

Plant Cysteine Oxidase Function in Early Plants and Algae

A thesis submitted to the faculty of Physical Sciences in partial fulfilment of the degree of Doctor of Philosophy in Chemical Biology

Leah Taylor-Kearney
Wolfson College
2020

Dedicated to Dan and Joan Taylor

Acknowledgements

I would like to thank Prof Emily Flashman and Prof Rosalind Rickaby for their unwavering support and guidance over the course of this project. I could not have asked for better mentors.

A big thank you to all current and previous Flashman group members. You are exceptional scientists and colleagues, and this process would not have been as fun without you.

I extend a huge thank you to Ms Elisabete Pires, without whom much of the proteomic results would not have been possible. I thank all members of the CJS and Kawamura groups, past and present, for helping me over the years. I would like to extend a special thank you to Dr Adam Hardy, who always knew where everything was and never failed to lend a helping hand. Thank you to Dr Anthony Tumber, who always went out of his way to help me with RapidFire. I thank Ms Anka Lucic, Dr Patrick Rabe and Ms Pauline Lang for their help with crystallography. I would like to thank Dr Chiara Maniaci for her help with antibody work and Dr Eidarus Salah for providing the 3C KnPCO pellet while we were dealing with phage. I would like to thank Dr Dona Gunawardana for her constant help in the new chemistry lab, and Ms Karen Heathcote for a list of deeds too long to mention – you are the *best* Karen.

I would like to thank Mr Louis Claxton, Ms Poppy Diver and Dr Qiong Zhang for their help with algal culturing. A special thank you to Dr Joseph Snow, who was instrumental in getting the algal culturing project off the ground. A big thank you to Dr Philip Holdship for his work on ICP-MS. Thank you to Dr Madhi El Bendif for his patience and help with phylogenetic analysis when I needed it most. Thank you to Prof Francesco Licausi for the MpPCO plasmid, help with phylogenetic analysis and for being a fountain of knowledge over the years. I would also like to thank Mr Denis Shutin for his work on native MS.

I wish to thank the CRL facilities managers Dr Deborah Palmer, Mr Ron Tarry and especially Mr Martin Wall, who were instrumental in setting-up my novel experimental platforms. I would like to thank Dr Karl Morten for putting up with me in his lab at the John Radcliffe. I would like to thank Prof Jane Barling, who took time out of her busy schedule to show me some neat proteomics preparation techniques.

I would like to thank Dr Michael McDonagh and Dr Alistair Scott for proofreading Chapter 5, Ms Karen Heathcote for proofreading Chapters 2 and 6.

I would like to thank my friends, Ms Barbora Sojkova, Mr Howard Friman, Ms Joanna Storr, Mr Tom Storr, Ms Leoma Williams, Ms Lily Goodyer-Sait, Ms Rosie Worster, Dr Anne Atiso, Dr Kaja Fenn, Dr Brooke Johnson, Dr Darragh Ennis, Dr Joan Gannon, Ms Anna Szyszko, Ms Jodie White, Ms Hannah Wilson and Ms Charlotte Newman for keeping me fed, watered and sane over the course of this PhD. I could not have done it without you.

A special mention to Mr Zach Jackson, who always made sure I was fed and never complained.

Thank you to my family, who never stopped believing in me.

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

All aerobic organisms require O_2 for survival. When O_2 is limited (hypoxia), a response is required to reduce demand and/or improve supply. While plants can survive temporary periods of acute hypoxia, prolonged exposure can result in plant damage or death. A concern over food security has driven higher plant hypoxia research, with little contributed to understanding marine or early land plants. The photosynthesis of phytoplankton is the dominant global process that replenishes atmospheric and oceanic O_2. In light of global warming, and the reported expansion of oceanic O_2 minimums, the impact oceanic deoxygenation will have on algae, or *vice versa,* must be assessed.

The higher plant hypoxia response is mediated by the group VII ETHYLENE RESPONSE FACTORs (ERF-VIIs). ERF-VII transcription factor stability is regulated by the Arginine/Cysteine (Cys/Arg) branch of the N-degron pathway. It has been shown that the O_2-dependent Plant Cysteine Oxidases (PCOs) catalyse oxidation of ERF-VII N-terminal cysteine residues to form cysteine-sulfinic acid. This renders the ERF-VII N-terminus a substrate for arginyl-tRNA protein transferase (ATE); arginylation promotes ubiquitination by E3 Ubiquitin ligase PRT6 (PRT6) and degradation by the proteasome. During hypoxia, the ERF-VII transcription factors are stabilised, resulting in upregulation of genes which promote survival. Genes encoding for putative PCOs are ubiquitous across the plant kingdom. However, the ERF-VII transcription factors are confined to the flowering plants.

Undoubtedly, there are metabolic challenges associated with living as single-celled aquatic organisms, and also with transitioning from single-celled phytoplankton to land-dwelling, multicellular plants. What role the PCOs might have played in this transition, or in what capacity the PCOs help regulate algal health with respect to O_2 has yet to be described. This thesis uses a combination of recombinant protein studies, PCO substrate discovery and algal culturing to investigate the evolution of the plant O_2 sensing pathway, the effect of hypoxia on algal health, and the role of algal PCOs. To address these questions, a series of model plant species were selected for study. *M. polymorpha* is a liverwort, with liverworts reputed to be the first land plants; possibly retaining features of its algal ancestors and extant land plants. *K. nitens* is an undifferentiated freshwater alga and is a model for understanding the plant transition from water to land. *E. huxleyi* is a globally distributed Haptophyte, with extraordinary adaptive capabilities and therefore a powerful tool for understanding algal adaptation to climate change-driven stresses. All plants contain putative PCOs, with *K. nitens* and *M. polymorpha* also containing ATE and PRT6 homologs – essential elements of the Cys/Arg branch of the N-degron pathway.

This thesis describes novel PCO enzymes from *Marchantia polymorpha* (MpPCO) and *Klebsormidium nitens* (KnPCO). MpPCO was characterised with an endogenous substrate, MpERF, with homology to the ERF-VII transcription factors. Despite being the only PCO isoform in *M. polymorpha*, MpPCO catalytic activity was relatively insensitive to changes to O_2 concentrations. By contrast, KnPCO had the greatest O_2 sensitivity of any PCO characterised thus far. KnPCO did not have as broad a substrate profile as other PCOs, failing to oxidise LITTLE ZIPPER 2, a known substrate of MpPCO and the higher plant PCOs. This thesis also established algal hypoxia culturing facilities and identified potential hypoxia relevant proteins in the *C. reinhardtii* proteome, which were verified as substrates of the PCOs *in vitro*.

This study takes the first steps towards understanding the relationship between the basal and higher plant O_2 sensing pathways. It characterises a functional PCO with an ERF-VII-related substrate in *M. polymorpha*, and describes the first algal PCO and substrates. It suggests an O_2 sensitive N-degron pathway exists in *K. nitens*, without any ERF-VII, ZPR2 or MC-initiating VRN2 substrates in the *K. nitens* genome. Substrate hunting experiments suggest an algal metabolic response to hypoxia, which could have far reaching consequences for global algal health.

Abbreviations

5-TAMRA; 5-Carboxytetramethylrhodamine

ACC; 1-aminocyclopropane-1-carboxylic acid

ADO; Cysteamine (2-aminoethanethiol) dioxygenase

AP2; APETALA 2

AS; asparagine synthetases

BeStSel; Beta Structure Selection

Blastp; Basic Local Alignment Search Tool, protein

CAPITO; CD Anaylsis and Plotting Tool

CDO; Cysteine dioxygenases

CDPKs; Ca^{2+}-dependent protein kinases

CODD; C-terminal oxygen degradation domain

$CysO_2H$; Cys sulfinic acid

DES; Desaturase

DIC; diisopropylcarbodiimide

DIEA; N,N-diisopropylethylamine

DMF; Dimethylformamide

DSBH; double-barrelled β-sheet

DSSP; Dictionary of Secondary Structure of Proteins

DTT; Dithiothreitol

E3; Ubiquitin protein ligase enzymes

EDTA; Ethylenediaminetetraacetic acid

ERF; ETHYLENE RESPONSE FACTOR

FA; Formic acid

FIRe; Fluorescence Induction and Relaxation

Fmoc; Fluorenylmethyloxycarbonyl

FPLC; Fast Protein Liquid Chromatography

GC; growth curve

GFAT; glutamine-fructose-6-phosphate transaminases

GOE; Great Oxidation Event

Hb; Haemoglobins

HD-ZIP III; homeodomain leucine zipper

HEPES; 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

hiPCO; Hypoxia-induced PCO

HRA 1; HYPOXIA RESPONSE ATTENUATOR 1

HRE 1 and 2; HYPOXIA-RESPONSIVE ERFs 1 and 2

HRG; Hypoxia-responsive genes

HRP; Horseradish peroxidase

IAM; 2-Iodoacetamide

ICP-MS; Inductively coupled plasma mass spectrometry

IgG; Immunoglobulin G

IL-32; Interleukin-32

IPTG; Isopropyl β-d-1-thiogalactopyranoside

I-TASSER; Iterative Threading ASSEmbly Refinement

LIC; Ligation independent cloning

LOES; Low-O_2 escape syndrome

LOMETS; LOcal MEta-Threading-Server

LOQS; Low-O_2 quiescence syndrome

LRP; Lateral root primordia

MAFFT; Multiple alignment using fast Fourier transform

MAPKs; Mitogen-activated protein kinases

MDO; Mercaptopropionate Dioxygenases

MEGA; Molecular Evolutionary Genetics Analysis

metaT; Metatranscriptomic

MicroPIXE; Particle-induced X-ray emission

MRE; mean residue ellipticity

MS; Mass spectrometry

MSDO; Mercaptosuccinate Dioxygenases

NEM; N-ethyl maleimide

nhiPCO; Non-hypoxia-induced PCO

NODD; N-terminal oxygen degradation domain

PCO; Plant cysteine oxidase

PCR; Polymerase chain reaction

PEG; Polyethylene glycol

PFK; Phosphofructokinase

PFP; Pyrophosphate fructose-6-phosphate-1-phophotransferase

PHD 1-3; prolyl hydroxylases 1-3

PHYRE2; Protein Homology/analogY Recognition Engine V 2.0

PMSF; phenylmethylsulfonyl fluoride

QTOF MS; Quadrupole Time-of-flight Mass Spectrometry

RCF; Relative centrifugal field

ROS; Reactive oxygen species

SDS-PAGE; Sodium dodecyl sulphate–polyacrylamide gel electrophoresis

SEC-MALS; size exclusion chromatography with multi-angle light scattering

SK 1 and 2; SNORKEL 1 and SNORKEL 2

SOP; Standard operating procedure

SOW; synthetic ocean water

SW; seawater

TCEP; Tris(2-carboxyethyl)phosphine

TFA; Trifluoroacetic acid

TIPS; triisopropylsilane

Trx; Thioredoxin

TSP; 3-(tri-methylsilyl)-propanoic

UPLC; Ultra-performance liquid chromatography

VRN2; VERNALISATION 2

WT; Wild-type

ZPR2; LITTLE ZIPPER 2

Table of Contents

Acknowledgements iv
Declaration vi
Abstract vii
Abbreviations viii

Chapter 1: Introduction 7

1.1 Plants, O_2 and signalling 7
1.1.1 Chronic hypoxia and development 8
1.1.2 Integrated signalling and acute hypoxia 10
1.1.3 Phenotypic responses to acute hypoxia during flooding 14

1.2 N-Degron Pathway–dependent activation of the plant anaerobic response 17
1.2.1 Transcriptional regulation of the anaerobic response in *Arabidopsis* 17
1.2.2 The N-degron pathways 18
1.2.3 AP2/ERF protein family 20
1.2.4 The putative molecular O_2 sensing pathway in *Arabidopsis thaliana* 21
1.2.5 ADO and the human hypoxia response 24

1.1 Cysteine Dioxygenases 26
1.3.1 Cysteine dioxygenases, structure and catalytic mechanism 26
1.3.2 The Plant Cysteine Oxidases structure and homology 29
1.3.3 Plant cysteine oxidases as O_2 sensors 33

1.2 The evolution of O_2 sensing 34
1.4.1 O_2 evolution on Earth 34
1.4.2 Modern oceans and O_2 sensing 36
1.4.3 Conservation of PCO and ADO-like sequences in the oceans 39
1.4.4 PCO sequence phylogeny 44
1.4.5 Thesis objectives 46

Chapter 2: Materials and Methods 49

2.1 Protein expression, purification and characterisation 49
2.1.1 Plasmid preparation and characterisation 49
2.1.2 Site-directed mutagenesis 51

2.1.3 Transformation 52
2.1.4 Culturing 53
2.1.5 Harvesting cells by centrifugation 54
2.1.6 Lysate preparation 55
2.1.7 Ni(II) affinity chromatography-5 mL HisTrap™ HP 55
2.1.8 His_6-tag removal 56
2.1.9 Gel filtration chromatography - HiLoad® 26/600 Superdex® 75 pg 57
2.1.10 Exact protein mass analysis by mass spectrometry 58
2.1.11 Determining protein activity 58
2.1.12 MS-MS 59
2.1.13 Determining the metal content of recombinant protein 61
2.2 Peptide synthesis, purification and characterisation 62
2.2.1 Solid phase peptide synthesis 62
2.2.2 Peptide purification 65
2.2.3 Peptide quantification *via* NMR 65
2.3 Steady-state kinetic analysis 66
2.3.1 Agilent RapidFire Q-TOF analysis of peptide oxidation 66
2.3.2 Determining PCO co-factor and pH dependence 66
2.3.3 Michaelis-Menten substrate kinetics 67
2.3.4 Michaelis-Menten O_2 kinetics 68
2.4 KnPCO and MpPCO structure and stability analysis 69
2.4.1 Thermal stability of recombinant protein using SYPRO™ Orange 69
2.4.2 Thermal stability and aggregation properties of MpPCO *via* the Nanotemper Prometheus NT.48 69
2.4.3 Size Exclusion Chromatography - Multi-Angle Light Scattering (SEC-MALS) 70
2.4.4 Circular dichroism of PCOs 70
2.4.5 Antibody detection of PCOs via dot blot 71
2.4.6 Screening crystallisation conditions for MpPCO and KnPCO 71
2.5 PCO substrate detection in algal lysates 72
2.5.1 Peptide cysteine-sulfinic acid labelling with N-ethyl and reporter maleimide 72
2.5.2 Culturing *C. reinhardtii* for substrate hunting experiments 73
2.5.3 Culturing *K. nitens* for substrate hunting 73
2.5.4 Exposing cells to hypoxia for 6 hours 73

2.5.5 Algal whole cell lysis and Bradford assay for $CysO_2H$ labelling 74
2.5.6 Labelling and detection of $CysO_2H$ moieties in cell lysates using TAMRA-5-maleimide 75
2.5.7 Testing substrates from *C. reinhardtii* and *V. carteri* with recombinant PCOs 76
2.6 Algal culturing under hypoxia and proteomics 76
2.6.1 Culturing *E. huxleyi* 76
2.6.2 Nuclear enriched protein extraction protocol 77
2.6.3 Algal cell lysis by sonication 78
2.6.4 Culturing *K. nitens* in the PhO_2xBox 79

Chapter 3: Biochemical characterisation of *Marchantia polymorpha* plant cysteine oxidase 80
3.1 Introduction 80
3.2 Expression, purification and characterisation of MpPCO 80
3.2.1 Plasmid preparation 80
3.2.2 Expression 82
3.2.3 Purification His_6-MpPCO 82
3.2.4 Purification of MpPCO and hexahistidine-tag cleavage 85
3.2.5 Determining MpPCO activity 86
3.2.6 Determining His_6-MpPCO metal-binding *via* Inductively Coupled Plasma Mass Spectrometry 90
3.3 Endogenous MpPCO substrate discovery and synthesis 91
3.3.2 Identification of *M. polymorpha* Hypoxia Response Element-like (MpERF) 91
3.3.3 Preparation of $MpERF_{2\text{-}15}$ 92
3.3.4 Initial activity of $MpERF_{2\text{-}15}$ 94
3.4 Probing the His_6-MpPCO active site 95
3.4.1 Site-directed mutagenesis of His_6-MpPCO 96
3.4.2 Activity of His_6-MpPCO_D196N and His_6-MpPCO_H184D 96
3.5 Kinetic studies of MpPCO 98
3.5.1 Determining the optimum conditions for MpPCO kinetic analysis .99
3.5.2 Michaelis-Menten kinetics of the His_6-MpPCO/$AtRAP2.12_{2\text{-}15}$ and His_6-MpPCO/$MpERF_{2\text{-}15}$ reactions 103
3.6 Discussion 114

Chapter 4: Biochemical characterisation of putative algal plant cysteine oxidases 118

4.1 Summary 118
4.2 Expression of putative algal PCOs 120
4.2.1 Identification of putative algal PCOs 120
4.2.2 Preparation of algal PCO sequences in the pET-28a expression vector 121
4.2.3 Expression trials of algal PCOs 122
4.2.4 Large scale expression of His_6-KnPCO 132
4.2.5 Purification of His_6-KnPCO 133
4.2.6 Determining homologous plant cysteine oxidase activity of KnPCO 135
4.2.7 Determining His_6-KnPCO metal-binding *via* Inductively Coupled Plasma Mass Spectrometry 137
4.3 Kinetic Studies of KnPCO 138
4.3.1 Determining the optimum conditions for KnPCO kinetic analysis 139
4.3.2 Michaelis-Menten kinetics of the His_6-KnPCO/AtRAP2.12_{2-15} and His_6-KnPCO/MpERF$_{2-15}$ reactions 142
4.3.3 Screening His_6-KnPCO with other known PCO substrates 145
4.4 Discussion 146

Chapter 5: Structural studies of MpPCO and KnPCO 151

5.1 Summary 151
5.2 Crystallisation trials and thermal stability studies of MpPCO and KnPCO 151
5.2.1 Introduction and protein preparation and crystallisation protocol 151
5.2.2 Crystallisation trials: KnPCO 154
5.2.3 Crystallisation trials: MpPCO 161
5.2.4 Thermal stability assays with KnPCO and MpPCO 165
5.3 Structural analyses of PCOs 169
5.3.1 Introduction 169
5.3.2 PCO analysis *via* antibody recognition 169
5.3.3 PCO secondary structure analysis using circular dichroism 173
5.2.3 PCO tertiary structure analysis via I-TASSER and Phyre2 178
Discussion 188

Chapter 6: Novel PCO substrates in algal cell lysates 192
6.1 Summary 192
6.2 N-ethyl maleimide as a cysteine sulfinic acid probe 192
6.2.1 Introduction 192
6.2.2 Preparation of AtRAP2.12$_{2\text{-}15}$CysO_2H 194
6.2.3 Peptide CysO_2H detection with N-ethyl maleimide 195
6.2.4 Increasing AtRAP2.12$_{2\text{-}15}$CysO_2H-NEM formation 198
6.2.5 Peptide CysO_2H detection with TAMRA-5-maleimide 200
6.3 Investigation of N-terminal CysO_2H proteins in Chlamydomonas reinhardtii lysates 203
6.3.1 Introduction 203
6.3.2 Culturing *C. reinhardtii* and exposure to hypoxia 203
6.3.3 Detecting N-terminal cysteine sulfinic acid proteins 205
6.3.4 Activity of PCOs towards *C. reinhardtii* peptides 214
6.4 Detection of N-terminal cysteine sulfinic acid in Klebsormidium nitens lysates 214
6.4.1 Introduction 214
6.4.2 Culturing *K. nitens* for substrate hunting 215
6.4.3 Detecting N-terminal cysteine sulfinic acid proteins 215
6.5 Discussion 217

Chapter 7: The proteome of algae grown under low O_2 220
7.1 Summary 220
7.2 Initial culturing of E. huxleyi and proteomic analysis 221
7.2.1 Introduction 221
7.2.2 Culturing *E. huxleyi* using the Invivo$_2$400 chamber 221
7.2.3 Establishing *E. huxleyi* culturing conditions in a PhO_2xBox 224
7.2.4 Trialling suitable protein extraction protocols for proteomics 225
7.2.5 Proteomics of *E. huxleyi* cells grown at 15 and 20.6 % O_2 229
7.3 Culturing of K. nitens and whole-cell proteomics 233
7.4.1 Introduction 233
7.4.2 Initial culturing of *K. nitens* cells 233
7.4 Discussion 235

Chapter 8: Conclusions and future directions 238

Appendix 247
References 293

Chapter 1: Introduction

1.1 Plants, O_2 and signalling

O_2 is essential to all aerobic life. Plants not only use O_2 as part of mitochondrial respiration, but produce O_2 via photosynthesis. In photosynthesis, O_2 is a product of the light-dependent lysis of H_2O in the chloroplast, converting light energy into chemical energy in the form of glucose and ATP. In addition to its essential role in ATP production *via* mitochondrial oxidative phosphorylation, O_2 is a substrate in many other essential plant pathways. O_2-dependent fatty acid desaturases play a key role in plant metabolism, membrane structural components and signalling molecule precursors [4-7]. O_2 is also an essential component in the synthesis of phytohormones such as ethylene [8], abscisic acid [9], and gibberellins [10]. O_2 is required for reactive oxygen species (ROS) production. While ROS can cause oxidative damage to DNA, lipids and proteins [11], they are also important chemicals in plant development [12] and signalling. Abiotic stresses such as intense light, heat, salinity, cold, or mechanical injury initiates an upregulation of ROS in affected local tissue, while also triggering a wave of ROS production from the affected tissue to the entire plant system [13-15]. Interestingly, an interplay between an autopropagating wave of ROS and a stress-specific signal for heat stress has been demonstrated in *Arabidopsis thaliana* [13].

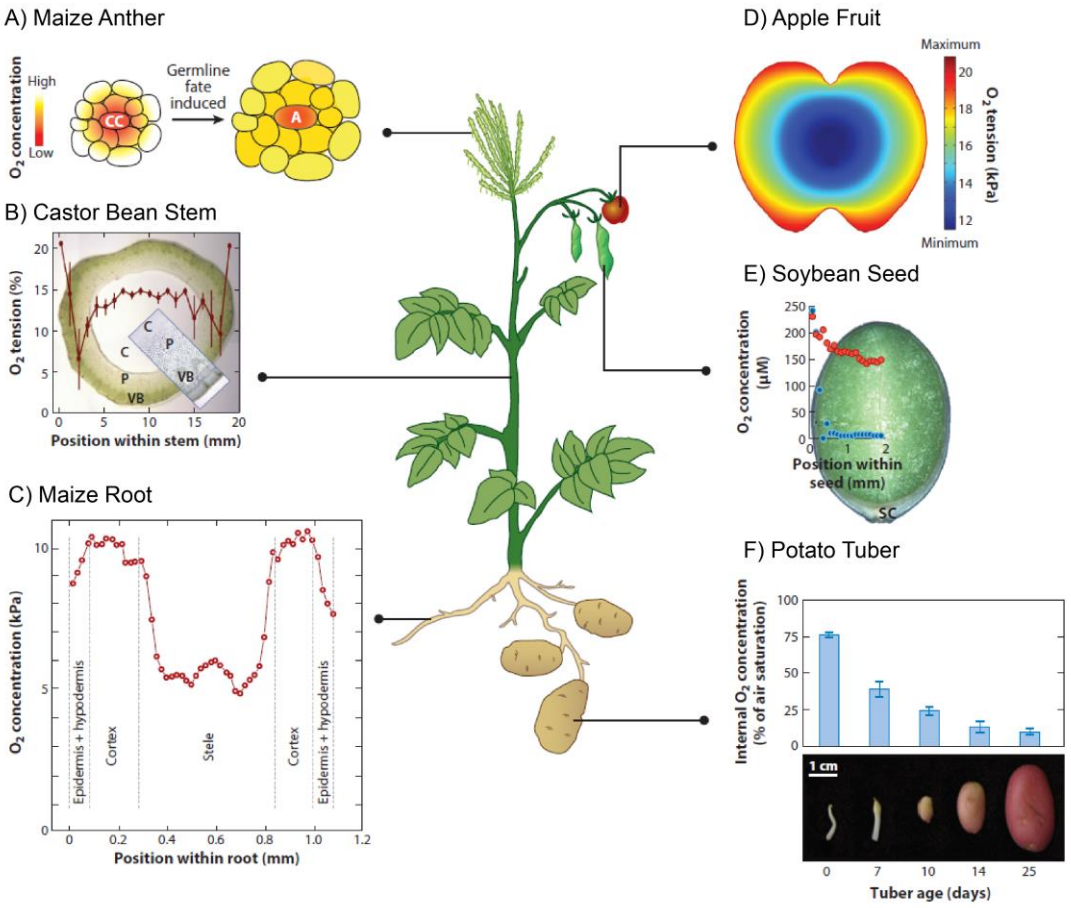


Fig. 1.1: Plant tissue O_2 gradients. Adapted from van Dongen *et al* [2]. Examples of the various O_2 gradients that can exist within plants tissues and throughout development. A) O_2 gradient in maize anthers. Hypoxia is thought to actively induce the differentiation of central cells (CC) into archespores (A). B) O_2 profile through a castor bean stem. The metabolically active cells of the vascular bundle (VB) experience hypoxic conditions. C) O_2 profile across the primary root of maize. Due to the lateral transport of O_2 from the shoot toward root tissues, O_2 concentrations are greater in the cortex. D) A model of the O_2 partial pressure (kPa) along the vertical axis of an apple. E) The distribution of O_2 within soybean seeds measured under light (red circles) or dark (blue circles) conditions. UV-dependent photolysis of H_2O by photosystem II increases O_2 availability. F) Internal O_2 levels in the cores of tubers at different developmental stages, with O_2 concentrations decreasing as tuber size increases.

1.1.1 Chronic hypoxia and development

Plants have no active circulatory system to distribute O_2. Tissue oxygenation therefore relies on passive gas exchange mechanisms such as diffusion [16] and convection [17]. As such, steep O_2 gradients are commonplace in plants (Fig. 1.1).

Hypoxia, or low O_2 availability, is a feature of developing seeds, large fruit [16, 18], roots [19] and bulky storage organs [20]. The partial pressure of atmospheric O_2 is measured at just over 21 kPa (Fig. 1.1, B). Measured O_2 concentrations under 5 kPa have been recorded in the developing seeds of maize, pea, sunflower and bean [21]. *In situ* O_2 profiles of growing normoxic (21% (v/v) *Ricinus communis* stem measured 7% (v/v) O_2 in the vascular region, limiting energy metabolism locally [20]. Despite limiting O_2, these conditions reflect a physiologically normoxic state for these tissues. The literature describes constitutive, homeostatic hypoxia as 'chronic' hypoxia, while any drop in O_2 as a result of increased O_2 consumption or limited gas exchange is called 'acute' hypoxia [2, 22]. Therefore these endogenous chronic hypoxic states can be distinguished from stress-associated acute hypoxia.

Studies have suggested hypoxia is necessary for key plant developmental processes. Hypoxia is thought to be actively induced within maize anthers by the activation of alternative respiration pathways, resulting in the differentiation of central cells (CC) into archespores (A, Fig 1.1, A) [23]. Chronic low O_2 conditions have been identified in the meristems of Arabidopsis and *Solanum lycopersicum,* areas that sustain plant growth and development. Endogenous and synthetic promoters activated by hypoxia were found to be active in both shoot apices and lateral root primordia (LRP) [24, 25]. Results suggest low O_2 conditions are important during the establishment of a root meristem, with hypoxia transcriptional activity maintained in Arabidopsis LRP even when the cells were open to the air [25]. When these tissues were exposed

to hyperoxia, meristem activity actually decreased, further reinforcing hypoxia in the meristem as a favourable state for plant development [24, 25].

O_2 dependent developmental transcriptional regulators were also identified in plants: VERNALISATION 2 (VRN2), LITTLE ZIPPER 2 (ZPR2) [24-26]. ZPR proteins have demonstrated post-translational repression of the class III homeodomain leucine zipper (HD-ZIP III) transcription factors involved in plant development including; leaf patterning, vascularisation, embryogenesis and root architecture [27-29]. VRN2 is involved in a whole host of developmental pathways [30-32], including vernalisation, the epigenetically regulated process by which long-term cold exposure promotes the transition from vegetative to reproductive development [33]. The upregulation of VRN2 and ZPR2 transcription factors relies on a low O_2 environment, with O_2 availability triggering their proteolysis [26, 34] (see Section 1.1.2).

1.1.2 Integrated signalling and acute hypoxia

Acute hypoxia represents a deviation from physiological constitutive hypoxia, to a stressed, adaptive state. Acute hypoxia responses in plants are induced not only by the O_2 concentration *per se* but also by a combination of different factors involved in mediating plant stress responses to low O_2. A number of regulatory pathways interact to form an integrated plant hypoxia signalling and response network, increasing O_2 diffusion to plant tissues, ameliorating ROS damage, regulating anaerobiosis and sustaining redox signalling.

In animals and yeast, mitochondrial ROS function in retrograde signalling during low O_2 stress to adjust energy management [35, 36]. Similarly, a mitochondrial ROS burst is observed upon a drop in O_2 levels in plants [37]. It is suggested that $^{\bullet}O_2^-$ might act as a transient signalling molecule for mitochondrial O_2 deprivation, with resulting $^{\bullet}O_2^-$ converted to H_2O_2 by dismutation [38, 39]. The H_2O_2 produced during this reaction either oxidizes mitochondrial proteins or exits the organelle to induce downstream signalling responses [40]. H_2O_2 has been shown to accumulate in plant tissues during hypoxia and re-oxygenation in the roots of rice, wheat seedlings and rhizomes of iris [41], and in the leaves and roots of barley [42].

As well as the transient mitochondrial oxidative burst, a plasma membrane-localized burst occurs through activation of Ca^{2+}- dependent NADPH oxidases [41 43]. Ca^{2+} has long been associated with plant stress responses [44], with a spike of cytosolic Ca^{2+} levels reported under hypoxia in Arabidopsis seedlings [45]. RESPIRATORY BURST OXIDASE HOMOLOG (RBOH)-dependent ROS production is a well-known generic stress response, with the mechanism of hypoxia-specificity unknown [40]. The physiological implications for H_2O_2 accumulation under hypoxia have begun to be rationalised. Three classes of kinases (MPKs, MPK kinases, and MPK kinase kinases) regulate each other in a cascade-like manner to transmit, amplify, and integrate different signals [46]. H_2O_2 signals were shown to influence Arabidopsis seedling survival under transient O_2 deprivation *via* the activation of the eukaryote-conserved signal

transduction modules – the mitogen-activated protein kinases MPK3, MPK4 and MPK6 [37].

Nitric oxide (NO) has also been identified as a hypoxia-signalling molecule in plants. NO synthase has been reported as absent from land plants, but some algae have been reported as containing sequences with homology to the mammalian NO synthase [47]. An NO synthase from *O. tauri* was shown to catalyse the production of NO *in vivo* and biochemically [48]. In land plants, the main source of NO is the reduction of NO_2^- to NO through a NADH-dependent nitrate reductase enzyme [49, 50]. NO production is upregulated by H_2O_2, and has been shown to promote or antagonize H_2O_2-mediated signalling [51]. The futile NO cycle has been proposed as an alternative pathway to recycle NAD^+ from NADH during hypoxia. Under low O_2 conditions, NO_2^- is reduced to NO by cytosolic nitrate reductase [49, 52], accumulating NO. NO is scavenged and converted back to NO_3^- by plant class-I non symbiotic haemoglobins (Hb) [53, 54], which are also upregulated with hypoxia [55-57]. This cycle requires NADH as a co-substrate and helps balance the antioxidant status of the cytosol [58].

Arabidopsis seedlings overexpressing a class 1 Hb showed increased hypoxia tolerance and promoted early growth. Overexpression of a Hb predicted to have lower affinity for O_2 resulted in reduced stress tolerance [59], indicating the protection demonstrated against hypoxia depends on the ability of the Hb to bind ligands, such as O_2, with high affinity. It has been suggested that non-symbiotic Hbs have O_2 sensing capabilities [60], however this seems at odds with

their generally high affinity for O_2 [61]. The NO cycle is also hypothesised to alleviate fermentation, with plants that produce large quantities of NO upon hypoxia experiencing a decrease in lactate accumulation [62, 63]. NO is also known to inhibit cytochrome c oxidase, and has been shown to positively regulate O_2 concentrations in tissues under hypoxia in mammals [64]. NO is emerging as a ubiquitous regulator of plant development and stress responses by posttranslational nitrosylation [65], and has been implicated in activating downstream Ca^{2+}-modulated signalling [66] and directly affecting developmental processes [67, 68].

Ethylene (C_2H_4) is another signalling component of the hypoxia response, driving some submergence survival strategies in flood-tolerant plants [69]. Flooding reduces O_2 and CO_2 diffusion to plant tissues, while also preventing outward diffusion of evolved C_2H_4, with C_2H_4 content increasing 25-fold within 24 hours of flooding within both deep and non-deep water rice varieties [70]. C_2H_4 production is also upregulated with hypoxia, by the specific increase of its precursor; 1-aminocyclopropane-1-carboxylic acid (ACC). Endogenous ACC is produced by ACC synthase, and converted to C_2H_4 by O_2 - dependent ACC oxidase [71]. Specific genes in the ACC synthase gene family are induced in the roots of tomato [72, 73] and rice [74] under hypoxia. Increased tissue porosity *via* the formation of interconnected air-filled voids (aerenchyma) can facilitate O_2 diffusion to anaerobic tissues [75]. Aerenchyma formation requires C_2H_4 [76], Ca^{2+} [75] and ROS signalling [77] and ultimately cell wall degradation, with hypoxia readily inducing aerenchyma formation in the roots of maize [75] and rice [71].

Interestingly, cross-talk between C_2H_4, Ca^{2+}-dependent protein kinases (CDPKs) (major primary Ca^{2+} sensors in plants) and mitogen-activated protein kinases (MAPKs) has been demonstrated in plant stress responses. Elevated CDPK signalling compromised stress-induced MAPK activation, with inhibition requiring C_2H_4 synthesis and sensing [78].

It has been shown that C_2H_4 positively regulates the expression of many ETHYLENE RESPONSE FACTOR VII (ERF-VII) transcription factors in *Arabidopsis* and rice [79] [80] [81] [70], which are associated with numerous anaerobic response genes [70, 79, 82-84] (See Section 1.2). It has been proposed that C_2H_4 – mediated positive regulation of ERF-VIIs might play an important role in hypoxia tolerance by building up a reservoir of mRNA ERF-VII transcripts, priming the cells for low O_2 [2, 85]. In Arabidopsis, NO has also been shown to negatively regulate ERF VII transcription factors, directly affecting seed germination and seedling establishment [67] (the N-degron pathway is discussed in Section 1.2). NO scavenging by PHYTOGLOBIN 1 was also shown to be increased by C_2H_4 signalling, accumulating ERF-VII transcription factors and pre-adapting Arabidopsis plants to hypoxia stress [86].

1.1.3 Phenotypic responses to acute hypoxia during flooding

Suboptimal O_2 and high C_2H_4 concentrations can occur in roots and aerial organs as a consequence of poor soil drainage, soil compaction or submergence [87-90]. During flooding, microbial and plant metabolic activities quickly consume the remaining O_2 in the soil, leading to hypoxic or anoxic

conditions [91], restricting aerobic respiration [90, 92, 93]. A reduced O_2 supply also affects microbial processes [94], which eventually causes changes in soil chemical composition. Changes in the soil redox potential results in an accumulation of phytotoxic elements including Mn^{2+}, Fe^{2+}, S^{2-} and its conjugate acids H_2S, HS^- [95-97].

Flood tolerant plants have adopted a spectrum of survival strategies to either improve gas exchange or reduce metabolic activity, at the extreme ends of which is the low-O_2 escape syndrome (LOES) and low-O_2 quiescence syndrome (LOQS) [69, 95]. LOES phenotypes include hyponasty (upward bending of leaves) [98], shoot elongation [70], aerenchyma formation [99], growth of barriers to counteract radial O_2 loss in roots due to increased tissue porosity, pressurized gas flow through porous tissues [100] development of adventitious roots [101], formation of gas films on leaf surfaces [102] and modifications to leaf anatomy [103]. LOQS strategies in rice include the negative regulation of genes associated with carbohydrate catabolism and cell elongation, the positive regulation of genes involved in fermentation [104], repressing gibberellin signalling [105] and ameliorating ROS and dehydration effects to facilitate post-flood recovery [106].

Flood tolerance strategies can vary widely, with differences demonstrated between cultivars, ecotypes, even sexes. Studies of poplar and willow trees have indicated that female willow and poplar trees are more tolerant of flooding than males [107, 108]. *Rumex palustris* and *Rumex acetosa* are two related wild dicot species with different flood adaptation strategies, reflecting their

respective hydrological niches [109]. Different rice cultivars can exhibit differential physiological and metabolic responses to hypoxia [70, 83]. Although enzymes regulating fermentative metabolism are ubiquitous in the plant kingdom, this does not necessarily confer flood tolerance. Differential carbohydrate consumption regulation between different rice cultivars can cause different hypoxia tolerances. In a hypoxia tolerant variety, it was demonstrated that fermentation rate was controlled by mutual regulation of the glycolytic enzymes phosphofructokinase (PFK) and pyrophosphate fructose-6-phosphate-1-phophotransferase (PFP). By contrast, an intolerant variety showed no correlation between PFK activity and fermentation rate [110]. Similarly, escape strategies can vary. The escape strategy of deep water rice during shallow, but prolonged floods, is regulated by two VII-ERFs, SNORKEL 1 (*SK1*) and SNORKEL 2 (*SK2*) [70]. During flooding, accumulating C_2H_4 induces the *SK1* and *SK2* expression, triggering internode elongation *via* gibberellin. The application of C_2H_4 along with the putative C_2H_4 receptor inhibitor 1-methylcyclopropene (1-MCP) [111], suppresses internode elongation [70]. The elongated shoot sits above the air water interface, enabling gas exchange. In contrast, lowland *indica* cultivars that face flash floods adopt a quiescence rather than escape strategy mediated by the C_2H_4 upregulation of the VII-ERF SUB1A-1 gene [83], where growth is reduced in order to conserve energy.

1.2 N-Degron Pathway–dependent activation of the plant anaerobic response

1.2.1 Transcriptional regulation of the anaerobic response in *Arabidopsis*

A number of genome-wide studies in *Arabidopsis thaliana* led to the identification of ~50 core hypoxia-responsive genes (HRGs), which are ubiquitously induced upon hypoxia [82, 112] and submergence [87]. These HRGs encode proteins associated with metabolic adaptation to fermentation and include: ATP-DEPENDENT-PHOSPHOFRUCTOKINASE, ALCOHOL DEHYDROGENASE, ALANINE AMINOTRANSFERASE, PYRUVATE DECARBOXYLASE 1, SUCROSE SYNTHASE 1 and PUTATIVE LACTATE TRANSPORTER [113]. Also expressed are HAEMOGLOBIN 1, LOB DOMAIN PROTEIN 41, THE HYPOXIA-RESPONSIVE UNKNOWN PROTEIN and HYPOXIA RESPONSE ATTENUATOR 1 (HRA1). The transcriptional upregulation of many of these Arabidopsis HRGs is due to the stabilization of ERF-VII proteins under hypoxia. There are five ERF-VIIs in Arabidopsis: RAP2.2, RAP2.12 and RAP2.3 are constitutively and ubiquitously expressed in Arabidopsis. By contrast, the HYPOXIA-RESPONSIVE ERFs (HRE 1 and HRE 2) are HRGs [82, 112]. Two core HRGs encode PLANT CYSTEINE OXIDASE (PCO1 and PCO2), key ERF-VII regulatory enzymes (See section 1.2.4) [34]. It has been suggested that HRE1 and HRE2 maintain HRG transcription under prolonged low-O_2 [82] which may be necessary due to RAP2.12 inactivation during hypoxia by direct interaction with the trihelix-domain transcription factor HRA1 [114]. Many of the HRGs identified encode proteins of an unknown function [112].

1.2.2 The N-degron pathways

Targeted protein degradation (proteolysis) has evolved as a mechanism of fundamental importance throughout nature. Targeted proteolysis is important in controlling the levels of regulatory proteins, degrading cytotoxic misfolded or damaged proteins and regulating protein complex formation. A key feature of proteolysis targeting in eukaryotes is ubiquitination. Individual ubiquitin molecules are activated by ubiquitin activating enzymes (E1s) and transferred to ubiquitin conjugating enzymes (E2s). The ubiquitin is transferred to the substrate (target protein), which is recognized by ubiquitin protein ligase 3 enzymes (E3s). The E3 ligases, in concert with the E2s, provide specificity in this system. They usually recognise a specific Lys residue on target substrates, trigger polyubiquitination and ultimately signal for substrate degradation by the ATP-dependent 26S proteasome [115]. The E3-ligases are a class of N-recognins. N-recognins target substrates for degradation by binding to the N-terminal amino acid of the target substrate [116]. Substrates of N-recognins, N-degrons, have an exposed N-terminus, an N-terminal destabilizing residue and a Lys for ubiquitination downstream of the N-terminal residue.

The N-degron pathways describe the half-life of proteins as a function of their N-terminal amino acids, with residues classified as either stabilising or de-

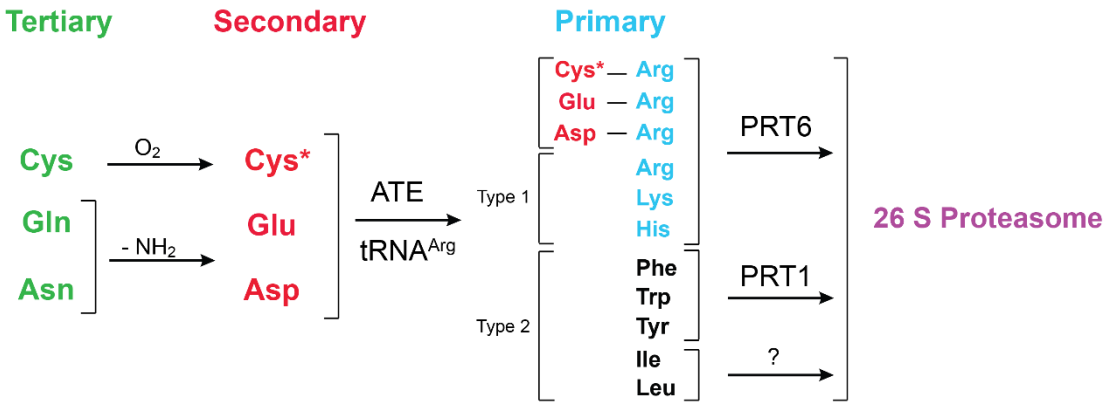


Fig. 1.2: Destabilising amino acid residues of the N-degron pathway. Primary, secondary and tertiary destabilizing residues of the N-degron pathway in plants.

stabilising [117, 118]. In yeast it was found that the N-terminal-acetylated Met residue could act as a primary destabilizing residue, targeted by the Doa10 ubiquitin ligase. Doa10 also recognized the N-terminal-acetylated Ala, Val, Ser, Thr, and Cys residues [119]. On the other hand, the E3 ligases of the Arg/N-end rule pathway directly recognize the α-amino group of the target proteins, with N-terminal acetylation of this α-amino group protecting against degradation [120]. In the classic Eukaryotic N-degron pathway, the amino acids are exposed after the post-translational cleavage of Met by the Met-aminopeptidases [121], or by sequence cleavage by non-processive endoproteases [122, 123]. Primary destabilizing residues have been categorised in two classes; type I basic residues (Arg, His, Lys) and type II hydrophobic residues (Trp, Phe Tyr, Leu, Ile) [124]. In plants, Type I and II residues are recognised directly by the E3 N-recognins; PROTEOLYSIS (PRT) 6 and 1 respectively [125, 126]. As part of the Arg branch of the N-degron pathway there are also secondary and tertiary destabilizing residues. Secondary destabilizing residues (Asp, Glu) are recognised by arginyl-tRNA transferases (ATEs) which transfer Arg from Arg-tRNA to the N-terminal

amino acid. Arginylated Asp and Glu are then primary destabilizing residues, and substrates for the N-recognins. Tertiary destabilizing residues (Cys, Gln, Asn) require post-translational chemical modifications to become secondary destabilising residues, and therefore substrates for N-terminal arginylation [127, 128] (Fig. 1.2). These modifications include, oxidation of Cys or deamidation of Asn or Gln by Asn- or Gln-specific N-terminal amidases to Asp and Glu respectively [129-133]. The Cys/Arg branch of the N-degron pathway for protein degradation is active in both mammals and plants [132, 134].

1.2.3 AP2/ERF protein family

The ERF gene family of transcription factors is part of the AP2/ERF superfamily that contain a single APETALA 2 (AP2) domain [135]. Phylogenetic and motif analysis of the AP2/ERF domain from the 122 ERF transcription factors from *Arabidopsis* divided the family into 12 subgroups. Rice, however, contains 139 ERFs with an additional 3 subgroups [136]. In addition to the 70 amino acid AP2 domain, the group VII ERFs are characterized by a conserved N-terminal motif (NH_2-MCGGAI/L) [137]. The ERF-VII transcription factors are regulated by the Cys/Arg branch of the N-degron pathway, with the O_2 sensitive Cys-2 necessary for N-degron formation [138] (Fig. 1.2). The angiosperm ERF-VII group can be divided into two classes by phylogenetic analysis. In *Arabidopsis*, group 1 proteins RAP2.12, RAP2.2 and HRE1 are thought to have evolved from a single ancestor, whereas group 2 proteins HRE2 and RAP2.3 grouped separately, indicating a second ancestral gene [139]. Not all ERF-VII transcription

factors in plants are substrates for the N-degron pathway. The VII-ERF SUB1A-1 gene [83], a transcription factor in rice which regulates a quiescence survival strategy when submerged, is not a substrate for the N-degron pathway. This is possibly due to the absence of a suitable Lys residue that could be used as a ubiquitin acceptor by an N-recognin [138, 140]. Other studies suggest the SUB1A-1 N-terminus interacts with the C-terminus, burying the Cys-2 necessary for N-degron formation [141].

1.2.4 The putative molecular O_2 sensing pathway in *Arabidopsis thaliana*

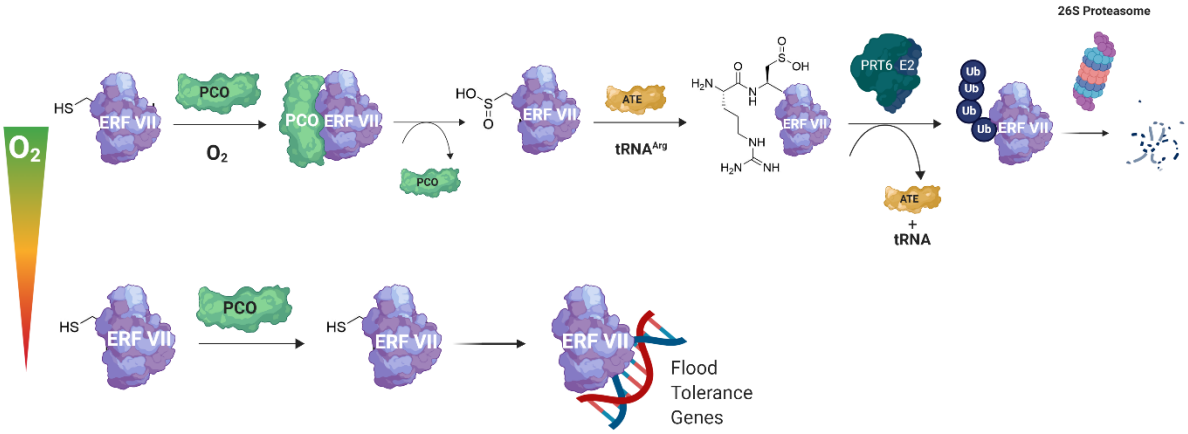


Fig. 1.3: PCO-mediated stability of the ERF-VIIs as part of the N-degron pathway. PCO-mediated regulation of the stability of ERF-VII transcription factors via the N-degron pathway in plants. Created with **BioRender.com**.

Several studies in Arabidopsis and rice have indicated that O_2, C_2H_4 and NO can regulate ERF-VII protein expression and proteolysis. C_2H_4 accumulation has been shown to pre-adapt Arabidopsis to hypoxia stress [86] and positively regulates ERF-VII transcription factors: RAP2.2 [79], RAP2.3 [80], HRE1 [81], and SK 1 and 2 in rice [70]. Low NO levels in *Arabidopsis* have been linked with the stabilization of ERF-VII proteins under aerobic conditions [67, 86]. The stabilization of ERF-VII proteins during normoxia ± an NO scavenger, suggested that the

reintroduction of O_2 following hypoxia is insufficient to destabilize ERFs, also requiring NO to induce ERF-VII proteolysis [67]. However, ERF-VII stability has been directly related to O_2 availability both *in vivo* and biochemically [133, 142].

The Cys/Arg branch of the N-degron pathway for protein degradation relies on the enzymatically catalysed di-oxidation of Cys-2 to Cys-sulfinic acid ($CysO_2H$). In *Arabidopsis*, $CysO_2H$ is essential for forming a functional N-degron for the group ERF-VIIs [67, 133, 138]. $CysO_2H$ renders the ERF-VII as a substrate for arginylation by arginyl transferase 1 and 2 [142]. These two enzymes add an arginine to the N terminus *via* a peptide bond, turning the protein into a substrate for E3 Ubiquitin ligase PRT6 [133], which triggers the covalent addition of ubiquitin molecules. After poly-ubiquitination, the protein is degraded by the 26S proteasome [118] (Fig. 1.3). As such, high O_2 (normoxia) results in negative ERF-VII regulation *via* the N-degron pathway, resulting in the suppression of HRGs. Under normoxia, some RAP2.12 protein was shown to be protected from degradation by binding to ACYL-COA BINDING PROTEIN 1 and 2 (ACBP1 and ACBP2) in the plasma membrane [34, 143]. This is thought to prevent RAP2.12 from moving into the nucleus under aerobic conditions and to protect it against proteasomal degradation in air [34]. As external O_2 levels fall below 10%, the pre-existing membrane bound and *de-novo* expressed RAP2.12 migrates to the nucleus to regulate HRGs [34, 144, 145].

The plant cysteine oxidases (PCOs) help control the lifetime of the ERF-VIIs by catalysing the oxidation of Cys-2 using O_2 as a co-factor. Five PCO isoforms

have been identified in *Arabidopsis*, with PCO1, 2, 4 and 5 shown to regulate the hypoxic response, and plant health, both during and after flooding in *Arabidopsis* [34, 133, 146]. PCO-overexpressing plants showed impaired tolerance to prolonged submergence compared to WT. This sensitivity to hypoxia was evidenced when *Arabidopsis* plants overexpressing PCO1 and PCO2 had impaired hypoxia gene upregulation after treatment with 5% O_2 compared to WT. However under 0% O_2 the plants overexpressing PCO1 and 2 had comparable hypoxia gene upregulation as the WT, indicating O_2 as a necessary co-factor of the PCOs in counteracting the induction of anaerobic genes [133]. A variant ERF-VII, without an N-terminal Cys and therefore unavailable to the PCOs and the N-degron pathway (ERF-$VII_{variant}$), accumulated in the nucleus under normoxia. Under hypoxia, these plants had a significantly faster transcriptional response compared to the WT, with the accumulated ERF-$VII_{variant}$ remaining after the re-introduction of O_2. However, these plants had reduced growth and recovery after flooding compared to the wild-type [34]. Plant health therefore depends on the PCO-mediated balance between the upregulation of HRGs during flooding, as well as the downregulation of these genes during recovery. The PCO-catalysed reaction between O_2 and the N-terminal Cys-2 of ERF-VII, VRN2 and ZPR2 proteins, is believed to be the central O_2 sensing pathway in plants.

1.2.5 ADO and the human hypoxia response

PCOs are classified as thiol dioxygenases (See Section 1.3). There are only two known thiol dioxygenase activities in mammals, cysteine dioxygenases (CDO) and an enzyme previously assigned as cysteamine (2-aminoethanethiol) dioxygenase (ADO) [147, 148]. Recombinant ADO catalysed the oxidation of small molecule cysteamine (a step in cysteamine catabolism), although with poor cysteamine affinity (K_M = 3.8 mM) [148] compared to CDO affinity for L-Cys (K_M = 0.45 mM) [149]. A recent study, however, established human ADO (HsADO) as an O_2 sensitive, N-terminal cysteine dioxygenase that regulates the stability of proteins *via* the N-degron pathway in human cells [150]. ADO catalyses the conversion of N-terminal Cys to $CysO_2H$ in a functionally homologous manner to the PCOs (See Section 1.2.4). While no crystal structure of ADO has been reported, sequence alignment suggests ADO is more PCO-like than CDO-like (See Section 1.3.3). The ADO-catalysed conversion of N-terminal Cys to $CysO_2H$ also demonstrated marked sensitivity to O_2 availability (K_M^{app} O_2 >500 μM).

Hypoxia and its role in human disease has been well documented [151-153]. In humans, HIF mediates the low-O_2 response *via* the regulation of specific hypoxia-response genes. Cellular HIF-1α expression levels are mediated by Fe(II)/2-oxyglutarate dependent HIF-hydroxylases in an O_2 dependent manner [154]: prolyl hydroxylases (PHD 1-3) and an asparaginyl hydroxylase (FIH) [155, 156]. Under normoxia, or normal cellular O_2 levels, the PHD-catalysed *trans*-4-hydroxylation of two prolyl residues targets HIF for proteosomal degradation.

Under hypoxia, HIF-1α accumulates, forms a heterodimer with HIF-β in the nucleus, and promotes HIF target gene expression. As the HIF hydroxylases act as an interface between the cell and O_2, signalling stress and regulating transcriptional adaptation, they have been the target of therapeutic studies for decades [153, 157, 158]. Much like ADO, PHD2, the most important human PHD isoform, has a relatively high sensitivity to changes in O_2 availability [159]. It is this sensitivity that is proposed to make PHD2 biochemically suited as a human hypoxia sensor.

The N-degron pathway is involved a number of human disease pathways, including HIV and Johanson–Blizzard syndrome [116]. HsADO was found to be an essential regulator of Cys/Arg N-degrons, RGS4, RGS5 and Interleukin-32 (IL-32) *in vivo*, in a parallel but distinct mechanism to that of the PHDs and HIF. Hypoxia-inducible [150] RGS4 and RGS5 and have been implicated in O_2 homeostasis in mammals, affecting cardiovascular development and angiogenesis [160-162]. IL-32 is a pro-inflammatory cytokine which has been connected to auto-immune diseases and cancer [163-165]. Interestingly, these HsADO substrates are also targets for Hypoxia-inducible factor (HIF) transcription [166-168]. It has been proposed that under cellular conditions in which both the HIF hydroxylase and ADO pathways are active, they will interact to facilitate physiological responses to hypoxia [150]. This places the thiol dioxygenases as part of a conserved O_2 sensing family across Eukaryota domain.

1.1 Cysteine Dioxygenases

1.3.1 Cysteine dioxygenases, structure and catalytic mechanism

Three types of small molecule thiol dioxygenases have been characterised to date; CDOs, Mercaptopropionate Dioxygenases (MDOs) and Mercaptosuccinate Dioxygenases (MSDOs) [169, 170]. Compared to protein thiol dioxygenases, small molecule thiol dioxygenases are well characterised. Cysteine dioxygenases are ubiquitous, found in bacteria [169], fungi [171], plants [133] and animals [150]. Cysteine dioxygenase (CDO) is a non-heme, Fe(II) and O_2 dependent dioxygenase, catalysing the oxidation of free L-Cys to L-CysO_2H. Cys oxidation is the first step in thiol oxidative breakdown, and has an essential role in cellular Cys maintenance and hypotaurine/taurine biosynthesis pathways [172-174]. CDO expression can be upregulated 45-fold and have its catalytic efficiency modified to suit metabolic need [173, 175].

CDO is part of the cupin superfamily. Despite the very low overall sequence similarity of cupin superfamily members, they all contain two short but partially conserved cupin sequence motifs, Gly...**His**...**His**...**Glu**...Gly (cupin motif 1) and Gly...Asp...Pro...Gly...**His**...Asn (cupin motif 2), separated by a less conserved intermotif region. The residues in bold represent the conserved active site residues. They also share a conserved 6-stranded β-barrel core with the active site generally located at the centre of the barrel [176]. In contrast to the His-His-D/E Fe(II)-binding triad found in most non-heme Fe(II) dependent enzymes, or the His-His-His-Glu metal-coordinating tetrad found in many cupin family

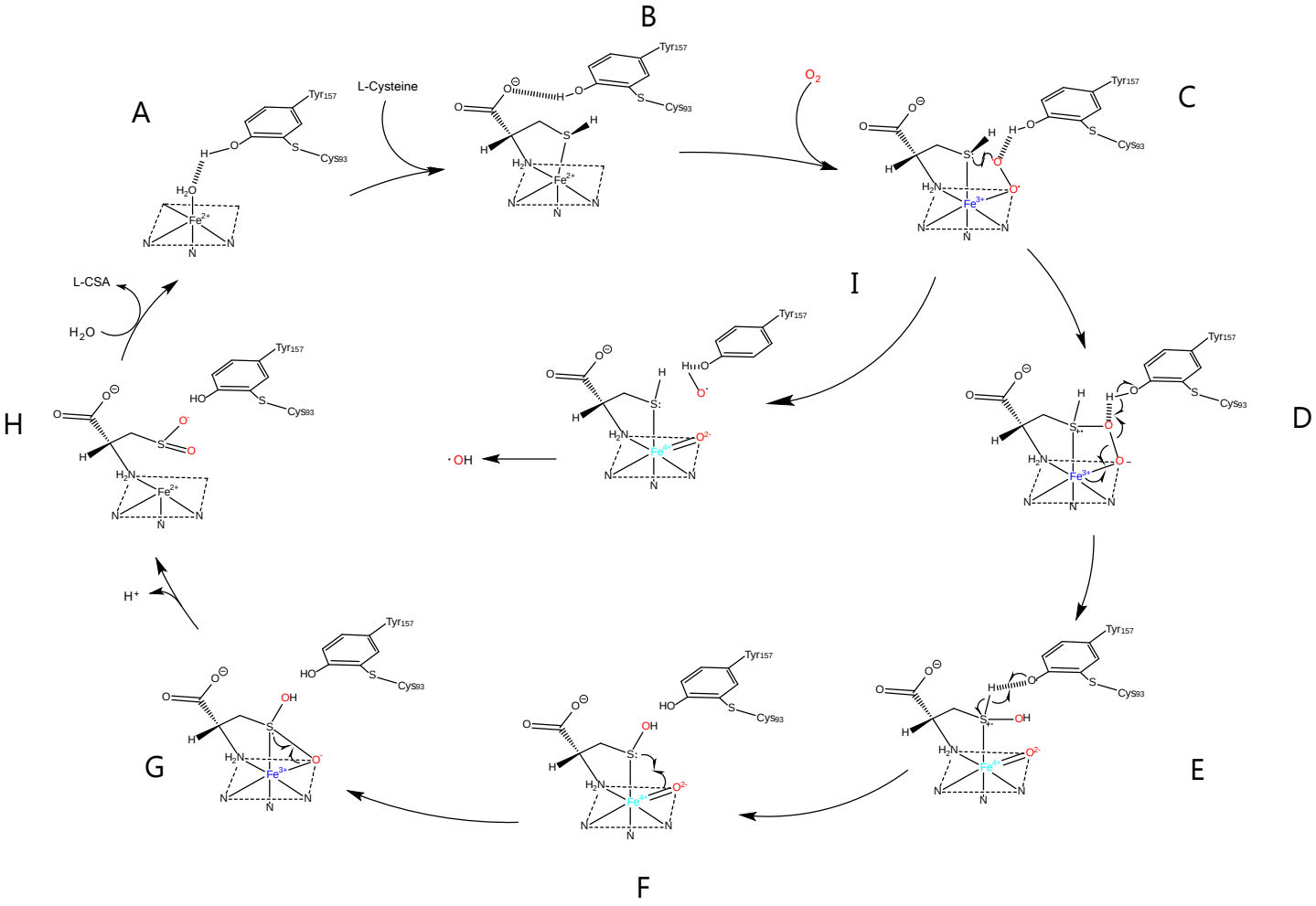


Fig. 1.4: The putative CDO mechanism of free L-Cys oxidation. The CDO catalytic mechanism. A) the His triad and the oxygen atom of a water molecule likely coordinates the Fe in a distorted tetrahedral geometry. B) The water is displaced upon Cys coordination to Fe(II), with Fe(II) binding directly to the thiol and to the amino group. C) The O_2 binds to Fe(II), with a hydrogen bond (H-bond) between the distal oxygen atom and the hydroxyl group on the Tyr157. This coordination shifts to a more favourable pseudo-octahedral coordination geometry. This H-bond favours the Fe(III)-superoxide form of the Fe(O_2) moiety in CDO. Coordinated in this fashion, the thiol group of L-Cys and the distal oxygen atom on the bound O_2 are now in close proximity. Radical coupling occurs between them, forming a peroxo intermediate with a sulfur radical cation. D) Cleavage of the O–O bond (of the O_2 co-factor) then occurs *via* abstraction of a hydrogen atom from the Tyr157 hydroxyl group by the distal oxygen atom 174 to form a tyrosyl radical. E) The tyrosyl radical removes a hydrogen atom from the S-H group. F) This returns the sulfur to a formal S(II) state. G-H) The Fe(IV) moiety is a powerful oxidizing agent, and a single S–O bond is formed as a likely intermediate that is then reductively eliminated from the Fe centre to give CysO_2H (CSA) and Fe(II)

proteins, CDO uses a His triad to coordinate Fe(II) [176, 177]. In the case of eukaryotic CDOs, the highly conserved Glu in cupin motif 1, mentioned above, is substituted by a Cys, which forms a thioether bond with a nearby Tyr residue. In mammals, the Cys93-Tyr157 (*Rattus norvegicus* CDO, RnCDO numbering) thioether cofactor is proposed to be important for maintaining ferrous Fe coordination (and therefore active protein) and reducing damaging radical formation (Fig. 1.4, I). It also credited with maintaining the strained ferrous and water coordination geometry seen in resting mammalian CDOs [178]. However, the lack of conservation of Cys93 in many bacterial CDOs suggests the thioether linkage is not crucial for activity [179]. It has been suggested that the Ser153-His155-Tyr157 H-bonding network is primarily responsible for increasing Try157 acidity [180] (Fig. 1.5).

Numerous studies have culminated in an understanding of the CDO catalytic mechanism [178, 181-184]. In the human CDO (HsCDO) and RnCDO resting states, the His triad and the oxygen atom of a water molecule likely coordinates the Fe in a distorted tetrahedral geometry (Fig. 1.4, A) [178, 180]. The water is displaced upon Cys coordination to Fe(II), with Fe(II) binding directly to the thiol and to the amino group (Fig. 1.4, B). The carboxyl group of L-Cys is H-bonded to several highly conserved residues in the second coordination shell; Arg-60, Tyr-58, Tyr-157 and His-155 [180]. The O_2 binds to Fe(II), with a hydrogen bond (H-bond) between the distal oxygen atom and the hydroxyl group on the Tyr157 (Fig. 1.4, C). This coordination shifts to a more favourable pseudo-octahedral coordination geometry. This H-bond favours the Fe(III)-superoxide form of the

Fe(O_2) moiety in CDO. Coordinated in this fashion, the thiol group of L-Cys and the distal oxygen atom on the bound O_2 are now in close proximity. Radical coupling occurs between them, forming a peroxo intermediate with a sulfur radical cation. Cleavage of the O–O bond (of the O_2 co-factor) then occurs *via* abstraction of a hydrogen atom from the Tyr157 hydroxyl group by the distal oxygen atom [185] to form a tyrosyl radical (Fig. 1.4, D). The tyrosyl radical removes a hydrogen atom from the S-H group (Fig. 1.4, E), returning the sulfur to a formal S(II) state (Fig. 1.4, F). The other electron in the O–O bond is used to form a bond with the Fe centre, thus forming the highly reactive Fe(IV) species (Fig. 1.4, E-F). The Fe(IV) moiety is a powerful oxidizing agent, and a single S–O bond is formed as a likely intermediate that is then reductively eliminated from the Fe centre to give $CysO_2H$ (CSA) and Fe(II) (Fig. 1.4, G-H). Non-crystallographic structural studies of ADO suggest the N-terminal Cys of the protein substrate is coordinated to the metal centre by the thiol group only, in sharp contrast with the dual thiol and amino group coordination at the CDO active site [186].

1.3.2 The Plant Cysteine Oxidases structure and homology

The formation of the ERF-VII N-degron is dependent on the PCO-mediated catalysis of N-terminal Cys to $CysO_2H$ as a function of cellular O_2 availability. The PCOs are Fe(II) dependent dioxygenases that are unique amongst previously described cysteine dioxygenases, and are reported to represent a new class. A study comparing RnCDO with *Arabidopsis thaliana* PCOs 4 and 5

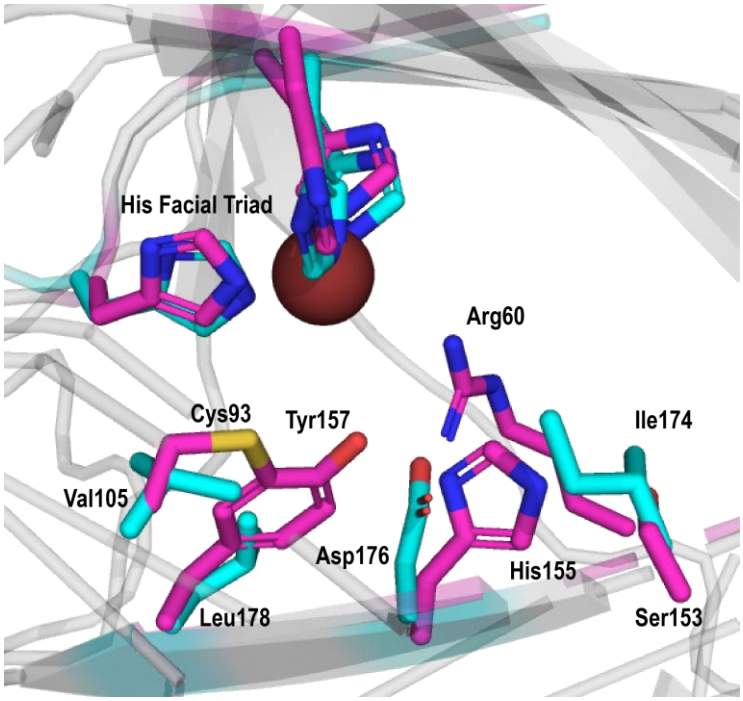


Fig. 1.5: The active sites of AtPCO 5 and RnDCO. AtPCO 4 (blue, PDB: 6S7E) and RnCDO (pink, PDB: 2GH2) active sites overlaid.

(AtPCO 4 and AtPCO 5) identifies distinct structural features that distinguish PCOs from other members of this family [1].

The structures of AtPCO 4 and AtPCO 5 consist of a core double-stranded beta-helix (DSBH) typical of the cupin family. The DSBH supports an active site metal ion involved in catalysis. Resting AtPCO 4 and AtPCO 5 coordinate mononuclear Fe(II) *via* a His triad and three H_2O molecules in an octahedral geometry typical of other metalloenzymes [172], but seemingly in contrast with the distorted tetrahedral geomtery of resting HsCDO and RnCDO [178, 180]. The AtPCOs also have a large active site cavity with the potential to accommodate large polypeptide substrates and a charged hairpin loop close to the active site that may have a role in substrate binding [1]. In contrast, RnCDO has a more restricted active site and no hairpin loop, likely due to the small nature of its

free L-Cys substrate. Similarly, the HsCDO active site was reported to be restricted and sunk in a deep cavity under the protein surface [178].

In previously described thiol dioxygenases, substrate specificity and reactivity are influenced by highly conserved Arg or Gln/Arg residues in the active site [177, 180]. In the CDOs, Arg60 forms a salt bridge with the carboxylate group of the free L-Cys substrate [180]. The PCOs lack this active site Arg, which is likely attributed to the substrate peptide bond that is in place *in lieu* of the charged carboxylate group of free L-Cys. Also, the Ser153-His155-Tyr157-Cys93 motif on DSBH β-strand-8 catalytic triad is not conserved in PCOs (Fig. 1.5). In the AtPCO 4/5 active sites, Ser153, Tyr157 and Cys93 are replaced with Ile174, Leu178 and Val105 at equivalent positions respectively. Therefore a thioether crosslink formation and a catalytic role is unlikely for these particular residues in AtPCOs. AtPCO 4 variants identified active site residues His164 and Asp176 residues as important for catalytic activity. His164 is part of the Fe(II)-binding triad, with a His164Asp variant showing no activity. The Asp176Asn variant had much reduced activity compared to WT, but activity was improved upon Fe(II) and ascorbate addition.

Alignment of AtPCO, HsADO and RnCDO sequences suggest that some of the features involved in PCO substrate binding and activation are conserved in ADO (Fig. 1.6). Like AtPCOs, the RnCDO Ser-His-Tyr catalytic triad, is not conserved in HsADO. Furthermore, AtPCO Asp176 is conserved in ADO, supporting an important role for this residue in PCO/ADO catalysis.

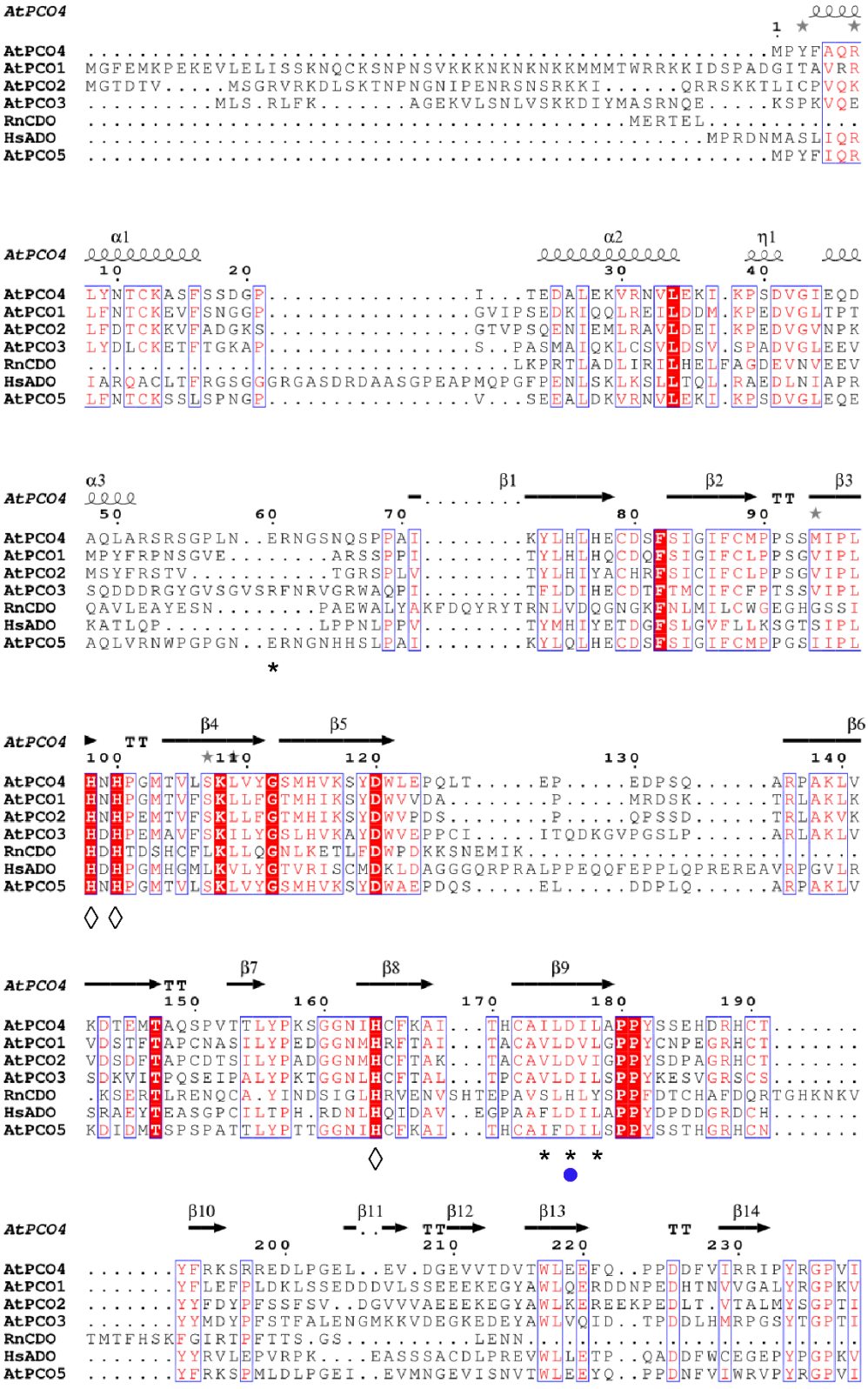


Fig. 1.6: Alignment of AtPCO 1-5, RnCDO and HsADO. Sequences were aligned using Clustal Omega and the secondary structure annotation added using ESPript 3.0. The secondary structure is based on AtPCO 4 (PDB: 6S7E). The His facial triad is denoted by ◊, the Asp176 residue that was important for AtPCO 4 activity is marked by •, and the Ser – His – Tyr CDO catalytic triad by *****. Conserved residues are highlighted in red.

Interestingly, a Cys-Tyr thioether crosslink was observed in ADO between Cys220 and Tyr222 [187]. These residues are conserved with the AtPCOs, but no crosslink was detected.

1.3.3 Plant cysteine oxidases as O_2 sensors

Arabidopsis has five PCO isoforms (AtPCO 1-5). All PCO genes are expressed throughout plant development, while AtPCO 1 and AtPCO 2 were also specifically induced by low O_2 as a result of stress-related hypoxia and at certain developmental stages [133]. Much like the PHDs in humans, the PCOs are kinetically tailored to act as O_2 sensors in plants [188], exhibiting catalytic activity across O_2 concentrations of physiological relevance [20, 188]. AtPCO 4 was found to be the most catalytically competent isoform, with the highest reported k_{cat} with the AtRAP2.12 peptide [188], a known ERF-VII substrate. The AtPCO 4/AtRAP2.12 reaction was also found to have the greatest sensitivity to changes in O_2 concentrations, with the highest reported K_M^{app} O_2. AtPCO 4 also did not demonstarate an AtERF-VII substrate preference, in contrast with the other AtPCO isoforms [188]. As a result of its catalytic competency, sensitivity to O_2 changes and constitutive expression in the cell [133], AtPCO 4 is thought to be the most important of the AtPCO isoforms. The PCO O_2-sensing role [188] suggests that different expression patterns [34], ERF-VII substrate selectivity [142, 188], and catalytic capability may enable the different PCO isoforms to have distinct biological functions. Individual PCOs could therefore be targeted to manipulate ERF-VII levels and improve stress tolerance in plants.

1.2 The evolution of O_2 sensing

1.4.1 O_2 evolution on Earth

For the first half of Earth's history, free O_2 was limited; with concentrations mostly less than 0.001% of current atmospheric levels [189]. The geological record suggests a permanent rise in the concentration of atmospheric O_2 occurred between 2.4 and 2.1 billion years ago [190]. This Great Oxidation Event (GOE) is evidenced by certain geological characteristics such as the appearance of oxidised rusty red soils on land, the disappearance of easily oxidized minerals such as FeS_2 from fluvial systems [191] and redox-sensitive pyrite and uraninite grains [192].

Photosynthesis is by far the dominant source of O_2 on Earth. However, the timing of the first emergence of O_2-producing photosynthesis is disputed [193]. There is also a question of whether or not photosynthesis predates the GOE, or is concurrent with it. There is evidence which suggests O_2-producing photosynthesis existed prior to the GOE, whereby biologically produced O_2 was buffered by reducing compounds produced by early tectonic activity [194, 195]. The concentrations of these reducing agents decreased over time, with O_2 production eventually surpassing reducing agent availability. It is this critical shift that is proposed to have triggered the GOE. However, it is widely accepted that the evolution of O_2-producing cyanobacteria forever altered the chemistry and breadth of life on Earth [191, 196].

It would seem obvious that life before cyanobacteria was largely anaerobic, with anoxic photosynthesis using light to synthesize ATP with reducing power obtained from reducing agents such as Fe^{2+}, H_2S, H_2, $S_2O_3^{2-}$, and S°. However, phylogenetic reconstructions of cytochrome oxidase proteins, which catalyse the reduction of O_2 to H_2O in aerobic respiration, suggest these enzymes predate the evolution of cyanobacteria, and may even predate the split between bacteria and archaea [197, 198].

The molecular transition from anaerobic to aerobic metabolism, where O_2 and sensing is functionally beneficial, is not well understood. Some primitive O_2 scavenging pathways exist that might hold some information about the transition from O_2 cytotoxicity to use in the respiratory chain. As strict anaerobes cannot grow in the presence of greater than 5 µM dissolved O_2 [199], machinery to help cope with fluctuations in O_2 availability are important for maintaining cellular health. Evolutionarily, such machinery would have been important for organism survival while transitioning from an anaerobic to an aerobic world. In strict prokaryotic anaerobes, terminal O_2 reductases in the cytochrome *bd* (quinol oxidases, di-heme site [200]) family have been reported as key in consuming cellular O_2 [201], and in low O_2 conditions even stimulating growth [202]. This machinery might well reflect vestiges of early O_2 use and tolerance, and is present in a variety of other anaerobic bacteria [202]. Cytochrome oxidases have been shown to be homologous to the bacterial NO reductase, which catalyses the conversion of NO to N_2O and H_2O. Interestingly, in oceanic suboxic zones (<10 µmol kg^{-1}), NO_3^- can become involved in

microbial respiration *via* denitrification, which utilises NO reductase [203, 204]. Phylogenetically, NO reductase is an outgroup of all cytochrome oxidases and it is therefore considered likely that NO reductase was present in the last universal common ancestor [197].

There is also consensus that the complexity of Eukaryotic organisms is linked with surface O_2 availability [205, 206]. O_2 is the only respiratory electron acceptor that can meet the metabolic demands required for attaining the large sizes and active lifestyles characteristic of early life [207]. The reduction of O_2 provides the largest free energy release per electron transfer of the terminal oxidants used in biological metabolism. Mitochondrial oxidative phosphorylation harnesses this energy capacity, and with mitochondrial DNA, increases the number of proteins that a cell can evolve, inherit and express by four to six orders of magnitude, potentially paving the way for increased complexity [208]. O_2 availability also supports the synthesis of membrane sterols [209] and structural proteins [210], and the availability of ozone to protect against UV damage.

1.4.2 Modern oceans and O_2 sensing

Due to climate change-driven processes, significant changes in global oceanic conditions have been observed. Ocean warming over the last 50 years [211] is reflected in an expansion of O_2 minimum zones in the central and eastern tropical Atlantic and equatorial Pacific Oceans [212]. Annual variations in O_2 concentrations have been observed in the upper 100 m [213], with further long-term oxygenic changes reported in the subpolar and subtropical regions [214,

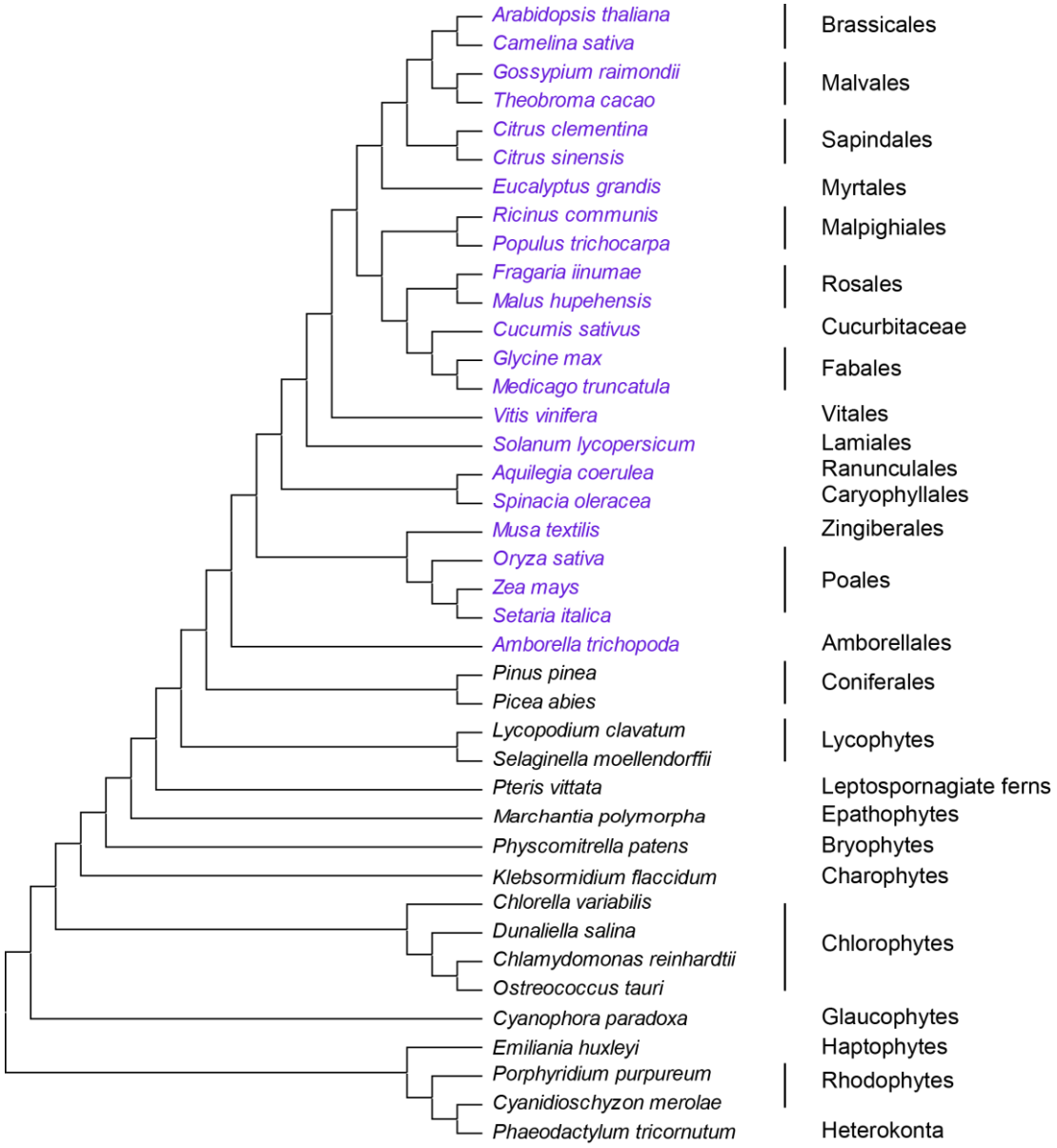


Fig. 1.7: Phylogenetic tree of the plant kingdom. Phylogenetic tree of the plant kingdom using the *rbcl,* gene. All species contain putative PCOs. Those in purple contain homologous ERF-VII proteins. Data for PCO conservation was provided by F. Licausi. ERF-VII conservation was determined using *A. thaliana* ERF-VII sequences as templates for species specific searches using the NCBI protein blast tool.

[215]. Global diatom populations have declined significantly, with a rate of -1.22% yr^{-1} reported between 1998 and 2012. This decline was associated with a significant shallowing of the mixed layer depth (MLD), an increase in photosynthetically available radiation and a decline in NO_3^- [216]. Great regional

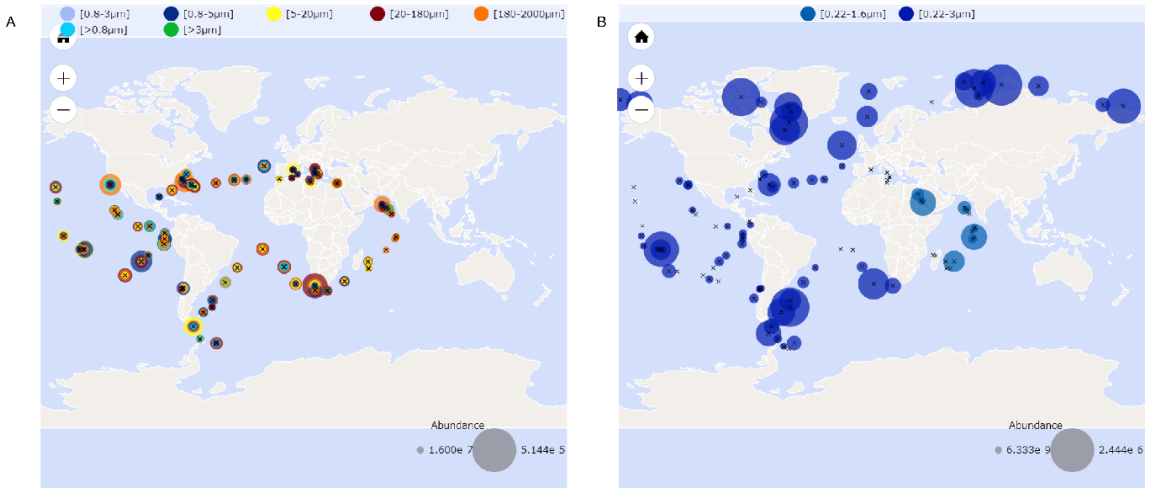
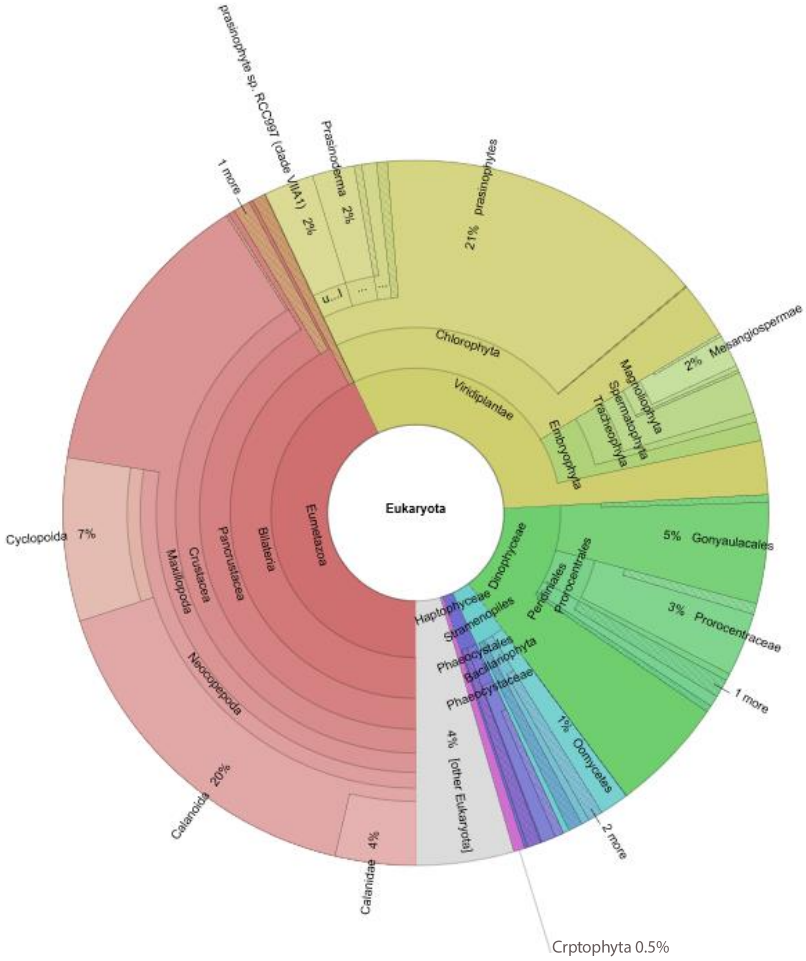


Fig. 1.8: The geographic distribution of PCO-like sequences. The geographic distribution of retrieved PCO and ADO-like sequences from the A) MATOUv1+T and B) OM-RGC.v2+T metatranscriptomes *via* the Tara Ocean Gene Atlas.

phytoplankton population trends were also observed, with a correlation between nutrient loss and phytoplankton decline.

Concern over food security has focussed hypoxia research on higher plants, with little contributed to understanding marine or early land plants. Phytoplankton, however, provide essential oceanic biomass and are central to global nutrient cycles. Furthermore, phytoplankton photosynthesis replenishes atmospheric and oceanic O_2, contributing approximately 50 % of the O_2 in the atmosphere [217]. As such, the impact oceanic deoxygenation will have on algal health, or *vice versa*, must be assessed. Understanding what role, if any, the plant O_2 sensing pathway might have on algal susceptibility to hypoxia will help us improve our understanding of algal responses to climate forcing.

A



B

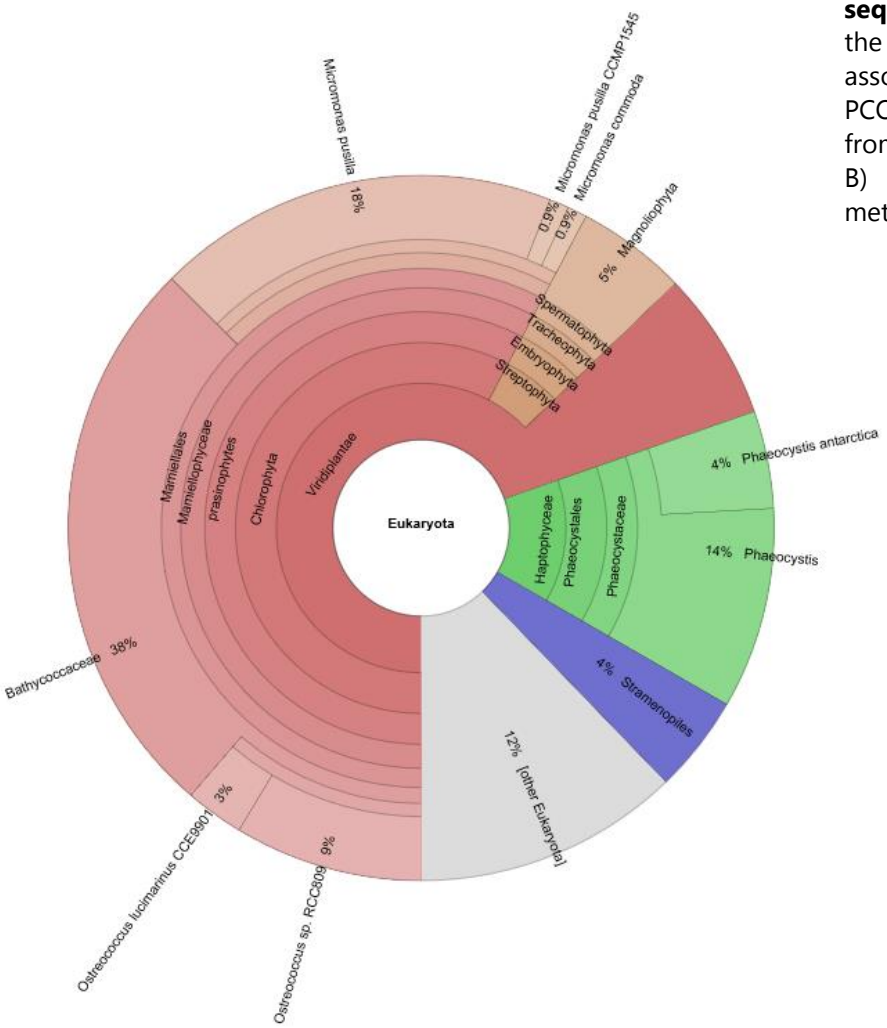


Fig. 1.9: The description of planktonic species associated with the retrieved PCO and ADO-like sequences. Krona charts of the planktonic species associated with the retrieved PCO and ADO-like sequences from the A) MATOUv1+T and B) OM-RGC.v2+T metatranscriptomes.

1.4.3 Conservation of PCO and ADO-like sequences in the oceans

ERF-VII transcription factors are defined by their conserved N-terminal MCGGAI and C-terminal DNA-binding LWSFDD motifs [34]. The ERF-VIIs are limited to higher plants. Homologous transcription factors containing the conserved N-terminal MCGGAI and C-terminal LWSFDD motifs are absent from early land plants and algae. While ERF-VII, ZPR2, and MC-initiating VRN2 substrates are not conserved in basal plants, PCO-like sequences can be found throughout the plant kingdom (Fig. 1.7) [146] [22, 26].

To explore the prevalence of PCO and ADO-like sequences in the oceans, the Tara Oceans Initiative metatranscriptomic (metaT) databases were screened for homologous AtPCO 4 sequences. The Tara Oceans databases contain metagenomic and metaT analysis of samples collected at 189 and 126 stations distributed worldwide respectively (Fig. 1.8) [218] [219]. Two databases exist: the Ocean Microbial Reference Gene Catalog examines the smallest Tara Oceans size fractions (from 0 to 3 µm, OMRGC), and the Marine Atlas of Tara Ocean Unigenes for larger size fractions (0.8 to 2000 µm, MATOU). The OMRGC database includes extended geographic coverage, particularly for the Arctic Ocean. The databases grant access to PCO/ADO sequence global distribution across multiple depth layers of the ocean.

Blastp searches of the OM-RGC.v2+T and MATOUv1+T metatranscriptomic reference gene catalogues *via* the Ocean Gene Atlas [220] database revealed PCO and ADO-like sequences also have a wide oceanic distribution (Fig. 1.8, A and

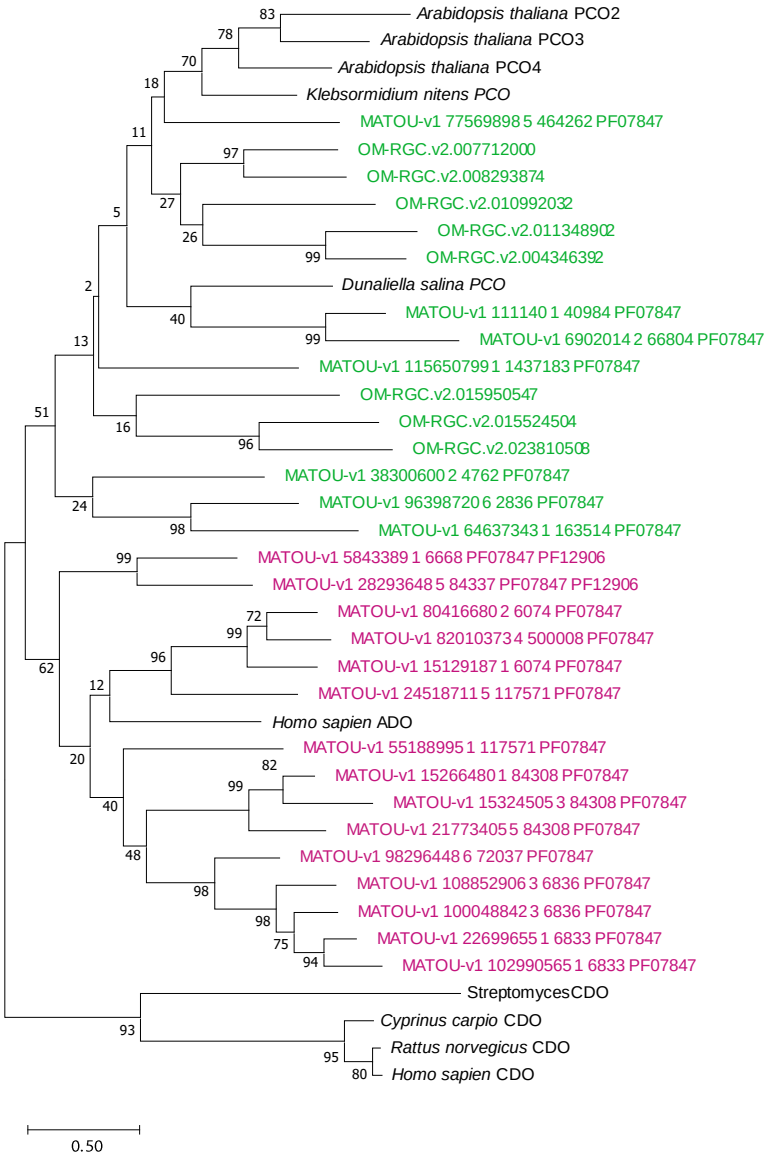


Fig. 1.10: Evolutionary analysis of some of the retrieved Tara Ocean Gene Atlas sequences with homology to AtPCO 4. Evolutionary analysis by the Maximum Likelihood method on MEGA X of a selection of protein sequences homologous to AtPCO 4 from **plants** and **animals** from ocean transcriptomic data.

B). The OM-RGC.v2+T and MATOUv1+T catalogues yielded 28 and 286 sequences homologous to AtPCO 4 respectively. Krona charts illustrate the distribution of these sequences across the animal and plant kingdoms (Fig. 1.9, A and B). Retrieved transcriptomic sequences were scrutinised for redundancy using the CD-HIT clustering package [221, 222] in the MAFFT 7 refinement suite. Sequences below an identity threshold of 0.75 were selected and manually

assessed for PCO conserved motifs. A Krona chart of the redundant sequences illustrates their distribution across Eukaryota (Appendix 1.1).

Phylogenetic and sequence analysis were used to determine if the retrieved sequences were PCO, ADO or CDO-like. The selection criteria included conserved PCO motifs; the HD/EH...H facial His-triad, and the C-terminal DVLAPPY motif. None of the sequences contained the Ser-His-Tyr CDO catalytic domain. Over 60% of the sequences retrieved from the Ocean Gene Atlas database were non-redundant and contained PCO conserved motifs.

To ensure the sequences were more PCO/ADO-like than CDO-like, some representative sequences were selected for phylogenetic analysis. 15 sequences were chosen from a broad range of Opisthokonta and therefore presumed to be ADO-like. 15 sequences were chosen from Viridiplantae, Alveolata and Stramenopiles and presumed to be PCO-like. The sequences were aligned using the MAFFT 7 suite and trimmed in BioEdit. Evolutionary analysis of the sequences was done *via* the Maximum Likelihood method [223, 224] on MEGA X.

Presumed ADO-like sequences grouped with human ADO (Fig. 1.10). Presumed PCO-like sequences grouped with *Arabidopsis* PCO2, 3 and 4, *Klebsormidium nitens* PCO and *Dunaliella salina* PCO (Fig. 1.10). None of the Tara Ocean retrieved metaT or PCO/ADO reference sequences grouped with the CDO sequences. This confirmed the selected sequences were more PCO/ADO-like than CDO-like, suggesting the initial Blast and subsequent

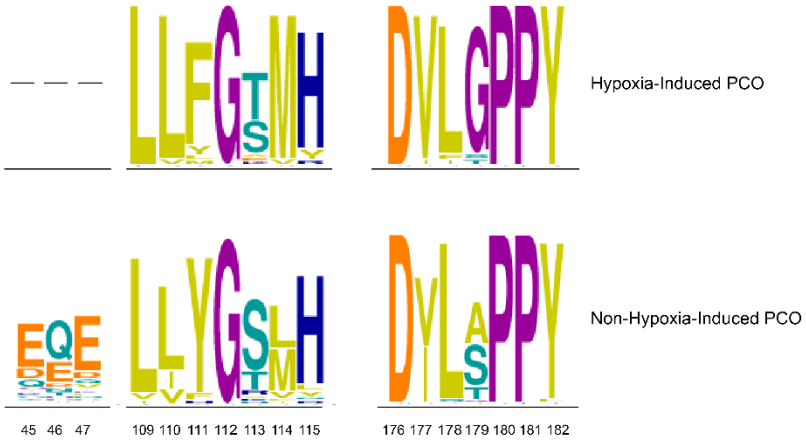


Fig. 1.11: putative hiPCO and nhiPCO motifs. hiPCO and nhiPCO conserved regions as determined by Weits *et al*. Letter height indicates relative abundance. The AtPCO 4 sequence is the reference for amino acid position.

analyses retrieved and selected for N-terminal thiol dioxygenases (PCO/ADO), rather than small molecule thiol dioxygenases (CDO).

The PCO-like sequences are largely from the green-plant lineage (with some Embryophyte contamination noted, Fig. 1.9, B) but with conservation also prevalent for the Haptophytes, Dinophytes and Stramenopiles (Fig. 1.9, A and B). However, less than 10% of metaT data was resolved at the species-level [219]. It is worth noting that a PCO-like sequence in metaT Tara Ocean data was identified as *Ostreococcus lucimarinus* (OM-RGC.v2.011348902, with the retrieved sequence found in the *O. lucimarinus* CCE9901 genome. Overall, the results suggest a wide conservation of PCO-like sequences across diverse algal lineages and geographic distributions.

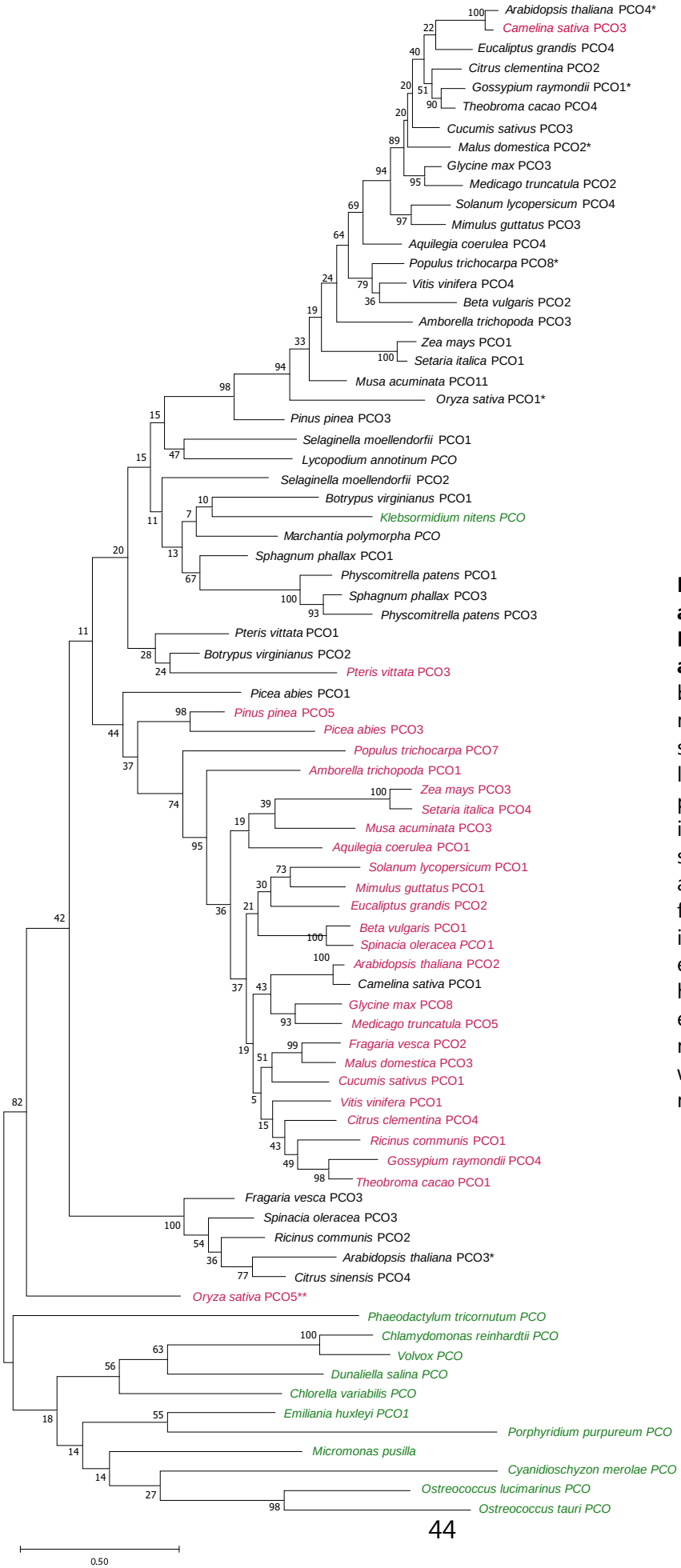


Fig. 1.12: Evolutionary analysis of a selection of PCOs from higher plants and algae. Evolutionary analysis by the Maximum Likelihood method on MEGA X of a selection of higher plant, early land plant and algal PCO protein sequences. Sequences in black are **nhiPCO**, sequences in red are **hiPCO** and sequences in green are from **algae**. Those with a * indicate confirmed experimental nhiPCO or hiPCO. Those with ** are experimentally confirmed nhiPCO or hiPCO, but for which the conserved motif did not correctly predict as such.

1.4.4 PCO sequence phylogeny

To establish phylogenetic relationships between PCOs from different species across the plant kingdom, a phylogenetic tree of putative PCO protein sequences was constructed. PCOs can be constitutively expressed or induced *in vivo* as part of a positive feedback loop during hypoxia [133]. F. Licausi provided PCO sequences of conserved hypoxia-induced PCO (hiPCO) and non-hypoxia-induced PCO (nhiPCO) from across the plant kingdom. Weits *et al* described conserved hiPCO and nhiPCO-specific sequence motifs which correlate well with current experimental data (Fig. 1.11) [146]. Where multiple PCO isoforms were available, both a hiPCO and a nhiPCO were selected where possible. Other proteins with homology to the AtPCO isoforms were identified *via* the NCBI Blastp suite. All sequences used in the phylogenetic tree contain the HD/EH...H facial His-triad, and the conserved C-terminal DVLAPPY motif. These protein sequences were aligned with AtPCO2, 3 and 4 using the MAFFT 7 suite and trimmed in BioEdit.

Evolutionary analysis by the Maximum Likelihood method [223, 224] on MEGA X showed algal PCO protein sequences, with the exception of *Klebsormidium nitens* PCO, group separately from the land plant PCOs (Fig. 1.12). Results indicated *Klebsormidium nitens* PCO and *Marchantia polymorpha* PCO are more land-plant like, grouping with the nhiPCOs. This is in keeping with the finding that *K. nitens* proteins are more land plant-like more than those of

other algae studied [225]. It would therefore be informative to also study a more algal-like PCO.

1.4.5 Thesis objectives

The conservation of PCO-like throughout the plant kingdom sequences suggests the PCOs are important. The paucity of ERF-VII, ZPR2 or MC-initiating VRN2 homologs in algae suggest the PCOs have a function that pre-dates ERF-VII gene regulation. The lack of sequences with homology to the higher plants substrates limits our understanding of the evolutionary origins of O_2 sensing, and the role of PCOs in algae. Undoubtedly, there are challenges associated with living as single-celled organism, and also with transitioning from a single-celled, aquatic species to a land-dwelling, complex, multicellular plant. However, what role the PCOs might play in this transition is largely unknown. The broad distribution of PCOs across the plant kingdom also offers an opportunity to study the evolution of PCOs and their substrates within the context of the many varied plant morphologies and environments.

- This thesis aims to produce novel recombinant algal and early plant PCOs. Biochemical characterisation will focus on determining homologous PCO function and establishing enzyme kinetics using the Michaelis-Menten model.
- This thesis aims to identify, synthesize and characterise previously unknown PCO substrates using a combination of substrate hunting in cell lysates, whole cell proteomics and public database searches.

- Algal cells will be grown under hypoxia, and proteomic analysis performed such that proteomic changes as a function of O_2 availability in algae can be understood.

The models used in this thesis are at important nodes in plant evolutionary history. *Marchantia polymorpha* is a liverwort, one of three lineages of Bryophyte and a basal lineage of land plants (Fig. 1.7). The fossil record suggests the first land plants can be morphologically classified as liverworts [226, 227]. Thus *M. polymorpha* possibly retains features of both its algal ancestors and extant land plants. *M. polymorpha* proteome contains both a PCO and an MC-initiating protein with homology to the ERF-VIIs, and is therefore an important model for the study of the higher plant hypoxia pathway.

Klebsormidium nitens is a single-celled, undifferentiated, Charophytic alga, but has acquired some genes fundamental to terrestrialization [225]. *K. nitens* is reputed to produce several plant hormones and homologues of some of the signalling intermediates required for hormone activity in higher plants. The *K. nitens* genome also encodes a primitive system to protect against high-intensity light damage [225]. Characterising *K. nitens* PCO with endogenous substrates would allow us to study the hypoxia response from the perspective of the water to land transition.

Emiliania huxleyi is a coccolithophoric microalgae (Haptophyta) and is an important model for evolutionary, cellular and physiological studies because haptophytes represent a distinct branch on the eukaryotic tree of life [228]. *E.*

huxleyi is also found in habitats ranging from the equator to the subarctic and can form large-scale episodic blooms under a wide variety of environmental conditions [228]. The genome of *E. huxleyi* is thought to be the key to understanding its global distribution and extraordinary adaptive capabilities [228]. NCBI Blastp searches with known PCOs retrieved two putative PCO isoforms from the *E. huxleyi* CCMP1516 genome. *E. huxleyi's* adaptive capacity, genomic plasticity and putative PCOs make it a powerful model for the study of algal O_2 sensing with respect to evolution and adaptations to global climate change. Proteomic studies of *E. huxleyi* grown under hypoxic conditions would yield valuable insight into the algal molecular hypoxia response, and whether or not the PCO/N-degron pathway plays a part.

Chapter 2: Materials and Methods

2.1 Protein expression, purification and characterisation

2.1.1 Plasmid preparation and characterisation

2.1.1.1 Insertion of DsPCO and MpPCO DNA fragment into a pET28a plasmid

DsPCO and MpPCO amplification primers were synthesised by Sigma Aldrich. The DsPCO and MpPCO DNA fragments were amplified from the template DNA (pDEST17 plasmid) using Q5® High-Fidelity DNA Polymerase (New England Biolabs) and PCR: 35 cycles of 98 °C for 10 s, 65.5 °C (average primer annealing temperature) for 30 s and 72 °C for 1 min. Template DNA was digested using *DpnI* (New England Biolabs) at 37 °C for 3 hours. A 2% agarose gel confirmed that observed DNA fragments corresponded to the MW of the MpPCO (792 bp) and DsPCO (702 bp) inserts. The PCR amplified DNA was purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific).

The pET-28a plasmid (Novagen) was linearized *via* the *NdeI* and *XhoI* cleavage sites at 37 °C overnight. The appropriate ratio of PCR product and DNA insert was calculated using the Molar ratio calculator for In-Fusion Cloning tool (Takara Bio USA). The DsPCO and MpPCO DNA inserts were incubated with the linearized pET-28a vector and the in-fusion HD enzyme (Takara Bio USA) at 50 °C for 15 min, alongside a no enzyme negative control. The reactions were then used to transform *E. coli* XL10 Gold ultracompetent cells (Agilent) according to SOP (Section 2.1.3). Colonies from the plates were selected for overnight growth (Section 2.1.4.2). The DNA was extracted and purified *via* the

GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) and Sanger sequencing. Sanger sequencing of the purified DNA confirmed that the colonies selected from the XL10 Gold cells agar plate had been transformed with the MpPCO-pET-28a and the DsPCO-pET-28a plasmids.

2.1.1.2 pNHTrxT plasmid preparation– Ligation Independent Cloning

Ligation independent cloning (LIC)-compatible EhPCO1 and EhPCO 2 primers were synthesised by Sigma Aldrich. The *E. huxleyi* PCO 1 and PCO 2 DNA fragments were amplified from the template DNA (pDEST17 plasmid) in the presence of Q5® High-Fidelity DNA Polymerase using PCR: 35 cycles of 98 °C for 10 s, 68 °C (average primer annealing temperature) for 30 s and 72 °C for 1 min. Template DNA was digested using *Dpn1* (New England Biolabs) at 37 °C for 3 hours. A 2% agarose gel confirmed that DNA fragments corresponded to the MW of EhPCO 1 (462 bp) and EhPCO 2 (618 bp). The amplified DNA was purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific).

A pNH-TrxT plasmid was generously provided by Dr. K. Lippl (Chemistry Deparment). The plasmid was amplified *via* transformation and overnight growth of *E. coli* XL1-Blue cells (Agilent, Sections 2.1.3 and 2.1.4.2). DNA was extracted and purified *via* the GeneJET Plasmid Miniprep Kit. Sanger sequencing confirmed the vector as pNH-TrxT. The plasmid was linearized *via* *BsaI* at 50 °C for 2.5 hours. A 1% agarose gel confirmed linearization. The plasmid and DNA inserts were amplified separately using T4 DNA polymerase (New England Biolabs) in the presence of BSA and dGTP (vector, New England

Biolabs) and dCTP (DNA inserts, New England Biolabs) for 30 min at RT, and then at 75 °C for 20 min. The T4-treated vector and insert were combined and incubated in ratios of 1:3 1:5 and 1:9 on ice for 30 min before transforming *E. coli* XL10-Gold cells with the vector/insert solutions (Section 2.1.3). Only the 1:9 vector/insert ratio worked for EhPCO2. No colonies were observed for the 1:3 and 1:5 combinations. A 1:3 vector/insert ratio was optimum for EhPCO1, with three colonies observed. Colonies from the plates were selected for overnight growth (Section 2.1.4.2). DNA was extracted and purified *via* the GeneJET Plasmid Miniprep Kit. Sanger sequencing of the purified DNA confirmed EhPCO1 and 2 DNA fragments had been inserted into the pNH-TrxT vector.

2.1.2 Site-directed mutagenesis

MpPCO_H184D and MpPCO_D196N mutated DNA sequences were generated using the QuikChange II site-directed mutagenesis kit (Agilent). Primers were designed according the Agilent SOP, with the desired mutation flanked by 14-19 bp of unmodified nucleotide sequence. The MpPCO pET-28a plasmid was mutated in the presence of forward and reverse primers, dNTPs and *PfuUltra* HF DNA polymerase (Agilent) using PCR: 18 cycles of 95 °C for 30 s, 55 °C for 1 min and 68 °C for 5 min. The first 9 PCR cycles were conducted with forward and reverse primers separate, followed by another 9 cycles with the primers combined. The DNA solution was cooled to 4°C following temperature cycling.

Transformation (Section 2.1.3) of 25 µL of *E. coli* XL10-Blue cells with 2, 5 and 10 ng of the mutated MpPCO DNA yielded colonies, with the *DpnI*-treated pET-28a negative control plasmid yielding no colonies. A series of colonies were selected for overnight growth (Section 2.1.4.2). The DNA was extracted and purified *via* the GeneJET Plasmid Miniprep Kit. Sanger sequencing of MpPCO_H184D and MpPCO_D196N DNA verified the successful mutation.

2.1.3 Transformation

Transformed *E. coli* cells were grown on Luria Bertani (LB) agar made in house. LB agar (0.25 L): Tryptone (2.5g), yeast extract (1.25 g), NaCl (2.5g), agar (5 g) to 0.25 L using ddH_2O. The solution was autoclaved prior to use. S.O.C. medium was provided by the manufacturer with the competent cells. Antibiotic was prepared in ddH_2O and filtered using a 0.22 µm filter prior to combining with LB agar.

All transformations were carried out according to the standard operating procedure outlined by the manufacturer. All cell lines were procured commercially. DNA was incubated with the desired *E. coli* cells for 30 min on ice. Heat shock was carried out at 42 °C for 30 s (XL10-Golds) or 45 s (BL21(DE3) or XL10-Blue) to induce DNA uptake. A tenfold volume of S.O.C. medium was added to the cells followed by incubation on ice for 2 min to allow for cell recovery. The cell/DNA solution was then incubated at 37 °C for 1 hour prior to spreading on an agar plate containing the desired antibiotic (30 µg/mL kanamycin for pET-28a and pNH-TrxT plasmids, with 30 µg/mL ampicillin for

the pDEST™17 plasmid). Agar plates were incubated overnight (approximately 15 hours) at 37 °C.

2.1.4 Culturing

E. coli cells were grown in 2 TY medium made in-house and autoclaved by Mr Herminio Manso Jubier (Chemistry Department). 2 TY Media (1 L): Tryptone (16g), yeast extract (10 g), NaCl (5g) to 1 L using ddH_2O. The solution was autoclaved prior to use. Antibiotic stock was prepared in ddH_2O and filtered using a 0.22 µm filter (final concentration of 30 µg/mL kanamycin for pET-28a and pNH-TrxT plasmids, with 30 µg/mL ampicillin for the pDEST™17 plasmid).

2.1.4.1 Starter and DNA amplification cultures and glycerol stock preparation

100 mL of 2 TY media and 30 µg/mL of antibiotic were inoculated with a single *E. coli* colony transformed with the desired DNA. The culture was grown at 37 °C overnight at 180 rpm. The resulting culture could be used to initiate cultures for protein expression, for the isolation of plasmid DNA or for the preparation of glycerol stocks. To prepare a glycerol stock (such that the transformation step could be avoided for future protein expression experiments), 1 mL of the starter culture was centrifuged and re-suspended in 400 µL of sterile 2 TY medium and 400 µL of sterile glycerol. The glycerol stock was frozen and stored at -80 °C. To initiate a starter culture from a glycerol stock instead of a colony, a loop scraping of the frozen glycerol stock was used to inoculate the culture, and grown at 37 °C overnight at 180 rpm.

2.1.4.2 Small scale protein expression trials

100 mL of 2 TY media and 30 µg/mL of antibiotic were inoculated with 1 mL of the *E. coli* starter culture (Section 2.1.4.2). If the culture required additional FeII, $(NH_4)_2Fe(SO_4)_2.6H_2O$ was added prior to inoculation to a final concentration of 100 µM. The culture was grown at 37 °C at 180 rpm until an optical density at 600 nm (OD_{600}) of 0.8 was reached. Protein expression was induced by adding Isopropyl β-d-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM (prepared in ddH_2O and filtered using a 0.22 µm filter prior to use). Protein expression continued overnight at 15 or 20 °C.

2.1.4.3 Large scale protein expression

600 mL of of 2 TY media and 30 µg/mL of antibiotic were inoculated with 10 mL of the *E. coli* starter culture. The culture was grown at 37 °C and 180 rpm until an OD_{600} of 0.8 was reached. Protein expression was induced by adding Isopropyl β-d-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. Protein expression continued overnight at 18 °C.

2.1.5 Harvesting cells by centrifugation

The cell pellet of small scale protein expression trials and starter cultures (for glycerol stocks) was harvested by centrifuging 1 mL of the culture using a Microfuge 22R (14,000 x g RCF, Beckman Coulter) at 4 °C. The pellet of protein expression trials were stored at -80 °C prior to analysis. The cell pellet of amplified DNA was harvested by centrifuging 10 mL of the culture using an Allegra X-30R (1,800 x g RCF, Beckman Coulter) at 4 °C for 20 min. The cell

pellet of large scale protein expressions was harvested by centrifugation using the Avanti J-25 (18,000 x g RCF, Beckman Coulter) at 4 °C for 8 min. The pellet was stored at -80 °C prior to cell lysis and purification.

2.1.6 Lysate preparation

The cell pellets of small scale cultures were re-suspended in 5 mL BugBuster®, UltraCruz® Protease Inhibitor Cocktail (ChemCruz) (1 tablet/50 mL) and DNAse1 (tip of a small spatula). To ensure maximum protein extraction, the solution was sonicated on ice using a Vibra-Cell™ VCX 150 sonicator fitted with a microtip (Sonics and Materials) at 50% amplitude on a 45 s on, 45 s off-cycle for 4 cycles. The lysed cell solution was centrifuged using a Microfuge 22R (14,000 x g RCF) at 4 °C, and the supernatant placed in a clean Eppendorf.

The cell pellets of large scale cultures were re-suspended in HisTrap™ binding buffer (Section 2.1.7), UltraCruz® Protease Inhibitor Cocktail and DNAse1. The solution was agitated at 4 °C for 1 hour prior to lysis to ensure dissolution of the pellet. The solution was sonicated on ice using on a Vibra-Cell™ VCX 500 sonicator (Sonics and Materials) at 60% amplitude on a 45 s on, 45 s off-cycle for 4 cycles. The supernatant was harvested using an Avanti J-25 (75,600 x g RCF, Beckman Coulter) at 4 °C for 20 min. The supernatant was filtered using a 0.45 µm filter (Sarstedt AG & Co.), to ensure any residual cell debris was removed.

2.1.7 Ni(II) affinity chromatography-5 mL HisTrap™ HP

Binding buffer: 50 mM HEPES, pH 7.5, 0.4 M NaCl, 20 mM imidazole

Elution buffer: 50 mM HEPES, pH 7.5, 0.4 M NaCl, 1 M imidazole

Enzymes were purified using an AKTA Pure Fast Protein Liquid Chromatography (FPLC) System (Cytiva) at 4 °C. The FPLC system was controlled using a PC running Unicorn v. 6.3. All buffers were prepared with ddH_2O, filtered through a 0.2 µm filter and de-gassed for a minimum of 30 min prior to use. A fixed volume of lysate was loaded at a rate of 1 mL/min on to a 5 mL HisTrap™ High Performance column. Once the lysate was loaded, the column was washed with binding buffer until a stable UV signal was reached (tolerance ±10 mAu, 2 min). The protein was eluted stepwise, with a 5% elution buffer wash for 3 column volumes (CV) to remove non-specifically bound proteins. The concentration of elution buffer was then increased linearly for 5 CV, with eluted protein collected in 2 mL fractions. The elution was completed with a 3 CV wash with 100% elution buffer to ensure all bound protein is eluted. The fractions containing pure protein was determined using sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE).

2.1.8 His_6-tag removal

The purest fractions after affinity chromatography were identified *via* SDS-PAGE, and concentrated using a 15 mL 10 kDa Amicon® ultra centrifugal filter (Merck) at 4 °C using an Allegra X-30R (1,800 x g RCF).

2.1.8.1 His_6-tag removal with thrombin protease

The fractions were pooled and incubated with thrombin protease (restriction grade, Novagen) overnight at 4 °C. The affinity purified protein/thrombin

solution was dialysed overnight into HisTrap™ binding buffer. After overnight His_6-tag cleavage, the recombinant protein/thrombin solution was manually purified using a 5 mL HisTrap™ column, separating His_6-tagged protein from His_6-cleaved protein (reverse HisTrap ™).

2.1.8.2 His_6-tag removal with 3C protease

The fractions were pooled and incubated with 2.5 mg of 3C protease (generously provided by Dr. J. Brem, Chemistry Department) overnight at 4 °C. The Ni(II) affinity purified protein/3C solution was dialysed overnight into HisTrap™ binding buffer. After overnight His_6-tag cleavage, the His-tagged protein/3C solution was manually purified using a 5 mL HisTrap™ column, separating His_6-tagged protein from His_6-cleaved protein (reverse HisTrap ™).

2.1.9 Gel filtration chromatography - HiLoad® 26/600 Superdex® 75 pg

Gel filtration buffer: 50 mM HEPES, pH 7.5, 0.4 M NaCl

Enzymes were purified by gel filtration using an AKTA Pure FPLC System at 4 °C. All buffers were prepared with ddH_2O, filtered through a 0.2 µm filter and de-gassed for a minimum of 30 min prior to use. The Ni(II) affinity purified fractions were concentrated using a 15 mL 10 k Amicon® ultra centrifugal filter (Merck) at 4 °C using an Allegra X-30R (1,800 x g RCF). Once a volume of 4 mL was reached, the solution was buffer exchanged to gel filtration buffer using a PD-10 de-salting column (Cytiva) at 4 °C. The buffer exchanged protein solution was injected on to a HighLoad™ 26/600 Superdex™ 75 pg column *via* a 5 mL capillary loop at a rate of 2.6 mL/min. The protein was filtered at a rate

of 2.6 mL/min, with 2 mL fractions collected by UV absorption at values above 10 mAu.

2.1.10 Exact protein mass analysis by mass spectrometry

The exact mass of purified protein was determined using the RapidFire 365 high-throughput MS System and Agilent 6545 Q-TOF system (Agilent) using a pre-set protein analysis method and a C4 cartridge. Protein mass was determined using Qualitative Analysis software v. B.07.00 (Agilent). The isotope signal was deconvoluted *via* the maximum entropy algorithm using a limited isotope m/z range.

2.1.11 Determining protein activity

2.1.11.1 Initial activity assays

Initial activity assays were carried out at 25 °C in a 1.5 mL Eppendorf in 50 mM HEPES, 50 mM NaCl, pH 7.5, with 1 mM tris(2-carboxyethyl)phosphine (TCEP), $Peptide_{2\text{-}15}$ and 0.1 µM enzyme. Temperature control was maintained using the ThermoMixer C (Eppendorf). The reaction was quenched using an equal volume of 1% (v/v) formic acid after 30 min. The samples were analysed using Xevo G2-S Q-TOF UPLC/MS (Section 2.1.12.1).

2.1.11.2 Mass-flow controlled assays – Brooks Instrument

The O_2 dependency of PCO-catalysed turnover was determined *via* atmospherically controlled assays using mass flow controllers (Brooks Instrument, UK). The peptide/buffer/TCEP solution was pre-equilibrated under the desired N_2:O_2 ratio for 10 min in septa sealed glass vials prior to enzyme

addition. Pure enzyme was added using gas tight syringes (Hamilton). Reactions were quenched with an equal volume of 1% (v/v) formic acid after 2 min. Initial assays were analysed using a Xevo G2-S Q-TOF UPLC/MS and MS/MS (Section 2.1.12.1).

2.1.12 MS-MS

2.1.12.1 Xevo G2-S Q-TOF UPLC/MS

UPLC-MS measurements were obtained using an Acquity UPLC system coupled to a Xevo G2-S Q-TOF mass spectrometer (Waters) in positive electrospray mode. 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 using a gradient of 95% deionized water supplement with 0.1% (v/v) formic acid to 95% acetonitrile. Fragmentation spectra of substrate and product peptide ions (MS/MS) were obtained using a targeted approach with a typical collision-induced dissociation energy ramp of 30 to 40 eV. Analysis was carried out with the same source settings, flow rate and column elution conditions as above

2.1.12.2 Nano LC-MS/MS

Nano LC-MS/MS (mass spectrometry) analyses were carried out by Ms. E. Pires (Chemistry Department). Sample peptides were analysed on a NanoAcquity-UPLC system (Waters) connected to an Orbitrap Elite mass spectrometer

(Thermo Fischer Scientific) possessing an EASY-Spray nano-electrospray ion source (Thermo Fischer Scientific). The peptides were trapped on an in-house packed guard column (75 µm i.d. x 20 mm, Acclaim PepMap C18, 3µm, 100 Å) using solvent A (0.1 % Formic Acid in water) at a pressure of 140 bar. The peptides were separated on an EASY-spray Acclaim PepMap® analytical column (75 µm i.d. × 50 mm, RSLC C18, 3 µm, 100 Å) using a linear gradient (length: 100 minutes, 3 % to 60 % solvent B (0.1% formic acid in acetonitrile), flow rate: 200 nL/min). The separated peptides were electrosprayed directly into the mass spectrometer operating in a data-dependent mode using a collision-induced dissociation (CID), higher energy collision dissociation (HCD) and electron-transfer dissociation (ETD) based method. Full scan mass spectrometry spectra (scan range 350-1500 m/z, resolution 120000, automated gain control (AGC) target 1e6, maximum injection time 250 ms) and subsequent CID/HCD and ETD MS/MS spectra (AGC target 5e4, maximum injection time 100 ms) of 10 most intense peaks were acquired in the Orbitrap. CID and HCD fragmentation were performed at 35 % of normalized collision energy and the signal intensity threshold was kept at 500 counts. The analysis was performed with Peaks Studio v. 8.5 (Bioinformatics Solutions Inc.). The raw mass spectrometry file was searched against the respective peptide sequence. The custom Oxi-N-terminal (on Cys, +32.00) was set as variable modification. Precursor mass tolerance was set as 80 ppm. Fragment mass tolerances for HCD were set to 0.8 Da respectively. All spectra were manually validated.

2.1.13 Determining the metal content of recombinant protein

2.1.13.1 Determining recombinant protein metal content via inductively coupled plasma mass spectrometry (ICP-MS)

The metal content of recombinant proteins was determined using ICP-MS by Dr. P. Holdsworth (Earth Sciences Department). Trace Element analyses of the samples were conducted using a PerkinElmer NexION 350D ICP-MS. The instrument was coupled with an Elemental Scientific prepFAST FIAS (Flow Injection Automation System) autosampler, in order to perform flow injection measurements. This methodology was adopted so that the samples could be analysed directly, in order to achieve better detection of the trace elements present and to preclude the standard requirement of diluting them below the 0.2% total dissolved solids (TDS) concentration limit [229]. The micro-flow volume aliquots of the sample solutions were injected over very small (milli-second) time frames, which ensured minimal matrix loading onto the plasma and therefore negligible matrix effects and signal drift.

2.1.13.2 Determining recombinant protein metal content via particle induced X-ray emission with a microfocused beam (microPIXE)

The metal content of His_6-KnPCO and His_6-MpPCO was determined by microPIXE [3] by Dr. P. Woodcock (Biochemistry Department). The appropriate concentration of protein for microPIXE analysis was determined using the ratio of total amino acids per sequence/the number of Met and Cys residues (Fig. 2.1) [3]. Protein samples were buffer exchanged *via* centrifugation using micro Bio-Spin 6 columns (BioRad) to approximately 50 µL final volume. The final concentrations and buffer used are outlined in Table 7.1.

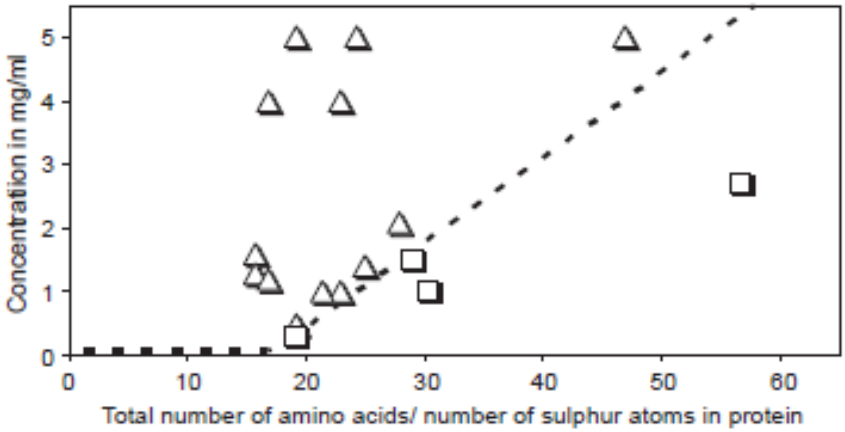


Fig. 2.1: Protein concentration for MicroPIXE analysis. The concentration of protein required for MicroPIXE is dependent on the number of sulphur atoms in the protein sequence. Protein concentrations must be above the line. Image taken from Garman *et al* [3].

2.1.13.3 Determining metal content by Native Mass Spectrometry

Native MS analyses were carried out by Mr. D. Shutin (Chemistry Department). The protein was desalted 3 times using a 5 mL Hi Trap Desalting column (Cytiva) pre-equilibrated with 1M ammonium acetate. The samples were run on a Q Exactive UHMR mass spectrometer (Thermo Scientific). Ions were generated in the positive ion mode from a static nanospray source using gold-coated capillaries prepared in-house as previously described [230]. The instrumental parameters included capillary temperature of 30 °C, spray voltage of 1.2 kV, 0 HCD energy and no in-source trapping to minimize gas-phase activation. Data was processed using MassLynx v.4.1.

2.2 Peptide synthesis, purification and characterisation

2.2.1 Solid phase peptide synthesis

PCO substrates were synthesised in-house, using 14-mer N-terminal peptides *in lieu* of full-length protein substrates. Peptides were synthesised using the

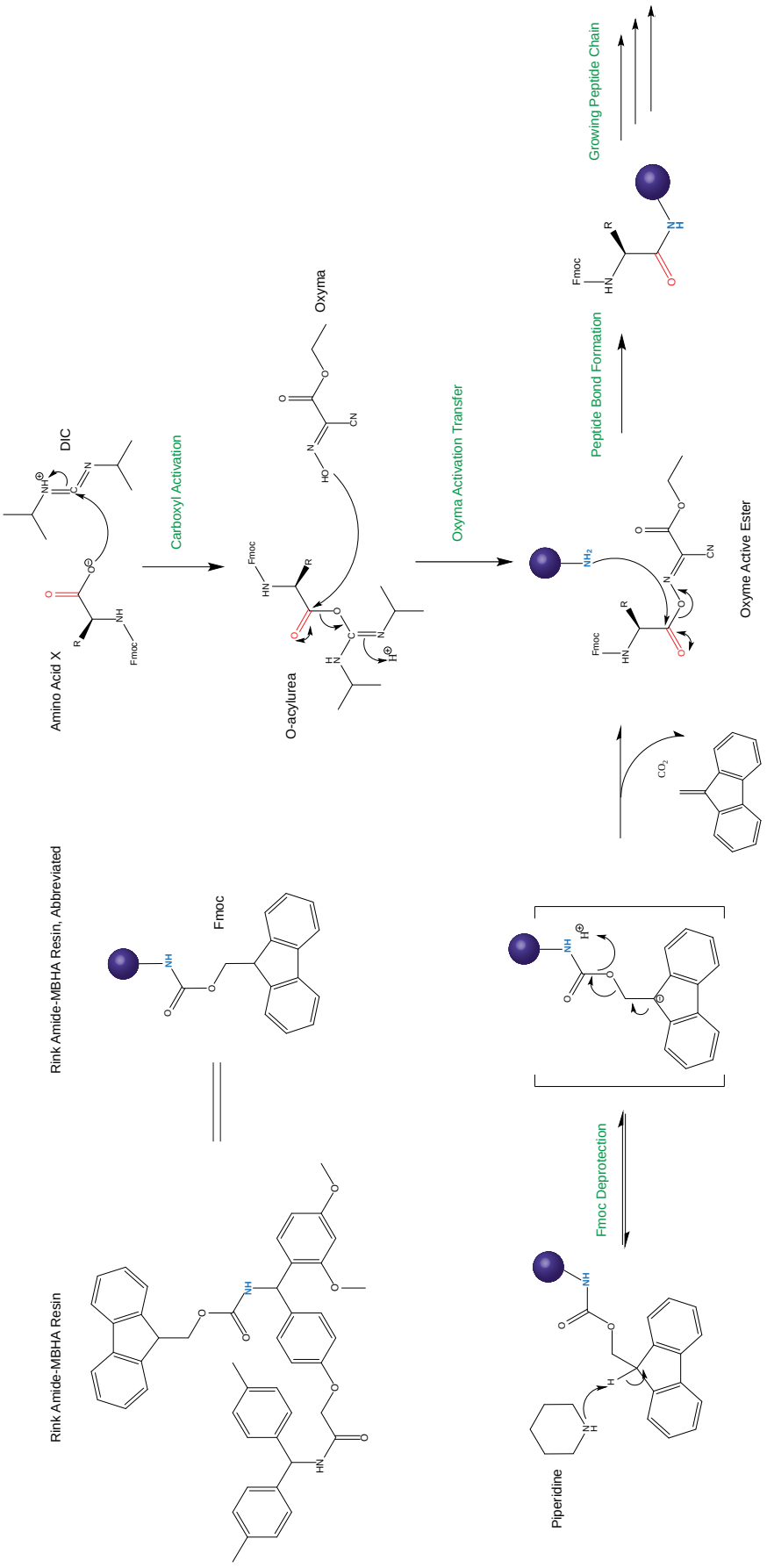


Fig. 2.2: Solid phase peptide synthesis scheme

Liberty Blue™ (CEM) Microwave Peptide Synthesiser, *via* fluorenylmethyloxycarbonyl (Fmoc) solid phase peptide synthesis (Fig. 2.2). All peptides were prepared as C-terminal amides and stored as stock solutions in ddH_2O at −20 °C. The peptides were synthesised on a 0.1 mM scale using a rink amide 4-methylbenzhydrylamine resin (MBHA, AGTC Bioproducts). N-Terminal Fmoc- and side-chain-protected amino acids were weighed-out and solubilised in ultrapure dimethylformamide (DMF) (Cambridge Reagents) prior to synthesis. Amide couplings were facilitated by activating reagents, 0.5 M diisopropylcarbodiimide (DIC) (Fluorochem) and 1 M Oxyma (Novabiochem®) with 0.1 M N,N-diisopropylethylamine (DIEA). Reagent solutions were prepared in ultrapure DMF (Cambridge Reagents). A standard amino coupling was done at 90 °C for 2 minutes. Cysteine and Histidine were coupled at 50 °C for 10 minutes. Addition of arginine to the peptide chain required two coupling reactions at 75°C for 2 minutes. The grown peptide chain was cleaved from the MBHA resin, and its side chain protecting groups removed by incubation in 92.5% (v/v) trifluoroacetic acid (TFA, Sigma Aldrich), 2.5% (v/v) water, 2.5% (v/v) triisopropylsilane (TIPS, Sigma Aldrich), and 2.5% (v/v) 1,3-dimethoxybenzene for 3 hours. The cleaved peptide was precipitated in diethyl ether at – 20 °C. The precipitated peptide was pelleted *via* centrifugation at 1,800 x g RCF for 10 min and the supernatant discarded. The pellet was air dried, and re-suspended in ddH_2O before freezing in LN_2 and freeze drying using the CoolSafe lyophiliser (Scanvac).

2.2.2 Peptide purification

The lyophilised peptide was re-suspended in 2 mL ddH_2O or a ddH_2O/acetonitrile mix, depending on peptide solubility. The peptide solution was purified by reverse-phase high-performance liquid chromatography (HPLC, Jasco) using a Gemini® 5 µm NX-C18 110 Å, 250 x 30 mm, AXIA packed C18 column (Phenomenex). Solvent A: 0.1% (v/v) TFA in ddH_2O, Solvent B: 0.1% (v/v) TFA in acetonitrile. The fractions containing pure peptide were identified by MALDI–TOF MS. Fractions containing pure peptide were pooled, frozen in LN_2, and lyophilised to a powder.

2.2.3 Peptide quantification *via* NMR

The pure peptide was quantified using ^1H NMR spectroscopy *via* a pre-set peptide quantification program on the AVB 500 (Bruker) with 1D-^1H-nuclear Over-hauser effect spectroscopy (1D-NOESY) solvent suppression. A 1:1 (v/v) mixture of a 10 mg/mL solution of TSP and 50 mM peptide stock were dissolved in deuterium oxide (D_2O, Sigma Aldrich). Aromatic or Met residues were used to determine the concentration of the peptide relative to a 3-(tri-methylsilyl)-propanoic acid (TSP) standard (1H chemical shift = 0 ppm). Spectra were generated in TopSpin 3.2 and integrated manually in MestReNova versions 12.0.0 - 14.2.0.

2.3 Steady-state kinetic analysis

2.3.1 Agilent RapidFire Q-TOF analysis of peptide oxidation

Peptide oxidation was analysed using the RapidFire 365 high-throughput MS System and Agilent 6530 Accurate-Mass Quadrupole time-of-flight (Q-TOF) mass spectrometer operated in the positive ion mode (Agilent, Wakefield, MA, USA). Prior to mass spectrometry analysis, the samples were aspirated under vacuum for 400 ms then loaded onto a C-4 solid phase extraction (SPE) cartridge to remove buffer salts from the samples. This was achieved by washing the cartridge with 0.1% (v/v) formic acid in water at a flow rate of 1.5 ml min−1 for 4.5 s and followed with an elution step (85% acetonitrile, 15% deionized water, 0.1% formic acid). The samples were loaded onto the MS at a flow rate of 1.25 mL min^{-1} for 4.5 s. The cartridge was re-equilibrated with water for 500 ms. Peptide oxidation was quantified using Quantitative Analysis software v. B.07.00. The charged ion with the highest intensity was chosen for deconvolution and peptide quantification. All spectra were assessed manually in Qualitative Analysis B.07.00 to ensure the correct ion was chosen for quantification.

2.3.2 Determining PCO co-factor and pH dependence

A combination of real-time and quenched assays were conducted to determine PCO co-factor and pH dependence. 50 mM BisTris propane, 50 mM NaCl buffers at pH 6-9 in 0.5 pH intervals were prepared. Assays were conducted with 4 μM $Peptide_{2\text{-}15}$, 1 mM TCEP, and 0.1 μM enzyme at n =2. Real-time assays were carried out on the RapidFire Q-TOF (Section 2.3.1).

Dithiothreitol (DTT), TCEP and metal screens were conducted using quenched assays. The enzyme reaction in the presence of DTT and TCEP was conducted in triplicate. Temperature control was maintained using the ThermoMixer C. The enzyme and peptide/reducing agent solutions were aliquoted on a 96 well 500 µL round bottomed plate (Agilent). All solutions were incubated at 25 °C for 3 min prior to mixing and initiation of the reaction. 45 µL aliquots were quenched in 5 µL 1% formic acid at regular time intervals.

Reaction conditions for the DTT and TCEP screens included: 50 mM BisTris propane, pH 8, 50 mM NaCl, 100 µM $Peptide_{2\text{-}15}$, and 0.1 µM enzyme. Either 1 mM TCEP or 1 mM DTT were added, with a no-added reducing agent control. Reaction conditions for the metal screens included: 50 mM BisTris propane, pH 8, 50 mM NaCl, 1 mM TCEP, 50 µM $Peptide_{2\text{-}15}$, and 0.1 µM enzyme. 25 – 200 µM of $(NH_4)_2Fe(SO_4)_2.6H_2O$, $ZnCl_2$ or $NiSO_4$ with and without 4 mM sodium ascorbate were incubated with the enzymes to assess PCO metal dependence.

2.3.3 Michaelis-Menten substrate kinetics

The K_M of the PCO substrates were determined using benchtop assays analysed on the RapidFire Q-TOF (Section 2.3.1). All assays were conducted at in triplicate. Temperature control was maintained using the ThermoMixer C. The enzyme and peptide/reducing agent solutions were aliquoted on a 96-well, 500 µL round bottomed plate (Agilent). All solutions were incubated at 25 °C for 3 min prior to mixing and initiation of the reaction. 45 µL aliquots were quenched in 5 µL 1% formic acid at regular time intervals. Reaction conditions

for substrate kinetics included: 50 mM BisTris propane, pH 8, 50 mM NaCl, 1 mM TCEP and 0.1 μM enzyme, with peptide concentrations varied until V_{max} was reached. Peptide oxidation was quantified as outlined in Section 2.3.1. Specific enzyme activity was calculated from the initial rate zone and as a function of active enzyme (μmoles/min/mg of active enzyme). The proportion of active enzyme was inferred from substoichiometric Fe-binding. The specific enzyme activity at different peptide concentrations were plotted against the peptide concentrations and fitted to a standard or inhibition Michaelis-Menten model using GraphPad Prism® v. 5.0.

2.3.4 Michaelis-Menten O_2 kinetics

The K_M^{app} O_2 of PCO enzymes was determined *via* atmospherically controlled assays using mass flow controllers (Brooks Instrument, UK) and analysed on the RapidFire Q-TOF (Section 2.3.1). O_2:N_2 ratios were varied from 0 to 80%. The peptide/buffer/TCEP solution was pre-equilibrated under the desired N_2:O_2 ratio for 10 min in septa sealed glass vials prior to enzyme addition. Temperature control was maintained using a water bath at 25 °C. Enzyme was added using gas tight syringes (Hamilton). Reaction conditions included: 50 mM BisTris propane, pH 8, 50 mM NaCl, 1 mM TCEP and 0.1 μM enzyme, with the concentration of peptide at the point of maximum enzyme activity chosen for the K_M O_2 experiments (using the results of Section 2.3.3). Reactions were quenched with an equal volume of 1% (v/v) formic acid at a time point known to be within the initial rate zone of that peptide concentration. Peptide oxidation was quantified as outlined in Section 2.3.1. Specific enzyme activity

was calculated from the initial rate zone and as a function of active enzyme (μmoles/min/mg of active enzyme). The specific enzyme activity at different peptide concentrations was plotted against the O_2 concentrations and fitted to a standard Michaelis-Menten model using GraphPad Prism® v. 5.0.

2.4 KnPCO and MpPCO structure and stability analysis

2.4.1 Thermal stability of recombinant protein using SYPRO™ Orange

The thermal stability of recombinant proteins was determined using the fluorescence of SYPRO® Orange as outlined by Niesen *et al* [231]. Fluorescence signal as a function of temperature was monitored using the C1000 Touch™ Thermal Cycler (BioRad) in a 96-well format, using the Hard-Shell® PCR plates (BioRad). The temperature was increased from 20°C to 95 °C at a rate of 1 °C/min. The appropriate SYPRO® Orange/protein ratio was determined by combining either 2, 3 or 5 μM of protein with 2, 3, 5 or 10 units of SYPRO® Orange. The dye/protein combination with the optimum fluorescence intensity (high enough to be consistently detected, but not so high that the detector is saturated) was selected. All subsequent analyses were carried out at this dye/protein ratio.

2.4.2 Thermal stability and aggregation properties of MpPCO *via* the Nanotemper Prometheus NT.48

The Prometheus NT.48 (NanoTemper) utilises differential scanning fluorimetry technology. Here, intrinsic tryptophan fluorescence is used to measure protein thermal stability and aggregation status is determined by measuring light intensity loss due to scattering. 10 μM His_6-MpPCO and MpPCO solutions were

analysed at in triplicate. The temperature was ramped from 20 to 95 °C at a rate of 1.6 °C/min.

2.4.3 Size Exclusion Chromatography - Multi-Angle Light Scattering (SEC-MALS)

The dispersity of recombinant proteins was determined using SEC-MALS by Dr. D. Staunton in the Molecular Biophysics Suite (Biochemistry Department). A 1 mg/mL protein solution was submitted for analysis. Protein dispersity was determined using a Dawn HELEOS-II 8-angle light scattering detector and Optilab rEX refractive index monitor (Wyatt Technology) linked to a HPLC system comprising LC-20AD pump, SIL-20A Autosampler and SPD20A UV/Vis detector (Shimadzu).

2.4.4 Circular dichroism of PCOs

Circular dichroism (CD) data was collected on the Chirascan spectrophotometer (Applied Photophysics). Experiments were performed at 23°C in a 0.1 cm pathlength cuvette using approximately 0.2 mg/mL protein in 450 µL 10 mM sodium phosphate buffer, 50 mM $ZnSO_4$, pH 8.0 (all samples were buffer exchanged *via* centrifugation using micro Bio-Spin 6 columns, BioRad). Two scans were collected for each sample and the baseline (buffer) corrected for after each scan. The mean residue ellipticity (MRE) was calculated using the Chirascan software, accounting for protein concentration and the number of amino acid residues.

2.4.5 Antibody detection of PCOs via dot blot

TBST: 0.2M tris 1.5M NaCl, pH 7.4, 1% Tween 20

0.340 pmol of each PCO protein was dotted onto a nitrocellulose blotting membrane (Amersham). After blocking the membrane with 5% milk in TBST, the membrane was incubated with anti-KnPCO (1:1000 dilution in 5% BSA), the primary antibody overnight at 4°C with gentle agitation. After overnight treatment with anti-KnPCO, the membrane was washed for 3 x 10 min with TBST. After washing, the membrane was incubated with anti-Rabbit IgG (H+L), HRP conjugate, 1:3300 dilution in 5% Skim Milk powder in PBS + 0.1 % Tween-20 for 1 hr. After incubation the membrane was washed for 3 x 10 min in TBST. For imaging, the membrane was incubated with ECL™ Prime chemiluminescent substrate (Amersham) for approximately 1 min and imaged on the ChemiDoc XRS+ (BioRad) using the Multichannel, DyeLight 680 and Chemi functions. The exposure time varied between 0.1 and 5 s.

2.4.6 Screening crystallisation conditions for MpPCO and KnPCO

Crystallisation screens and protein samples were dispensed using a Phoenix Crystallization Robot (Art Robbins Instruments). A sitting drop vapour diffusion method was used in 96-well format (Art Robbins Instruments, 96 well low profile Intellipates™), trialling three separate protein/precipitant solution ratios (2:1, 1:1, 1:2) per well. The concentrated protein was centrifuged using a Microfuge 22R (14,000 x g RCF) at 4 °C prior to laying crystallisation screens to remove any potential aggregated protein and particulates to avoid blocking

the robot. The supernatant was carefully removed for use in the robot. The crystallisation screens were checked under a stereo zoom microscope (Leica M205 C). All crystallisation screens were purchased commercially unless stated otherwise.

2.5 PCO substrate detection in algal lysates

2.5.1 Peptide cysteine-sulfinic acid labelling with N-ethyl and reporter maleimide

0.25 mg/mL of AtRAP2.12-CysO_2H and AtRAP2.12-Cys were solubilised in 6 M urea $Na_2PO_4H.2H_2O$/Citric acid at pH 7.4 and incubated with 0.25 mg/mL of free L-Cys and 2.5 mM DTT for 30 min. This step was followed by alkylation with 50 mM IAM, 2% (v/v) SDS and 10 mM EDTA in the dark (black Eppendorfs and lidded Thermomixer) for 1 hour at 25 °C. Excess reagents were removed using ZipTip® (Millipore). The purified solutions were speed Vac-ed for 20 min to remove remaining acetonitrile from the ZipTip® purification.

Each sample was re-solubilized in 6 M urea $Na_2PO_4H.2H_2O$/Citric acid buffer at pH 4.5 and 31.5 mM NEM and incubated overnight at 900 rpm (~19 hours). When labelling peptide CysO_2H with 5-TAMRA-maleimide, the same steps were used above, but instead of NEM, 23 mM 5-TAMRA-maleimide was incubated with the alkylated peptides overnight at 900 rpm in the dark, as the TAMRA molecule is light sensitive. Excess reagents were removed using ZipTip® prior to analysis on the RapidFire Q-TOF (Section 2.3.1).

2.5.2 Culturing *C. reinhardtii* for substrate hunting experiments

HSMA media (1 L): 200 mg/L NaAcetate, 68 µM $CaCl_2.2H_2O$, 80 µM $MgSO_47H_2O$, 8 mM K_2HPO_4, 5.3 mM KH_2PO_4 and 1 mL of Hutner's Trace [232]. Prepared media was filtered using a 0.22 µm filter prior to use. *C. reinhardtii* stationary cultures were grown in HSMA media, at 20 °C on a 12 hour light/dark cycle (light: 1350 lux) under atmospheric conditions. Cultures were initiated by adding 1 mL of a stock *C. reinhardtii* culture to 30 mL of fresh HSMA media. The cell culture flasks were Falcon® 25cm^2 Rectangular Canted Neck Cell Culture Flask with Vented Cap (Corning).

2.5.3 Culturing *K. nitens* for substrate hunting

NIES C media (100 mL): 15 mg $Ca(NO_3)_2.4H_2O$, 10 mg KNO_3, 5 mg β–Na_2glycerophosphate.$5H_2O$, 4 mg $MgSO_47H_2O$, 0.01 µg Vitamin B12, 0.01 µg Biotin, 1 µg Thiamine HCl, 0.3 mL PIV metals [233], 50 mg Tris (hydroxymethyl) aminomethane. Prepared media is filtered using a 0.22 µm filter prior to use. *K. nitens* (NIES 2285) stationary cultures were grown using NIES C media, at 20 °C on a 12 hour light/dark cycle (light: 1350 lux) under atmospheric conditions. Cultures were initiated by adding 1 mL of a stock *K. nitens* culture (in HSMA) to 30 mL of fresh NIES C media at pH 7.5. The cell culture flasks were Falcon® 25cm^2 Rectangular Canted Neck Cell Culture Flask with Vented Cap (Corning).

2.5.4 Exposing cells to hypoxia for 6 hours

15 mL cell cultures were exposed to hypoxia for 6 hours under an atmosphere of 100% N_2 using a mass flow controller. The gas was bubbled through the

culture for the duration of the experiment. Culture temperature was maintained at 20 °C using a N°Ice cooling bucket (Techne). The cells were kept in the dark for the duration of the experiment. As a control, a second 15 mL culture was also kept in the dark, exposed to the atmosphere by a gas permeable membrane. After 6 hours the hypoxia and normoxia-treated cells were pelleted by centrifugation using the Allegra X-30R (1,800 x g RCF, 30 min). This does not create enough force to pellet the cells properly, so the pellet was removed by pipetting, and centrifuged again using the Microfuge 22R at 14,000 x g RCF. The supernatant was gently removed *via* pipetting and discarded. The cell pellets were flash frozen using LN_2 and stored at -80 °C.

2.5.5 Algal whole cell lysis and Bradford assay for CysO_2H labelling

Hypoxia and normoxia-treated cells were resuspended in 50 mM HEPES.NaOH pH 8, 0.1 % triton X-100 and protease inhibitor cocktail (1 tablet/50 mL of buffer). *C. reinhardtii* cells were lysed using a Vibra-Cell™ VCX 150 sonicator fitted with a microtip on a 1 s on, 1 s off cycle, amplitude: 20% for 60s. *K. nitens* cells were lysed using a Vibra-Cell™ VCX 150 sonicator fitted with a microtip on a 1 s on, 1 s off cycle, amplitude: 40% for 280s. All cells were kept on ice during sonication. The lysates was centrifuged for 15 min using the Microfuge 22R at 14,000 x g RCF to remove cell debris. The concentration of protein in the resulting supernatant was determined by Bradford assay (BioRad), by measuring the absorbance at 595 nm using the CLARIOstar Plus plate reader (BMG Technology). The standard curve was generated using BSA standards (Thermo Fisher). All analyses were done in duplicate.

2.5.6 Labelling and detection of $CysO_2H$ moieties in cell lysates using TAMRA-5-maleimide

Bradford assay determined the protein concentration within the normoxia and hypoxia cell lysates. The volume of lysate taken forward for labelling was adjusted such that the same quantity of protein was used from both lysates. Each cell lysate was incubated with 35 µg of recombinant protein and 50 mM NaCl for 80 min at 25 °C to convert any N-terminal Cys to $CysO_2H$. Temperature control was maintained using the ThermoMixer C. The reaction was quenched in 6M urea $Na_2PO_4H.2H_2O$/Citric acid at pH 7.4, and buffer exchanged into the same using an Amicon® Ultra 3k 0.5 mL centrifugal filter (Millipore) on the Microfuge 22R (14,000 x g RCF). After buffer exchange, alkylation and TAMRA-5-maleimide labelling proceeded as described in Section 2.5.1. However, instead of a ZipTip® purification step between the alkylation and maleimide-labelling steps, a chloroform/methanol precipitation was performed.

After overnight incubation with the TAMRA-5-maleimide probe, the labelled lysate was run neat on a 4-12% NuPAGE precast 10 lane gel (Thermo Fisher) in MOPS buffer (Thermo Fisher) at 140 V until the molecular weight ladder was resolved. The gel rig was kept in the dark to maintain the integrity of the fluorescent probe for the duration of the SDS-PAGE procedure. The gel was imaged on the ChemiDoc XRS+ using the UV transilluminator with the 605/50 nm filter (TAMRA excitation maxima 575 nm in 0.1 M phosphate pH 7.0). Any fluorescent bands observed were excised and sent for proteomic analysis and Trypsin digestion with Ms E. Pires (Chemistry Department). The excised gel

bands were digested with Trypsin and analysed by MS according to the protocol outlined in Section 2.1.12.2.

2.5.7 Testing substrates from *C. reinhardtii* and *V. carteri* with recombinant PCOs

14-mer peptides representing the N-termini (residues 2-15) of proteins identified in the hypoxia and normoxia fluorescent band proteomic data were synthesised and purified by GL Biochem Ltd (Shanghai). An MC-initiating protein from *V. carteri* with homology to one of the *C. reinhardtii* proteins was also synthesised. All stocks were solubilised in ddH_2O or ACN depending on solubility. Reaction conditions: 50 µM peptide in 50 mM BisTris propane, pH 8, 50 mM NaCl, 1 mM TCEP and 0.1 µM His_6-KnPCO, His_6-MpPCO and His_6-AtPCO 4, at n = 3 and quenched with an equal volume of 5% (v/v) formic acid. The assays were analysed on the RapidFire Q-TOF (Section 2.3.1).

2.6 Algal culturing under hypoxia and proteomics

2.6.1 Culturing *E. huxleyi*

2.6.1.1 Culturing in the InvivO$_2$400 chamber

Aquil medium is a synthetic ocean water medium, designed for the culturing of marine algae. The full recipe [234, 235] used for the following experiments can be found here:

https://ncma.bigelow.org/PDF%20Files/NCMA%20algal%20medium%20Aquil.pdf

E. huxleyi (RCC 1242) stationary cultures were grown using Aquil media pH 8.2, at approximately 20 °C on a 12 hour light/dark cycle (light: 2700 lux) under

atmospheric conditions. Temperature control was maintained using a N°Ice cooling bucket. Cultures were initiated by adding 1 mL of a stock *E. huxleyi* culture (in Aquil) to 80 mL of fresh Aquil media. The cell culture flasks were Falcon® 75cm^2 Rectangular Canted Neck Cell Culture Flask with Vented Cap (Corning). Cultures were agitated every 2-3 days.

2.6.1.2 Culturing in the PhO2xBox chamber

Synthetic ocean water (SOW) f/2 pH 8.2: f/2 medium prepared as described by Ryther *et al* [236] and Orlowska *et al* [237], with SOW prepared as described by Kester *et al* [238]. Seawater (SW) was collected and aged by the Marine Biological Association. f/2 was prepared as described in the references above.

E. huxleyi (RCC 1242) stationary cultures were grown in either 30 or 80 mL of SW f/2 or SOW f/2 media, pH 8.2, at 15 °C on a 12 hour light/dark cycle (light: 2700 lux). Cultures were initiated by adding 1 mL of a stock *E. huxleyi* culture (in seawater) to 30 or 80 mL of fresh SOW or SW f/2 media. The cell culture flasks were Falcon® 75 cm^2 or 25 cm^2 Rectangular Canted Neck Cell Culture Flask with Vented Cap (Corning). Cultures were agitated every 2-3 days. A plumbed air and N_2 supply was used for controlling the atmosphere within the box. O_2 and CO_2 concentrations were reduced by increasing the N_2 ratio.

2.6.2 Nuclear enriched protein extraction protocol

Protocol provided by Prof J. Mellor (Department of Biochemistry). 80 mL algal cultures were centrifuged using an Allegra X-30R (1,800 x g RCF). The algal pellet was re-suspended in 100 µL of 10 mg/mL Zymolase and Viscozyme, 1 M

sorbitol solution. The solution was aspirated for homogeneity and once homogenous was made-up to 1 mL using the 10 mg/mL Zymolase and Viscozyme, 1 M sorbitol solution. Aspiration must be limited and quick to avoid rupture of the cell membrane after cell wall removal. Cell wall digestion followed for 7 min. Once digestion was complete, the mixture was spun gently at 12,000 x g RCF for 30 s to avoid disrupting the cell membranes. The enzyme supernatant mix was removed and 1 mL of 1 M sorbitol was added to the pellet. The solution was centrifuged, being careful to face the pellet towards the centre of the rotor, such that the pellet be washed. This was repeated 3x to remove excess enzyme mix. The sorbitol was removed each time without re-suspending the pellet. The sorbitol-washed pellet was re-suspended in 100 µL of 18% ficoll, 20 mM HEPES, pH 8. The tip of the pipette was used to break up the pellet. Following this, 100 µL 9% ficoll 20 mM HEPEs, pH 8 was added and the mixture inverted, stabilizing the nuclei. The mixture was centrifuged at 16,800 × g for 30 s – 1 min. The nuclei pellet was re-suspended in 100 µL 0.42 M NaCl, 20 mM HEPEs, pH 8, removing protein bound to DNA. The mixture was spun-down after Buffer C treatment. Supernatant and pellet stored at -80 °C.

2.6.3 Algal cell lysis by sonication

Algal cell pellet was re-suspended in a 50 mM Tris/HEPES, pH 7.5 0.1% triton X-100 solution. Cells were lysed using a Vibra-Cell™ VCX 150 sonicator fitted with a microtip on a 1 s on, 1 s off cycle, amplitude: 40% for 280s. All cells were

kept on ice during sonication. The lysates was centrifuged for 15 min using the Microfuge 22R at 14,000 x g RCF to remove cell debris.

2.6.4 Culturing *K. nitens* in the PhO_2xBox

NIES C media was prepared as outlined in Section 2.5.3.

K. nitens (NIES 2285) stationary cultures were grown using NIES C media, at 15 °C on a 12 hour light/dark cycle (light: 2700 lux) under atmospheric conditions. Cultures were initiated by adding 1 mL of a stock *K. nitens* culture (in HSMA) to 30 mL of fresh NIES C media at pH 7.5. The cell culture flasks were Falcon® 25cm² Rectangular Canted Neck Cell Culture Flask with Vented Cap (Corning). A plumbed air and N_2 supply was used for controlling the atmosphere within the box. O_2 and CO_2 concentrations were reduced by increasing the N_2 ratio.

Chapter 3: Biochemical characterisation of *Marchantia polymorpha* plant cysteine oxidase

3.1 Introduction

M. polymorpha occupies an important node in plant evolutionary history, possibly retaining features of both its algal ancestors and extant land plants. The predicted *M. polymorpha* proteome contains one putative PCO, with a single ERF-VII transcription factor homolog, and PRT6 and ATE homologs – essential proteins in the N-degron pathway. Studying homologous plant hypoxia regulatory proteins from the *M. polymorpha* proteome will help understand the relationship between O_2 sensing enzymes and land plant evolution. This chapter describes the expression, purification and biophysical characterisation of *M. polymorpha* plant cysteine oxidase (MpPCO), and its activity towards a putative endogenous substrate *M. polymorpha* ethylene response factor-like (MpERF).

3.2 Expression, purification and characterisation of MpPCO

3.2.1 Plasmid preparation

The DNA sequence encoding for MpPCO was provided by F. Licausi (affiliation), in a pDEST17 plasmid. To facilitate protein purification by Ni(II) affinity chromatography, the MpPCO sequence was subcloned into the pET28a plasmid (Chapter 2, Section 2.1.1.1). PCR amplification of the template DNA

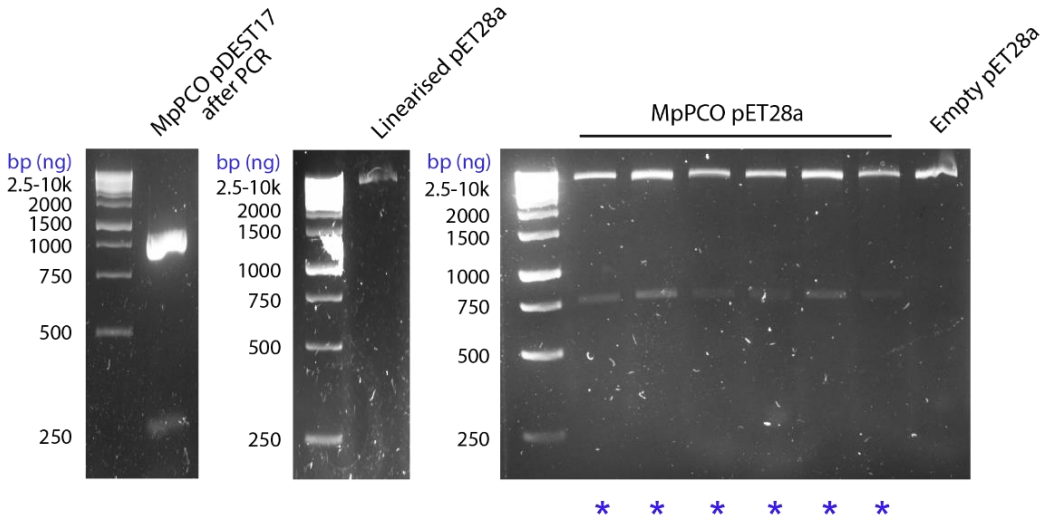


Fig. 3.1: Subcloning the MpPCO sequence into the pET-28a plasmid. Ligation of the MpPCO sequence (831bp) into the pET28a plasmid, after amplification from the pDEST17 plasmid.

was verified by 2% agarose gel (Fig. 3.1), followed by *DpnI* digestion of the remaining template DNA. The pET28a plasmid was linearized *via* the *NdeI* and *XhoI* restriction sites (Appendix 3.1, and Fig. 3.1). The DNA fragment was amplified and then inserted into the pET28a vector using an In-Fusion® HD cloning kit. *E. coli* XL10 Gold® ultra-competent cells were transformed with the resulting recombinant DNA. Resultant colonies were selected at random for DNA amplification and purification *via* the GeneJET Plasmid Miniprep Kit. All six colonies selected contained plasmid DNA (Fig. 3.1). Sanger sequencing of the purified DNA plasmids verified the pET28a plasmid contained the MpPCO coding sequence. An alternative pET28a plasmid for the expression of His_6-MpPCO with a His_6-3C cleavage site was synthesised by Genscript®, and verified by Sanger sequencing. 3C protease is more efficient at His_6-cleavage at 4 °C than thrombin protease, and was considered a better alternative for

structural studies, where yield and purity are important factors (Chapter 5, Section 5.2.3.2 - 5.2.3.3).

3.2.2 Expression

Both recombinant His_6-MpPCO plasmids were expressed in *Escherichia coli (E.coli)* BL21(DE3) competent cells (Chapter 2, Sections 2.1). Initially, a small-scale expression trial was attempted, but was unsuccessful (Appendix 3.2). 600 mL cultures of *E. coli* transformed with the His_6-MpPCO plasmid were grown at 37 °C to an optical density of ~0.8 at 600 nm before inducing protein expression with 0.5 mM IPTG. Induced cultures were agitated at 180 rpm and grown overnight at 18 °C. After overnight growth, the cells were harvested by centrifugation Cell pellet yields varied between 6 and 10 g/L culture and pellets stored at – 80 °C.

3.2.3 Purification His_6-MpPCO

Soluble His_6-MpPCO (expressed from both plasmids) was purified using HisTrap™ HP affinity chromatography (Chapter 2, Section 2.1.7) and HiLoad 26/600 Superdex 75pg gel filtration chromatography (Chapter 2, Section 2.1.10): The lysate was prepared by re-suspending the harvested cells in HisTrap™ HP binding buffer, DNAse I and protease inhibitor cocktail, followed by sonication, centrifugation and filtration of soluble proteins prior to loading on to the HisTrap™ HP column (Chapter 2, Section 2.1). Bound protein was eluted over a 0-100% gradient of elution buffer. Protein began to elute at approximately 300 mM imidazole. His_6-MpPCO-containing fractions were

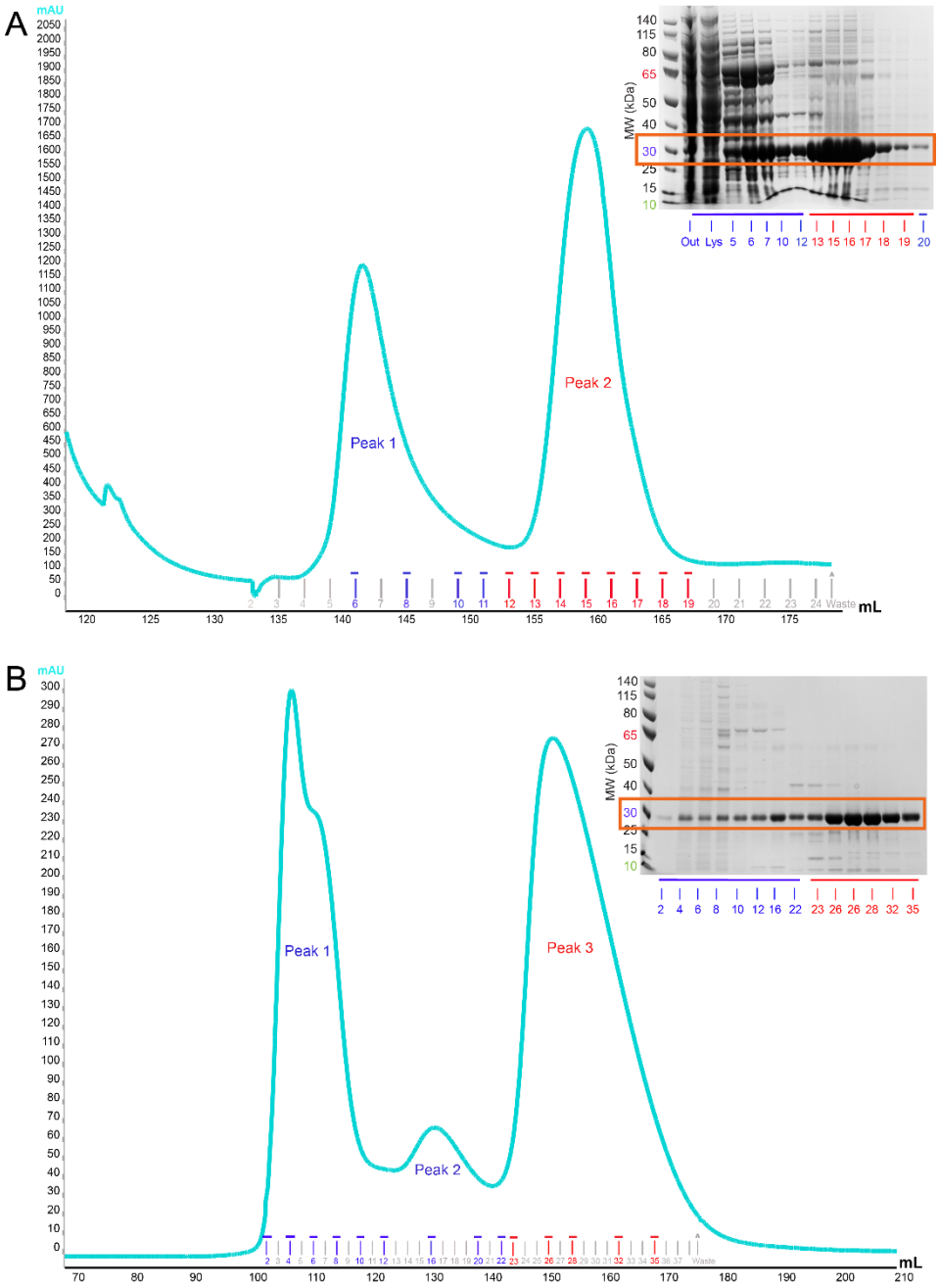


Fig. 3.2: Purification of recombinant His_6-MpPCO. A) Chromatogram and Coomassie stained SDS-PAGE of HisTrap™ purified *E. coli* BL21 DE3 lysate with overexpressed, recombinant His_6-MpPCO. B) Chromatogram and Coomassie stained SDS-PAGE of gel filtered HisTrap™ Peak 2 fractions. Protein bands of interest are highlighted

identified using SDS PAGE (Fig. 3.2, A). HisTrap™ Peak 2 fractions had fewer *E. coli* contaminants than those of HT Peak 1, and were therefore selected for further purification by gel filtration chromatography.

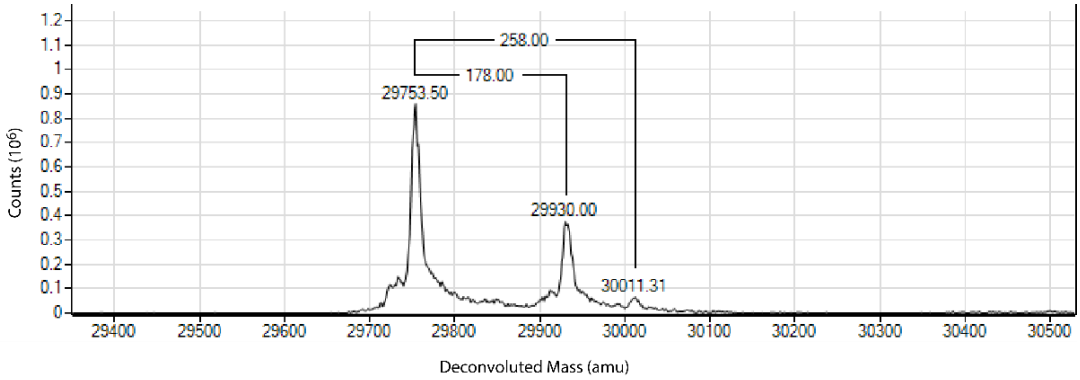


Fig. 3.3: Exact mass of His_6-MpPCO using RapidFire® Q-TOF MS. Exact His_6-MpPCO mass is 29753.5 Da, with post-translational alpha-N-6-phosphogluconoylation (+258 Da) and subsequent de-phosphorylation (+178 Da) of the His_6-tag by *E. coli* proteins.

HisTrap™ Peak 2 fractions 13-19 were pooled and buffer exchanged into gel filtration buffer using a PD-10 desalting column prior to gel filtration purification with the HiLoad 26/600 Superdex 75pg column (Chapter 2, Section 2.1.9). Fractions from gel filtration Peak 3 (GF Peak 3) had the highest proportion of pure His_6-MpPCO (Fig. 3.2, B). All subsequent protein characterisation used His_6-MpPCO from GF Peak 3.

Exact protein mass was determined using the Agilent RapidFire® auto-sampler and C4 cartridge coupled to a 6550 iFunnel Q-TOF MS. His_6-MpPCO exact mass was determined to be 29753.5 Da (Fig. 3.3), in agreement with the predicted mass of 29753.57 Da. The +178 and +258 Da species are likely due to post-translational alpha-N-6-phosphogluconoylation and subsequent de-phosphorylation of the His_6-tag by *E. coli* proteins [239]. Size exclusion chromatography with multi-angle light scattering (SEC-MALS) showed the purified His_6-MpPCO was monomeric in solution, as SEC-MALS eluted protein mass was within 4% of the exact His_6-MpPCO mass (Fig. 3.4, Chapter 2, Section 2.4.3). This is in agreement with previously characterised PCOs [142].

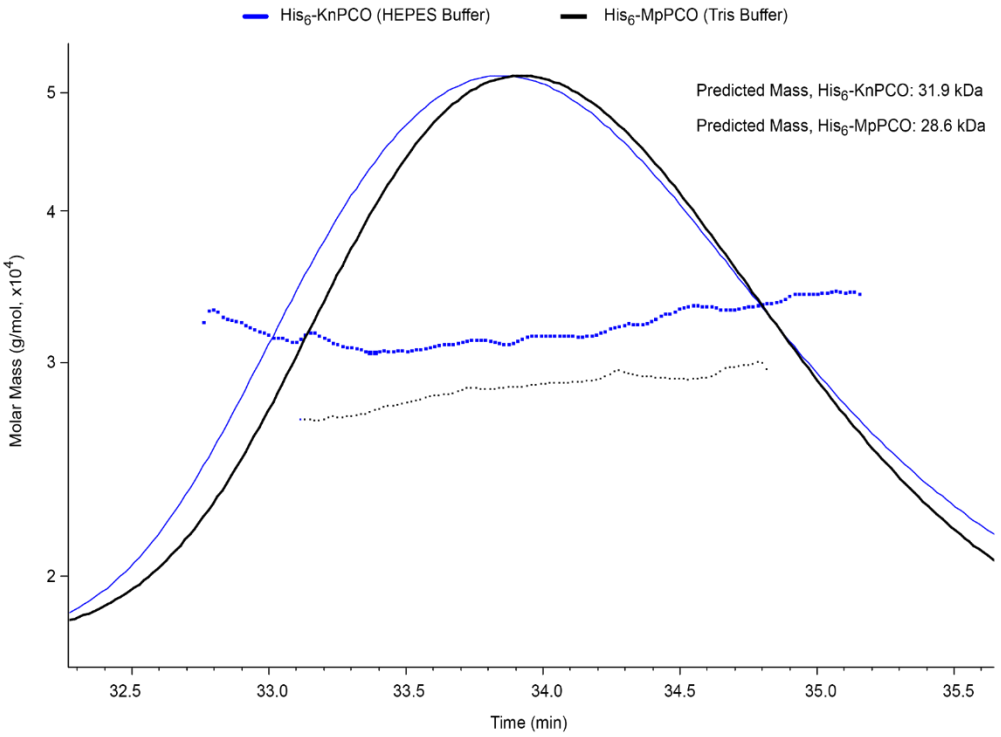


Fig. 3.4: SEC-MALS of His_6-MpPCO and His_6-KnPCO. UV-trace of His_6-MpPCO and His_6-KnPCO proteins eluting from the Superdex 75 HR10/30 column. Protein concentration was determined using the Wyatt Optilab rEX refractive index monitor. Intensity of light scattering was measured on the Wyatt Dawn HELEOS-II 8-angle light scattering detector. Data collected by Dr D. Staunton.

3.2.4 Purification of MpPCO and hexahistidine-tag cleavage

MpPCO without the His_6-fusion tag was prepared by incubation with either thrombin or 3C protease, depending on the integrated cleavage site between the protein and the His_6-fusion tag (Chapter 2, Section 2.1.8). 3C-cleaved MpPCO was used exclusively for structural or thermal stability and aggregation studies. Thrombin-cleaved was MpPCO favoured for kinetic characterisation, as 3C-cleaved MpPCO was initially found to be inactive (Appendix 5.5). A second 3C-cleaved MpPCO preparation using a different pellet, showed that a significant proportion of 3C-cleaved MpPCO activity was lost after gel filtration, but that some activity remained (Appendix 5.6). This suggested high batch to

batch protein activity variation and the gel filtration process could be detrimental to protein activity.

The purification procedures of His_6-cleaved MpPCO differ only to include the His_6-tag cleavage step prior to gel filtration. The chosen proteases were added to the eluted HisTrap™ fractions and the protein/protease solution dialysed overnight into HisTrap™ binding buffer (low imidazole) at 4˚C. Manual HisTrap™ purification of the recombinant protein/protease solution after cleavage separated His_6-tag cleaved from His_6-tag un-cleaved protein.

The cleaved-protein fractions were identified by SDS-PAGE and concentrated for gel filtration purification. Exact protein mass was determined using the Agilent RapidFire® auto-sampler and C4 cartridge coupled to a 6550 iFunnel Q-TOF MS. Thrombin-cleaved MpPCO exact mass was determined to be 28003.0 Da (Appendix 3.3.1), also in agreement with the predicted mass of 28002.71 Da. 3C-cleaved MpPCO exact mass was determined to be 27875.9 Da (Appendix 3.3.2), in agreement with the predicted mass of 27875.61. Neither the 3C nor thrombin cleaved MpPCO solutions had the previously observed +178 Da or +258 Da species. This validates the assumption that the observed +178 Da and +258 Da peaks are likely post-translational modifications of the His_6-tag.

3.2.5 Determining MpPCO activity

PCO enzymes are defined by their ability to catalyse the oxidation of N-terminal cysteine (Cys) to cysteine-sulfinic ($CysO_2H$) acid in the presence of O_2.

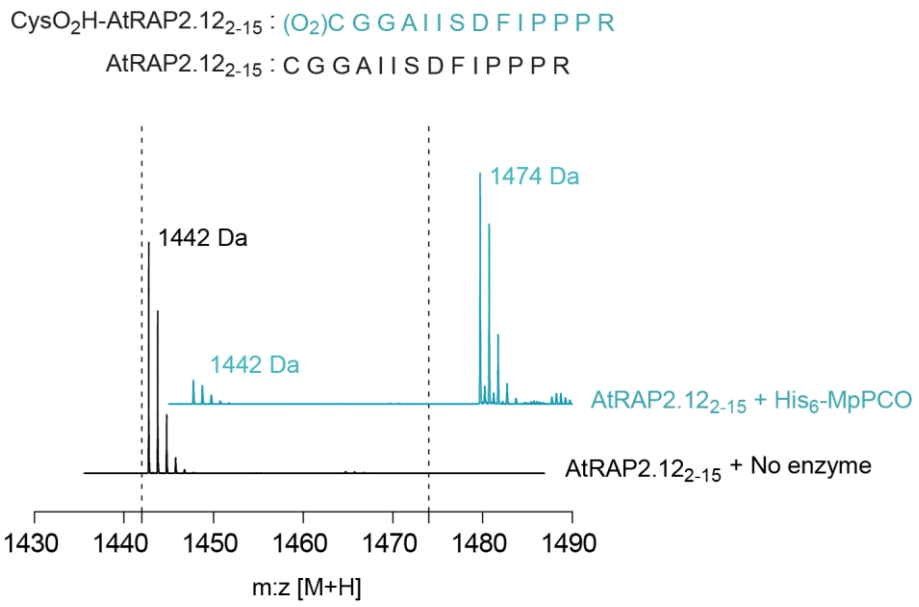


Fig. 3.5: Activity of His_6-MpPCO with AtRAP2.12$_{2-15}$. Mass spectrum showing AtRAP2.12 $_{2-15}$ incubated for 60 min with and without His_6-MpPCO, demonstrating a +32 Da shift in the presence of enzyme.

To be considered functionally homologous to known PCOs [142], MpPCO had to exhibit Cys to $CysO_2H$ catalysis in an O_2 dependent manner. Initial His_6-MpPCO activity was determined by incubation with 14-mer *Arabidopsis thaliana* RAP2.12$_{2-15}$ (AtRAP2.12$_{2-15}$) peptide under atmospheric O_2 for 1 hour at 25 ˚C, the documented substrate for *Arabidopsis thaliana* PCOs (AtPCOs) (Chapter 2, Section 2.1.11.1). Analysis on the Xevo G2-S Q-TOF UPLC/MS with a Chromolith® RP-18e 100-2 mm HPLC column showed a +32 Da mass shift on the AtRAP2.12$_{2-15}$ peptide upon the addition of His_6-MpPCO, consistent with di-oxidation (Fig. 3.5).

The O_2 dependence of His_6-MpPCO catalysed oxidation of AtRAP2.12$_{2-15}$ was determined using a Brookes Mass Flow Controller (Chapter 2, Section 2.1.11.2) and Xevo G2-S Q-TOF UPLC/MS analysis. AtRAP2.12$_{2-15,}$ TCEP and buffer were pre-equilibrated under the desired atmosphere for 10 min at 25 ˚C before the

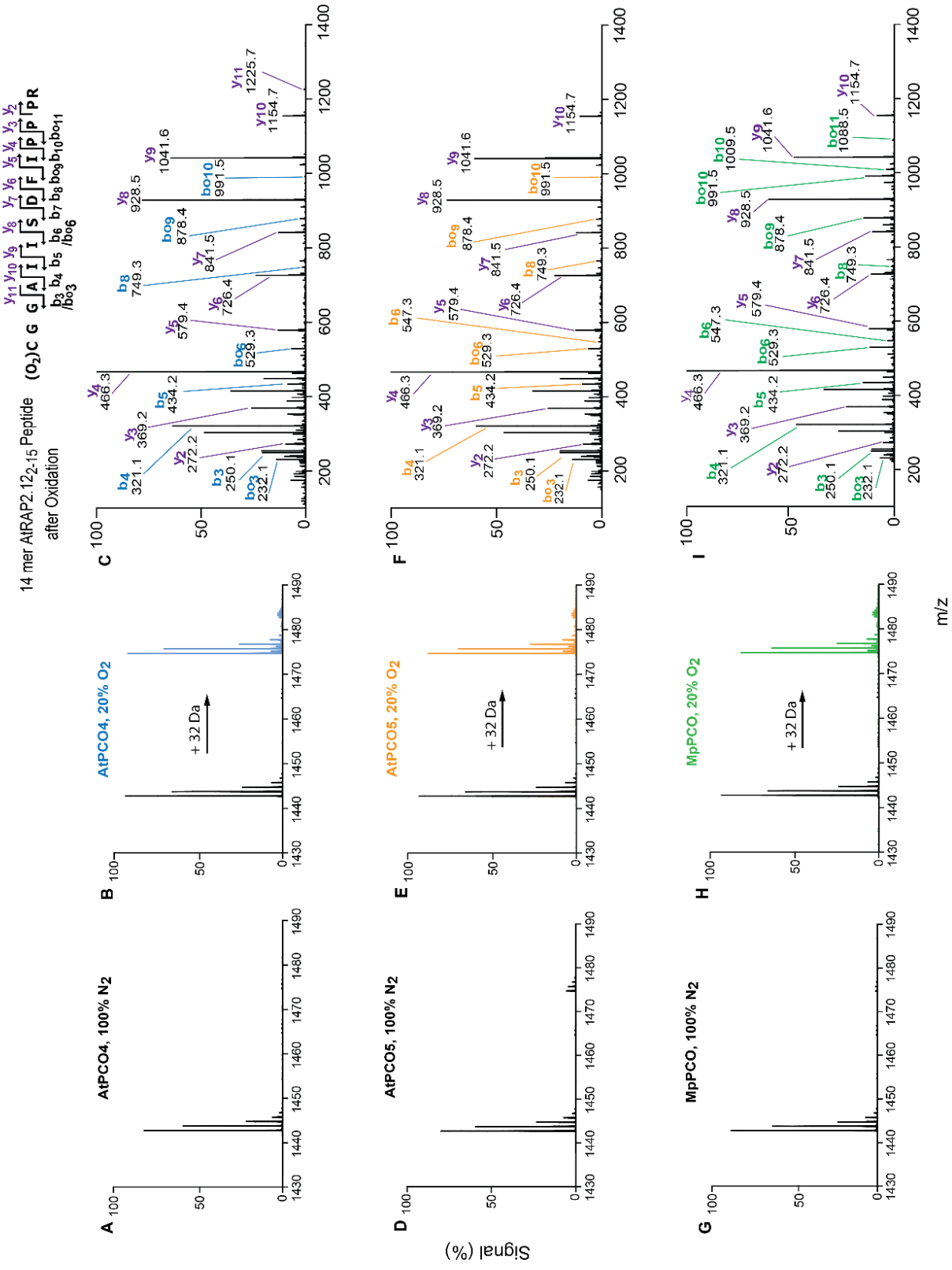
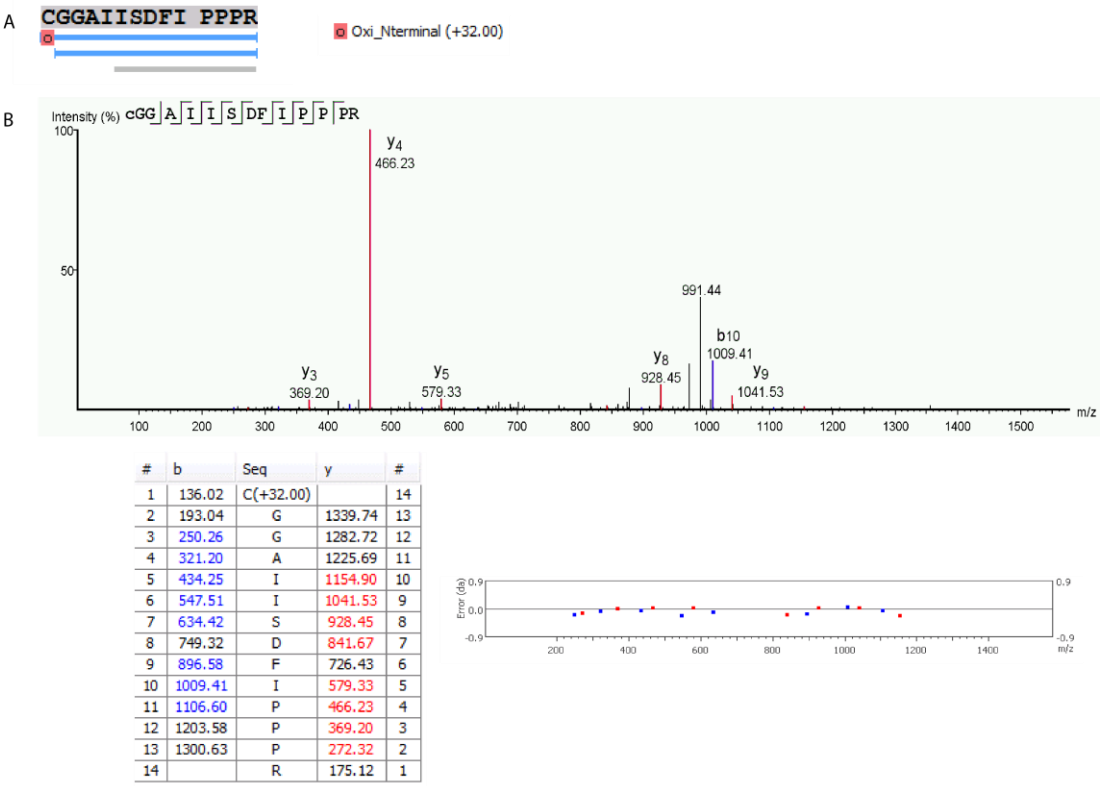


Fig. 3.6: Tandem MS analysis of AtRAP2.12 peptide oxidation. Tandem MS confined the +32 Da modification to the N-terminal CGG fragment, with the same results for His_6-AtPCO 4 and AtPCO 5/AtRAP2.12$_{2\text{-}15}$ reactions (A-F). The O_2 dependence of the His_6-MpPCO AtRAP2.12$_{2\text{-}15}$ reaction (G – I).

addition of His_6-MpPCO, quenching with 1% (v/v) formic acid after 2 min. Under an atmosphere of 100% N_2, a +32 Da shift was not observed on the AtRAP2.12_{2-15} peptide in the presence of any enzymes (Fig. 3.6, A, D, G). Similarly, under an atmosphere of 20% O_2 and no enzyme, a +32 Da shift is not observed on the AtRAP2.12_{2-15} peptide (Fig. 3.6, B, E, H, **black** spectra). However in the presence of 20% O_2 and His_6-MpPCO, His_6-AtPCO 4 and His_6-AtPCO 5 a +32 Da shift is observed on the AtRAP2.12_{2-15} peptide (Fig. 3.6, H, **green** spectra) demonstrating the O_2 dependence of MpPCO-catalysed oxidation. These results are in agreement with the AtPCO isoforms, where both O_2 and enzyme are necessary for the oxidation of the AtRAP2.12_{2-15} peptide [142].

Tandem MS of AtRAP2.12_{2-15} peptide on the Xevo G2-S Q-TOF UPLC/MS after incubation with His_6-MpPCO and 20% O_2 showed the +32 Da modification on the Cys-Gly-Gly $\mathbf{b_{o3}/b_3}$fragments (Fig. 3.6, I, Chapter 2, Section 2.2.12.1). The complementary $\mathbf{y_{11}}$ Ala-Ile-Ile-Ser-Asp-Phe-Ile-Pro-Pro-Pro-Arg fragment is visible here for AtPCO 4 only (Fig. 3.6, C), but has been identified in the spectra for both AtPCO 5 and MpPCO in lower abundance.

The +32 Da modification was further confined to the N-terminal Cys of AtRAP2.12_{2-15} using nanoLC-MS/MS and PEAKS 8.5 sequencing software by E. Pires (Chemistry Department, Fig. 3.7, Chapter 2, Section 2.2.12.2). These results suggest MpPCO catalyses N-terminal Cys oxidation in a manner homologous to that of the AtPCOs.



#	b	Seq	y	#
1	136.02	C(+32.00)		14
2	193.04	G	1339.74	13
3	250.26	G	1282.72	12
4	321.20	A	1225.69	11
5	434.25	I	1154.90	10
6	547.51	I	1041.53	9
7	634.42	S	928.45	8
8	749.32	D	841.67	7
9	896.58	F	726.43	6
10	1009.41	I	579.33	5
11	1106.60	P	466.23	4
12	1203.58	P	369.20	3
13	1300.63	P	272.32	2
14		R	175.12	1

Fig. 3.7: Tandem MS of the AtRAP2.12$_{2\text{-}15}$ peptide after His$_6$-MpPCO-catalysed oxidation. A) Full Sequence coverage (100%) of AtRAP2.12 (C(+32)GGAIISDFIPPPR) was achieved. Peptide sequencing software, PEAKS 8.5, listed the RAP2.12$_{2\text{-}15}$ N-terminal Cys as a "confident modification site" and present a -10lgP> 20. All peptides presenting any modification, spectra were manually checked and validated or disqualified. For C1 the identification is very reliable, as full fragmentation in the HCD is observed. Data collected by Ms E. Pires.

3.2.6 Determining His$_6$-MpPCO metal-binding *via* Inductively Coupled Plasma Mass Spectrometry

Determining MpPCO metal content is important for informing MpPCO kinetic characterisation conditions and concentrations of active enzyme. The metal content of His$_6$-MpPCO was determined by Inductively Coupled Plasma Mass Spectrometry (ICP-MS) in the Trace Element ICP-MS Small Research Facility, Earth Sciences (Chapter 2, Section 2.2.13.1, Table 3.1). Fe, Zn and Ni were bound in substoichiometric levels, with 0.17 (± 0.031) Fe, 0.1 (± 0.014) Zn and 0.09 (± 0.008) Ni atoms per monomer (Appendix 3.4 A-C), compared with 0.3

Protein	Bound Metal		
	Fe	Zn	Ni
His_6-MpPCO	0.17 (± 0.03)	0.08 (±0.01)	0.4 (± 0.03)
MpPCO	0.36 (± 0.00)	0.01 (± 0.01)	0.09 (± 0.01)

Table 3.1: Purification and iron content of MpPCO and His_6-MpPCO. The metal content of each protein was quantified using ICP-MS. All proteins had substoichiometric iron levels. Error is shown by standard deviation (n = 2). Data provided by Dr P. Holdship (earth Sciences Department)

(± 0.12) Fe atoms/molecule of AtPCO1 and 0.31 (± 0.015) Fe atoms/molecule for AtPCO 4 [142]. The Ni content could be as a result of the hexahistidine purification tag, which has a known affinity for Ni [240]. While MicroPIXE data for AtPCO 1 and 4 found Ni to be bound, the crystal structures of the AtPCOs did not have any bound Ni suggesting perhaps the Ni peak is as a result of buffer contamination [1]. Fe(II)-dependent PCO-catalysed di-oxygenation is in keeping with other known PCOs [142], and the (His)3 "facial triad" motif is synonymous with Fe binding in PCOs. The role of Zn is unclear. Zn(II) is known to inhibit Arabidopsis PCOs [241], with similar results for His_6-MpPCO (Fig. 3.13, E). These results will inform enzyme activity assays as a function of metal availability, and further align MpPCO with previously described PCOs.

3.3 Endogenous MpPCO substrate discovery and synthesis

3.3.2 Identification of *M. polymorpha* Hypoxia Response Element-like (MpERF)

To generate more biologically meaningful data, an endogenous substrate for MpPCO was necessary. The predicted proteome of *M. polymorpha* was accessed *via* the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute, Phytozome 12. The sequences of known PCO

1: MpERF	100.00	29.37	30.85	29.81	30.77	**32.80**
2: AtHRE1	29.37	100.00	40.68	38.52	38.89	51.61
3: AtRAP2.12	30.85	40.68	100.00	66.19	40.00	45.96
4: AtRAP2.2	29.81	38.52	66.19	100.00	41.15	45.24
5: AtRAP2.3	30.77	38.89	40.00	41.15	100.00	44.05
6: **AtHRE2**	32.80	51.61	45.96	45.24	44.05	100.00

Fig. 3.8: ID matrix of the Arabidopsis ERF-VII transcription factors with MpERF. The ERF VII substrates from Arabidopsis share approximately 30% identity with MpERF. Data collected using Clustal Omega.

substrates were used as search templates for potential MpPCO substrates (e.g. AtZPR2, AtRAP2.12 and 2.3). One Met-Cys initiating protein homologous to the ERF-VII substrates was identified in the *M. polymorpha* proteome (Transcript ID: Mapoly0293s0001). The protein is a putative transcription factor, AP2/ERF, and shares the highest percentage identity with *Arabidopsis thaliana* Hypoxia Response Element 2 (AtHRE2) at 32.8% (Fig. 3.8). The protein was named 'MpERF' to reflect its homology.

3.3.3 Preparation of $MpERF_{2\text{-}15}$

To investigate the activity of the MpPCO towards the MpERF protein, a 14-mer peptide of N-terminal MpERF residues 2-15 was synthesised, purified and quantified (Chapter 2, Section 2.2). The N-terminal Met_1 was omitted to expose the Cys_2. 14-mer ERF peptides were optimally active with Arabidopsis PCO enzymes [188], and as such this peptide length was chosen for MpERF analysis. MpERF was synthesised *via* solid state peptide synthesis on the CEM Liberty Blue microwave peptide synthesiser using MBHA Rink Amide resin (Chapter 2, Section 2.2). The crude peptide was purified on the JASCO HPLC

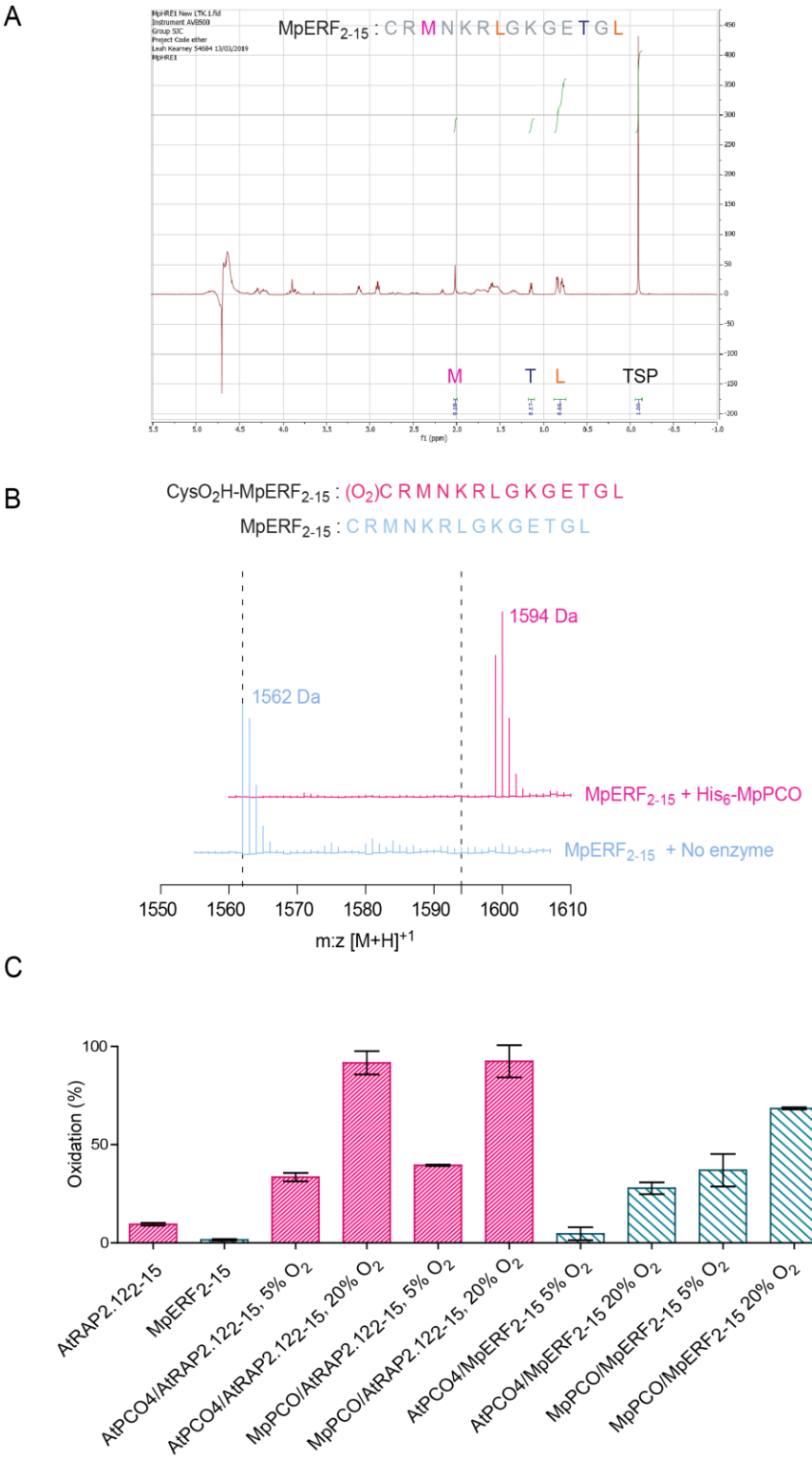


Fig. 3.9: NMR of MpERF, activity and O_2 dependence of the $MpERF_{2\text{-}15}$ peptide with MpPCO and AtPCO 4. A) 1H NMR with NOESY water suppression of the $MpERF_{2\text{-}15}$ peptide B) Mass spectrum showing $MpERF_{2\text{-}15}$ with and without incubation with His_6-MpPCO. A +32 Da modification on the $MpERF_{2\text{-}15}$ peptide is observed in the presence of His_6-MpPCO after 1 hour at 25 °C. C) O_2 dependence of the His_6-MpPCO/$MpERF_{2\text{-}15}$ reaction, compared to the MpPCO/$AtRAP2.12_{2\text{-}15}$ and His_6-AtPCO 4/$AtRAP2.12_{2\text{-}15}$ and His_6-AtPCO 4/$MpERF_{2\text{-}15}$ reactions. Error bars represent standard deviation, n=3.

with a Gemini® 5 µm NX-C18 110 Å, 250 x 30 mm, AXIA packed C18 column. Peptide-containing HPLC fractions were identified *via* MALDI-TOF MS, and combined for quantification by ^{1}H NMR. A 1:1 solution of purified MpERF$_{2\text{-}15}$ peptide and trimethylsilylpropanoic acid (TSP) standard in D_2O was run on the AVB500 with a 1D NOESY water suppression technique. The 9 TSP protons gave a standard signal by which the concentration of MpERF was determined using the signal from Met, Leu and Thr in the polypeptide chain (Fig. 3.9, A).

3.3.4 Initial activity of MpERF$_{2\text{-}15}$

200µM MpERF$_{2\text{-}15}$ peptide was incubated with and without 1 µM His$_6$-MpPCO at 30°C for 60 min under atmospheric O_2. Xevo G2-S Q-TOF UPLC/MS analysis of the peptide confirmed a +32 Da mass shift on the MpERF$_{2\text{-}15}$ peptide in the presence of His$_6$-MpPCO (Fig. 3.9, B).

The O_2 dependence of His$_6$-MpPCO catalysed oxidation of AtRAP2.12$_{2\text{-}15}$ was determined using a Brookes Mass Flow Controller, with peptide oxidation analysis performed on the Xevo G2-S Q-TOF UPLC/MS. 200 µM MpERF$_{2\text{-}15}$ and AtRAP2.12$_{2\text{-}15}$ with 1 mM tris(2-carboxyethyl)phosphine (TCEP) were incubated with and without 0.5 µM His$_6$-MpPCO or His$_6$-AtPCO 4, under 5 and 20% O_2 (Fig. 3.9, C). TCEP was added to ensure the N-terminal Cys was reduced, and to avoid peptide dimerization as this would make the N-terminal Cys of the peptide inaccessible to the enzyme.

Tandem MS studies of the CysO$_2$H MpERF$_{2\text{-}15}$ peptide could not ascertain the location of the +32 Da mass shift as there was no discernible ion fragmentation

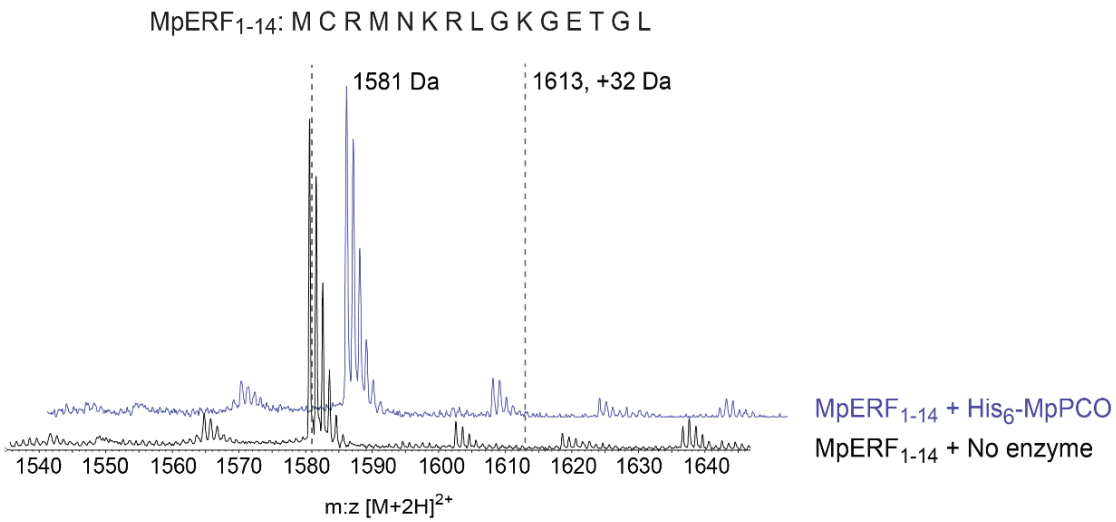


Fig. 3.10: MpERF$_{1-14}$ peptide activity with His$_6$-MpPCO. MpERF$_{1-14}$ with and without His$_6$-MpPCO, with no observed +32 Da shift in the presence of enzyme after 1 hour at 25 °C.

pattern, likely due to lability of the oxidised peptide. However a variant MpERF$_{1-14}$ peptide, with an N-terminal Met blocking the cysteine, did not exhibit any activity when incubated with His$_6$-MpPCO (Fig. 3.10). This indicates that an N-terminally available Cys on the MpERF$_{2-15}$ peptide is necessary for activity. This is consistent with the activity of other PCOs towards their endogenous substrates.

3.4 Probing the His$_6$-MpPCO active site

To better understand which MpPCO residues are catalytically relevant, the active site was probed using site-directed mutagenesis. MpPCO shares the greatest homology with PCO 4 of the Arabidopsis PCOs. AtPCO 4 residues, Asp_196 and His_164, are known to be associated with catalysis [1]. As such, corresponding MpPCO residues, His184 and Asp196 were selected for site-directed mutagenesis. His184 is believed to be part of the Fe-binding triad,

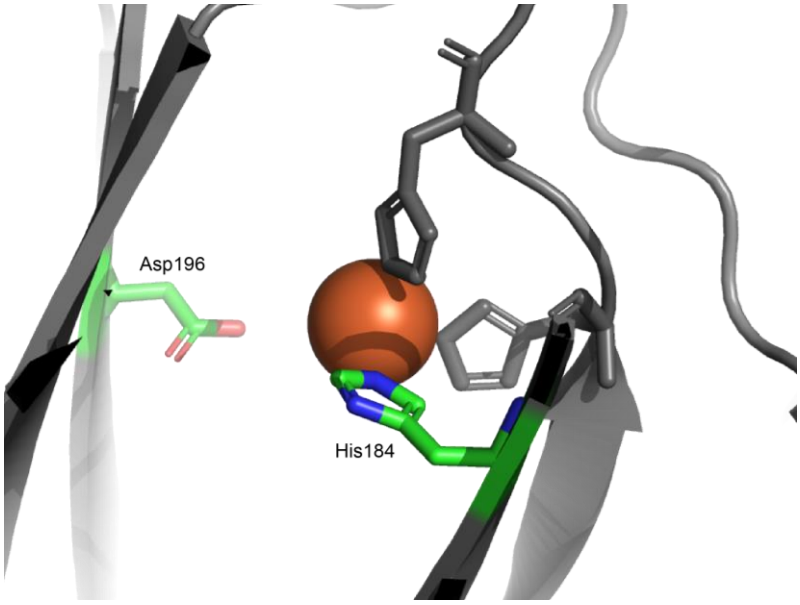


Fig. 3.11: Crystal structure of the AtPCO 4 active site. The AtPCO 4 crystal structure (PDB ID: 6S7E) with the corresponding MpPCO_Asp196 (AtPCO 4_Asp176) and MpPCO_His184 (AtPCO 4_His164) residues coloured by element.

with Asp196 in close proximity to the active site in the AtPCO 4 crystal structure (Fig. 3.11).

3.4.1 Site-directed mutagenesis of His_6-MpPCO

The thrombin His_6-MpPCO plasmid was mutated *via* the Agilent QuikChange® II Site Directed Mutagenesis Kit (Chapter 2, Section 2.1.2). The D196N and H184D mutations were confirmed by Sanger sequencing. The variants were expressed and purified according to the standard operating procedures outlined in Chapter 2, Sections 2.1. Exact protein masses for MpPCO_D196N and MpPCO_H184D were 29752.58 Da and 29731.51 Da respectively as determined by Xevo G2-S Q-TOF UPLC/MS (Appendix 3.5).

3.4.2 Activity of His_6-MpPCO_D196N and His_6-MpPCO_H184D

H184D and D196N His_6-MpPCO variant activity was tested using a 10 minute end-point assay under atmospheric O_2 to determine the ability of the variants

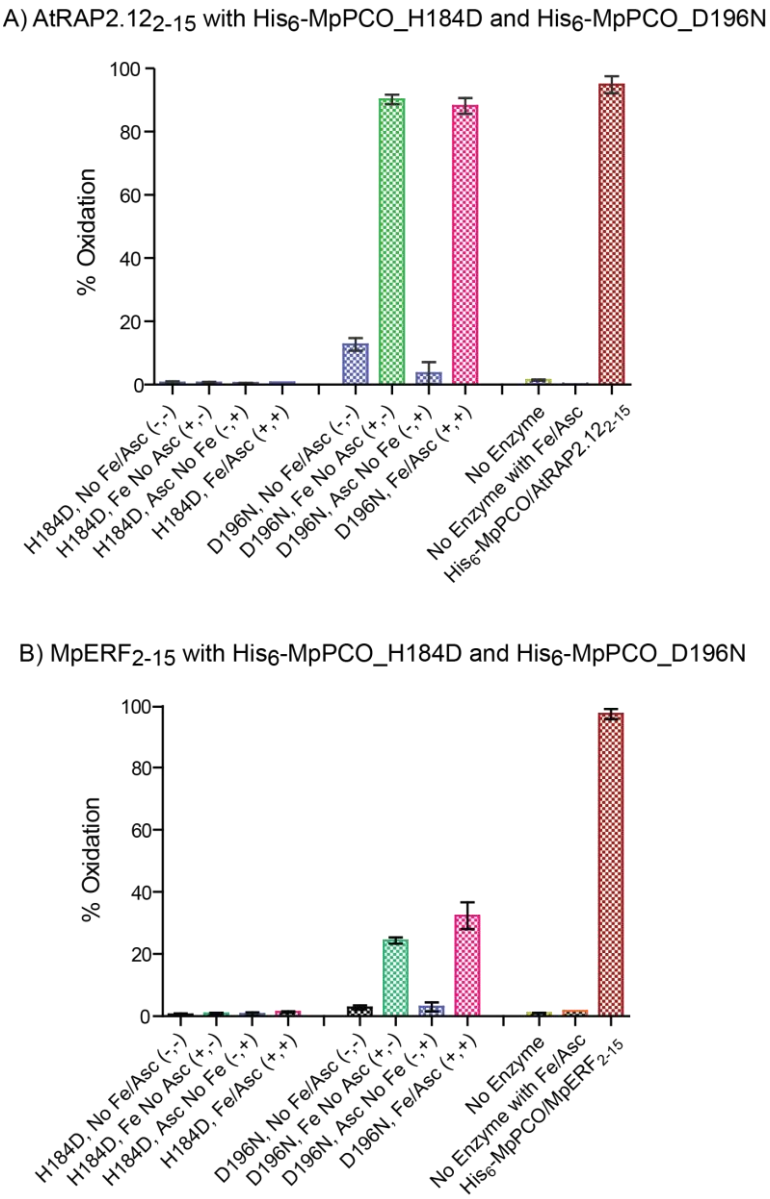


Fig. 3.12: Activity of MpPCO_D196N and MpPCO_H184D variants. Activity profiles His$_6$-MpPCO_D196N and His$_6$-MpPCO_H184D, with A) AtRAP2.12$_{2-15}$ and B) MpERF$_{2-15}$ after 10 min of incubation at 25 °C. Error bars represent standard deviation, n=3.

to catalyse N-terminal Cys oxidation. 0.5 μM enzyme 1mM TCEP, 200 μM AtRAP2.12$_{2-15}$ and MpERF$_{2-15}$ were incubated with and without Fe(II) and ascorbate at 25 °C for 10 minutes, followed by quenching with 1% (v/v) formic acid. Xevo G2-S Q-TOF UPLC/MS analysis of the resulting peptides showed the H184D mutation had a deleterious effect on His$_6$-MpPCO activity independent of substrate, Fe(II) or ascorbate addition (Fig. 3.12). This result is in keeping

with the hypothesis that His184 forms part of the Fe(II)-binding triad in MpPCO.

His_6-MpPCO_D196N activity depended on both substrate choice and Fe(II) addition. His_6-MpPCO_D196N activity was greatly reduced compared to the wild-type (WT) control. Activity towards both AtRAP2.12_{2-15} and $MpERF_{2-15}$ could be restored by Fe(II) addition, with the degree of restoration greater for AtRAP2.12_{2-15} than $MpERF_{2-15}$. This would suggest that D196 is involved, directly or indirectly, with substrate binding. Fe(II) addition may help to stabilize D196N_MpPCO in a catalytically active conformation for substrate binding or *vice-versa*.

3.5 Kinetic studies of MpPCO

To compare the catalytic efficiency and O_2 sensitivity of MpPCO with other characterised PCOs, MpPCO kinetics with the $MpERF_{2-15}$ and His_6- AtRAP2.12_{2-15} peptides were studied. His_6-MpPCO kinetics were assessed *via* quenched assays on the Agilent RapidFire® 6550 iFunnel Q-TOF MS (C4 cartridge) system. The RapidFire® was chosen for the analysis of peptides following kinetic assays as it can reliably handle high-throughput analysis, with sampling time reduced from ~7 min/sample to <10 s/sample compared to the Xevo G2-S Q-TOF UPLC/MS. Real-time assays were used for generating relative activity data for the pH dependence of MpPCO catalysis, with quenched assays preferred for metal and TCEP/DTT screens and Michaelis-Menten kinetics. Peptide oxidation was quantified using Agilent Quantitative Analysis software.

Variability among replicates was quantified by the standard deviation. Enzyme aliquot concentrations were checked using the Nanodrop™ spectrophotometer (protein absorbance at 280 nm) prior to any experiment to ensure protein concentration consistency.

3.5.1 Determining the optimum conditions for MpPCO kinetic analysis

This section focuses on establishing the ideal conditions for studying the His_6-MpPCO/$MpERF_{2\text{-}15}$ and His_6-MpPCO/$AtRAP2.12_{2\text{-}15}$ reactions prior to establishing Michaelis-Menten kinetics (Chapter 2, Section 2.3).

3.5.1.1 Determining TCEP and DTT effect onHis_6-MpPCO/substrate turnover

As a result of the N-terminal Cys on target ERF peptides, a reducing atmosphere is required to minimise peptide di-sulfide bond formation. Peptide dimerization would likely render the N-terminal cysteine unavailable for PCO-catalysed di-oxygenation (Section 3.3.4), resulting in varying levels of 'bioavailable' peptide per reaction. TCEP and dithiothreitol (DTT) were screened for their effect on His_6-MpPCO activity (Fig. 3.14 A, B). For both His_6-MpPCO/$MpERF_{2\text{-}15}$ and His_6-MpPCO/$AtRAP2.12_{2\text{-}15}$ reactions, a greater degree of oxidation is observed in the presence of TCEP than DTT or without a reducing agent (Fig. 3.13 A, B). TCEP was therefore selected for use in all kinetic assays.

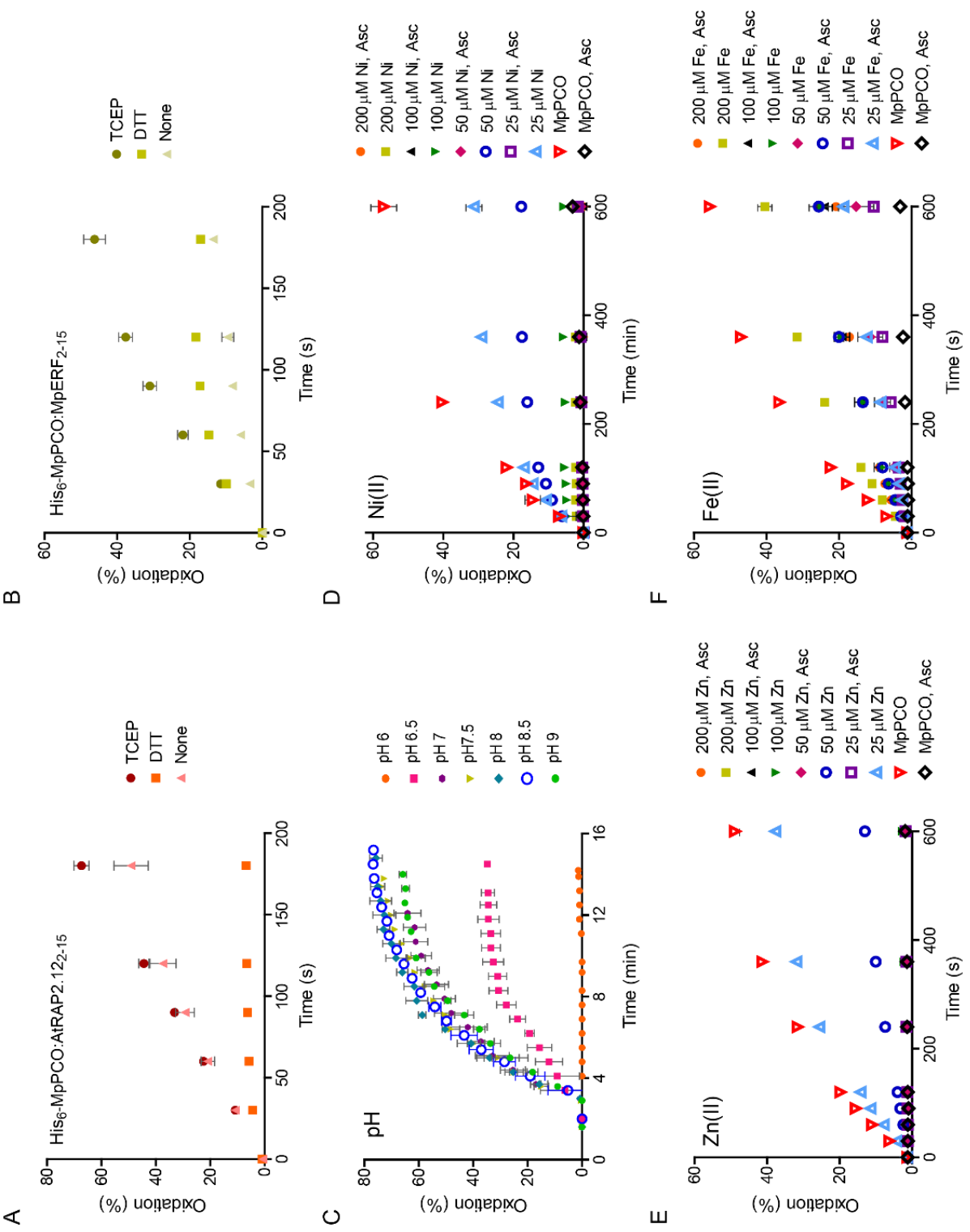


Fig. 3.13: Determining optimum His_6-MpPCO kinetic conditions. Activity profile of the A) His_6-MpPCO/AtRAP2.12_{2-15} and B) His_6-MpPCO/$MpERF_{2-15}$ complexes in the presence of DTT and TCEP. Activity of the His_6-MpPCO/$MpERF_{2-15}$ reaction C) as a function of pH D) Ni(II) and ascorbate E) Zn(II) and ascorbate and F) Fe(II) and ascorbate. Error bars represent standard deviation, n=3.

3.5.1.2 His_6-MpPCO pH dependence

AtPCO 4 demonstrated pH dependent activity [188], with enzyme activity almost eliminated at pH 6 and 6.5. ^{1}H NMR determined the pK_a of the N-terminal thiol on the Cys of the AtRAP2.12$_{2-15}$ peptide to be 6.78. It is therefore believed that peptide deprotonation is important for substrate binding and/or turnover [188]. To explore this pH dependency of the His_6-MpPCO/MpERF$_{2-15}$ reaction, 1 mM TCEP, 0.1 μM His_6-MpPCO and 4 μM MpERF$_{2-15}$ were incubated at pH range 6-9 at room temperature (~29 ˚C), monitoring real-time peptide turnover on the RapidFire® (Chapter 2, Section 2.3.1). The His_6-MpPCO/MpERF$_{2-15}$ reaction has a wide pH tolerance, with near indistinguishable activity levels between pH 7 and 9 (Fig. 3.13, C). pH may have an effect on activity but it is likely to be negligible in the 25-29 ˚C temperature range. While activity loss was not as dramatic at pH 6.5 compared to the AtPCO 4/AtRAP2.12$_{2-15}$ reaction, the pH change caused a ~55% drop in activity compared to pH 8. The N-terminus of MpERF$_{2-15}$ (CysArgMetAsn) is markedly different to that of AtRAP2.12$_{2-15}$ (CysGlyGlyAla), with a charged Arg group adjacent to the Cys. The observed activity at pH 6.5 could be a reflection of a lower Cys pK_a in the MpERF$_{2-15}$ peptide. The time lag in Fig 3.13, C could be a combination of slower protein diffusion due to the much greater reaction volume (1 mL) compared to quenched assays with same peptide concentrations (500 μL). A quenched assay at 4.8 μM MpERF$_{2-15}$ did not show any turnover until 60 s (Appendix 3.6). Further kinetic studies were done at pH 8, in keeping with previous Arabidopsis PCO kinetic studies [188].

3.5.1.3 His_6-MpPCO metal dependence

ICP-MS studies of His_6-MpPCO showed Zn, Fe and Ni bound in substoichiometric levels (Section 3.2.6). As Fe(II) is catalytically relevant, supplementing the assays with Fe(II) could improve turnover. As Zn(II) role is not understood, it was considered important to explore Zn(II) effect on activity (Chapter 2, Section 2.3.2). Ni(II) addition was studied to see if it affected activity, an important consideration when co-purifying enzymes with Ni(II), and for comparing kinetics and activity between protein batches. Between 200 and 25 μM $ZnCl_2$, $NiSO_4$ and $Fe(NH_4)_2(SO_4)_2.6H_2O$ were incubated with 1 mM TCEP, 0.1 μM His_6-MpPCO and 50 μM $MpERF_{2\text{-}15}$ with and without 4 mM ascorbate at 25 °C The reaction was quenched in 1% (v/v) formic acid In every instance, additional metal and ascorbate inhibited His_6-MpPCO activity (Fig. 3.13 D-F). His_6-MpPCO/$MpERF_{2\text{-}15}$ activity was inversely proportional to $ZnCl_2$ and $NiSO_4$ concentrations. Additional Fe(II) also had an inhibitory effect on His_6-MpPCO activity, suggesting MpPCO Fe(II) uptake to the catalytic site is poor.

The study of His_6-MpPCO activity as a function of pH, metal availability and a reducing environment ascertained the ideal conditions for further kinetic analysis. The buffer chosen was 50 mM bis-tris propane, 50 mM NaCl pH 8 (the same pH used for the AtPCOs [188]), TCEP was selected as the peptide reducing agent, and no additional metal or ascorbate was added to the reaction.

3.5.2 Michaelis-Menten kinetics of the His_6-MpPCO/AtRAP2.12$_{2-15}$ and His_6-MpPCO/MpERF$_{2-15}$ reactions

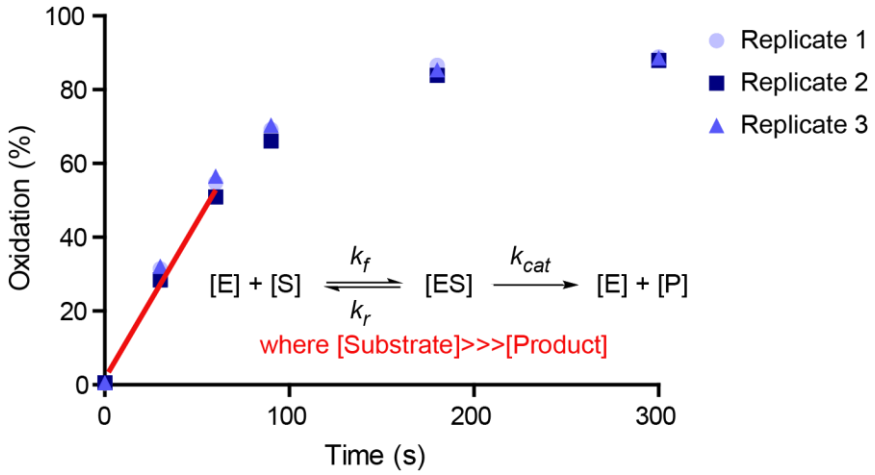


Fig. 3.14: The initial zone, as determined by Michaelis-Menten kinetics. The rate of turnover for His_6-MpPCO is calculated from the 'initial zone', where the concentration of substrate is much greater than product.

Michaelis-Menten kinetics are a useful tool for understanding the binding efficiencies of substrates and co-factors of any given enzymatic reaction. This data will inform further studies for understanding the molecular details underpinning substrate binding and co-factor uptake. Under the Michaelis-Menten model, steady-state kinetics can be approximated when the concentration of substrate [S] is much greater than that of product [P]. Under these conditions the rate of formation of the intermediate complex [ES] does not change with product formation and product formation can be assumed to be irreversible. These conditions are met in the initial zone, from which a rate of turnover can be calculated (Fig. 3.14). These calculated rates were divided by the concentration of active enzyme available in solution (as determined by the proportion of MpPCO-bound Fe(II) in the ICP-MS studies). These rate/mg of active enzyme values were plotted versus substrate concentration until the

maximum reaction velocity, V_{max}, was reached. The Michaelis-Menten constant, K_M, is calculated as the substrate concentration at half V_{max}.

The Michaelis-Menten equation is defined as follows, where the velocity (ν) is proportional to the maximum velocity (V_{max}), substrate concentration ([S]) and the Michaelis-Menten constant, (K_M)

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

where

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

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

3.5.2.1 K_M (AtRAP2.12$_{2\text{-}15}$) and K_M (MpERF$_{2\text{-}15}$)

All assays were conducted in triplicate at 25 °C, in 50 mM bis-tris propane, 50 mM NaCl, pH 8 buffer, 1 mM TCEP and 0.1 μM His$_6$-MpPCO. Assay samples were quenched in 1% (v/v) formic acid at regular time intervals. AtRAP2.12$_{2\text{-}15}$ concentrations varied from 2-210 μM, with the resulting rate of peptide oxidation per mg of enzyme fit to a substrate inhibition model (Fig. 3.15, A).

$$Y = \frac{V_{max}\,X}{K_M + X(1 + {}^{X}/_{K_i})}$$

where

$$K_i = \frac{k_{off}}{k_{on}}$$

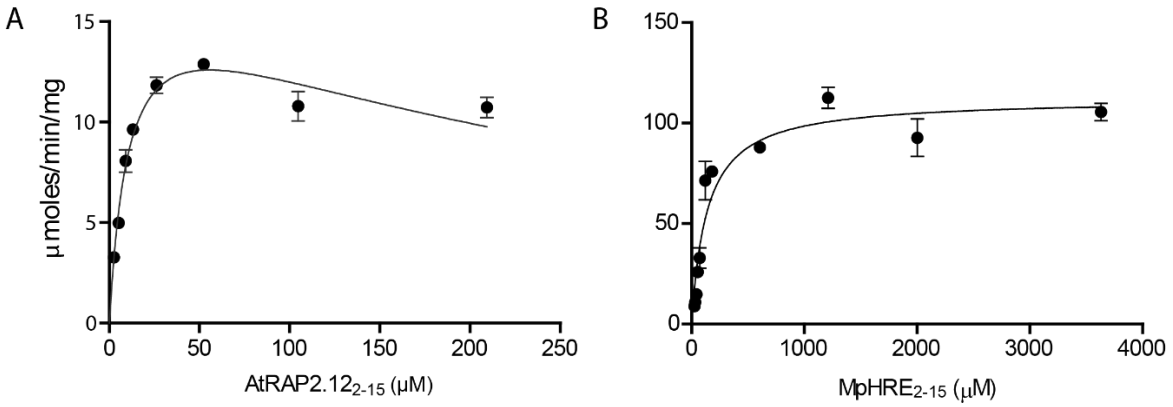


Fig. 3.15: Kinetics of His-MpPCO activity towards the AtRAP2.12$_{2-15}$ and MpERF$_{2-15}$ peptides. A) The K_M AtRAP2.12$_{2-15}$ for His$_6$-MpPCO was 11.0 ± 1.1 µM, with an inhibition constant, Ki, of 279.9 ± 62.9 µM, SD and B) the K_M MpERF$_{2-15}$ for His$_6$-MpPCO was 137.8 ± 19.54 µM. Error bars represent standard deviation, n=3.

The calculated K_M for AtRAP2.12$_{2-15}$ was 11.0 ± 1.1 µM, with an inhibition constant, K_i, of 279.9 ± 62.9 µM. MpERF$_{2-15}$ concentrations varied from 0.016 to 3.63 mM, with resulting rate of peptide oxidation per mg of enzyme fit to a standard Michaelis-Menten model (Fig. 3.15, B). The K_M for the MpERF$_{2-15}$ peptide was 137.8 ± 19.54 µM. The K_M for His$_6$-MpPCO/AtRAP2.12$_{2-15}$ is lower than the lowest calculated K_M for the AtPCOs by a factor of 8 (AtPCO2) [188]. The calculated K_M values indicate MpPCO has a greater affinity for the A tRAP2.12$_{2-15}$ peptide than for the MpERF$_{2-15}$ peptide.

3.5.2.2 K_M^{app} O_2 of the His$_6$-MpPCO/AtRAP2.12$_{2-15}$ and His$_6$-MpPCO/MpERF$_{2-15}$ reactions

To understand how sensitive the His$_6$-MpPCO/AtRAP2.12$_{2-15}$ and His$_6$-MpPCO/MpERF$_{2-15}$ reactions were to O_2 availability, the K_M app O_2 for the two reactions were calculated. All assays were conducted in triplicate at 25 °C, in 50 mM bis-tris propane, 50 mM NaCl, pH 8 buffer, 1 mM TCEP and 0.1 µM His$_6$-

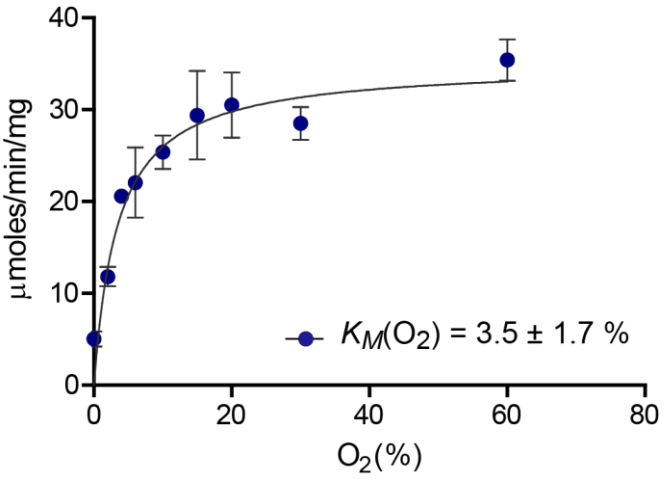


Fig. 3.16: K_M^{app} O_2 curve for the His$_6$-MpPCO/AtRAP2.12$_{2\text{-}15}$ reaction. The K_M O_2 curve for the His$_6$-MpPCO/AtRAP2.12$_{2\text{-}15}$ reaction, K_{Mapp} O_2 3.5 ± 0.8 %. Error bars represent standard deviation, n=3.

MpPCO. Initially a substrate concentration of approximately three-times the determined K_M was used in experiments to determine the K_M app O_2. However these substrate concentrations did not necessarily promote maximal enzyme activity, therefore this approach was abandoned in favour of AtRAP2.12$_{2\text{-}15}$ and MpERF$_{2\text{-}15}$ concentrations at the point of maximum enzyme activity in the K_M substrate experiments, 50 µM and 1.2 mM respectively (Fig. 3.15, A and B). Reactions were quenched in 1% (v/v) formic acid at a time known to be within the initial rate zone of the chosen peptide concentration (Appendix 3.7). O_2 concentrations varied between 2 and 80%.

The K_M app O_2 determined for the His$_6$-MpPCO/AtRAP2.12$_{2\text{-}15}$ reaction was 3.5 ± 0.8 %, lower than the reported values for the AtPCOs [188] (Fig. 3.16). In contrast, the K_M^{app} O_2 for the His$_6$-MpPCO/MpERF$_{2\text{-}15}$ reaction at 600 µM MpERF$_{2\text{-}15}$ was found to be 37.4 ± 9.4 % (Fig. 3.17, A). This experiment was repeated at 1200 µM MpERF$_{2\text{-}15}$ with the high K_M app O_2 verified at 34.9 ± 13.2

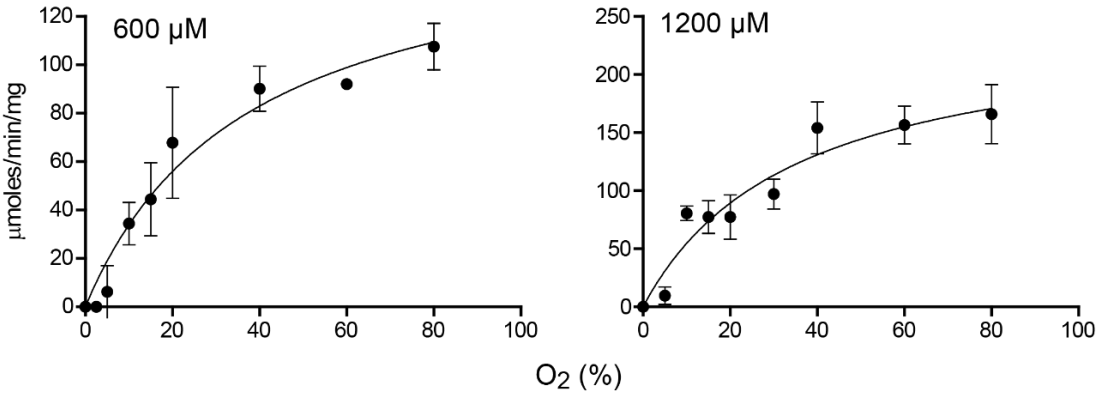


Fig. 3.17: The $K_M{}^{app}O_2$ curve for the MpPCO/MpHRE$_{2\text{-}15}$ reaction. The $K_M{}^{app}O_2$ curve for the MpPCO/MpHRE$_{2\text{-}15}$ reaction at 600 and 1200 µM peptide, with a K_MO_2 of 37.4 ± 9.4 %, SD and 34.9 ± 13.2 %. Error bars represent standard error of the mean, n=3.

% (Fig. 3.17, B). Despite repeated attempts, the $K_M{}^{app}\ O_2$ values for the His$_6$-MpPCO/MpERF$_{2\text{-}15}$ reaction could not be reproduced. The calculated $K_M{}^{app}\ O_2$ values varied between 3.5 and 37.4 %, despite using the same conditions for each experiment. Furthermore, the data within experiments was prone to large error, resulting in a high number of outlying data points (Fig. 3.18). The variability could not be attributed to the peptide stocks (MpERF$_{2\text{-}15}$ peptide synthesised and purified using the same protocols, but at different times), nor could it be attributed to the buffer (Fig. 3.18). Once the peptide and the buffer were ruled-out as sources of error, His$_6$-MpPCO stability was analysed.

3.5.2.3 Examining the stability of MpPCO and His$_6$-MpPCO

To investigate whether His$_6$-MpPCO itself contributed to the kinetic variability of the His$_6$-MpPCO/MpERF$_{2\text{-}15}$ reaction, His$_6$-MpPCO was analysed by thermal shift assay on a BioRad C1000 Touch™ Thermal Cycler and Nanotemper Prometheus NT.48 (Chapter 2, Section 2.4.1 and 2.4.2).

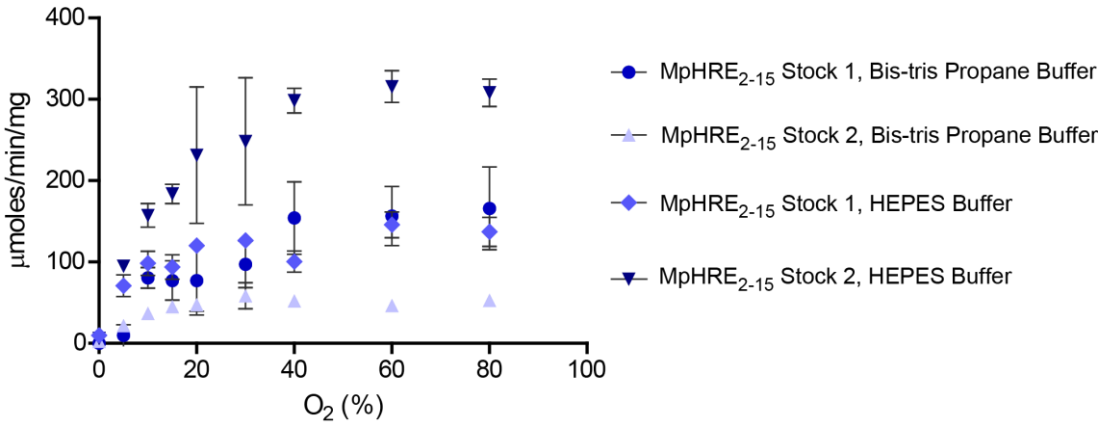


Fig. 3.18: The K_M^{app} O_2 Michaelis-Menten curves for the His_6-MpPCO/$MpERF_{2-15}$ reaction. The $K_M^{app}O_2$ values could not be repeated, with a great deal of variability within and between data sets. Error bars represent standard deviation, n=3.

Initial thermal studies were carried out on the BioRad using SYPRO™ Orange fluorescence to monitor protein conformational changes as a function of temperature, with an increase in fluorescent signal indicative of increased protein folding [231]. 3C protease His_6-cleaved MpPCO was used for comparison as it was known to generate reliable thermal shift data using this method (Chapter 5, Section 5.1.3). Results, collected by Jack Wilson, indicated this technique was not suitable for studying His_6-MpPCO as the fluorescent trace indicated the protein was unfolded at 20 °C (Fig. 3.19, A), which contradicts assay results (Fig. 3.19, B).

The Prometheus NT.48 provided a dye-free alternative route to thermal protein stability data, exploiting the intrinsic emission fluorescence of Trp at 350 and 330 nm. Protein aggregation as a function of temperature was assessed by measuring the scattering of incident light upon back-reflection by aggregated

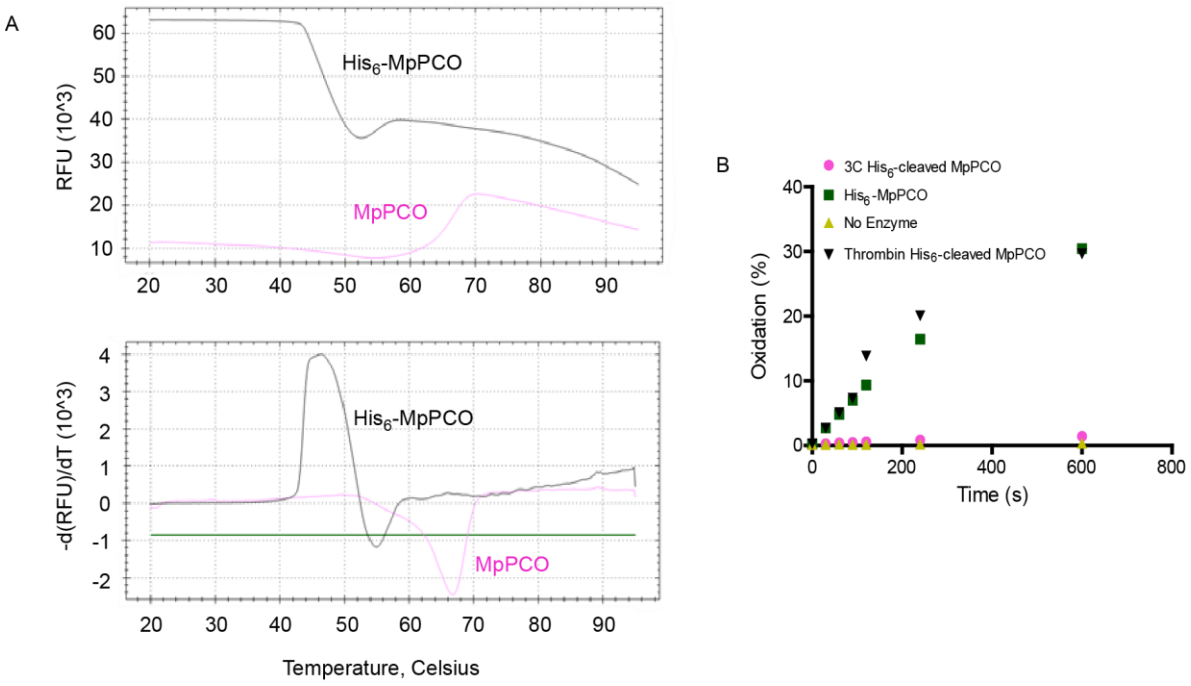


Fig 3.19: Thermal shift data for His_6-MpPCO and MpPCO using SYPRO® Orange dye. A) Thermal shift data of MpPCO with and without the hexahistadine tag (His_6) using SYPRO® orange dye B) comparative quenched assay showed His_6-MpPCO was active. Data collected by J Wilson.

particles. 10 µM of protein in ddH_2O was loaded into each capillary (~10 µL), and the temperature increased at a rate of 7 °C/min from 20 °C to 95 °C.

The ratio of florescence emission at 350 and 330 nm, F350/F330, suggest that His_6-MpPCO and MpPCO exhibit similar thermal stability (Fig. 3.20, A). However, calculated onset for scattering was 62.8 ± 1.6 °C and 47.0 ± 1.9 °C for MpPCO and His_6-MpPCO respectively (Fig. 3.20, B), with His_6-MpPCO aggregating almost 16 ˚C cooler than His_6-cleaved MpPCO. Polyhistidine purification tags are often reported to cause protein aggregation. The first step in the Michaelis-Menten derivation applies the law of mass action, which is reliant on free diffusion. It is possible that in an environment with very high peptide concentrations such as those in the His_6-MpPCO/$MpERF_{2\text{-}15}$ K_M^{app} O_2

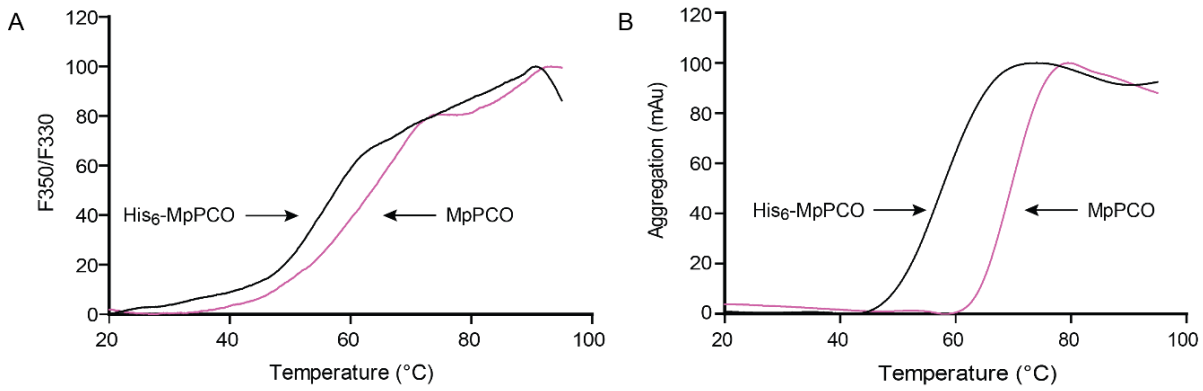


Fig 3.20: Thermal shift data for His_6-MpPCO and MpPCO using intrinsic Trp fluorescence. The A) ratio of F350 and F330 fluorescence and B) the light scattering data of His_6-MpPCO and MpPCO. This data was collected jointly with Jack Wilson.

experiment, His_6-MpPCO could aggregate. His_6-MpPCO aggregation would induce heterogeneity within reactions, altering diffusivity and potentially compromising the reproducibility of His_6-MpPCO kinetics.

3.5.2.4 Reproducing MpPCO kinetics

The results from Section 3.4.3.3 suggested that removing the His_6-tag would reduce MpPCO aggregation potential which might increase the reproducibility of kinetic analyses. Thrombin-cleaved MpPCO was expressed and purified according to the protocols in Section 2.1, Appendix 3.3.1 as 3C-cleaved MpPCO was found to be inactive (Appendix 5.5). ICP-MS was repeated for His_6-tag cleaved MpPCO revealing a ratio of 0.36 (± 0.001) Fe to protein (Table 3.1). Protein activity was assessed by a quenched assay, 100 µM $MpERF_{2\text{-}15}$, 0.1 µM MpPCO, 1 mM TCEP at 25 °C with aliquots quenched in 1% (v/v) formic acid (Fig. 3.19, B).

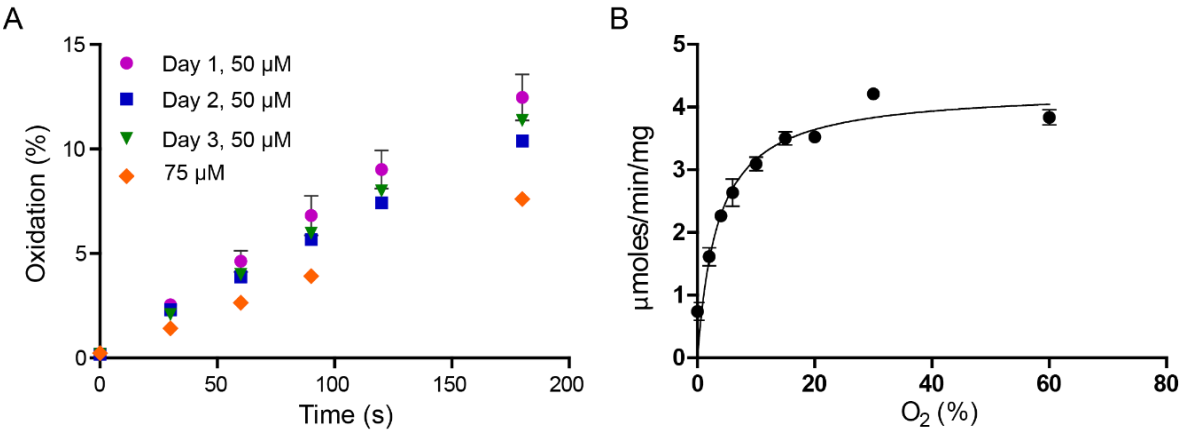


Fig. 3.21: Reproducing MpPCO/AtRAP2.12$_{2\text{-}15}$ kinetics. The A) reproducibility of MpPCO/AtRAP2.12$_{2\text{-}15}$ data at 50 µM, demonstrating substrate inhibition at 75 µM and B) the K_M^{app} O_2 curve for the MpPCO/AtRAP2.12$_{2\text{-}15}$ reaction, reproducing the previous K_M^{app} O_2 of 3.5 ± 0.6 %. Error bars represent standard deviation, n=3.

As the His$_6$-MpPCO protein had the potential to aggregate, it was important to assess whether or not removing the His$_6$-tag affected kinetics for both the AtRAP2.12$_{2\text{-}15}$ and MpERF$_{2\text{-}15}$ peptides. To verify the use of 50 µM AtRAP2.12$_{2\text{-}15}$ for K_M^{app} O_2 studies, the activity of thrombin-cleaved MpPCO with 50 µM AtRAP2.12$_{2\text{-}15}$ was measured on three different days. Furthermore, to ensure 50 µM of peptide was not limiting, MpPCO was incubated with 75 µM AtRAP2.12$_{2\text{-}15}$ (Fig. 3.21, A). Reproducibility was achieved with the 50 µM benchtop assays, with clear inhibitory effects at 75 µM AtRAP2.12$_{2\text{-}15}$. Inhibitory effects at concentrations above 50 µM AtRAP2.12$_{2\text{-}15}$ are consistent with previous kinetic data (Fig. 3.15, A), and undesirable for K_M^{app} O_2 studies. As such, 50 µM

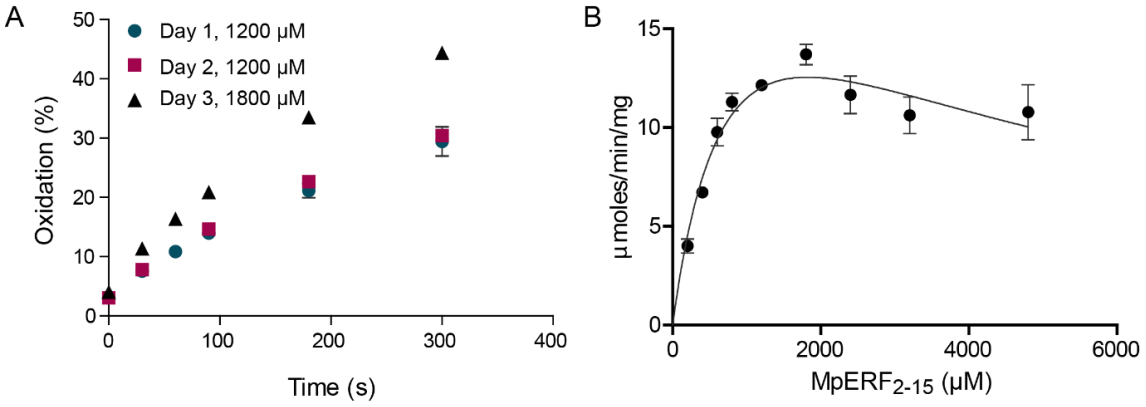


Fig. 3.22: Reproducing MpPCO/MpERF$_{2\text{-}15}$ kinetics. The A) reproducibility of MpPCO/MpERF$_{2\text{-}15}$ data at 1200 µM was established, yielding greater activity at 1800 µM and B) the K_{Mapp} MpERF$_{2\text{-}15}$ curve with a K_{Mapp} MpERF$_{2\text{-}15}$ of 694.1 ± 155.8 µM. Error bars represent standard deviation, n=3.

Protein	Substrate	k_{cat}	K_M	V_{max}
		s^{-1}	µM	µ$mol\ min^{-1}\ mg^{-1}$
His$_6$-MpPCO	AtRAP2.12$_{2\text{-}15}$	8.7 ± 0.57	11 ± 1.1	17.6 ± 1.1
MpPCO	MpERF$_{2\text{-}15}$	11.6 ± 2.1	896 ± 268	24.88 ± 4.4

Table 3.2: Steady-state kinetic parameters of MpPCO and His$_6$-MpPCO on AtRAP2.12$_{2\text{-}15}$ and MpERF$_{2\text{-}15}$ availability under atmospheric O_2. The turnover numbers (k_{cat}), Michaelis constants (K_M), and related maximum velocities (V_{max}) of MpPCO with respect to AtRAP2.12$_{2\text{-}15}$ and MpERF$_{2\text{-}15}$ Substrate inhibition was observed for MpPCO with both AtRAP2.12$_{2\text{-}15}$ and MpERF$_{2\text{-}15}$, with inhibition constants calculated as 0.28 ± 0.06 and 3.71 ± 1.3 mM, respectively.

AtRAP2.12$_{2\text{-}15}$ was used for the subsequent $K_M{}^{app}\ O_2$ experiment. The $K_M{}^{app}\ O_2$ for the MpPCO/AtRAP2.12$_{2\text{-}15}$ reaction was determined to be 3.5 ± 0.6 % (Fig. 3.21, B, Table 3.3), reproducing the previous measurement (Fig. 3.16).

To verify kinetic reproducibility, the activity of MpPCO with 1200 µM MpERF$_{2\text{-}15}$ was measured on two different days. To ensure 1200 µM of peptide was not limiting, MpPCO was also incubated with 1800 µM MpERF$_{2\text{-}15}$ (Fig. 3.22, A). Reproducibility was achieved with the 1200 µM benchtop assays, but it was

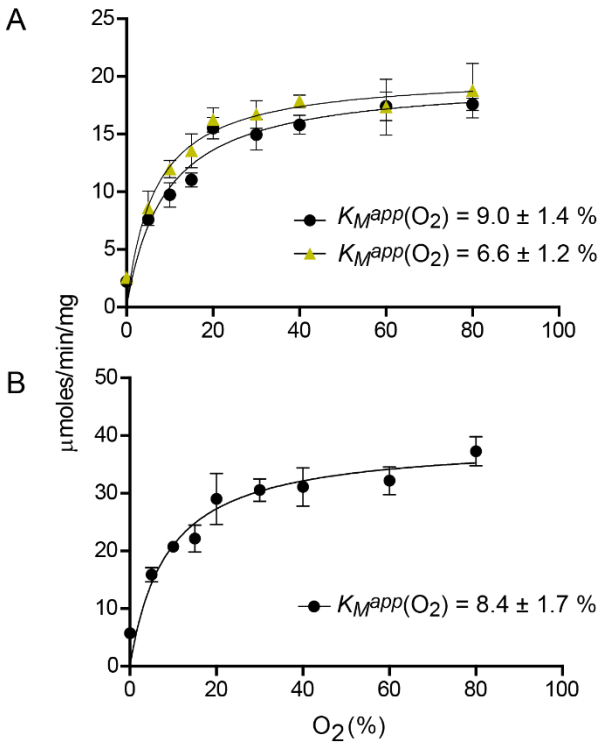


Fig. 3.23: Reproducing the O_2 kinetics of the MpPCO/MpERF$_{2\text{-}15}$ reaction. The A) K_M O_2 curve for the MpPCO/MpERF$_{2\text{-}15}$ reaction at 1200 µM MpERF$_{2\text{-}15}$, with a $K_M O_2$ of 9.1 ± 1.4 %, and reproduced at 6.6 ± 1.1 % and B) the $K_M O_2$ curve for the MpPCO/MpERF$_{2\text{-}15}$ reaction at 1800 µM MpERF$_{2\text{-}15}$, with a $K_M O_2$ of 8.4 ± 1.7%. Error bars represent standard deviation, n=3.

evident 1200 µM of peptide was a limiting concentration, with better activity at 1800 µM. It was therefore necessary to repeat the K_M for the MpERF$_{2\text{-}15}$ peptide. Prior to repeating the K_M for the MpERF$_{2\text{-}15}$ peptide, it was important to ensure reproducible K_M^{app} O_2 could be generated using the His$_6$-cleaved MpPCO protein. The K_M^{app} O_2 for the MpPCO/MpERF$_{2\text{-}15}$ reaction at 1200 µM peptide was 9.1 ± 1.4 %. Upon repetition the value was determined to be 6.6 ± 1.1 % (Fig. 3.23, A). This data showed for the first time, reproducible kinetics for the MpPCO/MpERF$_{2\text{-}15}$ reaction and O_2.

While the kinetics were reproducible, it was also apparent that the previous concentration of 1200 µM MpERF$_{2\text{-}15}$ peptide was no longer the peptide concentration at which maximum MpPCO activity was achieved. To determine the point of maximum activity, the K_M for the MpERF$_{2\text{-}15}$ peptide was repeated using the His$_6$-tag cleaved MpPCO.

Protein	Substrate	k_{cat}	K_M^{app}	V_{max}
		s^{-1}	%	$\mu mol\ min^{-1}\ mg^{-1}$
MpPCO	AtRAP2.12$_{2-15}$	2.0 ± 0.1	3.5 ± 0.6	4.3 ± 0.2
	MpERF$_{2-15}$	18.2 ± 0.9	8.7 ± 1.7	39.1 ± 2.0

Table 3.3: Steady-state kinetic parameters of MpPCO with O_2. The turnover numbers (k_{cat}), Michaelis constants (K_M^{app}), and related maximum velocities (V_{max}) of MpPCO with respect to AtRAP2.12$_{2-15}$ and MpERF$_{2-15}$.

MpERF$_{2-15}$ concentrations were varied from 0.2 to 4.8 mM, and fit to a Michaelis-Menten inhibition model (Fig. 3.22, B). The K_M MpERF$_{2-15}$ was determined to be 740.3 ± 171.6 µM, over five-fold greater than previously measured (Table 3.2). Maximum activity was observed at 1800 µM MpERF$_{2-15}$, and was therefore the chosen peptide concentration for the K_M^{app} O_2 experiment. The K_M^{app} O_2 for the MpPCO/MpERF$_{2-15}$ reaction at 1800 µM peptide was 8.4 ± 1.7% (Fig. 3.23, B, Table 3.3), reproducing the K_M^{app} O_2 values previously observed. The k_{cat} of MpPCO with the MpERF$_{2-15}$ peptide from the K_M^{app} O_2 experiment was higher than that obtained when measuring substrate dependence, indicating that MpPCO/MpERF$_{2-15}$ activity is submaximal at atmospheric O_2 concentrations.

3.6 Discussion

This chapter discusses the expression, purification and biophysical characterisation of MpPCO, a single PCO isoform in the *M. polymorpha* proteome. His$_6$-MpPCO was shown to catalyse the oxidation of an N-terminal Cys to CysO_2H in an O_2 dependent manner. His$_6$-MpPCO has both Fe(II), Zn(II) and Ni(II) ions bound in solution, a comparable result to previously

characterised PCOs; AtPCO1 and 4 [188]. When the putative His_6-MpPCO Fe-binding His triad is disrupted, mutating a single His to Asp, all enzymatic activity is lost, suggesting Fe(II) is bound and is essential for His_6-MpPCO activity. Additional Fe(II), Zn(II) and Ni(II) metals in the His_6-MpPCO/peptide assay inhibit activity. The observed Fe(II) inhibition suggests His_6-MpPCO metal uptake to the active site is likely difficult. His_6-MpPCO activity is pH dependent, with activity maintained as low as pH 6.5 with the $MpERF_{2\text{-}15}$ peptide.

His_6-MpPCO has a high affinity for the $AtRAP2.12_{2\text{-}15}$ peptide, a contrast with the high K_M $MpERF_{2\text{-}15}$. Initial K_M^{app} O_2 studies of the His_6-MpPCO /$MpERF_{2\text{-}15}$ reaction proved error-prone and irreproducible. Thermal shift assays of His_6-MpPCO suggested removing the His_6-tag could reduce aggregation potential and increase assay reproducibility. Thrombin-cleaved MpPCO yielded reliable K_M^{app} O_2 data. The calculated K_M^{app} O_2 for the MpPCO/$AtRAP2.12_{2\text{-}15}$ reaction was reproduced using both MpPCO and His_6-MpPCO. The reproducibility of the MpPCO/$AtRAP2.12_{2\text{-}15}$ reaction compared to the His_6-MpPCO/$AtRAP2.12_{2\text{-}15}$ reaction suggests the His_6-protein aggregation effects are minimal at low peptide concentrations, and do not affect kinetics.

Based on the experimentally determined K_M of 896 ± 268 μM and k_{cat} = 11.6 ± 2.1 s^{-1} values for the MpPCO/$MpERF_{2\text{-}15}$ reaction, and the K_M of 11.0 ± 1.1 μM and k_{cat} = 8.7 ± 0.57 s^{-1} for the MpPCO/$AtRAP2.12_{2\text{-}15}$ reaction, we can surmise that MpPCO has a greater affinity for the $AtRAP2.12_{2\text{-}15}$ peptide, than the $MpERF_{2\text{-}15}$ peptide. The low affinity for the $MpERF_{2\text{-}15}$ peptide was seemingly

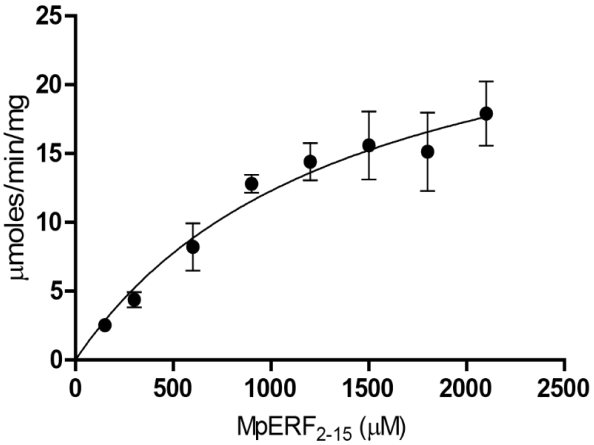


Fig. 3.24: The incomplete K_M curve for the MpERF$_{2\text{-}15}$ /AtPCO 4 reaction. K_M MpERF$_{2\text{-}15}$ curve for the His$_6$-AtPCO 4/MpERF$_{2\text{-}15}$ reaction, The K_M for the MpERF$_{2\text{-}15}$ peptide was 1097 ± 257 µM. Error bars represent standard deviation, n=3.

reflected in the His$_6$-AtPCO 4/MpERF$_{2\text{-}15}$ reaction. The K_M data set is incomplete as the concentration of peptide predicted to be required to achieve V_{max} was so high the experiment had to be abandoned due to time and cost constraints (Fig. 3.24). The predicted His$_6$-AtPCO 4/MpERF$_{2\text{-}15}$ K_M of 1388 ± 382 µM and k_{cat} = 1.4 ± 0.2 s^{-1} is in keeping with the lower affinity of MpPCO for MpERF$_{2\text{-}15}$, with a distinct loss of catalytic efficiency compared to the endogenous substrate: His$_6$-AtPCO 4/AtRAP2.12$_{2\text{-}15}$ K_M of 690 ± 110 µM and k_{cat} = 31 ± 3.49 s^{-1} [188].

The MpERF$_{2\text{-}15}$ N-terminus is bulky (CysArgMetAsn), charged (Arg) and polar (Asn), a contrast to the sterically unhindered, apolar CysGlyGlyAla N-terminus of AtRAP2.12$_{2\text{-}15}$. The differences between the local chemical environments of the N-terminal Cys may be responsible for the MpERF$_{2\text{-}15}$ and AtRAP2.12$_{2\text{-}15}$ respective kinetics with MpPCO. MpPCO and MpERF are the only observed PCO and ERF-like proteins in the Marchantia proteome with homology to the higher plant hypoxic pathway, linking gene regulation and cellular O_2. Homologous ERF VII substrates with CysGlyGlyAla N-termini, are not observed in the *M. polymorpha* proteome.

Of the Arabidopsis PCOs, MpPCO shares the greatest sequence homology with PCO4. AtPCO 4 is considered the most important Arabidopsis PCO isoform, with the highest reported K_{Mapp} O_2 of 17.3 ± 1.7 % [188]. Interestingly, both the MpPCO/AtRAP2.12$_{2-15}$ and MpPCO/MpERF$_{2-15}$ reactions are relatively less sensitive to changes in O_2 concentrations, especially above 20% O_2. This may be reflective of an alternative PCO function, one which does not require a great deal of sensitivity to changes in O_2. It is also possible that the O_2 sensitivity of MpPCO catalysis will vary with substrate, or perhaps the O_2 sensitivity of MpPCO-catalysis determined biochemically is not representative of reactions *in vivo* with full length substrates.

Chapter 4: Biochemical characterisation of putative algal plant cysteine oxidases

4.1 Summary

Algae provide an interesting avenue for investigating plant O_2 sensing as many live in aquatic habitats, and contain protein sequences homologous to the higher plant PCOs in their genomes. Although their genomes contain putative PCO sequences, they lack proteins with homology to the ERF-VII and ZPR2 transcription factors or MC-initiating proteins with homology to the VRN2 proteins. The conservation of PCO-like sequences throughout the plant kingdom suggests they have a function that pre-dates oxidation of the ERF transcription factors or other known substrates. This poses the question of what role PCOs have in algae, and what this role might tell us about how the plant O_2 sensing pathway evolved.

Blastp alignment searches revealed PCO-like sequences in a number of oceanic and terrestrial algal genomes (Fig. 4.1), but a paucity of ERF-like sequences outside of the flowering plants (Fig. 1.7). *Klebsormidium nitens* is a single-celled, undifferentiated, Charophytic algae, but has acquired some genes fundamental to terrestrialization, including UV protection and hormone signalling [225]. Phylogenetic analysis of various land plant and algal PCO sequences suggest there are sequences which are "land plant-like" and "algal-like" (See Chapter 1, Section 1.4.4). Phylogenetic analysis of the *K.*

Algae with sequenced genomes	Type	Number of isoforms	BLASTp with AtPCO 4	Identity with AtPCO4 (%)
Volvox carteri (taxid:3067)	Green	1	Significant simliarity	-
Micromonas pusilla CCMP1545 (taxid:564608)	Green	1	Significant simliarity	-
Ostreococcus lucimarinus CCE9901 (taxid:436017)	Green	1	Significant simliarity	-
Ostreococcus tauri (taxid:70448)	Green	2	Significant simliarity	27.9/26.7
Bathycoccus prasinos (taxid:41875)	Green	1	Significant simliarity	-
Chlamydomonas reinhardtii (taxid:3055)	Green	1	Significant simliarity	32.2
Chlorella variabilis (taxid:554065)	Green	1	Significant simliarity	41.3
Coccomyxa subellipsoidea C-169 (taxid:574566)	Green	1	Significant simliarity	-
Klebsormidium nitens (taxid:105231)	Green	1	Significant simliarity	42.6
Dunaliella salina (taxid:3046)	Green	1	Significant simliarity	32
Chondrus crispus (taxid:2769)	Red	1	No significant similarity	-
Porphyridium purpureum (taxid:35688)	Red	1	No significant similarity	-
Cyanidioschyzon merolae strain 10D	Red	1	Significant simliarity	-
Galdieria sulphuraria (taxid:130081)	Red	1	Significant simliarity	-
Pyropia yezoensis (taxid:2788)	Red	1	No significant similarity	-
Ectocarpus siliculosus (taxid:2880)	Brown	1	Significant simliarity	-
Emiliania huxleyi (taxid:2903)	Cocolithophore	2	Significant simliarity	38.9/25.3
Phaeodactylum tricornutum CCAP 1055/1 (taxid:556484)	Diatom	1	Significant simliarity	27.7

Table 4.1: List of algae with fully sequenced genomes (as of 2016). Those selected for expression trials are highlighted in green.

nitens PCO (KnPCO) suggested that KnPCO is more land plant-like in sequence (Fig. 1.10). As such, this chapter aims to produce land plant-like and algal-like PCO sequences, to explore algal PCO properties.

4.2 Expression of putative algal PCOs

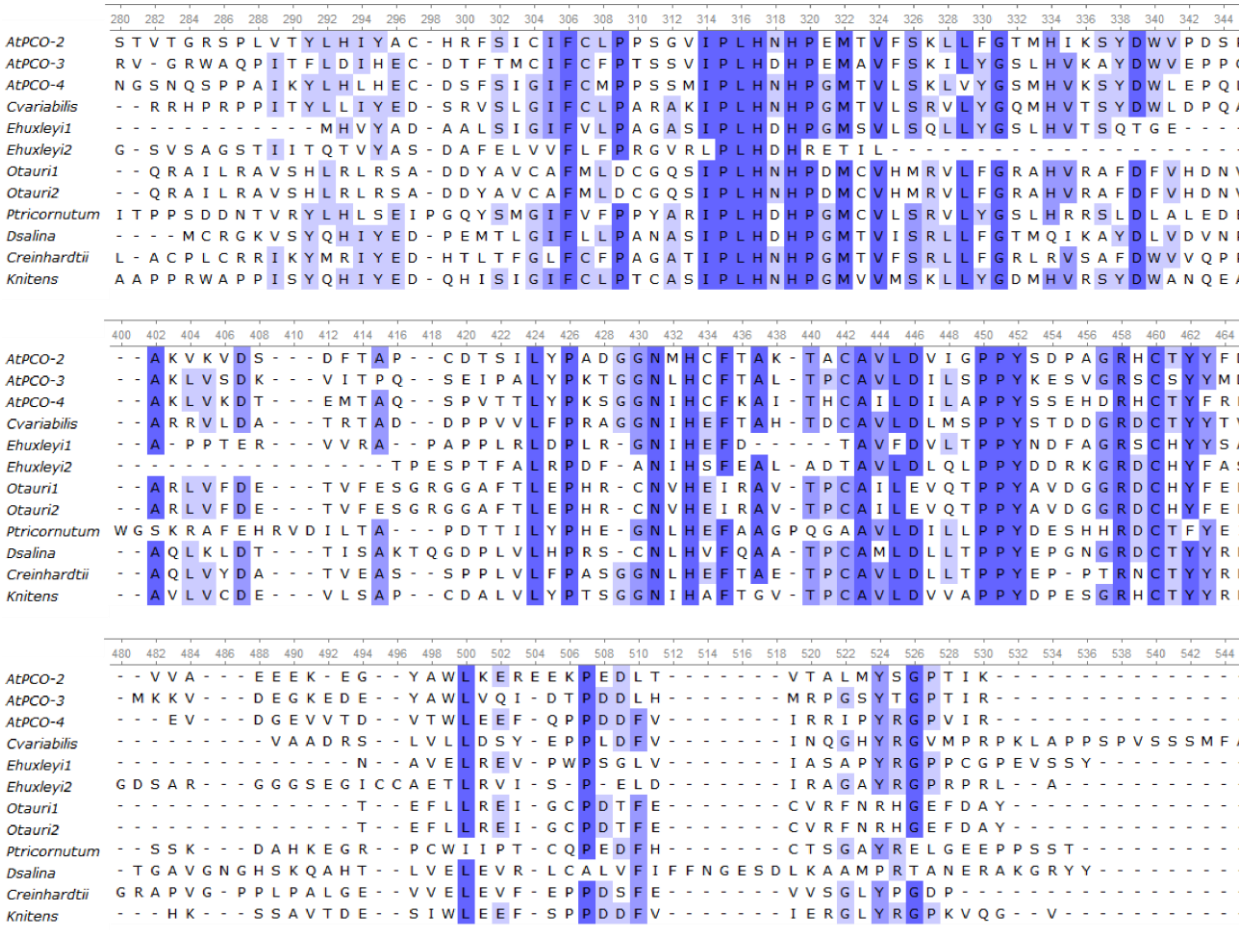


Fig. 4.1: Alignment of putative algal PCOs. PCO sequences are from *E. huxleyi* (EhPCO1/2), *O. tauri* (OtPCO 1/2), *P. tricornutum* (PtPCO), *D. salina* (DsPCO), *C. variabilis* (CvPCO), *K. nitens* (KnPCO) with PCOs from Marchantia (MpPCO) and Arabidopsis (AtPCO2, 3 and 4) for comparison.

4.2.1 Identification of putative algal PCOs

Novel algal PCOs were identified *via* transcriptomic and genomic online Blastp alignment search tools, exploiting the high degree of sequence conservation exhibited by the PCOs (Fig. 4.1). At the time of this search, only 18 algal genomes were fully sequenced (Table 1). Most genomes were screened using The Basic Local Alignment Search Tool (BLAST) provided by the National Centre for Biomedical Information (NCBI,

https://blast.ncbi.nlm.nih.gov/Blast.cgi), with *K. nitens* transcriptomic data accessed *via* the *K. nitens* Genome Project Portal:

http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/klebsormidium_blast.html

The BLASTp suites enabled high throughput genomic screens using previously described AtPCO sequences [133], namely AtPCO 2, 3 and 4. The chosen AtPCO sequences represent constitutively expressed (AtPCO 4), hypoxia-induced (AtPCO2) and a kinetically and structurally distinct form of the AtPCOs (AtPCO3). 15 of the 18 algal genomes contained partial or full protein sequences with homology to the AtPCOs. 9 full length proteins, from 7 algal species were selected for gene synthesis and recombinant expression trials (Table 4.1, highlighted). These sequences were chosen as they appeared to be full-length proteins, not partial sequences. The *Dunaliella salina* PCO (DsPCO) sequence was provided by F. Licausi. See Appendix 4.1 in the supplementary digital files for putative algal PCO sequences and ID matrix.

4.2.2 Preparation of algal PCO sequences in the pET-28a expression vector

With the exception of DsPCO, all protein sequences were synthesised and subcloned into the pET-28a vector by Genscript®. Sanger sequencing verified the pET28a plasmid contained the respective algal PCO coding sequence fused to a His_6-tag with a thrombin cleavage site on the N-terminus for affinity purification.

The DNA sequence encoding for DsPCO was provided in a pDEST17 plasmid. To facilitate protein purification by affinity chromatography, the DsPCO sequence was subcloned into the pET28a plasmid (Chapter 2, Section 2.1.1.1). PCR amplification of the template DsPCO DNA was verified by 2% agarose gel (Appendix 4.2), followed by *DpnI* digestion of the remaining template DNA. The pET28a plasmid was linearized *via* the *NdeI* and *XhoI* restriction sites (Appendix 3.1 and 4.2). The DsPCO DNA fragment was amplified and then inserted into the pET28a vector using an In-Fusion® HD cloning kit. *E. coli* XL10 Gold® ultra-competent cells were transformed with the resulting recombinant DNA. Resultant colonies were selected at random for DNA amplification and purification *via* the GeneJET Plasmid Miniprep Kit. Four of the six colonies selected contained plasmid DNA (Appendix 4.2). Sanger sequencing of the purified DNA plasmids verified the pET28a plasmid contained the DsPCO coding sequence.

4.2.3 Expression trials of algal PCOs

A series of small and large scale expression trials examined whether recombinant algal PCOs could be expressed in *E. coli*, as recombinant land plant-like PCOs express easily in bacteria. All algal PCOs were examined using a small scale expression trial with DsPCO, CrPCO and EhPCO 1 and 2 taken forward for large scale trials. The small scale protocol is outlined below. The protocol is included as it is an important consideration when examining the results, and may rationalise why many of the expression trials were unsuccessful. The large scale expression protocols follow those outlined in

Chapter 3, Section 3.1. The protocols are described in full in Chapter 2, Section 2.1.4.

4.2.3.1 Small scale expression trial protocol

4.2.3.1.1 Transformation

E.coli BL21(DE3) was the cell-type of choice unless otherwise stated. Transformations were carried out using a standard operating procedure (Chapter 2, Section 2.1.3). As a negative control, BL21(DE3) cells were also transformed with an empty pET28a vector. Glycerol stocks were prepared by incubating a transformed colony overnight at 37 °C in 2TY media and 30 µg/mL kanamycin sulfate. The cells were harvested by centrifugation and re-suspended in 2TY media and 50% glycerol. Glycerol stocks were stored at -80 °C.

4.2.3.1.2 Pre-culture

To propagate the transformed cells from glycerol stocks in preparation for recombinant protein expression, transformed cells were incubated overnight 37 °C in 2TY media and 30 µg/mL kanamycin sulfate. Cells transformed with an empty pET28a vector were grown alongside cells transformed with a PCO-encoding plasmid.

4.2.3.1.3 Protein expression

100 mL expression trials were conducted using *E.coli* BL21(DE3) competent cells unless otherwise stated. 30 µg/mL kanamycin sulfate was used as an

antibiotic against non-transformed cells. Fe(II) addition was thought to positively affect AtPCO 4 expression and as such three protein expression conditions were tested: 1) with 100 µM $Fe(NH_4)_2(SO_4)_2.6H_2O$, 2) No additional $Fe(NH_4)_2(SO_4)_2.6H_2O$ and 3) 100 µM of $Fe(NH_4)_2(SO_4)_2.6H_2O$ added after IPTG addition. Cultures were grown at 37 °C to an optical density of ~0.8 at 600 nm before inducing protein expression with 0.5 mM IPTG. These cultures were then grown overnight at 15 and 20 °C.

4.2.3.1.4 Cell lysis

1 mL aliquots of the cultures were sampled before IPTG induction of protein expression and after overnight incubation for lysis and analysis. The cultures were centrifuged at 10,000 x g for 10 min at 4 °C to harvest the cells. Cells were lysed using sonication in 50 mM Tris, pH 7.5 containing DNAse I and EDTA-free protease inhibitor cocktail. The lysed cell solutions were centrifuged at 16,000 x g for 20 min at 4 °C to separate the cell debris and insoluble protein from the soluble proteins. Both the pellet and the supernatant were analysed for the presence of recombinant PCO-like proteins.

4.2.3.1.5 Analysis

Soluble (supernatant) and insoluble (pellet) samples were analysed using SDS-PAGE, staining the resulting SDS-PAGE gels for proteins (Coomassie stain) and His_6-tags (InVision™ His-Tag In-Gel stain, Thermo Fisher). His_6-tag Stain Reagent is a fluorescent stain for preferential detection of His-tagged recombinant fusion proteins in protein polyacrylamide gels. As the proteins are

fused to a His_6-tag, this will allow for their visualisation on an SDS-PAGE gel. Stained gels were imaged on the Bio-Rad ChemiDoc™ XRS+ system. Initially, in-house cast Bis-tris polyacrylamide gels were used for analysing samples (see Appendix 4.3, A and B). Invitrogen 4-12% NuPAGE™ Bis-Tris protein precast gels were used for all subsequent analyses.

4.2.3.2 Small scale expression trial: O. tauri PCO 1 and 2 (OtPCO1/OtPCO2)

The *O. tauri* genome nominally contains two PCO-like sequences, but both are identical save for an additional 12 amino acid insert near the N-terminal region (Appendix 4.1). The expression of both isoforms was attempted. The expected molecular weight (MW) of His_6-OtPCO 1 and His_6-OtPCO 2 were 27412.54 Da and 26183.16 Da respectively. Expression trials suggested that both proteins were expressed, but were not soluble (Appendix 4.3 and 4.4).

4.2.3.3 Small scale expression trial: C. variabilis PCO (CvPCO)

The *C. variabilis* genome contains one putative PCO isoform. The expected MW of His_6-CvPCO is 32533.46 Da. SDS-PAGE analysis suggested no protein of that MW was expressed in the soluble or insoluble phase (Appendix 4.5). The quality of the insoluble analysis was low, so this result is difficult to interpret.

4.2.3.4 Small scale expression trial: P. tricornutum PCO (PtPCO)

The *P. tricornutum* genome contains one putative PCO isoform. The expected MW of His_6-PtPCO is 40275.80 Da. Analyses suggests His_6-PtPCO was expressed, but was not soluble (Appendix 4.6). Large quantities of insoluble

His_6-PtPCO were expressed, and could therefore be an excellent candidate for expression trials with the addition of solubility tag.

4.2.3.5 Expression trials: E. huxleyi PCO 1 and 2 (EhPCO 1/ EhPCO 2)

4.2.3.5.1 Small scale expression

The *E. huxleyi* genome contains two putative PCO isoforms. The expected MW of His_6-EhPCO1 is 17758.03 Da, and His_6-EhPCO2 has an expected MW of 22914.70 Da. Initial analyses suggested neither soluble nor insoluble His_6-EhPCO 1 was overexpressed (Appendix 4.7). SDS-PAGE results suggest His_6-EhPCO 2 was not expressed in either the soluble phase, but the insoluble results cannot be confirmed due to the low quality of the SDS-PAGE results (Appendix 4.13). *E. huxleyi* is of particular scientific interest due to its ecological success and influence on the carbon balance of the atmosphere and oceans. Studying PCOs from *E. huxleyi* could give great insights into algal PCO function, and as such EhPCO 1/2 were taken forward for more in-depth expression trials. A solubility enhancer tag known as thioredoxin (TRX) was fused to the EhPCO 1 and EhPCO 2 proteins *via* a pNHTrxT plasmid which encodes for an N-terminal His_6-TRX-TEV cassette (Appendix 4.8). TRX tags can help with minimizing inclusion body formation, possibly making EhPCO 1 and EhPCO 2 proteins available for expression, purification and characterisation.

4.2.3.5.2 pNHTrxT plasmid preparation

The EhPCO 1/2 sequences were ligated into the pNHTrxT plasmid *via* ligation independent cloning (LIC, Chapter 2, Section 2.1.1.2). PCR amplification of the template DNA was verified by 2% agarose gel (Appendix 4.8), followed by *DpnI* digestion of the remaining template DNA. The pNHTrxT plasmid was linearized *via* the *BsaI* restriction site (Appendix 4.9). The EhPCO1/2 DNA fragments were inserted into the linearized pNHTrxT vector using T4 DNA polymerase. *E. coli* XL10 Gold® ultra-competent cells were transformed with the resulting recombinant DNA. Resultant colonies were selected at random for DNA amplification and purification *via* the GeneJET Plasmid Miniprep Kit. Sanger sequencing of the purified DNA plasmids verified the pNHTrxT plasmid contained the EhPCO1/2 coding sequences.

4.2.3.5.3 Western blot of His_6-TRX-EhPCO 1 small scale expression

Another cell type, Rosetta 2(DE3) pLysS, was trialled for EhPCO expression as observation of *E. coli* cultures after overnight growth indicated BL21(DE3) cells were dying. This was likely due to the bacteriophage infection that remained undiagnosed for another year. At the time of these trials, it was thought perhaps the EhPCOs might be toxic to *E. coli*. A small scale expression trial was carried out in BL21(DE3) and Rosetta 2(DE3) pLysS cells. The expected MW of His_6-TRX-EhPCO 1 is 29608.66 Da. SDS-PAGE analysis of the Rosetta 2(DE3) pLysS expression did not indicate successful His_6-TRX-EhPCO 1 expression (Appendix 4.10, A and B). To ascertain if any His_6-TRX-EhPCO 1 was expressed,

the small scale expression trial was analysed by western blot using an antibody for the His_6-tag (Appendix 4.10, C and D). The antibody will detect expressed levels of the His_6-tag and therefore the His_6-TRX-EhPCO 1 protein. The western blot suggested His_6-TRX-EhPCO 1 expression was more successful in the Rosetta 2(DE3) pLysS cell type, than the BL21(DE3). Results also suggest that Fe(II) addition might be inhibiting protein expression. To test this result, two large scale expressions of His_6-TRX-EhPCO 1 were conducted, with and without Fe(II) addition using the Rosetta 2(DE3) pLysS cell type.

4.2.3.5.4 Large scale expression of His6-TRX-EhPCO 1 with added Fe(II)

His_6-TRX-EhPCO 1-transformed Rosetta 2(DE3) pLysS cells were grown in the presence of 100 µM $Fe(NH_4)_2(SO_4)_2.6H_2O$. 600 mL cultures were grown to an optical density of ~0.8 at 600 nm before inducing protein expression with 0.5 mM IPTG. Induced cultures were agitated and grown overnight at 20 °C. The cell pellet yield was 8.6 g/L. The pellets were stored at – 80 °C. The lysate was prepared by re-suspending a third of the harvested pellet in HisTrap™ HP binding buffer, DNAse I and protease inhibitor cocktail, followed by sonication, centrifugation and filtration prior to loading on to the HisTrap™ HP column (Chapter 2, Sections 2.1.6-2.1.7). The protein was eluted over a 0-100% gradient of elution buffer. Fractions containing His_6-TRX-EhPCO 1 were identified using SDS PAGE and western blot (Appendix 4.11). Western blot results suggest some of the His_6-TRX-EhPCO 1 protein was expressed, but this

quantity of protein is not enough to pursue gel filtration. Gel filtration is a crucial step in obtaining pure protein for kinetic and structural analysis.

4.2.3.5.5 Large scale expression of His_6-TRX-EhPCO 1 no added Fe(II)

To improve the yield of His_6-TRX-EhPCO 1, a second expression trial in the absence of Fe(II) was carried out. This expression trial was conducted as described above, but without the additional 100 µM $Fe(NH_4)_2(SO_4)_2.6H_2O$. The cell pellet yield was 8.3 g/L Results suggest an appreciable level of His_6-TRX-EhPCO 1 expression, but with a high degree of degradation (Appendix 4.12, A and B). This was thought to be due to the handling of the SDS-PAGE samples prior to analysis, with samples not being refrigerated adequately or analysed quickly enough post-purification; it was also thought that perhaps the purification protocol was too long, compromising protein integrity. A second, quicker purification protocol in the presence of 1 mM phenylmethylsulfonyl fluoride (PMSF) protease inhibitor (Thermo Scientific) and protease inhibitor during lysate preparation showed far less degradation post-purification, but His_6-TRX-EhPCO1 expression levels remained low, comparable to expression levels in the presence of Fe(II) (Appendix 4.12, C and D). While protein degradation was reduced, the quantity of protein produced was not enough to pursue gel filtration purification.

4.2.2.5 Large scale expression trial: D. salina PCO (DsPCO)

The sequence for DsPCO was received at a later stage in the expression trials, after some expression results had suggested expression trial outcomes

improve when a larger expression scale is used (see Chapter 3, Section 3.2.2). 3 x 600 mL cultures of *E. coli* transformed with the His_6-DsPCO plasmid were grown in 2 TY media were grown overnight at 18 °C alongside His_6-MpPCO. It was noted that the cultures expressing His_6-DsPCO had a lower cell density compared to the His_6-MpPCO cultures, although the cell pellet yield was not noted.

Following cell lysis, the soluble lysate was purified by HisTrap™ HP affinity chromatography (Appendix 4.14, A). The intensity of the second elution peak on the HisTrap™ UV-chromatogram suggested soluble His_6-DsPCO expression levels were low. This observation was supported by subsequent SDS-PAGE analysis; no overexpressed His_6-DsPCO band at 26857.57 Da was observed on the SDS-PAGE gel. Nevertheless, HisTrap™ Peak 2 fractions were buffer exchanged into gel fitration buffer using a PD-10 desalting column prior to purification with the HiLoad 26/600 Superdex 75pg column (Appendix 4.14, B, Chapter 2, Section 2.1.9). The gel filtration UV-chromatogram did not have the familiar double purification peak typical of PCO purifications. SDS-PAGE results also suggest that if His_6-DsPCO is present, expression levels are low.

4.2.2.6 Large scale expression trial: C. reinhardtii PCO (CrPCO) and $CrPCO_{149-406}$

6 x 600 mL cultures of *E. coli* transformed with the His_6-CrPCO and His_6-$CrPCO_{149-406}$ plasmids in 2 TY media were grown overnight at 18 °C. Two CrPCO constructs were chosen for expression trials as the CrPCO WT N-terminus had strings of repeating amino acid residues and it was thought a shorter construct

omitting this region would be easier to express and purify. Following cell lysis, the soluble lysate was purified by HisTrap™ HP affinity chromatography manually (Appendix 4.15). The lysate was loaded onto a 5 mL HisTrap™ column and washed with binding buffer. The column was then washed with 50 mM imidazole buffer, as it is as this concentration of imidazole that the observed HT Peak 1 elutes. This peak often contains *E. coli* impurities, and is therefore never taken forward for gel filtration. After washing with 50 mM imidazole buffer, the column was washed with elution buffer. SDS-PAGE analysis suggested neither His_6-CrPCO (Appendix 4.15, A, ProtParam predicted mass: 45699.02 Da) nor His_6-$CrPCO_{149-406}$ (Appendix 4.15, B, ProtParam predicted mass: 30206.66 Da) was overexpressed.

Even though no strong bands were visible on the SDS-PAGE analysis that indicated protein overexpression, it was uncertain as to whether or not *any* His_6-CrPCO or His_6-$CrPCO_{149-406}$ had expressed. It was thought an activity assay would help indicate if any PCO protein had expressed. The soluble eluted fractions were pooled, concentrated and buffer exchanged by centrifugation to gel filtration buffer. An activity assay with His_6-CrPCO WT and His_6-$CrPCO_{149-406}$ with $AtRAP2.12_{2-15}$ peptide for 1 hour showed no oxidation of the $AtRAP2.12_{2-15}$ peptide. These results suggest it is unlikely either soluble His_6-CrPCO WT or His_6-$CrPCO_{149-406}$ was expressed.

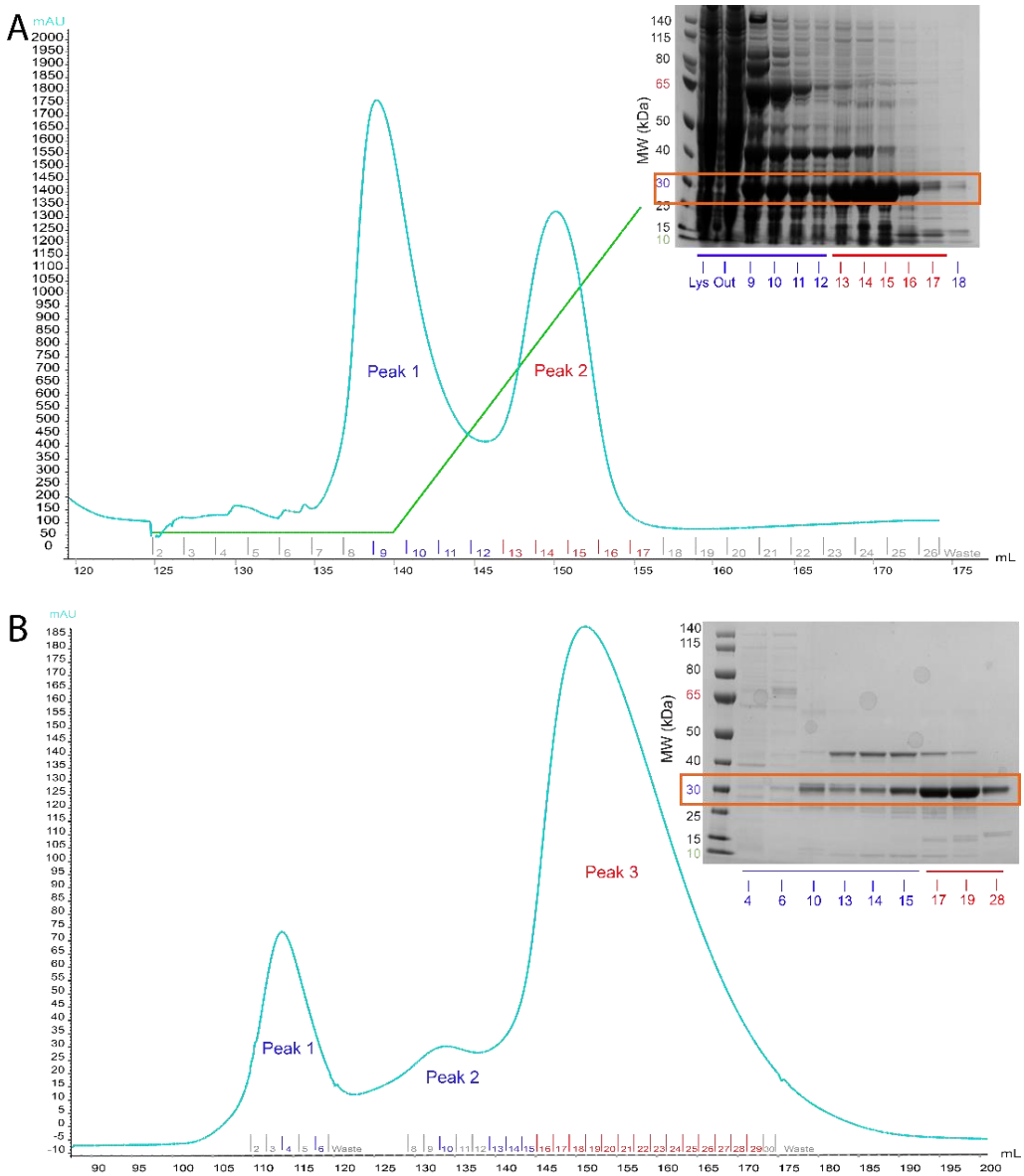


Fig. 4.2: Purification of recombinant His_6-KnPCO. A) Chromatogram and Coomassie stained SDS-PAGE of HisTrap™ purified *E. coli* BL21 DE3 lysate with overexpressed, recombinant His_6-KnPCO. B) Chromatogram and Coomassie stained SDS-PAGE of gel filtered HisTrap™ Peak 2 fractions. The protein bands of interest are highlighted

4.2.4 Large scale expression of His_6-KnPCO

The following sections describe the first successful expression, purification and characterisation of an algal PCO, KnPCO. Recombinant His_6-KnPCO with both thrombin and 3C His_6-tag cleavage sites were expressed in *E.coli* BL21(DE3) competent cells (Chapter 2, Sections 2.1.3-2.1.4). 3C protease is more active at

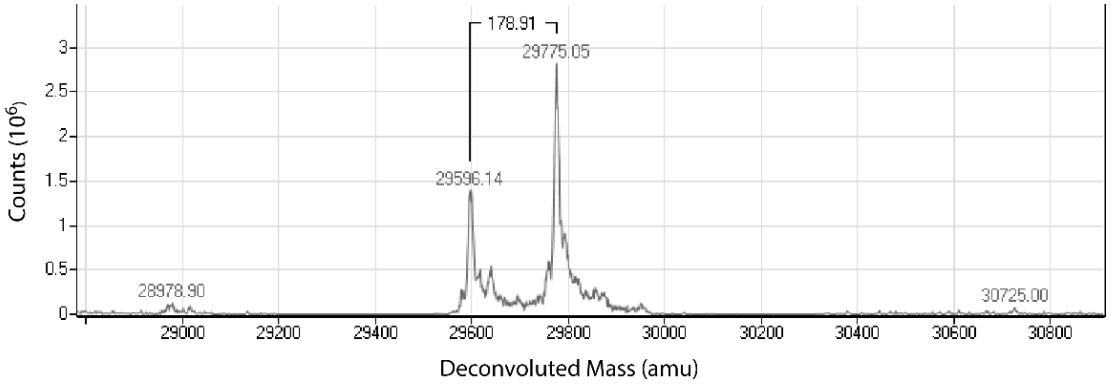


Fig. 4.3: Exact mass of His_6-KnPCO using RapidFire® Q-TOF MS. Exact His_6-KnPCO mass is 29596.1 Da, with post-translational alpha-N-6-phosphogluconoylation and subsequent de-phosphorylation (+178 Da) of the His_6-tag by *E. coli* proteins.

4 °C than thrombin protease, and as such was considered likely to be a better choice for the production of His_6-tag cleaved KnPCO for crystallisation trials. 600 mL cultures of *E. coli* transformed with the His_6-KnPCO plasmid in 2 TY media were grown overnight at 18 °C Cultures were grown to an optical density of ~0.8 at 600 nm before inducing protein expression with 0.5 mM IPTG. Induced cultures were agitated and grown overnight at 18 °C. Cell pellet yields varied between 2.9 and 9.2 g/L. Pellets were stored at – 80 °C.

4.2.5 Purification of His_6-KnPCO

Thrombin and 3C His_6-KnPCO were purified using HisTrap™ HP affinity chromatography and HiLoad 26/600 Superdex 75pg gel filtration chromatography. The lysate was prepared by re-suspending the harvested cells in HisTrap™ HP binding buffer, DNAse I and protease inhibitor cocktail, followed by sonication, centrifugation and filtration prior to loading on to the HisTrap™ HP column (Chapter 2, Section 2.1.6 - 2.1.7). The protein was eluted over a 0-100% gradient of elution buffer. Protein began to elute at

approximately 300 mM imidazole. Fractions containing His_6-KnPCO were identified using SDS PAGE (Fig. 4.2, A). HT Peak 2 fractions had fewer *E. coli* contaminants than HT Peak 1 fractions, and were therefore selected for gel filtration chromatography. HisTrap™ Peak 2 fractions 13-17 were buffer exchanged into gel filtration buffer using a PD-10 desalting column prior to purification with the HiLoad 26/600 Superdex 75pg column (Chapter 2, Section 2.1.9). Fractions from Peak 3 on the gel filtration chromatogram (GF Peak 3) had the highest proportion of pure His_6-KnPCO (Fig. 4.2, B). All subsequent protein characterisation used His_6-KnPCO from GF Peak 3.

Exact protein mass was determined using the Agilent RapidFire® auto-sampler and C4 cartridge coupled to a 6550 iFunnel Q-TOF MS. His_6-KnPCO exact mass was determined to be 29596.1 Da (Fig. 4.3), in agreement with the ProtParam predicted mass of 29597.39 Da. The +178 species is likely due to post-translational alpha-N-6-phosphogluconoylation and subsequent de-phosphorylation of the His_6-tag by *E. coli* proteins [239]. Size exclusion chromatography with multi-angle light scattering (SEC-MALS) showed GF Peak 3-harvested His_6-KnPCO was monomeric in solution, as the SEC-MALS eluted protein mass was within 7.8% of the exact His_6-KnPCO mass (Fig. 3.4, Chapter 2, Section 2.4.3). This is in agreement with MpPCO and previously characterised AtPCOs [142].

4.2.6 Determining homologous plant cysteine oxidase activity of KnPCO

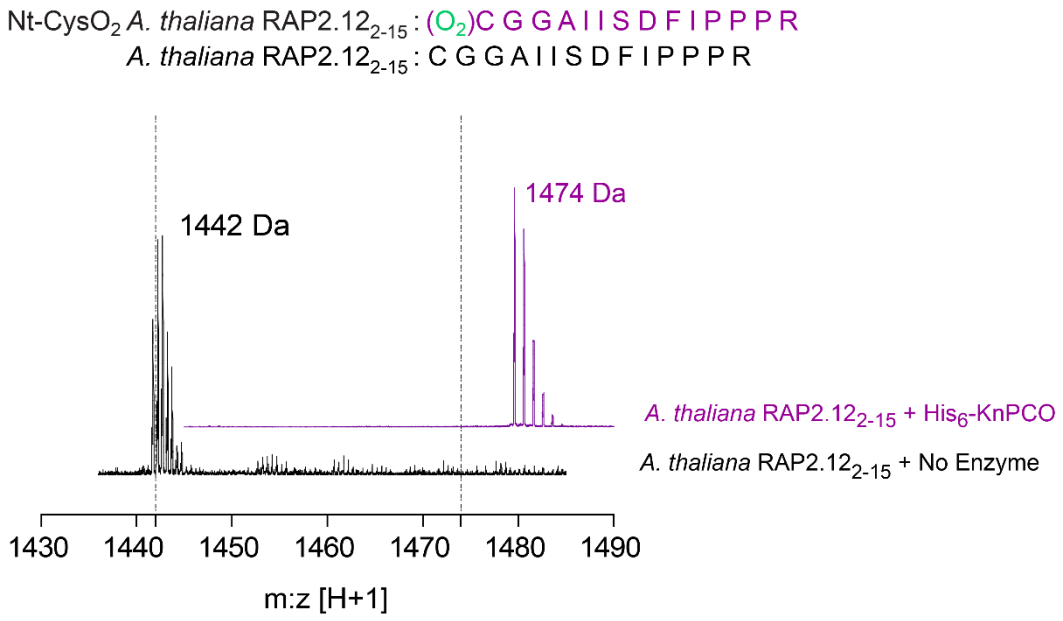


Fig. 4.4: Activity of His_6-KnPCO with AtRAP2.12_{2-15}. Mass spectrum showing AtRAP2.12 $_{2-15}$ incubated for 60 min with and without His_6-KnPCO, demonstrating a +32 Da shift in the presence of enzyme.

To be considered functionally homologous to known PCOs [142], His_6-KnPCO had to demonstrate Cys to $CysO_2H$ catalysis in an O_2 dependent manner. Initial His_6-KnPCO activity was determined by incubation with 14-mer AtRAP2.12_{2-15} peptide under atmospheric O_2 for 1 hour at 25 ˚C. Analysis on the Xevo G2-S Q-TOF UPLC/MS with a Chromolith® RP-18e 100-2 mm HPLC column showed a +32 Da mass shift on the AtRAP2.12_{2-15} peptide upon the addition of His_6-KnPCO, consistent with KnPCO-catalysed peptide di-oxidation (Fig. 4.4).

The O_2 dependence of His_6-KnPCO catalysed oxidation of AtRAP2.12_{2-15} was determined using a Brookes Mass Flow Controller, with peptide oxidation analysis performed on the Xevo G2-S Q-TOF UPLC/MS. AtRAP2.$12_{2-15,}$ TCEP and buffer were pre-equilibrated under the desired atmosphere for 10 min at

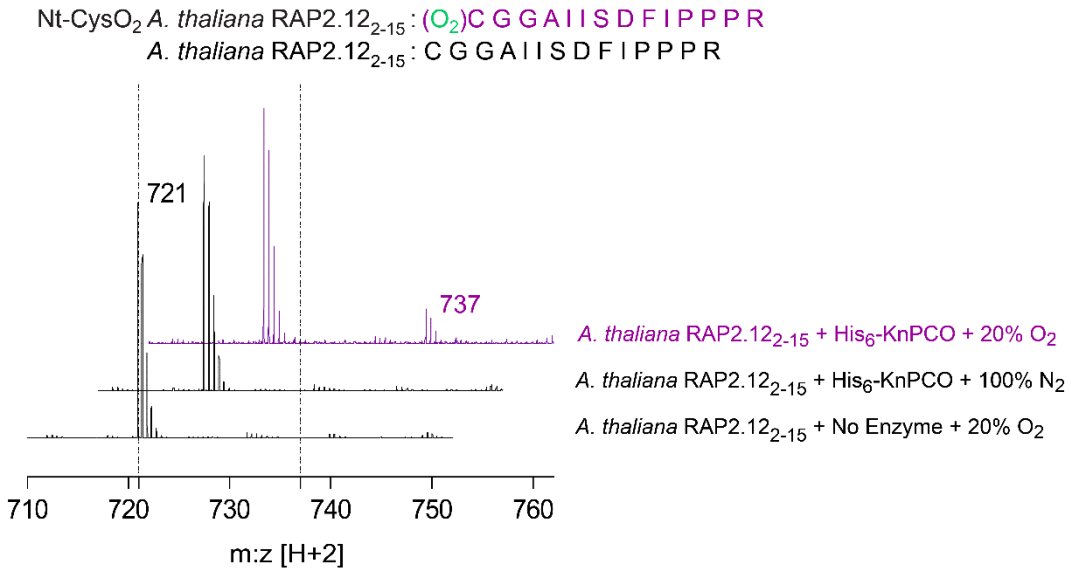
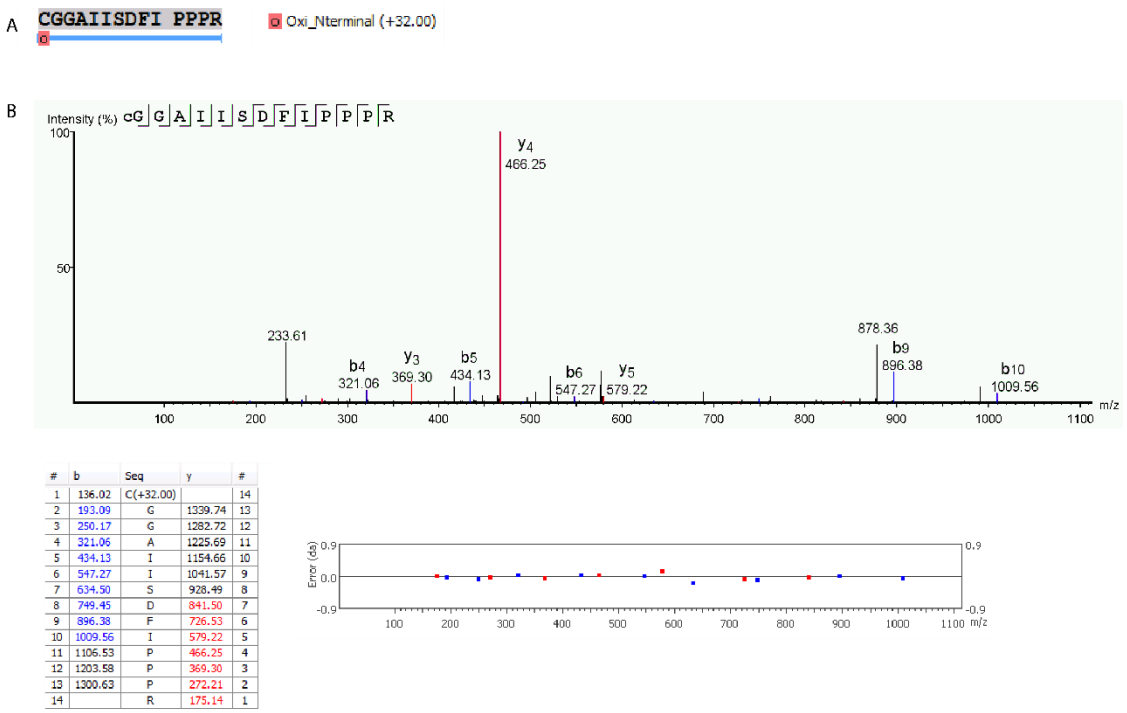


Fig. 4.5: O_2 dependence of His_6-KnPCO catalysis. MS analysis confined the O_2 dependence of the His_6-KnPCO AtRAP2.12_{2-15} reaction after 2 min. Here, only the +2 charged ion species was observed, so the 16 m:z [H+2] shift observed here corresponds to a + 32 Da mass change on the singly charged peptide.

25 ˚C before the addition of His_6-KnPCO, quenching with 1% (v/v) formic acid after 2 min. Under an atmosphere of 100% N_2, a +32 Da shift is not observed on the AtRAP2.12_{2-15} peptide in the presence of enzyme (Fig. 4.5). Similarly, under an atmosphere of 20% O_2 and no enzyme, a +32 Da shift is not observed on the AtRAP2.12_{2-15} peptide (Fig. 4.5). However in the presence of 20% O_2 and His_6-KnPCO, a +32 Da shift is observed on the AtRAP2.12_{2-15} peptide (Fig. 4.5) demonstrating the O_2 dependence of the His_6-KnPCO catalysed reaction (Chapter 2, Section 2.1.11.2). These results are in agreement with MpPCO and the AtPCO isoforms, where both O_2 and enzyme are necessary for the oxidation of the AtRAP2.12_{2-15} peptide [142], albeit with far less oxidation observed with the addition of His_6-KnPCO compared to His_6-MpPCO (Fig. 3.6).

Finally, the +32 Da modification was further confined to the N-terminal Cys of AtRAP2.12_{2-15} using nanoLC-MS/MS and PEAKS 8.5 sequencing software by E.



#	b	Seq	y	#
1	136.02	C(+32.00)		14
2	193.09	G	1339.74	13
3	250.17	G	1282.72	12
4	321.06	A	1225.69	11
5	434.13	I	1154.66	10
6	547.27	I	1041.57	9
7	634.50	S	928.49	8
8	749.45	D	841.50	7
9	896.38	F	726.53	6
10	1009.56	I	579.22	5
11	1106.53	P	466.25	4
12	1203.58	P	369.30	3
13	1300.63	P	272.21	2
14		R	175.14	1

Fig. 4.6: Tandem MS of the AtRAP2.12$_{2\text{-}15}$ peptide after His$_6$-KnPCO-catalysed oxidation. A) Full Sequence coverage (100%) of AtRAP2.12 (C(+32)GGAIISDFIPPPR) was achieved. Peptide sequencing software, PEAKS 8.5, listed the RAP2.12$_{2\text{-}15}$ N-terminal Cys as a "confident modification site" and present a -10lgP> 20. All peptides presenting any modification, spectra were manually checked and validated or disqualified. For C1 the identification is very reliable, as full fragmentation in the HCD is observed.

Pires (Fig. 4.6, Chapter 2, Section 2.2.12.2). This further demonstrates that KnPCO functions in a manner homologous to that of previously characterised PCOs, catalysing N-terminal Cys oxidation and confirms the first identification of a PCO in an algal species.

4.2.7 Determining His$_6$-KnPCO metal-binding *via* Inductively Coupled Plasma Mass Spectrometry

PCOs are Fe-dependent, but are known to co-purify with metal contaminants [1]. This section looks to determine KnPCO metal content for the purposes of informing kinetic characterisation conditions and homology to known PCOs. The metal content of His$_6$-KnPCO was determined by ICP-MS in the Trace Element ICP-MS Small Research Facility (Dr Phil Holdship, Earth Sciences

Chapter 2, Section 2.1.13.1). Fe, Zn and Ni were all found to bind in substoichiometric levels, with 0.27 (± 0.013) Fe, 0.08 (± 0.023) Zn and 0.23 (± 0.014) Ni atoms per monomer (Appendix 3.4 A-C). The ratio of bound Fe atoms to KnPCO is comparable with the 0.17 (± 0.031) Fe atoms/molecule of His_6-MpPCO and 0.31 (± 0.015) Fe atoms/molecule measured for AtPCO 4 [142]. These results will inform enzyme activity assays as a function of metal availability (Section 3.4.2.3), and further align KnPCO characteristics with previously described PCOs. A His_6-KnPCO_H172D variant was cloned to examine the contribution made by the putative His-facial triad to His_6-KnPCO metal binding and catalytic activity (Appendix 4.16) but unfortunately did not express successfully, suggesting a negative effect of this mutation on the structural integrity of the variant enzyme, potentially linked to its Fe-binding role.

4.3 Kinetic Studies of KnPCO

To compare the catalytic efficiency and O_2 sensitivity of KnPCO with other characterised PCOs, KnPCO kinetics were studied. The *K. nitens* genome does not contain any sequences with homology to the ERF or ZPR2 transcription factors, or an MC-initiating VRN2 homolog. This necessitated the use of non-endogenous substrates; MpERF and AtRAP2.12. Both *M. polymorpha* and *Arabidopsis* substrates were used, such that KnPCO kinetics could be compared to MpPCO (Chapter 3) and the AtPCOs [188].

His_6-KnPCO kinetics were assessed *via* quench assays and mass spectrometry such that initial rates of peptide turnover could be generated (Fig. 3.13 and

Appendix 4.17), satisfying the Michaelis-Menten model. All assays were analysed *via* the Agilent RapidFire® 6550 iFunnel Q-TOF MS (C4 cartridge) system. Real-time assays generated relative activity data for assay development, and quenched assays used for kinetics. Peptide oxidation was quantified using Agilent Quantitative Analysis software. Variability among replicates was quantified by the standard deviation. Enzyme aliquot concentrations were checked using the Nanodrop™ spectrophotometer prior to any experiment to ensure protein concentration consistency. All assays were conducted at n=3, unless otherwise stated.

4.3.1 Determining the optimum conditions for KnPCO kinetic analysis

This section focuses on establishing 'ideal' conditions for studying the His_6-KnPCO/$MpERF_{2\text{-}15}$ and His_6-KnPCO/$AtRAP2.12_{2\text{-}15}$ reactions for Michaelis-Menten kinetics.

4.3.1.1 Determining how TCEP and DTT affect His_6-KnPCO/substrate turnover

Peptide reducing agents, TCEP and DTT, were screened to see how they affected His_6-KnPCO activity relative to no added reducing agent (Fig. 4.7 A, B, Chapter 2, Section 2.3.2). For both His_6-KnPCO/$MpERF_{2\text{-}15}$ and His_6-KnPCO/$AtRAP2.12_{2\text{-}15}$ reactions, a greater degree of oxidation is observed in the presence of TCEP than DTT or without a reducing agent. TCEP was therefore selected for use in all kinetic assays.

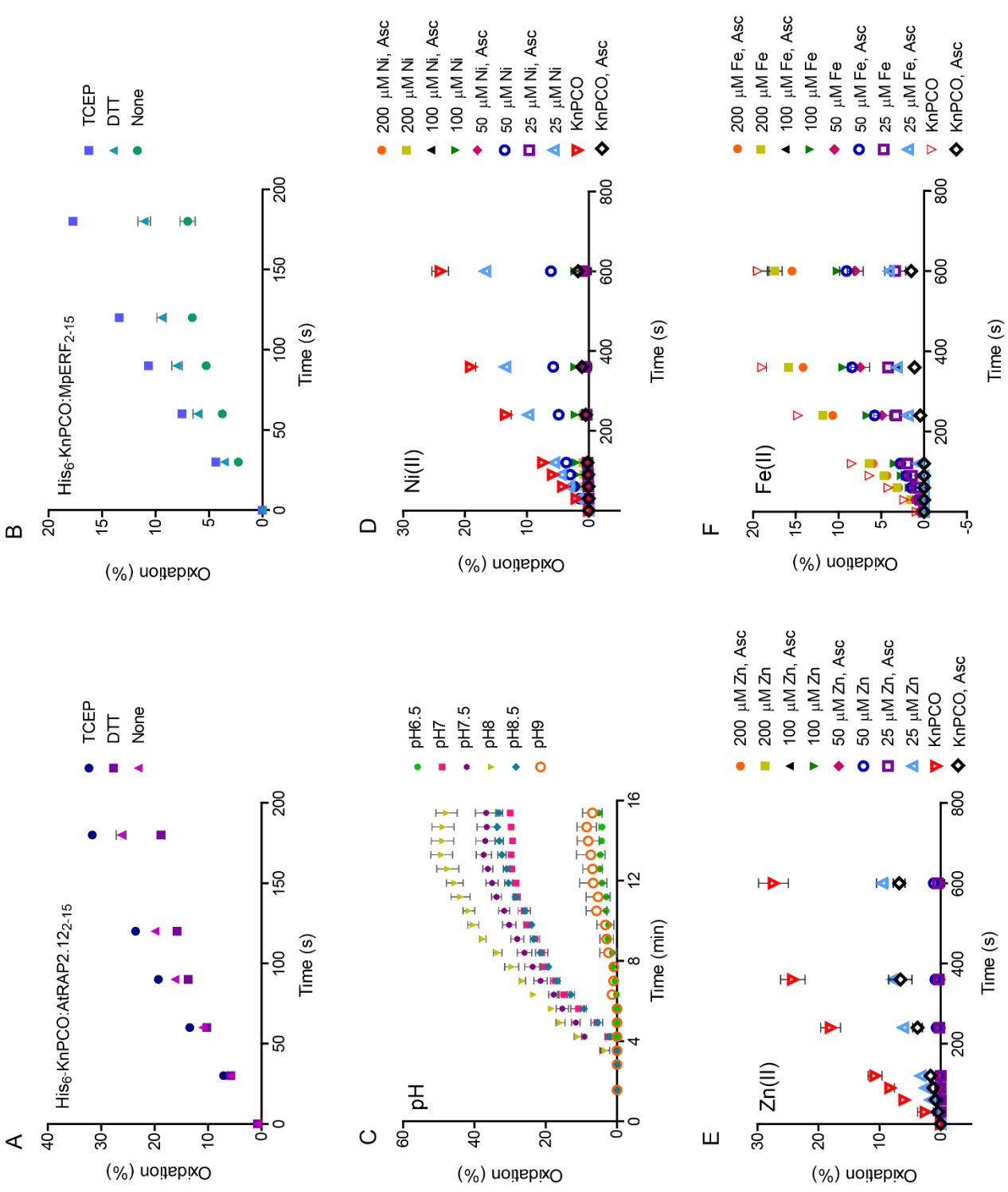


Fig. 4.7: Determining optimum His_6-KnPCO kinetic conditions. Activity profile of the A) His_6-KnPCO/AtRAP2.12_{2-15} and B) His_6-KnPCO/$MpERF_{2-15}$ complexes in the presence of DTT and TCEP. Activity of the His_6-KnPCO/$MpHRE_{2-15}$ reaction C) as a function of pH D) Ni(II) and ascorbate E) Zn(II) and ascorbate and F) Fe(II) and ascorbate. Error bars represent standard deviation, n=3.

4.3.1.2 Determining His_6-KnPCO pH dependence

The $MpERF_{2-15}$ peptide was chosen as the substrate for pH and metal dependency assays such that the data could be compared to His_6-MpPCO. Real-time peptide turnover as a function of pH was analysed on the RapidFire® (Chapter 2, Section 2.3.2). This method is unsuitable for kinetic studies, but is a useful technique for qualitative analyses such catalysis as a function of pH. 0.1 µM His_6-KnPCO was incubated with 4 µM MpERF and 1 mM TCEP, between pH 6 and 9, at room temperature (~29 ˚C) as the RapidFire® has no temperature control capabilities. The His_6-KnPCO/$MpERF_{2-15}$ reaction does not have as wide a pH tolerance as His_6-MpPCO/$MpERF_{2-15}$ (Fig 4.7, C). Activity levels drop steeply at pH 6.5 and 9, comparable with the considerable loss of activity of the AtPCO 4/$AtRAP2.12_{2-15}$ reaction at pH 6.5. Further kinetic studies were therefore done at pH 8, in keeping with previous AtPCO and MpPCO kinetic studies [188].

4.3.1.3 Determining His_6-KnPCO metal dependence via ICP-MS and activity profiling

ICP-MS studies of His_6-KnPCO showed Zn, Fe and Ni bound in substoichiometric levels (Section 4.2.5). Understanding how metals may affect catalytic activity is important for developing ideal assay conditions. $ZnCl_2$, $NiSO_4$ and $Fe(NH_4)_2(SO_4)_2.6H_2O$ were incubated at concentrations ranging from 25 to 200 µM, with and without 4 mM ascorbate, with 0.1 µM His_6-KnPCO, 50 µM $MpERF_{2-15}$ and 1 mM TCEP at 25 °C , with the reaction quenched in 1% (v/v) formic acid (Chapter 2, Section 2.3.2). In every instance, additional

metal and ascorbate inhibited His_6-KnPCO activity (Fig. 4.7 D-F), with 25 μM of Zn enough to severely affect turnover. Additional Fe(II) also had an inhibitory effect on His_6-KnPCO activity, similar to His_6-MpPCO.

These studies of His_6-KnPCO activity determined the ideal conditions for further kinetic analysis. The buffer chosen was 50 mM bis-tris propane, 50 mM NaCl pH 8 (the same pH used for the AtPCOs [188] and MpPCO), TCEP was selected as the peptide reducing agent, and no additional metal or ascorbate was added.

4.3.2 Michaelis-Menten kinetics of the His_6-KnPCO/AtRAP2.12_{2-15} and His_6-KnPCO/$MpERF_{2-15}$ reactions

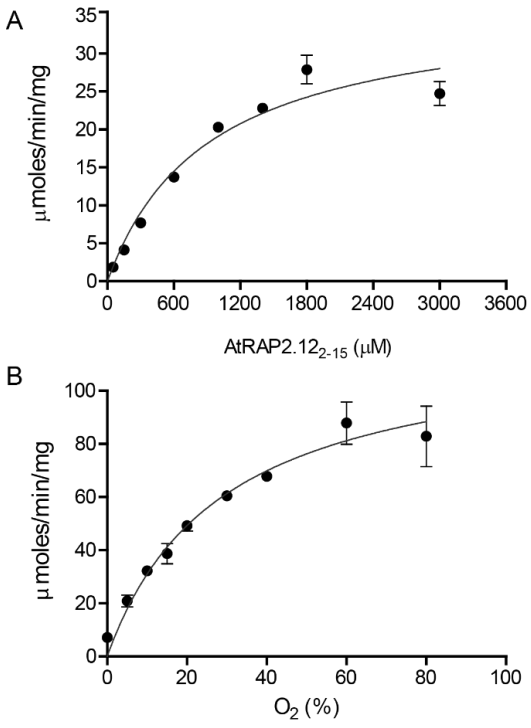


Fig. 4.8: Kinetics of His_6-KnPCO with the AtRAP2.12_{2-15} peptide. A) The K_M AtRAP2.12_{2-15} for His_6-KnPCO using a standard Michaelis-Menten fit was 879 ± 153 μM B) the K_M^{app} O_2 for the His_6-KnPCO/ AtRAP2.12_{2-15} reaction was 28.9 ± 4.1 % O_2. Error bars represent standard deviation, n=3

4.3.2.1 K_M AtRAP2.12_{2-15} and K_M $MpERF_{2-15}$

Kinetic assays were performed under optimised conditions to determine the relative catalytic efficiency of KnPCO towards both the *Arabidopsis* and *M. polymorpha* peptides. All assays were conducted in triplicate at 25 °C, in 50 mM bis-tris propane, 50 mM NaCl, pH 8 buffer, 1 mM TCEP and 0.1 μM

His_6-KnPCO. Assay samples were quenched in 1% (v/v) formic acid at regular time intervals.

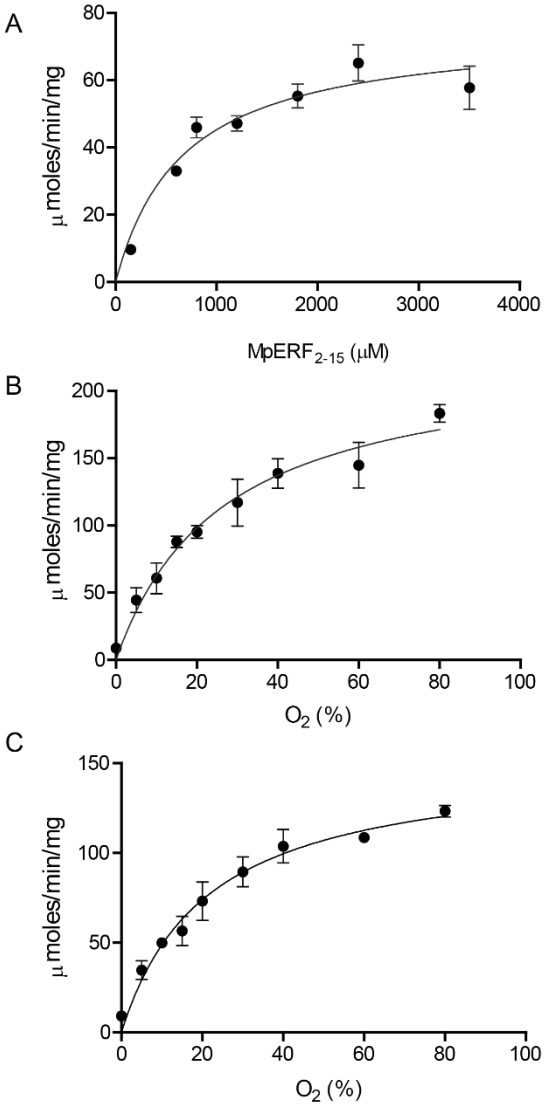


Fig. 4.9: Kinetics of His_6-KnPCO activity towards the $MpERF_{2-15}$ peptide. A) The K_{Mapp} $MpHRE_{2-15}$ for His_6-KnPCO was 830 ± 225 μM, error bars represent standard deviation, n=3 and B) the K_{Mapp} O_2 $MpHRE_{2-15}$ for His_6-KnPCO was 26 ± 2.6 % and C) reproduced at 21.6 ± 2.6 % O_2. Error bars represent standard deviation, n=3.

$AtRAP2.12_{2-15}$ concentrations varied from 50-3000 μM. Initial rates of activity are calculated per mg of Fe-occupied enzyme (active enzyme). The resulting rates of $AtRAP2.12_{2-15}$ turnover were fitted to a standard Michaelis-Menten model (Fig 4.8, A). The calculated K_M for $AtRAP2.12_{2-15}$ was 879 ± 153 μM (k_{cat} = 4.9 ± 0.34 s^{-1}). The point of maximum His_6-KnPCO activity was observed at 1.8 mM $AtRAP2.12_{2-15}$. This was therefore the chosen peptide concentration for the following K_M O_2 experiments. $MpERF_{2-15}$ concentrations were varied from 150 to 3500 μM, fit to a standard Michaelis-Menten model (Fig. 4.9, A). The calculated K_M for the $MpERF_{2-15}$ peptide was 830 ± 225 μM, with max activity at 2.4 mM peptide and a k_{cat} = 10.7 ± 1.04 s^{-1}. The K_M for both the

$MpERF_{2\text{-}15}$ and $AtRAP2.12_{2\text{-}15}$ peptides are high, indicating His_6-KnPCO has poor affinity for the peptides.

4.3.2.2 K_M^{app} O_2 of the His_6-KnPCO:$AtRAP2.12_{2\text{-}15}$ and His_6-KnPCO:$MpERF_{2\text{-}15}$ reactions

Protein	Substrate	k_{cat}	K_M	V_{max}
		s^{-1}	μM	$\mu mol\ min^{-1}\ mg^{-1}$
His_6-KnPCO	$AtRAP2.12_{2\text{-}15}$	18.0 ± 1.2	914 ± 144	36.55 ± 2.4
	$MpERF_{2\text{-}15}$	37.2 ± 2.3	664 ± 131	75.4 ± 4.6

Table 4.2: Steady-state kinetic parameters of His_6-KnPCO with $AtRAP2.12_{2\text{-}15}$ and $MpERF_{2\text{-}15}$ under atmospheric O_2. The turnover numbers (k_{cat}), Michaelis constants (K_M), and related maximum velocities (V_{max}) of KnPCO with respect to $AtRAP2.12_{2\text{-}15}$ and $MpERF_{2\text{-}15}$.

To understand how sensitive the His_6-KnPCO/$AtRAP2.12_{2\text{-}15}$ and His_6-KnPCO/$MpERF_{2\text{-}15}$ reactions were to O_2 availability, the K_M^{app} O_2 for both reactions was calculated. All assays were conducted in triplicate at 25 °C, in 50 mM bis-tris propane, 50 mM NaCl, pH 8 buffer, 1 mM TCEP with 0.1 µM His_6-KnPCO. The chosen concentrations of $AtRAP2.12_{2\text{-}15}$ and $MpERF_{2\text{-}15}$ are at the point of maximum enzyme activity in the K_M substrate experiments (Fig. 4.8 and 4.9, A), 1.8 mM and 2.4 mM respectively. This was to ensure peptide was not limiting during the K_M^{app} O_2 reactions. Reactions were quenched in 1% (v/v) formic acid after 1 min, a time known to be within the initial rate zone of the chosen peptide concentration as outlined in Fig. 3.14 (Appendix 4.17). O_2 concentrations varied between 2 and 80%. Peptide, TCEP and buffer were incubated for 10 min under the desired atmosphere using the Brookes Mass Flow Controller (MFC) prior to enzyme addition and reaction initiation.

Protein	Substrate	k_{cat}	K_M^{app}	V_{max}
		s^{-1}	%	$\mu mol\ min^{-1}\ mg^{-1}$
His_6-KnPCO	$AtRAP2.12_{2-15}$	59.3 ± 3.7	28.9 ± 4.1	120.2 ± 7.6
	$MpERF_{2-15}$	112.2 ± 6.8	26.3 ± 3.8	227.5 ± 13.8

Table 3: Steady-state kinetic parameters of KnPCO with O_2. The turnover numbers (k_{cat}), Michaelis constants (K_M^{app}), and related maximum velocities (V_{max}) of KnPCO with respect to $AtRAP2.12_{2-15}$ and $MpERF_{2-15}$.

The $K_M^{app}O_2$ for the His_6-KnPCO/$AtRAP2.12_{2-15}$ reaction was 28.9 ± 4.1 % O_2 (Fig 4.8, B). The $K_M^{app}O_2$ for the His_6-KnPCO/$MpERF_{2-15}$ reaction was 26 ± 2.6 % O_2 (Fig 4.9, B). Due to the reproducibility issues observed for the His_6-MpPCO/$MpERF_{2-15}$ $K_M^{app}O_2$ experiment (Chapter 3, Sections 3.4.3.2-3.4.3.3), the His_6-KnPCO/ $MpERF_{2-15}$ $K_M^{app}O_2$ was re-examined over 5 months later. The second calculated $K_M^{app}O_2$ value was within range of the first, at 21.6 ± 2.6 % O_2 (Fig 4.9, C), and therefore considered reproducible.

4.3.3 Screening His_6-KnPCO with other known PCO substrates

His_6-KnPCO, His_6-MpPCO and His_6-AtPCO 4 were incubated with $AtZPR2_{2-15}$. All assays were conducted in triplicate at 25 °C, in 50 mM bis-tris propane, 50 mM NaCl, pH 8 buffer, 1 mM TCEP with 0.1 μM enzyme. Analysis on the RapidFire Q-TOF indicated $AtZPR2_{2-15}$ was not oxidised after 2 min in the presence of His_6-KnPCO (Fig 4.10, A). To validate this finding, and ensure time was not limiting, a yes/no assay was conducted under the same conditions, but incubated for 1 hour with $AtRAP2.12_{2-15}$, $AtRAP2.3_{2-15}$, $AtVRN2_{2-15}$ and $AtZPR2_{2-15}$ (Fig. 4.10, B). This verified the previous observation, that $AtZPR2_{2-15}$ is not a

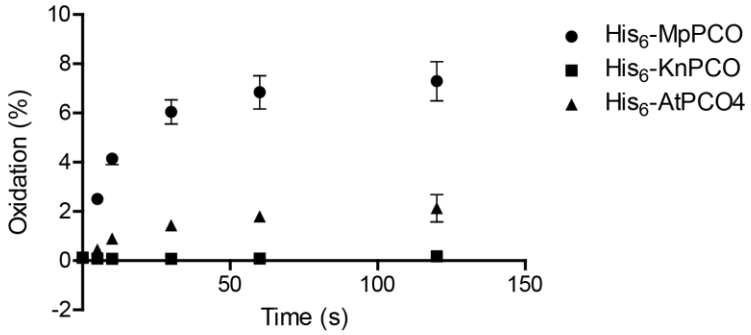
substrate of His_6-KnPCO. This is the first documented failure of a PCO to oxidise a 14-mer peptide with an N-terminal Cys residue.

4.4 Discussion

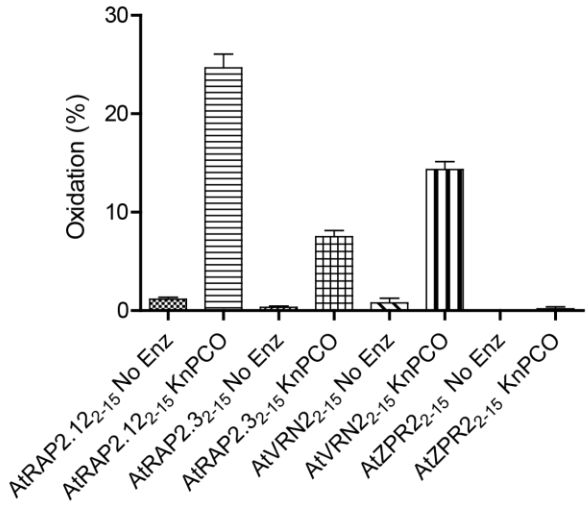
This chapter presents the expression trials of various algal-like PCOs. While many of the expression trials had limited success, there remains a number of promising avenues by which algal-like PCO proteins could be produced. Analysis suggested His_6-PtPCO and His_6-OTPCO 1 and 2 were successfully expressed, but were confined to inclusion bodies. These proteins look like promising candidates for solubility-tag expression trials. Some moderate success was had with the His_6-TRX-EhPCO 1 fusion protein, producing small quantities of soluble protein. This avenue has yet to be exhausted, and should be pursued further. Due to time constraints the His_6-TRX-EhPCO 2 expression trials were not attempted. These expression trials should also be considered in the context of the facilities used for these experiments. The fermentation facilities tested positive for a bacteriophage infection shortly after completing these studies. It is possible this infection affected expression capabilities and yield. As such, future studies could still successfully express soluble enzymes.

His_6-KnPCO, a single PCO isoform in the *K. nitens* genome, was successfully expressed, purified and characterised. His_6-KnPCO was shown to catalyse the oxidation of an N-terminal Cys to $CysO_2H$ in an O_2 dependent manner. His_6-KnPCO has substoichiometric Fe(II), Zn(II) and Ni(II) ions bound in solution, a comparable result to previously characterised PCOs; MpPCO, AtPCO1 and 4 [188].

A) Incubation of recombinant His_6-tagged PCOs with $AtZPR2_{2-15}$



B) Incubation of recombinant His_6-KnPCO with $AtZPR2_{2-15}$, $AtVRN2_{2-15}$ and $AtRAP2.13_{2-15}$



C) $AtZPR2_{2-15}$ seqeunce

$AtZPR2_{2-15}$: CLTTSEPPFPDTDT

Fig. 4.10: Incubating the $AtVRN2_{2-15}$, $AtZPR2_{2-15}$ and $AtRAP2.3_{2-15}$ peptides with His_6-KnPCO. A) Initial experiments indicated AtZPR2 was not a substrate of His_6-KnPCO. B) This result was repeated. C) the $AtZPR2_{2-15}$ sequence.

When the putative His_6-KnPCO Fe-binding His triad was disrupted, mutating a single His to Asp, the resulting His_6-KnPCO_H172D protein proved difficult to express and purify. Additional Fe(II), Zn(II) and Ni(II) metals in the His_6-KnPCO/peptide assay inhibit activity. The observed Fe(II) inhibition of His_6-KnPCO activity suggests metal uptake to the active site is likely difficult, a result

reflected by similar studies with MpPCO and AtPCO 4. Inhibition of activity due to additional Zn(II) and Ni(II) is also consistent with the AtPCOs [241], but are not thought to be catalytically relevant.

His_6-KnPCO activity is pH dependent, with activity with the $MpERF_{2\text{-}15}$ peptide severely impaired at pH 6.5 and 9. The His_6-MpPCO/$MpERF_{2\text{-}15}$ reaction had much greater activity at pH 6.5 compared to both AtPCO 4 and KnPCO. Maximum activity is observed at pH 8, a feature which is consistent across all PCOs studied thus far. Both $MpERF_{2\text{-}15}$ and $AtRAP2.12_{2\text{-}15}$ peptides have poor affinity for His_6-KnPCO, with K_M values above 800 μM. This stands in contrast with MpPCO, which also had a poor affinity for the $MpERF_{2\text{-}15}$ peptide (896 ± 268 μM), but had a great affinity for $AtRAP2.12_{2\text{-}15}$ with a K_M of 11 ± 1.1 μM. K_M for the $AtRAP2.12_{2\text{-}15}$ peptide varies significantly across the AtPCO isoforms, with AtPCO 3 and 4 having the highest K_M values 560 ± 60 μM and 690 ± 110 μM, respectively. The isoform with the lowest K_M value was for AtPCO 2 at 90 ± 10 μM. While the difference in peptide preference for KnPCO relative to MpPCO is interesting, it is not necessarily biologically relevant as both are non-endogenous substrates.

Despite poor affinity for the tested peptides, KnPCO has the greatest sensitivity to O_2 of any PCO enzymes tested thus far. KnPCO has a greater O_2 sensitivity than MpPCO, which had low $K_M^{app}O_2$ for both the $AtRAP2.12_{2\text{-}15}$ (3.5 ± 0.6 %) and $MpERF_{2\text{-}15}$ (8.7 ± 1.7 %) peptides, or AtPCO 4 with $AtRAP2.12_{2\text{-}15}$ (17.3 ± 1.69 %), the Arabidopsis PCO isoform with the highest $K_M^{app}O_2$ [188]. KnPCO

could provide insights into the molecular basis of PCO O_2 sensitivity and inform experiments to tune or modify PCO O_2 sensitivity. Sensitivity to O_2 here means sensitivity of the KnPCO-catalysed reaction to changes in O_2 concentrations. The reactions studied here, however, used exogenous substrates and may not be representative of KnPCO O_2 sensitivity with natural substrates.

Assuming the KnPCO O_2 sensitivity observed here is representative of its endogenous O_2 sensitivity, it implies that *K. nitens*, a terrestrial freshwater algae, has a use for an O_2 sensitive PCO. This is especially interesting in light of the presence of PRT and ATE homologs in the *K. nitens* genome – essential proteins in the higher plant N-degron pathway. This implies a highly O_2 sensitive PCO-regulated protein degradation pathway in *K. nitens*. It is unclear what this would mean for *K. nitens* from a physiological perspective, but considering its aquatic lifestyle, it seems at odds with the higher plant function of mediating cellular responses to acute hypoxia (often a as result of flooding). It is possible *K. nitens* may experience periods of chronic or acute hypoxia in an aquatic environment, for example as a result of overcrowding, or during periods of high metabolic demand.

While the *K. nitens* genome does not contain VRN2 or ZPR2 homologs (both involved in plant development), it is possible KnPCO has a role to play in mediating *K. nitens* development and growth as a function of O_2 availability. The finding that the AtZPR2$_{2\text{-}15}$ peptide is not a substrate of His$_6$-KnPCO is interesting. AtZPR2 is involved in Arabidopsis shoot meristem development –

a process for which *K. nitens* has no need as its cells are undifferentiated. This is the first documented failure of a PCO to oxidise a peptide with an N-terminal Cys residue. This finding could help us better understand PCO substrate evolution, and also help us better understand PCO substrate selectivity.

Chapter 5: Structural studies of MpPCO and KnPCO

5.1 Summary

With X-ray crystal structures now available for the Arabidopsis PCOs 4 and 5 [1], the structures of MpPCO and KnPCO could help inform both PCO structure evolution, and develop a better understanding of the PCO structure-function relationship. This chapter describes experiments aimed at generating structural information about MpPCO and KnPCO. Crystallisation trials were unsuccessful, but the process yielded useful protein preparation protocols for future use. Assessments of MpPCO and KnPCO thermal stability were made, alongside secondary structure determination *via* circular dichroism and PCO epitope assessments using antibody recognition. A series of modelling platforms were used to predict MpPCO and KnPCO secondary and tertiary structure using the AtPCO 5 crystal structure as a template.

5.2 Crystallisation trials and thermal stability studies of MpPCO and KnPCO

5.2.1 Introduction and protein preparation and crystallisation protocol

This section describes various KnPCO and MpPCO preparation methods, incorporating metal and His_6-tag removal, and preliminary crystallisation screens performed with the resulting proteins. The basic protein preparation and crystallisation screening methods used in Sections 5.2.2 and 5.2.3 are outlined below. If any deviations are made from the methods outlined here, they will be described accordingly. A diverse selection of commercial screens

were trialled to identify potential crystallisation conditions for both proteins. These screens, along with protein thermal stability assays, help inform the conditions necessary to crystallise target proteins. Neither MpPCO nor KnPCO crystallised successfully, and so additional protein preparation protocols were developed. These protocols focussed not only on generating protein in a specific desirable state, but also upon improving protein yields at each step.

All protein was expressed and purified as described in Chapter 2, Sections 2.1. Protein was overexpressed in BL21(DE3) *E. coli* cells, and lysed *via* sonication. Filtered lysate was loaded on to a 5 mL HisTrap ™ column for affinity purification. Purified protein fractions of interest were identified *via* SDS-PAGE, and desalted using a PD-10 gravity desalting column (Cytiva). The desalted fractions were loaded on to the HiLoad™ 26/600 Superdex™ 75 pg for size exclusion/gel filtration purification. Fractions containing protein were identified *via* SDS-PAGE and the purest fractions were used for subsequent structural work.

His_6-tag removal steps were inserted between the affinity purification and gel filtration steps. Either thrombin protease or 3C protease was used to cleave the His_6-tag from the recombinant protein *via* an integrated cleavage site between the His_6-tag and the protein. After overnight His_6-tag cleavage, the His-tagged protein/thrombin solution was manually purified using a 5 mL HisTrap™ column, separating His_6-protein and His_6-thrombin from His_6-cleaved protein (reverse HisTrap ™).

Metal removal steps were either incorporated as part of the gel filtration procedure, or done after gel filtration. Overnight dialysis steps followed gel filtration, either to remove any bound metal or to remove the metal chelation reagents used during gel filtration. For native MS, excess NaCl was removed from the protein samples during dialysis and using a Micro Bio-Spin® P-6 (BioRad) and a HiTrap™ IEX HP cation exchange column (Cytiva) after gel filtration.

The screening of crystallisation conditions is outlined in Chapter 2, Section 2.4.6. Crystallisation screens and protein samples were dispensed using a Phoenix Crystallization Robot (Art Robbins Instruments). A sitting drop vapour diffusion method was used in 96-well format (Art Robbins Instruments, 96 well low profile Intellipates™), trialling three separate protein/precipitant solution ratios (2:1, 1:1, 1:2) per well. The concentrated protein was centrifuged prior to laying crystallisation screens to remove any potential aggregated protein and particulates to avoid blocking the robot. The supernatant was carefully removed for use in the robot. The crystallisation screens were checked under a stereo zoom microscope (Leica M205 C) once a day for seven days, then every three days for two weeks and then once a week for three weeks to check for potential crystals.

5.2.2 Crystallisation trials: KnPCO

5.2.2.1 His_6-KnPCO crystallisation screens

His_6-KnPCO was expressed and purified as described in Chapter 2, Sections 2.1 (Appendix 5.1). The purest fractions were concentrated to 32 mg/mL and used for crystallisation trials without freezing. Room temperature screens were prioritised such that an anaerobic chamber could be used to screen anaerobic crystallisation conditions with the same crystallisation solution if necessary. His_6-KnPCO at a concentration of 32 mg/mL was screened using PACT premier ™ [242] (Molecular Dimensions) and SaltRx ™ HT (Hampton Research) crystallisation screens. Initial observations of the wells indicated this concentration was suitable for the SaltRx ™ HT screen, with heavy precipitate observed in over 50% of the drops. However, heavy protein precipitation in a high number of the wells indicated a 32 mg/mL concentration was too high for the polyethylene glycol (PEG)-containing screen, PACT premier ™. Another set of PACT Premier ™ and JCSG+ plates were laid using a protein concentration of 21.8 mg/mL. His_6-KnPCO was diluted again to 11.9 mg/mL and a second PACT premier ™ screen setup. The remaining 11.9 mg/mL His_6-KnPCO was used to setup JCSG+ [242, 243] (Molecular Dimensions), Morpheus ® I and II [244, 245] (Molecular Dimensions), Wizard™ Classic 6, Index ™ HT (Hampton Research), PEG-Rx ™ HT (Hampton Research) crystallisation screens. No crystals or clear promising crystallisation conditions were identified after 6 weeks of observation.

5.2.2.2 KnPCO preparation and crystallisation screens

Hexahistidine purification tags may contribute to protein heterogeneity in solution, resulting in reduced crystallisation success. As such, the His_6-tag was removed from KnPCO as an attempt to improve its crystallisability. Preparation of His_6-cleaved KnPCO was first attempted using a His_6-KnPCO protein with an integrated cleavage site that facilitates His_6-cleavage *via* 3C protease. The overall protein yield was too low to proceed to gel filtration after His_6-cleavage (Appendix 5.2). Subsequently, purification of a thrombin protease-cleaved KnPCO construct was prepared as described in Chapter 2, Sections 2.1.8.1 (Appendix 5.3). The His_6-cleaved KnPCO mass obtained using Q-TOF MS (27847.45 Da) was equivalent to the Protparam predicted mass (27846.53Da, Appendix 5.3, D), indicating that the His_6-cleavage was successful.

Following His_6-tag cleavage and gel filtration to remove the His_6-tag and thrombin protease, KnPCO protein was concentrated to 15 mg/mL and crystallisation trials were performed using JCGS+, PACT Premier™ and Morpheus II crystallisation screens. 16 days after setting-up the screens, it was observed that well D11 of the JCSG+ screen, had a pyramidal shape crystal and wells E2 and E4 of the Morpheus II screen had potential microcrystals. The conditions which resulted in formation of a pyramidal crystal were considered most promising for generating crystals that could be subjected to X-ray diffraction (0.14 M $CaCl.2H_2O$, 0.07 M Sodium acetate, pH 4.6, 14% v/v 2-propanol, 30% glycerol). However it appeared as though the original crystal had partially degraded (16 days after laying the plate) and was therefore not a

viable candidate for X-ray diffraction. Instead, an optimization plate formatted around this condition was set up to try and re-create and optimise the crystals.

The optimization plate varied both pH and 2-propanol concentration and used the degraded crystal as a seed for KnPCO crystal growth. The final optimization conditions included; 0.14 M $CaCl_2.2H_2O$, 0.07 M Sodium acetate, 30% glycerol, pH 4.4, 4.6, 4.8, 2-propanol 0, 4, 8, 12, 16 and 20% v/v. The original D11 crystal was transferred to a seed bead kit microfuge tube containing a small plastic ball (Hampton Research) and 0.5 µL crystallisation condition mother liquor (conditions in which the crystal was found) and vortexed to break-up the crystal. The resulting seed stock was used to manually seed a 24 well hanging drop vapour diffusion optimization plate (Hampton Research) with glass cover slips. The final drop volumes were 4 µL, with 1:1 protein/crystallisation solution. The plate was checked every day for two weeks and every three days for the following 6, no crystals were observed.

5.2.2.3 Metal-free His_6-KnPCO preparation and crystallisation screens

Due to a lack of viable crystals using His_6-KnPCO and His_6-cleaved KnPCO, preparation of a His_6-tag cleaved metal-free KnPCO ($KnPCO_{MF}$) protein was considered to be worthwhile. His_6-KnPCO is known to bind Fe, Ni and Zn in substoichiometric quantities (Appendix 3.4). In an effort to improve protein homogeneity for crystallisation, a metal-free His_6-KnPCO (His_6-$KnPCO_{MF}$) preparation protocol was trialled. Previous attempts at removing bound metals from PCOs have been unsuccessful - it is thought the metal binding site is

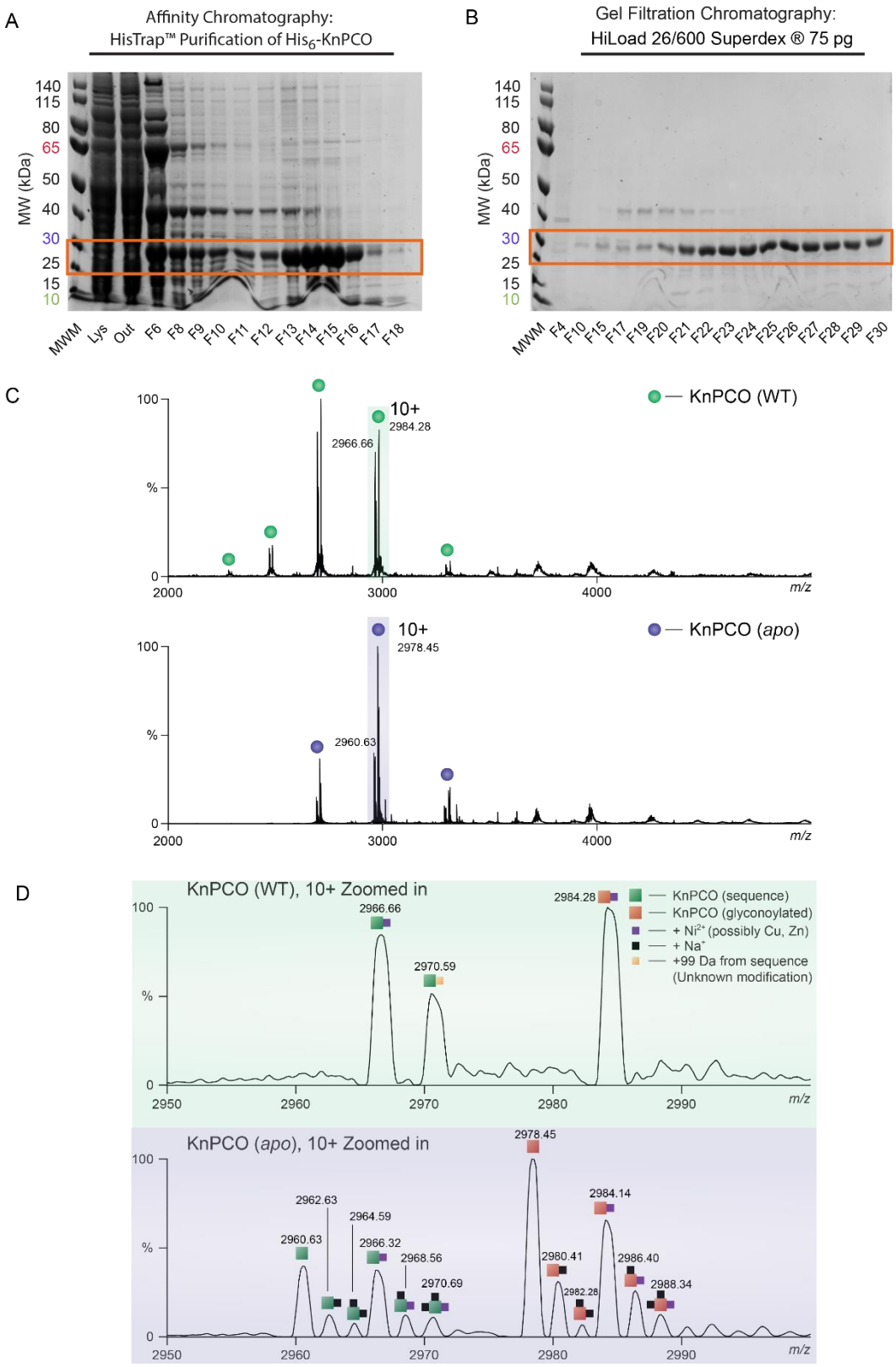


Fig. 5.1: Expression of His_6-KnPCO for crystallography and metal-free expression trials. A) SDS-PAGE of HisTrap™ purified *E. coli* BL21 DE3 lysate with overexpressed, recombinant His_6-KnPCO. B) SDS-PAGE of gel filtered HisTrap™ Peak 2 fractions. Fractions 23-36 were concentrated to a protein concentration of 21.8 mg/mL. Boxes indicate the protein bands of interest. PACT Premier and JCSG+ plates were laid at 21.8 mg/mL His_6-KnPCO. The remaining protein was treated with EDTA and Chelex overnight to remove bound metal. C-D) The "metal-free" His_6-KnPCO was analysed using native MS, alongside untreated His_6-KnPCO to determine the protein metal content post-treatment. Native MS data provided by D. Shutin.

difficult to access [1]. The following protocol aimed to produce metal-free His_6-$KnPCO_{MF}$ ($KnPCO_{MF}$) and analyse the success of the metal removal step using native MS.

His_6-KnPCO was expressed and purified as described in Chapter 2, Sections 2.1 (Fig 5.1 A and B). The gel-filtered His_6-KnPCO was dialysed overnight into 50 mM HEPES, 50 mM NaCl, 100 mM Ethylenediaminetetraacetic acid (EDTA), pH 8. EDTA is a known metal chelating ligand, and can be used to successfully remove protein-bound metals [246]. The EDTA-treated protein was dialysed overnight into 50 mM HEPES, pH 8 containing Chelex® 100 exchange resin (BioRad) to remove the NaCl (for native MS) and any remaining metal ions.

Native MS was carried out and the data processed by Mr D. Shutin (Chemistry Department, Fig 5.1, C and D). His_6-$KnPCO_{MF}$ was run alongside untreated His_6-KnPCO (His_6-$KnPCO_M$). Native MS indicated both His_6-$KnPCO_{MF}$ and His_6-$KnPCO_M$ were monomeric, in agreement with SEC-MALS analysis (Fig. 3.4). The spectrum resolution was too low to distinguish between Ni, Cu, Fe or Zn His_6-KnPCO metalation, but treatment with EDTA and Chelex ® 100 appeared to reduce the amount of metalated protein (Fig. 5.1, D). Despite the EDTA/Chelex® 100 treatment, a significant amount of transition metal ions remained bound to His_6-KnPCO. As such another metal removal protocol was trialled.

5.2.2.4 KnPCO$_{MF}$ preparation and ICP-MS

The native MS results from Section 5.2.2.3 suggest a high proportion of transition metal ions and Na ions bound to His$_6$-KnPCO after treatment with EDTA and Chelex® 100. 1, 10-Phenanthroline is a strong metal ion-coordinating ligand [247], and has been used successfully in removing protein-bound metals. Upon recommendation from Ms. P. Lang (Chemistry Department), a metal removal procedure using phenanthroline was incorporated into the gel filtration step of the purification protocol. SDS-PAGE analysis of the gel filtration fractions indicated soluble KnPCO survived gel filtration, even with the added phenanthroline step (Fig 5.2, C).

The gel filtration fractions were pooled, concentrated and dialysed overnight into 50 mM HEPES, 0.4 M NaCl, pH 7.5 to remove phenanthroline. Phenanthroline absorbs UV light, with a maximum at 260 nm wavelength, and as such can interfere with protein concentration measurements at 280 nm wavelength. The UV profile of dialysed KnPCO$_{MF}$ showed the main absorption peak was a 280 nm protein peak, with no peak observed at 260 nm (data not shown). As such, the measured protein concentration was thought to be accurate. However, the overall protein yield of approximately 2 mg was insufficient for crystallisation screens, compared with the 8 mg KnPCO yield from the same *E. coli* pellet used in Section 5.2.2.2.

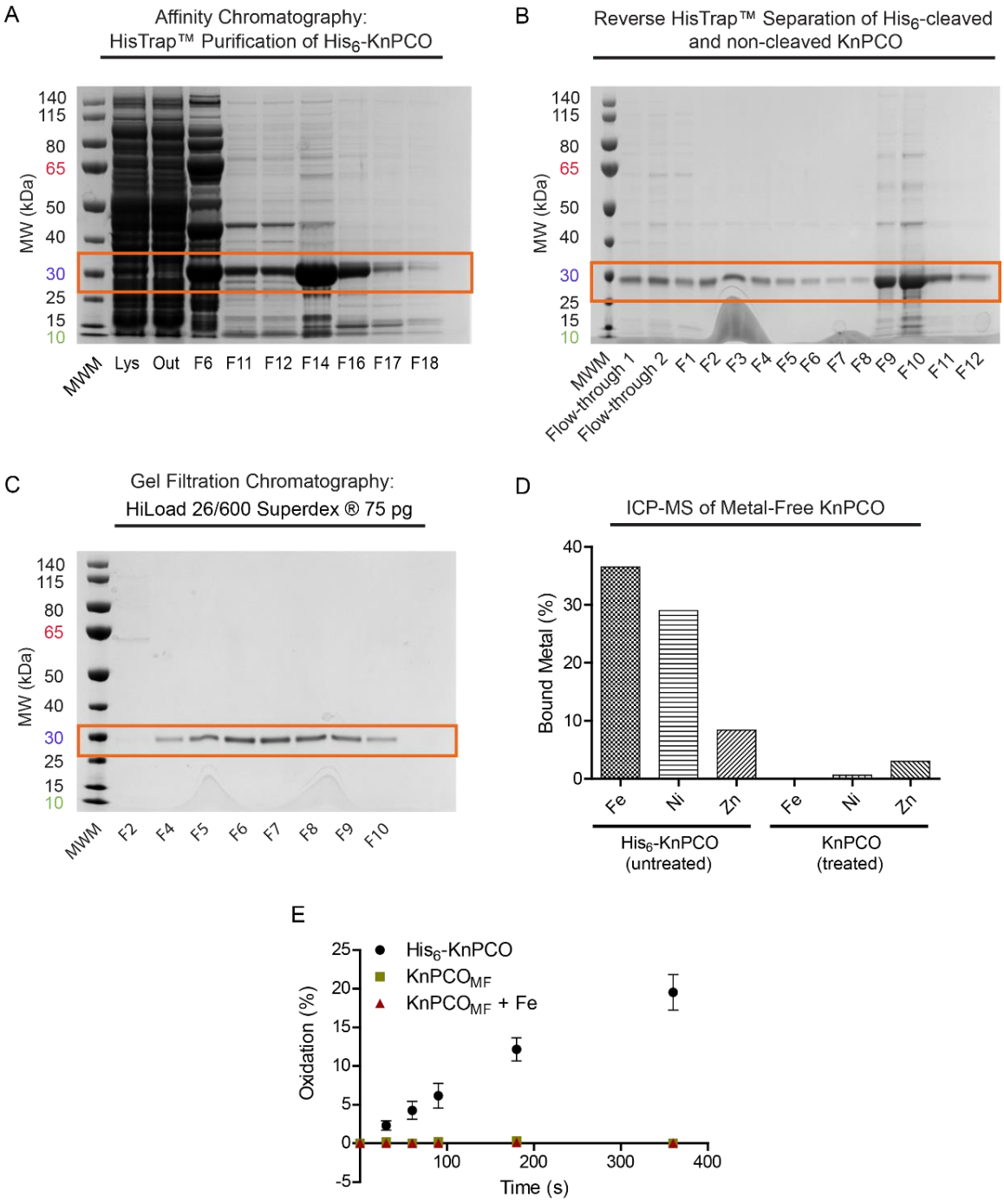


Fig. 5.2: Expression of His_6-cleaved metal-free KnPCO for crystallography. A) SDS-PAGE of HisTrap™ purified *E. coli* BL21 DE3 lysate with overexpressed, recombinant His_6-KnPCO. B) Reverse HisTrap™ of thrombin protease-treated recombinant His_6-KnPCO, separating His_6-cleaved KnPCO from non-cleaved (cleaved protein is assumed to elute prior to the column washes with high imidazole elution buffer, fractions 9-12). C) Reverse HisTrap™ flow-through fractions 1 and 2 and F1-F8 were purified by gel filtration. The gel filtration buffer contained 10 mM phenanthroline to remove bound metal. SDS-PAGE analysis confirmed some soluble KnPCO remained after treatment with phenanthroline D) The metal content of phenanthroline-treated, His_6-cleaved KnPCO was analysed alongside untreated His_6-KnPCO at n=1. ICP-MS data provided by Dr P. Holdsworth (Earth Science Department). E) Activity assay with $KnPCO_{MF}$ with the $AtRAP2.12_{2-15}$ peptide.

ICP-MS analysis of untreated His_6-KnPCO and $KnPCO_{MF}$ suggested the phenanthroline treatment successfully removed bound Fe, and significantly reduced bound Ni and Zn (Fig 5.2, D). This is the first protocol describing the successful removal of Fe from a PCO and could be a useful step to include in future purification protocols. Unfortunately, an activity assay with $KnPCO_{MF}$ indicated that enzymatic activity could not be reconstituted with additional Fe(II) (Fig. 5.2, E). This observation is in keeping with analyses of the AtPCO 4 and 5 crystal structures, which suggest the active site is buried and difficult to access [1].

5.2.3 Crystallisation trials: MpPCO

5.2.3.1 His_6-MpPCO crystallisation screens

As KnPCO failed to crystallise, it was thought prudent to try crystallising MpPCO. His_6-MpPCO was prepared as described in Chapter 2, Sections 2.1 (Appendix 5.4). Fractions with the fewest impurities were selected for screening crystallisation conditions. Selected fractions were concentrated to 60 mg/mL and used to set-up crystallisation screens. The JCSG+ screen was observed to have heavy precipitate in over 50% of the wells with His_6-MpPCO at 60 mg/mL. The JCSG+™ HT screen with His_6-MpPCO at 40 mg/mL had heavy precipitate in over 30% of the wells, which is slightly more than recommended for screens. However, the JCSG+™ HT screen with His_6-MpPCO at 30 mg/mL had no precipitate whatsoever, indicating 30 mg/mL was too low a concentration of protein for crystallisation to occur. The PACT premier ™ [242] HT, PEGRx ™ HT and SaltRx ™ HT crystallisation screens were trialled with His_6-MpPCO at 30

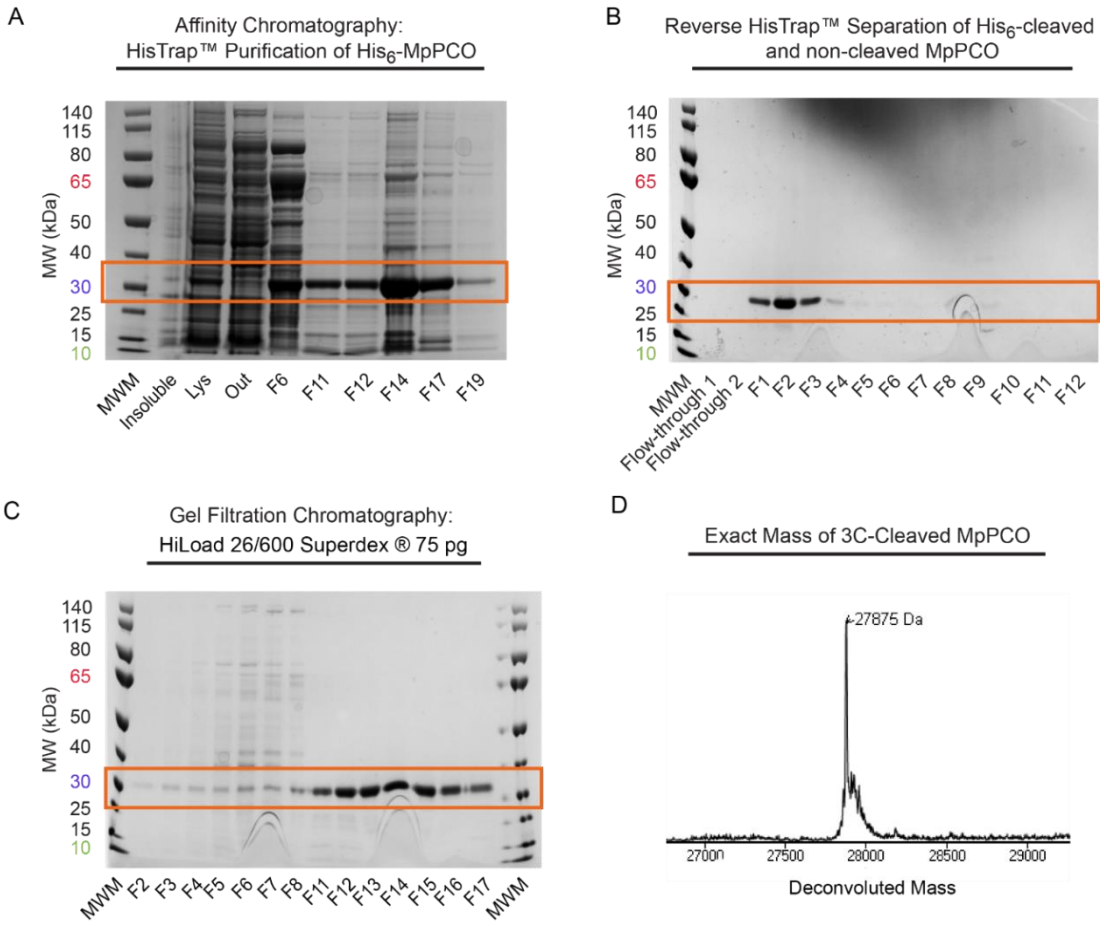


Fig. 5.3: Expression of His_6-cleaved MpPCO for crystallography. A) SDS-PAGE of HisTrap™ purified *E. coli* BL21 DE3 lysate with overexpressed, recombinant His_6-MpPCO. B) Reverse HisTrap™ of 3C protease-treated recombinant His_6-MpPCO, separating His_6-cleaved MpPCO from non-cleaved. C) Reverse HisTrap™ fractions F1-F6 were purified by gel filtration (GF). D) The predicted mass (sequence analysis by Protparam) of His_6-cleaved MpPCO was verified by Q-TOF MS.

mg/mL. The plates were checked every day for a week. By day 5 it was apparent 60 mg/mL was a better concentration for the JCSG+™ HT screen, with much of the initial precipitate re-solubilizing. Plates laid at 30 mg/mL His_6-MpPCO also had much reduced precipitate compared to initial observations (below 20%). No crystals or promising crystallisation drops were observed in any of the screens within a 6 week period.

5.2.3.2 MpPCO crystallisation screens

Light scattering experiments demonstrated that His_6-MpPCO was more prone to aggregation than MpPCO (Chapter 3, Section 3.4.3.3). It was rationalised that to remove possible sources of protein heterogeneity, the His_6-tag cleaved MpPCO should be trialled in crystallisation screens. His_6-tag cleaved MpPCO was prepared *via* incubation of His_6-MpPCO with 3C protease as described in Chapter 2, Sections 2.1.8.2 (Fig. 5.3). Subsequently, gel filtered MpPCO was concentrated to 8 mg/mL and screened using the JCGS+™, PACT Premier ™ and PEG/Ion ™ crystallisation screens. ~1 mm crystals (shape unknown) were harvested almost 10 weeks later from the A1 PEG/Ion ™ (0.2 M sodium fluoride, 20% w/v Polyethylene glycol 3,350, pH 7.3) sub-well positions 1-3, however the diffraction pattern indicated the crystals were salt. No crystal hits were observed after 6 weeks.

5.2.3.3 Improving MpPCO yield and additional crystallisation screens

The overall protein yield from methods in Section 5.2.3.2 was low, at approximately 1.4 mg (Fig. 5.3). A large amount of precipitate was often observed after desalting His_6-MpPCO using a PD-10 desalting column (Chapter 2, Section 2.1.9). The following protocol aimed to improve the yield of MpPCO in order to perform additional crystallisation screens. His_6-cleaved MpPCO was prepared *via* incubation with 3C protease as described in Chapter 2, Section 2.1.8.2 (Appendix 5.5) with the following modifications.

To avoid the large protein losses during de-salting, the de-salting step *via* a PD-10 column was avoided, opting instead for a gentler, dialysis-based approach. Following HisTrap™ purification, the fractions containing His_6-MpPCO (UV chromatogram, peak 2) were incubated with 1.25 mg of 3C protease at 4 °C overnight and dialysed into HisTrap ™ binding buffer using a 10 mL Thermo Fisher dialysis cassette. Cleaved protein was purified by gel filtration according as described in Chapter 2, Section 2.19. This protocol successfully tripled the MpPCO yield to approximately 4.1 mg when compared to the PD-10 desalting protocol (Section 5.2.3.2). The purest fractions were pooled and concentrated to 11.3 mg/mL and crystallisation trials were performed using JCSG+™, PACT Premier ™, Morpheus, PEG Rx ™, Salt Rx ™ and Wizard Classic crystallisation screens. Unfortunately, no crystal hits were observed after 6 weeks.

5.2.3.4 Metal-free His_6-MpPCO preparation and Native MS

His_6-MpPCO had previously been observed to bind metal in substoichiometric amounts (Appendix 3.4). In an effort to improve protein homogeneity, and similar to the investigations of KnPCO, a metal-free His_6-MpPCO preparation protocol was trialled. This protocol aimed to produce metal-free His_6-MpPCO (His_6-$MpPCO_{MF}$) and analyse its metalation and using native MS to assess whether or not the metal removal protocol was successful. His_6-MpPCO was expressed and purified as previously described (Appendix 5.7). The purest gel-filtered fractions were pooled and incubated overnight in 50 mM HEPEs, pH

7.5, 100 mM NaCl, 100 mM EDTA, 10 mM phenanthroline. As an initial trial, this process was considered more suitable than the "on column" approach used with KnPCO, as MpPCO is prone to precipitation and as such may compromise the integrity of the gel filtration column. Some light precipitation was noted after overnight EDTA/phenanthroline treatment. The treated protein was dialysed to remove any EDTA/phenanthroline and NaCl (necessary for successful native MS) and any remaining metal ions. A significant amount of precipitation was observed at this step. When assessing the concentration of His_6-MpPCO after phenanthroline treatment and dialysis, the most intense UV absorption peak was at 260 nm, not 280 nm wavelength, indicating phenanthroline contamination. Buffer exchange by centrifugation removed the contaminating phenanthroline, but a significant amount of precipitate was observed. Precipitate was removed by centrifugation. The remaining protein (100 µL at 0.75 mg/mL) was sent for native MS (Mr. D. Shutin, Chemistry Department), but precipitated during the necessary additional desalting process.

5.2.4 Thermal stability assays with KnPCO and MpPCO

Before continuing with crystallisation screens, it was thought prudent to examine protein thermal stability as a function of different salts and pH values, with an aim to garner insights into potential suitable crystallisation conditions. The Durham pH and Salt screens allows for the assessment of the effects of different buffers and pH on protein properties such as solubility, isothermal stability, and thermal denaturation point [248]. Buffer molecules that influence

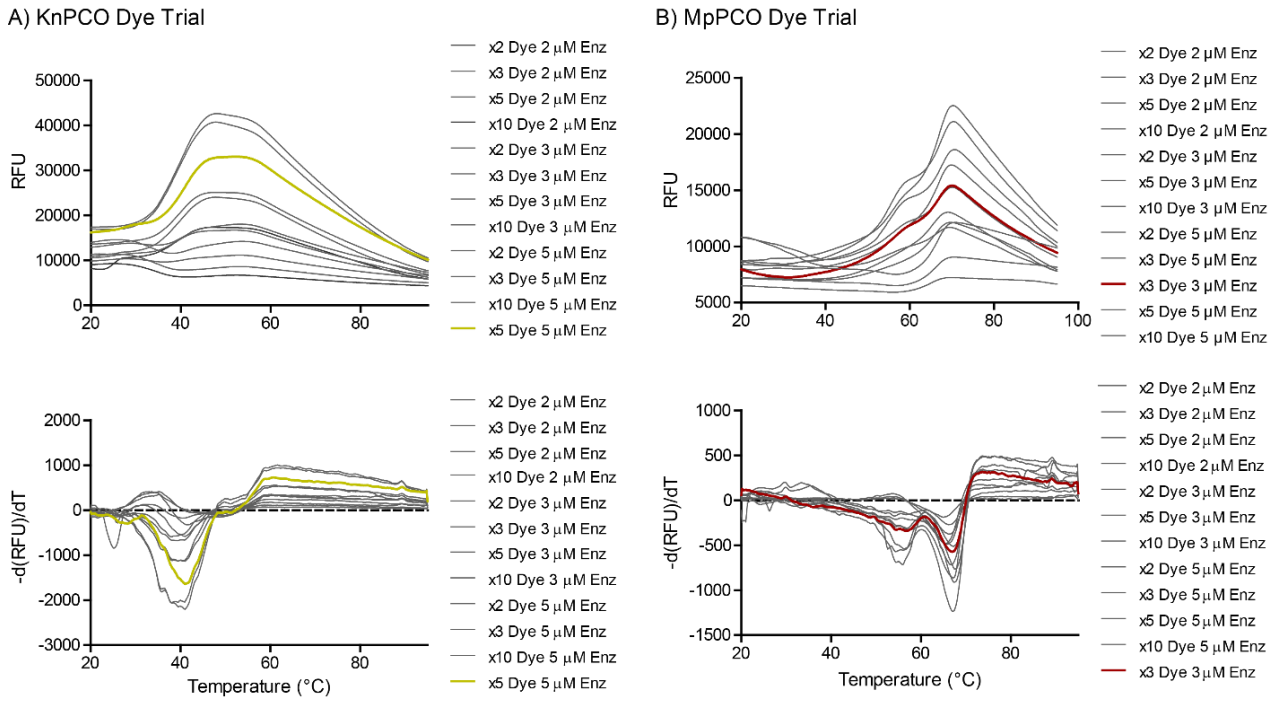


Fig. 5.4: Determining the optimum protein:dye ratio for thermal stability analysis. A) A selection of His_6-KnPCO and SYPRO® Orange dye concentrations. The protein:dye ratio chosen for salt and pH screens is highlighted in yellow. B) A selection of His_6-cleaved MpPCO and SYPRO Orange dye concentration ratios. The protein:dye ratio chosen for salt and pH screens is highlighted in red.

the thermal denaturation point usually also influence properties such as crystallisability and solubility. Thermal stabilisation studies were carried out using a BioRad C1000 Touch™ Thermal Cycler and SYPRO® Orange dye. The temperature was ramped from 20 to 95 °C (1 °C/min), with an increase in dye fluorescent signal indicating protein denaturation.

His_6-cleaved MpPCO was chosen for thermal stability assays as His_6-MpPCO did not yield a typical fluorescence pattern [231] using SYPRO® orange dye (Chapter 3, Section 3.4.3.3). Before the thermal stability of MpPCO and KnPCO with the Durham screens was assessed, first a suitable SYPRO® orange: protein ratio was determined. This was to ensure a good fluorescent signal was

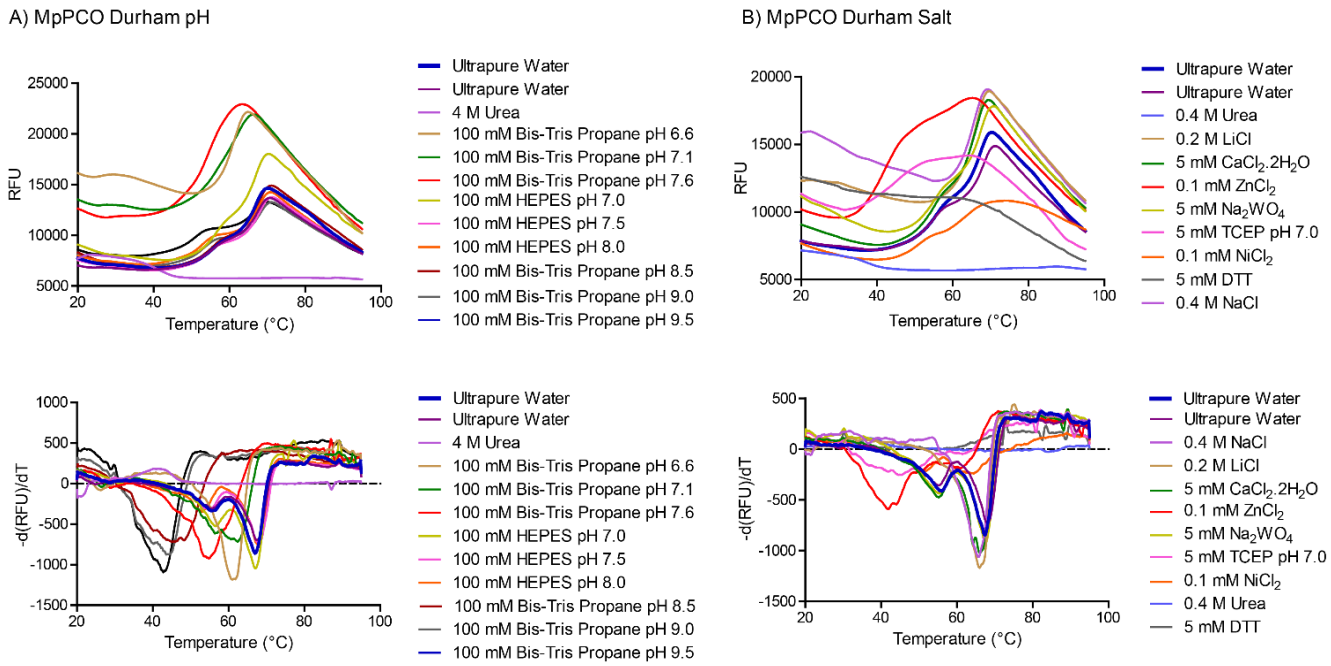


Fig. 5.5: Durham pH and salt screens with MpPCO. His_6-cleaved MpPCO thermal stability as a function of the A) Durham Salt and B) pH Screen. Salts and buffers used in MpPCO purification, storage and assays are shown. Also shown are any salt and buffer that improved MpPCO thermal stability relative to ultrapure water.

obtained, without overloading the detector. 3 µM of MpPCO and 3 units of SYPRO® orange and 5 µM of KnPCO and 5 units of SYPRO® orange were determined to be suitable protein: dye ratios (Fig. 5.4, B and C). Only traces from the Durham screens with the highest thermal stability, and traces with reagents of interest (storage and assay buffers, reducing agents, metals known to be bound to the protein) are presented here.

The Durham salt and pH screens for MpPCO suggest the protein is thermally stable up to approximately 70 °C in ultrapure water (Fig. 5.5). The trace for MpPCO fluorescence in 4 M urea is used as a reference for denatured protein. MpPCO with NaCl, LiCl or $CaCl_2$ salts had comparable thermal stability with ultrapure water. Due to time constraints, KnPCO was only incubated with the

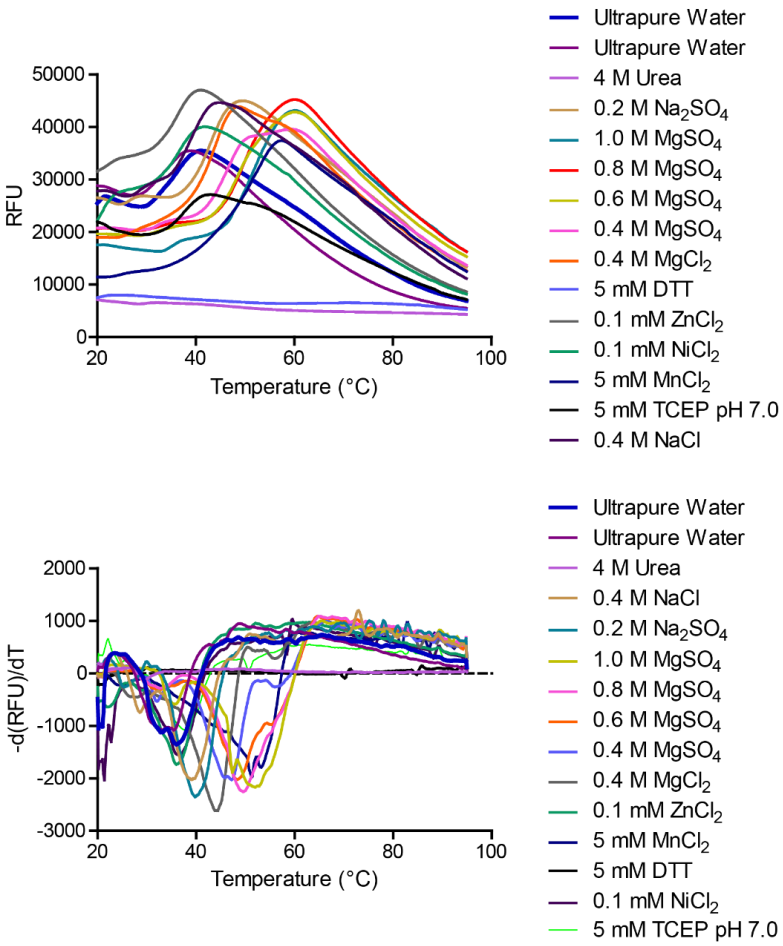


Fig. 5.6: Durham salt screens with His_6-KnPCO. KnPCO thermal stability as a function of the A) Durham Salt and B) pH Screen. Salts and buffers used in MpPCO purification, storage and assays are shown. Also shown are any salt and buffer that improved MpPCO thermal stability relative to ultrapure water.

salt screen. KnPCO, in contrast with MpPCO, denatured at approximately 38 °C in ultrapure water (Fig. 5.6). Significant improvements in thermal stability could be made with the addition of Mn or Mg II ions, but even with supplementary metal ions KnPCO was not as thermally stable as MpPCO.

5.3 Structural analyses of PCOs

5.3.1 Introduction

In lieu of high resolution structures, further experimental techniques and computational modelling were used to infer how closely the KnPCO and MpPCO structures might align with those of known PCOs. This was done using a combination of antibody recognition, circular dichroism and secondary and tertiary structure modelling.

5.3.2 PCO analysis *via* antibody recognition

Epitopes from macromolecules, such as proteins, are classified into two categories. Linear epitopes are composed of a continuous string of residues along the amino acid chain. Conformational epitopes are composed of amino acids that although not adjoining in primary sequence, are brought into close proximity within the folded 3-dimensional protein structure [249, 250]. Conformational epitopes correspond to the vast majority of epitopes found in proteins [250-254]. This suggests that specific surface conformational features of the antigen, rather than a sequential string of amino acids, are more likely to be the antigenic determinant. Antibody studies of the PCOs could indicate the similarity of MpPCO and KnPCO epitopes, relative to PCOs with known resolved structures.

A polyclonal antibody was raised against KnPCO (anti-KnPCO) in rabbits. The antibody was affinity purified using the KnPCO antigen. The His_6-tag cleaved KnPCO protein was prepared according to protocols in Chapter 2, Section 2.1

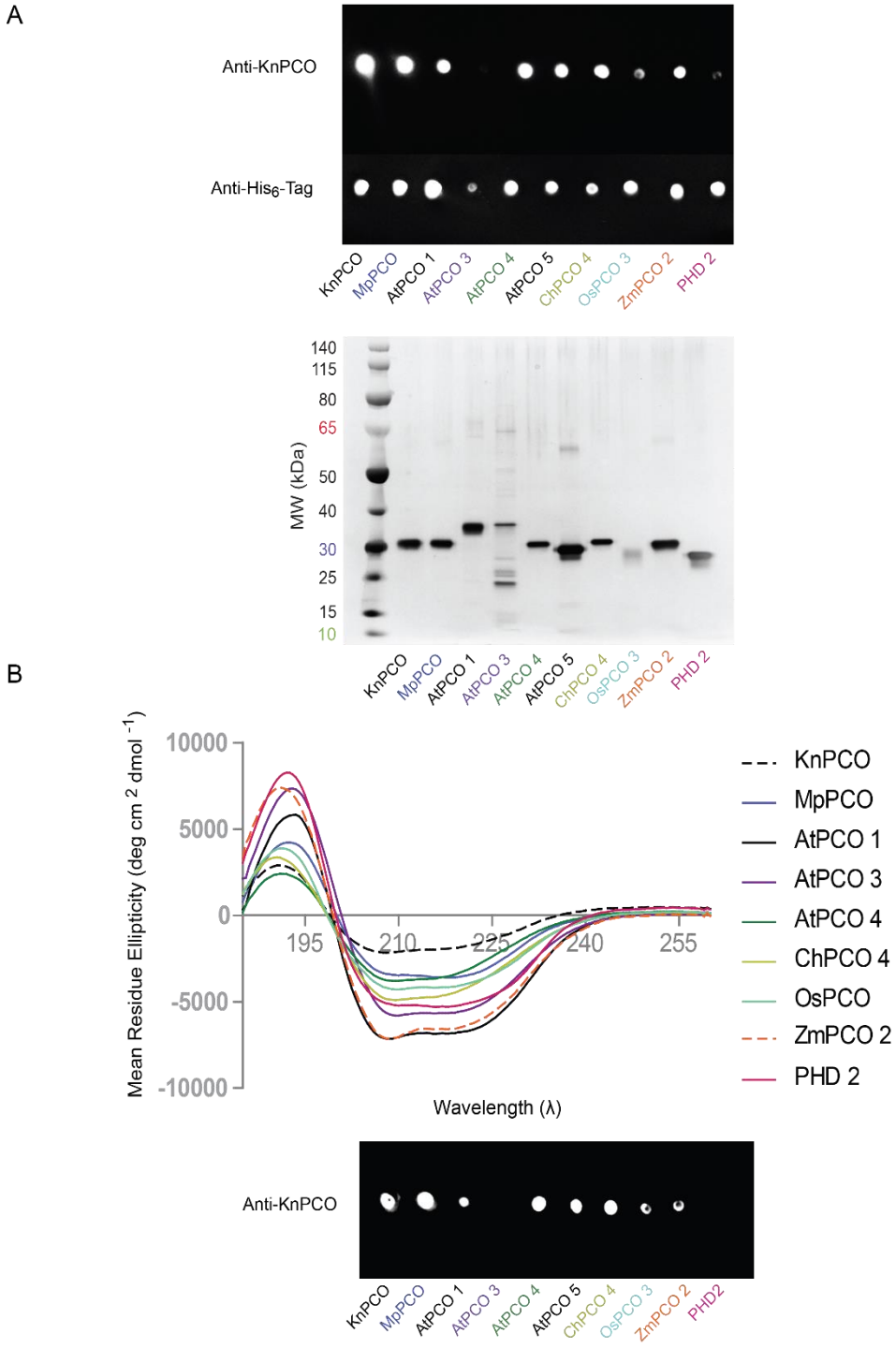


Fig. 5.7: Structural analysis of PCOs from across the plant kingdom using circular dichroism and an antibody raised against KnPCO. A) A dot blot was performed using both the α-KnPCO antibody and a α-His_6 antibody. The presence of protein was confirmed using a silver stain of an SDS-PAGE gel. B) Circular dichroism (CD) studies of His_6-tagged fusion PCOs from various plant species: KnPCO, MpPCO, AtPCOs 1, 3-5, ChPCO 4), ZmPCO 2), and OsPCO 3) and prolyl hydroxylase domain 2 (a human O_2 sensing enzyme). A dot blot of the same protein aliquots used for CD analysis showed varying degrees of "recognisability" using the α-KnPCO antibody, with KnPCO and Prolyl hydroxylase domain 2 as positive and negative blot controls respectively.

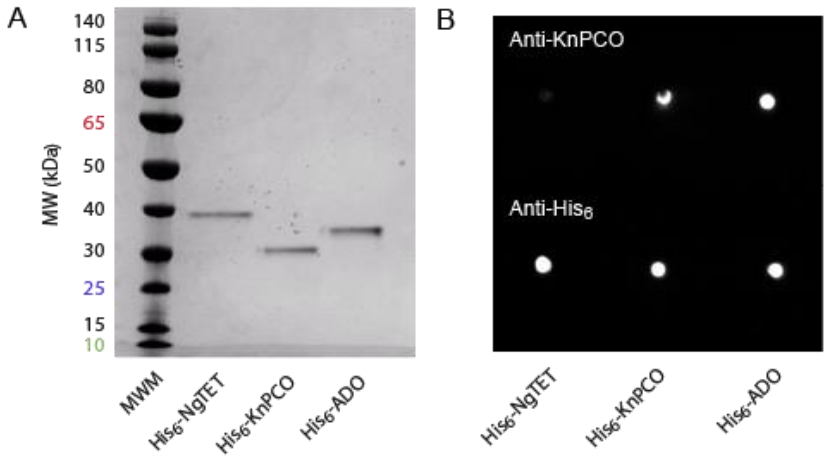


Fig. 5.8: Analysis of ADO using the α-KnPCO antibody. A) An SDS-PAGE gel stained with Coomassie of His_6-tagged KnPCO and ADO with Ng-TET as a negative control. B) A dot blot of the same ADO and KnPCO samples as used for the SDS-PAGE gel, with Ng-TET as a negative control. Both the α-KnPCO antibody and a α-His_6 antibody were used to detect the blotted proteins.

and sent to Genscript to raise the antibody (Section 5.2.3.3, Appendix 5.3). Anti-KnPCO antibody aliquots received from Genscript were tested to determine whether they could successfully identify KnPCO. The anti-KnPCO antibody recognised His_6-KnPCO (Fig. 5.7), despite the antibody being raised against His_6-tag cleaved KnPCO – indicating the epitope(s) was not greatly affected by the His_6-tag. To test whether the KnPCO antibody could recognise PCO enzymes from other organisms, His_6-HsADO, His_6-MpPCO, His_6-KnPCO, His_6-AtPCO 5, His_6-AtPCO 3, His_6-AtPCO 1, His_6-AtPCO 4, *Orza sativa* PCO 3 (His_6-OsPCO 3), *Zea mays* PCO 2 (His_6-ZmPCO 2) and *Crucihimalaya himalaica* PCO 4 (His_6-ChPCO 4) as well as a His_6-HsPHD $2_{181-407}$ negative control (provided by Dr. M. Abboud, Department of Chemistry) were exposed to the antibody.

Each His_6-PCO protein and the negative control was dotted onto a nitrocellulose blotting membrane and incubated with anti-KnPCO overnight at

4°C (Chapter 2, Section 2.4.5). After overnight treatment, the anti-KnPCO membrane was washed and incubated with an anti-Rabbit Immunoglobulin G (IgG) horseradish peroxidase (HRP) conjugate, secondary antibody. After a 1 hour incubation, the membrane was washed and incubated with a HRP-binding chemiluminescent reagent and imaged on the BioRad ChemiDoc XRS+. The anti-KnPCO membrane was exposed for 3s, and the anti-His_6-tag membrane exposed for 5s. Both the blots verified the presence of protein in all samples tested (Fig. 5.7 A). The anti-KnPCO successfully bound to His_6-KnPCO, and exhibited broad PCO recognition capabilities (Fig. 5.7 A). His_6-OsPCO had a less intense chemiluminescent signal compared to His_6-KnPCO, His_6-MpPCO, His_6-AtPCO 1, 4 and 5, His_6-HsADO (Fig. 5.8), His_6-ChPCO and His_6-ZmPCO, with His_6-AtPCO 3 having a lower signal than the HsPHD $2_{181-407}$ negative control. An anti-His_6-tag antibody confirmed the presence of His_6-AtPCO 3 protein on the blotting membrane (Fig. 5.7, A). A silver stain of the blotted protein solutions confirmed a band for His_6-AtPCO 3 at the expected molecular weight (33166.84 Da, Fig. 5.7, A).

His_6-AtPCO3 consistently failed to bind anti-KnPCO, despite having similar percentage sequence identity to KnPCO and AtPCOs 1, 4 and 5, all of which bound anti-KnPCO (Appendix 5.9). The PCO silver stain and the anti-His_6-tag recognition study confirmed the presence of His_6-AtPCO 3, with circular dichroism studies (Section 5.3.3) suggesting His_6-AtPCO 3 was in a folded conformation.

5.3.3 PCO secondary structure analysis using circular dichroism

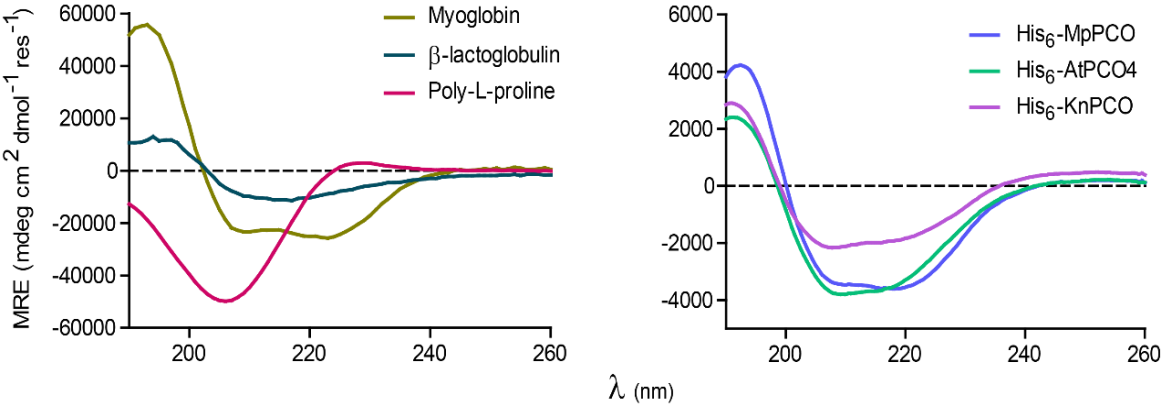
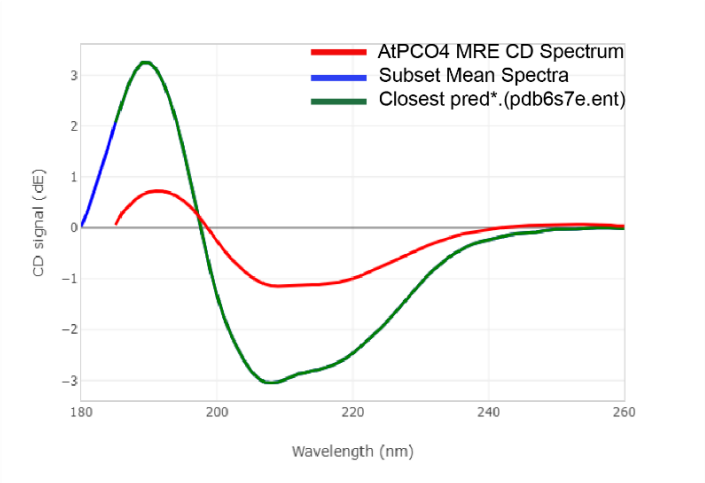


Fig. 5.9: CD spectrum of His_6-tagged MpPCO, KnPCO and AtPCO 4. The far-UV CD spectra of His_6-tagged KnPCO MpPCO and AtPCO 4 compared to standards; Myoglobin (PDB ID; 1YMB), β-lactoglobulin (PDB ID; 1B8E) and Poly-L-proline. All CD standard spectra were taken from The Protein Circular Dichroism Data Bank (https://pcddb.cryst.bbk.ac.uk/).

CD spectroscopy can be used to estimate the fraction of a protein that is alpha-helical, beta-sheet, or random coil conformations. Analysis of the CD spectra for KnPCO and MpPCO, relative to that of AtPCO 4 (PDB: 6S7E) could give insights into KnPCO and MpPCO secondary structure. Circular dichroism (CD) data of the same blotted His_6-tagged proteins from Section 5.3.2 showed the samples had polarised light absorption behaviour typical of folded protein (Fig. 5.7 B). The CD spectra of AtPCO 4, MpPCO and KnPCO are more similar to each other, than to AtPCO 3, which had a signal more comparable to ZmPCO 2. There is no CD data is available for AtPCO 5 as the protein precipitated upon exchange prior to CD.

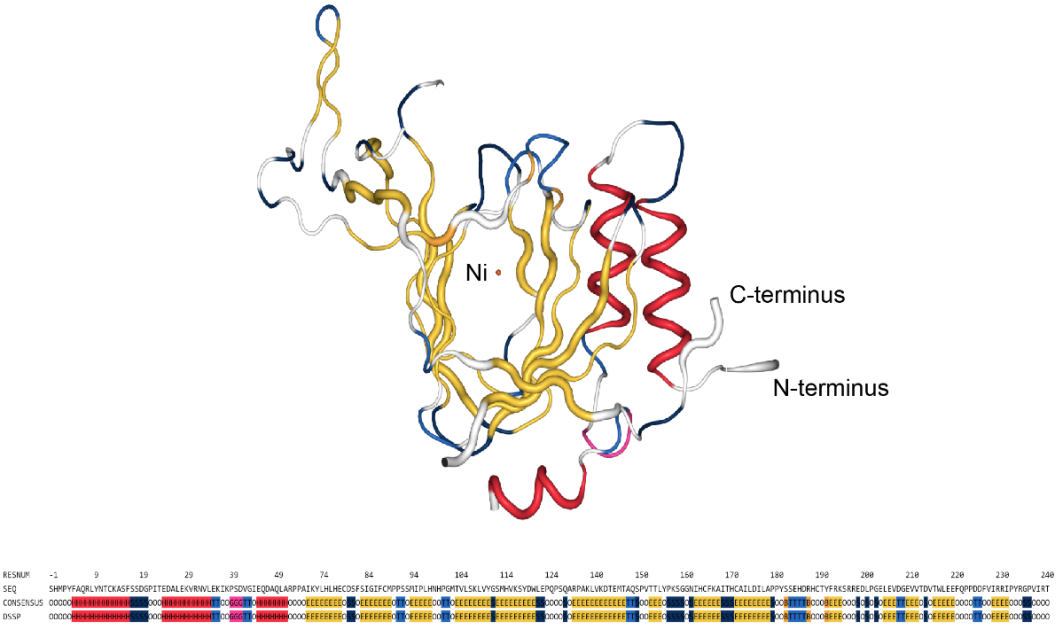
Most fully folded proteins display mean residue ellipticity (MRE) signals of the order 10^4 mdeg cm^2 $dmol^{-1}$ res^{-1} (Fig. 5.9). The signal intensities for all protein samples were lower than anticipated despite verifying a protein concentration

A) PDBMD2CD



helix	β-strand	turn	irregular
0.13	0.32	0.07	0.48

B) DSSP

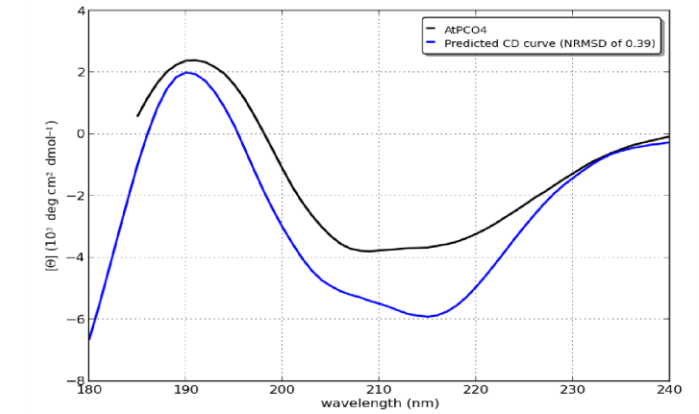


helix	β-strand	other
0.13	0.32	0.55

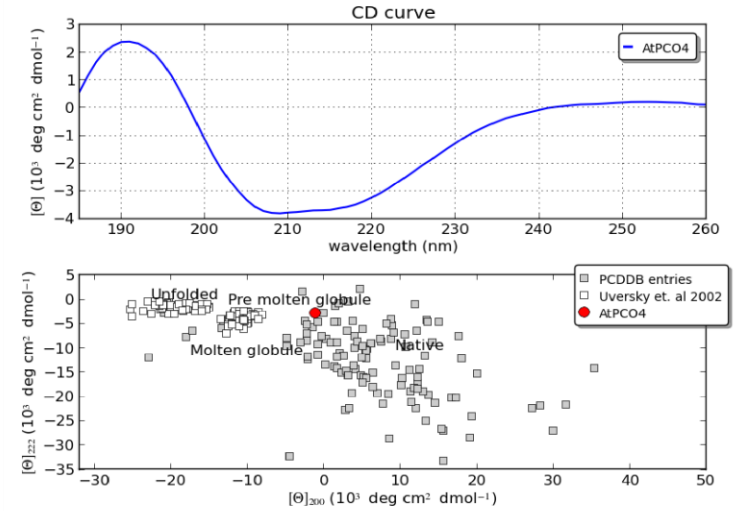
Fig. 5.10: Secondary structure assignment of the AtPCO 4 crystal structure. A) PDBMD2CD CD and secondary structure (SS) prediction tool, using the AtPCO 4 crystal structure (PDB ID: 6S7E, green). For comparison, the experimental CD spectrum for AtPCO 4 is included (red). B) Dictionary of Secondary Structure of Proteins (DSSP) assigns secondary structure using PDB entries (*via* the 2StrucCompare server). Both prediction models are in agreement on the SS assignment for AtPCO 4.

Secondary structure prediction of AtPCO4 using CD data

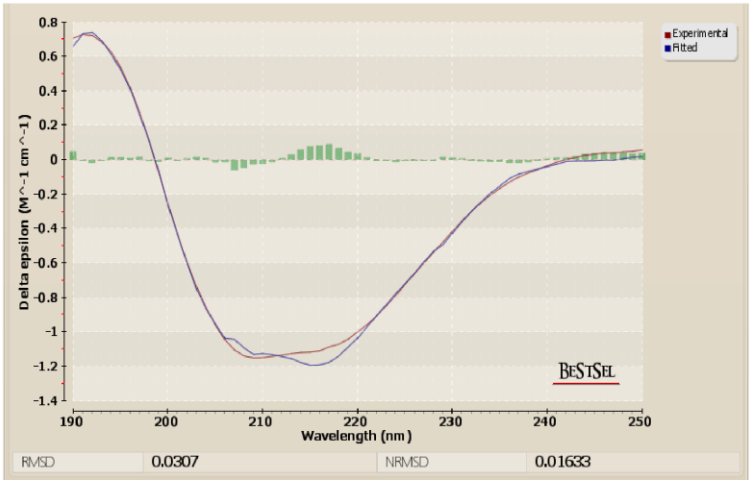
A) CAPITO



helix	β-strand	irregular
0.05	0.40	0.54



B) BeStSel

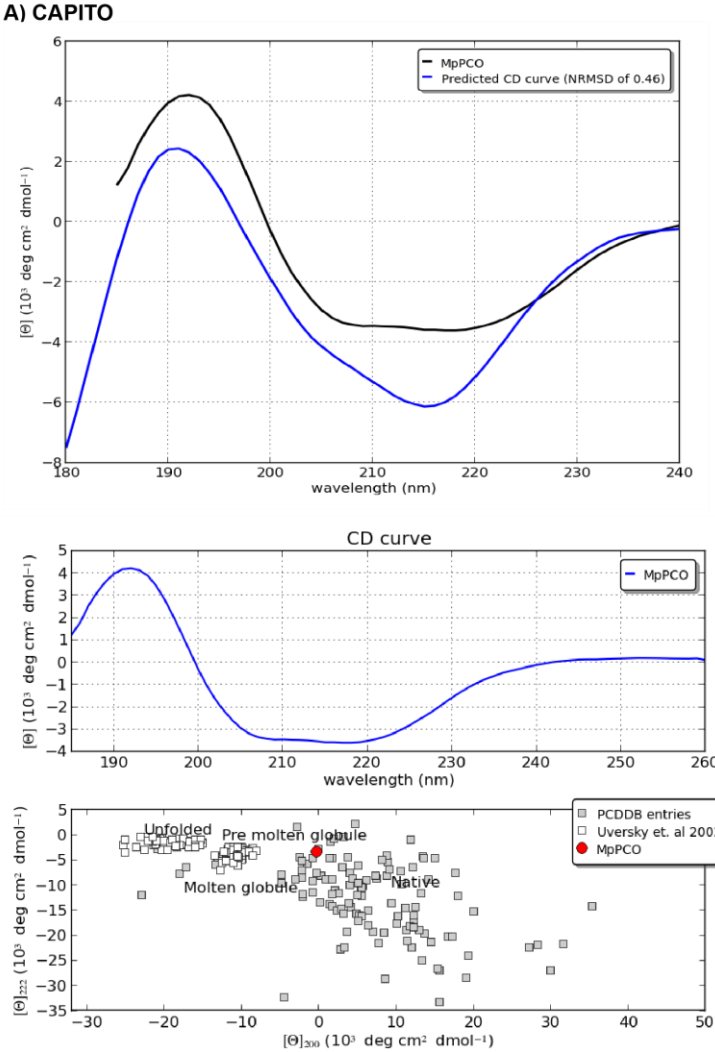


helix	β-strand	irregular
0.04	0.41	0.55

Fig. 5.11: Secondary structure prediction using the AtPCO 4 CD spectrum. A) CD Analysis and Plotting Tool (CAPITO) prediction of AtPCO 4 SS. B) Beta Structure Selection (BeStSel) prediction of AtPCO 4 SS. Both models have complementary AtPCO 4 SS assignment.

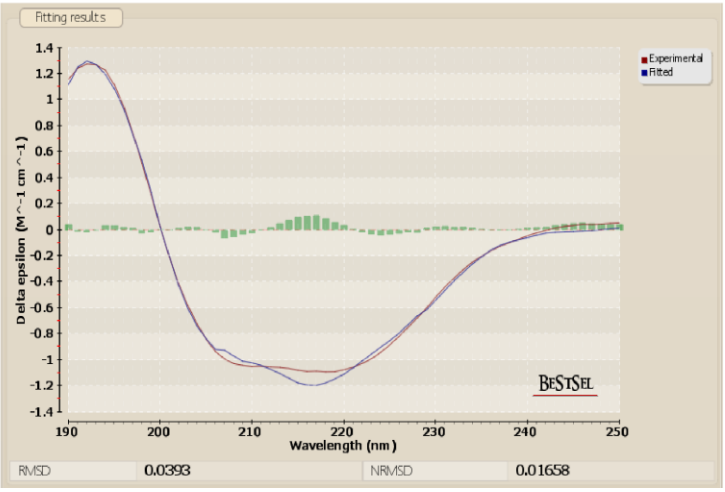
Secondary structure prediction of MpPCO using CD data

A) CAPITO



helix	β-strand	irregular
0.04	0.45	0.55

B) BeStSel



helix	β-strand	irregular
0.03	0.42	0.55

Fig. 5.12: Secondary structure prediction using the MpPCO CD spectrum. A) CD Analysis and Plotting Tool (CAPITO) prediction of MpPCO SS. B) Beta Structure Selection (BeStSel) prediction of MpPCO SS. Both models have complementary MpPCO SS assignment.

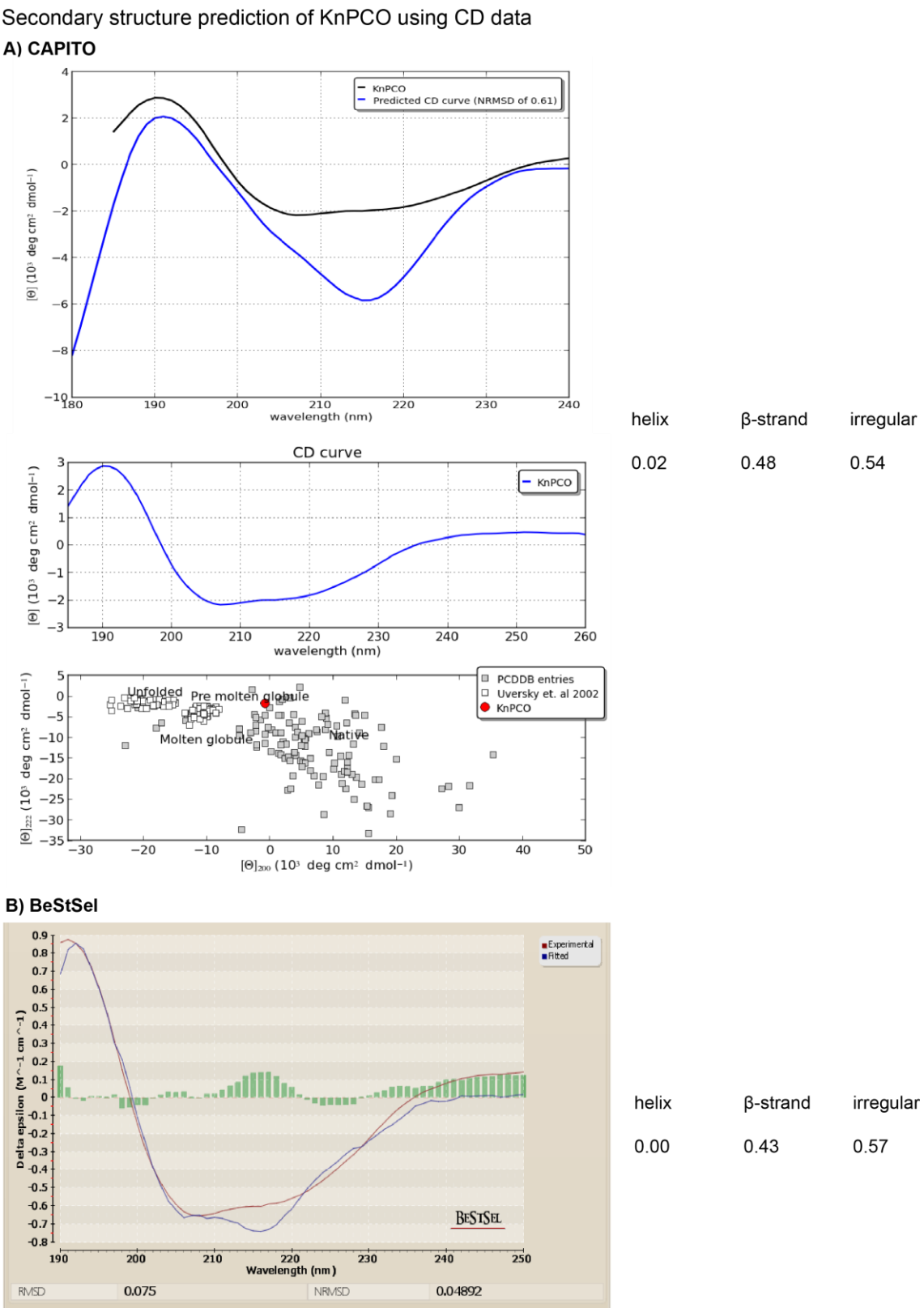
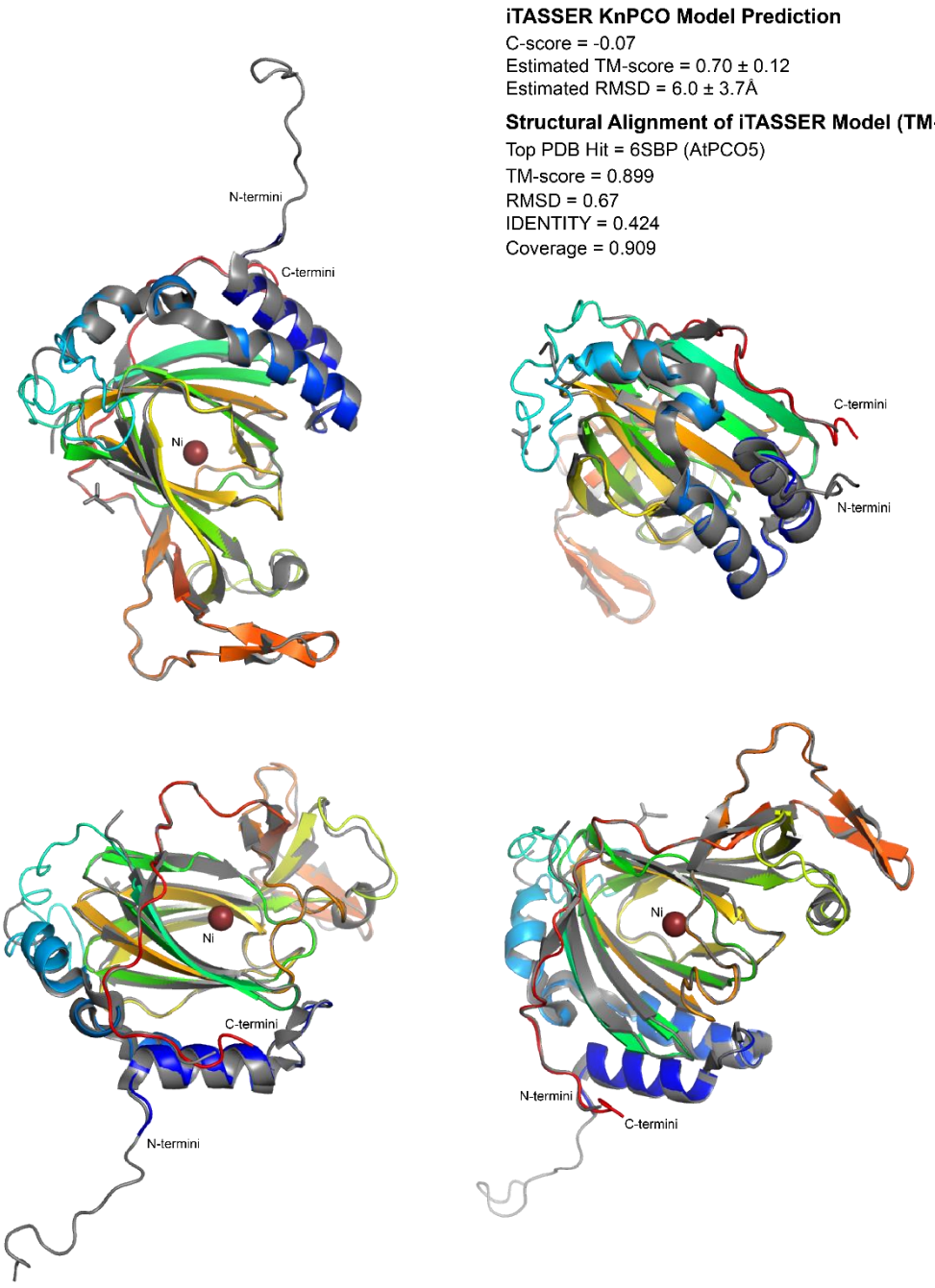


Fig. 5.13: Secondary structure prediction using the KnPCO CD spectrum. A) CD Analysis and Plotting Tool (CAPITO) prediction of KnPCO SS. B) Beta Structure Selection (BeStSel) prediction of KnPCO SS. Both models have complementary KnPCO SS assignment.

of 2 mg/mL using absorption at 280 nm prior to experimentation. This could negatively affect the accuracy of secondary structure prediction from the data. To determine the accuracy of the secondary structure prediction using CD data, the secondary structure assignment from the AtPCO 4 (PDB: 6S7E) X-ray crystal structure was compared with the AtPCO 4 experimental CD secondary structure assignment (Fig. 5.10 and 5.11). PDBMD2CD [255] and Dictionary of Secondary Structure of Proteins (DSSP) [256] web prediction tools estimated from the AtPCO 4 X-ray crystal structure that AtPCO 4 had a secondary structure ratio comparable to that of the CD secondary structure assignment tools (CD Anaylsis and Plotting Tool, CAPITO [257] and Beta Structure Selection, BeStSel [258], Fig. 5.10 and 5.11). On the assumption that the same is true for MpPCO and KnPCO, it appears that the CD data from these enzymes can therefore be used as a crude secondary structure assignment for MpPCO and KnPCO. Secondary structure assignment predicted a lower ratio of helical protein for KnPCO and MpPCO (0.02 and 0.035 respectively, Fig. 5.12 and Fig. 5.13), compared to AtPCO 4 (0.045). AtPCO 4, KnPCO and MpPCO contain comparable ratios of the β-sheet and irregular loop conformations (0.545, 0.555 and 0.55 respectively).

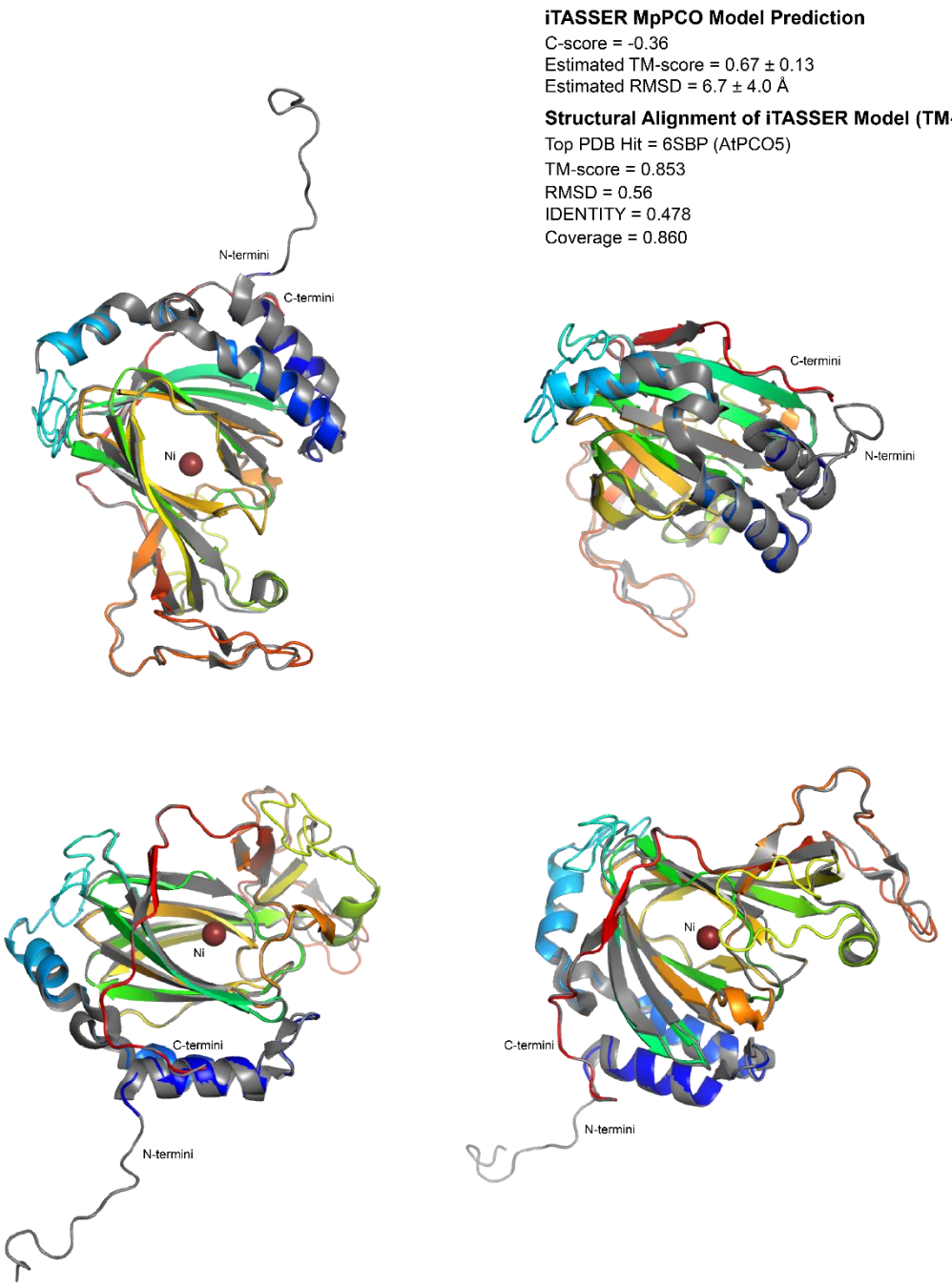
5.2.3 PCO tertiary structure analysis via I-TASSER and Phyre2

This section discusses a series of protein structure computational modelling techniques: namely protein homology modelling and threading. The purpose of this section is to use these techniques to gain further knowledge (predictions) of KnPCO and MpPCO structure. Threading is a method for fold



(a) iTASSER KnPCO model in rainbow, while the 6SBP analog is in grey.
(b) C-score is a confidence score for estimating the quality of predicted models by I-TASSER.
C-score is typically in the range of [-5,2], with a higher C-score indicating greater confidence in the model
(c) TM-score measures the similarity between the KnPCO model and 6SBP structures.
A TM-score > 0.5 indicates a model of correct topology and a TM-score < 0.17 means a random similarity.
(d) RMSD is the root-mean-square deviation between residues that are structurally aligned by TM-align.
(d) IDENTITY is the percentage sequence identity in the structurally aligned region.
(e) Coverage represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

Fig. 5.14: KnPCO structure prediction using I-TASSER. The Iterative Threading ASSEmbly Refinement (I-TASSER) is a hierarchical approach to protein structure prediction and structure-based function annotation. Here, I-TASSER has predicted the most likely KnPCO tertiary structure. The TM-align structural alignment program predicts the I-TASSER KnPCO tertiary structure model to be most comparable to that of AtPCO 5 (PDB ID: 6SBP). Both the I-TASSER model (rainbow) and 6SBP (grey) are aligned here, illustrating the double-barrell jelly roll motif synonymous with PCO structure [1]. This image is four different angles of the overlaid I-TASSER model (rainbow) and 6SBP (grey).



(a) iTASSER MpPCO model in rainbow, while the 6SBP analog is in grey.
(b) C-score is a confidence score for estimating the quality of predicted models by I-TASSER.
C-score is typically in the range of [-5,2], with a higher C-score indicating greater confidence in the model
(c) TM-score measures the similarity between the MpPCO model and 6SBP structures.
A TM-score > 0.5 indicates a model of correct topology and a TM-score < 0.17 means a random similarity.
(d) RMSD is the root-mean-square deviation between residues that are structurally aligned by TM-align.
(d) IDENTITY is the percentage sequence identity in the structurally aligned region.
(e) Coverage represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

Fig. 5.15: MpPCO structure prediction using I-TASSER. I-TASSER predicted the most likely MpPCO tertiary structure. The TM-align structural alignment program predicts the I-TASSER MpPCO tertiary structure model to be most comparable to that of AtPCO 5 (PDB ID: 6SBP). Both the I-TASSER model (rainbow) and 6SBP (grey) are aligned here, illustrating the double-barrell jelly roll motif synonymous with PCO structure (ref). This image is four different angles of the overlaid I-TASSER model (rainbow) and 6SBP (grey).

recognition, comparing the target protein sequence with a library of known structures. The focus of threading is to try multiple folds and alignments and evaluate a number of crude models. Only the models with the highest score (C-score) are discussed in this chapter. The Iterative Threading ASSEmbly Refinement (I-TASSER) identifies structural templates from the PDB by the multiple threading approach *via* the LOcal MEta-Threading-Server (LOMETS), with full-length atomic models constructed by iterative template-based fragment assembly simulations [259-261]. The final, lowest energy predicted model (highest C-score) of the protein is then used to search the protein data bank (PDB) for the protein with the greatest structural homology to the I-TASSER model using the TM-align program.

The I-TASSER models with the highest C-score for KnPCO and MpPCO have an N-terminal α-helical region and the conserved Cupin family double-barrelled β-sheet (DSBH) fold (Fig. 5.14 and 5.15). The TM-align program found both the KnPCO and MpPCO models had the closest structural similarity to that of AtPCO 5 (PDB: 6SBP [1]). The LOMETS threading results also found RnCDO and *Ralstonia eutropha* (ReCDO) structures to be suitable templates for MpPCO, with only RnCDO selected for KnPCO. These results suggest both KnPCO and MpPCO adhere to the DSBH Cupin family fold.

Model building by protein homology is a technique for predicting the tertiary fold of a target protein sequence, when the target protein is related to a protein of known sequence and structure. The modelling tool will first identify

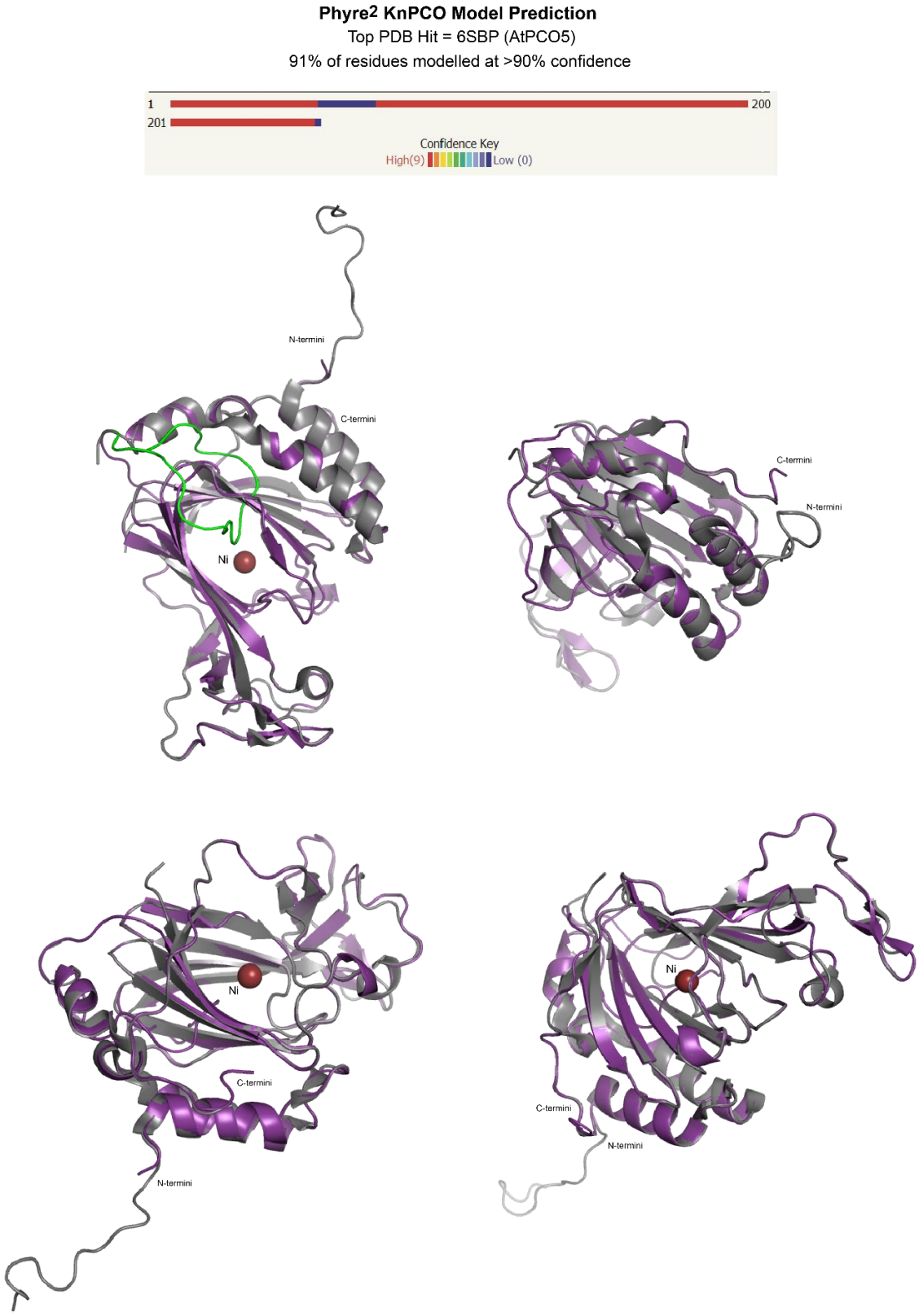


Fig. 5.16: KnPCO structure prediction using Phyre^2. The Protein Homology/analogY Recognition Engine V 2.0 (Phyre^2) 16. Here, Phyre^2 has predicted the most likely KnPCO tertiary structure, using AtPCO 5 (PDB ID: 6SBP) as a template. Both the Phyre^2 model (purple) and 6SBP (grey) are aligned here, illustrating the double-barrell jelly roll motif synonymous with PCO structure (ref). The region highlighted in green in the top LH image is the region predicted by *ab initio* folding prediction methods, and are considered unreliable. This image is four different angles of the overlaid Phyre2 model (purple) and 6SBP (grey).



Fig. 5.17: KnPCO secondary structure prediction using Phyre[2]. The SS assignment of KnPCO *via* the Phyre[2] platform.

sequence homologues based on the KnPCO and MpPCO sequences, then determine the optimum alignment and optimise a single model. The quality of the predicted model will rely on the similarity of the target sequence with the homologous sequence. The Protein Homology/analogY Recognition Engine V 2.0 (Phyre2) [262] tool uses remote homology detection methods to build tertiary protein structure models. It examines template hits, selecting for maximum

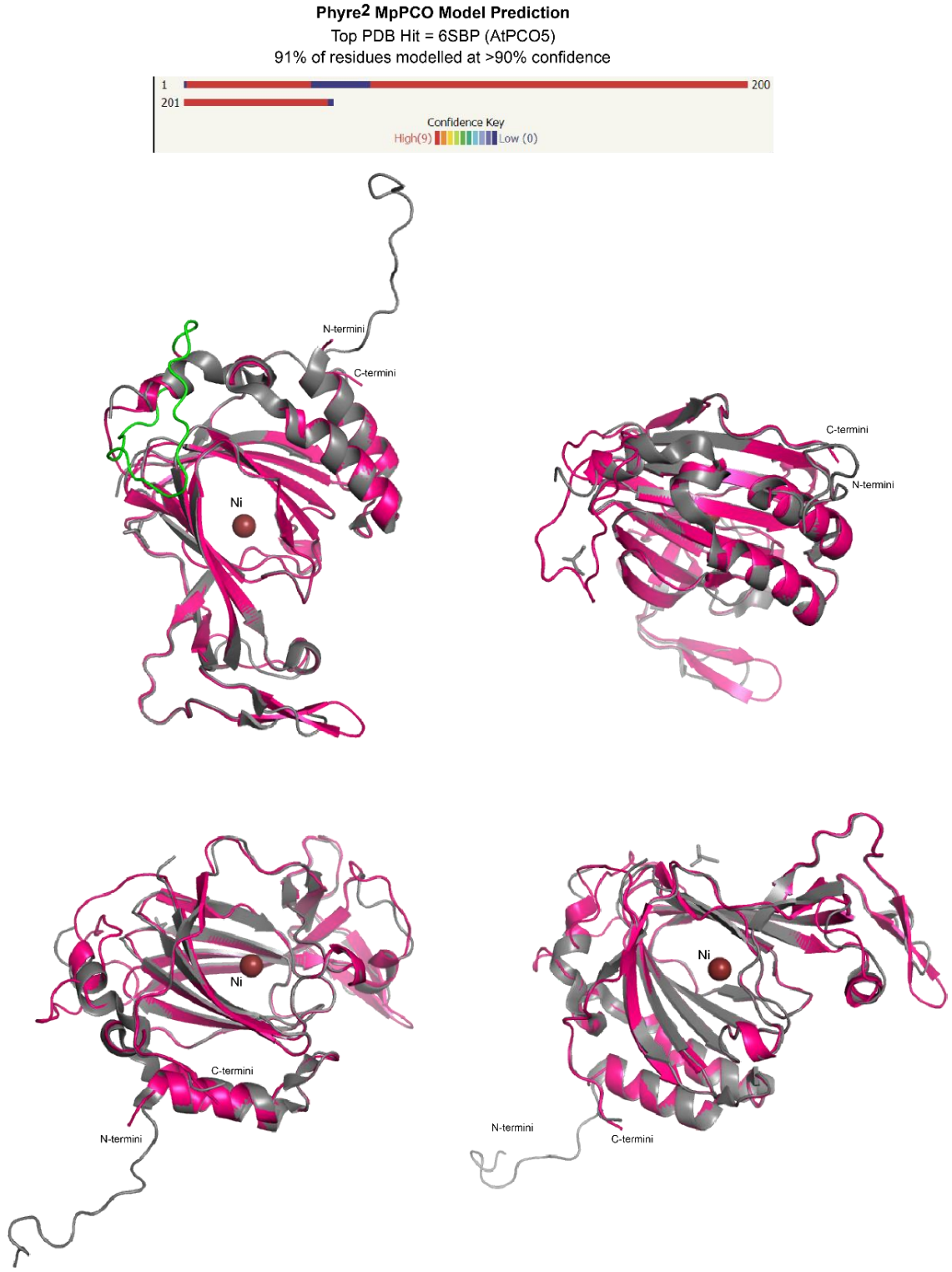


Fig. 5.18: MpPCO structure prediction using Phyre[2]. Phyre[2] has predicted the most likely MpPCO tertiary structure, using AtPCO 5 (PDB ID: 6SBP) as a template. Both the Phyre[2] model (pink) and 6SBP (grey) are aligned here, illustrating the double-barrell jelly roll motif synonymous with PCO structure (ref). The region highlighted in green in the top LH image is the region predicted by *ab initio* folding prediction methods, and are considered unreliable. This image is four different angles of the overlaid Phyre2 model (pink) and 6SBP (grey).

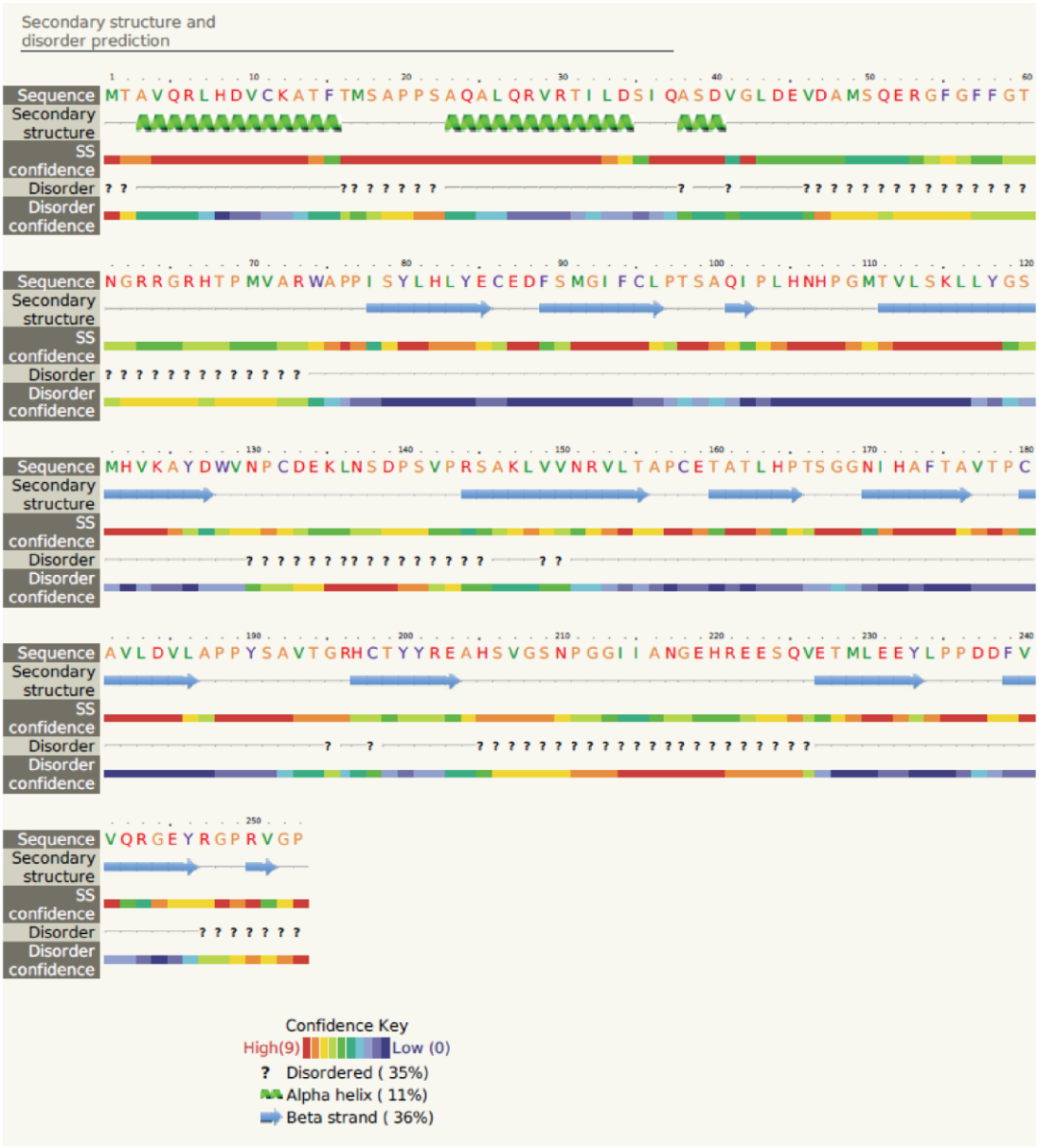


Fig. 5.19: MpPCO secondary structure prediction using Phyre[2]. The SS assignment of MpPCO *via* the Phyre[2] platform.

sequence coverage and confidence. Models are constructed based on the template hits and provide pairwise distance constraints for the Poing *ab initio* and multi-template modelling tool [262]. Poing builds the query protein in the context of these distance constraints, modelled by springs. Regions for which there is no template information are modelled by the *ab initio* simplified

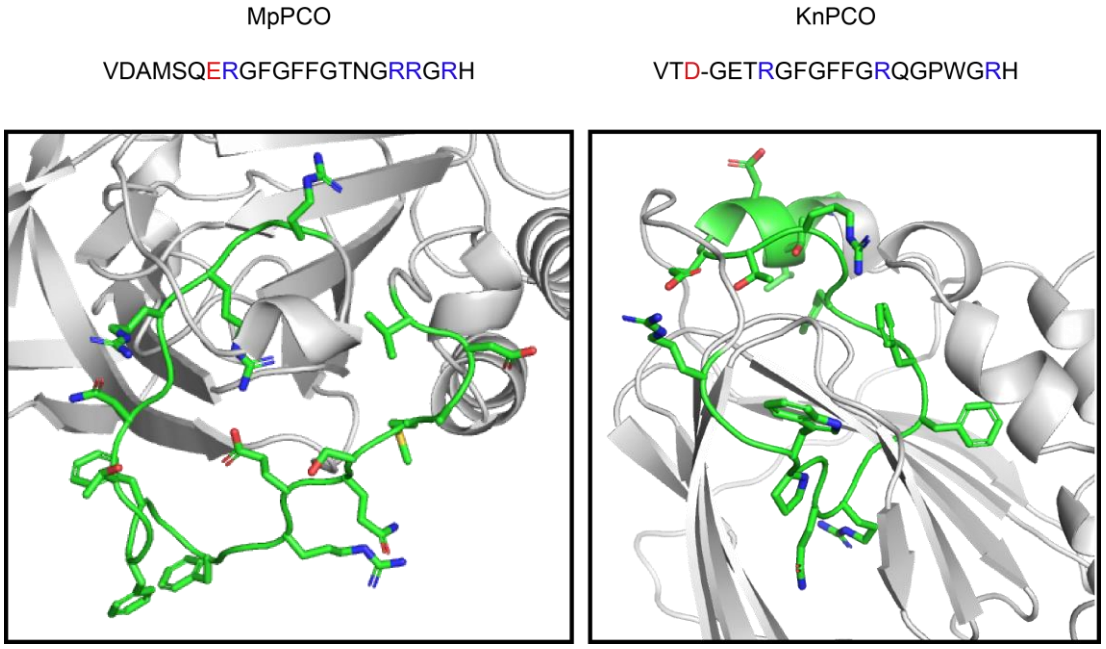


Fig. 5.20: MpPCO and KnPCO *ab initio* structure prediction using Phyre[2]. This non-conserved 22 amino acid region is rich in GlyPhe repeats and Arg residues.

physics model of Poing. The complete model generated by Poing is combined with the original templates as input to MODELLER.

Phyre2 modelled 91% of KnPCO residues at over 90% confidence using AtPCO5 (PDB: 6SBP, Fig. 16). KnPCO and MpPCO share 41 and 47% sequence identity with AtPCO5 respectively (Appendix 5.10). Phylogenetic analysis suggests MpPCO and KnPCO group with the hypoxia non-induced PCOs, of which AtPCO5 is also a member (Fig. 1.10). The high percentage sequence identity and phylogenetic analysis suggests AtPCO5 would likely be a suitable template for accurately predicting MpPCO and KnPCO structure (Appendix 5.10).

The Phyre2-predicted KnPCO model has a double-barrelled β-sheet (DSBH) fold, and an N-terminal helical region comparable to that of AtPCO 5 and I-

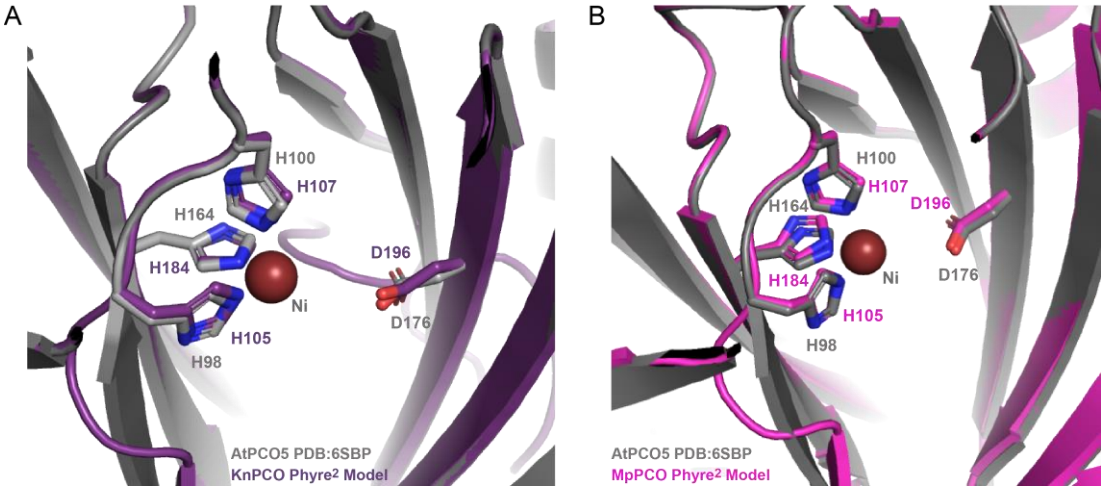


Fig. 5.21: MpPCO and KnPCO Phyre^2 modelled active sites overlaid with AtPCO 5 (PDB:6SBP). A) KnPCO active site residues in **purple**. B) MpPCO active site residues in **pink**. AtPCO 5 (PDB:6SBP) residues are coloured **grey** in both A and B.

TASSER model. A GlyArgPhe-rich region between the N-terminal helices and DSBH lacked template information, and as such was modelled *ab initio* (highlighted in green in Fig. 5.16 and 5.20). The MpPCO Phyre2 model had 91% of residues modelled at over 90% confidence, and features the same DSBH and helical folds (Fig. 5.18 and 5.20). MpPCO also had the same GlyArgPhe-rich region between the N-terminal helices and DSBH folds. This region is not conserved in the higher plant PCOs (Appendix 5.9). The Phyre^2 secondary structure prediction tool suggested with good confidence levels that this region is disordered in MpPCO (Fig 5.19). For KnPCO, this region has a less confident assignment of disorder compared to MpPCO, with secondary structure prediction also suggesting the possibility of a helical region (Fig. 5.17). The metal-binding His facial triad and conserved Asp residue at the active site were conserved in both KnPCO and MpPCO Phyre2 models (Fig. 5.21).

Discussion

The aim of the work described in this chapter was to gain structural information on KnPCO and MpPCO. Despite not obtaining any crystals suitable for diffraction, some useful metal-free and His_6-tag cleaved protein preparation protocols were developed in the process. Due to a lack of crystals suitable for diffraction, KnPCO and MpPCO structures had to be inferred using a series of experimental CD data analysis and tertiary structure predictive tools.

KnPCO crystallisation screens yielded some promising results, with JCSG™+ and Morpheus II screen conditions appearing worthy of optimization. Despite the inability to obtain diffracting protein crystals, metal-free KnPCO was produced successfully. This is the first reported protocol for the successful removal of metal ions from a PCO. However, MpPCO readily precipitated in a low-salt, metal-free environment. This could be a result of sub-optimum salt and metal removal protocols, and optimization or rapid metalation with a specific uniform metal may improve the outcome and provide highly active crystallisable protein.

Thermal shift assays using the Durham salt and pH screens showed MpPCO is highly stable in ultrapure water, with most additional salts negatively affecting thermal stability. KnPCO, by contrast, denatured below 40 °C in ultrapure water. KnPCO thermal stability could be improved somewhat with the addition of Mn or Mg II ions, but even with supplementary metal ions, KnPCO was far less

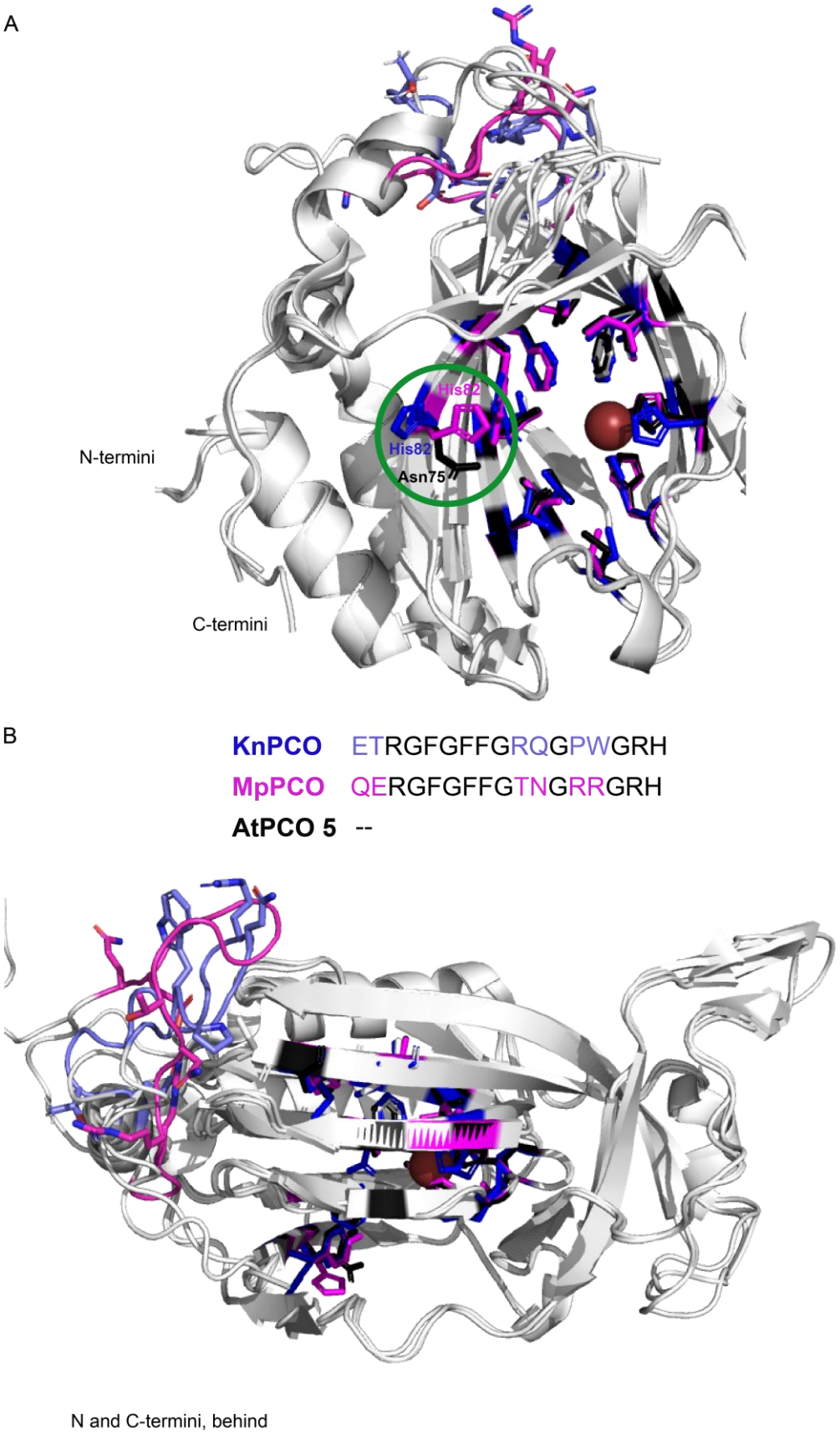


Fig. 5.22: MpPCO Phyre[2] and KnPCO I-TASSER modelled active sites overlaid with AtPCO5 (PDB:6SBP). KnPCO active site residues and non-conserved GRF rich region in **blue**. MpPCO active site residues and non-conserved GRF rich region in **pink**. AtPCO5 (PDB:6SBP) residues are coloured **black.** A) view of the active site B) view of the GRF rich regions

thermally stable than MpPCO. This suggests KnPCO may be quite challenging to crystallise.

Anti-KnPCO polyclonal antibody cross reactivity suggests somewhat conserved epitopes across the PCO family. CD studies suggested KnPCO and MpPCO had a comparable secondary structure content to that of AtPCO 4 and 5 (Appendix 5.8). The secondary structure predictions align well with the computational structure prediction tools, with I-TASSER and Phyre2 platforms predicting very similar secondary and tertiary folds for KnPCO and MpPCO to those of AtPCO 5, including: a DSBH and an N-terminal helical region.

A GlyArgPhe rich region between the N-terminal helices and the DSBH was found to be conserved between MpPCO and KnPCO, but not with higher plant PCOs (Appendix 5.9). It is not clear what impact these regions will have on the respective KnPCO and MpPCO folds, but CD spectra suggest some helices form part of the MpPCO (and to a lesser degree) the KnPCO secondary structure. Computational modelling predicts KnPCO and MpPCO helical regions to be N-terminal based on sequence alignments with AtPCO 5 (PDB: 6SBP). Overall, it is likely KnPCO and MpPCO have similar folds to other known PCOs.

The active sites of both KnPCO and MpPCO closely match that of AtPCO 5 (Fig 5.21 A and B), with the metal-binding His facial triad and conserved Asp residues in close proximity to the metal centre. This structure is in keeping with activity assays using the MpPCO_H184D variant (Fig. 3.12), and AtPCO 4_ H164D [1], where enzyme activity was ablated. The MpPCO_D196N variant

exhibited very low specific activity, with some activity restored upon the addition of Fe II (Fig 3.12), again consistent with the AtPCO 4_D176N variant [1]. Interestingly, the increase in activity of the MpPCO_D196N variant was substrate dependent, with AtRAP2.12$_{2\text{-}15}$ seeing a greater increase in activity compared to MpERF$_{2\text{-}15}$. Unfortunately the KnPCO_H184D variant failed to express (Appendix 4.15).

KnPCO has a different substrate profile to that of MpPCO, with activity assays indicating that AtZPR2$_{2\text{-}15}$, which was successfully oxidised by MpPCO, is not a substrate of KnPCO (Fig. 4.10). To explore which structural features might contribute to KnPCO substrate selectivity, the KnPCO I-TASSER, MpPCO Phyre[2] and AtPCO 5 (PDB: 6SBP) structures were overlaid (Fig 5.22). The active site does not contain any obvious, structure disrupting amino acids. Asn 75 in AtPCO 5 is mutated to His82 in KnPCO and MpPCO, with the His82 amino acids having slightly different orientations in both KnPCO and MpPCO models (Fig. 5.22, A). The GlyArgPhe-rich region has a number of mutations which could affect substrate binding (Fig. 5.22, B), however due to a lack of homology of this region to AtPCO 5, the modelling tools used could not accurately predict the structure of this region. Overall, it is currently unclear why KnPCO has a substrate selectivity previously unobserved in PCOs.

Chapter 6: Novel PCO substrates in algal cell lysates

6.1 Summary

Our knowledge of PCO substrate scope is narrow; limited thus far to proteins homologous to the higher plant ERF-VII and ZPR2 and VRN2 transcirption regulators [24, 26, 133, 142]. This limits our understanding of PCO function in algae, which contain no ERF, ZPR2, or MC-initiating VRN2-related proteins in their genomes (Fig. 1.7). This chapter outlines the search for new and hypoxia relevant PCO protein substrates. Cells exposed to hypoxia and normoxia were lysed and incubated with recombinant protein to generate the N-terminal $CysO_2H$ moiety, and then labelled with a reporter molecule for protein-$CysO_2H$ detection.

6.2 N-ethyl maleimide as a cysteine sulfinic acid probe

6.2.1 Introduction

To target N-terminal Cys sulfinic acid ($CysO_2H$) moieties in cell lysates (generated by incubating with recombinant PCO), a $CysO_2H$ chemical probe had to be selected and a suitable $CysO_2H$ labelling protocol established. Akter *et al* [263] described the use of BioDiAlk as an effective $CysO_2H$ probe. An aliquot of BioDiAlk was kindly sent by Dr. K. Carroll. The BioDiAlk probe was analysed by Q-TOF MS and NMR (data not shown) to determine its mass, integrity and concentration. MS analysis suggested there was a molecule with a mass similar to that of the BioDiAlk (Appendix 6.1), but NMR determined a significant proportion of the probe had degraded, possibly during its transport to Oxford.

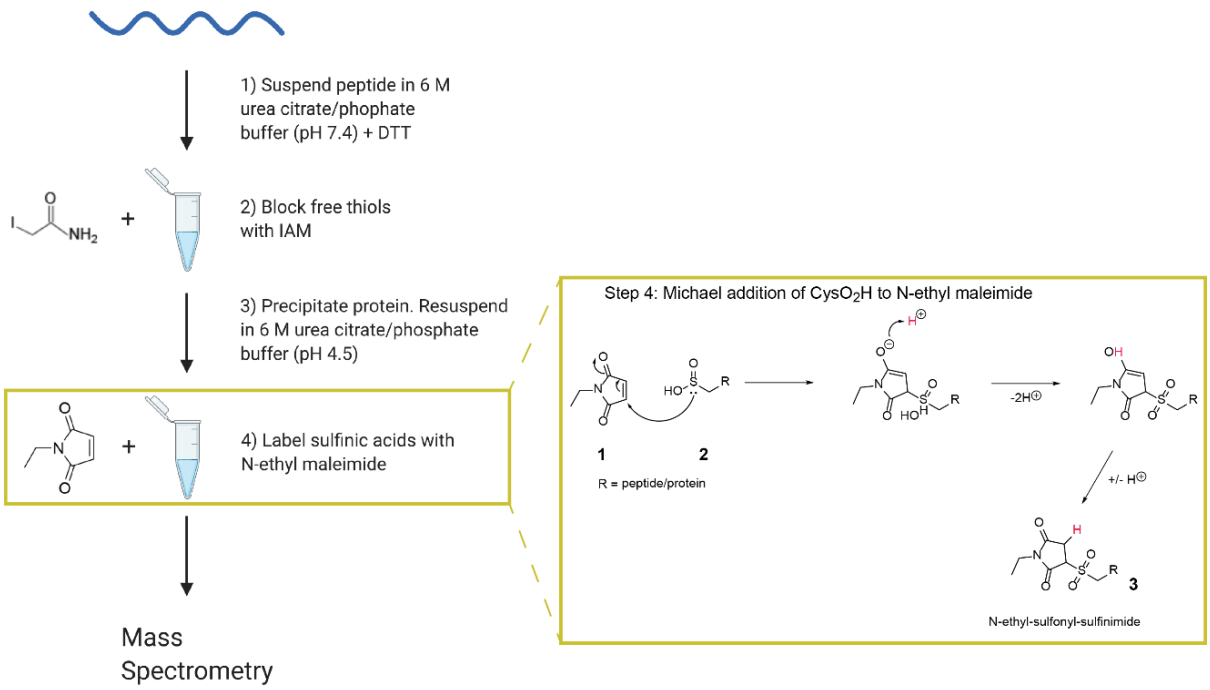


Fig. 6.1: Cys-sulfinic acid labelling scheme. Step 1) Lysed cells or pure peptide are re-suspended in 6M urea citrate/phosphate buffer pH 7.4 and DTT. Step 2) Free thiols are blocked using iodoacetamide (IAM). Step 3) protein or peptide is re-suspended in 6M urea citrate/phosphate buffer pH 4.5. Step 4) N-ethyl maleimide (**1**) undergoes nucleophilic attack by the sulphur on the sulfinic acid (**2**) group to form N-ethyl-sulfonyl-sulfinimide (**3**). Partially created with **BioRender.com**

After incubation with AtRAP2.12$_{2\text{-}15}$-CysO$_2$H, there were no peaks in the deconvoluted spectrum corresponding with the expected AtRAP2.12$_{2\text{-}15}$-CysO$_2$H-BioDiAlk mass (Appendix 6.1). It is possible the moiety was not visible on the Q-TOF or the labelling protocol was not optimum.

The use of maleimide probes for targeting CysO$_2$H has been reported in the literature [264]. This seemed a more promising probe for protocol development for a number of reasons. Firstly, pure N-ethyl maleimide is inexpensive and easy to purchase, whereas the BioDiAlk is not commercially available. Maleimide-reporter probes are also readily available and their use in targeting thiols well-reported [265-267]. It was also desirable that the CysO$_2$H-labelled

moiety be visible on the Q-TOF for protocol development purposes. As such, the BioDiAlk probe was abandoned in favour of maleimide-based probes.

Kuo *et al*. outlined a bioconjugation assay for labelling S-sulfinated cysteine using reporter-maleimide reagents in cell lysates [264] (Fig. 6.1). In this protocol, the proteins in cell lysates are denatured in 6 M urea and treated with DTT. Afterwards, the lysate is treated with iodoacetamide (IAM), to block free thiols. The $CysO_2H$ moieties are then covalently modified using N-ethyl maleimide at pH 4.5 (**1**). This modification step is pH dependent and thought to proceed by a Michael addition (Fig 6.1). The maleimide group undergoes nucleophilic attack by the lone pair of the sulfinic acid (**2**) group to form N-ethyl-sulfonyl-sulfinimide (**3**). The maleimide group is attached to a reporter molecule that allows for labelled proteins to be detected as desired.

To ensure N-terminal $CysO_2H$ moieties could be successfully targeted using maleimide-based reporter molecules, initial experiments examined the labelling of the known PCO substrate, AtRAP2.12_{2-15}, with N-ethyl maleimide (NEM). This section outlines the protocol for thiol-capping and $CysO_2H$ labelling of peptide to ensure the experimental conditions were conducive to $CysO_2H$ labelling prior to cell lysate trials (Fig. 6.1).

6.2.2 Preparation of AtRAP2.$12_{2-15}$$CysO_2H$

To develop a $CysO_2H$ labelling protocol, a suitable $CysO_2H$ target for the NEM probe had to be made. To do this, AtRAP2.12_{2-15} was incubated with AtPCO 4 for 1 hour at 25 °C. The oxidised peptide was separated from AtPCO 4 using a

10 k MWCO concentrator-with peptide separated into the flow-through. Analysis of the peptide-containing flow-through on RapidFire QTOF verified the presence of oxidised AtRAP2.12$_{2\text{-}15}$ peptide (Appendix 6.2). The percentage oxidised peptide was quantified using the Agilent Quantitative Analysis software. The flow-through was determined to be 99.5% AtRAP2.12$_{2\text{-}15}$CysO_2H, 0.5 % AtRAP2.12$_{2\text{-}15}$. This N-terminally oxidised peptide stock was used for all following peptide labelling experiments.

6.2.3 Peptide CysO_2H detection with N-ethyl maleimide

The protocol described in Kuo *et al.* [264] is optimised for targeting CysO_2H in cell lysates. Consequently, it incorporates a thiol-capping step with iodoacetamide (IAM), which blocks reduced Cys from maleimide labelling (Chapter 2, Section 2.5.1). The AtRAP2.12$_{2\text{-}15}$ (1440.81 Da, Fig. 6.2, A) and AtRAP2.12$_{2\text{-}15}$CysO_2H (1472.78 Da, Fig. 6.2, B) peptides were therefore incubated separately with IAM in 6 M urea $Na_2PO_4H.2H_2O$/Citric acid pH 7.4 for 1 hour. Afterwards, the peptides were buffer exchanged into 6 M urea $Na_2PO_4H.2H_2O$/Citric acid pH 4.5 and incubated separately with NEM for 1 hour. All steps were carried out at room temperature. Excess NEM was removed by ZipTip® prior to RapidFire QTOF MS analysis.

The MS spectrum of AtRAP2.12$_{2\text{-}15}$Cys when incubated with IAM and NEM, has a peak consistent with the AtRAP2.12$_{2\text{-}15}$Cys-acetamide modification (1497.84 Da, Fig. 6.2, C), but no peaks consistent with AtRAP2.12$_{2\text{-}15}$Cys-NEM formation (1597.84 Da). The MS spectrum of AtRAP2.12$_{2\text{-}15}$-CysO_2H incubated with IAM

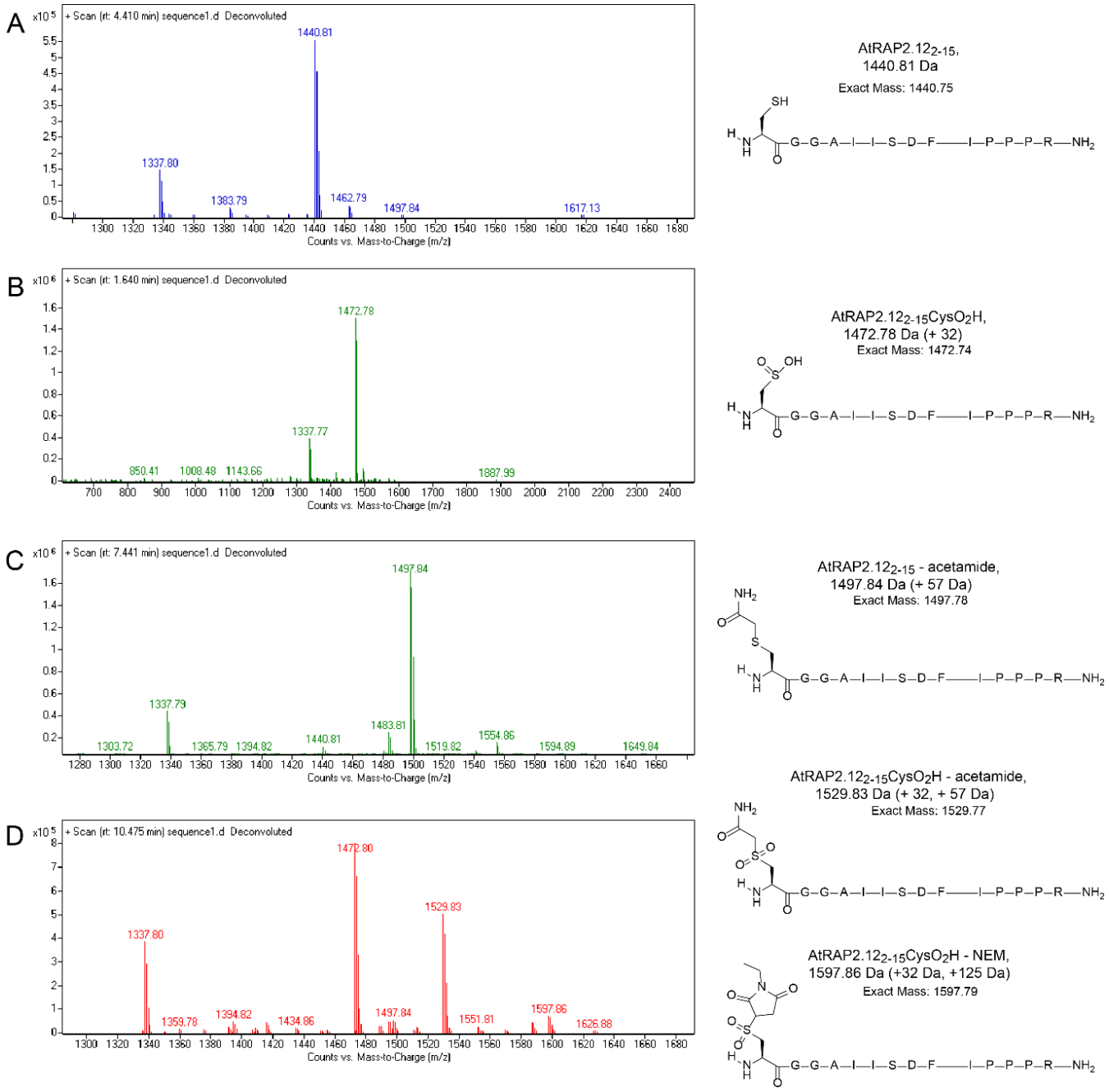


Fig. 6.2: AtRAP2.12$_{2\text{-}15}$CysO$_2$H detection using N-ethyl maleimide (NEM). A) Reduced, untreated AtRAP2.12$_{2\text{-}15}$ peptide has a mass of 1440.81 Da. B) oxidised, untreated AtRAP2.12$_{2\text{-}15}$CysO$_2$H peptide has a mass of 1472.78 Da C) The MS spectrum of AtRAP2.12$_{2\text{-}15}$ –Cys when incubated with iodoacetamide (IAM) and NEM, has a peak consistent with the AtRAP2.12$_{2\text{-}15}$-Cys-acetamide modification. D) the MS spectrum of AtRAP2.12$_{2\text{-}15}$-CysO$_2$H incubated with IAM and NEM has peaks consistent with both the AtRAP2.12$_{2\text{-}15}$-CysO$_2$H-acetamide and AtRAP2.12$_{2\text{-}15}$-CysO$_2$H-NEM modifications.

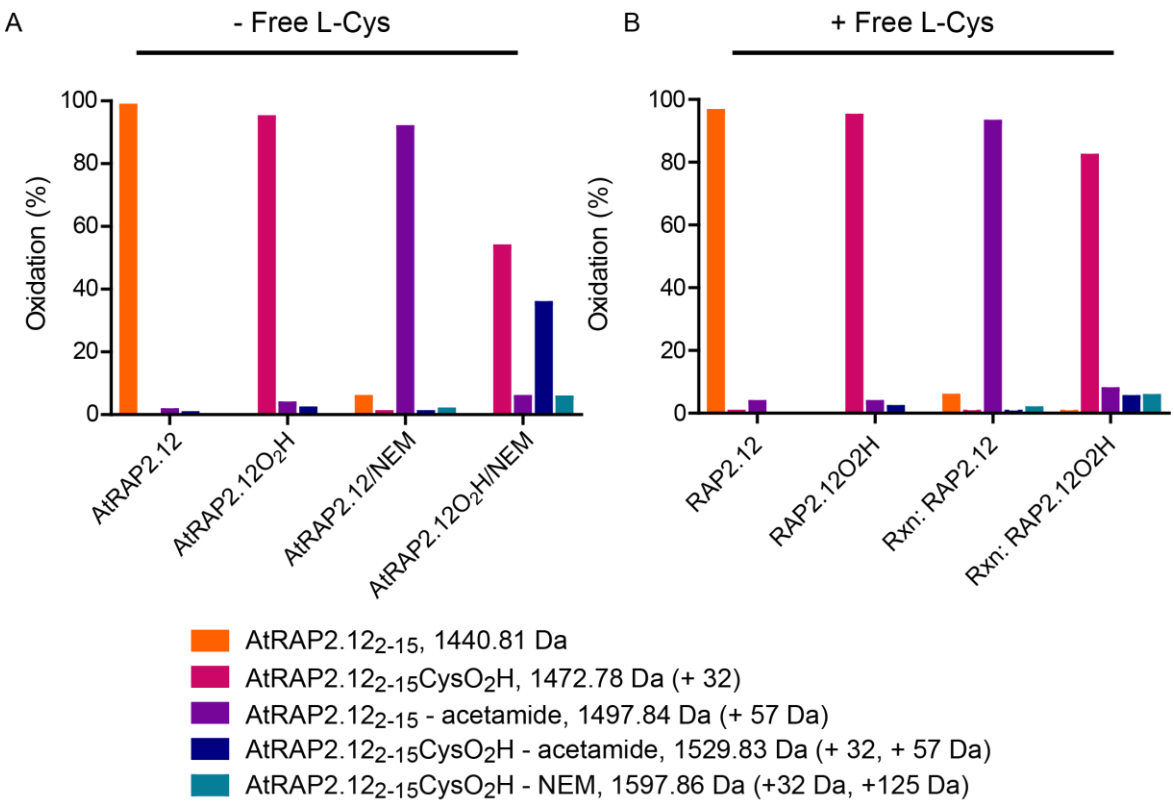


Fig. 6.3: AtRAP2.12$_{2\text{-}15}$CysO$_2$H detection using N-ethyl maleimide (NEM) with and without free L-Cys. A) AtRAP2.12$_{2\text{-}15}$CysO$_2$H incubated with IAM and NEM in the absence of free L-Cys has a MS peak consistent with AtRAP2.12$_{2\text{-}15}$CysO$_2$H-acetamide at approximately 40%. B) AtRAP2.12$_{2\text{-}15}$CysO$_2$H incubated with IAM and NEM in the presence of free L-Cys sees a significant reduction in the intensity of the AtRAP2.12$_{2\text{-}15}$CysO$_2$H-acetamide peak.

and NEM has peaks consistent with both the AtRAP2.12$_{2\text{-}15}$-CysO$_2$H-acetamide and AtRAP2.12$_{2\text{-}15}$-CysO$_2$H-NEM modifications. These MS results suggest that the thiol-capping IAM not only binds to AtRAP2.12$_{2\text{-}15}$Cys to form AtRAP2.12$_{2\text{-}15}$Cys-acetamide, but can also label AtRAP2.12$_{2\text{-}15}$CysO$_2$H. This could block the CysO$_2$H moiety from targeted NEM labelling, and may explain the relatively low intensity of the AtRAP2.12$_{2\text{-}15}$-CysO$_2$H-NEM peak compared to the AtRAP2.12$_{2\text{-}15}$-CysO$_2$H-acetamide peak (Fig. 6.2, D).

Consequently the experiment was repeated, with the addition of free L-Cys to compete with CysO$_2$H for reaction with IAM. The results are presented as a bar

graph, to show the relative intensities of each species calculated from the spectra using Agilent Quantitative Analysis. Mass spectrometry (MS) results showed a large decrease in the AtRAP2.12$_{2\text{-}15}$CysO_2H-acetamide MS peak in the presence of free L-Cys (1529.83 Da, Fig. 6.3, B). In the AtRAP2.12$_{2\text{-}15}$ MS spectrum, the AtRAP2.12$_{2\text{-}15}$-acetamide peptide peak is present in excess of 90% (1497.84 Da, Fig. 6.3, A). These results suggest that IAM has a preference for reduced Cys over oxidised CysO_2H. Therefore the addition of L-Cys is necessary to ensure N-terminal CysO_2H remains free for maleimide labelling. However, no increase in the AtRAP2.12$_{2\text{-}15}$CysO_2H-NEM peak intensity was observed upon addition the addition of L-Cys (5.3% versus 5.4%, Fig. 6.3, B).

The samples were analysed by E. Pires (Dept. of Chemistry) using tandem MS to determine the location of the NEM group on the AtRAP2.12$_{2\text{-}15}$CysO_2H-NEM peptide (1597.86 Da). Results detected the expected mass of the maleimide-AtRAP2.12$_{2\text{-}15}$CysO_2H peptide, but concentrations were too low to visualise the CysO_2H-NEM fragment (Appendix 6.3).

6.2.4 Increasing AtRAP2.12$_{2\text{-}15}$CysO_2H-NEM formation

To determine the cause of the low levels of AtRAP2.12$_{2\text{-}15}$CysO_2H-NEM formation (approximately 5%), the pH of the NEM-labelling buffer was examined. The pH of the NEM-labelling buffer can drift below the suggested pH for the maleimide–labelling step [264] as the pH of the buffer (Na_2HPO_4/citric acid) is highly temperature dependent. It was thought that this might explain the apparent low levels of AtRAP2.12$_{2\text{-}15}$CysO_2H-NEM formation observed

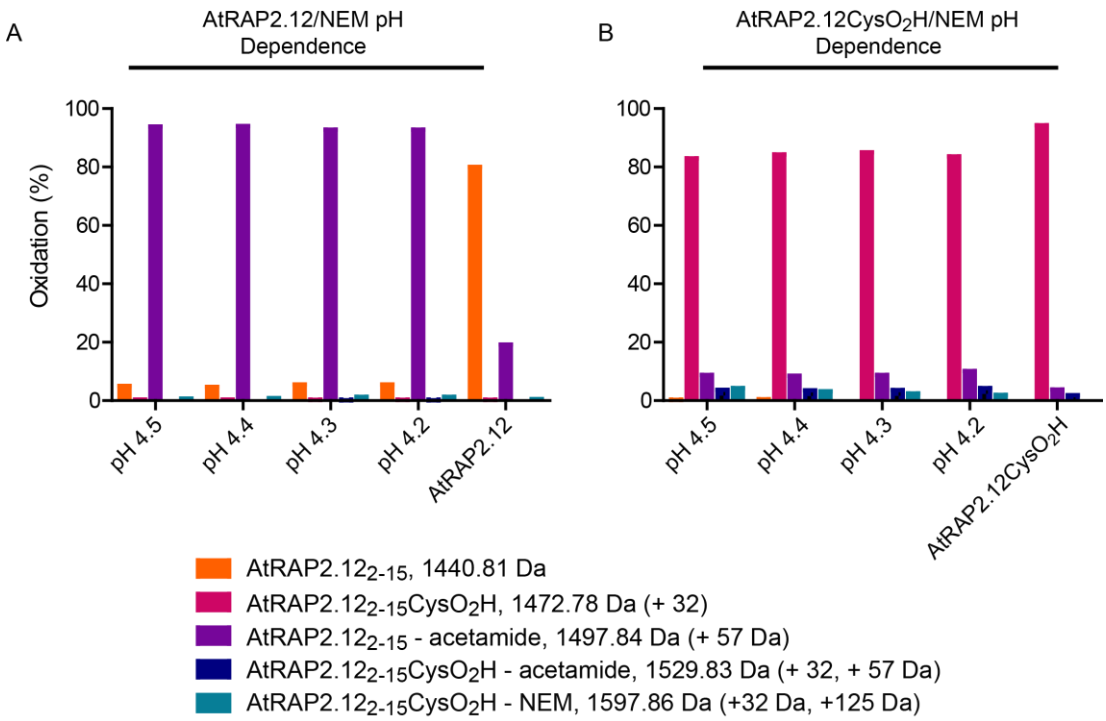


Fig. 6.4: AtRAP2.12$_{2\text{-}15}$CysO$_2$H detection using N-ethyl maleimide (NEM) as a function of pH. A) As predicted, the pH of the NEM step did not affect the formation of AtRAP2.12$_{2\text{-}15}$-acetamide. B) AtRAP2.12$_{2\text{-}15}$CysO$_2$H-NEM formation had robust pH tolerance, although the overall yield remained low (approximately 5%)

previously (Fig. 6.2 and 6.3). A pH-dependency screen showed that the AtRAP2.12$_{2\text{-}15}$CysO$_2$H-NEM modification step had good pH tolerance (Fig. 6.4). By a small margin, the highest percentage of AtRAP2.12$_{2\text{-}15}$CysO$_2$H-NEM formation was found at pH 4.5 (5%) compared with buffers of pH 4.4-4.2. The percentages of AtRAP2.12$_{2\text{-}15}$CysO$_2$H-NEM formation remained low in all samples, with pH drift likely not contributing very much to the overall low yield.

As neither the pH nor the free L-Cys addition increased the formation of AtRAP2.12$_{2\text{-}15}$CysO$_2$H-NEM, the NEM-peptide incubation time was increased from 1 hour to overnight (approximately 15 hours). This increased the yield of AtRAP2.12$_{2\text{-}15}$CysO$_2$H-NEM to 67% (Fig. 6.5, B). This process was also repeated

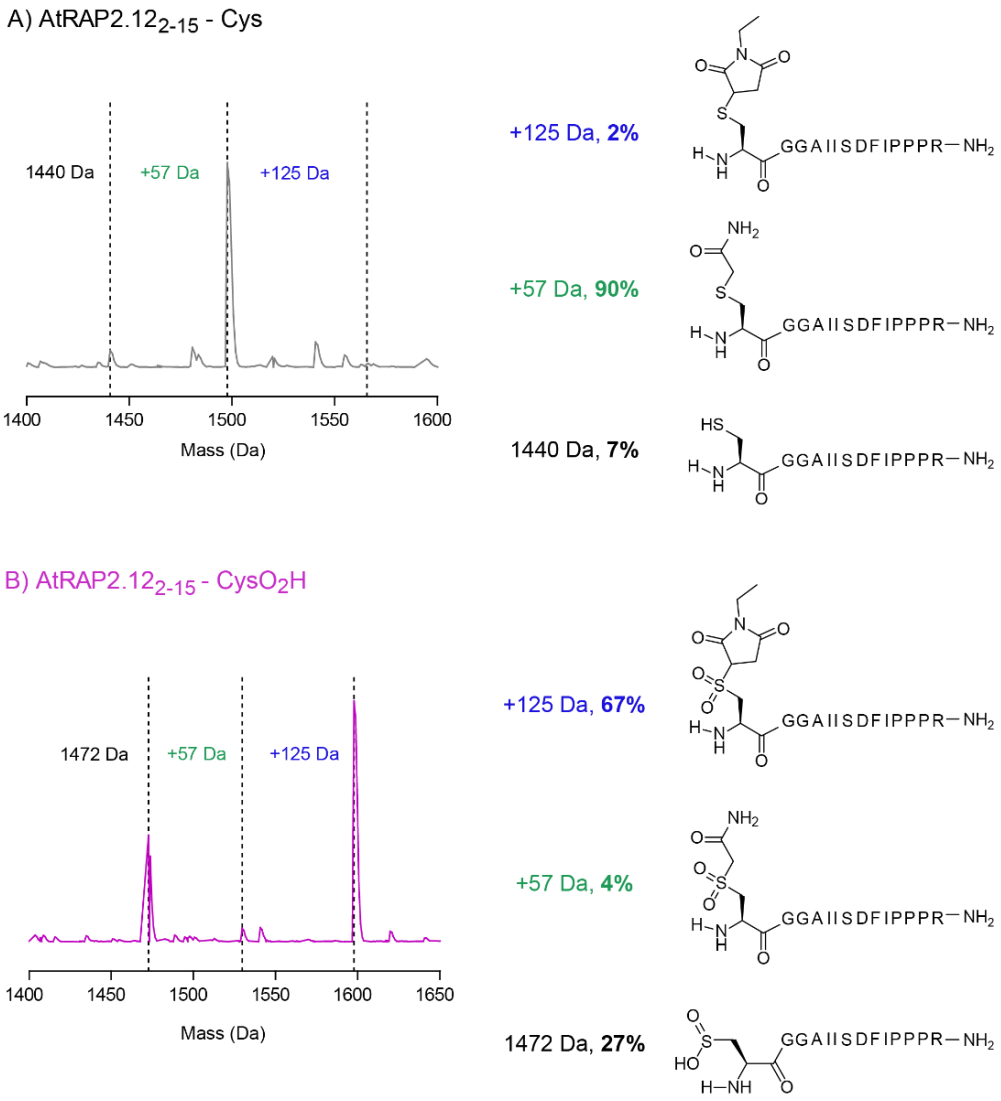


Fig. 6.5: Increasing the yield of AtRAP2.12$_{2-15}$CysO$_2$-NEM formation. A) AtRAP2.12$_{2-15}$-Cys-acetamide formation as expected with an IAM thiol-blocking step. B) the yield of AtRAP2.12$_{2-15}$-CysO$_2$H-NEM increased significantly with increased incubation time

with fresh buffers, but this failed to further improve upon the yield of AtRAP2.12$_{2-15}$CysO$_2$H-NEM (Fig. 6.6).

6.2.5 Peptide CysO$_2$H detection with TAMRA-5-maleimide

It was decided that these conditions (prolonged incubation time with the maleimide probe) were sufficient to trial the reporter-maleimide reagent, 5-Carboxytetramethylrhodamine – 5-maleimide (TAMRA-5-maleimide, Chapter

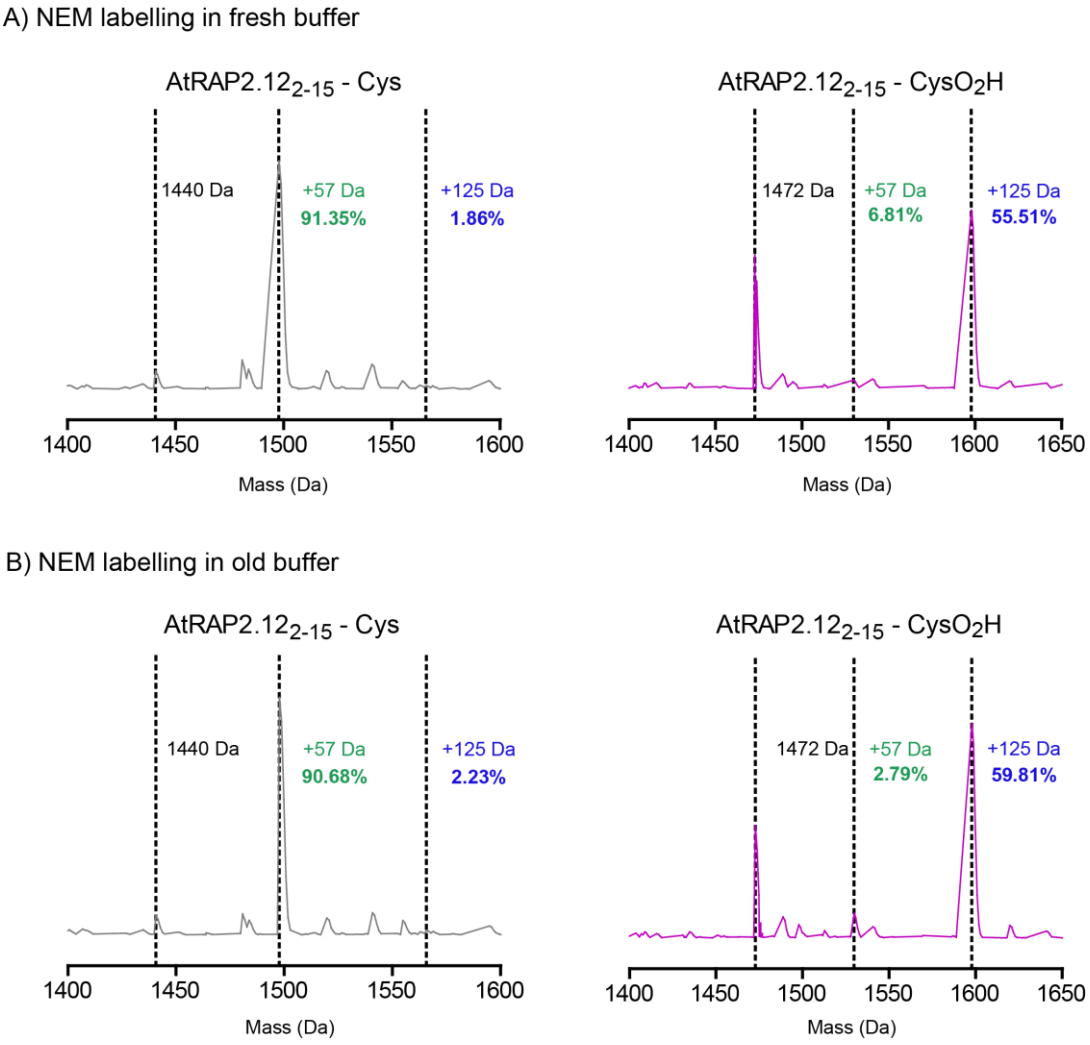


Fig. 6.6 The effect of fresh Na_2HPO_4, citric acid, 6M urea buffer on the yield of AtRAP2.12$_{2-15}$CysO$_2$H-NEM formation A) The yield of AtRAP2.12$_{2-15}$Cys-NEM was <2%. The yield of AtRAP2.12$_{2-15}$CysO$_2$H-NEM formation in fresh buffer was >55%. B) The yield of AtRAP2.12$_{2-15}$CysO$_2$H-NEM formation in old buffer was consistent with previous results.

2, Section 2.5.1). TAMRA is a widely used fluorophore in the preparation of bioconjugates, and its conjugation to the maleimide moiety was anticipated to help visualise potential PCO substrate proteins in cell lysates. As the TAMRA-5-maleimide probe is light sensitive, AtRAP2.12$_{2-15}$CysO$_2$H was incubated with TAMRA-5-maleimide overnight in the dark at room temperature. Analysis of the peptide by MS post-incubation suggested a 48% yield of the AtRAP2.12$_{2-}$

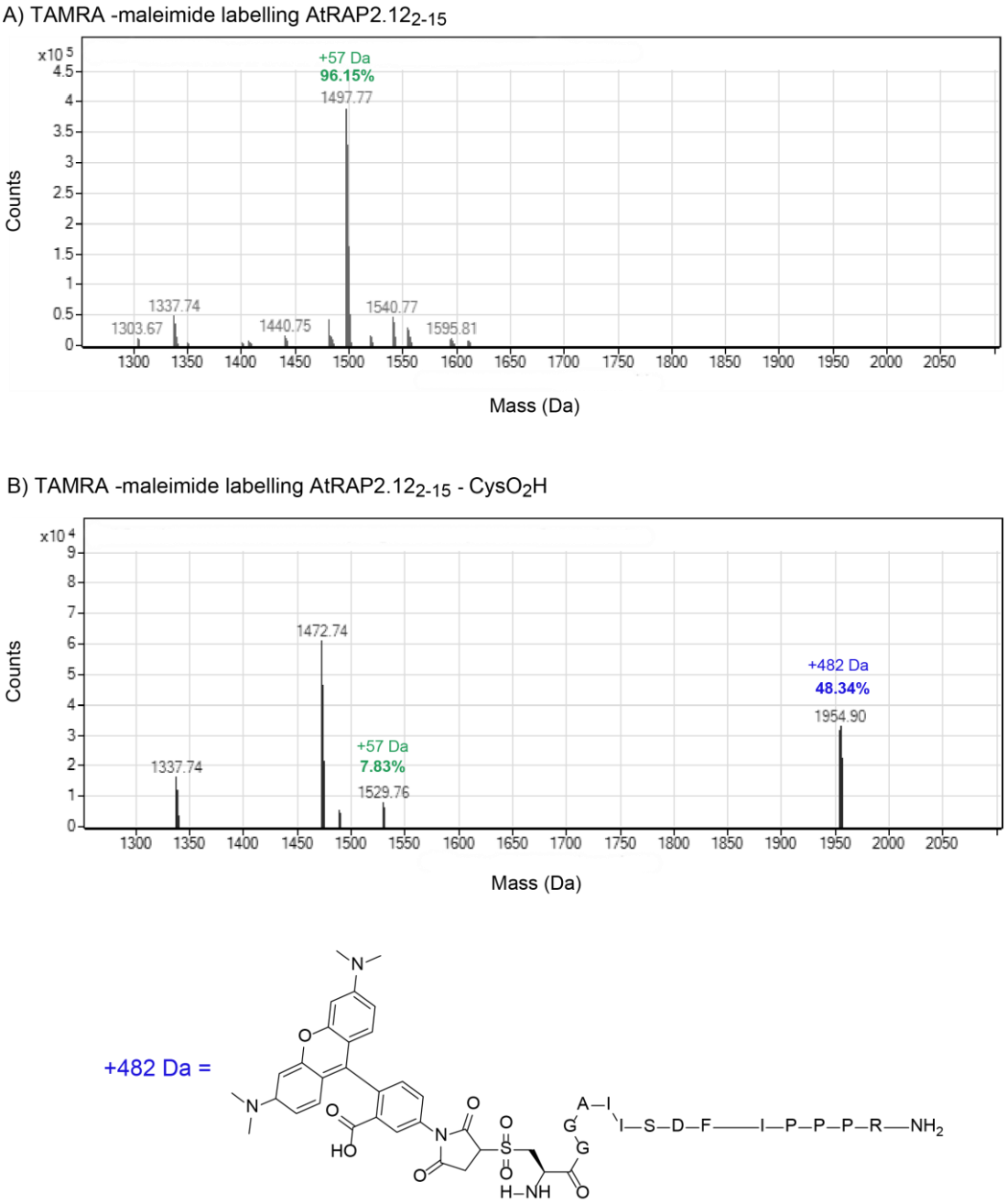


Fig. 6.7: AtRAP2.12$_{2-15}$CysO$_2$H detection using 5-Carboxytetramethylrhodamine – maleimide (5-TAMRA-maleimide). A) The yield of AtRAP2.12$_{2-15}$-acetamide formation was reproduced, with comparable yields to previous experiments B) The yield of AtRAP2.12$_{2-15}$CysO$_2$H-TAMRA formation was approximately 50%.

$_{15}$CysO$_2$H-TAMRA peptide (Fig. 6.7). This was considered satisfactory evidence that the TAMRA-5-maleimide reporter could bind CysO$_2$H to a degree that was likely sufficient to enable CysO$_2$H detection in cell lysates.

6.3 Investigation of N-terminal $CysO_2H$ proteins in Chlamydomonas reinhardtii lysates

6.3.1 Introduction

This section describes the use of the TAMRA-5-maleimide labelling technique in cell lysates (Fig. 6.8), and the subsequent identification of hypoxia-relevant PCO substrates in *C. reinhardtii*. This cell line was chosen for the labelling trial as *C. reinahardtii* cells are easy to grow (exponential phase in 3-4 days [268]) compared to *K. nitens* which is much slower (exponential phase in 2-3 weeks, Chapter 7).

6.3.2 Culturing *C. reinhardtii* and exposure to hypoxia

C. reinhardtii cells were grown in HSMA media at 20 °C under a light/dark cycle (illuminance of 1350 lx) of 12 hours under atmospheric conditions (*C. reinhardtii* cells provided by Q. Zhang, Chapter 2, Section 2.5.2). In Experiment A (Fig. 6.9, A), *C. reinhardtii* cells were grown for 14 days, with fresh media added prior to hypoxia treatment to ensure trace elements were not limiting. Experiment A was the first time this labelling technique was attempted. Due to the complexity of the protocol, it was thought prudent to practise the entire method before using precious cell lines. As such, priority was given to a high biomass rather than the phase of growth of the cells; harvesting the cells in stationary phase. In the biological repeat, experiment B (Fig. 6.9, B), a new *C. reinhardtii* culture was grown to exponential phase prior to hypoxia treatment to ensure that the cells were healthy.

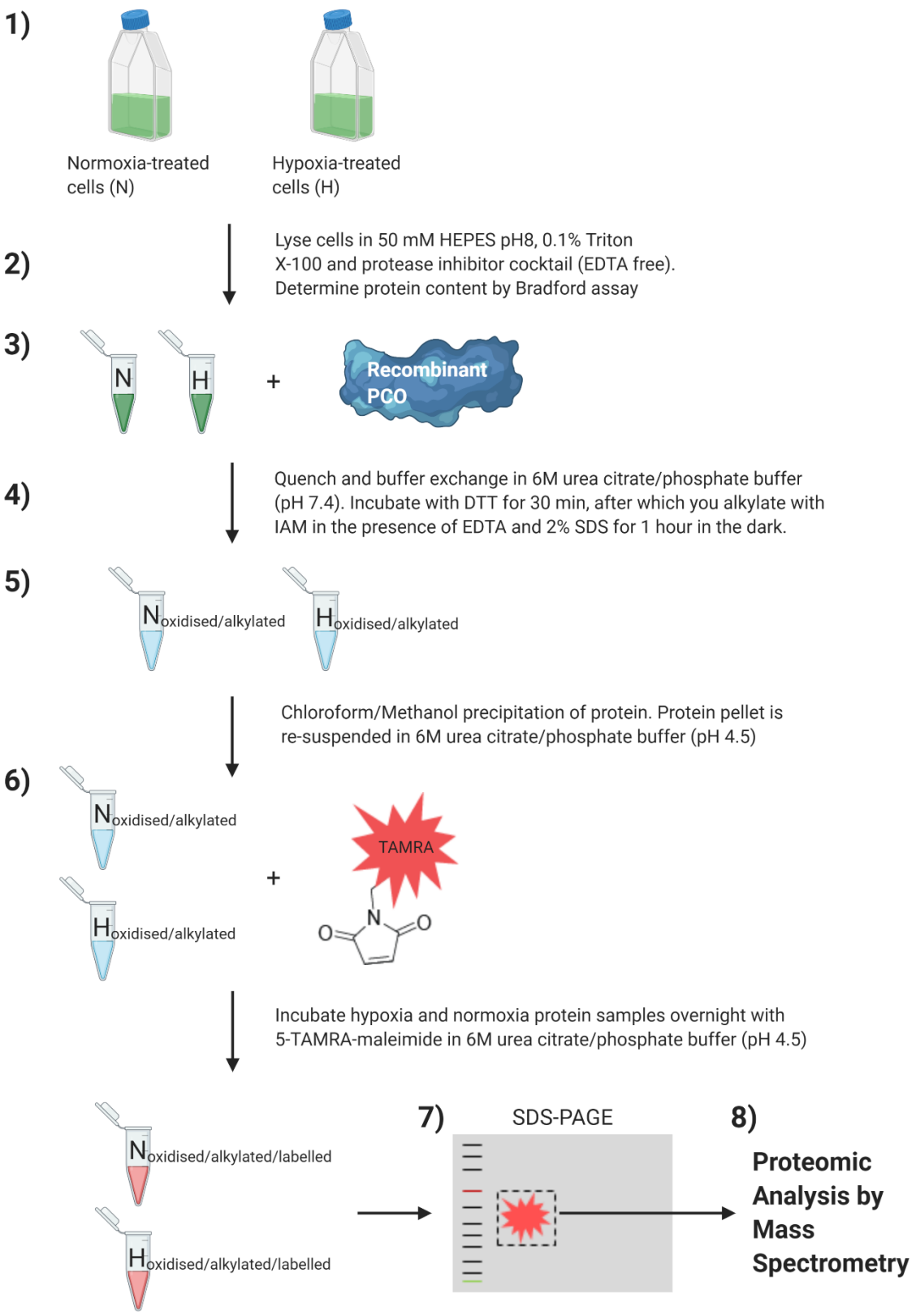


Fig. 6.8: Protocol for the detection of $CysO_2H$ with TAMRA fluorescence in cell lysates. 1) Cells are incubated under normoxia and hypoxia for 6 hours 2) Treated cells are lysed, and protein content determined by Bradford assay. 3) Lysates are incubated with recombinant PCOs 4) the reaction is quenched by buffer exchange in 6 M urea citrate/phosphate buffer pH 7.4 and incubated with IAM. 5) oxidised/alkylated lysates are precipitated and re-suspended in 6 M urea citrate/phosphate buffer pH 4.5. 6) 5-TAMRA maleimide was added to the processed lysates and incubated overnight. 7) the labelled lysates were analysed by SDS-PAGE. 8) Fluorescent gel bands were excised and analysed by mass spectrometry. Created with **BioRender.com**.

Once grown, the cell culture was split into two flasks. Hypoxia was achieved by bubbling N_2 through one of the cultures for 6 hours (*C. reinhardtii*$_{Hypoxia}$), with the second culture exposed to atmospheric conditions (*C. reinhardtii*$_{Normoxia}$). The hypoxia exposure time was selected based on experiments previously conducted with *C. reinhardtii* cells [269]. Both cultures were kept at 20 °C in the dark during the treatment. It is during the dark cycle the cells would respire, and O_2 would be required for cellular processes.

6.3.3 Detecting N-terminal cysteine sulfinic acid proteins

6.3.3.1 In-gel detection of cysteine sulfinic acid proteins

The protocol for labelling proteins containing $CysO_2H$ is outlined in Fig. 6.8. The cells were harvested and lysed and protein content was determined by Bradford assay at n=2 (Chapter 2, Section 2.5.5). This ensured that the same quantity of protein was processed from both the *C. reinhardtii*$_{Hypoxia}$ and *C. reinhardtii*$_{Normoxia}$ lysates. The lysates were incubated with and without recombinant His_6-KnPCO (Experiment A) and with and without His_6-KnPCO and His_6-MpPCO (Experiment B) for 90 min to convert the N-terminal Cys of any PCO substrates present in the *C. reinhardtii* lysates to $CysO_2H$.

The PCO-treated lysates were processed using the IAM and TAMRA-5-maleimide steps outlined above (Fig. 6.8 and Chapter 2, Section 2.5.1). The TAMRA labelled lysate was analysed *via* SDS-PAGE and the TAMRA fluorescence detected using the BioRad ChemiDoc XRS+ UV transilluminator

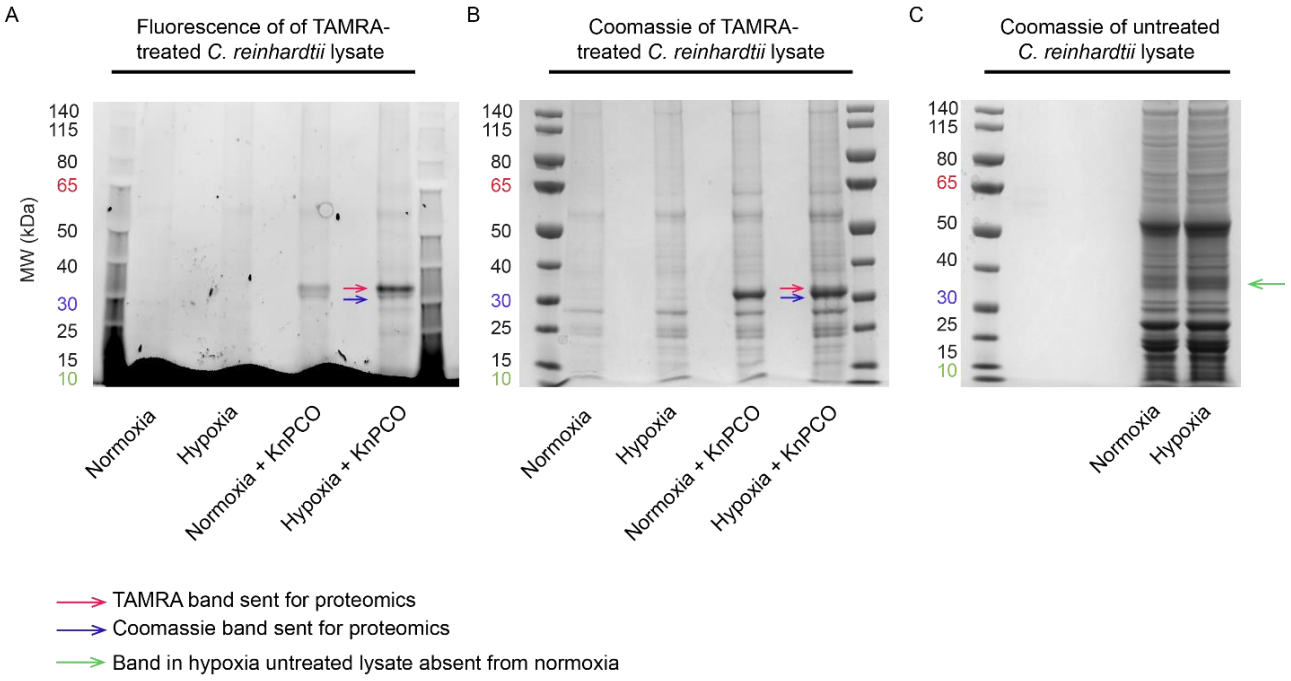


Fig. 6.9: High resolution detection of $CysO_2H$ with TAMRA fluorescence in *C. reinhardtii* cells exposed to hypoxia (Experiment A). A) A clear increase in the intensity of the 5-TAMRA fluorescent band is observed at approximately 35 kDa in the hypoxia/KnPCO treated *C.reinhardtii* lysate, compared to the normoxia/KnPCO treated *C.reinhardtii* lysate with a fluorescence intensity of 1:3 normoxia/KnPCO: hypoxia/KnPCO. B) the same gel, but stained using Coomassie general protein stain and C) untreated normoxia and hypoxia *C. reinhardtii* lysates, with a band at the 35 kDa mark in the hypoxia lane, that is absent from the normoxia lane

with a 605/50 nm filter (TAMRA excitation emission maxima at 546/579 nm, Chapter 2, Section 2.5.6).

The results from both Experiment A and its biological repeat, B, suggest a potential upregulation of a KnPCO substrate in *C. reinhardtii* cells after exposure to hypoxia (Fig. 6.9-6.10). The fluorescent signal at 35 kDa was approximately three times more intense in the KnPCO-treated *C. reinhardtii*$_{Hypoxia}$ lysate compared to the *C. reinhardtii*$_{Normoxia}$ lysate for both Experiments A and B. It is not clear why the recombinant MpPCO-treated lysates failed to produce the same upregulation pattern as those seen in with

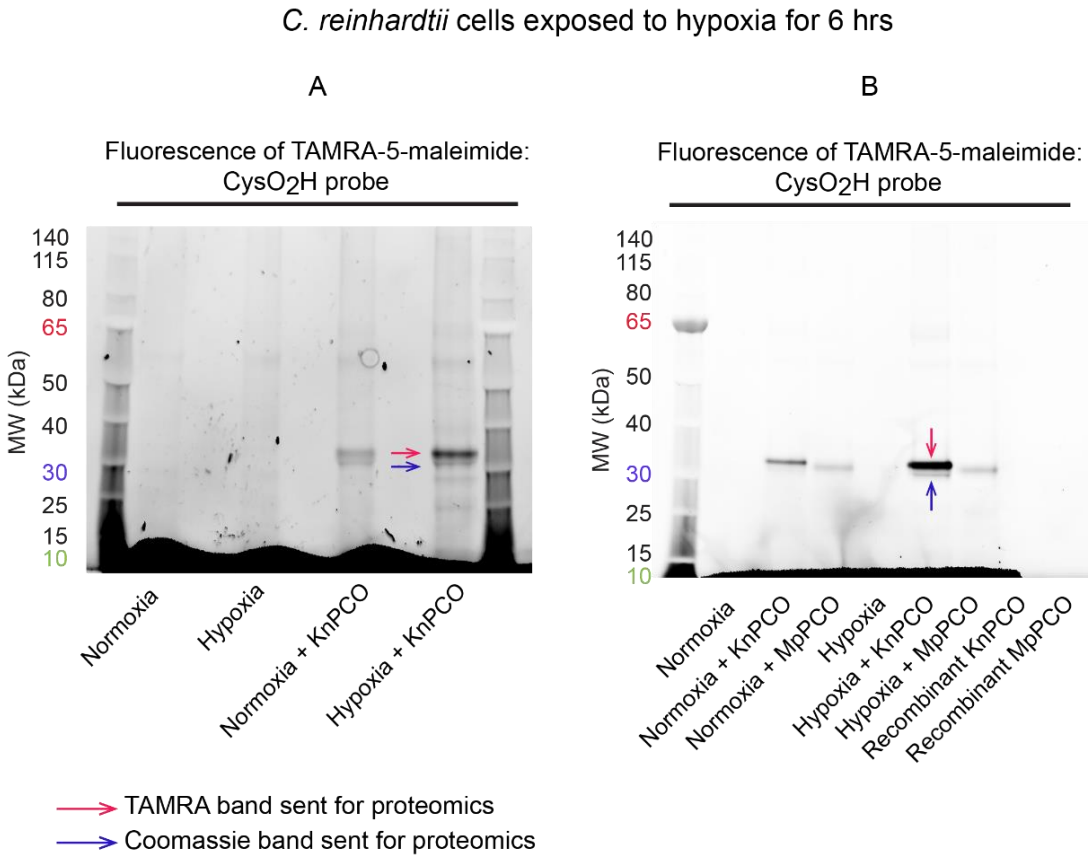


Fig. 6.10: Biological repeat of $CysO_2H$ detection with TAMRA fluorescence in *C. reinhardtii* cells exposed to hypoxia (Experiment A and B). A) A clear increase in the intensity of the 5-TAMRA fluorescent band is observed at approximately 35 kDa in the hypoxia/KnPCO treated *C.reinhardtii* lysate, compared to the normoxia/KnPCO treated *C.reinhardtii* lysate. B) this result was repeated a second time, with a fluorescence intensity of 1:3 normoxia/KnPCO: hypoxia/KnPCO for both experiment A and B.

KnPCO-treated lysates. It is worth noting that KnPCO and MpPCO do not have the same substrate profiles. For example, $ZPR2_{2\text{-}15}$ is a substrate of MpPCO, but not of KnPCO (Fig. 4.10).

While these results seem promising, caution must be taken when asserting upregulation of a substrate with hypoxia using these fluorescent gel results. His_6-KnPCO has a MW of ~30 kDa and deposits in the same region we see the intense fluorescent bands and seems to non-specifically bind TAMRA-5-maleimide (Fig. 6.14). It could be that simple processing error gives the

impression of hypoxia upregulation, so these findings must be verified in the future using another technique. It is also worth noting that the probe likely does not differentiate between N-terminal and internal $CysO_2H$ moieties, and could be detecting a general upregulation of $CysO_2H$ with hypoxia as opposed to a KnPCO substrate.

6.3.3.2 Proteomic analysis of excised gel fragments

Proteomic analysis was carried out on the TAMRA-labelled normoxia and hypoxia KnPCO-modified bands from the SDS-PAGE gel in experiment A (Fig. 6.9, A) to try and identify the labelled proteins. The TAMRA-labelled hypoxia and normoxia lysates were run on a 4-12% polyacrylamide bis-tris SDS-PAGE gel. The top intense fluorescent band was excised separately from the bottom band as the TAMRA pink/orange band could been seen with the naked eye, and sat above the Coomassie band (Fig. 6.9 A and B). Only one normoxia band was taken, as no TAMRA band could be seen with the naked eye, as in the

				MC-initiating proteins in *C. reinhardtii* proteomic data			
SDS-PAGE Resolution	**Data File (Appendix)**	**PEAKS Peptide ID**	**Lysate Type**	**Excised Band**	**Q2HWK7**	**A4GRC6**	**Q09JZ4**
High	6.5	Fig. 6.11	Hypoxia	TAMRA	Yes	No	No
	6.6	Fig. 6.12	Hypoxia	Coomassie	No	Yes	Yes
	6.7	Fig. 6.13	Normoxia	-	No	Yes	Yes
Low	6.8	Appendix 6.11	Hypoxia	TAMRA	Yes	Yes	No
	6.9	Appendix 6.12	Hypoxia	Coomassie	Yes	Yes	Yes
	6.10	Appendix 6.13	Normoxia	-	Yes	Yes	Yes

Table 6.1: Detection of MC-initiating proteins in the *C. reinhardtii*$_{Normoxia}$ and *C. reinhardtii*$_{Hypoxia}$ cell lysates using mass spectrometry (Experiment A). Proteomic analysis of bands excised from high (Fig 6.9) and low (A 6.4) resolution SDS-PAGE analysis of Experiment A.

High Resolution SDS-PAGE: Mass spectrometry identifies one MC-initiating protein in the excised hypoxia TAMRA fluorescent band

Q2HWK7 (CrDES)

Unique Peptide Identifer 1 (-10lgP = 18.58):

#	b	Seq	y	#
1	164.07	Y		7
2	301.13	H	725.39	6
3	400.20	V	588.34	5
4	513.28	L	489.27	4
5	627.33	N	376.18	3
6	742.35	N(+.98)	262.14	2
7		K	147.11	1

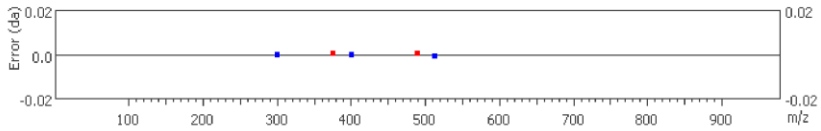


Fig. 6.11: PEAKS identified one MC-initiating protein Q2HWK7 (CrDES) in the hypoxia TAMRA fluorescent band excised from the high resolution SDS-PAGE. The annotated mass spectrometry peaks for the Q2HWK7 (CrDES) unique identifier peptide.

hypoxia lane. These well resolved TAMRA/Coomassie bands were sent for proteomic analysis separately (Digital Appendix 6.5-6.7). Another SDS-PAGE gel of the same lysates, but with poorer resolution between the TAMRA/Coomassie bands were also sent for mass spectrometry analysis (Digital Appendix 6.8-6.10).

The excised proteins were digested using Trypsin prior to proteomic analysis by E. Pires (Chemistry Department). The resulting peptide fragments were analysed on an Orbitrap Elite ETD nLC-MS. Peptides were identified and quantified using the PEAKS Studio platform (Appendix 6.5-6.10). Only peptides with a -10logP of 15 or above were reported. Each protein discussed here had at least one unique identifier peptide with a –10logP of 18 or above (Table 6.1). Proteomics identified only one MC-initiating protein in the TAMRA band from the high resolution SDS-PAGE gel of the *C. reinhardtii*$_{\text{Hypoxia}}$ lysate (Fig. 6.9, **pink** and Fig. 6.11): acyl-lipid omega-13 desaturase (accession number: Q2HWK7

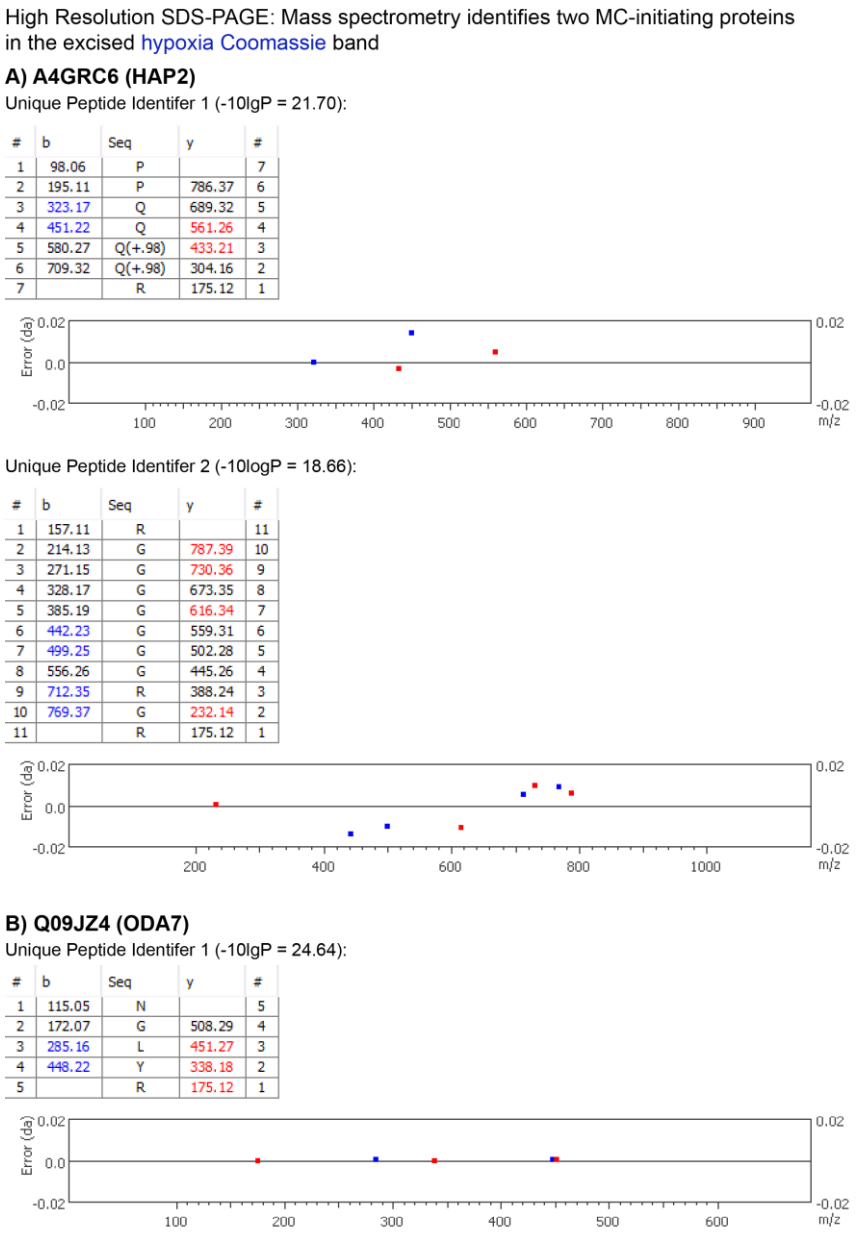


Fig. 6.12: PEAKS identified two MC-initiating proteins, A4GRC6 (HAP2) and Q09JZ4 (ODA7), in the hypoxia Coomassie band excised from the high resolution SDS-PAGE. The annotated mass spectrometry peaks for the A4GRC6 (HAP2) and Q09JZ4 (ODA7) unique identifier peptides.

[270]). Proteomics of the Coomassie band from the high resolution SDS-PAGE gel of the *C. reinhardtii*$_{\text{Hypoxia}}$ lysate (Fig. 6.9, **blue** and Fig. 6.12) identified two MC-initiating proteins; cellular fusion protein, hapless 2 (HAP2, accession number:

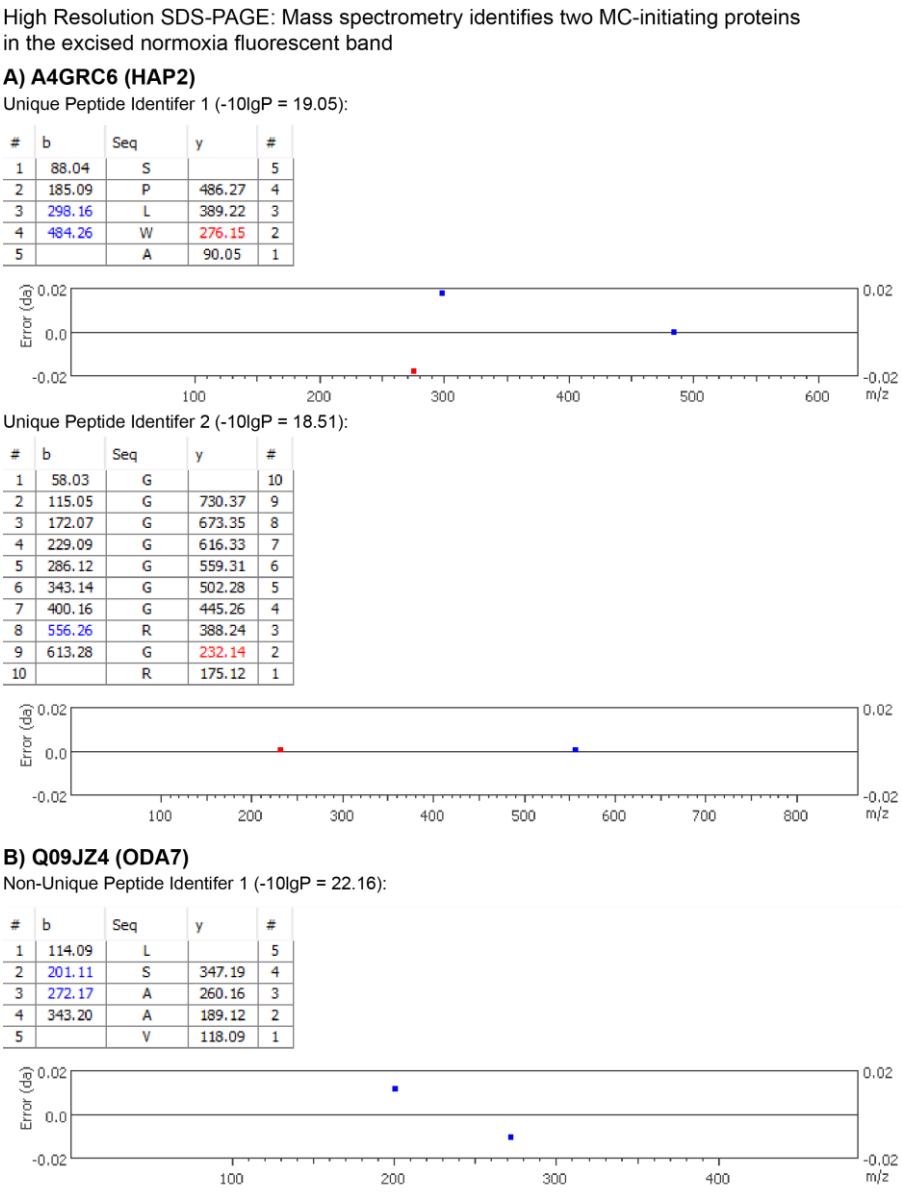


Fig. 6.13: PEAKS identified two MC-initiating proteins, A4GRC6 (HAP2) and Q09JZ4 (ODA7), in the normoxia fluorescent band excised from the high resolution SDS-PAGE. The annotated mass spectrometry peaks for the A4GRC6 (HAP2) and Q09JZ4 (ODA7) identifier peptides.

A4GRC6[270]) and predicted leucine-rich repeat-containing protein ODA7 (accession number: Q09JZ4 [270]) implicated in bend propagation in *C. reinhardtii* flagella [271]. The acyl-lipid omega-13 desaturase (Q2HWK7) was not found in the proteomic data of the Coomassie band. The band from the high resolution SDS-PAGE gel of the *C. reinhardtii*$_{\text{Normoxia}}$ gel band was found to also contain

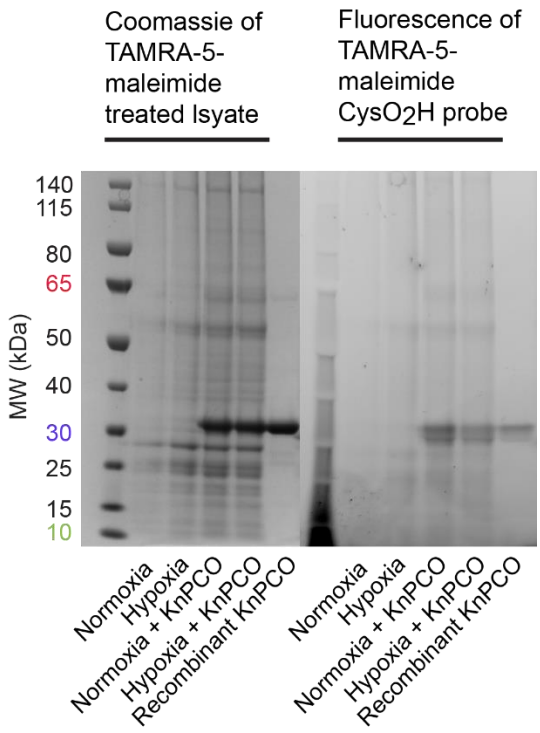
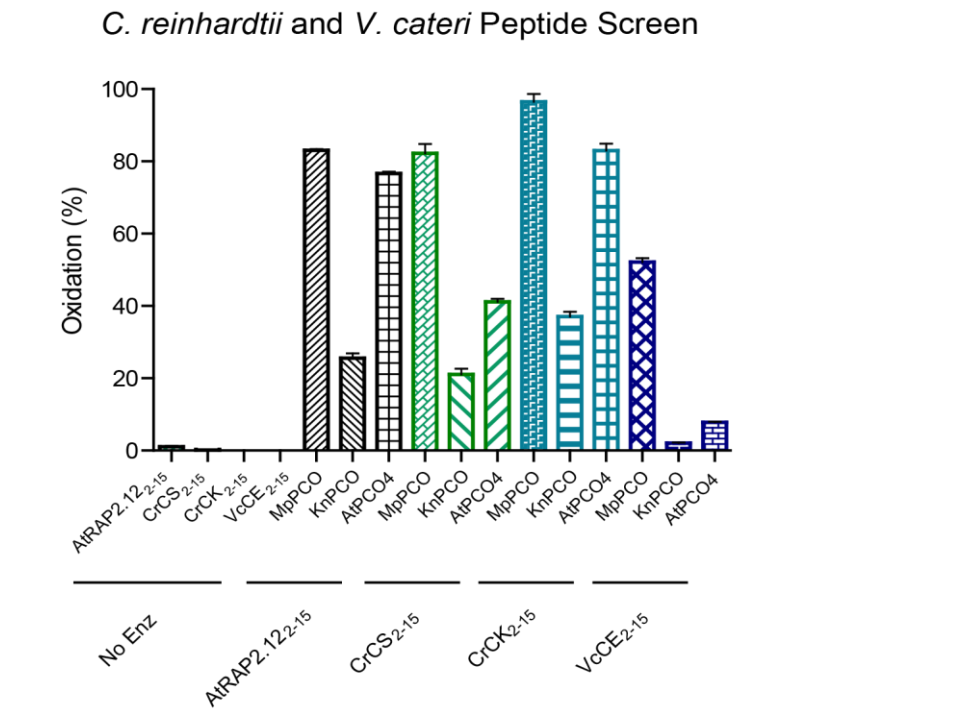


Fig. 6.14: His_6-KnPCO interference with TAMRA labelling. This SDS-PAGE results illustrates how added recombinant His_6-KnPCO interferes with TAMRA labelling.

the A4GRC6.2 and Q09JZ4 proteins, but not the Q2HWK7 protein found in the TAMRA *C. reinhardtii*$_{Hypoxia}$ fluorescent gel band (Fig. 6.13). The same MC-initiating proteins were found in the gel bands excised from the low resolution SDS-PAGE gel (Appendix 6.11-6.13).

To understand if the MC-N termini of the proteins identified in the proteomic data are conserved across plant species, and therefore could be a conserved PCO substrate, the protein sequences were used as templates for a protein BLAST a. BLASTp searches revealed that the acyl-lipid omega-13 desaturase identified in the HRA of the *C. reinhardtii*$_{Hypoxia}$ gel band (accession number: XP_001698534.1 [270]) has an MC-initiating homolog in *Volvox carteri*, (accession number: XP_002951536.1 [272]). This is important as the conservation of the MC-initiating N-terminus between *C. reinahardtii* and *V. carteri* suggests the MC-initiating N-terminus is biologically and possibly evolutionarily relevant. The identified protein is a *V. carteri* hypothetical protein, and does not have any specific published literature alluding to its function. BLASTp results indicated that neither HAP2 nor the leucine-rich repeat-containing protein had MC-



CrDES = CrCS$_{2-15}$: CRPTDSDSGPALPS, 1402.54 Da

Q09JZ4 = CrCK$_{2-15}$: CACMTKEALLEVCK, 1540.7 Da

V. cateri CrDES homolog = VcCE$_{2-15}$: CQLSESSAPAAVPE, 1388.55 Da

CrHAP2 = CrCQ$_{2-15}$: CRAIAVALIVYLAQ, 1502.8 Da (predicted)

Fig. 6.15: Screening of *C. reinhardtii* and *V. carteri* peptides with recombinant PCOs. With the exception of CrHAP2$_{2-15}$, which was insoluble and not visible on the RapidFire QTOF, all screened peptides were oxidised in the presence of recombinant PCOs. Analysed at n =3.

initiating homologs in other plant species. The *C. reinhardtii* MC-initiating proteins from the *C. reinhardtii* proteomic data had no MC-initiating homologs in *K. nitens* or *M. polymorpha*. The excised gel bands from the biological repeat, experiment B, have been sent for proteomic analysis with E. Pires (Dept. of Chemistry), but the results have yet to be received.

6.3.4 Activity of PCOs towards *C. reinhardtii* peptides

The proteomic analysis of the TAMRA fluorescent bands identified three MC-initiating proteins: CrDES – an acyl-lipid omega-13 desaturase, CrHAP2 - gamete fusion protein and Q09JZ4 – a leucine-rich repeat-containing protein. To verify these proteins as substrates of the PCOs, 14-mer peptides of the discovered proteins were incubated with recombinant PCOs; His_6-KnPCO, His_6-MpPCO and His_6-AtPCO 4 (Chapter 2, Section 2.5.7). This experiment was carried out in triplicate, and repeated. Results confirmed that the $CrDES_{2\text{-}15}$, $Q09JZ4_{2\text{-}15}$ and $VcCE_{2\text{-}15}$ peptides are substrates of the PCOs (Fig. 6.15). The $CrHAP2_{2\text{-}15}$ was insoluble under standard assay conditions and not visible on the RapidFire QTOF. $CrDES_{2\text{-}15}$, $Q09JZ4_{2\text{-}15}$ and $VcCE_{2\text{-}15}$ are the first reported potential PCO-substrates from algae.

6.4 Detection of N-terminal cysteine sulfinic acid in Klebsormidium nitens lysates

6.4.1 Introduction

Expression trials of CrPCO were unsuccessful – consistent with other algal-like PCOs (Chapter 4) so it was not possible to test whether the peptides were substrates for PCOs from the same algal species. However, recombinant KnPCO was successfully expressed (Chapter 4), and therefore the *K. nitens* system provides an opportunity to study potential new substrates with an endogenous PCO. This section describes the identification of hypoxia-relevant PCO substrates in *K. nitens,* using the TAMRA-5-maleimide labelling technique.

6.4.2 Culturing *K. nitens* for substrate hunting

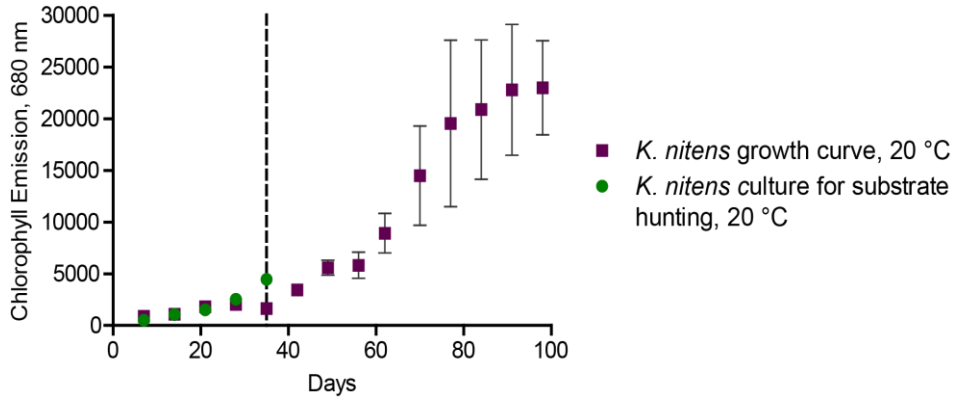


Fig. 6.16: Culturing *K. nitens* for PCO substrate hunting. *K. nitens* (NIES: 2286) cells were grown at 20 °C to early exponential phase prior to treatment with hypoxia.

K. nitens cells were grown in NIES C media at 20 °C under a light/dark cycle (light illuminance of 1350 lx) of 12 hours (*K. nitens*: NIES 2286, Chapter 2, Section 2.5.3). The *K. nitens* culture was grown to early exponential phase prior to hypoxia treatment to ensure the cells were healthy (Fig. 6.16). Once grown to early exponential, the culture was split into two flasks. Hypoxia was achieved by bubbling N_2 through one of the cultures for 6 hours (*K. nitens*$_{Hypoxia}$), with the second culture exposed to atmospheric conditions (*K. nitens*$_{Normoxia}$). Both cultures were kept at 20 °C in the dark during the treatment.

6.4.3 Detecting N-terminal cysteine sulfinic acid proteins

The same protocol for labelling $CysO_2H$-containing proteins was used for *K. nitens* cells as used in Section 6.3.3 (Fig. 6.8). The cells were harvested and lysed and protein content was determined by Bradford assay at n=2 (Chapter 2, Section 2.5.5). This ensured that the same quantity of protein was processed from both the *K. nitens*$_{Hypoxia}$ and *K. nitens*$_{Normoxia}$ lysates. The lysates were

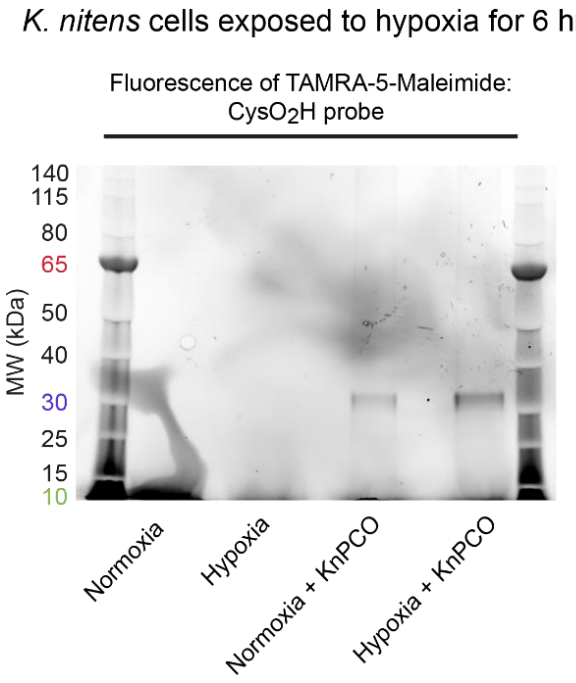


Fig. 6.17: Detection of $CysO_2H$ with TAMRA fluorescence in *K. Nitens* cells exposed to hypoxia. A higher intensity TAMRA fluorescent band is observed at approximately 35 kDa in the hypoxia/KnPCO treated *K. nitens* lysate, compared to the normoxia/KnPCO treated *K. nitens* lysate

incubated with and without recombinant His_6-KnPCO to convert any N-terminal Cys to $CysO_2H$ for labelling.

The PCO-treated lysates were processed using the IAM and TAMRA-5-maleimide steps outlined above (Fig. 6.8 and Chapter 2, Section 2.5.6). The TAMRA labelled lysate was analysed *via* SDS-PAGE and the TAMRA fluorescence detected using the BioRad ChemiDoc XRS+ UV transilluminator with a 605/50 nm filter (TAMRA excitation emission maxima at 546/579 nm. The fluorescent band in the hypoxia lane was 3 times more intense than the fluorescent band in the normoxia lane. This could be recombinant protein interference, a possibility described for the *C. reinhardtii* experiments (Fig. 6.14), but the consistent higher intensity of the band observed in the hypoxia treated experiment and proteomic data of the TAMRA-band is suggestive of a hypoxia-upregulated protein. (Fig. 6.17). To assess the protein content of the bands, the TAMRA and Coomassie bands were sent for proteomic analysis with E. Pires (Chemistry Department, results pending).

6.5 Discussion

This chapter reports the use of a maleimide-reporter probe for the identification of PCO substrates in cell lysates. Proteomic analysis of the *C. reinhardtii*$_{\text{Normoxia}}$ and *C. reinhardtii*$_{\text{Hypoxia}}$ SDS-PAGE gel bands identified three MC-initiating proteins. SDS-PAGE results suggest a higher concentration of $CysO_2H$ in the hypoxia band, with proteomic data suggesting this could be CrDES; CrDES – an acyl-lipid omega-13 desaturase. The proteomic data also identified; CrHAP2 - gamete fusion protein and Q09JZ4 – a leucine-rich repeat-containing protein. The upregulation with hypoxia will need to be verified, as it was shown His_6-KnPCO can interfere with TAMRA-5-maleimide labelling, likely *via* non-specific binding, nevertheless the consistent relative intensities of the bands in the hypoxia-treated samples is encouraging.

Mass spectrometry analysis of the high resolution SDS-PAGE TAMRA band of the *C. reinhardtii*$_{\text{Hypoxia}}$ lysate identified only CrDES, suggesting CrDES is responsible for the increased fluorescence relative to *C. reinhardtii*$_{\text{Normoxia}}$. CrDES was not detected in the high resolution SDS-PAGE Coomassie band of the *C. reinhardtii*$_{\text{Hypoxia}}$ lysate. It is recognised that this does not necessarily indicate that CrDES is upregulated under hypoxia, as simple proteomic processing discrepancies could account for the greater fluorescent signal in the hypoxia versus normoxia. While CrDES was the only MC-initiating protein identified in the HRS TAMRA band, the mass spectrometry results could also be masking a proteolytic cleavage event, where an upregulated protein is

cleaved to expose an N-terminal Cys for oxidation. CrDES has been reported upon previously in the literature [273], and contains a cytochrome b5 domain at the N-terminus and is reputed to have O_2 dependent desaturase activity for both linoleic acid and α-linolenic acid - a key step for betain lipid formation. This is in keeping with previous observations that desaturated fatty acids are upregulated under dark anoxia in *C. reinhardti*i cells [274]. CrDES also has an MC-initiating homolog in *V. carteri* (VcCE), which is also a substrate of the PCOs (Fig 6.15).

Assays with 14-mer peptides of the MC-proteins from *C. reinhardtii* and *V. carteri* confirmed a mass change of +32 Da in the presence of the PCOs, consistent with peptide oxidation. Further experiments are required to verify these proteins as true substrates, not just using biochemical analysis of CrPCO catalysis, but also verifying the proteins as substrates *in vivo*. As recombinant CrPCO has proven difficult to express, a synthetic *in vitro* system such as those described by Licausi *et al.* [275], could prove useful for determining if the full-length proteins are substrates of the endogenous CrPCO. Further work is currently underway in the group to produce CrPCO using truncated sequences and MBP solubility tags. Further analysis of the proteomic data collected here is required, not just focussing on MC-initiating proteins. Analysis of all the proteins identified could further contribute to understanding the *C. reinhardtii* hypoxia response [269, 274, 276-278]. Overall, these results suggest an algal metabolic response to hypoxia which appears to involve MC-initiating N-terminal

proteins. This is the first reported experimental link between the higher plant O_2 sensing pathway and the algal hypoxia response.

Chapter 7: The proteome of algae grown under low O_2

7.1 Summary

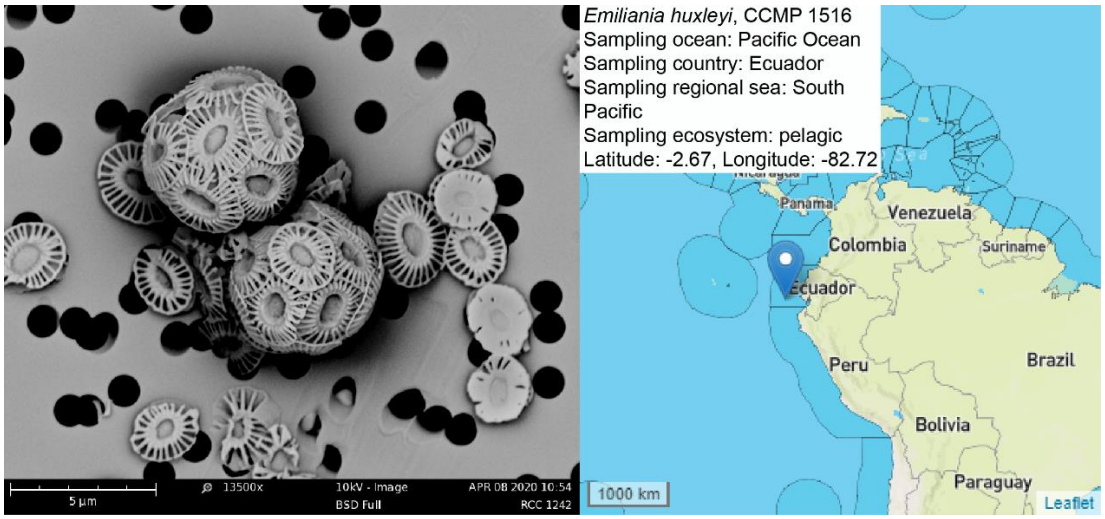


Fig. 7.1: Reference strain of *E. huxleyi*, CCMP 1516 and sampling location. This project uses a non-calcifying, non-motile variation of this reference strain for study, RCC 1242.

Emiliania huxleyi is a globally distributed coccolithophore and can be found in a diverse range of habitats [228]. The genome of *E. huxleyi* is thought to be the key to understanding its global distribution and extraordinary adaptive capabilities [228]. The *E. huxleyi* reference genome contains two putative PCO homologs, implying there may be a land plant-homologous O_2 sensing pathway in *E. huxleyi*. *E. huxleyi's* adaptive capacity, genomic plasticity and putative PCOs make it a powerful model for the study of algal O_2 sensing with respect to evolution and adaptations to global climate change. This chapter aimed to establish facilities and techniques for culturing *E. huxleyi* under different O_2 conditions, with a view to analysing the proteome of *E. huxleyi* cells grown under different O_2 concentrations. This chapter discusses progress

made in establishing suitable culturing facilities and uses preliminary proteomic data to identify MC-initiating proteins from the *E. huxleyi* proteome.

This chapter also focusses on culturing *Klebsormidium nitens* under various O_2 concentrations, again with a view to understanding proteomic changes as a function of O_2 availability. The *K. nitens* genome contains a functional, O_2 sensitive PCO and as such could shed some light on algal PCO function especially with respect to hypoxia. As a result of its acquisition or adaptation of genes fundamental to terrestrialization [225] and its aquatic habitat, *K. nitens* provides an opportunity to study hypoxia and PCO function from the perspective of the water to land transition. This chapter discusses establishing *K. nitens* culturing facilities under low O_2.

7.2 Initial culturing of E. huxleyi and proteomic analysis

7.2.1 Introduction

This section outlines initial culturing and proteomic experiments with *E. huxleyi*, RCC 1242, which is a non-calcifying (non-motile) variation of the reference strain, CCMP 1516 (Fig. 7.1) with a view to understanding the algal hypoxia response. Early experiments focussed on establishing the optimum cell growth conditions, suitable lysate preparation for proteomic analysis and initial proteomic analyses.

7.2.2 Culturing *E. huxleyi* using the Invivo$_2$400 chamber

A Baker-Ruskinn Invivo$_2$400 hypoxia chamber in the Chemistry Research Laboratory was chosen for algal culturing. The greatest obstacle to overcome

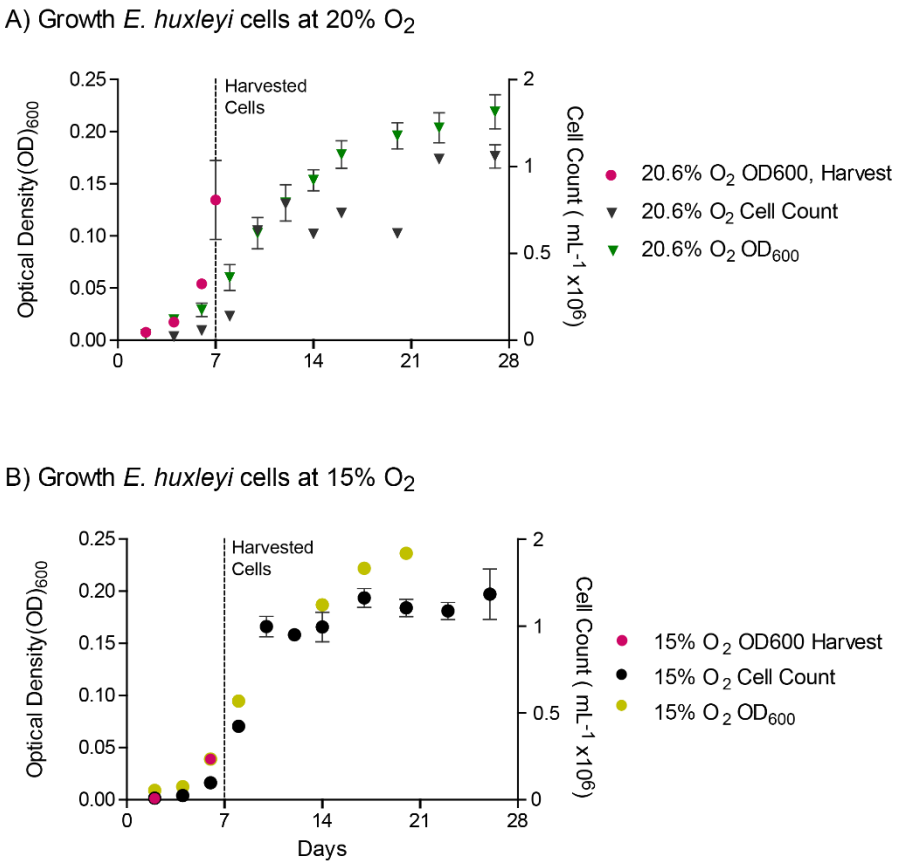


Fig. 7.2: Growing *E. huxleyi* under 15 and 20.6% O_2 using an Invivo$_2$400 chamber. A) The growth curve of *E. huxleyi cells* at 20.6% O_2 suggested mid-exponential growth was reached after 7 days. A follow-up growth at 20.6% O_2 was harvested at this point, and the cells stored at -80 °C for proteomics. B) The same procedure was used to grow *E. huxleyi* cells at 15% O_2.

while using this chamber was maintaining a stable temperature, as both the light source and the temperature control unit had to operate within the chamber, and the InvivO$_2$400 chamber itself had no temperature control capability. Furthermore, the room in which the InvivO$_2$400 chamber was housed was used for fermentation, and as such the internal temperature was often above 30 °C. A N°ICE Electronic Ice Bucket was used to maintain the temperature of the algae below ambient. An LED lamp was used as a light source, with an illuminance of 2700 lx.

E. huxleyi cells were sub-cultured from a culture collection RCC 1242 stock and grown as batch cultures at approximately 20 °C, under a 12 hour light/dark cycle in 80 mL synthetic ocean water (SOW) Aquil [234, 235] medium (Chapter 2, Section 2.6.1.1). The atmospheric O_2 concentration was varied using the Invivo$_2$400. The cells were sampled and analysed using optical density measurements at 600 nm (OD_{600}), with cell counts determined using a Beckman Coulter Counter (CC, cell size, 2-8 µm). The growth curves are based on parallel OD_{600} and cell count measurements. This was to ensure the OD_{600} measurements are representative of cell growth (Fig. 7.2) and could subsequently be used *in lieu* of cell count measurements, which were time consuming.

Once the growth phases of the cells were established *via* the growth curve, a second set of cultures were inoculated and grown at either 20.6 or 15% O_2. This second set of cultures were harvested during the exponential growth phase (Fig. 7.2) as this is when the culture is healthy and the cells are dividing. The cells were pelleted by centrifugation at 4,000 x g. The supernatant was discarded and the cells flash frozen in liquid nitrogen. The harvested cells were stored for proteomic analysis (Section 7.2.4).

Cultures were successfully grown at 10% O_2 (Appendix 7.1, A), but due to a power failure and a loss of atmospheric control from approximately day 10, the culture was restarted. A second growth at 10% O_2 was attempted, but the algae failed to grow (Appendix 7.1, B). When growth at 20.6% O_2 was repeated, cell

death was also observed (Appendix 7.2), indicating problems with the experimental set-up. Given concurrent problems with phage in the same laboratory, it was decided to establish an algal culturing facility using a different experimental set-up.

7.2.3 Establishing *E. huxleyi* culturing conditions in a PhO_2xBox

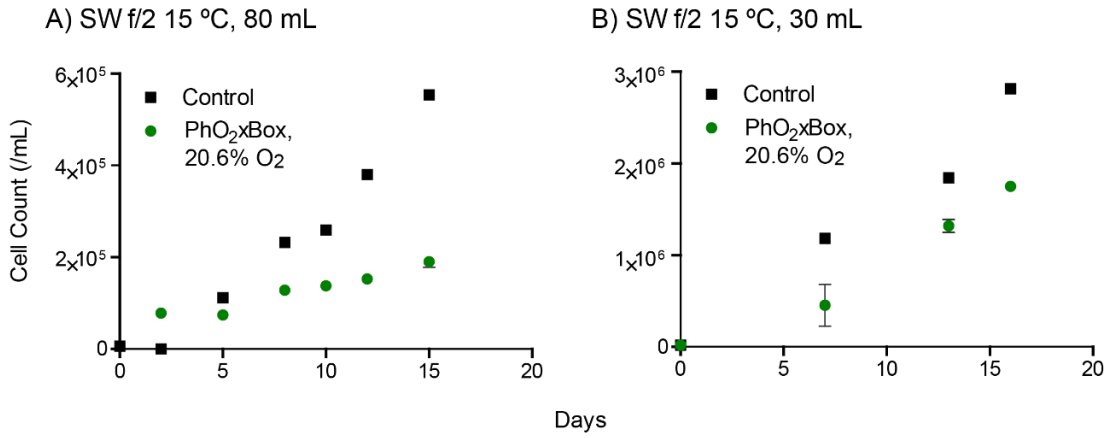


Fig. 7.3: Growing *E. huxleyi* in a PhO_2xBox. A) The culturing medium changed to seawater (SW) f/2. The culture volume was maintained at 80 mL, but growth rates were still poor. B) The volume was reduced to 30 mL and the cells began to more closely match control growth outcomes.

A Baker-Ruskinn PhO_2xBox for atmospheric control was established in the Dept. Earth Sciences Geobiology algal culturing facility and subsequent *E. huxleyi* (RCC 1242) cultures were grown in this environment (Chapter 2, Section 2.6.1). *E. huxleyi* cells were subcultured from a culture collection RCC 1242 stock and grown as batch cultures. The cultures grew in an atmospherically controlled Perspex box in a 15 °C incubator under a 12 hour light/dark cycle (light illuminance of 2700 lx). A variety of different mediums and volumes thereof were trialled, including; SOW f/2 [236, 237] and seawater (SW) f/2 medium. Cell growth was determined using cell counts on a coulter counter.

E. huxleyi cells failed to grow in 80 mL SOW f/2 cultures (Appendix 7.3). As the higher volume was necessary to produce a good biomass for proteomic analysis, SW f/2 was trialled. An 80 mL culture grew very slowly compared to the control (Fig. 7.3, A). Adjusting the volume from 80 to 30 mL increased the cell yield by an order of magnitude (Fig. 7.3, B). It is unclear why the volume had an effect on culture health. Unfortunately the measurement made on day 16 of the growth was the last measurement made before quarantining with COVID19 symptoms. Further lockdown restrictions blocked access to the algal culturing facilities. The Marine Biological Association had a limited seawater supply as a result of lockdown, and as such these experiments could not continue post-lockdown.

7.2.4 Trialling suitable protein extraction protocols for proteomics

Despite being unable to harvest hypoxia-grown *E. huxleyi* cultures, and assess the regulation of *E. huxleyi* proteins with respect to hypoxia, some preliminary proteomic analyses could be carried out. The resulting proteomic data was used to identify MC-initiating proteins in the *E. huxleyi* proteome. Proteins must have Cys residues at their N-termini to be substrates for PCO-oxidation and therefore regulation by the N-degron pathway. Understanding which MC-proteins are found in the *E. huxleyi* proteome will help us identify potential PCO substrates. This could be especially interesting as *E. huxleyi* is known to contain two proteins sequences with homology to the higher plant PCOs. This section also explores the use of different protein extraction techniques, to assess the best method by which to assess the proteome and identify the MC-

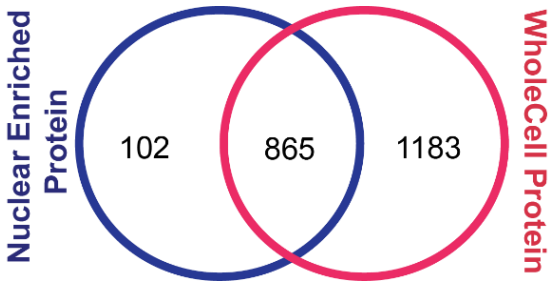


Fig. 7.4: The number of identified proteins in the nuclear enriched and whole cell protein extraction of *E. huxleyi* cells grown at atmospheric O_2. The proteomic data identified 102 unique proteins in the nuclear enriched extracted protein, and 1183 proteins in the whole cell protein extraction. The two extractions shared 865 identified proteins.

initiating proteins within and the whole-cell proteomics of *E. huxleyi* cells grown under 15 and 20.6 % O_2.

7.2.4.1 Nuclear enrichment versus whole-cell protein extraction

This section discusses two cellular protein extraction techniques; whole cell and nuclear enriched extraction. For the purposes of this experiment, *E. huxleyi* cells grown under 20.6% O_2 in Aquil medium as outlined in Section 7.2.2 were used. With our interest in gene-regulatory proteins, it was thought prudent to try a nuclear enriched protein extraction protocol to assess whether or not this gave an advantage over a whole cell protein profile. Prof. J. Mellor (Biochemistry Department) provided a protocol for generating nuclear protein enriched lysates (Chapter 2, Section 2.6.2).

The whole cell and nuclear enriched lysates were trypsin digested and analysed on a QExactive LC-MS/MS interfaced with a Dionex U3000 UPLC by Dr. J. Snow (Earth Sciences Department) in the Prof. S. Mohammed Proteomics Facility (Digital Appendix 7.4).

Data Set	UniProt ID*	MC-Initiating Homologs in Other Plants	Annotation	Unique Peptides Nuclear	Unique Peptides WholeCell	Unique Peptides 20.6% versus 15% O_2	Average Number of Unique Peptides per Measurement (15% O_2 Proteome, n=3)	Average Number of Unique Peptides per Measurement (20% O_2 Proteome n=3)
MC-Initiating Proteins Unique to the Nuclear vs Whole Cell Extraction Proteomic Analysis	R1CZH8;R1DZK9	No	Signal recognition particle 9 kDa protein	2;2	0	-	-	-
	R1FFP7	No	Uncharacterized protein	0	2	-	-	-
	R1F057	No	Uncharacterized protein	0	1	-	-	-
	R1E8J2	No	Uncharacterized protein	0	1	-	-	-
	R1DG38	No	Uncharacterized protein	0	4	-	-	-
	R1CCR2	No	Uncharacterized protein	0	1	-	-	-
	R1BUY0;R1CHC8	No	Uncharacterized protein	0	2;2	-	-	-
	R1BQQ8	No	Uncharacterized protein	0	1	-	-	-
	R1BLG2	No	ANK_REP_REGION domain-containing protein	0	1	-	-	-
	A0A0D3KSU5	No	Uncharacterized protein	0	1	-	-	-
	R1BFN3;R1FQK5	No	GMP synthase (glutamine-hydrolyzing)	1	4;4	14;12	4.0	4.0
MC-Initiating Proteins Shared by Nuclear vs Whole Cell Extraction and 20.6 vs 15% O_2 Proteomic Analysis	R1BRP2;R1BJ30	No	cobW domain-containing protein	0	2;2	2;2	0.3	1.0
	R1BL17	No	PDZ domain-containing protein	3	3	4	0.3	0.0
	R1CKT7;R1BM71	No	NAD(P)-bd_dom domain-containing protein	3	8;8	8;8	2.3	1.3
	R1C2J9	No	Uncharacterized protein	1	2	2	1.0	0.7
	R1DPC7;R1DF86	No	Metalloprotease 1	0	2;1	1;1	1.3	2.3
	R1E3D7	No	Uncharacterized protein	2	3	3	0.0	0.3
	R1EIN3	No	Uncharacterized protein	5	10	5	9.0	5.0
MC-Initiating Proteins Unique to the 20.6 vs 15% O_2 Proteomic Analysis	R1FFU5;R1FWR2	No	Inosine-5'-monophosphate dehydrogenase	0	2;1	9;1	1.3	1.0
	Q4G363	No	50S ribosomal protein L23, chloroplastic	-	-	5	1.0	3.0
	R1DZI7	No	Uncharacterized protein	-	-	3	5.3	5.7
	R1E297	No	AIR carboxylase	-	-	1	1.7	1.7
	R1EC69;R1E393	Yes	Amidophosphoribosyltransferase	-	-	6;6	0.3	0.3
	R1EJR3	No	AIR carboxylase	-	-	1	1.3	1.3
	R1EYS6;R1CKD9	No	Phosphoribosylaminoimidazolesuccinocarboxamide synthase	-	-	6;3	1.0	2.0
	R1FCY0	No	Uncharacterized protein	-	-	1	5.7	5.7

Table 7.1: UniProt identifiers of MC-initiating proteins identified in the *E. huxleyi* proteome from the nuclear vs whole cell extraction proteomic data and the 20.6 versus 15% O_2 proteomic data. The two proteomic data sets overlapped, sharing identified proteins. Only one protein identified had MC-initiating homologs in other organisms. Where two UniProt IDs are used, it means the results yielded enough peptides to know the protein is one of the two identified proteins, but not enough to know which one.

Both whole cell and nuclear enriched extractions satisfied the requirements for reliable proteomic analysis (Appendix 7.5). Where two UniProt IDs are used, it means the results yielded enough peptides to know the protein is one of the two identified proteins, but not enough to discriminate between the two possibilities.

Results showed a significant overlap between nuclear enriched and whole cell protein extractions, with only approximately 100 unique nuclear proteins observed (Digital Appendix 7.4 and Fig. 7.4). A single MC-initiating protein was identified from the 100 unique nuclear proteins: a signal recognition particle 9 kDa protein (UniProt ID: R1DZK9). From its homology it was inferred this protein is likely a ribonucleoprotein involved in the negative regulation of peptide elongation. An NCBI BLASTp analysis indicated the MC-initiating N-terminus of the R1DZK9 protein is not conserved amongst homologous proteins in different organisms. This suggests the MC-initiating N-terminus of R1DZK9 is not evolutionarily relevant in terms of PCO regulation. Eighteen MC-initiating proteins were found in the whole-cell protein extraction, the majority of which were uncharacterised proteins (Table 7.1). An NCBI Blastp analysis of the whole-cell MC-initiating proteins indicated the MC N-termini of these proteins are not conserved amongst other homologous proteins from other organisms. Nevertheless, the greater yield of MC proteins validated the use of a whole cell protein extraction method for the subsequent proteomic analysis of the *E. huxleyi* cells grown under 15 and 20.6 % O_2.

7.2.5 Proteomics of *E. huxleyi* cells grown at 15 and 20.6 % O_2

7.2.5.1 Trialling whole-cell protein extraction protocols

Subsequent proteomic analysis of algal cultures would focus on whole-cell protein extractions, therefore the efficiency of suitable whole-cell protein extraction methods was examined, namely P-PER plant protein extraction kit and sonication with 0.1% triton X-100 in TEAB and Tris buffers. The P-PER kit relies on an organic solvent and mechanical cell lysis to extract the plant protein. Sonication in 0.1% triton X-100 in TEAB and Tris buffers are a more classic protein extraction technique and were performed as outlined in Chapter 2, Section 2.6.3. Following protein extraction using these methods, protein was quantified using a standard BCA assay for the extractions *via* triton X-100 and sonication. A BCA reducing agent compatible kit was used to quantify the P-PER plant kit extracted protein. The P-PER plant kit extracted the greatest amount of protein from the *E. huxleyi* cells (146.4 ± 53.6 µg/mL), compared to intense sonication in 0.1 % triton X-100 (89.1 ± 45.0 µg/mL, Appendix 7.6). Despite the higher protein yields using the P-PER kit, the proteomics facility requested the 0.1 % triton X-100 sonication extraction as it was thought the P-PER reagents might interfere with proteomic analysis. The yield of protein using intense sonication in 0.1 % triton X-100 was considered acceptable without further optimisation.

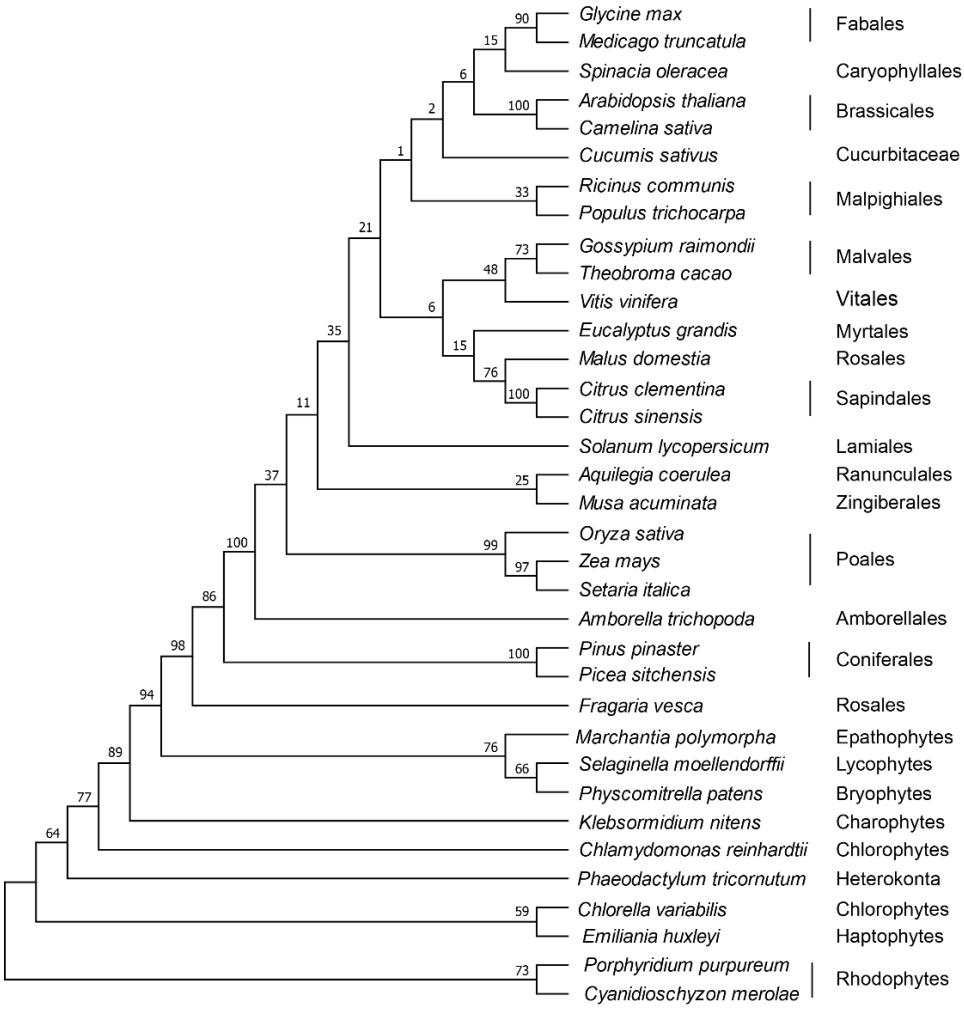
7.2.5.2 The proteome of E. huxleyi grown under 15 and 20.6 % O_2

Analysis of the proteome of *E. huxleyi* cells grown under 15 and 20.6 % O_2 identified fifteen MC-initiating proteins, eight of which were previously

observed in the whole cell extraction (Table 7.1). Seven of the MC-initiating proteins were unique to the 15 and 20.6 % proteomic analysis. That is, 7 MC-initiating proteins observed here were not previously observed in the nuclear versus whole cell extraction proteomic data. Further culturing is required to assess the regulation of these proteins with respect to O_2 availability, as a 5 and 10% t-test of the identified proteins in the 15 and 20.6 % O_2 proteomes, suggested the expression levels of these proteins did not change significantly between 15 and 20.6 % O_2 (provided by Dr, J. Snow, Earth Sciences, Digital Appendix 7.7).

Of all the MC-initiating proteins analysed using the NCBI BLASTp tool, only two had homologous proteins with MC-initiating proteins in other organisms (UniProt IDs: R1EC69 and R1E393). Using the R1EC69 and R1E393 protein sequences as templates, NCBI Blastp analyses identified two families of MC-initiating proteins with homology to R1EC69 and R1E393 in other plants, animals, fungi, bacteria and archaea: asparagine synthetases (AS, Fig. 7.5) and glutamine-fructose-6-phosphate transaminases (GFAT, Fig. 7.6). Both the Asn and GFAT families exhibit a fully conserved MCGI N-terminal sequence in all sequences analysed. This is significant as it suggests the MCGI N-terminus is biologically relevant. The biological relevance of this finding requires further investigation.

A) Phylogenetic tree of plant asparagine synthetases



B) N-termini of plant asprargine synthetases

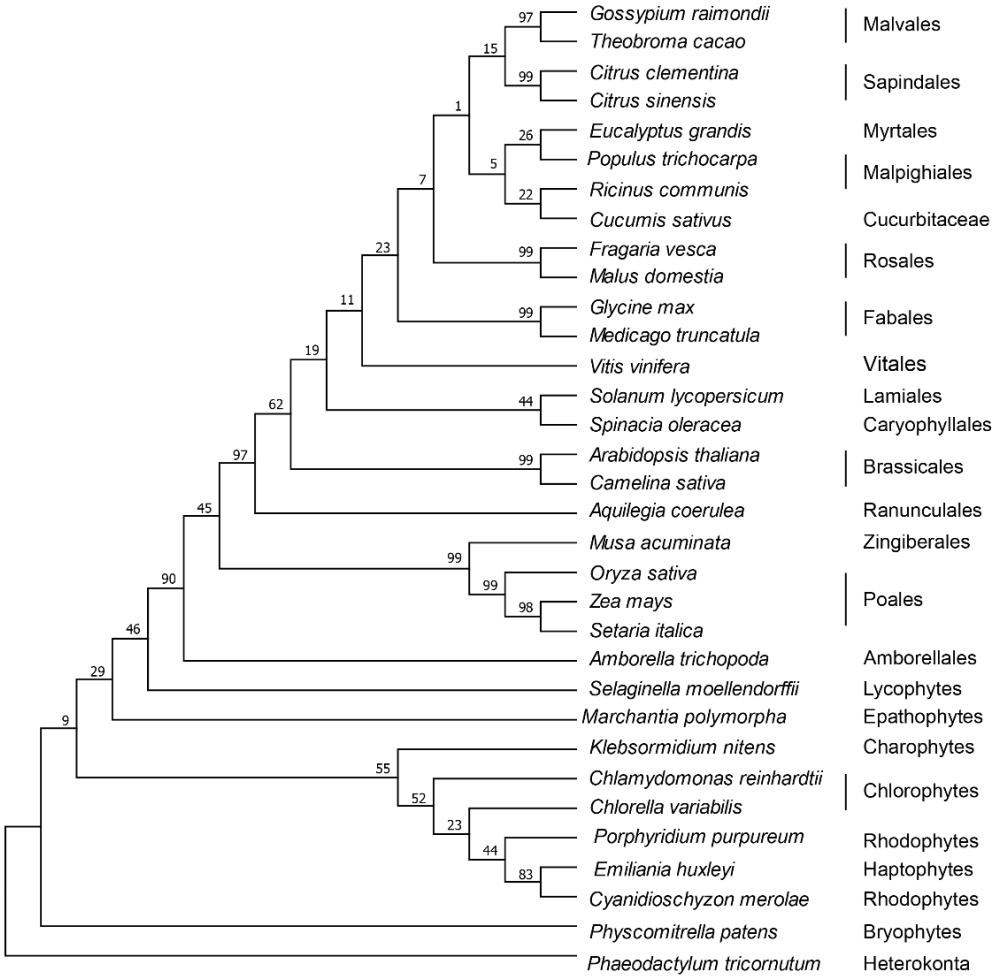


C) N-termini of asparagine synthetases from across the tree of life

```
Arabidopsis thaliana                MCGILAVLGCVDNSQ
Chlamydomonas reinhardtii           MCGILAVLNTTDDSQ
Homo sapiens                        MCGIWALFGSDDCLS
Candidatus Woesearchaeota archaeon  MCGIIGIFEFKNRSK
Escherichia coli                    MCSIFGVLDIKTDAV
Saccharomyces cerevisiae            MCGIFAAFKHEDIHN
Drosophila melanogaster             MCGIFAIFSRDGEPI
Danio rerio                         MCGIWALFGSDECLS
                                    **.* . :
```

Fig. 7.5: Conservation of asparagine synthetases across the plant kingdom. Two proteins were identified in the 20.6 versus 15% O_2 *E. huxleyi* proteomic data that had homologous MC-initiating proteins in other organisms. A) The sequence is strongly conserved in plants B) the N-termini have a conserved MCGI N-terminal motif and C) MC-initiating homologs were found in plants, animals, fungi, bacteria and archaea.

A) Phylogenetic tree of plant glutamine-fructose-6-phosphate transaminases



B) N-termini of plant glutamine-fructose-6-phosphate transaminases



C) N-termini of glutamine-fructose-6-phosphate transaminases from across the tree of life

```
Arabidopsis thaliana          MCGIFAYLNFHANKE
Chlamydomonas reinhardtii     MCGIVGYYTFNVQRD
Homo sapiens                  MCGIFAYMNYRVPRT
Thermoplasmatales archaeon    MCGIVGYIGKRIAVP
Escherichia coli              MCGIVGAIAQRDVAE
Saccharomyces cerevisiae      MCGIFGYCNYLVERS
Drosophila melanogaster       MCGIFAYLNYLTPKS
Danio rerio                   MCGIFAYLNYHVPHT
                              ****..
```

Fig. 7.6: Conservation of glutamine-fructose-6-phosphate transaminases across the plant kingdom. Two proteins were identified in the 20.6 versus 15% O_2 *E. huxleyi* proteomic data that had homologous MC-initiating proteins in other organisms. A) The sequence is strongly conserved in plants B) the N-termini have a conserved MCGI N-terminal motif and C) MC-initiating homologs were found in plants, animals, fungi, bacteria and archaea.



Fig. 7.7: *K. nitens* (NIES 2285). *K. nitens* grown in NIES C media at 15 °C and imaged on a Zeiss Axioscope.A1 Light Microscope at x630 magnification

7.4.1 Introduction

This section discusses the culturing of *K. nitens* (NIES 2285, Fig. 7.7) under different O_2 concentrations. Initial experiments focussed on establishing the appropriate media for optimum cell growth, before attempting cultures at low O_2.

7.4.2 Initial culturing of *K. nitens* cells

The following *K. nitens* cultures were grown in the Earth Sciences Geobiology algal culturing facility using a Baker-Ruskinn PhO_2xBox for atmospheric control (Chapter 2, Section 2.6.4). The cultures were grown in the atmospherically controlled Perspex box in a 15 °C incubator. *K. nitens* cells were subcultured from a culture collection NIES 2285 stock in HSMA media and grown as batch cultures, under a 12 hour light/dark cycle in NIES C media (light illuminance of

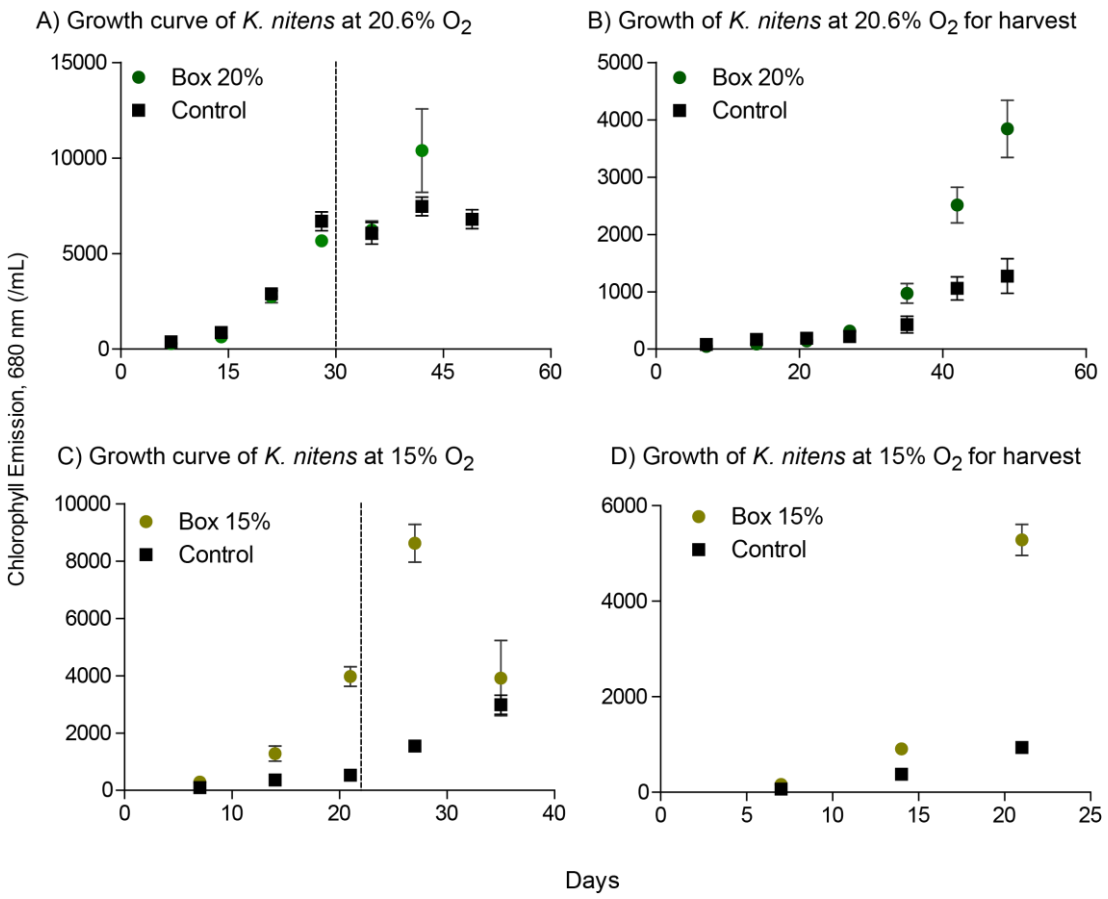


Fig. 7.8: *K. nitens* growth curve in NIES C media at 15 °C in the PhO_2xBox atmosphere control unit in the Rickaby Geobiology Lab. *K. nitens* stationary culture grown in C media at 15 °C. Growth determined by chlorophyll emission at 680 nm on the Tecan Spark® microplate reader. A) *K. nitens* growth curve at 20.6% O_2 B) a second culture at 20.6% O_2 was harvested at mid-exponential phase. C) Growth curve at 15% O_2 D) a second culture at 15% O_2 was harvested at mid-exponential phase.

2700 lx). Growth was monitored using chlorophyll emission measurements at 680 nm. To measure chlorophyll fluorescence, the cells were lysed using a 1:4 sample to DMSO ratio. After 90 min, the cells were centrifuged for 30 min at 4,000 x g to pellet any cell debris. Chlorophyll fluorescence emission was measured using the Tecan Spark® microplate reader. The coulter counter could not be used to count *K. nitens* cells as the cells form filaments. The cells were harvested by centrifugation at 4,000 x g. The supernatant was discarded and the cells flash frozen in liquid nitrogen. Cells were stored at -80 °C for

proteomic analysis. Cultures grown at 20.6 % O_2 reached exponential phase after 21 days (Fig 7.8, A). A second set of cultures were set up at this O_2 concentration and harvested once they reached exponential phase (Fig 7.8, B). This process was repeated at 15% O_2 (Fig. 7.8 C and D).

7.4 Discussion

This chapter outlines the necessary conditions for *E. huxleyi* and *K. nitens* culturing under low O_2. Preliminary proteomic analysis of *E. huxleyi* established whole cell protein and nuclear enriched protein extraction methods had significant protein ID overlap. However, a distinct loss of MC-initiating protein IDs in the nuclear enriched sample justified the use of the whole cell extraction method. In all data sets analysed, only one group of the identified MC-initiating proteins in the *E. huxleyi* proteome had MC-initiating homologs in other organisms. This protein group was homologous to the highly conserved AS and GFAT families, both of which have conserved MCGI N-termini and can be found across the tree of life.

AS proteins catalyse an ATP-dependent transfer of ammonia to aspartate yielding asparagine in the presence of a magnesium ion. Two families of AS proteins exist, AS-A and AS-B. AS-A uses ammonia exclusively as an amide donor to produce asparagine from aspartate and ATP in prokaryotes [279, 280]. By contrast AS-B can use either ammonia or glutamine as a nitrogen donor. Due to the large size of AS-B, and its ability to use either glutamine or ammonia as the nitrogen donor [280], it has been speculated that AS-B was formed by the

fusion of an ancestral glutamine amidotransferase gene with an ancestral ammonia-dependent asparagine synthetase gene, homologous to AS-A [281].

An ASN gene in Arabidopsis was shown to be in control of nitrogen assimilation and export during vegetative growth. Arabidopsis ASN knockout lines displayed a visible delayed leaf senescence phenotype [282]. Not much is understood of GFAT plant function, but in Arabidopsis it is thought to play a role in UDP-N-acetylglucosamine biosynthesis and a decrease in AtGFAT1 transcripts impaired vegetative and reproductive development [283]. Any relationship between these functions and O_2 regulation is unknown.

To understand if these MC-initiating proteins have the potential to be substrates of the PCOs, peptide oxidation assays must be performed. Asn and GFAT peptide oxidation with HsADO, the human O_2 sensitive enzyme with homology to the PCOs, has been reported upon in the literature [150]. Results showed that HsADO did not catalyse the conversion of N-terminal Cys to $CysO_2H$ on the $HsAS_{2-15}$ peptide [150]. Therefore the role of the MC N-terminus of the AS in humans is not known. Interestingly, HsADO did catalyse the oxidation of the $HsGFAT_{2-15}$ peptide [150]. Despite the AS and GFAT peptides identical amino acids in the 1-4 positions, HsADO has a preference for GFAT over AS. The PCO-catalysed conversion of N-terminal Cys to $CysO_2H$ on the AS and GFAT peptides has not yet been assessed and could be an interesting area for future investigation of ancient PCO function. Further proteomic analysis will

have to be carried out to understand how any of the MC-initiating proteins identified here are regulated with respect to O_2 in *E. huxleyi* cells.

Chapter 8: Conclusions and future direction

It has been demonstrated both *in vivo* and biochemically that the higher plant hypoxia response is controlled by the N-degron protein degradation pathway [34, 133, 142]. Under normal O_2 levels, PCOs catalyse the oxidation of N-terminal Cys residues to $CysO_2H$ on the ERF-VII transcriptions factors (as well as ZPR2 and VRN2) [24-26, 133]. This modification targets the ERF-VIIs for arginylation [142], ubiquitination and subsequent proteomic degradation. However, when O_2 is limiting, the ERF-VIIs stabilise and regulate a number of genes responsible for mediating plant development and the acute hypoxia response [70, 79, 82-84, 87, 112]. The PCO enzymes are highly conserved throughout the plant kingdom, with homologous N-terminal Cys oxidases even found in animals (ADO). However, the ERF-VII and other known MCGGAI-initiating PCO substrates are confined to the flowering plants, with no homologous ERF sequences found in algae. Without homologous substrates, this greatly limits our understanding of PCO function in algae. This thesis aimed to understand algal and early plant PCO function and what this might tell us about how the plant O_2 sensing pathway evolved.

Chapter 3 describes the expression, purification and biophysical characterisation of MpPCO, a single PCO isoform from the proteome of the liverwort, *M. polymorpha*. It also describes an ERF-VII homolog from the predicted *M. polymorpha* proteome ($MpERF_{2\text{-}15}$). MpPCO was found to have a low K_M for the $AtRAP2.12_{2\text{-}15}$ peptide, a contrast with the high K_M for the

$MpERF_{2\text{-}15}$ peptide. The differences between the local chemical environments of the N-terminal Cys on the $MpERF_{2\text{-}15}$ and $AtRAP2.12_{2\text{-}15}$ peptides may be responsible for the different binding affinities of the peptides with MpPCO. The $MpERF_{2\text{-}15}$ N-terminus is bulky (CysArgMetAsn), charged (Arg) and polar (Asn), a contrast to the sterically unhindered, apolar CysGlyGlyAla N-terminus of $AtRAP2.12_{2\text{-}15}$. The high K_M of MpPCO with the MpERF peptide is a characteristic shared with other PCOs, with some K_M values for the AtPCOs with the $AtRAP2.12_{2\text{-}15}$ peptide in a similar range. This apparent poor substrate affinity may be an artefact of using peptides over full-length proteins, or perhaps low substrate binding affinity is biologically relevant. Both the MpPCO/$AtRAP2.12_{2\text{-}15}$ and MpPCO/$MpERF_{2\text{-}15}$ reactions are relatively less sensitive to changes in O_2 concentrations, with the $K_M^{app}\ O_2$ values determined to be 3.5 ± 0.6 % and 8.4 ± 1.7% respectively. These $K_M^{app}\ O_2$ values are lower compared to the AtPCO 4/$AtRAP2.12_{2\text{-}15}$ reaction, which was determined to have a K_M^{app} O_2 of 17.3 ± 1.7 % (AtPCO 4 is the isoform believed to be the most important in Arabidopsis). The insensitivity of MpPCO to changes in O_2 may be reflective of an alternative function, one which does not require a great deal of sensitivity to O_2.

PCO function in early plants is still poorly understood. The Arabidopsis PCOs have K_M O_2 values in physiologically relevant ranges (AtPCO 1-5, $K_M^{app}\ O_2$ values: 15.7 ± 1.9, 7.37 ± 0.7, 11.1 ± 1, 17.3 ± 1.7 and 5.45 ± 0.6 % O_2 respectively [188]), with AtPCO function linked to enhanced plant hypoxia tolerance [133]. The link between *M. polymorpha* hypoxia tolerance and PCO

function has yet to be determined. It is possible that PCO O_2 sensitivity in early plants cannot be accurately used as proxy for plant hypoxia tolerance. Furthermore, MpPCO sensitivity to changes in O_2 has been determined using only the $MpERF_{2\text{-}15}$ and exogenous $AtRAP2.12_{2\text{-}15}$ substrates. It is possible O_2 sensitivity may vary with substrate, and that MpPCO has other, not yet identified endogenous substrates. As such these results may not be representative of biological MpPCO O_2 sensitivity. Furthermore, peptide activity is not a substitute for full length proteins. *In vivo* hypoxia experiments [275] could not only establish if MpERF is a true substrate of MpPCO, but also if it stabilises in the *M. polymorpha* proteome under low O_2. The fossil record suggests the first land plants were morphologically classified as liverworts [226, 227]. The presence of a functional MpPCO, plus ERF-VII, PRT6 and ATE homologs (essential proteins in the higher plant N-degron pathway) suggests the molecular components of the hypoxia response pathway are important, even for early land plants.

This is especially convincing in light of the results described in Chapter 4. Here, the expression, purification and characterisation of the first algal PCO, KnPCO, is described. KnPCO is the most O_2 sensitive PCO ever characterised. The determined $K_M^{app}O_2$ values were 26 ± 2.6 % O_2 and 28.9 ± 4.1 % O_2 for the His_6-KnPCO/$MpERF_{2\text{-}15}$ the His_6-KnPCO/$AtRAP2.12_{2\text{-}15}$ reactions respectively. HsADO, however, is more O_2 sensitive with a $K_M^{app}O_2$ of 55.1 ± 6.6 % with the $RGS4_{2\text{-}15}$ peptide [150]. KnPCO is also the first PCO to demonstrate significant substrate selectivity in that it does not oxidise $AtZPR2_{2\text{-}15}$ at all. Structural and

site-directed mutagenesis studies of KnPCO could enable a better understanding of the molecular mechanism of PCO O_2 sensing and the factors that influence substrate selectivity. The existence of an O_2 sensitive PCO in an alga is exciting, especially in light of the PRT 6 and ATE homologs present in the *K. nitens* genome. This implies the existence of a highly O_2 sensitive PCO-regulated protein degradation pathway in *K. nitens*. As no ERF-VII homologs are found in the *K. nitens* genome, understanding the role of such a pathway is difficult. The identification of MC-initiating proteins in algal cell lysates or using bioinformatics may help elucidate alternative PCO substrates (Chapter 6). Further algal culturing and proteomics of *K. nitens* cells grown under hypoxia could also shed light on the role of KnPCO in *K. nitens* and on the algal hypoxia response as a whole (Chapter 7).

Recombinant KnPCO expression was straightforward, a characteristic shared with most other land plant-like PCOs. However, the expression of most algal-like PCOs was difficult, with many proving insoluble or failing to express outright. While many of the expression trials had limited success, there remains a number of promising avenues by which more algal-like PCO proteins could be produced. Analysis suggested His_6-PtPCO and His_6-OtPCO 1 and 2 were successfully expressed, but were confined to inclusion bodies. These proteins could be promising candidates for solubility-tag expression trials. Expressing a more algal-like PCO could help us better understand algal PCO behaviour such as O_2 sensitivity and substrate selectivity.

Structural analyses of KnPCO and MpPCO may shed some light on KnPCO substrate selectivity relative to MpPCO and the AtPCOs. Although attempts to crystallise these enzymes were unsuccessful, CD and modelling studies suggest both proteins likely retain the Cupin family DSBH fold. The active sites of both KnPCO and MpPCO also closely match that of AtPCO 5 with the metal-binding His facial triad and conserved Asp residues in close proximity to the metal centre. A GlyArgPhe rich region between the N-terminal helices and the DSBH was found in MpPCO and KnPCO, but not in the higher plant PCOs. These regions are in close proximity to the active site and may influence substrate binding. As the active sites of the KnPCO and MpPCO models are very similar to that of the AtPCO 5 crystal structure, it may be that the observed amino acid substitutions to the GlyArgPhe-rich region might account for KnPCO substrate selectivity compared to MpPCO. Site-directed mutagenesis studies could shed light on the role of this region. Mutating the KnPCO GlyArgPhe-rich region (ETRGFGFFGRQGPWGRH) to that of MpPCO (QERGFGFFGTNGRRGRH) would be a logical next step in trying to understand KnPCO substrate selectivity.

Our knowledge of PCO substrate scope is narrow; limited thus far to proteins homologous to the higher plant ERF-VII and ZPR2 transcription factors and VRN2. This limits our understanding of PCO function in algae, which contain no ERF, ZPR2, or MC-initiating VRN2-related proteins in their genomes. This thesis reports the use of a maleimide-reporter probe for the identification of PCO substrates in cell lysates, and provides the first reported experimental link between the higher plant O_2 sensing pathway and the algal hypoxia response.

C. reinhardtii cells were exposed to hypoxia for 6 hours. The hypoxia-exposed cells and a normoxia control were incubated with recombinant KnPCO to convert N-terminal Cys to $CysO_2H$, and subsequently labelled using a $CysO_2H$ fluorescent probe. SDS-PAGE analysis suggested there was a higher proportion of $CysO_2H$ in the hypoxia band than in the normoxia (3:1). Proteomic analysis of the *C. reinhardtii*$_{Normoxia}$ and *C. reinhardtii*$_{Hypoxia}$ SDS-PAGE fluorescent gel bands identified three MC-initiating proteins. Proteomic data suggest the intense fluorescent band in the hypoxia sample is CrDES; an acyl-lipid omega-13 desaturase. This is in keeping with studies in higher plant species, which suggest lipid desaturation is upregulated under hypoxia [284-286]. The proteomic data also identified; CrHAP2, a gamete fusion protein and ODA7, a leucine-rich repeat-containing protein, implicated in flagella movement [271]. Blastp searches using CrDES as a template showed that CrDES has an MC-initiating homolog in *V. carteri* (VcCE), but not in *K. nitens* or *M. polymorpha*. This suggests that the MC-Nterminus of CrDES could be biologically relevant for both *C. reinhardtii* and *V. carteri*. Assays with N-terminal peptides of the MC-proteins from *C. reinhardtii* and the *V. carteri* CrDES homolog confirmed the peptides as substrates of the PCOs. A biological repeat of the *C. reinhardtii*$_{Normoxia}$ and *C. reinhardtii*$_{Hypoxia}$ labelling experiment generated an upregulated hypoxia fluorescent band of similar relative molecular mass (proteomic analysis pending).

Further experiments are required to verify these proteins as true substrates using the endogenous CrPCO. Recombinant CrPCO has proven difficult to

express; a synthetic *in vitro* system such as those described by Licausi *et al* [275], could determine if the full length proteins are indeed substrates of the endogenous CrPCO. The results in Chapter 6 would suggest an algal metabolic response to hypoxia which appears to involve MC-initiating N-terminal proteins. This is especially interesting in light of the presence of putative PCO, ATE and PRT6 homologs in the *C. reinhardtii* genome. The same hypoxia exposure and $CysO_2H$ labelling procedure was repeated using *K. nitens* cells. A similar upregulated fluorescent band was seen in the *K. nitens*$_{Hypoxia}$ lane compared to *K. nitens*$_{Normoxia}$ (3:1, proteomic analysis pending). It will be of significant interest to see if similar proteins are identified in the proteomic analysis of the *K. nitens* oxidised Cys proteins (Chapter 6).

Chapter 7 outlines the necessary conditions for culturing *E. huxleyi* and *K. nitens* under low O_2 for analysis of their global proteomic changes at different O_2 concentrations. Preliminary proteomic analysis of *E. huxleyi* identified one MC-initiating protein group that had MC-initiating homologs in other organisms. This protein group was homologous to the highly conserved asparagine synthetase (AS) and glucosamine—fructose-6-phosphate aminotransferase (GFAT) families, both of which have conserved MCGI N-termini and can be found across the tree of life. Activity of the PCOs towards the AS and GFAT peptides is not known. However, the human O_2 sensitive N-terminal Cys dioxygenase with homology to the PCOs, HsADO, does not catalyse the oxidation of the $HsAS_{2-15}$ peptide, but has some activity towards the $HsGFAT_{2-15}$ peptide [150]. Further proteomic analysis is required to

understand how any of the MC-initiating proteins identified here are regulated with respect to O_2 in *E. huxleyi* cells. Unfortunately, due to issues with culturing facilities and limited access to seawater as a result of the COVID-19 pandemic, the growth and proteomics of *E. huxleyi* and *K. nitens* cells grown under low O_2 did not progress beyond preliminary analyses.

Overall this thesis characterises the first PCO from early plants with a putative endogenous ERF-homolog. This thesis also describes characterisation of the first recombinant PCO from algae, which also happens to be the most O_2 sensitive PCO characterised thus far and the first PCO to demonstrate marked substrate selectivity. The first putative algal PCO substrates are also described, which preliminary data suggests are upregulated under hypoxia. Hypoxia-upregulated proteins identified using established algal culturing techniques and whole-cell proteomics may provide future avenues for the identification of novel substrates in algae. Results presented here suggest PCO catalysed oxidation of N-terminal Cys to $CysO_2H$ is biologically relevant in algae and early plants independent of ERF-VII and development-related substrates. While the necessary components for a functional N-degron pathway are present in *K. nitens* and *M. polymorpha*, the *E. huxleyi* genome only contains PCO-like sequences, with ATE and PRT6 sequences notably absent. This alludes to another function for the PCOs in other algae.

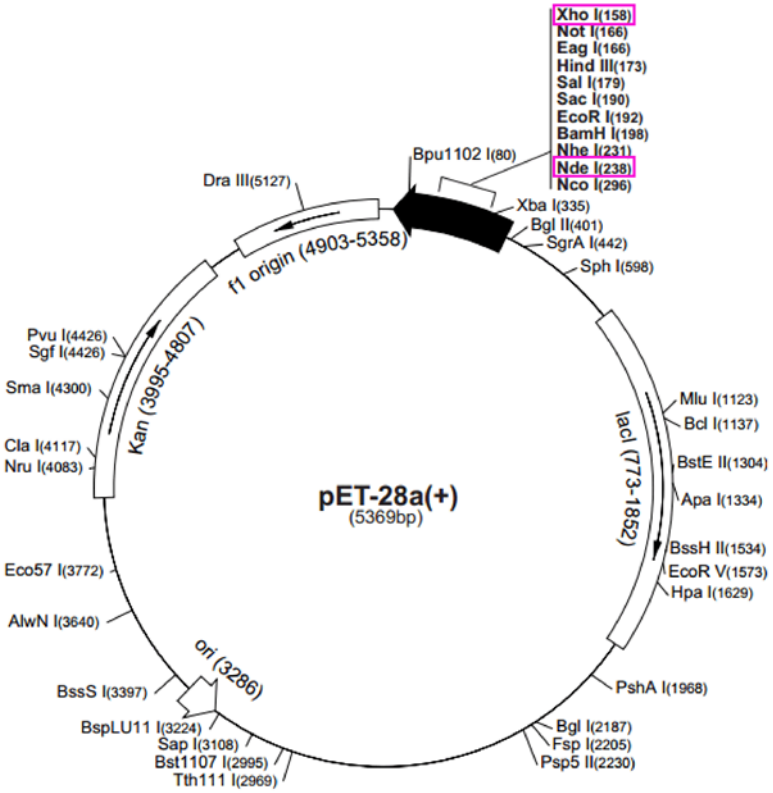
The results presented here suggest that the PCO-mediated Cys/Arg branch of the N-degron pathway is biologically relevant for *M. polymorpha*, *K. nitens* and

C. reinhardtii, pre-dating the emergence of ERF-VII transcription factors (N-termini: MCGGAI). PCO substrate hunting experiments suggest algae experience a response to hypoxia-induced stress, appearing to upregulate MC-initiating proteins under low O_2. The identified proteins in the *C. reinhardtii* proteome seem not to be involved directly with gene regulation, but potentially link algal hypoxia with lipid synthesis, gamete fusion and flagella propagation. This could mean that in simple organisms such as algae, PCOs, rather than regulating the genetic response to hypoxia stress, instead directly regulate the expression levels of proteins important for adaptation to changes in O_2, such as metabolism and movement. Further proteomic and transcriptomic analysis is required to fully understand the nature of PCO function in algae, and how the enzymes contribute to adaptation and survival under low O_2.

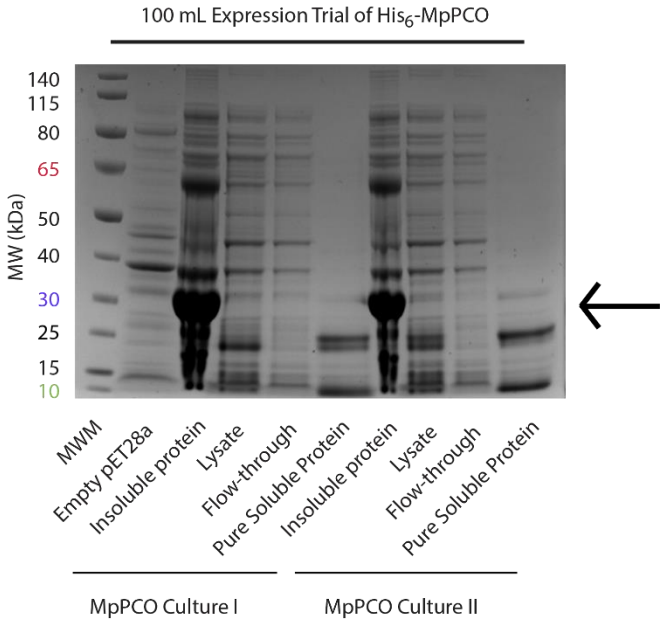
Appendix



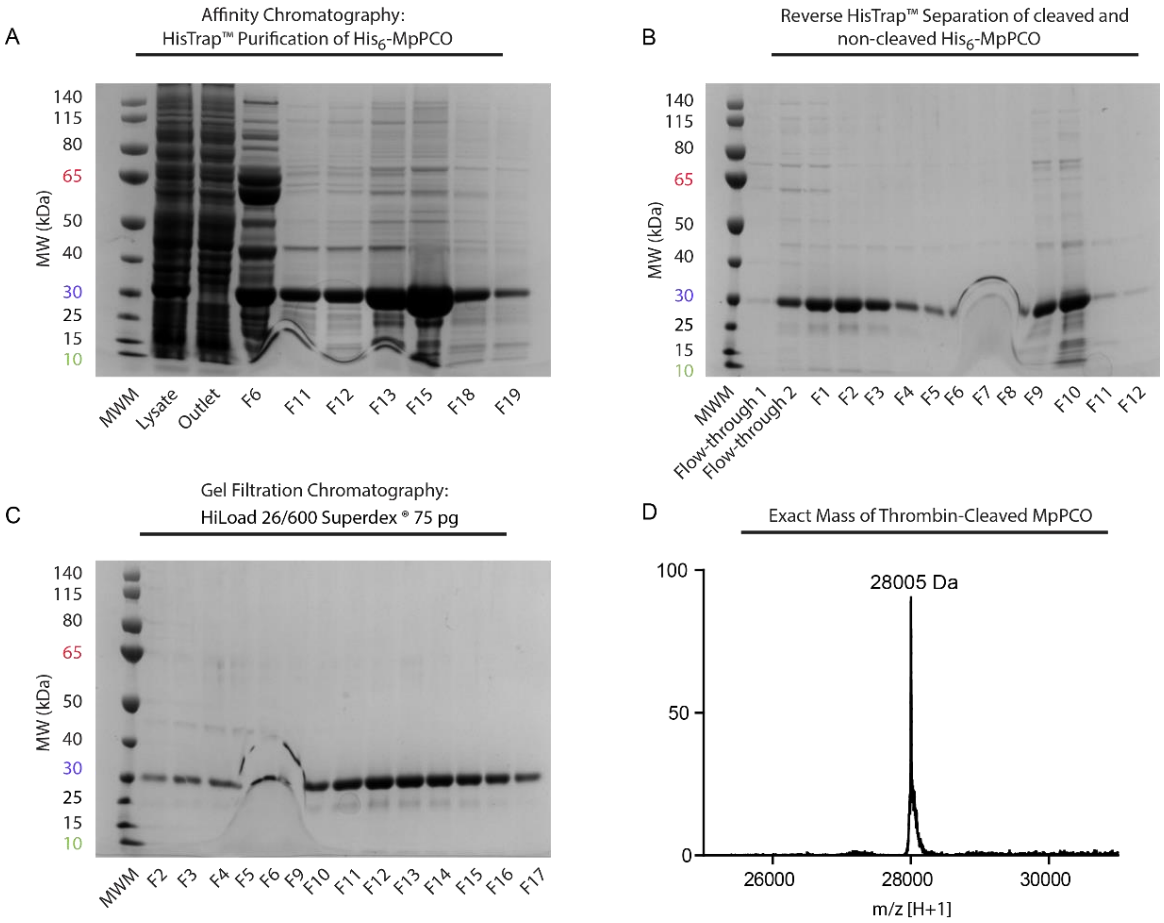
A. 1.1: The description of planktonic species associated with the retrieved PCO and ADO-like redundant sequences. Krona chart of the distribution of redundant PCO and ADO-like sequences from the MATOUv1+T and OM-RGC.v2+T metatranscriptomes.



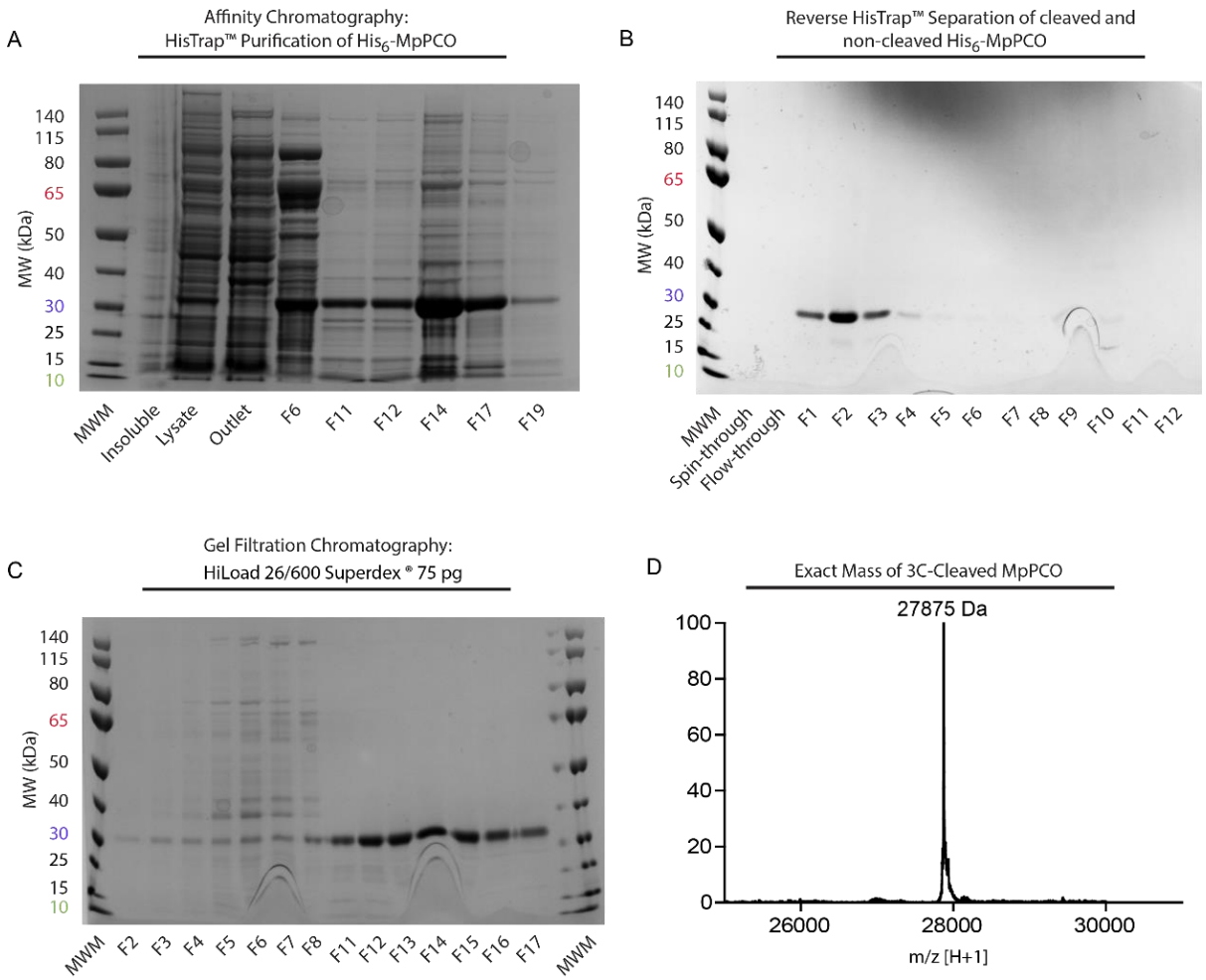
A 3.1: Map of the pET28a(+) bacterial vector for the expression of N-terminally His_6-tagged proteins with a thrombin cleavage site. The sites at which the plasmid was linearised are highlighted. The amplified DNA fragment encoding for MpPCO was inserted here using the In-Fusion® HD Enzyme.



A 3.2: Small scale expression of His_6-MpPCO. *E. coli* BL21(DE3) cells were transformed according to the standard operating procedure outlined in Chapter 2, Section 2.1.3. Two colonies were selected and grown overnight in the presence of 30 µg/mL kanamycin. These starter cultures were used to inoculate fresh 2TY media containing 30 µg/mL kanamycin and 100 µM $(NH_4)_2Fe(SO_4)_2(H_2O)_6$. Once OD_{600} = ~0.8 protein expression was induced with 0.5 mM IPTG and grown overnight at 18 °C (SOP: Chapter 2, Section 2.1.4.1). OD_{600} after overnight growth was ~1, indicating the cells failed to grow/and or they died. The cells were spun-down, and re-suspended in HisTrap™ binding buffer and DNAse I before sonication, spin-down and lysate filtration. The lysate was manually loaded onto a 1 mL HisTrap™ column, collecting the flow-through. The column was washed with 5 column volumes of binding buffer, before 3 column volumes of elution buffer. The proposed MW of His_6-MpPCO is 29752 kDa, so a band is expected at ~30 kDa. The results indicate a significant quantity of insoluble protein, but no significant soluble bands are observed.



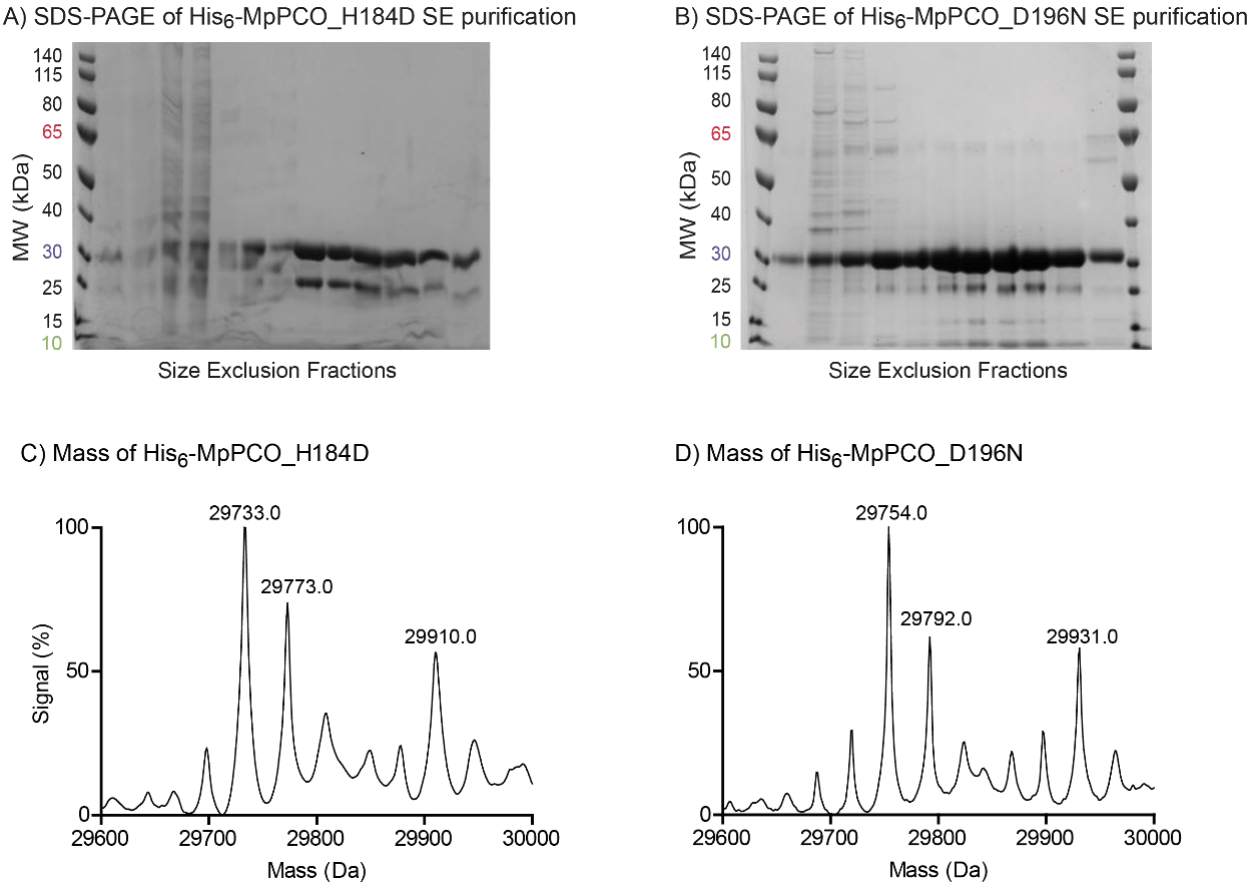
A 3.3.1: Purification of His_6-cleaved MpPCO using thrombin protease. A) Soluble His_6-MpPCO was purified using HisTrap™ HP affinity chromatography, eluting over a 0-100% gradient of elution buffer. His_6-MpPCO containing fractions were identified using SDS PAGE. Protein-containing fractions were pooled and thrombin protease added to ~0.3 U/mg of purified enzyme. The solution was dialysed overnight into HisTrap™ binding buffer (low imidazole) at 4˚C. **B)** After overnight incubation the protein/protease solution was loaded onto a 5 mL HisTrap™ column, collecting the flow-through. The column was washed with 8 column volumes (CV) of binding buffer and collected as fractions 1-8. The column was then washed with 4 CV of elution buffer, collected as fractions 1-4. Thrombin-cleaved-MpPCO-containing fractions were identified using SDS PAGE. **C)** Fractions 1-8 were concentrated in a MWCO 10 kDa concentrator tube before injecting onto the HiLoad 26/600 Superdex ® 75 pg gel filtration column. Fractions containing MpPCO were identified using SDS PAGE, concentrated, aliquoted, flash frozen in $N_2(l)$ and stored at -80 °C.



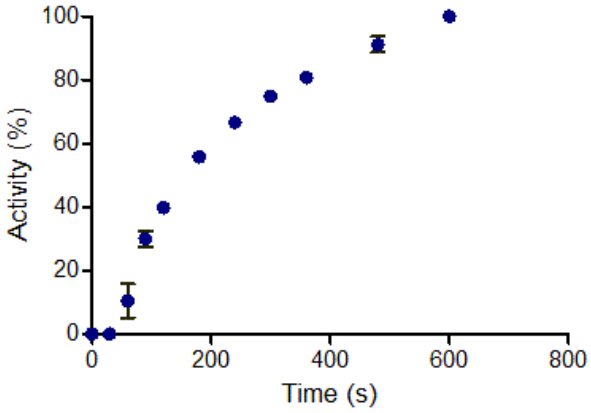
A 3.3.2: Purification of His_6-cleaved MpPCO using 3C protease. A) Soluble His_6-MpPCO was purified using HisTrap™ HP affinity chromatography, eluting over a 0-100% gradient of elution buffer. His_6-MpPCO containing fractions were identified using SDS PAGE. Protein-containing fractions were pooled and 50 µL of a 50 mg/mL 3C protease stock was added to the purified enzyme. The solution was dialysed overnight into HisTrap™ binding buffer (low imidazole) at 4˚C. **B)** After overnight incubation the protein/protease solution was loaded onto a 5 mL HisTrap™ column, collecting the flow-through. The column was washed with 8 column volumes (CV) of binding buffer and collected as fractions 1-8. The column was then washed with 4 CV of elution buffer, collected as fractions 1-4. 3C-cleaved-MpPCO-containing fractions were identified using SDS PAGE. **C)** Fractions 1-8 were concentrated in a MWCO 10 kDa concentrator tube before injecting onto the HiLoad 26/600 Superdex ® 75 pg gel filtration column. Fractions containing MpPCO were identified using SDS PAGE, concentrated, aliquoted, flash frozen in $N_2(l)$ and stored at -80 °C.

Sample Id	Fe-1 56 Helium KED (ppb)	Ni 60 Helium KED (ppb)	Zn-1 66 Helium KED (ppb)	Fe 56 Ammonia DRC (ppb)
LeahTrisBlank_1	**0.085**	**0.471**	1.466	**0.809**
LeahHEPESBlank_1	**-0.660**	**0.685**	5.249	**1.615**
LeahAtPCO4_1	**3.064**	**7.984**	5.271	**3.875**
LeahKnPCO_1	**8.356**	**8.279**	**8.701**	**9.153**
LeahMPPCO_1	**4.201**	**2.604**	4.345	**5.480**
LeahTrisBlank_2	**-0.250**	**0.487**	1.249	**0.902**
LeahHEPESBlank_2	**-0.990**	**0.710**	5.164	**1.816**
LeahAtPCO4_2	**1.245**	**7.282**	**8.062**	**4.061**
LeahKnPCO_2	**7.793**	**7.692**	7.495	**8.740**
LeahMPPCO_2	**3.165**	**2.365**	3.629	**5.325**

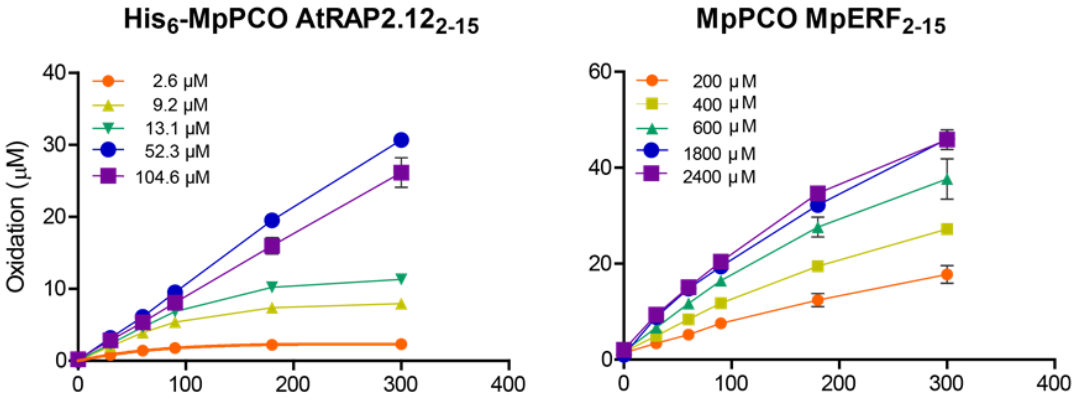
A 3.4: ICP-MS raw data in ppb. ICP-MS raw data for 0.39 µM His_6-MpPCO, 0.53 µM His_6-KnPCO and 0.26 µM His_6-AtPCO 4



A 3.5: Purification and mass of His_6-MpPCO_H184D and His_6-MpPCO_D196N. A) Soluble His_6-MpPCO_H184D was purified using HisTrap™ HP affinity and size exclusion chromatography (SEC). Fractions from SEC containing His_6-MpPCO_H184D are shown here on SDS PAGE. **B)** Soluble His_6-MpPCO_D196N was purified using HisTrap™ HP affinity and size exclusion chromatography (SEC). Fractions from SEC containing His_6-MpPCO_D196N are shown here on SDS PAGE. **C)** His_6-MpPCO_H184D predicted mass of 29731.51 Da was confirmed using LC-MS. **D)** His_6-MpPCO_D196N predicted mass of 29752.58 Da was confirmed using LC-MS. Fractions containing target proteins were concentrated, aliquoted, flash frozen in $N_2(l)$ and stored at -80 °C.

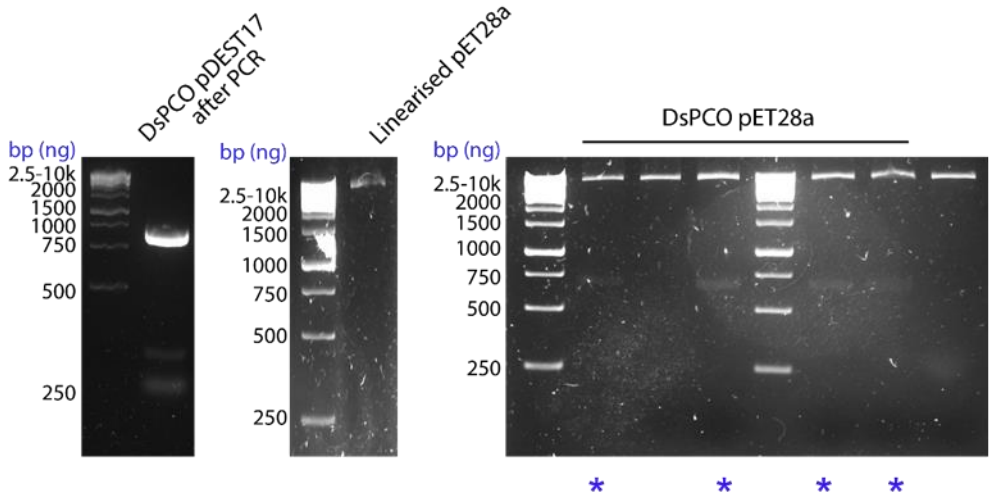


A 3.6: Quenched assay with 4.8 µM $MpHRE_{2-15}$ and His_6-MpPCO showing slight lag phase. Quenched assay with 0.1 µM His_6-MpPCO, 4.8 µM $MpHRE_{2-15}$ and 1 mM TCEP at 25 °C in 50 mM bis-tris propane, 50 mM NaCl, pH 8.

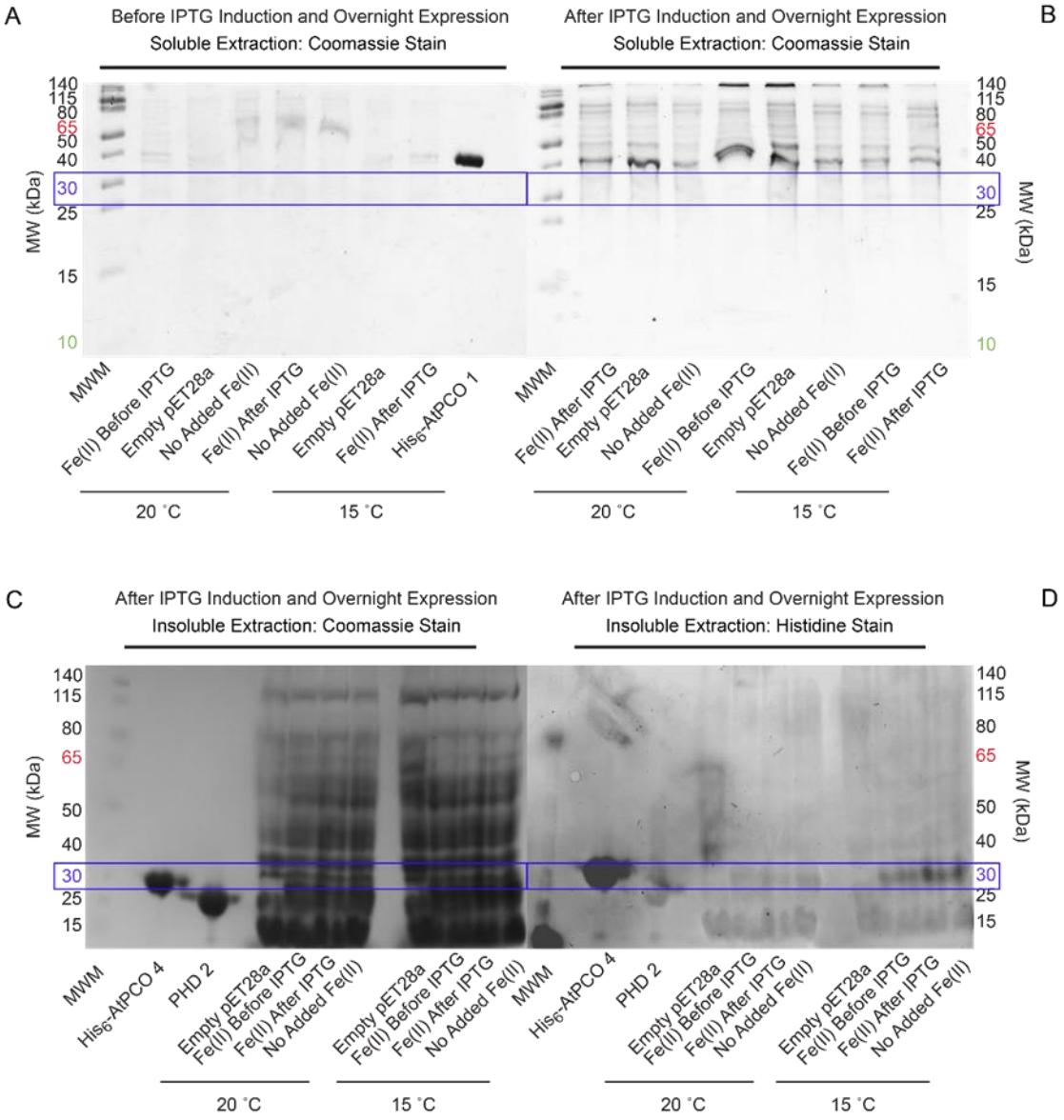


A 3.7: Determining the linear range of peptide turnover: Initial rate assays at different $MpERF_{2-15}$ and $AtRAP2.12_{2-15}$ concentrations at atmospheric O_2. Assays were conducted at 25 °C using 50 mM Bis tris propane, 50 mM NaCl, and 1 mM TCEP, pH 8.0, as buffer. Error bars represent standard deviation, n = 3.

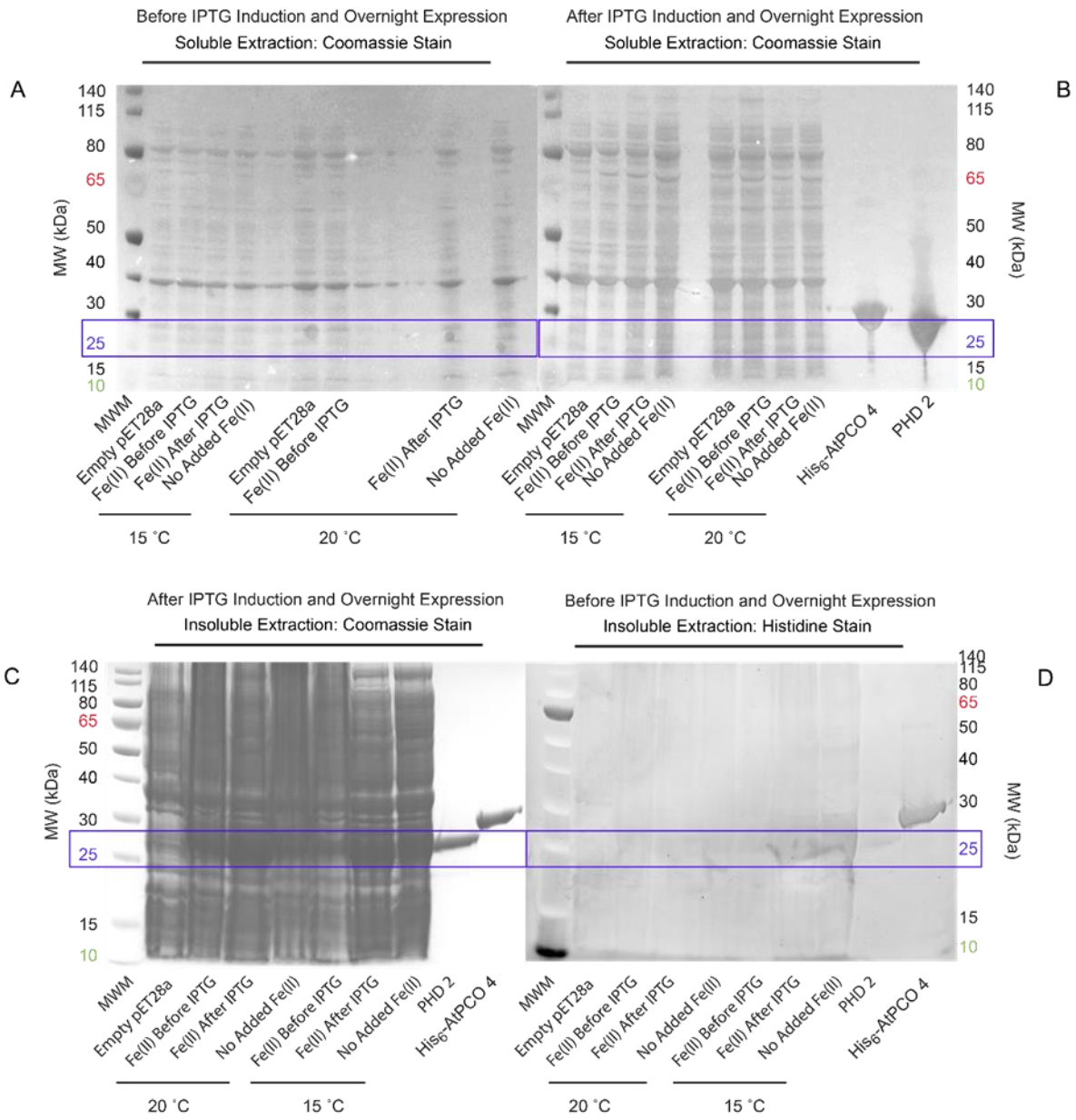
A 4.1: See supplementary files



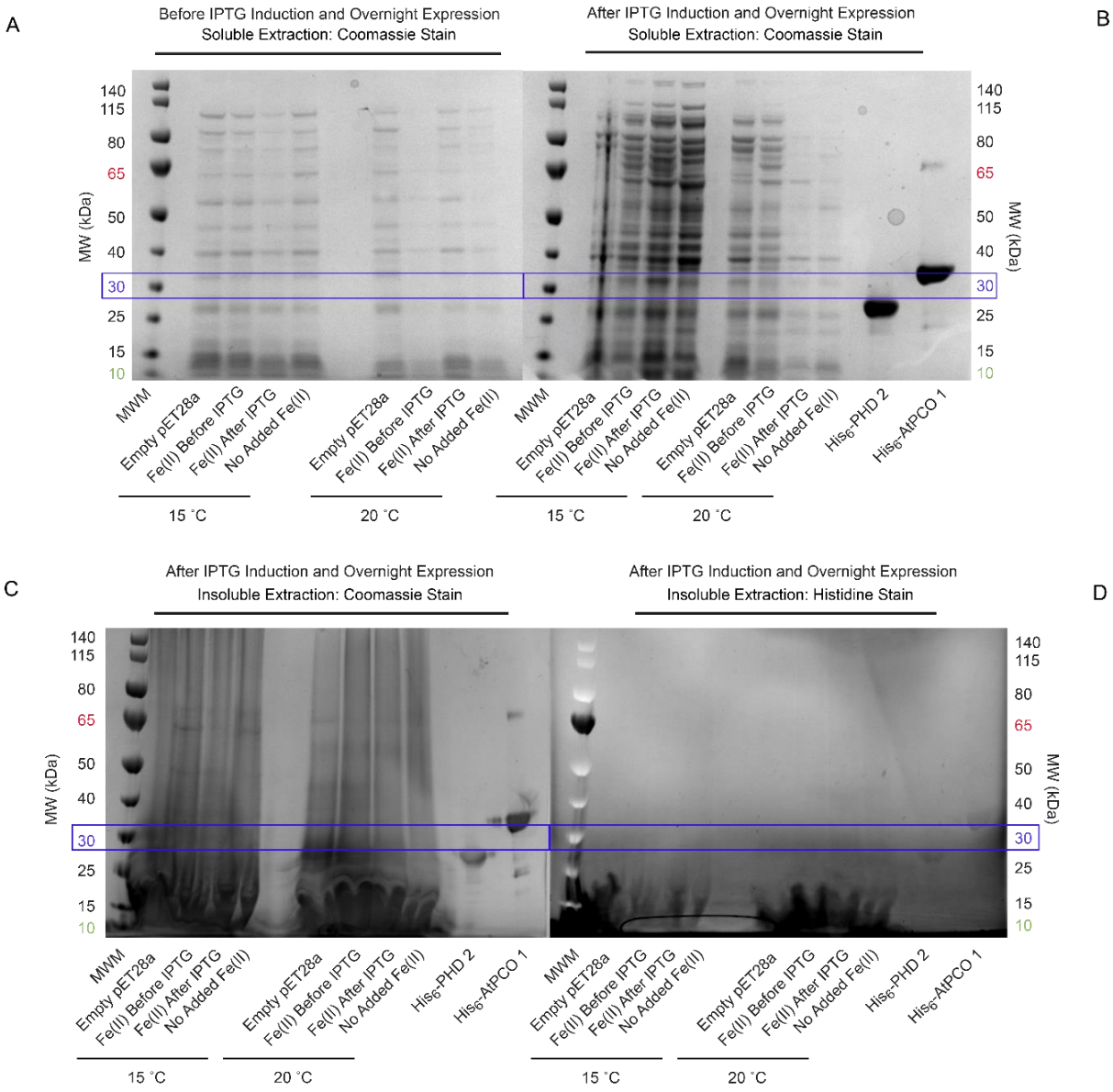
A 4.2: Subcloning the DsPCO sequence into the pET-28a plasmid. Ligation of the DsPCO sequence (702 bp) into the pET28a plasmid, after amplification from the pDEST17 plasmid.



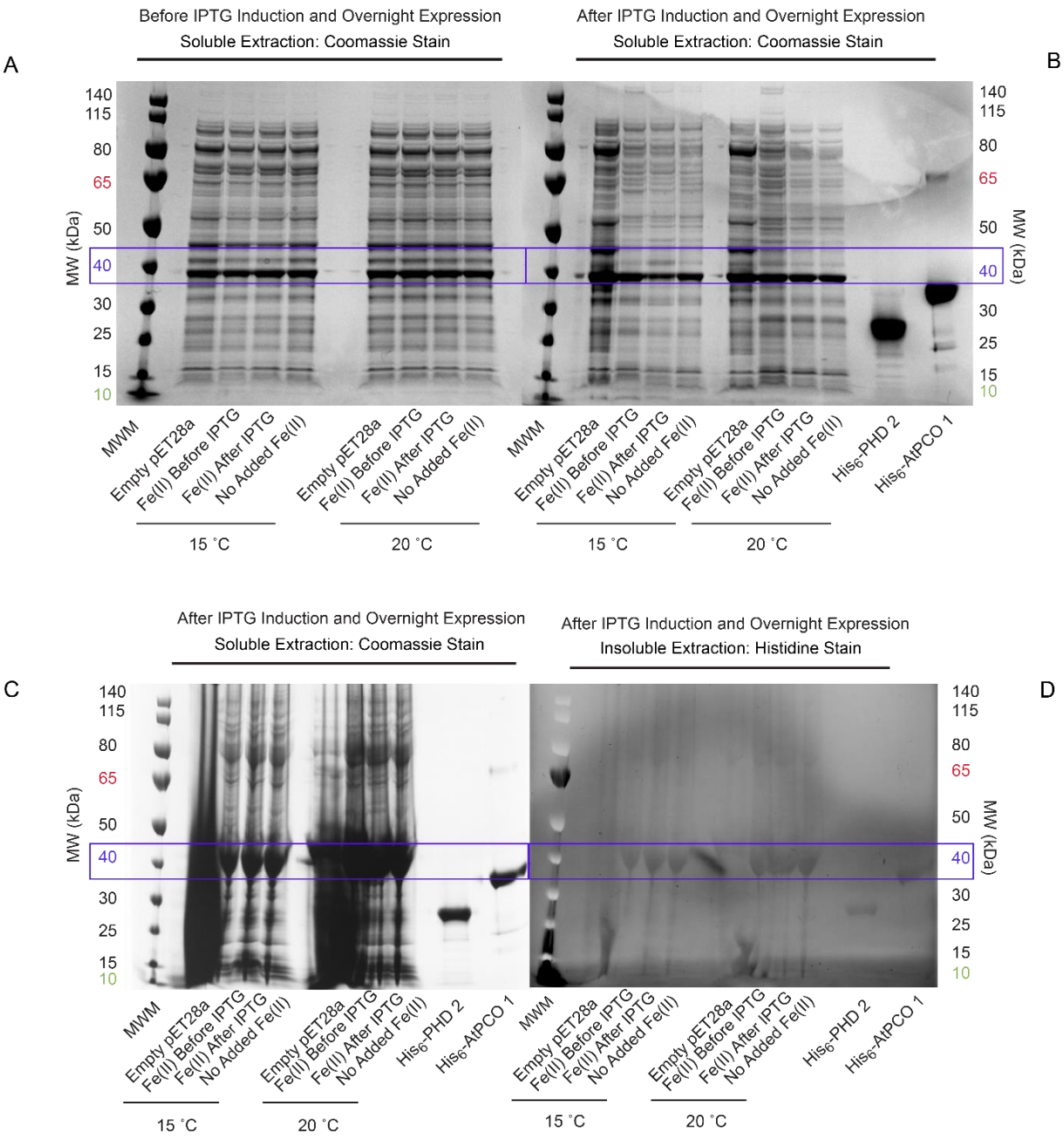
A 4.3: Small scale expression trial of *O. tauri* PCO1 (OtPCO 1). BL21(DE3) cells were transformed with the OtPCO 1 pET 28a plasmid. The transformed cells were grown overnight at 37 °C in the presence of 30 µg/mL kanamycin sulfate. This pre-culture was used to set-up different expression conditions 1) with 100 µM $Fe(NH_4)2(SO_4)2.6H_2O$, 2) with no additional $Fe(NH_4)2(SO_4)2.6H_2O$ and 3) with 100 µM of $Fe(NH_4)2(SO_4)2.6H_2O$ added after IPTG addition. Once these cultures reached an optical density of ~0.8 at 600 nm, protein expression was induced with 0.5 mM IPTG and the cultures incubated overnight at 15 and 20 °C. These cells were sampled for SDS-PAGE analysis **A)** analysis of the OtPCO 1 expression trial before IPTG induction and **B)** after IPTG induction and overnight expression using in-house cast 12% Bis-Tris polyacrylamide gels. The pellet of the cells sampled after IPTG induction and overnight growth was analysed using 4-12% acrylamide NuPAGE gels for insoluble protein. These gels were stained with a **C)** general protein stain (Coomassie) and **D)** a His_6-fusion tag fluorescent stain. His_6-AtPCO4 and His_6-HsPHD $2_{181-407.}$ were used as controls. These gels were imaged using the Bio-Rad ChemiDoc™ XRS+ system. Results suggest OtPCO1 was expressed but was confined to insoluble inclusion bodies under these conditions.



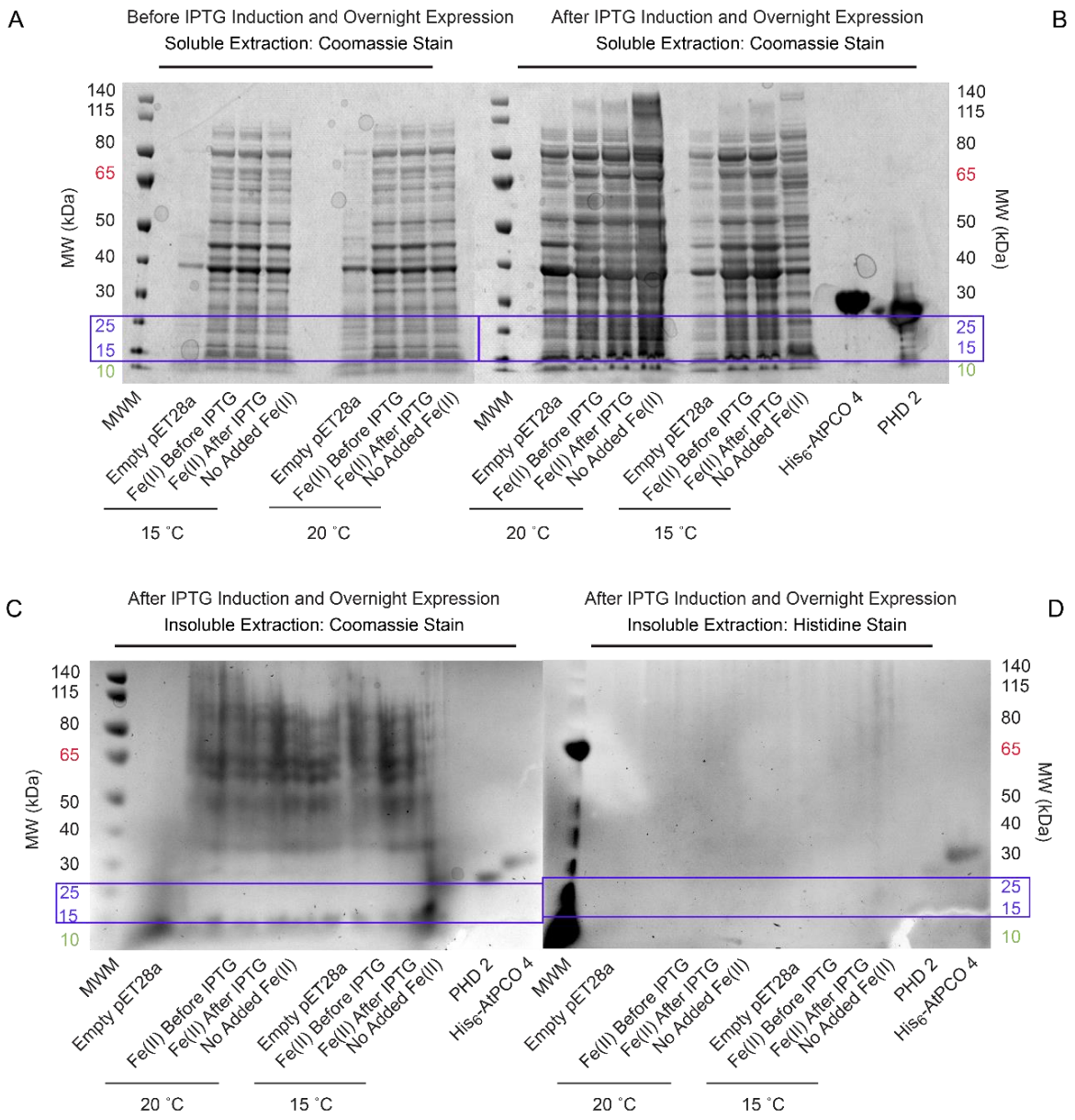
A 4.4: Small scale expression trial of OtPCO 2. BL21(DE3) cells were transformed with the *O. tauri* PCO 2 pET 28a plasmid. The transformed cells were grown overnight at 37 °C in the presence of 30 µg/mL kanamycin sulfate. This pre-culture was used to set-up different expression conditions 1) with 100 µM $Fe(NH_4)2(SO_4)2.6H_2O$, 2) with no additional $Fe(NH_4)2(SO_4)2.6H_2O$ and 3) with 100 µM of $Fe(NH_4)2(SO_4)2.6H_2O$ added after IPTG addition. Once these cultures reached an optical density of ~0.8 at 600 nm, protein expression was induced with 0.5 mM IPTG and the cultures incubated overnight at 15 and 20 °C. These cells were sampled for SDS-PAGE analysis **A)** before IPTG induction and **B)** after IPTG induction and overnight expression. The pellet of the cells sampled after IPTG induction and overnight growth was analysed by SDS-PAGE for insoluble protein. These gels were stained with a **C)** general protein stain (Coomassie) and **D)** a His_6-fusion tag fluorescent stain. His_6-AtPCO 4 and His_6-HsPHD $2_{181-407.}$ were used as controls. These gels were imaged using the Bio-Rad ChemiDoc™ XRS+ system. Results suggest OtPCO 2 was expressed but was confined to insoluble inclusion bodies under these conditions.



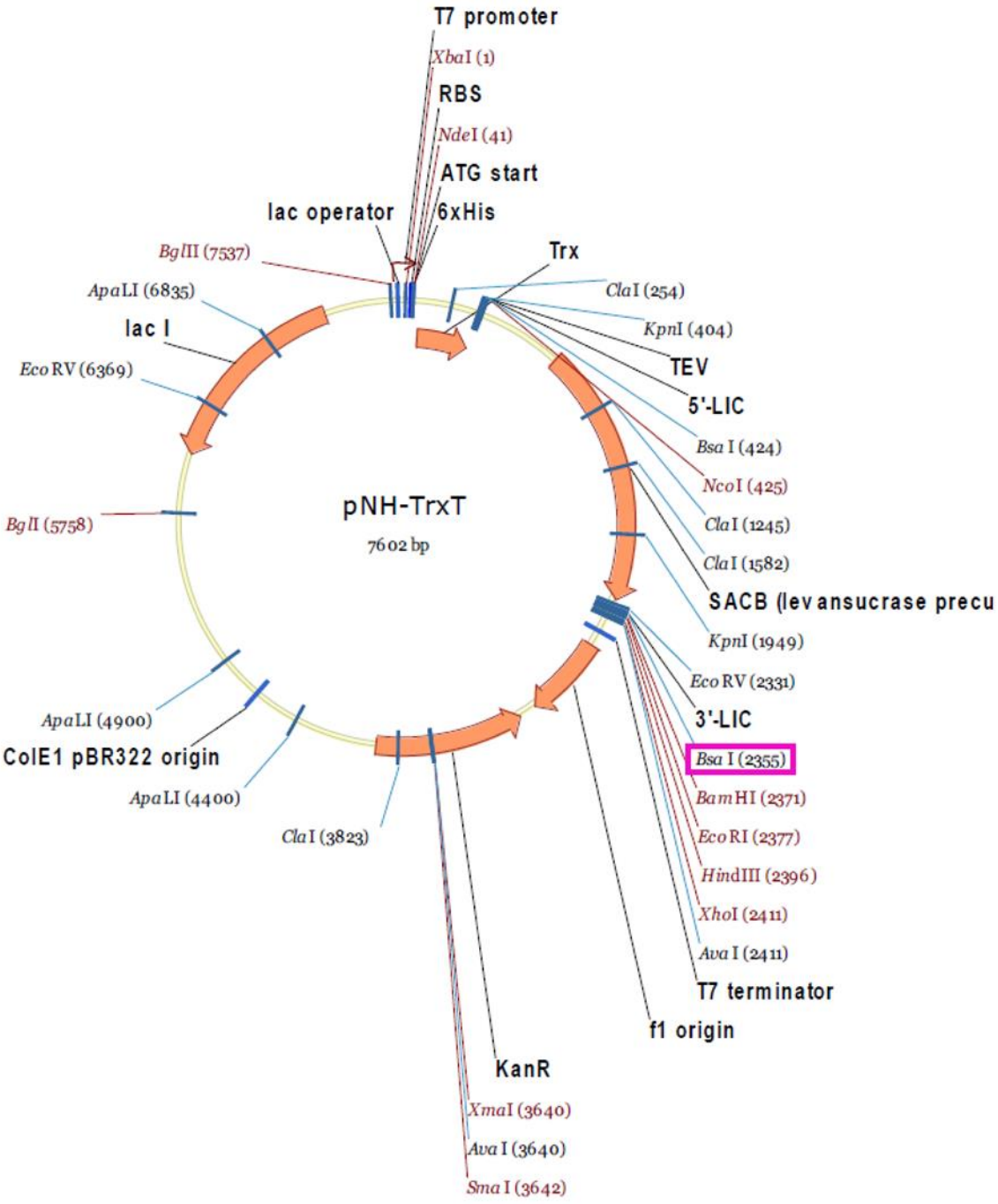
A 4.5: Small scale expression trial of *C. variabilis* PCO (CvPCO). BL21(DE3) cells were transformed with the *C. variabilis* PCO pET 28a plasmid. The transformed cells were grown overnight at 37 °C in the presence of 30 µg/mL kanamycin sulfate. This pre-culture was used to set-up different expression conditions 1) with 100 µM $Fe(NH_4)2(SO_4)2.6H_2O$, 2) with no additional $Fe(NH_4)2(SO_4)2.6H_2O$ and 3) with 100 µM of $Fe(NH_4)2(SO_4)2.6H_2O$ added after IPTG addition, incubated overnight. Once these cultures reached an optical density of ~0.8 at 600 nm, protein expression was induced with 0.5 mM IPTG and the cultures incubated overnight at 15 and 20 °C. These cells were sampled for SDS-PAGE analysis **A)** before IPTG induction and **B)** after IPTG induction and overnight expression. The pellet of the cells sampled after IPTG induction and overnight growth was analysed by SDS-PAGE for insoluble protein. These gels were stained with a **C)** general protein stain (Coomassie) and **D)** a His_6-fusion tag fluorescent stain. His_6-AtPCO 1 and His_6-HsPHD $_{2181-407.}$ were used as controls. These gels were imaged using the Bio-Rad ChemiDoc™ XRS+ system. Results suggest CtPCO was not expressed.



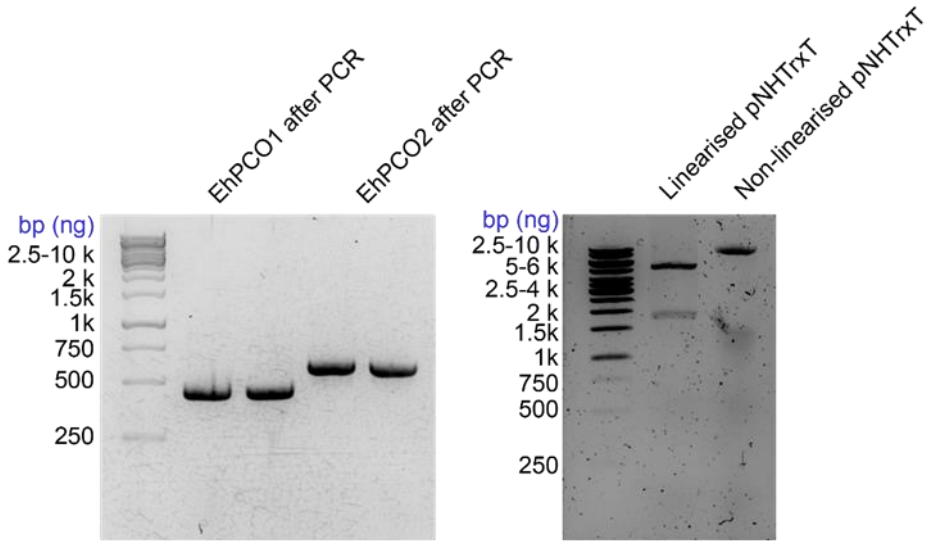
A 4.6: Small scale expression trial of *P. tricornutum* PCO (PtPCO). BL21(DE3) cells were transformed with the *P. tricornutum* PCO pET 28a plasmid. The transformed cells were grown overnight at 37 °C in the presence of 30 µg/mL kanamycin sulfate. This pre-culture was used to set-up different expression conditions 1) with 100 µM $Fe(NH_4)2(SO_4)2.6H_2O$, 2) with no additional $Fe(NH_4)2(SO_4)2.6H_2O$ and 3) with 100 µM of $Fe(NH_4)2(SO_4)2.6H_2O$ added after IPTG addition. Once these cultures reached an optical density of ~0.8 at 600 nm, protein expression was induced with 0.5 mM IPTG and the cultures incubated overnight at 15 and 20 °C. These cells were sampled for SDS-PAGE analysis **A)** before IPTG induction and **B)** after IPTG induction and overnight expression. The pellet of the cells sampled after IPTG induction and overnight growth was analysed by SDS-PAGE for insoluble protein. These gels were stained with a **C)** general protein stain (Coomassie) and **D)** a His_6-fusion tag fluorescent stain. These gels were imaged using the Bio-Rad ChemiDoc™ XRS+ system. Results suggest PtPCO was expressed, but was confined to the insoluble phase.



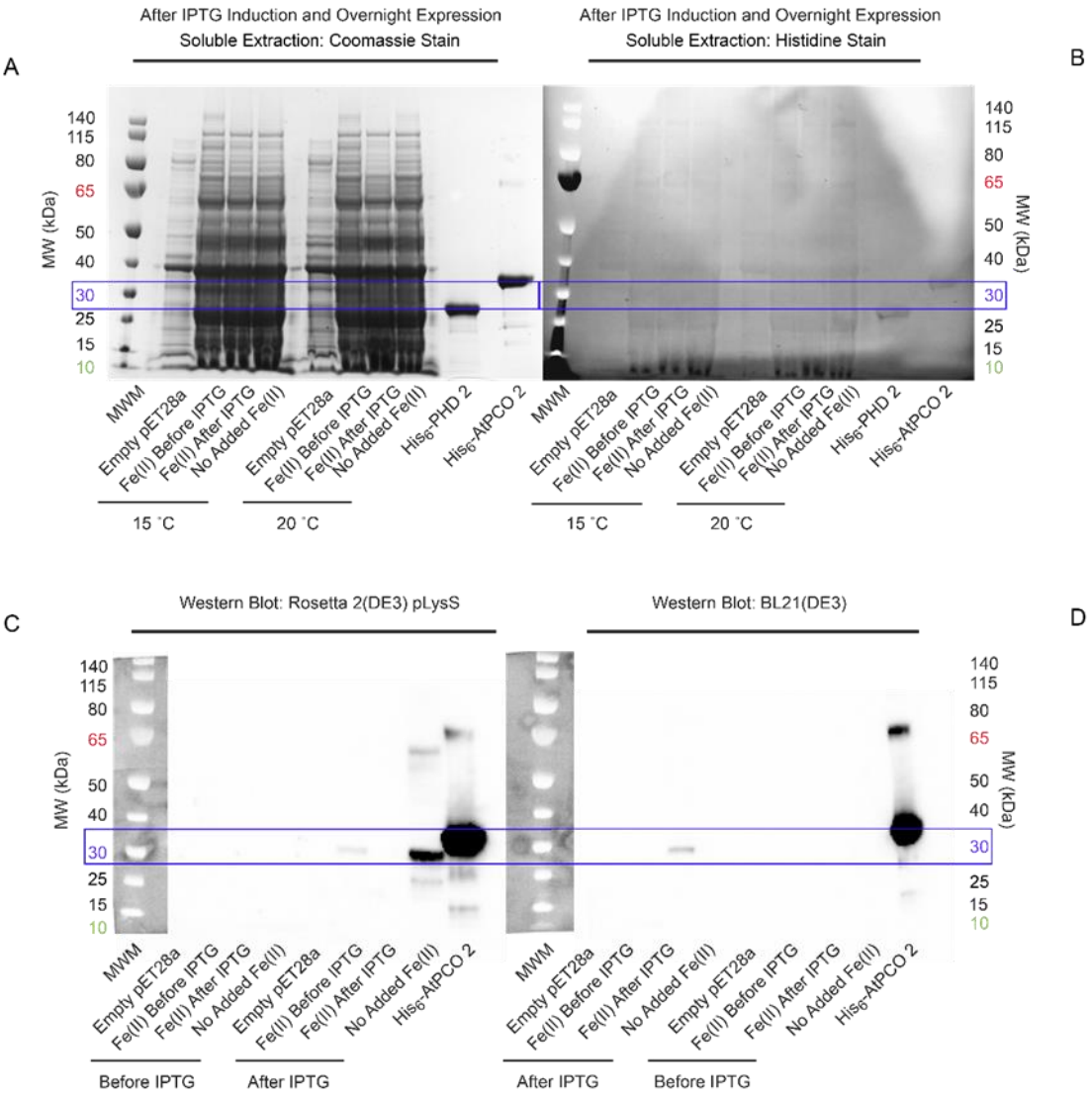
4.7: Small scale expression trial of *E. huxleyi* PCO1 (EhPCO 1). BL21(DE3) cells were transformed with the *E. huxleyi* PCO1 pET 28a plasmid. The transformed cells were grown overnight at 37 °C in the presence of 30 µg/mL kanamycin sulfate. This pre-culture was used to set-up different expression conditions 1) with 100 µM $Fe(NH_4)2(SO_4)2.6H_2O$, 2) with no additional $Fe(NH_4)2(SO_4)2.6H_2O$ and 3) with 100 µM of $Fe(NH_4)2(SO_4)2.6H_2O$ added after IPTG addition. Once these cultures reached an optical density of ~0.8 at 600 nm, protein expression was induced with 0.5 mM IPTG and the cultures incubated overnight at 15 and 20 °C. These cells were sampled for SDS-PAGE analysis **A)** before IPTG induction and **B)** after IPTG induction and overnight expression. The pellet of the cells sampled after IPTG induction and overnight growth was analysed by SDS-PAGE for insoluble protein. These gels were stained with a **C)** general protein stain (Coomassie) and **D)** a His_6-fusion tag fluorescent stain. His_6-AtPCO4 and His_6-HsPHD $_{2181-407.}$ were used as controls. These gels were imaged using the Bio-Rad ChemiDoc™ XRS+ system. Results suggest EhPCO1 was not expressed.



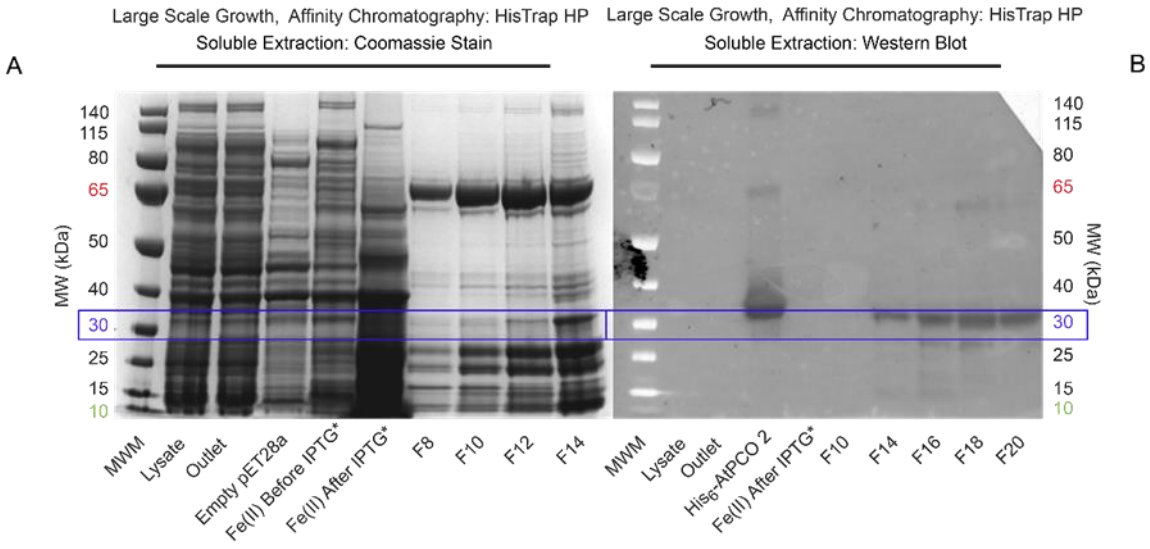
A 4.8: pNHTrxT vector map. Map of the pNHTrxT bacterial vector for the expression of N-terminally His_6-TRX-tagged proteins with a TEV cleavage site. The plasmid was linearized via the Bsa I cleavage site (**highlighted**). The amplified DNA fragment encoding for EhPCO 1 and EhPCO 2 were inserted here using the ligation independent cloning and T4 DNA Ligase.



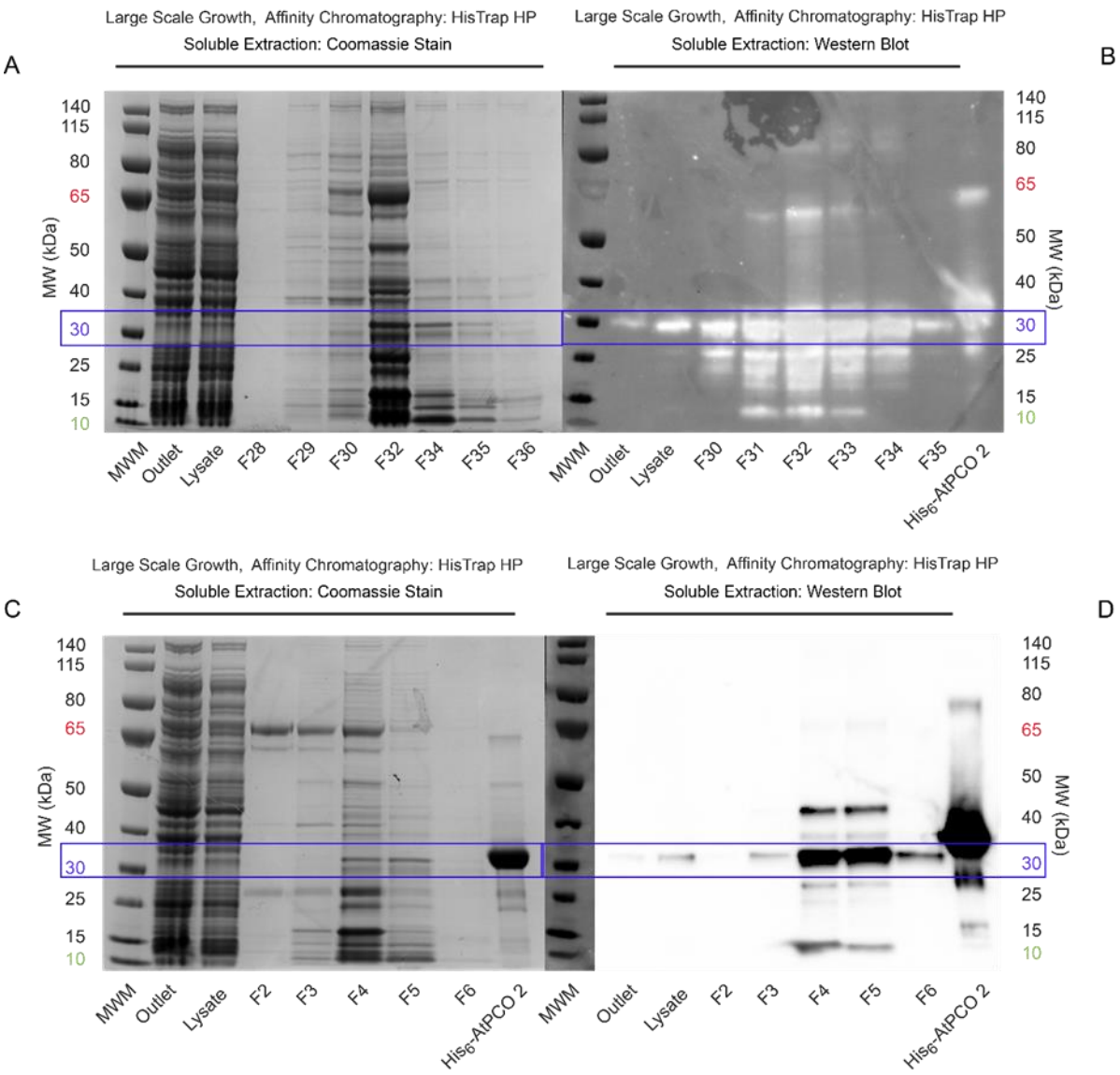
A 4.9: EhPCO 1 and 2 amplification and pNHTrxT linearization. Amplification of the EhPCO 1/2 sequences from their pET28a plasmids and the linearization of the pNHTrxT plasmid on a 2% agarose gel.



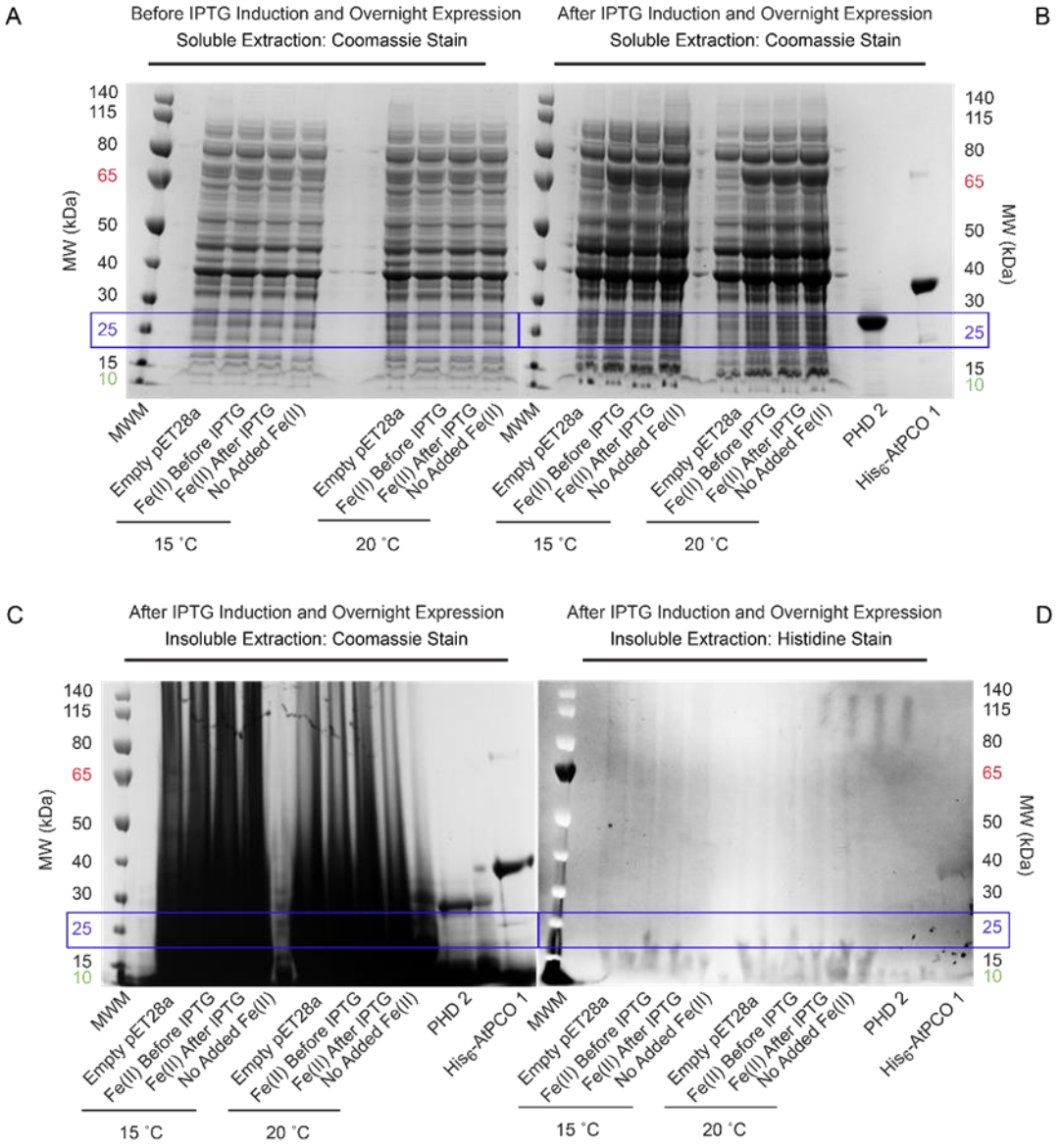
A 4.10: Small scale expression trial of EhPCO 1 in a pNHTrxT plasmid. BL21(DE3) and Rosetta 2(DE3) pLysS cells were transformed with the *E. huxleyi* PCO1 pNHTrxT plasmid. The transformed cells were grown overnight at 37 °C in the presence of 30 µg/mL kanamycin sulfate. This pre-culture was used to set-up different expression conditions 1) with 100 µM $Fe(NH_4)2(SO_4)2.6H_2O$, 2) with no additional $Fe(NH_4)2(SO_4)2.6H_2O$ and 3) with 100 µM of $Fe(NH_4)2(SO_4)2.6H_2O$ added after IPTG addition. Once these cultures reached an optical density of ~0.8 at 600 nm, protein expression was induced with 0.5 mM IPTG and the cultures incubated overnight at 15 and 20 °C. These cells were sampled for SDS-PAGE analysis **A)** by Coomassie stain after IPTG induction and overnight expression and **B)** by His_6-fusion tag fluorescent stain. These expression trials were then analysed by western blot (WB), using an antibody raising against a His_6-fusion tag. **C)** the WB of the Rosetta 2(DE3) pLysS expression and **D)** the WB of the BL21(DE3) expression trial. His_6-AtPCO 2 was used as control. Results suggest His_6-TRX-EhPCO1 was best expressed in Rosetta 2(DE3) pLysS cells in the absence of Fe(II). The gels and WB membranes were imaged using the Bio-Rad ChemiDoc™ XRS+ system.



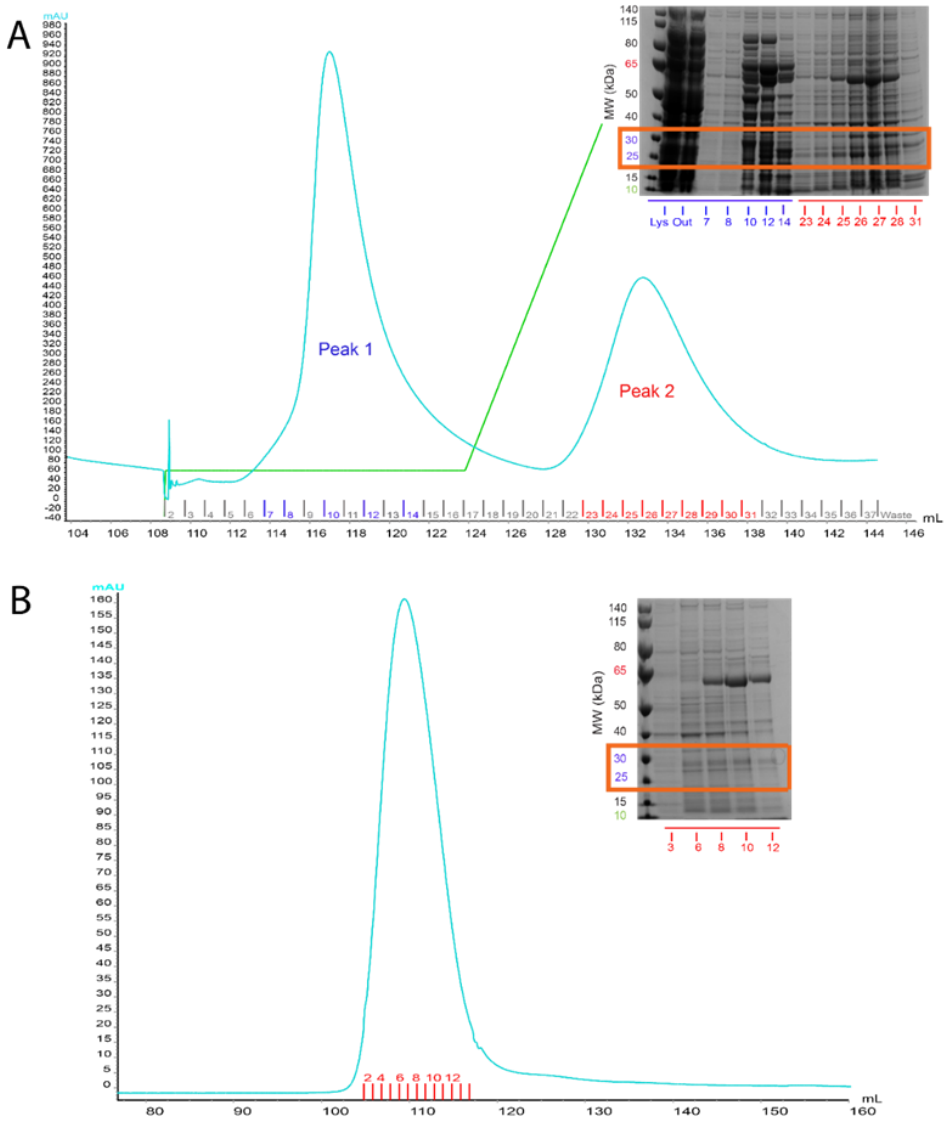
A 4.11: Small scale expression trial of EhPCO 1 in a pNHTrxT plasmid with additional Fe. Rosetta 2(DE3) pLysS cells were transformed with the *E. huxleyi* PCO1 pNHTrxT plasmid. The transformed cells were grown overnight at 37 °C in the presence of 30 µg/mL kanamycin sulfate. This pre-culture was used to set-up expression conditions with additional 100 µM $Fe(NH_4)2(SO_4)2.6H_2O$. Once these cultures reached an optical density of ~0.8 at 600 nm, protein expression was induced with 0.5 mM IPTG and the cultures incubated overnight at 20 °C. These cells were sampled for SDS-PAGE analysis **A)** by Coomassie after overnight expression and purification by HisTrap™ and **B)** by WB, using an antibody raising against a His_6-fusion tag. His_6-AtPCO 2 was used as control. The gels and WB membranes were imaged using the Bio-Rad ChemiDoc™ XRS+ system. Results suggest His_6-TRX-EhPCO1 was expressed in low levels in the presence of Fe(II).



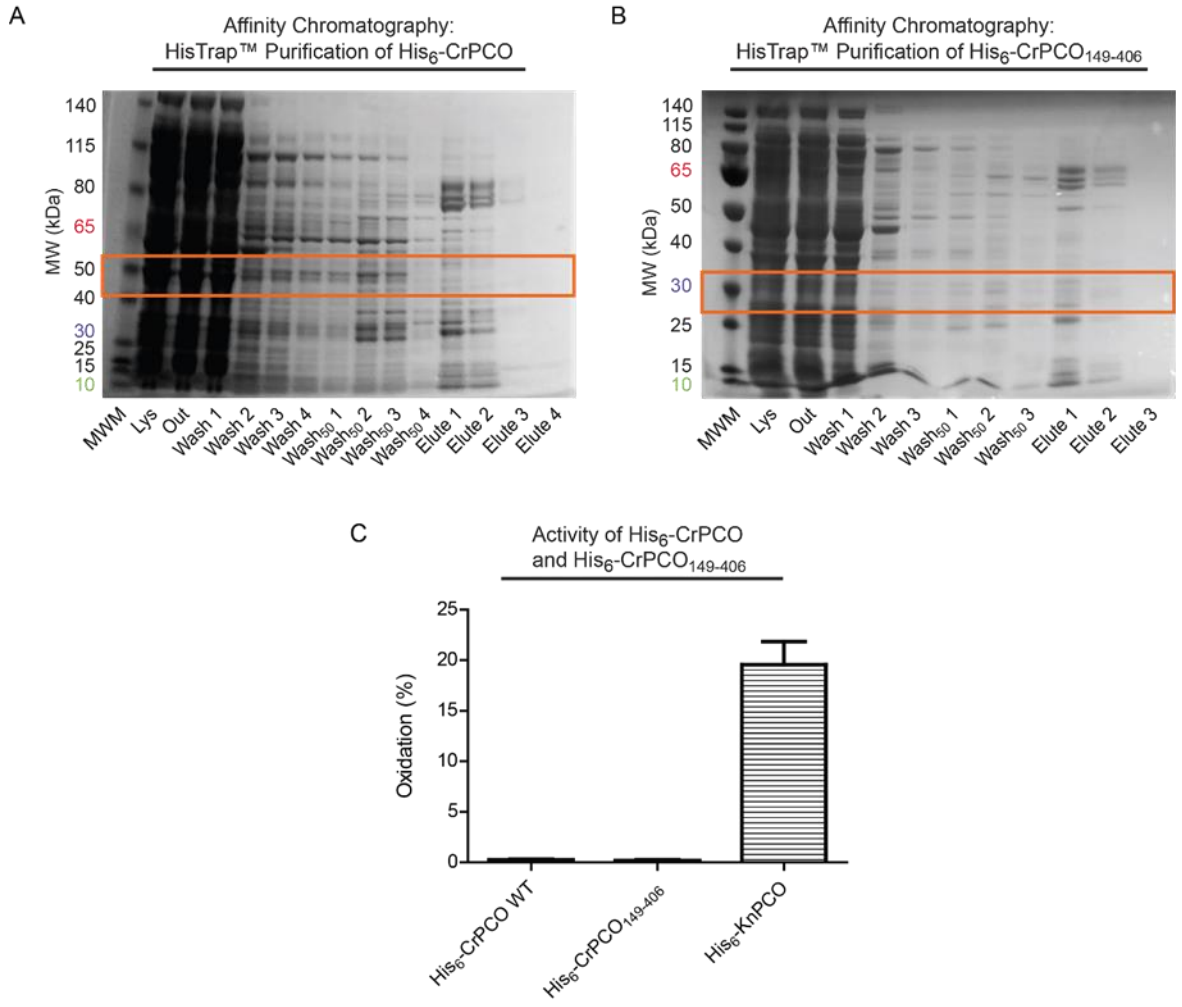
A 4.12: **Small scale expression trial of EhPCO 1 in a pNHTrxT plasmid with no additional Fe(II).** Rosetta 2(DE3) pLysS cells were transformed with the *E. huxleyi* PCO1 pNHTrxT plasmid. The transformed cells were grown overnight at 37 °C in the presence of 30 µg/mL kanamycin sulfate. This pre-culture was used to set-up expression conditions with no additional Fe(II). Once these cultures reached an optical density of ~0.8 at 600 nm, protein expression was induced with 0.5 mM IPTG and the cultures incubated overnight at 20 °C. These cells were sampled for SDS-PAGE analysis **A/C)** by Coomassie stain after overnight expression and purification by HisTrap™ and **B/D)** by WB, using an antibody raising against a His_6-fusion tag. His_6-AtPCO 2 was used as control. The gels and WB membranes were imaged using the Bio-Rad ChemiDoc™ XRS+ system. Results suggest His_6-TRX-EhPCO1 was expressed in low levels in without Fe(II), and that His_6-TRX-EhPCO1 degradation occurs quickly (overnight).



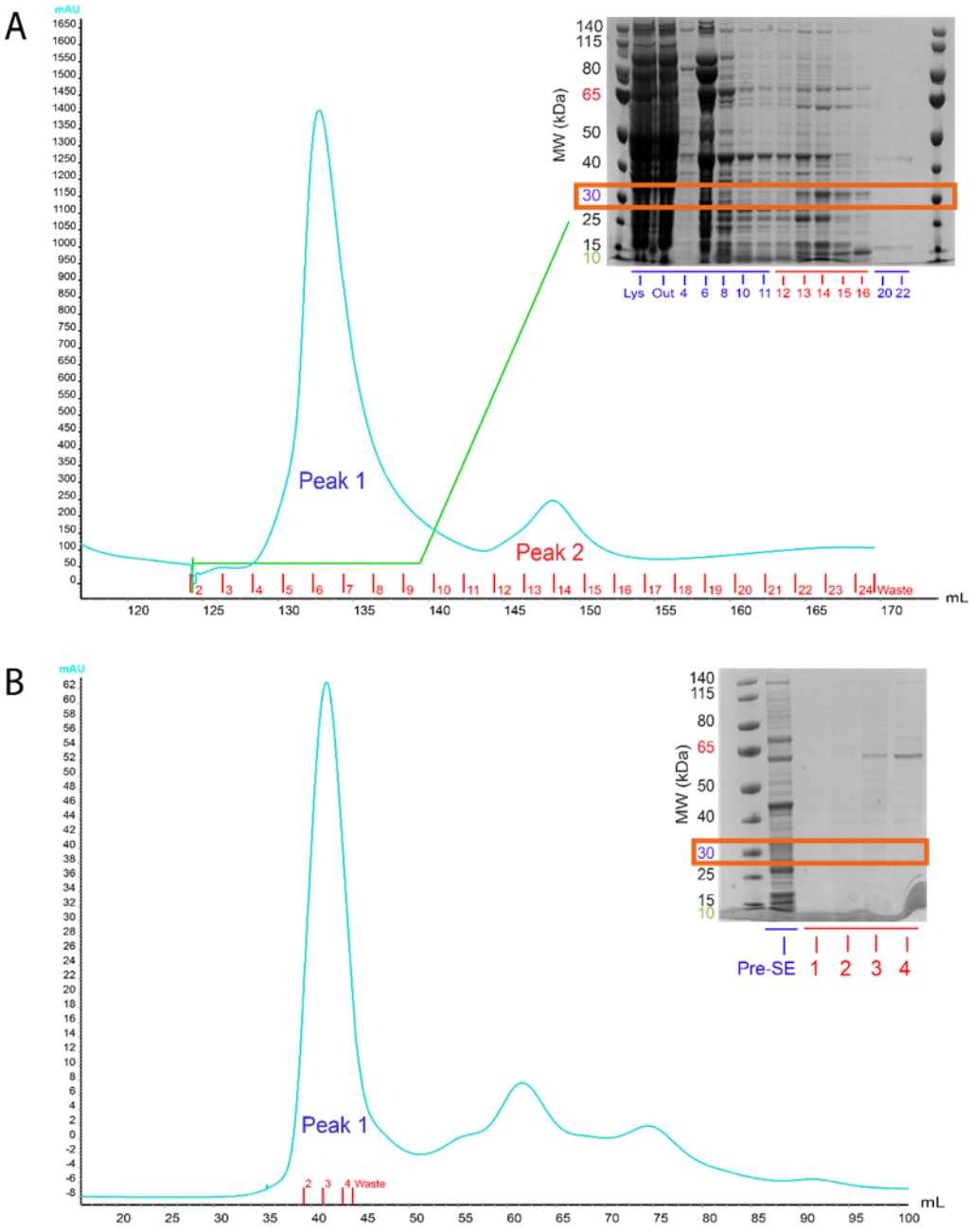
A 4.13: Small scale expression trial of EhPCO 2. BL21(DE3) cells were transformed with the *E. huxleyi* PCO 2 pET 28a plasmid. The transformed cells were grown overnight at 37 °C in the presence of 30 µg/mL kanamycin sulfate. This pre-culture was used to set-up different expression conditions 1) with 100 µM $Fe(NH_4)2(SO_4)2.6H_2O$, 2) with no additional $Fe(NH_4)2(SO_4)2.6H_2O$ and 3) with 100 µM of $Fe(NH_4)2(SO_4)2.6H_2O$ added after IPTG addition. Once these cultures reached an optical density of ~0.8 at 600 nm, protein expression was induced with 0.5 mM IPTG and the cultures incubated overnight at 15 and 20 °C. These cells were sampled for SDS-PAGE analysis **A)** before IPTG induction and **B)** after IPTG induction and overnight expression. The pellet of the cells sampled after IPTG induction and overnight growth was analysed by SDS-PAGE for insoluble protein. These gels were stained with a **C)** general protein stain (Coomassie) and **D)** a His_6-fusion tag fluorescent stain. His_6-AtPCO 1 and His_6-HsPHD $2_{181-407}$. were used as controls. These gels were imaged using the Bio-Rad ChemiDoc™ XRS+ system. Results suggest EhPCO2 was not expressed.



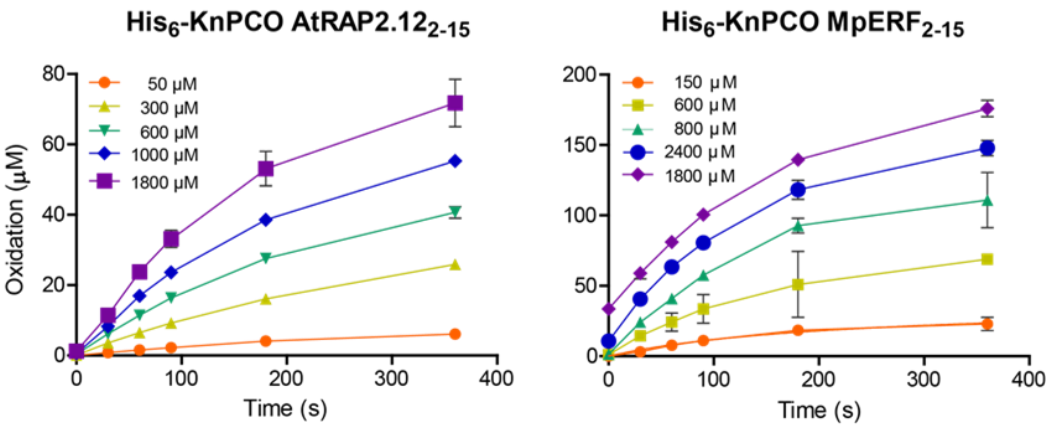
A 4.14: Large scale expression trial of DsPCO. A) Soluble His_6-DsPCO was purified using HisTrap™ HP affinity chromatography, eluting over a 0-100% gradient of elution buffer. His_6-DsPCO containing fractions were identified using SDS PAGE. Protein-containing fractions were buffer swapped using PD-10 desalting column at 4˚C prior to gel filtration chromatography. **B)** Fractions 23-31 were concentrated in a MWCO 10 kDa concentrator tube before injecting onto the HiLoad 26/600 Superdex ® 75 pg gel filtration column. Fractions were analysed for His_6-DsPCO using SDS PAGE. Results suggest His_6-DsPCO was not successfully purified.



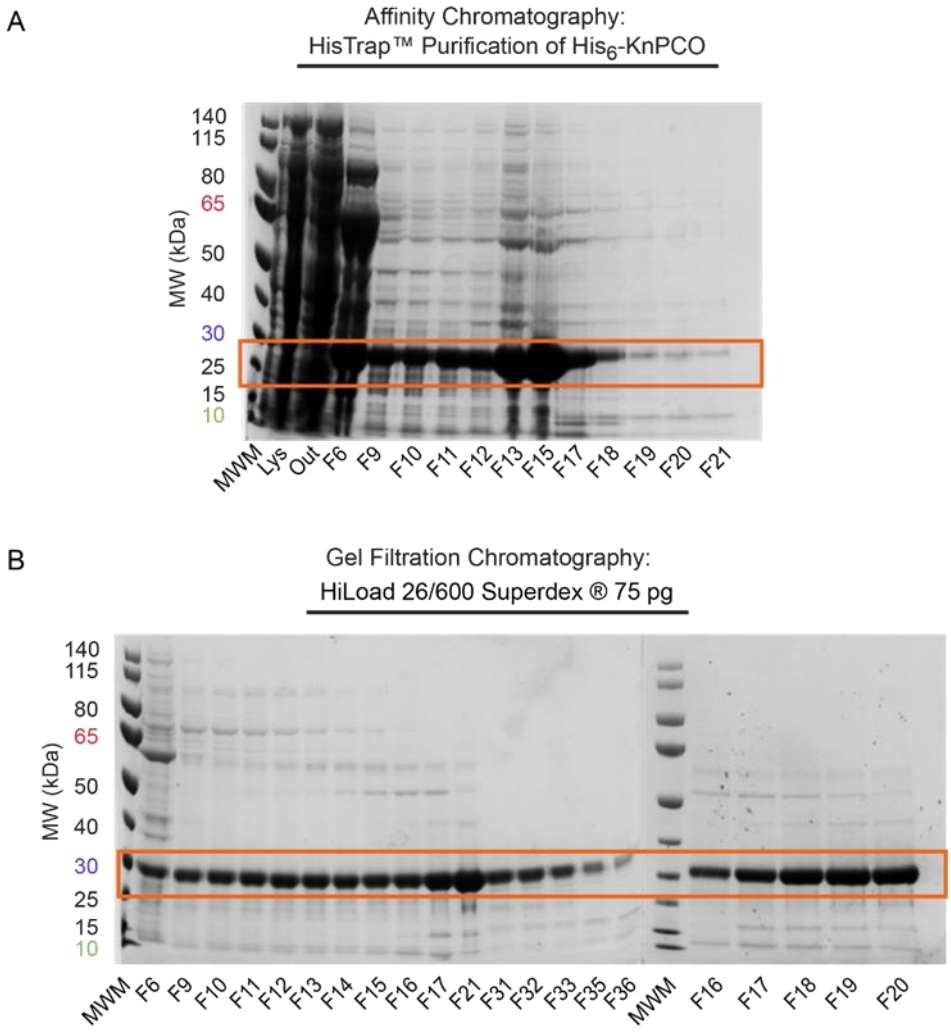
A 4.15 Expression and purification of CrPCO WT and CrPCO$_{149-406}$. BL21(DE3) cells were transformed with the *C. reinhardtii* PCO (CrPCO) and CrPCO$_{149-406}$ pET 28a plasmids. The transformed cells were grown overnight at 37 °C in the presence of 30 µg/mL kanamycin sulfate. This pre-culture was used to set-up large scale growths in the presence of 30 µg/mL kanamycin sulfate. The cultures were grown at 37 °C to an optical density of ~0.8 at 600 nm. Protein expression was induced with 0.5 mM IPTG and the cultures incubated overnight at 18°C. A) SDS-PAGE of HisTrap™ purified His$_6$-CrPCO pellet. The column was washed with 3 column volumes (CV) of binding buffer, followed by 3 CV of 50 mM imidazole buffer and 3 CV of elution buffer. Expected mass His$_6$-CrPCO WT = 45.7 kDa. **A)** SDS-PAGE of HisTrap™ purified His$_6$-CrPCO and **B)** His$_6$-CrPCO$_{149-406}$ pellets. Expected mass of His$_6$-CrPCO is 45699.02 Da and His$_6$-CrPCO$_{149-406}$ is 30075.47 Da. **C)** Activity assay of His$_6$-CrPCO WT and His$_6$-CrPCO$_{149-406}$ with the AtRAP2.12$_{2-15}$ peptide for 1 hour, and His$_6$-KnPCO with AtRAP2.12$_{2-15}$ peptide for 6 min as a control.



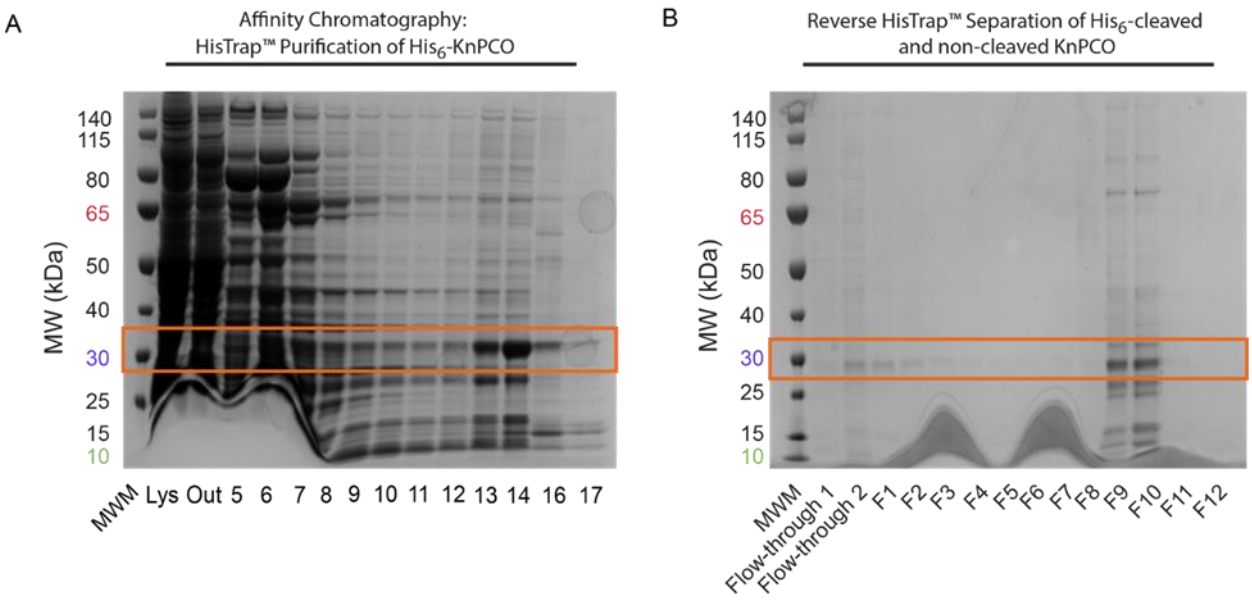
A 4.16: Large scale expression trial of His_6-KnPCO_H184D. A) Soluble His_6-KnPCO_H184D was purified using HisTrap™ HP affinity chromatography, eluting over a 0-100% gradient of elution buffer. His_6-KnPCO_H184D containing fractions were identified using SDS PAGE. Protein-containing fractions were buffer swapped using PD-10 desalting column at 4˚C prior to gel filtration chromatography. **B)** Fractions 12-16 were concentrated in a MWCO 10 kDa concentrator tube before injecting onto the HiLoad 26/600 Superdex ® 75 pg gel filtration column. Fractions were analysed for His_6-KnPCO_H184D using SDS PAGE. Results suggest small amounts of His_6-KnPCO_H184D was purified by HisTrap™, but was not successfully purified by gel filtration.



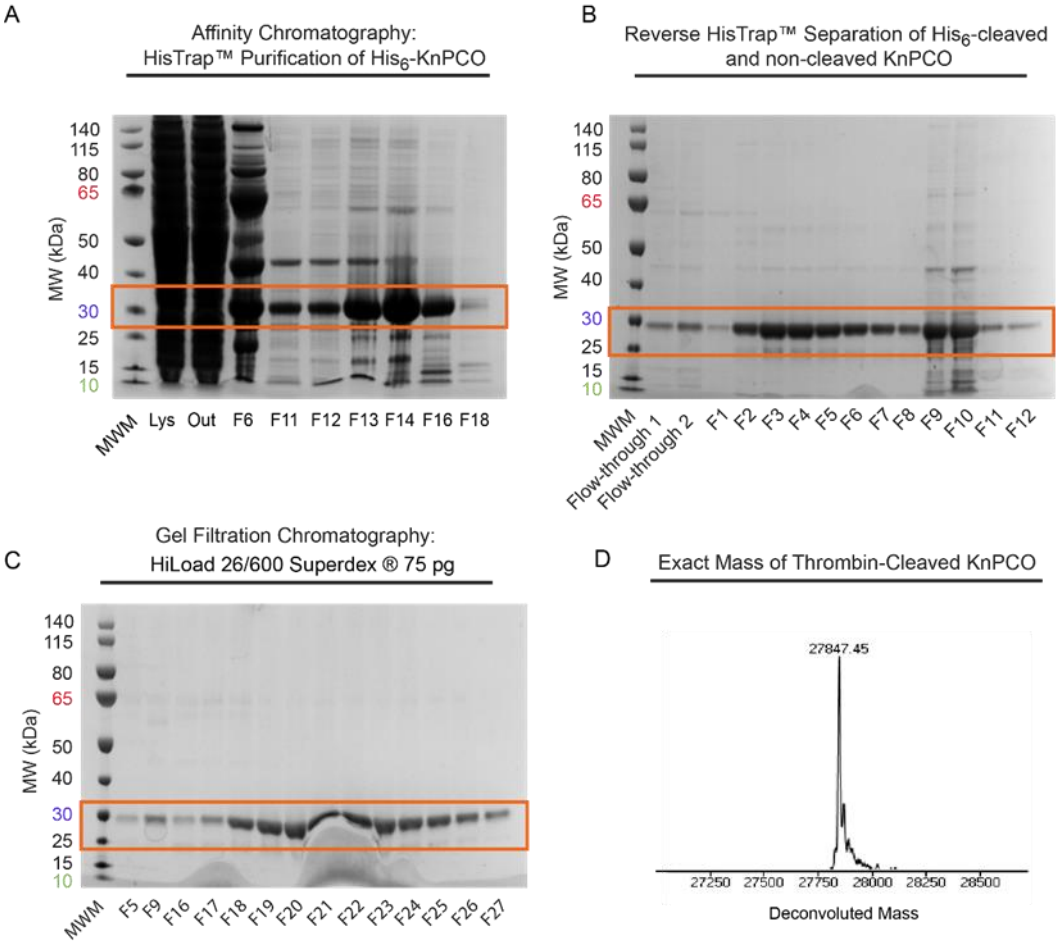
A 4.17: Determining the linear range of peptide turnover: Initial rate assays at different $MpERF_{2\text{-}15}$ and $AtRAP2.12_{2\text{-}15}$ concentrations at atmospheric O_2. Assays were conducted at 25 °C using 50 mM Bis tris propane, 50 mM NaCl, and 1 mM TCEP, pH 8.0, as buffer. Error bars represent standard deviation, n = 3.



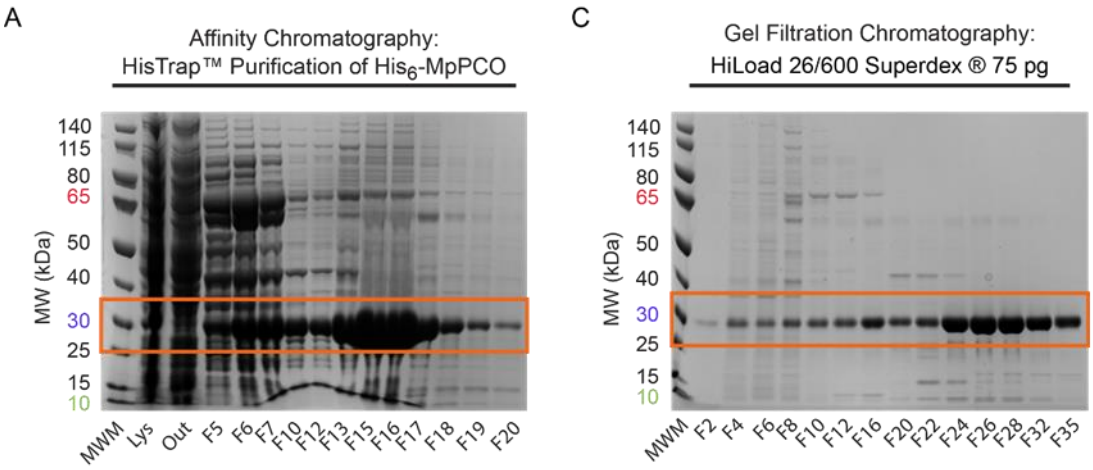
A 5.1: Expressing His_6-KnPCO for crystallography. A) SDS-PAGE of HisTrap™ purified *E. coli* BL21 DE3 lysate with overexpressed, recombinant His_6-KnPCO. B) SDS-PAGE of gel filtered HisTrap™ Peak 2 fractions using the HiLoad 26/600 Superdex® 75 pg column. Fractions 21 – 36 were concentrated to 32 mg/mL as they contained the fewest impurities. Fractions 16-20 were concentrated separately.



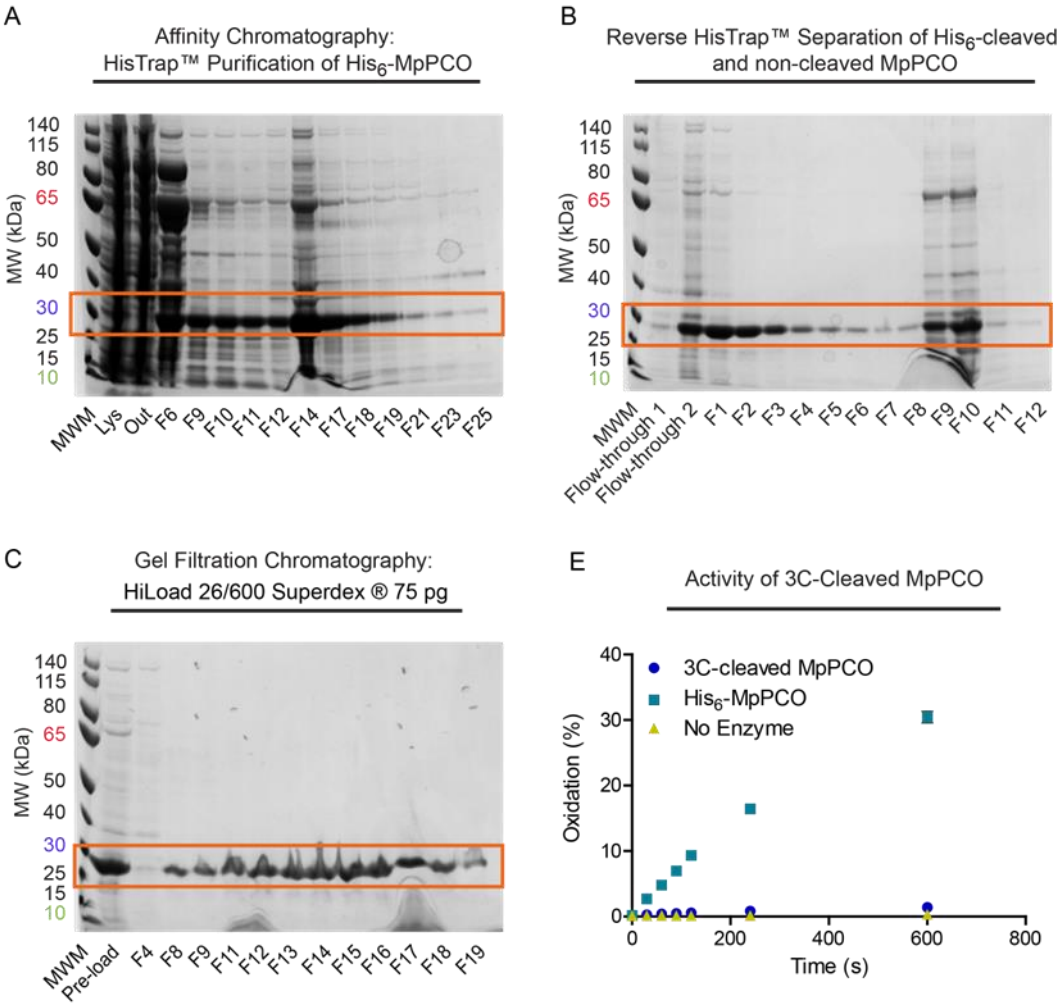
A 5.2: Preparing His_6-cleaved KnPCO using 3C protease. A) SDS-PAGE of HisTrap™ purified *E. coli* BL21 DE3 lysate with overexpressed, recombinant His_6-KnPCO. B) SDS-PAGE of HisTrap™-purified KnPCO treated with 3C protease overnight, separating His_6-cleaved KnPCO from uncleaved. Fractions 1-8 were concentrated and purified by gel filtration. The UV trace was lower than the 10 mAu cut-off for fractions collection, and as such no gel filtered protein was collected.



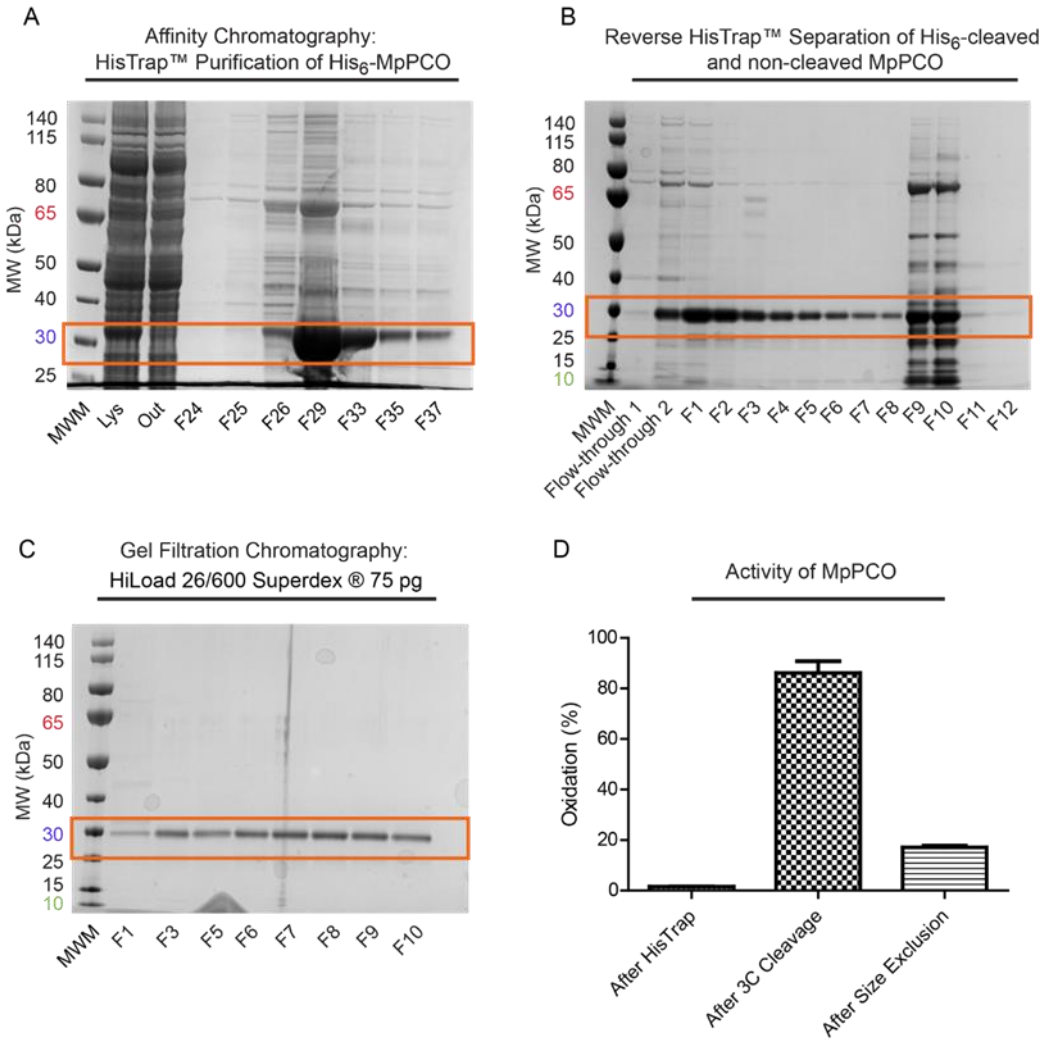
A 5.3: Preparing His_6-cleaved KnPCO using thrombin protease for crystallisation and antibody production. A) SDS-PAGE of HisTrap™ purified *E. coli* BL21 DE3 lysate with overexpressed, recombinant His_6-KnPCO. B) SDS-PAGE of reverse HisTrap™ Peak 2 fractions treated with thrombin protease overnight, separating His_6-cleaved KnPCO from uncleaved. C) Fractions 1-8 were concentrated and purified by gel filtration using the HiLoad 26/600 Superdex® 75 pg column. D) Exact KnPCO mass was determined by mass spectrometry. This protein was used to raise an antibody against KnPCO, and to lay some crystallisation screens.



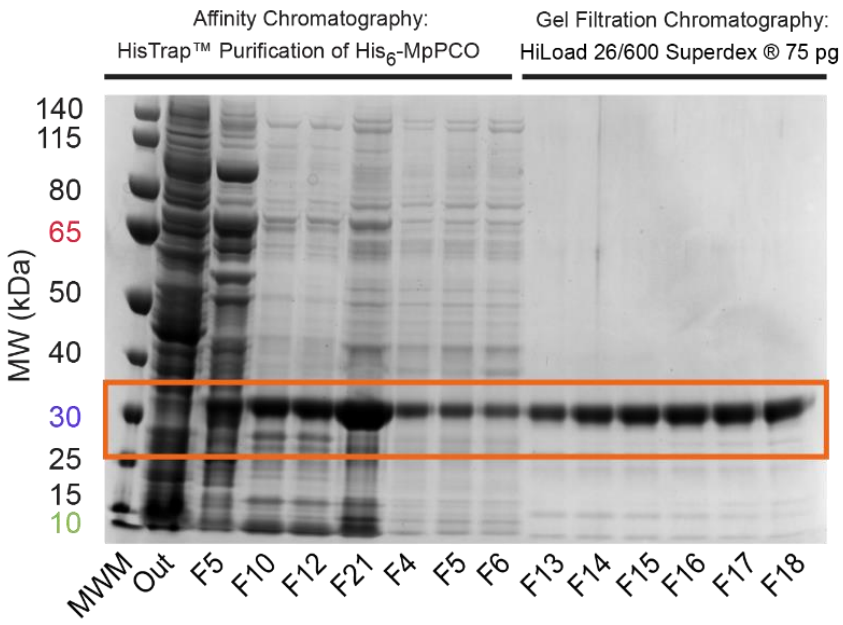
A 5.4: Preparing His_6-MpPCO for crystallography trials. A) SDS-PAGE of HisTrap™ purified *E. coli* BL21 DE3 lysate with overexpressed, recombinant His_6-MpPCO; fractions 10-20 were selected for further purification. B) The HisTrap™-purified fractions were purified by gel filtration using the HiLoad 26/600 Superdex® 75 pg column. Gel-filtered fractions 24 – 35 were chosen for crystallisation trials



A 5.5: Expression of His_6-cleaved MpPCO for crystallography. A) SDS-PAGE of HisTrap™ purified *E. coli* BL21 DE3 lysate with overexpressed, recombinant His_6-MpPCO. B) Reverse HisTrap™ of 3C protease-treated recombinant His_6-MpPCO, separating His_6-cleaved MpPCO from non-cleaved. C) Reverse HisTrap™ fractions flow through 2 and F1-F6 were purified by gel filtration. D) An activity assay with the $MpERF_{2\text{-}15}$ peptide suggested the 3C-cleaved MpPCO protein was not active (assay data collected by Mr J. Wilson, Chemistry Department).

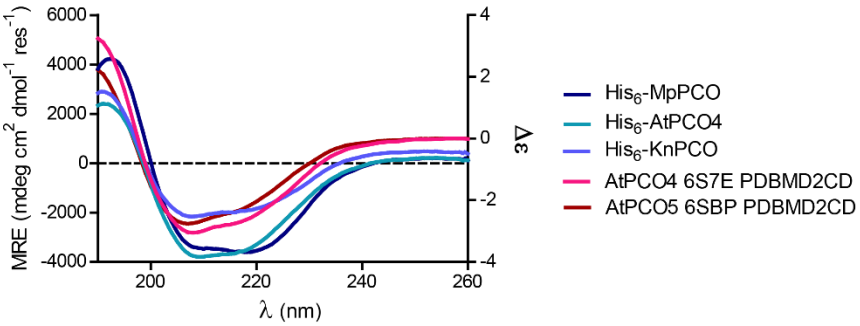


A 5.6: Expression of His_6-cleaved MpPCO for crystallography. A) SDS-PAGE of HisTrap™ purified *E. coli* BL21 DE3 lysate with overexpressed, recombinant His_6-MpPCO. B) Reverse HisTrap™ of 3C protease-treated recombinant His_6-MpPCO, separating His_6-cleaved MpPCO from non-cleaved. C) Reverse HisTrap™ fractions flow-through 2 and F1-F8 were purified by gel filtration. D) Protein samples after HisTrap™ purification, after 3C cleavage and manual HisTrap™ purification and after gel filtration were buffer-swapped using a micro Bio-Spin 6 column from BioRad and incubated with $MpERF_{2\text{-}15}$ peptide for 30 min at 25 °C.



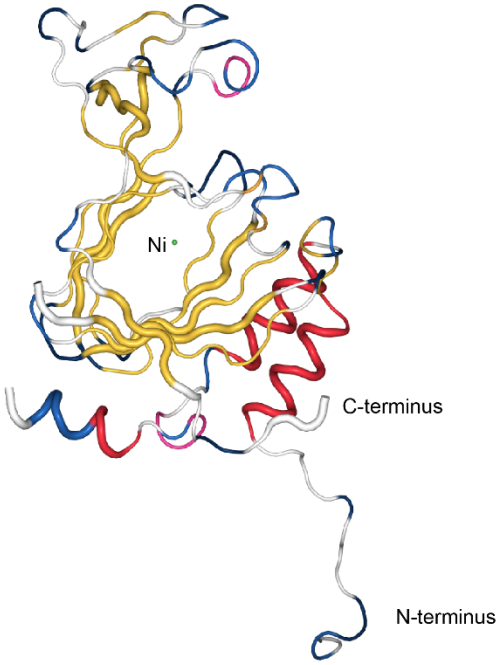
A 5.7: Expression of His_6-cleaved MpPCO for metal removal. A) SDS-PAGE of HisTrap™-purified *E. coli* BL21 DE3 lysate with overexpressed, recombinant His_6-MpPCO and gel filtered His_6-MpPCO.

A) PDBMD2CD vs. experimental data



helix	β-strand	turn	irregular
0.13	0.34	0.11	0.42

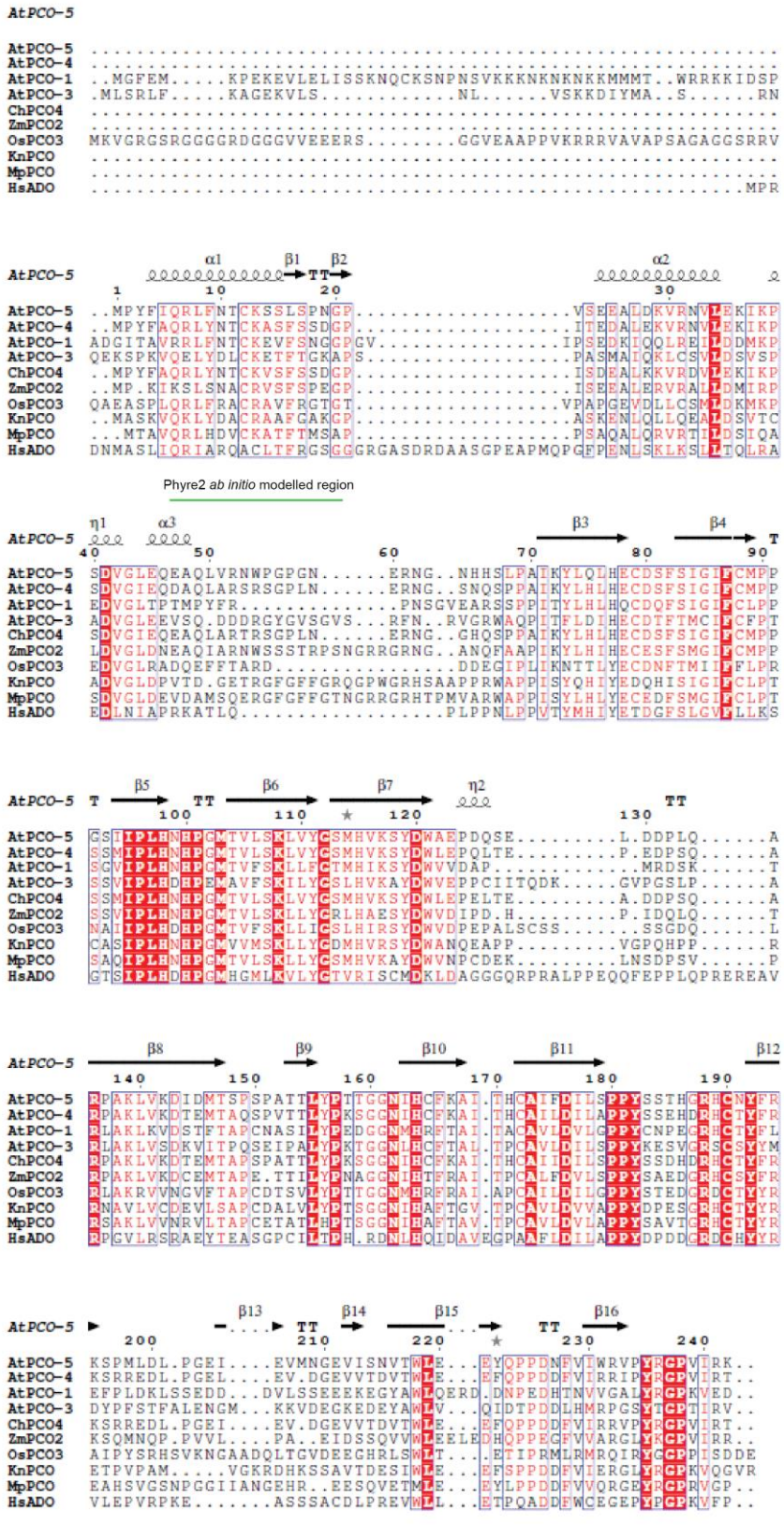
B) DSSP





helix	β-strand	other
0.14	0.37	0.49

A 5.8: Secondary structure assignment of the AtPCO5 crystal structure. A) PDBMD2CD CD and secondary structure (SS) prediction tool, using the AtPCO5 crystal structure (PDB ID: 6SBP [1], green). For comparison, the experimental CD spectrum for AtPCO4, KnPCO and MpPCO are included. B) Dictionary of Secondary Structure of Proteins (DSSP) assigns secondary structure using PDB entries (*via* the 2StrucCompare server). Both models are in agreement on the secondary structure prediction for AtPCO5.

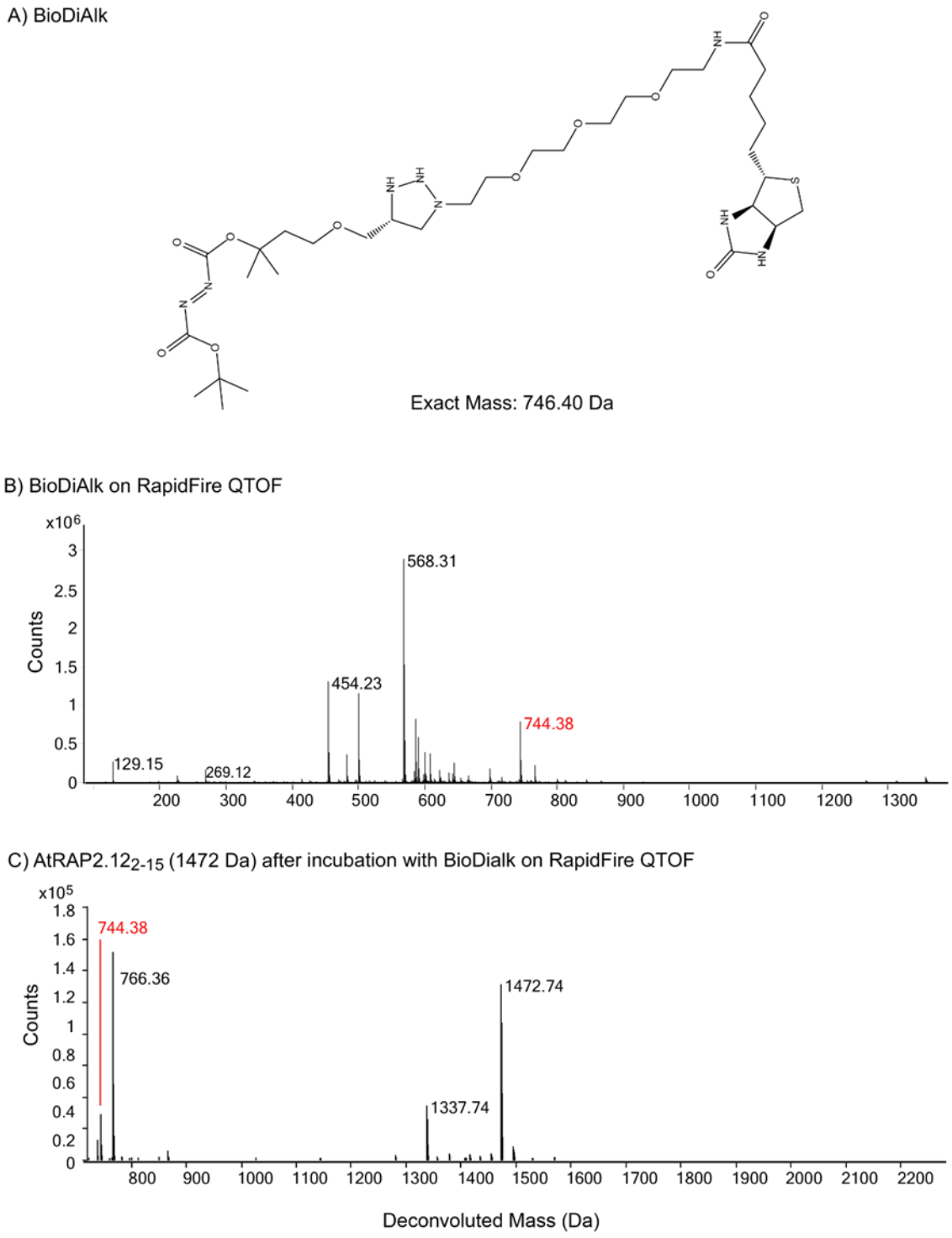


A 5.9: Alignment of PCO and HsADO sequences with the AtPCO4 crystal structure. The protein sequences were aligned using Clustal Omega. The secondary structure annotation was added using ESPript 3.0. The secondary structure is based on AtPCO4 (PDB: 6S7E [1]). The non-conserved region is marked in green. Conserved residues are highlighted in red.

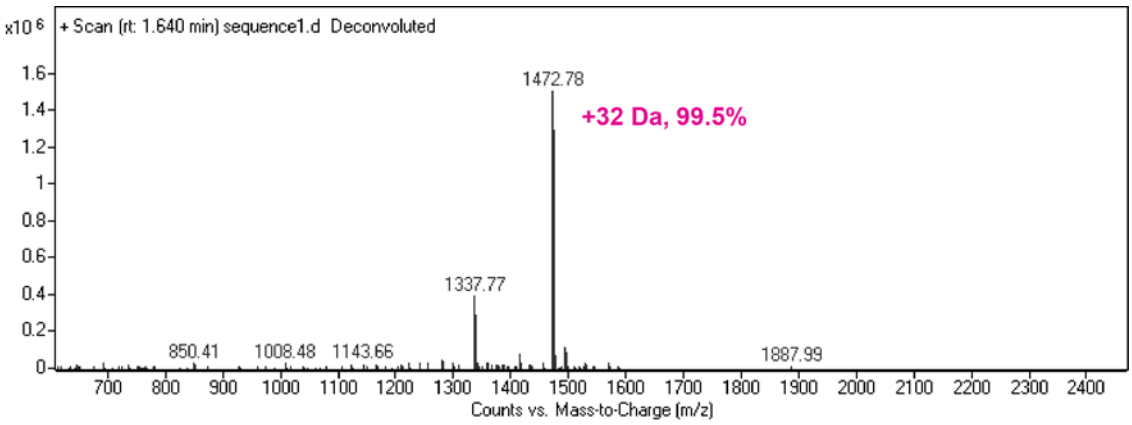
1: AtPCO-4	100.00	91.70	58.23	39.83	46.29	41.42	76.76	42.37	50.00	30.53
2: ChPCO4	91.70	100.00	58.23	38.53	46.29	41.00	80.08	43.22	50.00	30.09
3: ZmPCO2	58.23	58.23	100.00	38.77	44.98	39.50	58.65	42.44	48.55	30.63
4: OsPCO3	39.83	38.53	38.77	100.00	38.91	38.58	39.22	38.63	41.03	29.24
5: AtPCO-1	46.29	46.29	44.98	38.91	100.00	38.85	45.65	43.86	47.62	27.90
6: AtPCO-3	41.42	41.00	39.50	38.58	38.85	100.00	39.58	41.32	44.26	26.05
7: AtPCO-5	76.76	80.08	58.65	39.22	45.65	39.58	100.00	40.51	48.12	29.07
8: KnPCO	42.37	43.22	42.44	38.63	43.86	41.32	40.51	100.00	54.25	33.92
9: MpPCO	50.00	50.00	48.55	41.03	47.62	44.26	48.12	54.25	100.00	31.72
10: HsADO	30.53	30.09	30.63	29.24	27.90	26.05	29.07	33.92	31.72	100.00

1: AtPCO-1	100.00	57.61	38.02	46.93	45.41	46.98	43.48
2: AtPCO-2	57.61	100.00	41.57	42.04	40.97	43.36	40.27
3: AtPCO-3	38.02	41.57	100.00	41.35	39.50	43.85	41.32
4: AtPCO-4	46.93	42.04	41.35	100.00	76.76	49.58	42.68
5: AtPCO-5	45.41	40.97	39.50	76.76	100.00	47.28	40.83
6: MpPCO	46.98	43.36	43.85	49.58	47.28	100.00	54.25
7: KnPCO	43.48	40.27	41.32	42.68	40.83	54.25	100.00

A 5.10: Identity matrix of AtPCOs with KnPCO and MpPCO. The percentage identities of the protein sequences were determined using Clustal Omega.

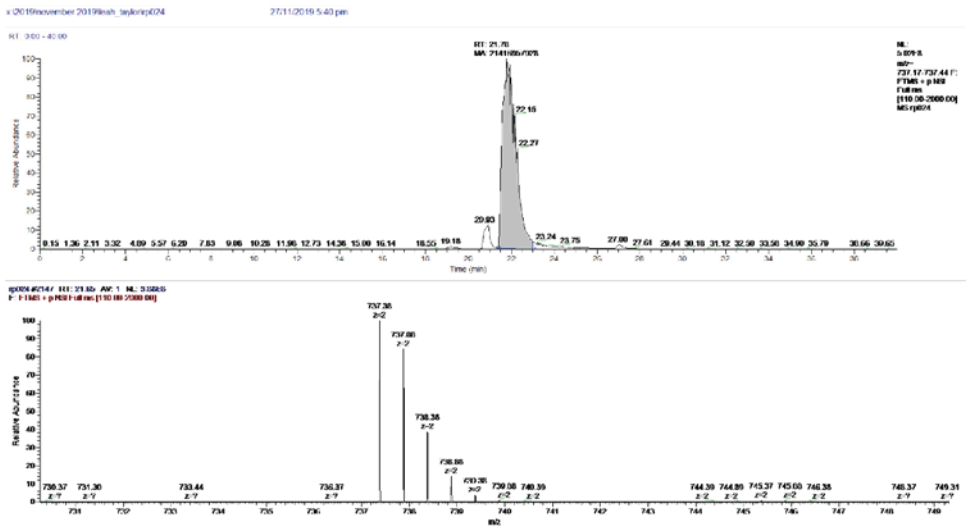


A 6.1: Incubation of AtRAP2.12$_{2-15}$CysO$_2$H with BioDialk probe. A) The BioDiAlk probe was analysed by Q-TOF MS to determine its mass and integrity. **B)** A mass similar to that of the BioDiAlk probe was observed (744.38 Da, expected 746.4 Da). **C)** AtRAP2.12$_{2-15}$ was incubated with 1 mM TCEP and 1 µM His$_6$-KnPCO in 50 mM BisTrisPropane, 50mM NaCl, pH 8 for an hour. AtRAP2.12$_{2-15}$CysO$_2$H (1472.74 Da) formation was confirmed by Q-TOF MS. 5 µM of the BioDiAlk probe was then incubated with AtRAP2.12$_{2-15}$CysO$_2$H at 4 °C for 30 min at 600 rpm. A mass corresponding to the AtRAP2.12$_{2-15}$CysO$_2$H + BioDiAlk probe was not observed in the deconvoluted spectrum (2119.06 Da).

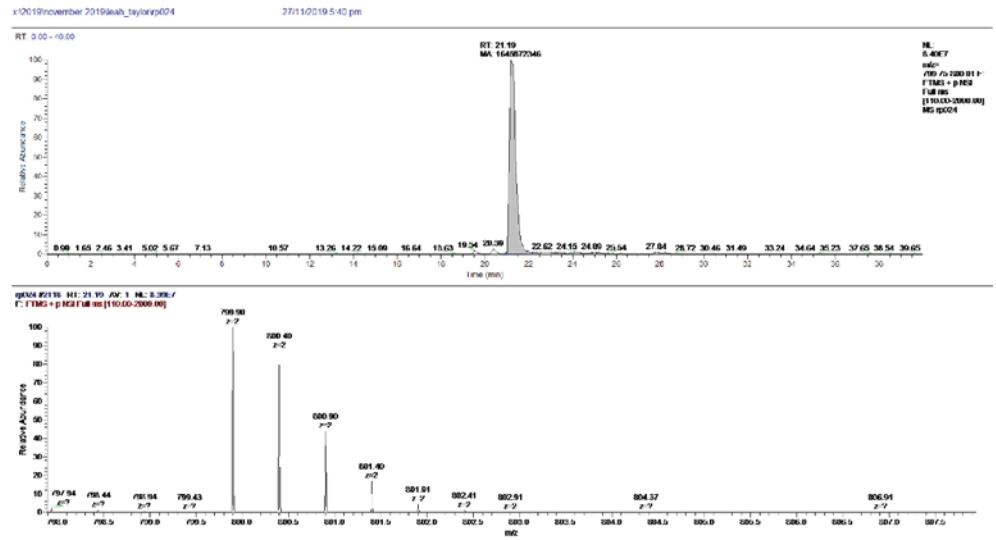


A 6.2: Preparation of AtRAP2.12$_{2\text{-}15}$CysO$_2$H for maleimide-probe screens. AtRAP2.12$_{2\text{-}15}$ (MW = 1440.78 Da) was oxidised to AtRAP2.12$_{2\text{-}15}$CysO$_2$H (MW = 1472.78 Da) with a yield of 99.5% using Agilent Quantitative Analysis.

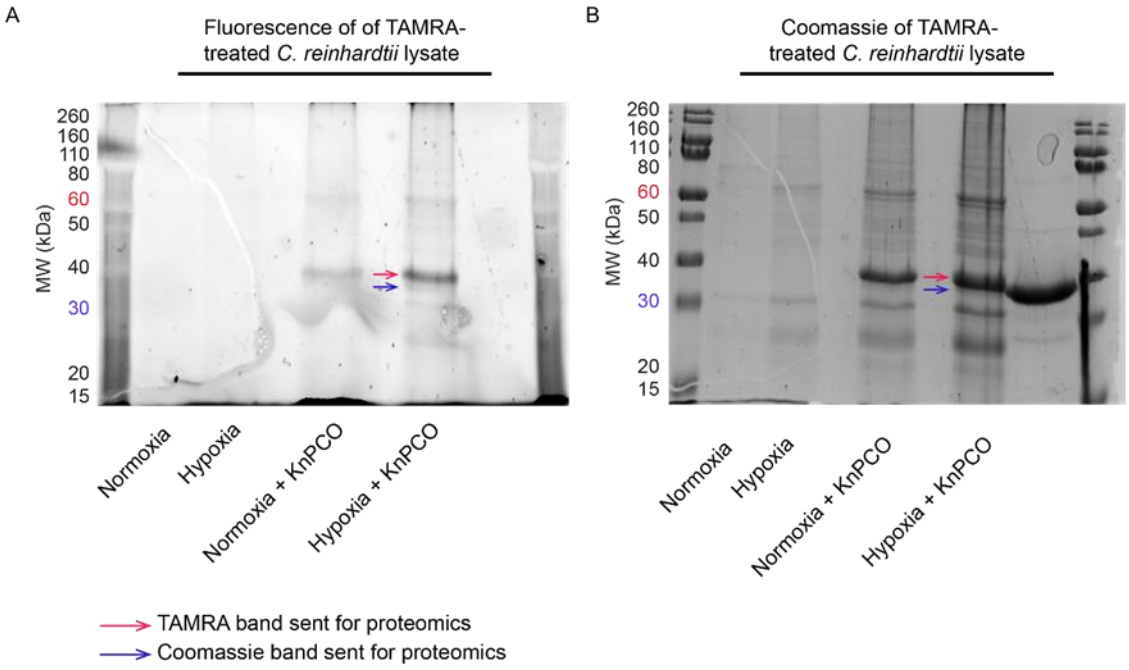
A) AtRAP2.12CysO_2H, + 32 Da



B) AtRAP2.12CysO_2H-NEM, + 32, +125 Da



A 6.3: Analysis of AtRAP2.12$_{2\text{-}15}$CysO_2H-NEM using nLC-MS. MS analysis by E. Pires confirmed the presence of a peak consistent with the addition of the NEM moiety to AtRAP2.12$_{2\text{-}15}$CysO_2H. Concentrations of the AtRAP2.12$_{2\text{-}15}$CysO_2H-NEM molecule were too low to identify the CysO_2H-NEM fragment.



A. 6.4: Experiment A: Low resolution detection of $CysO_2H$ with 5-TAMRA in *C. reinhardtii* cells exposed to hypoxia. A) A clear increase in the intensity of the 5-TAMRA fluorescent band is observed at approximately 35 kDa in the hypoxia/KnPCO treated *C.reinhardtii* lysate, compared to the normoxia/KnPCO treated *C.reinhardtii* lysate B) the same gel, but stained using Coomassie general protein stain. The TAMRA and Coomassie bands were sent separately for proteomics, as in the high resolution gel, but it was thought this gel was not as well resolved.

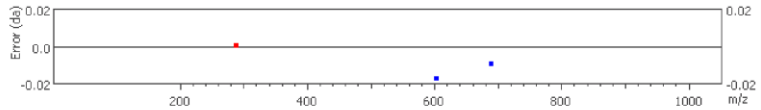
A 6.5 – 6.10: See digital files

Low Resolution SDS-PAGE: Mass spectrometry identifies two MC-initiating proteins in the excised hypoxia TAMRA fluorescent band

A) A4GRC6 (HAP2)

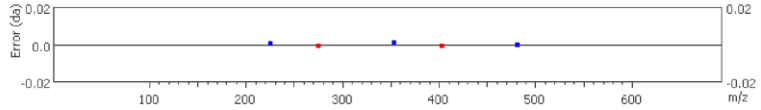
Unique Peptide Identifer 1 (-10lgP = 22.54):

#	b	Seq	y	#
1	58.03	G		11
2	157.10	V	1035.49	10
3	286.14	E	936.42	9
4	343.16	G	807.38	8
5	456.25	L	750.36	7
6	603.30	M(+15.99)	637.27	6
7	690.32	S	490.24	5
8	804.36	N	403.20	4
9	861.38	G	289.16	3
10	918.40	G	232.14	2
11		R	175.12	1



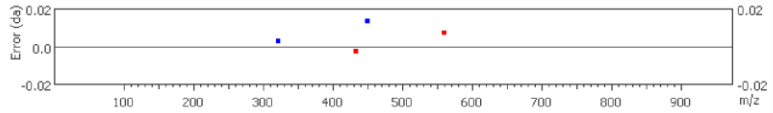
Unique Peptide Identifer 2 (-10lgP = 21.80):

#	b	Seq	y	#
1	98.06	P		5
2	226.12	Q	531.25	4
3	354.18	Q	403.19	3
4	482.24	Q	275.14	2
5		Q	147.08	1



Unique Peptide Identifer 3 (-10lgP = 19.59):

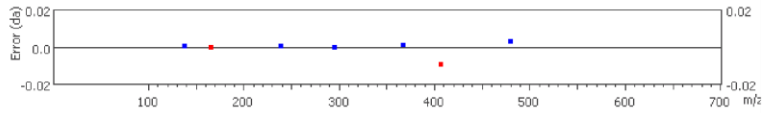
#	b	Seq	y	#
1	98.06	P		7
2	195.11	P	786.37	6
3	323.17	Q	689.32	5
4	451.22	Q	561.26	4
5	580.27	Q(+.98)	433.21	3
6	709.32	Q(+.98)	304.16	2
7		R	175.12	1



B) Q2HWK7 (CrDES)

Unique Peptide Identifer 1 (-10lgP = 27.73):

#	b	Seq	y	#
1	138.07	H		6
2	239.11	T	508.28	5
3	296.14	G	407.24	4
4	367.17	A	350.21	3
5	480.25	L	279.17	2
6		F	166.09	1



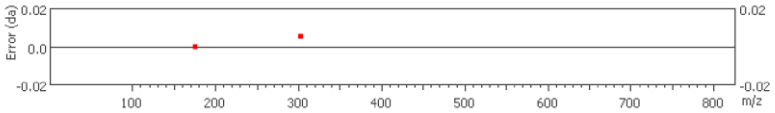
A 6.11: PEAKS identified two MC-initiating proteins, A4GRC6 (HAP2) and Q2HWK7 (CrDES), in the hypoxia 5-TAMRA band excised from the low resolution SDS-PAGE. The annotated mass spectrometry peaks for the A4GRC6 (HAP2) and Q2HWK7 (CrDES) unique identifier peptides.

Low Resolution SDS-PAGE: Mass spectrometry identifies three MC-initiating proteins in the excised hypoxia Coomassie band

A) A4GRC6 (HAP2)

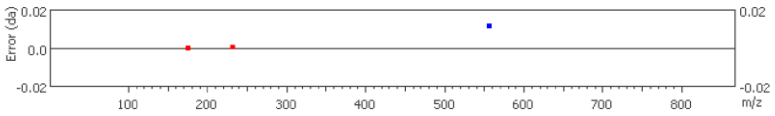
Unique Peptide Identifer 1 (-10lgP = 21.20):

#	b	Seq	y	#
1	148.04	M(+15.99)		5
2	334.12	W	602.34	4
3	447.21	L	416.26	3
4	575.27	Q	303.17	2
5		R	175.12	1



Unique Peptide Identifer 2 (-10lgP = 18.08):

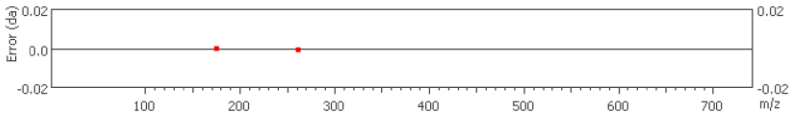
#	b	Seq	y	#
1	58.03	G		10
2	115.05	G	730.37	9
3	172.07	G	673.35	8
4	229.09	G	616.33	7
5	286.12	G	559.31	6
6	343.14	G	502.28	5
7	400.16	G	445.26	4
8	556.25	R	388.24	3
9	613.28	G	232.14	2
10		R	175.12	1



B) Q2HWK7 (CrDES)

Unique Peptide Identifer 1 (-10lgP = 21.13):

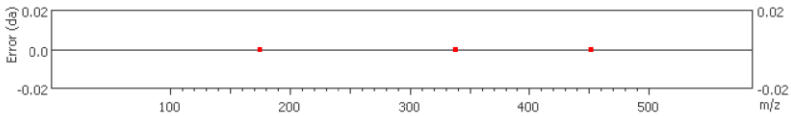
#	b	Seq	y	#
1	148.08	F		5
2	263.10	D	524.25	4
3	410.17	F	409.22	3
4	497.20	S	262.15	2
5		R	175.12	1



C) Q09JZ4 (ODA7)

Unique Peptide Identifer 1 (-10lgP = 16.57):

#	b	Seq	y	#
1	115.05	N		5
2	172.07	G	508.29	4
3	285.16	L	451.27	3
4	448.22	Y	338.18	2
5		R	175.12	1



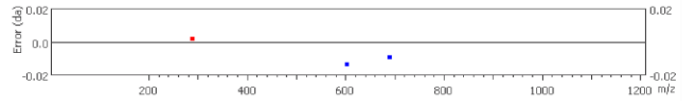
A 6.12: PEAKS identified three MC-initiating proteins, A4GRC6 (HAP2), Q2HWK7 (CrDES) and Q09JZ4 (ODA7) in the hypoxia Coomassie band excised from the low resolution SDS-PAGE. The annotated mass spectrometry peaks for the A4GRC6 (HAP2), Q2HWK7 (CrDES) and Q09JZ4 (ODA7) unique identifier peptides.

Low Resolution SDS-PAGE: Mass spectrometry identifies three MC-initiating proteins in the excised normoxia fluorescent band

A) A4GRC6 (HAP2)

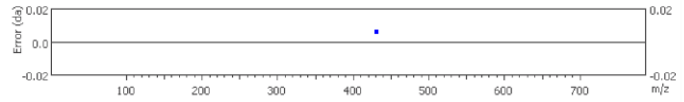
Unique Peptide Identifer 1 (-10lgP = 19.59):

#	b	Seq	y	#
1	58.03	G		11
2	157.10	V	1035.49	10
3	286.14	E	936.42	9
4	343.16	G	807.38	8
5	456.25	L	750.36	7
6	603.29	M(+15.99)	637.27	6
7	690.32	S	490.24	5
8	804.36	N	403.20	4
9	861.38	G	289.16	3
10	918.40	G	232.14	2
11		R	175.12	1



Unique Peptide Identifer 2 (-10lgP = 17.91):

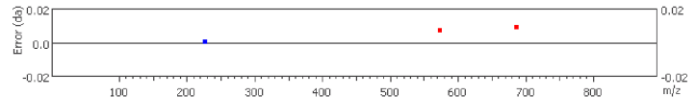
#	b	Seq	y	#
1	114.09	L		6
2	229.12	D	600.21	5
3	316.15	S	485.19	4
4	431.17	D	398.16	3
5	594.24	Y	283.13	2
6		T	120.07	1



B) Q2HWK7 (CrDES)

Unique Peptide Identifer 1 (-10lgP = 17.55):

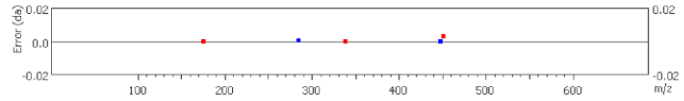
#	b	Seq	y	#
1	114.09	L		6
2	227.18	I	687.31	5
3	358.22	M	574.23	4
4	521.28	Y	443.19	3
5	635.32	N	280.13	2
6		F	166.09	1



C) Q09JZ4 (ODA7)

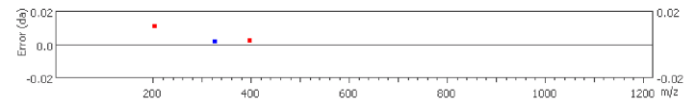
Unique Peptide Identifer 1 (-10lgP = 25.23):

#	b	Seq	y	#
1	115.05	N		5
2	172.07	G	508.29	4
3	285.16	L	451.26	3
4	448.22	Y	338.18	2
5		R	175.12	1

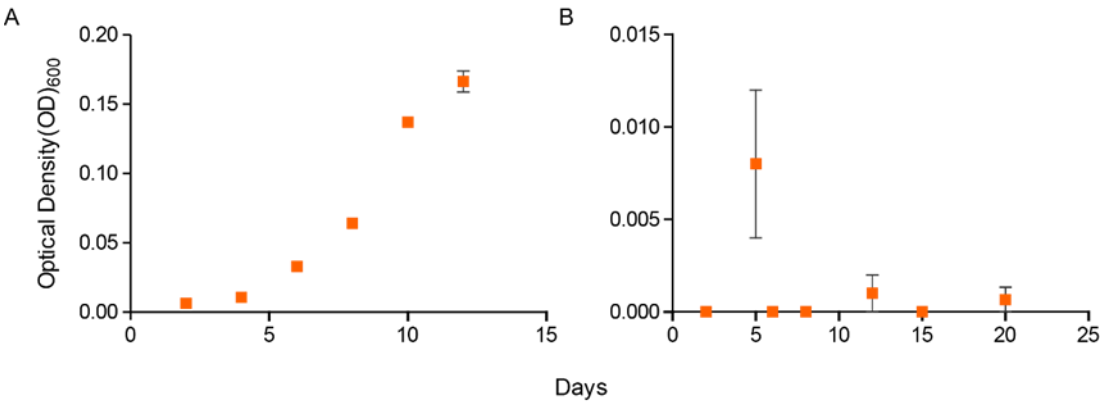


Unique Peptide Identifer 1 (-10lgP = 17.66):

#	b	Seq	y	#
1	115.05	N		13
2	214.12	V	1251.72	12
3	327.20	L	1152.65	11
4	456.25	E	1039.57	10
5	557.29	T	910.52	9
6	670.38	L	809.48	8
7	799.42	E	696.39	7
8	856.44	G	567.35	6
9	969.53	L	510.33	5
10	1066.58	P	397.24	4
11	1163.63	P	300.19	3
12	1276.72	L	203.13	2
13		A	90.05	1

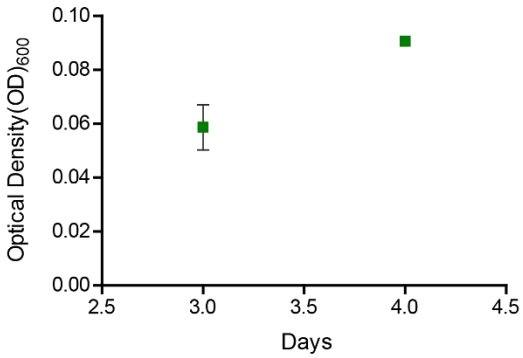


A 6.13: PEAKS identified three MC-initiating proteins, A4GRC6 (HAP2), Q2HWK7 (CrDES) and Q09JZ4 (ODA7) in the normoxia Coomassie band excised from the low resolution SDS-PAGE. The annotated mass spectrometry peaks for the A4GRC6 (HAP2), Q2HWK7 (CrDES) and Q09JZ4 (ODA7) unique identifier peptides.

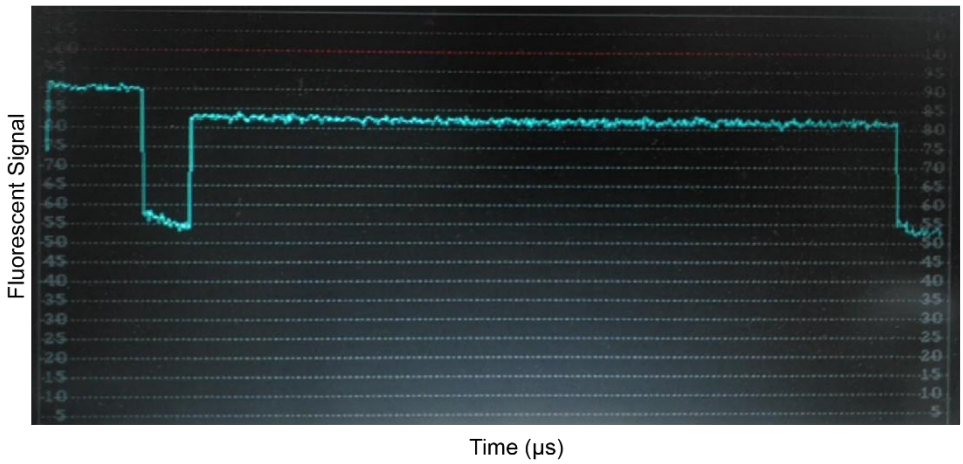


A 7.1: Growth of *E. huxleyi* cells under 10% O_2 in an Invivo$_2$400 anaerobic chamber. A) The first growth curve of *E. huxleyi* cells at 10% O_2 was successful, but a power cut near day 10 necessitated a repeat of the curve. B) the second culturing attempt was not successful, with the *E. huxleyi* cells dying.

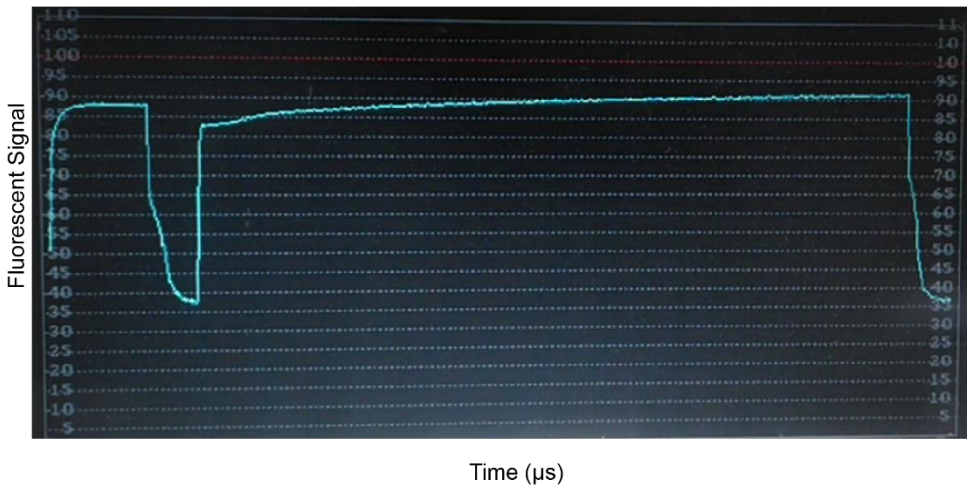
A



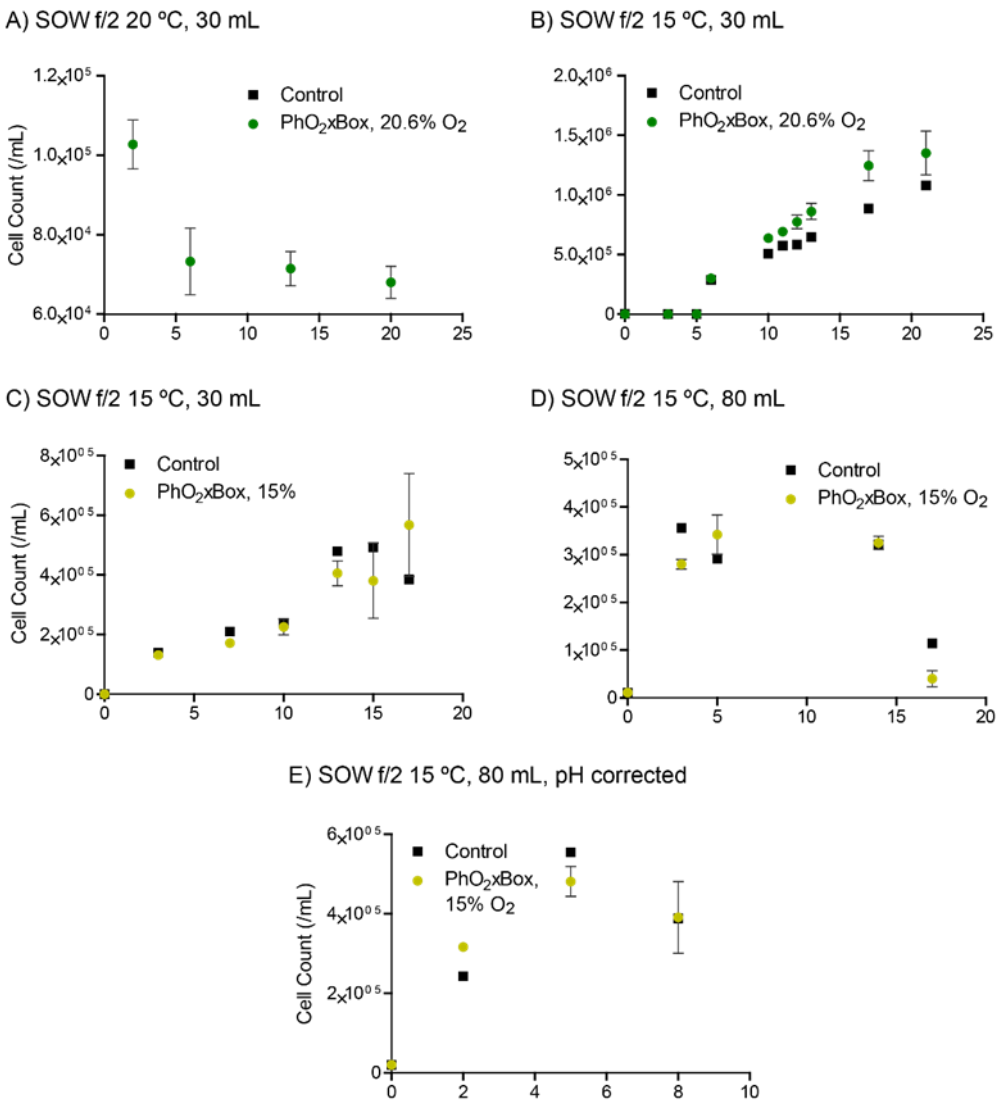
B



C

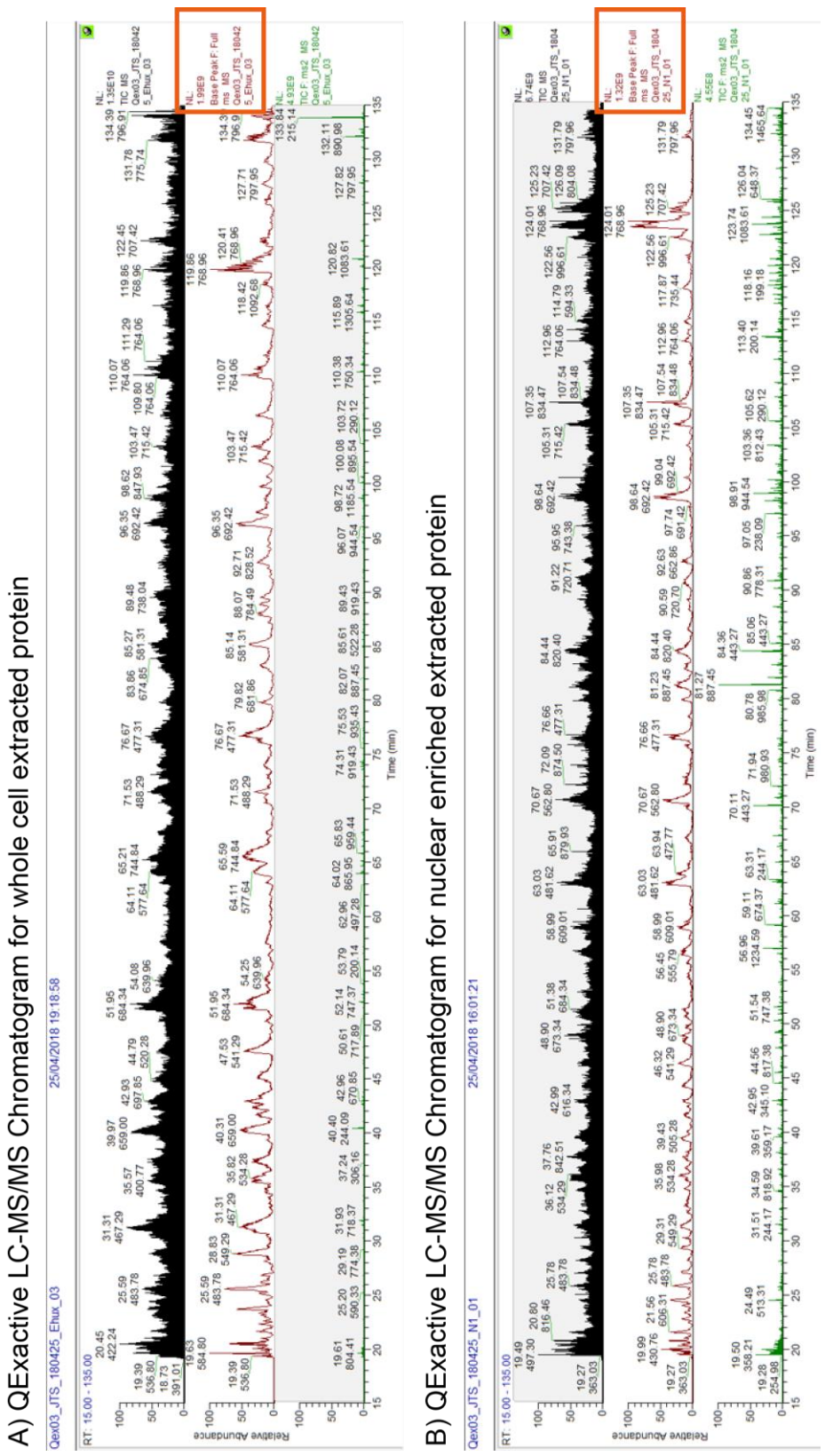


A 7.2: Growth of *E. huxleyi* cells under 20.6% O_2 in an Invivo$_2$400 anaerobic chamber. The Satlantic FIRe (Fluorescence Induction and Relaxation) System was used to detect chlorophyll fluorescence yields, and can be used to evaluate the health of photosynthesizing cells. A) OD600 measurements suggested *E. huxleyi* cell growth between day 3 and 4. B) analysis of the chlorophyll fluorescence yield on the Satlantic FIRe System suggested they were not photosynthesizing compared to a C) control growth in the Rickaby Geobiology Lab.

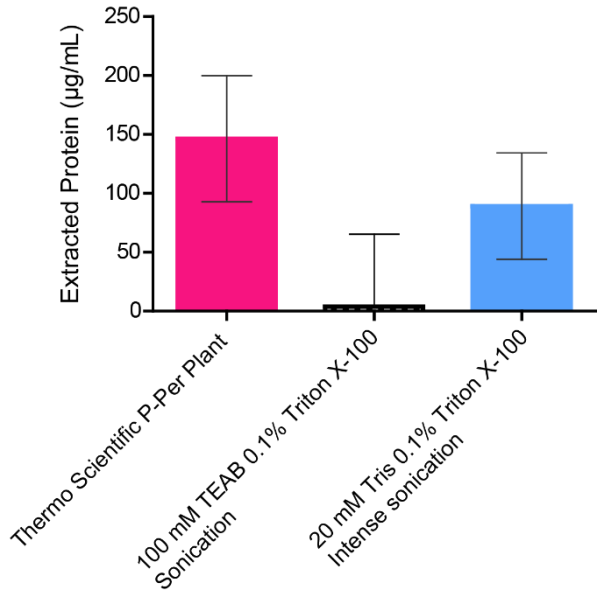


A 7.3: Growing *E. huxleyi in* SOW f/2 under 15 and 20.6% O_2 using a PhO_2xBox. A) initial growths of *E. huxleyi* in synthetic ocean water (SOW) f/2 at 20 °C were unsuccessful. B) *E. huxleyi* cells grew successfully in SOW f/2 at 15 °C, 20.6% O_2 C) and in SOW f/2 at 15% O_2 D) in an effort to increase biomass, the volume of the culture was increased to 80 mL, but the cells failed to grow. E) the pH of the growth media was a little lower than desired (target pH 8.1-8.2), but correction of the discrepancy did not improve growth outcomes

A 7.4 see supplementary digital files



A 7.5: Chromatogram of extracted protein on the QExactive LC-MS/MS. A) The baseline chromatogram (in red) of the whole cell extracted protein has an intensity of 1.99 e9 B) with the nuclear enriched extraction baseline at 1.32 e9. The proteomic facility require an intensity of e9 or higher, so both extractions satisfy this requirement.



A 7.6: The quantities of extracted protein using different whole cell extraction techniques. This graph shows the yield of extracted protein from *E. huxleyi* cells using sonication and the Thermo Fisher P-PER kit.

A 7.7 see supplementary digital files

References

1. M. D. White, L. Dalle Carbonare, M. Lavilla Puerta, S. Iacopino, M. Edwards, K. Dunne, E. Pires, C. Levy, M. A. McDonough, F. Licausi and E. Flashman, 2020, **117**, 23140-23147.
2. J. T. van Dongen and F. Licausi, *Annual Review of Plant Biology, Vol 66*, 2015, **66**, 345-367.
3. E. F. Garman and G. W. Grime, *Prog Biophys Mol Bio*, 2005, **89**, 173-205.
4. J. Ohlrogge and J. Browse, *Plant Cell*, 1995, **7**, 957-970.
5. J. L. Harwood, *Bba-Lipid Lipid Met*, 1996, **1301**, 7-56.
6. J. Okuley, J. Lightner, K. Feldmann, N. Yadav, E. Lark and J. Browse, *Plant Cell*, 1994, **6**, 147-158.
7. H. Weber, *Trends Plant Sci*, 2002, **7**, 217-224.
8. K. L. C. Wang, H. Li and J. R. Ecker, *Plant Cell*, 2002, **14**, S131-S151.
9. S. H. Schwartz, B. C. Tan, D. A. Gage, J. A. D. Zeevaart and D. R. McCarty, *Science*, 1997, **276**, 1872-1874.
10. S. Yamaguchi, *Annu Rev Plant Biol*, 2008, **59**, 225-251.
11. K. Apel and H. Hirt, *Annu Rev Plant Biol*, 2004, **55**, 373-399.
12. H. Tsukagoshi, W. Busch and P. N. Benfey, *Cell*, 2010, **143**, 606-616.
13. N. Suzuki, G. Miller, C. Salazar, H. A. Mondal, E. Shulaev, D. F. Cortes, J. L. Shuman, X. Z. Luo, J. Shah, K. Schlauch, V. Shulaev and R. Mittler, *Plant Cell*, 2013, **25**, 3553-3569.
14. R. Mittler, S. Vanderauwera, N. Suzuki, G. Miller, V. B. Tognetti, K. Vandepoele, M. Gollery, V. Shulaev and F. Van Breusegem, *Trends Plant Sci*, 2011, **16**, 300-309.
15. G. Miller, K. Schlauch, R. Tam, D. Cortes, M. A. Torres, V. Shulaev, J. L. Dangl and R. Mittler, *Sci Signal*, 2009, **2**.
16. Q. T. Ho, P. Verboven, B. E. Verlinden, E. Herremans, M. Wevers, J. Carmeliet and B. M. Nicolai, *Plant Physiology*, 2011, **155**, 1158-1168.
17. J. Armstrong, R. E. Jones and W. Armstrong, *New Phytol*, 2006, **172**, 719-731.
18. B. Biais, B. Beauvoit, J. W. Allwood, C. Deborde, M. Maucourt, R. Goodacre, D. Rolin and A. Moing, *J Plant Physiol*, 2010, **167**, 242-245.
19. W. Armstrong, M. E. Strange, S. Cringle and P. M. Beckett, *Ann Bot-London*, 1994, **74**, 287-299.
20. J. T. van Dongen, U. Schurr, M. Pfister and P. Geigenberger, *Plant Physiology*, 2003, **131**, 1529-1543.
21. L. Borisjuk and H. Rolletschek, *New Phytol*, 2009, **182**, 17-30.
22. D. A. Weits, J. T. van Dongen and F. Licausi, *New Phytol*, 2020, DOI: 10.1111/nph.16424.
23. T. Kelliher and V. Walbot, *Plant J*, 2014, **77**, 639-652.
24. D. A. Weits, A. B. Kunkowska, N. C. W. Kamps, K. M. S. Portz, N. K. Packbier, Z. N. Venza, C. Gaillochet, J. U. Lohmann, O. Pedersen, J. T. van Dongen and F. Licausi, *Nature*, 2019, **569**, 714-+.

25. V. Shukla, L. Lombardi, S. Lacopino, A. Pencik, O. Novak, P. Perata, B. Giuntoli and F. Licausi, *Mol Plant*, 2019, **12**, 538-551.
26. D. J. Gibbs, H. M. Tedds, A. M. Labandera, M. Bailey, M. D. White, S. Hartman, C. Sprigg, S. L. Mogg, R. Osborne, C. Dambire, T. Boeckx, Z. Paling, L. A. C. J. Voesenek, E. Flashman and M. J. Holdsworth, *Nature Communications*, 2018, **9**.
27. S. Wenkel, J. Emery, B. H. Hou, M. M. S. Evans and M. K. Barton, *Plant Cell*, 2007, **19**, 3379-3390.
28. Y. S. Kim, S. G. Kim, M. Lee, I. Lee, H. Y. Park, P. J. Seo, J. H. Jung, E. J. Kwon, S. W. Suh, K. H. Paek and C. M. Park, *Plant Cell*, 2008, **20**, 920-933.
29. F. D. Ariel, P. A. Manavella, C. A. Dezar and R. L. Chan, *Trends Plant Sci*, 2007, **12**, 419-426.
30. P. Roszak and C. Kohler, *P Natl Acad Sci USA*, 2011, **108**, 20826-20831.
31. M. Ikeuchi, A. Iwase, B. Rymen, H. Harashima, M. Shibata, M. Ohnuma, C. Breuer, A. K. Morao, M. de Lucas, L. De Veylder, J. Goodrich, S. M. Brady, F. Roudier and K. Sugimoto, *Nat Plants*, 2015, **1**, 1-7.
32. M. de Lucas, L. Pu, G. Turco, A. Gaudinier, A. K. Morao, H. Harashima, D. Kim, M. Ron, K. Sugimoto, F. Roudier and S. M. Brady, *Plant Cell*, 2016, **28**, 2616-2631.
33. A. R. Gendall, Y. Y. Levy, A. Wilson and C. Dean, *Cell*, 2001, **107**, 525-535.
34. F. Licausi, M. Kosmacz, D. A. Weits, B. Giuntoli, F. M. Giorgi, L. A. C. J. Voesenek, P. Perata and J. T. van Dongen, *Nature*, 2011, **479**, 419-422.
35. Robert B. Hamanaka, Samuel E. Weinberg, Colleen R. Reczek and Navdeep S. Chandel, *Cell Reports*, 2016, **15**, 451-459.
36. R. D. Guzy, M. M. Mack and P. T. Schumacker, *Antioxid Redox Sign*, 2007, **9**, 1317-1328.
37. R. Chang, C. J. H. Jang, C. Branco-Price, P. Nghiem and J. Bailey-Serres, *Plant Mol Biol*, 2012, **78**, 109-122.
38. C. Bowler, M. Vanmontagu and D. Inze, *Annu Rev Plant Phys*, 1992, **43**, 83-116.
39. I. E. Santosa, P. C. Ram, E. I. Boamfa, L. J. J. Laarhoven, J. Reuss, M. B. Jackson and F. J. M. Harren, *Planta*, 2007, **226**, 193-202.
40. R. R. Schmidt, D. A. Weits, C. F. J. Feulner and J. T. van Dongen, *Plant Physiology*, 2018, **176**, 1131-1142.
41. O. B. Blokhina, T. V. Chirkova and K. V. Fagerstedt, *J Exp Bot*, 2001, **52**, 1179-1190.
42. M. Skutnik and A. M. Rychter, *J Plant Physiol*, 2009, **166**, 926-937.
43. A. Baxter-Burrell, Z. B. Yang, P. S. Springer and J. Bailey-Serres, *Science*, 2002, **296**, 2026-2028.
44. S. Gilroy, N. Suzuki, G. Miller, W. G. Choi, M. Toyota, A. R. Devireddy and R. Mittler, *Trends Plant Sci*, 2014, **19**, 623-630.
45. J. C. Sedbrook, P. J. Kronebusch, G. G. Borisy, A. J. Trewavas and P. H. Masson, 1996, **111**, 243-257.
46. L. P. Hamel, M. C. Nicole, S. Sritubtim, M. J. Morency, M. Ellis, J. Ehlting, N. Beaudoin, B. Barbazuk, D. Klessig, J. Lee, G. Martin, J. Mundy, Y.

Ohashi, D. Scheel, J. Sheen, T. Xing, S. Q. Zhang, A. Seguin and B. E. Ellis, *Trends Plant Sci*, 2006, **11**, 192-198.
47. N. Foresi, N. Correa-Aragunde, G. Parisi, G. Calo, G. Salerno and L. Lamattina, *Plant Cell*, 2010, **22**, 3816-3830.
48. N. Foresi, M. L. Mayta, A. F. Lodeyro, D. Scuffi, N. Correa-Aragunde, C. Garcia-Mata, C. Casalongue, N. Carrillo and L. Lamattina, *Plant J*, 2015, **82**, 806-821.
49. H. Yamasaki and Y. Sakihama, *Febs Lett*, 2000, **468**, 89-92.
50. R. Desikan, R. Griffiths, J. Hancock and S. Neill, *P Natl Acad Sci USA*, 2002, **99**, 16314-16318.
51. A. H. Lin, Y. Q. Wang, J. Y. Tang, P. Xue, C. L. Li, L. C. Liu, B. Hu, F. Q. Yang, G. J. Loake and C. C. Chu, *Plant Physiology*, 2012, **158**, 451-464.
52. P. Rockel, F. Strube, A. Rockel, J. Wildt and W. M. Kaiser, *J Exp Bot*, 2002, **53**, 103-110.
53. A. U. Igamberdiev, C. Seregelyes, N. Manac'h and R. D. Hill, *Planta*, 2004, **219**, 95-102.
54. M. Perazzolli, P. Dominici, M. C. Romero-Puertas, E. Zago, J. Zeier, M. Sonoda, C. Lamb and M. Delledonne, *Plant Cell*, 2004, **16**, 2785-2794.
55. C. Dordas, J. Rivoal and R. D. Hill, *Ann Bot-London*, 2003, **91**, 173-178.
56. C. Dordas, B. B. Hasinoff, A. U. Igamberdiev, N. Manac'h, J. Rivoal and R. D. Hill, *Plant J*, 2003, **35**, 763-770.
57. S. Kundu, J. T. Trent and M. S. Hargrove, *Trends Plant Sci*, 2003, **8**, 387-393.
58. A. Igamberdiev, M. Stoimenova, C. Seregelyes and R. D. Hill, *Planta*, 2006, **223**, 1041-1046.
59. P. W. Hunt, E. J. Klok, B. Trevaskis, R. A. Watts, M. H. Ellis, W. J. Peacock and E. S. Dennis, *P Natl Acad Sci USA*, 2002, **99**, 17197-17202.
60. C. A. Appleby, D. Bogusz, E. S. Dennis and W. J. Peacock, *Plant Cell Environ*, 1988, **11**, 359-367.
61. R. Arredondo-Peter, M. S. Hargrove, J. F. Moran, G. Sarath and R. V. Klucas, *Plant Physiology*, 1998, **118**, 1121-1125.
62. H. C. Oliveira, L. Freschi and L. Sodek, *Plant Physiol Bioch*, 2013, **66**, 141-149.
63. H. C. Oliveira, I. Salgado and L. Sodek, *Planta*, 2013, **237**, 255-264.
64. M. T. Gladwin and S. Shiva, *Circ Res*, 2009, **104**, 1136-1138.
65. J. Feng, L. Chen and J. Zuo, *J Integr Plant Biol*, 2019, **61**, 1206-1223.
66. A. Besson-Bard, A. Pugin and D. Wendehenne, *Annu Rev Plant Biol*, 2008, **59**, 21-39.
67. D. J. Gibbs, N. M. Isa, M. Movahedi, J. Lozano-Juste, G. M. Mendiondo, S. Berckhan, N. Marin-de la Rosa, J. V. Conde, C. S. Correia, S. P. Pearce, G. W. Bassel, B. Hamali, P. Talloji, D. F. A. Tome, A. Coego, J. Beynon, D. Alabadi, A. Bachmair, J. Leon, J. E. Gray, F. L. Theodoulou and M. J. Holdsworth, *Mol Cell*, 2014, **53**, 369-379.
68. M. L. Lanteri, A. M. Laxalt and L. Lamattina, *Plant Physiology*, 2008, **147**, 188-198.

69. L. A. C. J. Voesenek and J. Bailey-Serres, *New Phytol*, 2015, **206**, 57-73.
70. Y. Hattori, K. Nagai, S. Furukawa, X.-J. Song, R. Kawano, H. Sakakibara, J. Wu, T. Matsumoto, A. Yoshimura, H. Kitano, M. Matsuoka, H. Mori and M. Ashikari, *Nature*, 2009, **460**, 1026-1030.
71. H. Mergemann and M. Sauter, *Plant Physiology*, 2000, **124**, 609-614.
72. T.-W. Wang and R. N. Arteca, 1992, **98**, 97-100.
73. O. Y. Shiu, J. H. Oetiker, W. K. Yip and S. F. Yang, 1998, **95**, 10334-10339.
74. T. I. Zarembinski and A. Theologis, 1993, **4**, 363-373.
75. M. C. Drew, C. J. He and P. W. Morgan, *Trends Plant Sci*, 2000, **5**, 123-127.
76. A. H. L. A. N. Gunawardena, D. M. Pearce, M. B. Jackson, C. R. Hawes and D. E. Evans, *Planta*, 2001, **212**, 205-214.
77. S. Neill, R. Desikan and J. Hancock, *Curr Opin Plant Biol*, 2002, **5**, 388-395.
78. A. A. Ludwig, H. Saitoh, G. Felix, G. Freymark, O. Miersch, C. Wasternack, T. Boller, J. D. G. Jones and T. Romeis, *P Natl Acad Sci USA*, 2005, **102**, 10736-10741.
79. M. Hinz, I. W. Wilson, J. Yang, K. Buerstenbinder, D. Llewellyn, E. S. Dennis, M. Sauter and R. Dolferus, *Plant Physiology*, 2010, **153**, 757-772.
80. M. Buttner and K. B. Singh, *P Natl Acad Sci USA*, 1997, **94**, 5961-5966.
81. N. Hess, M. Klode, M. Anders and M. Sauter, *Physiol Plantarum*, 2011, **143**, 41-49.
82. F. Licausi, J. T. Van Dongen, B. Giuntoli, G. Novi, A. Santaniello, P. Geigenberger and P. Perata, *The Plant Journal*, 2010, **62**, 302-315.
83. K. Xu, X. Xu, T. Fukao, P. Canlas, R. Maghirang-Rodriguez, S. Heuer, A. M. Ismail, J. Bailey-Serres, P. C. Ronald and D. J. Mackill, *Nature*, 2006, **442**, 705-708.
84. C. Papdi, E. Ábrahám, M. P. Joseph, C. Popescu, C. Koncz and L. Szabados, *Plant Physiology*, 2008, **147**, 528-542.
85. L. A. C. J. Voesenek and R. Sasidharan, *Plant Biology*, 2013, **15**, 426-435.
86. S. Hartman, Z. G. Liu, H. van Veen, J. Vicente, E. Reinen, S. Martopawiro, H. T. Zhang, N. van Dongen, F. Bosman, G. W. Bassel, E. J. W. Visser, J. Bailey-Serres, F. L. Theodoulou, K. H. Hebelstrup, D. J. Gibbs, M. J. Holdsworth, R. Sasidharan and L. A. C. J. Voesenek, *Nature Communications*, 2019, **10**.
87. S. C. Lee, A. Mustroph, R. Sasidharan, D. Vashisht, O. Pedersen, T. Oosumi, L. A. C. J. Voesenek and J. Bailey-Serres, *New Phytol*, 2011, **190**, 457-471.
88. W. Armstrong, in *Advances in Botanical Research*, ed. H. W. Woolhouse, Academic Press, 1980, vol. 7, pp. 225-332.
89. J. Kreuzwieser and H. Rennenberg, *Plant Cell Environ*, 2014, **37**, 2245-2259.
90. P. Geigenberger, *Curr Opin Plant Biol*, 2003, **6**, 247-256.
91. M. C. T. Trought and M. C. Drew, *Plant and Soil*, 1980, **54**, 77-94.
92. T. Fukao and J. Bailey-Serres, *Trends Plant Sci*, 2004, **9**, 449-456.

93. J. Gibbs and H. Greenway, *Functional Plant Biology*, 2003, **30**, 1-47.
94. I. M. Unger, A. C. Kennedy and R. M. Muzika, *Appl Soil Ecol*, 2009, **42**, 1-8.
95. J. Bailey-Serres and L. A. C. J. Voesenek, *Annu Rev Plant Biol*, 2008, **59**, 313-339.
96. L. Lamers, L. Govers, I. Janssen, J. Geurts, M. Van der Welle, M. Van Katwijk, T. Van der Heide, J. Roelofs and A. Smolders, 2013, **4**.
97. F. R. Zeng, L. Shabala, M. X. Zhou, G. P. Zhang and S. Shabala, *Front Plant Sci*, 2013, **4**.
98. J. K. Polko, L. A. C. J. Voesenek, A. J. M. Peeters and R. Pierik, *Aob Plants*, 2011.
99. M. B. Jackson and W. Armstrong, *Plant Biology*, 1999, **1**, 274-287.
100. T. D. Colmer, *Plant Cell Environ*, 2003, **26**, 17-36.
101. M. Sauter, *Curr Opin Plant Biol*, 2013, **16**, 282-286.
102. T. D. Colmer and O. Pedersen, *New Phytol*, 2008, **177**, 918-926.
103. L. Mommer and E. J. W. Visser, *Ann Bot-London*, 2005, **96**, 581-589.
104. T. Fukao, K. N. Xu, P. C. Ronald and J. Bailey-Serres, *Plant Cell*, 2006, **18**, 2021-2034.
105. T. Fukao and J. Bailey-Serres, *P Natl Acad Sci USA*, 2008, **105**, 16814-16819.
106. T. Fukao, E. Yeung and J. Bailey-Serres, *Plant Cell*, 2011, **23**, 412-427.
107. K. R. Hultine, S. E. Bush, A. G. West and J. R. Ehleringer, *Oecologia*, 2007, **154**, 85-93.
108. J. L. Nielsen, S. B. Rood, D. W. Pearce, M. G. Letts and H. Jiskoot, *Tree Physiol*, 2010, **30**, 1479-1488.
109. H. van Veen, A. Mustroph, G. A. Barding, M. Vergeer-van Eijk, R. A. M. Welschen-Evertman, O. Pedersen, E. J. W. Visser, C. K. Larive, R. Pierik, J. Bailey-Serres, L. A. C. J. Voesenek and R. Sasidharan, *Plant Cell*, 2013, **25**, 4691-4707.
110. J. Gibbs, S. Morrell, A. Valdez, T. L. Setter and T. Greenway, *J Exp Bot*, 2000, **51**, 785-796.
111. E. C. Sisler and M. Serek, *Physiol Plantarum*, 1997, **100**, 577-582.
112. A. Mustroph, M. E. Zanetti, C. J. H. Jang, H. E. Holtan, P. P. Repetti, D. W. Galbraith, T. Girke and J. Bailey-Serres, *P Natl Acad Sci USA*, 2009, **106**, 18843-18848.
113. C. Branco-Price, K. A. Kaiser, C. J. H. Jang, C. K. Larive and J. Bailey-Serres, *Plant J*, 2008, **56**, 743-755.
114. B. Giuntoli, S. C. Lee, F. Licausi, M. Kosmacz, T. Oosumi, J. T. van Dongen, J. Bailey-Serres and P. Perata, *PLoS biology*, 2014, **12**.
115. E. K. Schrader, K. G. Harstad and A. Matouschek, *Nat Chem Biol*, 2009, **5**, 815-822.
116. A. Varshavsky, *Protein Sci*, 2011, **20**, 1298-1345.
117. A. Bachmair, D. Finley and A. Varshavsky, *Science*, 1986, **234**, 179-186.
118. T. Tasaki, S. M. Sriram, K. S. Park and Y. T. Kwon, *Annu Rev Biochem*, 2012, **81**, 261-289.

119. C. S. Hwang, A. Shemorry and A. Varshavsky, *Science*, 2010, **327**, 973-977.
120. K. T. Nguyen, S. H. Mun, C. S. Lee and C. S. Hwang, *Exp Mol Med*, 2018, **50**.
121. R. A. Bradshaw, W. W. Brickey and K. W. Walker, *Trends in biochemical sciences*, 1998, **23**, 263-267.
122. H. Rao, F. Uhlmann, K. Nasmyth and A. Varshavsky, *Nature*, 2001, **410**, 955-959.
123. A. Varshavsky, *Nat Cell Biol*, 2003, **5**, 373-376.
124. J. Y. An, J. W. Seo, T. Tasaki, M. J. Lee, A. Varshavsky and Y. T. Kwon, *P Natl Acad Sci USA*, 2006, **103**, 6212-6217.
125. M. Garzon, K. Eifler, A. Faust, H. Scheel, K. Hofmann, C. Koncz, A. Yephremov and A. Bachmair, *Febs Lett*, 2007, **581**, 3189-3196.
126. S. Stary, X. J. Yin, T. Potuschak, P. Schlogelhofer, V. Nizhynska and A. Bachmair, *Plant Physiology*, 2003, **133**, 1360-1366.
127. S. M. Sriram and Y. T. Kwon, *Nat Struct Mol Biol*, 2010, **17**, 1164-1165.
128. A. Mogk, R. Schmidt and B. Bukau, *Trends Cell Biol*, 2007, **17**, 165-172.
129. J. S. Park, J. Y. Lee, Y. T. K. Nguyen, N. W. Kang, E. K. Oh, D. M. Jang, H. J. Kim, D. D. Kim and B. W. Han, *Biomolecules*, 2020, **10**.
130. H. Q. Wang, K. I. Piatkov, C. S. Brower and A. Varshavsky, *Mol Cell*, 2009, **34**, 686-695.
131. S. M. Sriram, B. Y. Kim and Y. T. Kwon, *Nat Rev Mol Cell Bio*, 2011, **12**, 735-747.
132. M. J. Lee, T. Tasaki, K. Moroi, J. Y. An, S. Kimura, I. V. Davydov and Y. T. Kwon, *P Natl Acad Sci USA*, 2005, **102**, 15030-15035.
133. D. A. Weits, B. Giuntoli, M. Kosmacz, S. Parlanti, H.-M. Hubberten, H. Riegler, R. Hoefgen, P. Perata, J. T. van Dongen and F. Licausi, *Nat Commun*, 2014, **5**.
134. E. Graciet, F. Mesiti and F. Wellmer, *The Plant Journal*, 2010, **61**, 741-751.
135. J. L. Riechmann, J. Heard, G. Martin, L. Reuber, C. Z. Jiang, J. Keddie, L. Adam, O. Pineda, O. J. Ratcliffe, R. R. Samaha, R. Creelman, M. Pilgrim, P. Broun, J. Z. Zhang, D. Ghandehari, B. K. Sherman and C. L. Yu, *Science*, 2000, **290**, 2105-2110.
136. T. Nakano, K. Suzuki, T. Fujimura and H. Shinshi, *Plant Physiology*, 2006, **140**, 411-432.
137. B. Tournier, M. T. Sanchez-Ballesta, B. Jones, E. Pesquet, F. Regad, A. Latche, J. C. Pech and M. Bouzayen, *Febs Lett*, 2003, **550**, 149-154.
138. D. J. Gibbs, S. C. Lee, N. M. Isa, S. Gramuglia, T. Fukao, G. W. Bassel, C. S. Correia, F. Corbineau, F. L. Theodoulou, J. Bailey-Serres and M. J. Holdsworth, *Nature*, 2011, **479**, 415-U172.
139. H. van Veen, M. Akman, D. C. L. Jamar, D. Vreugdenhil, M. Kooiker, P. van Tienderen, L. A. C. J. Voesenek, M. E. Schranz and R. Sasidharan, *Plant Cell Environ*, 2014, **37**, 2421-2432.

140. J. Bailey-Serres, T. Fukao, D. J. Gibbs, M. J. Holdsworth, S. C. Lee, F. Licausi, P. Perata, L. A. C. J. Voesenek and J. T. van Dongen, *Trends Plant Sci*, 2012, **17**, 129-138.
141. C. C. Lin, Y. T. Chao, W. C. Chen, H. Y. Ho, M. Y. Chou, Y. R. Li, Y. L. Wu, H. A. Yang, H. Hsieh, C. S. Lin, F. H. Wu, S. J. Chou, H. C. Jen, Y. H. Huang, D. Irene, W. J. Wu, J. L. Wu, D. J. Gibbs, M. C. Ho and M. C. Shih, *P Natl Acad Sci USA*, 2019, **116**, 3300-3309.
142. M. D. White, M. Klecker, R. J. Hopkinson, D. A. Weits, C. Mueller, C. Naumann, R. O'Neill, J. Wickens, J. Y. Yang, J. C. Brooks-Bartlett, E. F. Garman, T. N. Grossmann, N. Dissmeyer and E. Flashman, *Nature Communications*, 2017, **8**.
143. H. Y. Li and M. L. Chye, *Plant Mol Biol*, 2003, **51**, 483-492.
144. M. Kosmacz, S. Parlanti, M. Schwarzlander, F. Kragler, F. Licausi and J. T. Van Dongen, *Plant Cell Environ*, 2015, **38**, 1094-1103.
145. R. R. Schmidt, M. Fulda, M. V. Paul, M. Anders, F. Plum, D. A. Weits, M. Kosmacz, T. R. Larson, I. A. Graham, G. T. S. Beemster, F. Licausi, P. Geigenberger, J. H. Schippers and J. T. van Dongen, *P Natl Acad Sci USA*, 2018, **115**, E12101-E12110.
146. D. A. Weits, L. Zhou, B. Giuntoli, L. D. Carbonare, S. Iacopino, L. Piccinini, V. Shukla, L. T. Bui, G. Novi, J. T. van Dongen and F. Licausi, 2020, DOI: Pre-print at: https://doi.org/10.1101/2020.06.24.169417.
147. K. P. Mccann, M. T. Akbari, A. C. Williams and D. B. Ramsden, *Bba-Protein Struct M*, 1994, **1209**, 107-110.
148. J. Dominy, C. R. Simmons, L. L. Hirschberger, J. Hwang, R. M. Coloso and M. H. Stipanuk, *J Biol Chem*, 2007, **282**, 25189-25198.
149. C. R. Simmons, L. L. Hirschberger, M. S. Machi and M. H. Stipanuk, *Protein Expres Purif*, 2006, **47**, 74-81.
150. N. Masson, T. P. Keeley, B. Giuntoli, M. D. White, M. L. Puerta, P. Perata, R. J. Hopkinson, E. Flashman, F. Licausi and P. J. Ratcliffe, *Science*, 2019, **365**, 65-+.
151. F. J. Giordano, *Arteriosclerosis, Thrombosis, and Vascular Biology*, 2010, **30**, 641-642.
152. P. Bert, *C R Acad Sci Paris* 1882, **94**, 805–807.
153. K. S. Hewitson and C. J. Schofield, *Drug Discov Today*, 2004, **9**, 704-711.
154. J. Aragones, P. Fraisl, M. Baes and P. Carmeliet, *Cell Metab*, 2009, **9**, 11-22.
155. R. K. Bruick and S. L. McKnight, *Science*, 2001, **294**, 1337-1340.
156. D. Lando, D. J. Peet, D. A. Whelan, J. J. Gorman and M. L. Whitelaw, *Science*, 2002, **295**, 858-861.
157. M. Sugahara, S. Tanaka, T. Tanaka, H. Saito, Y. Ishimoto, T. Wakashima, M. Ueda, K. Fukui, A. Shimizu, R. Inagi, T. Yamauchi, T. Kadowaki and M. Nangaku, *J Am Soc Nephrol*, 2020, **31**, 560-577.
158. C. Loinard, E. Berra, J. Vilar, A. Gmoucve, C. Cochan, Y. Zouggari, A. Recalde, M. Duriez, B. Levy, P. Jacques and J. S. Silvestre, *Circulation*, 2008, **118**, S550-S550.

159. E. Flashman, L. M. Hoffart, R. B. Hamed, J. M. Bollinger, C. Krebs and C. J. Schofield, *Febs J*, 2010, **277**, 4089-4099.
160. J. Hamzah, M. Jugold, F. Kiessling, P. Rigby, M. Manzur, H. H. Marti, T. Rabie, S. Kaden, H. J. Grone, G. J. Hammerling, B. Arnold and R. Ganss, *Nature*, 2008, **453**, 410-U467.
161. C. Arnold, E. Demirel, A. Feldner, G. Genove, H. J. Zhang, C. Sticht, T. Wieland, M. Hecker, S. Heximer and T. Korff, *Faseb J*, 2018, **32**, 2021-2035.
162. Y. T. Kwon, A. S. Kashina, I. V. Davydov, R. G. Hu, J. Y. An, J. W. Seo, F. Du and A. Varshavsky, *Science*, 2002, **297**, 96-99.
163. J. T. Hong, D. J. Son, C. K. Lee, D. Y. Yoon, D. H. Lee and M. H. Park, *Pharmacol Therapeut*, 2017, **174**, 127-137.
164. M. B. Khawar, M. H. Abbasi and N. Sheikh, *Mediat Inflamm*, 2016, **2016**.
165. T. Xin, M. Chen, L. W. Duan, Y. L. Xu and P. Gao, *Resp Res*, 2018, **19**.
166. S. W. Z. Olechnowicz, A. O. Fedele and D. J. Peet, *Plos One*, 2012, **7**.
167. Y. Jin, X. J. An, Z. L. Ye, B. Cully, J. P. Wu and J. Li, *J Biol Chem*, 2009, **284**, 23436-23443.
168. M. Zahoor, M. Westhrin, K. R. Aass, S. H. Moen, K. Misund, K. M. Psonka-Antonczyk, M. Giliberto, G. Buene, A. Sundan, A. Waage, A. M. Sponaas and T. Standal, *Blood Adv*, 2017, **1**, 2656-2666.
169. E. P. Tchesnokov, M. Fellner, E. Siakkou, T. Kleffmann, L. W. Martin, S. Aloi, I. L. Lamont, S. M. Wilbanks and G. N. L. Jameson, *J Biol Chem*, 2015, **290**, 24424-24437.
170. U. Brandt, M. Schurmann and A. Steinbuchel, *J Biol Chem*, 2014, **289**, 30800-30809.
171. A. Kasperova, J. Kunert, M. Horynova, E. Weigl, M. Sebela, R. Lenobel and M. Raska, *Mycoses*, 2011, **54**, E456-E462.
172. M. H. Stipanuk, C. R. Simmons, P. A. Karplus and J. E. Dominy, *Amino Acids*, 2011, **41**, 91-102.
173. M. H. Stipanuk, I. Ueki, J. E. Dominy, C. R. Simmons and L. L. Hirschberger, *Amino Acids*, 2009, **37**, 55-63.
174. K. Yamaguchi and Y. Hosokawa, *Methods in enzymology*, 1987, **143**, 395-403.
175. Y. H. Kwon and M. H. Stipanuk, 2001, **280**, E804-E815.
176. J. M. Dunwell, A. Culham, C. E. Carter, C. R. Sosa-Aguirre and P. W. Goodenough, *Trends in biochemical sciences*, 2001, **26**, 740-746.
177. S. Aloi, C. G. Davies, P. A. Karplus, S. M. Wilbanks and G. N. L. Jameson, *Biochemistry-Us*, 2019, **58**, 2398-2407.
178. S. Ye, X. Wu, L. Wei, D. M. Tang, P. Sun, M. Bartlam and Z. H. Rao, *J Biol Chem*, 2007, **282**, 3391-3402.
179. J. E. Dominy, C. R. Simmons, P. A. Karplus, A. M. Gehring and M. H. Stipanuk, *J Bacteriol*, 2006, **188**, 5561-5569.
180. C. R. Simmons, Q. Liu, Q. Q. Huang, Q. Hao, T. P. Begley, P. A. Karplus and M. H. Stipanuk, *J Biol Chem*, 2006, **281**, 18723-18733.

181. D. A. Proshlyakov, M. A. Pressler and G. T. Babcock, *P Natl Acad Sci USA*, 1998, **95**, 8020-8025.
182. J. G. McCoy, L. J. Bailey, E. Bitto, C. A. Bingman, D. J. Aceti, B. G. Fox and G. N. Phillips, *P Natl Acad Sci USA*, 2006, **103**, 3084-3089.
183. J. D. Gardner, B. S. Pierce, B. G. Fox and T. C. Brunold, *Biochemistry-Us*, 2010, **49**, 6033-6041.
184. C. R. Simmons, K. Krishnamoorthy, S. L. Granett, D. J. Schuller, J. E. Dominy, T. P. Begley, M. H. Stipanuk and P. A. Karplus, *Biochemistry-Us*, 2008, **47**, 11390-11392.
185. D. A. Proshlyakov, M. A. Pressler and G. T. Babcock, *J Inorg Biochem*, 1999, **74**, 267-267.
186. Y. Wang, I. Davis, Y. Chen, S. G. Naik, W. P. Griffith and A. Liu, *J Biol Chem*, 2020, DOI: 10.1074/jbc.RA120.013915.
187. Y. F. Wang, W. P. Griffith, J. S. Li, T. Koto, D. J. Wherritt, E. Fritz and A. M. Liu, *Angew Chem Int Edit*, 2018, **57**, 8149-8153.
188. M. D. White, J. J. A. G. Kamps, S. East, L. J. T. Kearney and E. Flashman, *J Biol Chem*, 2018, **293**, 11786-11795.
189. A. A. Pavlov and J. F. Kasting, *Astrobiology*, 2002, **2**, 27-41.
190. J. Farquhar, H. M. Bao and M. Thiemens, *Science*, 2000, **289**, 756-758.
191. D. E. Canfield, *Annu Rev Earth Pl Sc*, 2005, **33**, 1-36.
192. J. E. Johnson, A. Gerpheide, M. P. Lamb and W. W. Fischer, *Geol Soc Am Bull*, 2014, **126**, 813-830.
193. J. L. Kirschvink and R. E. Kopp, *Philos T R Soc B*, 2008, **363**, 2755-2765.
194. F. Gaillard, B. Scaillet and N. T. Arndt, *Nature*, 2011, **478**, 229-U112.
195. L. R. Kump and M. E. Barley, *Nature*, 2007, **448**, 1033-1036.
196. D. E. Canfield and R. Raiswell, *Am J Sci*, 1999, **299**, 697-723.
197. J. Castresana and M. Saraste, *Trends in biochemical sciences*, 1995, **20**, 443-448.
198. M. M. Pereira, M. Santana and M. Teixeira, *Bba-Bioenergetics*, 2001, **1505**, 185-208.
199. P. G. Engelkirk, J. Duben-Engelkirk and V. Dowell, *Principles and practice of clinical anaerobic bacteriology: a self-instructional text and bench manual*, Star, 1992.
200. V. B. Borisov, R. B. Gennis, J. Hemp and M. I. Verkhovsky, *Bba-Bioenergetics*, 2011, **1807**, 1398-1413.
201. R. S. Lemos, C. M. Gomes, M. Santana, J. LeGall, A. V. Xavier and M. Teixeira, *Febs Lett*, 2001, **496**, 40-43.
202. A. D. Baughn and M. H. Malamy, *Nature*, 2004, **427**, 441-444.
203. S. W. A. Naqvi, H. Naik, A. Pratihary, W. D'Souza, P. V. Narvekar, D. A. Jayakumar, A. H. Devol, T. Yoshinari and T. Saino, *Biogeosciences*, 2006, **3**, 621-633.
204. B. Kraft, M. Strous and H. E. Tegetmeyer, *J Biotechnol*, 2011, **155**, 104-117.
205. J. R. Nursall, *Nature*, 1959, **183**, 1170-1172.

206. C. T. Reinhard, N. J. Planavsky, S. L. Olson, T. W. Lyons and D. H. Erwin, *P Natl Acad Sci USA*, 2016, **113**, 8933-8938.
207. D. C. Catling, C. R. Glein, K. J. Zahnle and C. P. Mckay, *Astrobiology*, 2005, **5**, 415-438.
208. N. Lane and W. Martin, *Nature*, 2010, **467**, 929-934.
209. R. E. Summons, A. S. Bradley, L. L. Jahnke and J. R. Waldbauer, *Philos T R Soc B*, 2006, **361**, 951-968.
210. K. M. Towe, 1970, **65**, 781-788.
211. S. Levitus, J. Antonov and T. Boyer, *Geophys Res Lett*, 2005, **32**.
212. L. Stramma, G. C. Johnson, J. Sprintall and V. Mohrholz, *Science*, 2008, **320**, 655-658.
213. H. E. Garcia, T. P. Boyer, S. Levitus, R. A. Locarnini and J. Antonov, *Geophys Res Lett*, 2005, **32**.
214. F. A. Whitney, H. J. Freeland and M. Robert, *Prog Oceanogr*, 2007, **75**, 179-199.
215. R. J. Matear, A. C. Hirst and B. I. McNeil, *Geochem Geophy Geosy*, 2000, **1**.
216. C. S. Rousseaux and W. W. Gregg, *Global Biogeochem Cy*, 2015, **29**, 1674-1688.
217. M. Bender, T. Sowers and L. Labeyrie, *Global Biogeochem Cy*, 1994, **8**, 363-376.
218. F. M. Ibarbalz, N. Henry, M. C. Brandao, V. Martini, G. Busseni, H. Byrne, L. P. Coelho, H. Endo, J. M. Gasol, A. C. Gregory, F. Mahe, J. Rigonato, M. Royo-Llonch, G. Salazar, I. Sanz-Saez, E. Scalco, D. Soviadan, A. A. Zayed, A. Zingone, K. Labadie, J. Ferland, C. Marec, S. Kandels, M. Picheral, C. Dimier, J. Poulain, S. Pisarev, M. Carmichael, S. Pesant, S. G. Acinas, M. Babin, P. Bork, E. Boss, C. Bowler, G. Cochrane, C. de Vargas, M. Follows, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels, L. Karp-Boss, E. Karsenti, F. Not, H. Ogata, S. Pesant, N. Poulton, J. Raes, C. Sardet, S. Speich, L. Stemmann, M. B. Sullivan, S. Sunagawa, P. Wincker, L. Bopp, F. Lombard, L. Zinger and T. O. Coordinators, *Cell*, 2019, **179**, 1084-+.
219. G. Salazar, L. Paoli, A. Alberti, J. Huerta-Cepas, H. J. Ruscheweyh, M. Cuenca, C. M. Field, L. P. Coelho, C. Cruaud, S. Engelen, A. C. Gregory, K. Labadie, C. Marec, E. Pelletier, M. Royo-Llonch, S. Roux, P. Sanchez, H. Uehara, A. A. Zayed, G. Zeller, M. Carmichael, C. Dimier, J. Ferland, S. Kandels, M. Picheral, S. Pisarev, J. Poulain, S. G. Acinas, M. Babin, P. Bork, E. Boss, C. Bowler, G. Cochrane, C. de Vargas, M. Follows, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels-Lewis, L. Karp-Boss, E. Karsenti, F. Not, H. Ogata, S. Pesant, N. Poulton, J. Raes, C. Sardet, S. Speich, L. Stemmann, M. B. Sullivan, S. Sunagawa, P. Wincker and T. O. Coordinators, *Cell*, 2019, **179**, 1068-+.
220. E. Villar, T. Vannier, C. Vernette, M. Lescot, M. Cuenca, A. Alexandre, P. Bachelerie, T. Rosnet, E. Pelletier, S. Sunagawa and P. Hingamp, *Nucleic Acids Research*, 2018, **46**, W289-W295.

221. Y. Huang, B. F. Niu, Y. Gao, L. M. Fu and W. Z. Li, *Bioinformatics*, 2010, **26**, 680-682.
222. W. Z. Li and A. Godzik, *Bioinformatics*, 2006, **22**, 1658-1659.
223. D. T. Jones, W. R. Taylor and J. M. Thornton, *Comput Appl Biosci*, 1992, **8**, 275-282.
224. S. Kumar, G. Stecher, M. Li, C. Knyaz and K. Tamura, *Mol Biol Evol*, 2018, **35**, 1547-1549.
225. K. Hori, F. Maruyama, T. Fujisawa, T. Togashi, N. Yamamoto, M. Seo, S. Sato, T. Yamada, H. Mori, N. Tajima, T. Moriyama, M. Ikeuchi, M. Watanabe, H. Wada, K. Kobayashi, M. Saito, T. Masuda, Y. Sasaki-Sekimoto, K. Mashiguchi, K. Awai, M. Shimojima, S. Masuda, M. Iwai, T. Nobusawa, T. Narise, S. Kondo, H. Saito, R. Sato, M. Murakawa, Y. Ihara, Y. Oshima-Yamada, K. Ohtaka, M. Satoh, K. Sonobe, M. Ishii, R. Ohtani, M. Kanamori-Sato, R. Honoki, D. Miyazaki, H. Mochizuki, J. Umetsu, K. Higashi, D. Shibata, Y. Kamiya, N. Sato, Y. Nakamura, S. Tabata, S. Ida, K. Kurokawa and H. Ohta, *Nature Communications*, 2014, **5**.
226. P. Steemans, A. Le Herisse, J. Melvin, M. A. Miller, F. Paris, J. Verniers and C. H. Wellman, *Science*, 2009, **324**, 353-353.
227. C. V. Rubinstein, P. Gerrienne, G. S. de la Puente, R. A. Astini and P. Steemans, *New Phytol*, 2010, **188**, 365-369.
228. B. A. Read, J. Kegel, M. J. Klute, A. Kuo, S. C. Lefebvre, F. Maumus, C. Mayer, J. Miller, A. Monier, A. Salamov, J. Young, M. Aguilar, J. M. Claverie, S. Frickenhaus, K. Gonzalez, E. K. Herman, Y. C. Lin, J. Napier, H. Ogata, A. F. Sarno, J. Shmutz, D. Schroeder, C. de Vargas, F. Verret, P. von Dassow, K. Valentin, Y. Van de Peer, G. Wheeler, J. B. Dacks, C. F. Delwiche, S. T. Dyhrman, G. Glockner, U. John, T. Richards, A. Z. Worden, X. Y. Zhang, I. V. Grigoriev, A. E. Allen, K. Bidle, M. Borodovsky, C. Bowler, C. Brownlee, J. M. Cock, M. Elias, V. N. Gladyshev, M. Groth, C. Guda, A. Hadaegh, M. D. Iglesias-Rodriguez, J. Jenkins, B. M. Jones, T. Lawson, F. Leese, E. Lindquist, A. Lobanov, A. Lomsadze, S. B. Malik, M. E. Marsh, L. Mackinder, T. Mock, B. Mueller-Roeber, A. Pagarete, M. Parker, I. Probert, H. Quesneville, C. Raines, S. A. Rensing, D. M. Riano-Pachon, S. Richier, S. Rokitta, Y. Shiraiwa, D. M. Soanes, M. van der Giezen, T. M. Wahlund, B. Williams, W. Wilson, G. Wolfe, L. L. Wurch and E. H. Annotation, *Nature*, 2013, **499**, 209-213.
229. E. M. a. W. Proper, *Spectroscopy*, 2014, **29**.
230. H. Hernandez and C. V. Robinson, *Nat Protoc*, 2007, **2**, 715-726.
231. F. H. Niesen, H. Berglund and M. Vedadi, *Nat Protoc*, 2007, **2**, 2212-2221.
232. S. H. Hutner, *J Gen Microbiol*, 1950, **4**, 286-293.
233. L. Provasoli, *The ecology of algae*, 1960, **2**, 84-96.
234. N. M. Price, G. I. Harrison, J. G. Hering, R. J. Hudson, P. M. V. Nirel, B. Palenik and F. M. M. Morel, *Biological Oceanography*, 1989, **6**, 443-461.
235. W. Sunda, N. Price and F. Morel, *Journal*, 2005.
236. J. H. Ryther and R. R. Guillard, *Can J Microbiol*, 1962, **8**, 437-&.

237. E. Orlowska, A. Roller, M. Pignitter, F. Jirsa, R. Krachler, W. Kandioller and B. K. Keppler, *Sci Total Environ*, 2017, **577**, 94-104.
238. D. R. Kester, I. W. Duedall, D. N. Connors and R. M. Pytkowicz, *Limnol Oceanogr*, 1967, **12**, 176-+.
239. K. F. Geoghegan, H. B. F. Dixon, P. J. Rosner, L. R. Hoth, A. J. Lanzetti, K. A. Borzilleri, E. S. Marr, L. H. Pezzullo, L. B. Martin, P. K. LeMotte, A. S. McColl, A. V. Kamath and J. G. Stroh, *Anal Biochem*, 1999, **267**, 169-184.
240. A. Spriestersbach, J. Kubicek, F. Schafer, H. Block and B. Maertens, *Method Enzymol*, 2015, **559**, 1-15.
241. L. D. Carbonare, M. D. White, V. Shukla, A. Francini, P. Perata, E. Flashman, L. Sebastiani and F. Licausi, *Plant Physiology*, 2019, **180**, 1614-1628.
242. J. Newman, D. Egan, T. S. Walter, R. Meged, I. Berry, M. Ben Jelloul, J. L. Sussman, D. I. Stuart and A. Perrakis, *Acta Crystallogr D*, 2005, **61**, 1426-1431.
243. R. Page, S. K. Grzechnik, J. M. Canaves, G. Spraggon, A. Kreusch, P. Kuhn, R. C. Stevens and S. A. Lesley, *Acta Crystallographica Section D-Biological Crystallography*, 2003, **59**, 1028-1037.
244. F. Gorrec, *J Appl Crystallogr*, 2009, **42**, 1035-1042.
245. F. Gorrec, *Acta Crystallogr F*, 2015, **71**, 831-837.
246. E. W. N. W. A. F. Holleman, *Inorganic Chemistry*, Academic Press, San Diego, CA, 2001.
247. W. W. Brandt, F. P. Dwyer and E. C. Gyarfas, *Chem Rev*, 1954, **54**, 959-1017.
248. M. K. Groftehauge, N. R. Hajizadeh, M. J. Swann and E. Pohl, *Acta Crystallogr D*, 2015, **71**, 36-44.
249. D. J. Barlow, M. S. Edwards and J. M. Thornton, *Nature*, 1986, **322**, 747-748.
250. J. Huang and W. Honda, *Bmc Immunol*, 2006, **7**.
251. D. C. Benjamin, *Adv Exp Med Biol*, 1995, **386**, 95-108.
252. A. D. Vinion-Dubiel, M. S. McClain, P. Cao, R. L. Mernaugh and T. L. Cover, *Infect Immun*, 2001, **69**, 4329-4336.
253. N. D. Rubinstein, I. Mayrose, D. Halperin, D. Yekutieli, J. M. Gershoni and T. Pupko, *Mol Immunol*, 2008, **45**, 3477-3489.
254. M. H. V. Van Regenmortel, in *Epitope Mapping Protocols: Second Edition*, eds. M. Schutkowski and U. Reineke, Humana Press, Totowa, NJ, 2009, DOI: 10.1007/978-1-59745-450-6_1, pp. 3-20.
255. E. D. Drew and R. W. Janes, *Nucleic Acids Research*, 2020, **48**, W17-W24.
256. W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577-2637.
257. C. Wiedemann, P. Bellstedt and M. Gorlach, *Bioinformatics*, 2013, **29**, 1750-1757.
258. A. Micsonai, F. Wien, E. Bulyaki, J. Kun, E. Moussong, Y. H. Lee, Y. Goto, M. Refregiers and J. Kardos, *Nucleic Acids Research*, 2018, **46**, W315-W322.
259. A. Roy, A. Kucukural and Y. Zhang, *Nat Protoc*, 2010, **5**, 725-738.

260. J. Y. Yang, R. X. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang, *Nat Methods*, 2015, **12**, 7-8.
261. Y. Zhang, *Bmc Bioinformatics*, 2008, **9**.
262. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass and M. J. E. Sternberg, *Nat Protoc*, 2015, **10**, 845-858.
263. S. Akter, L. Fu, Y. Jung, M. Lo Conte, J. R. Lawson, W. T. Lowther, R. Sun, K. K. Liu, J. Yang and K. S. Carroll, *Nat Chem Biol*, 2018, **14**, 995-+.
264. Y. H. Kuo, A. M. Konopko, N. B. Borotto, J. D. Majmudar, S. E. Haynes and B. R. Martin, *Chembiochem*, 2017, **18**, 2028-2032.
265. M. Z. Chen, N. S. Moily, J. L. Bridgford, R. J. Wood, M. Radwan, T. A. Smith, Z. G. Song, B. Z. Tang, L. Tilley, X. H. Xu, G. E. Reid, M. A. Pouladi, Y. N. Hong and D. M. Hatters, *Nature Communications*, 2017, **8**.
266. T. Matsumoto, Y. Urano, T. Shoda, H. Kojima and T. Nagano, *Org Lett*, 2007, **9**, 3375-3377.
267. J. K. Doh, S. J. Tobin and K. E. Beatty, *Bio-Protocol*, 2019, **9**.
268. T. M. Wittkopp, *Bio-protocol*, 2018, **8**, e2838.
269. D. Strenkert, S. Schmollinger, S. D. Gallaher, P. A. Salome, S. O. Purvine, C. D. Nicora, T. Mettler-Altmann, E. Soubeyrand, A. P. M. Weber, M. S. Lipton, G. J. Basset and S. S. Merchant, *P Natl Acad Sci USA*, 2019, **116**, 2374-2383.
270. S. S. Merchant, S. E. Prochnik, O. Vallon, E. H. Harris, S. J. Karpowicz, G. B. Witman, A. Terry, A. Salamov, L. K. Fritz-Laylin, L. Marechal-Drouard, W. F. Marshall, L. H. Qu, D. R. Nelson, A. A. Sanderfoot, M. H. Spalding, V. V. Kapitonov, Q. H. Ren, P. Ferris, E. Lindquist, H. Shapiro, S. M. Lucas, J. Grimwood, J. Schmutz, P. Cardol, H. Cerutti, G. Chanfreau, C. L. Chen, V. Cognat, M. T. Croft, R. Dent, S. Dutcher, E. Fernandez, H. Fukuzawa, D. Gonzalez-Ballester, D. Gonzalez-Halphen, A. Hallmann, M. Hanikenne, M. Hippler, W. Inwood, K. Jabbari, M. Kalanon, R. Kuras, P. A. Lefebvre, S. D. Lemaire, A. V. Lobanov, M. Lohr, A. Manuell, I. Meir, L. Mets, M. Mittag, T. Mittelmeier, J. V. Moroney, J. Moseley, C. Napoli, A. M. Nedelcu, K. Niyogi, S. V. Novoselov, I. T. Paulsen, G. Pazour, S. Purton, J. P. Ral, D. M. Riano-Pachon, W. Riekhof, L. Rymarquis, M. Schroda, D. Stern, J. Umen, R. Willows, N. Wilson, S. L. Zimmer, J. Allmer, J. Balk, K. Bisova, C. J. Chen, M. Elias, K. Gendler, C. Hauser, M. R. Lamb, H. Ledford, J. C. Long, J. Minagawa, M. D. Page, J. M. Pan, W. Pootakham, S. Roje, A. Rose, E. Stahlberg, A. M. Terauchi, P. F. Yang, S. Ball, C. Bowler, C. L. Dieckmann, V. N. Gladyshev, P. Green, R. Jorgensen, S. Mayfield, B. Mueller-Roeber, S. Rajamani, R. T. Sayre, P. Brokstein, I. Dubchak, D. Goodstein, L. Hornick, Y. W. Huang, J. Jhaveri, Y. G. Luo, D. Martinez, W. C. A. Ngau, B. Otillar, A. Poliakov, A. Porter, L. Szajkowski, G. Werner, K. M. Zhou, I. V. Grigoriev, D. S. Rokhsar, A. R. Grossman, C. Annotation and J. A. Team, *Science*, 2007, **318**, 245-251.
271. J. Freshour, R. Yokoyama and D. R. Mitchell, *J Biol Chem*, 2007, **282**, 5404-5412.

272. S. E. Prochnik, J. Umen, A. M. Nedelcu, A. Hallmann, S. M. Miller, I. Nishii, P. Ferris, A. Kuo, T. Mitros, L. K. Fritz-Laylin, U. Hellsten, J. Chapman, O. Simakov, S. A. Rensing, A. Terry, J. Pangilinan, V. Kapitonov, J. Jurka, A. Salamov, H. Shapiro, J. Schmutz, J. Grimwood, E. Lindquist, S. Lucas, I. V. Grigoriev, R. Schmitt, D. Kirk and D. S. Rokhsar, *Science*, 2010, **329**, 223-226.
273. M. Kajikawa, K. T. Yamato, Y. Kohzu, S. Shoji, K. Matsui, Y. Tanaka, Y. Sakai and H. Fukuzawa, *Plant Cell Physiol*, 2006, **47**, 64-73.
274. A. Hemschemeier, D. Casero, B. S. Liu, C. Benning, M. Pellegrini, T. Happe and S. S. Merchant, *Plant Cell*, 2013, **25**, 3186-3211.
275. F. Licausi and B. Giuntoli, *New Phytol*, 2020, DOI: 10.1111/nph.16441.
276. J. M. Quinn, M. Eriksson, J. L. Moseley and S. Merchant, *Plant Physiology*, 2002, **128**, 463-471.
277. A. Hemschemeier, M. Duener, D. Casero, S. S. Merchant, M. Winkler and T. Happe, *P Natl Acad Sci USA*, 2013, **110**, 10854-10859.
278. M. Duner, J. Lambertz, C. Mugge and A. Hemschemeier, *Plant J*, 2018, **93**, 311-337.
279. L. J. Reitzer and B. Magasanik, *J Bacteriol*, 1982, **151**, 1299-1313.
280. R. Humbert and R. D. Simoni, *J Bacteriol*, 1980, **142**, 212-220.
281. L. Gaufichon, M. Reisdorf-Cren, S. J. Rothstein, F. Chardon and A. Suzuki, *Plant Sci*, 2010, **179**, 141-153.
282. L. Gaufichon, C. Masclaux-Daubresse, G. Tcherkez, M. Reisdorf-Cren, Y. Sakakibara, T. Hase, G. Clement, J. C. Avice, O. Grandjean, A. Marmagne, S. Boutet-Mercey, M. Azzopardi, F. Soulay and A. Suzuki, *Plant Cell Environ*, 2013, **36**, 328-342.
283. K. V. Vu, C. Y. Jeong, T. T. Nguyen, T. T. H. Dinh, H. Lee and S. W. Hong, *J Exp Bot*, 2019, **70**, 1775-1787.
284. J. Klinkenberg, H. Faist, S. Saupe, S. Lambertz, M. Krischke, N. Stingl, A. Fekete, M. J. Mueller, I. Feussner, R. Hedrich and R. Deeken, *Plant Physiology*, 2014, **164**, 570-583.
285. R. G. Upchurch, *Biotechnol Lett*, 2008, **30**, 967-977.
286. L.-J. Xie, Y. Zhou, Q.-F. Chen and S. Xiao, *Progress in Lipid Research*, 2021, **81**, 101072.