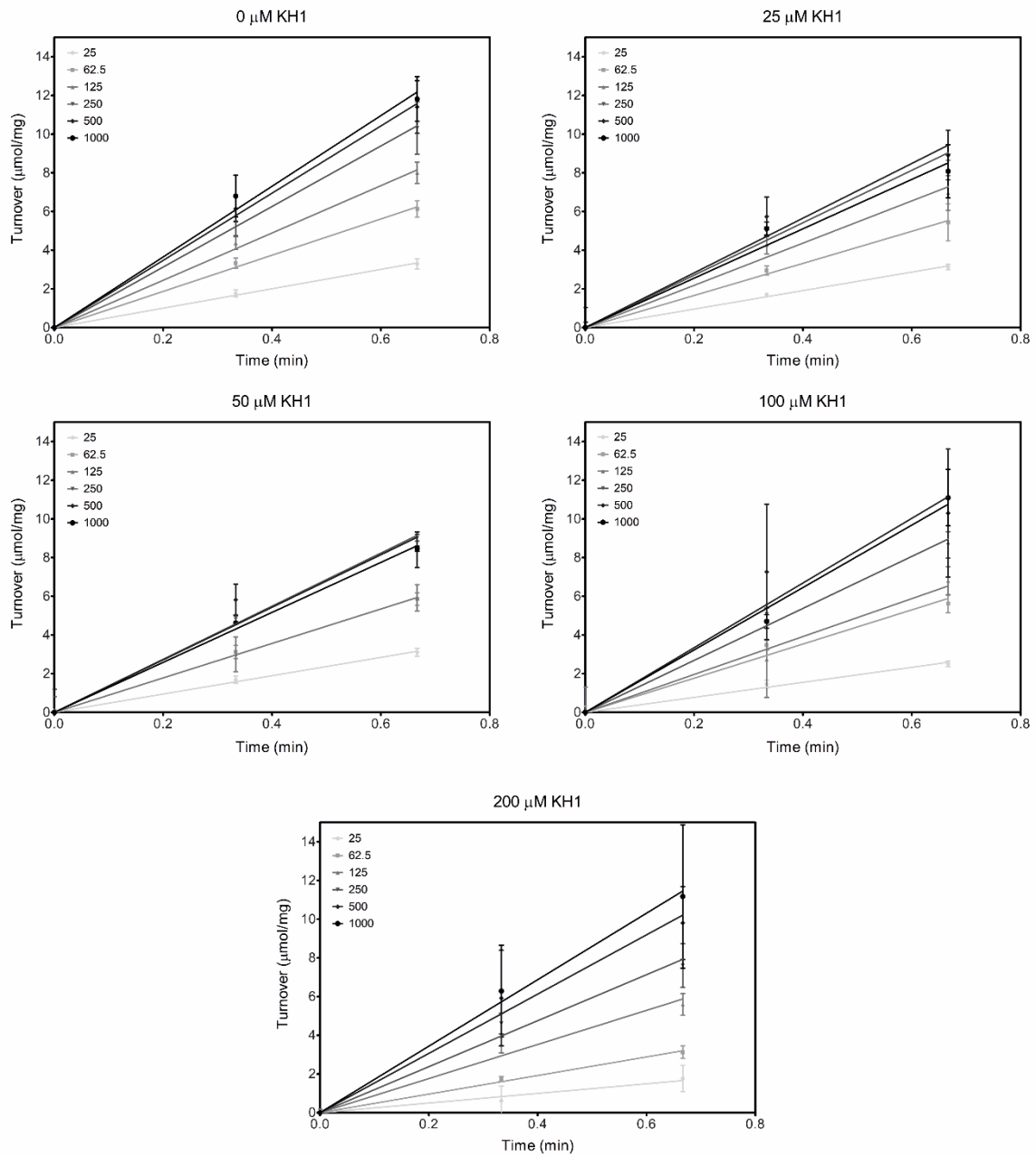
Ox-MTPN$_{2-15}$



Deconvoluted mass spectra of MTPN$_{2-15}$ peptide (1 mM) after 1 hour incubation with *Hs*ADO (10 μM) at 37 °C. MTPN$_{2-15}$ MW = 1669.93; Ox-MTPN$_{2-15}$ MW = 1701.92

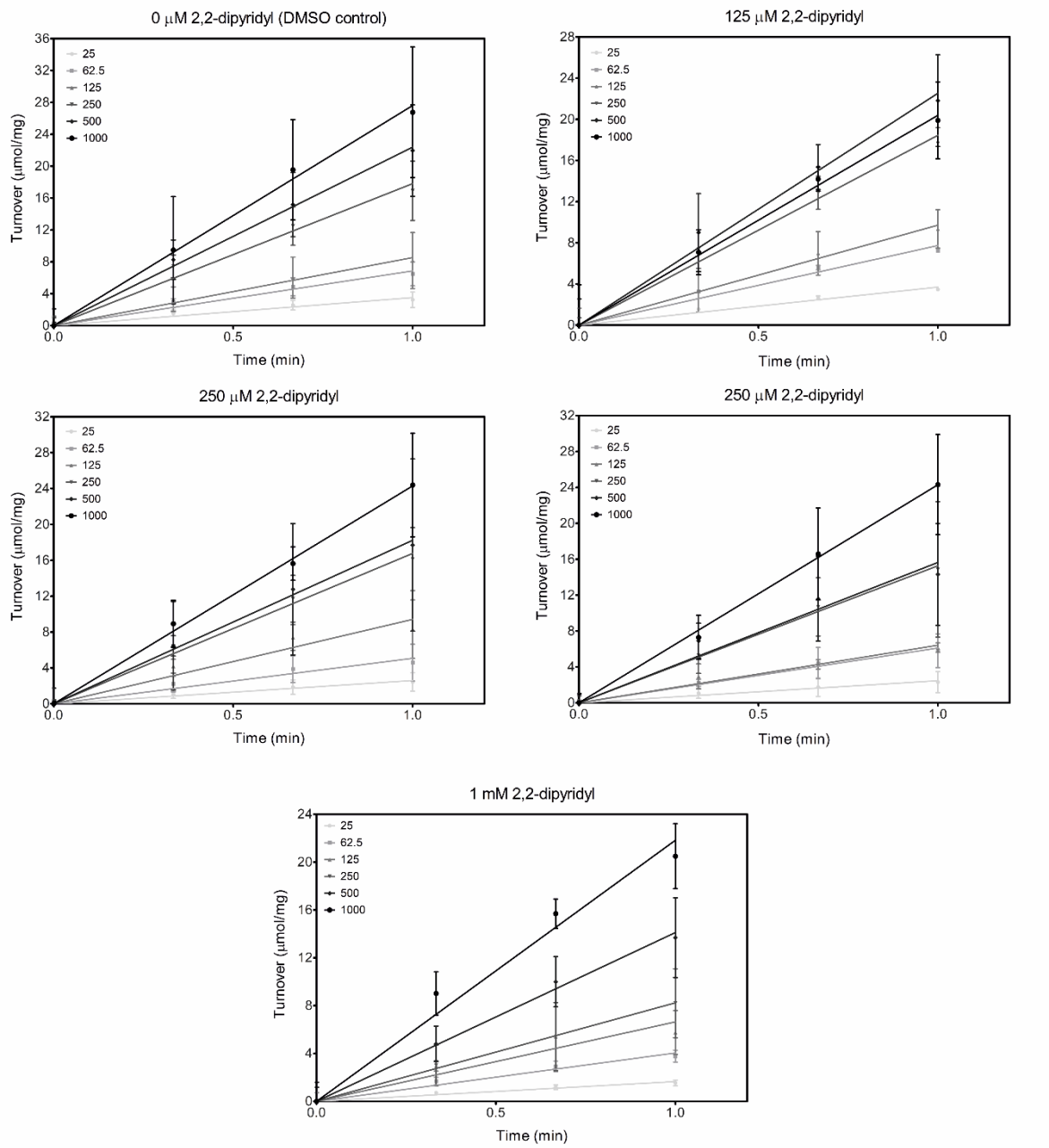
Appendix 11: Raw data from mode of inhibition assays

A: KH1 peptide inhibitor



Data obtained from high throughput MS analysis of reaction of HsADO (0.05 μM) with $RGS4_{2-15}$ at 37 °C with increasing KH1 peptide inhibitor concentrations. Graph legends refer to concentration of $RGS4_{2-15}$ in μM. Error bars show standard deviation, reactions carried out in triplicate; due to variability in data often only two replicates are represented.

B: Dipyridyl small molecule inhibitor



Data obtained from high throughput analysis MS of reaction of HsADO (0.05 μM) with $RGS4_{2\text{-}15}$ at 37 °C with increasing 2,2-dipyridyl inhibitor concentrations. Graph legends refer to concentration of $RGS4_{2\text{-}15}$ in μM. Error bars show standard deviation, reactions carried out in triplicate.

Appendix 12: Crystallisation screening trays

24 Well 1

Sample: ADO Concentration: 16 mg/ml Tray Number: 1 Experiment ID: KH Karen Date: Thursday, August 27, 2020 5:30 am PST

	1	2	3	4	5	6
A	12 % PEG 3350 0.1 M Ammonium nitrate	14 % PEG 3350 0.1 M Ammonium nitrate	16 % PEG 3350 0.1 M Ammonium nitrate	18 % PEG 3350 0.1 M Ammonium nitrate	20 % PEG 3350 0.1 M Ammonium nitrate	22 % PEG 3350 0.1 M Ammonium nitrate
B	12 % PEG 3350 0.2 M Ammonium nitrate	14 % PEG 3350 0.2 M Ammonium nitrate	16 % PEG 3350 0.2 M Ammonium nitrate	18 % PEG 3350 0.2 M Ammonium nitrate	20 % PEG 3350 0.2 M Ammonium nitrate	22 % PEG 3350 0.2 M Ammonium nitrate
C	12 % PEG 3350 0.3 M Ammonium nitrate	14 % PEG 3350 0.3 M Ammonium nitrate	16 % PEG 3350 0.3 M Ammonium nitrate	18 % PEG 3350 0.3 M Ammonium nitrate	20 % PEG 3350 0.3 M Ammonium nitrate	22 % PEG 3350 0.3 M Ammonium nitrate
D	12 % PEG 3350 0.4 M Ammonium nitrate	14 % PEG 3350 0.4 M Ammonium nitrate	16 % PEG 3350 0.4 M Ammonium nitrate	18 % PEG 3350 0.4 M Ammonium nitrate	20 % PEG 3350 0.4 M Ammonium nitrate	22 % PEG 3350 0.4 M Ammonium nitrate

24 Well 2

Sample: ADO Concentration: 16 mg/ml Tray Number: 2 Experiment ID: KH Karen Date: Thursday, August 27, 2020 5:38 am PST

	1	2	3	4	5	6
A	10 % PEG 4000 0.1 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	12 % PEG 4000 0.1 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	14 % PEG 4000 0.1 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	16 % PEG 4000 0.1 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	18 % PEG 4000 0.1 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	20 % PEG 4000 0.1 M Sodium acetate trihydrate 0.1 M Tris pH 8.5
B	10 % PEG 4000 0.2 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	12 % PEG 4000 0.2 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	14 % PEG 4000 0.2 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	16 % PEG 4000 0.2 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	18 % PEG 4000 0.2 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	20 % PEG 4000 0.2 M Sodium acetate trihydrate 0.1 M Tris pH 8.5
C	10 % PEG 4000 0.3 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	12 % PEG 4000 0.3 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	14 % PEG 4000 0.3 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	16 % PEG 4000 0.3 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	18 % PEG 4000 0.3 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	20 % PEG 4000 0.3 M Sodium acetate trihydrate 0.1 M Tris pH 8.5
D	10 % PEG 4000 0.4 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	12 % PEG 4000 0.4 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	14 % PEG 4000 0.4 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	16 % PEG 4000 0.4 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	18 % PEG 4000 0.4 M Sodium acetate trihydrate 0.1 M Tris pH 8.5	20 % PEG 4000 0.4 M Sodium acetate trihydrate 0.1 M Tris pH 8.5

24 Well 3

Sample: ADO Concentration: 16 mg/ml Tray Number: 3 Experiment ID: KH Karen Date: Thursday, August 27, 2020 5:34 am PST

	1	2	3	4	5	6
A	16 % PEG 3350 0.1 M Sodium fluoride	18 % PEG 3350 0.1 M Sodium fluoride	20 % PEG 3350 0.1 M Sodium fluoride	22 % PEG 3350 0.1 M Sodium fluoride	24 % PEG 3350 0.1 M Sodium fluoride	26 % PEG 3350 0.1 M Sodium fluoride
B	16 % PEG 3350 0.2 M Sodium fluoride	18 % PEG 3350 0.2 M Sodium fluoride	20 % PEG 3350 0.2 M Sodium fluoride	22 % PEG 3350 0.2 M Sodium fluoride	24 % PEG 3350 0.2 M Sodium fluoride	26 % PEG 3350 0.2 M Sodium fluoride
C	16 % PEG 3350 0.3 M Sodium fluoride	18 % PEG 3350 0.3 M Sodium fluoride	20 % PEG 3350 0.3 M Sodium fluoride	22 % PEG 3350 0.3 M Sodium fluoride	24 % PEG 3350 0.3 M Sodium fluoride	26 % PEG 3350 0.3 M Sodium fluoride
D	16 % PEG 3350 0.4 M Sodium fluoride	18 % PEG 3350 0.4 M Sodium fluoride	20 % PEG 3350 0.4 M Sodium fluoride	22 % PEG 3350 0.4 M Sodium fluoride	24 % PEG 3350 0.4 M Sodium fluoride	26 % PEG 3350 0.4 M Sodium fluoride

6 Well 1

	1	2	3
A	0.1 m Bis-Tris HCl, 15 % PEG 3350, 0.2 M Li_2SO_4,	0.1 m Bis-Tris HCl, 20 % PEG 3350, 0.2 M Li_2SO_4,	0.1 m Bis-Tris HCl, 25 % PEG 3350, 0.2 M Li_2SO_4,
B	0.1 m Bis-Tris HCl, 20 % PEG 3350, 0.1 M Li_2SO_4,	0.1 m Bis-Tris HCl, 20 % PEG 3350, 0.2 M Li_2SO_4,	0.1 m Bis-Tris HCl, 20 % PEG 3350, 0.3 M Li_2SO_4,

6 Well 2

	1	2	3
A	0.1 M Tris pH 8, 10 % PEG 3350, 0.3 M $MgCl_2$	0.1 M Tris pH 8, 15 % PEG 3350, 0.3 M $MgCl_2$	0.1 M Tris pH 8, 20 % PEG 3350, 0.3 M $MgCl_2$
B	0.1 M Tris pH 8, 15 % PEG 3350, 0.2 M $MgCl_2$	0.1 M Tris pH 8, 15 % PEG 3350, 0.3 M $MgCl_2$	0.1 M Tris pH 8, 15 % PEG 3350, 0.4 M $MgCl_2$

Appendix 13: MODELLER scripts and .ali files

ADO.ali:

```
>P1;ADO
sequence:ADO:::::::0.00: 0.00

MPRDNMASLIQRIARQACLTFRGSGGGRGASDRDAASGPEAPMQPGFPENLSKLKSLLTQLRAEDLNIAPRK
ATLQPLPPNLPPVTYMHIYETDGFSLGVFLLKSGTSIPLHDHPGMHGMLKVLYGTVRISCMDKLDAGGGQRP
RALPPEQQFEPPLQPREREAVRPGVLRSRAEYTEASGPCILTPHRDNLHQIDAVEGPAAFLDILAPPYDPDD
GRDCHYYRVLEPVRPKEASSSACDLPREVWLLETPQADDFWCEGEPYPGPKVFP*
```

align2d.py:

```
from modeller import *

env = environ()
aln = alignment(env)
env.io.hetatm = True

mdl = model(env, file='6s7e', model_segment=('FIRST:A','LAST:A'))
aln.append_model(mdl, align_codes='6s7e', atom_files='6s7e.pdb')
aln.append(file='ADO.ali', align_codes='ADO')
aln.align2d()
aln.write(file='ADO-6s7e.ali', alignment_format='PIR')
aln.write(file='ADO-6s7e.pap', alignment_format='PAP')
```

ADO-6s7e.ali:

```
>P1;6s7e
structureX:6s7e.pdb:  -1 :A:+223 :A:MOL_ID  1; MOLECULE  PLANT CYSTEINE
OXIDASE 4; CHAIN  A; EC  1.13.11.20; ENGINEERED  YES:MOL_ID  1;
ORGANISM_SCIENTIFIC  ARABIDOPSIS THALIANA; ORGANISM_COMMON  THALE CRESS;
ORGANISM_TAXID  3702; GENE  PCO4, AT2G42670, F14N22.6; EXPRESSION_SYSTEM
ESCHERICHIA COLI 'BL21-GOLD(DE3)PLYSS AG EXPRESSION_SYSTEM_TAXID
866768; EXPRESSION_SYSTEM_PLASMID  PET 28A: 1.82: 0.19
---SHMPYFAQRLYNTCKASF--------SSD-----G------
PITEDALEKVRNVLEKIKPSDVGIEQDAQLA
RP-----PAIKYLHLHECDSFSIGIFCMPPSSMIPLHNHPGMTVLSKLVYGSMHVKSYDWL-----------
---
---EP--QP---SQARPAKLVKDTEMTAQSPVTTLYPKSGGNIHCFKAIT-
HCAILDILAPPYSSEHDRHCTYFR
KSRREDLPGELEVDG-EVVTDVTWLEEFQPPDDFVIRRIPYRGPVIRT.*

>P1;ADO
sequence:ADO:     : :     : ::: 0.00: 0.00
MPRDNMASLIQRIARQACLTFRGSGGGRGASDRDAASGPEAPMQPGFPENLSKLKSLLTQLRAEDLNIAPRK
ATL
QPLPPNLPPVTYMHIYETDGFSLGVFLLKSGTSIPLHDHPGMHGMLKVLYGTVRISCMDKLDAGGGQRPRAL
PPE
QQFEPPLQPREREAVRPGVLRSRAEYTEASGPCILTPHR-
DNLHQIDAVEGPAAFLDILAPPYDPDDGRDCHYYR
VLE-PVRPKEASSSACDLPREVWLLETPQA-DDFWCEGEPYPGPKVFP.*
```

model-ligand.py:

```
# Comparative modeling with ligand transfer from the template
from modeller import *              # Load standard Modeller classes
from modeller.automodel import *    # Load the automodel class

log.verbose()    # request verbose output
env = environ()  # create a new MODELLER environment to build this model
in

# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']

# Read in HETATM records from template PDBs
env.io.hetatm = True

a = automodel(env,
              alnfile  = 'ADO-6s7e.ali',  # alignment filename
              knowns   = '6s7e',              # codes of the templates
              sequence = 'ADO',              # code of the target
              assess_methods=(assess.DOPE,
                              assess.GA341))
a.starting_model= 1                 # index of the first model
a.ending_model  = 5                 # index of the last model
                                    #  (determines  how  many  models  to
calculate)
a.make()                            # do the actual comparative modeling
```

From model-ligand.log:

```
>> Summary of successfully produced models:
Filename                          molpdf     DOPE score    GA341 score
----------------------------------------------------------------------
ADO.B99990001.pdb              1673.42871   -23663.00781        1.00000
ADO.B99990002.pdb              1653.41333   -23674.63867        1.00000
ADO.B99990003.pdb              1798.59729   -23448.36133        1.00000
ADO.B99990004.pdb              1692.21252   -23839.90820        1.00000
ADO.B99990005.pdb              1685.18774   -23671.26758        1.00000
```

Model 4 selected

evaluate_model.py:

```
from modeller import *
from modeller.scripts import complete_pdb

log.verbose()    # request verbose output
env = environ()
env.libs.topology.read(file='$(LIB)/top_heav.lib') # read topology
env.libs.parameters.read(file='$(LIB)/par.lib') # read parameters

env.io.hetatm = True

# read model file
mdl = complete_pdb(env, 'ADO.B99990004.pdb')

# Assess with DOPE:
s = selection(mdl)   # all atom selection
s.assess_dope(output='ENERGY_PROFILE                              NO_REPORT',
file='ADOmodel4.profile',
              normalize_profile=True, smoothing_window=15)
```

evaluate_template.py:

```
from modeller import *
from modeller.scripts import complete_pdb

log.verbose()    # request verbose output
env = environ()
env.libs.topology.read(file='$(LIB)/top_heav.lib') # read topology
env.libs.parameters.read(file='$(LIB)/par.lib') # read parameters

# directories for input atom files
env.io.atom_files_directory = './:../atom_files'

env.io.hetatm = True

# read model file
mdl = complete_pdb(env, '6s7e.pdb', model_segment=('FIRST:A', 'LAST:A'))

s = selection(mdl)
s.assess_dope(output='ENERGY_PROFILE NO_REPORT', file='6s7e.profile',
              normalize_profile=True, smoothing_window=15)
```

plot_profiles.py:

```
import pylab
#import matplotlib
import modeller

def r_enumerate(seq):
    """Enumerate a sequence in reverse order"""
    # Note that we don't use reversed() since Python 2.3 doesn't have it
    num = len(seq) - 1
    while num >= 0:
        yield num, seq[num]
        num -= 1

def get_profile(profile_file, seq):
    """Read `profile_file` into a Python array, and add gaps corresponding
to
       the alignment sequence `seq`."""
    # Read all non-comment and non-blank lines from the file:
    f = open(profile_file)
    vals = []
    for line in f:
        if not line.startswith('#') and len(line) > 10:
            spl = line.split()
            vals.append(float(spl[-1]))
    # Insert gaps into the profile corresponding to those in seq:
    for n, res in r_enumerate(seq.residues):
        for gap in range(res.get_leading_gaps()):
            vals.insert(n, None)
    # Add a gap at position '0', so that we effectively count from 1:
    vals.insert(0, None)
    return vals

e = modeller.environ()
a = modeller.alignment(e, file='ADO-6s7e.ali')

template = get_profile('6s7e.profile', a['6s7e'])
model = get_profile('ADOmodel4.profile', a['ADO'])

# Plot the template and model profiles in the same plot for comparison:
pylab.figure(1, figsize=(10,6))
pylab.xlabel('Alignment position')
pylab.ylabel('DOPE per-residue score')
pylab.plot(model, color='red', linewidth=2, label='Model 4')
pylab.plot(template, color='green', linewidth=2, label='Template (PCO4)')
pylab.legend()
pylab.savefig('dope_profile.png', dpi=65)
```

loop_refine.py:

```
# Loop refinement of an existing model
from modeller import *
from modeller.automodel import *

log.verbose()
env = environ()

# directories for input atom files
env.io.atom_files_directory = './:../atom_files'

env.io.hetatm = True

# Create a new class based on 'loopmodel' so that we can redefine
# select_loop_atoms (necessary)
class MyLoop(loopmodel):
    # This routine picks the residues to be refined by loop modeling
    def select_loop_atoms(self):
        # 10 residue insertion
        return selection(self.residue_range('X', 'Y'))

m = MyLoop(env,
           inimodel='ADO_basic4.pdb', # initial model of the target
           sequence='ADO')          # code of the target

m.loop.starting_model= 1           # index of the first loop model
m.loop.ending_model  = 10          # index of the last loop model
#m.loop.md_level = refine.very_fast # loop refinement method; this yields
m.loop.md_level = refine.slow # models quickly but of low quality;
                                   # use refine.slow for better models

m.make()
```

model_energies.py:

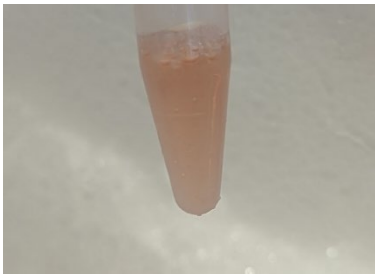
```
from modeller import *
from modeller.scripts import complete_pdb

log.verbose()    # request verbose output
env = environ()
env.libs.topology.read(file='$(LIB)/top_heav.lib') # read topology
env.libs.parameters.read(file='$(LIB)/par.lib') # read parameters

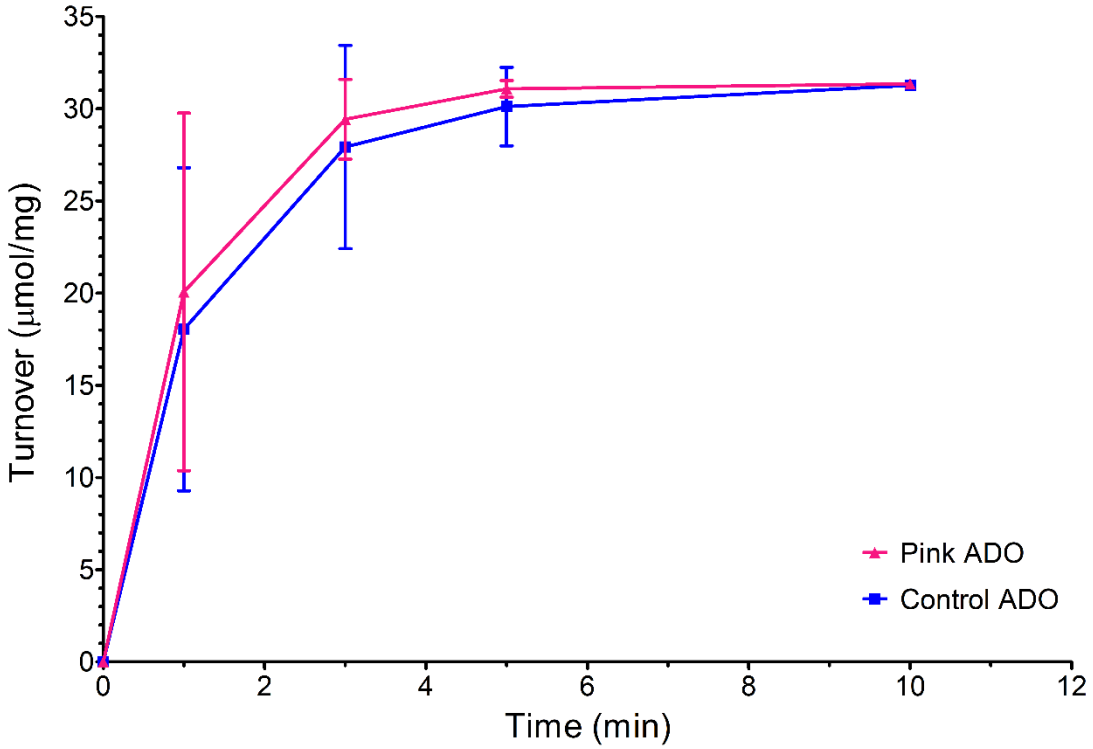
for i in range(1, 11):
    # read model file
    code = "ADO.BL%04d0001.pdb" % i
    mdl = complete_pdb(env, code)
    s = selection(mdl)
    s.assess_dope(output='ENERGY_PROFILE NO_REPORT', file='ADO.profile',
                  normalize_profile=True, smoothing_window=15)
```

Appendix 14: 1,10-phenanthroline contamination during ADO purification

A - Purified and concentrated recombinant *Hs*ADO (as used in Chapter 6, Section 6.4) – 2.55 mM *Hs*ADO in 400 mM NaCl, 50 mM Tris, pH 7.5:



Contaminated *Hs*ADO (pictured above) was able to oxidise $RGS4_{2-15}$ peptide at the same rate as a control sample of *Hs*ADO (kindly provided by Roberta Sher (Flashman Group)):



Assay conditions: 100 µM $RGS4_{2-15}$, 0.1 µM *Hs*ADO, 5 mM TCEP, 20 µM Fe(II), 1 mM L-ascorbate, 50 mM Bis-Tris Propane, 50 mM NaCl, pH 7.5, 37 °C

Turnover of $RGS4_{2-15}$ peptide was measured by high-throughput MS as described in Thesis.

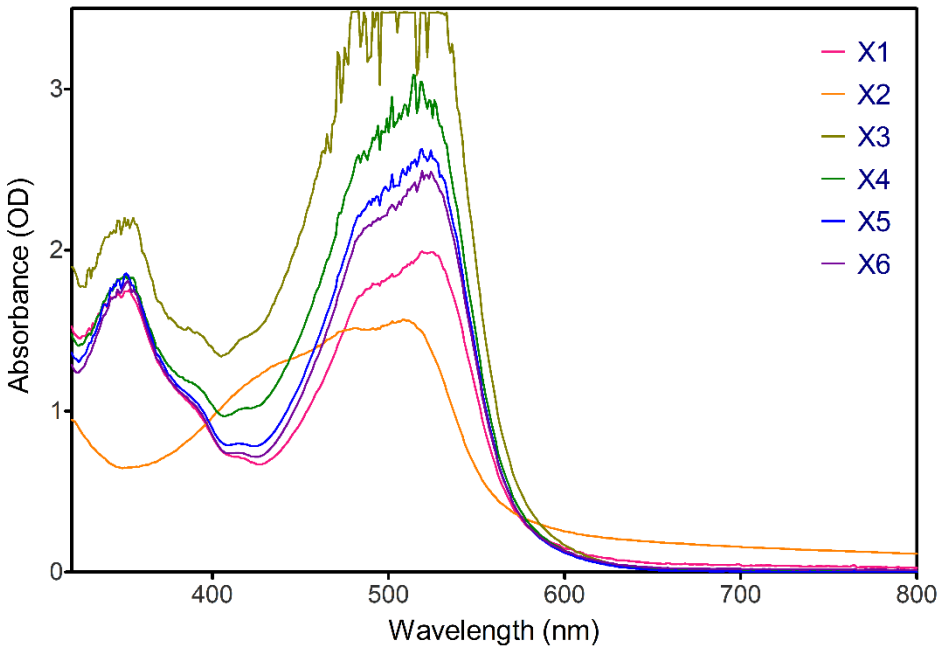
This result demonstrated that a substrate peptide was able to outcompete any contaminant that might be bound in the active site of *Hs*ADO.

B - *Hs*ADO sample (**A**) after addition of iron chelator 2,2-dipyridyl – 2 mM *Hs*ADO, 10 mM 2,2-dipyridyl, 4 mM Fe (II) in 400 mM NaCl, 50 mM Tris, pH 7.5:



In order to confirm that the colour of the *Hs*ADO sample was due to 1,10-phenanthroline contamination, titration experiments were conducted using Fe(II), 2,2-dipyridyl, and 1,10-phenanthroline.

No colour change was observed when 1,10-phenanthroline was added to a solution of 4 mM Fe_2SO_4 (Fe(II)) and 10 mM 2,2-dipyridyl (final concentrations: 2 mM Fe(II), 5 mM 2,2-dipyridyl, 5 mM 1,10-phenanthroline). However, addition of 2,2-dipyridyl to a solution of 1,10-phenanthroline and Fe(II) did produce a change in solution colour:



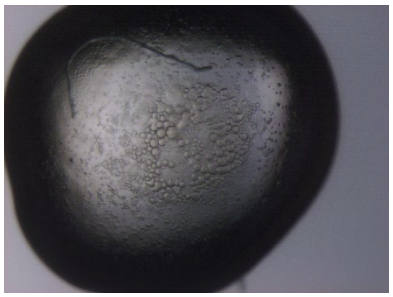
X1 = 2 mM *Hs*ADO, 10 mM 2,2-dipyridyl, 4 mM Fe (II) (**B**)
X2 = 5 mM Fe(II) + 5 mM 1,10-phenanthroline
X3 = 5 mM Fe (II) + 5 mM 1,10-phenanthroline + 10 mM 2,2-dipyridyl
X4 = 5 mM Fe (II) + 2.5 mM 1,10-phenanthroline + 10 mM 2,2-dipyridyl
X5 = 5 mM Fe (II) + 1 mM 1,10-phenanthroline + 10 mM 2,2-dipyridyl
X6 = 5 mM Fe (II) + 0.5 mM 1,10-phenanthroline + 10 mM 2,2-dipyridyl
(All solutions in 400 mM NaCl, 50 mM Tris, pH 7.5)

Absorbance was measured using CLARIOstar (details).

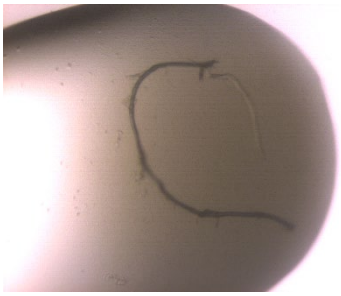
*Hs*ADO sample above (**B**) had the same absorbance profile as solutions of Fe(II), 2,2-dipyridyl, and 1,10-phenanthroline. It was therefore concluded that the sample contained 1,10-phenthroline, likely at a concentration of < 0.5 mM.

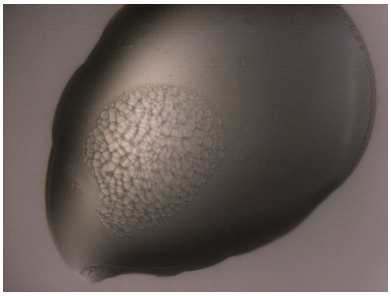
Appendix 15: ADO + KH1 co-crystallisation trial microscope pictures

6 Well 2
2 days after trays laid. B1:



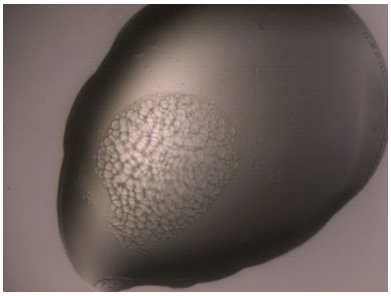
Crystal Screen™
6 days after trays laid, A3, D1, D3:

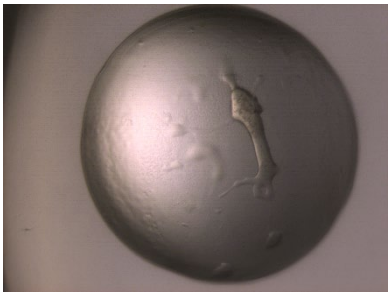




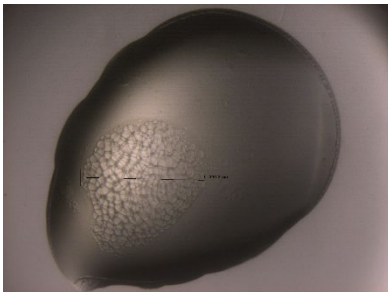


8 days after trays laid, D1, D3:





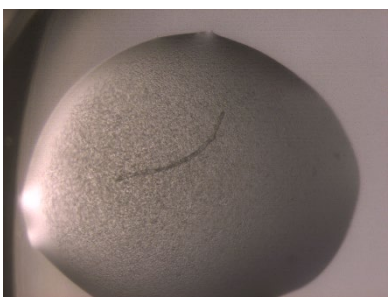
11 days after trays laid, D1, D3:



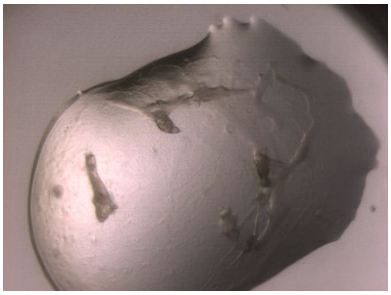


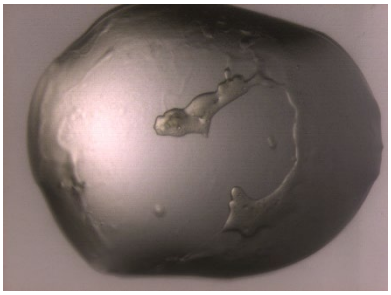
JCSG Plus™

6 days after trays laid, D3:



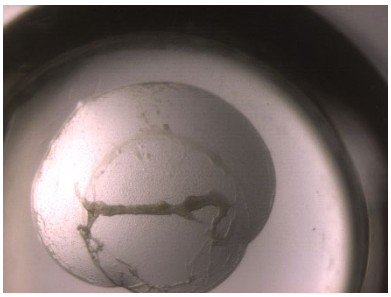
8 days after trays laid, B2, G4:





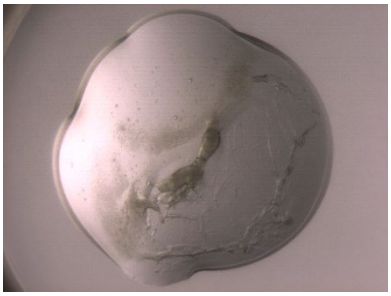
Morpheus®

11 days after trays laid, D4:

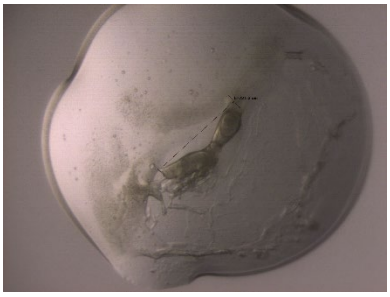


PACT Premier™

8 days after trays laid, D12:



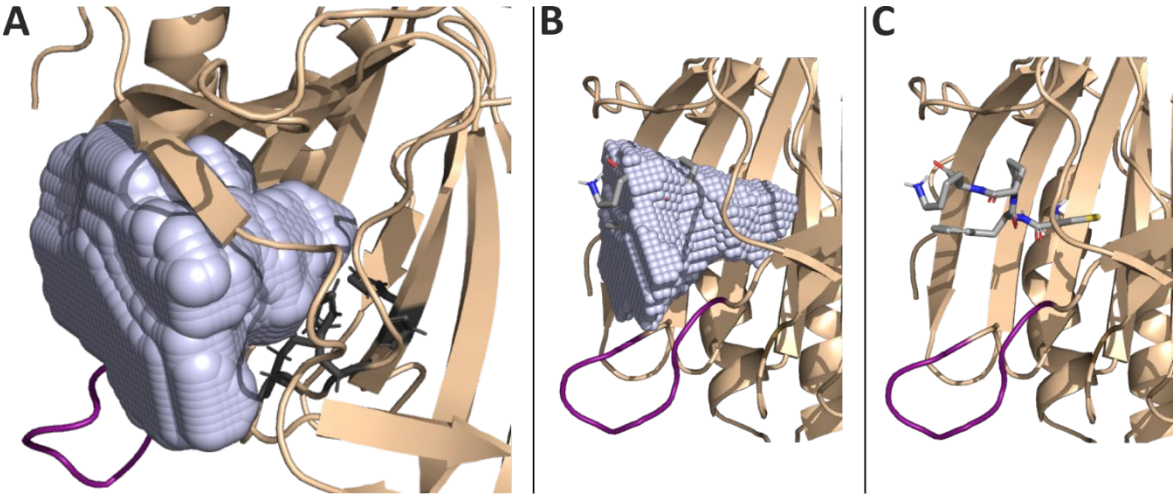
15 days after trays laid, D12:



Appendix 16: CavityPlus analysis of *Mm*ADO

Cavity detection using CavityPlus server[263] and chain D of *Mm*ADO (PDB: 7LVZ)[6].

CavityPlus results (as seen in Figure 6.7):



A: Cavity 1 (vacant space) in *Mm*ADO, as predicted by CavityPlus.
B: Cavity 1 (vacant space) viewed with IL32 AutoDock4 result (run 67, Figure 6.7), showing IL32 peptide residues 2-5 (CFPK).
C: IL32 AutoDock4 result (run 67, Figure 6.7) viewed without Cavity 1.
Hairpin loop in purple, iron-binding His residues in dark grey, cavity in blue-white, peptide residues in light grey (O atoms red, N atoms blue, S atoms yellow).

Cavity Results:

#	Pred Max pKd	Pred Ave pKd	DrugScore	Druggability	Surface	More
1	9.07	6.85	1429.00	Strong	☑	...

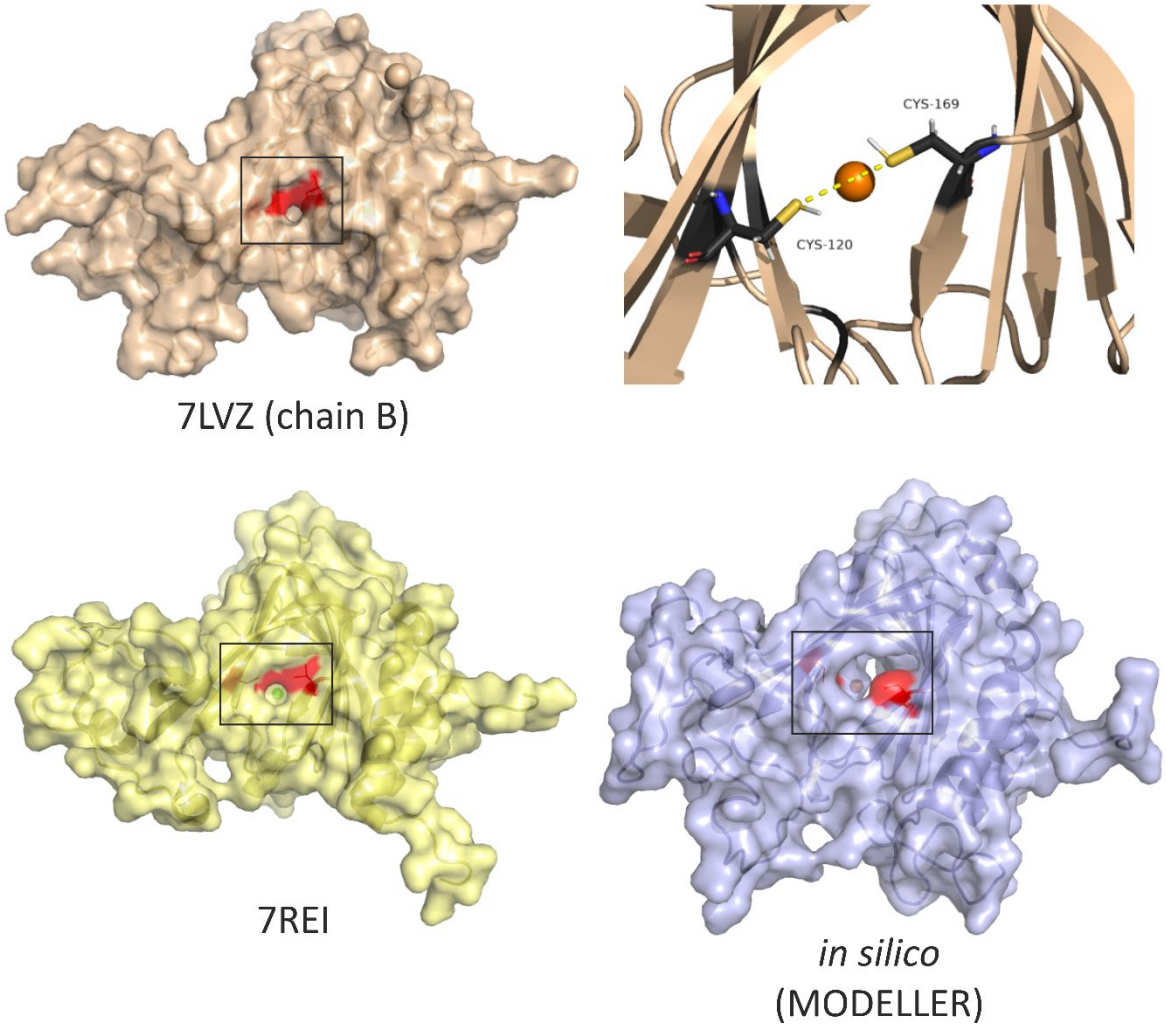
Surface Area (Å^2)	529.00
Volume (Å^3)	791.25
Box Size (Å)	16.5 14.5 14.0
Box Center (Å)	38.75 80.75 49.0

*Mm*ADO residues on the surface of the predicted cavity (residues highlighted in alignment Figure 6.9 are highlighted in the same colour):

Green = Fe-binding His residues
Blue/black = Phe-Asp-Leu motif
Pink = Polar pocket residues
Orange = aromatic active site residues
Tyr75 and Phe190 (from polar pocket and Phe-Asp-Leu motif, respectively) are also aromatic active site residues.

Residues on Protein Surface in Cavity 1				
VAL73	SER85	HIS102	LEU194	HIS207
THR74	LEU86	MET105	PRO196	TYR208
TYR75	GLY87	LEU109	PRO197	GLN238
MET76	VAL88	HIS179	TYR198	PHE242
HIS77	PHE89	ILE181	ASP199	TRP243
TYR79	ILE97	PHE190	PRO200	CYS244
GLU80	PRO98	LEU191	ARG204	GLU245
THR81	HIS100	ASP192	ASP205	GLY246
PHE84	ASP101	ILE193	CYS206	GLU247

Appendix 17: Potential O_2 access tunnel in ADO structures



PDB codes shown below published crystal structures.

Cys120 and Cys169 (*Mm*ADO numbering) shown in red on full protein structures.

Opening to proposed backside 'O_2 access' tunnel highlighted with black square.

Top right image shows close-up view of Cys120 and Cys169 in structure of *Mm*ADO (PDB: 7LZV chain D).