Characterisation of Plant Cysteine Dioxygenase

Brooke Hayes

A thesis presented for the degree of Master of Science

Department of Biochemistry
University of Otago
New Zealand
April 2018
Abstract

Many organisms possess a non-heme, mononuclear iron enzyme termed cysteine dioxygenase (CDO) that aids in homeostasis. This enzyme irreversibly adds molecular oxygen to the thiol group of the N-terminal cysteine residue of proteins in plants, acting as an oxygen sensor. Plant CDO targets are transcription factors which permit transcription of enzymes required for anaerobic metabolism. Under hypoxic conditions, the molecular oxygen co-substrate is not present at a high concentration therefore these transcription factors cannot be dioxidised and remain able to elicit an anoxic response. In normoxia, plant CDO dioxidises the N-terminal cysteine of these transcription factors, tagging them for degradation. This study evaluated recombinant expression and purification of CDO from a three plant species. CDO from Arabidopsis thaliana and Zea mays were then characterised both structurally and kinetically.

Plant CDO’s were expressed in Escherichia coli cells and purified using Strep-tag® affinity chromatography. Plant cysteine oxidase 1 (PCO1) and plant cysteine oxidase 2 (PCO2) isoforms from Arabidopsis thaliana co-purified with chaperones, and DNA. Two shortened variants of PCO1 were designed to abolish these interactions and improve homogeneity. A variant, ΔN, had residues 2-51 removed and produced contaminant-free product. When residues 2-52 and 247-293 were removed in the variant termed ΔNΔC, more co-purifying contaminants were present than with full-length PCO1. Plant CDO from Zea mays was expressed and purified both as a full length construct and a variant missing the first 32 residues. The full length Z. mays construct co-purified with a range of contaminants, and removal of the N-terminal did not improve protein homogeneity. A PCO2 isoform from Orzya sativa was attempted to be produced, resulting in no expression.

Homology models of PCO1 were produced to assess structural characteristics. Surface charge distribution, disulfide bonding and accessibility of the active site was explored in these models. Mass spectrometry (MS) showed that disulfide bonding was present in PCO1. PCO1, PCO2, ΔN, ΔNΔC and both maize constructs were subjected to a range of high-throughput crystallography screens. Promising conditions were optimised, but no
diffracting crystals were produced. Metal ion binding to plant CDO was assessed using intact MS, and showed that the protein may also bind zinc in vitro. Nuclear magnetic resonance showed that plant CDO is not able to dioxidise free cysteine. Other colorimetric kinetic assays were performed to show that plant CDO is able to act on N-terminal cysteine as part of a di-, tri- or penta-peptide having the N-terminal of target molecules. Plant CDO has greater affinity for longer peptides.

DNA binding is predicted to be via an electrostatic interaction with N-terminal, which also appears to also contain a nuclear localisation signal. Nuclear localisation, followed by DNA binding could permit localisation to targets in vivo. This would allow plant CDO to quickly bind transcription factor targets as they are also DNA binding proteins. Disulfide bonding may also play a role in modulating protein activity. As disulfide bonding relates to cell oxidation state, this could permit conformational changes that allow the protein activity to be increased or decrease accordingly.
Acknowledgements

First, I must thank Sigurd. I am thankful for all of the time and effort that you have put in to this project. I am grateful for having to opportunity to learn from such a great scientist. Next, I would like to thank Malcolm. You have a depth of knowledge of almost everything science, and I thank you for sharing that. Pauline, you have become a great friend and I wish you luck with continuing work with plant CDO.

To Mum and Brendan, I appreciate all of your visits to Dunedin. Your support has and will always be invaluable. Dad, I have always strived to better myself as a result of your encouragement and expectation. For that, I am thankful. To my sisters and my grandparents, again I am thankful for your support.

Hannah, you were one of the better things to come of studying at Otago. I am eternally indebted to you for pushing me to study for exams, and ensuring I was always working toward my goals. You are a great friend, and I will miss you much. To Ted, thanks for saving me from insanity at Studholme. Our coffee (and beer) dates will be missed. Rob, cheers for the great chat and entertaining game nights.

To Sarah, you are the only reason I am glad I studied undergraduate Microbiology. I came to this city not knowing what lay ahead, and I cannot say I am disappointed with the outcome. I leave here with you, and the two degrees are just a bonus.
# Contents

Abstract .................................................. i  
Acknowledgements ........................................ iii  
Contents .................................................. iv  
List of Figures ............................................ viii  
List of Tables ............................................. xii  
Abbreviations ............................................. xiii  
Amino acids ............................................... xvii  

## Chapter One

**Introduction** ........................................... 1  
1.1 Cysteine Dioxygenase ................................... 1  
1.2 Mammalian CDO ........................................ 2  
1.3 Plant CDO ................................................ 3  
  1.3.1 Target proteins ..................................... 4  
  1.3.2 RAP proteins ...................................... 7  
  1.3.3 HRE proteins ...................................... 7  
  1.3.4 Localisation and expression ....................... 8  
1.4 Bacterial CDO ........................................... 9  
1.5 Structure of CDO ....................................... 10  
  1.5.1 Active site structure ............................. 12  
1.6 Substrate binding ....................................... 16  
  1.6.1 Proposed Mammalian CDO Mechanism .......... 17  
1.7 CDO kinetics ........................................... 19  
1.8 Project significance ................................... 21  
1.9 Project aims ........................................... 21  

## Chapter Two

**Materials and Methods** ................................. 22  
2.1 Materials ............................................... 22  
2.2 DNA work ................................................ 34  
  2.2.1 DNA buffers ......................................... 34
2.2.2 Plasmid design 35
2.2.3 Plasmid purification 36
2.2.4 Horizontal TAE agarose gels 36

2.3 Bacterial Growth 36
2.3.1 Bacterial work reagents 36
2.3.2 Competent cell preparation 37
2.3.3 Transformation 37
2.3.4 Induction of protein expression 38

2.4 Protein work 40
2.4.1 Protein work buffers 40
2.4.2 Homogenisation and lysis 43
2.4.3 Syringe purification 43
2.4.4 FPLC purification 44
2.4.5 Dialysis and concentrating 45
2.4.6 Gel filtration chromatography 46
2.4.7 Cation exchange chromatography 46
2.4.8 SDS-PAGE 47
2.4.9 Solubility testing 47
2.4.10 Improving protein solubility during lysis 48
2.4.11 Improving protein solubility post-purification 48
2.4.12 Refolding assay 49
2.4.13 TCA precipitation 49
2.4.14 Dynamic light scattering (DLS) 49

2.5 Crystallisation 50
2.5.1 High throughput screening 50
2.5.2 Varying condition plates 50
2.5.3 Mounting crystals 51
2.5.4 X-ray diffraction 51

2.6 Assays 52
2.6.1 Ellman’s assay 52
2.6.2 Ferrozine 53
2.6.3 Limited proteolysis 55

2.7 Nuclear Magnetic Resonance (NMR) Spectroscopy 56
2.7.1 $^1$H NMR spectroscopy

2.8 Computational work
   2.8.1 Bioinformatics
   2.8.2 Modelling
   2.8.3 Protein sequence-based prediction
   2.8.4 Plasmid sequencing
   2.8.5 Other

2.9 Mass spectrometry (MS)
   2.9.1 Identification of proteins
   2.9.2 Identification of disulfide bonds
   2.9.3 Intact mass MS

Chapter Three

*Arabidopsis thaliana* CDO expression and purification

3.1 PCO1 and PCO2
   3.1.1 PCO1 and PCO2 expression
   3.1.2 PCO1 purification
   3.1.3 PCO2 purification

3.2 Purified PCO1 assessment
   3.2.1 PCO1 chaperone removal
   3.2.2 PCO1 instability

3.3 PCO1 truncations
   3.3.1 Expression and solubility of PCO1 truncations
   3.3.2 Purification of $\Delta N \Delta C$ truncations of PCO1

3.4 PCO1 nucleic acid binding

3.5 Summary

Chapter Four

*Zea mays* and *Orzya sativa* CDO expression and purification

4.1 *Zea mays* constructs
   4.1.1 Maize A construct
   4.1.2 Maize B construct

4.2 *Orzya savita* construct

4.3 Summary
Chapter Five

Structural modelling and analysis 113

5.1 Homology models 114
  5.1.1 SWISS-MODEL homology model 116
  5.1.2 I-TASSER homology model 119
5.2 Disulfide bonding 124
5.3 Domain analysis 130
5.4 Crystallography 131
  5.4.1 Arabidopsis thaliana constructs 132
  5.4.2 Zea mays constructs 140
5.5 Summary 143

Chapter Six

Functional characterisation: Iron binding and kinetics 144

6.1 Iron binding 144
6.2 Kinetics 153
6.3 Summary 159

Conclusions 161
Future directions 167
References 169
Appendices 16981
# List of Figures

1.1 Reaction catalysed by CDO. 1
1.2 Metabolic pathway of Mammalian CDO catalytic activity. 2
1.3 Plant CDO activity scheme. 4
1.4 Scheme of RAP2.12 mediated response. 5
1.5 ERFVII N-terminal consensus sequence. 7
1.6 Structure of Mammalian CDO. 11
1.7 Comparison of bacterial and mammalian crystal structures. 12
1.8 Fitted electron density map for the key active site features of Mammalian CDO. 13
1.9 Multiple sequence alignment of CDO from various origins. 14
1.10 Proposed mechanisms for the production of cysteinesulfinate by mammalian CDO. 18
2.1 Schematic of Strep-affinity method. 44
2.2 Ellman’s reaction. 52
2.3 L-cysteine standard curve. 53
2.4 Ferrozine compound. 54
2.5 Ferrozine standard curve. 55
3.1 Multiple sequence alignment of PCO1 and PCO2 from various plant origins. 61
3.2 PCO1 and PCO2 expression at 18°C. 63
3.3 PCO1 and PCO2 expression at 28°C. 64
3.4 PCO1 solubility. 65
3.5 PCO2 solubility. 66
3.6 PCO1 bench-top syringe purification. 67
3.7 PCO1 Strep-tag® FPLC purification. 68
3.8 PCO2 Strep-tag® FPLC purification. 70
3.9 PCO2 Strep-tag® FPLC purification gel. 71
3.10 PCO1 post-ATP incubation size exclusion. 72
3.11 PCO1 post-ATP incubation size exclusion peak one.
3.12 PCO1 post-EDTA dialysis size exclusion.
3.13 PCO1 post-EDTA dialysis size exclusion peak one.
3.14 PCO1 disorder prediction using PONDR® VLXT.
3.15 PCO1 limited proteolysis.
3.16 ΔN expression.
3.17 ΔNΔC expression.
3.18 ΔN solubility.
3.19 ΔNΔC solubility.
3.20 ΔN Strep-tag® FPLC purification.
3.21 ΔN Strep-tag® FPLC purification peak one.
3.22 ΔN Strep-tag® FPLC purification peak two.
3.23 ΔNΔC Strep-tag® FPLC purification.
3.24 ΔNΔC Strep-tag® FPLC purification gel.
3.25 Agarose gel of PCO1 and ΔN.
3.26 Agarose gel of PCO1 with nucleases.
3.27 Summary of all CDOs purified from A. thaliana.
4.1 Maize A expression.
4.2 Maize A solubility at 18°C.
4.3 Maize A solubility at 28 and 37°C.
4.4 Maize A Strep-tag® FPLC purification.
4.5 Maize A Strep-tag® FPLC purification gel.
4.6 Cation exchange chromatography of maize A.
4.7 Maize A cation exchange chromatography gel.
4.8 Maize B expression.
4.9 Maize B solubility at 18°C and 28°C.
4.10 Maize B solubility at 37°C.
4.11 Maize B Strep-tag® FPLC purification.
4.12 Maize B Strep-tag® FPLC purification gel.
4.13 ΔNR expression in BL21(DE3).
ΔNR expressed in BL21(DE3) solubility at 28°C and 37°C.

ΔNR expressed in BL21(DE3) solubility at 18°C.

ΔNR expressed in BL21(DE3) Strep-Tag® FPLC purification.

ΔNR expressed in BL21(DE3) Strep-Tag® FPLC purification gel.

ΔNR expression in BL21(DE3) pLysS.

ΔNR expression in Origami™ (DE3).

Summary of all purified Maize CDO constructs.

Comparison of homology models.

SWISS-MODEL homology model of PCO1.

SWISS-MODEL homology model of PCO1 surface topology.

SWISS-MODEL homology model of PCO1 with catalytic residues.

I-TASSER homology model of PCO1.

I-TASSER homology model of PCO1 superimposed on mammalian CDO.

I-TASSER homology model of PCO1 surface topology.

PCO1 and RAP2.12 surface topology with charge distribution.

I-TASSER homology model of ΔN surface topology.

SWISS-MODEL homology model of PCO1 with cysteine residues mapped.

PCO1 ± BME.

Alternative labelling of PCO1 cysteines followed by MS.

PCO1 refolding samples ± BME.

PCO1 Superdex 200 gel filtration.

PCO1 Superdex 200 gel filtration SDS-PAGE gel.

Putative PCO1 crystal from the second Crystal Screen HT attempt.

Putative PCO1 crystals after incubation with various metal ions.

Drops containing sea urchin crystals.

Maize B spherulites from Crystal Screen HT.

Optimised maize B spherulites from PEGRx HT with calcium.
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.21</td>
<td>Large drop maize B spherulites from PEGRx HT with calcium.</td>
<td>144</td>
</tr>
<tr>
<td>6.1</td>
<td>Intact mass spectrometry data for ΔN.</td>
<td>145</td>
</tr>
<tr>
<td>6.2</td>
<td>PCO1 expression with iron and reduction of IPTG.</td>
<td>147</td>
</tr>
<tr>
<td>6.3</td>
<td>Bench-top purification of PCO1 with varying IPTG induction concentrations.</td>
<td>148</td>
</tr>
<tr>
<td>6.4</td>
<td>Bench-top purification of PCO1 with varying IPTG induction concentrations and iron addition.</td>
<td>149</td>
</tr>
<tr>
<td>6.5</td>
<td>Iron saturation levels of purified protein ± iron addition.</td>
<td>150</td>
</tr>
<tr>
<td>6.6</td>
<td>Iron bound to refolding experiment samples.</td>
<td>151</td>
</tr>
<tr>
<td>6.7</td>
<td>Velocity of substrate depletion using refolding experiment samples.</td>
<td>152</td>
</tr>
<tr>
<td>6.8</td>
<td>$^1$H NMR spectra of standards and cysteine incubated with CDO.</td>
<td>154</td>
</tr>
<tr>
<td>6.9</td>
<td>PCO1 titration with dipeptide substrate.</td>
<td>156</td>
</tr>
<tr>
<td>6.10</td>
<td>Michaelis-Menten curve for PCO1 with dipeptide.</td>
<td>157</td>
</tr>
<tr>
<td>6.11</td>
<td>Substrate depletion of tri-peptide by PCO1.</td>
<td>158</td>
</tr>
</tbody>
</table>
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Chemical suppliers.</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>Enzymes and associated buffers.</td>
<td>25</td>
</tr>
<tr>
<td>2.3</td>
<td>Chromatography columns, resins and commercial eluent.</td>
<td>26</td>
</tr>
<tr>
<td>2.4</td>
<td>General consumables.</td>
<td>27</td>
</tr>
<tr>
<td>2.5</td>
<td>Crystallography consumables.</td>
<td>27</td>
</tr>
<tr>
<td>2.6</td>
<td>Equipment.</td>
<td>28</td>
</tr>
<tr>
<td>2.7</td>
<td><em>E. coli</em> strains used.</td>
<td>31</td>
</tr>
<tr>
<td>2.8</td>
<td>Plasmids used.</td>
<td>31</td>
</tr>
<tr>
<td>2.9</td>
<td>Antibiotic stocks.</td>
<td>32</td>
</tr>
<tr>
<td>5.1</td>
<td>DLS of <em>A. thaliana</em> PCO1 derivatives.</td>
<td>132</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>ACBP1</td>
<td>Acyl-coA binding-domain containing protein 1</td>
<td></td>
</tr>
<tr>
<td>ADH1</td>
<td>Alcohol Dehydrogenase</td>
<td></td>
</tr>
<tr>
<td>amp</td>
<td>Ampicillin</td>
<td></td>
</tr>
<tr>
<td>APS</td>
<td>Ammonium persulfate</td>
<td></td>
</tr>
<tr>
<td>ATE1</td>
<td>Arginyltransferase 1</td>
<td></td>
</tr>
<tr>
<td>ATE2</td>
<td>Arginyltransferase 2</td>
<td></td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
<td></td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic local alignment search tool</td>
<td></td>
</tr>
<tr>
<td>BME</td>
<td>$\beta$-mercaptoethanol</td>
<td></td>
</tr>
<tr>
<td>CDO</td>
<td>Cysteine Dioxygenase</td>
<td></td>
</tr>
<tr>
<td>CPR</td>
<td>Centre for protein research</td>
<td></td>
</tr>
<tr>
<td>CSA</td>
<td>Cysteine sulfinic acid</td>
<td></td>
</tr>
<tr>
<td>CV</td>
<td>Column volume</td>
<td></td>
</tr>
<tr>
<td>$\Delta N$</td>
<td>PCO1 lacking residues 2-53</td>
<td></td>
</tr>
<tr>
<td>$\Delta N\Delta C$</td>
<td>PCO1 lacking residues 2-53 and 247-293</td>
<td></td>
</tr>
<tr>
<td>DLS</td>
<td>Dynamic light scattering</td>
<td></td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double stranded DNA</td>
<td></td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single stranded DNA</td>
<td></td>
</tr>
<tr>
<td>DOC</td>
<td>Deoxycholate</td>
<td></td>
</tr>
<tr>
<td>DTNB</td>
<td>5,5-dithio-bis-2-nitrobenzoic acid</td>
<td></td>
</tr>
<tr>
<td>DTPA</td>
<td>Diethylenediamine penta-acetic acid</td>
<td></td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
<td></td>
</tr>
<tr>
<td>EEO</td>
<td>Electroendosmosis</td>
<td></td>
</tr>
<tr>
<td>EF-Tu</td>
<td>Elongation factor thermo-unstable</td>
<td></td>
</tr>
<tr>
<td>ERF-VII</td>
<td>Ethylene response factor group VII</td>
<td></td>
</tr>
<tr>
<td>FPLC</td>
<td>Fast protein liquid chromatography</td>
<td></td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
<td></td>
</tr>
<tr>
<td>GMQE</td>
<td>Global model quality estimation</td>
<td></td>
</tr>
<tr>
<td>GSH</td>
<td>Reduced Glutathione</td>
<td></td>
</tr>
<tr>
<td>GSSG</td>
<td>Oxidised Glutathione</td>
<td></td>
</tr>
<tr>
<td>HABA</td>
<td>4'-hydroxyazobenzene-2-carboxylic acid</td>
<td></td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid</td>
<td></td>
</tr>
<tr>
<td>HP</td>
<td>High performance</td>
<td></td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
<td></td>
</tr>
<tr>
<td>IAA</td>
<td>Iodoacetic acid</td>
<td></td>
</tr>
<tr>
<td>IAM</td>
<td>Iodoacetamide</td>
<td></td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
<td></td>
</tr>
<tr>
<td>I-TASSER</td>
<td>Iterative threading assembly refinement</td>
<td></td>
</tr>
<tr>
<td>KAN</td>
<td>Kanamycin</td>
<td></td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny broth</td>
<td></td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix assisted laser desorption/ionisation-time of flight</td>
<td></td>
</tr>
<tr>
<td>MetAP</td>
<td>Methionine aminotransferase</td>
<td></td>
</tr>
<tr>
<td>MME</td>
<td>Mono-methyl ether</td>
<td></td>
</tr>
<tr>
<td>3-MPA</td>
<td>3-mercaptopropionic acid</td>
<td></td>
</tr>
<tr>
<td>MPD</td>
<td>2-Methyl-2,4-pentanediol</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
<td></td>
</tr>
<tr>
<td>MS/MS</td>
<td>Tandem mass spectrometry</td>
<td></td>
</tr>
<tr>
<td>MWCO</td>
<td>Molecular weight cut off</td>
<td></td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
<td></td>
</tr>
<tr>
<td>NEB</td>
<td>New England Biolabs</td>
<td></td>
</tr>
<tr>
<td>NLS</td>
<td>Nuclear localisation signal</td>
<td></td>
</tr>
<tr>
<td>OD</td>
<td>Optical Density</td>
<td></td>
</tr>
<tr>
<td>PAGE</td>
<td>Poly-acrylamide gel electrophoresis</td>
<td></td>
</tr>
<tr>
<td>PCO1</td>
<td>Plant Cysteine Oxidase 1</td>
<td></td>
</tr>
<tr>
<td>PCO2</td>
<td>Plant Cysteine Oxidase 2</td>
<td></td>
</tr>
<tr>
<td>PDC1</td>
<td>Pyruvate decarboxylase</td>
<td></td>
</tr>
<tr>
<td>PEG</td>
<td>Poly-ethylene glycol</td>
<td></td>
</tr>
<tr>
<td>pI</td>
<td>Isoelectric point</td>
<td></td>
</tr>
<tr>
<td>PRT6</td>
<td>Proteolysis 6</td>
<td></td>
</tr>
<tr>
<td>QDO</td>
<td>Oxoquinoline 2,4-dioxygenase</td>
<td></td>
</tr>
<tr>
<td>RAP 2</td>
<td>Related to apetala 2</td>
<td></td>
</tr>
<tr>
<td>R_f</td>
<td>Relative motility</td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td>Room temperature</td>
<td></td>
</tr>
<tr>
<td>SAP</td>
<td>Shrimp alkaline phophatase</td>
<td></td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td>Sulfopropyl</td>
<td></td>
</tr>
<tr>
<td>SUS4</td>
<td>Sucrose synthase 4</td>
<td></td>
</tr>
<tr>
<td>TAE</td>
<td>Tris acetate-EDTA</td>
<td></td>
</tr>
<tr>
<td>TCA</td>
<td>Trichloroacetic acid</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>TCEP</td>
<td>Tris(2-carboxyethyl)phosphine</td>
<td></td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA</td>
<td></td>
</tr>
<tr>
<td>TEMED</td>
<td>Tetramethylethylenediamine</td>
<td></td>
</tr>
<tr>
<td>TNB</td>
<td>2-nitro-5-thiobenzoate</td>
<td></td>
</tr>
<tr>
<td>Tris</td>
<td>Trisaminomethane</td>
<td></td>
</tr>
<tr>
<td>TSP</td>
<td>Trisodium phosphate</td>
<td></td>
</tr>
<tr>
<td>uHPLC</td>
<td>Ultra-high performance liquid chromatography</td>
<td></td>
</tr>
<tr>
<td>UV</td>
<td>Ultra Violet</td>
<td></td>
</tr>
</tbody>
</table>
# Amino acids

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Single Letter</th>
<th>Three Letter</th>
<th>One Letter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Aspartic Acid</td>
<td>Asp</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
<td>Q</td>
</tr>
<tr>
<td>Glutamic Acid</td>
<td>Glu</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
<td>K</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>
Chapter One

Introduction

1.1 Cysteine Dioxygenase

Almost any living organism to survive, must be capable of adapting to a range of environments. Many organisms possess a non-heme, mononuclear-iron enzyme termed cysteine dioxygenase (CDO- E.C 1.13.11.20) that aids in homeostasis. This enzyme is responsible for the irreversible addition of molecular oxygen to the thiol group of cysteine residues. This reaction leads to the incorporation of both oxygen atoms to produce cysteine sulfinic acid (CSA) as depicted in Figure 1.1.

\[
\begin{align*}
\text{HOOC} & \quad \text{S}^- \\
\text{NH}_2 & \quad \text{CDO} \\
\text{O}_2 & \quad \text{HOOC} \\
\text{NH}_2 & \quad \text{S}^=O \\
L-\text{Cysteine (Cys)} & \quad L-\text{Cysteine Sulfinic Acid (CSA)}
\end{align*}
\]

*Figure 1.1 Reaction catalysed by CDO.* Cysteine is dioxidised to form cysteine sulfinic acid. Note the requirement for molecular oxygen as a co-substrate (red). Figure taken from Davies (2016).

This enzyme was first described in 1966 by Ewetz and Sorbo, who determined that it functioned as a general oxidase (Ewetz and Sörbo, 1966). Lombardini and colleagues showed that CDO was in fact a dioxygenase, and that it did not require external redox cofactors as originally anticipated (Lombardini et al., 1969). Upon purification from rat liver in the 1970s by Yamaguchi, the first major steps in mammalian CDO characterisation were performed (Yamaguchi et al., 1978). This showed that mammalian CDO was a dioxygenase that was important for cysteine metabolism (see 1.2). Since then, many other researchers have produced experimental data for both the activity and structure of mammalian CDO. The role of CDO in plants was not elucidated until recently. Weits et al. showed that plant CDO dioxidised targets like mammalian CDO and permitted a response to low oxygen levels (see 1.3) (Weits et al., 2014). CDO was
identified in bacteria in 2006 by sequence alignment, however the physiological role has not yet been elucidated (see 1.4). Although many studies have been conducted to derive the general mechanism of CDO, it remains elusive (Ye et al., 2007; Kumar et al., 2011). It is currently best understood in mammalian CDO.

1.2 Mammalian CDO

Within mammals, CDO is involved in the initial steps of cysteine catabolism (Stipanuk, 2004). After free cysteine is dioxidised to cysteine sulfinate (CSA), the reaction can branch in two directions. Decarboxylation of CSA leads to the production of hypotaurine, which is then able to be further oxidised in order to produce an important amino acid, taurine (Stipanuk et al., 2011). The second branch enables the production of free sulfate groups and pyruvate by a transamination reaction (Figure 1.2).

![Figure 1.2. Metabolic pathway of Mammalian CDO catalytic activity.](image)

Mammalian CDO regulation is cysteine mediated and is solely post-translational (Bella et al., 2002; Joseph and Maroney, 2007). Degradation is rapid, and ubiquitin facilitated as the 26S proteasome is responsible for its breakdown. This mechanism has been shown to have remarkable responsiveness, and CDO activity is determined primarily by dietary
intake of cysteine (Stipanuk et al., 2006). Such responsiveness indicates that homeostasis of cysteine is important for cell viability. This is further highlighted by the cytotoxic and neurotoxic effects of increased cysteine levels (Stipanuk et al., 2009). A decrease in CDO activity has also been observed in many types of cancer (Heafield et al., 1990; Brait et al., 2012; Prabhu et al., 2014). A CDO knockout mouse model resulted in excess production of toxic hydrogen sulfide (H₂S) and build-up of cysteine. This product was excreted in urine and affected mouse growth substantially. The long term effects were not elucidated, as these mice were fed hypotaurine and taurine 6-weeks post birth (Roman et al., 2013).

1.3 Plant CDO

Controlling oxygen usage in plants is crucial to different growth stages such as seed development and maturation. During these various growth stages oxygen diffusion is limited, therefore a means of oxygen sensing is critical for plant survival. Oxygen sensing is also required during submergence-tolerance and waterlogging, a factor essential for the cultivation of food such as rice (Xu et al., 2006). Tolerance to submergence is also important in flooding, where timely adaptation to this hypoxic environment permits survival (Bailey-Serres and Voesenek, 2008). The activity of CDO within plants permits a response to molecular oxygen concentration. This is because CDO activity is only possible in the presence of molecular oxygen, as it is a co-substrate. This has been demonstrated in rice, Arabidopsis and barley (Kosmacz et al., 2015; Mendiondo et al., 2016).

More specifically, plant CDO catalyses the N-terminal dioxidation of penultimate cysteine residues to form cysteine sulfinic acid under normoxia (>1% v/v O₂)(White and Flashman, 2016). This dioxidation leads to arginylation by arginyl transferases (ATE1/2) which decreases target protein stability by the pathway that executes the N-end rule (Bachmair et al., 1986; Weits et al., 2014). A single subunit E3 ligase, PROTEOLYSIS 6 (PRT6), recognises the N-terminal arginine and polyubiquitylates the protein. Target degradation then occurs via the 26S proteasome (Licausi et al., 2011; Bailey-Serres et al., 2012). For this to occur, methionine aminopeptidase (MetAP) is required to remove
the N-terminal methionine residue of the nascent target to expose the N-terminal cysteine (Figure 1.3).

Figure 1.3. Plant CDO activity scheme. Depicted is the process plant CDO is involved in to permit target degradation. Green oval represents target transcription factor RAP2.12, single letters before the green oval are N-terminal amino acids, and PCO1/2 are plant CDO isoforms from Arabidopsis. Subscript ‘ox’ refers to the cysteine being dioxidised. RAP2.12: Related to Apetala 2.12, MetAP: methionine aminopeptidase, ATE1/2: arginyl transferase 1/2, PRT6: PROTEOLYSIS 6. Figure taken from Weits et al. (2014).

White et al. conducted a range of experiments to explore the specifics of plant CDO’s activity. Using mass spectrometry and NMR, it was demonstrated that plant CDO catalyses the formation of species from 10mer peptides (with an N-terminal cysteine) that are consistent with the addition of an extra 2 or 3 oxygen atoms. It was deduced that both oxygen atoms from molecular oxygen were incorporated into the product by using isotopically labelled molecular oxygen (18O2). It was later shown that trioxidation was an artefact of their experimental procedure and was a result of laser exposure during Matrix assisted laser desorption/ionisation mass spectrometry (MALDI-MS) (White et al., 2017).

Studies were conducted to evaluate whether plant CDO was not active on targets, but instead enhanced activity of ATE1/2 or PRT6. It was also hypothesised that the response could be as a result of plant CDO acting as a transcriptional repressor. Neither of these hypotheses were proven to be true, and evidence instead supported the idea of the protein acting as a dioxygenase (Weits et al., 2014).

1.3.1 Target proteins

Although there are almost 200 proteins in Arabidopsis that contain a cysteine at position 2 (Licausi et al., 2011), as yet plant CDO has been shown to act only on group VII ethylene response factor (ERF-VII) transcription factors (Weits et al., 2014; White et al., 2017). Arabidopsis has 5 transcription factors within this group, Hypoxia Responsive1
(HRE1), Hypoxia Responsive2 (HRE2) and the constitutively expressed Related to Apetala 2.12 (RAP2.12), RAP2.2 and RAP2.3.

**Target protein activities**

These transcription factors are characterised by a single conserved ethylene response factor (ERF) domain which binds *cis*-elements in DNA. They are involved in up-regulating approximately 25 hypoxia associated genes, including the transcription of enzymes that utilise anaerobic fermentation to produce energy such as alcohol dehydrogenase 1 (ADH1), lactate dehydrogenase (LDH) and pyruvate decarboxylase 1 (PDC1) (**Figure 1.4**) (Gasch *et al.*, 2016). They also permit transcription of enzymes required for enhancing carbohydrate mobilisation and glycolysis for fermentation such as sucrose synthase 4 (SUS4) (Liu *et al.*, 2005; Gibbs *et al.*, 2011). These processes prevent excessive oxygen usage during metabolism, permitting oxygen conservation (Weits *et al.*, 2014). This switch in metabolism is essential for plant survival during anoxia, but limits plant growth (Paul *et al.*, 2016).

![Hypoxia](image)

**Figure 1.4. Scheme of RAP2.12 mediated response.** Figure taken from Paul *et al.* (2016). Red indicates inhibition, green indicates activation and the dotted arrow indicates an uncharacterised interaction. RAP2.12: Related to Apetala 2.12, ADH: alcohol dehydrogenase, PDC: Pyruvate decarboxylase, LDH: Lactose Dehydrogenase, AlaAT: Alanine Aminotransferase, TCA: Tricarboxylic acid cycle.
During hypoxia, the enzyme alanine aminotransferase (alaAT) is also upregulated (Klok et al., 2002) (Figure 1.4). It is proposed that this upregulation is related to the enhanced stability of the RAP proteins, although this interaction has not been characterised. Under hypoxic conditions, this alaAT converts pyruvate into alanine. This presents an alternative to the lactate and ethanol fermentation pathways. This reaction does not provide cofactor reduction such as with other pathways, thus the exact role of this upregulation is not clear. Alanine does however provide a non-toxic molecule into which excess pyruvate can be converted to for the removal of glycolytic products. It has been suggested that this aids in maintaining cellular pH and permits a balance between carbon and nitrogen during hypoxia (Miyashita et al., 2007). This also allows for downregulation of the TCA cycle as it is not required to utilise pyruvate from glycolysis. The downregulation of the TCA cycle then reduces the need for oxygen utilisation during cofactor recycling by the electron transport chain.

ERF-VII proteins contain a short (approximately 10 amino acid) conserved N-terminal sequence (Figure 1.5). The presence of an N-terminal cysteine is crucial for catalysis. Mutating this residue to alanine slowed oxygen consumption substantially, suggesting that dioxidation by plant CDO was no longer occurring (Weits et al., 2014). Licausi et al. showed that by removing the first 13 amino acids of RAP2.12 (termed Δ12RAP2.12), it was no longer degraded under normoxic conditions (21% v/v oxygen) (Licausi et al., 2011). This is because Δ13RAP2.12 is not recognised by plant CDO due to the absence of the N-terminal recognition sequence. Over expression of Δ13RAP2.12 during normoxia led to large changes in the plant metabolome. In particular, there was an accumulation of fermentation products which slowed plant development. Interestingly, over expression of RAP2.12 led to decreased survival during anoxia (Paul et al., 2016).
Figure 1.5. ERFVII N-terminal consensus sequence. Black represents polar residues, green non-polar residues and red negatively charged residues. Size of single amino acid code dictates likelihood of the residue at that position, with 1.5 bits being full conservation. Created used software developed by Schneider and Stephens (1990), and Crooks et al. (2004). Sequences of all 5 predicted CDO targets were used (RAP2.2, RAP2.3, RAP2.12, HRE1 and HRE2). Sequences were taken from Uniprot.

1.3.2 RAP proteins

Transcriptome studies show that RAP2.2, 2.3 and 2.12 elicit the strongest response during anoxia, and therefore are important transcription factors in the ERF-VII group. All three are constitutively expressed, and exhibit similar expression profiles. RAP2.2 and RAP2.12 are close homologues with 62% percent identity (Appendix 1). They are thought to be functionally redundant, acting on the same target genes (Bui et al., 2015).

Whilst RAP2.2 appears to be a solely nuclear protein, fluorescent labelling showed that RAP2.12 and RAP2.3 can be translocated to the plasma membrane. RAP2.12 and RAP2.3 are translocated in complex with an acyl-CoA binding domain containing protein 1 (ACBP1). This prevents transcription factor degradation as plant CDO is not found near the cell membrane. Although the transcription factors are present, it does not permit the hypoxic response as they must be interacting with DNA in the nucleus to elicit a transcriptional response. As the transcription factor is readily available for translocation back into the nucleus when oxygen levels deplete (below 10% v/v oxygen), this permits timely adaption to hypoxic conditions (Licausi et al., 2011; Kosmacz et al., 2015).

1.3.3 HRE proteins

In contrast to RAP proteins, these transcription factors are expressed in low levels during normoxia, and are upregulated during anoxia. Licausi et al. showed that over expression of these proteins enhanced survival during anoxia, yet decreased plant survival during
normoxic conditions. Micro-arrays suggest HRE1 is the main HRE protein responsible for the anaerobic response in vivo, as more anoxic genes were upregulated by over expression of HRE1 than HRE2. However, knockouts showed that HRE2 is able to compensate for the loss of HRE1 as loss of only HRE1 had no effect on plant survival, only influencing a small number of non-lethal factors (Licauisi et al., 2010). This suggests that HRE1 and HRE2 are partially redundant. Experiments show regulation of HRE1 is at the expression level, whilst regulation of HRE2 is post-transcriptional. As HRE1 is regulated at the expression level, this means that plant CDO is only required post-anoxia to remove excess HRE1 proteins. HRE2 however appears to be regulated by plant CDO under normal conditions.

1.3.4 Localisation and expression

There are 5 different plant CDO isoforms within Arabidopsis. These isoforms were termed plant cysteine oxidase (PCO) 1-5 by Weits et al. The activities of PCO1 and PCO2 are predicted to be redundant, as are the activities of PCO3, PCO4 and PCO5. This was demonstrated by single double gene knockouts whereby removal of either PCO1 or PCO2 had little to no impact, but removal of both led to a decrease in plant survival under normoxia. Analysis of the genes transcribed as a result of these knockouts showed that the same genes were being over-expressed in the absence of PCO1 or PCO2 during normoxia. Using fluorescent labelling, PCO1 and PCO2 were shown to localise exclusively to the nucleus, whereas PCO3, PCO4 and PCO5 are cytosolic proteins (Weits et al, 2014).

Weits et al. analysed the expression profile of all PCOs and found that all genes were expressed during development. PCO1 and PCO2 expression was greatly upregulated during growth stages, such as seed development, which are characterised by low oxygen levels. Over-expression of PCO1/2 by Weits et al. (2014) unsurprisingly led to impaired tolerance during submergence and negatively impacted plant growth.

Plant CDO upregulation during anoxia is due the increased presence of RAP proteins. The upstream promoter region of PCO1 and PCO2 comprises a hypoxia-responsive promoter element (HRPE) 5′-AAACCA(G/C)(G/C)(G/C)GC-3′ which contains a highly conserved AC-rich start, loosely conserved linker and GC end. This element was shown
to be important for transcriptional transactivation by RAP2.12 and RAP2.2 as they can directly bind to this element. As these CDO targets have been shown to upregulate transcription of the PCO genes themselves, this creates a negative feedback loop that primes the anoxic cell for target degradation on return to normoxia (Weits et al., 2014; Gasch et al., 2016).

1.4 Bacterial CDO

With the aid of crystallographic data and alignment tool searches, Dominy et al. identified four potential CDO enzymes within Bacillis and Streptomyces genera. Overall identity was low (12-21%), yet active site residue conservation indicated similar function. Enzyme kinetic data showed that the four proteins tested exhibited a strong substrate preference for cysteine. These data confirmed that these enzymes were in fact bona fide CDOs of prokaryotic origin. Further phylogenetic analyses indicate that CDO is also found within many other eubacteria phyla. The use of these enzymes within bacteria is proposed to be for the production of elemental sulfur. It may also play a role in regulation of disulfide bonds and formation of endospores (Dominy et al., 2006).

Sequence alignment by Tchesnokov et al. also showed a CDO homologue from Pseudomonas aeruginosa (Tchesnokov et al., 2015). This enzyme was shown to be expressed under basal conditions using mass spectrometry. Assays showed the protein was able to oxidise cysteine in vivo, but showed a preference for 3-mercaptopropionic acid (3-MPA). Mössbauer spectroscopy suggested that substrate binding via the thiol group and crystallography showed similar active site geometry to mammalian CDO. The role of 3-MPA oxidation was suggested to provide a means of substrate assimilation. This would permit removal and subsequent utilisation of the elemental sulfur of the thiol group. As this enzyme prefers 3-MPA as opposed to cysteine, it is a thiol dioxygenase not a CDO.
1.5 Structure of CDO

Crystal structures of CDO have been analysed from mammalian and bacterial sources. These crystal structures, alongside sequence homology, show that CDO is a member of the very diverse cupin superfamily (Dunwell et al., 2001). As a member of this superfamily, CDO contains two short characteristic cupin motifs (GX₅HXHₓ₆EXₓG and GX₇PXGXₓ₂HXₓ₃N) which are separated by a variable linker region of 15-20 amino acids (Chad R Simmons et al., 2006). These conserved sequence motifs are responsible for providing ligands for metal binding. These ligands are the side chains of histidine residues, which explains the histidine residue conservation for the cupin motif. The central structure of CDO consists of a canonical six-stranded β-barrel, comprised of β-sheets. This core is where Fe²⁺ is coordinated by three histidine residues (Figure 1.6) (Joseph and Maroney, 2007; Stipanuk et al., 2011). As seen in Figure 1.6, only three α-helices are present within the structure and of these, none are near the active site.
Figure 1.6. Structure of Mammalian CDO. Blue is used to colour α-helices, whereas β-strands are tan. Iron is shown as an orange sphere bound to the active site histidine residues, which are depicted as sticks with nitrogen atoms coloured blue. PDB accession: 2B5H.

When superimposing bacterial and mammalian CDO structures, no secondary structure elements vary between the two structures by more than ~0.5 Å. This suggest that although mammalian and bacterial CDO exhibit low sequence similarity (12-21%), they maintain similar structural characteristics (Figure 1.7). Thus these structures likely reflect that of plant CDO with significant homology, even though plant CDO also shares low sequence similarity with mammalian CDO. Structural characteristics of plant CDO have yet to be confirmed due to the absence of a crystal structure. The production of homology models could also give insight into structural features, and allow structural comparison.
Figure 1.7. Comparison of bacterial and mammalian crystal structures. Iron coordinating histidine residues are shown. Tan represents bacterial thiol dioxygenase (from *Pseudomonas aeruginosa*), blue represents mammalian CDO (*Rattus norvegicus*). Bacterial CDO accession: 3USS.

1.5.1 Active site structure

The active site of CDO differs from many other non-heme, mono-iron enzymes. The archetype is a 2-His-1-carboxylate triad of iron ligands. CDO lacks this configuration, and instead contains a neutrally charged 3-His facial triad that coordinates the central iron. The change may be related to the fact that many similar non-haeme mono-iron enzymes catalyse reactions with non-thiol substrates. The coordinating histidine residues are present at positions 86, 88 and 140 in mammalian CDO, as shown in Figure 1.8.

To understand why such a variation is present, de Visser and Stragnaz conducted computation modelling where they replaced the 3-His coordination with a 2-His/1-Asp triad within mammalian CDO (de Visser and Straganz, 2009). The results of this indicate that although the reaction was able to proceed, a larger energy barrier was present. Therefore, the rare coordination was not required for activity but aided in optimal
cysteine dioxygenation (Tchesnokov et al., 2012). This is likely due to a stronger interaction between the iron and thiol group, as well as the variation influencing the iron spin state. This in turn affects dioxygen activation (Souness et al., 2013).

Figure 1.8. Fitted electron density map for the key active site features of Mammalian CDO. Taken from Simmons et al. (2006). Water is portrayed as red spheres, whereas iron is portrayed with a larger orange sphere. A crosslink is present between Tyr-157 and Cys-93, which is shown in orange. All residues and their positions are labelled. Nitrogen coloured blue, oxygen red, orange sulfur and carbon is grey.

Other important features of the active site include the residues Tyr-157, Cys-93 and His-155 (Figure 1.8). The hydrogen bonding networks of these residues aid in stabilising catalysis (Aluri and de Visser, 2007). It has also been proposed that Tyr-157 is a catalytic acid/base due to its hydroxyl group (Dominy et al., 2006; Davies et al., 2014).

Sequence alignment suggests that the 3-His triad is present in plant CDO, yet indicates that other core stabilising amino acids are not conserved (Figure 1.9). This includes Tyr-157, Cys-93 and His-155. Due to such difference in the active site, an overall sequence identity of 24% between mammalian and plant CDO is not surprising (not shown, comparison done using BLAST). Comparison of plant CDO and bacterial CDO shows
even less sequence similarity with little more than only the iron co-ordinating histidine residues being conserved. The production of crystal data would enable the active site configuration to be determined.

| Arachnopus | MGEMFPEKEKEYLELILLSHQCKSNPNSVKKKHKNKHKKMTCWRRKIDSPADITAVRR | 60 |
| Bacillus | ------------------------------- | ME | 2 |
| Rattus | ------------------------------- | MERTELLKPR----TLAD | 14 |
| Mus | ------------------------------- | MERTELLKPR----TLAD | 14 |
| Pan | ------------------------------- | MEQTEVLKPR----TLAD | 14 |
| Homo | ------------------------------- | MEQTEVLKPR----TLAD | 14 |
| Arachnopus | LPNTCKEVSNGGVGIPEDEQKQLELIDMDPEDVGLTPMTFFYFRPSQVARSSPP | 120 |
| Bacillus | LYECIQDFGGK----------NP8VSDLATSLLQIPNAAKL---SQPIKEFDQY---Y | 48 |
| Rattus | LIRILHELFGD---VN----VEEVTLEAVAEY--SNPA---EMALYKFDQYR-------Y | 58 |
| Mus | LIRILHELFGD---VN----VEEVTLEAVAEY--SNPA---EMALYKFDQYR-------Y | 58 |
| Pan | LIRILHELFGD---VN----VEEVTLEAVAEY--SNPA---EMALYKFDQYR-------Y | 58 |
| Homo | LIRILHELFGD---VN----VEEVTLEAVAEY--SNPA---EMALYKFDQYR-------Y | 58 |
| Figure 1.9. Multiple sequence alignment of CDO from various origins. Arabidopsis thaliana (thale cress or mouse-ear cress), Bacillus subtilis (non-pathogenic, soil dwelling, gram positive bacterium), Rattus norvegicus (brown rat), Mus musculus (house mouse), Pan troglodytes (common chimpanzee) and Homo sapien (modern human). Blue indicates histidines in the iron coordination sphere and green the proposed catalytic tryptophane (not present in PCO1. One dot indicates similar amino acid, two dots very similar amino acids and a star indicates identical amino acids. Alignment generated on Clustal Omega, and sequences were taken from UniProt. |
Metal iron centre

Tchesnokov et al. revealed that exceptionally tight, stoichiometric binding was present between the central iron and 3-His imidazole groups. This iron contains a four-coordinate pseudo-tetrahedral geometry (Tchesnokov et al., 2012). As three of the iron coordination sites are occupied by histidine, this leaves the fourth coordination site vacant to a water molecule (pictured as Wat4 in Figure 1.8) (Souness et al., 2013). A second water molecule then interacts via H-bonding with the first.

The active site Fe$^{2+}$ of CDO is crucial for enzymatic activity (Joseph and Maroney, 2007). The necessity for iron was shown experimentally by Tchesnokov et al. by use of an oxygen electrode to follow oxygen consumption. Results disclosed that the amount of iron binding was proportional to enzymatic activity (Tchesnokov et al., 2012). White et al. was able to co-purify plant CDO with 0.3 ± 0.12 atoms of Fe$^{2+}$ per CDO suggesting that plant CDO also contains a metal iron centre (White and Flashman, 2016). Iron occupancy of ~30% may be as a result of in vitro purification, as free iron availability is likely to be different in plant cells. Mammalian CDO is able to be purified with ~20% iron occupancy, and this can be increased to greater than 90% by loading purified protein with iron (Tchesnokov et al., 2012). Attempts to increase plant CDO iron binding post-purification could be achieved by iron addition such as with mammalian CDO.

Crosslink formation

Another key feature of the mammalian CDO is a crosslink which provides a structural variant. This crosslink can be formed between Tyr-157 and Cys-93, both residues which are buried deep in the active site. This thioether bond (depicted in orange within Figure 1.8) crosslinks two β-strands which are present within the active site. This crosslink is not a ligand for iron binding as the distance is too large (4.16 Å) (Joseph and Maroney, 2007). Crosslinking was proposed to facilitate enhanced catalysis due to the constant orientation of the tyrosine aromatic ring within the active site (de Visser and Straganz, 2009). More recently it has been shown by enzyme kinetics and crystallography that the increase in catalysis is instead due to the removal of the thiol of Cys-93 out of the active site (Davies et al., 2014). Model chemistry also suggests that it is important for increased catalysis as it decreases both the redox potential and p$K_a$ of the tyrosine (Shinobu Itoh et al., 1997).
Covalent crosslinking was only able to be achieved in the presence of oxygen, and as such this is a condition in which CDO enzyme maturation is important. Among naturally occurring CDO, crosslinking has only been experimentally demonstrated in mammalian CDO. Crystal structure analysis shows that variant prokaryotic sources are also able to form a crosslink (Siakkou et al., 2011; Fellner et al., 2016). Sequence alignment suggests that both Cys-93 and Tyr-157 are not present within plant CDO, therefore the likelihood of an active site crosslink is low. Crystal structure determination would aid in verifying if this assumption is true.

Bioinformatic based analysis has shown that although the tyrosine present in the mammalian crosslink is found within bacterial CDOs, Cys-93 is substituted for glycine (Joseph and Maroney, 2007). Most bacteria contain a glycine residue in this position, which does not support the formation of a crosslink. Some bacteria have a cysteine close to this position, therefore it has been speculated that this could permit formation of the crosslink. Solved bacterial CDO crystal structures did not contain a cross link, so it was suggested that this hypothesis is not correct (Driggers et al., 2015). However, bacterial CDO has not been crystallised in the presence of substrate, so this has not been extensively tested.

1.6 Substrate binding

Mammalian and bacterial CDO dioxidise free cysteine whereas plant CDO dioxidises cysteine as part of a peptide. Changes to the substrate will have large influences on substrate co-ordination and order of substrate entry into the active site. It was determined by Dominy et al. that cysteine-CDO binding is highly specific, thus no alternate substrates react or compete with native substrate (Dominy et al., 2006). This specificity is a characteristic of all known CDOs, and is contrasting to other known non-heme mononuclear iron enzymes (Tchesnokov et al., 2015). Electron paramagnetic resonance (EPR) showed that binding of the dioxygen substrate occurs after cysteine is bound (Stipanuk et al., 2011; Souness et al., 2013). This ordered binding occurs in both mammalian and bacterial CDO. However, it is not known if this binding order applies to plant CDO as it has not been experimentally tested. It seems likely that oxygen will enter the active site before the target peptide as the peptide is large, and may block co-substrate
entry. Multiple spectroscopic approaches could be harnessed to test this, such as EPR as mentioned above.

Cysteine binds to mammalian CDO in a bidentate fashion via both the amide and sulfide groups (Souness et al., 2013). The cysteine carboxyl interacts with the Arg-60 guanidinium group via a salt bridge. This interaction allows further orientation stabilisation and substrate specificity (Aluri and de Visser, 2007; Kumar et al., 2011; Tchesnokov et al., 2012). This was demonstrated experimentally by Fellner et al. (Fellner et al., 2016). This reaction is not possible in plant CDO as the carboxylate is contained within the amide bond within the peptide. The formation of a substrate-bound crystal would enable affirmation of how peptide binds in the active site.

1.6.1 Proposed Mammalian CDO Mechanism

Most of what is known of the potential mechanism of mammalian CDO is derived computationally. This is because the reaction proceeds very quickly, therefore intermediates are exceptionally hard to trap and characterise. Stopped-flow spectrophotometry by Davies (unpublished) shows that the reaction proceeds to completion within 30 ms. However, using isotopically labelled molecular oxygen it was shown that both of the dioxygen atoms are integrated to form the sulfinate (Aluri and de Visser, 2007). This gives some indication as to how the mechanism proceeds. There are currently two favoured models for the mechanism of mammalian CDO.

**Mechanism I**

With the aid of both quantum mechanics/molecular mechanics (QM/MM) and density function theory (DFT), de Visser and Straganz were able to predict a potential CDO mechanism (Mechanism I of Figure 1.10)(de Visser and Straganz, 2009). After cysteine binding, the initial step of the reaction is proposed to be dioxygen activation. Upon coordination of dioxygen, the Fe$^{2+}$ is oxidised to Fe$^{3+}$ and a superoxide is formed (Figure 1.10, A). This activation leads to radical addition on the distal oxygen atom to the cysteine sulfur. This forms a 4 membered heterocycle where a new O-S bond is present and the sulfur become a radical cation (Figure 1.10, B). This structure weakens the O-O bond, causing bond cleavage (Figure 1.10, C). An iron(IV)-oxo complex is then formed with a cis-bound sulfoxide. A rotation of the sulfoxide oxygen enables the proximal
oxygen atom to be transferred whereby cysteine sulfinate is formed. The iron is then reduced back to the +2 oxidation state (Figure 1.10, D). Modelling conducted by Aluri and de Visser concluded that the rate limiting step of the reaction was that preceding the sulfoxyl radical formation (Aluri and de Visser, 2007).

**Figure 1.10. Proposed mechanisms for the production of cysteinesulfinate by mammalian CDO.** Taken from Scheme 2 of Kumar *et al.* (2011). Blue is the distal oxygen, whereas the proximal oxygen is coloured red.

**Mechanism II**

The second proposed mechanism was described by Kumar *et al.* (Kumar *et al.*, 2011). It was projected that the initial step involves the radical addition of the proximal oxygen atom to the cysteine sulfur to form a persulfinate (Mechanism II of Figure 1.10, E). Through atomic rearrangement, the distal oxygen inserts into the persulfinate forming a 4-membered, heterocycle (Figure 1.10, F). This ring structure permits a break in the dioxygen bond, enabling a sulfoxide intermediate to form (Figure 1.10, G). The last step involves transfer of an oxygen atom to yield cysteine sulfinate (Figure 1.10, D). This mechanism has been predicted to have higher activation energies and is therefore not the most likely mechanism. Instead, it was proposed that this mechanism may present a means of enzymatic inhibition as it blocks the active site.

Although there has been significant research into the mechanism of mammalian CDO, the reaction proceeds too quickly for intermediates to be determined. Given this, the true mechanism of mammalian, bacterial or plant CDO is still not known. As there is a current plateau in formulating the correct mammalian CDO mechanism, analysing the active site
of plant CDO may offer some benefit. As many active site residues are lacking in plant CDO, yet the same thiol dioxidation reactions take place, characterisation of plant CDO shows what features are essential for the reaction.

1.7 CDO kinetics

Several studies have been conducted to characterise kinetic properties of thiol dioxygenases. To do this, a multitude of different methodologies have been employed. These include the use of oxygen electrodes to measure co-substrate usage (Njeri and Ellis, 2014), whereas HPLC (Weits et al., 2014) and mass spectrometry (MS) measure product formation (White et al., 2017). NMR has also been used to look at both substrate depletion, and product formation (Bruland et al., 2009). Unfortunately, all of the methods described have deficiencies. Although oxygen electrodes are highly specific for oxygen, data does show what product is being formed during oxygen consumption. HPLC and MS allow end point analysis but do not easily allow kinetic parameters to be assessed. Much like HPLC and MS, NMR provides means to specifically assess product formation but does not permit routine kinetic measurements.

A number of kinetic parameters have been reported for mammalian CDO. Fellner et al. (2014) used the Ellman's assay to suggest a $K_m$ of 2.8 mM for cysteine in presence of saturating $O_2$ and $k_{cat}$ of 0.63 s$^{-1}$ (Fellner et al., 2014). Simmons et al. (2006) used HPLC to determine a $K_m$ of 0.45 mM for cysteine in presence of saturating $O_2$ and $k_{cat}$ of 0.72 s$^{-1}$ (Chad R. Simmons et al., 2006). Although many kinetic values have been determined there are inconsistencies in results. This may be due to variation in assay pH and temperature as well as heterogeneity of protein samples due to the crosslink and/or iron occupancy (Davies et al., 2014). Differing kinetic values may also be due to the use of different assay methods.

Kinetic parameters from bacterial thiol dioxygenases have also been explored. Tchesnokov et al. (2015) used HPLC to determine kinetic properties of *Pseuodmonas aeruginosa* thiol dioxygenase under the presence of saturating $O_2$. This study suggested a $K_m$ of $1.0 \pm 0.4$ mM and $k_{cat}$ of $0.11 \pm 0.002$ s$^{-1}$ (Tchesnokov et al., 2015). Dominy et al. (2006) also used HPLC under the presence of saturating $O_2$ and reported $K_m$ of 3.0 mM and $k_{cat}$ of 0.39 s$^{-1}$ for *Bacillus subtilis* thiol dioxygenase. Dominy et al. also determined
a $K_m$ of 5.7 mM and $k_{cat}$ of 2.0 s$^{-1}$ for *Bacillus cereus* (Dominy *et al*., 2006). Again, although many kinetic values have been determined there are inconsistencies in results. These changes may appear due to factors mentioned above.

A colorimetric assay that measures substrate usage have also been used to study thiol dioxygenase kinetics (Davies *et al*., 2014). This method presents a means of standardising thiol dioxygenase kinetic studies. This assay uses Ellman’s reagent to measure free thiol groups, and as all substrates contain a thiol prior to dioxidation, one can follow substrate depletion. As this assay can be conducted in a 96-well plate, data can be obtained in a high-throughput manner. As data points can be obtained at short time intervals, it is suitable for determining kinetic parameters. This method presents only one issue- that substrate depletion is measured as opposed to product formation. This can be accounted for by the use of other more specific techniques such as NMR and MS to assess if the correct product is being formed. This method was utilised in research conducted for this thesis.

Little is known of plant CDO’s kinetic properties. Weits *et al.* (2014) used HPLC to determine that plant CDO is able to catalyse the formation of CSA from free cysteine. However, the no enzyme control appeared to produce almost as much CSA as samples that contained enzyme and this result is contradicted by a later study (White *et al*., 2017). These authors also used an oxygen electrode to show that plant CDO is active with peptides that contain an N-terminal cysteine residue. However, data in this paper suggested that plant CDO shows some activity with peptides that do not have an N-terminal cysteine which is thought to be extremely unlikely (Weits *et al*., 2014). White *et al.* (2017) used MS to show specifically that plant CDO is able to dioxidise short peptides (10 amino acids) that have an N-terminal cysteine residue. The data produced also suggested that the product peptides included CSA. No kinetic parameters were reported in either of these studies. Kinetic properties of plant CDO will be further explored in this thesis.
1.8 Project significance

This research presents both agricultural significance, and aids mammalian CDO research which may have clinical relevance. Characterisation of plant CDO could aid in determining the mechanism of mammalian CDO. As the substrate differs greatly to that of both mammalian and bacterial CDO, it is intriguing that the same dioxidation reaction takes place. It therefore offers a unique perspective on the minimal requirements for CDO activity. As mammalian CDO is involved in the metabolism of a cytotoxic amino acid, the determination of its mechanistic activity has clinical relevance. This could then permit the production of treatments for some cysteine related disorders involving cysteine accumulation, including Alzheimer’s and Parkinson’s disease.

Agricultural applications arise when considering the effect of CDO on plant survival. As CDO facilitates both submergence-tolerance and seed development in plants, manipulation may reap benefits in terms of crop yield. This is of vital importance when considering increasing food scarcity and population size. Climate change is also impacting weather conditions, and severe weather events such as flooding are becoming increasingly common (Lesk et al., 2016). Strategies to address the survival of crops in such adverse conditions is vital for maintaining the economy of countries reliant on agriculture, and to ensure enough crops are cultivated for food.

1.9 Project aims

It is clear that little research has been conducted to characterise plant CDO. This project aimed to establish expression and purification conditions for different CDO isoforms from a range of plant species. Extensive crystallographic based research was attempted, with a range of variation screens and optimisation of selected conditions. Four Arabidopsis thaliana, two Zea mays and one Orzya sativa CDO protein/s were employed to aid in crystallisation. Homology models were also employed to assess structural features. Parameters of enzymatic activity and metal binding properties were also explored. The use of different substrate peptides allowed probing of substrate recognition and catalytic efficiency. Characterisation will give not only insight into plant CDO, but may aid in determining the mechanism of mammalian and bacterial CDO.
Chapter Two

Materials and Methods

2.1 Materials

All water (unless stated otherwise) is sourced from a Milli-Q® filtration system produced by Merck. Water from this system results in Type 1 ultrapure water with a resistivity of 18.2 MΩ·cm at 25 °C. For a list of all chemicals used and the supplier, see Table 2.1. A list of all enzymes and associated buffers can be found in Table 2.2.

Table 2.1 Chemical suppliers.

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetic acid, glacial</td>
<td>Merck</td>
</tr>
<tr>
<td>Acrylamide, 30% acrylamide/bis solution 37:5:1</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>Agar Bacteriology grade</td>
<td>Lab Supply</td>
</tr>
<tr>
<td>Agarose low endoelectrosmosis (EEO)</td>
<td>Lab Supply</td>
</tr>
<tr>
<td>Antifoam A concentrate</td>
<td>Sigma</td>
</tr>
<tr>
<td>Ampicillin (amp)</td>
<td>Goldbio</td>
</tr>
<tr>
<td>Ammonium acetate</td>
<td>Scharlau Chemie</td>
</tr>
<tr>
<td>Ammonium iron (II) sulfate</td>
<td>Thermofisher</td>
</tr>
<tr>
<td>Ammonium persulfate (APS)</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>L-Arginine</td>
<td>Sigma</td>
</tr>
<tr>
<td>L-Ascorbic acid</td>
<td>Thermofisher</td>
</tr>
<tr>
<td>Adenosine triphosphate (ATP)</td>
<td>Sigma</td>
</tr>
<tr>
<td>Chemical</td>
<td>Supplier</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>β-mercaptoethanol (BME)</td>
<td>Sigma</td>
</tr>
<tr>
<td>Bromophenol blue</td>
<td>VWR</td>
</tr>
<tr>
<td>Calcium chloride</td>
<td>VWR</td>
</tr>
<tr>
<td>Calcium acetate</td>
<td>Sigma</td>
</tr>
<tr>
<td>Charcoal, activated</td>
<td>Riedel-deHaen</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>Goldbio</td>
</tr>
<tr>
<td>Citric acid</td>
<td>Thermofisher</td>
</tr>
<tr>
<td>Cobalt Chloride</td>
<td>Sigma</td>
</tr>
<tr>
<td>Coomassie blue G250</td>
<td>VWR</td>
</tr>
<tr>
<td>L-Cysteine</td>
<td>Sigma</td>
</tr>
<tr>
<td>Deoxycholate (DOC)</td>
<td>Sigma</td>
</tr>
<tr>
<td>5,5-dithio-bis-2-nitrobenzoic acid (DTNB)</td>
<td>Sigma</td>
</tr>
<tr>
<td>Diethylenediamine penta-acetic acid (DTPA)</td>
<td>Thermofisher</td>
</tr>
<tr>
<td>Dithiothreitol (DTT)</td>
<td>Roche</td>
</tr>
<tr>
<td><strong>DNA ladders</strong></td>
<td></td>
</tr>
<tr>
<td>1Kb</td>
<td>NEB</td>
</tr>
<tr>
<td>1Kb plus</td>
<td>Thermofisher</td>
</tr>
<tr>
<td>Ethanol, bulk and analytical grade</td>
<td>Scharlau Chemie</td>
</tr>
<tr>
<td>Ethylenediaminetetraacetic acid (EDTA)</td>
<td>VWR</td>
</tr>
<tr>
<td>Ferrozine</td>
<td>Sigma</td>
</tr>
<tr>
<td>Glutathione, reduced (GSH)</td>
<td>Sigma</td>
</tr>
<tr>
<td>Glutathione, oxidised (GSSG)</td>
<td>Sigma</td>
</tr>
<tr>
<td>Chemical</td>
<td>Supplier</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Glycerol</td>
<td>Merck</td>
</tr>
<tr>
<td>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)</td>
<td>Sigma</td>
</tr>
<tr>
<td>Hydrochloric acid (HCl)</td>
<td>VWR</td>
</tr>
<tr>
<td>Iodoacetic acid, 2-C\textsubscript{13} labelled</td>
<td>Sigma</td>
</tr>
<tr>
<td>Isopropyl β-d-1-thiogalactopyranoside (IPTG)</td>
<td>Lab Supply</td>
</tr>
<tr>
<td>Iron (III) chloride</td>
<td>Sigma</td>
</tr>
<tr>
<td>Kanamycin (KAN)</td>
<td>Goldbio</td>
</tr>
<tr>
<td>Magnesium chloride</td>
<td>VWR</td>
</tr>
<tr>
<td>Molecular weight marker, broad range</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>Nickel chloride</td>
<td>Merck</td>
</tr>
<tr>
<td><strong>Peptide substrates</strong></td>
<td></td>
</tr>
<tr>
<td>Cys-Gly</td>
<td>Genscript</td>
</tr>
<tr>
<td>Cys-Gly-Gly</td>
<td></td>
</tr>
<tr>
<td>Cys-Gly-Gly-Ala-Ile</td>
<td></td>
</tr>
<tr>
<td>Peptone from casein, tryptone</td>
<td>Merck</td>
</tr>
<tr>
<td>Potassium chloride</td>
<td>Sigma</td>
</tr>
<tr>
<td>Polyethylene glycol (PEG) 8,000</td>
<td>Sigma</td>
</tr>
<tr>
<td>PEG Mono-Methyl Ether (MME) 5000</td>
<td>Molecular Dimensions</td>
</tr>
<tr>
<td>Propan-2-ol</td>
<td>Scharlau Chemie</td>
</tr>
<tr>
<td>Protease inhibitor cocktail (cOmplete, EDTA free)</td>
<td>Roche</td>
</tr>
<tr>
<td>Sodium acetate</td>
<td>Merck</td>
</tr>
<tr>
<td>Sodium cacodylate</td>
<td>Scharlau Chemie</td>
</tr>
</tbody>
</table>
Sodium chloride | Merck
Sodium citrate | Scharlau Chemie
Sodium dodecyl sulfate (SDS) | Merck
Di-sodium hydrogen phosphate | Scharlau Chemie
Sodium di-hydrogen phosphate | Scharlau Chemie
Sodium hydroxide pellets | Scharlau Chemie
D-Sucrose | Lab Supply
Sulphuric acid | VWR
Trichloroacetic acid (TCA) | VWR
Tetramethylethylenediamine (TEMED) | VWR
Trisaminomethane (Tris) | Lab Supply
TWEEN® 20 | Sigma
Xylene cyanol FF | Sigma
Yeast extract, granulated | Merck

All chemicals used were analytical grade unless otherwise specified.

**Table 2.2 Enzymes and associated buffers.**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type II restriction enzyme BamHI</td>
<td>NEB</td>
</tr>
<tr>
<td>BspHI</td>
<td>NEB</td>
</tr>
<tr>
<td>DnaseI, bovine pancreas</td>
<td>Roche</td>
</tr>
<tr>
<td>Mung Bean Nuclease</td>
<td>Roche</td>
</tr>
<tr>
<td>NcoI</td>
<td>NEB</td>
</tr>
<tr>
<td>Product</td>
<td>Supplier</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Proteinase K</td>
<td>Sigma</td>
</tr>
<tr>
<td>RNase A, bovine pancreas</td>
<td>Roche</td>
</tr>
<tr>
<td>Shrimp alkaline phosphatase (SAP)</td>
<td>Roche</td>
</tr>
<tr>
<td>T4 DNA ligase</td>
<td>Roche</td>
</tr>
<tr>
<td>T4 DNA ligase buffer</td>
<td>Roche</td>
</tr>
<tr>
<td>NEB buffer 4</td>
<td>NEB</td>
</tr>
</tbody>
</table>

**Table 2.3 Chromatography columns, resins and commercial eluent.**

<table>
<thead>
<tr>
<th>Product</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chelex® 100 resin, styrene divinylbenzene copolymer</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>Sulfopropyl (SP) high performance (HP) 5 mL cation exchange column, HiTrap®</td>
<td>GE Healthcare</td>
</tr>
<tr>
<td><em>Strep</em>-Tactin® Superflow High Capacity resin, streptavidin</td>
<td>IBA</td>
</tr>
<tr>
<td><strong>Strep-Tag® protein purification buffers</strong></td>
<td>IBA</td>
</tr>
<tr>
<td>Elution buffer</td>
<td>IBA</td>
</tr>
<tr>
<td>4'-hydroxyazobenzene-2-carboxylic acid (HABA)</td>
<td>IBA</td>
</tr>
<tr>
<td>StrepTrap™ HP, HiTrap®, streptavidin resin</td>
<td>Roche</td>
</tr>
<tr>
<td>Superdex 75, HiLoad 16/60, sepharose resin</td>
<td>GE Healthcare</td>
</tr>
<tr>
<td>Superdex 200, HiLoad 26/600, sepharose resin</td>
<td>GE Healthcare</td>
</tr>
</tbody>
</table>

All columns mentioned in Table 2.3 were stored in 20% ethanol, and were washed with water before being equilibrated with running buffer. All columns were purchased pre-poured.
### Table 2.4 General consumables.

<table>
<thead>
<tr>
<th>Material</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dialysis cassette, Slide-A-Lyzer™, 10K molecular weight cut off (MWCO)</td>
<td>Thermofisher</td>
</tr>
<tr>
<td>Dialysis tubing, Snakeskin® pleated, 10K MWCO</td>
<td>Thermofisher</td>
</tr>
<tr>
<td>E.Z.N.A® plasmid extraction kit</td>
<td>VWR</td>
</tr>
<tr>
<td>Filters, circular 47 mm diameter, cellulose acetate 0.2 μm</td>
<td>Lab Supply</td>
</tr>
<tr>
<td>Plate, clear, 96-well flat bottom Corning Co-star®</td>
<td>Sigma</td>
</tr>
<tr>
<td>Syringe filter, ReliaPrep™ 0.2 and 0.45 μm, 30 mm diameter</td>
<td>Lab Supply</td>
</tr>
<tr>
<td>T7 bacteriophage promoter forward and reverse primers</td>
<td>Sigma</td>
</tr>
<tr>
<td>Amicon® concentrator, 10K MWCO (0.5, 6 and 15 mL)</td>
<td>Merck</td>
</tr>
</tbody>
</table>

### Table 2.5 Crystallography consumables.

<table>
<thead>
<tr>
<th>Material</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bases, copper (with vials)</td>
<td>Mitegen</td>
</tr>
<tr>
<td>Cover slides, siliconized glass 18 mm circle</td>
<td>Hampton</td>
</tr>
<tr>
<td><strong>Loops</strong></td>
<td></td>
</tr>
<tr>
<td>Cryoloops- Nylon</td>
<td>Hampton</td>
</tr>
</tbody>
</table>
MiTeGen® loops | Mitegen

**Screens**

Crystal screen HT | Hampton
JCSG plus™ | Molecular Dimensions
Morpheus® Additive Screen | Hampton
PACT Premier™ | Molecular Dimensions
PEG/ion HT | Hampton
PEGRx HT | Hampton
Pre-crystallisation | Hampton

Mosquito micro-reservoir strips | TTP LabTech

**Plates**

Swissci 96-well 3-drop | Hampton
Pre-greased 24-well | Molecular Dimensions

Plate seals (96-well), clear ViEWSEAL SEALER | Lab Supply

---

**Table 2.6 Equipment.**

<table>
<thead>
<tr>
<th>Equipment</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Centrifuges</strong></td>
<td></td>
</tr>
<tr>
<td>Avanti J-25 (standing)</td>
<td>Beckman Coulter</td>
</tr>
<tr>
<td>Avanti JXN-26 (standing)</td>
<td>Beckman Coulter</td>
</tr>
<tr>
<td>Microfuge® 16 (bench-top)</td>
<td>Beckman Coulter</td>
</tr>
<tr>
<td>Centrifuge 5810 R (bench-top)</td>
<td>Eppendorf</td>
</tr>
<tr>
<td>Dynamic light scattering (DLS), Dynapro-99-E-50</td>
<td>Wyatt technology</td>
</tr>
<tr>
<td><strong>Electrophoresis powerpacks</strong></td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--</td>
</tr>
<tr>
<td>PowerPac Junior</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>PowerPac 300</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>Fermenter, Microgen</td>
<td>New Brunswick Scientific</td>
</tr>
<tr>
<td>Fast protein liquid chromatography (FPLC) ÄKTA™ pure with Unicorn® Software</td>
<td>GE Healthcare</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Gel casting</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoefer Mighty Small II SE250 Multiple gel castor (4 8x7 cm gels)</td>
<td>GE Healthcare</td>
</tr>
<tr>
<td>Horizontal DNA gel castor, Mini-Sub® Cell GT Cell</td>
<td>Bio-Rad</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Gel imaging</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gel Doc™ XR+ (with Image Lab™ Software)</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>LabScanner III, LabScan™ 6.0 Software</td>
<td>GE healthcare</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Gel running</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoefer SE250 Mighty Small II mini vertical gel electrophoresis unit</td>
<td>GE Healthcare</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Heating blocks</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Boekel</td>
<td>Thermofisher</td>
</tr>
<tr>
<td>Select BioProducts</td>
<td>Lab Supply</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Homogenisers</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Glass pestle, 30 mL (Dounce style)</td>
<td>Thermofisher</td>
</tr>
<tr>
<td>Micropestle (plastic)</td>
<td>Eppendorf</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Microscopes</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Leica MZ75</td>
<td>VWR</td>
</tr>
<tr>
<td>Nikon SMZ-U</td>
<td>VWR</td>
</tr>
<tr>
<td><strong>pH meter</strong></td>
<td>Thermofisher</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Orion 3 star</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Shakers</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>New Brunswick Innova™ 2300 (platform)</td>
<td>Eppendorf</td>
</tr>
<tr>
<td>New Brunswick Innova™ 4300 (incubator)</td>
<td>Eppendorf</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Sonicators</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sonics vibra cell™ CV33</td>
<td>Sonics and Materials</td>
</tr>
<tr>
<td>Sonifier® Branson</td>
<td>Thermofisher</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Spectrophotometers</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharmacia ultrospec II</td>
<td>GE Healthcare</td>
</tr>
<tr>
<td>Multiskan™ GO plate reader</td>
<td>Thermofisher</td>
</tr>
<tr>
<td>NanoDrop® ND-1000 UV-Vis</td>
<td>Thermofisher</td>
</tr>
<tr>
<td>UV light platform, Hoefer Mighty Bright</td>
<td>GE Healthcare</td>
</tr>
<tr>
<td>Vortex, MS1 minishaker</td>
<td>IKA</td>
</tr>
<tr>
<td>Water purification, Milli-Q®</td>
<td>Merck</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>X-ray crystallography equipment</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Liquid handler, Mosquito (v 3.4.1 43716)</td>
<td>TTP LabTech</td>
</tr>
<tr>
<td>Rock Imager® R1-1000 with Rock Maker® Software</td>
<td>Formulatrix</td>
</tr>
<tr>
<td>X-ray detector, Rigaku R-Axis IV ++</td>
<td>Rigaku</td>
</tr>
<tr>
<td>X-ray generator, MicroMax-007 HF with CrystalClear 2.0 Software</td>
<td>Rigaku</td>
</tr>
</tbody>
</table>

For sources of general consumables see Table 2.4 and crystallography consumables see Table 2.5. A list of equipment used is found in Table 2.6.
Table 2.7 *E. coli* strains used.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Antibiotic resistance</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL21(DE3)</td>
<td>None</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>BL21(DE3) pLysS</td>
<td>Chlormaphenicol</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>DH5α</td>
<td>None</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>Origami™ (DE3)</td>
<td>Kanamycin, Tetramycin, Streptomycin</td>
<td>Day Lab</td>
</tr>
</tbody>
</table>

Almost all *E. coli* cell lines from Table 2.7 were sourced from Invitrogen, a subsidiary of Thermofisher. Origami™ (DE3) was sourced from the Day Lab of the Biochemistry Department at the University of Otago.

Table 2.8 Plasmids used.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Antibiotic resistance</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET-21a+ with <em>A. thaliana</em> PCO1</td>
<td>Ampicillin</td>
<td>GenScript</td>
</tr>
<tr>
<td>pET-21a+ with <em>A. thaliana</em> PCO2</td>
<td>Ampicillin</td>
<td>GenScript</td>
</tr>
<tr>
<td>pET-21a+ with <em>A. thaliana</em> Δ2-51 PCO1</td>
<td>Ampicillin</td>
<td>GenScript</td>
</tr>
<tr>
<td>pET-21a+ with <em>A. thaliana</em> Δ2-51, Δ247-293 PCO1</td>
<td>Ampicillin</td>
<td>GenScript</td>
</tr>
<tr>
<td>pET-21a+ with <em>Z. mays</em> PCO2</td>
<td>Ampicillin</td>
<td>GenScript</td>
</tr>
<tr>
<td>pET-21a+ with <em>Z. mays</em> Δ1-32 PCO2</td>
<td>Ampicillin</td>
<td>GenScript</td>
</tr>
<tr>
<td>pET-21a+ with <em>O. sativa</em> Δ1-34 PCO2</td>
<td>Ampicillin</td>
<td>GenScript</td>
</tr>
<tr>
<td>pET-24d+</td>
<td>Kanamycin</td>
<td>Poulter Lab</td>
</tr>
</tbody>
</table>

All constructs were purchased in pET-21a+ plasmids (Table 2.8). Plant origin and isoform is stated for each construct. All constructs were optimised for *E. coli* codon usage. Proteins with ΔX-X denotes what residues were removed for shortened constructs. The plasmid pET-24d+ (Table 2.8) was sourced from the Poulter Lab of the Biochemistry Department at the University of Otago.
Table 2.9 Antibiotic stocks.

<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Preparation (1000 ×)</th>
<th>Working concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin (amp)</td>
<td>1 g of amp powder 10 mL of water</td>
<td>100 µg/mL</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>0.5 g of chloramphenicol powder 10 mL of 100 % ethanol</td>
<td>50 µg/mL</td>
</tr>
<tr>
<td>Kanamycin (KAN)</td>
<td>0.5g of KAN powder 10 mL of water</td>
<td>50 µg/mL</td>
</tr>
</tbody>
</table>

Details on antibiotic stock preparation can be found in Table 2.9.

Supplier details

**Beckman Coulter Inc. USA.**
Website: www.beckmancoulter.com
Email: sales_aust_nz@beckman.com

**Merck, USA.**
Website: www.merckmillipore.com
Email: orders.nz@merckgroup.com

**Bio-Rad Laboratories, Inc. USA.**
Website: www.bio-rad.com
Email: sales.nz@bio-rad.com

**MiTeGen, USA.**
Website: www.mitegen.com
Email: info@mitegen.com

**Eppendorf, Germany.**
NZ subsidiary: Eppendorf South Pacific Pty.
Website: www.eppendorf.com/AU-en
Email: info@eppendorf.com.au

**Molecular Dimensions, USA.**
Website: www.moleculardimensions.com
Email: enquiries@moleculardimensions.com
<table>
<thead>
<tr>
<th>Company</th>
<th>Website</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formulatrix, USA.</td>
<td><a href="http://www.formulatrix.com">www.formulatrix.com</a></td>
<td><a href="mailto:sales@formulatrix.com">sales@formulatrix.com</a></td>
</tr>
<tr>
<td>New England BioLabs, Inc. (NEB) USA.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE Healthcare Ltd. USA.</td>
<td><a href="http://www.GEhealthcare.com">www.GEhealthcare.com</a></td>
<td><a href="mailto:sales.nz@ge.com">sales.nz@ge.com</a></td>
</tr>
<tr>
<td>Rigaku Corporation, Japan.</td>
<td><a href="http://www.rigaku.com/en">www.rigaku.com/en</a></td>
<td><a href="mailto:info@rigaku.com">info@rigaku.com</a></td>
</tr>
<tr>
<td>Genscript Biotech, USA.</td>
<td><a href="http://www.genscript.com">www.genscript.com</a></td>
<td><a href="mailto:info@genscript.com">info@genscript.com</a></td>
</tr>
<tr>
<td>Roche Ltd. NZ.</td>
<td><a href="http://www.roche.co.nz">www.roche.co.nz</a></td>
<td><a href="mailto:nz.info@roche.com">nz.info@roche.com</a></td>
</tr>
<tr>
<td>Gold Biotechnology, Inc. USA.</td>
<td><a href="http://www.goldbio.com">www.goldbio.com</a></td>
<td><a href="mailto:contactgoldbio86@goldbio.com">contactgoldbio86@goldbio.com</a></td>
</tr>
<tr>
<td>Scharlau Chemie, Spain.</td>
<td><a href="http://www.scharlab.com">www.scharlab.com</a></td>
<td><a href="mailto:sales@chemsupply.com.au">sales@chemsupply.com.au</a></td>
</tr>
<tr>
<td>Hampton Research, USA.</td>
<td><a href="https://www.hamptonresearch.com/">https://www.hamptonresearch.com/</a></td>
<td><a href="mailto:info@hrmail.com">info@hrmail.com</a></td>
</tr>
<tr>
<td>Sigma-Aldrich, USA.</td>
<td><a href="http://www.sigmaaldrich.com">www.sigmaaldrich.com</a></td>
<td><a href="mailto:anzcs@sial.com">anzcs@sial.com</a></td>
</tr>
<tr>
<td>Hoefer Inc. USA.</td>
<td><a href="http://www.hoeferinc.com">www.hoeferinc.com</a></td>
<td><a href="mailto:sales@hoeferinc.com">sales@hoeferinc.com</a></td>
</tr>
<tr>
<td>Sonics and Materials, Inc. USA.</td>
<td><a href="http://www.sonics.com">www.sonics.com</a></td>
<td><a href="mailto:info@sonics.com">info@sonics.com</a></td>
</tr>
</tbody>
</table>
2.2 DNA work

All DNA sequencing was performed by Genetic Analysis Services, based at the University of Otago. Primers used were purchased T7 forward and reverse primers. Much of the work within this section uses methods adapted from Sambrook et al., (1989).

2.2.1 DNA buffers
Tris Acetate-EDTA (TAE) buffer (50×)
242 g tris base
57.1 mL acetic acid
100 mL of 0.5 M EDTA (pH 8)
Made up to 1 L with water
This buffer was used in the preparation of agarose gels, as well as the gel running buffer.

**Agarose gel (1%)**
- 0.5 g agarose
- 50 mL 1× TAE buffer

**DNA gel sample buffer (6×)**
- 25 mg bromophenol blue
- 25 mg xylene cyanol FF
- 2.5 g sucrose
- 7 mL water

### 2.2.2 Plasmid design

Synthetic plant CDO genes were purchased from GenScript with a C-terminal Strep-Tag® II within expression plasmid, pET-21a+. Strep-Tag® II is a synthetic octapeptide (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) with an intrinsic affinity toward Strep-tactin® resin, which contains an active group consisting of engineered streptavidin (refer to 2.4.2).

Constructs ordered consisted of full length PCO1 (At5g1512) and PCO2 (At5g3989) from A. thaliana. Shortened constructs of A. thaliana PCO1 were ordered which had residues 2-51 removed for one construct, and these residues alongside residues 247-293 removed for the second construct. A predicted PCO2 isoform from Zea mays (maize) was ordered as a full length construct. Shortened constructs from maize and Oryza sativa (rice) were ordered with residues 1-32 and 1-34 removed respectively. These constructs also had restriction sites which could permit the removal of the Strep-Tag®, and subsequent addition of a hexa-histidine tag.

As plasmids arrived lyophilised, tubes were centrifuged at 6000 g for 1 min at room temperature. Subsequently 20 μL of Tris-EDTA (TE) buffer was then added and the plasmids were resuspended by vortexing.
2.2.3 Plasmid purification

All plasmids were amplified and purified following transformation into DH5α cells. In a growth tube, 5 mL of 1× LB was inoculated with a single colony from the transformation using a wire loop, and the appropriate stock antibiotics were added. This was incubated overnight (16-20 h) at 37°C on a shaker at 200 rpm. Three millilitres of overnight culture was centrifuged down 1 mL at a time, and preparation was carried out using an Omega bio-tek ENZA® plasmid mini kit I as per the kit instructions. The wash step was repeated, and 40 μL of elution buffer was added. This was left to incubate for 30-40 min before the final elution.

2.2.4 Horizontal TAE agarose gels

Agarose was prepared by heating the solution in the microwave for around 1 min (until bubbling), then 2.5 μL of 10 mg/mL ethidium bromide was added. The solution was swirled and let to cool for approximately 5 minutes before being poured in the Bio-Rad® gel caster. The comb was added once it was poured, and any bubbles were removed using a sterile pipette tip. Once set, (around 20 min later) the comb was removed, running buffer was added and samples were loaded. Gels were run at 100 V for 45 min, and imaged on a BioRad® Gel Doc system. The ladder used was BioRad® 1 kb plus, which was prepared as per the manufacturer’s instructions.

2.3 Bacterial Growth

2.3.1 Bacterial work reagents

Lysogeny broth (LB) (1x)
1% (w/v) peptone
1% (w/v) sodium chloride
0.5% (w/v) yeast

This was dissolved in water, and autoclaved before use.

Agar plates
1.5% (w/v) agar was added to 1× LB before autoclaving
Stock antibiotics were added to plates once the liquid agar had cooled to a temperature of 45-50°C. Plates without antibiotics could be poured without the need for cooling. Plates were poured under flame for sterility.

### 2.3.2 Competent cell preparation

In a growth tube, 5 mL of 1× LB was inoculated using either glycerol stocks which were scraped with a sterile toothpick whilst still frozen, or inoculated using 50 μL of thawed competent cells. The appropriate antibiotics for each bacterial strain were added. This was left overnight at 37°C on a 200 rpm shaker.

A 250 mL flask containing 50 mL of 1× LB was inoculated with 500 μL of the overnight culture, and again, the appropriate antibiotics were added. This was incubated at 37°C on a 200 rpm shaker until an optical density (OD) reading at 600 nm had an absorbance value of approximately 0.5-0.6. The culture was removed from the shaker and immediately transferred to an ice bath.

The culture was poured into a pre-chilled 50 mL Falcon tube and centrifuged at 4°C, and 3220 g for 5 min. The supernatant was discarded and the pellet resuspended in 25 mL of ice cold 0.1 M calcium chloride. This was incubated in an ice slurry for 30 min, then centrifuged again (4°C, 3220 g for 5 min). Supernatant was again discarded and the pellet was resuspended in 2 mL of 0.1 M calcium chloride 20% glycerol. Two-hundred microlitres of cells were dispensed into 1.7 mL microcentrifuge tubes before being snap frozen using methanol and dry ice. Cells were stored at -80°C until required. The method was derived from Dagert and Ehrlich (1979).

### 2.3.3 Transformation

Both methods described below are variations on protocols described by Sambrook et al. (1989).

**Chemical**

Transformation begun with adding 0.5 or 1 μL of stock plasmid to 1.7 mL microcentrifuge tubes, followed by the addition of between 50 and 100 μL of competent BL21 (DE3) *E. coli* cells. DH5α, BL21(DE3) pLysS and BL21(DE3) origami strain
E. coli cells were also used for certain preparations. This was then incubated on ice for 30 min. Samples were subjected to a 42°C heat block for 45 s before being put back on ice for 2 min. LB was added (500 μL of 1×) and all tubes were again incubated, this time on a 200 rpm shaker at 37°C for 60 min. Samples (100 μL of each) were then spread on agar plates with the appropriate antibiotics, using a sterilised spreader. Plates were inverted then left at 37°C overnight. All work was conducted under flame for sterility, and used aseptic technique. Positive controls were produced by plating stock cells on antibiotic free agar to confirm cell viability. Negative controls were produced by plating stock cells on agar that contained the selective antibiotic. This insured that all colonies on the selective media contained the plasmid of interest.

Electro-competent
Once the competent cells (prepared by Malcolm Rutledge) had thawed on ice, 0.5 μL of plasmid was combined with 50 μL of competent cells. This was incubated on ice for 15 min, then the cells were transferred to cuvette and electroporated. One mL of 1× LB was added immediately, and the cells were transferred to a microcentrifuge tube for a 60 min incubation on a 200 rpm shaker at 37°C. Samples were spread as per chemical transformation. Again, both positive and negative control plates were produced, and all work was conducted under a flame where possible.

2.3.4 Induction of protein expression
An overnight culture (generally 50 mL of 1× LB media with 50 μL of the appropriate antibiotic) was inoculated with a single colony from a recent transformation. The newly inoculated media was then left to incubate overnight shaking at 200 rpm at 37°C. Inoculation was performed aseptically.

Large scale expression was then conducted using one of two platforms, either a fermenter (10 L) or in 2 L baffled flasks (each flask contained 1 L of culture). Smaller scale expressions were used for expression trials. Flask volumes here were 100-125 mL in 250 mL flasks and volumes of culture, were scaled appropriately, concentration of antibiotics and IPTG kept constant. During expression trials, expression was conducted at 18, 28 or 37°C for 3, 6 or 22 hours.
The New Brunswick SF-116 Microgen fermenter was heat sterilised while containing 8 L of distilled water. Once sterilised, the temperature was dropped to 37°C and 2 L of sterile 5× LB was added to the 8 L of sterilised distilled water to make 10 L of 1× LB. For this volume, and when using pET-21+ expression vectors with BL21(DE3) cells, amp was added followed by 100 mL of overnight culture. One millilitre of sterile Antifoam A concentrate was then pipetted into the fermenter. This was left at 37°C with the sparger set at 1-2 L/min aeration, and the agitator at 400 rpm until an OD$_{600}$ of 0.6-0.7 was reached. The culture was cooled to 18 or 28°C before 1 mL of 1 M IPTG was added to a final concentration of 100 μM. The culture was left stirring at 400 rpm, and the sparger was set at 1-2 L/min aeration for 6 h.

To express in flasks, 1 L of 1× LB was sterilised in baffled 2 L flasks. Flasks were stored at 37°C overnight before use to bring the broth to temperature for bacterial growth. Ten millilitres of overnight culture was used to inoculate each flask, alongside the addition of the appropriate antibiotics for selectivity. Flasks were left at 37°C on a 200 rpm shaker until an OD$_{600}$ of 0.6-0.7 was reached. Flasks were moved to the expression temperature (either 18, 28 or 37°C depending on the protein), induced with 1 mL of 1 M IPTG and incubated for 6 h.

Cells were pelleted using 1 L centrifuge bottles in the Beckman-Coulter standing centrifuge with the JLA-8.1 rotor. The samples were centrifuged at 4°C and 6000 g for 15 min. The pellet was stored in a plastic bag at -20 °C until required.

**Expression with iron**

In an attempt to increase iron bound to the protein post-purification, changes were made to the expression protocol. Here, four 2 L flasks containing 400 mL of 1× LB were prepared. An iron (III) chloride stock (80 mM) was prepared, and used at a final concentration of 200 μM. The iron supplement was added immediately before incubation, and this flask was induced with 100 μM IPTG. The other 3 cultures were used to test varying IPTG concentrations, inducing with 100, 1 and 0.1 μM once an OD$_{600}$ of ~0.6-0.7 was reached. All cultures were left to express for 6h at 18°C before being
pelleted by centrifugation. Expression levels were analysed via SDS-PAGE. Each pellet was subsequently subjected to protein purification and assessment of iron occupancy.

2.4 Protein work

2.4.1 Protein work buffers
All purification buffers for use with the ÄKTA™ pure were filtered through 0.2 μM cellulose acetate membranes and degassed before use.

**Cell lysis buffer (1×)**
- 4.1 mL 10× strep wash buffer
- 203 μL TWEEN® 20
- 660 μL 300 mM DTPA
- 22.7 mL water
- 1 × cOmplete™, Mini, EDTA free Protease Inhibitor Cocktail

**Strep-Tactin® wash buffer (1×)**
- 100 mM tris pH 8
- 150 mM sodium chloride
- 1 mM EDTA

**High salt Strep-Tactin® wash buffer (1×)**
- 400 mM tris pH 8
- 150 mM sodium chloride
- 1 mM EDTA

**Strep-Tactin® elution buffer (10×)**
- 1 M tris
- 1.5 M sodium chloride
- 10 mM EDTA
- 20 mM biotin

The elution buffer was purchased prepared from IBA.
Bench top regeneration buffer (1×)
HABA was mixed from a 10 × stock purchased from IBA.

ÄKTA™ regeneration buffer (1×)
0.5 M sodium hydroxide

Refolding buffer (2×)
40 mM tris pH 8
1 cOmplete EDTA free protease inhibitor tablet
200 μM ammonium iron (II) acetate
0.8 M L-arginine

This buffer should be prepared immediately before use to prevent iron (III) formation, and to ensure optimal activity of protease inhibitors.

SDS-PAGE buffers
For two 12% acrylamide sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gels:

Resolving gel
1.25 mL 0.5 M tris pH 6.8
0.667 mL 30% acrylamide/Bis solution 37:5:1
50 μL 10% (w/v) SDS
3.033 mL water
50 μL 10% (w/v) APS
5 μL TEMED

Stacking gel
2.5 mL 1.5 M tris pH 8.8
4 mL 30% acrylamide/Bis solution 37:5:1
100 μL 10% (w/v) SDS
3.4 mL water
100 μL 10% (w/v) APS
10 μL TEMED
**SDS-PAGE running buffer (1×)**
195 mM glycine
50 mM tris
1% (w/v) SDS

**SDS-PAGE sample buffer (5×)**
60 mM tris pH 6.8
2% (w/v) SDS
0.1% (w/v) bromophenol blue
32.5% (v/v) glycerol
5% (v/v) BME

**SDS-PAGE marker preparation (1×)**
5 μL of stock Bio-Rad® broad range molecular weight marker
20 μL 5× sample buffer
75 μL sterile water

After combining materials, samples were boiled at 95 °C for 15 min and centrifuged at 13000 g for 1 min. Aliquots (5.2 μL) were prepared and frozen at -20 °C for later use.

**Coomassie Blue stain (1×)**
250 mL propan-2-ol
100 mL acetic acid
0.5 g Coomassie Brilliant Blue R250

This was made up to 1 L using water, and filtered through 0.2 μm Whatman paper.

**Destain (1×)**
100 mL propan-2-ol
140 mL acetic acid
50 mL glycerol
1.71 L water
2.4.2 Homogenisation and lysis
A cell pellet was resuspended in 1× cell lysis buffer before being homogenised using a glass homogeniser (30 mL of lysis buffer was used per 10 g of pellet). This was followed by sonication to disrupt cell membranes using a Sonifier® at 40% amplitude; 1 s interval pulsing for 30 s to 5 min. This was followed by centrifugation at 4°C and 6000 g for 10-45 min. Purification then continued on one of two platforms, a bench top method that uses gravity flow or using FPLC on an ÄKTA™ pure.

2.4.3 Syringe purification
Purification harnesses the Strep-Tag® present on the protein as it binds to Strep-Tactin® Sepharose® resin with high affinity (Figure 2.1). For further details on Strep-Tag® purification see (Skerra and Schmidt, 2000; Schmidt and Skerra, 2007). After homogenate centrifugation, supernatant was bound to Strep-Tactin® Sepharose® resin (new resin was washed with 1× Strep-Tag® wash buffer before use). The supernatant was incubated with the resin on a rotating wheel at 30 rpm, 4°C for 30 minutes. This was then washed twice using 1× Strep-Tag® wash buffer. Protein was eluted using 1× Strep-Tag® elution buffer, the active component of which is desthiobiotin. As the resin has a higher affinity for desthiobiotin than the Strep-Tag®, protein was eluted from the column. 1× HABA was used to regenerate the column by removing anything still bound to the resin, as the resin has an even higher affinity for this compound. HABA is removed by washing the resin with 1.5 M Tris pH 10.5 as the high pH interferes with the HABA-resin interaction. Flow-through can then be reapplied to the column if necessary. This method was derived from IBA Life Sciences.
Figure 2.1. Schematic of Strep-affinity method. Recombinant proteins (green sphere) with the Strep-tag® bind Strep-Tactin resin (purple spheres) whilst host proteins are washed out. Desthiobiotin (black tear drop) permits elution of the protein of interest as it has a higher affinity for the resin than the Strep-Tag®. HABA (4'-hydroxyazobenzene-2-carboxylic acid, orange/red square) is then used to regenerate the resin as it disrupts desthiobiotin. HABA is removed by increasing the pH of the wash buffer. Taken from Schmidt and Skerra (2007).

2.4.4 FPLC purification

Once centrifugation of homogenate was complete, the supernatant was filtered using 0.45 μM ReliaPrep™ syringe filters. This was then applied to a 5 mL StrepTrap™ HiTrap™ column. At all times the flow rate was set at 5 mL/min.

The column was prepared by washing with at least 5 column volumes (CV) of water until the conductivity and UV absorption stabilised. The column was then equilibrated with 5 CV of 1× Strep-Tactin® wash buffer. Sample application occurred either with a sample pump (volumes of greater than 4 mL) or using a static loop. Once the sample had been
applied, 10 CV of 1× Strep-Tactin® wash buffer was run through the column. Protein was eluted using 1× Strep-Tactin® Elution buffer for 6 CV. Fraction sizes were generally 2 mL.

Column regeneration begun by washing the column with 5 CV of water, followed by 5 CV of 0.5 M sodium hydroxide. This was washed again with 5 CV of water, or until the conductivity and UV absorption stabilised. The column was often then used for further purification. If the column was to be stored, it was washed with 20% ethanol until the conductivity and UV absorption stabilised. The column was stored in 20% ethanol at 4°C. The method was derived from GE Healthcare Life Sciences Recombinant Protein Purification Principles and Methods handbook.

2.4.5 Dialysis and concentrating
Once protein was eluted from bench top columns, or in fractions from the FPLC, it was dialysed using SnakeSkin 10,000 molecular weight cut off (MWCO) 22 mm inner diameter dialysis tubing. Dialysis was against 20 mM Tris and 50 mM NaCl at pH 8 in a 3 L container with a stirring flea. Approximately 3 buffer changes occurred, at least 2 hours apart but at times overnight. The method was adapted from Bollag et al. (1996).

Dialysed protein was concentrated using 10 kDa Amicon® Ultra spin filters of various volumes (0.5, 6 and 15 mL). Membranes were first wet by running a small volume of water through the filter. The sample was then applied, and centrifuged at 4°C until the desired final volume was met. Purified and concentrated protein samples were snap frozen in liquid nitrogen and stored at -80°C. Protein concentration was determined using the NanoDrop spectrophotometer which scanned wavelengths between 220 and 360 nm. The absorbance value at 280 nm was recorded and used with the predicted extinction coefficient for each protein as determined by Expasy (see 2.8.3). Results from the NanoDrop spectrophotometer were confirmed using a dilution series SDS-PAGE gel.
2.4.6 Gel filtration chromatography

This type of chromatography separates proteins based on size. The column matrix consists of hollow beads with varying channel sizes that act as a ‘molecular sieve’. For further detail about this technique see Hagel (2001).

**Superdex 75**

In an attempt to separate CDO and the co-purifying chaperones, gel filtration was used. Two millimolar adenosine triphosphate (ATP) was added to the protein sample, and this was incubated on ice for 15 min. This sample was then applied to a HiLoad 16/60 Superdex 75 column which had been equilibrated with filtered and degassed 50 mM sodium chloride, 20 mM tris-HCl pH 8 buffer. The gel filtration was also conducted with this buffer with a flow rate of 1 mL/min.

A second gel filtration was conducted with varying sample preparation. In this instance, the protein was prepared by dialysing overnight using a Slide-A-Lyzer in 3 mM EDTA, 50 mM NaCl and 20 mM Tris-HCl pH 8. A single buffer change occurred as the same volume was only 3 mL and the dialysis buffer volume was 3 L. A charcoal slug was present during dialysis to remove excess ions chelated by EDTA or nucleotides. This sample was then loaded on the Superdex 75 with the same buffer and flow rate as the previous protocol.

**Superdex 200**

This was run to determine if purified protein was forming larger complexes or aggregating. The column was run in 20 mM Tris pH 8 and 50 mM sodium chloride with a flow rate of 1 mL/min. Once equilibrated with this buffer, 1 mL of purified PCO1 protein was added to the column. Fractionation volume was 2 mL and samples were run on a gel once eluted.

2.4.7 Cation exchange chromatography

First, a Dnase I digestion was conducted using 1 μL of a 1 mg/mL Dnase I stock and 20 mM magnesium chloride. This was left to incubate at room temperature (RT) for 1h. The cation exchange column employed was an SP HP 5 mL HiTrap. The high salt buffer
consisted of 50 mM HEPES pH 7, 1 M sodium chloride, whilst the low salt buffer consisted of 50 mM HEPES pH 7, 10 mM sodium chloride. The flow rate was 1 mL/min.

The column was washed with water, followed by low salt buffer. The gradient of high salt buffer was then increased until the column was completely in the high salt buffer. A blank run was conducted to ensure the column was clean and establish the population of counter ions on the resin. After this, the protein sample was loaded onto the column and the same protocol run again. Peak fractionation was selected, and fractions were 1.8 mL in size.

2.4.8 SDS-PAGE
Protein gels were traditional Laemmlı-style which were cast and run using a Hoefer electrophoresis system. APS and TEMED were added immediately before pouring the resolving and stacking gel. Approximately 1 mL of propan-2-ol was pipetted on the top of the freshly poured resolving gel. The isopropanol was washed out with water once the gel was set and excess water was blotted with paper towels before the stacking gel was then poured. Combs were then added. Gels were stored at 4°C for up to 2 weeks before use. This method is derived from Bollag *et al.* (1996).

All samples were prepared with a 1 in 5 addition of 5x loading dye and boiled at 95°C for at least 5 min prior to loading (unless stated otherwise). Gels were run in 1x SDS-PAGE running buffer at 100 V until the dye front had run through the stacking gel, and at 200 V thereafter for approximately 45 min. Once the dye front had run off, gels were stopped and transferred to Coomassie Blue stain. After 2 hours, the Coomassie Blue stain was removed, the gel was washed with water and destain was added. Destain was removed once bands were visible and background stain was no longer present. Gel pictures were taken using LabScan.

2.4.9 Solubility testing
Before large scale expression and purification occurred, smaller scale trials were performed. Once samples were produced (2.2.2 and 2.2.3), and expression levels were evaluated using SDS-PAGE (2.4.8), conditions with promising results were subjected to
solubility testing. This was to ensure that the resulting protein was soluble in those conditions.

Testing begun by centrifuging 1 mL samples at 13000 g for 5 min at 4°C. Supernatant was discarded, and the pellet was resuspended using a micropestle in 500 μL of 100 mM sodium phosphate buffer pH 7.2. Cells were lysed using sonication for 30 secs, 30% amplitude, pulsing at 1 sec intervals, using a Sonics VibraCell with a microtip. A 30 μL sample of each homogenate was kept for SDS-PAGE analysis, and the rest of the samples were spun at 13000 g for 5 min at 4°C. A 30 μL sample was taken of each supernatant, and pellets were resuspended in 440 μL of 100 mM sodium phosphate buffer pH 7.2. A 20 μL sample was taken of all pellets, and samples were run on SDS-PAGE to analyse results. Gel samples were prepared by combining 30 μL of each sample with 7 μL of loading dye, and 15 μL of this was loaded onto the gel.

2.4.10 Improving protein solubility during lysis

For the N-terminal truncation of the maize PCO2, six additional lysis conditions were trialled for effect on solubility of expressed protein: regular lysis buffer (2.4.1), regular lysis buffer with 400 mM sodium chloride or 125 mM L-arginine, 100 mM sodium citrate pH 4 and 5 or 100 mM Tris pH 7.1 all with 300 mM sodium chloride and 0.7% TWEEN® 20. Five-hundred microlitres of each buffer was used to resuspend 6 samples of ΔN Maize pellet. Samples were then prepared as stated in 2.4.9.

2.4.11 Improving protein solubility post-purification

Six tubes were prepared with 60 μL of freshly purified protein, and a range of reagents were added at varying concentrations. They are as follows (final concentration listed): ammonium iron (II) acetate 200 μM, sodium chloride 0.4 M, TWEEN® 20 0.7%, BME 100 μM, magnesium chloride 200 mM and DTPA 2 mM. Each sample was left to incubate on ice for 15 min before being centrifuged at 10000 g and 4°C for 5 min. The amount of pellet was visually assessed.
2.4.12 Refolding assay
Stock protein was prepared by combining 70 μL of 150 μM PCO1 with 98 μL of 2× refolding buffer. The stock solution was then split by adding 40 μL to 4 microcentrifuge tubes. Different ratios of oxidised to reduced glutathione were then employed (stock GSSG and GSH concentrations were 50 mM). Tube 1 add: 5 mM GSSG, 0.5 mM GSH, tube 2: 5 mM GSH, 0.5 mM GSSG, tube 3: 2 mM GSSG, 2 mM GSH and tube 4: no glutathione control. Potassium chloride was added to all tubes for a final concentration of 62.5 mM. Samples were incubated at 4°C overnight on a vortex with low speed shaking.

2.4.13 TCA precipitation
Inducing protein precipitation helps one determine if protein is present within a sample, even if the amount of protein present is very low. This is because the protein can be pelleted and resuspended in a smaller volume for analysis. To begin, samples were brought up to a volume of 1 mL. One hundred microlitres of 100% TCA was added to each tube and the samples were vortexed. The samples were incubated at -20°C for 20 min, then spun at 10000 g for 5 min. The supernatant of each sample was removed, and all pellets were resuspended in 20 μL of 0.1 N sodium hydroxide. The resuspension was run on a gel.

When the protein concentration is predicted to be below 5 μg/mL (the sensitivity of the TCA precipitation), but greater than 1 μg/mL a variation of the TCA precipitation can be employed. Once the sample was brought up to 1 mL, 100 μL of 0.15% DOC was added and the tube vortexed. Incubation occurred at RT for 10 min, then 50 μL of 100 % TCA was added, the tube was vortexed and the instructions were followed from the TCA addition as specified above. Both methods were taken from Bollag et al. (1996).

2.4.14 Dynamic light scattering (DLS)
DLS instruments are used to determine particle size and distribution in solution. One hundred microlitres of 2 mg/mL protein solution was prepared. Protein concentration was judged by absorbance at 280 nm, and SDS-PAGE gels. Protein was centrifuged at 13000 g and 4°C for 30 min. Water used as a control was spun down in the same way.
The cuvette used in the machine was cleaned thoroughly with 20% ethanol followed by water. Once rinsed, 80 μL was added to the cuvette, the cuvette exterior was wiped clean with lint free tissues and data were recorded for 4 min at RT. Results were then assessed and recorded.

2.5 Crystallisation

2.5.1 High throughput screening

Before high throughput screening, pre-screens were conducted. This is used to evaluate whether or not large scale screening would be worthwhile with the protein preparation. The PCT™ Pre-Crystallization Test produced by Hampton Research was used, and the manufacturer’s instructions were used to perform the test.

Many high-throughput screens were used for initial crystal screening. This includes: Hampton Research HT, PEG/ion HT, PEGRx HT, JCSG Plus and PACT Premier screens. Protein preparation for crystallography varied. Stock protein concentration was using 5 mg/mL and a range of different metals including nickel, cobalt and iron were added to the stocks with varying stoichiometry. Peptide substrate was also included in some of the protein stocks used for setting up plates.

To set up screens, 25 μL of each solution was added to the reservoirs of a Swissci MRC 3 drop 96 well plate. TTP Labtech’s Mosquito® liquid handling robot was used for protein and solute addition. Generally, the drop sizes were approximately 600 nL and each of the three drops had varying ratios of well solution and protein. All drops prepared using the Mosquito® were sitting drops and employed vapour diffusion. Drops were imaged using the daily/weekly/nightly schedule of the Rock Imager® software. Images were taken using visible light, UV light and under cross polarised conditions.

2.5.2 Varying condition plates

Any drops that showed interesting results with commercial screens were attempted to be reproduced using lab reagents. Conditions of interest were not only copied but varied in an attempt to form crystals or try improve crystallisation. Factors that were varied were
generally buffer pH range, and concentration of precipitant. A Morpheus® additive screen was also used by adding it to a condition of interest at a concentration of 10% v/v.

Both 24- and 96-well plate formats were used. Twenty-four well plates were prepared manually and the drop size was 1 μL of protein mixed with 1 μL of the solute of interest. The Mosquito was used to set up 96-well plates with conditions of interest. The 24-well varying plate solutions were prepared on the plate, while 96-well plates were prepared on a separate 96-well plate so excess could be prepared. Twenty-four well plates used a hanging drop vapour diffusion technique, whilst the 96-well plate format used sitting drop vapour diffusion.

2.5.3 Mounting crystals
Any crystals that were formed and were of a large enough size were attempted to be mounted. A range of loop sizes were used, but generally those smaller than 100 μm. MiTeGen MicroLoops™ were the preferred style. Both RT and frozen crystals were prepared for assessment on the Otago University Biochemistry Department home source. At room temperature, plastic sleeves (Mitogen) were used to minimize evaporation. Crystals were frozen by plunging into liquid nitrogen immediately after mounting in loops. All crystals that were taken to the Australian synchrotron were frozen. Addition of cryoprotectant was always considered, with 10-20% glycerol generally being used when required.

2.5.4 X-ray diffraction
At the University of Otago Biochemistry department, crystals were manually placed on the goniometer and aligned. Single-crystal diffraction data was collected using Cu Kα radiation at the University of Otago (see Table 2.6 for details of equipment). At 93 K and a wavelength of 1.542 Å, 140 images were collected at an exposure time of ten minutes, using a detector distance of 250 mm.

At the Australian Synchrotron the MX2 beam-line was used for crystal screening. MX2 featured a silicon double-crystal monochromator and an Eiger direct detector. Radiation of wavelength 0.9537 Å (13 keV), usual exposure, 1 second/degree of oscillation and beam attenuation of 80%. The detector was set 250 mm from the crystal. A crystal
mounting robot placed bases with mounted crystals on the goniometer. Blu-Ice software was used to direct a crystal mounting robot orientate the crystal and direct data collection.

2.6 Assays

2.6.1 Ellman’s assay

Ellman’s reagent
23.1 mg DTNB
2.9 mg EDTA
50 mL of 100 mM sodium phosphate at pH 7.5

The substrate of plant CDO contains an N-terminal cysteine, which is characterised by a thiol group. This thiol group becomes dioxidised during catalysis, forming a sulfonyl group. Ellman’s assay measures free thiol groups which react with DTNB to form a coloured product (Figure 2.2). As the substrate of CDO contains a free thiol group, this assay reports substrate depletion. A standard curve is employed to determine unknown free thiol concentrations at a given absorbance (Figure 2.3).

In a 96 well plate, 97.5 μL of reagent was added to each well. Sample volume added to each well for varying time points was 2.5 μL. Sample addition to the Ellman’s reagent quenched the reaction. Assays were conducted in a dark box and samples taken every 2 minutes for between 12 and 48 minutes. Samples were read at 412 and 750 nm on a MultiSkan GO 96-well plate spectrophotometer. Ellman’s reagent absorbs maximally at 412 nm and 750 nm can show if there any air bubbles or other inconsistencies in the samples. Data were analysed using Microsoft Excel and GraphPad prism. For further details of this assay, see Fellner et al. (2014).

Figure 2.2. Ellman’s reaction. The presence of a free thiol (A) breaks the colourless DTNB (B) compound into a mixed disulfide (C) and yellow coloured 2-nitro-5-thiobenzoate (TNB) product (D). Adapted from Nieri et al. (2017).
Figure 2.3. L-cysteine standard curve. Line slope 0.0817 ± 0.0002. Graphed are three technical replicates, error bars representing standard error of the mean are shown but are not visible outside symbol. The standard curve reports the concentration of cysteine in the cuvette.

2.6.2 Ferrozine

Ferrozine reagent
200 µL 37 mM ferrozine
200 µL 1.14 M ascorbic acid
1 mL 2 M ammonium acetate pH 9

Ammonium iron (II) sulfate stock (30mM)
11.8 mg ammonium iron (II) sulfate
1 mL water

This was prepared fresh every time, as iron (II) is readily oxidised to become iron (III) in the environment.

Chelex resin preparation
A 50 mL falcon tube was filled to the 25 mL mark with resin. This was then topped up to 50 mL with 100 mM Tris-HCl pH 8 and the tube was shaken vigorously by hand many times. This was then spun at 3220 g for 5 min at room temperature, separating the beads and buffer. The buffer was poured off and the process was repeated twice. Once washed
three times, the beads were topped up to 50 mL with fresh 100 mM Tris-HCl pH 8 and stored up to 30 days at 4°C.

The ferrozine reagent measures the presence of Fe\(^{2+}\) by binding and forming a coloured complex (Figure 2.4). It forms a stable, soluble complex with ferrous iron that absorbs strongly at 562 nm without a large influence from other divalent cations (Stookey, 1970). In this assay, a standard curve is employed to determine an unknown iron concentration at a given absorbance (Figure 2.5). For the preparation of samples, 5 μL of protein, 5 μL TRIS buffer (100 mM, pH 8) and 5 μL of concentrated sulphuric acid were combined. Samples were incubated at 97°C for 30 minutes, after which the samples were centrifuged at 4°C, 13000 \(g\) for 1 minute. After centrifugation, 140 μL of ferrozine reagent was added, samples were vortexed and centrifuged again. In a 96 well plate, 120 μL of each sample was added to each well and the results were read at 562 and 750 nm. Data was analysed using Microsoft Excel, and GraphPad Prism (version 7). This method follows that used previously in the Wilbanks Laboratory (Tchesnokov et al., 2012).

![Ferrozine compound](image)

**Figure 2.4. Ferrozine compound.** Nitrogen that chelate the ferrous iron are shown in red. Iron has four binding sites, so two of these molecules come together to co-ordinate the metal ion.

Purified protein samples that required iron loading, had a stoichiometric amount of stock 30 mM ammonium iron (II) sulfate added. This was added to the protein and was left to incubate on ice for 10 min. Samples were applied to chelex resin to remove all unbound iron. Within 0.2 μM spin filters pin filters, 1 mL of resin mix (1:1 resin to buffer) was spun until dry. The incubated sample was then added to the resin (usually no more than 200 μL to ensure all excess Fe\(^{2+}\) was removed) and this was centrifuged at 13000 \(g\) for 2
min at 4°C. Once the sample are loaded with iron, a ferrozine assay was performed to analyse the occupancy.

![Figure 2.5. Ferrozine standard curve.](image)

Figure 2.5. Ferrozine standard curve. Line slope 0.00030 ± 0.000002. Graphed are two technical replicates, error bars representing standard error of the mean are shown. Standard curve does not account for dilution factor.

### 2.6.3 Limited proteolysis

This was conducted to determine if there were poorly folded domains of PCO1 (Fontana et al., 2004). To begin, 60 μL of 200 μM PCO1 was aliquoted into 5 microcentrifuge tubes. Four microlitres of 1 mg/mL proteinase K was added to four tubes (one tube without proteinase K was included as a control), which were then mixed by inversion. Each sample was left to incubate for either 5, 20, 60 or 120 min.

At the end of each incubation time, the reaction was quenched with 100 μL of TCA and was left at -20°C for greater than 20 min. Samples were centrifuged at 12500 g for 10 min at 4°C. The supernatant was discarded and the pellet was resuspended in 100 μL of 100% ethanol. Centrifugation was repeated at the same specifications and again the supernatant was discarded. The pellet was left to dry at 65°C for 5 min with the microcentrifuge lid open for evaporation. After drying, the pellet was resuspended in 20 μL of 5× SDS sample buffer and 0.450 μL of 1 M sodium hydroxide to adjust the pH. Samples were incubated at 95°C for 5 min before being run on a gel.
2.7 Nuclear Magnetic Resonance (NMR) Spectroscopy

2.7.1 \(^1\)H NMR spectroscopy

A 10 mM L-cysteine and 15 μM plant CDO sample was prepared by incubation for 2 h at room temperature. Standard samples of L-cysteine, cystine, Tris and CSA at 10 mM were also prepared. All samples were in 100 mM sodium phosphate at pH 7.2, with H\(_2\)O as the solvent. Once the samples were loaded, \(^1\)H NMR was conducted using a 500 MHz magnet. Trimethylsilylpropanoic acid (TSP) was used as the reference and solvent suppression was conducted using pre-saturation.

2.8 Computational work

2.8.1 Bioinformatics

To formulate multiple amino acid sequence alignments Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo) was used, employing the function clustal w/ numbers (Sievers et al., 2014). The protein basic local alignment search tool (blastp) (blast.ncbi.nlm.nih.gov/Blast.cgi) was used to find similar amino acid sequences from an extensive database (Altschul et al., 1990). FASTA format was used to input sequences for both sites (Lipman and Pearson, 1985). The uniprot sever (www.uniprot.org) was used to search for protein sequences using only the gene or protein name (Consortium, 2017).

2.8.2 Modelling

Both swiss-model (swissmodel.expasy.org) (Arnold et al., 2006) and iterative threading assembly refinement (I-TASSER) (zhanglab.ccmb.med.umich.edu/I-TASSER) (Zhang, 2008) were used to create homology models. Both sites are fully automated, and use similar protein sequences and structures as a model. No templates were specified and FASTA format was used to input sequences for both sites.

UCSF-chimera software was used for the visualisation and analysis of homology models (Pettersen et al., 2004). Ribbon diagrams were mainly produced, with some surface
topology analysis. Commands were also used to permit the mapping of surface charge residues, and the mapping of surface hydrophobicity.

2.8.3 Protein sequence-based prediction

To predict general characteristics based solely on amino acid sequence Expasy ProtParam (web.expasy.org/protparam) (Gasteiger et al., 2005) was used. This tool computes a pI, molecular weight and other physical and chemical parameters. Predictor of Natural Disordered Regions (PONDR) (www.pondr.com) VLXT function (Obradovic et al., no date) was used to determine the likelihood and protein instability and disorder. This was based on known disordered proteins and their sequences. Secondary structure was predicted using PSIPRED v3.3 (bioinf.cs.ucl.ac.uk/psipred) (Buchan et al., 2013) with raw amino acid sequence as the input. This software uses known protein sequences and structures to determine the likelihood of secondary structure elements within the given protein sequence. Scanning the amino acid sequence for predicted motifs (based off known sequence motifs) was conducted using Expasy ScanProsite (prosite.expasy.org) (Sigrist et al., 2012). Motifs with a high probability of being present were excluded.

2.8.4 Plasmid sequencing

Plasmids that were sequenced (sequencing done by GAS at The University of Otago), were examined using 4Peaks software. Sequence quality was visually assessed and the nucleotide sequence was copied into the Expasy translate tool (web.expasy.org/translate). This tool converts nucleotide sequences to amino acid sequences. The amino acid sequence could then be compared to the proposed sequence using either a pBLAST search (2.8.1) or a direct comparison using clustal omega (2.8.1). If the reverse primer sequence was being analysed, the raw nucleotide sequence was converted to the reverse complement (www.bioinformatics.org/sms/rev_comp) (Stothard, 2000) before being translated from a nucleotide to amino acid sequence.

2.8.5 Other

ImageJ was used to label and manipulate SDS-PAGE gels.
2.9 Mass spectrometry (MS)

All equipment and chemicals listed below are not found in the materials section as they were provided by the centre for protein research (CPR) at The University of Otago. All work was conducted by the CPR unless stated otherwise.

2.9.1 Identification of proteins

To identify bands from SDS-PAGE, trypsin and/or chymotrypsin digestions were performed using automated liquid handling. Liquid chromatography mass spectrometry (LC-MS) was used as it allows for physical separation based on size (LC), followed by determination of fragment size using a mass spectrometer. This protein profiling at the CPR uses an UltiMate™ 3000 RSLCnano ultra-high performance liquid chromatography (uHPLC) system inline coupled to nanospray LTQ-Orbitrap XL™ mass spectrometry. Once fragments were identified, protein identification was done using sequence database dependent search engines, including MASCOT and SEQUEST.

To determine N-terminal truncation sites of the full length maize construct, N-termini were tagged in gel using sodium cyanoborohydride to dimethylate N-termini (lysines will also be dimethylated). After in gel N-terminal labelling of peptides, tryptic digestion was used to produce peptide fragments. Once fragmented, peptides were identified using liquid chromatography-tandem mass spectrometry (LC-MS/MS). MS/MS first separates ionised fragments using a mass to charge ratio, followed by collision-induced dissociation and which causes ion separation that is detected by a second mass spectrometer. This allows for increased sensitivity. Peptides with an N-terminal label were then identified, permitting determination of where the N-terminal region is being cleaved.

2.9.2 Identification of disulfide bonds

To identify disulfide bonds within proteins, differential labelling followed by MS was used. The first label forms bonds only with reduced cysteines as they have a free thiol group. The label is not able to interact with cysteines that have formed a disulfide bond as it is protected. A reducing agent is added to break these disulfide bonds, and a second heavy label is added which can now interact with these free thiol groups. Once both labels
have been added, the peptide mass is determined. One can then tell the difference between the light and heavy labelling, and distinguish which state each cysteine was in.

After labelling occurred, samples were digested using chymotrypsin and trypsin in parallel. Initial samples were labelled first with light iodoacetamide (IAM). Excess light IAM was removed using a spin filter and washed three times with buffer. Once washed, the sample was reduced using tris(2-carboxyethyl)phosphine (TCEP) and alkylated with heavy iodoacetic acid (IAA). Samples were subjected to tandem mass spectrometry (MS/MS).

2.9.3 Intact mass MS

Intact mass determination was used to decipher which metal ion plant CDO binds. As different metal ions have different weights, the mass difference between protein with no ion bound and protein with an ion bound should indicate which metal is present. For the experiment, PCO1 and Arabidopsis ΔN protein were diluted to a final concentration of 5-10 pM/μL. This was then buffer exchanged from post-dialysis Tris and sodium chloride concentrations, into 5 mM ammonium bicarbonate with trace amounts of sodium chloride. The intact mass was then determined using matrix assisted laser desorption/ionisation-time of flight (MALDI-TOF) mass spectrometry and compared to a calculated mass. MALDI-TOF uses a laser energy absorbing matrix to form ions from intact molecules present, which shift from a liquid to a gas phase. Ions fly toward the detector and their time of flight is indicative of the fragment mass (charges must be taken into account).
Chapter Three

*Arabidopsis thaliana* CDO expression and purification

To characterise plant CDO, it first must be expressed and purified. As cellular levels of plant CDO are usually low, recombinant over-expression was employed. As plant CDO does not appear to require any post-translation modifications, an *E. coli* prokaryotic cell line was used. This allows for speed, ease and low cost of production (Baneyx, 1999). After a period of over-expression in *E. coli*, cells were lysed, releasing plant CDO as it is an intracellular protein. Plant CDO was expressed fused with a *Strep*-tag® made of amino acids to permit affinity chromatography based purification. This tag selectively bound to *Strep*-tactin® resin, separating it from other cellular proteins. Once the resin was washed, plant CDO was eluted and used for further experiments.

*A. thaliana* is a well-studied plant that is recognised as a model organism (Meinke *et al.*, 1998). It is commonly used in plant biology as it is small, easy to grow and phenotypic changes are easily identified (Somerville and Koornneef, 2002). Two isoforms of CDO (PCO1 and PCO2) from *A. thaliana* were the first constructs to be expressed and purified. *A. thaliana* CDO was chosen as Weits *et al.* (2014) worked on both of these isoforms and was able to detect expression *in vivo* and purify these proteins to show catalytic activity *in vitro*. It also appeared to represent CDO from other plant origins well as seen in Figure 3.1.
Figure 3.1. Multiple sequence alignment of PCO1 and PCO2 from various plant origins.
**Figure 3.1. Multiple sequence alignment of PCO1 and PCO2 from various plant origins.** BLAST was used to source predicted PCO1 and PCO2 sequences from various plants. Clustal Omega produced the alignment and numbers correspond to last amino acid on the line. A star represents the same amino acid, whilst two dots represents very similar amino acids and one dot represents similar amino acids. The first 240 residues of Liverwort CDO are not included. Yellow shows conserved cysteines, green predicted catalytic tyrosine residue and blue shows predicted iron co-ordinating histidine residues. Accession numbers are as follows- Apple: XP_008341544, Liverwort: OAE21248, Tomato: XP_004232237 Arabidopsis: Q9LXG9, Rice: XP_015642636 Maize: XP_008673605.

### 3.1 PCO1 and PCO2

First, both constructs were designed using the genes described by Weits *et al.* (2014) and ordered from GenScript. Plasmids containing either the PCO1 or PCO2 gene were transformed into BL21(DE3) *E. coli* cells, which resulted in several colonies. This was used to prepare an overnight culture and a subsequent trial expression was conducted at 3 different temperatures, with samples taken at 3 different time points. The time points and temperatures with the most promising expression levels were subjected to solubility testing. Varying temperatures were tested as this often results in varying levels of protein expression, and also influences protein solubility.
3.1.1 PCO1 and PCO2 expression

Figure 3.2. SDS-PAGE of PCO1 and PCO2 expression at 18°C. Equal number of cells as judged by optical density at 600 nm (OD$_{600}$), were loaded into each lane. The negative sign indicates sample before induction and the numbers following this represents number of hours after induction. Predicted mobility of PCO1 and PCO2 are indicated with black arrows. The lane labelled M contains molecular weight marker. Marker masses are indicated at left in kDa and each band contains 500 ng of protein.

Overall, PCO1 (predicted mass of 34 kDa) expression was greater than of PCO2 (predicted mass of 32 kDa) expression at 18°C (Figure 3.2). Based on the intensity of the bands, optimal levels of PCO1 expression are 6 h or 20 h post-induction. As 6 h of induction is less time intensive yet the same level of expression is achieved, this condition was chosen for solubility testing. PCO2 expression levels appeared low and did not change dramatically between 6 and 20 h post-induction.
Figure 3.3. PCO1 and PCO2 expression at 28°C. Equal number of cells as judged by OD$_{600}$, were loaded into each lane. The negative sign indicates sample before induction and the numbers following this represents number of hours after induction. Predicted mobility of PCO1 and PCO2 are indicated with black arrows. The lane labelled M contains molecular weight marker. Marker as in Figure 3.2.

PCO1 expression at 28°C was greater than that of PCO2 (Figure 3.3). Band density of PCO1 after 6 hours induction indicated that this induction time produced the greatest amount of protein, even with a greater number of cells loaded. Twenty hours of induction also produced a significant amount of PCO1. As 6 h and 20 h both produced reasonable amounts of protein, these samples were subjected to solubility testing. PCO2 expression did not seem to change dramatically based on the amount of time after induction. It appears as though 3 h produced the most protein, but this lane seems to have a greater amount of cells loaded as all bands in this lane are darker.
Figure 3.4. PCO1 solubility. Amount loaded was constant to allow comparison. H indicates homogenate, S represents supernatant and P indicates pellet. Predicted mobility of PCO1 is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 3.2. Bacterial cells lysed using sonication.

The homogenate analysed in Figure 3.2 and Figure 3.3 is a sample of the cell lysate, so indicates total protein content irrespective of solubility. After cell lysis, insoluble protein is spun down with cellular debris and is thus found in the pellet. Protein found in the supernatant is in solution after centrifugation, and therefore it is soluble. There was little protein in the pellet after expression at 18°C for 6 h (Figure 3.4). Both time points at 28°C contained an insoluble fraction of PCO1 as bands appeared in the pellet lanes. The 6 h time sample contained much more insoluble PCO1 as the band is much darker compared to the 20 h sample. This band may contain substantial unlysed cells due to poor sonication, as well as insoluble protein. The 20 h sample still contained a relatively large amount of PCO1 in the supernatant.
Figure 3.5. PCO2 solubility. Amount loaded was constant to allow comparison. H indicates homogenate, S represents supernatant and P indicates pellet. Predicted mobility of PCO2 is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 3.2.

Both 28°C samples resulted in darker bands at the mass of PCO2 in the soluble fraction compared to the insoluble fraction (Figure 3.5). This means that expression under this temperature produces protein that is mostly soluble. The 20 h sample appears to have a slightly greater proportion of protein that is soluble. Expression at 37°C for 6 h resulted in a large band of PCO2 in the pellet lane, indicating that it is mostly insoluble protein under these conditions.

Based on the solubility and expression gels, PCO1 was expressed at 18°C for 6 h. PCO2 was expressed at 28°C for 20 h. The increase in solubility at lower temperatures in not at all surprising as the temperature growth range of A. thaliana is 16-25°C (Arabidopsis Biological Resource Centre, 2018). This suggests that most proteins within A. thaliana work optimally within this temperature range. Previous studies have also shown that expression of recombinant proteins at lower temperatures improves protein solubility (Schein and Noteborn, 1988). However, lowering the temperature comes at a cost as it slows growth and reduces yield. It may be of interest in the future to test the solubility of PCO2 at 18°C to see if solubility is improved compared to expression at 28°C. The 18°C samples were not tested in these experiments due to the lower expression at 18°C compared to 28°C. However, if the solubility of PCO2 increases substantially at the lower expression temperature then more protein may be able to be purified overall.
3.1.2 PCO1 purification

Once ideal expression conditions were determined, both proteins were subjected to trial purification using Strep-tag® affinity purification. Bacterial pellet was lysed using sonication, and the supernatant after centrifugation was purified using Strep-tag® resin or was loaded onto a Strep-tag® HP column on an FPLC.

![Image of gel with bands](image)

**Figure 3.6. PCO1 bench-top syringe purification.** Predicted mobility of PCO1 is indicated with a black arrow. The lane labelled H contains homogenate, S contains supernatant, F contains column flow-through, P contains purified protein and M contains molecular weight marker. Marker as in Figure 3.2.

For the bench-top purification of PCO1, the dark band at ~34 kDa shows that a large amount of PCO1 was present in the homogenate (**Figure 3.6**). A band of a similar intensity was found at ~34 kDa in the supernatant, which means that PCO1 was soluble under these conditions. The dense band of PCO1 in the flow-through shows that not all of the protein bound to the resin. This is likely due to too much PCO1 being present for the resin to bind. Alternatively, this could be due to proteolysis removing the tag required for purification or misfolding resulting in tag inaccessibility.

The protein lane (P) contains a large band suggesting protein that is reasonably pure. However, several contaminant bands can be seen at ~60 and 70 kDa, as well as a number of bands below the PCO1. The bottom bands may contain PCO1 that has been partially degraded by proteases at the N-terminal end. As the tag is found on the C-terminal end, cleavage here would result in protein that is not able to be purified by affinity
chromatography. The top two bands were predicted to contain molecular chaperones that have bound the protein and were pulled out during purification. These bands were excised and sent for identification by MS.

The results showed the top band to be 69kDa *E. coli* DnaK with 78% coverage ([Appendix 2](#)) and the lower band to be 57kDa *E. coli* groEL with 94% coverage ([Appendix 3](#)). Both of these proteins are chaperones, which are some of the most abundant proteins within *E. coli* cells. Under normal conditions DnaK makes up 1.4% of total protein within *E. coli* ([Herendeen et al., 1979](#)). These proteins bind other proteins that are misfolded to prevent intracellular aggregation and encourage proper folding (reviewed by Gasser *et al.* 2008 and Gottesman and Hendrickson 2000). The binding of chaperones to PCO1 was indicative of poorly folded subdomains or unfolded termini ([Hellebust *et al.*, 1990](#)). As these proteins prevented the purified protein from being homogenous, this posed a potential problem for crystallography. Attempts were made to remove these chaperones from PCO1 ([3.2.1](#)).

![Figure 3.7. PCO1 Strep-tag® FPLC purification.](#)

The chromatogram in **Figure 3.7** had an initial flow-through peak (4-20 mL) in the absorbance up to 3250 mAu. This peak likely comprised cellular proteins that do not interact with the resin. The absorbance then decreased as the column was washed. The second peak at about 70 mL elution volume contained PCO1 as determined by SDS-
PCO1 was eluted in a relatively sharp peak. Changes to conductivity were due to changes in buffer conditions.

PCO1 has been purified previously by Weits et al. and White et al. 2014. Weits used a pDEST17 vector and PCO1 was N-terminally tagged with a cleavable hexahistidine tag. Expression was in Rosetta E. coli cells at an unspecified temperature for 18 h. Lysis buffer consisted of 50 mM HEPES-Na pH 7.5, 300 mM NaCl and 5 mM MgCl₂. The protein was purified using a nickel affinity chromatography column. The elution buffer was similar to the lysis buffer but contained 150 mM imidazole and was pH 7. No SDS-PAGE gels were reported, and there was no mention of protein purity. The protein was frozen in 50% glycerol, suggesting some difficulty with protein solubility.

White et al. used the same expression plasmid, but expressed PCO1 in BL21(DE3) cells and specified that expression was at 18 °C. Expression time was the same as Weits et al. Cells were lysed by sonication and the protein was purified using the same nickel affinity chromatography. There was no mention of the purification buffers, but it was declared that the protein was buffer exchanged into 250 mM NaCl and 50 mM Tris (pH 7.5) post-purification. An SDS-PAGE gel of purified protein showed pure (>90%) PCO1 with a very thin band of protein at ~70 kDa. This band could be DnaK, but it was not described. A few bands were present below the full length protein. This suggests these authors also had difficulty with truncations during purification as these bands may be comprised of cleaved full length protein.

Overall the published protocols were similar to that reported here. The expression vector (pDEST17) differed slightly from the one used in these experiments (pET21+). Both vectors contained a T7 promoter which is induced with IPTG, carried amp resistance and permitted the ability to tag the protein of interest either at the N- or C-terminus. The main difference is that the PCO1 used in these experiments had a C-terminal Strep-Tag® as opposed to an N-terminal His-tag. The buffers used also differed, the pH changed by 0.5 pH units, the NaCl concentration was higher and HEPES was used as opposed to Tris (for Weits et al.) As the type of tag and what termini the tag is fused to influences protein solubility the method used above to those
published may aid in producing a more stable protein. Changing buffer conditions can also influence this.

3.1.3 PCO2 purification

**Figure 3.8. PCO2 Strep-tag® FPLC purification.** The blue line represents absorbance at 280 nm in mAU (left axis) and the red line represents conductivity in mS/cm (right axis). Fractions numbers relate to the gel in Figure 3.9. Sample application was followed by a wash with running buffer until 55 mL elution volume, elution buffer was added until 90 mL elution volume, followed by a water wash until 115 mL, then a sodium hydroxide wash from 115-140 mL elution. The last step consisted of a wash with water.

The initial peak in the absorbance (4-18 mL) in Figure 3.8 was the flow-through, and likely comprised cellular proteins that did not interact with the resin. The absorbance then decreased as the column was washed. The second peak at about 60 mL elution volume contained PCO2, which was eluted in a relatively sharp peak with a tail. The tail implies that there may be protein in a different state compared to that which eluted quickly. Changes to conductivity were due to changes in buffer conditions.
There was no dense band of PCO2 present in the homogenate, implying that expression levels were not high (Figure 3.9). There is also no clear band of PCO2 in the supernatant. This may be because this purification used less pellet than the purification seen in Figure 3.7, and PCO2 expression has been shown to be less than that of PCO1 in the past. Fraction A6 contained the most PCO2. This is not surprising run as this is where there is a peak in the absorbance in the chromatogram (Figure 3.8). Fraction A6 contains protein that is relatively pure, yet again contaminant bands are present. Four bands are seen at ~60, 70, 45 and 36 kDa. The upper two bands could be the same contaminants found during PCO1 purification.

3.2 Purified PCO1 assessment
3.2.1 PCO1 chaperone removal
As there is a significant size difference between PCO1 and both chaperones, size exclusion chromatography could be used to separate the proteins. However, as the proteins co-purified during affinity purification the proteins may be interacting. This interaction must first be disrupted before separation by size exclusion chromatography as these bands may comprise cleaved full length protein. Two different methods were used in an attempt to do this. The first method was incubation with ATP and the second
method was removal of all nucleotide by dialysis in the presence of EDTA and activated charcoal.

The addition of ATP for the removal of DnaK has been used before (Hellebust et al., 1990; Rial and Ceccarelli, 2002). The substrate binding and release cycle of DnaK is regulated by ATP binding. When DnaK has ATP bound it releases the bound substrate, and prepares to bind a new substrate. ATP hydrolysis occurs when the new substrate is bound, and DnaK continues to interact with the bound protein until the new ADP is replaced with an ATP (McCarty et al., 1995). A heat shock protein, GrpE, induces the removal of the ADP from DnaK allowing recycling of DnaK (Packschies et al., 1997). Unfortunately, in the case of GroEL the addition of nucleotide will induce substrate binding (Ishida et al., 2018). Nonetheless, removal of one chaperone is preferable to having both bound.

Figure 3.10. PCO1 post-ATP incubation size exclusion. Sample was separated by a Superdex 75 column. The blue line represents absorbance at 280 nm in mAU (left axis) and the red line represents conductivity in mS/cm (right axis). Standard proteins elution position as shown at top: Thyroglobulin (660 kDa), Bovine Serum Albumin (67 kDa), Ovalbumin (43 kDa), Myoglobin (17 kDa) and Vitamin B12 (2 kDa). Standard values from (Moir, 2014) were aligned to these data using void volume. Fractions numbers relate to Figure 3.11.

Separation over a Superdex 75 column resulted in two main peaks (Figure 3.10). The first peak at 45 mL elution volume contains protein that elutes quickly, and therefore is made of larger proteins, oligomers or large aggregates. This peak was expected to contain
the aggregate plant CDO, or plant CDO bound to chaperones. The two shoulders implied there were a range of species, perhaps different ratios of DnaK to PCO1 molecules. The second peak that begins at 100 mL elution volume elutes later, and thus comprises smaller molecules. This peak likely comprised excess ATP. Gels were run to determine composition of major peaks (Figure 3.11).

**Figure 3.11.** PCO1 post-ATP incubation size exclusion peak one. Predicted mobility of PCO1 is indicated with a black arrow. Lanes labelled with a letter and a number are fraction numbers. Lane labelled M contains molecular weight marker. Marker masses as in Figure 3.2. The chromatogram that relates to this gel is seen in Figure 3.10.

SDS-PAGE suggested that fraction A4 contained PCO1 bound to groEL (59 kDa) and DnaK (69 kDa) (Figure 3.11). Fraction A5 contained PCO1 bound to DnaK (69 kDa) and a second unknown protein. Fractions A6 and B1 appeared to contain PCO1 with little chaperones bound. These later fractions have much less contamination as the ratio of PCO1:chaperones has improved. This method is able to separate chaperone bound PCO1 and monomeric PCO1 but yield is low as much remains bound. Reproducing the experiment with a longer incubation with ATP may improve yield.

Fractions from the second peak were also run on an SDS-PAGE gel. There was no protein detected in any of the fractions from C2 – D1 as no bands were present on the gel. It is likely that this peak was excess ATP eluting. As this molecule is so small, late elution is to be expected. The maximum absorbance of ATP is at 260 nm. However, it does have some absorption at 280 nm so it can be detected as a peak in the chromatogram.
One of the groups who has used this method in the past followed ATP addition with a second incubation with denatured *E. coli* proteins (Rial and Ceccarelli, 2002). This would encourage DnaK that bound the protein of interest to bind other proteins. This approach could have improved the proportion of PCO1 released by DnaK. However, this would make the purified PCO1 preparation contaminated by many other proteins, and would require a further affinity purification step.

Past students in the Wilbanks laboratory have also experienced issues with chaperones co-purifying with their protein of interest. To disrupt the interaction, dialysis with EDTA and activated charcoal was used. DnaK requires magnesium for high affinity binding of the nucleotide to the active site (Neuhofen *et al.*, 1996). As EDTA is a metal ion chelator, EDTA dialysis strips DnaK of Mg$^{2+}$ and therefore lowers affinity of the protein for bound nucleotide. The activated charcoal absorbs free nucleotide so that it cannot bind to DnaK again (Roy, 1995). The release of the nucleotide appears to also induce release of the substrate. GroEL also binds Mg$^{2+}$, which is required for catalysis. The removal of Mg$^{2+}$ is predicted to decrease the stability of the quaternary structure, which would cause GroEL to release its substrate (Gorovits and Horowitz, 1995; Yan *et al.*, 2018).
PCO1 was dialysed in the presence of EDTA and activated charcoal overnight, then was applied to a size exclusion column. After sample application, there is an initial peak in absorbance from 40 mL elution volume (Figure 3.12). This peak eluted quickly after sample addition, suggesting this peak comprises large molecules. A second double peak is then eluted at 90-100 mL elution volume. This double peak suggested that two slightly different molecules or the same molecule in varying states was present. As this is the second peak of the run, the molecule/s that comprise this peak are smaller than that of the first peak. Gels were run to determine composition of major peaks.
Figure 3.13. PCO1 post-EDTA dialysis size exclusion peak one. Predicted mobility of PCO1 is indicated with a black arrow. Lanes labelled with a letter and a number are fraction numbers. Lane labelled M contains molecular weight marker. Marker as in Figure 3.2. The chromatogram that relates to this gel is seen in Figure 3.12.

SDS-PAGE showed that fraction A10 contained considerable PCO1 and GroEL (Figure 3.13). This fraction appears to contain proportionally less DnaK than the sample before separation by size exclusion. Based on the chromatogram peak, and resulting fractions on either side of A11, it is unlikely this fraction truly contains no protein. The lack of protein in fraction A11 may be due to under-loading, or issues with sample preparation. Fractions B1-3 contained an unknown protein with a very small amount of PCO1. Fractions C2-C4 contained only DnaK, suggesting that the attempt to separate PCO1 and DnaK was partially successful. Fractions from the second peak were also run on SDS-PAGE but no bands resulted. It is likely the second peak comprised ATP that was not removed by the activated charcoal.

Neither of the methods described above were completely successful in separating a large portion of PCO1 from the chaperones without loss of PCO1. Given this, no extra steps were added into the purification protocol for PCO1. Instead, the focus was shifted on understanding why the chaperones were bound, and designing constructs that would result in protein that folded correctly.
3.2.2 PCO1 instability

The presence of chaperones during purification of PCO1 was the first suggestion that the protein was not fully folded. PCO1 also tended to precipitate within a few days at 4°C. Although this is not unusual, it does suggest the preparation could have been improved. To begin investigating this, PONDR® software was used to predict disordered regions of the protein in silico.

![PONDR score graph](image)

**Figure 3.14. PCO1 disorder prediction using PONDR® VLXT.** A PONDR score of 1 indicates an ideal prediction of disorder, and 0 indicates an ideal prediction of order. The threshold for assigning disorder is ≥ 0.5.

The N-terminal region from the first residue to position 120 was predicted to be mostly disordered (Figure 3.14). The central region of the protein was predicted to be ordered, which is unsurprising as this is the proposed catalytic region of the protein. The C-terminal end from ~250 – 300 was also predicted to be disordered. This supports the sequence alignment in Figure 1.9 where *A. thaliana* CDO has ~45 N-terminal and ~30 C-terminal residues that extended beyond other CDO sequences. Results for the prediction of each individual residue can be found in Appendix 4.

PONDR® and previous sequence alignments suggested that the N- and C- termini were disordered. To experimentally determine if the termini of the recombinant protein were disordered, a limited proteolysis was conducted. This method uses partial proteolytic digestion to determine whether or not an element is folded, unfolded or disordered based
on backbone accessibility to the protease (reviewed by Fontana et al. 1997). Regions of the protein that are folded into secondary structure elements are more slowly cleaved as the backbone lacks flexibility and displays steric hindrance to protease binding. Regions that are unfolded are readily cleaved as a flexible peptide bond acts as a better substrate for the protease (Fontana et al., 1986, 2004).

**Figure 3.15. PCO1 limited proteolysis.** Enzyme used was a relatively non-specific serine protease, proteinase K. PCO1 digestion products are indicated with a right brace. Lanes differ by incubation time as indicated by numbers in minutes. Lane labelled M contains molecular weight marker. Marker as in Figure 3.2.

Before the incubation was conducted, there was already some degradation of PCO1 as seen in the 0 min control sample (Figure 3.15). After 5 min of incubation, four bands were produced at ~36, 34, 32 and 30 kDa. A change of ~2 kDa is the weight of around 18 amino acids based on the average weight being 110 Da. The same four bands are seen after 20 min of incubation. A fifth band may be present in the 5 and 20 min samples, but cannot be seen due to how the gel ran. Bands formed after 20 min of incubation were sent for peptide identification using MS in an attempt to determine where the termini cleavage sites were. Contamination of full length protein in each of the samples resulted in an unsuccessful experiment. Contamination is due to the smear characteristic seen on the gel, as the bands are not cleanly separated. Given this poor result, future N-terminal identification experiments were conducted using an N-terminal labelling technique prior to digestion.
Banding patterns on the gel after only 5 min suggested that there are some unfolded or disordered regions of the protein. It is likely that the disordered regions are the termini as a poorly folded catalytic domain would result in bands smaller than those observed. The computational and experimental results all imply that it is the disordered termini that are causing chaperone binding and general instability. These results were used to design truncations that removed any disordered regions of PCO1 (section 3.3). Removal of disordered regions was predicted to improve protein stability, and disrupt chaperone binding.

3.3 PCO1 truncations

Due to issues with the stability and purity of PCO1 and PCO2, truncated versions were designed. Both constructs were designed from PCO1 as PCO2 expression was not as plentiful and there were initial problems with purifying PCO2. As experiments and PONDR® analysis showed that it was the N and C terminal ends of PCO1 that appeared unstable and disordered, these regions were removed. One construct had residues 2-53 of the N-terminal removed, and was thus termed ΔN. The second construct had the same N-terminal residues removed, as well as residues 248-293 of the C-terminal. This construct was termed ΔNΔC. Plasmids with genes for both constructs were supplied by GenScript and transformed into BL21(DE3) cells. Expression and solubility was tested at a range of temperatures for a range of different times.
3.3.1 Expression and solubility of PCO1 truncations

Figure 3.16. ΔN expression. Equal number of cells as judged by OD$_{600}$, were loaded into each lane. The negative sign indicates sample before induction and the numbers following this represent number of hours after induction. Band identified as ΔN by MS is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 3.2. Note ΔN runs higher than expected on SDS-PAGE gels.

At 18°C there appears to be little expression of ΔN (predicted mass of 28 kDa) until 20 h post induction (Figure 3.16). The same is to be said of 28°C, although a small band can be seen after 6 h of induction. There is little to no expression of ΔN at 37°C as no bands of ΔN are visible. As expression after 6 h and 20 h in all temperatures looked resulted in reasonable expression levels, all of these samples were subjected to solubility testing.

Figure 3.17. ΔNΔC expression. Equal number of cells as judged by OD$_{600}$, were loaded into each lane. The negative sign indicates sample before induction and the numbers following this represents number of hours after induction. Predicted mass of ΔNΔC is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 3.2.
For ΔNΔC (predicted mass of 23 kDa) all temperatures tested resulted in reasonable expression. (Figure 3.17). At 18°C expression increases with time post induction, with the most being expressed after 20 h. A different pattern is seen at 28°C as little is expressed until 20 h post induction. Expression at 37°C appears similar to that of 18°C, with increasing expression with increasing induction time. All 6 and 20 h samples were subjected to solubility testing, as all of these samples contained enough ΔNΔC for later purification.

**Figure 3.18. ΔN solubility.** Amount loaded was constant to allow comparison. H indicates homogenate, S represents supernatant and P indicates pellet. Band identified as ΔN by MS is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 3.2.

Samples after 6 h of induction did not produce as much soluble protein as the 20 h induction samples (Figure 3.18). However, both 6 h samples resulted in a higher proportion of soluble ΔN than insoluble ΔN. It also appears that the 6 h samples contain a high proportion of ΔN as opposed to other cellular proteins. Of the 6 h samples, 28°C produced the most protein thus 28°C for 6 h was chosen as the optimal conditions for expression of ΔN. The solubility of ΔN at 37°C was analysed with no promising results and is included in the appendix (Appendix 5).
ΔNΔC is soluble under most conditions (Figure 3.19). It appears to be almost completely soluble at 18°C after expression for 6 h, and 20 h. There is a small amount of insoluble protein in the 20 h sample yet as the 20 h sample loading is darker. When considering the samples at 28°C, the same trend is seen as in the 18°C samples. It is almost completely soluble after 6 h induction, and appears to be less soluble after 20 h induction. This could again only be due to greater loading. Expression at 18°C for 6 h appeared to produce the greatest proportion of soluble protein, thus these conditions were chosen for future expression. The solubility of ΔNΔC at 37°C was analysed but was not found useful, and thus is included in the appendix (Appendix 5).

3.3.2 Purification of ΔN and ΔNΔC truncations of PCO1

Once expression conditions were selected, trial purification of both proteins was conducted using Strep-tag® affinity purification. Bacterial pellet was produced using the conditions specified in 3.1.3. The bacteria were lysed using sonication, and the supernatant after centrifugation was loaded onto a Strep-tag® HP column on an FPLC. Below are examples of purifications for both constructs.
Figure 3.20. ΔN Strep-tag® FPLC purification. The blue line represents absorbance at 280 nm in mAU (left axis) and the red line represents conductivity in mS/cm (right axis). This chromatogram is representative of the chromatogram that produced Figure 3.21 and Figure 3.22. Sample application was followed by a wash with running buffer until 70 mL elution volume, elution buffer was added until 115 mL elution volume, followed by a water wash until 130 mL, then a sodium hydroxide wash from 130-155 mL elution. The last step consisted of a wash with water.

The initial peak in the absorbance (2-25 mL) likely comprises cellular proteins that do not interact with the resin (Figure 3.20). The second peak at about 70 mL elution volume contained ΔN as shown by SDS-PAGE. Two insignificant peaks were seen during the regeneration steps of the protocol. These peaks are likely ΔN that bound tightly to the resin, therefore the interaction was not disrupted by the elution buffer. These interactions were instead disrupted by washing the column with water and sodium hydroxide associated with changes in conductivity.
Fractions A7 and B12 appears to have no dense bands around where ΔN would be found on a gel (Figure 3.21). This implies that all the protein was binding to the column. Fraction E11 corresponds to the shoulder of the flow-through peak. This protein runs approximately where ΔN would be, suggesting the shoulder may be ΔN that had failed to bind to the column. This failure to bind may have been due to protein folding in a way that prevents the tag from interacting with the resin.

ΔN eluted over several fractions as many dense bands can be seen on the gel (Figure 3.22) Band density shows that a large amount of the protein was able to be purified. The
resulting protein was pure and no contaminants were seen. The lack of chaperones or other contaminants implied that the protein was likely to be folded correctly. Together, the observation suggests that the N-terminal truncation has removed the disordered region but has not compromised the ability of the protein to fold correctly.

**Figure 3.23. ΔNΔC Strep-tag® FPLC purification.** The blue line represents absorbance at 280 nm in mAU (left axis) and the red line represents conductivity in mS/cm (right axis). This chromatogram is similar to the chromatogram that produced samples used in **Figure 3.24.** Sample application was followed by a wash with running buffer until 50 mL elution volume, elution buffer was added until 100 mL elution volume, followed by a water wash until 115 mL, then a sodium hydroxide wash from 115-130 mL elution. The last step consisted of a wash with water.

The initial peak in the absorbance comprised cellular proteins that do not interact with the resin (**Figure 3.23**). The second peak at about 60 mL elution volume contained ΔNΔC, which was eluted in a peak with a tail. This tail implies that a fraction of the protein was able to bind with the resin with a slightly stronger affinity. This could mean that this protein is in a different state to the protein that eluted immediately. Two other small peaks can be seen during the regeneration steps of the protocol. These peaks are likely ΔNΔC that bound tightly to the resin, therefore the interaction was not disrupted by the elution buffer. These interactions were instead disrupted by washing the column due to changes conductivity.
Figure 3.24. ΔNΔC Strep-tag® FPLC purification gel. Lanes labelled with a letter and a number are fraction numbers. M indicates molecular weight marker. Marker as in Figure 3.2. Where ΔNΔC is predicted to run is indicated with a black arrow. Fraction A6 eluted at 55 mL elution volume.

The homogenate and supernatant lane of Figure 3.24 shows that ΔNΔC was present and soluble before the purification was carried out. Fractions A6-B8 all contained ΔNΔC, yet many contaminants were present. The same chaperones that were seen with PCO1 and PCO2 appear to be present, indicating issues with protein folding. Two other main contaminants were present at ~43 and 36 kDa. These proteins could also be chaperones, but this was not assessed. One of these contaminant proteins present could be DnaJ, a 40 kDa heat shock protein that interacts with DnaK (Liberek et al., 1991). Mass spectrometry would aid in determining if this is the case.

When looking at the resulting purified protein for both truncations, it is clear that removal of the N-terminal improved protein folding and thus solubility. Removal of both termini had a negative impact on the protein, as there could be several chaperones bound. This suggests that the C-terminal is important for protein folding, whereas the N-terminal is not.
3.4 PCO1 nucleic acid binding

After the first purification of PCO1, the concentration of purified protein was attempted to be measured using the absorbance at 280 nm. A spectrum was taken from 220 to 360 nm to assess absorbance. Almost all proteins have a maximum absorbance at 280 nm due to the presence of aromatic amino acids (tyrosine, phenylalanine and tryptophan in particular) (Layne, 1957). However, PCO1 had a maximum absorbance at 260 nm.

A lack of absorbance at 280 nm has in past been the result of specific interaction of buried aromatic amino acids (Lakowicz, 1999). This was tested by taking native PCO1 spectra and comparing it to spectra of PCO1 that was denatured using urea. There was no change in the spectrum (not shown). It was also considered that there was a lack of aromatics or a higher ratio of phenylalanine to the other aromatics. This is not the case as PCO1 has 8 tyrosine, 12 phenylalanine and 4 tryptophan residues. The shift in absorbance could also have been due to the presence of chaperones, as DnaK contains only 1 tryptophan and groEL contains none (sequences from Uniprot database analysed on Expsay). This would heavily influence the spectral properties as tryptophan contributes the most absorbance to the 280 nm reading (Held, 2003).

The alternative explanation is the presence of nucleic acid. The presence of the heterocyclic rings which make up the bases causes nucleic acids to have a maximum absorbance at 260 nm (Layne, 1957). Nucleic acids are negatively charged molecules so will readily interact with positive charges. In the first 47 residues of the N-terminal region of PCO1 there are 13 lysine and 2 arginine residues. As the pKa of the side chains of these amino acids are 10.5 and 12.5 respectively, these residues are positively charged at the pH the protein was purified in (pH 8) (Grimsley et al., 2009). These positively charged residues would readily interact electrostatically with the negative nucleic acid molecules. To determine if PCO1 bound DNA, purified protein was run on an agarose gel and stained with ethidium bromide. ΔN was included to assess if the N-terminal region had an influence on nucleic acid binding.
The first lane containing PCO1 shows a large amount of nucleic acid which appears to be a range of lengths (Figure 3.25). There are two distinct bands at ~2000 and 150 bp. The top distinct band is likely DNA bound to protein and the bottom band is free DNA. When PCO1 was boiled, only one distinct band was present at just less than 100 bp. This suggests all DNA is now free as heat has disrupted the interaction. ΔN had distinct bands at approximately the same lengths as PCO1. Unlike PCO1 samples, boiling had no effect on these DNA bands. Thus suggests DNA present may not be bound to ΔN. Alternatively, ΔN could be very heat stable and thus heat does not interrupt the interaction between the protein and DNA. To test this definitively, one could perform heat denaturation of ΔN and assess this with circular dichroism.

As distinct bands were observed, this implied that the interaction between PCO1 and DNA may be physiological and not an artefact. If binding was unspecific one would see DNA of a range of sizes as opposed to distinct bands. The removal of the N-terminal resulted in substantially less or no nucleic acid binding even though ΔN was more concentrated than PCO1. Thus, it was clear that the removal of the N-terminal decreased the amount of nucleic acid that PCO1 bound. This supports the theory that DNA binding
may be mediated by charged residues of the N-terminal. Experiments that followed focused on determining with what type of nucleic acid PCO1 was interacting.

**Figure 3.26. Agarose gel of PCO1 with nuclease.** Gel contains PCO1 incubated with DNase I or RNase A.

When comparing the samples digested with DNase and RNase to the control lane, both bands had a reduction in intensity (**Figure 3.26**). As DNase cleaves DNA molecules, the loss in intensity suggests that DNA is present. RNase digests RNA, indicating that RNA was also present. There was a complete digestion of sample in the DNase lane, whilst only a portion of the RNase lane was digested. This implies that there was a greater proportion of DNA to RNA. The DNase used was ostensibly RNase free so the result observed was as a result of DNase activity. The RNase stock did contain some DNase as it was grade II, thus reduction of DNA intensity in the RNase lane could be due to DNase activity.

Overall, nucleic acid digestion suggested that PCO1 binds DNA. Further experiments should be conducted with more specific nucleases to determine which type of DNA PCO1 interacts with. DNase I is able to cleave both double stranded (dsDNA) and single stranded DNA (ssDNA), but has a strong preference for dsDNA (Suck, 1994). Whereas Mung Bean nuclease is highly specific for single stranded nucleic acids (Johnson & Laskowski, 1968). Thus digestion with these nucleases, should help determine if the DNA is single stranded or double stranded. If PCO1 binds dsDNA, it is likely that the interaction is physiologically relevant. As PCO1 acts on transcription factors that are
found in the nucleus (close to the double stranded genomic DNA), PCO1 binding dsDNA would localise the protein to its target. PCO1 being in close proximity to its targets will enable catalysis to happen quicker, permitting a faster response to oxygen deprivation.

3.5 Summary

All *A. thaliana* constructs that were designed were able to expressed and purified. All purified proteins were subjected to mass spectrometry for identification. This was to ensure that the correct protein was purified, and that no mutations were present in the protein. All proteins purified were identified with high confidence. See Appendix 6 (PCO1), Appendix 7 (PCO2), Appendix 8 (ΔN) and Appendix 9 (ΔNΔC).

![Figure 3.27. Summary of all CDOs purified from *A. thaliana*. The lanes labelled M contain molecular weight marker. Marker masses are indicated in kDa and each band contains ~500 ng of protein. Construct name is above each gel lane.](image)

With the exception of ΔN, all of the proteins co-purified with chaperones (Figure 3.27). PCO1 co-purifies with two chaperones which were identified as DnaK and groEL. ΔNΔC and PCO2 also co-purify with these chaperones, and two other main contaminant proteins; one of which could be DnaJ. Attempts were made to remove these chaperones from PCO1 using ATP incubation or EDTA dialysis followed by gel filtration, with limited success.
Both PCO1 and PCO2 have truncated protein as seen in the bands below the full length protein band (**Figure 3.27**). Limited proteolysis, sequence alignments and *in silico* work suggests that these truncations are likely N- and C-terminal. Removal of the N-terminal end resulted in a protein that had no chaperones bound, indicating that solubility and stability had greatly improved. Removing both the N- and C-terminal ends resulted in a protein which co-purified with chaperones and two other contaminants. This suggests that removing the C-terminal end decreases protein stability and thus solubility. An ethidium bromide stained agarose gel showed that PCO1 binds DNA, potentially electrostatically via its N-termini. Although this may be physiologically relevant, DNA binding increases protein heterogeneity and thus decreases protein quality for crystallography. Given these results, ΔN was judged the best candidate for crystallography.
Chapter Four

Zea mays and Orzya sativa CDO expression and purification

As PCO1 and PCO2 both co-purify with chaperones, this compromises crystallography due to the lack of homogeneity (McPherson, 1999). Two shortened constructs of PCO1 were produced, one which resulted in pure protein without obvious contaminants, and a second protein that was predicted to bind even more chaperones than PCO1 and PCO2. To extend attempts to crystallise plant CDO, the closest PCO1 homologue from two additional species (termed PCO2 in both cases) was expressed in E. coli and purified using Strep-tag® affinity chromatography. These constructs were designed based on experimental results with constructs from A. thaliana.

4.1 Zea mays constructs

Zea mays, or maize, is cultivated worldwide to produce cereal grain. It is a vital source of food globally, and it also used as a fuel source (Ranum et al., 2014). One billion tonnes were produced in 2016, 100 million tonnes greater than that of wheat and rice (FAOSTAT, 2017). The increasing production and improvement in maize yield is of exceptional importance with increasing population size (Godfray et al., 2010). Manipulation of cellular proteins that influence crop survival could enhance yield, providing more food. Two versions of maize CDO were expressed and purified, and designated “Maize A” and “Maize B”.

4.1.1 Maize A construct

Maize A is the full length predicted PCO2 isoform. The plasmid was obtained from GenScript and was overexpressed in BL21(DE3) E. coli cells. Three expression temperatures were tested for 3 different lengths of time. The solubility of the protein at these different time points was tested, and pellet which was produced from the most suitable expression parameters was subjected to trial purification.
**Figure 4.1. Maize A expression.** Equal number of cells as judged by OD$_{600}$, were loaded into each lane. The negative sign indicates sample before induction and the numbers following this represent hours after induction. Band identified as Maize A by MS is indicated with a black arrow. The lane labelled M contains the molecular weight marker. Marker masses are indicated in kDa and each band contains 500 ng of protein.

There does not appear to be any expression of maize A (predicted mass of 31 kDa) until 20 h after induction at 18°C (**Figure 4.1**). At 28°C, there was noticeable expression after only 3 h. The level of expression did not seem to increase dramatically after 6 or 20 h of induction. Induction at 37°C appeared to express the most protein as these bands are the darkest. Strangely, 3 and 20 h seemed to produce the greatest amount of protein. All 6 and 20 h samples were subjected to solubility testing. After purification and mass spectrometry-based identification it was concluded that Maize A runs higher on an SDS-PAGE gel than expected (at ~35 kDa). This same phenomenon is seen with the ΔN construct.
Figure 4.2. Maize A solubility at 18°C. Loading was kept constant to allow comparison. H indicates homogenate, S represents supernatant and P indicates pellet. Band identified as Maize A by MS is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 4.1.

At 18°C a minimal amount of maize A was expressed, however all that was expressed appears to be soluble (Figure 4.2). A band can be seen in the homogenate at where maize A runs for both 18°C samples. A band of the same intensity in the same position can be seen in the supernatant lane. This implies that the protein is soluble under these conditions.

Figure 4.3. Maize A solubility at 28 and 37°C. Loading was kept constant to allow comparison H indicates homogenate, S represents supernatant and P indicates pellet. Where maize A is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 4.1.
Induction at 28°C led to almost completely soluble protein, regardless of induction time (Figure 4.3). This was determined as bands of maize A in the homogenate are mirrored by bands of similar intensities in the supernatant. Only light bands of maize A bands were seen in the pellet fraction for these samples. In the 37°C samples, almost all maize A protein found in the homogenate was in the pellet fraction after centrifugation. Although there appeared to be a greater amount of protein expressed at 37°C, this protein was not suitable for purification as it was not soluble. Given the results for the expression trial, it was decided that maize A would be expressed at 28°C for 20 h.

Figure 4.4. Maize A Strep-tag® FPLC purification. The blue line represents absorbance at 280 nm in mAU (left axis) and the red line represents conductivity in mS/cm (right axis). Fractions in this chromatogram relate to Figure 4.5. After sample application, the column was washed with running buffer from 10-50 mL. Elution buffer ran through the column from 50-90 mL until a water wash from 90-110 mL. This was followed by a sodium hydroxide wash from 110-135, and finished with a water wash.

The initial peak in the absorbance was the flow-through which comprised cellular proteins that did not interact with the resin (Figure 4.4). The second peak at about 55 mL elution volume contained maize A, which was eluted in a peak with a front shoulder. This shoulder could comprise maize A that was in a different state, as it interacted with the resin with slightly less affinity than maize A that eluted in the main peak.
Figure 4.5. Maize A *Strep-tag®* FPLC purification gel. The lane labelled M contains molecular weight marker and lanes labelled with a letter and a number are fraction numbers. Fractions present are from one FPLC run. Where maize A is expected to run is indicated with a black arrow. Marker as in Figure 4.1. Gel corresponds to chromatogram in Figure 4.4.

Fractions A2-A5 from the purification contained flow-through which comprised cellular proteins (Figure 4.5). As no distinct bands of maize A were in these lanes, it appears all protein present was able to bind to the resin. Fraction C6 contained a very small amount of protein whilst fraction C7 contained most of the eluted maize A. A range of fractions around the elution peak were run as the peak contained a front shoulder, however no dense bands resulted.

The protein eluted was relatively pure, with a few contaminants present. A light band can be seen at 22 kDa, and a number of other light bands can be seen directly below maize A. Proteins under the dense maize A band may be truncated versions of the full-length protein. Like PCO1, sequence analysis shows that maize A appears to have disordered N- and C-terminal ends. Truncations must be at the N-terminus as C-terminal truncations would remove the *Strep-tag®* and prevent maize A from interacting with the resin. As distinct bands were formed, it is probable that proteases were cleaving the N-terminal at sites of enhanced sensitivity to proteases such as disordered loop regions. The cells used to express this protein were deficient in two of the main *E. coli* proteases (Lon and OmpT) which prevented excessive proteolysis during expression (New England Biolabs, 2018). Protease inhibitor cocktails were also present in the lysis buffer.
Bands below the protein were analysed by MS coupled with an N-terminal labelling technique to determine where the protein was being cleaved. Bands from the gel in Figure 4.27 were excised for this experiment, including the band of full length at 31 kDa, and the predicted truncations at ~29, 27 and 21 kDa. The results showed that the full length protein had been purified, and that the N-terminal methionine remained intact. The 29 kDa band was truncated at residue 30-33, and the 27 kDa band was truncated at residue 77-80. The last band at 21 kDa led to an inconclusive result, as peptides did not appear to match those of maize A. This suggests that it may be a contaminant. The results show that the N-terminal is susceptible to truncation, and demonstrates that previous construct designs are plausible as cleavage at residue 30-33 corresponds closely with position 51 of A. thaliana PCO1. Data for this experiment can be found in Appendix 10.

**Improving purified Maize A homogeneity**

The observation of distinct bands below maize A on an SDS-PAGE gel was amplified by sample concentration (see 4.3). An attempt was made to separate these bands from the full length maize A protein. As the N-terminal of the protein contains a large number of positively charged amino acids (9 of the first 30 residues), removal of regions of the N-terminal should influence the pI of the protein dramatically. A sample of concentrated protein was applied to a cation exchange resin, SP HP (a negatively charged resin that binds positively charged molecules). The predicted pI of maize A is 8.5 (as determined by Expasy). When the pH is less than the pI the molecule is positively charged, so the cation exchange was conducted at pH 7. The predicted pI of maize A when the N-terminal is truncated decreases. Thus, the truncated versions of the protein should be negatively charged and may not interact with the resin.
Figure 4.6. Cation exchange chromatography of maize A. The blue line represents absorbance at 280 nm in mAU (left axis) and the red line represents conductivity in mS/cm (right axis). Fractions produced the samples used for Figure 4.7. The column was run using a pH 7 HEPES buffer.

A large peak in initial absorbance suggests that truncated maize A did not interact with the cation exchange resin as predicted (Figure 4.6). However, the full length construct also did not interact with the resin as no other peaks are seen in the chromatogram. This could be due to the theoretical pI of the protein being wrong, or local charge concentration. If DNA was bound to maize A, such as with PCO1, this would also influence pI. There could have also been an issue with the resin, as the column used was several years old. Fractions from the cation exchange chromatography were analysed on an SDS-PAGE gel.

Figure 4.7. Maize A cation exchange chromatography gel. The lane labelled M contains molecular weight marker and lanes labelled with a letter and a number are fractions. Marker as in Figure 4.1. Gel corresponds to chromatogram in Figure 4.6.
Protein eluted from the column was contained within fraction F2 (Figure 4.7). The dense band in this fraction ran at ~ 64 kDa. This is not the size of maize A, but is very close to the size of maize A if it had dimerised. Dimerisation could also explain why maize A did not interact with the cation exchange resin as expected. The change in buffer conditions could have encouraged dimerisation of the protein, as all other work with plant CDO had been conducted in tris at pH 8. Some larger products are also present as seen by the smear at the top of the lane. These could have been aggregates of maize A or related to the contaminants that were present after purification. Although the dense band was not confirmed as maize A, it is unlikely to be another protein as the sample that was loaded onto the column was relatively pure maize A. This could be confirmed using MS in the future.

As this is a reducing gel, it is interesting that the protein remains dimerised. Before samples are loaded they are mixed with a range of denaturing compounds and boiled. If the protein remains a dimer after this process then it is likely a covalent interaction between the two monomers. PCO1 has also been shown to form higher order structures in certain conditions. This is explored in Chapter 5. This interaction could be an artefact of in vitro purification, but could also be physiologically relevant as dimerisation of proteins is a common mechanism used to control protein activity.

4.1.2 Maize B construct
As maize A was prone to truncation which prevented a homogenous sample from being produced, a truncated version was synthesised. This construct, termed maize B, had residues 1-39 removed based on results from PCO1 truncations. A plasmid was purchased from GenScript, and was transformed into BL21(DE3) E. coli cells. Expression trials were conducted at a range of temperatures and time points, followed by trial purification.
Expression levels of maize B (predicted mass of 27 kDa) under all conditions were promising (Figure 4.8). When inducing the protein at 18°C, a band was present after 3 h which becomes denser as the induction time increases. This is also true of 28°C and 37°C. All 6 and 20 h samples were subjected so solubility testing.
The general trend for solubility was that most of maize B was insoluble (Figure 4.9). Both 18°C time samples contained almost all of maize B in the pellet fraction. There was a small amount that was soluble, as a modest band of maize B could be seen in the supernatant for both time points. The same is true of the 28°C time points. It appeared as though maize B was also generally insoluble after expression at 37°C (Figure 4.10). There was only a small amount of protein in the soluble fraction after 6 h of induction at 37°C. A 20 h induction time at 37°C did not improve solubility.

![Figure 4.10. Maize B solubility at 37°C. Loading was kept constant to allow comparison. H indicates homogenate, S represents supernatant and P indicates pellet. Where maize B is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 4.1.](image)

After analysis of the expression trial results, it was determined that under these lysis conditions the greatest proportion of soluble protein was produced with induction at 18°C for 6 h. Protein produced under these conditions was purified to determine whether or not a large enough amount of protein could be produced for further experimentation.
Figure 4.11. Maize B Strep-tag® FPLC purification. The blue line represents absorbance at 280 nm in mAU (left axis) and the red line represents conductivity in mS/cm (right axis). This chromatogram is representative of the chromatogram that produced the samples used in Figure 4.12. After sample application, the column was washed with running buffer from 20-65 mL. Elution buffer ran through the column from 65-100 mL until a water wash from 100-115 mL. This was followed by a sodium hydroxide wash from 115-145, and finished with a water wash.

The initial peak in the absorbance comprised cellular proteins that did not interact with the resin (Figure 4.11). The second peak at about 70 mL elution volume contained maize B, which was eluted in a peak with a slight front shoulder. Maize B that eluted in this shoulder was likely in a different state as it interacted with the resin with slightly less affinity than maize B that eluted in the main peak.

Figure 4.12. Maize B Strep-tag® FPLC purification gel. Lanes labelled with a letter and a number are fraction numbers. Where maize B is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 4.1.
The first 3 fractions contain many cellular proteins, confirming that the initial absorbance was flow-through (Figure 4.12). It appears as though no or little maize B was in these fractions. Thus, all maize B that was loaded onto the column was able to interact with the resin, and that the column was not overloaded. Most of maize B eluted in the fraction B12. There are many contaminants present in this sample, with 5 main contaminant bands. As the column was washed thoroughly before elution, it is likely many of these contaminants interact with maize B or have some affinity for the resin. Three of the bands were predicted to be chaperones including DnaK (69 kDa), groEL (60 kDa), and DnaJ (40 kDa).

The gel in section 4.3 had all five bands (including the band predicted to contain maize B) sent for MS to be identified. The results showed that maize B had been purified, with 70% sequence coverage. The top two bands were comprised of DnaK (27% coverage) and groEL (53% coverage) as expected. The bottom two bands contained ribosomal protein S1 (47 kDa, 30% coverage) and elongation factor thermo-unstable (EF-Tu) (43 kDa, 67% coverage) (Appendix 11).

The presence of chaperones implies the protein was not folding correctly. Chaperone binding is also seen with other plant CDOs including PCO1 and PCO2 (See Chapter 3). The ribosomal protein S1 is an acidic, RNA binding protein that is part of the 30S ribosomal subunit (Draper and Hippel, 1978). It unwinds mRNA structures before entry into the ribosome, so is an important protein for translation (Subramanian, 1983). The reason for the interaction between maize A and the ribosomal protein S1 is not clear. EF-Tu is a GTP requiring enzyme that promotes binding of aminoacyl-tRNA to the A-site of the ribosome, aiding in translation. EF-Tu is the most abundant protein in E. coli (Weijland et al., 1992), which may explain its presence in the purified protein solution.

**Improving Maize B solubility**

Not only did the protein appear to bind chaperones, but it appeared cloudy after elution. It was concluded that maize B had solubility issues, therefore, attempts were made to improve the solubility. A range of different lysis conditions were trialled. The theoretical pI was computationally calculated, and this was used to decipher what conditions to use. As the pI was predicted to be 6.1, buffers were produced to cover two pH units either
side of this value. Sodium citrate pH 4 and 5, and Tris pH 7.1 and 8 were employed. Excess salt, L-arginine and TWEEN®20 have in the past improved the solubility of other proteins so these solutions were also included in certain samples (Leibly et al., 2012; Lebendiker and Danieli, 2014). Further details of this experiment can be found in 2.4.10.

Although a range of conditions were trialled in order to improve protein solubility, it is difficult to say whether or not any were successful. No clear bands of maize B can be seen in either of the gels (Appendix 12) except for the pellet fraction of lysis condition 3. There is no obvious reason for the lack of protein in any of the samples. When the samples were centrifuged before being loaded on the gel, it is possible aggregate maize B was spun down to the bottom of the tube and not loaded. To prevent this in the future, the samples could be boiled for longer.

A further expression trial of maize B was also conducted. This time, a different E. coli cell was used. Origami® cells are generally used for proteins that have difficulty folding as they contain disulfide bonds. Little maize B appeared to be expressed in this cell line so solubility testing or a trial purification was not conducted. The expression gel can be found in Appendix 13. Overall, expressing maize B in BL21(DE3) cells at 18°C for 6 h produced enough soluble protein for further experiments. In the future, it may be of interest to try expressing in another prokaryotic cell line. Eukaryotic cells lines could also be evaluated.

4.2 Orzya sativa construct

Orzya sativa or more commonly, rice, is a staple food source for over 3.5 billion people worldwide, and is of particular importance to those living in developing parts of Asia and sub-Saharan Africa. It plays a vital role in food security and the economy, providing over 200 million jobs (Muthayya et al., 2014). Furthermore, it is fast becoming a model organism like Arabidopsis (Rensink and Buell, 2004). During cultivation, rice is often subjected to flooding conditions, therefore plant CDO may be very important for plant growth and crop yield.
A plasmid containing the PCO2 isoform from rice was purchased. The construct was designed with residues 1-34 removed in an effort to improve protein stability, as this was proven to be the case with *A. thaliana* CDO (see Chapter 3). This construct was transformed into BL21(DE3) *E. coli* cells and was subsequently used to perform expression trials. Three temperatures were tested for 3 different lengths of time. The protein was termed ΔNR (N-terminal removed rice construct). For these expression trials, a clearly induced Coomassie stained band was analysed however the identity was never confirmed by MS.

**Figure 4.13. ΔNR expression in BL21(DE3).** Equal number of cells as judged by OD$_{600}$ were loaded into each lane. The negative sign indicates sample before induction and the numbers following this represents number of hours after induction. Where ΔNR is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 4.1.

At 18°C there appeared to be little expression of ΔNR (predicted mass of 27 kDa) until 6 h, and the most expression was after 20 h (Figure 4.13). When expressing at 28°C there was similar expression after 3 and 6 h, with the most ΔNR expressed after 20 h. Expression seemed to increase with induction time in the 37°C samples as the bands got progressively darker from 3 to 20 h. All 6 and 20 h samples were subjected to solubility testing.
Figure 4.14. ΔNR expressed in BL21(DE3) solubility at 28°C and 37°C. Loading was constant to allow for comparison. H indicates homogenate, S represents supernatant and P indicates pellet. Where ΔNR is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 4.1.

It appeared as though ΔNR was almost completely soluble under all conditions (Figure 4.14). At 28°C the ΔNR band in the homogenate was the same intensity as the band found in the supernatant under this condition. This was also true of the 28°C 20 h sample, although the bands were more prominent suggesting more protein is present. The samples at 37°C showed the same trend.

Figure 4.15. ΔNR expressed in BL21(DE3) solubility at 18°C. Loading was constant to allow for comparison. H indicates homogenate, S represents supernatant and P indicates pellet. Where ΔNR is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 4.1.
There appeared to be somewhat more insoluble ΔNR at 18°C compared to 28 and 37°C (Figure 4.15). After 6 h of induction, the band in the homogenate at ~27 kDa was found to be of similar intensity in the supernatant. A small band of this protein was found in the pellet under this condition. At 20 h, there appears to be slightly more protein present in the pellet. This implies that a greater portion of the protein was insoluble.

Given the expression and solubility testing results, it was decided that expression would be conducted at 28°C for 20 h. Although there was comparable expression and solubility of ΔNR at 37°C for 20 h, 28°C had worked well in the past for other similar constructs. Pellet produced under these conditions was subjected to a trial protein purification.

Figure 4.16. ΔNR expressed in BL21(DE3) Strep-Tag® FPLC purification. The blue line represents absorbance at 280 nm in mAU (left axis) and the red line represents conductivity in mS/cm and the X-axis shows elution volume in mL. Fractions numbers are included as this is the chromatogram that produced the samples analysed in Figure 4.17. After sample application, the column was washing with running buffer until 60 mL, and this was followed by elution buffer from 60-90 mL. A water wash occurred from 90-110 mL followed by sodium hydroxide from 110-140 mL. The last wash was with water.

The initial peak in the absorbance was comprised of cellular proteins that do not interact with the resin (Figure 4.16). This peak had a long tail, suggesting there is a protein/s present that is interacting weakly with the resin. During elution there was no peak in absorbance, suggesting that no ΔNR had bound the column. A second possibility is that the interaction between the protein and the resin was not being disrupted by the elution buffer. This would be surprising as the elution buffer from the same batch was used
successfully with other constructs. If this was the case, there would also be a large elution peak during column regeneration. It was thought that the protein weakly interacting with the column during the flow-through tail was ΔNR. This was tested by running these fractions on a gel.

![Figure 4.17](image)

**Figure 4.17.** ΔNR expressed in BL21(DE3) Strep-Tag® FPLC purification gel. Lanes labelled with a letter and a number are fraction numbers. Where ΔNR is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 4.1. Chromatogram that produced these fractions is found in Figure 4.16.

There is a prominent band where ΔNR is predicted to run in the flow-through (Figure 4.17). This suggests that the protein may not have interacted with the resin. There was no other protein found in the remainder of the peak. The tail of the flow-through appears to comprise a non-protein molecule with an absorbance at 280 nm.

As no protein was purified, it was thought that no protein had been expressed. The bands interpreted as ΔNR could have been another cellular protein. In an effort to produce ΔNR, it was expressed in two other cell lines. Samples of the expression trials were taken at various time points, and various temperatures were tested. All samples with acceptable expression levels were subjected to solubility testing.
Expression of ΔNR in BL21(DE3) pLysS cells

BL21(DE3) pLysS cells are canonically used for producing proteins that are toxic to bacterial cells. These cells contain the pLys plasmid which prevents any uninduced expression by inhibiting basal levels of the T7 RNA polymerase (Invitrogen, 2018). Although toxicity is not predicted to be an issue ΔNR, changing cell lines can change the experimental outcome.

![Figure 4.18. ΔNR expression in BL21(DE3) pLysS.](image)

Equal number of cells as judged by OD\textsubscript{600}, were loaded into each lane. The negative sign indicates sample before induction and the numbers following this represent hours after induction. Where ΔNR is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Figure 4.1.

Induction at 18°C showed a prominent band after 3 h, with expression increasing slightly after 6 and 20 h (Figure 4.18). At 28°C the same trend is seen. Expression at 37°C showed bands after 3 and 6 h induction, but no protein after 20 h. The absence of a band in this lane is likely due to under-loading as no other cellular proteins are visible in this lane. All pre-induction samples have no protein expressed, but this may be because many of the pre-induction lanes appear under-loaded. All 6 and 20 h samples were subjected to solubility testing (Appendix 14 and Appendix 15). Given the results from the expression and solubility testing, pellet produced at 28°C for 6 h was used for a trial purification (Appendix 16 and Appendix 17). No protein was successfully purified.
Expression of ΔNR in Origami™ (DE3) cells

These cells are designed to aid proteins that have difficulty folding in *E. coli* cells as they contain disulfide bonds. Origami™ contains mutations in the two genes, thioredoxin reductase and glutathione reductase. These mutant proteins aid in the formation of disulfide bonds in the cytoplasm, which is highly reducing under usual conditions. As with the other cell lines used, they are inducible with IPTG and compatible with plasmid containing amp resistance (Bessette et al., 1999). An attempt was made to express ΔNR in Origami™ cells as both computational and experimental evidence suggested that plant CDO may contain disulfide bonds (Chapter 5).

![Figure 4.19. ΔNR expression in Origami™ (DE3).](image)

The bands present where ΔNR would run become increasingly dense over the induction period (Figure 4.19). At 18°C there was limited expression until 20 h, which contained the most protein. However this lane does appear to be over-loaded as other cellular proteins were also denser. Induction at 28°C was more successful than that of 18°C. There was expression after 3 h of induction, but the amount of protein produced did not seem to increase substantially for the next two time points. The sample at 37°C lacked any expression. All 6 and 20 h samples were subjected to solubility testing (Appendix 18 and Appendix 19). Given these results, pellet produced from induction at 28°C for 6 h was used for a trial purification (Appendix 20 and Appendix 21). No protein was purified.
After attempts were made to express ΔNR in three different cell lines were unsuccessful, it was thought that the protein was not expressing at all. To determine if the bands analysed were ΔNR, the band at ~27 kDa from the 28°C 6 h supernatant lane of Appendix 14 was sent for identification using MS. The results showed that the protein was a 26 kDa E. coli chloramphenicol acetyltransferase with 35% identity (Appendix 22). As there were no peptides of ΔNR present, it was confirmed that the protein was not expressing in BL21(DE3) pLysS, and potentially the other cell lines tested.

The lack of protein expression could be due to a range of factors. It was considered that the plasmid may have an early stop codon or the wrong inducer was being used for the plasmid. The plasmid was sequenced by GAS and the results showed that both the insert and the plasmid were correct (Appendix 23). Other factors considered were protein toxicity, ribosomal stalling and poor folding (Rosano and Ceccarelli, 2014). Protein toxicity was ruled out as the protein was able to be efficiently transformed into BL21(DE3) cells which do not prevent uninduced expression.

Ribosomal stalling was considered as this prevents the protein from being translated, and can lead to degradation of the mRNA (Sunohara et al., 2004). Stalling is often due to the presence of a string of rare codons or a specific sequence that the ribosome recognises (Krasheninnikov et al., 1991; Peil et al., 2013). As it is widely known that E. coli has different codon bias to eukaryotic cells (Buchan and Stansfield, 2005), maize B was codon optimised for E. coli in the synthetic gene used. Given this, maize B does not appear to contain any likely ribosomal stalling sites.

The most likely reason is poor protein folding. If a protein is folded incorrectly, it can be tagged for degradation in the cell by chaperones (Gething and Sambrook, 1992). If this was the case, it would mean that the protein is expressing as is expected, but the translated product is being degraded quickly after. As it appears that expression levels were low in Origami™ cells, it is likely that the failure to fold correctly is not disulfide bond related.

For future experiments, it is suggested that the C-terminal Strep-tag® is removed and replaced with a hexahistidine tag. As this is already built into the construct, simple cloning would yield a different construct that may behave differently during expression.
4.3 Summary

All *Zea mays* constructs were able to expressed and purified. All purified proteins were subjected to mass spectrometry for identification. This was to ensure that the correct protein was purified, and that no mutations were present in the protein. All proteins purified were identified with high confidence. See Appendix 10 (Maize A) and Appendix 11 (Maize B).

![Figure 4.20. Summary of all purified Maize CDO constructs.](image)

The lane labelled M contains molecular weight marker. Marker as in Figure 4.1. Ma indicates maize A, and Mb indicates maize B.

Both constructs purified from maize did not produce chaperone free protein (Figure 4.20). Maize A binds what appears to be DnaK, and has a range of degradation products that arise post-purification. These truncated products could not be removed using cation exchange chromatography. Maize B interacts with 5 contaminant proteins, two of which have been confirmed to be chaperones. This implies there are folding issues with the construct. Unlike maize A, maize B does not lead to the production of many truncated versions of the protein post-purification. This implies that the truncation has prevented proteolysis of maize B’s termini.

The construct of *Oryza sativa* origin, ΔNR, was not able to be expressed and therefore was not able to be purified. Future experiments should focus on trying to remedy this. As two maize constructs were able to purified, these constructs were utilised for crystallography.
Chapter Five

Structural modelling and analysis

One of the most informative elements of protein characterisation is determining the structure. Gaining structural insights not only makes physical properties clear, but it also provides an idea of protein function and dynamics. Atomic resolution models can be experimentally determined using several methods including NMR spectroscopy, X-ray crystallography and electron microscopy. The most common method is X-ray crystallography (Smyth and Martin, 2000), with 86% of all structures in the protein data bank (PDB) determined by this method in 2011 (Dessau and Modis, 2011).

Protein crystals are formed by supersaturating purified protein, which slowly precipitates to form an ordered crystalline structure under certain conditions. When an X-ray beam hits the protein crystal, it diffracts to form a characteristic pattern of spots termed a diffraction pattern. If a clear diffraction pattern is produced, it can be analysed to produce an electron density map. This map is then used to fit the amino acid sequence using the preferred geometry of each amino acid, and knowledge of the chemical nature of the polypeptide. Once the sequence is fitted to the map and the structure is deemed energetically favourable, the atomic model is produced (Nurachman, 2004; Drenth and Mesters, 2007).

Unfortunately, it can be difficult to crystallise certain proteins for reasons such as flexibility and heterogeneity. Instead, a computational method termed homology modelling can be employed to predict the structure. This method uses the assumption that similar amino acid sequences will have a similar structure. This is often correct as sequences evolve quickly, whilst the structure retains the same general topology as certain folds are highly energetically favourable (Chothia and Lesk, 1986).
This technique uses the amino acid sequence of the protein of interest and a known homologous protein structure to produce a 3-D model of the protein of interest. *In silico* structural predictions are not as precise and reliable as experimentally determined structures, but they provide rapid results that are inexpensive. The process begins with recognising a template by aligning the two sequences. A backbone is generated using the template co-ordinates, and loops are modelled in the gaps. Side chains are then orientated and the model is optimised to ensure energetic favourability. See review by (Dorn et al., 2014) for more information on approaches to modelling.

### 5.1 Homology models

Models of PCO1 were produced using two servers, SWISS-MODEL and I-TASSER. SWISS-MODEL used bacterial CDO (3USS) as the template to which the server found 16% sequence similarity to PCO1. I-TASSER utilised mammalian CDO (2B2H) as the template, which shares 24% sequence identity with PCO1 (calculated using BLAST). A protein with greater than 80 amino acids requires a template with greater than 25% homology for confidence in the resulting model (Sander and Schneider, 1991). Both values are outside the suggested minimum sequence identity, but nevertheless only a few homologues are available with a solved structure. The automated function was used for the SWISS-MODEL tool, see (Swiss Institute of Bioinformatics, 2018) for details on modelling framework. The automated function of I-TASSER was also employed with no nominated restraints used; see (Zhang Lab) for details of how the model was produced.
Figure 5.1. Comparison of homology models. The side chains of histidine residues that co-ordinate the iron (orange sphere in the left model) are shown as sticks. For the SWISS-MODEL (left), core structure was modelled. PDB accession number 3USS was used as the template and Gobal Model Quality Estimation (GMQE) = 0.33 and QMEAN = -4.47. The I-TASSER model (right) showed the full structure, and the template used was PDB accession number 2B2H. Estimated TM-score = 0.37±0.13 and C-score = -3.04. Chimera was used to produce these images.

Both models produced contained the same overall architecture (Figure 5.1). The central component of the structures comprise a β-barrel with iron co-ordinating histidine residues. The SWISS-MODEL was able to orientate the histidines with iron present in the model, whereas the I-TASSER model does not have the histidines in the correct orientation for this to occur. The histidines however, are in close enough proximity to interact if side chains are reoriented. Two α-helices in the central structure are also present on both models. Two further α-helices are present on the C-terminal end of the I-TASSER model as it modelled the ends. The I-TASSER model does appear to contain more loop regions than the SWISS-MODEL. Quality assessment of each model is discussed in Section 5.1.1 for the SWISS-MODEL and Section 5.1.2 for the I-TASSER model.
5.1.1 SWISS-MODEL homology model

**Figure 5.2. SWISS-MODEL homology model of PCO1.** The left depicts a homology model of PCO1 (tan). The side chains of histidine residues that co-ordinate the iron (orange sphere) are shown as sticks. The right shows PCO1 (blue) superimposed on mammalian CDO (tan, PDB accession number 2B5H). N- and C-terminal residues of PCO1 are not modelled. PDB accession number 3USS was used as the template for the PCO1 model and Chimera was used to produce these images. Quality measures as in Figure 5.1.

In regards to overall structure, PCO1 is comprised of a β-barrel with two prominent α-helices (**Figure 5.2, left**). As seen in mammalian CDO, there is a predicted cupin-like motif also present in PCO1. The only elements of the structure that differ are within loops. Slight changes in loop sizes may be as a result of the need to accommodate different substrates. The 3-His triad which co-ordinates the central iron overlays well, with slight differences in one of the histidines (**Figure 5.2, right**). The His-triad conservation in PCO1 predicts that PCO1 binds iron the same way, suggesting that iron is important for the catalytic mechanism. The secondary structure of the PCO1 model is plausible as it contains predicted cupin motifs and β-sheets. The lack of modelling for the ends of PCO1 is due to the lack of homologous sequences in bacterial CDO.
When considering model quality, two main parameters are assessed in SWISS-MODEL, GMQE and QMEAN. GMQE is a measure that considers alignment of the target and template, as well as the template search method. This value is between 0 and 1, where a higher value indicates a better quality model. The GMQE value for this model was reported as 0.33 therefore suggesting the model is not of high quality. This is because the target does not fit the template, which is unsurprising with only 16% sequence similarity. A QMEAN value is an estimate formed by combining a range of geometric properties that assess both local and global folding quality. It compares features of the model to features seen in other structures to determine how likely they are. A QMEAN value of 0 indicates good agreement between the model and other reported structures. When the score drops below -4.0, the model is designated as low quality. This model had a score of -4.47, placing it in the low quality category. Given this, the structure should be assessed with this in mind.

Figure 5.3. SWISS-MODEL homology model of PCO1 surface topology. Blue shading refers to the nitrogen atoms of the iron co-ordinating histidine residues. Model is the same as in Figure 5.2.

The theoretical substrate entry site is large and lines up well with the His-triad and iron (Figure 5.3). The substrate entry site is large compared to that of mammalian CDO (not
This is likely due to the need for PCO1 to accommodate a much larger substrate, as it is a polypeptide. Interestingly, the active site is open on both sides of the structure. One possible reason for entry on both sides of the structure is that it permits entry of the co-substrate, molecular oxygen, after the target peptide has entered the active site. If there were only one entry site, and the peptide entered first this could block molecular entry, and thus prevent catalysis. This active site model allows a random rather than sequential order of substrate binding.

**Figure 5.4.** SWISS-MODEL homology model of PCO1 with catalytic residues. Tyrosine 229 is shown with a red side chain (OH group) and phenylalanine 213 is also shown. Model is the same as in Figure 5.2.

In mammalian CDO, a tyrosine residue within the active site is proposed to play an important role in catalysis (See 1.5.1). This tyrosine is completely conserved among CDO from other origins (Figure 1.9). PCO1 does not have a tyrosine residue in this position, instead it is replaced by phenylalanine. As the His-triad is conserved, it is proposed that the iron catalytic mechanism is also conserved. This would mean that a hydroxyl group in close proximity to the active site would also be required in PCO1. It is proposed that a tyrosine 229 could act as the catalytic tyrosine (Figure 5.4). No other hydroxyl groups are predicted to be in close proximity to the active site. A tyrosine in
this position is completely conserved in plant CDO from all species in the sequence alignment in Figure 3.1.

In this model the distance between the phenylalanine and central metal ion was determined to be 6.6 Å (Figure 5.4). The tyrosine present in the model was shown to be 4.2 Å from the central metal ion. Although this does not assess whether or not the molecule is in the correct orientation, it appears that the tyrosine is close enough spatially to aid in catalysis. To determine if this theory is correct, a Y228F mutant could be produced. Phenylalanine would occupy the space in the same manner, but the proposed catalytic hydroxyl group would not be present.

5.1.2 I-TASSER homology model

Figure 5.5. I-TASSER homology model of PCO1. The side chains of histidine residues that co-ordinate the iron are shown as sticks. Tyrosine 229 is also shown with a red side chain (OH group). PDB accession number 2B5H was used as a reference and Chimera was used to produce the image. Quality assessment as in Figure 5.1.

This model predicts a structure comprising mostly β-strands (Figure 5.5). There are two α-helices close to the core β-barrel, and two further α-helices are suggested at the C-terminal end. The N-terminal end appears to be an irregular loop, aligning with previous
experimental data suggesting it is disordered. The histidines of the iron co-ordination sphere are present in close proximity as expected. However, the side chains would require re-orientation to co-ordinate a metal ion. This is unsurprising as the model was energy minimised without iron present. Tyrosine 229 in this model is close to the active site showing that a role in catalysis is spatially plausible.

A C-score is a value to describe the confidence in the quality of the model. The number range is from -5 to 2, where the larger the number the better the quality. This model has a C-score of -3.04, which implies little confidence in the model quality. This is likely due to the lack of sequence similarity between plant and mammalian CDO. The TM-score is a measure of how closely related the model structure and the query structure are post-modelling. This is used as it is not as sensitive to local error as the estimated root-mean-square deviation of atomic positions (RMSD). A TM-score of 0.37±0.13 shows that the models show relatively similar geometry. Overall, the model assessment parameters suggest the model is of average quality, and should be analysed with caution.

Figure 5.6. I-TASSER homology model of PCO1 superimposed on mammalian CDO. PCO1 (blue) is superimposed on mammalian CDO (tan). The side chains of histidine residues that co-ordinate the iron (orange sphere) are shown as sticks. PDB accession number 2B5H was used as a reference and Chimera was used to produce the image. Quality assessment as in Figure 5.1.
The core structure of the I-TASSER model and the mammalian CDO structure share many similarities (Figure 5.6). They both contain a series of β-strands and the cupin motif that forms the characteristic β-barrel fold. Only two α-helices are present in the central structure for both mammalian and PCO1. The only element that differs between the two core structures is the loop formations. This is unsurprising as the sequences are not similar, and the substrate size differs extensively. The N- and C-terminal regions of the two proteins are very different. PCO1 has extended ends that are not present on mammalian CDO. The N-terminal appears to be a disordered loop, whilst the C-terminus appears to contain two α-helices as well as irregular regions. This aligns with previous experiments that suggested these regions were disordered (Chapter 3). The C-terminus is more ordered than that of the N-terminal, which is perhaps why C-terminal removal resulted in several co-purifying contaminants, including at least two chaperones.

Figure 5.7. I-TASSER homology model of PCO1 surface topology. Iron co-ordinating histidine residues are shown as sticks. Model as in Figure 5.5. Quality assessment as in Figure 5.1.

The iron co-ordinating residues align well with the proposed substrate entry site (Figure 5.7). This substrate entry site does not penetrate to the other side of the protein as with the SWISS-MODEL homology model. The I-TASSER model seems to predict sequential
binding. It has been shown experimentally that the \( \text{O}_2 \) molecule must enter the active site first in mammalian CDO for catalysis to occur. This sequential binding, with \( \text{O}_2 \) first, is consistent with the arrangement of this active site. The exterior of the protein may aid in recognising the substrate, as the peptide would extend out of the protein and interact with surface residues. Surface charges were mapped to see if this suggested any means of substrate interaction.

![Figure 5.8. PCO1 and RAP2.12 surface topology with charge distribution.](image)

**Figure 5.8. PCO1 and RAP2.12 surface topology with charge distribution.** RAP2.12 (left), PCO1 (right). Blue represents a positive charge, and red represents a negative charge. I-TASSER was used to produce both models. PDB accession number 2B5H was used as a reference for PCO1 and Chimera was used to produce the image. Quality assessment for PCO1 as in Figure 5.1. Active site of PCO1 is facing forward.

The positively charged region at the bottom of the RAP2.12 structure is the DNA binding domain (**Figure 5.8, left**). This allows electrostatic interactions with the negatively charged DNA. This is important for the role of RAP2.12 as it is a transcription factor, and must interact with the DNA to perform its function. RAP2.12 also has a negatively charged region at the top of the structure. There are no known functions for this negatively charged region, although it is a feature of many transcription factors. This may be present to balance the pI of the protein, or it could be present to permit an interaction with PCO1. The model of RAP2.12 mostly comprises loops, and no similar transcription factor structures have been solved. This resulted in a structure with low confidence, and thus suggests the model should be analysed with caution.
When the charges are mapped on the exterior of PCO1, it is clear that there are two distinct regions (Figure 5.8, right). The left side (as shown) is positively charged which would be permit both a mutually exclusive electrostatic interaction with DNA or with RAP2.12. The negatively charged region on the right may be present to balance the pI of the protein, or to allow another unknown interaction. Charges were also mapped on mammalian CDO to determine if this was characteristic of all CDOs. The charge distribution was even on mammalian CDO (Appendix 24), therefore the charge distribution of PCO1 is specific and likely reflects its function. Production of a docking scheme using molecular dynamic tools could aid in assessing this relationship further.

Figure 5.9. 1-TASSER homology model of ΔN surface topology. Red shading refers to the histidines of the iron co-ordinating histidine residues. PDB accession number 2B5H was used as a reference and Chimera was used to produce the image. Image was produced by removing the first 40 residues of the model in Figure 5.1. Quality assessment as in Figure 5.1.

When the model of PCO1 was produced in Figure 5.7, the substrate entry site appeared to be smaller than expected. Although a single amino acid could easily fit in this space, a peptide is much larger. The N-terminus sits close to the substrate entry site, and so it was thought that it may move for substrate binding. The lack of regular structure in this element also suggested that movement is possible. This is directly supported by limited
proteolysis conducted in Figure 3.15. Given this theory, a ΔN model was produced to see if this would change the size of the active site.

The active site in Figure 5.9 is indeed much larger than the active site of the full length model. The histidine residues are still aligned well within the active site, as expected. Below the active site, there is a concave element to the structure. This could act as a cleft for binding of the amino acids of peptide targets that protrude outside the active site, aiding in substrate recognition. This cleft could be responsible for recognising the end of the ~10 amino acid N-terminal recognition sequence of all PCO1 targets.

5.2 Disulfide bonding

When analysing the sequence alignment of plant CDOs from various plant origins, it was observed that there were an array of conserved cysteines (yellow in Figure 3.1). PCO1 cysteine residues 129, 138 and 218 were completely conserved between the species analysed. Cysteine residues 21 and 230 were not at all conserved and residues 65, 197, and 235 were mostly conserved. The presence of highly conserved cysteines can highlight important structural or catalytic roles. As cysteine residues can participate in covalent disulfide bonding, they contribute to structural rigidity (Fass, 2012). Given this, disulfide bonding is crucial for protein folding and stability (Anfinsen, 1973).
Mapping of PCO1 cysteine residues (N-terminal C21 was not included) on a homology model suggested a number of pairs in close enough proximity to interact (Figure 5.10). In particular, C65 and C129 (left) were predicted to interact, along with C138 and C218 (top right). Although the side chains appeared to lack the correct orientation to mediate an interaction, the orientation is based solely on a model and therefore may not be correct. If there are disulfide bonds present in PCO1 as suggested, then it is possible that the presence of chaperones during purification may be due to misfolding as a result of incorrect disulfide bonding.

This result aligns well with the sequence comparison data, as it also suggests C65, 129, 138 and 218 are important structurally. Both the conservation of these residues and placement on the structure suggest a role in disulfide bonding. The presence of disulfide bonds was tested experimentally by running an SDS-PAGE gel of PCO1 in the presence and absence of BME.
**Figure 5.11. PCO1 ± BME.** Where PCO1 is expected to run is indicated with a black arrow. The lane labelled M of the left SDS-PAGE gel contains molecular weight marker. Marker masses are indicated in kDa and each band contains 500 ng of protein.

BME is present in SDS-PAGE loading dye to reduce cysteine residues and promote protein denaturation. This ensures that protein runs as a clean band, and it migrates as expected on an SDS-PAGE gel. PCO1 ran as expected when denatured by BME as a clean band at 36 kDa (**Figure 5.11**). When no BME was present, the band of PCO1 was smeared and protein ran both higher and lower than expected. A band is also found at the top of the gel, suggesting potential higher order structures with intermolecular disulfide bonds are present. DnaK and groEL contaminants run as discrete bands independently of reductant. Both are deficient in cysteine, groEL completely so. As changes were observed in the band of PCO1 in the absence of BME, this suggests that there are intramolecular and/or intermolecular disulfide bonds present in the purified, recombinant protein.

To determine which residues are forming disulfide bonds, reduction and alkylation experiments were conducted followed by MS. Iodoacetamide (IAM) and iodoacetic acid (IAA) are chemicals that are able to attack free thiol groups, and thus can be used to tag reduced cysteine residues. Light and heavy versions of these chemicals were used to enable differentiation based on molecular weight. For the experiment, PCO1 was exposed to light IAM, then reduced and exposed to heavy IAA. This means that any residues with light IAM were not protected and were unlikely to be participating in disulfide bonding. Residues with heavy IAA were protected from IAM until reduced, and therefore are likely to participate in disulfide bonding.
Cysteine residues 21, 138 and 218 were not detected (Figure 5.12). This may due to the formation of large fragments post-digestion that are not able to be detected by the mass spectrometer. Cysteine 65 showed both labels, which suggests that it could be participating in disulfide bonding such as predicted. Cysteine 129, 197 and 236 were not protected, and therefore do not appear to be participate in disulfide bonding. Only one cysteine residue, C230 was confirmed to form a disulfide bond. This residue was not predicted to form a disulfide bond, thus this was an unexpected result.

As there were only two residues that appeared to form disulfide bonds, C230 and C65, these residues were assessed spatially in the structure in Figure 5.10 to determine if they could interact. Cysteine 65 is constrained by an α-helix, whilst C230 is part of a long loop region. The flexible loop region of C230 could permit the movement required to allow these residues to come into close proximity.

Unfortunately, two residues of interest (C218 and C138) were not able to be detected. Cysteine C65 also give rise to an ambiguous result. It would be preferable to repeat this experiment to ensure the reliability of these results, and to resolve the uncertainty of specific residues. Using chemical labels that have a greater mass difference such as
proprietary isotope coded affinity tag reagents would aid in improved data as it is easier to distinguish between the two labels. Digestion with a different enzyme could also improve the result, as different peptides would result. A further experiment that could be conducted is mutating each of the cysteine residues and assessing protein folding. Removal of cysteines participating in disulfide bonds should decrease protein stability, and prevent proper folding.

Intriguingly, the presence of disulfide bond may permit a conformational change in relation to redox state. The role of cysteine residues as redox sensitive switches has been seen with other proteins (Barford, 2004). As plant CDO acts as an oxygen sensing protein, this cysteine mediated conformational change may switch the protein from an ‘off’ to an ‘on’ state or vice versa. For example, if normoxia oxidised the cysteine residues, this would encourage disulfide bonds and lead to plant CDO being switched ‘on’. The ‘on’ state would then allow the protein to act on targets, as they are not required in normoxia. Conversely, when oxygen levels are low the cysteine residues would be reduced and not able to form disulfide bonds. This would switch the protein ‘off’ by encouraging a conformational state that prevents or decreases substrate binding. There is also a possibility that oxidation reduces plant CDO activity, and reduction increases it. This relationship could be explored further using reactive oxygen species and reductants in vitro.

As MS experiments showed that PCO1 may contain disulfide bonds, a refolding experiment was conducted to determine if this could improve protein stability. Refolding was induced by incubating PCO1 overnight with varying ratios of reduced (GSH) and oxidised (GSSG) glutathione. GSH encourages the reduction of disulfide bonds, whilst GSSG aids in the formation of disulfide bonds. Presence of a mixture of GSH and GSSG allows for the re-organisation of disulfide bonds, to ensure the correct bonds are formed (Gilbert, 1990). The correct disulfide bond formations permit the protein to fold properly, enhancing protein stability.
Figure 5.13. PCO1 refolding samples ± BME. Lane labelled C contains control, 1, 5 mM GSH, 0.5 mM GSSG, 2, 5 mM GSSG, 0.5 mM GSH and 3, 2 mM of each. Where PCO1 is expected to run is indicated with a black arrow. The grey arrow indicates the band that had relative motility calculated. The lane labelled M contains molecular weight marker. Marker as in 5.11.

The results could not be assessed under the conditions of BME as this reduces any newly formed disulfide bonds. A reducing SDS-PAGE gel was run for comparison, and to assess protein concentration (Figure 5.13, left). The control sample, as well as sample 3 (5 mM GSSG, 0.5 mM GSH) appeared to have less protein loaded, even though all samples were prepared from the same stock (Figure 5.13, right). All samples in the presence of glutathione appear to have formed higher order structures.

The higher order structures may comprise PCO1, or PCO1 and multiple DnaK monomers. To determine this, the bands at ~68, 140 and 180 kDa were excised and subjected to identification using MS. The results showed that the ~68 kDa band comprised DnaK, and the other two bands were higher order structures of PCO1 as only background levels of DnaK was found (Appendix 25). The size of the ~140 kDa molecule was determined using Rf (Appendix 26) to estimate how many PCO1 monomers were present. The results indicated that this band was a tetramer (138 kDa, allows for four 34 kDa PCO1 monomers), and it is likely that the above bands are simply the addition of further PCO1 monomers.
The formation of higher order structures of PCO1 in a non-reducing gel suggest that intermolecular disulfide bonds may be present. Strangely, no dimer or trimers were observed. This means that several CDO molecules are required to form stable higher order structures. Interestingly, intracellular proteins usually lack disulfide bonds as the cytosol is a reducing environment. However, these experiments suggest there are intermolecular disulfide bonds, and thus when difficulty with solubility and expression arose with some constructs Origami™ cells were used (Chapter 4). The samples from this experiment were further analysed by testing iron binding efficiency and performing kinetic assays; these results are found in Chapter 6.

5.3 Domain analysis

PCO1 was submitted to a range of online servers to decipher if the protein contained any motifs. ScanProsite showed several potential glycosylation and myristylation sites as well as a multitude of potential phosphorylation sites (Appendix 27). However, all of these potential post-translational modifications are often found with high frequency, and therefore significance is difficult to assess lacking biological information. The final motif hit suggested an N-terminal nuclear localisation signal (NLS).

NLS tags allow entry of the protein into the nucleus of the cell. Two classic types of NLS exist, monopartite and bipartite. Monopartite consists of either a series of basic amino acids (>4) (Kalderon et al., 1984) or the consensus sequence K(K/R)X(K/R) (Dang and Lee, 1988). Bipartite is a series of basic amino acids followed by a 10-12 amino acid linker, followed by a second string of basic amino acids (K/R)(K/R)X_{10-12}(K/R)_{3/5}. The size of linker sequence is optimal for the binding groove of importin, the nuclear localisation mediating protein (Kosugi et al., 2009).

The predicted NLS of PCO1 was scored with low confidence, suggesting that this should be assessed with caution. This proposed NLS is within the N-terminal domain of the protein, which is thought to be disordered. The sequence was reported to be bipartite, conforming to the proposed consensus sequence (K/R)(K/R)X_{10-12}(K/R)_{3/5}. Of the nuclear proteins, 56% contain a NLS whilst only 4% of cytoplasmic proteins contain this motif. This means that almost all proteins with this motif are transported into the nucleus.
This result is consistent with protein size, as only proteins smaller than 40 kDa are able to be transported through the nuclear pores (Lange et al., 2006). This would also explain why the N-terminus lacks structure as it needs to be exposed for protein-protein interactions that allow for nuclear import. This is consistent with data by Weits et al. which used fluorescent labelling to show that PCO1 is a nuclear protein (Weits et al., 2014).

For 90% of the proteins with a known NLS and DNA binding domain, these regions overlapped in the protein sequence (Cokol et al., 2000). This suggests a dual purpose for the sequence, and reflects that NLS are comprised of many positively charged basic residues. These regions allow for an electrostatic interaction with the negative charged DNA, as well as regulation by nuclear transport mediating proteins. The linker region of bipartite NLS has also been shown to interact with DNA, mediating DNA binding (Choo and Klug, 1993). This result further suggests that DNA binding by PCO1 is not an artefact of purification, and that it may be important for protein activity in vivo (Chapter 3).

To experimentally determine if PCO1 does contain an N-terminal NLS, fusion to a green fluorescent protein (GFP) could be used. If PCO1 and ΔN were both tagged with GFP, one could assess protein localisation within the cell. If the N-terminal domain does contain a NLS then PCO1 would enter the nucleus, whilst ΔN would stay in the cytoplasm.

5.4 Crystallography

After producing computational homology models, crystallography was attempted to produce an experimental model. Although homology models provide a detailed view of predicted protein structure, generating an experimental model is much more reliable. Forming an experimental model would also allow the presence of disulfide bonds to be explored, as well as suggesting which residues are important for catalysis based on proximity to the active site.
5.4.1 *Arabidopsis thaliana* constructs

Before crystallography was performed, purified *A. thaliana* protein samples were assessed. Ideally, protein samples are monodispersed, as this implies that the sample is homogenous. Homogenous protein samples are more likely to form crystals as each of the protein units are identical, allowing for optimal crystal packing. The technique used was dynamic light scattering (DLS) to determine the size and distribution of particles in solution. Several parameters were assessed.

**Table 5.1. DLS of *A. thaliana* PCO1 derivatives.**

<table>
<thead>
<tr>
<th>Construct</th>
<th>R (nm) 1</th>
<th>Polydispersity (%) 2</th>
<th>MW (kDa)</th>
<th>Mass (%) 3</th>
<th>Counts 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCO1</td>
<td>9.9</td>
<td>36.1</td>
<td>721</td>
<td>96.6</td>
<td>27</td>
</tr>
<tr>
<td>ΔN</td>
<td>2.6</td>
<td>12.7</td>
<td>33</td>
<td>95</td>
<td>27</td>
</tr>
<tr>
<td>ΔNΔC</td>
<td>7.3</td>
<td>23</td>
<td>351</td>
<td>77</td>
<td>29</td>
</tr>
</tbody>
</table>

1 R refers the hydrodynamic radius, and is a further measure of molecule size.
2 Polydispersity <20% is indicative of monodispersity.
3 Mass shows what percentage of the total solution is comprised of the stated molecular weight.
4 Counts indicates the number of readings taken to obtain the values listed.

DLS suggested that the purified PCO1 assessed was not a homogenous sample as polydispersity is 36% (**Table 5.1**). The molecular weight of PCO1 was predicted to be ~34 kDa, therefore a molecular weight of 721 kDa implies that large aggregates were present in solution, which may have been comprised of PCO1 alone or PCO1 in complex with chaperones. DNA bound to PCO1 may have also influenced the results. ΔN had low polydispersity, and the molecular weight was close to the predicted molecular weight of ΔN of 28 kDa. This shows that the purified ΔN assessed was homogenous. ΔNΔC had 23% polydispersity, which is above the recommended value for homogenous protein samples. The molecular weight reported, 351 kDa, was much higher than the predicted molecular weight of ΔNΔC of 23 kDa. This suggested that large aggregates of ΔNΔC or ΔNΔC in complex with chaperones were present such as with PCO1.
DLS results for PCO1 suggested very large aggregates were present in solution. DLS can at times be problematic, giving false readings due to dust present in the samples or cuvette. Given this, a further experiment was conducted to ensure the reliability of the DLS data for PCO1. Gel filtration was used as it enables one to determine the size of proteins components within solution. If large aggregates are present, then a peak will be present early in the chromatogram.

Figure 5.14. PCO1 Superdex 200 gel filtration. The blue line represents absorbance at 280 nm in mAU (left axis) and the red line represents conductivity in mS/cm (right axis). Standard elution volumes indicated by a black arrow, with molecular weight indicated in kDa. Standard proteins: Thyroglobulin (660 kDa), Ovalbumin (43 kDa), Myoglobin (17 kDa) and Vitamin B12 (2 kDa). Standard values from (Hsu, 2009). Fractions relate to the samples used for Figure 5.15. Resin was sephadex.

Gel filtration resulted in three clear peaks (Figure 5.14). The first peak at 120 mL eluted quickly, and therefore was made of larger protein/s. The standard indicates this peak was comprised of proteins slightly larger than 670 kDa. This is likely to be the large aggregates of PCO1 detected by DLS. There are two very small peaks, one at 185 and one at 220 mL elution volume. The sizes of molecules eluting in this peak were ~65 kDa and ~35 kDa, which is consistent with DnaK and PCO1 respectively. The second main peak that begins at 270 mL, overlapped with a peak at 300 mL elution volume. As these peaks eluted later, they comprised smaller molecules (around 2 kDa). Gels were run to determine what the peaks comprised.
Figure 5.15. PCO1 Superdex 200 gel filtration SDS-PAGE gel. A letter followed by a number indicates a fraction, and numbers above indicates the first number of the fraction e.g. 1.C.2. The lane labelled M contains molecular weight marker. Where PCO1 is expected to run is indicated with a black arrow. Marker as in Figure 5.10. Chromatogram that relates found in Figure 5.14. Ten microlitres of each fraction was loaded.

The first peak from the chromatogram in Figure 5.14 produced a band on an SDS-PAGE gel at ~36 kDa (Figure 5.15). This suggests that the peak comprised PCO1 aggregates or higher order structures. As this peak eluted early and comprised PCO1 this suggested the DLS was correct. The first of the insignificant peaks resulted in a thin band at ~ 69 kDa, suggesting that the peak comprised DnaK. The second insignificant peak, as well as the two peaks at the end of the run showed no bands on a gel. This implies that a non-protein molecule with an absorbance at 280 nm is eluting, or that the concentration is below that which can detected by SDS-PAGE.

Overall, it appears that co-purification with chaperones, DNA binding and higher order structures have affected the homogeneity of the PCO1. Co-purification with chaperones also has influenced homogeneity of ΔNΔC. This suggests that both PCO1 and ΔNΔC are not optimal for protein crystallography. Unfortunately, no means were produced to remove the chaperones and improve protein homogeneity (3.2.1). Therefore, these proteins were used for crystallography experiments regardless. In the past, protein crystals have been formed from impure samples, with some protein purification methods even employing crystallography to separate the protein of interest from other cellular proteins. ΔN protein samples appear to be ideal for crystal screening post-purification.
This is due to the fact ΔN is monodisperse, and monomeric. Given this, crystallographic efforts focused primarily on ΔN. All protein samples were incubated with Fe$^{2+}$ before plating unless stated otherwise.

**Pre-crystallisation screen**

All initial protein samples were subjected to a pre-crystallisation screen. This gives one an idea of what protein concentration to use for crystal screens. To do this, protein samples were mixed with reagents and precipitation type was observed. Heavy precipitate suggested lowering protein concentration, whilst clear drops suggested increasing protein concentration. The ideal concentration for all constructs was below 8 mg/mL, as higher concentrations resulted in heavy, amorphous precipitation.

**High throughput screening**

Crystal Screen HT was the first high-throughput screen used, and consisted of drops with 400 nL of 5 mg/mL protein and 550 nL of well solution. Almost all resulting drops had some form of precipitation but no crystals were formed and no interesting conditions were determined from this plate. This experiment was followed up by a second trial using the same screen with fresher protein and 3 different protein concentrations (2, 3.5 and 5 mg/mL). The method was also changed, as sitting drops were used. Again, nearly all drops resulted in some form of precipitation (not pictured). However, with a protein concentration of 2 mg/mL, a potential crystal was obtained after 4 days (**Figure 5.16, left**). The proposed crystal has a symmetrical edge and optical activity in cross polarised light (**Figure 5.16, right**).
Figure 5.16. Putative PCO1 crystal from the second Crystal Screen HT attempt. Left is an image taken under visible light, right is an image taken under cross polarised light. Conditions include 0.5 M ammonium sulfate, 30% v/v 2-methyl-2,4-pentanediol (MPD) and 0.1 M 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, pH 7.5). Protein concentration was 2 mg/mL.

Although a potential crystal was obtained during high throughput screening (Figure 5.16), it was not large enough to use for X-ray diffraction. Therefore, the conditions in which the proposed crystal were obtained were optimised to see if a larger crystal could be produced. Although a range of drops formed precipitate, no crystals were produced.

The next attempts focused on ΔN and ΔNΔC, and changing the metal ions used. Other metal ions were added as mammalian CDO was first crystallised with Ni$^{2+}$ bound, and at the time no experimental evidence existed to prove that plant CDO’s preferred metal ion is iron. PCO1 was incubated with either Ni$^{2+}$, Co$^{2+}$ or Fe$^{2+}$ and used with crystal screen HT. The shortened constructs were used with crystal screen HT after incubation with only Fe$^{2+}$ as the addition of the other metal ions caused immediate precipitation. The final protein concentration for all constructs were 8, 6 and 2 mg/mL. Promising results were found in several drops of the PCO1 plates, as well as a drop prepared using ΔN. No crystals were formed by ΔNΔC under these conditions.
Figure 5.17. Putative PCO1 crystals after incubation with various metal ions. Left, crystals imaged under visible light formed under condition 0.2 M Magnesium chloride, 0.2 M HEPES sodium pH 7.5 and 30% v/v 2-propanol with 2 mg/mL of PCO1 incubated with cobalt. This is representative of other crystals formed in these conditions. Right, diffraction pattern of a frozen crystal taken from drop to the left.

Condition A12 (0.2 M Magnesium chloride, 0.2 M HEPES sodium pH 7.5 and 30% v/v 2-propanol) from the Crystal Screen HT resulted in the formation of crystals for the 8, 6 and 2 mg/mL samples of PCO1 with Fe$^{2+}$ and Co$^{2+}$ (Figure 5.17, left). Crystals only formed in the 2 mg/mL sample of PCO1 with Ni$^{2+}$. Crystals from the 2 mg/mL PCO1 with Co$^{2+}$ drop were mounted at both room temperature and frozen in liquid nitrogen. For the frozen crystal, glycerol was added as a cryoprotectant. A diffraction pattern was then collected for both crystals mounted. The diffraction pattern for both indicated that all crystals under this condition comprised salt (Figure 5.17, right).
**Figure 5.1. Drops containing sea urchin crystals.** Both drops are under the conditions: 0.2 M ammonium sulfate, 0.1 M sodium cacodylate trihydrate pH 6.5 and 30% w/v polyethylene glycol (PEG) 8,000. Left contains 6 mg/mL of PCO1 incubated with Fe$^{2+}$, whereas right contains 2 mg/mL of ΔN also incubated with Fe$^{2+}$. Images taken under visible light.

Two drops under the conditions of 0.2 M ammonium sulfate, 0.1 M sodium cacodylate trihydrate pH 6.5 and 30% w/v PEG 8,000 also lead to the formation of crystals (Figure 5.18). These crystals differ to those in Figure 5.17 in their morphology. As these conditions differ greatly compared to those that formed salt crystals, and as the morphology also differs, it may have indicated that these crystals were made of protein. “Sea urchin” crystals such as those seen in Figure 5.18 are difficult to use for diffraction as they generally result in single crystals that are too small. However, the results from these conditions gave rise to the opportunity to optimise the conditions to produce larger, single crystals that may be suitable for diffraction.

To optimise these conditions, a range of PEG 8,000 and PEG MME 5,000 concentrations were used from 24 – 34%, with other solutes kept constant. Final protein concentrations of PCO1 ranged from 2-8 mg/mL and ΔN ranged from 0.5-4 mg/mL. An additive screen was also used with conditions of Figure 5.18 for optimisation. This was a Morpheus additive screen at 10% v/v which was added to the well solution. A further plate was prepared that varied all solutes. Ammonium sulfate was used at 0.1-0.4 M whilst sodium cacodylate was used at 0.1 M with pH ranging from 5 to 7.4. Drops prepared were both sitting and hanging, and ranged from 0.3-2 μL in size. Unfortunately, all variation screens produced were unsuccessful, and the initial experiment could not be reproduced.
As varying the conditions from Figure 5.18 did not result in any crystals, a range of other screens were employed. Three plates were prepared using the following screens: JCSG, PACT and INDEX. PCO1 and ΔN concentrations used were 6, 4 and 2 mg/mL, and drops prepared were 0.5 μL sitting drops. As the addition of substrate to enzymes has aided with crystallisation in the past (Hassell et al., 2007), these screens were produced both with and without the addition of a penta-peptide substrate. The peptide sequence (CGGAI) was derived from the N-terminal motif of PCO1 targets. The peptide was added after incubation with iron, with a 2 fold more peptide than protein. These twelve plates did not produce any promising results, nor suggest conditions that could be optimised to produce crystals.

As no improved crystals were produced in the JCSG, PACT and INDEX screens, the sea urchin crystals in Figure 5.18 were broken into single crystals and mounted. These crystals were small in size (< 50 μM) so were shot using the MX2 beam line at the Australian Synchrotron. These crystals did not diffract, but showed no evidence of being a salt crystal. Thus it was predicted that these crystals were formed of poorly ordered protein. As the sea urchins could not be reproduced, and the crystals that were formed did not diffract, further crystallography screens were trialled.

As PEG showed some success in the past, 96-well PEGRx HT and PEG ion HT screens were purchased. These screens were prepared with both PCO1 and ΔN at 6, 4 and 2 mg/mL. Drops prepared were sitting, and 0.3- 0.6 μL in size. A further PEG screen (MD1-14 clear strategy screen 1) was also trialled in a 24-well format with 10, 8, 6 and 4 mg/mL of both PCO1 and ΔN. Drops were sitting, and were 2 μL in size. None of the above PEG screens were successful in producing promising conditions or crystals.

Attempts were also made to crystallise PCO2. Three screens were employed for a wide range of coverage. These screens were Crystal Screen HT, PEG ion HT and PEG Rx HT. The protein was used at 6, 4 and 2 mg/mL concentrations in sitting drops that were 0.3-0.6 μL in size. No crystals resulted and the screens gave little insight into optimisation to produce crystals.
5.4.2 *Zea mays* constructs

After the lack of success with all *A. thaliana* constructs, plant CDO from maize was employed. Both a full length maize (maize A) and N-terminally truncated version of maize (maize B) were able to be expressed and purified. Maize A and maize B were used with 2 screens: Crystal Screen HT and PEG Rx HT. Drops were sitting and volumes ranged from 300 – 400 nL. Protein concentrations for both maize constructs were 6, 4 and 2 mg/mL. No crystals or drops of interest were formed from maize A, but maize B resulted in two drops of interest.

![Figure 5.19. Maize B spherulites from Crystal Screen HT. Both images taken under visible light. Left are spherulites formed in 0.2 M calcium chloride, 0.1 M HEPES pH 7.5 and 28% v/v PEG 400. Right are spherulites formed in 0.2 M calcium acetate, 0.1 M sodium cacodylate and 18% w/v PEG 8000. Spherulites were formed in these conditions under all maize B concentrations tested.](image)

Maize B produced small spherulites in two different conditions (Figure 5.19). The spherulites clustered around areas of dense protein precipitation. Spherulites from the right drop did not have any absorbance at 280 nm (Appendix 28). This does not give an indication that the spherulites were made of protein, but does not rule it out. Spherulites from the right drop showed absorbance at 280 nm and had activity in cross polarised light (Appendix 29). This features suggest that the spherulites comprised protein. As the only similarity between the two drop solutions was the presence of calcium and PEG, it was predicted that calcium was inducing the production of the spherulites. Given this, a screen
was set up with a range of PEG and buffer conditions (Rows A-D of the PEGRx HT screen) with 0.05, 0.1 and 0.2 M calcium chloride or calcium acetate. The maize B concentration used was 3 mg/mL.

Figure 5.20. Optimised Maize B spherulites from PEGRx HT with calcium. Left image was taken under cross polarised light, whilst right image was taken under visible light. Spherulites formed in 0.1 M calcium chloride, 0.1 M Tris pH 8 and 30% v/v PEG 400. Maize B concentration was 3 mg/mL.

Larger spherulites were formed in many of the drops on the PEGRx HT with calcium plate such as those in Figure 5.20. The spherulites pictured absorbed at 280 nm and had activity in cross polarised light suggesting that they may comprise protein (Appendix 30). Drops that contained calcium chloride were more likely to form larger spherulites, therefore calcium chloride was used in future experiments. Although this is an improvement on the small spherulites observed in the past, no crystals were produced. It was proposed that making larger crystal drops may slow nucleation and enhance the likelihood of crystal formation, thus this was performed. The larger drops were 2.4 μL in size, with a final protein concentration of 3 mg/mL.
Two drops of interest were produced. One drop had sea urchin crystals under visible light (Figure 5.21, left), but lacked any cross polarised activity and had no absorbance at 280 nm elements. The lack of these features suggests that the sea urchin crystals may not be comprised of protein. The other drop contained crystalline looking features (Figure 5.21, right). These putative crystals not only had activity under cross polarised light but also had absorbance at 280 nm (Appendix 31), this suggest that they comprised protein. Given the results for this, a further optimisation plate was produced.

For the variation screen buffer pH, PEG concentration and protein concentration were varied. Citric acid was used at 0.1 M pH 3, 3.5, 4 and 4.5, and PEG 200 was varied from 40-20 % in 4% increments. BIS-TRIS propane was used at 0.1 M, pH 8, 8.5, 9 and 9.5, whilst PEG 6000 was varied from 34-14% in 4% increments. The maize B concentrations used were 4, 3 and 2 mg/mL and drops sizes were 300-400 nL. Calcium chloride was added to the protein stock for a final concentration of 0.1 M. Unfortunately, no crystals were produced, and the original results were not able to be replicated. Seeding could be attempted in the future.
5.5 Summary

Homology models of plant CDO were produced from two different servers. Both models gave varying ideas on protein structure and substrate entry. The histidine triad was present in a similar orientation to mammalian CDO in both models. Both models suggested that the N- and C-termini lack order, supporting results of previous experiments. Modelling also suggested that PCO1 may bind its target using ionic interactions. After all cysteine residues were mapped onto one of the models, it was predicted that plant CDO contained disulfide bonds. A range of experiments were then conducted which were able to support the presence of disulfide bonds in recombinant CDO. Although it is not known if the bonds are inter- or intra-protein, future experiments to determine this were outlined.

Crystallography began with assessing PCO1 and both A. thaliana shortened constructs. DLS suggested that ΔN was homogenous and monodisperse, whilst ΔNΔC and PCO1 were not. This was predicted to be due to the presence of chaperones and/or DNA binding in the case of PCO1. DLS suggested that PCO1 was forming large aggregates, so this was further tested and confirmed by the use of size exclusion chromatography. All A. thaliana constructs were used in an attempt to produce crystals. ΔN produced sea urchins, however they did not diffract. No other A. thaliana constructs formed diffracting crystals, thus constructs from maize were employed. Maize B produced promising results but no crystals suitable for X-ray diffraction were formed.
Chapter Six

Functional characterisation: Iron binding and kinetics

6.1 Iron binding
It is predicted that approximately 40% of all proteins bind metal ions, and are termed metalloproteins (Thomson and Gray, 1998). Around 25-35% of these proteins require a metal ion for catalysis, whilst almost all require metal ions for protein folding and stability (Waldron and Robinson, 2009). These ions are usually co-ordinated by oxygen, sulfur or nitrogen groups of amino acid side chains (Yamashita et al., 1990). Most often, transition metals (e.g Mn$^{2+}$, Fe$^{2+/3+}$, Co$^{2+}$, Ni$^{2+}$, Cu$^{2+}$ and Zn$^{2+}$) are bound as they have preferable physiochemical properties such as a small atomic radius and low density. These metals also allow for strong electromagnetic and electrostatic interactions, and are readily bioavailable (Williams, 1997).

Previous experiments with CDO from other origins showed CDO is an iron-binding protein and that iron is required for catalysis (Tchesnokov et al., 2012). Plant CDO also contains the same 3-His iron binding co-ordinate sites found in other CDO proteins. The binding stoichiometry of mammalian CDO was shown to be 1:1 experimentally by the ferrozine assay (Tchesnokov et al., 2012). The same experiment was attempted with PCO1 but the results were inconclusive. It was originally thought that this may be an interaction between the negatively charged DNA bound to the N-terminal, and the positively charged iron. The C-terminus also has a number of negative charges that could interact with the iron electrostatically. However, binding stoichiometry was also inconclusive with ΔN and ΔNΔC. Given this, it was predicted that plant CDO may not bind iron, or have several iron binding sites. Intact MS was employed to assess this.
Figure 6.1. Intact mass spectrometry data for ΔN. M refers to molecular weight (in Da) determined for each of the prominent peaks (left) and ΔM indicates the difference in molecular weight (in Da) between the two prominent peaks (right). The charge for both of these figures is +20.

The intact mass experiment for both PCO1 and ΔNΔC did not result in interpretable data. Therefore, only data produced by the ΔN was analysed. The results showed two prominent peaks; one peak with a mass of 27972 Da and a second peak with a mass of 28036 Da (Figure 6.1, left). The predicted average mass of the ΔN without the N-terminal methionine is 27974.28 Da, which corresponds well with the first large peak. The second large peak is believed to be ΔN with a metal ion bound, which has been sequestered from within *E. coli* during expression. The two insignificant peaks between the larger peaks are likely the protein co-ordinated with 2 and 3 water molecules respectively. The difference between the two large peaks is 63.63 Da (Figure 6.1, right) which matches the addition of zinc (65.38 u ± 0.002 u) minus two protons (to give no net change in charge). The weight of iron is 55.845 ± 0.002, which is substantially different than the change in mass observed.

This result is unexpected as other CDOs co-ordinate iron. Previous experiments with mammalian CDO show that it is able to bind zinc, but the resulting protein lacks catalytic activity (Chad R Simmons *et al.*, 2006). If plant CDO does bind zinc, this would not allow the redox chemistry to permit dioxidation of targets. This suggests that this may be an artefact. Further intact mass experiments and competition assays between zinc and
iron should be conducted to corroborate this result. Trace analysis could also be performed by the Centre for Trace Element Analysis within the Otago University Department of Chemistry.

This result is partially inconsistent with particle-induced X-ray emission with a micro-focussed beam experiments conducted by White et al. Both PCO1 and PCO4 were analysed, and results showed there were 0.3 ± 0.12 iron atoms per molecule for PCO1 and 0.31± 0.015 iron atoms per molecule for PCO4 after purification from E. coli. Both PCO1 and PCO4 also showed high levels of nickel binding (greater than that of iron), but this was predicted to be an artefact of His-tag affinity purification (White et al., 2017). Although the paper suggests that iron is the main metal ion co-ordinated, PCO4 did show trace amounts of zinc (0.13 ± 0.02 atoms per molecule).

As White et al. also showed some zinc binding, the result from Figure 6.1 was somewhat validated. However, this does not explain why iron was not shown to be bound. The result may be due to a change in intracellular metal concentrations between bacterial and plant species. Free or loosely bound metal ion levels must be kept low due to cytotoxicity, but the pool must be large enough to accommodate newly formed proteins and other cellular processes. In an E. coli the most common metal ion present is iron, however only 1% of total iron concentration comprised free iron (~10 μM) (Keyer and Imlay, 1996; Semsey et al., 2006). Zinc is the second most common metal ion in E. coli (Porcheron et al., 2013), but again free zinc is not readily available as it is in femtomolar concentrations (Outten and O’Halloran, 2001). As it is proposed that more iron is present in a typical E. coli cell compared to zinc, this does not explain these results. Another proposed theory was that the amount of metal ions found in media was likely to influence what is bound to the protein. As peptone is derived from an animal source the amount of metal ions vary with each batch. In LB, iron is often the most common metal present, followed closely by zinc (Outten and O’Halloran, 2001). This variability could explain differences in results here as compared to White et al.

As Figure 6.1 shows the intact mass of ΔN, this could explain the change in metal binding. PCO1 and ΔN could have different metal affinities, although this is unlikely as the active site is within the domain shared by both constructs. Zinc is not able to support
the redox chemistry required to perform the proposed catalytic reaction. This again suggested that zinc is either an artefact of expression in *E. coli* or as a non-native construct was analysed (Porcheron *et al.*, 2013). Given this, and that literature suggests that iron is the preferred central metal ion in other CDOs, the 3-His triad is present, and White *et al.* showed that plant CDO does bind iron, attempts were made to improve iron binding.

After the purification of plant CDO, iron occupancy was tested using a colorimetric ferrozine-based assay (Tchesnokov *et al.*, 2012). Without loading the sample with iron, occupancy usually remained below the accuracy of the assay post-purification. As metal ion binding usually enhances the stability of ion binding proteins, an experiment was performed to increase iron occupancy. This experiment consisted of expressing the protein with iron (III) in the media. This was used as opposed to Fe$^{2+}$ as bacterial iron transporters prefer the Fe$^{3+}$ oxidation state under aerobic conditions (Braun, 2003). A reduction in IPTG concentration was also tested as this could improve iron occupancy. As one would expect less protein to be expressed, there would then be a larger ratio of iron to protein in the media.

![Figure 6.2. PCO1 expression with iron and reduction of IPTG. Loading was normalised. Amount of IPTG used to express indicated at the top in μM, and the flask where iron was added is indicated by the addition of +Fe$^{3+}$. The negative symbol shows expression before IPTG addition and the + symbol shows expression 6 h post-IPTG addition. Where PCO1 is expected to run is indicated with a black arrow. The lane labelled M of the SDS-PAGE gel contains molecular weight marker. Marker masses are indicated in kDa and each band contains 500 ng of protein.](image-url)
Expression of PCO1 using 100 μM of IPTG resulted in a clear band of protein (Figure 6.2). Using only 1 and 0.1 μM did not lead to prominent levels of expression, as a dense band of protein is not found on the gel where PCO1 would be expected to run. There is a slight band for both IPTG concentrations but it is insignificant compared to the 100 μM samples. The flask that contained iron resulted in a dense band, suggesting that PCO1 was over-expressed in this sample. As it appeared that all samples expressed produced at least a small amount of PCO1, all cultures were spun into a pellet and PCO1 was attempted to be purified.

![Image of SDS-PAGE gel](image)

**Figure 6.3. Bench-top purification of PCO1 with varying IPTG induction concentrations.** Amount of IPTG used to express indicated at the top in μM. H shows homogenate, S- supernatant, F- flow-through and P- purified protein. Where PCO1 is expected to run is indicated with a black arrow. The lane labelled M of the SDS-PAGE gel contains molecular weight marker. Marker as in Figure 6.2.

There is a large band of PCO1 in the homogenate lane of the 100 μM sample in Figure 6.3. A band is also seen in the supernatant, suggesting the protein was mostly soluble. A band in the protein lane shows that PCO1 was able to be purified, but a band still seen in the flow-through suggests that a second purification could have yielded more protein. Purification of the 1 μM sample shows minimal PCO1 in the homogenate, supernatant or flow-through. Given this, it is unsurprising that no PCO1 was present in the purified lane. Other samples are shown in Figure 6.4.
**Figure 6.4.** Bench-top purification of PCO1 with varying IPTG induction concentrations and iron addition. Amount of IPTG used to express indicated at the top in \( \mu \text{M} \), and flask expressed with iron is indicated with \( \text{Fe}^{3+} \). H shows homogenate, S- supernatant, F- flow-through and P- purified protein. Where PCO1 is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in **Figure 6.2**.

The 0.1 \( \mu \text{M} \) IPTG sample (**Figure 6.4**) shows the same trend as the 1 \( \mu \text{M} \) sample from **Figure 6.3**. It appears there was no band of PCO1 in the homogenate, supernatant or flow-through lanes. Thus, the lack of purified protein is not surprising. This implies that little PCO1 was expressed, which concurs with the absence of a dense band in **Figure 6.2**. The 100 \( \mu \text{M} \) sample expressed with iron in the media shows considerable bands in the homogenate and the supernatant. As these bands are the same density, this implies that all protein present was soluble. A clean band is seen in the protein lane showing that PCO1 was successfully purified. The presence of a band in the flow through shows that there was an excess of PCO1 for the amount of resin used, and that more protein could be purified with a further round of purification.

Overall, a reduction in IPTG concentration lowered the levels of expression dramatically. This then prevented the purification of detectable protein levels from these samples. Protein was able to be purified from both samples that were induced with 100 \( \mu \text{M} \) of IPTG. The addition of iron to the media did not change the amount of protein that was expressed or purified. Both of the purified protein samples were tested to determine the amount of iron bound to the protein.
Figure 6.5. Iron saturation levels of purified protein ± iron addition. The control sample was induced with 100 μM IPTG and did not have iron added to the media. The growth with iron sample (+Iron) contained 200 μM of Fe^{3+} within the media and was induced with 100 μM IPTG. Error bars shown with purified protein samples depict the assay detection limit whilst error bars shown with post-purification iron addition show standard error of the mean.

Both purified protein samples had low levels of iron binding, below the assay detection limit (Figure 6.5). However, a large increase in iron saturation was able to be achieved when iron was added to purified protein for both samples. The control sample occupancy levels rose to around 60% and the +iron sample rose to around 70%. Although the mean implies that a higher level of saturation was achieved in the +iron sample, this is not statistically significant. In conclusion, adding iron to the medium during growth did not improve the initial iron saturation levels or iron uptake after purification. Thus, growth with iron was not a worthwhile addition to the expression protocol.

As previous experiments with mammalian CDO showed that iron binding was essential for catalysis (Tchesnokov et al., 2012), it was postulated that this is also true of plant CDO. To determine if iron binding was essential for catalysis, the refolding experiment samples from 5.2 were used. These samples were predicted to form a range of folded states, therefore iron binding efficiency of each sample was predicted to be different. Iron occupancy of each sample was tested using the ferrozine assay. Each sample was then subjected to a kinetic assessment using Ellman’s assay. Iron occupancy and catalytic efficiency were also used to determine the effectiveness of each of the refolding glutathione ratios.
Figure 6.6 Iron bound to refolding experiment samples. Amount of iron bound is in μM and the concentration of PCO1 was 100 μM. GSH refers to reduced glutathione, whilst GSSG refers to oxidised glutathione. Data averages shown as mean ± standard error of the mean. One way ANOVA used for statistical significance, P-value of star shown is 0.0194.

The control sample (i.e protein as purified) resulted in ~80 μM of iron bound, whilst the 5 mM GSH/0.5 mM GSSG sample appeared to have more bound with ~90 μM of iron (Figure 6.6). However, the difference between the two is not statistically significant. The 5 mM GSSG/0.5 mM GSH sample resulted in ~40 μM of iron bound, and 2 mM of each enzyme resulted in ~70 μM bound. Both of these samples bind significantly less iron than the control and 5 mM GSH/0.5 mM GSSG samples. Thus, the addition of >0.5 mM of GSSG resulted in reduced iron binding.

As less iron binding was seen with the addition of 2 mM GSSG and the occupancy lowered further with 5 mM GSSG, there is an inverse relationship between GSSG addition and iron occupancy. Increased iron occupancy is thought to be a measure of native protein folding, thus these results suggested that the protein is better folded with GSH present. As GSH reduces the exposed cysteine residues in PCO1, this indicates that the protein is in a more stable conformation state with reduced cysteine residues. To
determine how conformational state and iron binding influenced catalytic efficiency, kinetic assays were conducted.

**Figure 6.7 Velocity of substrate depletion using refolding experiment samples.** The concentration of PCO1 was 5 μM as determined by the ferrozine, thus all protein measured had iron bound. The substrate used was pentapeptide (sequence CGGAI) at 25 mM to ensure saturation. GSH refers to reduced glutathione, whilst GSSG refers to oxidised glutathione. Raw data with error bars found in Appendix 32. Data averages shown as mean ± standard error of the mean. T-test showed a P-value of 0.0359 for samples with a star.

With a PCO1 concentration of 5 μM, the control sample showed an observed velocity of 0.3 mM/min, whilst the 5 mM GSH/0.5 mM GSSG sample showed upwards of 0.4 mM/min (Figure 6.7). The sample with 5 mM GSSG/0.5 mM GSH showed little activity with only ~0.05 mM/min. The 2 mM of GSSG and GSH sample achieved an observed velocity of ~0.35mM/min- with slightly higher activity than that of the control sample. The 5 mM GSH/0.5 mM GSSG sample proceeded with the highest velocity, followed by the 2 mM each and control samples. The 5 mM GSSG/0.5 mM GSH sample was significantly slower than the other samples. Again, this data suggests that >0.5 mM GSSG results in decreased activity.
A clear correlation is seen between iron occupancy and velocity of substrate depletion. The sample that contained the most iron proceeded the quickest. The next two samples had similar iron levels, and thus catalysis was at a similar rate. The sample with the lowest iron concentration resulted in the slowest substrate depletion, but at an even slower rate than one would predict. This implies a catalytic threshold for measured iron occupancy, where <50% occupancy results in a relatively inactive protein species. It is possible that \textit{in vivo} iron occupancy differs to \textit{in vitro} results. This may be because artefactual zinc may offer some competition for binding.

The change in iron binding/activity can also be linked to the redox state of cysteine residues present within the protein. When increased amounts of reducing GSH is present, catalytic efficiency increases. As the amount of GSSG increases past 0.5 mM, substrate depletion slows. This implies that the protein is more active when cysteine residues are reduced, and less active when cysteine residues are oxidised. This may offer a mechanism for control of plant CDO as it is a redox protein. Reactive oxygen species could be tested to determine how they influence the catalytic activity of PCO1. The flooding response, like other stress responses, is mediated in part by reactive oxygen species (Fukao \textit{et al.}, 2011). This is consistent with oxidising conditions turning CDO “off”.

6.2 Kinetics

Gaining an understanding of how enzymes operate is essential for enzyme characterisation. Kinetic characterisation of an enzyme results in a range of parameters which describe the enzyme’s speed ($V_{\text{max}}$) or ($k_{\text{cat}}$), and affinity for substrate ($K_{\text{m}}$) (Michaelis and Menten, 1913). Characterisation can also be performed in the presence of inhibitors to determine their effects. Enzymes catalyse reactions by binding to substrate and lowering the activation energy of product formation. Most enzymes are optimised to bind one substrate in the correct orientation, while some can bind multiple substrates productively. Substrate or transition state analogues can also be employed to determine the catalytic mechanism, or active site features.

The native substrates of plant CDO are group VII ethylene response factor proteins. These proteins contain an N-terminal motif for recognition, which begins with a cysteine
residue. Experiments using an oxygen electrode and high performance liquid chromatography (HPLC) by Weits et al. (2014) showed that PCO1 was able to dioxidise free cysteine as opposed to an N-terminal cysteine as part of a peptide. As this substrate is much different compared to that of the native plant CDO substrate, the reported result was unexpected. In order to assess whether plant CDO is indeed able to catalyse the formation of CSA from free cysteine, $^1$H NMR was conducted.

Within the spectra of Figure 6.8, a significant number of peaks presented with scalar coupling causing further splitting. The CSA standard spectrum (green) consisted of two split doublets at 4.12 and 2.79 ppm, as well as a triplet with some splitting at 2.68 ppm. The cysteine standard (purple) produced a triplet at 3.94 ppm and a split quartet at 3.04 ppm. The peaks produced by cystine (red) are less prominent, but are visible with a quartet 4.08 ppm, a split doublet at 3.37 ppm and a second quartet 3.18 ppm. Other peaks within the red spectra are composed of cysteine that has not yet formed cystine.
Within the cysteine incubated with CDO sample (blue), a quartet is seen at 4.08 ppm, followed by a triplet with some splitting at 3.94 ppm. A large singlet composed of Tris is visible at 3.70 ppm, alongside a split doublet at 3.37 ppm and a second quartet ~3.18 ppm. A third quartet with splitting is present at 3.04 ppm. The peaks produced by the sample with CDO are slightly broader than the other samples due to the presence of paramagnetic iron.

When comparing the NMR spectra of cysteine incubated with plant CDO to the standard spectra, it is clear that plant CDO did not produce CSA under these conditions (blue, Figure 6.8). A triplet at 3.94 ppm and a split quartet at 3.04 ppm indicates the presence of cysteine, representing unutilised substrate. A quartet at 4.08 ppm, a split doublet at 3.37 ppm and a further quartet at 3.18 ppm is indicative of cystine. This implies that plant CDO forms cystine instead of CSA whilst utilising cysteine as substrate. As the cystine peaks produced in two hours with plant CDO are comparable in size to those produced overnight without enzyme (not shown), it is likely this reaction is enzyme catalysed. Mammalian CDO also shows this activity in some conditions.

Overall, proton NMR data showed that PCO1 is not able to dioxidise free cysteine. This means that plant CDO is not a cysteine dioxygenase, but a cysteinyl dioxygenase. Given these results, a dipeptide substrate (cysteine followed by a glycine) was used to determine if the addition of a second amino acid would enable dioxidation to occur. A glycine was chosen as the second residue based on the N-terminal motif of proposed plant CDO targets. The activity of PCO1 with this substrate was then characterised using Ellman’s assay.

In order for kinetic characterisation to begin, a titration was conducted to assess the relationship between enzyme concentration and substrate at a set concentration. If the graph is linear, it suggests that the reaction is limited only by the amount of enzyme present. A linear dependence also shows that there are no limiting factors to catalysis, such as product inhibition or lack of co-factors. Given this, the initial evaluation is crucial to ensure an appropriate enzyme concentration is used for Michaelis-Menten kinetics.
Figure 6.9. PCO1 titration with dipeptide substrate. Each data point was produced in duplicate using technical replicates and error bars represent standard error of the mean. Data produced using Ellman’s reagent in a 96-well plate format. Substrate concentration was set at 28 mM. Raw data can be found in Appendix 33.

Substrate depletion occurs at a rate of 0.04 mM/min without any enzyme present, suggesting a basal rate of reaction (Figure 6.9). Product formation in this case is likely dipeptide forming intermolecular disulfide bonds. This prevents interacting thiol groups from being measured and thus depletes substrate levels just as does CSA formation. This may be due to the pH of the reaction, as an increasing pH encourages disulfide bond formation due to decreasing hydrogen ion availability. This is further highlighted by the pKa of the thiol group of cysteine being noticeably lower than the pH of the reaction (pKa of 5.14 as opposed to reaction pH of 7.5). When the pKa is lower than that of the pH, the thiol group will be deprotonated to form a sulfur anion, which is free to act as a nucleophile in disulfide bond formation. The addition of DTT or ascorbic acid was considered, however dismissed as this would interfere with the results by reacting with the Ellman’s reagent, obscuring signal from substrate cysteine.

Velocity increases linearly from the basal rate based on increasing enzyme concentration until 10 μM of PCO1. Within this range, the reaction is not limited by product inhibition or by the lack of co-substrate, molecular oxygen. This data suggested that any of the enzyme concentrations tested above were suitable for Michaelis-Menten based analysis, and has identified a background level of cysteine formation. To ensure that the enzyme
concentration was easily measurable, and that it was well within the linear range, 6 μM of PCO1 was chosen for these experiments.

![Figure 6.10. Michaelis-Menten analysis of PCO1 with dipeptide substrate.](image)

**Figure 6.10. Michaelis-Menten analysis of PCO1 with dipeptide substrate.** Enzyme concentration was 6 μM and error bars represent standard error of the mean of technical triplicates. Data produced using Ellman’s reagent in a 96-well plate format. Raw data can be found in Appendix 34. Fit was done via GraphPad Prism Michaelis-Menten function. Data averages are shown as mean ± standard error of the mean. Curve represents a fit of Michaelis-Menten equation to data of 0-30 mM substrate. Fit has an R² of 0.98 and absolute sum of squares 0.0014.

The data were well fit by the Michaelis-Menten model as suggested by an R² of 0.98 and an absolute sum of squares of 0.0014. Above a substrate concentration of 25 mM no further rate enhancement is observed (Figure 6.10). \( V_{\text{obs}} \) plateaus as more active sites become occupied due to approaching substrate saturation. A \( V_{\text{max}} \) of 0.5 ± 0.1 mM/min was fit. Catalysis at 40 mM dipeptide concentration proceeds slower than one would expect. This may be because the reaction pH is heavily influenced by addition of a large volume of the stock peptide solution which has a pH of ~3. After the experiment, this was tested by the use of pH paper, which confirmed this theory. A resulting drop of pH could slow the enzyme, therefore this point was excluded from the fit.

The fit yielded a \( K_m \) of approximately 30 ± 10 mM. This \( K_m \) is high, reflecting that the enzyme does not have a high affinity for this substrate. This is not surprising as the dipeptide used is not the native substrate of PCO1. To explore the relationship between
the enzyme and substrate further, a larger peptide was employed. This peptide was a tripeptide, consisting of the amino acids (in order) CGG. This is again derived from the N-terminal motif of native PCO1 targets.

The same enzyme concentration was used so that the tripeptide data could be directly compared to the dipeptide data. Thus, experiments were conducted using a set PCO1 concentration of 6 μM, and used a range of substrate concentrations. However, under these conditions the reaction appeared to proceed too quickly to be quantitated in the same way. Substrate was 50% depleted before collection of the first sample at 4 mins. Given this, an experiment was conducted to further characterise the slope of the line.

![Figure 6.11. Substrate depletion of tri-peptide by PCO1.](image)

**Figure 6.11. Substrate depletion of tri-peptide by PCO1.** The concentration of enzyme used was 6 μM and initial tripeptide concentration was 15 mM. This was conducted once. Data analysis was produced using a 96-well plate format using Ellman’s reagent. Samples were taken every 15 sec for 6 min.

The trend seen in **Figure 6.11** shows a steep drop followed by a plateau. This repeats data seen in previous experiments with the tripeptide. It appears as though the reaction proceeds quickly in the first 2 min of the reaction until 5 mM substrate is reached. This is consistent with the $K_m$ of the interaction between PCO1 and this substrate being greater than 5 mM. Given this, further studies using the tripeptide should be done using a lower enzyme concentration to slow the linear phase. This would enable one to better assess the efficiency of PCO1 with the tripeptide substrate. Conversely, one could keep the enzyme concentration at 6 μM and characterise all samples within the first 4 min of the
reaction, only using the initial linear phase to plot the slope and determine reaction velocity.

This preliminary result suggests that PCO1 has a higher affinity for the tripeptide compared to CG as the predicted $K_m$ has decreased. As an apparent increase in substrate affinity with a peptide larger by only one amino acid has occurred, this may suggest that increasing peptide size may further increase substrate affinity. Given this relationship, one can hypothesise that substrate affinity would increase up until 10 amino acids where the N-terminal recognition motif of plant CDO targets end. A plateau would then likely occur, as amino acids outside this region are unlikely to interact directly with or near the active site of plant CDO. This interaction could be explored further using target N-terminal peptides of varying length as substrate. Future kinetic experiments should use tools like NMR and MS to determine whether or not the correct product is being formed as Ellman’s based assays only assess substrate depletion.

6.3 Summary

Intact mass spectrometry data suggested that plant CDO may bind Zn$^{2+}$ as opposed to iron as predicted. As this experiment was conducted with ΔN instead of full length PCO1, the change in metal affinity may be an artefact of truncation, or zinc binding could be an artefact of in vitro purification. Published data by White et al. suggested that PCO1 does in fact bind iron, thus experiments were conducted in an attempt to improve iron binding. Dropping the level of inducer reduced expression levels, and addition of iron to the media had no effect on purified protein iron occupancy. Instead, post-purification iron addition supported the greatest iron occupancy. Samples from the refolding experiment in Chapter 5 showed increased iron binding under mildly reducing conditions. This may imply improved protein folding in this state, and kinetic analysis suggests that the reduction and oxidation of disulfide bonds in PCO1 may control a redox switch manipulating catalytic activity.

Proton NMR showed that PCO1 is not able to catalyse the formation of CSA from free cysteine as can mammalian CDO. These results disputed data published by Weits et al. Kinetic analysis of PCO1 with a dipeptide substrate gave a $V_{\text{max}}$ of $0.5 \pm 0.1$ mM/min.
and a \( K_m \) of 30 ± 10 mM. These values suggest PCO1 is capable of catalysing dipeptide substrate depletion, but shows that substrate affinity is low. Preliminary analysis of PCO1 with tripeptide imply that catalysis is faster, and that the \( K_m \) is lower – suggesting increased substrate affinity. This may mean that larger substrates, closer to those of plant CDO’s native substrate, permit better recognition, catalytic efficiency and substrate affinity. Further kinetic assays using Ellman’s reagent should be conducted to explore this.
Conclusions

Conditions were established for expression and purification for plant CDO from a range of species. Expression trials of PCO1 and PCO2 from *A. thaliana* showed that optimal expression resulted from induction at 18°C for 6 h and 28°C for 20 h respectively. SDS-PAGE showed that PCO1 co-purified with two main contaminants, identified by MS as chaperones, implying that expressed protein is not fully folded. PCO1 produced by White *et al.* appeared to contain little heterogeneity, perhaps as a result of using a His-tag instead of a *Strep-tag®* (White *et al.*, 2017). PCO2 co-purified with 3 contaminants, two of which appeared to be the same contaminants found in purified PCO1. This suggested that PCO2 also presents with protein folding issues. Incubation of PCO1 with ATP followed by size exclusion chromatography was used in an attempt to remove bound chaperones. This was able to separate plant CDO from DnaK but with low yield. A second method was then attempted; EDTA dialysis followed by size exclusion. This resulted in less DnaK binding but had no influence on binding by groEL, just as with the ATP incubation.

As the chaperones could not be removed efficiently, two shortened variants of PCO1 were designed. These designs were informed by *in silico* disorder predictions and experimental definition of domain boundaries by limited proteolysis. PONDR® predicted both the N- and C-termini are disordered. This aligns well with results of the limited proteolysis, which showed cleavage of termini. Expression trials for these constructs showed that optimal expression of ΔN and ΔNΔC resulted from induction at 28°C for 6 h and 18°C for 6 h respectively. The ΔN variant resulted in protein with no contaminants bound, and thus was homogenous. Conversely, the ΔNΔC variant bound even more contaminants than PCO1. This suggests that removal of the N-terminals aids in stabilising the protein, or removes the binding site of the chaperones. Removal of the C-terminal destabilises the remaining catalytic domain.

Throughout the purification of PCO1, it was noted that the maximum absorbance was at 260 nm as opposed to 280 nm. This suggested that PCO1 was binding nucleic acid. An agarose gel was run and stained with ethidium bromide which confirmed the presence of nucleic acid. This novel finding was further explored by treatment with nucleases
suggesting that PCO1 was binding DNA. This binding could be physiologically relevant as the transcription factor targets of plant CDO also localise to DNA. Localisation would improve speed of substrate binding, which is important as environmental factors that cause anoxia, such as flooding, arise unpredictably.

To extend attempts to crystallise plant CDO, PCO2 from two additional species were explored. These constructs were designed based on experimental results with constructs from *A. thaliana* and analysis of disorder and structural predictions. Expression trials suggested that for the *Z. mays* constructs maize A and maize B, optimal expression resulted from induction at 28°C for 20 h and 18°C for 6 h respectively. *O. sativa* expression and purification was not as straightforward. Expression trials were attempted in BL21(DE3), BL21(DE3) pLysS and Origami™ (DE3) cells. It appeared as though ΔNR expressed in all cell lines based on SDS-PAGE gels, but purification resulted in no protein produced. Thus it appeared that the protein does not accumulate, or it was not expressed. The plasmid was sequenced, showing that the correct plasmid and insert were used. Other possibilities were explored, with the most likely cause being poor folding which resulted in post-translational degradation. Expression of rice CDO in *E. coli* was not pursued further.

Two different servers were used to produce models of PCO1, I-TASSER and SWISS-MODEL. Both models were similar, suggesting that PCO1 contains a core β-barrel homologous to mammalian and bacterial CDO. I-TASSER modelled the ends of PCO1, proposing that the N-terminal was a completely disordered loop region whilst the C-terminal had an α-helix alongside large loop regions. Terminal modelling was analysed with caution due to low confidence in the model termini due to lack of a template. SWISS-MODEL did not model the ends as the template mammalian CDO does not contain extended ends as in PCO1. Both models had the iron binding site at the same position as mammalian CDO as shown by superimposing the structures. This predicts that plant CDO is able to co-ordinate a metal ion.

Surface representations of these models predicted different modes of substrate entry. The SWISS-MODEL had an active site that transverses the entire catalytic domain, implying that molecular oxygen could enter one side, whilst the peptide is already bound to the
other. This would prevent the peptide from blocking co-substrate entry. The model produced by I-TASSER had a small entry site on only one side of the molecule. This opened up dramatically after removal of the N-terminal. This may mean that the N-terminal must move before the substrate can enter the catalytic site. This would allow for modulation of activity. Surface charges were also mapped to PCO1, showing that charge distribution was uneven. This may permit interaction with DNA on the positively charged side and RAP2.12 on the negative side as this protein also contains uneven charge distribution.

Other features of the models were explored such as cysteine arrangement and assessing potential catalytic residues. All cysteine residues present in the SWISS-MODEL were highlighted to evaluate whether or not disulfide bonding could be present. This was considered as there are a number of conserved cysteine residues in CDO across a range of plant species. It was also considered that a requirement for disulfide bonds may relate to why there are issues with protein folding in the full length constructs. Given where the residues were spatially, C65 and C129 were predicted to interact, along with C138 and C218. Furthermore, all of these residues are highly conserved across a range of plant species.

To experimentally determine this, differential labelling followed by MS was used. These results showed that C230 and C65 may interact, although a number of peptide fingerprints were not found during data assessment. These data showed that disulfide bonds are likely present in the structure. This may present a means of modulating activity as plant CDO is a redox protein. The formation of disulfide bonds could be a reflection of oxidation state in the environment, and thus modulate structure and protein activity as a result.

Further in silico experiments scanning motifs within PCO1 revealed that a NLS is present. This signal is found on the N-terminal of the protein and tags PCO1 for entry into the nucleus. This is an important feature of the protein, as transcription factor targets act within the nucleus and thus plant CDO may also be in the nucleus to dioxidise these targets when required. The NLS detected is a bipartite signal, which is characterised by the consensus sequence \((K/R)(K/R)X_{10-12}(K/R)_{3/5}\). It is found in PCO1 as:
KKNKNKNKKMMMTWRRK, with the spacer region being 12 amino acids. This aligns with localisation experiments conducted by Weits et al. who showed that PCO1 and PCO2 are nuclear proteins (Weits et al., 2014).

To determine the quality of the models produced, attempts were made to crystallise plant CDO. Several high through-put screens were prepared with PCO1, PCO2, ΔN, ΔNΔC, maize A and maize B. A greater number of screens were prepared with ΔN as this protein was shown to be homogenous by DLS. Other constructs assessed by DLS were predicted to form aggregates, or contain contaminant proteins. PCO1 aggregation was confirmed by size exclusion chromatography. Peptide substrate, and differing metal ions were used as additives to assess different ways of stabilising structure and promoting crystal growth. Sea urchin crystals were produced under a condition of crystal screen HT, with PCO1 and iron in the drop. Optimisation of these conditions was attempted but no further crystals were produced. The sea urchin crystals were mounted and assessed using a micro-focused beam at the Australian Synchrotron, but resulted in no diffraction pattern. Maize B also produced an interesting result from a condition of the crystal screen HT plate. A number of small spherulites formed after a few days. These conditions were extensively optimised, resulting in larger spherulites but no crystals were formed.

A number of difficulties presented when attempting to determine iron binding stoichiometry of PCO1. This raised questions about metal ion binding, including whether or not PCO1 binds iron at all. This question was explored using intact MS. Surprisingly, the results suggested that ΔN was binding zinc. This may have been an artefact of in vitro purification, or removing the N-terminal could have resulted in a differing metal affinity to PCO1. Published data by White et al. showed that PCO1 binds predominantly iron. Given this, attempts were made to increase iron binding of PCO1. This was attempted by lowering the IPTG concentration used to induce to lessen production of PCO1 or adding iron to the media. Neither of these methods worked, and instead the optimal way of increasing iron occupancy was determined to be loading of iron post-purification.

Kinetic studies were also performed. Previous studies by Weits et al. showed that PCO1 was able to dioxidise free cysteine to form CSA (Weits et al., 2014). This was tested by NMR, which showed that PCO1 did not catalyse the formation of CSA, but instead
catalysed the formation of cysteine. Peptide substrates were also explored. A di-peptide substrate designed from the N-terminal consensus of target proteins produced an approximate $K_m$ of $30 \pm 10$ mM and an approximate $V_{\text{max}}$ of $0.5 \pm 0.1$ mM/min. When a tri-peptide was used, the reaction proceeded much faster and proposed a lower $K_m$. This suggested that PCO1 has a higher affinity for longer substrates, which is unsurprising given the length of the natural peptide substrate. The refolding experiment also showed that PCO1 is able to dioxidise penta-peptide but parameters were not assessed. White et al. conducted NMR experiments which also show that PCO1 is able to dioxidise the N-terminal cysteine of peptides. These peptides were also modelled off target N-termini, and data produced proved that dioxidation was occurring (White et al., 2017). This aligns well with data presented in this thesis, which also suggests dioxidation is occurring on peptides modelled from the same N-terminal motif.

Overall, CDO from two plant species were able to be expressed and purified. This included 3 full length constructs and 3 shortened variants. A rice shortened variant did not accumulate to a detectable level, or was not able to be expressed. Removal of the N-terminal of A. thaliana PCO1 aided in protein homogeneity. This alongside sequence alignment, homology models and limited proteolysis suggested the N-terminal domain is disordered. This region also contains an NLS, and provides a charged region which contributes to DNA binding. These features may allow regulation of plant CDO. Furthermore, the disorder in this region may act as a regulatory switch as removal increased the size of the substrate entry site dramatically in a homology model. Removal of the C-terminal was detrimental suggesting it contributes to protein folding.

Homology modelling and MS data suggested that PCO1 contains disulfide bonds. Disulfide bonds are an important feature of protein structure and absence of these bonds in vitro may explain difficulties with protein folding. These disulfide bonds may also play a role in regulation of activity as plant CDO is a redox protein. Thus, as oxidation state of the environment can influence disulfide bonding of the protein, this may allow a switch that can turn plant CDO from an “off” to an “on” state or vice versa. $\Delta N$ was shown to bind zinc in vitro but this was not demonstrated with a native construct. Following this, kinetic data was produced which suggest that PCO1 is faster with larger peptide substrates, which is unsurprising given the size of the native target peptide.
Kinetic data also showed that PCO1 was not able to diooxidise free cysteine residues. Characterisation of plant CDO has given rise to a range of potential features that can be further explored. Understanding these features may eventually lead to manipulation that can enhance crop yield.
Future directions

This preliminary investigation established expression and purification conditions that open up a range of future directions for this study. A number of constructs were able to be purified, however almost all co-purified with chaperones. Further attempts to remove these chaperones may be beneficial to crystallographic and kinetic studies. One construct, ΔNR, was not able to be purified. It was determined that this may be due to the protein not expressing or accumulating as one would expect. This result could also be due to the truncation preventing proper protein folding. It would be of interest to explore this further: a first step would be attempting to purify the full length version of rice PCO2. This is of interest as it may aid in determining the structure of plant CDO using crystallography.

Exploration into the role of the N-terminus of PCO1 is also of interest. To determine if DNA binding is specific, a pool of randomly cleaved genomic DNA from *A. thaliana* could be incubated with nuclease-treated, purified PCO1. After recovery of PCO1 by affinity chromatography, one could sequence all co-purifying DNA. If over represented sequences are found, then synthetic DNA could be used to establish a consensus binding sequence. A NLS is also predicted to be present on the N-terminus of PCO1. This element could be studied further by fusing a GFP with PCO1 or ΔN, and expressing in plant tissue. If the NLS is present in the first 52 amino acids, PCO1 will translocate into the nucleus, whereas ΔN will not. In a complementary experiment it can be determined whether the first 52 residues tagged with GFP would translocate into the nucleus in a plant cell.

Future structural studies should be aimed to probe disulfide bonding, either inter- or intra-molecular. These bonds may be important not only structurally, but could also permit a means of catalytic regulation. This could be explored by treating PCO1 with reactive oxygen species or reductants prior to conducting enzymatic assays. Crystallography may be best continued with the use of further constructs from different species, both full length and those analogous to ΔN. Crystallography could also be continued with those already purified yet with chaperones removed. These chaperones could be removed by
protein dialysis in the presence of EDTA followed by gel filtration, even though low yield results.

Iron binding could be further explored by use of Mössbauer spectroscopy to explore oxidation state and binding orientation. The ferrozine assay could be used for determining stoichiometry of binding, yet needs to be further optimised. As DNA is negatively charged, nuclease treatment of PCO1 may remove potential unspecific iron binding and improve ferrozine data. This could also be achieved by exploring the iron binding stoichiometry of ΔN.

Further kinetic studies could also be conducted. Extending analysis of the relationship between substrate peptide size and catalytic properties would be of interest. This could be achieved by further Ellman’s assays which chart Michaelis-Menten kinetics with a range of peptide substrates. Other techniques like NMR and MS could be used to characterise products. Understanding these factors could aid in determining how substrate recognition occurs. These various future approaches could permit an understanding of structure, mechanism and regulation of plant CDO.
References


Arabidopsis Biological Resource Centre (2018) *Seed Handling*. Available at: https://abrc.osu.edu/seed-handling


Stipanuk, M. H. (2015) Stipanuk Lab Research. Available at:
http://www.human.cornell.edu/dns/stipanuk/research2.cfm


Zhang Lab (2018) I-TASSER server for protein structure and function prediction. Available at: https://zhanglab.ccmb.med.umich.edu/I-TASSER/about.html
Appendices

Appendix 1. RAP2.2 sequence aligned to RAP2.12 using BLAST. Uniprot used to find sequences, RAP2.12 accession no. Q9SSA8 and RAP2.2 accession no. Q9LUM4. RAP2.2 is the query, whilst RAP2.12 is the subject.
### Appendix 2. Mass spectrometry results for DnaK.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>ΣCover age</th>
<th>Σ# Proteins</th>
<th>Σ# Unique Peptides</th>
<th>Σ# Peptides</th>
<th>Σ# PS Ms</th>
<th>Area</th>
<th># AAs</th>
<th>MW [kDa]</th>
<th>cal c. pI</th>
</tr>
</thead>
<tbody>
<tr>
<td>seq1_K22E</td>
<td>(K22E_HUMAN) Keratin</td>
<td>80.00</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>4.961E6</td>
<td>15</td>
<td>1.9</td>
<td>9.0</td>
</tr>
<tr>
<td>DnaK2_ECOLI</td>
<td>(DnaK2) DnaK E. coli, S423-C</td>
<td>78.37</td>
<td>11</td>
<td>3</td>
<td>63</td>
<td>427</td>
<td>1.620E8</td>
<td>63</td>
<td>8</td>
<td>69.1</td>
</tr>
<tr>
<td>Seq1_10175</td>
<td>(10175_DnaK) E coli chaperone protein mutant</td>
<td>76.02</td>
<td>9</td>
<td>4</td>
<td>64</td>
<td>430</td>
<td>1.620E8</td>
<td>63</td>
<td>8</td>
<td>69.1</td>
</tr>
<tr>
<td>16045_Seq1</td>
<td>(Seq1_16045)</td>
<td>67.11</td>
<td>3</td>
<td>21</td>
<td>21</td>
<td>84</td>
<td>1.223E7</td>
<td>30</td>
<td>1</td>
<td>33.9</td>
</tr>
<tr>
<td>15048_A7ZV12</td>
<td>(15048_CH60_ECO2 4_Ecoli) 60kDa_chaperonin_E coli</td>
<td>36.68</td>
<td>1</td>
<td>16</td>
<td>16</td>
<td>58</td>
<td>4.461E6</td>
<td>54</td>
<td>1</td>
<td>33.9</td>
</tr>
<tr>
<td>ALBU_BOVIN</td>
<td>(P02769) Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4</td>
<td>28.50</td>
<td>1</td>
<td>15</td>
<td>15</td>
<td>35</td>
<td>1.570E6</td>
<td>60</td>
<td>7</td>
<td>69.2</td>
</tr>
<tr>
<td>1123PVNZ_12137</td>
<td>(12137_1123PVNZ) 1123PVNZ</td>
<td>2.53</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>6.424E5</td>
<td>27</td>
<td>7</td>
<td>31.4</td>
</tr>
<tr>
<td>APOB_HUMAN</td>
<td>(APOB_HUMAN) Apolipoprotein B-100 OS=Homo sapiens GN=APOB PE=1 SV=1</td>
<td>0.18</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1.032E6</td>
<td>45</td>
<td>63</td>
<td>515.2</td>
</tr>
</tbody>
</table>

Trypsin was the enzyme used for digestion.
## Appendix 3. Mass spectrometry results for groEL.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Coverage</th>
<th># Proteins</th>
<th># Unique Peptides</th>
<th># Peptides</th>
<th># PSMs</th>
<th>Area</th>
<th># As</th>
<th>MW [kDa]</th>
<th>cal c. pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>15048_A7ZV12</td>
<td>(15048_CH60_ECO24_Ecoli) 60kDa_chaperonin_Ecoli</td>
<td>93.80</td>
<td>1</td>
<td>57</td>
<td>57</td>
<td>665</td>
<td>5.219E8</td>
<td>54</td>
<td>57.3</td>
<td>4.9</td>
</tr>
<tr>
<td>seq1_K22E</td>
<td>(K22E_HUMAN) Keratin</td>
<td>80.00</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>3.672E6</td>
<td>15</td>
<td>1.9</td>
<td>9.0</td>
</tr>
<tr>
<td>16045_Seq1</td>
<td>(Seq1_16045)</td>
<td>65.78</td>
<td>3</td>
<td>18</td>
<td>18</td>
<td>69</td>
<td>7.296E6</td>
<td>30</td>
<td>33.9</td>
<td>6.4</td>
</tr>
<tr>
<td>Seq1_1017</td>
<td>(10175_DnaK) E.coli chaperone protein mutant</td>
<td>30.56</td>
<td>13</td>
<td>17</td>
<td>17</td>
<td>48</td>
<td>1.888E6</td>
<td>63</td>
<td>69.1</td>
<td>4.9</td>
</tr>
<tr>
<td>ATPsb_Ba</td>
<td>(ATPsb_BAC) gi</td>
<td>7.14</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>7.008E6</td>
<td>16</td>
<td>19.5</td>
<td>8.7</td>
</tr>
<tr>
<td>ATPsb_BAC</td>
<td>gi</td>
<td>AAQ10086.1</td>
<td>ATP synthase subunit b [Bacillus sp. TA2.A1]</td>
<td>4.87</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3.487E6</td>
<td>22</td>
<td>24.8</td>
</tr>
<tr>
<td>atpE_0815</td>
<td>(SeqE_08157) atpE</td>
<td>9.07</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5.676E5</td>
<td>72</td>
<td>81.4</td>
<td>8.6</td>
</tr>
<tr>
<td>13047_Seq1</td>
<td>(Seq1_13047)</td>
<td>0.97</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1.042E6</td>
<td>93</td>
<td>106.1</td>
<td>8.9</td>
</tr>
<tr>
<td>08179_Seq13</td>
<td>(Seq13_08179) B4XRZ5_9TOMB p106 OS=Cocksfoot mild mosaic virus PE=4 SV=1</td>
<td>0.86</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.042E6</td>
<td>93</td>
<td>106.1</td>
<td>8.9</td>
</tr>
</tbody>
</table>

Trypsin was the enzyme used for digestion.
Appendix 4. PCO1 single residue disorder prediction using PONDR® VLXT. D indicates residues predicted to be disordered, the lack of a D indicates residues predicted to be ordered. Scores associated with these predictions can be found in Figure 3.X. Residue number is to the left, and VLXT is the mode used on PONDR®.

Appendix 5. ΔNΔC and ΔN solubility at 37°C. Amount loaded was kept constant to allow comparison. H indicates homogenate, S represents supernatant and P indicates pellet. Where ΔNAC and ΔN are expected to run is indicated with black arrows. The lane labelled M contains molecular weight marker. Marker masses are indicated in kDa and each band contains 500 ng of protein.
Appendix 6. PCO1 sequence coverage from MS identification. Residues in bold were peptide matched. Sequence coverage is 71%, the enzyme used for digestion was trypsin.

Appendix 7. PCO2 sequence coverage from MS identification. Residues in bold were peptide matched. Sequence coverage is 41%, the enzyme used for digestion was chymotrypsin.

Appendix 8. ΔN sequence coverage from MS identification. Residues in bold were peptide matched. Sequence coverage is 40%, the enzyme used for digestion was trypsin.
Appendix 9. ΔNΔC sequence coverage from MS identification. Residues in bold were peptide matched. Sequence coverage is 53%, the enzyme used for digestion was trypsin.
Appendix 10. MS of N-terminal labelling of maize A. Band 1 is full length protein, band 2 runs at 29 kDa on a gel, band 3 at 27 kDa and band 4 at 21 kDa. Digestion was with chymotrypsin. Gel the bands are excised from can be found in Figure 4.26, Ma.

<table>
<thead>
<tr>
<th>Annotated Sequence</th>
<th>No Enzyme</th>
<th>Modifications</th>
<th># PSMs</th>
<th>Positions in Protein</th>
<th>Abundances</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-].MRVESSGGLPLELERQQAGAK.R</td>
<td>1x Dimethyl [K22]; 1x Dimethyl [N-Term]</td>
<td>6</td>
<td>[1-23]</td>
<td>4.57E+07</td>
<td></td>
</tr>
<tr>
<td>[-].MRVESSGGLPLELERQQAGAK.R</td>
<td>1x Deamidated [Q]; 1x Dimethyl [K22]; 1x Dimethyl [N-Term]</td>
<td>2</td>
<td>[1-23]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[-].MRVESSGGLPLELERQQAGAK.R</td>
<td>1x Deamidated [Q]; 1x Dimethyl [K22]</td>
<td>4</td>
<td>[1-22]</td>
<td>1.53E+06</td>
<td></td>
</tr>
<tr>
<td>[-].MRVESSGGLPLELERQQAGAK.R</td>
<td>1x Deamidated [Q]; 1x Dimethyl [N-Term]</td>
<td>13</td>
<td>[1-22]</td>
<td>2.51E+08 1.65E+05</td>
<td></td>
</tr>
<tr>
<td>[-].MRVESSGGLPLELERQQAGAK.R</td>
<td>1x Deamidated [Q]; 1x Dimethyl [K22]; 1x Dimethyl [N-Term]</td>
<td>6</td>
<td>[1-22]</td>
<td>3.55E+06</td>
<td></td>
</tr>
<tr>
<td>[-].MRVESSGGLPLELERQQAGAK.R</td>
<td>1x Acetyl [N-Term]; 1x Dimethyl [K22]</td>
<td>2</td>
<td>[1-22]</td>
<td>2.70E+06</td>
<td></td>
</tr>
<tr>
<td>[-].MRVESSGGLPLELERQQAGA.K</td>
<td>1x Dimethyl [N-Term]</td>
<td>7</td>
<td>[1-21]</td>
<td>5.33E+07 2.62E+04 2.08E+04</td>
<td></td>
</tr>
<tr>
<td>[-].MRVESSGGLPLELERQQAGA.K</td>
<td>1x Dimethyl [N-Term]</td>
<td>2</td>
<td>[1-20]</td>
<td>2.86E+06</td>
<td></td>
</tr>
<tr>
<td>[-].MRVESSGGLPLELERQA.G</td>
<td>1x Dimethyl [N-Term]</td>
<td>3</td>
<td>[1-19]</td>
<td>8.91E+06</td>
<td></td>
</tr>
<tr>
<td>[-].MRVESSGGLPLELERQA.B</td>
<td>1x Dimethyl [N-Term]</td>
<td>6</td>
<td>[1-18]</td>
<td>7.95E+06 2.24E+04 1.80E+04</td>
<td></td>
</tr>
<tr>
<td>[-].MRVESSGGLPLELERQ.Q</td>
<td>1x Dimethyl [N-Term]</td>
<td>6</td>
<td>[1-17]</td>
<td>1.72E+07 5.74E+03 2.94E+03</td>
<td></td>
</tr>
<tr>
<td>[-].MRVESSGGLPLEL.R</td>
<td>1x Dimethyl [N-Term]</td>
<td>2</td>
<td>[1-16]</td>
<td>4.76E+06</td>
<td></td>
</tr>
<tr>
<td>[-].MRVESSGGLPLEL.E</td>
<td>1x Dimethyl [N-Term]</td>
<td>2</td>
<td>[1-15]</td>
<td>6.66E+05</td>
<td></td>
</tr>
<tr>
<td>[M].MRVESSGGLPLELERQQAGAK.R</td>
<td>1x Dimethyl [K21]; 1x Dimethyl [N-Term]</td>
<td>3</td>
<td>[2-22]</td>
<td>4.84E+06</td>
<td></td>
</tr>
<tr>
<td>[R].MRVESSGGLPLELERQQAGAK.R</td>
<td>1x Dimethyl [K20]; 1x Dimethyl [N-Term]</td>
<td>2</td>
<td>[3-22]</td>
<td>4.88E+06</td>
<td></td>
</tr>
<tr>
<td>[E].SSGGLPLELERQQAGAK.R</td>
<td>1x Dimethyl [K18]; 1x Dimethyl [N-Term]</td>
<td>2</td>
<td>[5-22]</td>
<td>1.44E+06</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Modification</td>
<td>Oxidation</td>
<td>Methylation</td>
<td>Precursor</td>
<td>Charge</td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>[S].SGGLPLPELERQQAGAK.[R]</td>
<td>1×Dimethyl [K17]</td>
<td>1</td>
<td>6-22</td>
<td>2.62E+05</td>
<td></td>
</tr>
<tr>
<td>[R].AAAAAAAAPRTM.[A]</td>
<td>1×Oxidation [M11]; 1×Dimethyl [N-Term]</td>
<td>2</td>
<td>30-40</td>
<td>5.20E+04</td>
<td></td>
</tr>
<tr>
<td>[A].AAAAAPRTMAAPATAVQ.[R]</td>
<td>1×Oxidation [M10]; 1×Dimethyl [N-Term]</td>
<td>2</td>
<td>31-48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[A].AAAAAPRTMA.[A]</td>
<td>1×Oxidation [M10]; 1×Dimethyl [N-Term]</td>
<td>1</td>
<td>31-41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[A].AAAAAPRTM.[A]</td>
<td>1×Oxidation [M10]; 1×Dimethyl [N-Term]</td>
<td>2</td>
<td>31-40</td>
<td>5.80E+04</td>
<td>1.74E+03</td>
</tr>
<tr>
<td>[A].AAAAAPRTMAAPATAVQ.[R]</td>
<td>1×Oxidation [M9]; 1×Dimethyl [N-Term]</td>
<td>2</td>
<td>32-48</td>
<td>7.19E+05</td>
<td></td>
</tr>
<tr>
<td>[A].AAAAAPRTM.[A]</td>
<td>1×Oxidation [M9]; 1×Dimethyl [N-Term]</td>
<td>1</td>
<td>32-41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[A].AAAAAPRTM.[A]</td>
<td>1×Oxidation [M9]; 1×Dimethyl [N-Term]</td>
<td>2</td>
<td>32-40</td>
<td>3.33E+05</td>
<td></td>
</tr>
<tr>
<td>[A].AAAAAPRTMAAPATAVQRLF.[E]</td>
<td>1×Oxidation [M8]; 1×Dimethyl [N-Term]</td>
<td>2</td>
<td>33-51</td>
<td>1.68E+05</td>
<td></td>
</tr>
<tr>
<td>[A].AAAAAPRTMAAPATAVQ.[R]</td>
<td>1×Oxidation [M8]; 1×Dimethyl [N-Term]</td>
<td>3</td>
<td>33-48</td>
<td>3.92E+06</td>
<td></td>
</tr>
<tr>
<td>[A].AAAAAPRTMAAPA.[T]</td>
<td>1×Oxidation [M8]; 1×Dimethyl [N-Term]</td>
<td>2</td>
<td>33-44</td>
<td>3.42E+05</td>
<td></td>
</tr>
<tr>
<td>[A].AAPRTMAAPATAVQ.[R]</td>
<td>1×Oxidation [M7]; 1×Dimethyl [N-Term]</td>
<td>2</td>
<td>34-48</td>
<td>1.61E+04</td>
<td>2.10E+05</td>
</tr>
<tr>
<td>[A].AAPRTMAAPATAVQRLF.[E]</td>
<td>1×Oxidation [M6]; 1×Dimethyl [N-Term]</td>
<td>2</td>
<td>35-51</td>
<td>6.81E+04</td>
<td></td>
</tr>
<tr>
<td>[A].AAPRTMAAPATAVQ.[R]</td>
<td>1×Oxidation [M6]; 1×Dimethyl [N-Term]</td>
<td>2</td>
<td>35-48</td>
<td>1.83E+06</td>
<td></td>
</tr>
<tr>
<td>Sequence</td>
<td>Modification</td>
<td>Start</td>
<td>End</td>
<td>Intensity</td>
<td>p-value</td>
</tr>
<tr>
<td>----------</td>
<td>--------------</td>
<td>-------</td>
<td>-----</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>[F].RRvDpyGTKity.[L]</td>
<td>1×Dimethyl [K10]</td>
<td>2</td>
<td>[96-108]</td>
<td>3.20E+05</td>
<td>2.06E+04</td>
</tr>
<tr>
<td>[F].SKLFGAMHIKSY.[D]</td>
<td>Dimethyl [K]; 1×Dimethyl [N-term]</td>
<td>2</td>
<td>[142-154]</td>
<td>5.87E+05</td>
<td>8.79E+03</td>
</tr>
<tr>
<td>[AT].MIKSY.[D]</td>
<td>1×Oxidation [M1]; 1×Dimethyl [K4]</td>
<td>2</td>
<td>[149-154]</td>
<td>8.84E+05</td>
<td>2.04E+04</td>
</tr>
<tr>
<td>[Q].LQAGPRLAKVGDGILTAPETAVLYP.DGGNMHCF.[T]</td>
<td>1×Carbamidomethyl [C35]; 1×Oxidation [M33]; 1×Dimethyl [N-term]; 1×Formyl [K]</td>
<td>4</td>
<td>[168-203]</td>
<td>5.02E+05</td>
<td></td>
</tr>
<tr>
<td>[Q].LQAGPRLAKVGDGILTAPETAVLYP.DGGNMHCF.[T]</td>
<td>1×Carbamidomethyl [C35]; 1×Oxidation [M33]; 1×Dimethyl [K]; 1×Dimethyl [N-term]</td>
<td>6</td>
<td>[168-203]</td>
<td>5.02E+05</td>
<td></td>
</tr>
<tr>
<td>[L].QAGPRLAKVGDGILTAPETAVLYP.DGGNMHCF.[T]</td>
<td>1×Carbamidomethyl [C34]; 1×Oxidation [M32]; 1×Dimethyl [N-term]; 1×Formyl [K]; 1×Dimethyl [K]</td>
<td>24</td>
<td>[169-203]</td>
<td>1.02E+07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modification</td>
<td>Charge</td>
<td>M+H</td>
<td>M+Na</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------</td>
<td>--------</td>
<td>-------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>[L].QAGPRLAKVDGILTAPHTAVLYPEDGGNMHCF.[T]</td>
<td>1xCarbamidomethyl [C34]; 1xOxidation [M32]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>22</td>
<td>169-203</td>
<td>1.03E+07</td>
<td></td>
</tr>
<tr>
<td>[L].QAGPRLAKVDGILTAPHTAVLYPEDGGNMHCF.[C]</td>
<td>1xDimethyl [N-Term]; 1xFormyl [K]; 1xDimethyl [K]</td>
<td>13</td>
<td>169-201</td>
<td>6.09E+06</td>
<td></td>
</tr>
<tr>
<td>[L].QAGPRLAKVDGILTAPHTAVLYPEDGGNMHCF.[C]</td>
<td>Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>17</td>
<td>169-201</td>
<td>6.09E+06</td>
<td></td>
</tr>
<tr>
<td>[L].QAGPRLAKVDGILTAPHTAVLYPEDGGNMHCF.[C]</td>
<td>1xOxidation [M32]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>12</td>
<td>169-201</td>
<td>1.79E+07 1.90E+04</td>
<td></td>
</tr>
<tr>
<td>[Q].AGPRLAVKVDGILTAPHTAVLYPEDGGNMHCF.[T]</td>
<td>1xCarbamidomethyl [C33]; 1xDimethyl [K]; 1xDimethyl [N-Term]</td>
<td>2</td>
<td>170-203</td>
<td>1.01E+06</td>
<td></td>
</tr>
<tr>
<td>[Q].AGPRLAVKVDGILTAPHTAVLYPEDGGNMHCF.[T]</td>
<td>1xCarbamidomethyl [C33]; 2xDimethyl [K7; K9]</td>
<td>3</td>
<td>170-203</td>
<td>3.51E+06 5.48E+03</td>
<td></td>
</tr>
<tr>
<td>[Q].AGPRLAVKVDGILTAPHTAVLYPEDGGNMHCF.[T]</td>
<td>1xCarbamidomethyl [C33]; 1xDimethyl [M31]; 1xDimethyl [N-Term]; 1xFormyl [K]; 1xDimethyl [K]</td>
<td>12</td>
<td>170-203</td>
<td>2.67E+06</td>
<td></td>
</tr>
<tr>
<td>[Q].AGPRLAVKVDGILTAPHTAVLYPEDGGNMHCF.[T]</td>
<td>1xCarbamidomethyl [C33]; 1xDimethyl [M31]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>12</td>
<td>170-203</td>
<td>2.70E+06</td>
<td></td>
</tr>
<tr>
<td>[R].LAKVKVDGILTAPHTAVLYPEDGGNMHCF.[T]</td>
<td>1xCarbamidomethyl [C29]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>17</td>
<td>174-203</td>
<td>1.07E+07</td>
<td></td>
</tr>
<tr>
<td>[R].LAKVKVDGILTAPHTAVLYPEDGGNMHCF.[T]</td>
<td>1xCarbamidomethyl [C29]; 1xDeamidated [N26]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>3</td>
<td>174-203</td>
<td>3.24E+05</td>
<td></td>
</tr>
<tr>
<td>[R].LAKVKVDGILTAPHTAVLYPEDGGNMHCF.[T]</td>
<td>1xCarbamidomethyl [C29]; 1xDimethyl [M27]; 1xDimethyl [N-Term]; 1xFormyl [K]; 1xDimethyl [K]</td>
<td>7</td>
<td>174-203</td>
<td>1.96E+07</td>
<td></td>
</tr>
<tr>
<td>[R].LAKVKVDGILTAPHTAVLYPEDGGNMHCF.[T]</td>
<td>1xCarbamidomethyl [C29]; 1xDimethyl [M27]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>28</td>
<td>174-203</td>
<td>2.51E+07</td>
<td></td>
</tr>
<tr>
<td>Peptide</td>
<td>Modifications</td>
<td>Charge</td>
<td>MZ mean</td>
<td>PE mean</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>---------------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>[R].LAKVKDILTAPHTAVLPEDDGNNM HC.[F]</td>
<td>1xCarbamidomethyl [C29]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>2</td>
<td>[174-202]</td>
<td>4.97E+05</td>
<td></td>
</tr>
<tr>
<td>[R].LAKVKDILTAPHTAVLPEDDGNNM HC.[F]</td>
<td>1xCarbamidomethyl [C29]; 1xOxidation [M27]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>3</td>
<td>[174-202]</td>
<td>5.61E+05</td>
<td></td>
</tr>
<tr>
<td>[R].LAKVKDILTAPHTAVLPEDDGNNM H.[C]</td>
<td>Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>18</td>
<td>[174-201]</td>
<td>2.38E+07</td>
<td>3.44E+03</td>
</tr>
<tr>
<td>[R].LAKVKDILTAPHTAVLPEDDGNNM H.[C]</td>
<td>1xOxidation [M27]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>19</td>
<td>[174-201]</td>
<td>1.76E+07</td>
<td>4.96E+04</td>
</tr>
<tr>
<td>[L].LAKVKDILTAPHTAVLPEDDGNNM CF.[T]</td>
<td>1xCarbamidomethyl [C28]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>6</td>
<td>[175-203]</td>
<td>1.08E+06</td>
<td></td>
</tr>
<tr>
<td>[L].LAKVKDILTAPHTAVLPEDDGNNM H.[C]</td>
<td>1xOxidation [M26]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>6</td>
<td>[175-201]</td>
<td>1.66E+05</td>
<td></td>
</tr>
<tr>
<td>[A].LKVKGILTAPHTAVLPEDDGNNM HC.F.[T]</td>
<td>1xCarbamidomethyl [C27]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>4</td>
<td>[176-203]</td>
<td>1.51E+06</td>
<td></td>
</tr>
<tr>
<td>[A].LKVKGILTAPHTAVLPEDDGNNM HC.F.[T]</td>
<td>1xCarbamidomethyl [C27]; 1xOxidation [M25]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>8</td>
<td>[176-203]</td>
<td>3.10E+06</td>
<td>8.24E+03</td>
</tr>
<tr>
<td>[A].LKVKGILTAPHTAVLPEDDGNNM HC.F.[C]</td>
<td>Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>4</td>
<td>[176-201]</td>
<td>5.37E+06</td>
<td></td>
</tr>
<tr>
<td>[A].LKVKGILTAPHTAVLPEDDGNNM HC.F.[C]</td>
<td>1xOxidation [M25]; Dimethyl [K]; 1xDimethyl [N-Term]</td>
<td>4</td>
<td>[176-201]</td>
<td>1.07E+06</td>
<td></td>
</tr>
<tr>
<td>[K].VKVDGILTAPHETAVLYPEDGGNMHC.F.</td>
<td>1×Carbamidomethyl [C26]; 1×Dimethyl [N-Term]; 1×Dimethyl [K2]</td>
<td>6</td>
<td>[177-203]</td>
<td>1.64E+06</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-------------------------------------------------</td>
<td>---</td>
<td>-----------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>[T]</td>
<td>1×Carbamidomethyl [C26]; 1×Oxidation [M24]; 1×Dimethyl [K2]; 1×Dimethyl [N-Term]</td>
<td>8</td>
<td>[177-203]</td>
<td>2.08E+06</td>
<td></td>
</tr>
<tr>
<td>[K].VKVDGILTAPHETAVLYPEDGGNMHC.F.</td>
<td>1×Dimethyl [K2]; 1×Dimethyl [N-Term]</td>
<td>6</td>
<td>[177-201]</td>
<td>1.94E+06</td>
<td></td>
</tr>
<tr>
<td>[C]</td>
<td>1×Carbamidomethyl [C22]; 1×Dimethyl [N-Term]</td>
<td>1</td>
<td>[181-203]</td>
<td>1.12E+06</td>
<td></td>
</tr>
<tr>
<td>[D].GILTAPHTAVLYPEDGGNHC.F.[T]</td>
<td>1×Carbamidomethyl [C22]; 1×Oxidation [M20]; 1×Dimethyl [N-Term]</td>
<td>2</td>
<td>[181-203]</td>
<td>1.39E+06</td>
<td></td>
</tr>
<tr>
<td>[Y].MGPKILDHWSHPQFEK.[-]</td>
<td>1×Oxidation [M1]; 2×Dimethyl [K4; K16]</td>
<td>6</td>
<td>[271-286]</td>
<td>1.03E+07</td>
<td></td>
</tr>
</tbody>
</table>

192
Appendix 11. MS of Maize B and contaminants. Table includes first 12 hits only. Gel in Figure 4.26 Mb, bands labelled 1-5 from the top molecular weight band. Bands were digested with chymotrypsin.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Protein Description</th>
<th>Sequest Score</th>
<th>Average peptide intensity: Band 1</th>
<th>Average peptide intensity: Band 2</th>
<th>Average peptide intensity: Band 3</th>
<th>Average peptide intensity: Band 4</th>
<th>Average peptide intensity: Band 5</th>
<th>Seque nce Coverage [%]</th>
<th># Unique peptides matching identified protein</th>
<th># all peptides matching identified protein</th>
<th># all Peptides</th>
<th># PS Ms</th>
<th># A As</th>
<th>MW [kDa]</th>
<th>cal c. pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>I0001</td>
<td>Maize CDO</td>
<td>504.39</td>
<td>2.48E+07</td>
<td>1.77E+07</td>
<td>5.36E+06</td>
<td>3.25E+06</td>
<td>1.13E+08</td>
<td>70</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>7</td>
<td>187</td>
<td>24</td>
<td>27</td>
</tr>
<tr>
<td>A7ZU2</td>
<td>Elongation factor Tu 2 OS=Escherichia coli O139 H28 (strain E24377A / ETDC) GN=nufl2 PE=3 SV=1</td>
<td>374.57</td>
<td>2.04E+05</td>
<td>2.52E+05</td>
<td>8.91E+05</td>
<td>4.55E+07</td>
<td>2.41E+05</td>
<td>67</td>
<td>4</td>
<td>31</td>
<td>150</td>
<td>39</td>
<td>4</td>
<td>43.3</td>
<td>4</td>
</tr>
<tr>
<td>B1X6D7</td>
<td>60 kDa chaperonin OS=Escherichia coli (strain K12 / DH10B) GN=nagD PE=3 SV=1</td>
<td>371.2</td>
<td>9.67E+04</td>
<td>9.71E+06</td>
<td>8.73E+05</td>
<td>8.58E+04</td>
<td>7.44E+03</td>
<td>53</td>
<td>24</td>
<td>27</td>
<td>141</td>
<td>54</td>
<td>8</td>
<td>57.3</td>
<td>4</td>
</tr>
<tr>
<td>Q0S8Z8X</td>
<td>Elongation factor Tu 1 OS=Shigella flexneri serotype 5b (strain 5401) GN=nufl1 PE=3 SV=1</td>
<td>369</td>
<td>2.53E+04</td>
<td>66</td>
<td>3</td>
<td>30</td>
<td>139</td>
<td>39</td>
<td>4</td>
<td>43.3</td>
<td>40</td>
<td>5</td>
<td>4</td>
<td>5.5</td>
<td>4</td>
</tr>
<tr>
<td>Q0A3G33</td>
<td>ATP synthase subunit beta OS=Shigella boydii serotype 18 (strain CDC 3083-94 / BS512) GN=suqD PE=3 SV=1</td>
<td>180.91</td>
<td>7.31E+04</td>
<td>2.31E+04</td>
<td>4.09E+04</td>
<td>3.59E+04</td>
<td>1.85E+04</td>
<td>52</td>
<td>13</td>
<td>20</td>
<td>63</td>
<td>46</td>
<td>0</td>
<td>50.3</td>
<td>5</td>
</tr>
<tr>
<td>P0A4G67</td>
<td>Transcription termination factor Rho OS=Shigella flexneri GN=suqE PE=3 SV=1</td>
<td>158.44</td>
<td>1.42E+04</td>
<td>3.05E+04</td>
<td>3.77E+06</td>
<td>2.05E+04</td>
<td>7.67E+03</td>
<td>43</td>
<td>22</td>
<td>22</td>
<td>59</td>
<td>41</td>
<td>9</td>
<td>47.2</td>
<td>5</td>
</tr>
<tr>
<td>P0A1G67</td>
<td>30S ribosomal protein S1 OS=Escherichia coli (strain K12) GN=suqA PE=1 SV=1</td>
<td>144.79</td>
<td>2.98E+04</td>
<td>1.52E+04</td>
<td>3.96E+04</td>
<td>7.67E+03</td>
<td>3</td>
<td>15</td>
<td>15</td>
<td>47</td>
<td>55</td>
<td>7</td>
<td>61.4</td>
<td>4.9</td>
<td>8</td>
</tr>
<tr>
<td>C3K3B8</td>
<td>Elongation factor Tu OS=Pseudomonas fluorescens (strain SBW25) GN=nufl2 PE=3 SV=1</td>
<td>134.86</td>
<td>7.94E+04</td>
<td>1.07E+05</td>
<td>4.44E+05</td>
<td>3.12E+05</td>
<td>26</td>
<td>6</td>
<td>11</td>
<td>62</td>
<td>39</td>
<td>7</td>
<td>43.5</td>
<td>3.4</td>
<td>5</td>
</tr>
<tr>
<td>B0K4NW</td>
<td>Succinyl-CoA ligase [ADP-forming] subunit beta OS=Pseudomonas putida (strain GB-1) GN=nucC PE=3 SV=1</td>
<td>132.05</td>
<td>1.45E+05</td>
<td>1.37E+05</td>
<td>2.90E+05</td>
<td>3.31E+05</td>
<td>22</td>
<td>7</td>
<td>7</td>
<td>44</td>
<td>38</td>
<td>8</td>
<td>41.1</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>A7ZV7V2</td>
<td>Chaperone protein DnaK OS=Escherichia coli O9:H4 (strain HS) GN=nuflD PE=3 SV=1</td>
<td>124.27</td>
<td>2.30E+06</td>
<td>4.77E+04</td>
<td>1.25E+04</td>
<td>4.93E+03</td>
<td>27</td>
<td>18</td>
<td>18</td>
<td>43</td>
<td>63</td>
<td>8</td>
<td>69.1</td>
<td>4.9</td>
<td>7</td>
</tr>
<tr>
<td>B117U3</td>
<td>50S ribosomal protein L1 OS=Escherichia coli (strain ATEC 8739 / DSM 1576 / Crooks) GN=nuflA PE=3 SV=1</td>
<td>118.33</td>
<td>3.03E+04</td>
<td>2.94E+04</td>
<td>2.79E+04</td>
<td>2.71E+04</td>
<td>9.75E+06</td>
<td>47</td>
<td>10</td>
<td>10</td>
<td>39</td>
<td>23</td>
<td>4</td>
<td>24.7</td>
<td>4</td>
</tr>
<tr>
<td>B2T4Y3</td>
<td>D-tagatose-1,6-bisphosphate aldolase subunit GatZ OS=Shigella boydii serotype</td>
<td>106.78</td>
<td>3.07E+04</td>
<td>5.12E+04</td>
<td>4.44E+06</td>
<td>7.51E+04</td>
<td>1.09E+04</td>
<td>28</td>
<td>9</td>
<td>9</td>
<td>43</td>
<td>42</td>
<td>0</td>
<td>47.8</td>
<td>5.8</td>
</tr>
</tbody>
</table>
Appendix 12. Improving solubility of maize B using lysis conditions. Lysis conditions are labelled with a number above. One-regular lysis buffer (2.4.1), 2-regular lysis buffer with 400 mM sodium chloride, 3-regular lysis buffer with 125 mM L-arginine and four-100 mM Tris pH 7.1 with 300 mM sodium chloride and 0.7% Tween20®. Samples with sodium citrate buffer did not run on the gel, perhaps due to the pH of these buffers. Number of *E. coli* cells has not been controlled for in normalising loading. H indicates homogenate, S represents supernatant and P indicates pellet. Where maize B is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Appendix 5.

Appendix 13. Maize B expression in Origami®. Equal number of cells as judged by OD$_{600}$, were loaded into each lane. The negative sign indicates sample before induction and the numbers following this represents number of hours after induction. Where Maize B is predicted to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Appendix 5.
Appendix 14. ΔNR expressed in BL21(DE3) pLysS solubility. Loading is kept constant for comparison. H indicates homogenate, S represents supernatant and P indicates pellet. Where ΔNR is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Appendix 5.

Appendix 15. ΔNR expressed in BL21(DE3) pLysS solubility at 37°C. Loading kept constant to allow for comparison. H indicates homogenate, S represents supernatant and P indicates pellet. Where rice is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Appendix 5.
Appendix 16. ΔNR expressed in BL21(DE3) pLysS Strep-tag® FPLC purification. The blue line represents absorbance at 280 nm in mAU (left axis) and the red line represents conductivity in mS/cm (right axis). Fractions numbers relate to Appendix 17. After sample application, wash buffer was run from 15-60 mL, and elution buffer was run from 60-100 mL. A water wash from 100-120 mL was followed by a sodium hydroxide wash from 120-140 mL. The final wash was with water.

Appendix 17. ΔNR expressed in BL21(DE3) pLysS Strep-tag® FPLC purification gel. Lanes labelled with a letter and a number are fraction numbers. Where ΔNR is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Appendix 5.
Appendix 18. ΔNR expressed in Origami™ (DE3) solubility. Loading was kept constant to allow for comparison. H indicates homogenate, S represents supernatant and P indicates pellet. Where ΔNR is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Appendix 5.

Appendix 19. ΔNR expressed in Origami® (DE3) solubility at 37°C. Loading kept constant to allow for comparison. H indicates homogenate, S represents supernatant and P indicates pellet. Where rice would run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Appendix 5.
Appendix 20. ΔNR expressed in Origami™ (DE3) Strep-tag® FPLC purification. The blue line represents absorbance at 280 nm in mAU (left axis) and the red line represents conductivity in mS/cm (right axis). Fractions relate to the samples analysed in Appendix 21. After sample application, wash buffer was run from 15-65 mL, and elution buffer was run from 65-95 mL. A water wash from 95-115 mL was followed by a sodium hydroxide wash from 115-145 mL. The final wash was with water.

Appendix 21. ΔNR expressed in Origami™ (DE3) Strep-tag® FPLC purification gel. Lanes labelled with a letter and a number are fraction numbers. Where ΔNR is expected to run is indicated with a black arrow. The lane labelled M contains molecular weight marker. Marker as in Appendix 5. Fractions relate to the chromatogram in Appendix 20.
Appendix 22. *E. coli* Chloramphenicol acetyltransferase sequence coverage from MS identification. Residues in bold were peptide matched. Sequence coverage is 35%, the enzyme used for digestion was chymotrypsin. Molecular weight is 26000 Da.

Appendix 23. Sequencing of the rice construct. The sequencing results are labelled sequencing, and the correct sequence is labelled Rice_PCO2. Sequence match is 100%. Clustal Omega was used to produce the alignment. A star is indicative of the same amino acid being present at the same position.
Appendix 24. Mammalian CDO surface topology with charge distribution. Blue represents a positive charge, and red represents a negative charge. PDB accession number 2B5H was used and Chimera was used to produce the image.
**Appendix 25. MS of higher order structures on non-reducing gel.** F1 is the top band (~180 kDa), F2 the middle (~140 kDa) and F3 the bottom (~68 kDa). Digestion was with trypsin. Gel the bands are excised from can be found in **Figure 5.13**.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Protein</th>
<th># all peptides matching identified protein</th>
<th># Unique peptides matching identified protein</th>
<th># Peptide Spectrum Matches for identified protein</th>
<th>Normalized abundance ratio (F1) / (F3)</th>
<th>Normalized abundance ratio (F2) / (F3)</th>
<th>Normalized abundance ratio (F1) / (F2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q9LXG9</td>
<td>PCO1</td>
<td>21</td>
<td>21</td>
<td>188</td>
<td>46.713</td>
<td>41.09</td>
<td>0.986</td>
</tr>
<tr>
<td>B1XDP7</td>
<td>GroEL</td>
<td>22</td>
<td>22</td>
<td>70</td>
<td>1.102</td>
<td>0.685</td>
<td>1.364</td>
</tr>
<tr>
<td>A7ZVV7</td>
<td>DnaK</td>
<td>62</td>
<td>62</td>
<td>295</td>
<td>0.023</td>
<td>0.011</td>
<td>1.901</td>
</tr>
</tbody>
</table>
Appendix 26. Molecular weight estimation of PCO1 refolding gel filtration. Line equation is \( y = -0.0013x + 0.0024 \). Quality of fit, \( R^2 = 0.9635 \).

Calculations
\( R_f \) of indicated band (grey arrow) = 0.18
\( \frac{0.18 - 0.0024}{-0.0013} = x \)
\( x = 138 \text{ Da} \)
\( 138 \text{ Da} / 36 \text{ Da (Mr of PCO1)} = 3.8 \text{ monomers of CDO} \)

Appendix 27. PCO1 sequence motifs.

<table>
<thead>
<tr>
<th>PS00005 PKC_PHOSPHO_SITE (pattern) Protein kinase C phosphorylation site [occurs frequently]</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 - 18</td>
</tr>
<tr>
<td>27 - 29</td>
</tr>
<tr>
<td>42 - 44</td>
</tr>
<tr>
<td>64 - 66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PS00006 CK2_PHOSPHO_SITE (pattern) Casein kinase II phosphorylation site [occurs frequently]</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 - 53</td>
</tr>
<tr>
<td>64 - 67</td>
</tr>
<tr>
<td>111 - 114</td>
</tr>
<tr>
<td>247 - 250</td>
</tr>
</tbody>
</table>
Headings include a ScanProsite accession number, followed by the type of modification then the specific modification. Frequency is also noted. Column below contains the amino acid number/s of the modification/s.

Appendix 28. Maize B spherulites from Crystal Screen HT. Image taken under UV light. Spherulites formed in 0.2 M calcium chloride, 0.1 M HEPES pH 7.5 and 28% v/v PEG 400. Spherulites were formed with 6, 4 and 2 mg/mL maize B.
Appendix 29. Maize B spherulites from Crystal Screen HT. Left images taken under cross polarised light, whilst right is under UV light. Spherulites formed in 0.2 M calcium acetate, 0.1 M sodium cacodylate and 18% w/v PEG 8000. Spherulites were formed with 6, 4 and 2 mg/mL maize B.

Appendix 30. Optimised Maize B spherulites from PEGRx HT with calcium. Image taken under UV light. Spherulites formed in 0.1 M calcium chloride, 0.1 M Tris pH 8 and 30% v/v PEG 400. Maize B concentration was 3 mg/mL. White arrow indicates a spherulite.
Appendix 31. Large drop Maize B spherulites from PEGRx HT with calcium. Left image was taken under UV light and right was taken under cross polarised light. Spherulites formed in 0.1 M calcium chloride, 0.1 M citric acid pH 3.5 and 34% v/v PEG 200. The maize B concentration was 3 mg/mL. White arrow indicates a spherulite.
Appendix 32. Raw kinetic data for refolding experiment. Slope indicative of observed velocity in mM/min, and shows substrate depletion. Peptide used was ~30 mM pentapeptide (amino acids in order were CGGAI). Enzyme used was PCO1 at a final concentration of 5 μM. Data produced using Ellman’s reagent in a 96-well plate format. Data was produced in duplicate and error bars represent standard error of the mean.
Appendix 33. Raw data for PCO1 dipeptide enzyme titration. Graphs show substrate depletion, and slope is indicative of velocity in mM/min. Peptide used was 5 mM dipeptide (amino acid were CG). PCO1 concentration is listed above each graph in μM. Data produced using Ellman’s reagent in a 96-well plate format.
Appendix 34. Raw data for varying substrate concentrations with PCO1. Enzyme concentration was 6 μM. Graphs show substrate depletion, and slope is indicative of velocity in mM/min. Peptide used was dipeptide (amino acids were CG). Substrate concentration is listed above each graph in mM. Data produced using Ellman’s reagent in a 96-well plate format.